Journal of Electronic Research and Application, 2026, Volume 10, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Transmission Line Defect Detection Algorithm Based on Improved RT-DETR Model

Qi Wu*

School of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, China

**Correspondence author: Qi Wu, 2858478839@qq.com*

Abstract: This paper addresses the urgent need for high-precision and high-efficiency visual perception technologies in power equipment operation and maintenance under the background of rapid development of smart grids. It points out the performance limitations of the existing real-time target detection framework RT-DETR when handling small targets, dense targets, and complex backgrounds in power inspection scenarios. To overcome this bottleneck, this study proposes an improved backbone network model, DETR-EVA, based on an efficient visual attention mechanism (EVA). This model innovatively designs an attention computation structure with linear complexity by deeply integrating the EVA mechanism with the C2f module in the RT-DETR backbone network, and combines local detail perception and global dependency modeling capabilities. Its core lies in the introduction of a gated fusion mechanism, which significantly enhances the model's ability to model long-distance contextual relationships and the adaptive adjustment efficiency of feature weights while retaining the advantages of multi-branch feature extraction and fusion of the C2f module. Experiments were conducted on an inspection image dataset containing typical power equipment targets. The results show that compared with the original RT-DETR model, DETR-EVA improves the overall accuracy index mAP50-95 by 2.5%, reduces computational complexity by 14%, and reduces the number of model parameters by 27%. This effectively verifies that the proposed method can significantly improve the detection accuracy of small targets and complex scenes while maintaining real-time detection speed, providing a better visual solution for intelligent operation and maintenance of power equipment.

Keywords: RT-DETR; Defect detection; Efficient vision attention; C2f; Small object detection

Online publication: February 13, 2026

1. Introduction

The rapid development of smart grids has raised higher demands on the intelligence level of power equipment operation and maintenance. High-voltage transmission lines, exposed to the natural environment for extended periods, inevitably suffer from potential defects such as insulator self-explosion, shock absorber slippage, and bird nest construction. If these defects are not detected and addressed in a timely manner, they can easily escalate into serious failures like line breaks and tower collapses, triggering widespread power outages and posing a severe threat to people's lives and property, as well as to socio-economic order. Therefore, regular and efficient inspection of high-voltage transmission lines is an indispensable part of the preventive maintenance system in the power system.

Traditional transmission line inspection primarily relies on manual patrols. This method is not only inefficient and labor-intensive, but also limited by the vision and experience of the inspectors, making it difficult to detect some concealed defects. In recent years, with the maturity of drone technology, intelligent inspection has gradually become the mainstream method for transmission line inspection.

In recent years, deep learning technology, particularly object detection algorithms based on convolutional neural networks (CNN), has achieved a qualitative leap in detection accuracy and generalization performance, owing to its powerful end-to-end feature learning and expression capabilities, thereby revolutionizing the situation where traditional methods suffered from poor performance [1]. Object detection algorithms are mainly divided into two-stage detectors (such as R-CNN, Fast R-CNN, Faster R-CNN) and single-stage detectors (such as YOLO[6][7], SSD) [2–9].

Although the two-stage detector has high accuracy, its computational complexity is high and inference speed is slow, making it difficult to meet the urgent real-time requirements of inspection tasks. The YOLO series relies on prior anchor boxes, which limits its generalization ability. The non-maximum suppression post-processing is non-differentiable and inefficient, and the CNN structure has weak modeling of global contextual relationships in images.

To overcome these limitations, detection transformer (DETR) abandoned anchor boxes and NMS, utilizing the global attention mechanism of transformer to achieve end-to-end set prediction, significantly enhancing its global reasoning capability [10,11]. However, DETR also introduces new issues, including slow training convergence, high computational cost, and poor performance in detecting small targets, which limit its application in real-time inspection scenarios.

RT-DETR, as a real-time high-performance variant in the DETR series, achieves a good balance between speed and accuracy through efficient hybrid encoder and intra-scale feature interaction design [12]. However, there is still room for improvement in detection performance in complex scenes, especially in small object detection, dense object detection, and complex background processing. The original RT-DETR backbone network mainly has the following limitations: insufficient modeling of long-distance feature dependencies; limited feature representation ability, which is prone to false positives or missed detections in complex backgrounds or situations where the contrast between the target and background is low; and small object features are easily overwhelmed.

In response, this paper aims to make targeted improvements to the RT-DETR model to enhance its detection accuracy for small target defects on transmission lines, while maintaining its real-time advantage, thereby meeting the needs of practical engineering applications.

2. Literature review

2.1. Research on transmission lines based on RT-DETR

Early detection methods were primarily based on image processing techniques, such as edge detection, threshold segmentation, and texture analysis. These methods were computationally simple but lacked robustness and were highly susceptible to environmental interference. With the continuous development of deep learning, some scholars have conducted targeted research on the RT-DETR model.

For example, Li *et al.* addressed the issues of difficult-to-capture small-sized insulator defect features, insufficient utilization of contextual information, and unstable matching by designing a multi-scale backbone network, introducing a Self-Attention Upsampling (SAU) module, and a dedicated Insulator Defect (IDIoU) loss function [13]. This improved the model's detection capability for small defects, significantly enhancing average precision and enhancing detection stability. Bai *et al.* addressed the issues of shallow measurement and difficulty in quantification in traditional defect detection methods by establishing a magnetic flux leakage detection method and analyzing signal characteristics [14]. This improved the quantitative detection capability for U-shaped suspension ring defects, achieving high-precision, low-error defect identification. Huang *et al.* addressed the issues of existing detection models relying on a large amount of labeled data, bulky parameters, and the difficulty in balancing lightweight and performance [15]. By adopting a federated knowledge distillation framework combined with asynchronous aggregation and model freshness mechanisms, they improved the model's deployment capability on resource-constrained devices, achieving lightweight model implementation while enhancing detection accuracy and training efficiency. Xie *et al.* addressed the challenges of small target sizes, similar shapes, and occlusion leading to detection difficulties in power line defect detection [16]. By introducing a Transformer-based Power-DETR network, combined with multi-scale feature enhancement, contrastive denoising training, and mixed label assignment strategies, they improved detection accuracy and training stability. Chen *et al.* addressed the detection challenges of small defects on ultra-high voltage transmission lines being easily obscured and subject to strong complex background interference [17]. By adopting a feature focused diffusion network (FFDN) and dynamic range histogram self-attention (DHSA) mechanisms to improve the RT-DETR model, they achieved simultaneous optimization of detection accuracy and missed detection rate. This not only improved inspection efficiency by 60% but also significantly reduced energy consumption and carbon emissions, providing key technical support for low-carbon operation and maintenance of transmission lines.

In summary, these studies have demonstrated significant advantages in detecting small targets and overcoming the interference of occlusion and complex backgrounds. However, they generally suffer from issues such as complex model structures, large parameter counts, and high computational costs. To address this, this paper designs a lightweight and efficient RT-DETR model that can better detect minor defects while reducing detection costs.

2.2. Research on attention mechanism

Attention mechanisms originate from the simulation of the human visual system, and their core lies in guiding the model to allocate limited computational resources to the more critical parts of the input information. The specific development process is as follows:

(1) The self-attention mechanism has been introduced, which directly computes the associations between all elements within the global scope through Query, Key, and Value operations, bringing powerful contextual modeling capabilities to the model [10];

(2) Multi-head attention further extends this idea by enabling the model to learn information collaboratively from different representation subspaces through parallel computation of multiple attention heads [10];

(3) CA (Cross-Attention) integrates features from different modalities or sources, with Query derived from one feature and Key and Value derived from another [10]. The vision transformer (ViT) demonstrated for the first time that splitting an image into a sequence of patches and directly applying a Transformer encoder can achieve performance on par with or even surpass that of the most advanced CNN models

on image classification tasks [18]. This verifies the powerful ability of attention mechanisms in modeling global contextual dependencies in images. More importantly, attention mechanisms have given birth to groundbreaking object detection frameworks such as DETR [10];

(4) SENet (Squeeze-and-Excitation Attention) learns the dependencies between channels through global average pooling and a two-layer fully connected network, focusing solely on the channel dimension while ignoring spatial position information [19]. To address this, CBAM is introduced, combining a hybrid mechanism of channel attention and spatial attention. It first weights the feature map through channel attention, and then focuses on important regions through spatial attention. However, the locality of convolution limits its ability to establish long-distance dependencies.

The adaptive hybrid encoder used in the RT-DETR real-time detection model is a representative design that dynamically integrates the efficient local feature extraction capability of CNNs with the global relationship modeling advantages of attention mechanisms. Our EVA module enhances the model's global context awareness and adaptive feature weight adjustment capabilities by integrating the EVA attention mechanism into the C2f structure, thereby improving its robustness in complex scenes and small object detection.

3. Improvement of the algorithm

3.1. RT-DETR model

RT-DETR is the first truly real-time end-to-end object detection framework proposed by Baidu Research [12]. Its core innovation lies in breaking through the speed bottleneck of the traditional DETR model while maintaining high accuracy. In this paper, RT-DETR-l is selected as the benchmark model. This framework is based on an encoder-decoder architecture, as shown in **Figure 1**.

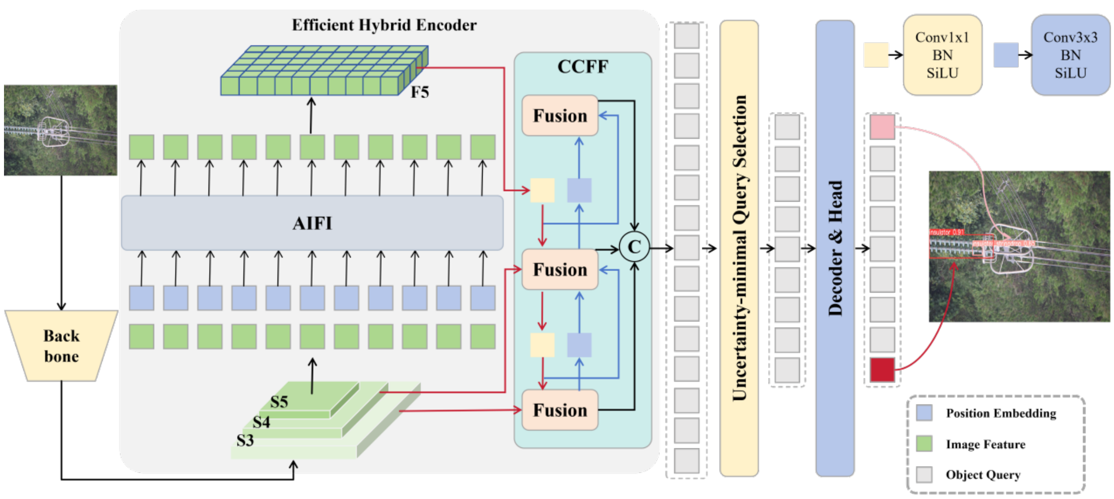


Figure 1. RT-DETR model.

RT-DETR, through a series of collaborative optimization designs, effectively balances accuracy and speed while maintaining the advantages of the end-to-end detection paradigm. Its core lies in an adaptive hybrid encoder, which innovatively integrates the local inductive bias of CNNs with the global modeling capabilities of transformers, and introduces an adaptive mechanism to dynamically allocate computational resources, thereby significantly reducing computational overhead while ensuring feature richness. The model employs a deeply optimized backbone network that extracts multi-scale feature maps through an efficient C2f module, providing feature representations for detection tasks that combine high semantic information and fine spatial details. Finally, an efficient query-based decoder utilizes a small number of learnable query vectors to directly interact with the features output by the encoder, achieving accurate object localization and classification. Its concise detection head design eliminates the need for complex post-processing, further ensuring inference efficiency.

However, there is still room for improvement in the detection performance of RT-DETR in complex scenarios, especially in small object detection, dense object detection, and complex background processing.

3.2. DETR-EVA model

To address the issues of the original RT-DETR backbone network, this paper proposes the DETR-EVA model. By deeply fusing efficient vision attention (EVA) with the C2f module, it enhances the ability to perceive global context and preserve local details, thereby improving the model's detection accuracy for small targets. The logical structure of EVA is illustrated in **Figure 2**.

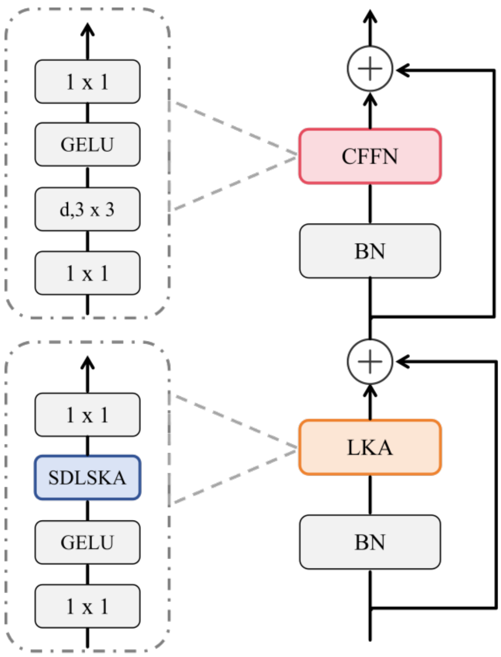


Figure 2. EVA framework.

The EVA module mainly consists of the following three parts:

(1) Sparse decomposition large kernel attention (SDLSKA): SDLSKA decomposes large convolutional kernels into local convolutions and two orthogonal band-dilated convolutions. After extracting local features using a 5×5 convolution, it captures long-range dependencies through 1×11 and 11×1 depthwise separable convolutions with a dilation rate of 3, effectively expanding the receptive field to 35×35. This design enhances the model's ability to model global semantics while reducing the number of parameters;

(2) Integrated nuclear selection mechanism (CKS): CKS further introduces a dual-path attention mechanism of channel and spatial attention to dynamically fuse multi-scale features. Channel attention generates weights through global pooling and fully connected layers, while spatial attention aggregates max and average pooling features and generates spatial weights through convolution. The two are multiplied

element-wise to achieve adaptive feature selection, thereby highlighting key regions in complex backgrounds;

(3) Convolutional feedforward network (CFFN): CFFN refines and enhances the channel dimension of the fused features through two pointwise convolutions and the GELU activation function, further improving the feature representation capability. The entire EVA module is embedded into the backbone network in a residual connection manner, which expands the receptive field and strengthens semantic understanding while maintaining the efficiency and practicality of the model.

The proposed DETR-EVA model addresses the challenges of small targets, complex backgrounds, and high real-time requirements in transmission line defect detection. By introducing a linearly complex EVA attention mechanism, it achieves efficient global context modeling under high-resolution features. The model integrates local and global attention and employs a gating mechanism to adaptively combine attention and convolutional features, significantly improving the ability to distinguish small target features.

4. Results and discussion

4.1. Dataset construction

The dataset used in this experiment is derived from images of transmission line defects captured by a drone from a certain company. The dataset comprises 7,612 images. In this experiment, LabelImg tool was employed to annotate the images as label files in XML format, which were then converted to the YOLO-specific txt format using the convert function. A total of seven different categories of abnormal defect images were annotated, namely: insulator, insulator string drop, insulator breakage, insulator flashover, damper, damper defect, and nest.

The resolution of the images is 640*640 pixels, and they are divided into training set, validation set, and test set in a ratio of 7:2:1, with 5327 images in the training set, 1523 images in the validation set, and 762 images in the test set. Some images from the dataset are shown in **Figure 3**.

Figure 3. Partial defect images.

4.2. Experimental hyperparameter settings

This experiment was developed based on the Python 3.9.24 and PyTorch 2.2.2 frameworks. The hyperparameters for the experiment are shown in **Table 1**. In the model training of this experiment, the key hyperparameters were set to prioritize the final accuracy and training stability of the model.

Table 1. Hyperparameter settings

Parameter	Value
Training epochs	300
Batch size	4
Image size	640*640
Optimizer	AdamW
Automatic Mixed Precision (AMP)	False

The parameters in this experiment were carefully designed to ensure training effectiveness, result reliability, and comparability with mainstream research paradigms. The training epochs (300 epochs) provide ample convergence space for object detection tasks, especially for Transformer-based models. A small batch size (batch size = 4) and a moderate input image size (640×640) maintain stable gradient estimation with limited hardware resources and effectively control memory usage. The optimizer AdamW was chosen, whose built-in weight decay mechanism helps alleviate overfitting and promotes model generalization. Automatic mixed precision training (AMP) was kept off to prioritize numerical stability and reproducibility during training. Overall, this parameter configuration balances algorithm performance, training efficiency, and experimental reproducibility, conforming to common settings in related research within the field.

4.3. Model evaluation metrics

In the research on defect detection in power transmission lines, to scientifically evaluate the overall performance of the improved RT-DETR algorithm, this study uses Precision, Recall, F1-Score, and mean Average Precision (mAP50, mAP50-95) as the core evaluation metrics. The specific description of the evaluation metrics is as follows.

Precision, which measures the accuracy of the model in classifying positive cases. Its mathematical expression is as follows:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

where TP (True Positives) represents the number of positive samples correctly predicted by the model, and FP (False Positives) represents the number of negative samples incorrectly predicted as positive by the model.

Recall, which measures the model's ability to identify and cover real positive samples. Its mathematical expression is as follows:

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

where FN represents real positive samples that the model incorrectly predicts as negative.

F1-Score, which is the harmonic mean of precision and recall, used to comprehensively evaluate the overall performance of a model, is defined as:

$$F1 - Score = \frac{2 \times Precition \times Recall}{Precition + Recall} \tag{3}$$

This metric combines precision and recall into a single unified measure, making it suitable for scenarios where insulator defects require an optimal balance between false positives and false negatives, providing a robust comprehensive benchmark for model performance.

Mean average precision (mAP), which evaluates the model's classification accuracy and localization ability comprehensively by calculating the average precision across all detection categories. Its calculation formula is:

$$AP = \int_0^1 P(R)dR \tag{4}$$

where P denotes Precision and R denotes Recall.

Then, the mean Average Precision (mAP) is obtained by taking the arithmetic mean of AP values across all categories:

$$mAP = \frac{1}{N} \times \sum_i AP_i \tag{5}$$

where N is the total number of categories.

In practical evaluation, mAP50 refers to the mAP value calculated with a fixed Intersection over Union (IoU) threshold of 0.5, which mainly evaluates the basic detection capability, mAP50-95 is the average of multiple mAP values calculated with IoU thresholds ranging from 0.5 to 0.95, more comprehensively reflecting the model's overall performance in both accurate recognition and precise localization.

4.4. Experimental results

The improved model was comprehensively evaluated on the dataset in this paper, and the experimental results comparing it with the baseline model RT-DETR are shown in **Table 2**.

Table 2. Experimental results

Model	P	R	F1	mAP50	mAP50-95	GFLops	Parameters	Model size
RT-DETR	0.918	0.868	0.892	0.913	0.634	57.0	19.8M	77.0MB
DETR-EVA	0.924	0.878	0.900	0.920	0.659	48.8	14.5M	56.6MB

Analysis of **Table 2** shows that the proposed DETR-EVA model significantly outperforms the benchmark RT-DETR model across all key performance indicators, achieving a synergistic optimization of accuracy and efficiency. Specifically, in terms of detection accuracy, the model's overall performance index mAP50-95 reaches 0.659, a significant improvement of 2.5% compared to the baseline. This directly verifies the effectiveness of the deep fusion of the EVA attention mechanism and the C2f module in enhancing the model's feature modeling capabilities, especially in complex scenes and small object detection. Meanwhile, the model's lightweight performance is even more remarkable, where the computational complexity (GFLOPs) is reduced to 48.8, a decrease of 14%; the number of parameters is compressed to 14.5M, a reduction of 27%. This is mainly due to the linear complexity attention design and gating fusion mechanism in the proposed method, which efficiently filters key features while introducing global context dependencies, avoiding redundant computation.

4.5. Comparative experiment

To comprehensively evaluate the overall performance of the improved model proposed in this paper, this study selected seven mainstream object detection algorithms, Faster R-CNN, Cascade R-CNN, YOLOv5n, YOLOv7-tiny, YOLOv8n, YOLOv10n, and YOLOv11n, as benchmarks for comparison and conducted systematic comparative experiments on the same transmission line defect dataset. The results are shown in **Table 3**.

Table 3. Comparative experiments

Model	P	R	mAP50	mAP50-95	F1
Faster R-CNN	0.802	0.736	0.792	0.516	0.783
Cascade R-CNN	0.826	0.749	0.813	0.523	0.796
Yolov5n	0.852	0.751	0.840	0.546	0.809
YOLOv7-tiny	0.856	0.781	0.834	0.544	0.810
Yolov8n	0.866	0.760	0.842	0.561	0.816
YOLOv10n	0.857	0.778	0.834	0.557	0.807
Yolov11n	0.863	0.770	0.848	0.565	0.812
RT-DETR	0.918	0.868	0.913	0.634	0.892
DETR-EVA	0.924	0.878	0.920	0.659	0.900

This demonstrates that the proposed improved model (EVA) exhibits comprehensive and significant advantages across all core metrics. It achieves the highest precision and recall, and its overall performance metrics, including mAP50, mAP50-95, and F1 score, significantly outperform all compared mainstream algorithms. This indicates that the model not only excels in detection accuracy but also maintains stronger robustness under a stricter intersection-union threshold (mAP50-95), achieving a better balance between precision and recall, thus validating its superior overall detection performance.

The performance of each model is visually compared and contrasted using horizontal bar charts and normalized radar charts, as shown in **Figure 4**.

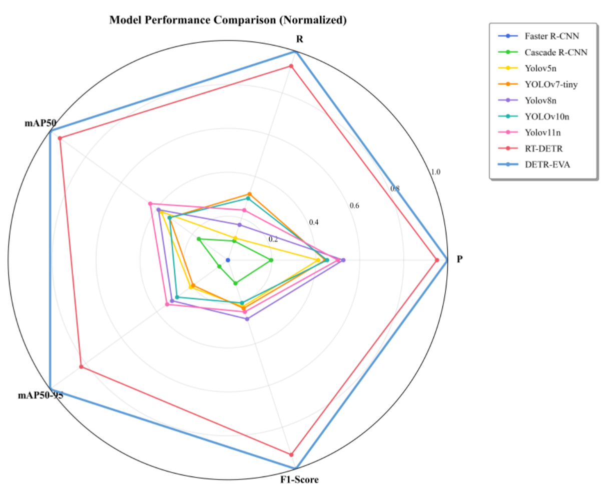
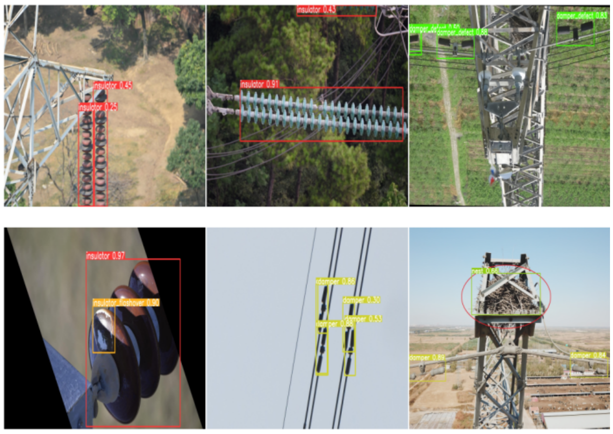


Figure 4. Normalized radar effect.

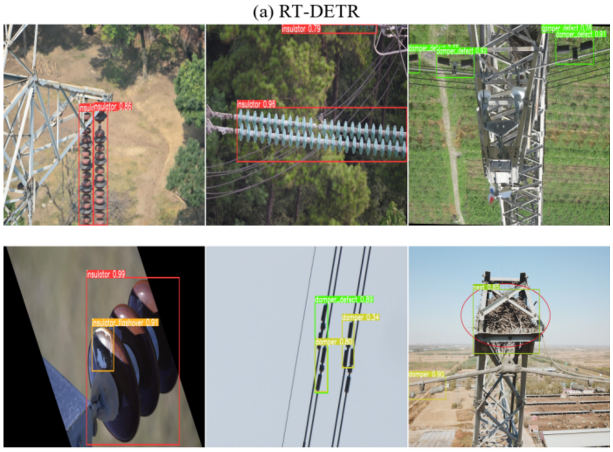
The DETR-EVA model comprehensively outperforms traditional and next-generation detectors, including Faster R-CNN, Cascade R-CNN, and RT-DETR, in radar image rendering. Its strategy of integrating EVA attention and C2f modules significantly improves the detection capability for small targets and complex backgrounds on power transmission lines by strengthening global context modeling and adaptive feature selection. While maintaining real-time inference, it achieves significant improvements in accuracy and robustness, laying a technological foundation for efficient and high-precision applications in power line inspection.

4.6. Visual analysis

To provide a more intuitive and qualitative assessment of the model's detection capabilities in complex real-world scenarios, beyond quantitative metrics, this study randomly selected six representative transmission line inspection images from the test set for inference visualization comparison. These images cover typical challenges such as small targets, multi-scale targets, cluttered backgrounds, uneven lighting, and target occlusion. **Figure 5** shows a comparison of the detection results of the unimproved RT-DETR model (**Figure 5a**) and the proposed DETR-EVA model (**Figure 5b**) on the same samples.



(a) RT-DETR



(b) DETR-EVA

Figure 5. Target detection results.

Direct observation reveals that the DETR-EVA model exhibits superior performance in detecting small-sized insulator defects, distinguishing dense targets, and controlling false alarms in complex backgrounds, intuitively verifying its stronger robustness and practicality in real-world scenarios.

5. Conclusion

This paper addresses the challenges of small targets, complex backgrounds, and high real-time requirements in power transmission line defect detection. It proposes a modified RT-DETR model, DETR-EVA, based on EVA. Through structural fusion of EVA and C2f and a gating adaptive strategy, the model achieves efficient collaboration between global context and local details, significantly enhancing its feature representation ability for small defects while maintaining linear complexity. Experiments show that this model comprehensively outperforms the original RT-DETR and mainstream lightweight models in terms of accuracy (mAP and recall), while further reducing computational overhead and parameter count, effectively balancing accuracy and speed. Future research will explore semi-supervised learning to utilize unlabeled data and improve the model's generalization ability in rare defects and unknown scenarios.

Disclosure statement

The author declares no conflict of interest.

References

[1] Tao T, Li Z, 2026, A Review of Deep Learning Application in Transmission Line Defect Detection. Electric Power Systems Research, 2026(250): 112193.

[2] Girshick R, Donahue J, Darrell T, et al., 2014, Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 580–587.

[3] Girshick R, 2015, Fast r-cnn. Proceedings of the IEEE International Conference on Computer Vision, 1440–1448.

[4] Ren S, He K, Girshick R, et al., 2016, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6): 1137–1149.

[4] Redmon J, Divvala S, Girshick R, et al., 2016, You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779–788.

[5] Redmon J, Farhadi A, 2017, YOLO9000: Better, Faster, Stronger. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 7263–7271.

[6] Redmon J, Farhadi A, 2018, Yolov3: An Incremental Improvement, arXiv preprint, https://doi.org/10.48550/arXiv.1804.02767

[7] Khanam R, Hussain M, 2024, Yolov11: An Overview of the Key Architectural Enhancements, arXiv preprint, https://doi.org/10.48550/arXiv.2410.17725

[8] Liu W, Anguelov D, Erhan D, et al., 2016, Ssd: Single Shot Multibox Detector. European Conference on Computer Vision, Springer International Publishing, 21–37.

[9] Carion N, Massa F, Synnaeve G, et al., 2020, End-to-End Object Detection with Transformers. European Conference on Computer Vision, Springer International Publishing, 213–229.

[10] Dosovitskiy A, 2020, An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv preprint, https://doi.org/10.48550/arXiv.2010.11929

[11] Lv Y, Chen W, Chen B, et al., 2023, RT-DETR: DETRs Beat YOLOs on Real-Time Object Detection. Proc. of the IEEE/CVF Int. Conf. on Computer Vision Workshops (ICCVW), 2420–2430.

[12] Li D, Yang P, Zou Y, 2024, Optimizing Insulator Defect Detection with Improved DETR Models. Mathematics, 12(10): 1507.

[13] Bai X, Ji H, Wang L, et al., 2024, Detection Method for Structural Defects of U-Shaped Hanging Ring for Power Fittings Based on Magnetic Leakage Principle. Engineering Research Express, 6(2): 025339.

[14] Huang X, Jia M, Tai X, et al., 2024, Federated Knowledge Distillation for Enhanced Insulator Defect Detection in Resource Constrained Environments. IET Computer Vision, 18(8): 1072–1086.

[15] Xie Z, Dong C, Zhang K, et al., 2024, Power-DETR: End-to-End Power Line Defect Components Detection Based on Contrastive Denoising and Hybrid Label Assignment. IET Generation, Transmission & Distribution, 18(20): 3264–3277.

[16] Chen W, Li S, Han X, 2025, IDD-DETR: Insulator Defect Detection Model and Low-Carbon Operation and Maintenance Application Based on Bidirectional Cross-Scale Fusion and Dynamic Histogram Attention. Sensors, 25(18): 5848.

[17] Vaswani A, Shazeer N, Parmar N, et al., 2017, Attention is All You Need. Advances in Neural Information Processing Systems, 2017: 30.

[18] Hu J, Shen L, Sun G, 2018, Squeeze-and-Excitation Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 7132–7141.

[19] Woo S, Park J, Lee J, et al., 2018, Cbam: Convolutional Block Attention Module. Proceedings of the European Conference on Computer Vision (ECCV), 3–19.