

RESEARCH ARTICLE

YOLO-PowerLite: A Lightweight YOLO Model for Transmission Line Abnormal Target Detection

CHUANYAO LIU¹, SHUANGFENG WEI¹, SHAOBO ZHONG², AND FAN YU¹

¹School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 102616, China

²Institute of Urban Systems Engineering, Beijing Academy of Science and Technology, Beijing 100035, China

Corresponding author: Shuangfeng Wei (weishuangfeng@bucea.edu.cn)


This work was supported in part by the Innovative Program of Beijing Academy of Science and Technology under Project 24CA004-02, in part by the National Key Research and Development Program of China under Grant 2021YFE0194700 and Grant 2021YFB2600101, and in part by the Research and Development Program of Beijing Municipal Education Commission (Research on the Quality Detection Method of Ground Cover Classification Based on Deep Confidence Neural Network) under Grant KM202010016010.

ABSTRACT The secure and stable operation of power transmission lines is essential for electrical systems. Given that abnormal targets such as bird's nests and defective insulators may lead to transmission failures, timely detection of these targets is imperative. This paper introduces the YOLO-PowerLite model, an advanced lightweight object detection model based on YOLOv8n, designed for efficient, real-time detection on resource-constrained unmanned aerial vehicles (UAVs) equipped with edge computing platforms. In the feature fusion module, YOLO-PowerLite incorporates the innovative C2f_AK module, significantly reducing the number of parameters and enhancing the adaptability and fusion capability of features at different scales. Meanwhile, the adoption of the Bidirectional Feature Pyramid Network (BiFPN) further optimizes the efficiency and effectiveness of feature processing. In addition, the newly designed lightweight detection head significantly reduces the number of parameters and computational requirements. The integration of the Coordinate Attention mechanism in the backbone network enhances the model's ability to focus on and recognize abnormal targets in complex backgrounds. Experimental results show that YOLO-PowerLite achieves a mAP@0.5 of 94.2%, maintaining the accuracy of the original YOLOv8n while significantly reducing parameters, FLOPs, and model size by 42.3%, 30.9%, and 40.4%, respectively. Comparative analysis shows that YOLO-PowerLite surpasses other mainstream lightweight models in detection accuracy and computational efficiency. Deployment on the NVIDIA Jetson Xavier NX platform demonstrates an average processing time of 31.2 milliseconds per frame, highlighting its potential for real-time applications in monitoring transmission lines.

INDEX TERMS Edge computing, lightweight structures, transmission lines, YOLO.

I. INTRODUCTION

Transmission lines, as a crucial component of the electrical power system, are entrusted with the vital task of conveying electricity generated at power stations to various consumption regions. The secure and stable operation of transmission lines is intrinsically linked to the reliability and economy of the entire power system. However, during the operation of transmission lines, certain abnormal targets may emerge, such as bird's nests on transmission towers and defective insulators.

The associate editor coordinating the review of this manuscript and approving it for publication was Diego Bellan .

These abnormal targets could result in short circuits, break-ages, or fires, thereby causing substantial damage to the power system [1]. Therefore, the timely detection and resolution of these abnormal targets are crucial for ensuring the safe operation of transmission lines.

Traditional methods for detecting abnormal targets in transmission lines primarily rely on manual inspection or ground-based sensor monitoring, yet these approaches exhibit considerable limitations [2]. Manual inspection is characterized by a large workload, high cost, and vulnerability to environmental challenges, making it difficult to achieve all-weather, all-round monitoring. Although sensor-based

monitoring offers a degree of automation, the installation and maintenance costs are substantial, and its stability and accuracy are often compromised in complex environments. With technological advancements, especially in drone and computer vision technologies, more efficient and accurate methods for detecting abnormalities in transmission lines have emerged [3]. Utilizing high-resolution cameras mounted on drones, in conjunction with sophisticated computer vision algorithms, this approach enables automated detection and rapid identification of abnormal targets along transmission lines.

Although current image analysis techniques can enhance detection accuracy, they often fall short in terms of real-time processing, struggling to meet the demands for rapid reaction in urgent situations [4]. Real-time detection of abnormal targets is critical for preventing accidents and reducing losses; it necessitates a detection system that can instantly recognize and address abnormalities, enabling timely action. Furthermore, with the widespread adoption of drones and edge computing devices in such applications, there are heightened demands for computational efficiency and reduced resource consumption in detection models. Model lightweighting becomes essential for achieving efficient operation in resource-constrained environments, particularly for applications that are deployed on mobile devices such as drones.

In light of this, this paper proposes the “YOLO-PowerLite” model, which is designed to address the dual challenges of real-time processing and lightweight design for the detection of abnormal targets on transmission lines. By strategically optimizing the latest YOLOv8 model, including improvements to the feature fusion module, the lightweight design of the detection head, and the integration of a lightweight attention mechanism, we aim to reduce the model’s computational complexity and storage demands while ensuring its accuracy and real-time detection capabilities. The development of this lightweight model not only enables efficient real-time detection of abnormalities on transmission lines by drones and other mobile devices under resource-limited conditions but also offers robust technical support for the safe operation of power systems, significantly enhancing the reliability and economic efficiency of the electrical infrastructure.

II. RELATED WORK

As the field of computer vision evolves, object detection has emerged as a pivotal area of research, aiming to identify and locate specific objects within images [5]. Early methods for detecting abnormalities in power transmission lines primarily relied on manually designed features and traditional machine learning algorithms. For instance, SIFT (Scale Invariant Feature Transform) [6] and HOG (Histogram of Oriented Gradient) [7], two classic feature descriptors, have been extensively utilized for object recognition and localization. They are first utilized to describe the object by extracting key points and edge information from the image, and then

machine learning algorithms such as SVM (Support Vector Machine) [8] are utilized for classification. Although these traditional methods perform well in simpler scenarios, they often falter in complex environments and diverse detection tasks. This is because manually designed features struggle to encompass all variations in objects, and these methods lack profound understanding of the deeper features of images [9].

With the rise of deep learning technologies, methods of object detection based on deep neural networks have started to garner widespread attention among researchers. Compared to traditional approaches, deep learning can autonomously learn complex feature representations from vast datasets, thereby significantly enhancing the accuracy and robustness of object detection [10]. Within deep learning techniques, object detection algorithms are generally categorized into two primary types: two-stage and one-stage algorithms. Two-stage algorithms, such as R-CNN [11] and its variants Fast R-CNN [12] and Faster R-CNN [13], initially generate a series of candidate regions and subsequently perform feature extraction and classification for each region. While these methods excel in precision, their significant computational cost limits their application in real-time scenarios.

In contrast to two-stage algorithms, one-stage algorithms such as YOLO (You Only Look Once) and SSD (Single Shot MultiBox Detector) [14] directly predict the class and location of objects in the image, eliminating the need for generating candidate regions, thus significantly accelerating the detection speed. As the quintessential representative of one-stage algorithms, the YOLO algorithm has undergone numerous iterations and enhancements since its inception, each version striving to achieve a superior equilibrium among detection speed, accuracy, and model complexity. YOLO partitions the input image into a grid of cells, with each cell responsible for predicting objects whose center points fall within it [15]. For each grid cell, YOLO predicts multiple bounding boxes along with their associated confidence scores and class probabilities. The confidence score reflects the likelihood that a bounding box contains an object and the accuracy of the prediction, while the class probabilities indicate the likelihood of the object belonging to various categories. Through this mechanism, YOLO provides a global view while achieving rapid and precise object detection.

Since the initial release of YOLO, subsequent versions including YOLOv2 (also known as YOLO9000) [16], YOLOv3 [17], YOLOv4 [18], YOLOv5, YOLOv6 [19], YOLOv7 [20], and the recent YOLOv8, have significantly improved and optimized the original framework. These iterations not only include adjustments and enhancements to the network architecture but also introduce new mechanisms, such as multi-scale prediction, enhanced feature extraction networks, and more precise methods for bounding box prediction, all aimed at enhancing the model’s performance and applicability.

However, regardless of being two-stage or one-stage deep learning models, they typically require substantial computational resources. This requirement renders them a bottleneck

for applications in resource-constrained environments such as drones equipped with edge computing platforms [21]. In the field of abnormality detection for power transmission lines, the development of lightweight models is crucial. Traditional detection methods, such as manual inspection and ground-based sensor monitoring, suffer from inefficiencies and high costs. Lightweight deep learning models can be deployed on mobile devices such as drones for efficient and cost-effective automated inspections.

Li et al. [22] achieved a substantial reduction in network parameters by using the more lightweight Mobilenetv2 network architecture in place of the original Darknet-53 structure used in YOLOv3, and by substituting standard 3×3 convolutions with depthwise separable convolutions in the detection head. Huang et al. [23] proposed an improved lightweight version of a power transmission line insulator defect detection algorithm based on YOLOv5. The primary strategy employed involves removing redundant convolution layers and reducing the number of channels; additionally, adaptive attention modules are integrated between adjacent residual blocks to assign greater weight to key features, thereby enhancing the model's learning capabilities. Chen et al. [24], in an effort to alleviate the computational demands of the network, integrated the lightweight GSCConv module into both the backbone and neck networks of YOLOv8. Additionally, they introduced the Content-Aware ReAssembly of Features (CARAFE) architecture into the neck structure, aimed at enhancing the efficiency of information utilization during the feature upsampling process and improving the model's capability for feature integration. Meanwhile, Zhang et al. [25] increased the accuracy of abnormality detection in power transmission lines by incorporating the Multi-Scale Large Kernel (MSLK) attention mechanism and utilizing the enhanced SIOU loss function. In addition, their GSC_C2f module, designed based on lightweight GSCConv module, successfully simplified the computational process of the model and significantly reduced the memory usage.

Despite progress in the field of abnormality detection for power transmission lines, current models still face challenges in balancing detection accuracy with model lightweighting. Since transmission line images often contain complex backgrounds such as trees, grass, and other power facilities, these factors increase the difficulty of accurate detection based on aerial images. Moreover, efficiently implementing abnormal target detection becomes challenging under the constraints of limited resources on edge computing device platforms. Existing research lacks specificity and is difficult to apply directly to the detection of abnormalities in power transmission lines. Consequently, further research is necessary to develop methodologies that enable effective identification and localization of potential hazards in these computationally constrained environments.

III. METHODS

In order to ensure the accuracy of abnormal target detection in transmission lines and simultaneously achieve a lightweight

model to reduce deployment costs, this study implemented a series of targeted improvements based on the YOLOv8n model. We refer to the final improved network model as YOLO-PowerLite, and the network structure of the improved model is shown in Fig. 1.

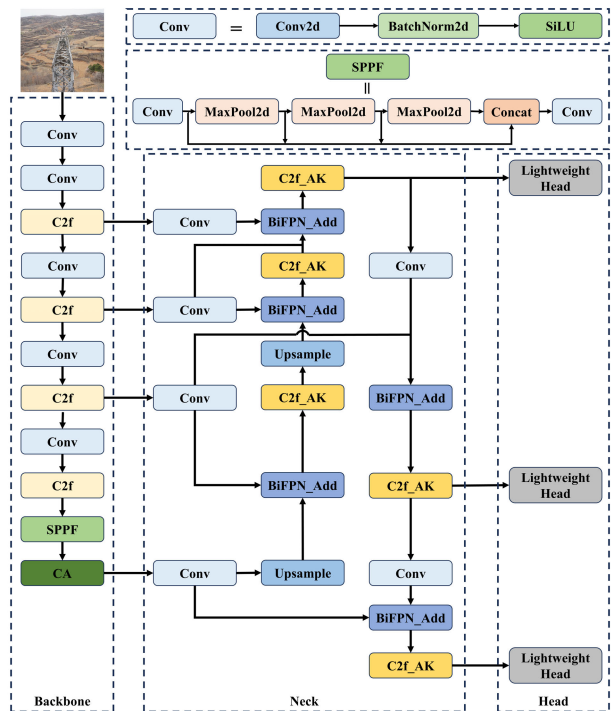


FIGURE 1. YOLO-PowerLite network structure.

First, in the feature fusion module, we implemented significant enhancements to the C2f module in YOLOv8. By replacing the standard Bottleneck structure within the C2f module with the AK_Bottleneck based on Alterable Kernel Convolution (AKConv), we created the new C2f_AK module. This enhancement significantly reduced the number of parameters and computational complexity, not only providing the model with advantages in terms of lightness and efficiency but also facilitating the deployment of the model on unmanned aerial vehicle (UAV) edge computing devices for real-time detection of abnormal targets on transmission lines. In addition, we reconstructed the feature fusion part based on the Bidirectional Feature Pyramid Network (BiFPN) to improve the efficiency and effectiveness of feature processing for transmission line abnormal targets.

Second, for the design of the detection head, we optimized the decoupled head structure used in the original YOLOv8 model. Although the decoupled head structure enhances model performance by providing dedicated network pathways for category prediction and bounding box regression, the resulting increase in parameter count and computational burden may become a bottleneck in resource-constrained environments. Therefore, we proposed to further reduce the total parameter count and computational demands of the model by sharing parameters at the forefront of category

prediction and bounding box regression tasks, while maintaining efficient feature representation by preserving task-specific terminal layers.

Finally, considering that transmission line images captured by UAVs often feature complex backgrounds, such as trees, grass, and other power facilities. All of these complex backgrounds can greatly affect the accurate detection of abnormal targets on transmission lines, so we introduced the Coordinate Attention (CA) mechanism into the backbone network of the model. The CA mechanism effectively improved the model's attention to key features and reduced sensitivity to background interference by emphasizing the channel and spatial information of the image.

A. IMPROVEMENTS IN FEATURE FUSION MODULE

1) C2F_AK MODULE

In YOLOv8, the C2f module is responsible for fusing high-level semantic features with low-level detailed features to enhance detection accuracy. To render YOLOv8 more lightweight, the original Bottleneck structure within the C2f module was replaced with the AK_Bottleneck, thereby creating the new C2f_AK module. The structures of C2f_AK and AK_Bottleneck are shown in Fig. 2a and 2b.

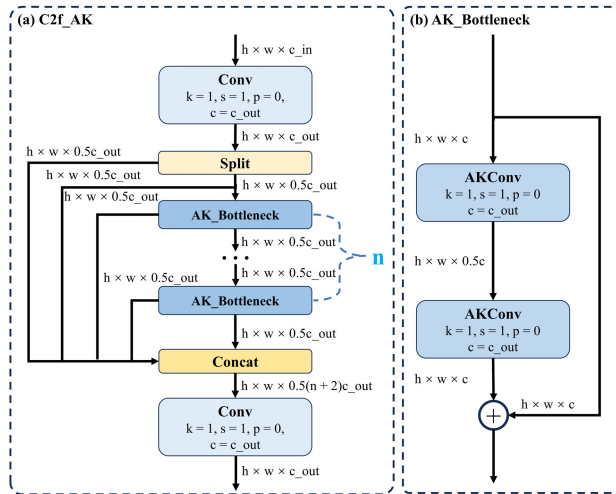


FIGURE 2. Structure diagram of C2f_AK in YOLO-PowerLite. (a) C2f_AK module. (b) AK_Bottleneck module.

The AK_Bottleneck structure employs an innovative lightweight network technology known as the Alterable Kernel Convolution (AKConv) module [26]. This module aims to enhance network performance and efficiency through dynamic adjustments in the shape and size of the convolutional kernels. The core principle of the AKConv module lies in its unique ability to allow the convolutional kernels to dynamically change their shape and parameter configurations based on the demands of the input features. This flexibility enables AKConv to not only adapt to different data characteristics but also to optimize the utilization of computational resources and achieve a linear reduction in the number of convolutional parameters, tailored to the specific

requirements of the task. The AKConv structure is shown in Fig. 3. AKConv utilizes a new coordinate generation algorithm to define initial positions for convolution kernels of arbitrary size. Moreover, to accommodate target changes, AKConv introduces offsets that adjust the shape of samples at each position. The introduction of this approach enables the C2f_AK module to significantly enhance its performance in lightweighting while maintaining the strong detection capability of the YOLOv8 model.

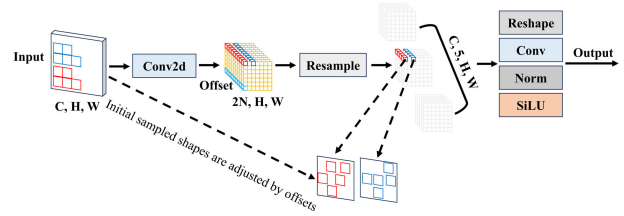


FIGURE 3. The structure of AKConv.

To specifically enhance the detection of abnormal targets in power transmission lines, the AKConv technology has been further refined to focus on the unique characteristics of these targets. Transmission lines often involve complex backgrounds and small, irregularly shaped anomalies. The AKConv module dynamically adapts to the input features' requirements, adjusting the shape and size of the convolution kernels, making it particularly suited for this application. Specifically, the offsets introduced by AKConv effectively modify the shape of the samples at each position, allowing for more precise capture of anomaly details. This enables the C2f_AK module to effectively distinguish between normal transmission line components and various types of anomalies, such as bird nests and insulator damage.

2) BIFPN

In the feature fusion module of YOLOv8, the Path Aggregation Network (PANet) [27] is utilized, exhibiting notable advantages over the traditional Feature Pyramid Network (FPN) [28]. As shown in Fig. 4b, PANet effectively enhances the fusion process between features at different scales through its innovative bidirectional information flow mechanism. This mechanism not only facilitates the transfer of low-level detail information to higher levels but also ensures that the semantic information at higher levels augments the representation of low-level features. This leads to superior performance of the model in processing images with rich details and complex backgrounds. Although PANet has made significant progress in multi-scale feature fusion, there is still room for improvement [29]. Initially, PANet does not fully manifest its potential when processing large-scale feature maps, possibly overlooking some crucial information, thereby impeding the overall performance of object detection. Additionally, the feature map loses some of its original information during the process of upsampling and downsampling,

and this loss of information reduces the reuse efficiency of the feature map.

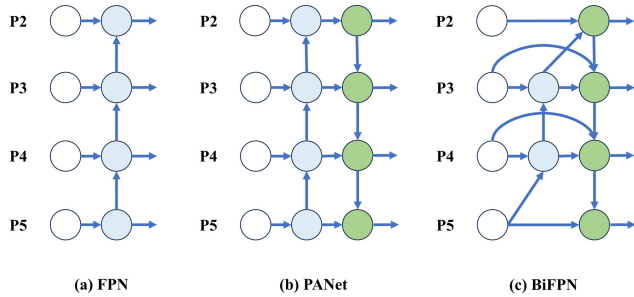


FIGURE 4. Feature network design: (a) FPN. (b) PANet. (c) BiFPN.

To effectively address the issues outlined, this study restructured the feature fusion module of YOLOv8 based on the Bidirectional Feature Pyramid Network (BiFPN) [30]. In comparison to PANet, BiFPN incorporates key optimizations and improvements. First, it eliminates nodes that have only a single input edge and do not participate in feature fusion. This approach is based on a simple idea: if a node is unable to contribute to the fusion of different features, it is relatively less useful to the structure of a network aimed at feature fusion. This adjustment streamlines the network structure, yielding a more efficient bidirectional network. Next, BiFPN adds additional connections between the original input nodes and output nodes in the same hierarchy, which enables efficient fusion of more features without significantly increasing the cost. Finally, unlike PANet, which only features a top-down and a bottom-up pathway, BiFPN treats each bidirectional pathway (both top-down and bottom-up) as a separate feature network layer and promotes more advanced feature fusion by repeating these layers multiple times.

B. LIGHTWEIGHT HEAD

The original YOLOv8 detection head employs a decoupled head structure, as shown in Fig. 5a. The decoupled head structure considers category prediction and bounding box regression as two separate tasks, each computed through its own independent convolutional block. The primary advantage of this design is that it provides dedicated network pathways for each task, enabling the model to learn more specialized feature representations for category classification and spatial localization, potentially enhancing the overall performance of the model.

However, this separation in the decoupled head structure also introduces significant drawbacks. First, the independent network pathways for each task increase the model’s total parameter count, subsequently raising computational demands and storage requirements. This issue becomes particularly problematic in environments constrained by computational resources, potentially limiting the model’s deployment and applicability. Second, while task specificity aids in enhancing performance, it also reduces the model’s

optimization flexibility, as the layers for category prediction and bounding box regression cannot share information.

In light of these considerations, our proposed improvement sought to address these challenges by introducing a shared parameter structure, while retaining key advantages of the decoupled head structure. The improved detection head is shown in Fig. 5b. By sharing parameters in the front part of the category prediction and bounding box regression branches, our design not only reduced the model’s overall parameter count but also its computational burden. Moreover, by preserving task-specific end layers, we ensured that the model is able to learn effective feature representations for each task. This approach improved the model’s computational efficiency and continued to provide sufficient task specificity to achieve high performance.

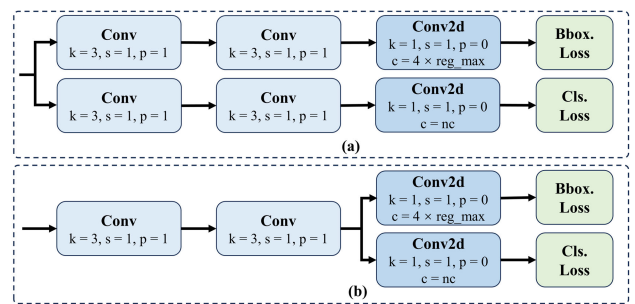


FIGURE 5. Head design: (a) Original YOLOv8 head. (b) Improved lightweight head.

C. COORDINATE ATTENTION MECHANISM

In the processing of aerial imagery of transmission lines, the complexity of the background poses a significant challenge. The presence of trees, grass, and other power facilities often interferes with model detection, resulting in decreased performance in tasks aimed at suppressing irrelevant background elements. To improve the model’s focus on key features and reduce its reliance on background information, we introduced the Coordinate Attention (CA) mechanism [31] into the model’s backbone network. The CA mechanism module is a lightweight attention mechanism that concentrates on the channel and spatial information of images, thus enhancing the model’s performance in handling complex and blurred backgrounds. The structure of the CA mechanism module is shown in Fig. 6.

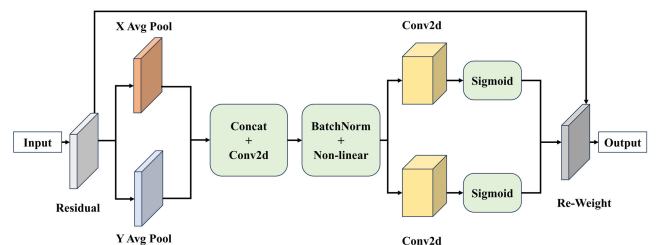


FIGURE 6. Structure of the CA mechanism module.

The implementation of the CA mechanism encompasses two principal stages: coordinate information embedding

and coordinate attention generation. During the coordinate information embedding phase, although global pooling is commonly utilized in channel attention to encode spatial information globally, it compresses this information into a single channel descriptor, thus making it challenging to preserve the integrity of positional information. However, preserving positional information is crucial for capturing spatial structure in visual tasks. For a given input X , the CA mechanism utilizes two scales of pooling kernels, $(H, 1)$ and $(1, W)$, to encode each channel along the horizontal and vertical directions, respectively, thereby enhancing the capture of spatial information. Moving into the coordinate attention generation phase, the feature maps obtained from the coordinate information embedding phase are first merged, followed by a convolutional transformation to reduce their dimensionality, thereby simplifying the model's complexity. Subsequently, after batch normalization and nonlinear activation, an intermediate feature map rich in information is produced. This feature map is then split along the spatial dimension into two separate tensors. Finally, each tensor is processed through a 1×1 convolution and a Sigmoid activation function, transforming them into tensors with the same number of channels as input X , thereby completing the construction of the entire CA mechanism.

In this study, we integrated the CA mechanism within the backbone network. One significant advantage of this approach is that the CA mechanism, being a lightweight attention mechanism, can significantly enhance the model's ability to focus on critical information about abnormal targets while considering the limited computational and storage resources of UAV hardware platforms. To maximize the utilization of this attention mechanism, we opted to employ the CA attention mechanism at the end of the model's backbone network, aiming to achieve superior performance.

IV. EXPERIMENTS

A. EXPERIMENTAL INTRODUCTION

This chapter first introduces the datasets used, subsequently describes the experimental environment and training strategies, and finally presents the evaluation metrics tied to the experimental results.

1) DATASET

The dataset used in this study is derived from three distinct sources: high-voltage transmission line images obtained through UAV field inspections; the publicly available transmission line bird's nest dataset [32]; and the publicly available Chinese Power Line Insulator Dataset (CPLID) [33]. Given the lack of a publicly available dataset that addresses multiple types of abnormalities on power transmission lines in the electric power sector, we combined existing data and supplemented it with our self-collected images. Thus, we constructed a comprehensive dataset of 1738 images aimed at covering two common types of abnormal targets on transmission lines: bird nests and insulator defects.

To systematically evaluate the performance of the YOLO-PowerLite model, we divided the entire dataset into training, validation, and test sets at a ratio of 7:2:1. This division ensured that the model is trained on a sufficiently diverse set of data, while providing separate validation and test sets for evaluating the model's generalization ability and performance in real-world applications.

Given the common issue of overfitting encountered during the training of deep learning models, particularly with limited training samples, we implemented a series of data augmentation techniques to enhance the model's robustness and generalization ability. Data augmentation operations include horizontal flipping, vertical flipping, random rotation, brightness adjustment, gaussian noise, and blurring. These techniques expand the diversity of the training samples and simulate various complex environmental conditions likely encountered in real-world applications, thereby ensuring the model's effectiveness and robustness. The effects of data augmentation are illustrated in Fig. 7, where images (a) through (h) represent magnifications of original images using data processing methods.

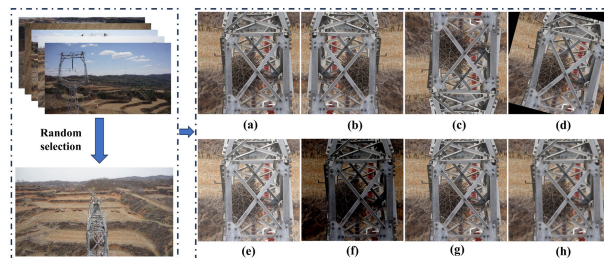


FIGURE 7. Data augmentation process and results. (a) Original image. (b) Horizontal flip image. (c) Vertical flip image. (d) Random rotation image. (e) High brightness image. (f) Low brightness image. (g) Gaussian noise image. (h) Blur image.

Additionally, the detailed information of the dataset is illustrated in Fig. 8. Fig. 8a displays the number of defect labels for bird nests and insulators within the dataset. Fig. 8b shows the size of the target bounding boxes in the dataset. Fig. 8c reveals the distribution of the center point coordinates of the labeled boxes, indicating that the center points are predominantly concentrated in the middle region of the images. Fig. 8d presents a scatter plot of the width-to-height ratio of the labeled boxes, with the deepest color in the lower-left corner, suggesting that the current dataset mainly contains small to medium-sized targets.

2) EXPERIMENTAL ENVIRONMENT AND TRAINING STRATEGIES

In this experiment, the configuration of our experimental setup included an Intel(R) Xeon(R) Platinum 8375C processor alongside an NVIDIA GeForce RTX 4090 graphics card. The software stack comprised PyTorch 2.1.1 and Python 3.10 for the deep learning framework, CUDA version 11.8, and Ubuntu 22.04 as the operating system. To ensure fairness and comparability in model performance, pretrained weights

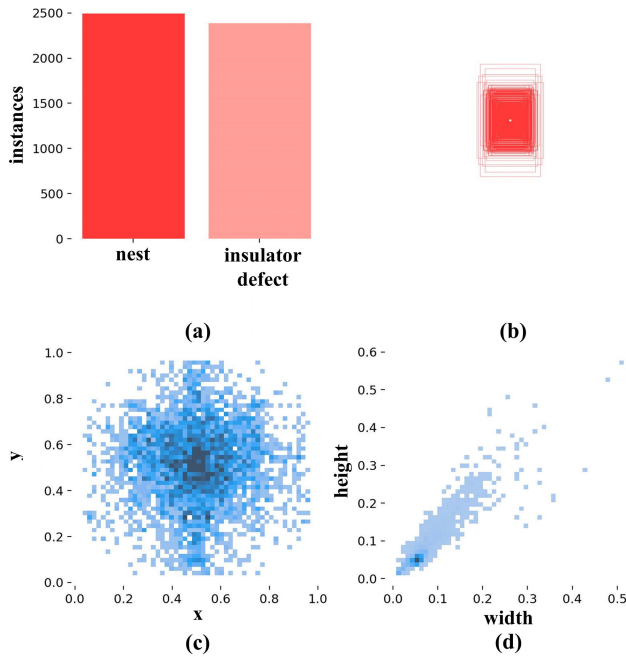


FIGURE 8. Dataset information.

were not utilized during the training of ablation and comparative experiments. The important hyperparameter settings for the models during the training phase are presented in Table 1.

TABLE 1. Model hyperparameter settings.

Parameters	Setup
Epoch	200
Batch size	32
Image Size	640×640
Initial Learning Rate	1×10 ⁻²
Final Learning Rate	1×10 ⁻⁴
Momentum	0.937
Weight-Decay	5×10 ⁻⁴
Optimizer	SGD

To accommodate various application environments and hardware requirements, the YOLOv8 model is engineered by adjusting two critical dimensions—depth and width. This process results in five scaled model variants, designated as YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. Each version exhibits a progressive increase in parameter count and resource consumption, meeting diverse performance needs for detection. Detailed specifications of depths, widths, and maximum channel counts for these models are presented in Table 2.

TABLE 2. Parameters corresponding to different sizes of YOLOv8.

Model	Depth	Width	Max Channels
YOLOv8n	0.33	0.25	1024
YOLOv8m	0.33	0.50	1024
YOLOv8s	0.67	0.75	768
YOLOv8l	1.00	1.00	512
YOLOv8x	1.00	1.25	512

Given that edge computing devices typically face limitations in computational resources and energy consumption, we selected the lightest YOLOv8n as the baseline model for lightening and preparing for deployment.

3) EVALUATION METRICS

To comprehensively evaluate the performance of abnormality detection models for transmission lines, this study focused not only on the accuracy of the models but also considered their lightweight requirements for deployment on edge computing devices. Therefore, precision, recall, average precision (AP), mean average precision (mAP), floating-point operations (FLOPs), Params, and model size were selected as key evaluation metrics.

Precision is the ratio of all predicted targets correctly identified by the model, and the calculation formula is shown in (1). Where true positive (TP) is the number of transmission line abnormal targets correctly identified by the model, while false positive (FP) and false negative (FN) represent the number of abnormal targets that actually exist but are incorrectly identified and missed by the model, respectively.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

Recall is the ratio of all actual targets correctly recognized by the model and is calculated as shown in (2).

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

AP is equal to the area under the precision-recall curve, and the closer its value is to 1 means the better the model performance, and the formula is shown in (3).

$$AP = \int_0^1 Precision(Recall) dRecall \quad (3)$$

mAP is the average of AP of all sample categories, which is the most commonly used evaluation metric in object detection and intuitively reflects the performance of the current model, and the calculation formula is shown in (4).

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (4)$$

FLOPs can reflect the computational complexity of the model, which is calculated as shown in (5).

$$FLOPs = 2 \times H \times W (C_{in}K^2 + 1) C_{out} \quad (5)$$

Model lightweighting is evaluated by the number of parameters (Params), which is calculated as shown in (6).

$$Params = C_{in} \times K^2 \times C_{out} \quad (6)$$

In addition, the model size can be used to measure the ease of deployment of the model and its applicability in lightweight devices [34].

B. EXPERIMENTAL RESULTS

1) ABLATION EXPERIMENTS

In order to validate the effectiveness of each improvement strategy proposed in this paper, we conducted ablation experiments on the baseline model, and the experimental results are displayed in Table 3.

TABLE 3. Detection results after the introduction of different improvement strategies.

Model	mAP@0.5 (%)	Params (M)	FLOPs (G)	Model Size (MB)	FPS
YOLOv8n	94.2	3.01	8.1	5.96	149.3
C2f-AK	93.0	2.73	7.6	5.44	169.5
C2f-AK+BiFPN	94.0	1.95	6.8	3.97	107.5
C2f-AK+BiFPN+Head	93.7	1.73	5.6	3.53	111.1
C2f-AK+BiFPN+Head+CA	94.2	1.73	5.6	3.55	101.0

From Table 3, it is evident that by replacing the C2f module in YOLOv8n with the C2f_AK module, we observed a reduction in the number of parameters (Params), floating-point operations (FLOPs), and model size by 9.1%, 6.2%, and 8.7% respectively. This enhancement significantly diminishes the model's size and computational cost, thereby improving the efficiency of feature extraction. However, the mAP@0.5 slightly decreased to 93.0%. Specifically, the C2f_AK module employs the AK_Bottleneck structure, dynamically adjusting the shape and size of convolution kernels to optimize the network's performance and efficiency, achieving a lightweight model while maintaining a certain level of accuracy.

Subsequently, we restructured the feature fusion component of YOLOv8 based on the bidirectional feature pyramid network (BiFPN). The results demonstrate further reductions in Params, FLOPs, and model size by 28.7%, 10.5%, and 27% respectively, with an increase in mAP@0.5 to 94.0%. BiFPN optimizes the feature fusion structure, enhancing the model's capability to process features of varying scales.

Through the introduction of a shared parameter structure in the original YOLOv8n detection head, while retaining some key advantages of the decoupled head structure, we effectively minimized redundant computations and memory access. Although the mAP@0.5 only decreased by 0.3%, Params, FLOPs, and model size were significantly reduced by 11.4%, 17.6%, and 11.1% respectively. This validates the efficacy of our lightweight detection head improvements.

Finally, we incorporated the Coordinate Attention (CA) module at the terminal part of the backbone network, enhancing the model's ability to locate and semantically understand anomalies in transmission lines, mitigating the impact of complex backgrounds on feature extraction. This improvement elevated the model's mAP to 94.2%, equivalent to the original YOLOv8n.

Through these enhancements, YOLO-PowerLite achieved reductions in Params, FLOPs, and model size by 42.3%, 30.9%, and 40.4% respectively compared to the baseline

model YOLOv8n. This effectively lowers the model's complexity and optimizes its efficiency, making it more suitable for low-performance devices and easier to integrate. From Table 3, we can observe that the decrease in FPS is primarily due to the reconstruction of YOLOv8's feature fusion component based on the BiFPN and the introduction of the CA attention mechanism. The inclusion of BiFPN adds feature processing steps, while the CA attention mechanism increases computational complexity, both contributing to the FPS decrease. However, despite the FPS dropping to 101, slightly lower than the baseline model's 149.3, it still meets the real-time detection requirements for anomalies in transmission lines [35]. These ablation experiment results clearly demonstrate the advantages and efficacy of our model in terms of accuracy and achieving a lightweight design, validating the feasibility of our approach in the domain of transmission line anomaly detection.

2) COMPARISON EXPERIMENTS

To further validate the performance of YOLO-PowerLite, we compared its performance with that of several mainstream lightweight target detection models. We focused on evaluating key metrics such as mAP@0.5, Params, FLOPs, and Model Size for each model. The experimental results are presented in Table 4.

TABLE 4. Results of each indicator for different models.

Model	mAP@0.5 (%)	Params (M)	FLOPs (G)	Model Size (MB)
Faster R-CNN	93.7	136.69	79.20	108.0
SSD	88.3	23.61	22.18	100.0
YOLOv3-tiny	92.5	8.67	12.90	16.6
YOLOv5s	93.7	7.02	15.80	13.7
YOLOv6n	93.4	4.63	11.34	9.97
YOLOv7-tiny	90.0	6.01	13.0	12.3
YOLOv8n	94.2	3.01	8.10	5.96
YOLO-PowerLite	94.2	1.73	5.60	3.55

From Table 4, we can see that the mAP@0.5 of YOLO-PowerLite and YOLOv8n were both 94.2%. YOLO-PowerLite demonstrated exceptional detection performance, significantly surpassing other lightweight models. This result confirmed that YOLO-PowerLite effectively captures abnormal targets in transmission lines while remaining lightweight.

In terms of model efficiency, YOLO-PowerLite excelled significantly. With merely 1.73M parameters, 5.6G FLOPs, and a model size of only 3.55MB, these metrics were significantly lower than those of other models in the comparison group. Compared to YOLOv7-tiny, YOLO-PowerLite had 71.2%, 56.9%, and 71.1% reductions in Params, FLOPs, and model size, respectively, and compared to YOLOv5s, the reductions were 75.4%, 64.6%, and 74.1%, respectively. This significant advantage indicated that YOLO-PowerLite is highly valuable in resource-limited deployment environments and is particularly suited for real-world application scenarios with stringent requirements on detection speed and model size.

3) VISUALIZATION ANALYSIS

Grad-CAM [36] is a widely utilized visualization method for highlighting the regions of an image that most significantly contribute to the model’s predicted results, thereby enhancing the transparency of the model’s decision-making process. We chose images of power transmission lines with complex backgrounds as our test cases. In these cases, the image backgrounds are complex and variable, containing trees, grass, and other power facilities, all of which could potentially impact the performance of the target detection model. By applying Grad-CAM to the output of the YOLOv8n and YOLO-PowerLite models, we produced heatmaps reflecting the focus of the model’s attention, as shown in Fig. 9.

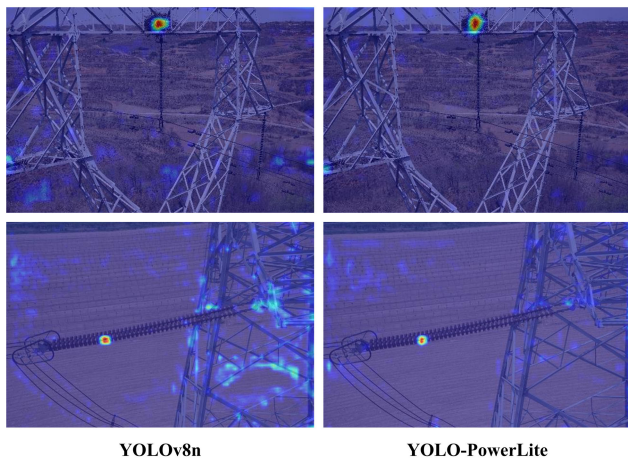


FIGURE 9. Grad-CAM visualization results.

The visualization results indicated that although YOLOv8n generally performs well in detection tasks, its focus is occasionally diverted to non-target regions due to significant background interference. In contrast, YOLO-PowerLite exhibits greater robustness in target detection and maintains a more intense focus on actual transmission line abnormalities, even amidst complex backgrounds.

Furthermore, we performed a visual comparison and analysis of the YOLO-PowerLite model’s performance against three other lightweight models—YOLOv8n, YOLOv5s, and YOLOv3-tiny—in detecting transmission line abnormalities. The detection results are shown in Fig. 10.

The experimental results showed that when confronted with drone imagery characterized by rich and complex backgrounds, both the YOLOv5s and YOLOv3-Tiny models exhibit significant issues of false detections and missed detections in identifying abnormalities in power transmission lines. Although these two models are lightweight and fast, their accuracy and robustness are insufficient when dealing with subtle targets in complex scenarios, leading to elevated rates of false positives and omissions.

In comparison, the detection performance of YOLO-PowerLite is significantly enhanced. Under equally complex background conditions, YOLO-PowerLite not only reduces the instances of false positives and omissions but also slightly



FIGURE 10. The detection results of four lightweight algorithms in abnormal target detection of transmission lines.

surpasses YOLOv8n in detection efficacy. This advantage is attributed to enhancements in model design, particularly through the optimization of feature extraction and fusion mechanisms. These optimizations enable the model to more effectively identify and focus on abnormal targets, even amid significant background interference.

4) MODEL DEPLOYMENT ON NVIDIA JETSON XAVIER NX

To assess the feasibility and efficiency of the YOLO-PowerLite model in practical scenarios, particularly on resource-constrained edge computing devices, we conducted tests and analyses using the NVIDIA JETSON Xavier NX as our chosen deployment platform. The JETSON Xavier NX, a module specifically engineered for edge AI computations, is ideally suited for performing complex computational tasks on mobile devices such as drones, owing to its compact design and low energy consumption. It is equipped with a six-core NVIDIA Carmel ARMv8.2 CPU and a 384-core NVIDIA Volta GPU, the latter of which includes 48 Tensor Cores specifically designed to accelerate deep learning tasks. These specifications enable the Jetson Xavier NX to deliver up to 21 TOPS of AI processing power, while maintaining a remarkably low energy consumption of just

10 to 15 watts, thus demonstrating an exceptional energy efficiency ratio.

To ensure the model runs efficiently on the Jetson Xavier NX platform, we first converted the YOLO-PowerLite model from its original framework format (such as PyTorch) to the ONNX (Open Neural Network Exchange) format. Subsequently, we utilized NVIDIA TensorRT tools to optimize the ONNX model and convert it into the TensorRT engine format (.engine), thereby fully leveraging the hardware acceleration capabilities of the Jetson Xavier NX. In this study, we used the test set from the dataset as input for model inference to simulate real-world scenarios where drones equipped with edge computing devices detect abnormal targets on transmission lines. The deployment of the YOLO-PowerLite model on the Jetson Xavier NX platform exhibited an average processing time of only 31.2ms per frame, highlighting its significant advantage in processing speed and demonstrating its immense potential for real-time applications.

Additionally, we provided detailed configuration information for the Jetson Xavier NX, including the Jetpack version and TensorRT version, as shown in Table 5.

TABLE 5. Detailed configuration information for NVIDIA jetson Xavier NX.

Configuration Item	Details
Jetpack Version	4.5.1
CUDA Version	10.2
TensorRT Version	8.2
OpenCV Version	4.1
Operating System	Ubuntu 18.04

V. CONCLUSION AND FUTURE WORK

To address the challenge of conducting efficient, real-time abnormality detection of power transmission lines on resource-constrained UAV-mounted edge computing platforms, we introduce the YOLO-PowerLite model. This model, by optimizing and enhancing the YOLOv8n framework, achieves significant success in model lightweighting while also ensuring high accuracy and real-time object detection. In our model improvement efforts, we implement several innovative optimizations to the YOLOv8 model, including the introduction of the C2f_AK module and BiFPN-based feature fusion strategy within the feature fusion module, as well as the design of the lightweight detection head and the incorporation of the CA mechanism.

Given the absence of publicly available datasets for multiple categories of transmission line abnormalities, we synthesized and augmented the existing datasets to construct a more comprehensive and challenging consolidated dataset. Ablation experiments conducted on this dataset show that the YOLO-PowerLite model achieved a mAP@0.5 of 94.2%, comparable to that of YOLOv8n. In terms of model efficiency, YOLO-PowerLite demonstrated significant

reductions in parameters, floating-point operations (FLOPs), and model size—42.3%, 30.9%, and 40.4%, respectively—with parameters reduced to 1.73 M, FLOPs to 5.6 G, and model size to 3.55 MB. Moreover, when compared with other mainstream lightweight object detection models, YOLO-PowerLite not only leads in detection accuracy but also exhibits significant advancements in model lightness. Notably, the deployment of YOLO-PowerLite on the NVIDIA Jetson Xavier NX edge computing platform demonstrated its exceptional real-world application performance, with an average inference time of just 31.2 milliseconds per frame. This remarkable processing speed underscores YOLO-PowerLite's significant advantage in real-time operations and highlights its immense potential and value in real-time abnormality detection applications for power transmission lines.

However, our current work continues to face several limitations that require attention. Firstly, due to the lack of publicly available datasets for multiple categories of transmission line abnormalities in the electrical field, we have integrated and supplemented existing datasets; however, challenges related to insufficient data samples and diversity persist. These issues could impact the model's generalizability and adaptability to various abnormalities. Therefore, it is imperative that we expand the dataset, collect a broader array of abnormal object images, and validate the model's generalizability and practicality across a wider range of application scenarios.

Additionally, the successful deployment of YOLO-PowerLite on the NVIDIA Jetson Xavier NX edge computing platform showcases the model's excellent real-time processing performance. We plan to deploy the model on an authentic UAV platform and conduct extensive field testing to validate its effectiveness and stability in practical applications.

Ultimately, although we have achieved significant progress in model lightweighting, our exploration of multitasking capabilities remains somewhat limited. Looking ahead, we might enhance the model's ability to concurrently learn multiple tasks, such as simultaneous abnormality detection and health assessment of power transmission lines, to offer a more comprehensive solution for monitoring electrical infrastructure.

REFERENCES

- [1] J. Yuan, X. Zheng, L. Peng, K. Qu, H. Luo, L. Wei, J. Jin, and F. Tan, "Identification method of typical defects in transmission lines based on YOLOv5 object detection algorithm," *Energy Rep.*, vol. 9, pp. 323–332, Sep. 2023, doi: [10.1016/j.egy.2023.04.078](https://doi.org/10.1016/j.egy.2023.04.078).
- [2] M. D. F. Ahmed, J. C. Mohanta, A. Sanyal, and P. S. Yadav, "Path planning of unmanned aerial systems for visual inspection of power transmission lines and towers," *IETE J. Res.*, vol. 70, no. 3, pp. 3259–3279, Mar. 2024, doi: [10.1080/03772063.2023.2175053](https://doi.org/10.1080/03772063.2023.2175053).
- [3] C. Chen, Z. Zheng, T. Xu, S. Guo, S. Feng, W. Yao, and Y. Lan, "YOLO-based UAV technology: A review of the research and its applications," *Drones*, vol. 7, no. 3, p. 190, Mar. 2023, doi: [10.3390/drones7030190](https://doi.org/10.3390/drones7030190).

- [4] H. Li, Y. Dong, Y. Liu, and J. Ai, "Design and implementation of UAVs for bird's nest inspection on transmission lines based on deep learning," *Drones*, vol. 6, no. 9, p. 252, Sep. 2022, doi: [10.3390/drones6090252](https://doi.org/10.3390/drones6090252).
- [5] V. K. Sharma and R. N. Mir, "A comprehensive and systematic look up into deep learning based object detection techniques: A review," *Comput. Sci. Rev.*, vol. 38, Nov. 2020, Art. no. 100301, doi: [10.1016/j.cosrev.2020.100301](https://doi.org/10.1016/j.cosrev.2020.100301).
- [6] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Sep. 1999, pp. 1150–1157, doi: [10.1109/ICCV.1999.790410](https://doi.org/10.1109/ICCV.1999.790410).
- [7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 886–893, doi: [10.1109/CVPR.2005.177](https://doi.org/10.1109/CVPR.2005.177).
- [8] A. Mammone, M. Turchi, and N. Cristianini, "Support vector machines," *WIREs Comput. Statist.*, vol. 1, no. 3, pp. 283–289, Nov. 2009, doi: [10.1002/wics.49](https://doi.org/10.1002/wics.49).
- [9] X. Wu, D. Sahoo, and S. C. H. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing*, vol. 396, pp. 39–64, Jul. 2020, doi: [10.1016/j.neucom.2020.01.085](https://doi.org/10.1016/j.neucom.2020.01.085).
- [10] Z. Li, Y. Wang, N. Zhang, Y. Zhang, Z. Zhao, D. Xu, G. Ben, and Y. Gao, "Deep learning-based object detection techniques for remote sensing images: A survey," *Remote Sens.*, vol. 14, no. 10, p. 2385, May 2022, doi: [10.3390/rs14102385](https://doi.org/10.3390/rs14102385).
- [11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587, doi: [10.1109/CVPR.2014.81](https://doi.org/10.1109/CVPR.2014.81).
- [12] R. B. Girshick. (Apr. 2015). *Fast R-CNN*. Accessed: Jan. 21, 2024. [Online]. Available: <https://www.semanticscholar.org/paper/Fast-R-CNN-Girshick/7ffdb358b63378f07311e883dddacc9faeaf4b>
- [13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- [14] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Computer Vision—ECCV 2016 (Lecture Notes in Computer Science)*, vol. 9905, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Cham, Switzerland: Springer, 2016, pp. 21–37, doi: [10.1007/978-3-319-46448-0_2](https://doi.org/10.1007/978-3-319-46448-0_2).
- [15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788, doi: [10.1109/CVPR.2016.91](https://doi.org/10.1109/CVPR.2016.91).
- [16] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525. Accessed: Apr. 28, 2024. https://openaccess.thecvf.com/content_cvpr_2017/html/Redmon_YOLO9000_Better_Faster_CVPR_2017_paper.html
- [17] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
- [18] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
- [19] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, Y. Li, B. Zhang, Y. Liang, L. Zhou, X. Xu, X. Chu, X. Wei, and X. Wei, "YOLOv6: A single-stage object detection framework for industrial applications," 2022, *arXiv:2209.02976*.
- [20] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475, doi: [10.1109/cvpr52729.2023.00721](https://doi.org/10.1109/cvpr52729.2023.00721).
- [21] J. Cao, W. Bao, H. Shang, M. Yuan, and Q. Cheng, "GCL-YOLO: A GhostConv-based lightweight Yolo network for UAV small object detection," *Remote Sens.*, vol. 15, no. 20, p. 4932, Oct. 2023, doi: [10.3390/rs15204932](https://doi.org/10.3390/rs15204932).
- [22] H. Li, L. Liu, J. Du, F. Jiang, F. Guo, Q. Hu, and L. Fan, "An improved YOLOv3 for foreign objects detection of transmission lines," *IEEE Access*, vol. 10, pp. 45620–45628, 2022, doi: [10.1109/ACCESS.2022.3170696](https://doi.org/10.1109/ACCESS.2022.3170696).
- [23] S. Huang, X. Dong, Y. Wang, and L. Yang, "Detection of insulator burst position of lightweight YOLOv5," in *Proc. 8th Int. Conf. Comput. Artif. Intell.*, Mar. 2022, pp. 573–578, doi: [10.1145/3532213.3532300](https://doi.org/10.1145/3532213.3532300).
- [24] Y. Chen, H. Liu, J. Chen, J. Hu, and E. Zheng, "Insu-YOLO: An insulator defect detection algorithm based on multiscale feature fusion," *Electronics*, vol. 12, no. 15, p. 3210, Jul. 2023, doi: [10.3390/electronics12153210](https://doi.org/10.3390/electronics12153210).
- [25] L. Zhang, B. Li, Y. Cui, Y. Lai, and J. Gao, "Research on improved YOLOv8 algorithm for insulator defect detection," *J. Real-Time Image Process.*, vol. 21, no. 1, p. 22, Jan. 2024, doi: [10.1007/s11554-023-01401-9](https://doi.org/10.1007/s11554-023-01401-9).
- [26] X. Zhang, Y. Song, T. Song, D. Yang, Y. Ye, J. Zhou, and L. Zhang, "LDConv: Linear deformable convolution for improving convolutional neural networks," 2023, *arXiv:2311.11587*.
- [27] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8759–8768, doi: [10.1109/CVPR.2018.00913](https://doi.org/10.1109/CVPR.2018.00913).
- [28] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944, doi: [10.1109/CVPR.2017.106](https://doi.org/10.1109/CVPR.2017.106).
- [29] X. Wang, H. Gao, Z. Jia, and Z. Li, "BL-YOLOv8: An improved road defect detection model based on YOLOv8," *Sensors*, vol. 23, no. 20, p. 8361, Oct. 2023, doi: [10.3390/s23208361](https://doi.org/10.3390/s23208361).
- [30] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10778–10787, doi: [10.1109/CVPR42600.2020.01079](https://doi.org/10.1109/CVPR42600.2020.01079).
- [31] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13708–13717, doi: [10.1109/CVPR46437.2021.01350](https://doi.org/10.1109/CVPR46437.2021.01350).
- [32] J. Li, D. Yan, K. Luan, Z. Li, and H. Liang, "Deep learning-based bird's nest detection on transmission lines using UAV imagery," *Appl. Sci.*, vol. 10, no. 18, p. 6147, Sep. 2020, doi: [10.3390/app10186147](https://doi.org/10.3390/app10186147).
- [33] X. Tao, D. Zhang, Z. Wang, X. Liu, H. Zhang, and D. Xu, "Detection of power line insulator defects using aerial images analyzed with convolutional neural networks," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 4, pp. 1486–1498, Apr. 2020, doi: [10.1109/TSMC.2018.2871750](https://doi.org/10.1109/TSMC.2018.2871750).
- [34] H. Jiang, F. Hu, X. Fu, C. Chen, C. Wang, L. Tian, and Y. Shi, "YOLOv8-peas: A lightweight drought tolerance method for peas based on seed germination vigor," *Frontiers Plant Sci.*, vol. 14, Sep. 2023, Art. no. 1257947, doi: [10.3389/fpls.2023.1257947](https://doi.org/10.3389/fpls.2023.1257947).
- [35] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy, "Speed/accuracy trade-offs for modern convolutional object detectors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3296–3297. Accessed: Jun. 12, 2024. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2017/html/Huang_Speed_Accuracy_Trade-Offs_for_CVPR_2017_paper.html
- [36] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626. Accessed: Mar. 6, 2024. [Online]. Available: https://openaccess.thecvf.com/content_iccv_2017/html/Selvaraju_Grad-CAM_Visual_Explanations_ICCV_2017_paper.html



CHUANYAO LIU is currently pursuing the M.E. degree with the School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing, China. His research interests include deep learning and point cloud processing.



SHUANGFENG WEI was born in Hubei, China, in 1979. He received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2007. Since 2007, he has been with Beijing University of Civil Engineering and Architecture, Beijing, China, where he is currently an Associate Professor. He has authored or co-authored about 60 papers. His research interests include point cloud processing and SLAM.



SHAOBO ZHONG was born in Hubei, China, in 1978. He received the Ph.D. degree in cartography and GIS from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 2006. From 2006 to 2018, he was an Educator and a Researcher with Tsinghua University, Beijing. He is currently a Researcher with Beijing Academy of Science and Technology, Beijing. He has authored or co-authored more than 140 papers. His research

interests include geoinformation for disasters, resilient and risk assessment, and spatial big data and machine learning for spatial and temporal problem solving, particularly in safety operations of smart city.



FAN YU was born in Xiaogan, Hubei, China, in 1982. He received the B.S. and M.S. degrees in photogrammetry and remote sensing from Wuhan University, in 2004 and 2007, respectively, and the Ph.D. degree in remote sensing from the Chinese Academy of Sciences, in 2010.

From July 2010 to 2018, he was an Associate Researcher and an Academic Secretary with the Key Laboratory of Geospatial Information Engineering, State Bureau of Surveying and Mapping, China Academy of Surveying and Mapping Sciences. Since July 2018, he is an Associate Professor with the School of Surveying and Mapping and Urban Spatial Information, Beijing University of Civil Engineering and Architecture. He has published over 30 SCI/EI journal articles, and hosted two National Natural Science Foundation projects. He has rich experience in the preprocessing of remote sensing images, research on segmentation and classification algorithms on machine learning/data mining for remote sensing information extraction.

...