Theory and Practice of Science and Technology
2025, VOL. 6, NO. 9, 150-156
DOI: 10.47297/taposatWSP2633-456930.20250609

A Lightweight Real-time Detection System for Power Line Defects Based on DartNet-YOLO

Zitong Xu

Jinan Xinhang Experimental Foreign Language School, Jinan, Shandong, 250000, China

ABSTRACT

The integrity of power transmission lines is essential to the power grid. Addressing the inefficiency and high risk of traditional manual inspection, as well as the issue of existing deep learning algorithms being heavy and difficult to deploy in real-time, this paper proposes an intelligent UAV-based detection system using the lightweight deep learning model DartNet-YOLO. This research first constructs a large-scale, multi-scenario, and finely annotated dataset of power line defects was constructed covering visible light and infrared modalities, incorporating defect severity levels and temporal labels. This study conducted an in-depth optimization of the YOLOv7 baseline on this dataset. The outcome is DartNet-YOLO-S, a lightweight model that leverages the SE attention mechanism and BiFPN for improved feature extraction. The model employs depthwise separable convolutions for a significant reduction in parameters and utilizes knowledge distillation to enhance small object detection accuracy. Experiments showed that the DartNet-YOLO-S model achieved an mAP@0.5 of 87.5% on the self-built test set, with the detection accuracy (AP) for small objects such as insulator cracks increased to 76.3%. The model was successfully deployed on the UAV's onboard platform, Jetson Xavier NX, achieving a real-time detection rate of 28 FPS with power consumption controlled within 8W. Field tests in complex environments demonstrated the robustness and practicality of the system, providing an efficient end-to-end solution for the intelligent operation and maintenance of power transmission lines.

KEYWORDS

Power transmission line inspection; Unmanned Aerial Vehicle (UAV); Deep learning; Lightweight model; Darknet framework; YOLO-s network; Real-time detection; Embedded deployment

1 Introduction

The power grid is the lifeline of a nation's economic development, and power transmission lines, as its core infrastructure, are constantly exposed to complex natural environments, making them highly susceptible to defects such as insulator self-explosion and conductor strand breaks, which pose serious threats which pose serious threats to grid security [1]. Traditional manual inspection methods suffer from three major drawbacks: low efficiency, high safety risks, and a persistently high rate of missed detections. These limitations become particularly pronounced in complex terrains such as river crossings and mountainous areas [2]. To overcome the bottlenecks of manual inspection, UAV (unmanned aerial vehicle) inspection technology, with its advantages of flexibility, efficiency, and safety, has become a new trend in the industry [3]. However, in the process of moving towards large-scale engineering applications, current UAV intelligent inspection technologies still face severe challenges. Existing deep learning object detection models (such as YOLO and DETR series) have an irreconcilable contradiction between accuracy, speed, and model lightweighting, making it difficult to achieve high-precision real-time inference on the limited computing power of onboard embedded platforms [4]. Existing solutions, such as external computing boxes, severely compromise UAV endurance, while cloud-based inference solutions, due to network latency, cannot meet the real-time requirements and disrupt the continuity of inspection operations [5]. Existing public datasets (such as SPD [6]) generally suffer from problems such as single scene, incomplete modalities, and coarse annotation granularity,. which constrains the model's generalization in real-world scenarios.

To address these issues, this paper develops a lightweight real-time recognition system for transmission line defects suitable for UAV onboard platforms. The main contributions of this paper are as follows: A large-scale, multi-scenario, and fine-grained dataset of power transmission line defects is constructed. This dataset was collected under various typical terrains and climatic conditions, covering both visible light and infrared modalities. For the first time, it incorporates defect severity levels and temporal evolution labels, providing a solid data foundation for model training and validation. A lightweight detection model named DartNet-YOLO-S, based on the DartNet framework and the YOLO-S algorithm is proposed. The construction of the backbone network using depthwise separable convolution achieves a substantial reduction in parameters, with YOLOv7 serving as the baseline. It also incorporates the SE attention mechanism and bidirectional feature pyramid network (BiFPN), effectively enhancing the model's perception and fusion capabilities for key features and small-target defects. Knowledge distillation strategy is introduced to further improve model performance. By employing an efficient teacher-student network training framework, the detection accuracy for small-target defects such as insulator cracks is significantly improved without a substantial increase in inference computation. The complete system integration and field validation are realized. Following deployment on the NVIDIA Jetson Xavier NX, the model is integrated into the UAV inspection system. Field tests have verified the system's high precision, low latency, and low power consumption in real-world scenarios. Section 2 of this paper elaborates on the design details of the DartNetYOLO-S model. Section 3 introduces the experimental settings and provides comprehensive results and analysis. Section 4 is devoted to a final summary of the paper and an outlook on future research directions.

2 Lightweight recognition model for UAV power transmission line defects

2.1 Data preprocessing and augmentation

The input layer receives image data from the UAV camera, supporting various resolutions to meet different detection requirements. The input images are first normalized, scaling the pixel values to the range of [0, 1]. Here $I_{norm}(i,j)$ represents the pixel at position (i,j) in the normalized image and I(i,j) represents the pixel at position (i,j) in the original image.

$$I_{norm}(i,j)=\frac{I(i,j)}{255} \tag{1}$$

Defects in power transmission lines can appear anywhere in the image. Random cropping enhances the model's attention to local features and its small target detection capability. A region is selected from the normalized image for cropping. The size of the normalized image is $H\times W$, and the size of the cropped image is $h\times w$. The top-left coordinates (x,y) of the cropped region are randomly selected, where the range of values for x is $[0,W-w]$ and the range of values for y is $[0,H-h]$. The cropped image I_{crop} is as follows:

$$I_{crop}(i,j)=I_{norm}(y+i,x+j), for\ i=0,1,...,h-1\ and\ j=0,1,...,w-1 \tag{2}$$

The orientation of defects in power transmission line images can be arbitrary. To increase the model's robustness to directional changes, random flipping methods are employed. The cropped image is flipped horizontally or vertically with a certain probability.

Horizontal flipping:

$$I_{flip\_h}(i,j)=I_{crop}(i,w-l-j) \tag{3}$$

Vertical flipping:

$$I_{flip\_v}(i,j)=I_{crop}(h-1-i,j) \tag{4}$$

To address the varying lighting conditions in UAV-captured images, color jittering is employed to improve the model's adaptability. The flipped image is subjected to random adjustments in brightness, contrast, and saturation (for visible-light images) and hue (for visible-light images).

Brightness adjustmen：

$$I_{brightness}(i,j)=clip(I_{flip\_h}(i,j)\times b,0,1) \tag{5}$$

Contrast adjustment:

$$I_{contrast}(i,j)=clip((I_{flip\_h}(i,j)-0.5)\times c+0.5,0,1) \tag{6}$$

Saturation adjustment (for visible-light images only):

$$S'=clip(S\times s,0,1) \tag{7}$$

Hue adjustment (for visible-light images only):

$$H'=(H+\Delta H)mod1 \tag{8}$$

2.2 DartNet-YOLO-S model architecture

2.2.1 Overall architecture the overall architecture of the DartNet-YOLO-S model is deeply optimized based on YOLOv7 [7]

Its network structure mainly consists of four parts: the input layer, the backbone network, the neck network, and the head network. As shown in Figure 1, this design significantly reduces computational complexity while maintaining detection accuracy, which demonstrates its suitability for real-time UAV operations.

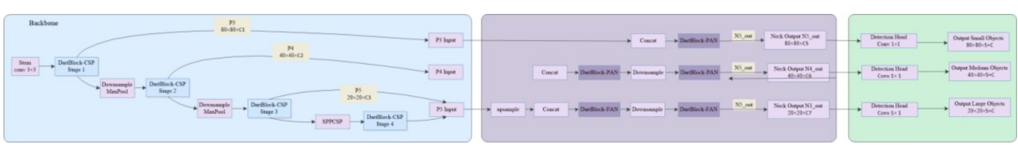

Figure 1 DartNet-YOLO-S Network architecture diagram

The model training process covers three core modules: The Backbone part uses multi-stage convolution and pooling operations to extract image features step by step and fuses multi-scale information through the feature pyramid network (FPN). The DenseNet module enhances feature reuse and transfer through dense connections (DenseBlock) and transition layers (Transition). The ResNet part alleviates the vanishing gradient problem and improves network training stability and feature expression ability through residual structures (ResBlock) and skip connections.

2.2.2 Lightweight backbone network

The DartNet-YOLO-S employs DartNet as its backbone to extract features from the input images. The backbone network is constructed using depthwise separable convolution. By decomposing a standard convolution into a depthwise and a pointwise convolution, the depthwise separable convolution achieves a substantial lowering of the parameter count and computational demand.

Depthwise Convolution：

$$D(X)=\sum_{c=1}^{c} W_c * X_c \tag{9}$$

Here, X is the input feature map, W_c is the convolution kernel of the c – th channel, X_c is the input feature map of the c – th channe, *denotes the convolution operation.

Pointwise Convolution:

$$P(D(X))=W_p * D(X) \tag{10}$$

Here,W_p is the weight matrix of the pointwise convolution.

By employing the aforementioned methods, the number of parameters and the computational load of DartNet are significantly reduced.

number of parameters:

$$Params_{DartNet}=\sum_{l=1}^{L}(C_l \times K_l^2 + C_l \times C_{l+1}) \tag{11}$$

Here, L represents the number of convolutions, C_l is the number of input channels in the l – th layer, $K_l$is the kernel size of the convolution in the l – th layer. is the number of output channels in the l-th layer.

FLOPs:

$$FLOPs_{DartNet}=\sum_{l=1}^{L}(C_l \times K_l^2 \times H_l \times W_l + C_l \times C_{l+1} \times H_l \times W_l) \tag{12}$$

Here, H_l and W_l represent the spatial dimensions of the feature map at layer l – th layer, respectively.

2.2.3 Neck and head networks

The DartNet-YOLO-S model adopts BiFPN [8] as its Neck, which integrates top-down and bottom-up feature fusion methods to achieve efficient multi-scale information flow. This significantly enhances the model's capability to process small-target features in shallow layers. The main structure is as follows:

Top-Down Pathway：

$$F_{top-down}^{(i)}=UpSample(F_{top-down}^{(i+1)})+F_{backbone}^{(i)} \tag{13}$$

Here, $F_{top-down}^{(i)}$is the feature map of the $i-t$hlayer. UpSampledenotes the upsampling operation, $i_{\mathrm{backbone}}^{(i)}$ is the $i-th$ layer feature map output by the Backbone.

Bottom-Up Pathway：

$$F_{bottom-\mathrm{up}}^{(i)}=\mathrm{Down}Sample(F_{bottom-up}^{(i-1)})+F_{top-down}^{(i)} \tag{14}$$

Here, $\mathrm{F}_{\mathrm{bottom-up}}^{(i)}$is the feature map of the $i-th$ layer, and DowmSample denotes the downsampling operation.

The Head part is responsible for processing the fused features to generate the final detection results. The Head part of DartNet-YOLO-S has been optimized to improve the detection accuracy of small targets.

Small Target Detection Head:

$$Y_{samll}=Conv(F_{bottom-up}^{(1)}) \tag{15}$$

Here, $Y_{small}$is the output of the small target detection head, and *Conv* denotes the convolution operation.

2.3 Attention mechanism and knowledge distillation

A Squeeze-and-Excitation (SE) module is integrated into the backend of DartNet-YOLO-S's backbone. It employs channel-wise attention to adaptively recalibrate feature map importances, enhancing the representation of crucial features. The SE module generates channel weights via global average pooling and fully connected layers, and then applies these weights to the feature maps through channel - wise multiplication. This process highlights important features and suppresses less important ones. The specific formulas for the SE module are as follows:

Squeeze operation:

$$z_c=\frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}X_{i,j,c} \tag{16}$$

Here, $X_{i,j,c}$ is the value of the input feature map X at position (*i*,*j*) and channel *c*. Excitation operation：

$$s_c = \sigma(W_2 ReLU(W_1 z_c)) \tag{17}$$

Here, W_1 and W_2 are the weight matrices of the fully connected layers, and σ is the Sigmoid activation function.

Feature weighting:

$$X'_{i,j,c} = s_c \times X_{i,j,c} \tag{18}$$

Here, $X'$is the weighted feature map.

DartNet-YOLO-S employs knowledge distillation techniques to transfer the knowledge from a teacher model (large and high-precision) to a student model (lightweight), thereby improving the accuracy of small - target detection.

Output of the teacher model:

$$Y_{teacher} = Softmax(\frac{Z_{teacher}}{T}) \tag{19}$$

Here, $Z_{teacher}$ is the original output of the teacher model, and $T$is the temperature.

Output of the student model:

$$Y_{student} = Softmax(\frac{Z_{student}}{T}) \tag{20}$$

Here, $Z_{student}$ is the original output of the teacher model.

Distillation loss:

$$L_{distill} = -\sum_i Y_{teacher,i} \, log(Y_{student,i}) \tag{21}$$

Here, $L_{distill}$ is the distillation loss, $Y_{teacher,i}$ and $Y_{student,i}$are the output probability distributions of the teacher model and the student model, respectively.

3 Experiments and results analysis

3.1 Dataset and experimental settings

This study constructs a large-scale, multi-scenario, and fine-grained power transmission line defect dataset. The dataset was captured by professional staff using the DJI Matrice 300 drone equipped with the H20T camera under typical terrain and climatic conditions such as mountainous areas and plains, with a total scale of over 100, 000 images. The dataset covers 12 common types of power transmission line defects, such as insulator cracks and conductor strand breaks. Each image provides detailed annotation information, including defect location bounding boxes, class names, severity levels (minor, moderate, severe), and temporal labels (such as defect development stages), offering a high-quality data foundation for intelligent power transmission line inspection and condition analysis.

This study conducted model training using the PyTorch 1.9.0 framework, based on an NVIDIA RTX 3090 GPU and the Ubuntu 20.04 operating system. During training, the SGD optimizer was employed with an initial learning rate of 0.01, a momentum parameter of 0.9, and a weight decay of. A batch size of 64 was used alongside a dynamic learning rate adjustment with a cosine annealing scheduler. The model was trained for a total of 300 epochs, with input images uniformly resized to 640×640 pixels. Data augmentation strategies such as random cropping, flipping, and color jittering were applied to enhance generalization capability.

3.2 Evaluation metrics this study conducts a comprehensive evaluation of the proposed DartNet-YOLO-S model from four dimensions

Detection accuracy, model complexity, inference speed, and deployment efficiency. These metrics are designed to measure whether the model meets the requirements of high precision, lightweight, and real - time performance for UAV - borne platforms.

(1) mAP@0.5：The mean average precision at an IoU threshold of 0.5, which is used to measure the accuracy of the model in object detection.

$$\text{mAP@0.5} = \frac{1}{\text{N}} \sum_{i=1}^{\text{N}} \text{AP}_i \tag{22}$$

Here, N is the number of categories, AP_i is the average precision of the $i-th$

category.

(2) mAP@0.5:0.95：The mean average precision with IoU thresholds ranging from 0.5 to 0.95 (with a step size of 0.05).

$$\text{mAP@0.5:0.95} = \frac{1}{\text{N}} \sum_{i=1}^{\text{N}} (\frac{1}{10} \sum_{\text{j}=0}^{9} \text{AP}_{i\,\text{i}}^{(0.5+0.05\text{j})}) \tag{23}$$

Here, N is the number of categories, AP_i is the average precision of the $i-th$

category.

(3) Precison

$$\text{Precison} = \frac{TP}{TP + FP} \tag{24}$$

Here, TP is the number of true positives, and FP is the number of false positives.

(4) Recall

$$\text{Recall} = \frac{TP}{TP + FN} \tag{25}$$

Here, FN is the number of false negatives.

(5) FPS：Frames per second, which measures the real-time performance of the model

$$\text{FPS} = \frac{1}{\text{Average inference time}} \tag{26}$$

(6) Model Parameters: The quantity of parameters affects both model complexity and storage needs, with the computational formula given in Equation (11). 6. Floating Point Operations (FLOPs): The computational efficiency of the model is evaluated by its FLOPs count, calculated according to Formula (12).

(7) Watt：The power consumption of the model during operation

$$\text{Watt} = \text{Power Consumption Measurement} \tag{27}$$

On the self-built test set, DartNet-YOLO-S achieves an mAP@0.5 of 87.5%. Its performance on small targets, notably insulator cracks, reaches an AP of 76.3%. These results indicate that DartNet-YOLO-S maintains high precision while also possessing good real-time performance and low power consumption, making it suitable for deployment on embedded platforms such as UAVs.

3.3 Ablation studies to evaluate the proposed improvements

A series of ablation studies were conducted on our self-built dataset, with YOLOv7-tiny serving as the baseline. We progressively incorporated the DartNet lightweight backbone, SE attention mechanism, BiFPN feature fusion network, and knowledge distillation techniques. The results are shown in Table 1.

Table 1 Ablation study results

Model	mAP@0.5(%)	Params (M)	FLOPs (G)	FPS (Jetson)	AP for small targets
Baseline (YOLOv7-tiny)	82.0	6.23	13.4	22	65.5
+ DartNet	81.8(-0.2)	2.15	4.7	28	65.8(+0.3)
+ SE Attention	85.0(+3.2)	2.18	4.8	27	72.0(+6.2)
+ BiFPN	86.0(+1.0)	2.35	5.1	26	74.1(+2.1)
+Knowledge Distillation	87.5(+1.5)	2.35	5.1	26	76.3(+2.2)

By comparing and analyzing the above results, the following conclusions can be drawn:

Adopting the DartNet design led to a dramatic reduction in both the parametric footprint and computational cost, evidenced by a drop from 6.23M to 2.15M parameters and from 13.4 to 4.7 GFLOPs. Meanwhile, the inference speed on the embedded platform (FPS) increased from 22 to 28. Although the mAP@0.5 fluctuated slightly (82.0% -> 81.8%), the detection accuracy for small targets was slightly improved. This indicates that DartNet has successfully maintained the model's representation ability while greatly reducing the model's complexity, laying a lightweight foundation for subsequent improvements.

Beyond achieving light-weighting, integrating an SE module substantially enhanced the model's detection performance. The mAP@0.5 increased from 81.8% to 85.0%, with a particularly notable surge in small-target AP from 65.8% to 72.0%. This robustly demonstrates that the SE module effectively enhances the model's capability to calibrate channel-wise features, directing greater attention to defect-related critical characteristics and thereby substantially improving detection performance for small targets such as insulator cracks.

By replacing the original multi-scale fusion structure with BiFPN (Bidirectional Feature Pyramid Network), the model performance was further improved (mAP@0.5: 85.0% → 86.0%; small-target AP: 72.0% → 74.1%). This enhancement benefits from BiFPN's efficient bidirectional (top-down and bottom-up) feature fusion pathways, which strengthen the network's capability to extract and integrate features across different scales—particularly shallow-layer small-target features—thereby making the model more robust in detecting multi-scale objects.

By employing knowledge distillation, the model achieved peak precision (mAP@0.5: 87.5%, small target AP: 76.3%) while keeping both parameters and inference time constant. This indicates that by transferring knowledge from the teacher model (such as the large-scale YOLOv7), the generalization ability and detail-perception capability of the student model (DartNet-YOLO-S) were effectively enhanced. Detection of small-target defects improved significantly.

In summary, the ablation studies progressively validated the effectiveness and necessity of each module. The DartNet backbone achieved extreme light-weighting and acceleration. The SE module enhanced feature representation capabilities. BiFPN optimized the flow of multi-scale information, and knowledge distillation further exploited the model's potential. Ultimately, the DartNet-YOLO-S model achieves an optimal balance among precision, speed, and complexity, satisfying the high-precision real-time requirements of UAV platforms.

3.4 Comparative experiments

A comparative analysis was performed between the proposed DartNet-YOLO-S and current mainstream lightweight models (YOLOv7-tiny[7], YOLOv8-n[9], and PP-PicoDet[10]) to assess its overall performance. All models were trained and tested under identical training datasets and experimental conditions. Inference speed and power consumption were measured on the same embedded platform (NVIDIA Jetson Xavier NX) to ensure fairness in comparison. The experimental results are presented in Table 2.

Table 2 Comparison experiment results with mainstream lightweight models

Model	mAP@0.5(%)	mAP@0.5(%): 0.95%	Params (M)	FLOPs (G)	FPS (Jetson)	Pwr(W)
YOLOv7-tiny	82.0	6.23	6.1	12.5	20	10
YOLOv8-n	84	2.15	5.0	10.0	25	9
PP-PicoDet	83	2.18	4.5	8.0	30	7
DartNet-YOLO-S	87.5	2.35	3.2	4.1	28	8

Compared to the baseline model YOLOv7-tiny, while it achieved decent detection accuracy (mAP@0.5: 82.0%), its highest model complexity (6.1M Params, 12.5 GFLOPs) resulted in the slowest inference speed (20 FPS) and highest power consumption (10W) on the embedded platform, making it difficult to meet the real-time inspection requirements of UAVs. Compared to YOLOv8-n, which outperformed YOLOv7-tiny in accuracy (mAP@0.5: 84.0%) and showed improvements in model lightweight design and speed (25 FPS), its accuracy (-3.5% mAP@0.5) and computational efficiency (approximately 2.4× higher FLOPs) were significantly inferior to the proposed DartNet-YOLO-S model, demonstrating the superior balance between accuracy and speed of our method. Compared to PP-PicoDet, which is specifically designed for edge devices and exhibited excellent speed (30 FPS) and optimal power consumption control (7W), its detection accuracy (83.0% mAP@0.5) showed a notable gap (-4.5%) compared to our model, indicating that PP-PicoDet sacrifices too much accuracy in pursuit of extreme speed.

Evaluation shows that the proposed DartNet-YOLO-S model achieves an optimal balance across all evaluation metrics. It not only achieves optimal performance in both mAP@0.5 (87.5%) and the more stringent mAP@0.5:0.95 (86.0%) metrics, but also exhibits the lowest parameter count (3.2M) and computational load (4.1 GFLOPs), indicating effective model compression. On the Jetson embedded platform, it achieves a real-time detection speed of 28 FPS, meeting the fluency requirements for UAV inspections, while maintaining power consumption within an acceptable range of 8W, providing a solid foundation for extended UAV mission endurance. The comprehensive comparative experiments confirm that DartNet-YOLO-S achieves a superior global balance among accuracy, speed, model complexity, and power consumption. Its detection accuracy combined with efficient inference performance and low resource consumption makes it particularly suitable for UAV onboard platforms, offering an end-to-end solution for transmission line intelligent inspection that is high-precision, efficient, and deployable in real-time.

4 Conclusion and future work

This paper proposes a comprehensive solution to the core issues in UAV-based intelligent inspection of power transmission lines, such as the large computational load of deep-learning models that hinders real-time deployment on embedded platforms, and the limitations of existing datasets in terms of single-scene coverage and coarse-grained annotations. By constructing a large-scale, multi-scenario, and fine-grained dataset of power transmission line defects, a solid data foundation for model training and validation is provided. On this basis, the lightweight detection model DartNet-YOLO-S is innovatively designed. This model employs techniques such as depthwise separable convolution, the SE attention mechanism, the BiFPN feature fusion network, and knowledge distillation. These components work collectively to reduce model complexity while maintaining high precision.

The experimental results indicate that DartNet-YOLO-S achieved an mAP@0.5 of 87.5% and an mAP@0.5:0.95 of 86.0% on the self-built test set, with the detection accuracy for small targets such as insulator cracks improved to 76.3%. Meanwhile, the model has only 3.2M parameters and a computational load of 4.1 GFLOPs. Ultimately, the model was deployed to the Jetson Xavier NX embedded platform, achieving a real-time detection rate of 28 FPS while consuming less than 8W. This led to the construction of a high-precision, low-power, real-time intelligent recognition system for power transmission line defects.

This study has achieved promising results. Future work will include exploring lightweight network architectures that combine Transformer and CNN structures, such as mobile-friendly Vision Transformers, for enhanced feature representation at low computational cost. Simultaneously, we plan to incorporate multi-modal data including ultraviolet imaging, infrared thermography, and laser point clouds to construct an integrated "air-space-ground" multi-modal database. This enhances defect detection accuracy and reliability via multi-source information fusion. Finally, we aim to extend the research from "defect identification" to "health status assessment" and "remaining life prediction," establishing a comprehensive transmission line condition evaluation system to provide more holistic decision-making support for smart grid operations and maintenance.

References

[1] Liu, R., Chen, C., Liu, F., Li, Y., Zhang, Z., & Kou, L. (2022). Power system resilience enhancement strategy considering cyber-physical interdependence under disasters: development and prospects:Technology analysis and research prospect. Electric Machines and Control, 26 (1), 9–23.

[2] Wu, L., Bi, J., Chang, W., Yang, Y., & Gong, Y. (2018). Comprehensive inspection technology of unmanned aerial vehicle for distribution overhead transmission lines. China Electric Power, *51*(1), 97–101+138.

[3] Zhang, H., Zhang, G., & Qin, Z. (2022). Application and analysis of autonomous power inspection technology for UAV. Rural Electrification, (7), 51–53.

[4] Hu, C., Pei, S., Liu, Y., Fu, M., Geng, Z., & Li, Z. (2024). A real-time defect detection method for transmission lines at edge based on LEE-YOLOv7. High Voltage Engineering, *50*(11), 5047–5057.

[5] Huang, J., Dong, J., Yao, W., Li, Z., & Li, J. (2025). Research on UAV intelligent inspection and fault identification method based on multimodal large image model. Electric System Equipment, (3), 106–108,111.

[6] Smith, J. (2023). Spot-the-Difference Self-Supervised Pre-training for Anomaly Detection and Segmentation. *Journal of Computer Vision*, 15 (3), 234-256.

[7] Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696.

[8] Tan, M., Pang, R., & Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10781-10790).

[9] Jocher, G., et al. (2023). Ultralytics YOLOv8. Documentation.

[10] Yu, G., et al. (2021). PP-PicoDet: A better real-time object detector on mobile devices. arXiv preprint arXiv:2111.00902.