ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

RBI-YOLO: A Defect Detection Algorithm for Transmission Lines

Yangwei Xia[1], *Shuqing Wang*[1,*] *and Haoyu Hu*[1]

[1] School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, Hubei, China

SUMMARY: *Aiming at the problems of low detection accuracy, difficult feature extraction and large amount of model calculation of small target defects in transmission lines under UAV inspection scenarios, a transmission line defect detection algorithm RBI-YOLO based on YOLOv8 n is proposed. The RFCAConv (Receptive Field Coordinate Attention Convolution) module integrating receptive field and coordinate attention is introduced to optimize the backbone network, strengthen defect feature extraction and suppress background interference. The neck network is reconstructed, and the P2 shallow prediction branch is added to make full use of the high-resolution fine-grained features. At the same time, the BiFPN (Bidirectional Feature Pyramid Network) is used to optimize the aggregation path of cross-scale features, so as to improve the fusion and perception ability of the model to the shallow features of small targets. The Inner-MPDIoU (Inner Multi-Patch Distance Intersection over Union) loss function is designed to optimize the regression accuracy and stability of dense small target bounding box. The experimental results on the transmission line defect dataset show that compared with the baseline model YOLOv8n, the accuracy, recall, mAP50, and mAP50-95 of RBI-YOLO are increased by 1.3, 5.2, 3.9, and 4.2 percentage points, respectively. At the same time, the number of model parameters is reduced by 23.9 %, and the volume is reduced by 14.4 %, which can be effectively applied to the transmission line defect detection task.*

KEYWORDS: *UAV inspection; defect detection; small target; improving YOLOv8n; loss function*

1 Introduction

The difficulty of recognizing transmission line defects does not exist in image collection, but in the small target sizes, strong background interferences, coexisting types, and scattered distributions that are met when drones carry out inspections. Related review documents point out that suspended foreign objects, damaged insulators, and abnormal fittings have become the main detection objects in visual power line inspections. Complex land forms, changing shooting distances, and unsTab. lighting situations further increase the possibility of small target omissions; among all these items, insulator defects still belong to one of the most deeply researched task types [1-3]. Methodological evolution displays that object detection has step by step moved from candidate region-dependent frameworks to end-to-end detection. Therefore, model improvements have also transferred from only chasing accuracy to balancing detection efficiency, lightweight deployment, and multi-scale representation capabilities thus [4]. In tasks of line inspection, methods which are based on improved YOLO can at present cover multiple target categories, for instance, foreign objects and electrical components. However, obvious

*ethan_xhide@163.com
https://doi.org/10.65102/is2026837
Published: 30 April, 2026

trade-off situations still exist between complicated backgrounds, dense distributions of small objects, and real-time inference carried out on board[5].

Early two-stage methods such as R-CNN and Fast R-CNN showed representative ability in feature expression and localization precision. However, because their inference processes are very long and parameter quantities are very large, hence they cannot satisfy real-time inspection requirements of UAVs[6, 7]. Therefore, later study has more and more concentrated on light-weight single-stage detection frameworks. Existing research mainly walks three roads: first, compress network architectures to raise edge device deploy ability; second, strengthen the transfer between shallow-layer details and high-layer semantics; third, optimize detection heads and feature fusion methods for small objects. Correspondingly, improved light-weight YOLOv5s variants have been used to insulator defect identification, with LiteYOLO-ID further strengthening light-weight deployment capabilities and MAP-YOLOv8 raising detection performance in insulator scenarios[8-11]. In recent days, related research keeps moving forward via the path of ‘small object enhancement + structural lightweighting’; this path includes YOLOv11n-based insulator defect detection, real-time bolt defect detection, and multi-module collaborative optimisation networks for insulator detection[12-14]. However, existing methods have two deficiencies: first, they make inadequate usage of detail response and spatial localisation information for distant, small-scale defects, which leads to missed detections and false positives; second, their multiscale fusion granularity is not sufficient, hence bounding box regression stability remains persistently weak in dense small-object scenarios.

To solve these limit problems, this paper puts forward RBI-YOLO, which is an improved YOLOv8 model made specially for transmission line defect detection. The backbone part adds an RFCACov module, therefore enhancing the ability of defect detail description and space position fixing. A P2 detection layer combined with BiFPN is added into the neck network, thus raising the efficiency of multi-scale feature fusion. Simultaneously an Inner-MPDIoU loss functio. is constructed as a contribution to bounding box regression sTab. limited to dense small targets. This finds preferable scale for detection, size of model and real-time capabilities for the use of unmanned aerial vehicle inspection.

2 Algorithm Principles

2.1 YOLOv8

YOLOv8, as a new generation of advanced object detection model, builds on the strengths of its predecessors with multiple new modules for dual uplift of capabilities and flexibility. Internal to YOLOv8, this manifests in three ways: a new backbone architecture, an Anchor-Free shift for detection heads, and a better optimised loss function. This enables the ability to run efficiently from CPU to GPU across heterogeneous hardware stacks while getting state-of-the-art detection performance that beats almost every mainstream algorithm. YOLOv8 continues the YOLOv5 design philosophy by constructing five scale variants, N/S/M/L/X, according to model scaling factors for deployment in any application scenario. The original C3 modules in the YOLOv5 backbone and neck networks have been replaced with, “more gradient efficient” C2f structures. This design serves a purpose in assisting the feature learning power of convolutional neural networks, with decent complexity and reasonable gradient streaming. By matching the number of channels across all scale models, the adaptability and general average capability of this architecture is (lightly improved). The detection head heavily rewritten adopting the current mainstream decoupled architecture to separate classification from localisation and achieving the last step of Anchor-Based to Anchor-Free detection. At the level of the loss function, we design a Task-Aligned Assigner for more accurate positive-negative

counting and propose Distribution Focal Loss to help boost accuracy of regression. YOLOX's disable Mosaic augmentation for last ten epochs of training is adopted, too. Following this can stabilise the training and even improve ultimate recogniton accuracy.

2.2 RBI-YOLO

In response to the problems encountered in power line inspections by drones, such as a small defect appearing in the images, the risk of losing feature detail after network prediction, and strong redundancy of network predictions caused by the complicated background, this study enhances the YOLOv8 model. The RBI-YOLO model improves the detection accuracy while reducing the model parameters, as depicted in Fig. 1. The specific method to enhance a conventional YOLOv8 is: to increase the network size with a RFCAConv module that integrates receptive field attention and coordinate attention, which can expand the effective perception of the network range and retain the location information of the object while guiding the network to pay more attention to defect information and suppress interference from background textures; then, to enhance the shallow features with P2 scale detection branch in the neck network, and moreover to combine with BiFPN, which achieves the bidirectional weighted fusion of the multi-scale features, thereby improving the transmission and aggregation of minute target features; finally, we construct an Inner-MPDIoU loss function to compensate for its position deviation for the close-packed small objects and the unsTab. bounding box regression, which integrates the constraint property of Inner-IoU with the geometric metric of MPDIoU, thus enhancing the accuracy of bounding box fitting and regression stability.

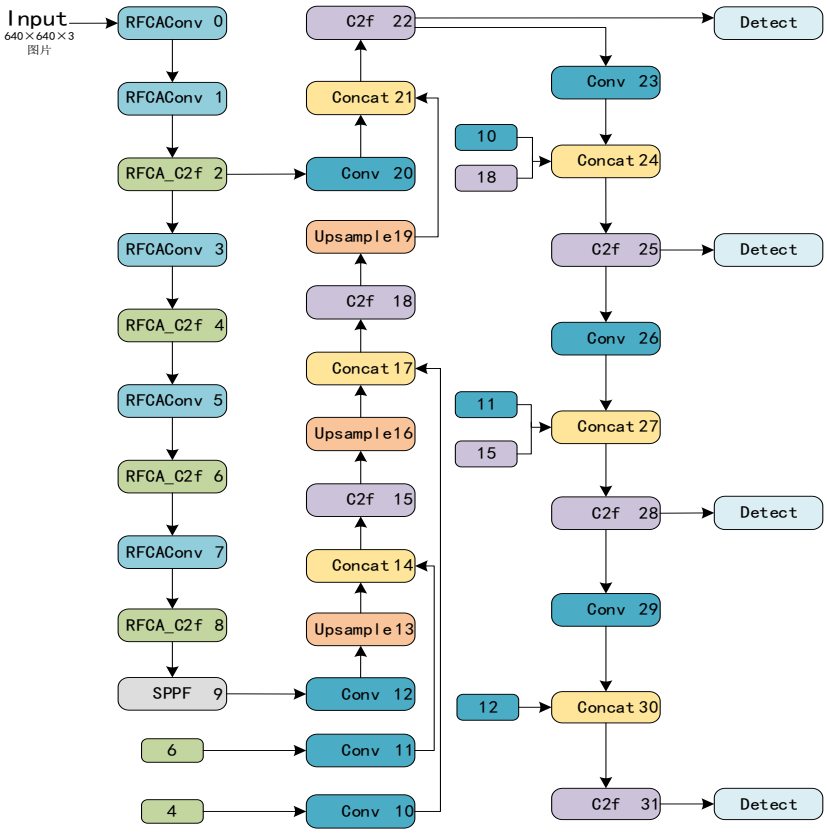


Figure 1: RBI-YOLO Architecture Diagram

2.3 RFCA-Enhanced Backbone Network

To tackle the YOLOv8 backbone network issue of limited small defect detection capabilities on power lines due to clear background interference (caused by information sharing in traditional convolutional channel sharing parameters), we propose the Receptive Field Coordinate Attention Convolution (RFCAConv) module, which merges the advantages of RFA with CA when they are created, then replaces the backbone network's native convolutional modules and C2f modification, improving its extraction capacity and thereby providing high-quality featuring support for subsequent small object detection.

2.3.1 RFCAConv

The intuition behind Receptive Field Attention (RFA) is to dynamically produce the spatial features as the receptive fields of convolution kernel size, and associate an independent weight to each receptive field to obtain a ’parameter-free’ convolution operation. These weights are learned by training, enabling the model to dynamically adapt its focus on the most salient aspects of the task while maximizing the context that helps it. RFA actually takes care of retrieving multi-scale receptive fields with batch convolutions and generates the spatial attention weights with a light interactivity module: it first compresses and aggregates the spatial information through a global average pooling layer, then it computes the weights through batch convolutions of size of 11 with a Softmax function. Thus enhancing the feature response on task relevant areas meanwhile suppressing the background noise interferences. As stated in the paper (1):

$$F = Soft\max(g^{1\times1}(Avg\text{pool(X)})) \times \text{Re}\,LU(Norm(g^{k\times k}(X))) \tag{1}$$

The essence of Coordinate Attention (CA) is explicitly encoding positional sensory awareness. By separately encoding along height and width, the model learns channel dependencies and target location information concurrently, thus allowing it to focus its attention more effectively on important areas. The CA architecture is visualised in Fig. 2(a). The Receptive Field Coordinate Attention Convolution (RFCAConv) used here is similar to receptive field attention but applies coordinate attention mechanisms to it, thereby further normalising convolutional operations and modeling sensory awareness.

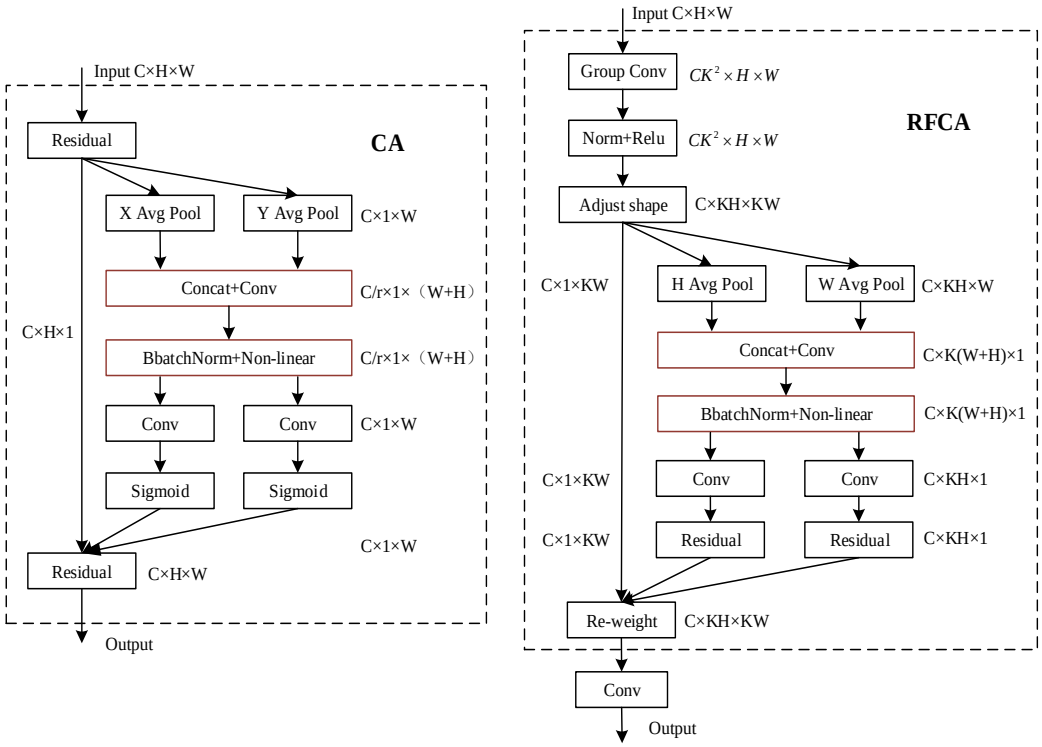
The specific fusion is as follows: this module performs feature extraction by batch convolutions, and then performs feature matching of the features in each channel with the features of the other channels and obtains the multi-receptive field spatial features. Subsequently, the pooling obtains the feature map, and then stack up and decompose it into a one-dimensional feature along the direction of the two spatial dimensions horizontally and vertically. Perform batch normalisation, and use the h-swish activation function to realise feature transformation into the second space by fusing the spatial information in the two’s extremes. After feature decomposition and dimensionality-increasing 1×1 convolutions, Sigmoid yields the attention weights in both spatial directions. Finally, the weights are weighted fused into the batch-convolved features and the receptive field spatial features are reconstructed and then sent to a convolutional layer.

$$F = C^{k}\left(CA\left(\text{Re}\,Lu\left(Norm\left(g^{k\times k}\right)\right)\right)\right) \tag{2}$$

In Equation (1)(2), the symbol $g^{\mathrm{k\times k}}$ denotes a batch convolution operation of size $\mathrm{k}\times\mathrm{k}$, where k represents the size of the convolutional kernel employed. C^{k} denotes a convolution

with stride k, Norm denotes the normalisation process, CA denotes the coordinate attention feature transformation operation, X denotes the input feature map, and F denotes the final feature map output from both operations.

Compared to standard convolutions having shared parameters, RFCAConv alleviates the degradation of feature representation, thus improving the modelling of spatial correlations. It demonstrates significant advantages in small object feature extraction and background suppression. The architecture of RFCAConv can be seen in Fig. 2(b).



(a) CA Structure Diagram (b) RFCAConv Structure Diagram

Figure 2: Schematic diagram of the CA and RFCAConv architecture

2.3.2 RFCA_C2F

Within the YOLOv8 backbone network, the C2f module assumes the core task of feature fusion and propagation by cascading multiple Bottleneck structures. However, when processing small defects on power transmission lines, the traditional convolutions relied upon by this module exhibit limited perceptual capability for fine-grained features, resulting in defect characteristics being easily obscured by complex backgrounds. To address this issue, this paper introduces RFCAConv to replace the convolutional layer within the Bottleneck structure, thereby constructing the RFCA_C2f module. This enhancement preserves the original structural framework of C2f while enabling the feature extraction process to benefit simultaneously from the guidance of receptive field attention and coordinate attention. This approach is expected to more effectively focus on the details and spatial locations of defect regions, thereby enhancing feature discriminative power and robustness against interference. Rfca_c2f structure diagram is shown in Fig. 3.

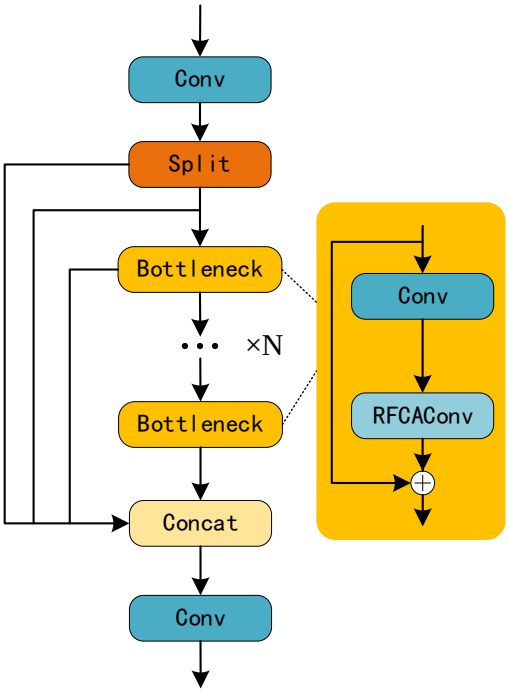


Figure 3: RFCA_C2f Structural Diagram

2.4 BiFPN Reconstructed Neck Network Architecture

The YOLOv8 model combines feature maps for objects at multiple scales through a Path Aggregation Network (PAN). Rather than relying on a standard Feature Pyramid Network (FPN), PAN introduces an extra bottom-up feature propagation branch that then interacts with the existing top-down feature fusion branch supplying the full path across the model for feature interaction. Given this PAN structure, feature maps at each level can fuse information with features of variable scales in the vicinity as well as extracting and outputting multi level features sequentially. Overall this strengthens the link between features of different levels for better object detection. However, where features are fused, all features at all resolutions are treated equally without weighting the influence different features levels have on the final fused features.

To overcome these limitations, in this paper, we equip a Bidirectional Feature Pyramid Network (BiFPN) on YOLOv8's feature fusion module. The architecture of bifpn is illustrated in Fig. 4. BiFPN inherits the advantages of both PAN's bottom-up and FPN's top-down bidirectional fusion, but also further optimises the feature propagation mechanism: introduces learnable weights for features at different resolutions, enabling adaptative adjustments to the contribution of various features from lower to higher levels; this allows high-resolution detail features and low-resolution semantic features to "actively" participate in fusion based on the specific task; adds cross-scale skip connections to allow comprehensive interaction of feature information across multi-scale channels. The weighted bidirectional fusion mechanism inherits the efficient multi-scale fusion capabilities of PAN while significantly improving the reasonableness and specificity of feature usage. It further boosts the model's detection performance of objects at different scales within a controllable computation cost.

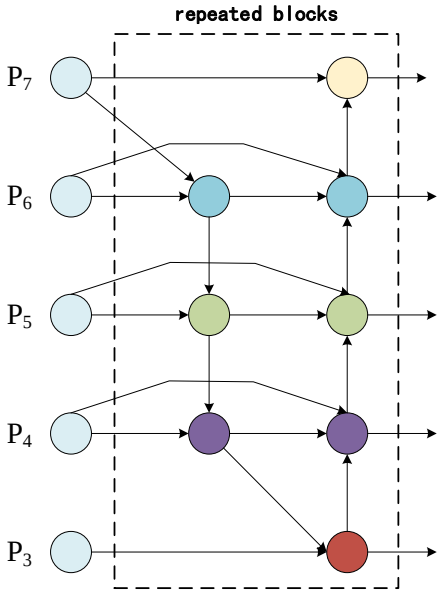


Figure 4: BiFPN Structural Diagram

2.5 Integration of the P2 Small Object Detection Layer

For the detection of tiny defects on high-voltage transmission lines, where insulators and other similar defects are extremely small, scarce in features, and heavily disturbed by background clutter, YOLOv8's native detection architecture is deficient (large loss of detail in small objects) given the requirement for being downsampled at the input layer and a high downsampling factor for deep features, rendering it incapable of achieving precise. In this paper we propose an P2 layer small object detection branch based on an improved feature pyramid structure, and rely on the use of high-resolution feature layer to accurately capture small object's details, while the RFCAConv attention mechanism enhances semantic feature information. This leads to an overall improvement in detection performance of the model for transmitting small defects.

The additions for the P2 small object detection layer are as follows. Features from the native C2f module after the second RFCAConv module in the backbone are defined as P2 layer features. After being convolutionally downsampled, P2 layer features are concatenated and fused with the features of the same layer in the native network and passed to detection head. To downsample the P2 layer feature map to a matching scale for points in the detection layers, two further convolutional modules are added that progressively remove spatial resolution from the features. In order to avoid excessive computational overhead caused by large feature maps and at the same time satisfy the practical requirements of deployment on UAV platforms, we perform yet another downsampling operation on the P2 layers processed features before passing the final features through to the detection head. With this approach, we essentially construct a four-layer detection head consisting of the P2, P3, P4, and P5 layers. This fully utilizes the shallow-layer features to improve detection capabilities for small defects on power transmission lines, while at the same time having a downsampling strategy to limit overall computational overhead, thus improving the model's applicability on realistic engineering tasks.

2.6 Inner-MPDIoU

The overall loss function of YOLOv8 comprises two components: classification loss and regression loss. The classification loss (L_{cls}) employs binary cross-entropy loss (BCE), whilst the regression loss utilises complete intersection with union loss (CIoU) and distributed focus

loss (DFL). CIoU measures the intersection-over-union loss (L_{box}) between predicted and ground-truth bounding boxes, whilst DFL quantifies the distance loss (L_{dfl}) between regression predictions and target bounding boxes. Consequently, YOLOv8's loss function L_{LOSS} may be described as:

$$L_{LOSS} = \lambda_1 L_{cls} + \lambda_2 L_{box} + \lambda_3 L_{dfl} \tag{3}$$

In the formula, λ_1, λ_2, and λ_3 are all weighting values.

Since the CIoU loss function only considers geometric factors such as the distance and aspect ratio between predicted boxes and ground truth, it lacks attention to the quality of annotated samples. This leads to slow convergence and poor generalization performance. Inner-IoU introduces auxiliary boxes with a scale factor ratio, calculating IoU loss using auxiliary bounding boxes of varying sizes to accelerate the bounding box regression process and overcome the limitations of existing methods in terms of generalization capability.

However, the previous approach only considers the overlap of boxes while ignoring their position and entire overlap. In the worst case, two boxes will have the same intersection-over-union ratio yet their respective positions may be completely different, e.g. having identical values in terms of being completely divergent. In order to more clearly assess how well a ground-truth box matches with a predicted box for the purposes of scoring MPDIoU, a distance penalty terms is added between ground-truth and predicted where the distance penalizes any difference in terms of their boxes' positions. By taking Inner-IoU and also utilizing this distance penalty, a more accurate metric results. This metric considers both the degree of overlap between boxes (via Inner-IoU) and their positional relationship (via distance penalty), enabling a more accurate assessment of the match between predicted and ground-truth boxes. This contributes to improving the performance of object detection algorithms. The inner_mpdiou structure is shown in Fig. 5, the combined Inner-MPDIoU expression is as follows:

$$L_{Inner-MPDIoU} = 1 - (Inner - MPDIoU) \tag{4}$$

The expression for $Inner - MPDIoU$ in Formula (4) is as follows:

$$Inner - MPDIoU = InnerIoU - \frac{d_1^2}{h^2 + w^2} - \frac{d_2^2}{h^2 + w^2} \tag{5}$$

The expression for $Inner - IoU$ in Formula (5) is as follows:

$$Inner - IoU = \frac{\text{int}\, er}{union} \tag{6}$$

$$\text{int}\, er = (\min(b_r^{gt}, b_r) - \max(b_l^{gt}, b_l)) * (\min(b_b^{gt}, b_b) - \max(b_t^{gt}, b_l)) \tag{7}$$

$$union = (w^{gt} * \text{h}^{\text{gt}}) * (\text{ratio})^2 + (w * h) * (ratio)^2 - \text{int}\, er \tag{8}$$

In Equation (7), b^{gt} and b represent the ground truth box and the predicted box, respectively. In Equation (8), w and h denote the width and height of the box, respectively. The subscripts and their meanings are as follows.

$$b_l = x_c - \frac{w*ratio}{2}, b_r = x_c + \frac{w*ratio}{2} \tag{9}$$

$$b_t = y_c - \frac{h*ratio}{2}, b_b = y_c + \frac{h*ratio}{2} \tag{10}$$

The top-left corner of the ground truth box is at coordinates (x_1, y_1), and the bottom-right corner is at (x_2, y_2). The top-left corner of the predicted box is at (x_3, y_3), and the bottom-right corner is at (x_4, y_4). The expression for d_1, d_2 in coordinate system (5) is as follows:

$$d_1^{\,2} = (x_3 - x_1)^2 + (y_3 - y_1)^2, d_2^{\,2} = (x_4 - x_2)^2 + (y_4 - y_2)^2 \tag{11}$$

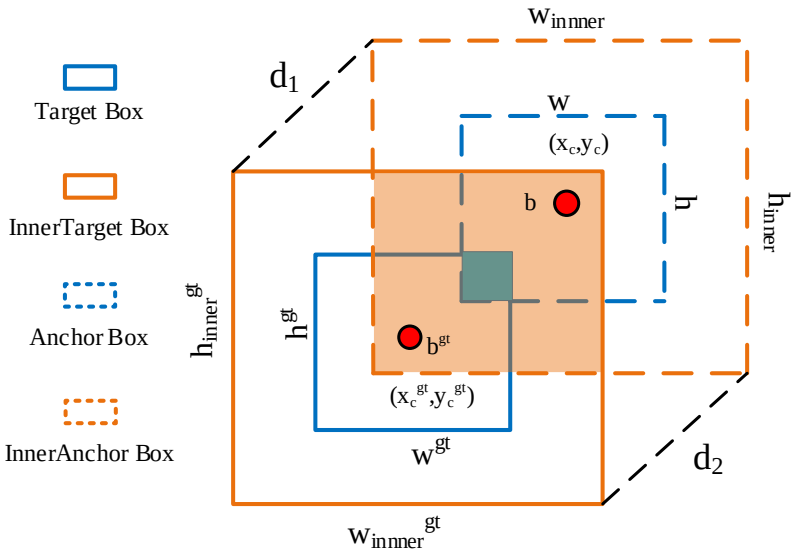


Figure 5: Inner_MPDIoU

3 Datasets and Experimental Setup

3.1 Datasets

The datasets used in this study comprise four components: the China Power Line Insulator Dataset (CPLID), the Insulator Defect Image Dataset (IDID), transmission line images sourced from the internet, and synthetic images generated using image synthesis techniques based on these images. Data annotation was performed using LabelImg, with five distinct label categories: insulator flashover, damage, missing insulator, bird nest, and fallen vibration damper. To aid the model in learning and convergence, we applied scale and cropping techniques to synthesise more example images from the original, yielding a total of 6,525 example images. For experimentation, the dataset was randomly split into training, validation and test sets in the ratio 7:2:1.

3.2 Evaluation Metrics

For the study analysis four metrics which are Precision (P) , Recall(R), Average Precision (AP) and Mean Average Precision (mAP)) based on these metrics will show the validity and

correctness of that multi-object detection system for power transmission lines. Higher the metrics will be more better. The respective formulas for these metrics are as follows.

$$P = \frac{TP}{TP + FP} \tag{12}$$

$$R = \frac{TP}{TP + FN} \tag{13}$$

Among these, TP denotes the number of correctly predicted positive samples, FN denotes the number of incorrectly predicted negative samples, and FP denotes the number of incorrectly predicted positive samples.

$$AP = \int_0^1 PRdr \tag{14}$$

$$mAP = \frac{1}{N} \sum_{i=0}^{n} APi \tag{15}$$

AP stands for average precision for a single class label (as seen in Formula). mAP is short for mean average precision across all classes, and is used to measure relative differences in performance between different models or parameters (as seen in Formula)(15).

3.3 Experimental Environment and Configuration

Experiment platform: Windows 11, Intel Core i5-12600KF CPU, 16GB VRAM NVIDIA RTX 4060 graphic card. Environment Configuration: Python 3.8.18, torch 2.1.1, CUDA 11.8, YOLO v8.0.202. Parameter Settings Ref. Tab. 1.

Table 1: Parameter Settings Tab.

Settings	Parameters
Optimizer	SGD
Batch	8
Epochs	300
Imgsz	640
Learning rate	0.01
Amp	true

4 Experimental Results and Analysis

4.1 Comparative Experiments on Improved Loss Functions

To further test the effectiveness of the ratio coefficient in Inner-MPDIoU on detection task performance, some comparison experiments were conducted, with the results shown in Tab. 2. One of the important characteristics of power line defect detection dataset is small target scale, so this experiment has some special research meaning. When ratio > 1, algorithm will generate an auxiliary bounding box larger than the gt. This design can improve the accuracy of regression for those low IoU samples.

During experimentation, the ratio coefficient was varied from 0.9 to 1.6 in increments of 0.1. Results indicate that model performance does not increase linearly with higher ratio values. Within the range of 0.9 to 1.4, performance shows a gradual improvement, peaking at ratio=1.4. Beyond this point, performance began to decline as the ratio value continued to increase. These results indicate that the optimal ratio coefficient value is highly dependent on the dataset characteristics and must be fine-tuned based on the target distribution features of the experimental dataset. Generally, for datasets containing a large number of large targets, the ratio should be less than 1; for datasets dominated by small targets, the ratio should be greater than 1.

Table 2: Ratio Experiment Results

ratio	P	R	mAP_{50}	map50-95
0.9	0.944	0.841	0.904	0.615
1	0.947	0.850	0.905	0.618
1.1	0.939	0.850	0.906	0.621
1.2	0.948	0.843	0.906	0.621
1.3	0.937	0.844	0.909	0.618
1.4	0.949	0.846	0.911	0.622
1.5	0.937	0.844	0.903	0.615
1.6	0.919	0.851	0.898	0.611

4.2 Ablation Studies

To validate the effectiveness of the proposed improvement strategies, YOLOv8n was adopted as the baseline model. By sequentially adding the enhancement modules, different combination models were constructed for comparative ablation studies. The experimental results are presented in Tab. 3. Group 1 represents the original YOLOv8n algorithm, serving as the performance reference baseline. Group 2 replaces selected convolutional modules in the backbone network with RFCAConv, including the backbone's Conv modules and some convolutional modules in C2f. This group achieves improvements of 0.1% in accuracy, 2.8% in recall, 1.7% in mAP_{50}, and 2.8% in mAP_{50-95}, at the cost of a slight increase in parameters and computational overhead. Group 3 optimizes the neck network using BiFPN. Group 4 adds a P2 small object detection layer on top of Group 3. Compared to Group 3, its accuracy, recall, mAP_{50}, and mAP_{50-95} improved by 0.6%, 3%, 0.7%, and 1.2% respectively, fully validating the effectiveness of the small object detection layer for shallow defect feature extraction. Group 5 introduces the Inner-MPDIoU loss function. After multiple ratio-adjustment experiments, optimal parameter configurations were identified, leading to steady performance improvements. Group 6 simultaneously improves the neck network architecture based on Group 2. Compared to Group 2, which only modified the backbone network, Group 6 significantly reduces parameter count and model size while achieving 1.1%, 2.2%, 1.4%, and 1% improvements in accuracy, recall, mAP_{50}, and mAP_{50-95} respectively, demonstrating the advantages of concurrent backbone and neck optimization. Group 7 represents the final model integrating all improvement strategies, achieving accuracy, recall, mAP_{50}, and mAP_{50-95} of 95.8%, 88.5%, 93.9%, and 65.2%, respectively. With only 2.29 million parameters, the model size was reduced by 14.3%. Compared to the baseline model, the improved model significantly enhances performance while reducing both parameter count and model size, effectively accomplishing the task of detecting defects in transmission lines.

Table 3: Ablation Study of the Improved Model

Group	RFCA	BiFPN	P2	Inner-MPDIoU	P	R	map50	map50-95	Parameters/M
1					0.945	0.833	0.900	0.61	3.01
2	1				0.946	0.861	0.917	0.639	3.14
3		1			0.944	0.835	0.905	0.613	1.99
4		1	1		0.950	0.865	0.912	0.625	2.23
5				1	0.949	0.854	0.911	0.622	3.01
6	1	1	1		0.957	0.873	0.931	0.649	2.29
7	1	1	1	1	0.958	0.885	0.939	0.652	2.29

4.3 Comparative Experiments

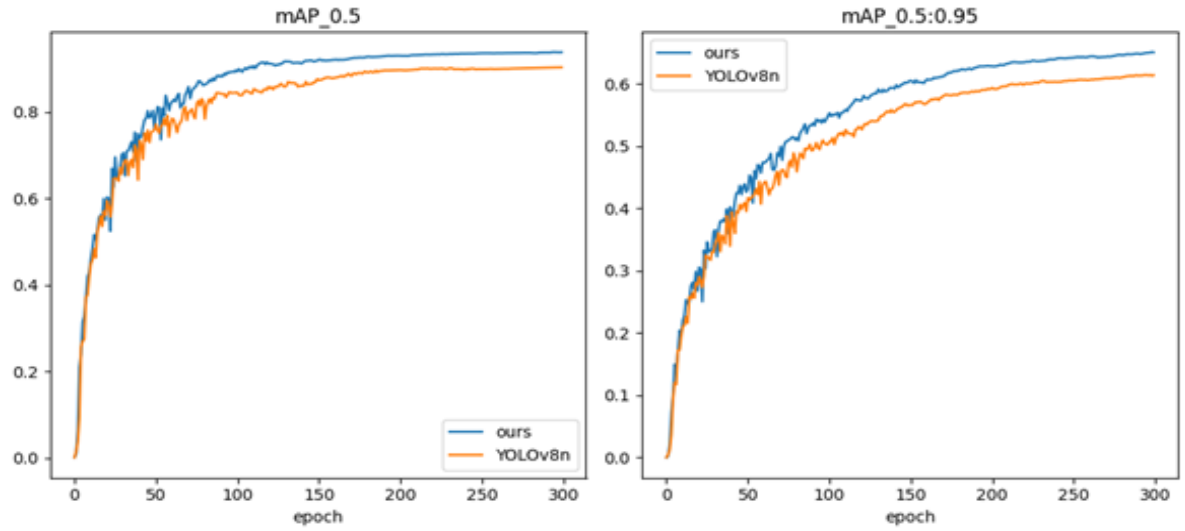
To comprehensively evaluate the performance of the RBI-YOLO algorithm in transmission line defect detection, comparative experiments were conducted against current mainstream lightweight models from the YOLO series. The results are presented in Tab. 4. The experiments demonstrate that the RBI-YOLO algorithm exhibits significant advantages in both detection accuracy and model lightweighting.

Table 4: Comparative Experiments

Model	P	R	map50	map50-95	Model Size/MB	Parameters/M
Yolov3n	0.882	0.865	0.890	0585	237	61.9
Yolov5n	0.917	0.852	0.904	0.607	5.30	2.503917
Yolov6n	0.927	0.788	0.859	0.561	8.29	4.234239
Yolov8n	0.949	0.833	0.900	0.610	5.96	3.006623
Yolov8s	0.952	0.892	0.940	0.687	21.4	11.127519
ASFYOLO	0.943	0.835	0.905	0.618	6.45	3.050
GOLDYOLO	0.951	0.824	0.909	0.614	12.5	6.001
Yolov11n	0.929	0.857	0.910	0.604	5.50	2.583129
ours	0.958	0.885	0.939	0.652	5.10	2.289192

For the core detection metrics, RBI-YOLO has accuracy of 95.8%, recall of 88.5% and mAP50 & mAP50–95 of 93.9% and 65.2% respectively. Compared with the mainstream object detection models YOLOv3, YOLOv5, YOLOv6, YOLOv8n, YOLOv11n, ASFYOLO, GOLDYOLO, RBI-YOLO is outstanding in all detection metrics, just falling short of YOLOv8s in mAP50-95 despite its model containing about 4.9 times more parameters! In terms of complexity, its parameters count is 2.29 million, 5.10 MB model size. Compared with YOLOv8n, counts and size are lower by 23.9 % and 14.4 % respectively. Although smaller in scale, RBI-YOLO still has detection accuracy exceeding lightweight benchmark models YOLOv5 and YOLOv11n. Compared with the YOLOv8s, there is still a slight gap in the mAP50-95 metrics (RBI-YOLO vs. YOLOv8s). However, the parameters are only 20.6% of the YOLOv8s and the model size is only 23.8% of YOLOv8s, and the aftermarket overhead is lower and relatively light. In addition, compared with the improved models proposed in recent years, the performance of RBI-YOLO still has certain advantages in both detection performance and model size. The comparative experiment of map50 and 50-95 is shown in Fig. 6. ASFYOLO enhances multi-scale feature representation by introducing an adaptive feature fusion mechanism. However, in this experiment, RBI-YOLO still achieves improvements of 1.5%, 6.1%, 3.0%, and 3.8% in accuracy, recall, mAP_{50} -score, and $\mathrm{m}AP_{50-95}$ -score, respectively. Concurrently, it reduces the number of parameters by 61.9% and shrinks the model size by

59.2%. GOLDYOLO improves detection efficiency through global-local information decoupling and efficient structural design. Nevertheless, RBI-YOLO outperforms it by achieving 1.5%, 6.1%, 2.9%, and 3.8% gains in accuracy, recall, mAP_{50}-score, and mAP_{50-95}-score, respectively, while further possessing a more significant advantage of lightweight model size. In summary, RBI-YOLO is a good trade-off method between detection accuracy and model complexity for power line defect detection tasks. It not only improves recognition performance for small defect targets but also achieves a small model size and low calculation overhead, making it suiTab. for deployment on resource-limited inspection platforms like drones and exhibiting high engineering application value.



(a) mAP50 comparative trial (b) mAP 50-95 comparison trial

Figure 6: mAP50 vs. mAP50-95 Comparison Experiment

5 Conclusions

In order to address the challenges of detecting small defect during drone inspections for power lines, including difficulties with identifying minor defects and deploying models. This paper proposes a lightweight way of detecting small defect called RBI-YOLO. Improvements to the original methodology will be addressed in three different areas:

(1) **Reconstructed backbone network**: To handle complex background interference, we propose a RFCA_C2f module where receptive field and coordinate attention are combined together. With the help of RFCA_C2f module, the network could learn to put attention in spatial and detail information of defect area which facilitates the feature extraction and removes background noise.

(2) **Enhanced Multi-Scale Feature Fusion**: To avoid false negatives for minute defects, a P2 small object detection layer is appended to neck network, fully extracting shallow, fine-grained features. Also, by using biFPN to help cross-layer feature propagation, our method strikes a balance between increasing small object rate and limiting total FLOPs so they run smoothly on edge devices on our drone.

(3) **Optimized Loss Function**: Designed specifically for this application, the Inner-MPDIoU loss function uses both internal anchor constraints as well as multidimensional geometric relationships to correct for the inherent bias in localized small-scale dense objects and stabilize their corresponding bounding box regression processes.

Experiments conducted on the power line defect dataset demonstrate that RBI-YOLO significantly outperform the mainstream methods in overall performance and achieves sufficient detection accuracy with a lightweight design to meet the practical requirements for real-time drone inspections. Future work will focus on the improvement of feature alignment strategies and attempt to transfer this algorithm to further power inspection tasks, such as foreign object intrusion and tower deformation.

Author's Profile

Jielun Zhou was born inTaibei, Taiwan. China, in 1979. He obtained a bachelor's degree from Wuhan University in China. He currently studying at the School of Economics and Management, Wuhan University. His main research direction is Western economics and macroeconomic management.

References

[1] Faisal M A A, Mecheter I, Qiblawey Y, et al. Deep learning in automated power line inspection: A review[J]. Applied Energy, 2025, 385: 125507.

[2] Luo Y, Yu X, Yang D, et al. A survey of intelligent transmission line inspection based on unmanned aerial vehicle[J]. Artificial Intelligence Review, 2023, 56: 173-201.

[3] Liu Y, Liu D, Huang X, et al. Insulator defect detection with deep learning: A survey[J]. IET Generation, Transmission and Distribution, 2023, 17(16): 3541-3558.

[4] Li Z, Dong Y, Shen L, et al. Development and challenges of object detection: A survey[J]. Neurocomputing, 2024, 598: 128102.

[5] Wang S, Tan W, Yang T, et al. High-Voltage transmission line foreign object and power component defect detection based on improved YOLOv5[J]. Journal of Electrical Engineering and Technology, 2024, 19: 851-866.

[6] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[7] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[8] Wei L, Jin J, Deng K, et al. Insulator defect detection in transmission line based on an improved lightweight YOLOv5s algorithm[J]. Electric Power Systems Research, 2024, 233: 110464..

[9] Li D, Lu Y, Gao Q, et al. LiteYOLO-ID: A lightweight object detection network for insulator defect detection[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-12.

[10] Xu Z, Tang X. Transmission line insulator defect detection algorithm based on MAP-YOLOv8[J]. Scientific Reports, 2025, 15(1): 10288.

[11] Sun S, Chen C, Yang B, et al. ID-Det: Insulator burst defect detection from UAV inspection imagery of power transmission facilities[J]. Drones, 2024, 8(7): 299.

[12] Zhao J, Miao S, Kang R, et al. Insulator defect detection algorithm based on improved YOLOv11n[J]. Sensors, 2025, 25(5): 1327.

[13] Peng L, Wang K, Zhou H, et al. YOLOv7-CWFD for real time detection of bolt defects on transmission lines[J]. Scientific Reports, 2025, 15(1): 1635.

[14] Zhang Q, Zhang J, Li Y, et al. ID-YOLO: A multi-module optimized algorithm for insulator defect detection in power transmission lines[J]. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 1-11.