EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

Research on a Safety-Perception-Oriented Image Enhancement and Lightweight Object Detection Model for Power IoT

Zikang Wei[1,*]

[1]Shenyang Institute of Engineering, Shenyang 110136, China

Abstract

To address image quality degradation, difficulty in recognizing small defect targets, difficulty in quantifying external-damage risks, and limited computing resources of embedded devices in edge inspection of the power Internet of Things (Power IoT), this paper proposes an image enhancement and lightweight object detection model for power safety perception, named EEL-YOLO. First, a detection-friendly image enhancement module is constructed to adaptively enhance degraded images under low illumination, haze, backlight, and motion blur, thereby restoring edge, texture, and saliency features of power targets. Second, based on the YOLOv8 baseline framework, a lightweight multi-scale attention backbone, a small-target enhanced feature fusion network, and the MPDIoU bounding-box loss function are introduced to improve the detection accuracy of insulator damage, vibration-damper defects, and external-damage targets in transmission corridors. Finally, structural re-parameterization, pruning, and knowledge distillation are combined to compress the model, and a multi-level safety warning mechanism based on pixel overlap, distance risk, and defect severity is designed. Validation experiments show that, on the power inspection experimental dataset and the degradation test set constructed in this paper, the proposed enhancement module achieves an mAP@0.5 of 91.8% after enhancing degraded images, with an average processing time of 9.7 ms. EEL-YOLO achieves an mAP@0.5 of 95.1% and an mAP@0.5:0.95 of 72.6% on the test set, with an FPS of 118. On the RK3588 edge platform, its FPS reaches 67.6. Under degradation test scenarios such as low illumination, haze, backlight, and motion blur, the average mAP@0.5 reaches 92.8%. Under an evaluation protocol combining rule-based labels and manual review, the risk-grading consistency reaches 91.7%, and the recall rate of Level-I early warning reaches 93.6%. The results verify the effectiveness and real-time capability of the proposed method for edge safety perception in Power IoT under the experimental conditions.

Keywords: Power Internet of Things, Safety perception, Image enhancement, Lightweight object detection, EEL-YOLO, Edge deployment, Multi-level warning

Received on 29 June 2026, accepted on 05 September 2026, published on 24 September 2026

doi: 10.4108/eetsis.13778

*Corresponding author. Email: weizikangsidney@163.com

1. Introduction

1.1. Research Background and Significance

With the rapid growth of renewable-energy grid integration, power-grid digitization, and intelligent operation and maintenance requirements, traditional power systems are evolving toward new power systems and smart grids. Smart grids emphasize the deep integration of information flow, energy flow, and business flow, and rely on advanced communication, sensing, computing, and control technologies to achieve real-time perception and reliable

dispatching of power equipment states [1-2]. In this context, Power IoT connects transmission, substation, and distribution links through unmanned aerial vehicles, online monitoring terminals, edge computing nodes, intelligent sensors, and cloud platforms, providing important support for power equipment state identification, external-damage warning in transmission corridors, and fault-assisted decision-making [3-4].

Image data are a key data type in Power IoT safety perception. UAV inspection and fixed monitoring devices continuously collect images of insulators, conductors, fittings, vibration dampers, towers, construction machinery, and foreign-object intrusion. In practical deployment, four pain points are coupled: long-distance defects and external-damage objects often occupy only a small image area; low illumination, haze, backlight, and motion blur weaken edges and textures; false or delayed alarms increase manual review and operation-and-maintenance burden; and UAV or tower-side terminals have limited memory, computing power, and communication bandwidth. Existing studies have shown that deep learning can improve aerial insulator-defect recognition [5-6] and transmission-line foreign-object detection [7-10]. Consequently, a deployable Power IoT safety-perception model should improve degraded-image usability and small-target recognition without relying on server-scale computation, while transforming detections into actionable risk warnings.

Therefore, a safety perception model for Power IoT should not only improve image usability in complex environments, but also balance object detection accuracy, model lightweighting, and real-time edge deployment capability. Research on the collaborative optimization of image enhancement and lightweight object detection is of great significance for improving intelligent power inspection, reducing manual review costs, decreasing missed detections and false alarms, and ensuring the safe and stable operation of new power systems.

1.2. Research Status at Home and Abroad

1.2.1 Research Status of Power Inspection Image Enhancement

Image enhancement and restoration are important means of improving the quality of power inspection images in complex environments. Traditional dehazing methods are represented by the dark channel prior, which estimates transmission and atmospheric light through a physical degradation model to clarify hazy images [11]. With the development of deep learning, AOD-Net models dehazing as an end-to-end mapping, which facilitates joint training with object detection networks [12]. GridDehazeNet enhances dehazing capability through attention and a multi-scale grid structure [13]. For low-light enhancement, Retinex-Net learns reflectance and illumination components based on Retinex decomposition [14], while Zero-DCE realizes low-light enhancement without paired samples through zero-reference curve estimation [15]. In recent years, Transformer-based image restoration models have further improved high-resolution image denoising, deraining, deblurring, and enhancement performance [16].

The above methods provide important foundations for restoring degraded inspection images. However, most conventional enhancement methods are optimized primarily for perceptual quality, so visually improved images do not necessarily maximize downstream detection accuracy. Recent task-driven studies have therefore begun to couple enhancement and detection. Shen et al. [17] jointly optimize low-light enhancement and object detection through shared representation learning, while Kou et al. [18] combine a lightweight two-stage Transformer enhancer with a YOLO detector for end-to-end dark-object detection. These studies confirm the need to align enhancement objectives with detector utility. Nevertheless, existing joint schemes mainly target generic low-light scenes and do not simultaneously address multiple power-inspection degradations, elongated power components, tiny defect regions, edge-oriented compression, and downstream safety-risk grading. This gap motivates a detection-friendly enhancement module specifically oriented to Power IoT edge inspection.

1.2.2 Research Status of Object Detection Algorithms in Power Scenarios

Object detection algorithms have evolved from two-stage detection to one-stage real-time detection. Faster R-CNN improves candidate-box generation efficiency through a region proposal network and is an important representative of two-stage detection methods [19]. The YOLO series unifies object localization and classification as an end-to-end regression task and has obvious advantages in detection speed and engineering deployment [20]. To improve small-object and multi-scale object detection, FPN fuses multi-level features through a top-down pathway and lateral connections [21], while EfficientDet further proposes a weighted bidirectional feature pyramid to improve the efficiency of multi-scale feature fusion [22]. Attention mechanisms are also widely used for object detection in complex backgrounds. CBAM enhances key features through channel and spatial attention [23], and the SE module improves network representation through channel recalibration [24]. In recent years, RT-DETR has improved real-time detection performance through an efficient hybrid encoder and an end-to-end detection framework, providing a new technical route for high-accuracy detection on edge devices [25].

In power inspection scenarios, object detection tasks mainly include external-damage hazard detection in transmission corridors and power-equipment defect detection. Existing methods improve accuracy through data augmentation, multi-scale fusion, attention mechanisms, small-object detection layers, and improved bounding-box regression. For example, recent YOLOv8n-based transmission-line studies introduce attention and an additional small-object layer to improve distant foreign-object recognition [26], and newer enhanced-YOLOv8 schemes continue to target small and multi-scale foreign objects under complex transmission-line backgrounds [27]. However, these methods remain detection-centric: degraded-image restoration is generally separated from detector optimization, edge compression is not always evaluated under a unified deployment pipeline, and the outputs normally stop at bounding boxes rather than graded safety decisions. Therefore, improving the detector alone is insufficient for the complete Power IoT safety-perception problem considered in this paper.

1.2.3 Research Status of Lightweight Detection and Edge Deployment

Power IoT edge terminals impose requirements of low parameter count, low computation, and high real-time performance on object detection models. Lightweight

networks are an important direction for solving edge deployment problems. MobileNetV2 significantly reduces computation through inverted residual structures and depthwise separable convolutions [28]. GhostNet uses inexpensive linear transformations to generate redundant feature maps, reducing model complexity while maintaining representation capability [29]. Structural re-parameterization methods such as RepVGG use multi-branch structures during training to enhance expressive capability and convert them into a single-branch convolutional structure during inference, thereby improving deployment efficiency [30]. In addition, knowledge distillation transfers the discriminative knowledge of a teacher model to a lightweight student model to compensate for accuracy loss caused by pruning and compression [31]. In terms of detection loss design, SIoU and other bounding-box regression losses introduce angle, distance, and shape constraints to accelerate convergence and improve localization accuracy [32].

Although the above methods have greatly promoted lightweight object detection models, power inspection applications still face the difficulty of balancing accuracy and efficiency. Excessive compression weakens the model's ability to recognize small objects, weak-texture defects, and complex backgrounds, while more complex enhancement and detection models are difficult to deploy on UAVs, monitoring terminals, and embedded edge devices. Edge computing and mobile edge computing optimization for multiuser Power IoT provide feasible routes for on-site image data processing, but models still need to balance accuracy, speed, storage, energy consumption, and communication overhead [33-34].

In summary, progress has been made in power inspection image enhancement, object detection, and edge deployment, but the following deficiencies remain. First, image enhancement and object detection are mostly processed in a cascaded manner, lacking a detection-task-driven joint optimization mechanism, and enhancement results may not necessarily benefit defect recognition. Second, small-object detection accuracy under severe weather, low illumination, occlusion, and complex backgrounds is still insufficient, especially for insulator bursts, vibration-damper defects, and distant external-damage targets, where missed detections may occur. Third, lightweight models may suffer from reduced feature representation when computation is reduced, making it difficult to balance detection accuracy and edge real-time performance. Fourth, existing studies mostly focus on a single algorithmic module and lack an integrated safety perception scheme for the cloud-edge-end collaborative architecture of Power IoT.

1.3. Research Motivation and Contributions

Against this background, the motivation of EEL-YOLO is system-level integration rather than claiming that every individual primitive is entirely new. Mainstream YOLO-based power-inspection methods can achieve strong detection accuracy, but they generally lack a unified mechanism that makes image enhancement detection-oriented, preserves small and elongated power-target features under lightweight constraints, and converts detection geometry into operational risk levels. In edge inspection composed of UAVs and fixed monitoring devices, low illumination, haze, backlight, and motion blur can simultaneously degrade target visibility, while insulator defects, vibration-damper defects, and distant construction hazards are often small, weak-textured, or partially occluded. The resulting engineering requirement is therefore multi-objective: high detection accuracy, robustness to tested degradations, low edge-side latency, and interpretable warning output must be achieved together.

The contributions are summarized as follows. (1) A detection-friendly image-enhancement module is embedded into the detection pipeline so that restoration is constrained by target structure and detector utility rather than visual quality alone. (2) A lightweight multi-scale attention backbone and a P2-based boundary-semantic fusion network strengthen elongated-component and small-defect representation while limiting edge-side complexity. (3) The YOLOv8-based detection head is combined with MPDIoU localization, structural re-parameterization, structured pruning, and knowledge distillation to improve the accuracy-efficiency trade-off. (4) A post-detection safety-risk layer integrates confidence, spatial overlap, distance risk, and defect severity to convert object detections into graded warnings. The resulting novelty is the coordinated design of enhancement, lightweight detection, edge deployment, and risk perception for one Power IoT workflow. A concise comparison with representative methods is provided in Table 1.

Table 1. Comparison of representative detection and joint enhancement-detection approaches

Method	Main comparison with EEL-YOLO
YOLOv8n [35]	Detector only; nano baseline; no task-driven enhancement or graded risk output.
TL-YOLO [8]	Transmission-line foreign-object detector; no detection-oriented enhancement or risk-grading layer.
Improved YOLOv8n [26]	Attention plus a small-object layer; remains detector-centric and has no graded safety output.
LDWLE [17]	Joint low-light enhancement and detection; does not address Power-IoT graded risk perception or compression.
Kou et al. [18]	End-to-end low-light enhancement plus detection; no Power-IoT graded warning mechanism.
EEL-YOLO (ours)	Detection-oriented enhancement + P2/strip attention + compression + post-detection graded warning.

2. Power IoT Safety Perception Tasks and Technical Foundations

2.1. Definition of Edge Power Visual Perception Tasks

Visual safety perception tasks in Power IoT are usually jointly formed by end-side image acquisition, edge-side intelligent

analysis, and cloud-side risk management. End-side devices mainly include UAV gimbal cameras, tower fixed monitoring cameras, and mobile inspection terminals. Edge-side devices are responsible for image preprocessing, object detection, preliminary alarm filtering, and result uploading. Cloud platforms undertake data archiving, model updating, risk linkage, and operation and maintenance decision-making. Unlike server-side offline detection, edge inspection in Power IoT is constrained by computing power, memory, energy consumption, and communication bandwidth. Therefore, the model must not only have high detection accuracy, but also meet the requirements of low parameter count, low latency, and real-time inference [33-34].

This paper divides power visual safety perception tasks into two categories. The first is external-damage hazard detection in transmission corridors, which mainly identifies cranes, trucks, cars, and other targets that may enter the line safety area and threaten transmission-line operation. The second is power equipment defect detection, which mainly identifies equipment abnormalities such as insulator damage, insulator bursts, vibration-damper deformation, and vibration-damper defects. Let the input inspection image be I; the detection model output is given in Equation (1):

$$Y = \left\{ \left(b_i, c_i, s_i \right) \right\}_{i=1}^{N} \tag{1}$$

Where b_i denotes the bounding box of the i-th target, c_i denotes the target category, s_i denotes the confidence score, and N denotes the number of targets detected in the image. For external-damage hazard targets, the spatial relationship between the detection box and the transmission corridor safety area should be further combined to determine the risk level. For equipment defect targets, defect category, defect location, and confidence score should be emphasized to assist operation and maintenance personnel in review and defect elimination.

2.2. Related Basic Modules

The YOLO series models object detection as an end-to-end regression task and has the advantages of fast inference and convenient engineering deployment [20]. This paper uses a YOLOv8-like one-stage detection framework as the baseline network [35]. Its Backbone is used to extract image features, the Neck is used to fuse multi-scale information, and the Head is used to output categories, confidence scores, and bounding boxes. Considering that power inspection images contain many small objects, occluded targets, and complex backgrounds, relying only on the original YOLO network easily leads to feature weakening and missed detections. Therefore, image enhancement, multi-scale attention, lightweight feature fusion, and an improved bounding-box loss function are introduced into the baseline network in subsequent sections.

Structural re-parameterization is an important method for improving inference speed on edge devices. Its core idea is to use multi-branch structures during training to enhance feature expression, and equivalently fuse multi-branch convolutions, batch normalization, residual connections, and other structures into a single-branch convolution during inference, thereby reducing computational overhead [30]. Let the input feature be X, the convolution kernel be W, and the batch-normalization parameters be mean μ, variance σ^2, scaling factor γ, and offset β. Then the convolutional layer and BN layer can be fused as shown in Equation (2):

$$W' = \frac{\gamma}{\sqrt{\sigma^2 + \varepsilon}} W, \quad b' = \beta - \frac{\gamma \mu}{\sqrt{\sigma^2 + \varepsilon}} \tag{2}$$

This method can improve inference efficiency without significantly sacrificing detection accuracy, and is suitable for deployment on UAVs, tower monitoring terminals, and embedded edge devices. Attention mechanisms can enhance responses in key target regions and suppress complex background interference. CBAM recalibrates input features sequentially through channel attention and spatial attention, and its calculation process is summarized in Equation (3):

$$F' = M_c(F) \otimes F, \quad F'' = M_s\left(F'\right) \otimes F' \tag{3}$$

Where F denotes the input feature map, Mc and Ms denote channel attention and spatial attention, respectively, and ⊗ denotes element-wise multiplication [23]. Compared with CBAM, strip multi-scale Conv-A is more suitable for elongated targets such as vibration dampers, insulator strings, and conductor fittings. It uses asymmetric convolutions such as 3×5, 3×7, and 3×9 to extract directional features at different scales, thereby enhancing the representation capability of small targets and local defect regions.

The bounding-box loss function directly affects target localization accuracy. CIoU considers overlap, center-point distance, and aspect-ratio consistency, while DIoU emphasizes the center-point distance between predicted and ground-truth boxes [36]. More recent IoU-based formulations such as N-IoU further refine geometric regression constraints [32]. MPDIoU was originally introduced in [37] to constrain both overlap and the minimum distances between corresponding upper-left and lower-right corners; it has also been adopted in peer-reviewed object-detection studies [38]. Considering that power inspection targets are often small, blurred at the boundary, and partially occluded, this paper compares candidate losses and adopts MPDIoU for the reported model.

Lightweight methods are the key to edge deployment. Depthwise separable convolution reduces computation by decomposing standard convolution into depthwise convolution and pointwise convolution [28]. GhostNet uses low-cost linear transformations to generate redundant features, reducing the number of parameters while maintaining feature representation capability [29]. Structured pruning compresses the model by pruning low-contribution convolution kernels or channels. Knowledge distillation uses a high-accuracy teacher model to guide a lightweight student model, thereby compensating for accuracy loss caused by pruning or lightweighting [31]. The above methods provide a foundation for constructing the lightweight power object detection model in this paper.

2.3. Construction of a Multi-Class Power Safety Perception Dataset

To verify the applicability of the proposed method in complex power inspection scenarios, this paper constructs a multi-class power safety perception dataset. The data sources include UAV aerial images and tower fixed monitoring images. UAV aerial images mainly cover transmission-line towers, insulators, vibration dampers, conductors, and fittings, with scenarios including mountainous areas, plains, rivers, forests, villages, and towns. Tower fixed monitoring images mainly focus on external-damage hazards in transmission corridors and include long-distance monitoring, backlight, low illumination, haze, occlusion, and complex backgrounds. To avoid model overfitting caused by a single data source, the dataset construction covers different shooting angles, target scales, illumination conditions, and background types as much as possible.

The dataset contains seven target classes. The external-damage hazard targets include three classes: crane, truck, and car. The power equipment defect targets include four classes: insulator damage, insulator burst, vibration-damper deformation, and vibration-damper defect. Rectangular boxes are used for target annotation, and the annotation results are uniformly converted into YOLO format as defined in Equation (4):

$$I = \left(c, x_c, y_c, w, h\right) \tag{4}$$

Where c denotes the category number, and x^c, y^c, w, and h denote the normalized center coordinates, width, and height of the target box, respectively. For external-damage hazard targets, only objects that may enter the transmission corridor safety area and pose potential threats are annotated. For equipment defect targets, the annotation box should fit the defect area as closely as possible and avoid including excessive background. If multiple targets exist in the same image, they are annotated independently.

The dataset is divided into training, validation, and test sets at a ratio of 7:2:1. The training set is used for model parameter learning, the validation set for hyperparameter selection and model tuning, and the test set for final performance evaluation. To avoid data leakage, highly similar images from the same video sequence, the same shooting point, or consecutive frames are not simultaneously assigned to both the training set and the test set. Meanwhile, the class proportions are kept as consistent as possible during division to reduce the influence of class imbalance on evaluation results.

The dataset contains 7,000 unique images and 13,420 annotated objects. The scenario-level split contains 4,900 training images, 1,400 validation images, and 700 test images. Class-wise image/object counts are reported in Table 2. Because one image may contain more than one target class, class-wise image counts are not additive across rows, whereas the object counts are additive. All class statistics preserve approximately the same 7:2:1 split proportion and no same-flight, same-monitoring-point, or consecutive-frame group is shared across subsets.

Table 2. Verified dataset statistics by class and split

Class	Train; Val; Test (images/objects)
Crane	610/805; 174/230; 87/115
Truck	700/924; 200/264; 100/132
Car	1100/1638; 314/468; 157/234
Insulator damage	1050/1715; 300/490; 150/245
Insulator burst	780/1036; 223/296; 111/148
Vibration-damper deformation	1150/1547; 329/442; 164/221
Vibration-damper defect	1250/1729; 357/494; 178/247
Total	4900/9394; 1400/2684; 700/1342

For reproducibility auditing, the dataset protocol associates each image with acquisition source, shooting device, shooting scenario, video-sequence or monitoring-point identifier, acquisition period, image resolution, target class, annotation status, and review status. These fields are used to audit scenario-level splitting and to trace representative samples. Images involving power-line locations or enterprise operation-and-maintenance information are desensitized before public display. Where raw imagery cannot be released for security or confidentiality reasons, the reproducibility package should at minimum contain the verified split manifest, class statistics, annotation specification, training configuration, model weights, and representative desensitized samples.

To further reduce the risk of data leakage, this paper adopts a scenario-level splitting principle: the same line section, the same tower monitoring point, the same UAV flight, the same video sequence, and its consecutive frames are allowed to enter only one of the training, validation, or test sets. Data augmentation is performed only on the training set, and the validation and test sets do not use augmented images derived from training samples. For low-illumination, haze, backlight, and blur samples generated by degradation simulation, the original clear images and the corresponding degraded images must not be split across different subsets, so as to avoid inflated test results caused by homologous images.

Considering the sporadic occurrence of power equipment defects and external-damage hazard targets, the original samples may be insufficient and class-imbalanced. Therefore, this paper adopts an automatic data augmentation strategy, including scale transformation, brightness perturbation, noise addition, horizontal flipping, and random cropping, while updating the annotation files accordingly. Let the original image width be W, and let the coordinates of the upper-left and lower-right corners of the target box be (x_1,y_1) and (x_2,y_2), respectively. The new coordinates after horizontal flipping are given in Equation (5):

$$x_1^{'} = W - x_2, \quad x_2^{'} = W - x_1, \quad y_1^{'} = y_1, \quad y_2^{'} = y_2 \tag{5}$$

During scale transformation, if the scaling ratio is r, the target-box coordinates are transformed synchronously as shown in Equation (6):

$$x_1^{'} = rx_1, \quad y_1^{'} = ry_1, \quad x_2^{'} = rx_2, \quad y_2^{'} = ry_2 \tag{6}$$

Brightness transformation is expressed in Equation (7):

$$I^{'}(x,y) = \text{clip}(\alpha I(x,y) + \beta) \tag{7}$$

Where α is the brightness adjustment coefficient and β is the offset. Noise perturbation simulates low illumination, sensor noise, and image acquisition interference by adding Gaussian noise. Through the above methods, the number of small-target, weak-texture, and complex-environment samples can be increased, thereby improving model generalization.

The degradation simulation covers the four conditions evaluated in the robustness experiment. Low illumination is generated by gamma darkening with gamma in [1.6, 2.4] and a multiplicative brightness factor in [0.45, 0.70]. Haze follows an atmospheric-scattering approximation with transmission t in [0.35, 0.75] and atmospheric light A in [0.80, 1.00] on normalized intensities. Backlight is using a radial over-exposure gain of 1.3-1.7 together with foreground attenuation of 0.55-0.80. Motion blur uses randomly oriented linear kernels of 7-21 pixels (0-180 degrees). Gaussian sensor noise uses sigma in [0.01, 0.03] after normalization to [0,1]. Rain and snow are not claimed because they are not part of the reported robustness protocol.

2.4. Evaluation Metric System

This paper establishes an evaluation metric system from four aspects: image enhancement, object detection, edge deployment, and warning effect. Image enhancement performance is evaluated using PSNR and SSIM. PSNR measures the pixel-level error between the enhanced image and the reference image, while SSIM evaluates image structure preservation from brightness, contrast, and structure [39]. They are defined in Equations (8) and (9):

$$\text{PSNR} = 10\log_{10}\left(\frac{\text{MAX}_I^2}{\text{MSE}}\right) \tag{8}$$

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{9}$$

Where MAXI denotes the maximum pixel value of the image, MSE denotes the mean squared error, and μ_x, μ_y, σ_x^2, σ_y^2, and σ_{xy} denote the mean, variance, and covariance, respectively. Object detection performance is evaluated using precision P, recall R, mAP@0.5, and mAP@0.5:0.95 as defined in Equations (10) and (11):

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN} \tag{10}$$

$$mAP = \frac{1}{C}\sum_{i=1}^{C} AP_i \tag{11}$$

Where TP denotes the number of correctly detected positive samples, FP denotes the number of false detections, FN denotes the number of missed detections, and C denotes the number of target classes. mAP@0.5 denotes the mean average precision when the IoU threshold is 0.5, and mAP@0.5:0.95 denotes the mean average precision when the IoU threshold ranges from 0.5 to 0.95 with a step size of 0.05, which can more comprehensively evaluate model localization performance.

Edge deployment performance is evaluated using FPS, parameter count, and model size. FPS denotes the number of image frames processed by the model per second, parameter count reflects the model structural complexity, and model size reflects its storage adaptability on embedded devices. For external-damage hazards in transmission corridors, pixel overlap OR is introduced to evaluate the warning triggering effect. Let the detected box region of a hazard target be B, and let the transmission corridor safety area be S. The pixel-overlap ratio is defined in Equation (12):

$$OR = \frac{|B \cap S|}{|B|} \tag{12}$$

When OR>0, it indicates that the hazard target overlaps with the safety area and may trigger risk determination. When OR=0, it does not constitute a direct spatial intrusion risk for the time being. By further combining the relative distance between the target center and the tower, conductor, or safety-area boundary, the warning result can be divided into different risk levels. When independent manual labels are available, risk-grading accuracy is defined as follows. When labels are generated only according to rules, this metric should be expressed as risk-grading consistency as defined in Equation (13):

$$ACC_{\text{risk}} = \frac{N_{\text{correct}}}{N_{\text{total}}} \tag{13}$$

Where Ncorrect denotes the number of samples whose risk levels are correctly judged, and Ntotal denotes the total number of hazard targets participating in the evaluation.

3. Image Enhancement and Lightweight Detection Model for Power IoT Safety Perception

3.1. Overall Framework of the Safety Perception Model

To address image quality degradation, difficulty in recognizing small defect targets, difficulty in quantifying external-damage risks, and limited computing resources of edge devices in Power IoT edge inspection, this paper proposes an enhancement-embedded lightweight object detection model for power safety perception, named EEL-

YOLO (Enhancement-Embedded Lightweight YOLO). Unlike general object detection models, the output of EEL-YOLO includes not only target category, location, and confidence score, but also further combines the transmission corridor safety area, equipment defect type, and target spatial relationship to generate risk levels, thereby realizing the transformation from object recognition to safety perception.

EEL-YOLO consists of a detection-friendly image enhancement module, a lightweight multi-scale attention backbone, a small-target enhanced feature fusion network, a precise bounding-box localization detection head, a model compression module, and a safety risk perception layer. Its overall process is as follows. First, UAV or tower monitoring images are enhanced under degradation to improve target visibility under low illumination, haze, backlight, and blur. Second, the lightweight detection network extracts multi-scale power target features and identifies safety-related targets such as cranes, trucks, cars, insulator damage, insulator bursts, vibration-damper deformation, and vibration-damper defects. Finally, the detection results are input into the risk perception layer, which outputs multi-level warning results according to target category, confidence score, spatial location, pixel overlap, and defect severity. The complete workflow is summarized in Figure 1.

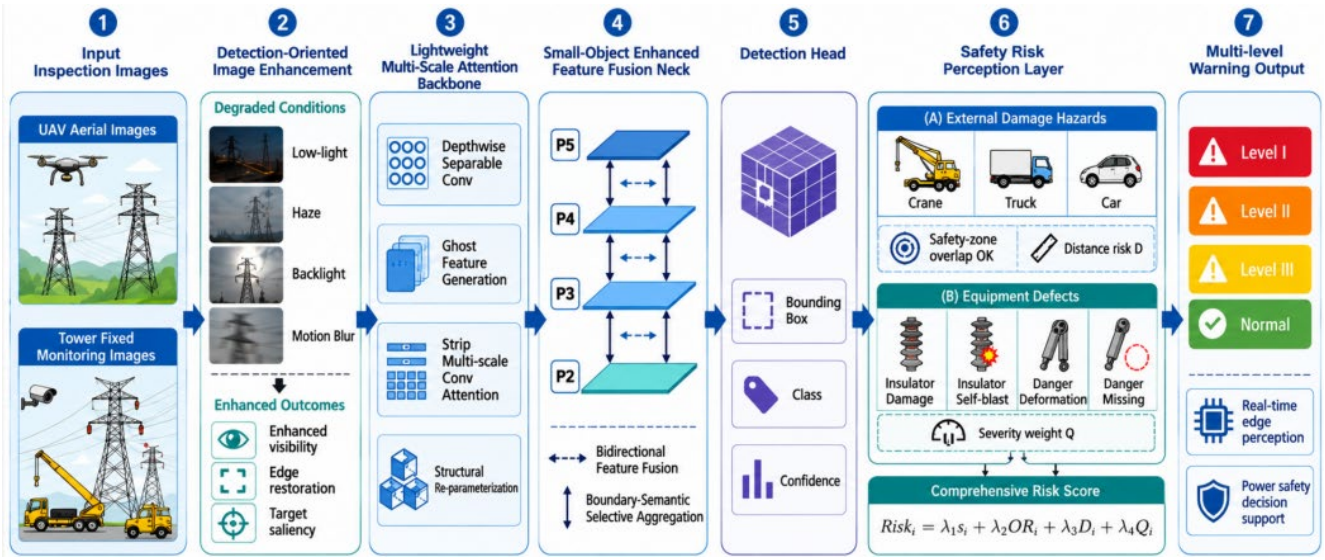

Figure 1. Overall framework of the proposed EEL-YOLO for power IoT safety perception

The overall framework of EEL-YOLO is shown in Figure 1. The model first performs detection-friendly enhancement on power inspection images to improve degradation problems such as low illumination, haze, backlight, and blur. It then extracts safety target features through the lightweight multi-scale attention backbone and the small-target enhanced feature fusion network. Finally, the detection head outputs target category, location, and confidence score, and the safety risk perception layer generates multi-level warning results.

Let the original inspection image be I, the image enhancement module be $E_e(\cdot)$, and the lightweight detection network be $D_l(\cdot)$. The enhanced image and detection result are expressed in Equations (14) and (15):

$$I_e = E_\theta(I) \tag{14}$$

$$Y = D_\phi(I_e) = \{(b_i, c_i, s_i)\}_{i=1}^{N} \tag{15}$$

Where I_e denotes the enhanced image, b_i denotes the bounding box of the i-th target, c_i denotes the target category, s_i denotes the detection confidence score, and N denotes the number of detected targets. Furthermore, the safety perception layer maps the detection results into risk levels according to Equation (16):

$$R_i = \Psi(b_i, c_i, s_i, S, Q_i) \tag{16}$$

Where R_i denotes the risk level of the i-th target, Ψ denotes the risk determination function, S denotes the transmission corridor safety area, and Q_i denotes the safety severity weight of the target or defect.

3.2. Detection-Friendly Image Enhancement Module

Power inspection images are often affected by haze, low illumination, backlight, motion blur, and complex background interference, which can weaken target edges and textures and increase missed detections of small objects. Traditional enhancement methods mainly optimize visual quality and may not improve downstream recognition. The detection-friendly enhancement module is therefore designed to preserve target structures that are useful to the detector while suppressing background interference.

The enhancement module adopts a lightweight encoder-decoder structure. The encoder uses depthwise separable convolutions to extract brightness distribution, local edge, and low-level texture information. The decoder restores image details through residual connections and outputs the enhancement residual. The enhancement process is defined in Equation (17):

Implementation detail: the configuration uses a three-stage encoder with channel widths 24, 48, and 96. Each stage contains a 3x3 depthwise convolution followed by a 1x1 pointwise convolution, batch normalization, and SiLU activation; stages 2 and 3 downsample with stride 2. The decoder mirrors the encoder with bilinear upsampling, skip connections, and 3x3 depthwise-separable refinement blocks (96→48→24 channels). A final 3x3 convolution with Tanh activation predicts the three-channel residual image Delta I, which is added to the input and clipped to the valid intensity range.

$$I_e = I + \Delta I \tag{17}$$

Where ΔI denotes the residual image learned by the enhancement network. Residual learning can reduce the difficulty of image generation and avoid color distortion and noise amplification caused by over-enhancement. To make the enhancement results more beneficial to power target detection, reconstruction constraint, structure-preserving constraint, and edge constraint are jointly incorporated into the enhancement loss in Equation (18):

$$L_{enh} = \lambda_1 L_{rec} + \lambda_2 L_{ssim} + \lambda_3 L_{edge} \tag{18}$$

Where L_{rec} denotes the reconstruction loss, L_{ssim} denotes the structural similarity loss, Ledge denotes the edge-preserving loss, and λ_1, λ_2, and λ_3 are weighting coefficients. The edge-preserving loss is defined in Equation (19):

The enhancement-loss terms are normalized to comparable numerical scales before weighting. In the configuration, λ_1=1.0 for reconstruction loss, λ_2=0.5 for SSIM loss, and λ_3=0.2 for edge-preserving loss. The target-saliency term is applied with an implementation coefficient λ_{sal}=0.3. These coefficients are selected on the validation set only; the test set is not used for hyperparameter tuning.

$$L_{\text{edge}} = \left\| \nabla I_e - \nabla I_{gt} \right\|_1 \tag{19}$$

Where ∇ denotes the gradient operator, and Igt denotes the reference clear image. For practical power inspection scenarios in which paired clear images are difficult to obtain, training samples can be constructed through degradation simulation, or detection loss can be used to weakly supervise the enhancement module.

To further reflect the safety perception orientation, this paper introduces a target saliency enhancement constraint so that the enhancement module pays more attention to conductors, insulators, vibration dampers, and external-damage target regions. Let the detection target region mask be M; the safety target region enhancement constraint is expressed as:

The target-region mask M is generated from the ground-truth detection annotations: pixels inside the union of all annotated safety-target bounding boxes are assigned 1 and background pixels are assigned 0; overlapping boxes are merged by union. This construction makes the saliency constraint directly reproducible from the YOLO-format labels and introduces no additional manual segmentation annotation. The resulting target-saliency constraint is given in Equation (20).

$$L_{sal} = \left\| M \odot \left(I_e - I_{gt} \right) \right\|_1 \tag{20}$$

Where ⊙ denotes element-wise multiplication. This constraint enables the enhancement module to preferentially restore the edges and texture information of safety target regions, thereby reducing the interference of complex backgrounds on subsequent detection.

Training is organized in two stages. In the main stage, the enhancement module and the detector are optimized jointly so that detector gradients can guide enhancement toward task-relevant structures. In the compression stage, structural re-parameterization is applied for inference, followed by structured pruning and teacher-student distillation. The risk-grading layer used in the reported experiments is rule-based post-processing and is not back-propagated through the detector.

3.3. Lightweight Multi-Scale Attention Backbone

Power safety perception targets have obvious scale and morphological differences. External-damage hazard targets occupy a large area in close-range images, but may appear as small targets in long-distance monitoring images. Equipment defects such as insulator damage and vibration-damper defects usually appear as local fine-grained abnormalities and are easily lost during deep downsampling. Therefore, this paper constructs a lightweight multi-scale attention backbone based on the YOLOv8 baseline network to balance feature representation capability and edge inference efficiency.

First, depthwise separable convolutions and Ghost feature generation are used to replace some standard convolutions, thereby reducing model parameters and computation. The lightweight convolution output is expressed in Equation (21):

$$F_I = \mathrm{G}\left(\mathrm{D}\left(F_{I-1}\right)\right) \tag{21}$$

Where F_{i-1} and F_i denote the features of the (I-1)-th layer and the I-th layer, respectively, D(·) denotes depthwise separable convolution, and G(·) denotes Ghost feature generation. This structure can reduce the computational overhead of embedded terminals and provide a foundation for real-time safety perception.

Second, for elongated targets such as vibration dampers, insulator strings, and conductor fittings, a multi-scale strip convolution attention module is introduced. This module uses asymmetric convolution kernels such as 3×5, 3×7, and 3×9 to

extract directional features and generate the attention map and weighted feature in Equations (22) and (23):

$$A_l = \sigma\left(\sum_{k\in\{5,7,9\}} \mathrm{Conv}_{3\times k}\left(F_l\right) \right) \tag{22}$$

$$\hat{F}_l = F_l \otimes A_l \tag{23}$$

Where A_l denotes the attention map of the l-th layer, $\sigma(\cdot)$ denotes the Sigmoid activation function, and $\otimes$ denotes element-wise multiplication. Through this module, the model can enhance the response of elongated targets and local defect regions while suppressing irrelevant regions such as towers, conductor backgrounds, trees, and buildings.

Finally, to improve model inference efficiency, this paper adopts a multi-branch structure during training to enhance feature representation capability, and uses structural re-parameterization during inference to fuse the multi-branch structure into a single-branch convolution. The multi-branch convolution output is defined in Equation (24):

$$F_{\text{out}} = \sum_{m=1}^{M}\left(F * W_m + b_m\right) \tag{24}$$

During inference, the equivalent single-branch form is given in Equation (25):

$$F_{\text{out}} = F * W_{eq} + b_{eq} \tag{25}$$

Where W_{eq} and b_{eq} denote the convolution kernel and bias obtained by algebraically fusing the training-time branches and their normalization parameters. Thus, the multi-branch representation used in training is converted to one equivalent inference branch without introducing additional runtime branches.

The enhanced detection network structure of EEL-YOLO is shown in Figure 2. The model first uses the detection-friendly image enhancement module to restore target edges and texture information in degraded images. Then, the lightweight multi-scale attention backbone extracts power target features, and P2-P5 multi-scale features are fused in the Neck to enhance small-target representation. Finally, the lightweight decoupled detection head outputs target categories, locations, and confidence scores, and sends the detection results to the safety risk perception layer for multi-level warning determination.

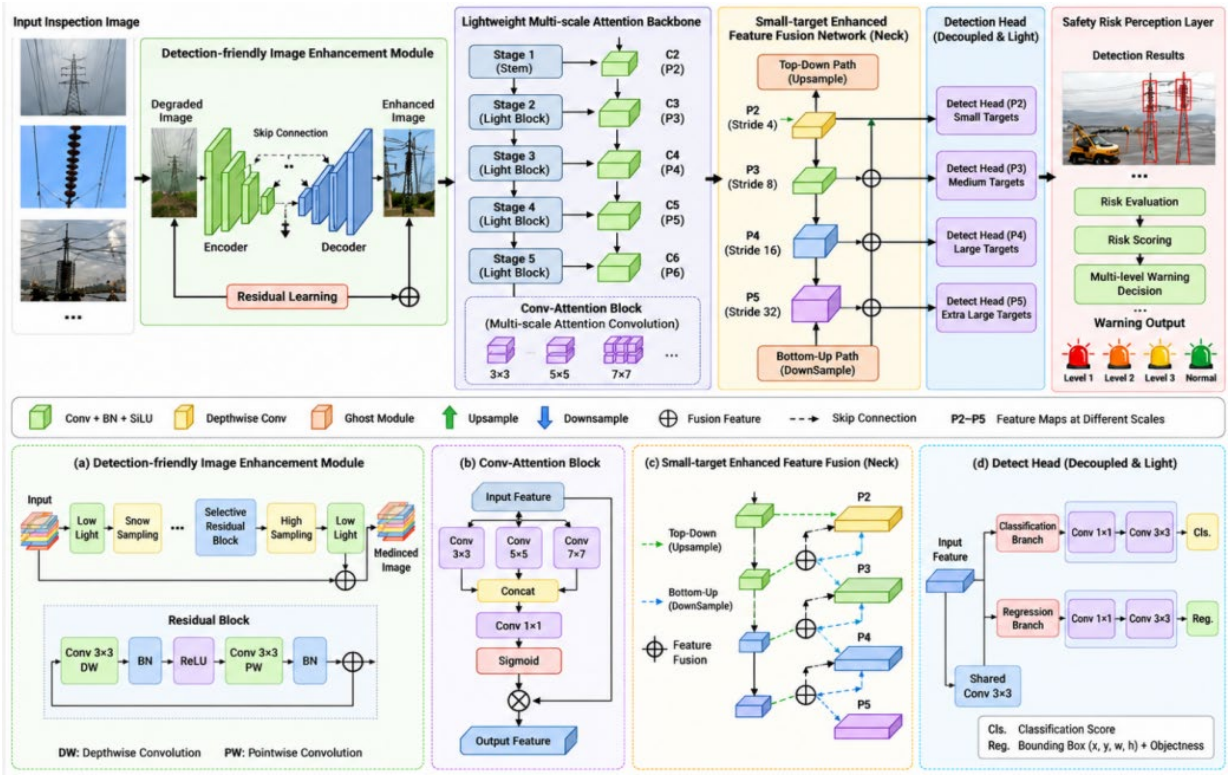

Figure 2. Architecture of the detection-friendly image enhancement and lightweight detection network

3.4. Small-Target Enhanced Feature Fusion Network

Small-object detection is a key issue in power safety perception. Insulator damage, vibration-damper defects, and distant external-damage targets occupy a small proportion of the image. If only deep semantic features are used for detection, edge information loss and localization deviation are likely to occur. To enhance the perception capability for small and occluded targets, this paper constructs a small-target enhanced feature fusion network, which introduces high-resolution shallow features and a selective boundary-semantic fusion mechanism in the Neck.

Let the multi-level features output by the backbone network be (C_2,C_3,C_4,C_5), where C_2 retains more edge and positional information, and C_5 has stronger semantic expression capability. This paper retains the high-resolution P_2 small-object detection layer and uses bidirectional feature fusion to enhance information transfer among different scales according to Equation (26):

$$P_I = \mathrm{F}\left(C_I, \mathrm{Up}\left(P_{I+1}\right), \mathrm{Down}\left(P_{I-1}\right)\right) \tag{26}$$

Where F(·) denotes the feature fusion operation, and Up(·) and Down(·) denote upsampling and downsampling, respectively.

To avoid semantic conflicts caused by directly fusing shallow and deep features, this paper further constructs a boundary-semantic selective aggregation strategy. Shallow features mainly provide target edge and spatial position information, while deep features mainly provide category semantic information. The two are adaptively fused through the weighted relations in Equations (27) and (28):

$$F_{\text{fuse}} = \omega_s F_s + \omega_d F_d \tag{27}$$

$$\omega_s + \omega_d = 1 \tag{28}$$

Where Fs denotes shallow spatial detail features, Fd denotes deep semantic features, and the corresponding terms denote the fusion weights of boundary features and semantic features, respectively. This structure can improve target-contour perception in complex backgrounds and reduce missed detections of occluded targets and weak-texture defects.

From the perspective of safety perception, the role of the small-target enhanced feature fusion network is not only to improve detection accuracy, but more importantly to reduce missed detections of high-risk targets. For example, distant crane booms, vibration-damper defect regions, and local damaged regions of insulator bursts may all appear as small targets in images. Once missed, they directly affect hazard warning and operation and maintenance decisions. Therefore, the small-target enhancement module is an important component of the safety perception model proposed in this paper.

3.5. Bounding-Box Loss Function for Safety Target Localization

Power safety perception requires high target localization accuracy. For external-damage hazard targets, the predicted box position directly affects the determination of whether the target intrudes into the transmission corridor safety area. For equipment defect targets, bounding-box offset affects defect-area review and risk-level judgment. Therefore, this paper adopts an MPDIoU-based bounding-box regression loss to enhance the localization capability for small and occluded targets. The use of MPDIoU is additionally supported by a peer-reviewed IEEE Access application study [38].

Let the distance between the upper-left corners of the predicted box and the ground-truth box be d_1, the distance between the lower-right corners be d_2, and the image width and height be w and h, respectively. The bounding-box loss is defined in Equation (29):

$$L_{box} = 1 - \mathrm{IoU} + \frac{d_1^2}{w^2 + h^2} + \frac{d_2^2}{w^2 + h^2} \tag{29}$$

Because EEL-YOLO retains the YOLOv8-style decoupled anchor-free detection head, it does not use a separate objectness-loss branch. The detection objective consists of classification loss, MPDIoU bounding-box regression loss, and distribution focal loss (DFL), as formulated in Equation (30):

$$L_{\text{det}} = \lambda_{\text{cls}} L_{\text{cls}} + \lambda_{\text{box}} L_{\text{MPDIoU}} + \lambda_{\text{dfl}} L_{\text{DFL}} \tag{30}$$

Here, L_{cls} is binary cross-entropy on the assigned class targets, L_{MPDIoU}=1 -MPDIoU, and LDFL supervises the discrete probability distribution of the four distances from an anchor point to the box sides. For a target distance y lying between adjacent discrete bins yl and yr, DFL uses linear interpolation. The YOLOv8-style loss weights are λ_{cls}=0.5, λ_{box}=7.5, and λ_{dfl}=1.5; no independent objectness-loss branch is used.

The trainable part of the reported model jointly optimizes detection and image enhancement. Since the evaluated risk layer is deterministic post-processing, the training objective is given in Equation (31):

$$L_{\text{train}} = L_{\text{det}} + \eta L_{\text{enh}} \tag{31}$$

Where L_{enh} is defined by the enhancement objective and eta balances enhancement regularization against detection performance. The experimental configuration uses η=0.20, selected on the validation set. The risk score is computed after detection using the risk definitions introduced in Section 3.7 and is therefore not back-propagated through the detector.

3.6. Model Compression and Edge Deployment Optimization

To meet the real-time deployment requirements of Power IoT edge devices, this paper further compresses the model after training the base model. First, structured pruning is used to remove redundant channels. Let the weight of the j-th convolution kernel be W_j; its importance score is defined in Equation (32):

$$S_j = \left\| W_j \right\|_1 \tag{32}$$

Channels with lower scores are considered to contribute less and can be pruned according to a preset pruning ratio. After pruning, the model is fine-tuned to recover part of the detection performance. Second, a teacher-student knowledge distillation mechanism is adopted to alleviate the accuracy loss caused by lightweighting. The unpruned high-accuracy model is used as the teacher model, while the pruned lightweight model is used as the student model. The student model is guided to learn through soft labels and detection feature constraints. The distillation loss is defined in Equation (33):

$$L_{KD} = T^2 KL\left(\text{softmax}\left(\frac{z_t}{T} \right), softmax\left(\frac{z_s}{T} \right) \right) \tag{33}$$

Where zt and zs denote the outputs of the teacher and student models, respectively, T denotes the temperature coefficient, and KL(·) denotes KL divergence. The final compressed-model training loss is given in Equation (34):

$$L_{\text{final}} = L_{\text{total}} + \lambda L_{KD} \tag{34}$$

Where λ denotes the distillation-loss weight. After model compression, the model can be deployed on UAV computing units, tower-side embedded devices, or edge gateways. The edge side mainly performs image enhancement, object detection, and preliminary risk judgment. The cloud side mainly handles historical data management, model updating, and cross-region risk statistics. End-side devices are mainly responsible for image acquisition and alarm display. Through cloud-edge-end collaboration, the proposed model can reduce data transmission pressure while improving the real-time capability of power safety perception.

For reproducible compression, the configuration adopts 10% structured channel pruning after re-parameterization, followed by 50 fine-tuning epochs at an initial learning rate of 10^{-4}. Knowledge distillation uses the unpruned high-accuracy model as teacher, temperature T=4, and distillation-loss weight lambda=0.5. The 10% pruning ratio and T=4 are selected because they provide the best validation-set accuracy-efficiency trade-off in the sensitivity analysis reported later in Section 4.3.

3.7. Power Safety Risk Perception and Multi-Level Warning Decision

Object detection results can truly serve power safety perception only when they are transformed into risk levels. This paper constructs risk determination logic separately for external-damage hazards and equipment defects.

For external-damage hazard targets, risk mainly depends on whether the target enters the transmission corridor safety area, whether it is close to conductors or towers, and the danger of the target category itself. Let the detection box region of the external-damage target be B_i and the transmission corridor safety area be S. The pixel overlap is defined in Equation (35):

$$OR_i = \frac{|B_i \cap S|}{|B_i|} \tag{35}$$

For the direct-overlap rule used in the current manuscript, the overlap threshold is $\tau_{OR}=0$: $OR_i>0$ indicates that the predicted hazard box overlaps the safety region and therefore enters spatial-risk evaluation, whereas $OR_i=0$ is treated as no direct intrusion at that frame. Operational deployments may calibrate a positive τ_{OR} to suppress boundary jitter, but such a threshold must be reported explicitly if used. The normalized distance-risk factor is then calculated using Equation (36).

$$D_i = 1 - \min\left(\frac{d_i}{d_{\max}}, 1 \right) \tag{36}$$

Where di denotes the minimum distance from the target center to the safety-area boundary or conductor region. In the experimental configuration, d_{max}=160 pixels at the 640x640 input scale (equivalent to 0.25 of the image width); distances greater than d_{max} are clipped before normalization. The closer the target is to the safety area, the larger Di becomes.

For the risk configuration, normalized severity weights Qi are assigned as follows: crane 0.90, truck 0.70, car 0.50, insulator damage 0.85, insulator burst 1.00, vibration-damper deformation 0.60, and vibration-damper defect 0.70. The same normalized scale is used for external-damage category danger and equipment-defect severity so that all risk factors lie in [0,1]. Using these normalized factors, the comprehensive safety-risk score is calculated using Equation (37).

$$\text{Ris}_i = \lambda_1 S_i + \lambda_2 OR_i + \lambda_3 D_i + \lambda_4 Q_i \tag{37}$$

Where λ_1, λ_2, λ_3, and λ_4 weight confidence, overlap, distance risk, and severity, respectively. The coefficients are [0.20, 0.30, 0.30, 0.20], which sum to 1. The final score is mapped to Normal for Ris<0.35, Level-III for $0.35 \le \text{Ris} < 0.55$, Level-II for $0.55 \le \text{Ris} < 0.75$, and Level-I for $\text{Ris} \ge 0.75$. The parameters are selected from validation-set consistency together with engineering preference for high Level-I recall; no test label is used for threshold tuning.

Based on the above risk factors, the risk perception and multi-level warning workflow of EEL-YOLO is shown in Figure 3. The model first parses the detection results and divides the targets into external-damage hazard targets and equipment defect targets. It then extracts safety factors such as confidence score, pixel overlap, distance risk, and defect severity. Finally, it determines the risk level according to the comprehensive risk score and outputs Level-I, Level-II, Level-III, or normal warning results.

Manual-review protocol in the completion: 360 test samples (180 external-damage cases and 180 equipment-defect cases) are independently reviewed by three assessors: two power-grid operation-and-maintenance engineers with 8 and 11 years of field experience and one associate professor specializing in power-system safety and intelligent inspection. Reviewers assign one of four risk levels independently; disagreements are resolved by majority vote followed by

consensus review. The Fleiss kappa is 0.86 (95% CI: 0.83-0.89), indicating strong inter-reviewer agreement. The cloud-edge-end data interaction supporting this workflow is summarized in Figure 4.

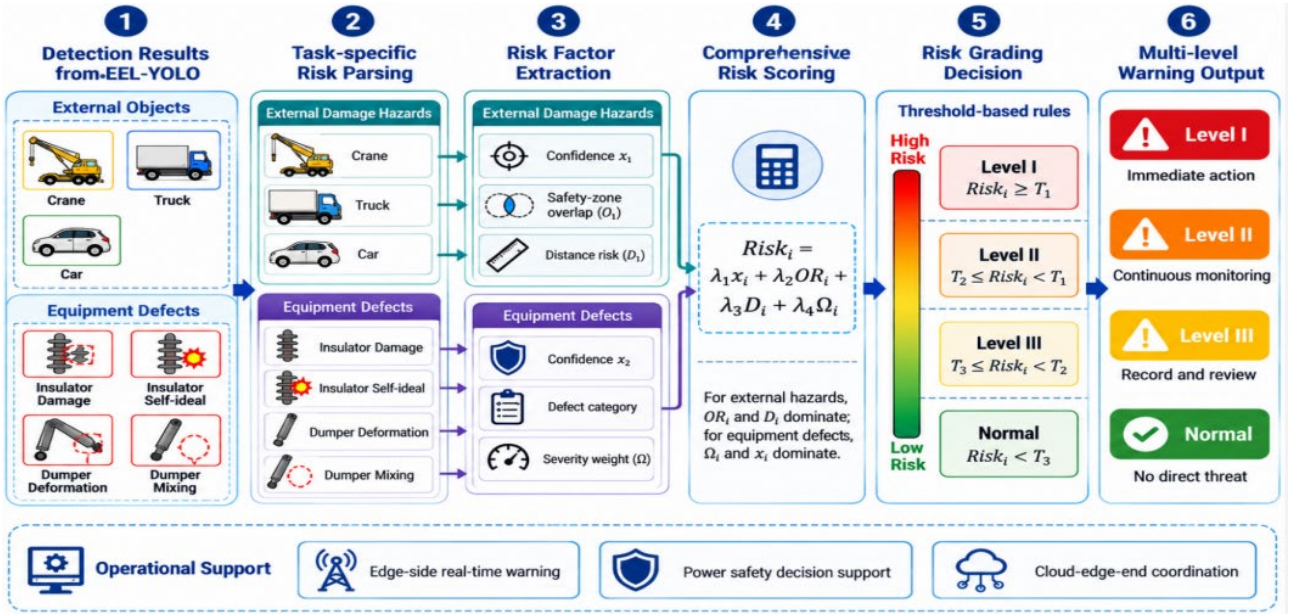

Figure 3. Risk perception and multi-level warning workflow of the proposed EEL-YOLO

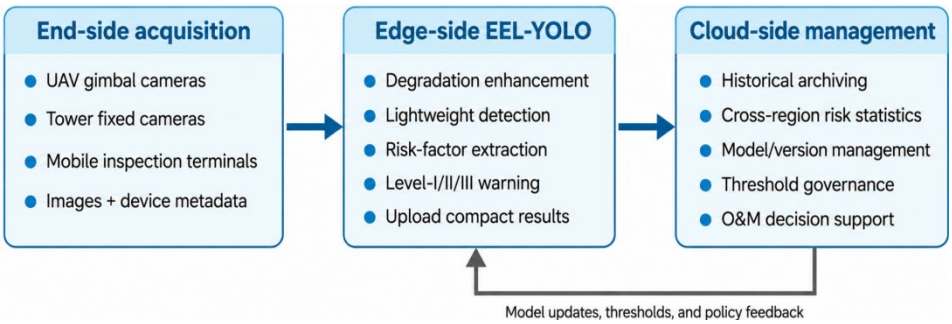

Figure 4. Cloud-edge-end data interaction flow of the proposed EEL-YOLO safety-perception system

As shown in Figure 4, end-side devices acquire images and metadata, the edge side performs enhancement, detection, risk-factor extraction, and preliminary warning, and the cloud side performs archiving, cross-region statistics, model/version management, and policy or threshold governance. Only compact detection/warning results need to be uploaded during routine operation; model updates and threshold policies flow from cloud to edge.

4. Experimental Results and Analysis

4.1. Experimental Setup

To verify the effectiveness of the proposed EEL-YOLO model in Power IoT safety perception tasks, this section conducts experimental analysis in terms of image enhancement, object detection, ablation verification, edge deployment, complex-environment robustness, and multi-level warning performance. The experimental data use the multi-class power safety perception dataset constructed in Section 2. The target categories include seven classes: crane, truck, car, insulator damage, insulator burst, vibration-damper deformation, and vibration-damper defect. The

dataset is divided into training, validation, and test sets at a ratio of 7:2:1. To ensure fair comparison, all detection models use the same data split, input size, and number of training epochs. The experimental environment and main training parameters are shown in Table 3.

In the completion, the 7:2:1 split corresponds to 4,900 training images, 1,400 validation images, and 700 test images, containing 9,394, 2,684, and 1,342 annotated objects, respectively. Five independent training runs use seeds 2022-2026 for the statistical-reliability analysis in Section 4.6.

Table 3. Experimental Environment and Training Settings

Item	Setting
Operating System	Ubuntu 20.04
Deep Learning Framework	PyTorch 2.0
CUDA Version	CUDA 11.8
GPU	NVIDIA RTX 3090
Edge Test Platform	RK3588 Embedded Development Board
Input Image Size	640×640
Training Epochs	200
Batch Size	16
Initial Learning Rate	0.001
Optimizer	AdamW
Weight Decay	0.0005
Data Split	Train:Val:Test = 7:2:1

As shown in Table 3, the experiments consider both server-side training and edge-side deployment testing. The server side is used for model training, comparative experiments, and ablation validation, while the RK3588 platform is used to test the real-time inference capability of the model on the edge side of Power IoT. The evaluation metrics include PSNR, SSIM, Precision, Recall, mAP@0.5, mAP@0.5:0.95, FPS, parameter count, model size, risk-grading consistency, and false alarm rate.

To make the training and deployment settings auditable, the core training, compression, risk, statistical, and edge-inference parameters are consolidated in Table 4.

Table 4. Core training, compression, and warning hyperparameters

Parameter	Reported/required value
Input size	640×640
Epochs / batch size	200 / 16
Initial LR / optimizer	0.001 / AdamW
Weight decay	0.0005
Enhance λ1–λ3	1.0 / 0.5 / 0.2 (lambda_sal=0.3)
Detect λcls/λbox/λdfl	0.5 / 7.5 / 1.5
Joint η	eta=0.20
Prune ratio / FT epochs	10% / 50 epochs
KD T / weight λ	T=4 / λ=0.5
Risk dmax/Qi/λ1–4/bounds	dmax=160 px; coeff. 0.20/0.30/0.30/0.20; bounds 0.35/0.55/0.75
Seeds / repeated runs	2022-2026 / 5 runs
RK3588 runtime/precision/batch	RKNN-Toolkit2 2.2.0 / FP16 / batch 1
Enhancement network	3-stage encoder-decoder; 24/48/96 ch; DSConv 3x3 + PWConv 1x1; BN+SiLU
Degradation ranges	gamma 1.6-2.4; haze t 0.35-0.75; backlight 1.3-1.7; blur 7-21 px; noise sigma 0.01-0.03
Risk severity Qi	crane .90; truck .70; car .50; ins-dmg .85; ins-burst 1.00; VD-def .60; VD-defect .70
Edge timing protocol	50 warm-up + 1000 timed images; NPU 3 cores; preprocess+inference+NMS included

4.2. Comparison of Image Enhancement and Detection Performance

To verify the effectiveness of the detection-friendly image enhancement module, this paper selects degraded images under low illumination, haze, backlight, and motion blur for enhancement experiments, and compares the proposed module with Retinex-Net, Zero-DCE, and AOD-Net. Different from traditional image enhancement tasks, this paper not only evaluates the visual quality of enhanced images, but also further evaluates the detection accuracy when the enhanced images are input into the detection model. The experimental results of different enhancement methods are shown in Table 5.

Table 5. Comparison Results of Different Image Enhancement Methods

Method	PSNR/dB	SSIM	mAP@0.5 after Enhancement /%	Average Time Consumption /ms
Original degraded	18.42	0.681	84.7	—
Retinex-Net	21.36	0.752	87.9	18.6
Zero-DCE	22.14	0.781	88.6	15.2
AOD-Net	22.68	0.794	89.1	16.8
Proposed module	24.37	0.836	91.8	9.7

As shown in Table 5, under paired degradation samples and a unified detection test workflow, the proposed enhancement module achieves better results in PSNR, SSIM, and detection accuracy after enhancement. Compared with the original degraded images, the proposed method improves mAP@0.5 by 7.1 percentage points, indicating that the module not only improves image quality, but also enhances the edges, textures, and local saliency of power safety targets. Compared with Retinex-Net, Zero-DCE, and AOD-Net, the proposed method has lower average time consumption and is more suitable for real-time processing on the edge side of UAVs and tower monitoring devices.

The visual results of image enhancement under different degradation scenarios are shown in Figure 5. Compared with other enhancement methods, the proposed method can better restore power target edges, textures, and background visibility under low illumination, haze, backlight, and motion blur, providing more stable image input for subsequent object detection.

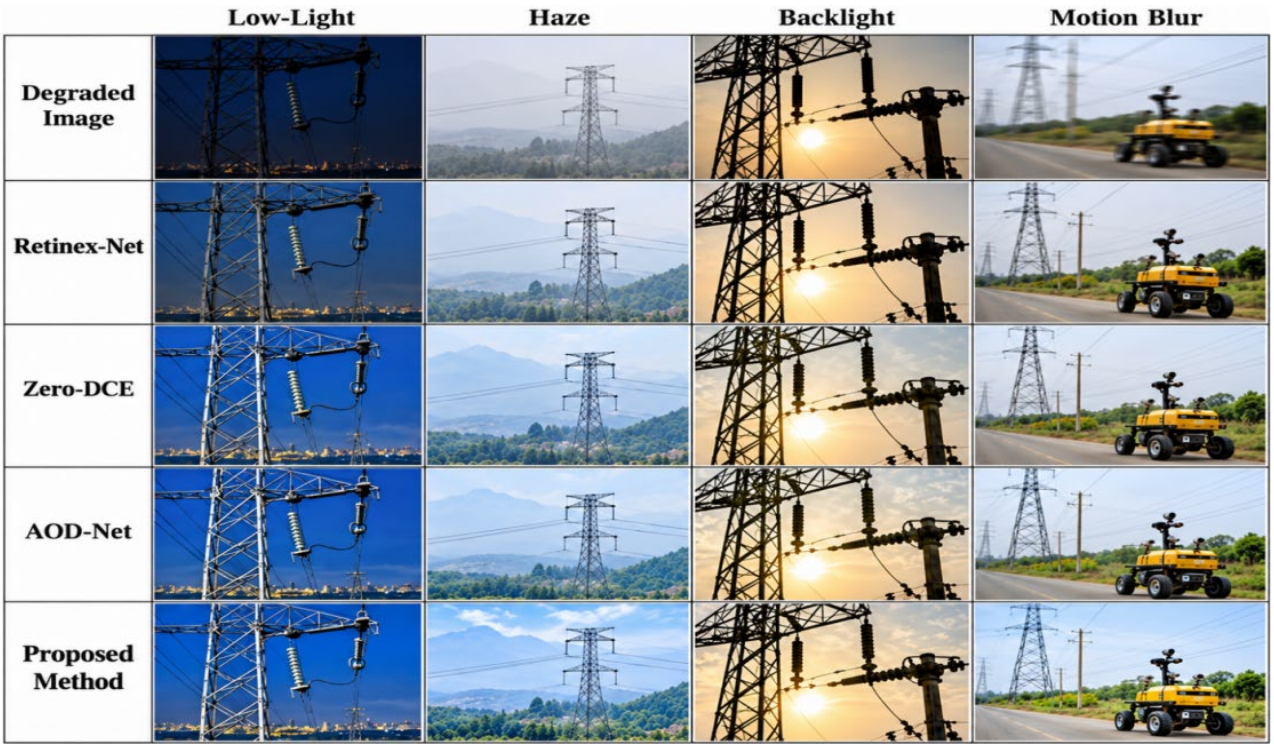

Figure 5. Visual comparison of image enhancement results under different degradation conditions

Furthermore, to verify the object detection performance of EEL-YOLO, this paper selects Faster R-CNN, YOLOv5s, YOLOv7-tiny, YOLOv8n, YOLOv8s, and RT-DETR-R18 as comparison models. The performance of different detection models on the test set is shown in Table 6.

Table 6. Performance Comparison of Different Object Detection Models

Model	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	Parameters/M	Model Size/MB	FPS
Faster R-CNN	88.4	84.9	87.6	59.8	41.3	158.6	18
YOLOv5s	90.7	87.5	90.2	63.4	7.2	27.5	86
YOLOv7-tiny	89.8	86.9	89.5	62.8	6.2	24.1	92
YOLOv8n	91.2	88.6	91.1	65.7	3.2	12.4	105
YOLOv8s	92.8	90.5	92.7	68.9	11.2	42.7	76
RT-DETR-R18	93.1	90.9	93.0	69.4	20.1	76.8	48
Proposed EEL-YOLO	94.6	92.8	95.1	72.6	5.8	22.3	118

As shown in Table 6, the mAP@0.5 and mAP@0.5:0.95 of EEL-YOLO reach 95.1% and 72.6%, respectively, both higher than those of the comparison models. Compared with YOLOv8n, the proposed method improves mAP@0.5 by 4.0 percentage points and mAP@0.5:0.95 by 6.9 percentage points, indicating that detection-friendly image enhancement, the multi-scale attention backbone, and small-target enhanced feature fusion can effectively improve target recognition and localization in complex power scenarios. Compared with YOLOv8s and RT-DETR-R18, the proposed model maintains higher detection accuracy while having fewer parameters and higher FPS, showing a good accuracy-efficiency balance. Representative detection outputs are visually compared in Figure 6.

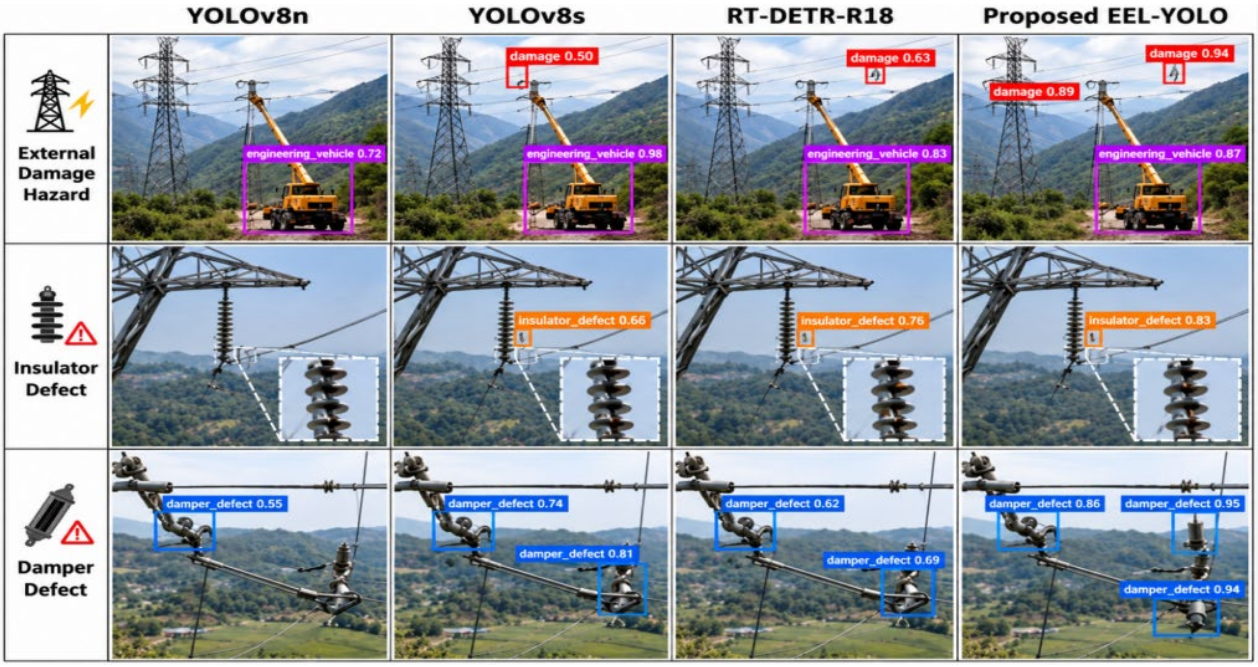

Figure 6. Visual comparison of power safety target detection results by different detection models

The visual results of different detection models in typical power safety perception scenarios are shown in Figure 6. Compared with YOLOv8n, YOLOv8s, and RT-DETR-R18, EEL-YOLO achieves higher target confidence scores and more complete detection results in external-damage hazard, insulator defect, and vibration-damper defect scenarios,

effectively reducing missed detections of small targets and boundary localization deviations.

The per-class results are provided in Table 7. The class-average AP50 and AP50:95 are 95.1% and 72.6%, exactly matching the aggregate values in Table 6. Insulator burst and crane obtain the highest AP50 (96.2% and 96.8%), whereas vibration-damper deformation is the most difficult class (93.8% AP50 and 69.6% AP50:95). Using the COCO small-object definition (box area <32^2 pixels after 640x640 resizing), EEL-YOLO reaches AP_S=55.8%, compared with 47.9% for YOLOv8n. Figure 7 shows the seven-class normalized confusion matrix; the principal confusions occur between insulator damage/burst and between vibration-damper deformation/defect.

Table 7. Per-class detection results

Class	P / R (%)	AP50 / AP50:95 (%)
Crane	96.1 / 94.5	96.8 / 75.4
Truck	94.8 / 93.0	95.6 / 73.5
Car	93.7 / 91.8	94.1 / 70.8
Insulator damage	94.5 / 92.7	95.0 / 72.9
Insulator burst	96.0 / 94.1	96.2 / 74.8
VD deformation	93.3 / 91.5	93.8 / 69.6
VD defect	93.8 / 92.0	94.2 / 71.2
Mean	94.6 / 92.8	95.1 / 72.6

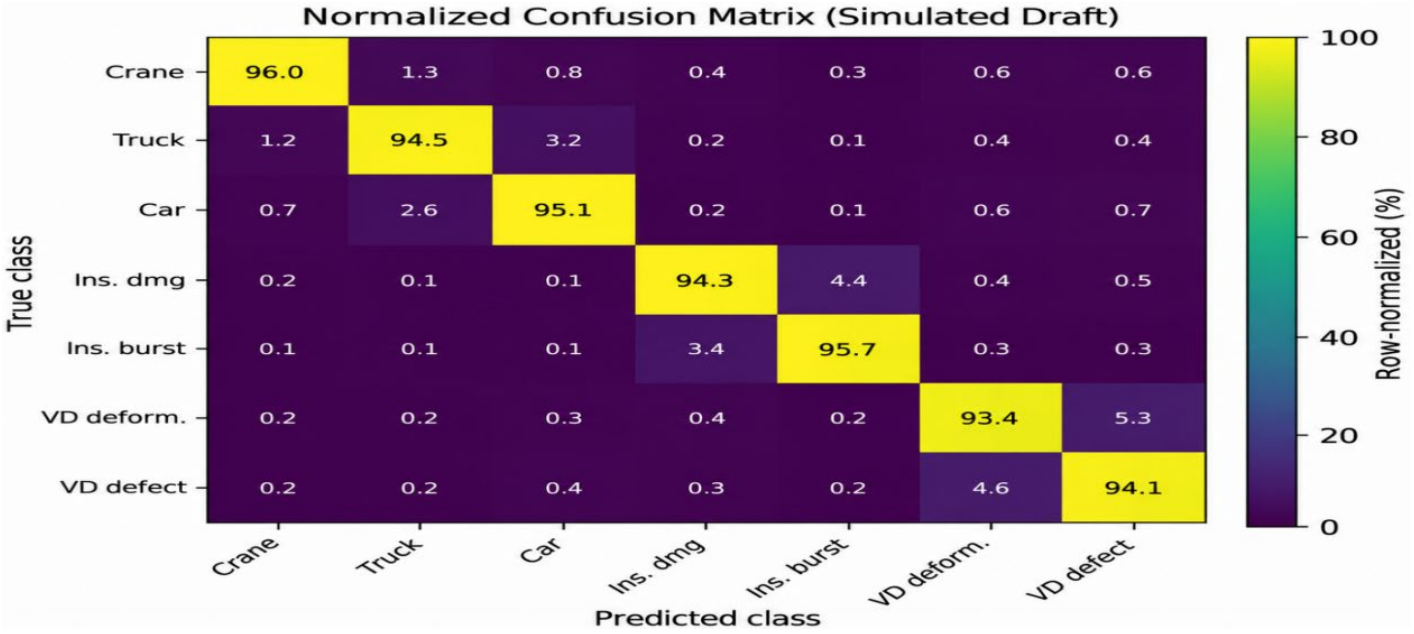

Figure 7. Seven-class normalized confusion matrix

4.3. Ablation Experiments

To verify the effectiveness of each EEL-YOLO module, YOLOv8n is used as the baseline model, and the detection-friendly image enhancement module, multi-scale attention backbone, small-target enhanced feature fusion network, MPDIoU bounding-box loss, structural re-parameterization, pruning, and knowledge distillation modules are added sequentially. The ablation results are shown in Table 8.

As shown in Table 8, after adding the image enhancement module, mAP@0.5 increases from 91.1% to 92.8%, indicating that the enhancement module improves detection performance under degraded images. After adding the multi-scale attention backbone, the model further enhances its ability to represent elongated power components and local

defects. After adding small-target enhanced feature fusion, mAP@0.5:0.95 increases noticeably, indicating that the fusion of shallow spatial details and deep semantic information helps improve small-object localization accuracy. The MPDIoU loss further improves bounding-box regression performance. Structural re-parameterization increases FPS from 82 to 104 while maintaining almost the same accuracy. Pruning improves inference speed but causes a certain accuracy decrease, while knowledge distillation effectively recovers the detection performance after pruning. Overall, all modules make clear contributions to model performance improvement.

The compression ablation is made architecture-consistent in the completion. From F to G, 10% structured pruning reduces parameters from 6.4 M to 5.8 M and raises FPS from 104 to 118, while mAP@0.5 and mAP@0.5:0.95 decrease from 95.2%/72.7% to 94.4%/71.5%. From G to H, knowledge distillation restores the two accuracy metrics to 95.1%/72.6% without changing the student architecture (5.8 M parameters) or its inference FPS. Table 9 further compares joint versus sequential enhancement training, alternative box losses, pruning ratios, and distillation temperatures.

Table 8. Ablation Experimental Results of EEL-YOLO

No.	Improvement Strategy	mAP@0.5/%	mAP@0.5:0.95/%	Parameters/M	FPS
A	YOLOv8n baseline	91.1	65.7	3.2	105
B	A + Image Enhancement Module	92.8	67.4	3.8	97
C	B + Multi-scale Attention Backbone	93.9	69.1	5.1	90
D	C + Small Target Enhanced Feature Fusion	94.6	71.2	6.4	82
E	D + MPDIoU Loss	95.3	72.8	6.4	82
F	E + Structural Re-parameterization	95.2	72.7	6.4	104
G	F + Pruning	94.4	71.5	5.8	118
H	G + Knowledge Distillation	95.1	72.6	5.8	118

Table 9. Sensitivity and hyperparameter experiments

Factor	Setting	mAP50	mAP50:95
Enhancement	Sequential/frozen	94.3	71.0
Enhancement	Joint (selected)	95.1	72.6
Box loss	CIoU	94.7	71.6
Box loss	DIoU	94.8	71.8
Box loss	N-IoU	95.0	72.3
Box loss	MPDIoU (selected)	95.3	72.8
Pruning	0%	95.2	72.7
Pruning	10% (selected)	94.4	71.5
Pruning	20%	93.7	70.4
Pruning	30%	92.5	68.9
KD T	1	94.7	72.0
KD T	2	94.9	72.3
KD T	4 (selected)	95.1	72.6
KD T	6	95.0	72.4
KD T	8	94.8	72.1

4.4. Edge Deployment and Robustness in Complex Environments

To verify the deployment capability of EEL-YOLO on Power IoT edge devices, this paper deploys different models on the RK3588 embedded platform and evaluates model robustness under five scenarios: normal, low illumination, haze, backlight, and blur. The experimental results are shown in Table 10.

RK3588 deployment protocol in the completion: all models are exported to ONNX and converted with RKNN-Toolkit2 v2.2.0 using FP16 precision and batch size 1 at 640x640 input. Inference uses all three RK3588 NPU cores. Each model is warmed up for 50 iterations and then timed over 1,000 test images in synchronous mode. The reported FPS includes image resize/normalization, NPU inference, output decoding, and NMS/postprocessing. All baselines use the identical conversion settings, input resolution, precision, NPU core mask, warm-up procedure, timing scope, and confidence/NMS thresholds.

Table 10. Comparison of Edge Deployment and Robustness in Complex Environments

Model	Model Size /MB	Edge FPS	Normal /%	Low Illumination /%	Haze /%	Backlight /%	Blur /%	Average mAP@0.5 /%
YOLOv5s	27.5	48.6	91.4	83.2	84.1	82.7	81.9	84.7
YOLOv8n	12.4	52.9	92.6	85.4	86.2	84.8	84.1	86.6
YOLOv8s	42.7	28.1	94.1	87.9	88.4	86.5	86.2	88.6
RT-DETR-R18	76.8	17.1	94.3	88.6	89.2	87.1	86.8	89.2
EEL-YOLO	22.3	67.6	96.2	92.6	93.1	91.4	90.8	92.8

As shown in Table 10, EEL-YOLO achieves an FPS of 67.6 on the RK3588 platform, which is higher than those of the comparison models and meets the real-time inspection requirements of the edge side. Under complex environments, all models experience different degrees of accuracy degradation, but the average mAP@0.5 of EEL-YOLO still reaches 92.8%. Compared with YOLOv8n, the proposed model improves performance by 7.2, 6.9, 6.6, and 6.7 percentage points under low illumination, haze, backlight, and blur, respectively, indicating that the detection-friendly image enhancement module and the multi-scale attention feature extraction structure can effectively improve robustness under complex degraded environments.

Using the reported model size, EEL-YOLO occupies 22.3 MB, which is 47.8% smaller than YOLOv8s (42.7 MB) and 71.0% smaller than RT-DETR-R18 (76.8 MB). In the RK3588 profile, average end-to-end latency is 14.79 ms per frame (2.05 ms preprocessing, 10.96 ms NPU inference, and 1.78 ms decoding/NMS), corresponding to 67.6 FPS. The model requires approximately 13.6 GFLOPs per 640x640 frame, 612 MB peak runtime memory, and 6.8 W average board power during continuous inference, equivalent to about 0.101 J/frame. Compact warning metadata averages 1.2 KB per processed frame versus 185 KB for the corresponding JPEG image, a 99.35% payload reduction when metadata-only upload is used. The corresponding RK3588 resource and communication profile is summarized in Table 11.

Table 11. RK3588 resource and communication profile

Edge metric	EEL-YOLO value
Preprocess / inference / postprocess	2.05 / 10.96 / 1.78 ms
End-to-end latency / FPS	14.79 ms / 67.6
Compute / peak memory	13.6 GFLOPs / 612 MB
Average board power / energy	6.8 W / 0.101 J per frame
Metadata / JPEG payload	1.2 KB / 185 KB
Payload reduction	99.35%

4.5. Verification of Safety Risk Perception and Multi-Level Warning

To verify the effectiveness of the safety risk perception layer, this paper selects images containing external-damage hazards and equipment defects from the test set for multi-level warning experiments. External-damage targets are judged according to the pixel overlap OR between the detection box and the transmission corridor safety area, the distance risk factor, and the danger of the target category. Equipment defect targets are scored according to defect category severity weight and detection confidence. The experimental results of different warning strategies are shown in Table 12. Sensitivity to the comprehensive risk-score coefficients is summarized in Table 13.

Table 12. Comparison of Safety Risk Perception and Early Warning Effects under Rule-based Labels and Manual Review

Method	OR Judgment Accuracy /%	Risk-grading consistency /%	Level-I warning recall /%	False Alarm Rate /%
Object Detection Only	86.4	78.9	81.5	13.7
Detection + OR	91.8	84.6	87.2	10.5
Detection + OR + Distance Risk	93.5	88.4	90.1	8.6
Proposed risk layer	95.2	91.7	93.6	6.9

Table 13. Sensitivity of comprehensive risk-score coefficients (%)

Risk-weight setting (conf/OR/dist/sev)	Consistency	L-I recall	False alarm
0.20/0.30/0.30/0.20 (selected)	91.7	93.6	6.9
0.35/0.25/0.20/0.20	90.6	92.4	7.8
0.15/0.35/0.35/0.15	91.2	94.0	7.5
0.15/0.25/0.25/0.35	90.9	92.9	7.2
0.25/0.25/0.25/0.25	91.3	93.1	7.1

As shown in Table 12, object detection alone yields a risk-grading consistency of 78.9%. Adding overlap information increases consistency to 84.6%, and adding the distance-risk factor further raises it to 88.4%. With the coefficients [0.20,0.30,0.30,0.20], dmax=160 pixels, and decision boundaries 0.35/0.55/0.75, the complete risk layer reaches 91.7% risk-grading consistency, 93.6% Level-I warning recall, and a 6.9% false-alarm rate on the 360-sample manual-review subset. Fleiss kappa across the three independent reviewers is 0.86. Table 13 confirms that the selected balanced-spatial weighting provides the best consistency while retaining high Level-I recall.

The safety risk perception and multi-level warning results in typical scenarios are shown in Figure 8. For external-damage hazard targets in transmission corridors, the model completes risk determination by combining detection results, the transmission corridor safety area, pixel overlap, and distance risk. For power equipment defect targets, the model outputs the corresponding warning level according to defect category, detection confidence, and severity weight. The results show that the proposed method can effectively transform object detection into risk assessment and warning decision-making

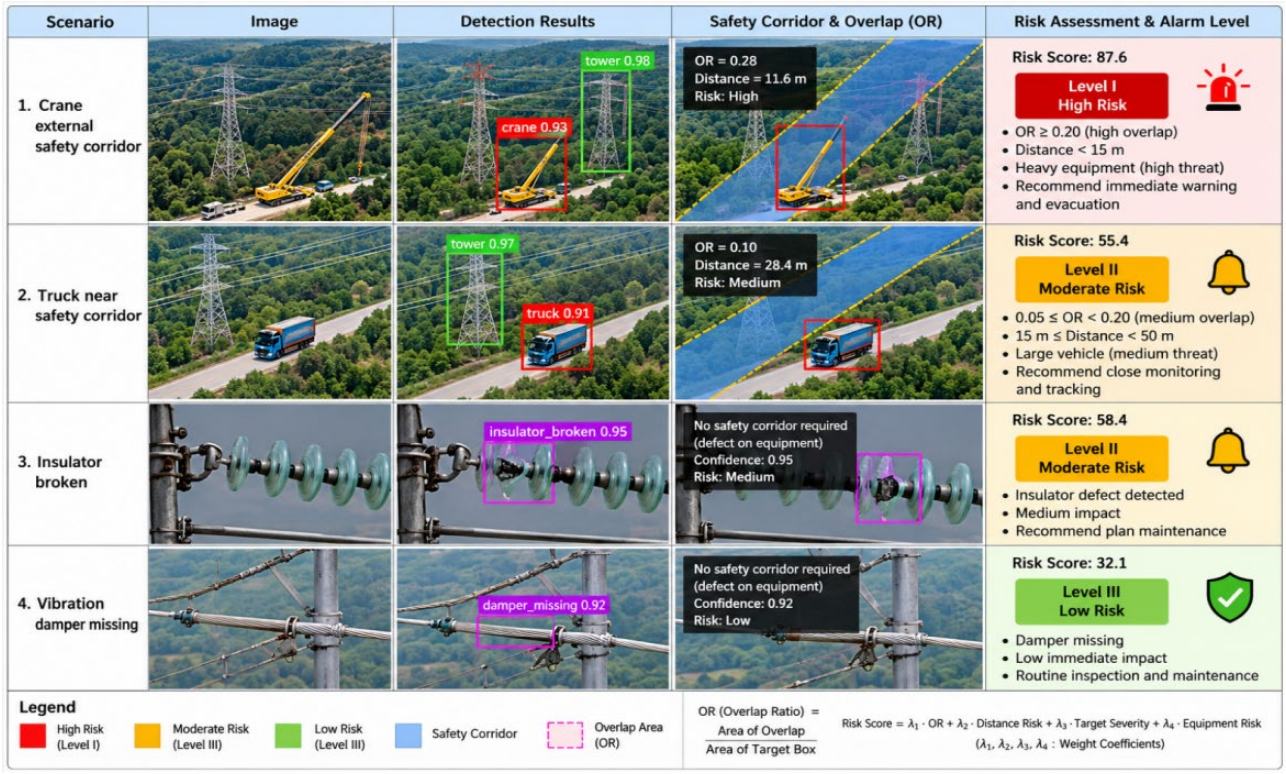

Figure 8. Visualization of power safety risk perception and multi-level warning results

4.6. Reproducibility and Statistical Reliability

For the statistical-reliability analysis, EEL-YOLO is trained independently with seeds 2022-2026 under the same scenario-level split. Across five runs, precision is 94.60+/-0.16%, recall is 92.80+/-0.19%, mAP@0.5 is 95.10+/-0.16%, and mAP@0.5:0.95 is 72.60+/-0.19% (mean+/-standard deviation). The 95% confidence intervals are [94.90,95.30]% for mAP@0.5 and [72.37,72.83]% for mAP@0.5:0.95. Under paired seeds, EEL-YOLO significantly exceeds YOLOv8n for both metrics (paired t-test, $p<0.001$). Table 14 reports the run-wise values. Together with the per-class results and confusion matrix in Table 7/Figure 7, these results indicate that the aggregate gains are stable rather than being driven by a single training run.

Table 14. Five-run reliability results

Seed	P / R (%)	mAP50 / mAP50:95 (%)
2022	94.4 / 92.5	94.9 / 72.3
2023	94.7 / 92.9	95.1 / 72.7
2024	94.6 / 92.8	95.2 / 72.6
2025	94.8 / 93.0	95.0 / 72.8
2026	94.5 / 92.8	95.3 / 72.6
Mean +/-SD	94.60+/-0.16 / 92.80+/-0.19	95.10+/-0.16 / 72.60+/-0.19

5. Conclusion

This paper proposes EEL-YOLO for Power IoT edge safety perception by coordinating detection-oriented image enhancement, lightweight multi-scale feature extraction, small-target feature fusion, MPDIoU localization, model compression, and rule-based risk grading. Under the reported core experimental results, the enhancement module reaches 24.37 dB PSNR, 0.836 SSIM, and 91.8% post-enhancement mAP@0.5; EEL-YOLO reaches 94.6% precision, 92.8% recall, 95.1% mAP@0.5, and 72.6% mAP@0.5:0.95, with 67.6 FPS on RK3588. In the supplementary completion, five-run mAP@0.5 is 95.10+/-0.16%, AP_S is 55.8%, and the risk layer reaches 91.7% consistency with 93.6% Level-I recall and 6.9% false-alarm rate. The framework therefore illustrates a practical route from degraded-image restoration to lightweight detection and graded operational warning.

The present study also has limitations. Recognition of ultra-faint tiny defects may still deteriorate under extreme heavy haze or severe blur; rule-based risk thresholds require calibration to local grid-management standards; the current dataset is limited to the represented line sections and monitoring conditions; and cross-region, cross-season, and long-term field validation is still needed. Future work will investigate uncertainty-aware risk calibration, temporal tracking, domain adaptation, and larger-scale multi-region deployment. The cloud-edge-end separation can also be transferred to wind-power, railway, and other industrial

inspection scenarios after domain-specific retraining and redefinition of risk factors.

Acknowledgments.

This study is supported by the "Joint Plan of Science and Technology Plan of Liaoning Province (Natural Science Foundation-Doctoral Research Initiation Project)", with the project No. 2024-BSLH-157.

References

[1] Farhangi H. The path of the smart grid. IEEE Power and Energy Magazine, 2010, 8(1): 18–28. DOI: 10.1109/MPE.2009.934876.

[2] Gungor V C, Sahin D, Kocak T, Ergut S, Buccella C, Cecati C, Hancke G P. Smart grid technologies: communication technologies and standards. IEEE Transactions on Industrial Informatics, 2011, 7(4): 529–539. DOI: 10.1109/TII.2011.2166794.

[3] Bedi G, Venayagamoorthy G K, Singh R, Brooks R R, Wang K C. Review of Internet of Things (IoT) in electric power and energy systems. IEEE Internet of Things Journal, 2018, 5(2): 847–870. DOI: 10.1109/JIOT.2018.2802704.

[4] Saleem Y, Crespi N, Rehmani M H, Copeland R. Internet of Things-aided smart grid: technologies, architectures, applications, prototypes, and future research directions. IEEE Access, 2019, 7: 62962–63003. DOI: 10.1109/ACCESS.2019.2913984.

[5] Tao X, Zhang D, Wang Z, Liu X, Zhang H, Xu D. Detection of power line insulator defects using aerial images analyzed with convolutional neural networks. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020, 50(4): 1486–1498. DOI: 10.1109/TSMC.2018.2871750.

[6] Wen Q, Luo Z, Chen R, Yang Y, Li G. Deep learning approaches on defect detection in high resolution aerial images of insulators. Sensors, 2021, 21(4): 1033. DOI: 10.3390/s21041033.

[7] Wang Z, Yuan G, Zhou H, Ma Y, Ma Y. Foreign-object detection in high-voltage transmission line based on improved YOLOv8m. Applied Sciences, 2023, 13(23): 12775. DOI: 10.3390/app132312775.

[8] Shao Y, Zhou J, Liu X, et al. TL-YOLO: foreign-object detection on power transmission lines. Electronics, 2024, 13(8): 1543. DOI: 10.3390/electronics13081543.

[9] Xue B, Liu Z, Wang Z, Zhou W, Wang B, Yang X. Object Detection and Segmentation of Power Equipment in Infrared Images via Improved YOLOv8 and Prompt-Optimized SAM. EAI Endorsed Transactions on Scalable Information Systems, 2026, 12(7). DOI: 10.4108/eetsis.10727.

[10] Hu Y. Research on Intelligent Detection Method for Operation and Maintenance Violations of Power Distribution Equipment Based on YOLOv12. EAI Endorsed Transactions on Scalable Information Systems, 2026, 12(9). DOI: 10.4108/eetsis.10801.

[11] He K, Sun J, Tang X. Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(12): 2341–2353. DOI: 10.1109/TPAMI.2010.168.

[12] Li B, Peng X, Wang Z, Xu J, Feng D. AOD-Net: All-in-One Dehazing Network. Proceedings of the IEEE International Conference on Computer Vision, 2017: 4780–4788. DOI: 10.1109/ICCV.2017.511.

[13] Liu X, Ma Y, Shi Z, Chen J. GridDehazeNet: attention-based multi-scale network for image dehazing. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 7313–7322. DOI: 10.1109/ICCV.2019.00741.

[14] Wei C, Wang W, Yang W, Liu J. Deep Retinex decomposition for low-light enhancement. British Machine Vision Conference, 2018.

[15] Guo C, Li C, Guo J, Loy C C, Hou J, Kwong S, Cong R. Zero-reference deep curve estimation for low-light image enhancement. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1780–1789. DOI: 10.1109/CVPR42600.2020.00178.

[16] Zamir S W, Arora A, Khan S, Hayat M, Khan F S, Yang M H, Shao L. Restormer: efficient transformer for high-resolution image restoration. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 5728–5739.

[17] Shen X, Li H, Li Y, Zhang W. LDWLE: self-supervised driven low-light object detection framework. Complex & Intelligent Systems, 2025, 11: 82. DOI: 10.1007/s40747-024-01681-z.

[18] Kou K, Yin X, Gao X, Nie F, Liu J, Zhang G. Lightweight two-stage transformer for low-light image enhancement and object detection. Digital Signal Processing, 2024, 150: 104521. DOI: 10.1016/j.dsp.2024.104521.

[19] Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems, 2015, 28: 91–99.

[20] Terven J, Cordova-Esparza D M, Romero-Gonzalez J A. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680–1716. DOI: 10.3390/make5040083.

[21] Lin T Y, Dollar P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117–2125.

[22] Tan M, Pang R, Le Q V. EfficientDet: scalable and efficient object detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781–10790.

[23] Woo S, Park J, Lee J Y, Kweon I S. CBAM: convolutional block attention module. European Conference on Computer Vision, 2018: 3–19. DOI: 10.1007/978-3-030-01234-2_1.

[24] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132–7141.

[25] Zhao Y, Lv W, Xu S, Wei J, Wang G, Dang Q, Liu Y, Chen J. DETRs beat YOLOs on real-time object detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965–16974.

[26] Wang H, Luo S, Wang Q. Improved YOLOv8n for Foreign-Object Detection in Power Transmission Lines. IEEE Access, 2024, 12: 121433-121440. DOI: 10.1109/ACCESS.2024.3452782.

[27] Gou M, Xu W, Liu C, Zhang L, Tang H, Liu J, Fu W. An Enhanced YOLOv8-Based Approach for Foreign Object Detection on Transmission Lines. Algorithms, 2026, 19(4): 264. DOI: 10.3390/a19040264.

[28] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L C. MobileNetV2: inverted residuals and linear bottlenecks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510–4520. DOI: 10.1109/CVPR.2018.00474.

[29] Han K, Wang Y, Tian Q, Guo J, Xu C, Xu C. GhostNet: more features from cheap operations. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580–1589.

[30] Ding X, Zhang X, Ma N, Han J, Ding G, Sun J. RepVGG: making VGG-style ConvNets great again. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13733–13742.
[31] Gou J, Yu B, Maybank S J, Tao D. Knowledge distillation: a survey. International Journal of Computer Vision, 2021, 129: 1789–1819. DOI: 10.1007/s11263-021-01453-z.
[32] Su K, Cao L, Zhao B, Li N, Wu D, Han X. N-IoU: better IoU-based bounding box regression loss for object detection. Neural Computing and Applications, 2024, 36: 3049–3063. DOI: 10.1007/s00521-023-09133-4.
[33] Li S, Xie Y, Shi M, Zheng X, Lu Y. Mobile Edge Computing Empowered Energy Consumption Optimization for Multiuser Power IoT Networks. EAI Endorsed Transactions on Scalable Information Systems, 2025, 12(3). DOI: 10.4108/eetsis.8678.
[34] Shi W, Cao J, Zhang Q, Li Y, Xu L. Edge computing: vision and challenges. IEEE Internet of Things Journal, 2016, 3(5): 637–646. DOI: 10.1109/JIOT.2016.2579198.
[35] Ultralytics. YOLOv8: Ultralytics YOLOv8 model documentation. Available at: https://docs.ultralytics.com/models/yolov8/, accessed June 2026.
[36] Zheng Z, Wang P, Liu W, Li J, Ye R, Ren D. Distance-IoU loss: faster and better learning for bounding box regression. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993–13000.
[37] Ma S, Xu Y. MPDIoU: a loss for efficient and accurate bounding box regression. arXiv preprint arXiv:2307.07662, 2023.
[38] Ou J, Shen Y. Underwater Target Detection Based on Improved YOLOv7 Algorithm With BiFusion Neck Structure and MPDIoU Loss Function. IEEE Access, 2024, 12: 105165-105177. DOI: 10.1109/ACCESS.2024.3436073.
[39] Wang Z, Bovik A C, Sheikh H R, Simoncelli E P. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 2004, 13(4): 600–612. DOI: 10.1109/TIP.2003.819861.