IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 28 June 2025, accepted 3 August 2025, date of publication 6 August 2025, date of current version 12 August 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3596300

RESEARCH ARTICLE

LARNet-SAP-YOLOv11: A Joint Model for Image Restoration and Corrosion Defect Detection of Transmission Line Fittings Under Multiple Adverse Weather Conditions

YUXIANG GONG[1], LIN GAO[1], (Member, IEEE), HAO ZHANG[1], HUAGUO LIU[2], YONGDAN ZHU[1], AND YU YANG[1]

[1]College of Intelligent Systems Science and Engineering, Hubei Minzu University, Enshi 445000, China
[2]College of Mechanical and Electrical Engineering, Guilin University of Aerospace Technology, Guilin 541004, China

Corresponding author: Lin Gao (2003001@hbmzu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 12464004 and Grant 61562025, in part by the Intramural Research Project of Hubei Minzu University under Grant XN2317, and in part by Guangxi University Young and Middle-Aged Teachers' Basic Scientific Research Capacity Improvement Project under Grant 2024KY0816.

ABSTRACT Adverse weather conditions such as haze, rain, and snow degrade images captured by uncrewed aerial vehicles (UAVs) during transmission line inspections and severely affect the detection of corrosion defects in transmission line fittings. To address this challenge, we propose LARNet-SAP-YOLOv11, a unified end-to-end model that integrates lightweight image restoration and defect detection. The proposed model comprises LARNet, a lightweight all-in-one image restoration network, and SAP-YOLOv11, an enhanced object detector based on YOLOv11. LARNet is built upon the DehazeFormer architecture and introduces a Triplet Attention Block (TAB) to improve adaptability to various weather degradations. SAP-YOLOv11 enhances the baseline YOLOv11n by incorporating a Shallow Robust Feature Downsampling (SRFD) module, an Adaptive Fine-Grained Channel Attention (AFGCAttention) mechanism, and a Pixel-level Cross-Attention Feature Fusion (PCAFFusion) module, significantly improving corrosion area perception. Experimental results show that LARNet achieves an average PSNR of 30.43 dB and SSIM of 0.951 across different conditions. For defect detection, SAP-YOLOv11 improves the mAP@50 by 2.1% compared to the original YOLOv11n. When jointly applied, LARNet-SAP-YOLOv11 achieves an mAP@50 of 88.6%, outperforming the baseline YOLOv11n by 12.1% in challenging weather conditions. This unified model offers an efficient and reliable solution for UAV-based intelligent inspection of transmission lines under diverse environmental conditions.

INDEX TERMS Adverse weather, corrosion defect detection, LARNet, transmission line fittings, UAV inspection, YOLOv11.

I. INTRODUCTION

Overhead transmission lines are critical infrastructure in power systems, responsible for delivering electricity from power stations to load centers, thereby meeting both industrial and residential energy demands. With the continuous expansion of the power grid, transmission lines now span vast geographic areas and are increasingly exposed to harsh environmental conditions such as high humidity, strong winds, and frequent rain or snow. Common transmission line fittings—such as suspension clamps, anti-vibration hammers, U-shaped hanging loops, and triangular joint boards—are particularly susceptible to corrosion due to long-term exposure to mechanical stress and environmental erosion. Corrosion significantly weakens the mechanical strength of these components, increasing the risk of failures such as conductor

The associate editor coordinating the review of this manuscript and approving it for publication was Hao Wang.

breakage and short circuits, which may severely compromise the safe and stable operation of the power system. Therefore, regular inspection of transmission lines, especially corrosion monitoring of critical fittings, is essential for infrastructure safety [1].

With advancements in smart grids and artificial intelligence, uncrewed aerial vehicles (UAV) and deep learning technologies have been increasingly applied in power line inspection tasks. UAV equipped with high-definition cameras can efficiently capture visual data of transmission line fittings, enhancing the automation and accuracy of inspections [2]. However, in regions prone to adverse weather like haze, rain, and snow, UAV-captured images suffer significant degradation—haze reduces contrast and blurs details, raindrops cause occlusions and refractions, and snowflakes obscure critical textures—severely impairing defect detection performance. Moreover, corroded transmission line fittings face higher risks of damage and failure in adverse weather, and relying solely on clear-weather inspections is insufficient for regions with prolonged exposure to haze, rain, or snow. Conducting inspections under adverse conditions is essential to ensure timely fault detection and reduce failure risks, necessitating robust image restoration to enable accurate defect identification.

To address the challenges posed by adverse weather conditions, researchers have proposed various methods that integrate image restoration with object detection. For instance, Qiu et al. [3] proposed the IDOD-YOLOv7 model, which combines an image dehazing module (IDOD) with the YOLOv7 detector to significantly enhance detection accuracy under low-light conditions. Wang et al. [4] focused on image degradation in battlefield scenarios and developed an image restoration method based on a physical imaging model, which was integrated with the YOLOv5 detector to enable clearer imaging and improved target detection under such conditions. In the field of transmission line defect detection under diverse weather conditions, Jia et al. [5] introduced an adaptive hybrid network based on the YOLOv8 model to improve the detection of foreign objects on transmission lines in adverse weather. However, the method lacked an image restoration module, limiting its ability to enhance input image quality for the detector. Song et al. [6] employed an improved dark channel prior algorithm to preprocess hazy images before inputting them into the YOLOv5 model, improving detection precision for fittings under hazy conditions; nevertheless, this two-stage approach involved cumbersome preprocessing. These studies offer valuable insights into the integration of image restoration and object detection. However, most of them focus on single-weather scenarios and lack adaptability to different weather conditions. Deng et al. [7] integrated multiple image restoration models with an enhanced YOLOv7 detector to improve the detection of insulator defects under hazy and rainy conditions. Their approach, however, required weather classification during preprocessing, resulting in a more complex pipeline and reduced scalability.

In summary, current approaches integrating image restoration with object detection still face significant limitations. For instance, existing methods primarily focus on combining single-weather restoration models with object detection models [3], [4], [7], lacking research on integrating multi-weather restoration models with object detection, which results in limited generalization across diverse adverse weather conditions. Additionally, current inspection tasks for transmission line fittings typically rely on acquiring clear images under sunny conditions [5], [6]. In regions with frequent haze, rain, or snow, this constraint severely impacts image quality and fault identification rates, leading to frequent missed and false detections.

To address these limitations, we propose LARNet-SAP-YOLOv11, an end-to-end joint model for image restoration and corrosion defect detection of transmission line fittings under multiple adverse weather conditions. The restoration module effectively removes haze, rain, and snow from images, providing clear inputs to an enhanced YOLOv11 model, which significantly improves detection accuracy across diverse weather conditions, thereby enhancing inspection reliability.

Main contributions of the paper are as follows:

1. We develop a lightweight all-in-one restoration network (LARNet) based on an improved DehazeFormer architecture, which demonstrates strong image enhancement capabilities across hazy, rainy, and snowy conditions. Additionally, we design a defect detection model—SAP-YOLOv11—based on the YOLOv11n backbone, incorporating Shallow Robust Feature Downsampling (SRFD) and Deep Robust Feature Downsampling (DRFD), adaptive fine-grained channel attention (AFGCAttention), and pixel-level cross-attention feature fusion (PCAFFusion) to boost the detection performance for various corroded transmission line fittings.

2. We construct an integrated LARNet-SAP-YOLOv11 model that jointly optimizes image restoration and defect detection in an end-to-end manner, eliminating the need for complex preprocessing. This pipeline ensures high-precision identification of transmission line fittings and efficient inference under adverse weather

3. We build a multi-weather corrosion defect dataset for transmission line fittings, covering three representative types of degradation: haze, rain, and snow. This dataset provides a solid foundation for model training and evaluation of generalization ability.

The remainder of this paper is organized as follows: Section II reviews related work on image restoration and object detection; Section III describes the proposed model and its key modules; Section IV presents the experimental setup and result analysis; and Section V concludes the paper and discusses future work.

II. RELATED WORK

A. IMAGE RESTORATION TASKS

1) RESTORATION METHODS IN SINGLE COMPLEX WEATHER CONDITIONS

Early studies on image restoration predominantly focused on modeling under specific weather conditions. Classic dehazing approaches typically rely on atmospheric scattering models and prior knowledge to estimate transmittance and scene radiance, enhancing visibility in hazy images [8], [9]. However, such methods are limited in handling haze with varying density and complexity. Rain removal techniques often involve image decomposition, frequency-domain filtering, and traditional machine learning methods [10], [11]. These approaches perform reasonably well in light rain but struggle with complex rain streaks and raindrops. Similarly, snow removal methods based on image-guided filtering and texture priors…)) [12], [13] can recover snow-covered regions, yet they show limited robustness when facing diverse snowflake shapes and distributions.

With the advancement of deep learning, substantial progress has been made in image restoration. For instance, the Dehazing Enhancement and Attention Network (DEA-Net) [14] and Dehazing Transformer (DehazeFormer) [15] incorporate atmospheric scattering priors and attention mechanisms into end-to-end neural networks, significantly improving both restoration quality and semantic understanding under hazy conditions. In the rain removal domain, models such as Dual Attention Mixed Network (DAMNet) [16], Deep Wavelet Transform Network (DWTN) [17], and Using Enhanced Transformer (UC-former) [18] combine Transformer structures with convolutional neural networks (CNNs), leveraging multi-scale feature fusion to adaptively suppress rain artifacts. For snow removal, models including DesnowNet [19], the Global Windowing Transformer Network for Snow Removal (SGNet) [20], and Wavelet Transform Frequency Snow Removal (WaveFrSnow) [21] introduce multi-stage architectures and global context modeling to effectively handle snow patterns while maintaining computational efficiency. Despite their strong performance in isolated scenarios, these methods lack the ability to generalize to mixed-weather environments and often involve large model sizes, which limit their deployment in real-world applications.

2) IMAGE RESTORATION UNDER MULTIPLE ADVERSE WEATHER CONDITIONS

To address the real-world challenges of UAV-based inspection under diverse adverse weather, researchers have proposed restoration methods tailored for multiple complex conditions. These methods can be broadly categorized into multi-stage strategies and unified restoration models.

In multi-stage strategies, weather-specific features are separately modeled and learned to enhance restoration across diverse degradation types. For example, Zhu et al. [22] proposed a unified model adopting a two-stage training strategy. The first stage captures shared weather features, while the second stage fine-tunes parameters for specific weather types, improving model adaptability and restoration accuracy. Chen et al. [23] introduced a multi-teacher multi-student model incorporating knowledge distillation and review mechanisms. Each teacher network is trained on a specific weather condition, and the student network learns a generalized restoration policy by integrating multi-source knowledge. Zhu et al. [24] proposed a multi-weather Transformer architecture that uses a hyper-network to extract weather features in the first stage, followed by condition-specific image restoration in the second stage. This design enables controllable and adaptive restoration across multiple weather types. Similarly, Patil et al. [25] first propose an instance-level domain translation with a multi-attentive feature learning approach. This method trains separate domain translation networks for different weather conditions to obtain various weather-degraded variants of the same scenario. Cheng et al. [26] propose a novel multi-weather distribution diffusion blind restoration model (WeaFU). Through a multi-stage strategy, the model trains a Latent Semantic Mapper, a Conditional Distribution-Aware Transformer, and a Diffusion Distribution Generator, followed by joint fine-tuning, creatively utilizing a diffusion model to perceive and extract different weather distributions.

Although multi-stage methods offer high restoration accuracy, they tend to involve complex network designs with substantial model sizes, resulting in low training and inference efficiency. Moreover, some of these approaches require explicit classification of input weather types, increasing deployment complexity.

In contrast, unified restoration approaches have attracted significant attention due to their simplicity and scalability. Models such as TransWeather [27], PromptIR [28], TANet [29], UVRNet [30] and MW-ConvNet [31], adopt single-architecture designs that learn both general and weather-specific degradation features, enabling effective restoration under various weather conditions within a single model. Furthermore, Xu et al. [32] propose a semi-supervised learning framework that leverages vision-language models to enhance restoration performance under diverse adverse weather conditions, utilizing vision-language models to assess pseudo-labels and incorporating weather prompt learning to improve image clarity across different weather scenarios.

These unified models exhibit excellent scalability and potential for real-world deployment. However, they still face challenges related to model size and parameter efficiency, necessitating further optimization to enhance generalization performance while maintaining adaptability to diverse application scenarios.

B. TRANSMISSION LINE DEFECT DETECTION TASKS

With the rapid advancement of deep learning-based object detection techniques, the field has evolved from two-stage

detection models [33], [34], [35], [36], to single-stage detectors such as SSD [37] and the YOLO series [38], [39], [40], [41], [42], [43], [44], which offer faster inference and superior real-time performance. These single-stage models have been widely adopted in practical applications, including UAV-based inspection tasks. More recently, the introduction of Detection Transformers (DETR) [45] has incorporated Transformer architectures into object detection, enabling end-to-end learning and expanding the boundaries of the field.

In the context of transmission line defect detection, Guo et al. [46] applied an enhanced Faster R-CNN model with feature amplification mechanisms to detect corrosion defects in transmission line fittings. Experimental results demonstrated the model's effectiveness in terms of precision and recall. However, its large computational complexity and model size limited real-time performance. Tan et al. [47] proposed a fast and efficient method for insulator classification and defect recognition based on YOLOv8n, which also performs well under complex scenarios. Yu et al. [48] integrated a novel lightweight backbone with YOLOv5, combining denoising and object detection for accurate identification of insulator defects in aerial images captured under challenging conditions. Wang et al. [49] enhanced the YOLOv7 model by incorporating Transformer architecture, triplet attention mechanisms, and a smooth Intersection over Union (SIoU) loss function, achieving real-time and high-accuracy detection of transmission line defects.

Despite these advancements, two major limitations remain: Due to the lack of dedicated visible-spectrum remote sensing datasets for corroded transmission hardware, most existing studies have focused on typical defects such as insulators, missing components, and foreign object intrusions, with limited research targeting corrosion-specific detection models.

Most detection models are designed for clear-weather conditions with high visibility. As a result, their performance degrades significantly under adverse weather conditions such as haze, rain, or snow, which severely restricts the applicability of UAV-based transmission line inspection systems.

To address these challenges, a dedicated dataset of aerial remote sensing images for corroded transmission line fittings was constructed. Furthermore, a lightweight all-in-one image restoration model was integrated with a defect detection model, enabling high-accuracy corrosion detection under multiple adverse weather conditions. This approach significantly enhances the robustness and practical utility of UAV inspection systems in real-world environments.

III. OVERALL MODEL ARCHITECTURE

We propose an end-to-end joint inference model named LARNet-SAP-YOLOv11, designed to achieve efficient corrosion defect detection of transmission line fittings under various adverse weather conditions. This model integrates a lightweight all-in-one restoration network (LARNet) with an enhanced object detection model (SAP-YOLOv11), enabling collaborative processing for image enhancement and defect identification. Specifically, the LARNet model is first employed to restore degraded images caused by haze, rain, and snow, generating images with clearer structures and sharper edges. These enhanced images are then passed into the SAP-YOLOv11 model to precisely locate and detect corroded fittings.

As illustrated in Fig. 1, the LARNet-SAP-YOLOv11 model consists of two primary stages. In the first stage, image restoration is conducted using a lightweight end-to-end network to reduce atmospheric interference and improve texture and edge details. In the second stage, high-quality features are extracted from the restored images by the detection module, allowing accurate and robust defect detection of corroded components. This joint optimization approach effectively mitigates the negative impacts of weather-induced degradation on inspection images. Moreover, the end-to-end architecture significantly improves inference efficiency and enhances adaptability in multi-scenario deployments.

A. LARNet DESIGN

To address efficient image restoration under multiple adverse weather conditions, we design a lightweight all-in-one restoration model named LARNet, based on the DehazeFormer architecture [15]. LARNet features a compact structure with low memory usage and fast inference speed, while also maintaining strong generalization capability for both rain and snow scenes. Additionally, its high compatibility with the YOLOv11 model facilitates seamless integration for end-to-end joint inference under harsh conditions.

The overall architecture of LARNet is illustrated in Fig. 2. As shown, the model is built upon the DehazeFormer-b baseline, enhanced by embedding a Triplet Attention Block (TAB) to extract semantic and texture features under various adverse weather conditions. In the encoder stage, LARNet first employs a convolutional layer to extract key features, followed by multiple DehazeFormer modules and downsampling operations to compress features, enhance channel information, and prepare for subsequent feature fusion. In the decoder stage, the first TAB module is incorporated at the beginning of the decoder, where the input features have the highest embedding dimension and the smallest spatial resolution, making it ideal for capturing global context and optimizing features. The second TAB module is added after a Selective Kernel Fusion (SKFusion) feature fusion step and a DehazeFormer module, at which point spatial details begin to recover. This placement enhances mid-level features, improving the model's focus on local weather artifacts, such as rain streaks or snow patches. Finally, at the end of the decoder, a soft reconstruction module facilitates image reconstruction and adjustment, producing clear images under diverse weather conditions.

The TAB module includes three sub-attention mechanisms: Local Pixel-wise Attention (LPA), Global Strip-wise Attention (GSA) and Global Distribution Attention (GDA). LPA and GSA are responsible for mitigating occlusions caused by spatially non-uniform degradations, while GDA addresses atmospheric-induced color distortions and contrast

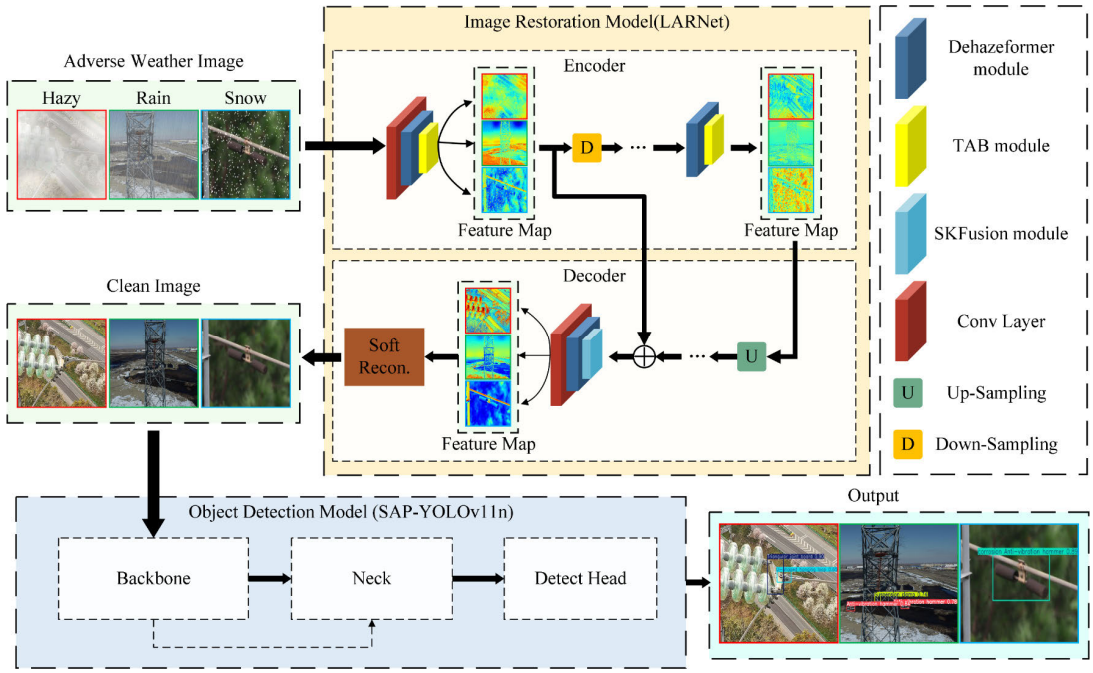


FIGURE 1. Working principle of LARNet-SAP-YOLOv11 model.

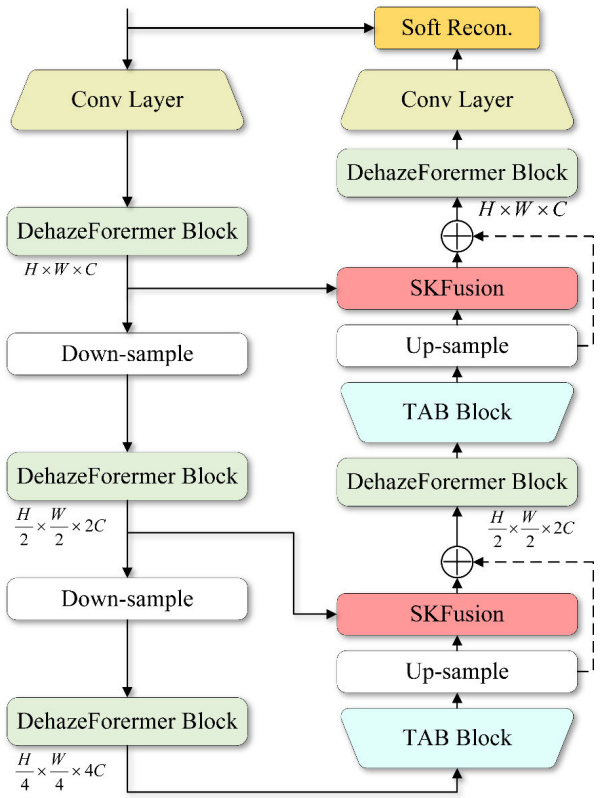


FIGURE 2. LARNet model architecture.

degradation [29]. By learning shared noise patterns across weather conditions, TAB enhances the model's robustness and provides stable, reliable features for corrosion defect detection.

In the TAB module, the input feature map is first processed through a convolutional layer, and then separately passed into LPA, GSA modules, and an additional convolutional branch. This produces three sets of sub-features, which enhance the model's perception of degraded regions.

$$\begin{cases} \boldsymbol{F}^{\mathrm{L}} = \mathrm{LPA}\left(\mathrm{Conv}\left(\mathrm{Conv}\left(\boldsymbol{F}\right)\right)\right) \\ \boldsymbol{F}^{\mathrm{G}} = \mathrm{GSA}\left(\mathrm{Conv}\left(\mathrm{Conv}\left(\boldsymbol{F}\right)\right)\right) \\ \boldsymbol{F}^{\mathrm{C}} = \mathrm{Conv}\left(\mathrm{Conv}\left(\boldsymbol{F}\right)\right) \end{cases} \tag{1}$$

where $\boldsymbol{F}$, $\boldsymbol{F}^{\mathrm{L}}$, $\boldsymbol{F}^{\mathrm{G}}$ and $\boldsymbol{F}^{\mathrm{C}} \in \mathbb{R}^{B\times C\times H\times W}$, The symbol $\mathbb{R}^{B\times C\times H\times W}$ represents feature information with a shape characterized by a batch size of B, channel number of C, height of H, and width of W. Among them, F is the input feature map; $\boldsymbol{F}^{\mathrm{L}}$, $\boldsymbol{F}^{\mathrm{G}}$ and $\boldsymbol{F}^{\mathrm{C}}$ denote the sub-feature branches, and Conv represents the convolution operation. These three sub-features are concatenated and then fused through a convolutional operation, followed by a residual connection to generate the multi-scale attention feature F^{M}, which is designed to address non-uniform degradation in the input image.

$$\begin{cases} \boldsymbol{F}^{\mathrm{M}} = \mathrm{Conv}\left(\mathrm{Concat}\left(\boldsymbol{F}^{\mathrm{L}}, \boldsymbol{F}^{\mathrm{G}}, \boldsymbol{F}^{\mathrm{C}}\right)\right) \\ \boldsymbol{F}^{\mathrm{D}} = \mathrm{GDA}\left(\boldsymbol{F}^{\mathrm{M}}\right) \end{cases} \tag{2}$$

where $\boldsymbol{F}^{\mathrm{M}} \in \mathbb{R}^{B\times C\times H\times W}$ is the fused feature; Concat denotes the concatenation operation across multiple feature maps; and GDA refers to the Global Distribution Attention. To address color distortion and contrast attenuation caused by scattering

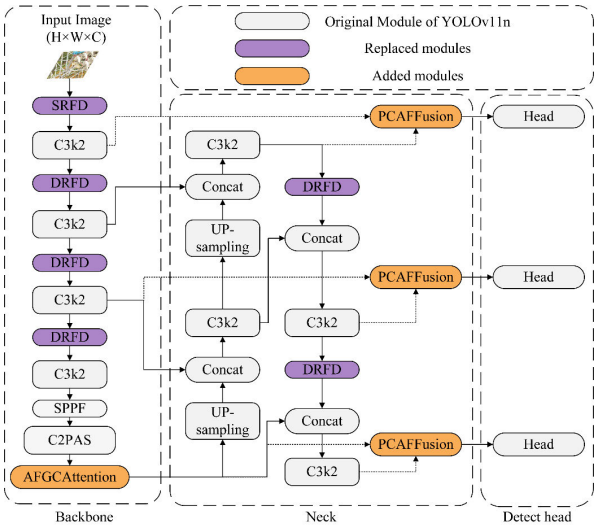


FIGURE 3. Architecture of the SAP-YOLOv11 model. Gray blocks denote original YOLOv11n modules, purple blocks indicate replaced modules, and orange blocks represent newly introduced modules.

effects, $\boldsymbol{F}^{\mathrm{M}}$ is passed through the GDA sub-module to obtain $\boldsymbol{F}^{\mathrm{D}}$. Then, a double residual connection is applied to globally adjust the feature distribution, further enhancing the restoration quality.

$$\boldsymbol{F}^{\mathrm{out}} = \boldsymbol{F}^{\mathrm{D}} + \boldsymbol{F}^{\mathrm{M}} + \mathrm{Conv}\,(\boldsymbol{F}) \tag{3}$$

where $\boldsymbol{F}^{\mathrm{out}} \in \mathbb{R}^{B\times C\times H\times W}$ is the final output of the TAB module.

B. SAP-YOLOv11 DESIGN

To improve the detection of corrosion defects on transmission line fittings under adverse conditions, we develop an improved object detection model named SAP-YOLOv11, based on the YOLOv11 model. As a representative lightweight detector, YOLOv11n demonstrates a well-balanced performance in terms of accuracy and efficiency on our corrosion dataset, making it suitable for deployment on resource-constrained embedded UAV systems. Thus, YOLOv11n is selected as the base structure, upon which we implement modular enhancements to construct the robust SAP-YOLOv11 model.

As shown in Fig. 3, the SAP-YOLOv11 architecture comprises three components: backbone, neck, and detection head. We introduce the following three major improvements:

(1) Shallow Robust Feature Downsampling (SRFD) and Deep Robust Feature Downsampling (DRFD) modules [50] are used to replace the original downsampling layers in both the backbone and neck, improving multi-level feature extraction.

(2) Adaptive Fine-Grained Channel Attention (AFGCAttention) [51] is introduced at the end of the backbone to enhance local detail extraction and global context awareness.

(3) Inspired by the Convolution and Attention Fusion Module (CAFM) from HCANet and PixelAttention from DEA-Net [14], [52], we design a novel Pixel-Level Cross-Attention Feature Fusion (PCAFFusion) module, positioned between the neck and the detection head. This module improves multi-scale feature fusion capabilities.

The SAP-YOLOv11 model significantly improves the detection accuracy and robustness for various types of transmission line fitting corrosion defects while maintaining efficient inference performance.

1) SRFD AND DRFD MODULES

To enhance feature extraction across different semantic levels, we replace the original YOLOv11n downsampling modules with the proposed SRFD and DRFD components. SRFD focuses on shallow-layer local detail enhancement, such as texture and edges, while DRFD strengthens deep-layer semantic abstraction and fusion.

The SRFD module adopts a two-stage downsampling structure. In the first stage, a large 7×7 convolutional kernel is applied to expand the receptive field and enhance the expressiveness of the input features. This is followed by a multi-branch combination of Cut-slice Downsampling (CutD), Depthwise Separable Convolutional Downsampling (DWConvD), and Group Convolution (GConv) to strengthen edge representation and maintain fine-grained texture information. The initial multi-branch downsampling stage processes the original feature map to produce a fused output with enhanced feature details. In the second stage, multi-scale features are further integrated, and a 3×3 convolution refines the fused information. Maxpooling Downsampling is then applied to preserve prominent spatial structures while reducing resolution, yielding the final downsampled feature map.

The DRFD module adopts a structure similar to the second stage of SRFD. The key difference lies in the use of the Gaussian Error Linear Unit (GELU) activation function within the DWConvD block, which enhances the non-linear modeling capacity of the feature extraction process.

These improvements collectively enable the model to better capture fine-grained corrosion features and enhance the robustness and precision of defect detection under diverse visual conditions.

2) AFGCATTENTION

To enhance the model's capability in identifying corrosion-prone areas on transmission line fittings, AFGCAttention module is incorporated at the end of the backbone network. This module operates based on a channel-level weighting mechanism, significantly improving feature selection and contextual modeling.

First, global average pooling is applied to compress spatial dimensions and concentrate on global channel-wise information:

$$\boldsymbol{x} = \mathrm{AvgPool}_{2\mathrm{d}}\,(\boldsymbol{X}) \tag{4}$$

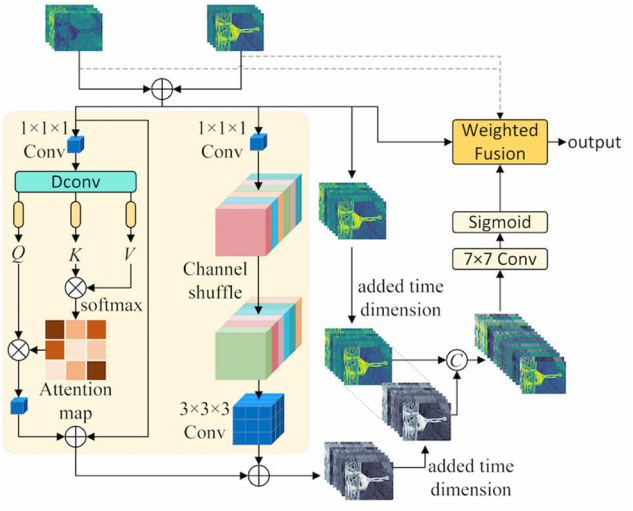


FIGURE 4. Schematic of the pixel-level cross-attention feature fusion(PCAFFusion) module, illustrating the integration of the CAFM unit and pixel-wise weighting for adaptive feature fusion.

where $\boldsymbol{X} \in \mathbb{R}^{B\times C\times H\times W}$ represents the input feature map, and $\boldsymbol{x} \in \mathbb{R}^{B\times C\times 1\times 1}$ denotes the feature obtained after global average pooling.

Next, the pooled feature is processed through two parallel paths to generate intermediate features. One path applies a one-dimensional convolutional layer with a dynamic kernel size, while the other employs a 1×1 convolutional layer after a transpose operation to align matrices, followed by the removal of singleton dimensions. These operations produce two intermediate feature representations. Forward and backward attention scores are then calculated through dual-direction attention mechanisms, using matrix multiplication to compute channel attention, followed by an unsqueeze operation to restore tensor dimensions for broadcasting. The resulting forward attention score and backward attention score are fused via a Mix operation, followed by a sigmoid activation to generate the final attention-guided output. The process can be expressed as (5):

$$\boldsymbol{Y} = \boldsymbol{X} \cdot \text{Sigmoid}\left(\text{Conv1d}\left(\text{Mix}\left(S_f, S_b\right)\right)^{\text{T}}\right) \tag{5}$$

where $\boldsymbol{Y} \in \mathbb{R}^{B\times C\times H\times W}$ represents the final output feature map after applying channel-wise attention weighting. S_f and S_b represent the forward and backward attention scores, respectively, T denotes the transpose operation applied to feature vectors, and Conv1d refers to a one-dimensional convolutional layer.

By introducing the AFGCAttention module, the YOLOv11n model exhibits significantly enhanced sensitivity to small corrosion regions, effectively reducing both missed detections and false positives.

3) PCAFFUSION

To optimize the feature transmission and fusion between the neck and the detection head of the network, we introduce the PCAFFusion module at this junction, as shown in Fig. 4. This module leverages a pixel-level cross-attention mechanism to integrate multi-level features, thereby enhancing the representation of corrosion regions.

Specifically, the PCAFFusion module first receives two feature maps at different scales and performs an initial fusion:

$$\boldsymbol{F} = \boldsymbol{X} + \boldsymbol{Y} \tag{6}$$

where $\boldsymbol{X}$ and $\boldsymbol{Y} \in \mathbb{R}^{B\times C\times H\times W}$ denotes the input feature maps; $\boldsymbol{F} \in \mathbb{R}^{B\times C\times H\times W}$ is the initially fused representation. ubsequently, the CAFM (Convolution and Attention Fusion Module) submodule employs a multi-head self-attention mechanism combined with deep convolutional layers to perform global modeling on $\boldsymbol{F}$:

$$\text{part}_1 = \text{CAFM}\left(\boldsymbol{F}\right) \tag{7}$$

where $\text{part}_1 \in \mathbb{R}^{B\times C\times H\times W}$ is the output feature map after processing x3 through the CAFM module. Next, pixel-wise Attention is applied to extract fine-grained local feature weights, highlighting the salient characteristics of the corroded regions. A sigmoid activation is then used to produce the adaptive weighting factor:

$$\text{part}_2 = \text{Sigmoid}\left(\text{PA}\left(\boldsymbol{F}, \text{part}_1\right)\right) \tag{8}$$

where $\text{part}_2 \in \mathbb{R}^{B\times C\times H\times W}$ is the weighted result produced by pixel attention and activation; PA denotes the pixel attention mechanism. The initially fused and weighted features are then adaptively combined:

$$\text{result} = \boldsymbol{F} + \text{part}_2 \cdot \boldsymbol{X} + (1 - \text{part}_2) \cdot \boldsymbol{Y} \tag{9}$$

where result $\in \mathbb{R}^{B\times C\times H\times W}$ is the final result after multi-scale feature integration. Finally, a 1×1 convolution is applied for channel adjustment and feature refinement, generating the final output:

$$\boldsymbol{Z} = \text{Conv}_{1\times 1}\left(\text{result}\right) \tag{10}$$

where $\boldsymbol{Z} \in \mathbb{R}^{B\times C\times H\times W}$ is the output feature map; $\text{Conv}_{1\times 1}$ represents the 1×1 convolution operation.

By integrating the PCAFFusion module, the object detection model can adaptively combine both local and global multi-scale information. This enhances the receptive field of the detection head, improves the quality of incoming features, and significantly strengthens the model's capability and robustness in detecting complex, textured corrosion defects.

IV. EXPERIMENTS AND RESULTS

To comprehensively evaluate the effectiveness and robustness of the proposed joint model in both image restoration and defect detection tasks, we trained the image restoration and object detection models independently and performed inference jointly during the test stage. Additionally, comparative experiments were conducted by combining various all-in-one image restoration methods with the baseline and improved detection models.

A. DATASET

The original dataset used in this study was constructed using high-resolution images of transmission lines captured by UAVs under clear weather conditions. To ensure comprehensive data collection, a two-side round-trip flight strategy was employed for UAVs, which avoids incomplete data acquisition caused by capturing images from a single angle while maximizing the reduction of occlusions and shadows. During image acquisition, to ensure the safety of UAV operations near transmission lines and to enhance the resolution of captured images, the distance between the UAV and the transmission lines or towers was dynamically adjusted between 5 and 25 meters based on different targets, with a flight speed maintained below 6 m/s.

1) MULTI-WEATHER DEGRADED TRANSMISSION LINE DATASET

This dataset was generated by applying image processing software to 4,156 original clear images to simulate weather effects such as haze, rain, and snow. Among them, 3,739 synthetic images were used for training, and 417 images were used to generate the test set for each of the three weather conditions, resulting in 1,251 test images under different weather conditions. The test set was completely excluded from training to ensure objectivity and support generalization assessment, yielding a total of 4,990 degraded images. Examples of the custom-built multi-weather degradation dataset are illustrated in Fig. 5.

As shown in Fig. 5, subfigures (a)–(c) present clear images with distinguishable features, enabling easy identification of key fittings and their potential corrosion defects. In contrast, subfigures (d)–(f) exhibit weather-degraded images affected by haze, rain, or snow, with evident occlusions, blurring, and feature loss, which hinder accurate recognition of defects and compromise detection effectiveness.

2) CORRODED TRANSMISSION LINE FITTINGS DATASET

This dataset comprises 4,156 UAV images containing corroded transmission line fittings, captured under clear weather conditions and manually annotated for defect types and corrosion conditions. The dataset was split into training, validation, and test sets at a ratio of 8:1:1. Sample images are shown in Fig. 6. Fig. 6 illustrates normal fittings in subfigures (a)–(d), which exhibit smooth surfaces and intact structural features. In contrast, the corroded fittings shown in subfigures (e)–(h) display visible surface degradation and structural damage, indicating potential safety hazards.

3) PUBLIC MULTI-WEATHER IMAGE RESTORATION DATASET

For the restoration tasks under haze, rain, and snow conditions, we selected subsets of three public datasets for restoring images affected by different weather conditions, namely: hazy (“RESIDE-OTS” [53]), rainy (“Rain1400” [54]), and snow (“Snow100k-L” [19]). To ensure sample balance and enhance the generalization of the restoration model, we uniformly selected 5,000 images from each of the three weather-specific datasets, totaling 15,000 images for training, and evaluation was conducted on the corresponding test sets: 1,000 for RESIDE-SOTS, 1,400 for Rain1400, and 1,681 for Snow100k-L.

4) PUBLIC OBJECT DETECTION DATASET

To validate the generalization performance of the object detection module, we selected the Visual Object Classes (VOC) (2007+2012) as the public object detection dataset. This dataset includes 16,551 images for training, 8,333 for validation, and 4,952 for final testing, covering 20 different object categories. The VOC dataset serves as an important benchmark in the field of computer vision, providing rich image and annotation data for tasks such as object detection and image segmentation.

B. EXPERIMENT DETAILS

All experiments in this paper were conducted on a Windows 10 operating system with an Intel i7-11700k CPU and an NVIDIA GeForce RTX 3090 GPU (24 GB VRAM). The deep learning framework was PyTorch 1.13; the programming language was Python 3.8; and the accelerated computing framework was CUDA 11.6. For the LARNet-SAP-YOLOv11 model, we separately trained the LARNet image restoration submodel and the SAP-YOLOv11 object detection submodel, with specific parameter settings as described in the following two subsections. In addition, implementation details of different experiments are elaborated in the corresponding parts of Subsection D.

1) PARAMETER CONFIGURATION OF IMAGE RESTORATION TASKS

The input size of images was set to 256×256 pixels, with a learning rate of 2×10.4. The optimizer was Adam with weight decay fix (AdamW), with a weight decay of 0.0005. The loss function adopted was L1Loss, with automatic mixed precision training (AMP) enabled. The batch size was 4, the total training duration was 400 epochs, and validation was performed after each epoch.

2) PARAMETER CONFIGURATION OF OBJECT DETECTION TASKS

The input size of images was set to 640×640 pixels. The learning rate was 0.01, and the optimizer was Stochastic Gradient Descent (SGD) with a momentum of 0.9 and a weight decay of 0.0005. The loss function consisted of BCEWithLogits loss (cross-entropy loss) and Complete Intersection over Union (CIoU) [55] loss function, where object confidence loss and classification loss used cross-entropy loss, and regression loss used CIoU loss, with automatic mixed precision training (AMP) enabled. The batch size was 32, and the training duration was 300 epochs, with validation performed after each epoch. The validation confidence threshold was set to 0.001, and the Non-Maximum Suppression (NMS)

Intersection over Union (IoU) threshold was set to 0.7. Additionally, Mosaic data augmentation was disabled in the final 10 epochs to enhance the fitting ability and accuracy on real data.

C. EXPERIMENTAL METRICS

1) IMAGE RESTORATION METRICS

We evaluated the performance of the image restoration models using multiple metrics, including Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Natural Image Quality Evaluator (NIQE), Learned Perceptual Image Patch Similarity (LPIPS), and Deep Image Structure and Texture Similarity (DISTS). In addition to the commonly used PSNR and SSIM, NIQE, LPIPS, and DISTS were introduced to provide a more comprehensive evaluation of perceptual quality. Specifically, NIQE measures the naturalness of restored images, with lower scores indicating better visual quality. LPIPS evaluates perceptual similarity based on deep features extracted from pre-trained networks, where lower values correspond to closer alignment with human perception. DISTS assesses both structural and textural similarity using deep features, with lower scores reflecting higher perceptual consistency. These metrics together enable an objective and perceptual evaluation of restoration quality under hazy, rainy, and snowy conditions.

In addition to restoration performance, we used Model Size, Parameters (Params), and Floating-Point Operations Per Second (FLOPS) to comprehensively assess the models' lightweight characteristics and computational efficiency.

2) OBJECT DETECTION METRICS

For evaluating the object detection models, we used the mean Average Precision (mAP) at IoU thresholds of 0.5 (mAP@0.5) and 0.5:0.95 (mAP@0.5:0.95), as well as Precision (P), Recall (R), and Parameters. mAP@0.5 quantifies detection performance at a fixed IoU threshold of 0.5, while mAP@0.5:0.95 averages performance across thresholds from 0.5 to 0.95 with a step of 0.05. Precision reflects the ability to avoid false positives(FP), and Recall measures the ability to detect all relevant instances. The number of parameters reflects the complexity of the network and its inference efficiency. These metrics provide a comprehensive evaluation of object detection performance.

D. EXPERIMENTAL RESULTS AND ANALYSIS

1) IMAGE RESTORATION RESULTS

To evaluate the performance of the proposed image restoration model, LARNet, we conducted comparative experiments with several state-of-the-art models, including all-in-one restoration models—TransWeather [27], PromptIR [28], TANet [29], MWFormer [24]. Additionally, the state-of-the-art dehazing model ChaIR [56] and the baseline model DehazeFormer [15] were included for comparison on the SOTS test set. All models were evaluated under identical initial conditions using the same training and test sets, with input images uniformly resized to 256 × 256 pixels. Each model was trained for 400 epochs following its respective original settings. The comparative results on the multi-weather degraded transmission line dataset are presented in Tables 1 and 2, while those on the public multi-weather image restoration dataset are provided in Tables 3 and 4.

As shown in Tables 1 and 2, LARNet demonstrates superior restoration capabilities across different weather degradation scenarios. Specifically, in Table 1, LARNet achieves the highest peak signal-to-noise ratio (PSNR) values in hazy, rainy, and snowy conditions, reaching 28.53 dB, 33.17 dB, and 29.59 dB, respectively, with an overall average of 30.43 dB. Moreover, the SSIM scores under haze and rain conditions reached the highest values of 0.950 and 0.962, respectively. Notably, the model size of LARNet is only 12.5 MB, with a parameter count of 2.968M, significantly smaller than MWFormer, PromptIR, and TransWeather, while its computational complexity (FLOPS) is considerably lower than that of PromptIR and TANet. These results confirm that LARNet delivers high-quality image restoration with low resource consumption and excellent lightweight characteristics, making it highly suitable for deployment on resource-constrained UAV inspection platforms.

In Table 2, we further present the performance comparison of LARNet and other models on perceptual metrics for image restoration on the multi-weather degraded transmission line dataset. As shown in Table 2, LARNet achieves superior values for NIQE, LPIPS, and DISTS under hazy, rainy, and snowy conditions. Ultimately, LARNet obtains the best overall LPIPS and DISTS scores across haze, rain, and snow conditions, with values of 0.0623 and 0.0607, respectively. These results indicate that LARNet's restorations align more closely with human visual perception, producing images more consistent with the human visual system.

To further assess the generalization ability of LARNet, we conducted experiments on the public multi-weather image restoration dataset described in Section IV, Subsection A, Point 3. The results are summarized in Tables 3 and 4.

As shown in Tables 3 and 4, LARNet exhibits consistently strong performance across all three public datasets, achieving superior values for multiple metrics on hazy, rainy, and snowy public datasets. These results further validate its domain generalization capability and adaptability in diverse weather conditions.

2) DEFECT DETECTION RESULTS

To evaluate the effectiveness of object detection models in the task of transmission line fitting corrosion detection, we conducted systematic experiments on our custom corrosion fitting dataset using various mainstream detection models. The evaluated models include lightweight versions from the YOLO series: YOLOv5n, YOLOv6n, YOLOv7-tiny, YOLOv8n, YOLOv9t, YOLOv10n, YOLOv11n, and our proposed SAP-YOLOv11. In addition, SSD, Faster R-

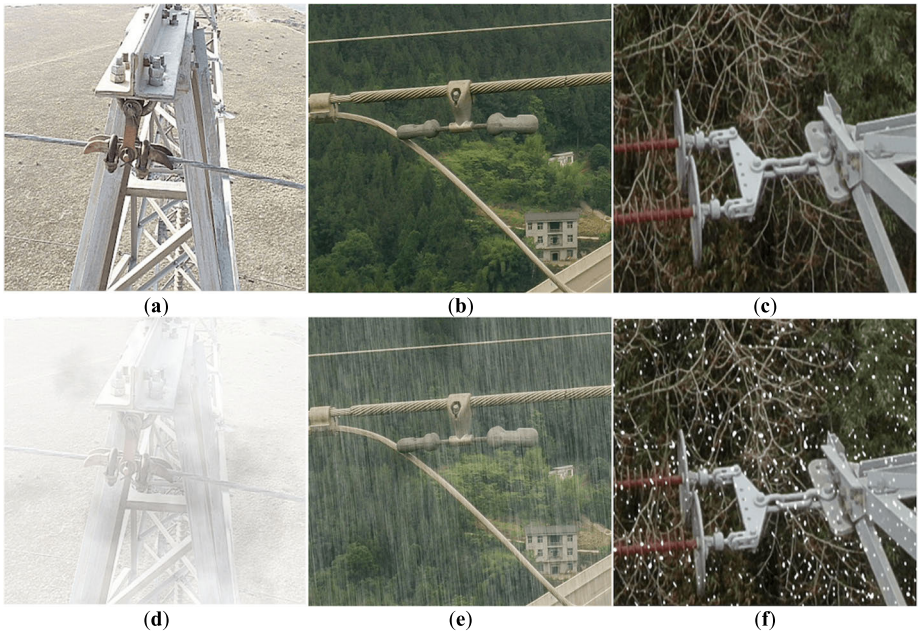


FIGURE 5. Sample images of degraded transmission lines under various weather conditions. Subfigures (a), (b), and (c) show clear images under haze-free, rain-free, and snow-free conditions, respectively, while subfigures (d), (e), and (f) illustrate the corresponding degraded images with added haze, rain, and snow effects.

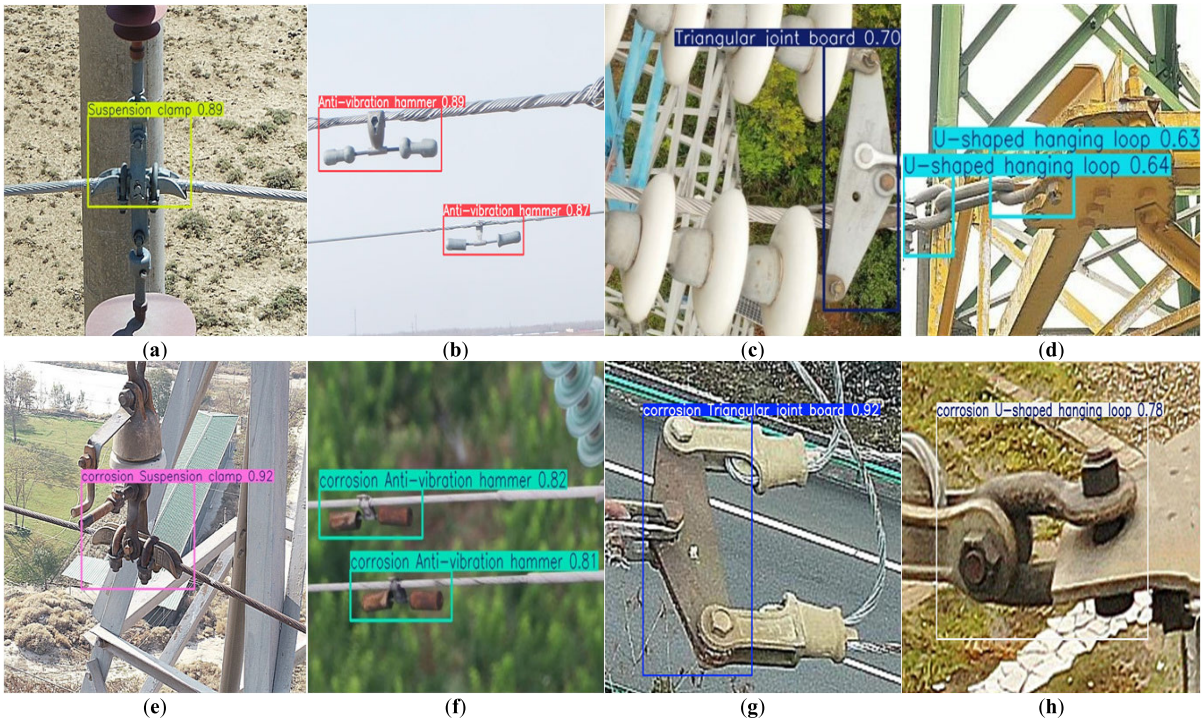


FIGURE 6. Sample images of corroded transmission line fittings. Subfigures (a), (b), (c), and (d) represent the normal conditions of suspension clamps, anti-vibration hammers, triangular joint boards, and U-shaped hanging loops, respectively. Subfigures (e), (f), (g), and (h) show the corresponding components in corroded conditions.

CNN, and RT-DETR-r18 were included as larger-model baselines for comparison. The experimental results are shown in Table 5.

As seen in Table 5, YOLOv11n achieves a good balance between detection performance and model efficiency, with a mAP@50 of 87.5%, precision of 83.9%, and model size

TABLE 1. Comparison of image restoration performance on the multi-weather degraded transmission line dataset.

Model	Haze		Rain		Snow		Mean value		Model Size	Params(M)	FLOPS(G)
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM			
Transweather	19.53	0.771	24.08	0.780	22.16	0.778	21.92	0.776	145.3 MB	38.05	**6.14**
PromptIR	22.99	0.864	29.51	0.906	24.23	0.925	25.58	0.898	406.3 MB	35.59	158.14
TANet	27.36	0.940	32.70	0.955	28.53	0.930	29.53	0.941	37.8 MB	8.97	50.06
Dehazeformer-b	28.00	0.948	32.55	0.958	**29.27**	0.935	29.94	0.947	**10.7 MB**	**2.51**	25.79
ChaIR	**28.09**	**0.952**	**33.35**	**0.968**	28.38	0.930	29.94	0.950	57.6 MB	15.02	141.04
MWFormer	28.01	0.949	**33.86**	0.961	29.17	**0.950**	**30.34**	**0.953**	649+47.7 MB	182.81	**20.87**
LARNet	**28.53**	**0.950**	33.17	**0.962**	**29.59**	**0.941**	**30.43**	**0.951**	**12.5 MB**	**2.97**	27.03

Note: Values that are both bold and underlined denote the best in each column; bold values indicate the second-best.

TABLE 2. Comparison of image restoration performance using perceptual metrics on the multi-weather degraded transmission line dataset.

Model	Haze			Rain			Snow			Meanvalue		
	NIQE	LPIPS	DISTS	NIQE	LPIPS	DISTS	NIQE	LPIPS	DISTS	NIQE	LPIPS	DISTS
Transweather	**5.3701**	0.3460	0.2865	5.5533	0.3944	0.3102	**5.5599**	0.4081	0.2803	**5.4944**	0.3828	0.2923
PromptIR	6.9981	0.1291	0.0959	7.6206	0.1269	0.0933	7.8828	0.1782	0.1000	7.5005	0.1447	0.0964
TANet	6.9876	0.0775	0.0899	7.5404	0.0718	0.0825	7.5158	**0.0726**	**0.0707**	7.3479	0.0739	0.0810
Dehazeformer-b	6.3819	**0.0528**	**0.0584**	6.1420	0.0491	0.0616	6.2720	0.0938	0.0745	6.2653	**0.0652**	**0.0648**
ChaIR	5.7822	0.0573	0.0603	**5.4626**	**0.0381**	**0.0422**	5.5001	0.1196	0.0759	5.5817	0.0717	0.0595
MWFormer	**4.5387**	0.1534	0.1404	**5.0556**	0.1337	0.1034	**4.6764**	0.1033	0.0792	**4.7569**	0.1301	0.1077
LARNet	6.4065	**0.0508**	**0.0573**	6.1536	**0.0465**	**0.0560**	6.2771	**0.0888**	**0.0692**	6.2791	**0.0620**	**0.0608**

Note: Values that are both bold and underlined denote the best in each column; bold values indicate the second-best.

TABLE 3. Image restoration results on the public multi-weather image restoration dataset.

Model	Haze1000		Rain1400		Snow1681		Multiple Weather4081		Model Size	Params(M)	FLOPS(G)
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM			
Transweather	19.56	0.680	26.41	0.829	23.11	0.722	23.37	0.749	145.3 MB	38.05	**6.14**
PromptIR	19.31	0.722	25.91	0.807	19.66	0.709	21.72	0.746	406.3 MB	35.59	158.14
TANet	22.07	0.704	31.64	0.937	**30.10**	**0.906**	**28.66**	**0.867**	37.8 MB	8.97	50.06
Dehazeformer-b	21.29	0.701	31.46	0.929	27.70	0.877	27.42	0.840	**10.7 MB**	**2.51**	25.79
ChaIR	22.17	**0.716**	30.69	0.917	30.08	**0.913**	28.35	0.866	57.6 MB	15.02	141.04
MWFormer	**22.78**	**0.770**	**32.59**	**0.921**	**30.20**	0.902	**29.20**	**0.876**	649+47.7 MB	182.81	**20.87**
LARNet	**22.20**	0.710	**32.30**	**0.938**	28.86	0.897	28.40	0.865	**12.5 MB**	**2.97**	27.03

Note: Values that are both bold and underlined denote the best in each column; bold values indicate the second-best.

TABLE 4. Image restoration results using perceptual metrics on the public multi-weather image restoration dataset.

Model	Haze1000			Rain1400			Snow1681			Multiple Weather4081		
	NIQE	LPIPS	DISTS	NIQE	LPIPS	DISTS	NIQE	LPIPS	DISTS	NIQE	LPIPS	DISTS
Transweather	4.5192	0.3457	0.2429	3.7966	0.3494	0.2736	3.7805	0.4256	0.3053	3.9726	0.3805	0.2796
PromptIR	7.4896	0.1998	0.0718	8.5058	0.2172	0.1190	7.6486	0.3534	0.1801	7.9149	0.2696	0.1329
TANet	**3.8385**	0.1673	0.0623	**3.3438**	0.0990	**0.0608**	**3.6513**	0.1424	**0.0788**	**3.5917**	0.1336	**0.0686**
Dehazeformer-b	4.3961	**0.1497**	0.0645	3.8725	**0.0566**	0.0689	4.2173	**0.1247**	0.1001	4.1428	**0.1075**	0.0807
ChaIR	4.1553	0.1557	**0.0607**	4.2475	0.0909	0.0802	4.1635	**0.1284**	**0.0813**	4.1309	0.1222	**0.0759**
MWFormer	**3.8681**	0.1676	**0.0622**	**3.6744**	0.1225	0.0712	**3.6453**	0.1602	0.0933	**3.7152**	0.1493	0.0777
LARNet	4.4412	**0.1552**	0.0704	3.7004	**0.0673**	**0.0682**	3.8916	0.1413	0.1116	3.9607	**0.1193**	0.0866

Note: Values that are both bold and underlined denote the best in each column; bold values indicate the second-best.

of 2.58M parameters—making it well-suited for deployment in engineering applications. Based on this model, the proposed SAP-YOLOv11 achieves significant performance gains: mAP@50 increases to 89.6% (a 2.1% improvement), precision reaches 93.0%, and mAP@50:95 rises to 65.9%, which is just slightly lower than RT-DETR-r18 (66.2%), despite the latter's much larger model size. These results indicate that SAP-YOLOv11 achieves superior detection accuracy and robustness while maintaining a lightweight architecture.

To further validate the effectiveness of the proposed improvements, we conducted ablation experiments on the

TABLE 5. Performance comparison of different object detection models on the corroded transmission line fittings dataset.

Model	mAP@50(%)	mAP@50:95(%)	P(%)	R(%)	parameters
Faster-RCNN	80.8	49.5	64.0	82.7	28,480,228
SSD	87.4	59.5	87.8	82.6	26,285,486
RT-DETR-r18	**88.5**	**66.2**	**91.6**	85.7	19,882,032
YOLOv5n	87.1	59.9	91.0	80.9	**1,774,741**
YOLOv6n	82.5	59.8	91.0	76.5	4,630,596
YOLOv7-tiny	86.7	59.4	83.9	**87.5**	6,026,538
YOLOv8n	87.3	64.1	91.4	80.8	2,685,928
YOLOv9t	84.5	60.4	86.6	82.0	2,662,016
YOLOv10n	87.3	63.3	88.4	83.9	2,697,536
YOLOv11n	87.5	64.0	83.9	**86.3**	**2,583,712**
SAP-YOLOv11	**89.6**	**65.9**	**93.0**	83.0	3,604,441

Note: Values that are both bold and underlined denote the best in each column; bold values indicate the second-best.

TABLE 6. Ablation study results of the SAP-YOLOv11 model on the corroded transmission line fittings dataset.

Model	mAP@50(%)	mAP@50:95(%)	P(%)	R(%)	Parameters
YOLOv11n	87.5	64.0	83.9	**86.3**	**2,583,712**
YOLOv11n+SRFD	87.7	64.3	91.3	82.9	**2,555,392**
YOLOv11n+PCAFFusion	87.5	64.2	88.8	82.4	3,507,827
YOLOv11n+AFGCAttention	87.6	63.5	91.5	81.3	2,649,510
YOLOv11n+SRFD+PCAFFusion	**88.9**	**66.3**	**92.0**	84.0	3,538,643
YOLOv11n+SRFD-AFGCAttention	88.7	65.0	88.4	**85.9**	2,621,190
YOLOv11n+PCAFFusion-AFGCAttention	88.2	65.7	90.0	83.6	3,573,625
SAP-YOLOv11	**89.6**	**65.9**	**93.0**	83.0	3,604,441

Note: Values that are both bold and underlined denote the best in each column; bold values indicate the second-best.

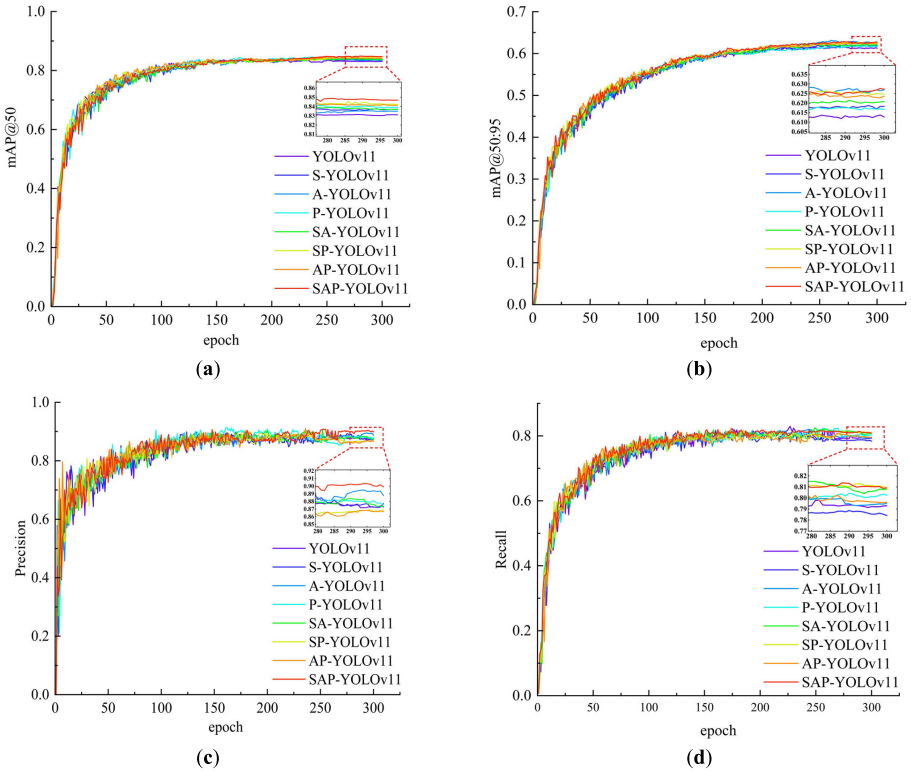


FIGURE 7. Comparison of training performance for different combinations of the proposed improvement modules. Subfigures (a), (b), (c), and (d) show the curves of mAP@50, mAP@50:95, precision, and recall, respectively.

three modules introduced in SAP-YOLOv11: SRFD, AFGCAttention, and PCAFFusion. A total of eight experiments were conducted, and the results are summarized in Table 6. As seen in Table 6, each of the three modules contributes to

TABLE 7. Ablation study results of the SAP-YOLOv11 model on the VOC dataset.

Model	mAP@50(%)	mAP@50:95(%)	P(%)	R(%)	Parameters
YOLOv11n	79.3	59.5	**80.9**	71.0	**2,583,712**
YOLOv11n+SRFD	79.4	59.4	80.4	71.9	**2,555,392**
YOLOv11n+PCAFFusion	**80.8**	**60.8**	79.7	**73.6**	3,507,827
YOLOv11n+AFGCAttention	79.2	59.1	79.7	71.7	2,649,510
YOLOv11n+SRFD+PCAFFusion	80.4	60.2	80.2	**73.8**	3,538,643
YOLOv11n+SRFD+AFGCAttention	80.0	60.3	80.3	71.9	2,621,190
YOLOv11n+PCAFFusion+AFGCAttention	**80.5**	**60.8**	79.7	73.3	3,573,625
SAP-YOLOv11	**80.8**	**60.6**	**80.5**	73.3	3,604,441

Note: Values that are both bold and underlined denote the best in each column; bold values indicate the second-best.

TABLE 8. Performance comparison of joint models under various weather conditions.

Model	Datasets(mAP@50(%))				Inference times(ms)	Detection distance
	Hazy	Rain	Snow	Multiple weather		
No Image Restoration -YOLOv11n	63.4	82.9	85.7	76.5	**10.12**	5m-25m
TransWeather-YOLOv11n	83.7	83.3	83.8	83.4	—	
PromptIR-YOLOv11n	85.4	86.5	86.5	85.7	—	
TANet-YOLOv11n	86.0	87.5	87.4	86.7	—	
MWFormer-YOLOv11n	86.0	87.7	87.0	86.6	—	
ChaIR-YOLOv11	85.3	86.5	86.3	85.9	—	
Dehazeformer-b-YOLOv11n	86.5	87.7	87.5	86.9	62.43	
LARNet-YOLOv11n	86.4	87.3	87.4	86.7	65.41	
No Image Restoration -SAP-YOLOv11	69.6	83.3	86.5	78.8	**15.93**	
TransWeather-SAP-YOLOv11	84.7	83.3	85.3	84.2	—	
PromptIR-SAP-YOLOv11	86.5	87.2	86.9	86.6	—	
TANet-SAP-YOLOv11	87.5	89.0.	88.2	88.2	—	
MWFormer-SAP-YOLOv11	87.3	**89.7**	**89.3**	**88.4**	—	
ChaIR-SAP-YOLOv11	87.0	87.9	88.5	87.7	—	
Dehazeformer-b-SAP-YOLOv11	**87.9**	**89.2**	88.6	88.3	69.46	
LARNet-SAP-YOLOv11	**88.0**	89.1	**88.9**	**88.6**	74.87	

Note: Values that are both bold and underlined denote the best in each column; bold values indicate the second-best.

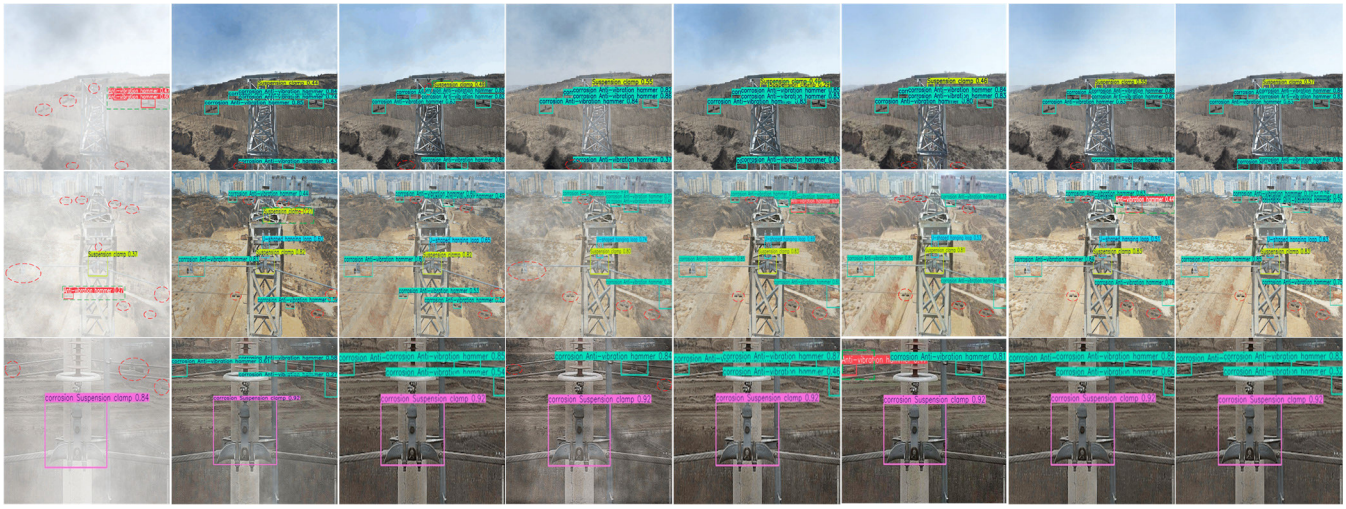
TABLE 9. Ablation study on the impact of individual components on final detection performance under adverse weather conditions.

Exp. No.	Model	mAP@50(%)	mAP@50:95(%)	P(%)	R(%)	Parameters
1	YOLOv11n	76.5	52.8	82.7	71.7	**2,583,712**
2	YOLOv11n+SRFD	77.1	53.7	84.0	72.3	**2,555,392**
3	YOLOv11n+PCAFFusion	76.5	53.3	86.1	68.8	3,507,827
4	YOLOv11n+AFGCAttention	76.0	52.5	86.8	67.6	2,649,510
5	YOLOv11n+SRFD+PCAFFusion	77.2	54.2	85.9	70.3	3,538,643
6	YOLOv11n+SRFD+AFGCAttention	78.0	53.8	86.5	72.0	2,621,190
7	YOLOv11n+PCAFFusion+AFGCAttention	78.7	54.3	85.0	73.5	3,573,625
8	SAP-YOLOv11	78.7	54.5	84.6	74.0	3,604,441
9	LARNet-YOLOv11n	86.8	62.7	85.7	**84.2**	5,551,982
10	LARNet-YOLOv11n+SRFD	86.6	63.1	89.9	82.4	5,523,662
11	LARNet-YOLOv11n+PCAFFusion	86.1	62.9	89.3	81.2	6,476,097
12	LARNet-YOLOv11n+AFGCAttention	86.0	62.0	90.8	79.6	5,617,780
13	LARNet-YOLOv11n+SRFD+PCAFFusion	**88.3**	**64.5**	**91.9**	81.8	6,506,913
14	LARNet-YOLOv11n+SRFD+AFGCAttention	87.9	63.4	**91.6**	83.3	5,589,460
15	LARNet-YOLOv11n+PCAFFusion+AFGCAttention	87.8	64.2	89.9	83.6	6,541,895
16	LARNet-SAP-YOLOv11	**88.6**	**64.7**	89.6	**84.0**	6,572,711

Note: Values that are both bold and underlined denote the best in each column; bold values indicate the second-best.

improvements in mAP and precision, enhancing the model's ability to detect corroded fittings. The full SAP-YOLOv11 model, which integrates all three modules, achieved the best performance across all metrics: mAP@50 of 89.6%, mAP@50:95 of 65.9%, and precision of 93.0%.

Fig. 7 clearly presents the training performance of various combinations of the three enhancement modules in terms of mAP@50, mAP@50:95, precision, and recall. As illustrated, the detection performance progressively improves with the inclusion of additional modules, and the complete



(a) YOLO | (b) PromptIR +YOLO | (c) TANet +YOLO | (d) Transweather +YOLO | (e) MWFormer +YOLO | (f) ChaIR +YOLO | (g) Dehazeformer +YOLO | (h) LARNet +YOLO(ours)

FIGURE 8. Detection results on hazy transmission line images after restoration using different restoration models. (Red dashed circles and green dashed rectangles denote missed and false detections, respectively. YOLO represents SAP-YOLOv11.)

(a) YOLO | (b) PromptIR +YOLO | (c) TANet +YOLO | (d) Transweather +YOLO | (e) MWFormer +YOLO | (f) ChaIR +YOLO | (g) Dehazeformer +YOLO | (h) LARNet +YOLO(ours)

FIGURE 9. Detection results on rainy transmission line images after restoration using different restoration models. (Red dashed circles and green dashed rectangles denote missed and false detections.)

SAP-YOLOv11 configuration achieves the most stable and superior results.

To further assess the generalization ability of the improvement modules in broader detection scenarios, the same ablation experiments were reproduced on the visual object classes (VOC) (2007+2012) public dataset. As shown in Table 7, the full model integrating all three modules also achieved the best performance in terms of comprehensive metrics, confirming its strong cross-dataset generalization capabilities. These ablation experiments confirm the effectiveness of the SRFD, AFGCAttention, and PCAFFusion modules integrated into SAP-YOLOv11.

3) JOINT INFERENCE PERFORMANCE COMPARISON

To further validate the effectiveness of image restoration in enhancing object detection performance, we conducted joint inference experiments on a test set of corroded transmission line fittings under adverse weather conditions. Specifically, we evaluated combinations of different image restoration models—No Image Restoration, TransWeather, PromptIR, TANet, MWFormer, DehazeFormer-b, and LARNet—with two detection models, YOLOv11n and SAP-YOLOv11, to assess their detection performance under various weather scenarios. In addition, to evaluate inference efficiency, we specifically compared the processing speeds

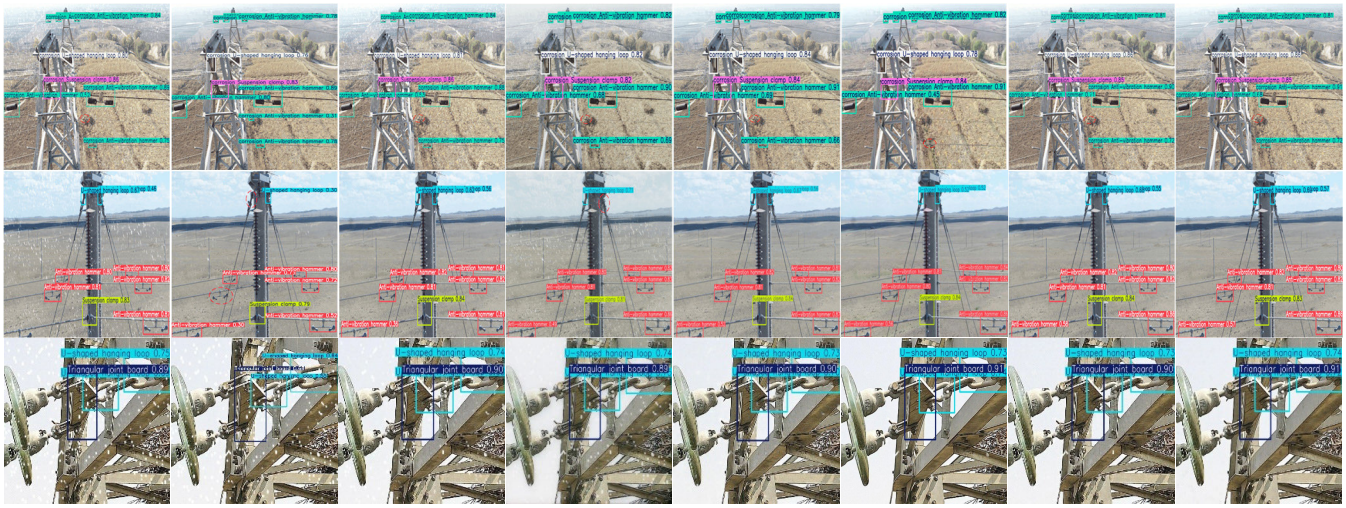


FIGURE 10. Detection results on snowy transmission line images after restoration using different restoration models. (Red dashed circles and green dashed rectangles denote missed and false detections.)

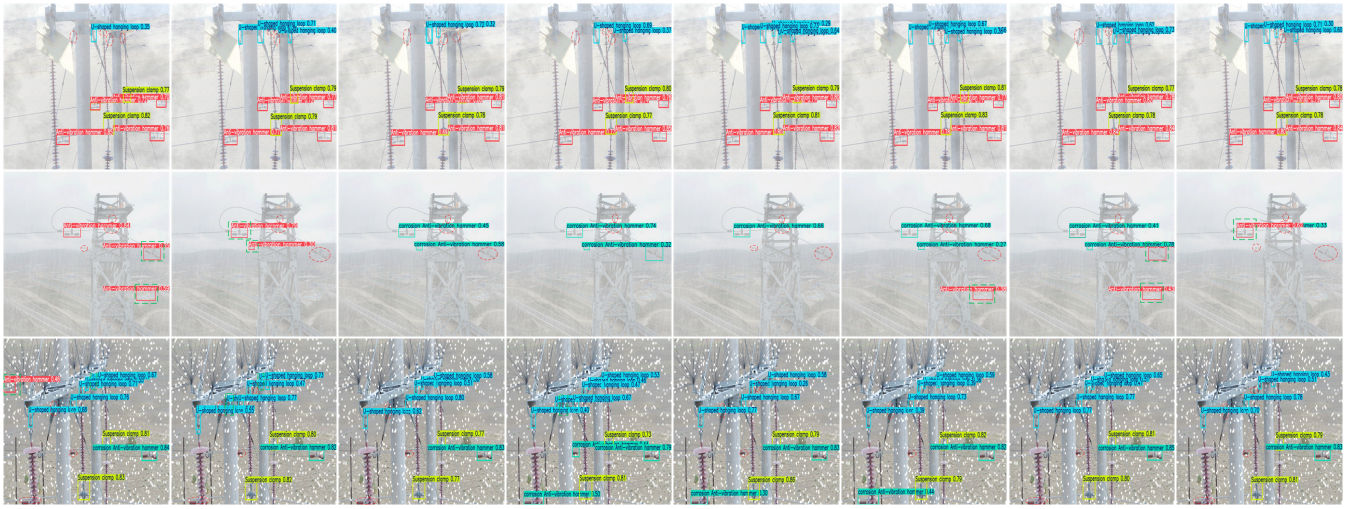


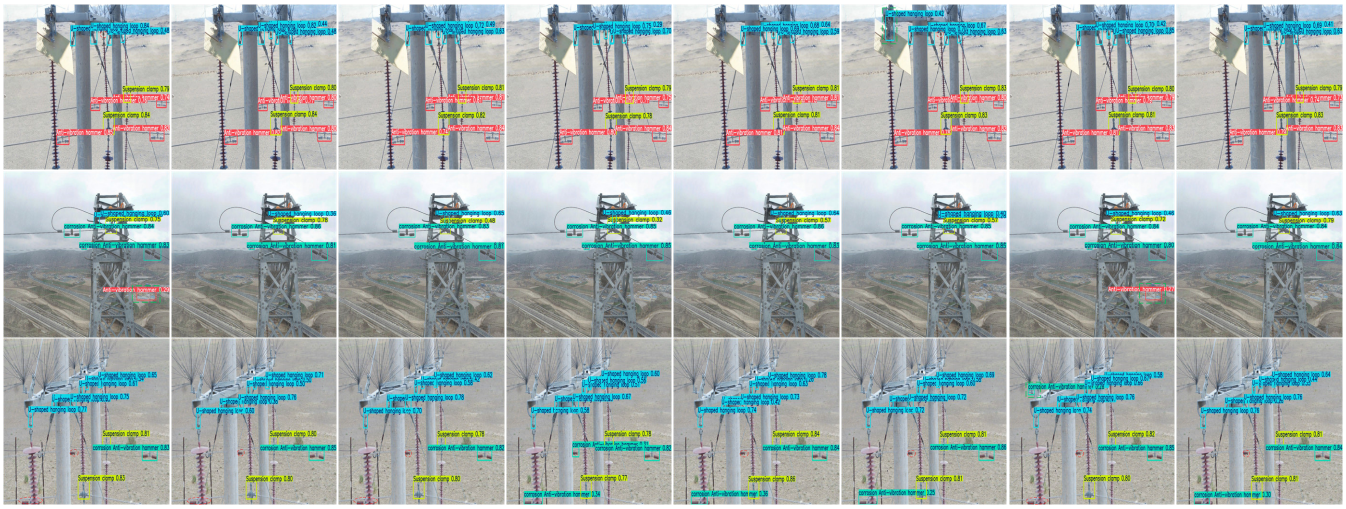
FIGURE 11. Impact of Individual components on final performance without considering the LARNet restoration submodel. (Red dashed circles and green dashed rectangles denote missed and false detections, respectively. In (e), (f), (g) and (h), S, P, and A represent SRFD, PCAFFusion, and AFGCAttention, respectively.)

of DehazeFormer and LARNet, both of which feature a relatively lightweight architecture.

The results of the joint inference experiments are presented in Table 8. As shown, with a data collection range of 5–25m, the LARNet-SAP-YOLOv11 achieved the best detection performance across the multi-weather test set, reaching a mAP@50 of 88.6%. Compared to the baseline YOLOv11n without any restoration module, this represents a 12.1% improvement, significantly reducing missed and false detections and demonstrating strong adaptability to various adverse weather conditions.

Although the restoration step in joint inference increases the overall inference time due to the added pre-processing, the delay remains within an acceptable range for practical engineering applications. Overall, this joint strategy effectively balances inference speed with detection accuracy, offering considerable deployment value and practical utility.

To clearly demonstrate the impact of each component on the overall performance of the joint model LARNet-SAP-YOLOv11, which exhibited the best comprehensive performance in Table 8, we conducted further ablation experiments on each component of LARNet-SAP-YOLOv11 using



(a) LARNet-YOLOv11 (b) LARNet-YOLOv11+SRFD (c) LARNet-YOLOv11+PCAF Fusion (d) LARNet-YOLOv11+AFGC Attention (e) LARNet-YOLOv11+SP (f) LARNet-YOLOv11+SA (g) LARNet-YOLOv11+AP (h) LARNet-YOLOv11+SAP

FIGURE 12. Impact of Individual components on final performance with the LARNet restoration submodel. (Red dashed circles and green dashed rectangles denote missed and false detections, respectively. In (e), (f), (g) and (h) S, P, and A represent SRFD, PCAFFusion, and AFGCAttention, respectively.)

the transmission line dataset under adverse weather conditions. The results are presented in Table 9. From experiments Exp. No. 1–8 in Table 9, it is evident that, without considering the LARNet image restoration submodel, adding the SRFD, AFGCAttention, and PCAFFusion modules improved various detection metrics. However, due to the absence of the image restoration submodule, the overall detection accuracy remained low, with relatively higher instances of missed and false detections. In experiments Exp. No. 9–12, we evaluated the impact of the LARNet image restoration submodel on final detection performance and conducted another set of ablation experiments on the SRFD, AFGCAttention, and PCAFFusion modules, fully considering the influence of each component in LARNet-SAP-YOLOv11 on overall performance. The results show that incorporating the LARNet image restoration submodel significantly enhanced the ability of the SRFD, AFGCAttention, and PCAFFusion modules to extract detailed features, substantially reducing false detections. Furthermore, in experiments Exp. No. 13–15, all metrics surpassed those achieved with individual modules, indicating that further combining modules continued to improve the model's detailed feature extraction capability. The final joint model, LARNet-SAP-YOLOv11, achieved optimal values for mAP@50 and mAP@50:95, reaching 88.6% and 64.7%, respectively, with a Precision of 89.6% and a near-optimal Recall of 84.0%. Compared to the YOLOv11n baseline model, these metrics improved by 12.1%, 11.9%, 6.9%, and 12.3%, respectively, demonstrating significant enhancements and confirming the impact of each component on detection performance.

In Figures 11 and 12, we selected one representative image for each of haze, rainy, and snowy weather conditions to demonstrate the impact of individual components on overall performance with and without the influence of the LARNet restoration submodel. As shown in Figure 11, without integrating the LARNet submodel, the overall detection performance is relatively low. In Figure 11(a), the baseline YOLOv11 model exhibits numerous missed and false detections under hazy and rainy conditions. In Figures 11(b)–(h), gradually adding improved modules effectively mitigates missed and false detections, though these issues persist. In Figure 12, it is evident that incorporating the LARNet restoration submodel significantly enhances the overall detection performance of individual components under hazy, rainy, and snowy conditions. In Figure 12(h), the final joint model, LARNet-SAP-YOLOv11, achieves high confidence scores for various targets across different weather conditions, validating the effectiveness of each component on overall performance and demonstrating the generalization and robustness of the LARNet-SAP-YOLOv11 model.

4) VISUALIZATION OF DETECTION RESULTS

The visual comparison of detection performance under hazy, rainy, and snowy conditions using LARNet-SAP-YOLOv11 is illustrated in Figs. 8, 9, and 10, respectively. These figures compare the detection results on degraded images before and after applying various restoration methods.

As observed in Figs. 8(a), 9(a), and 10(a), the original degraded images suffer from issues such as blurred object edges and missing details due to haze, rain streaks, or snow particles. These degradations lead to significantly fewer detection boxes and generally low confidence scores, mostly below 0.7. Numerous missed and false detections are evident—especially in hazy conditions where small objects are almost completely undetectable—confirming the adverse impact of weather on detection accuracy.

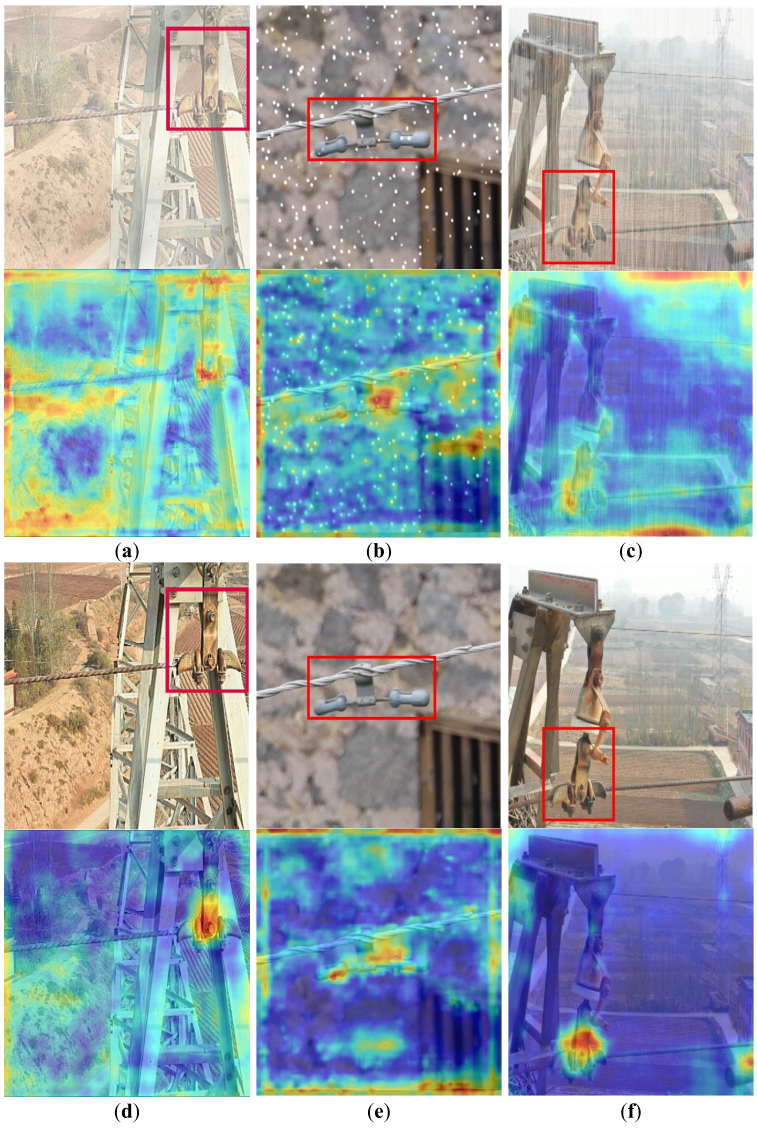


FIGURE 13. Comparison of feature focus before and after restoration under hazy, rain, and snow conditions. Red boxes indicate the target feature regions. Subfigures (a), (b), and (c) show the visible images and corresponding heatmaps under hazy, rainy, and snowy conditions, respectively, while (d), (e), and (f) show the restored visible images and heatmaps under the same weather conditions.

After image restoration, as seen in Figs. 7(b–g), 8(b–g), and 9(b–g), the target boundaries become clearer, texture details are recovered, and the number of missed and false detections is drastically reduced. The results are more complete and accurate, with some confidence scores exceeding 0.8. These improvements demonstrate that image restoration effectively enables the detection model to perform at its full potential. The LARNet-SAP-YOLOv11 model shows excellent performance across all three weather conditions, further validating its robustness and efficiency in complex environments.

These visual results confirm that the image restoration module significantly mitigates the negative impact of weather-induced degradation on feature extraction, enhancing the stability and reliability of the detection system in real-world inspection scenarios. For UAV-based transmission line inspections, this approach offers substantial improvements in detection reliability and safety. Additionally, the restoration-enhanced detection capability demonstrates broad adaptability to other complex environments, providing a promising direction for future system optimization.

5) VISUALIZATION OF FEATURE ATTENTION

To further investigate the effect of image restoration on the feature extraction process in object detection, we conducted a visual analysis of the attention maps generated by the LARNet-SAP-YOLOv11, as shown in Fig. 13.

As seen in the figure, the attention maps generated from original degraded images exhibit scattered and unfocused patterns, with significant deviations from the actual target locations. This is especially noticeable under hazy and snowy conditions, where the attention regions deviate from the corroded fitting targets—indicating that weather-related interference hinders effective feature learning.

In contrast, after image restoration, the object edges and texture features become significantly clearer. The high-response regions in the heatmaps concentrate around the actual targets, reflecting improved attention focus and enhanced discriminative ability of the detection model. These findings affirm the positive role of image restoration in improving detection performance and strengthening feature representation.

In conclusion, the joint inference model not only improves detection accuracy but also refines the model's focus on critical features. This demonstrates significant advantages in terms of efficiency and reliability for detection tasks in complex and adverse environments.

V. CONCLUSION

To address the problem of image degradation during transmission line inspection under adverse weather conditions (haze, rain, and snow), we propose a lightweight and unified joint inference model: LARNet-SAP-YOLOv11. This model integrates an enhanced image restoration network, LARNet, and a high-accuracy object detection model, SAP-YOLOv11, for efficient and robust detection of corroded fittings. Specifically, the LARNet model is based on the lightweight dehazing network DehazeFormer, and achieves excellent restoration performance across various weather scenarios. On the multi-weather test dataset, the model achieves an average PSNR of 30.43 dB and SSIM of 0.951, indicating strong robustness and generalization ability.

SAP-YOLOv11 is built upon the YOLOv11n model and incorporates three critical modules—SRFD, AFGCAttention, and PCAFFusion—which significantly improve the model's accuracy in detecting corrosion defects. On the corrosion fitting test set, SAP-YOLOv11 achieves top results with mAP@50 of 89.6%, mAP@50:95 of 65.9%, and precision of 93.0%, outperforming existing methods.

The proposed end-to-end joint inference model—LARNet-SAP-YOLOv11—effectively performs simultaneous restoration and detection under adverse weather conditions. It achieves a mAP@50 of 88.6% across multi-weather test scenarios, showcasing strong adaptability and superior detection capability.

Despite its effectiveness, the proposed LARNet-SAP-YOLOv11 model has certain limitations. While the model supports end-to-end inference, the training phase requires separate training of the image restoration (LARNet) and object detection (SAP-YOLOv11) submodels, which increases the initial workload. This separate training process involves additional efforts in hyperparameter tuning and data alignment for the restoration and detection datasets, potentially complicating model optimization. Addressing this limitation could enhance training efficiency and model scalability.

Future work will focus on addressing the limitations of the LARNet-SAP-YOLOv11 model and further enhancing its performance. To streamline the training process, we plan to explore balanced loss function designs to optimize the integration of restoration and detection tasks, as well as efficient data input strategies for adverse weather and detection datasets. Additionally, we aim to leverage restoration model weights as pre-trained weights for the detection phase to achieve fully end-to-end training, validation, and testing. Beyond training optimizations, we will address challenges posed by continuous rainfall and snowfall regions, where accumulated water and snow on transmission lines can impact defect detection. Accumulated water often alters the refractive index of light, imposing stricter requirements on data capture angles, while snow tends to obscure larger areas of feature information, demanding higher restoration capabilities from the image restoration model. These efforts will further enhance the intelligence and reliability of transmission line inspection systems.

REFERENCES

[1] B. Li, W. Li, L. Zhang, H. Yang, Z. Chen, T. Lei, and X. Wang, "Ranging algorithm of UAV inspection of transmission line based on monocular vision and SURF method," *J. Phys., Conf. Ser.*, vol. 2452, no. 1, Mar. 2023, Art. no. 012028, doi: 10.1088/1742-6596/2452/1/012028.

[2] H. Liu, Y. Sun, J. Cao, S. Chen, N. Pan, Y. Dai, and D. Pan, "Study on UAV parallel planning system for transmission line project acceptance under the background of industry 5.0," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5537–5546, Aug. 2022, doi: 10.1109/TII.2022.3142723.

[3] Y. Qiu, Y. Lu, Y. Wang, and H. Jiang, "IDOD-YOLOV7: Image-dehazing YOLOV7 for object detection in low-light foggy traffic environments," *Sensors*, vol. 23, no. 3, p. 1347, Jan. 2023, doi: 10.3390/s23031347.

[4] X. Wang, X. Chen, F. Wang, C. Xu, and Y. Tang, "Image recovery and object detection integrated algorithms for robots in harsh battlefield environments," in *Proc. Intell. Robot. Appl.*, Hangzhou, China, Oct. 2023, pp. 575–585, doi: 10.1007/978-981-99-6492-5_49.

[5] X. Jia, C. Ji, F. Zhang, J. Liu, M. Gao, and X. Huang, "AdaptoMixNet: Detection of foreign objects on power transmission lines under severe weather conditions," *J. Real-Time Image Process.*, vol. 21, no. 5, Sep. 2024, Art. no. 1347, doi: 10.1007/s11554-024-01546-1.

[6] Z. Song, X. Huang, C. Ji, and Y. Zhang, "Double-attention YOLO: Vision transformer model based on image processing technology in complex environment of transmission line connection fittings and rust detection," *Machines*, vol. 10, no. 11, p. 1002, Nov. 2022, doi: 10.3390/machines10111002.

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

[8] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3522–3533, Nov. 2015, doi: 10.1109/TIP.2015.2446191.

[9] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011, doi: 10.1109/TPAMI.2010.168.

[10] L.-W. Kang, C.-W. Lin, and Y.-H. Fu, "Automatic single-image-based rain streaks removal via image decomposition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1742–1755, Apr. 2012, doi: 10.1109/TIP.2011.2179057.

[11] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown, "Rain streak removal using layer priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2736–2744, doi: 10.1109/CVPR.2016.299.

[12] Y. Wang, S. Liu, C. Chen, and B. Zeng, "A hierarchical approach for rain or snow removing in a single color image," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3936–3950, Aug. 2017, doi: 10.1109/TIP.2017.2708502.

[13] J. Xu, W. Zhao, P. Liu, and X. Tang, "An improved guidance image based method to remove rain and snow in a single image," *Comput. Inf. Sci.*, vol. 5, no. 3, Apr. 2012, Art. no. 3, doi: 10.5539/cis.v5n3p49.

[14] Z. Chen, Z. He, and Z.-M. Lu, "DEA-net: Single image dehazing based on detail-enhanced convolution and content-guided attention," 2023, *arXiv:2301.04805*.

[15] Y. Song, Z. He, H. Qian, and X. Du, "Vision transformers for single image dehazing," *IEEE Trans. Image Process.*, vol. 32, pp. 1927–1941, 2023, doi: 10.1109/TIP.2023.3256763.

[16] R. Thatikonda, R. Cheruku, and P. Kodali, "DAMNet: Lightweight dual attention mixed network for efficient image deraining," *Neural Comput. Appl.*, vol. 37, no. 17, pp. 10557–10575, Jun. 2025, doi: 10.1007/s00521-024-10569-5.

[17] W. Tao, X. Yan, Y. Wang, and M. Wei, "DWTN: Deep wavelet transform network for lightweight single image deraining," *Adv. Continuous Discrete Models*, vol. 2024, no. 1, Sep. 2024, Art. no. 42, doi: 10.1186/s13662-024-03843-2.

[18] W. Zhou and L. Ye, "UC-former: A multi-scale image deraining network using enhanced transformer," *Comput. Vis. Image Understand.*, vol. 248, Nov. 2024, Art. no. 104097, doi: 10.1016/j.cviu.2024.104097.

[19] Y.-F. Liu, D.-W. Jaw, S.-C. Huang, and J.-N. Hwang, "DesnowNet: Context-aware deep network for snow removal," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 3064–3073, Jun. 2018, doi: 10.1109/TIP.2018.2806202.

[20] L. Shan, H. Zhang, and B. Cheng, "SGNet: Efficient snow removal deep network with a global windowing transformer," *Mathematics*, vol. 12, no. 10, p. 1424, May 2024, doi: 10.3390/math12101424.

[21] X. Dai, Y. Zhou, X. Qiu, H. Tang, T. Tan, Q. Zhang, and T. Tong, "WaveFrSnow: Comprehensive perception wavelet transform frequency separation transformer for image snow removal," *Digit. Signal Process.*, vol. 155, Dec. 2024, Art. no. 104715, doi: 10.1016/j.dsp.2024.104715.

[22] Y. Zhu, T. Wang, X. Fu, X. Yang, X. Guo, J. Dai, Y. Qiao, and X. Hu, "Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 21747–21758, doi: 10.1109/CVPR52729.2023.02083.

[23] W.-T. Chen, Z.-K. Huang, C.-C. Tsai, H.-H. Yang, J.-J. Ding, and S.-Y. Kuo, "Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 17632–17641, doi: 10.1109/CVPR52688.2022.01713.

[24] R. Zhu, Z. Tu, J. Liu, A. C. Bovik, and Y. Fan, "MWFormer: Multi-weather image restoration using degradation-aware transformers," *IEEE Trans. Image Process.*, vol. 33, pp. 6790–6805, 2024, doi: 10.1109/TIP.2024.3501855.

[25] P. W. Patil, S. Gupta, S. Rana, S. Venkatesh, and S. Murala, "Multi-weather image restoration via domain translation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Paris, France, Oct. 2023, pp. 21639–21648, doi: 10.1109/ICCV51070.2023.01983.

[26] B. Cheng, J. Li, J. Shi, Y. Fang, G. Zhang, Y. Chen, T. Zeng, and Z. Li, "WeaFU: Weather-informed image blind restoration via multi-weather distribution diffusion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 12, pp. 13530–13542, Dec. 2024, doi: 10.1109/TCSVT.2024.3450971.

[27] J. M. Jose Valanarasu, R. Yasarla, and V. M. Patel, "TransWeather: Transformer-based restoration of images degraded by adverse weather conditions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 2343–2353, doi: 10.1109/CVPR52688.2022.00239.

[28] V. Potlapalli, S. Waqas Zamir, S. Khan, and F. Shahbaz Khan, "PromptIR: Prompting for all-in-one blind image restoration," 2023, *arXiv:2306.13090*.

[29] H.-H. Wang, F.-J. Tsai, Y.-Y. Lin, and C.-W. Lin, "TANet: Triplet attention network for all-in-one adverse weather image restoration," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, Dec. 2024, pp. 3–19, doi: 10.1007/978-981-96-0911-6_1.

[30] A. Kulkarni, P. W. Patil, S. Murala, and S. Gupta, "Unified multi-weather visibility restoration," *IEEE Trans. Multimedia*, vol. 25, pp. 7686–7698, 2023, doi: 10.1109/TMM.2022.3225712.

[31] C. Li, F. Sun, H. Zhou, Y. Xie, Z. Li, and L. Zhu, "Multi-weather restoration: An efficient prompt-guided convolution architecture," *IEEE Trans. Circuits Syst. for Video Technol.*, vol. 35, no. 2, pp. 1436–1450, Feb. 2025, doi: 10.1109/TCSVT.2024.3469190.

[32] J. Xu, M. Wu, X. Hu, C.-W. Fu, Q. Dou, and P.-A. Heng, "Towards real-world adverse weather image restoration: Enhancing clearness and semantics with vision-language models," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2024, pp. 147–164, doi: 10.1007/978-3-031-72649-1_9.

[33] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587, doi: 10.1109/CVPR.2014.81.

[34] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1440–1448, doi: 10.1109/ICCV.2015.169.

[35] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[36] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," 2017, *arXiv:1703.06870*.

[37] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2016, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.

[38] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[39] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[40] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[41] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, Y. Li, B. Zhang, Y. Liang, L. Zhou, X. Xu, X. Chu, X. Wei, and X. Wei, "YOLOv6: A single-stage object detection framework for industrial applications," 2022, *arXiv:2209.02976*.

[42] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Vancouver, BC, Canada, Jun. 2023, pp. 7464–7475, doi: 10.1109/CVPR52729.2023.00721.

[43] C.-Y. Wang, I.-H. Yeh, and H.-Y. Mark Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," 2024, *arXiv:2402.13616*.

[44] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding, "YOLOv10: Real-time end-to-end object detection," 2024, *arXiv:2405.14458*.

[45] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "DETRs beat YOLOs on real-time object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2024, pp. 16965–16974, doi: 10.1109/CVPR52733.2024.01605.

[46] Z. Guo, Y. Tian, and W. Mao, "A robust faster R-CNN model with feature enhancement for rust detection of transmission line fitting," *Sensors*, vol. 22, no. 20, p. 7961, Oct. 2022, doi: 10.3390/s22207961.

[47] G. Tan, Y. Ye, J. Chu, Q. Liu, L. Xu, B. Wen, and L. Li, "Real-time detection method of intelligent classification and defect of transmission line insulator based on LightWeight-YOLOv8n network," *J. Real-Time Image Process.*, vol. 22, no. 2, Jan. 2025, Art. no. 53, doi: 10.1007/s11554-025-01627-9.

[48] Z. Yu, Y. Lei, F. Shen, S. Zhou, and Y. Yuan, "Research on identification and detection of transmission line insulator defects based on a lightweight YOLOv5 network," *Remote Sens.*, vol. 15, no. 18, p. 4552, Sep. 2023, doi: 10.3390/rs15184552.

[49] M. Wang, L. Sun, J. Jiang, J. Yang, and X. Zhang, "Enhancing hyperspectral power transmission line defect and hazard identification with an improved YOLO-based model," *Optoelectronics Lett.*, vol. 20, no. 11, pp. 681–688, Nov. 2024, doi: 10.1007/s11801-024-3213-3.

[50] W. Lu, S.-B. Chen, J. Tang, C. H. Q. Ding, and B. Luo, "A robust feature downsampling module for remote-sensing visual tasks," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4404312, doi: 10.1109/TGRS.2023.3282048.

[51] H. Sun, Y. Wen, H. Feng, Y. Zheng, Q. Mei, D. Ren, and M. Yu, "Unsupervised bidirectional contrastive reconstruction and adaptive fine-grained channel attention networks for image dehazing," *Neural Netw.*, vol. 176, Aug. 2024, Art. no. 106314, doi: 10.1016/j.neunet.2024.106314.
[52] S. Hu, F. Gao, X. Zhou, J. Dong, and Q. Du, "Hybrid convolutional and attention network for hyperspectral image denoising," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024, doi: 10.1109/LGRS.2024.3370299.
[53] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single image dehazing and beyond," 2017, *arXiv:1712.04143*.
[54] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1715–1723, doi: 10.1109/CVPR.2017.186.
[55] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," 2019, *arXiv:1911.08287*.
[56] Y. Cui and A. Knoll, "Exploring the potential of channel interactions for image restoration," *Knowl.-Based Syst.*, vol. 282, Dec. 2023, Art. no. 111156, doi: 10.1016/j.knosys.2023.111156.

YUXIANG GONG received the B.S. degree in electrical engineering from Changsha University, Changsha, China, in 2022. He is currently pursuing the M.S. degree with Hubei Minzu University, Enshi, China. His research interests include deep learning, image processing, and power-related object detection technology.

LIN GAO (Member, IEEE) received the B.S. degree in material forming and control engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2002, the M.S. degree in control engineering from Wuhan University of Technology, Wuhan, in 2007, and the Ph.D. degree in engineering from Sichuan University, Chengdu, China, in 2020. He is currently an Associate Professor with Hubei Minzu University, Enshi, China. His research interests include digital image processing, deep learning, and embedded systems.

HAO ZHANG received the B.S. degree in electrical engineering from Jinling Institute of Technology, Nanjing, China, in 2023. He is currently pursuing the M.S. degree with Hubei Minzu University, Enshi, China. His research interests include deep learning, computer vision, and power-related object detection.

HUAGUO LIU received the B.S. degree in mechanical engineering from West China University, Chengdu, China, in 2006, and the M.S. degree in mechanical engineering and the Ph.D. degree in engineering from Sichuan University, Chengdu, in 2009 and 2020, respectively. His research interests include e-government and safety evaluation of large equipment.

YONGDAN ZHU received the B.S. degree from Hubei Minzu University, Enshi, China, in 2004, the M.S. degree in engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2007, and the Ph.D. degree in engineering from Wuhan University, China, in 2013. His research interests include electrical applications, electronic information, optoelectronic information, and applied physics, with interdisciplinary research experience in these fields.

YU YANG received the B.S. degree in ship electronic and electrical engineering from Chongqing Jiaotong University, Chongqing, China, in 2022. He is currently pursuing the M.S. degree with Hubei Minzu University, Enshi, China. His research interests include deep learning, computer vision technology, and power target detection.