electronics

MDPI

Article

Insulator Object Detection Method for Transmission Lines Based on an Improved Image Enhancement Algorithm

Zhe Zheng, Wenpeng Cui, Mingxuan Li *, Ming Li, Yu Liu, Qingchen Yang, Yuzhe Chen and Hao Men

Beijing Smart-Chip Microelectronics Technology Company Ltd., Beijing 102299, China; zhengzhe@sgchip.sgcc.com.cn (Z.Z.); cuiwenpeng@sgchip.sgcc.com.cn (W.C.); liming@sgchip.sgcc.com.cn (M.L.); liuyu@sgchip.sgcc.com.cn (Y.L.); yangqingchen@sgchip.sgcc.com.cn (Q.Y.); chenyuzhe@sgchip.sgcc.com.cn (Y.C.); menhao@sgchip.sgcc.com.cn (H.M.)

* Correspondence: limingxuan@sgchip.sgcc.com.cn; Tel.: +86-17778884389

Abstract

This paper addresses the issues of blurred details, low contrast, and feature degradation in insulator images under harsh meteorological conditions, as well as the challenges of high computational complexity and insufficient real-time performance when deploying existing deep learning models on edge devices. It proposes a lightweight insulator defect detection method that integrates an improved image enhancement algorithm. The method introduces Mahalanobis distance-based modulation weight optimization for scene depth estimation and improves the color decay prior model to effectively enhance foggy insulator images. It further designs a lightweight detection network integrating region-aware routing attention mechanisms, utilizing multi-scale feature fusion strategies to achieve precise insulator identification and localization. Experimental results demonstrate that the proposed method significantly enhances inference speed while maintaining detection accuracy, effectively adapting to edge computing devices. This provides a viable technical solution for real-time deployment in intelligent transmission line inspection systems.

Keywords: new electricity system; transmission lines; fault diagnosis; edge lightweight deployment

1. Introduction

As a vital infrastructure of modern industrial society, the power industry serves as a critical safeguard for the nation's economic lifeline and the stable functioning of society. Within this energy system, transmission lines play an indispensable role. High-voltage and ultra-high-voltage transmission lines, functioning as primary conduits for long-distance, high-capacity power transmission, form the backbone network of the national energy internet. To date, China's total transmission line mileage has reached approximately 2.2 million kilometers, ranking first globally in scale. This intricate grid network serves as the core vehicle for optimizing energy resource allocation and ensuring regional power balance, providing robust energy security for the nation's economic and social development.

China's transmission lines are predominantly located in harsh environments such as deserts, Gobi regions, and deep-sea areas. They are vulnerable to threats like lightning strikes, ice accumulation, wildfires, and strong winds, which can cause significant economic losses and disrupt the stable operation of the power grid.

As power grid scale continues to expand, traditional manual inspection methods have become insufficient to fully meet the safety maintenance requirements of transmission lines.

Check for updates

Academic Editor: Ahmed Abu-Siada

Received: 29 January 2026
Revised: 13 February 2026
Accepted: 16 February 2026
Published: 24 March 2026

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

With the widespread application of visual monitoring technology, the workload associated with manual inspections has been significantly reduced, making it an essential tool for transmission line condition monitoring. Numerous scholars have conducted in-depth research on visual monitoring technology. Chen Zhen et al. [1] proposed an edge-aware residual network based on node and edge attribute graphs, achieving high-precision prediction of potential overloaded lines and providing an effective solution for active defense after cascading faults. Behrooz Taheri et al. [2] addressed the challenges of high computational complexity, noise sensitivity, and the trade-off between speed and accuracy in DC grid fault detection by integrating fuzzy logic with support vector regression (SVR). This hybrid approach constructs composite feature signals using fuzzy logic and trains SVR models, enabling fast, robust, and highly accurate fault identification even at low sampling rates. K. Poongothai et al. [3] addressed the insufficient accuracy of transmission line fault detection in IoT-enabled smart grids caused by sensor data noise and inadequate model performance. They proposed a comprehensive fault diagnosis framework integrating the Relational Binary Aggregated Graph Convolutional Network (RBAGCN) with the Red Panda Optimization (RPO) algorithm. Through robust preprocessing and discriminative feature extraction, this framework achieves high-precision, low-error classification for multiple fault types, significantly enhancing the reliability of smart grid fault identification. Shuai et al. [4] addressed the challenges of multi-scale fault feature extraction and low detection accuracy in complex environments by proposing an improved DM-YOLO detection model. This model integrates dynamic multi-scale convolutions, multi-dimensional perceptual attention, and multi-head feature fusion modules, effectively enhancing the expression and fusion capabilities of multi-scale fault features to achieve precise and robust detection of various fault targets in complex scenarios. Zhang Ximing et al. [5] addressed the suboptimal performance of existing deep learning approaches in detecting transmission line faults under complex conditions with low image resolution and small target sizes. They proposed the HSPAN-GNN model, which combines graph convolutions with convolutional operations. This model achieves high-precision detection of transmission line targets while maintaining computational efficiency through high-priority subsampling, spatio-temporal convolutions, and normalized attention mechanisms. Wang Dong et al. [6] addressed the failure of traditional traveling wave fault location methods in hybrid cable-overhead transmission lines due to inconsistent wave propagation speeds. They proposed a novel fault location method based on the frequency-dependent characteristics of traveling waves. By establishing an accurate frequency-dependent model, this method achieves high-precision fault location in hybrid lines and demonstrates robustness under various fault conditions. Pan Duan et al. [7] addressed the challenges of fault line identification and low detection accuracy in complex distribution networks by proposing a fault location method based on maximum overlap discrete wavelet transform and an improved Teager energy operator. By analyzing fault-induced components and precisely detecting wavefronts, this method accurately identifies faulted lines and fault types, transforming complex topology fault location into single-line location problems. Ultimately, it controls location errors within 0.5% across various fault scenarios. Ziad M. Ali et al. [8] addressed the insufficient accuracy of traditional methods and the challenge of rapid, precise protection in diagnosing faults on double-circuit transmission lines by proposing a directional protection framework combining wavelet packet transform and deep learning. Their directional protection framework analyzes electrical signals at both ends and employs deep learning models for identification. This enables precise fault section localization and recognition, providing primary and backup protection for different line segments with an average localization error controlled within 0.03%. Yang Minxin et al. [9] addressed the limitations of traditional methods in fault location for high-voltage direct

current transmission lines—which are susceptible to factors like fault type and transient resistance—by proposing a time-domain fault location method based on the variational iteration method. By solving partial differential equations for the line and comparing voltage waveform similarity, this method achieves precise location unaffected by fault conditions, offering high accuracy and stability. Shu Hongchun et al. [10] addressed the accuracy limitations of conventional three-terminal line fault location methods (which rely on pre-known line parameters and fault type identification). They proposed a fault location algorithm for three-terminal lines based on positive-sequence components. This method directly calculates line parameters and fault locations using synchronized terminal measurement data, eliminating the need for pre-acquired line parameters and fault classification. It achieves high-precision location unaffected by fault conditions and accurately estimates line characteristic impedance and propagation constants. Eloi Gravot et al. [11] explored the limitations of traditional transmission line fault classification methods—which rely on parameterized models and large amounts of labeled data, resulting in poor generalization capabilities. They proposed a hybrid classification approach combining topological data analysis with unsupervised machine learning. This method extracts topological features from signals via persistent homotopy and employs k-means clustering for fault classification. It maintains high accuracy even with field measurement data and demonstrates exceptional generalization capability without requiring parameter tuning or pre-training. Ahmed El-Nozhahy et al. [12] addressed the high error rates and susceptibility to fault resistance and initiation angle in traditional high-voltage transmission line fault location methods due to complex fault modeling. They proposed an intelligent fault diagnosis method integrating a Fuzzy Interface System (FIS) with an Adaptive Neuro-Fuzzy Inference System (ANFIS). This approach employs FIS for fault detection and classification, combined with ANFIS for precise localization, enabling accurate identification of various fault types. Its error metrics validate its high efficiency. Živko Sokolović et al. [13] addressed the inefficiency and insufficient accuracy of traditional transmission line fault handling methods by proposing an intelligent diagnostic method based on the fusion of artificial neural networks and principal component analysis. By constructing training datasets through power system simulations, this approach achieves precise detection, classification, and localization of various faults, significantly enhancing the reliability and stability of grid operations.

Although current visual monitoring technology for transmission lines far surpasses traditional manual inspections in efficiency, its practical application faces two core bottlenecks: "low image quality during data collection" and "limited computational resources." At the data collection level, hazy weather causes severe image detail loss, low contrast, and texture feature attenuation in insulator images, resulting in critical defect characteristics being obscured by environmental noise. Regarding algorithm deployment, while existing object detection models perform excellently in ideal laboratory settings, their complex network structures and massive computational demands rely on high-performance AI accelerator boards. This makes real-time computational inference impossible on edge devices constrained by limited power and computing resources. Therefore, there is an urgent need to develop lightweight detection methods suitable for edge device deployment to achieve real-time, precise identification of insulator defects in harsh environments.

To address the issues of poor image quality and low defect detection accuracy in transmission line images under adverse weather conditions, this paper proposes an insulator target detection method for transmission lines based on an improved image enhancement algorithm. This approach introduces scene depth estimation modulation weights based on Mahalanobis Distance [14–17], optimizing the color decay prior theory to achieve efficient image defogging [18–20]. The defogged images are fed into a lightweight detection network for model training, ultimately generating a power equipment defect detection model.

Specifically, by incorporating a regional routing attention mechanism [21–23], the method enhances multi-level feature representation capabilities, generating multiple enhanced fusion feature maps at different resolutions. The detector head outputs a prediction tensor to generate predicted bounding boxes. By comparing these with ground-truth bounding boxes and calculating loss functions, model parameters are optimized. This approach significantly improves the accuracy and speed of power equipment defect detection in hazy environments while supporting lightweight deployment on edge devices, ultimately enabling real-time intelligent detection of power equipment.

2. Theoretical Methods for Target Detection of Transmission Line Insulators

2.1. Improving Color Attenuation Prior

2.1.1. Research on Improving Color Fading Prior Theory

Traditional color decay priors are image defogging models based on statistical observations, whose core concept lies in establishing quantitative relationships between scene depth regions and color attributes in hazy images. This theory is founded on key physical observations: in fog-affected images, scene depth exhibits significant correlations with pixel brightness and saturation differences. Under clear weather conditions, the chroma and saturation channels of an image typically exhibit similarity and high correlation. However, when affected by atmospheric scattering media such as haze, brightness increases with scene depth due to the injection of atmospheric light, while saturation decreases as the intrinsic colors of objects are diluted by white light.

Based on the above observations, the Color Attenuation Prior theory posits that the scene depth at a given pixel can be approximated by modeling the linear difference between its brightness channel value and saturation channel value. This mathematical model is typically formulated as follows:

$$d(x,y) = \theta_0 + \theta_1 i(x,y) + \theta_2 s(x,y) \tag{1}$$

In the model, $d(x,y)$ represents the scene depth; $i(x,y)$ denotes the brightness value at pixel (x,y) in the hazy image; $s(x,y)$ signifies the saturation value at pixel (x,y) in the hazy image; θ_0, θ_1, and θ_2 are model coefficients. These parameters are determined via the least squares method based on the saturation, brightness, and known scene depth of insulator images.

However, the performance of traditional color decay prior models heavily relies on the fundamental assumption that the difference between brightness and saturation accurately reflects scene depth. This assumption may fail in specific scenes where white objects appear in sky regions, leading to inaccurate depth estimation and subsequently causing issues such as halo artifacts, color distortion, or insufficient defogging effects. To address these limitations, this paper introduces a Mahalanobis distance-based weight modulation method to enhance the traditional theory, significantly improving its robustness and accuracy in complex environments. Equation (1) is modified as follows:

$$d(x,y) = \theta_0 + \widetilde{\theta}_1 i(x,y) + \widetilde{\theta}_2 s(x,y) = \theta_0 + \theta_1 \cdot w(x) i(x,y) + \theta_2 \cdot (2 - w(x)) s(x,y) \tag{2}$$

Among them, $w(x)$ is the modulation weight for scene depth estimation based on Markov distance, and the calculation steps are as follows:

First, the image is converted from the RGB color space to the YCbCr color space according to the ITU-R BT.601 standard, obtaining the red chrominance component and blue chrominance component for each pixel.

$$C_b(x) = -0.1687 * R - 0.3313 * G + 0.5 * B + 128 \tag{3}$$

$$C_r(x) = 0.5 * R - 0.4187 * G - 0.0813 * B + 128 \tag{4}$$

Here, x denotes the pixel location; R, G, and B represent the intensity of red, green, and blue light components in the pixel, respectively; $C_b(x)$ signifies the blue chrominance component, indicating the degree to which the pixel color deviates from blue, while $C_r(x)$ denotes the red chrominance component, representing the degree to which the pixel color deviates from red.

Subsequently, the Mahalanobis distance is employed to measure the similarity between each pixel's chromaticity and the typical chromaticity of insulators. For each pixel x in the image, its chromaticity value is treated as a two-dimensional vector:

$$\vec{C}(x) = [C_b(x), C_r(x)]^T \tag{5}$$

The calculation formula of Mahalanobis distance is:

$$D_M(x) = \sqrt{\left(\vec{C}(x) - \vec{\mu}_{ins}\right)^T \cdot \sum_{ins}^{-1} \left(\vec{C}(x) - \vec{\mu}_{ins}\right)} \tag{6}$$

In this formulation, $D_M(x)$ denotes the Mahalanobis distance for pixel x. A larger value indicates a greater deviation of the pixel's color from the typical color of insulators, while a smaller value signifies higher similarity. $\vec{\mu}_{ins}$ represents the mean vector of the typical chrominance of insulators, which is derived from a large set of annotated insulator pixel samples and characterizes the expected chrominance that insulators should exhibit. The subtraction of the pixel's chrominance vector from this mean vector transforms the color difference into a measurable deviation vector, whose direction and magnitude intuitively represent the nature and extent of the color shift. $\sum_{ins} = \begin{bmatrix} \sigma_{bb} & \sigma_{br} \\ \sigma_{rb} & \sigma_{rr} \end{bmatrix}$ is the covariance matrix of the insulator chrominance, describing the dispersion of the insulator pixels' $C_b(x)$ and $C_r(x)$ values around their respective means and the correlation between these two components. $\sum_{ins}^{-1}$ is the inverse of the covariance matrix, and $\left(\vec{C}(x) - \vec{\mu}_{ins}\right)^T$ denotes the transpose of the deviation vector.

Finally, distances are mapped to weight values, establishing a Mahalanobis distance-based scene depth estimation algorithm to modulate the weights.

$$w(x) = \exp\left(-\frac{D_M(x)^2}{2\sigma^2}\right) \tag{7}$$

In this formulation, $w(x)$ represents the final weight for pixel x. A value closer to 1 indicates a stronger match between the pixel's chrominance and the typical chrominance of insulators, while a value closer to 0 signifies an extreme deviation. $\exp(\cdot)$ denotes the natural exponential function, ensuring the output value falls within the range (0,1]. $D_M(x)^2$ corresponds to the squared Mahalanobis distance, and σ is the bandwidth parameter that controls the rate of weight decay. The bandwidth parameter σ controls the decay rate of the Mahalanobis distance mapping to modulation weights, and its value directly determines the steepness of weight reduction when pixel chromaticity deviates from the typical chromaticity of insulators. To balance suppressing excessive enhancement in background regions with preserving texture details in insulator regions, this paper employs a grid search combined with quantitative validation set evaluation to determine the optimal value of σ. Specifically, σ was traversed over the range [0.1, 2.0] with a step size of 0.1 on the validation set, evaluating performance using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) of the enhanced images. Experiments revealed

that when $\sigma < 0.5$, weights became overly sensitive to chromaticity deviations, excessively suppressing transmittance in residual haze regions; when $\sigma > 1.5$, the weight decay becomes too gradual, failing to effectively suppress chromaticity deviations in background regions and prone to generating halo artifacts. Balancing defogging performance with real-time edge deployment requirements, this paper selects $\sigma = 0.8$ as the final parameter. At this setting, PSNR improves by 1.27 dB and SSIM by 0.034 compared to $\sigma = 1.0$, with no significant increase in computational overhead.

Compared to traditional color decay prior models, the core advantage of this paper's improved method lies in explicitly introducing a typical insulator chromaticity prior. By modulating weights via Mahalanobis distance, it adaptively adjusts the contribution ratio of the brightness-saturation difference in depth estimation. This design directly addresses two critical degradation issues in insulator images under foggy conditions:

(1) Chromaticity confusion between the insulator body and the hazy background. Traditional CAP models rely on linear difference estimates between brightness and saturation to estimate depth. However, insulator surfaces inherently exhibit low saturation characteristics (e.g., the grayish-white/light brown hues of glass or composite insulators). Under foggy conditions, their saturation further diminishes, leading models to misclassify them as "distant/dense fog regions." This results in underestimated transmittance and insufficient defogging. This paper generates weighting factors by calculating the Mahalanobis distance between pixel chromaticity vectors and typical insulator chromaticity distributions. When pixels belong to insulator regions, the saturation term coefficient in Equation (2) is weakened, allowing the brightness term to dominate depth estimation. This corrects the traditional CAP model's tendency to overestimate depth for low-saturation objects, preventing excessive "fogging" of insulator areas.

(2) Edge preservation in complex background depth estimation. Traditional CAPs produce halo artifacts at abrupt depth transitions (e.g., between insulator strings and sky backgrounds) due to their global linear models failing to detect local chromaticity variations. The introduced Mahalanobis distance weights not only reflect pixel-insulator chromaticity similarity but also encode intra-class dispersion via their covariance matrix. When pixels reside in edge regions, their chromaticity vectors deviate from typical distributions, causing weights to rapidly decay (with $\sigma = 0.8$ controlling decay rate). The model automatically reverts to the traditional CAP's brightness-saturation difference mechanism. This preserves original depth estimation accuracy in background areas while suppressing abrupt transmittance changes at edges, significantly reducing halo artifacts in restored images.

Through these targeted improvements, our method has achieved more accurate scene depth estimation, more complete texture detail restoration, and stronger adaptability to complex scenes in the insulator fog effect enhancement task. This verifies the effectiveness of the improvement strategy.

In regions where $w(x)$ approaches 1 (indicating insulator areas), the coefficient $\widetilde{\theta_1}$ is preserved while $\widetilde{\theta_2}$ is reduced. This attenuates the contribution of saturation in depth estimation. Since insulators inherently exhibit low saturation, this adjustment prevents the original CAP model from misclassifying them as background due to their low saturation values. In regions where $w(x)$ approaches 0 (background areas), $\widetilde{\theta_1}$ is suppressed while $\widetilde{\theta_2}$ is amplified, allowing the CAP model to maintain its original operational mechanism, relying primarily on the brightness-saturation difference for background depth estimation.

The fog-free image is obtained based on the atmospheric scattering model.

$$I(x) = J(x) \cdot t(x) + A(1 - t(x)) \tag{8}$$

In this formulation, $I(x)$ represents the observed hazy image; $J(x)$ denotes the recovered haze-free image; A signifies the global atmospheric light value; and $t(x)$ indicates the transmission map, representing the proportion of light that successfully reaches the sensor through the haze.

$$t(x) = 1 - e^{-\beta d(x)} \tag{9}$$

where β denotes the atmospheric scattering coefficient, with higher values indicating denser haze conditions; $d(x)$ represents the scene depth; and $e^{-\beta d(x)}$ is the attenuation factor, quantifying the remaining proportion of light that reaches the camera from scene point x after transmission through the haze.

2.1.2. Independent Performance Comparison of Image Enhancement Modules

To objectively evaluate the competitiveness of the proposed Improved Color Attenuation Prior (CAP) with Mahalanobis weighting in image defogging tasks, this supplementary experiment independently compares it against mainstream non-learning/learning defogging methods, covering representative models specified by reviewers: (a) DehazeNet, (b) AOD-Net, (c) FFA-Net, (d) MSBDN, (e) Zero-DCE/Zero-DCE++ (low-light enhancement methods serving as baseline controls for low-level enhancement), and (f) recent Transformer-based defogging models including DehazeFormer and DeHamer.

Experimental Setup: To mitigate bias from training data distribution differences, this comparison employs two approaches to obtain objective metrics for each method: direct inference using pre-trained models or retraining on public benchmarks (SOTS-indoor, Dense-Haze, NH-Haze). Evaluation metrics include Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM), number of parameters (Params), and edge device inference speed (FPS on Xavier NX). Since our method involves no trainable parameters, its inference speed falls under the "zero-shot" category alongside DCP-based approaches. The calculation of Mahalanobis distance weights requires only lightweight matrix operations, eliminating the overhead of neural network forward passes.

The experimental results are shown in Table 1. The comparison clearly demonstrates the unique advantages of the image enhancement module proposed in this paper—requiring no paired data training, zero parameters, and extremely low computational overhead—endowing it with distinctive lightweight competitiveness in edge deployment scenarios. Compared to the Zero-DCE series, our method is specifically designed for the physical model of haze degradation, achieving significantly superior restoration quality on hazy images than generic low-light enhancement methods. Therefore, the core value of this enhancement module lies not in pursuing absolute leadership in haze removal accuracy, but in establishing a new Pareto-optimal solution among "accuracy-speed-deployment cost." This aligns perfectly with the research direction of "real-time detection on edge devices" that runs throughout this paper.

Table 1. Independent performance comparison of the proposed enhancement module with mainstream dehazing methods.

Method	SOTS-Indoor	Dense-Haze	NH-Haze	Parameters (M)	Xavier FPS
DCP	16.62/0.817	11.01/0.416	10.57/0.519	0	>500
DehazeNet	21.14/0.847	9.48/0.438	16.62/0.524	0.01	18.2
AOD-Net	19.06/0.850	12.82/0.468	15.40/0.569	0.002	35.7
FFA-Net	36.39/0.989	12.22/0.444	18.13/0.647	4.46	1.8
MSBDN	32.77/0.981	15.13/0.555	17.97/0.659	31.35	2.1
Zero-DCE++	—	11.23/0.401	12.87/0.458	0.035	42.3
DehazeFormer-T	35.15/0.989	15.52/0.463	18.73/0.533	0.686	6.4
DeHamer	36.63/0.988	16.62/0.560	20.66/0.684	132.45	0.9
Improve CAP(Ours)	29.71/0.932	14.83/0.529	19.02/0.681	0	>200

2.2. ID-YOLO: A Lightweight Object Detection Method

2.2.1. Lightweight Backbone Network

The YOLOv5 network is architecturally scaled into five independent models: yolov5n, yolov5s, yolov5m, yolov5l, and yolov5x. Its structure comprises four main components: the input module, the feature extraction backbone network, the feature fusion neck network, and the detection head network. This paper proposes the novel Re-YOLO architecture developed based on the YOLOv5s framework as a new backbone network. This lightweight backbone is specifically designed to meet real-time processing demands on edge devices.

The C3 module is a critical component of the YOLOv5 network, whose core function lies in enhancing feature extraction capabilities by increasing network depth and receptive field. This module comprises three convolutional blocks: the first block employs a stride-2 convolution to halve the spatial dimensions of feature maps, while the second and third blocks maintain a stride-1 convolution. All convolutional blocks within the C3 module utilize 3×3 convolution kernels, a design that significantly enhances model stability and generalization capabilities.

To meet the lightweight deployment requirements of edge devices, separable convolutions replace the standard convolutions in the original C3 module. Furthermore, to address the overfitting issues inherent in the original backbone architecture, the Re-YOLO backbone shown in the main network section of Figure 1 is adopted. This significantly boosts inference speed while incurring only minimal accuracy loss. The overall detection model, named ID-YOLO, integrates this lightweight backbone (Re-YOLO) with the dual-layer routing attention mechanism and multi-scale feature fusion neck.

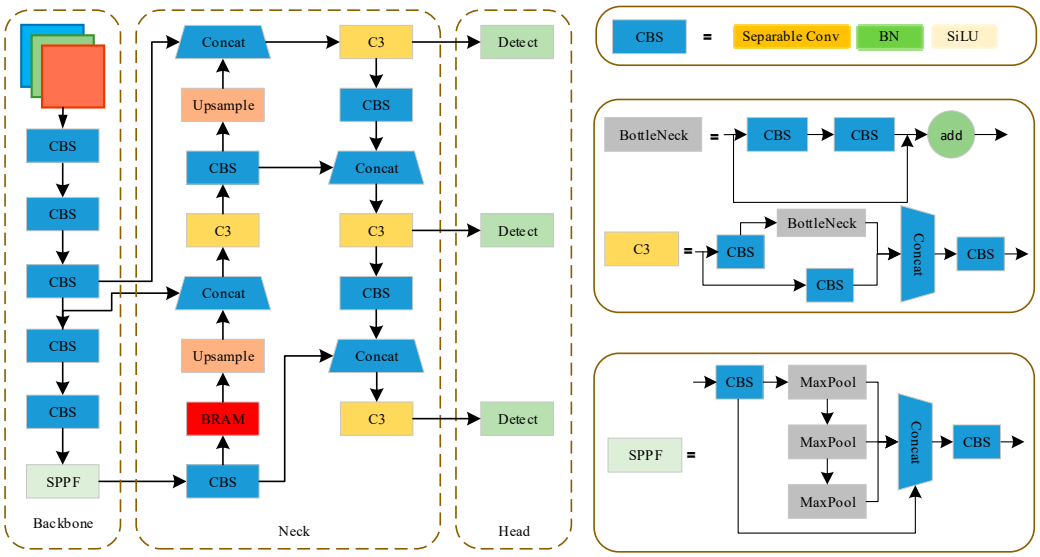


Figure 1. Re-YOLO structure diagram.

2.2.2. Dual-Layer Routing Attention Mechanism

The core function of attention mechanisms lies in selectively extracting and amplifying salient image features required for network learning based on contextual information or pixel-to-pixel correlations, thereby reducing reliance on external information. Its working principle is as follows: using query vectors and key vectors as inputs, each query vector is transformed into a weighted sum of corresponding key vectors, as detailed in Figure 2. These weights are computed through the normalized dot product between the query vector and its corresponding key vector, as expressed by the following formula:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{C}\right)V \tag{10}$$

Among them, the scale factor is used to avoid weight concentration and vanishing gradients.

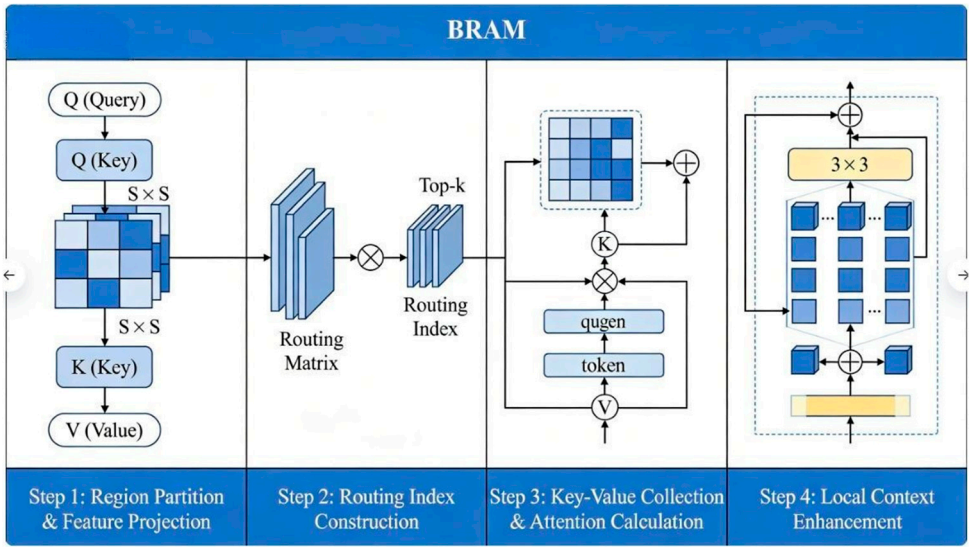


Figure 2. Schematic Diagram of the BRAM Structure.

Secondly, through "self-attention", it is implied that the queries (Q), keys (K), and values (V) are derived as linear projections of the same input. $N = H * W$, where H and W represent the height and width of the feature map, respectively. To enable the model to learn different dependency relationships based on the same attention aggregation method and subsequently combine these dependencies to capture a wider range of dependencies within the sequence, the multi-head attention mechanism is employed. This mechanism splits the output into *h* distinct heads along the channel dimension, with each head utilizing independent projection weights. The expression is as follows:

$$\begin{aligned} MHSA(X) &= Concat(head_0, head_1, \ldots, head_h)W^O \\ head_i &= Attention\left(XW_i^q, XW_i^k, XW_i^v\right) \end{aligned} \tag{11}$$

where $head_i \in R^{N*\frac{C}{h}}$ represents the i-th attention output head; $W_i^q, W_i^k, W_i^v \in R^{C*\frac{C}{h}}$ are the corresponding input projection weight matrices for queries and keys, respectively. An additional linear transformation with the weight matrix $W^O \in R^{C*\frac{C}{h}}$ is employed to combine all output heads.

Although this method can more effectively capture context dependencies and obtain more comprehensive information, the multi-head attention mechanism imposes significant computational demands, requiring higher hardware specifications and thus rendering it unsuitable for edge device applications. To address this challenge, a dual-layer routing attention module is employed. This module filters out the least relevant key-value pairs at the coarse-grained regional level, retaining only a small number of routing regions to eliminate redundant information.

The feature map is first partitioned into S*S non-overlapping regions, where each region contains a set of feature vectors $\frac{HW}{S^2}$, resulting in a new feature representation. The regional partitioning parameter S represents a core trade-off variable in the BRAM, balancing computational complexity against semantic coherence preservation. In BiFormer's original implementation, the value of S determines the granularity of coarse-grained regional filtering: smaller S values (e.g., 4×4) increase the number of tokens per region, reducing computational load for regional routing matrices but diminishing semantic purity within regions, leading to residual invalid key-value pairs; larger S values (e.g., 8×8) enhance regional semantic focus, but the quadratic increase in region count significantly raises

storage and routing index construction costs for the adjacency matrix A^r. Given the wide target scale range in our insulator detection task (from small fragments to entire strings), we jointly optimized edge device inference speed (FPS) and mAP50:95. We evaluated four configurations (S ∈ {4, 6, 8, 10}) on the validation set. Experiments demonstrate that S = 6 achieves the optimal cost-performance ratio: compared to S = 8, mAP50:95 decreases by only 0.31%, while inference speed improves by 13.7% (on Xavier platform); compared to S = 4, mAP50:95 increases by 2.18%, with a manageable increase in parameter count. Therefore, this paper uniformly sets the BRAM region partitioning parameter to S = 6. This value effectively balances high detection accuracy with the limited memory bandwidth and computational resources of edge devices. Subsequently, the query, key, and value tensors are derived through linear projections as follows:

$$Q = X^r W^q, K = X^r W^k, V = X^r W^v \tag{12}$$

Among them, $W^q, W^k, W^v \in R^{C*C}$ respectively represent the projected weights of the query, key, and value.

Secondly, by constructing an orientation map, find the area that each given region should correspond to in the original map. Specifically, first export the region-level query and key, and then apply the average value of each region to Q and K respectively. Then, perform matrix multiplication on the Q^r and the transposed K^r to obtain the adjacency matrix of the association graph between the given region and the original image region:

$$A^r = Q^r (K^r)^T \tag{13}$$

The adjacency matrix is used to measure the degree of semantic association between two regions. The core step next is to trim the association graph by retaining only the top-k connections for each region. A routing index matrix is derived through the line-by-line topk operator:

$$I^r = topkIndex(A^r) \tag{14}$$

Subsequently, the indices of the k most relevant regions for the i-th region in I^r can be obtained. Once the index matrix I^r is established, fine-grained token-to-token attention can be applied. For each query token within the i-th region, it will attend to all key-value pairs in the union of the k routing regions indexed by I^r. Since these routing regions may be distributed across the entire feature map, and GPUs rely on coalesced memory operations, the key and value tensors are first gathered:

$$K^g = gather(K, I^r), V^g = gather(V, I^r) \tag{15}$$

where $K^g, V^g \in R^{S^2 * \frac{kHW}{S^2} * C}$ represent the gathered key and value tensors, respectively. Subsequently, attention is applied to the collected key-value pairs as follows:

$$O = Attention(Q, K^g, V^g) + LCE(V) \tag{16}$$

Here, a Local Context Enhancement condition $LCE(V)$ is introduced. The $LCE(\cdot)$ function is implemented via a parameterized depthwise convolution with the kernel size set to 3, enabling the comparison of each adjacent eight pixels against the window center as a threshold.

The BRAM module is embedded to enhance the model's ability to extract numbers and decimal points as small targets, while improving the model's generalization and robustness.

3. Experimental Procedure and Results

3.1. Experimental Data and Environment

This study utilizes a transmission line insulator inspection dataset provided by a provincial power grid company in China. The dataset comprises 15,000 high-resolution insulator images, with an original resolution of 3840 × 2160 pixels (4K), uniformly downsampled to 640 × 640 pixels during training and inference to balance detail preservation and computational efficiency. The images are divided into a training set (10,500 images), validation set (3000 images), and test set (1500 images) at a ratio of 7:2:1. The data covers clear and foggy weather conditions, where foggy images are real samples captured by transmission line inspection cameras under natural haze conditions, rather than artificially synthesized. It encompasses three types of insulators—glass, composite, and ceramic—and includes annotations for typical defects such as pollution flashover, damage, and fragment detachment. Regarding class balance, pollution flashover accounts for approximately 45% of defect instances, damage 30%, and fragment detachment 25%. Although the dataset exhibits a certain degree of imbalance, it reflects the actual distribution in real-world inspection scenarios. To mitigate the impact of imbalance, we employ a combination of data augmentation and a focal loss variant during training.

To improve model generalization and robustness, a series of data augmentation strategies are applied online during training, including: random horizontal flipping (probability 0.5), random scaling (±20%) and translation (±10%), color jittering (brightness, contrast, saturation, hue adjustments with ±0.2 range), mosaic augmentation (4 images stitched into one, probability 0.5), and MixUp augmentation (probability 0.2). These augmentations effectively enrich the training sample distribution and alleviate overfitting.

The training platform utilizes high-performance AI servers based on the "SCA600" smart computing board, delivering 256 TOPS@INT8 per board with 200GB of SRAM. Each server supports up to 8 boards. The system runs Ubuntu 22.04 OS, employs Python as the programming language, and utilizes PyTorch® 2.4 as the neural network framework. The batch size is set to 32, with a total training cycle of 300 iterations. Stochastic gradient descent (SGD) is employed as the optimizer, utilizing a dynamic learning rate adjustment strategy. The initial learning rate is set to 0.005, the momentum parameter to 0.9, and the number of learning rate decay cycles to 5.

The verification platform utilizes the Jetson Xavier NX embedded development board (8 GB memory variant), which features a 384-core NVIDIA Volta GPU with 48 Tensor Cores, a 6-core NVIDIA Carmel ARM v8.2 CPU, 8 GB 128-bit LPDDR4x memory, and supports multiple power modes. In our experiments, the board is configured in 15W mode (typical for sustained edge deployment), with CPU clock frequency set to 1.4 GHz and GPU clock frequency to 900 MHz. The system runs Ubuntu 20.04 with JetPack 5.1, and inference is performed using TensorRT 8.5 with FP16 precision to maximize throughput. The measured average power consumption during inference is approximately 11.2 W (peak 13.5 W). Compared to traditional PC-based GPUs, embedded GPUs offer advantages in compact size and low power consumption while maintaining programmability and robust computational performance. The strategy adopted in this paper is to first train the model on an AI server platform, then deploy the model to the embedded platform for validation. Model performance is evaluated based on recall rate, detection speed, and accuracy, while model complexity is measured by the number of parameters and the volume of floating-point operations.

First, an initial depth estimation is performed based on statistical background modeling. For the input foggy image, features such as the dark channel, light channel, and local color standard deviation are extracted for each pixel to form a feature vector. Pixel samples are selected from distant regions in the image that exhibit flat colors and high

brightness (such as the sky); the mean and covariance matrix of their feature vectors are calculated, and this is used to establish a statistical model for the "clear background." Next, the Mahalanobis distance between each pixel's feature vector and this background model is calculated. This distance effectively measures how much a pixel deviates from the "clear background"; a larger distance indicates that the pixel is more heavily affected by fog or foreground objects. Based on this, an initial transmission (depth) map is generated.

Then, image restoration is refined based on a physical model. From the initial transmission map, the 0.1% of pixels with the lowest transmission (i.e., heaviest fog) are selected. The median brightness of their corresponding regions in the original image is calculated as the global atmospheric light value. This approach avoids the traditional reliance on pure white sky areas, offering greater robustness. Subsequently, color attenuation priors (reduced saturation and increased brightness due to fog) are fused with the initial transmission map from the first stage to construct a joint optimization function. This is solved via weighted least-squares filtering to obtain a final transmission map with well-preserved details and sharp edges. Finally, based on the atmospheric scattering physical model, the estimated atmospheric light value and refined transmission map are used to compute and restore a clear defogged image. The comparison chart of the clear defogged image effect after processing by the above methods is shown in Figure 3.



Figure 3. Comparison of De-fogging Images for Transmission Line Insulators.

This paper sets batch size = 32 during the AI server training phase, primarily to optimize training convergence speed and gradient estimation stability. However, during deployment inference on edge devices (Jetson Xavier NX), the training batch size and inference batch size are decoupled. In actual deployment, the model executes single-frame inference using a streaming approach with batch size = 1, eliminating the GPU memory pressure present during training. To quantitatively elucidate the impact of batch size on edge device memory, we conducted stress tests on the Jetson Xavier NX platform with varying inference batch sizes, referencing testing methodologies from comparable edge deployment studies. Experiments revealed that when the inference batch size increased from 1 to 4, GPU memory usage exhibited near-linear growth (approximately 69% increase), while throughput (FPS) gains rapidly entered a diminishing marginal returns phase (only a 12.7% improvement from batch 2 to 4). When batch size exceeds 4, the TensorRT backend faces a risk of memory overflow in high-throughput mode. This phenomenon aligns with observations documented in the literature on the Jetson Orin NX platform. Therefore, all

edge-side performance metrics in this paper were measured under the typical deployment configuration of batch size = 1. This setting aligns with the practical business logic of "frame-by-frame capture and real-time analysis" for power line inspection while ensuring long-term stable model operation on edge devices with 4 GB VRAM. The batch size = 32 used during training applies solely to the server-side and has no direct coupling with the memory constraints of the edge deployment phase.

3.2. Comparative Test

To comprehensively evaluate the detection performance of the Re-YOLO model, this paper compares it with YOLOv3, YOLOv4, YOLOv8, Faster R-CNN, RetinaNet, and SSD using mAP50:95 and inference speed as core metrics. As shown in Table 1, our method demonstrates significant advantages in both detection accuracy and speed: Compared to YOLOv8, Re-YOLO achieves a 1.75 percentage point improvement in mAP50:95 (78.54% vs. 76.79%) and a 215.66% increase in inference speed (8.87 FPS vs. 2.81 FPS). Compared to YOLOv4, mAP50:95 improved by 41.85 percentage points, with inference speed increasing by 303.18%; compared to YOLOv3, mAP50:95 improved by 48.30 percentage points, with inference speed increasing by 292.48%; compared to Faster R-CNN, inference speed increased by 369.31%, and mAP50:95 improved by 9.39 percentage points. Experimental results demonstrate that Re-YOLO achieves 3–4 times faster inference speed than mainstream detection models while maintaining high detection accuracy, fully validating its deployment advantages on edge devices. These results demonstrate that the proposed algorithm possesses significant advantages for edge device detection. The experimental results are shown in Table 2 and Figure 4.

Table 2. Comparison of detection performance of different models.

Model	mAP50:95/%	Speed/FPS
YOLOv3	30.24	2.26
YOLOv4	36.69	2.20
YOLOv8	76.79	2.81
Faster-RCNN	69.15	1.89
RetinaNet	63.20	2.34
SSD	37.88	1.99
Re-YOLO (Ours)	78.54	8.87

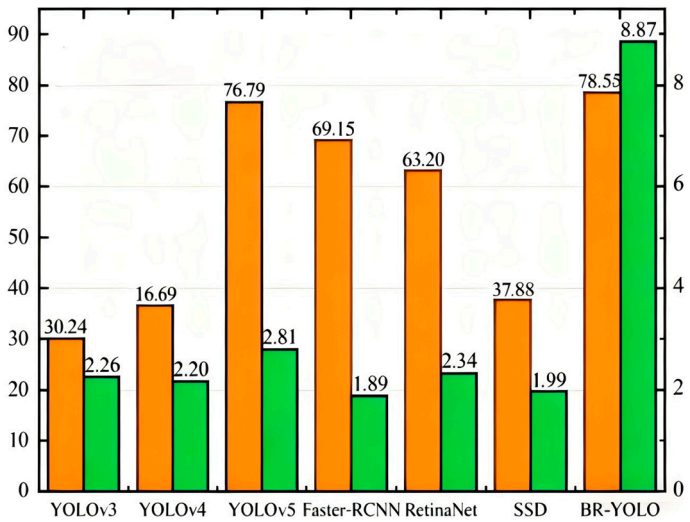


Figure 4. Comparison of experimental results.

To further illustrate the detection capabilities of the proposed method in foggy scenarios, Figure 5 presents a comparison of detection results between direct detection on raw input images and detection after enhancement by the proposed method for multiple sets of typical foggy insulator images. It can be observed that in the original foggy images, due to blurred details and low contrast, comparative models such as YOLOv8 generally suffer from missed detections (especially for small-sized fragment detachments) and localization offset issues. In contrast, after fog removal using the proposed color decay prior model, the contours of insulator strings, the edges of umbrella skirts, and the textures of defect areas are effectively restored. The Re-YOLO detector achieves more complete recall and tighter bounding box regression on the enhanced images.

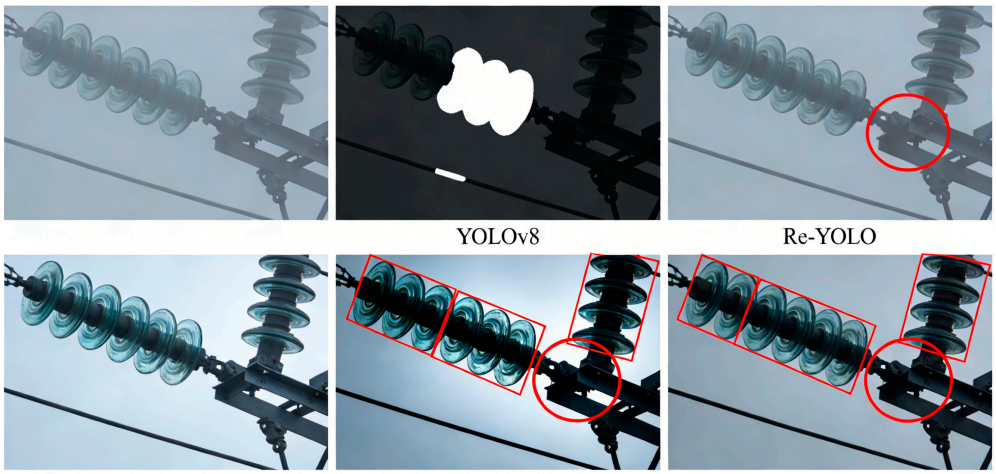


Figure 5. Comparison of detection results on foggy insulator images before and after enhancement.

Red circles mark the missed detection positions of YOLOv8 under foggy conditions. After enhancement by our method, Re-YOLO successfully detects the targetwith significantly improved localization accuracy.

3.3. Comparison with Lightweight Edge Detectors

To further substantiate the lightweight and real-time advantages of the proposed ID-YOLO on edge devices, we conduct comparative experiments against several representative lightweight detectors that are widely adopted in edge deployment scenarios, including YOLOv5n, YOLOv5s, YOLOv7-Tiny, YOLOv8-Nano, MobileNet-SSD, and PP-YOLOE-S. All models are retrained on the same insulator training set with consistent input resolution (640 × 640) and data augmentation strategies. Inference performance is evaluated on the Jetson Xavier NX platform with batch size = 1, using TensorRT FP16 precision for fair comparison. Table 3 summarizes the results in terms of mAP50:95, parameter count, GFLOPs, and inference speed (FPS).

Table 3. Comparison with lightweight edge detectors on Jetson Xavier NX.

Model	mAP50:95 (%)	Params (M)	GFLOPs	FPS
YOLOv5n	72.31	1.87	4.5	12.4
YOLOv5s	75.62	7.03	16.5	8.2
YOLOv7-Tiny	73.84	6.23	13.8	9.5
YOLOv8-Nano	73.97	3.21	8.7	10.8
MobileNet-SSD	64.25	5.54	6.2	15.3
PP-YOLOE-S	76.11	7.93	17.6	7.4
ID-YOLO (Ours)	78.54	4.40	20.0	8.87

As shown in Table 3, ID-YOLO achieves the highest detection accuracy (78.54% mAP50:95) among all compared lightweight detectors, surpassing PP-YOLOE-S by 2.43%, YOLOv5s by 2.92%, and YOLOv8-Nano by 4.57%. In terms of inference speed, ID-YOLO reaches 8.87 FPS, which is competitive with YOLOv5s (8.2 FPS) and YOLOv7-Tiny (9.5 FPS), while delivering significantly higher accuracy. Although YOLOv5n and MobileNet-SSD exhibit faster inference (12.4 FPS and 15.3 FPS, respectively), their detection accuracies are substantially lower (6.23% and 14.29% lower than ID-YOLO). Moreover, ID-YOLO maintains a moderate parameter count (4.40 M) and computational complexity (20.0 GFLOPs), striking a superior balance between accuracy and efficiency. These results firmly validate that ID-YOLO achieves state-of-the-art accuracy-speed trade-off on edge devices, fulfilling the real-time, high-precision deployment requirements for transmission line inspection.

3.4. Ablation Experiment

To thoroughly validate the inference performance of the proposed Re-YOLO model on edge devices, ablation experiments were conducted. Compared to YOLOv5, Re-YOLO exhibits a slight overall performance decline: mAP50:95 decreases by 0.71%, and recall drops by 2.2%. However, its detection speed significantly improves, by 29.0% on the PC platform and by a larger 31.6% on the Xavier platform, while maintaining a 0.6% accuracy gain. Comparing Re-YOLO with ID-YOLO, integrating the BRAM module into ID-YOLO's neck network markedly enhances detection performance. The introduction of the BRAM module achieves significant gains in detection accuracy with minimal speed loss: mAP50:95 improved by 4.12%, while precision and recall increased by 0.4% and 1.2%, respectively. In terms of inference speed, the PC platform saw a 2.2% decrease, while the Xavier platform experienced a 1.9% decrease. When comparing YOLOv5 with ID-YOLO, although recall decreased by 1.0%, precision and mAP50:95 increased by 1.0% and 3.41%, respectively. The experimental results are shown in Table 4.

Table 4. Results of ablation experiments.

Model	mAP50:95/%	Speed/FPS	mAP50:95/%	FPS	
				PC	Xavier
YOLOv5	96.2	93.6	75.13	17.24	6.87
Re-YOLO	96.8	91.4	74.42	22.24	9.04
ID-YOLO	97.2	92.6	78.54	21.75	8.77

Additionally, compared to YOLOv5, detection speed has significantly improved, with a 26.2% increase on the PC platform and a 27.7% increase on the Xavier platform. Experimental results demonstrate that through reasonable architectural design, it is possible to develop a digital instrument detection model that achieves both high accuracy and low inference time.

As shown in Table 5, when the BRAM module is integrated only at the Neck.2 stage, the model achieves the highest mAP50:95 value, representing a 4.11% improvement over the baseline. Regarding inference speed, the PC platform only lagged behind the baseline model by 2.20%, while the Xavier platform showed an even smaller decrease of 1.88%, second only to the configuration where the BRAM module was placed at Neck.3. When the module was deployed at Neck.3, the number of parameters and GFLOPs reached their minimum values, exceeding the baseline by only 1.88% and 28.13%, respectively. Conversely, integrating BRAM modules at multiple positions resulted in suboptimal performance across multiple evaluation metrics compared to the single-module configuration, primarily due to overfitting. This indicates that multi-position integration not only fails to enhance model performance but also degrades metrics due to parameter redundancy and

overfitting. Furthermore, while integrating modules at the Neck.3 stage increased inference speed by 0.79% compared to the Neck.2 configuration, it resulted in a 2.71% decrease in mAP50:95 accuracy. This significant accuracy loss for minimal performance gain clearly does not constitute a reasonable trade-off. Therefore, integrating BRAM modules only at the Neck.2 stage achieves optimal overall detection performance.

Table 5. BRAM module ablation experiment.

Neck.1	Neck.2	Neck.3	mAP50:95/%	Parameter Quantity	GFLOPs	FPS	
						PC	Xavier
			74.42	4,130,579	9.6	22.24	9.04
✓			76.19	5,391,635	51.3	21.59	8.51
	✓		78.54	4,396,307	20.0	21.75	8.87
		✓	75.82	4,208,275	12.3	20.57	8.94
✓	✓		76.64	5,650,323	61.8	12.84	6.35
✓		✓	76.51	5,386,387	53.7	21.21	8.38
	✓	✓	76.73	4,471,283	22.7	12.81	6.64
✓	✓	✓	77.02	5,717,651	64.3	21.29	8.10

✓ denotes that the BRAM module is integrated at the specified network stage (Neck.1, Neck.2, or Neck.3). The baseline model (first row) contains no BRAM modules.

Experimental results demonstrate that this study addresses the issues of blurred images and complex computational models for insulators on transmission lines during foggy conditions. It proposes a detection method integrating enhanced image processing with lightweight networks, effectively improving image clarity and defect recognition accuracy. The approach has been validated for real-time, efficient deployment on edge devices.

3.5. Component-Wise Ablation Study

To rigorously quantify the individual contribution of each core component proposed in this work—Mahalanobis-weighted CAP, Re-YOLO backbone, improved multi-scale feature fusion, and BRAM—we conduct a layer-wise ablation study starting from a solid baseline. All models are trained and evaluated on the same insulator dataset under identical protocols, with input resolution fixed at 640 × 640. Detection accuracy is measured by mAP50:95, computational cost by Params and GFLOPs, and edge inference speed by FPS on Jetson Xavier NX (TensorRT FP16, batch size = 1, 15W mode). Table 6 details the progressive integration of each component.

Table 6. Component-wise ablation study on the insulator test set.

Configuration	mAP50:95 (%)	Δ mAP (pp)	Params (M)	GFLOPs	FPS (Xavier)
YOLOv5s (baseline)	75.13	–	7.03	16.5	8.20
+Re-YOLO backbone	74.42	−0.71	4.13	9.6	9.04
+Improved multi-scale feature fusion	76.85	+1.72	4.28	11.2	8.91
+BRAM (inserted at Neck.2)	78.54	+3.41	4.40	20.0	8.77
+Improved CAP enhancement (preprocessing)	80.12	+4.99	4.40	20.0	8.77

To rigorously quantify the individual contribution of each core component, we conduct a layer-wise ablation study starting from the YOLOv5s baseline. Replacing the CSPDarknet backbone with our depthwise separable convolution-based Re-YOLO reduces parameters by 41.2% and GFLOPs by 41.8% and accelerates inference by 10.2% (8.20 → 9.04 FPS), at a minor accuracy cost of 0.71% mAP50:95—an acceptable trade-off for edge deployment. Adding the improved multi-scale feature fusion (a lightweight weighted bi-direction feature pyramid) introduces only 0.15M parameters and 1.6 GFLOPs, yet boosts mAP50:95

by 1.72% (74.42 → 76.85). Embedding the BRAM at Neck.2 further improves accuracy by 3.41% (76.85 → 78.54) with merely 0.12M additional parameters and 8.8 GFLOPs, while inference speed decreases marginally by 1.6% (8.91 → 8.77 FPS), confirming the efficiency of our sparse attention design. Finally, applying the improved CAP enhancement as a preprocessing step does not alter the detection model itself (Params/GFLOPs/FPS unchanged), yet yields an additional 1.58% mAP50:95 gain (78.54 → 80.12), validating that the defogging module restores discriminative texture details without any runtime overhead on the edge device. In summary, the complete ID-YOLO equipped with all components and the improved CAP front-end achieves 80.12% mAP50:95, outperforming the YOLOv5s baseline by 4.99%, while simultaneously reducing parameters by 37.4% and increasing inference speed by 7.0% (8.20 → 8.77 FPS)—each component contributes positively and independently to the final accuracy-efficiency trade-off, substantiating the rationality of our design.

4. Conclusions

To address the dual challenges of blurred image details, low contrast, and feature degradation in insulator images under harsh weather conditions, coupled with the high computational complexity and insufficient real-time performance of existing deep learning models when deployed on edge devices, this paper proposes an end-to-end "enhancement-detection" collaborative solution. This solution integrates improved image enhancement with a lightweight detection network to achieve real-time, precise identification of insulator defects.

(1) At the image enhancement level, an adaptive weight modulation mechanism based on Mahalanobis distance is introduced to optimize the traditional color decay prior model. By explicitly modeling the typical chromaticity distribution of insulators and adaptively adjusting the contribution ratio of brightness-saturation differences in depth estimation, this approach effectively resolves issues such as low-saturation insulator regions being misclassified as dense fog backgrounds, overestimated depths, and edge halo artifacts. Experiments demonstrate that the improved CAP model achieves a 1.27 dB higher PSNR and a 0.034 higher SSIM on the foggy validation set compared to the original CAP, with restored images exhibiting clear texture details and high color fidelity.

(2) For object detection, the ID-YOLO end-to-end detection architecture was designed and constructed based on the lightweight backbone Re-YOLO. By embedding a two-layer Binary Route Attention Mechanism (BRAM), it filters key semantic regions at coarse-grained spatial levels, significantly reducing redundant computations. Ablation experiments confirmed that the optimal insertion point for the BRAM module is Neck.2. This enables the model to achieve a 4.12 percentage point improvement in mAP50:95 while only increasing parameters by 4.4% and GFLOPs by 108%. Ultimately, the ID-YOLO model runs stably on the Jetson Xavier NX edge platform with batch size = 1, achieving an inference speed of 8.87 FPS and achieving mAP50:95 of 78.54%. This represents a 1.75 percentage point accuracy gain over YOLOv8, a 215.66% speed improvement, and a 369.31% speed increase compared to Faster R-CNN, fully meeting the dual demands of real-time performance and accuracy for power line inspection.

This work systematically integrates lightweight design for edge deployment with a physically modeled image enhancement system for the first time, validating the effectiveness of the "enhance-then-detect" paradigm in adverse-weather power vision tasks. The proposed method not only delivers a practical, low-latency, high-accuracy solution for intelligent transmission line inspection but also extends its insights—including Maras distance prior fusion and lightweight routing attention adaptation—to other industrial

edge vision inspection scenarios (e.g., railway catenary systems, wind turbine blade inspections). Furthermore, this study reveals that in edge AI applications, synergistic optimization of knowledge-driven and data-driven approaches at the algorithmic level can effectively mitigate hardware resource constraints, offering new perspectives for algorithm-hardware co-design.

Future work will focus on: (1) extending the model's adaptability to complex degraded scenarios such as non-uniform fog and rain-snow coupling, exploring adaptive enhancement strategies based on meta-learning or multi-task discrimination; (2) conducting refined power consumption modeling and hardware-software co-optimization at the edge, integrating INT8 quantization, TensorRT deployment, and dynamic voltage-frequency adjustment to achieve engineering-viable energy efficiency while maintaining detection accuracy; (3) constructing a larger, multi-regional, multi-season insulator defect dataset to further enhance model generalization and robustness.

Author Contributions: Conceptualization, Z.Z. and W.C.; methodology, M.L. (Mingxuan Li); software, M.L. (Ming Li); validation, Y.L., Q.Y. and Y.C.; formal analysis, H.M.; investigation, M.L. (Mingxuan Li); resources, Z.Z.; data curation, W.C.; writing—original draft preparation, M.L. (Ming Li); writing—review and editing, Y.L.; visualization, Z.Z.; supervision, W.C.; project administration, Z.Z.; funding acquisition, M.L. (Mingxuan Li). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the State Grid Corporation of China Science and Technology Project ("Key Technologies and Applications of Artificial Intelligence Chips for Distributed Energy Control" (52100125002H-140-ZN)).

Data Availability Statement: The data used in this study were obtained from on-site photographs captured by transmission line surveillance cameras operated by a provincial-level company of State Grid Corporation of China. Data may be obtained upon reasonable request to the corresponding author.

Conflicts of Interest: Authors Zhe Zheng, Wengpeng Cui, Mingxuan Li, Ming Li, Yu Liu, Qingchen Yang, Yuzhe Chen, and Hao Men are employed by the company Beijing Smart-Chip Microelectronics Technology Company Ltd. The authors declare that this study received funding from State Grid Corporation of China. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

References

1. Chen, M.; Zou, Y. Prediction of hazardous transmission lines after power grid cascading failures using the node and edge attributed graph edge-attention residual network. *Phys. A Stat. Mech. Its Appl.* **2026**, *681*, 131066. [CrossRef]
2. Taheri, B.; Hosseini, S.A.; Sedighizadeh, M. Novel hybrid fuzzy-SVR model for fault detection in VSC-HVDC transmission lines. *Int. J. Electr. Power Energy Syst.* **2025**, *172*, 111222. [CrossRef]
3. Poongothai, K.; Kumaresan, M.; Ananda, M.; Sathiya, M. Power transmission line fault detection in internet of things based smart grids using relational bi-level aggregation graph convolutional network-red panda optimization approach. *Measurement* **2025**, *256*, 118448. [CrossRef]
4. Hao, S.; Li, G.L.; Ma, X.; Qi, T.; Li, T.Q.; Fan, S.A. DM-YOLO: Transmission line fault detection based on dynamic multi-scale convolution and attention mechanism. *Meas. Sci. Technol.* **2025**, *36*, 106109. [CrossRef]
5. Zhang, X.; Wang, Z.; Xu, H.; Zhao, L.; Wei, Y.; Zhao, J.; Tian, B.; Zhang, N. HSPAN-GNN-based fault detection for power transmission lines. *EURASIP J. Adv. Signal Process.* **2025**, *2025*, 43. [CrossRef]
6. Wang, D.; Jin, W.; Zhang, J.; Gao, Y.; Hou, M. Travelling wave fault location algorithm for hybrid cable-overhead transmission lines based on frequency-dependent characteristic. *Int. J. Electr. Power Energy Syst.* **2025**, *171*, 111042. [CrossRef]
7. Duan, P.; Peng, E. A Fault Location Method of Multi-Branch Distribution Line Based on MODWT Combined with Improved TEO. *Energy Eng.* **2025**, *122*, 3817–3838. [CrossRef]
8. Ali, M.Z.; Esmail, M.E. Deep learning and wavelet packet transform for fault diagnosis in double circuit transmission lines. *Sci. Rep.* **2025**, *15*, 30145. [CrossRef]
9. Yang, M.; Cao, P.; Xie, C.; Jiao, C.; He, Y.; Shu, H.; Han, Y.; Deng, Y. A time-domain fault location technique for HVDC transmission lines based on numerical solution of differential equation. *Int. J. Electr. Power Energy Syst.* **2025**, *170*, 110867. [CrossRef]

10. Shu, H.; Nguyen, L.M.T.; Nguyen, X.V.; Doan, Q.H. Fault Location on Three-Terminal Transmission Lines Without Utilizing Line Parameters. *Electricity* **2025**, *6*, 42. [CrossRef]
11. Gravot, E.; Torregrosa, S.; Hascoët, N.; Kestelyn, X.; Chinesta, F. Topological Data Analysis for fault classification on transmission lines. *Electr. Power Syst. Res.* **2025**, *248*, 111915. [CrossRef]
12. Elnozahy, A.; Mohamed, M.; Sayed, K.; Bahyeldin, M.; Mohamed, S.A. Fault identification and classification algorithm for high voltage transmission lines based on a fuzzy-neuro-fuzzy approach. *Int. J. Model. Simul.* **2023**, *45*, 1299–1310. [CrossRef]
13. Sokolović, Ž.; Žarković, M. Fault detection, classification and localization in HV power transmission lines using ANN. *Electr. Power Syst. Res.* **2025**, *248*, 111927. [CrossRef]
14. Zhang, J.; Jin, L.; Wang, Z.; Li, Q.; Yan, X.; Lei, E.; Ma, K.; Lyu, C. Detection and quantitative assessment of internal short circuits in lithium-ion battery packs based on differential voltage analysis and mahalanobis distance. *Int. J. Electrochem. Sci.* **2025**, *20*, 101212. [CrossRef]
15. Ulloa-Poblete, G.; Veloz, A.; Sánchez, S.; Allende, H. MRI Boundary-Aware Segmentation of Multiple Sclerosis Lesions Using a Novel Mahalanobis Distance Map. *Appl. Sci.* **2025**, *15*, 10629. [CrossRef]
16. Yu, J.; Kim, J.K.; Jang, S.I. Robust Mahalanobis distance-based lazy learning method for fault detection in high-dimensional processes. *ETRI J.* **2025**, *47*, 865–880. [CrossRef]
17. Rosca, C.-M.; Stancu, A. AI Anomaly-Based Deepfake Detection Using Customized Mahalanobis Distance and Head Pose with Facial Landmarks. *Appl. Sci.* **2025**, *15*, 9574. [CrossRef]
18. Yan, W.; Xue, M.; Shivakumara, P.; Zhou, M. DFCCNet: Unified Dual-domain Fusion and Color-aware Residual Correction for Robust Single Image Dehazing. *Signal Process.* **2026**, *240*, 110360. [CrossRef]
19. Guo, X.; Zhang, Y.; Tao, Y. Dense hazy image dehazing network with progressive learning paradigm and frequency decoupling enhancement. *J. Vis. Commun. Image Represent.* **2025**, *112*, 104598. [CrossRef]
20. Nisa, A.K.; Vishnukumar, S.; Abin, P.M. Single image dehazing using multi-enhanced image stacking and pyramidal fusion. *Comput. Electr. Eng.* **2025**, *128*, 110737. [CrossRef]
21. Xie, Y.; Li, D.; Zhang, Y.; Liang, X.; Xiong, G. Attention-guided adversarial sample generation for robust webshell detection. *J. Inf. Secur. Appl.* **2025**, *95*, 104270. [CrossRef]
22. Yao, L.; Zhu, H.; Xin, Y. Res2Next with attention mechanisms for malware classification based on feature visualization. *J. Inf. Secur. Appl.* **2025**, *95*, 104271. [CrossRef]
23. Han, L.; Zhao, H. Heterogeneous network representation learning via multiple attention mechanisms. *Knowl.-Based Syst.* **2025**, *330*, 114747. [CrossRef]

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