frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 10 February 2026
DOI 10.3389/fenrg.2026.1745369

OPEN ACCESS

EDITED BY
Feng Liu,
Nanjing Tech University, China

REVIEWED BY
Man Wu,
Guangxi University, China
Weichao Pan,
Shandong Jianzhu University, China

*CORRESPONDENCE
Liang Zhang,
✉ liangzhangswpu@gmail.com

RECEIVED 13 November 2025
REVISED 05 January 2026
ACCEPTED 14 January 2026
PUBLISHED 10 February 2026

CITATION
Liu W, Yang T, Shi D, Li Y, Zhang L and Liu J (2026) YOLOv8-FOD: a lightweight foreign object detection algorithm for power transmission lines.
Front. Energy Res. 14:1745369.
doi: 10.3389/fenrg.2026.1745369

YOLOv8-FOD: a lightweight foreign object detection algorithm for power transmission lines

Wei Liu[1], Tao Yang[1], Dao Shi[1], Yufang Li[1], Liang Zhang[2]* and Jiansheng Liu[2]

[1]State Grid Sichuan Electric Power Company Guangyuan Power Supply Company, Guangyuan, China, [2]Southwest Petroleum University, Chengdu, China

Timely and accurate detection of foreign objects is crucial for the safe operation of transmission lines in power grid. Currently, object detection models have more and more parameters and their calculations are becoming increasingly complex. Therefore, sufficient computing power is usually required. To deploy models on resource-constrained edge devices, this paper proposed a lightweight detection algorithm based on improved YOLOv8, which is called YOLOv8-FOD (Foreign Object Detection). Firstly, the backbone network is optimized by incorporating large kernel block (LarK Block) into c2f module, forming a new C2f_LarK module. This achieves a wider receptive field, captures more contextual information, and effectively reduces network redundancy. Secondly, a lightweight detection head, Det_Tiny, is proposed. Through adaptive feature redundancy compression and hardware-friendly computational optimization, computational complexity is significantly reduced. Finally, a new feature fusion network structure (Fusion) is designed, utilizing the CGAFusion module to fuse high-dimensional and low-dimensional features. This captures important information at different semantic layers and improves the detection of edge details, effectively enhancing detection accuracy. Experimental results show that, compared with standard YOLOv8n, the proposed model reduces parameter count by 36.6%, model size by 31.1%, and computational complexity (GFLOPs) by 40.7%. While maintaining detection accuracy of mAP@0.5 and improving mAP@[0.5:0.95] by 0.3%, the model is more lightweight and has high practicality for deployment on edge devices.

KEYWORDS

CGAFusion, FasterNet, lightweight object detection, UniRepLKNet, YOLOv8

1 Introduction

The safe operation of transmission lines is important for power system. However, due to long-term exposure to the outdoors, they are susceptible to interference from foreign objects. Common foreign objects are kites, balloons, floating objects, bird nests, *etc.* When attached to transmission lines, they can easily cause arc discharges, short circuit, and even fires or power grid failures (Wang Z. et al., 2023). This seriously endangers the reliability and stability of power system. Therefore, it is necessary to detect and eliminate potential safety hazards at an early stage by accurately and efficiently detecting foreign objects.

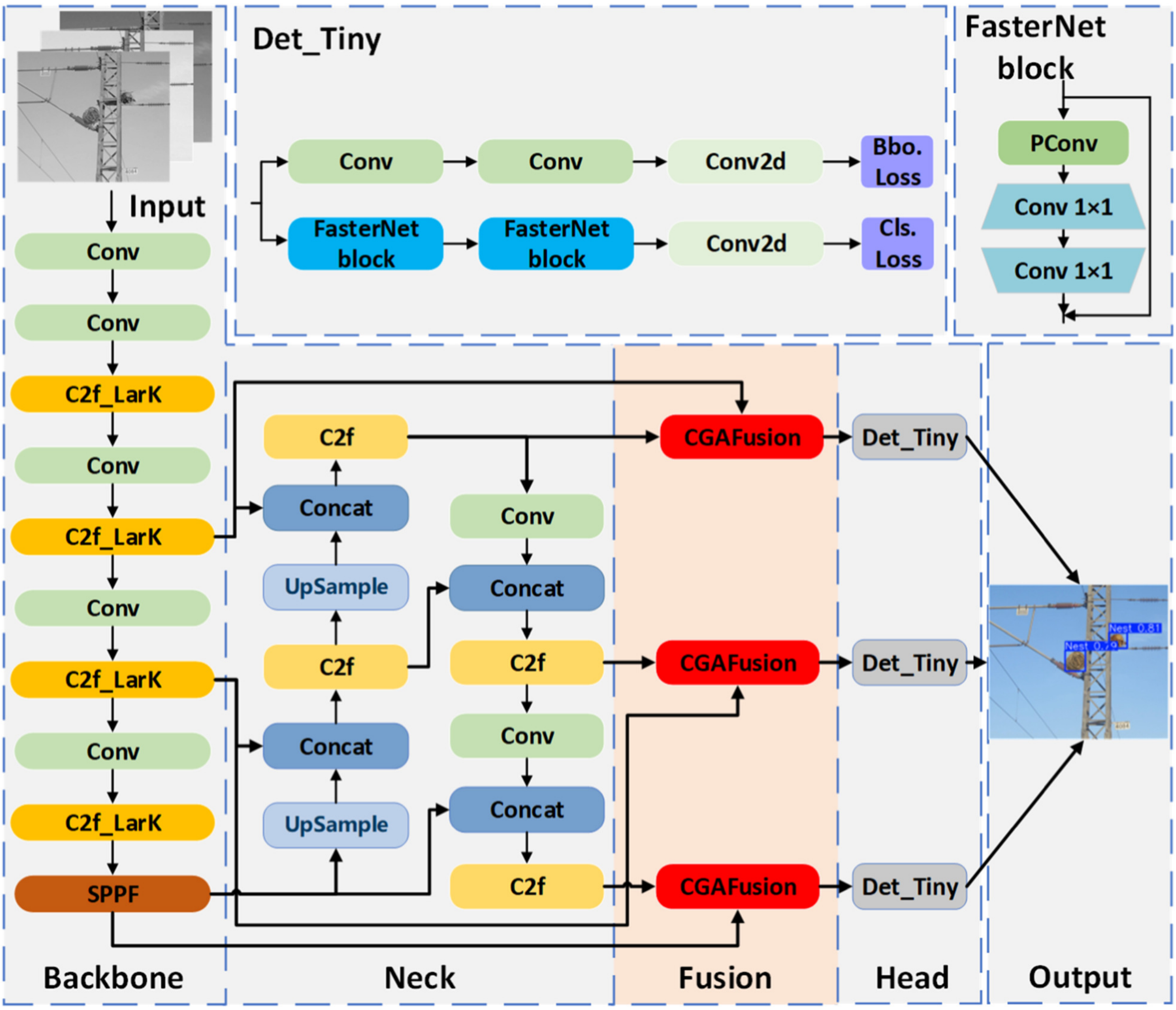

FIGURE 1
Yolov8-FOD model structure diagram.

Traditional foreign object detection on power transmission lines mainly relies on manual inspection. But it has the problems of low efficiency and high cost. Consequently, deep learning-based real-time object detection technology has emerged as a pivotal approach to addressing this challenge (Faisal et al., 2025). The continuous development of deep learning has shown significant advantages in foreign object detection. However, the parameter count and computational complexity continue to increase, resulting in large model size and limited inference speed. When deployed on actual edge devices, they are often constrained by memory capacity, computing resources, and energy consumption. This leads to problems such as model mismatch and high latency, which seriously hinders real-time monitoring and early warning capabilities. Compared with other detection models such as SSD, Faster R-CNN and Retina Net, the YOLO series detection models require relatively small computing resources (Liu et al., 2016; Maduako et al., 2022; Ren et al., 2017; Yan et al., 2025; Lin et al., 2017). They are more suitable for deployment on edge devices.

Representative YOLO models for foreign object detection on power transmission lines include YOLOv4, YOLOv5, YOLOv7, YOLOv8, and YOLO11 (Bochkovskiy et al., 2020; Peng et al., 2025; Wang C. Y. et al., 2023; Bin et al., 2025; Shao et al., 2024). Song et al. (2021) proposed a foreign object detection algorithm for high-voltage transmission lines based on improved YOLOv4 algorithm. By using k-means clustering and DIoU-NMS method, the detection accuracy on kite, balloon, plastic, wildfire, and smog reaches 81.72%. Zhou et al. (2024) proposed a foreign body detection algorithm based on the improved YOLOv5. The algorithm introduced an efficient channel attention (ECA) module in the backbone network and adopted bilinear interpolation in the neck network to improve the detection accuracy of the model, achieving a 3.9% increase in mAP@0.5 accuracy. Liu et al. (2023) combined two attention mechanisms and an additional detection layer to improve its ability to identify small defects and distant objects, achieving a detection speed 16.3% faster than YOLOv5 and a detection progress 3.3% higher than YOLOv7. Liu et al. (2021) improved the overall

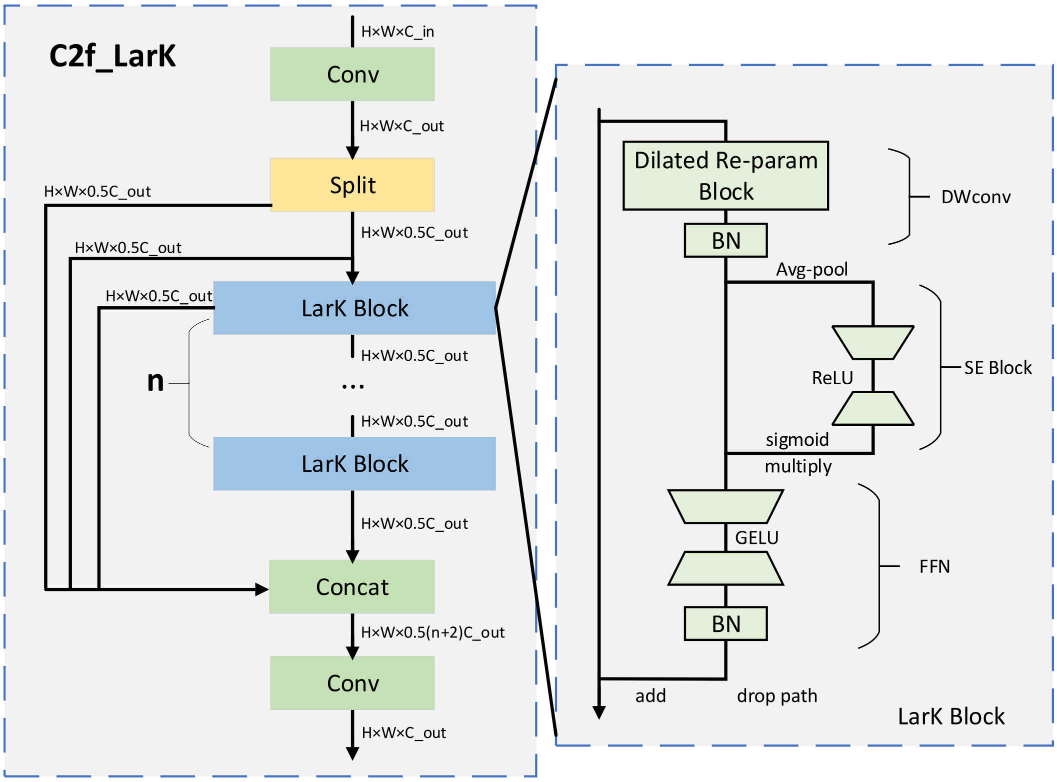

FIGURE 2
Structure design of C2f_UniRepLKNet.

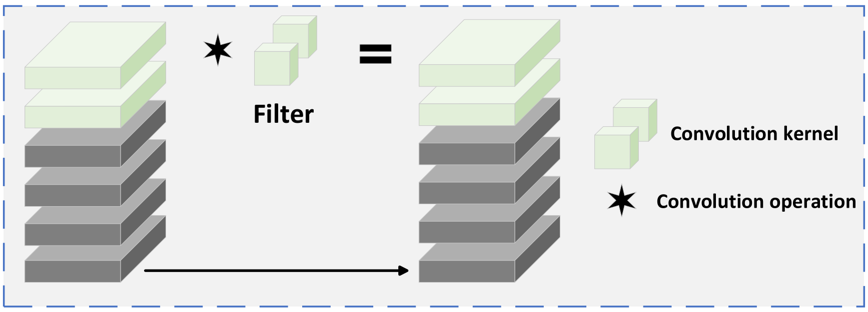

FIGURE 3
Structure design of PConv.

learning capability of the network by aggregating spatial and channel information in the feature map, achieving an 88.5% detection accuracy for foreign objects such as bird nests, hanging debris, and wildfires. However, this increase in detection performance came at the cost of a significant rise in parameters. To address the challenges of small scale, high density, and deformability of obstacles in high-voltage lines, Pan et al. (2024) proposed an improved YOLO algorithm based on multi-scale feature fusion. By incorporating data augmentation strategies such as image rotation and cropping, their approach significantly enhanced model accuracy and robustness in complex scenarios. Wang Z. et al. (2023) constructed an SPPCSPC module and added a global attention module to the stem to focus on occluded foreign objects to improve the feature extraction ability of YOLOv8 at multiple scales. The model achieved an accuracy of 95.5% in detecting foreign objects, while the number of parameters increased to 50.6M.

The models mentioned above all aim to improve the ability to detect targets by increasing the number of parameters. This poses challenges for deployment on edge devices. To overcome existing bottlenecks, recent research has pivoted toward lightweight architectures and efficient edge-side inference. For instance, Lu et al. (2023) reconstructed YOLOv5 using GhostNetV2 for transmission line defect detection. Their model achieved a mean Average Precision (mAP) of 94.3% at 63 FPS on edge devices,

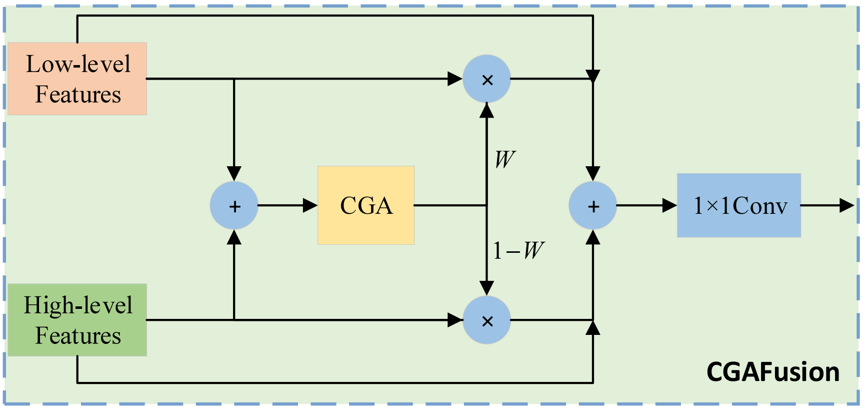

FIGURE 4
Structure design of CGAFusion.

substantially bolstering real-time monitoring and generalization capabilities in harsh environments (Bin et al., 2025). Furthermore, Sankuri et al. (2025) introduced a hybrid DETR framework combined with semi-supervised learning for insulator defect detection. By optimizing the encoder and introducing a hierarchical hybrid matching strategy, the model attained a high-speed inference of 85 FPS, markedly improving detection precision and robustness against complex backgrounds (Shao et al., 2024). Additionally, Zhu et al. (2022) developed Fast-PLDN, a real-time semantic segmentation network that leverages low-pass and high-pass filter blocks along with an edge-attention fusion module to tackle the challenges of curved power line detection. This model achieves an exceptional speed of 189.6 FPS with an mIoU of 71.3%, demonstrating superior real-time perception performance.

Therefore, how to effectively reduce model size and computational complexity while maintaining high detection accuracy for complex foreign objects remains a key research focus in this field. In light of this, this paper proposes a lightweight foreign object detection algorithm for transmission lines based on YOLOv8-FOD, with the following main contributions:

1. The large kernel block (LarK Block) of UniRepLKNet (Ding et al., 2024) is integrated into the c2f module to optimize the backbone network to form a new C2f_LarK module. This achieves a wider receptive field without increasing the depth of the model. By adopting a larger convolution kernel, LarK Block captures more contextual information without the need for additional network layers, effectively reducing the redundancy of the network.
2. A lightweight detection head, Det_Tiny, is proposed. The standard convolution of classification loss branch in original detector head is replaced by partial convolution (PConv) of FasterNet (Chen et al., 2023). Through adaptive compression of feature redundancy and hardware-friendly computation optimization, the computational complexity is significantly reduced while ensuring classification accuracy.
3. A new feature fusion network Fusion is designed. Based on CGAFusion (Chen et al., 2024a), high-dimensional features and low-dimensional features are fused. Content-guided attention (CGA) is used to assign a unique Spatial Importance Map (SIM) to each channel. This allows it to focus on more useful information in the features and improve the detection of edge details, effectively increasing detection accuracy.

2 YOLOv8-FOD lightweight detection model

This paper uses YOLOv8n as the baseline model and proposes an improved YOLOv8-FOD lightweight target detection model tailored for resource-constrained edge devices. The improved network structure is shown as Figure 1.

The model first incorporates the Large Kernel Block of UniRepLKNet into the backbone network, reconstructing the C2f module to create a new C2f_LarK module. This modification effectively expands the receptive field without increasing network depth, significantly enhancing the context perception ability for small targets at a distance. Secondly, a lightweight detection head, Det_Tiny, based on Partial Convolution (PConv) from FasterNet is designed to compress redundant features through channel-selective calculations. This reduces the computational load of the detection head while maintaining classification accuracy through a dynamic channel adjustment mechanism. Furthermore, a multi-scale feature fusion network (Fusion) based on CGAFusion is proposed. A dual-path attention mechanism and dynamic weight fusion strategy is used to collaboratively optimize high-dimensional semantic features alongside low-dimensional detail features, thereby improving the positioning accuracy of foreign objects in complex backgrounds.

2.1 Backbone network reconstruction

In the task of detecting foreign objects on power transmission lines, overcoming complex background interference and acquiring small target features are the key points. Although YOLOv8 achieves effective feature extraction in the backbone network through C2f module, the traditional small-scale convolution

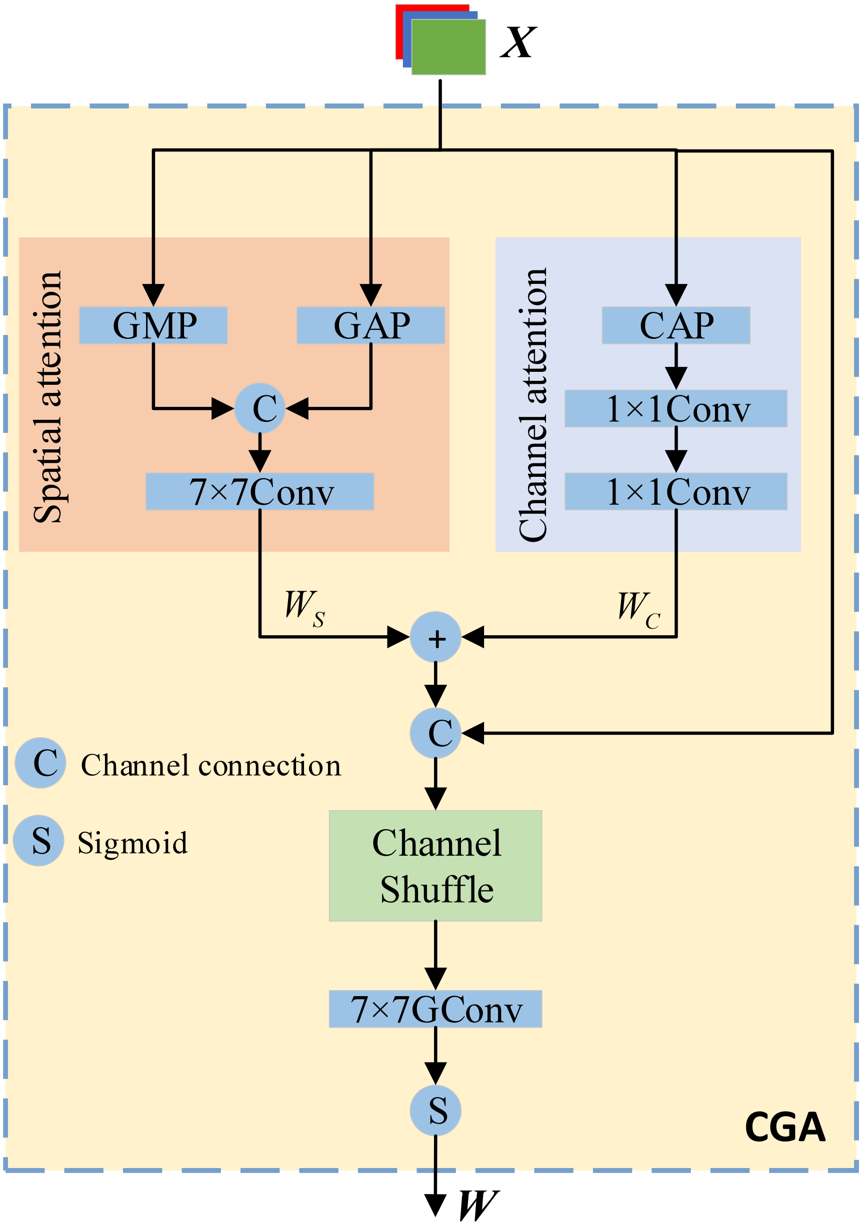

FIGURE 5
Structure design of CGA.

kernel (3 × 3) in its backbone network has problems such as limited receptive field and insufficient long-range dependency modeling capabilities. It is difficult to capture global contextual information in scenarios with large transmission line spans. This paper integrates the LarK block module of UniRepLKNet into C2f module to construct the C2f_LarK module. The LarK block module captures more contextual information through a larger receptive field, enabling the backbone network to capture richer features. At the same time, the re-parameterized nature of the LarK block can significantly improve the efficiency of the model during inference, making it more conducive to deployment on edge devices. The structure design of C2f_UniRepLKNet is shown as Figure 2.

In Figure 2, the LarK Block module achieves efficient feature extraction through a dual-path design of dilated reparameterization and channel attention enhancement (Hu et al., 2018). Its core is to use dilated convolution to reconstruct the parameter space of small-kernel convolutions, re-parameterizing the multi-branch structure into a single sparse large-kernel convolution through mathematical equivalence. Specifically, during the training phase, the module deploys non-dilated large-kernel convolutions (e.g., 13 × 13) and dilated small-kernel convolutions (e.g., 3 × 3, with a dilation rate of d = 4) in parallel, complementing their parameter spaces through gradient co-optimization. During the inference phase, leveraging the additivity principle of convolution, the multi-branch structure is mathematically equivalently reparameterized.

TABLE 1 Configurations for the experimental environment.

Items	Types
Operating system	Windows 11
CPU	Intel core i5-124900F
GPU	Nvidia RTX 4060TI GPU
Language	Python 3.9.18
Platform	PyTorch2.1.1
CUDA	11.8

TABLE 2 Training parameters.

Items	Parameters
Input image size	3*640*640
Batch size	16
Epochs	200
Optimizer	Adam
Momentum	0.937
Initial learning rate	0.01
Learning rate decay strategy	Cos annealing

Specifically, in physical implementation, we deploy it as a depth-wise dilated convolution. This implementation directly utilizes the underlying operator optimizations of dilated and depth-wise convolutions in mainstream hardware (such as GPUs and NPUs), achieving a receptive field equivalent to a 31 × 31 large core without explicit sparse matrix operations. This avoids the hardware efficiency degradation issues caused by unstructured sparsity. Furthermore, the module integrates the Squeeze-and-Excitation (SE) attention mechanism. Global Average Pooling (GAP) is used to compress the spatial dimensions to generate channel description vectors. Channel attention weights are then generated through a fully connected layer and sigmoid activation to dynamically calibrate feature channel responses.

Regarding the selection of the core kernel size for this module, this study strictly adheres to the ablation conclusions from the original UniRepLKNet (Ding et al., 2024) and designates 31 × 31 as the optimal kernel dimension. Research has substantiated that this size represents a "saturation point" for balancing receptive field gains and inference latency. Compared to larger kernels (e.g., 51 × 51), a 31 × 31 kernel provides a large-scale receptive field sufficient to cover the wide-span scenarios of transmission lines with negligible increases in Memory Access Cost (MAC). This enables efficient feature capture while effectively circumventing excessive computational overhead.

2.2 Lightweight target detection head

In traditional YOLOv8 object detection model, 25% of its parameters come from the detection head. Conventional convolutional layers (Conv + BN + SiLU) are often used for category classification and bounding box regression, but the introduced parameter amount and computational complexity limit their deployment capabilities on edge devices. To optimize this problem, this paper proposes to replace the standard convolution operation in original classification branch with Partial Convolution (PConv) of FasterNet. This allows for the construction of a new lightweight object detection head, Det_Tiny. Through adaptive feature redundancy compression and hardware-friendly computational optimization, the algorithm significantly reduces computational complexity while maintaining classification accuracy. The PConv structure design is shown in Figure 3.

Traditional standard convolution performs intensive calculations on all input channels in the spatial dimension. However, in the scenario of foreign object detection on power transmission lines, the feature distribution of background areas (such as the sky and vegetation) and target areas (kites and plastic bags) is significantly spatially sparse. Direct application of standard convolution will lead to a large number of redundant calculations. Especially in classification tasks, high-frequency detail features (such as foreign object edge textures) contribute much more to category judgment than smooth background areas. PConv performs spatial convolution only on some channels through channel grouping and selective calculation. The remaining channels are processed using low-cost point-by-point convolution (PWConv) to reduce redundant operations. Its mathematical expression is as Equation 1.

$$\mathbf{F}_{out} = \mathrm{Concat}\left(\mathrm{Conv}_{k\times k}\left(\mathbf{F}_{in}^{[1:g]}\right), \mathrm{PWConv}\left(\mathbf{F}_{in}^{[g+1:C_{in}]}\right)\right) \quad (1)$$

Where g is the number of channels involved in spatial convolution. By separating high-frequency feature channels (required for spatial detail extraction) from low-frequency channels (required for channel information fusion), a balance between computational efficiency and feature expression is achieved.

In this study, adhering to the ablation experimental conclusions from the original FasterNet authors (Chen et al., 2023), we set this parameter to 1/4 of the total input channels (g = 1/4). Research has demonstrated that this ratio serves as the optimal balance point between computational efficiency and feature representation capability. It ensures that a sufficient number of channels are dedicated to extracting critical high-frequency spatial features of foreign objects, while simultaneously maximizing the utilization of the remaining channels for low-cost fusion to reduce Giga Floating-point Operations Per Second (GFLOPs). An excessively low ratio leads to insufficient feature extraction, whereas an overly high ratio diminishes the lightweight advantages. Thus, the configuration

TABLE 3 Model performance with different modules.

Model	mAP@0.5 (%)	mAP@[0.5:0.95] (%)	Precision (%)	Recall (%)	Model size (MB)	Parameters (million)	GFLOPs
Baseline	94.0	83.7	91.2	87.8	6.1	3.0	8.1
+C2f_LarK	94.2	84.5	91.6	88.2	5.7	2.5	7.4
+Det-tiny	93.6	83.7	91.7	87.3	4.4	2.4	5.2
+Fusion	94.1	84.0	91.6	87.3	6.2	3.1	8.4
YOLOv8-FOD	94.0	84.0	91.8	87.4	4.1	1.9	4.8

of g = 1/4 effectively satisfies the dual requirements for both accuracy and speed.

2.3 Multi-scale feature fusion network

To address the accuracy loss caused by lightweight model, this paper proposes a multi-scale feature fusion network, Fusion, based on CGAFusion. Structure design of CGAFusion is shown in Figure 4.

The CGAFusion module fuses shallow features with deep features. The Content-Guided Attention (CGA) module, shown in Figure 5, assigns a unique SIM, $W \in R^{C\times H\times W}$, to each channel. This coarse-to-fine fusion of channel and spatial information enhances feature representation. The spatial attention module captures the spatial information of the feature map through global average pooling and global max pooling operations. The channel attention module utilizes adaptive average pooling and convolutional layers to highlight the importance of different channels. Finally, by combining the input feature map and the outputs of the first two attention modules, the SIM in the feature space is finely adjusted through channel shuffling.

The Fusion network proposed in this study is composed of three CGAFusion modules, which respectively fuse the low-level features from the final three C2f_LarK layers of the backbone with the high-level features from the final three C2f layers of the neck. Given the backbone low-level features F_{low} ($F_{low} \in R^{C\times H\times W}$) and the neck high-level features F_{high} ($F_{high} \in R^{C\times H\times W}$) the module first constructs ground-state features via initial feature superposition $X = F_{low} + F_{high}$ ($X \in R^{C\times H\times W}$) to serve as the input for the CGA module. Finally, the fusion results are fed into the Det-Tiny detection head. The specific steps are detailed below:

1. The spatial attention modeling of the CGA module captures local salient regions through dual-path feature compression. The calculation is expressed as shown in Equation 2.

$$W_S = \sigma\left(\mathrm{Conv}_{7\times7}^{\mathrm{reflect}}(\mathrm{Concat}(\mathrm{AvgPool}(X), \mathrm{MaxPool}(X)))\right) \quad (2)$$

Specifically, AvgPool and MaxPool compress the data along the channel dimension to produce 2-channel features. A 7 × 7 reflected padded convolution (Padding = 3) and a Sigmoid activation function are then used to generate a spatial weight map $W_S \in R^{C\times H\times W}$, highlighting the potential spatial distribution of foreign objects.

2. The channel attention modeling of the CGA module is based on the Squeeze-and-Excitation (SE) framework of SENet, which is utilized to model inter-channel dependencies. The calculation is expressed as shown in Equation 3.

$$W_C = \mathrm{Conv}_2(\mathrm{ReLU}(\mathrm{Conv}_1(\mathrm{AdaptiveAvgPool2d}(X, 1)))) \quad (3)$$

3. The CGA module dynamically generates refined Spatial Importance Maps through a fusion process. By incorporating grouped convolutions, the module facilitates feature interaction while effectively preventing cross-channel information interference. The calculation is expressed as shown in Equation 4.

$$W = \sigma\left(\mathrm{Conv}_{7\times7}^{\mathrm{Group=C}}(\mathrm{Concat}(X, W_S \oplus W_C))\right) \quad (4)$$

Specifically, the spatial and channel attention maps are fused via element-wise addition. The resulting integrated attention map is then concatenated with the initial features. A 7 × 7 grouped convolution (groups = C) is subsequently applied to extract local correlations, generating refined spatial importance weights $W \in R^{C\times H\times W}$. These weights are utilized to dynamically balance the contributions from the backbone and neck features.

4. Synergetic generation via the CGA-based Multi-scale Feature Fusion Network (CGAFusion). The calculation is expressed as shown in Equation 5.

$$F_{\mathrm{fuse}} = C_{1\times1}\left(X + F_{low} \odot W + F_{high} \odot (1 - W)\right) \quad (5)$$

Where $C_{1\times1}$ represents the channel recalibration layer implemented via a 1 × 1 convolution, and ⊙ denotes element-wise multiplication. The ground-state features serve to preserve the original information, while the Spatial Importance Maps dynamically allocate weights between the two components based on local complexity.

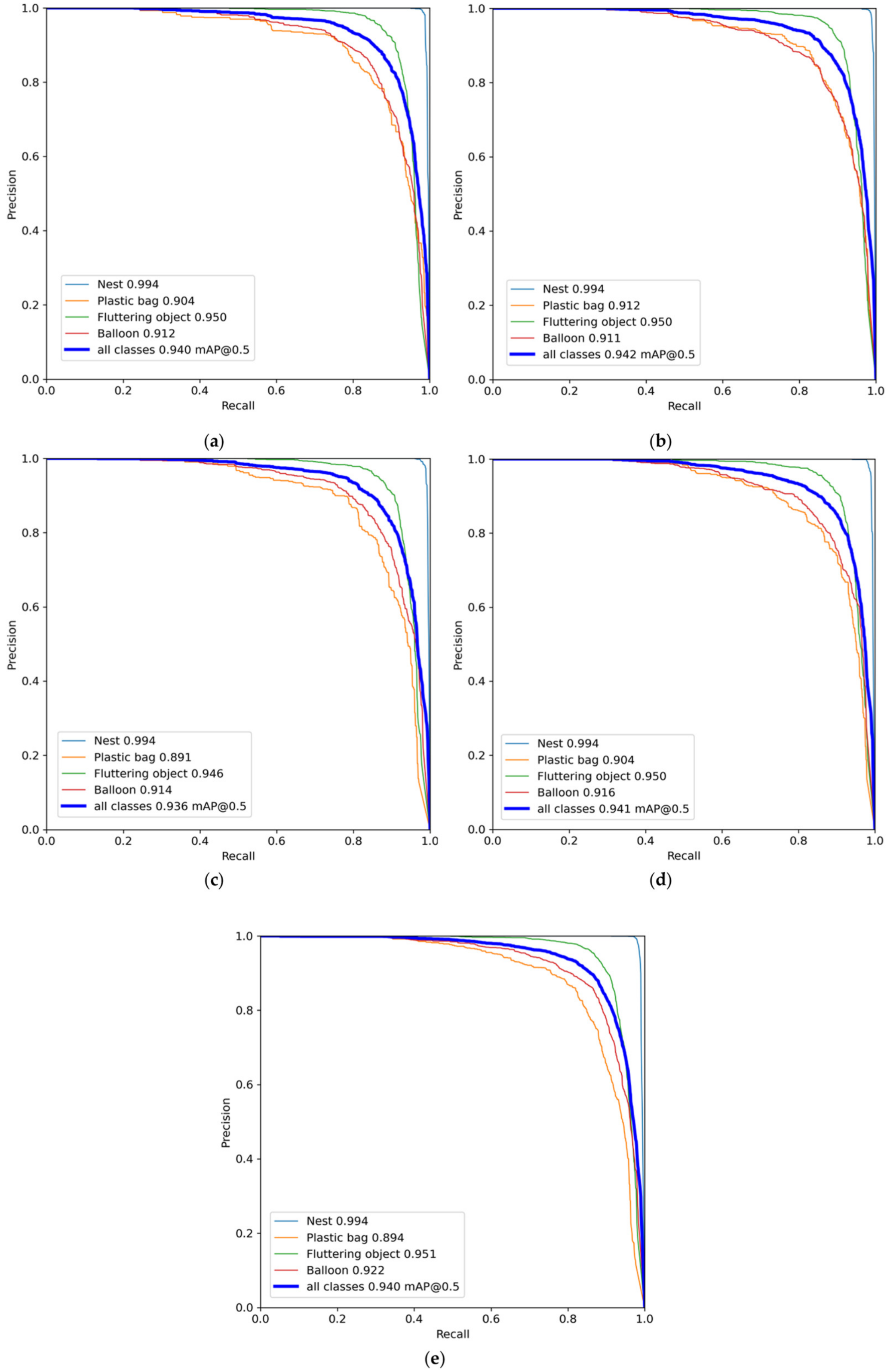

FIGURE 6
P-R curves of each model. **(a)** Baseline; **(b)** +C2f_LarK; **(c)** +Det-Tiny; **(d)** +Fusion; **(e)** YOLOv8-FOD.

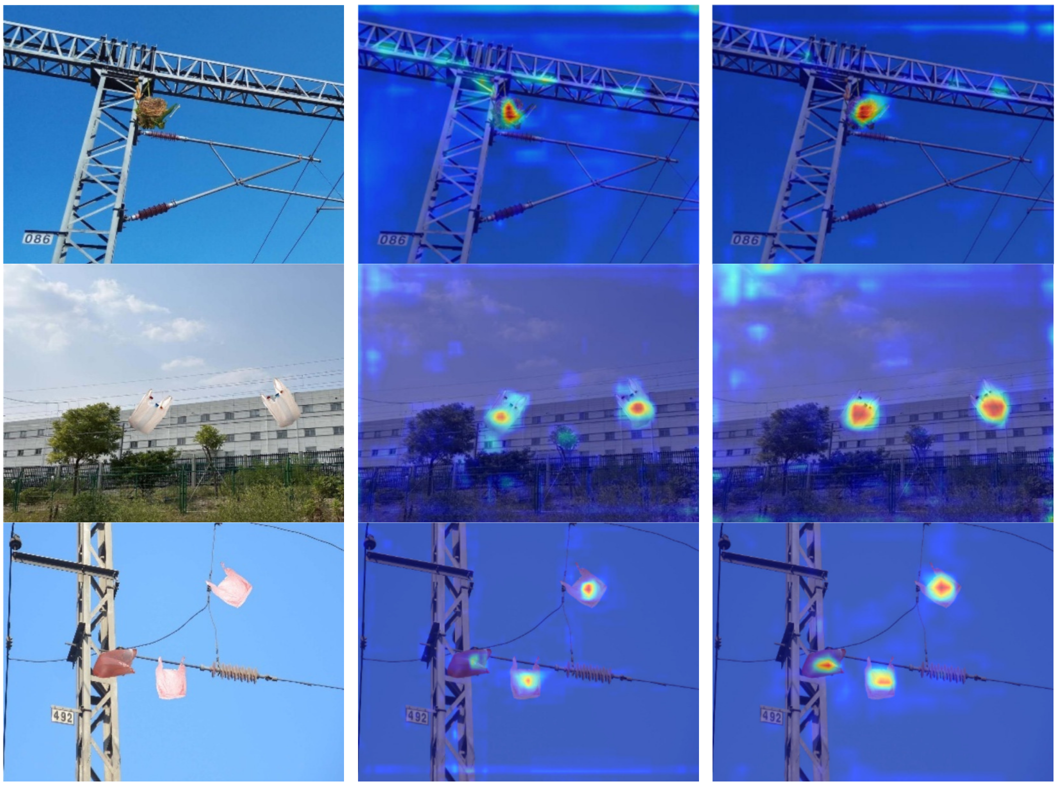

FIGURE 7
Comparison of heatmaps before and after model improvement.

TABLE 4 Performance of different model.

Model	mAP@0.5 (%)	mAP@[0.5:0.95] (%)	Precision (%)	Recall (%)	Model size (MB)	Parameters (million)	GFLOPs
Faster-RCNN	89.5	77.6	87.6	85.6	108.9	137.1	370.2
YOLOv5n	94.0	81.6	91.0	87.2	3.7	1.7	4.2
YOLOv7-tiny	94.5	82.0	91.4	87.5	12.3	6.0	11.7
YOLO11n	94.1	83.8	91.1	89.7	5.2	2.6	6.3
YOLO12n	93.8	82.6	91.0	87.5	5.2	2.6	6.3
YOLOv8-FOD	94.0	84.0	91.8	87.3	4.1	1.9	4.8

This fusion approach achieves information complementarity, enabling the model to capture high-level semantic information while retaining underlying details. Furthermore, the CGAFusion module's adaptive attention mechanism highlights key features and suppresses redundant information, effectively enhancing the model's ability to detect objects of varying scales. This improves the model's overall detection performance in complex scenarios.

3 Experimental setup

3.1 Experimental environment

Table 1 provides the main configurations for the experimental environment, including the operating system, CPU, GPU, Python, Cuda, and PyTorch version.

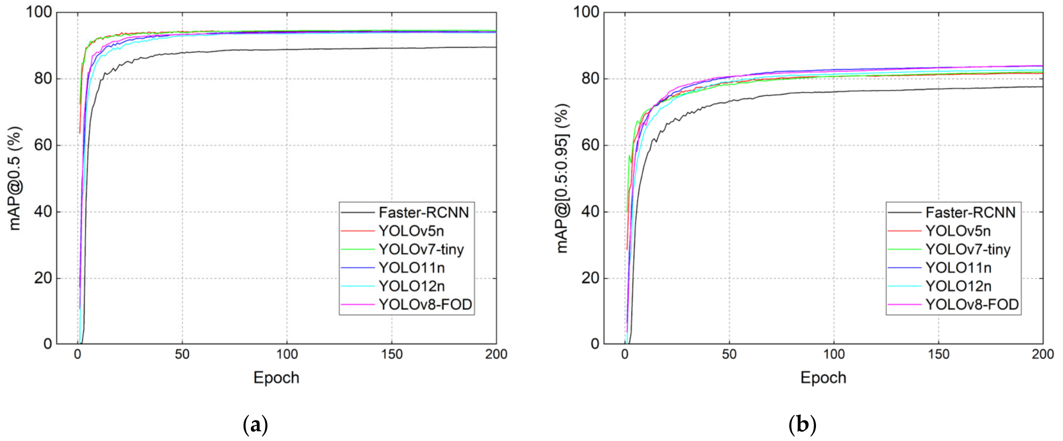

FIGURE 8
Comparison of mAP curves of different models: **(a)** mAP@0.5; **(b)** mAP@[0.5:0.95].

The hyperparameter settings for model training are shown in Table 2.

The input image size was uniformly set to 3 × 640 × 640. The number of training epochs was 200. The training batch size was 16. The initial learning rate was 0.01. And the cosine annealing decay algorithm was used to gradually change the current learning rate.

3.2 Experimental dataset

In view of the pronounced scarcity of publicly available datasets for power line foreign object detection, this study employs the RailFOD23 (Chen et al., 2024b) benchmark to evaluate the generalization capability of the proposed lightweight algorithm in complex scenarios. This dataset, which has been widely adopted in the literature for assessing the performance of power line detection models (Hao et al., 2024), comprises video frames sourced from multi-national railway surveillance systems. It spans a broad spectrum of illumination conditions, weather patterns, and object categories. Notably, its structural and environmental characteristics share a high degree of scene commonality and challenge with power line inspection tasks, providing a rigorous platform for validating model robustness.

The RailFOD23 dataset contains 14,615 images and 40,541 annotated objects. The annotated objects include four common foreign objects: bird nests, balloons, plastic bags, and floating objects. It is randomly divided into three group sets with a 7:2:1 ratio. The training set consists of 10,230 images. The validation set consists of 2,923 images. And the test set consists of 1,462 images.

3.3 Evaluation indicators

This study uses recall (R), precision (P), and Mean Average Precision (mAP) to evaluate the accuracy of the model in detecting small objects. Calculations are shown from Equations 6–9. To comprehensively evaluate the computational efficiency of the model, other metrics introduced include the total number of parameters, model size, and number of floating-point operations (GFLOPs).

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{7}$$

$$AP = \int_0^1 P(R)\mathrm{d}R \tag{8}$$

$$mAP = \frac{\sum_{i=1}^{K} AP_i}{K} \times 100\% \tag{9}$$

Where TP represents the true positive that is correctly predicted. FP represents the false positive that is incorrectly predicted as positive. FN represents the false negative that is incorrectly predicted as negative. And K represents the number of classes.

4 Results and analysis

Ablation experiment and comparison experiment are conducted to validate the effectiveness of proposed model.

4.1 Results of ablation experiment

C2f_LarK module, Det_Tiny module, and CGAFusion module are considered in ablation experiments.

4.1.1 Model performance with different modules

The experimental results of model performance with different modules are shown in Table 3 below. The baseline is original YOLOv8n model.

Table 3 shows that when C2f_LarK module is added alone, mAP@0.5 increases by 0.2%, mAP@ [0.5:0.95] increases by 0.8%, model size decreases by 0.4MB, parameter count decreases by 16.6%, and computational complexity also decreases by 8.6%.

While the standalone integration of the Det-Tiny module leads to a marginal 0.4% decrease in mAP@0.5, it yields substantial improvements in efficiency. The model size is reduced by 1.7 MB, the parameter count decreases by 20%, and the computational complexity (GFLOPs) is simultaneously lowered by 35.8%, thereby offering a superior overall advantage for the model.

When Fusion module is added alone, mAP@0.5 and mAP@[0.5:0.95] increases slightly by 0.1% and 0.3% respectively. Model size also increases by 0.1 MB. Parameter count increases by 3.3% and GFLOPs increases by 3.7%. All parameters are close to that of the baseline.

When C2f_LarK, Det_Tiny, and Fusion modules are added simultaneously, mAP@0.5 remains unchanged and mAP@[0.5:0.95] increases by 0.3%. Meanwhile, model size decreases by 31.1%. Parameter count decreases by 36.6%, and GFLOPs decreases by 40.7%. This demonstrates the effectiveness of the improved YOLOv8-FOD model in achieving lightweight design.

4.1.2 Comparison of precision-recall curve

The Precision-Recall (P-R) curves of five different target detection experiments conducted on the RailFOD23 dataset are shown in Figure 6. The larger the area of P-R curve around the coordinate axis, the higher the model's mAP value and the better its performance. It can be seen that mAP@0.5 change little for detecting nest, plastic bag, fluttering object and balloon. The improved model achieves lightweighting without sacrificing detection accuracy.

4.2 Comparative visualization analysis

To visually verify the effectiveness of proposed method and explain the underlying mechanism of performance improvement, we used Gradient-weighted Class Activation Mapping (Grad-CAM) technology to perform a feature map visualization comparison analysis of the models before and after the improvement, as shown in Figure 7. As can be seen from the figure, the original YOLOv8n model, limited by its small effective receptive field, often struggles to capture the complete features of the target, resulting in weak response or even feature loss in the heatmap at the target region.

The improved YOLOv8-FOD model exhibits stronger feature extraction capabilities, with its high-response regions in the heatmap more densely and completely covering the foreign object target. Notably, compared to the original model, the improved model demonstrates stronger focusing ability in the target edge region, enabling more accurate delineation of the object's outline. This complete perception of the target's shape is mainly due to the large receptive field of the C2f_LarK module, ensuring that the model can capture the target's contextual information from a global perspective. Simultaneously, the introduction of the Fusion module effectively enhances the preservation of details during feature fusion, allowing the model to keenly capture object boundary information, thereby achieving accurate localization of the foreign object from its center to its edge.

4.3 Comparative analysis with other models

To further verify overall effectiveness of the improved model proposed in this paper, we conducted a comprehensive comparative experiment with other models.

4.3.1 Comparison of model performance

As shown in Table 4, the performance is compared with Faster-RCNN, YOLOv5n, YOLOv7-tiny, YOLO11n, and YOLO12n.

As shown in Table 4, the proposed model significantly improves accuracy compared with two-stage Faster-RCNN. It also significantly reduces parameter count and computational complexity. Compared with YOLOv5n baseline, the improved model achieves a 2.4% increase in mAP@ [0.5:0.95], with concurrent improvements in both Precision and Recall (Bin et al., 2025). Furthermore, the model exhibits only a marginal increase in parameter count and computational complexity. Compared with YOLOv7-tiny, YOLO11n and YOLO12n, the proposed model significantly reduces both number of parameters and computational complexity while maintaining accuracy.

Overall, the proposed model in this paper outperforms other models in terms of comprehensive performance of lightweight and detection accuracy.

4.3.2 Comparison of mAP curves

Figure 8 shows the mAP@0.5 and mAP@[0.5:0.95] accuracy curves of different models tested on the RailFOD23 dataset. It show that the proposed model significantly outperforms other models in terms of mAP@ [0.5:0.95], while reducing the number of model parameters and computational complexity. Furthermore, the mAP@0.5 accuracy does not significantly decrease compared with the baseline model.

5 Conclusion

To ensure efficient deployment of foreign object detection algorithm for power transmission lines on resource-constrained edge devices, this paper proposed a lightweight algorithm based on improved YOLOv8. Firstly, the large kernel block of UniRepLKNet is integrated into C2f module to construct the C2f_LarK structure. This architecture expands the receptive field and reduces redundant computation without increasing network depth. Secondly, a Det_Tiny detection head based on FasterNet partial convolution (PConv) is employed to achieve hardware-friendly optimization of the classification branch through adaptive feature redundancy compression. Finally, the feature fusion module (Fusion) is designed based on CGAFusion. It enhances edge detail detection capabilities through cross-semantic feature interaction. Experimental results show that compared with standard YOLOv8n model, the proposed algorithm reduces parameter count by 36.6%, model size by 31.1%, and computational complexity (GFLOPs) by 40.7%. It maintains the same detection accuracy of mAP@0.5 and improves mAP@[0.5:0.95] by 0.3%. A balance between accuracy and efficiency is achieved. It provides a feasible lightweight solution for intelligent inspection of transmission lines.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

WL: Funding acquisition, Methodology, Writing – original draft, Writing – review and editing. TY: Investigation, Methodology, Writing – original draft. DS: Conceptualization, Investigation, Writing – original draft. YL: Project administration, Writing – original draft. LZ: Conceptualization, Formal Analysis, Writing – original draft, Writing – review and editing. JL: Project administration, Software, Writing – review and editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was funded by Science and Technology Project of State Grid Sichuan Electric Power Company: Development of a portable intelligent detection device for tree and bamboo hidden dangers in transmission lines (No. B7190725006G).

Conflict of interest

Authors WL, TY, DS, and YL were employed by State Grid Sichuan Electric Power Company Guangyuan Power Supply Company.

The remaining author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that this work received funding from State Grid Sichuan Electric Power Company. The funder had the following involvement in the study: study design, collection, analysis, interpretation of data, the writing of this article, and the decision to submit it for publication.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bin, F., He, J., Qiu, K., Hu, L., Zheng, Z., and Sun, Q. (2025). CI-YOLO: a lightweight foreign object detection model for inspecting transmission line. *Measurement* 242, 116193. doi:10.1016/j.measurement.2024.116193

Bochkovskiy, A., Wang, C. Y., and Liao, H. Y. M. (2020). YOLOv4: optimal speed and accuracy of object detection. *arXiv preprint, arXiv:2004.10934*. doi:10.48550/arXiv.2004.10934

Chen, J., Kao, S. H., He, H., Zhuo, W., Wen, S., Lee, C. H., et al. (2023). "Run, don't walk: chasing higher FLOPs for faster neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (Vancouver, Canada). 17–23 June 2023.

Chen, Z., He, Z., and Lu, Z. M. (2024a). DEA-Net: single image dehazing based on detail-enhanced convolution and content-guided attention. *IEEE Trans. Image Process.* 33, 1002–1015. doi:10.1109/TIP.2024.3354108

Chen, Z., Yang, J., Feng, Z., and Zhu, H. (2024b). RailFOD23: a dataset for foreign object detection on railroad transmission lines. *Sci. Data* 11 (1), 72. doi:10.1038/s41597-024-02918-9

Ding, X., Zhang, Y., Ge, Y., Zhao, S., Song, L., Yue, X., et al. (2024). "UniRepLKNet: a universal perception large-kernel convnet for audio, video, point cloud, time-series, and image recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (Seattle, USA), 16–22.

Faisal, M. A. A., Mecheter, I., Qiblawey, Y., Fernandez, J. H., Chowdhury, M. E., and Kiranyaz, S. (2025). Deep learning in automated power line inspection: a review. *Appl. Energy* 385, 125507. doi:10.1016/j.apenergy.2025.125507

Hao, J., Yan, G., Wang, L., Pei, H., Xiao, X., and Zhang, B. (2024). A lightweight transmission line foreign object detection algorithm incorporating adaptive weight pooling. *Electronics* 13 (23), 4645. doi:10.3390/electronics13234645

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Salt Lake City, UT, USA).

Lin, T. Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision ICCV* (Venice, Italy), 22–29.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., et al. (2016). "SSD: single shot multibox detector," in *Proceedings of the European conference on computer vision*. Amsterdam, Netherlands: ECCV. 11–14 October 2016.

Liu, B., Huang, J., Lin, S., Yang, Y., and Qi, Y. (2021). "Improved YOLOX-S abnormal condition detection for power transmission line corridors," in *Proceedings of the IEEE 3rd international conference on power data science ICPDS*. Harbin, China, 17–19 December 2021.

Liu, C., Ma, L., Sui, X., Guo, N., Yang, F., Yang, X., et al. (2023). YOLO-CSM-based component defect and foreign object detection in overhead transmission lines. *Electronics* 13 (1), 123. doi:10.3390/electronics13010123

Lu, L., Chen, Z., Wang, R., Liu, L., and Chi, H. (2023). Yolo-inspection: defect detection method for power transmission lines based on enhanced YOLOv5s. *J. Real-Time Image Process.* 20 (5), 104. doi:10.1007/s11554-023-01360-1

Maduako, I., Igwe, C. F., Abah, J. E., Onwuasaanya, O. E., Chukwu, G. A., Ezeji, F., et al. (2022). Deep learning for component fault detection in electricity transmission lines. *J. Big Data* 9 (1), 81. doi:10.1186/s40537-022-00630-2

Pan, W., Huan, W., and Xu, L. (2024). "Improving high-voltage line obstacle detection with multi-scale feature fusion in YOLO algorithm," in *Proceedings of 6th international conference on electronics and communication, network and computer technology (ECNCT)* (Guangzhou, China), 19–21.

Peng, L., Wang, K., Zhou, H., Ma, Y., and Yu, P. (2025). YOLOv7-CWFD for real-time detection of bolt defects on transmission lines. *Sci. Rep.* 15 (1), 1635. doi:10.1038/s41598-024-81386-y

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Analysis Mach. Intell.* 39 (6), 1137–1149. doi:10.1109/TPAMI.2016.2577031

Sankuri, R. S., Sristy, N. B., and Karri, S. P. K. (2025). Accurate insulator defect detection in power transmission lines using semi-supervised hybrid DETR with advanced loss methods. *J. Real-Time Image Process.* 22 (5), 1–14. doi:10.1007/s11554-025-01760-5

Shao, Y., Zhang, R., Lv, C., Luo, Z., and Che, M. (2024). TL-YOLO: foreign-Object detection on power transmission line based on improved YOLOv8. *Electronics* 13 (8), 1543. doi:10.3390/electronics13081543

Song, Y., Zhou, Z., Li, Q., Chen, Y., Xiang, P., Yu, Q., et al. (2021). "Intrusion detection of foreign objects in high-voltage lines based on YOLOv4," in *Proceedings of the 6th international conference on intelligent computing and signal processing ICSP* (China: Xi'an), 9–11.

Wang, C. Y., Bochkovskiy, A., and Liao, H. Y. M. (2023). "YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (Vancouver, Canada). 17–23 June 2023.

Wang, Z., Yuan, G., Zhou, H., Ma, Y., and Ma, Y. (2023). Foreign-object detection in high-voltage transmission line based on improved YOLOv8m. *Appl. Sci.* 13 (23), 12775. doi:10.3390/app132312775

Yan, X., Wang, W., Lu, F., Fan, H., Wu, B., and Yu, J. (2025). GFRF R-CNN: object detection algorithm for transmission lines. *Comput. Model. Eng. and Sci.* 82 (1), 1439–1458. doi:10.32604/cmc.2024.057797

Zhou, L., Li, S., Zhu, Z., Chen, F., Liu, C., and Dong, X. (2024). Improved YOLOv5 foreign object detection for transmission lines. *Optoelectron. Lett.* 20 (8), 490–496. doi:10.1007/s11801-024-3218-y

Zhu, K., Xu, C., Wei, Y., and Cai, G. (2022). Fast-PLDN: fast power line detection network. *J. Real-Time Image Process.* 19 (1), 3–13. doi:10.1007/s11554-021-01154-3