

A Precision Detection Method for Key Components of Power Transmission Towers Oriented to UAV Autonomous Localization

Luqi Zhang¹, Yunzuo Zhang¹✉, Song Tang²✉, Wei He², Tianliang Zhang², Yubo Hu¹

(1. School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang 050043, China; 2. The Institute of Applied Mathematics, Hebei Academy of Sciences, and the Hebei Information Security Authentication Technology Innovation Center, Shijiazhuang 050081, China)

Abstract: To address the challenges of multi-scale differences, complex background interference, and unstable small target positioning in visual inspection of power towers, the existing methods still face issues such as insufficient feature interaction and unstable confidence estimation, which lead to performance degradation in complex backgrounds and occlusion conditions. This paper proposes a precise inspection method for key power tower components using autonomous drone positioning. To this end, this paper makes three key improvements to the you only look once version 11 (YOLOv11) framework. First, it constructs C3k2-adaptive multi-receptive field block (C3k2-AMRB), combining multiple dilated convolutions with a reparameterization mechanism to significantly expand the receptive field and enhance multi-scale feature extraction. Second, it designs a hierarchical wavelet interaction unit (HWIU), which leverages high- and low-frequency decomposition and reconstruction of wavelet transform (WT) to achieve cross-scale semantic alignment, enhancing feature discriminability in complex backgrounds. Third, it proposes a distribution-aware confidence refinement head (DACR-Head), which adaptively calibrates classification confidence based on the statistical characteristics of the predicted bounding-box corner distribution, improving the localization stability and accuracy of small targets. Experiments on the inspection of power line assets dataset (InsPLAD) dataset show that the integrated approach achieves a component detection accuracy at intersection over union (IoU)=0.5 (CDA_{50}) of 88.3% and a component detection robustness ($CDR_{50:95}$) of 69.8%, respectively, improvements of 4.4% and 7.0% over the baseline.

Keywords: unmanned aerial vehicle (UAV) autonomous localization; power transmission tower; object detection; wavelet-based feature interaction; confidence calibration

1 Introduction

In modern society, electric power is an indispensable resource, and its stable supply depends on the secure operation of high-voltage overhead transmission lines [1]. As the supporting struc-

tures of these lines, high-voltage transmission towers are exposed to harsh environments over long periods and therefore require regular inspection [2]. Traditional manual inspection is hindered by the remote distribution of towers and complex terrain, leading to low efficiency, high cost, and safety risks. Moreover, in field or mountainous environments, global positioning system signals often suffer from drift or occlusion, degrading unmanned aerial vehicle (UAV) localization accuracy and increasing flight risk. Among multimodal sensing modalities, vision

Manuscript received Oct. 31, 2025; revised Nov. 12, 2025; accepted Nov. 18, 2025. The associate editor coordinating the review of this manuscript was Dr. Lijuan Jia. This work was supported by the National Natural Science Foundation of China (No. 61702347), Hebei Academy of Sciences Basic Research Operating Fund Project (No. 2025PF21).

✉ Corresponding author. Email: zhangyunzuo888@sina.com, tangsong@tju.edu.cn

DOI: [10.15918/j.jbit1004-0579.2025.082](https://doi.org/10.15918/j.jbit1004-0579.2025.082)

offers the most favorable cost-performance ratio while providing the richest target information, and thus acts as a key determinant of the performance ceiling for UAV-based transmission-line inspection systems [3]. Benefiting from their low cost, high maneuverability, and multi-view imaging capability, UAVs have emerged as an effective platform for automated inspection [4]. Notably, because transmission towers exhibit substantial morphological variability, holistic tower recognition is challenging; by contrast, the categories of their primary functional components and their combinatorial relationships are relatively consistent. Consequently, detecting key components to achieve tower recognition and condition assessment constitutes a more generalizable technical pathway [5].

In multi-sensor fusion localization frameworks, the visual detection module provides feature observations and component priors for geometric modeling, serving as a critical front end for achieving precise localization. In recent years, visual detection research for power-inspection scenarios has focused primarily on multi-scale feature fusion, small-object recognition, and robustness in complex backgrounds.

In terms of structural feature extraction, Li et al. [6] proposed a cross-scale spatial attention detector (CSSAdet) that effectively identifies subtle components within mechanical connectors of transmission lines. Tian et al. [7] improved you only look once version 3 (YOLOv3) by incorporating ResNet-50 and an attention mechanism, thereby enhancing insulator detection accuracy in complex backgrounds. Tan et al. [8] adopted oriented object detection coupled with deformable convolutions to achieve high-precision recognition of transmission towers in remote sensing imagery.

In instance segmentation and small object detection, Ma et al. [9] enhanced segmentation only look once version 2 (SOLOv2) by introducing a path aggregation path aggregation feature

pyramid network (PaFPN) and a mask intersection over union (MaskIoU) branch, thereby improving the segmentation accuracy of transmission lines and tower components. Cheng et al. [10] combined large and lightweight models and employed multimodal information fusion to strengthen collapsed-tower detection in few-shot conditions. Xu et al. [11] built on YOLOv7 with saliency enhancement and self-attention mechanisms, achieving higher robustness in multi-scale equipment detection.

Moreover, Zhao et al. [12] proposed a globally optimized feature pyramid network (GOFPN) that leverages multi-branch dilated convolutions and a graph reasoning mechanism to significantly enhance multi-scale feature representation in high-resolution images. Peterlevitz et al. [13] combined image-to-image translation with a mixed training strategy to effectively mitigate detection bias arising from the scarcity of aerial data. Liu et al. [14] employed a self supervised pretraining approach, tower masking masked image modeling (MIM), to improve the detection of transmission-line components in aerial imagery. The convolutional support vector machine (CSVM) network proposed by Bazi and Melgani offers a new perspective for UAV imagery detection in data-scarce conditions [15]. Wang et al. [16] introduced double prediction head (DPH)-YOLOv8, which adopts a dual-head prediction structure and coordinate attention to further optimize the accuracy and efficiency of small-object detection.

Overall, existing studies have made notable progress in multi-scale feature fusion and attention mechanism design, yet they still suffer from insufficient feature interaction and unstable confidence estimation, with pronounced performance degradation in complex backgrounds and occlusions [16]. Moreover, due to the diverse structural morphologies of transmission towers, holistic recognition remains challenging, whereas the categories and functions of their key compo-

nents are relatively consistent. Building on this insight, detecting these universal components and their compositional relationships can effectively bypass shape variability and enable more intrinsic tower recognition and condition assessment. To address the above issues, we developed a precise detection method for power-tower components based on the YOLOv11 framework, tailored to UAV autonomous localization and providing stable visual observations for subsequent perspective- n point (P n P)-based pose estimation. The method achieves multi-dimensional optimization through three structural improvements:

1) **C3k2-AMRB** We designed an adaptive multi-receptive-field fusion block that jointly models dilated convolutions and re-parameterization, enabling dynamic perception and efficient fusion of multi-scale features, which markedly improves the detection accuracy of large-size power-tower components.

2) **HWIU** We proposed a hierarchical wavelet interaction unit that leverages Haar wavelet decomposition and reconstruction to fuse high- and low-frequency features while preserving details, thereby enhancing robustness in complex backgrounds.

3) **DACR-Head** Based on statistical modeling of the predicted bounding-box distribution, this head reshapes confidence scores and down-weights low-quality boxes, significantly improving the stability and accuracy of small-object localization.

2 Improved Algorithm

To address the issues in UAV perspectives, such

as complex backgrounds, large scale variations, and mismatches between bounding box confidence and positioning accuracy, this paper conducts structural improvements [17]. Firstly, in the backbone network (gray), a C3k2-adaptive multi-receptive field block (C3k2-AMRB) is designed. This module integrates multi-dilated convolution with reparameterization, significantly expanding the receptive field and enhancing the multi-scale feature extraction capability. Secondly, in the feature fusion layer (neck, blue), a Haar wavelet-based feature interaction unit (hierarchical wavelet interaction unit, HWIU) is proposed. It realizes high-low frequency feature interaction through Haar wavelet decomposition and reconstruction, effectively alleviating the semantic mismatch problem in complex backgrounds and improving the quality of feature fusion. Finally, in the detection head (detect, green), a distribution-aware confidence refinement head (DACR-Head) is constructed. Based on the statistical features of prediction distribution, it adaptively optimizes confidence estimation, thereby improving the positioning accuracy and detection stability of small targets. The network structure of the enhanced algorithm is illustrated in Fig. 1.

2.1 C3k2-AMRB

The C3k2 module in you only look once (YOLO) series networks is designed with lightweight as the core objective, yet its single-scale convolution structure restricts the capability of global feature modeling [18]. To address this limitation, this paper embeds a AMRB based improved unit into the C3k2 framework, constructing the C3k2-

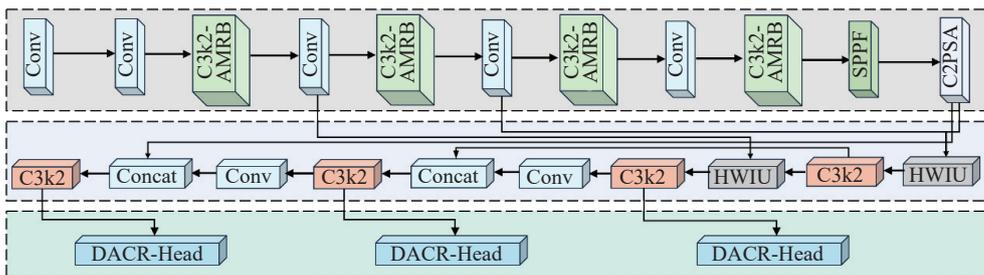


Fig. 1 The network of the proposed algorithm

AMRB module. Aiming to solve the problems of multi-scale target detection and fine-grained component recognition in power tower inspection tasks, the C3k2-AMRB inherits the C3k2 framework while replacing the internal computing units entirely with AMRB. The AMRB comprises three key components, each customized according to the characteristics of the power tower inspection scenario. Its architecture is illustrated in Fig. 2.

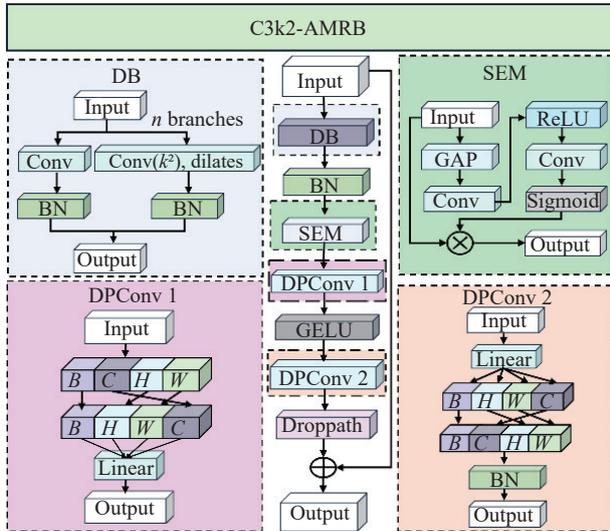


Fig. 2 Structures of C3k2-AMRB

We designed a dilated re-parameterization block (DB). In view of the large structural span of transmission towers and the pronounced scale variation among their components, this module adopts a dilated large-kernel convolution design. When the convolution kernel size ≥ 7 , the dilated re-parameterization block is employed. Through a multi-branch dilated convolution architecture, the training stage uses branches with different dilation rates to enlarge the receptive field, thereby capturing both the global tower structure and fine-grained components such as insulators and bolts.

We designed a scale enhancement module (SEM). Attention mechanisms emulate the human ability to selectively focus on informative cues; channel attention, in particular, strengthens responses on key feature channels. In power-tower inspection, background and target cues are

often interleaved across channels. Therefore, an SEM is designed. The module first applies global average pooling (GAP) to the feature map to compress spatial dimensions, and then uses a convolution followed by Sigmoid activation to generate channel weights. These weights are multiplied with the original feature map to adaptively amplify channels corresponding to towers and their components while suppressing irrelevant background channels, thereby improving the efficiency of fine-grained component feature extraction.

We designed an efficient dimension transformation (dynamic patch convolution 1, DPConv 1/DPConv 2). To balance efficiency and representational capacity, we designed an efficient feed-forward module with explicit layout conversion. Specifically, the feature tensor is converted from B, C, H, W to B, H, W, C to match the computation pattern of linear layers; a linear layer then expands the channel dimension and a Gaussian error linear unit (GELU) activation enhances nonlinearity, followed by another linear layer for dimensional reduction. The B, C, H, W represents batch, channel, height and width respectively. Finally, the layout is converted back to NCHW, and batch normalization (BN) is applied to stabilize training. This design preserves sufficient feature transformation while reducing computational overhead through appropriate layout/dimension adaptation for the power-tower inspection task.

The module employs a residual skip connection and applies stochastic depth (DropPath) regularization before the residual addition to promote cross-layer information flow and mitigate gradient vanishing, thereby improving model robustness in complex inspection scenarios.

2.2 Hierarchical Wavelet Interaction Unit

To achieve hierarchical interaction and fine-grained fusion of multi-scale features, we proposed the HWIU. The core idea is to leverage the wavelet transform (WT) to fuse high- and low-

frequency features across different scales, thereby meeting the feature-extraction requirements of power-tower inspection from global structure to fine-grained components. Its architecture is illustrated in Fig. 3. The execution workflow and structures are as follows.

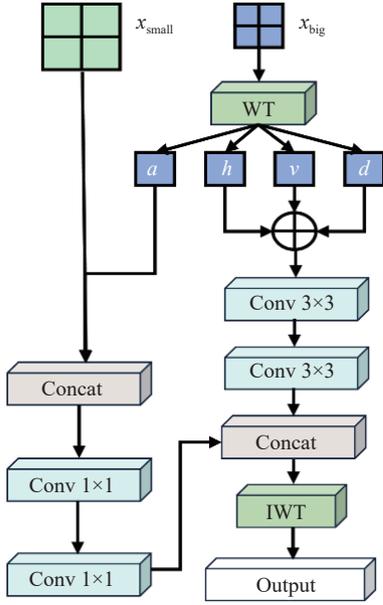


Fig. 3 Structures of HWIU

HWIU takes dual-scale features as input. It first applies a WT to the large-scale feature (x_{big}), using multi-directional wavelet kernels (horizontal, vertical, diagonal) to decompose it into the approximate subband a (low-frequency, LL) and the detail subbands h (horizontal detail, HL), v (vertical detail, LH), and d (diagonal detail, HH); here, a preserves low frequency structural information of the tower, while h , v , or d captures high-frequency cues such as component edges, achieving hierarchical decoupling of features. During the feature interaction stage, h , v , and d are fused and enhanced via residual blocks, where consecutive 3×3 convolutions strengthen cross-directional dependencies among detail features; meanwhile, a and the small-scale feature (x_{small}) are fed into a channel transformation module, in which 1×1 convolutions perform cross-scale, channel-level interaction and redundancy compression to produce a representation adapted for reconstruction. Finally, the enhanced detail

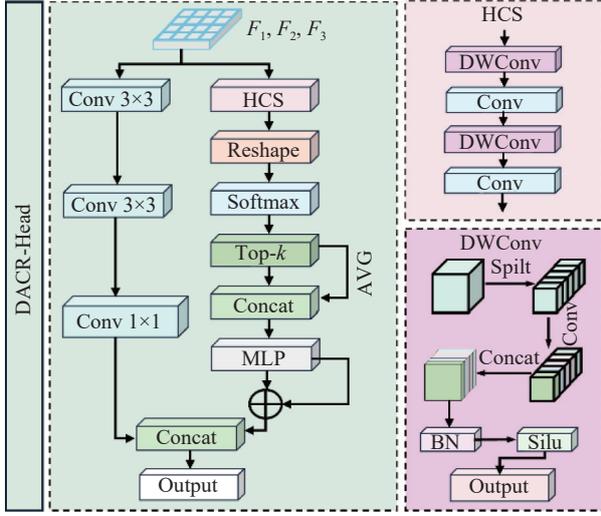
subbands are concatenated with the output of the channel transformation and passed to an inverse wavelet transform (IWT) module, which uses a transposed convolution operation to recombine the interacted features into a unified-scale output, thereby effectively integrating global low-frequency and local high-frequency information.

In transmission tower inspection, HWIU leverages the hierarchical feature separation afforded by wavelet decomposition to match the tower’s multiscale characteristics, covering the global structure as well as fine components. The interaction between residual blocks and the channel transformation module strengthens dependencies across scales and orientations, preserving global structural cues while emphasizing details of small components, thereby providing a representation with improved hierarchical discriminability for subsequent detection tasks.

2.3 Distribution Aware Confidence Refinement Head

To address the mismatch between bounding-box confidence and localization accuracy in transmission-tower inspection, this work proposes a DACR-Head that performs statistical analysis of bounding-box distributions to refine confidence estimates. The module comprises three core components, each tailored to distribution awareness and confidence optimization. The central idea is to exploit the probabilistic characteristics of bounding-box distributions to adjust the initial classification scores with fine granularity, thereby improving the reliability of the detection results. The processing pipeline is designed with the substantial scale variation of tower components and strong background clutter in mind, and the overall architecture is illustrated in Fig. 4.

Processing in the DACR-Head begins at the feature-input stage, where the input features are first processed by a hierarchical convolutional stream (HCS). The HCS adopts a hierarchical design that combines depthwise separable and



Note: AVG represents average, DWConv represents depthwise separable convolution, and Conv represents convolution.

Fig. 4 Structures of DACR-Head

standard convolutions, maintaining a lightweight profile while encoding multi-scale features and producing the initial classification scores. This stage provides the foundational features for subsequent confidence refinement, particularly accommodating the wide variation in component sizes in transmission towers, and ensures that the initial scores reflect the target category.

After obtaining the initial classification scores, the module proceeds to the distribution aware analysis stage. The predicted bounding box corner features are first reshaped into probability distributions for the four corners of the box, and a softmax activation is applied to compute probabilities at each location. To address the localization sensitivity of fine grained tower components, the module extracts the top- k probabilities from each corner distribution and computes their mean; these are concatenated to form a distributional statistical feature. This feature effectively captures the concentration of the predicted distributions. For components that are precisely localized, the probability mass is more concentrated and the gap between the top- k values and their mean is small, whereas for ambiguous or cluttered regions the distribution is more dispersed and the feature differences are more pronounced.

Based on the distributional statistics, the module applies a multilayer perceptron (MLP) to perform a nonlinear transformation and generate a quality score adjustment. This adjustment is added to the initial classification scores from the HCS branch on a per-element basis to obtain the refined confidences. The procedure jointly calibrates the classification scores and the distribution quality. For tower components with clear class identity but imprecise localization, the adjustment decreases the confidence; for small components with moderate classification confidence but precise localization, the adjustment increases the confidence, aligning the final confidence more closely with the actual localization accuracy. Finally, the refined confidences are concatenated with the bounding box coordinate distribution features after convolutional processing to produce the final output.

3 Experiment Analysis

3.1 Experimental Environment

1) Dataset

To evaluate the effectiveness of the improved model in power line inspection, we conducted experiments on the inspection of power line assets dataset (InsPLAD) [19, 20]. InsPLAD contains 10607 images collected by unmanned aerial vehicles (UAVs) during actual transmission line inspections and provides annotations for 17 categories of power line components, including towers, insulators, conductors, fittings, and connecting structures. Targets related to transmission towers exhibit pronounced multi-scale variation, large viewpoint changes, and susceptibility to background clutter, which makes detection challenging. The dataset faithfully reflects practical inspection conditions, such as complex backgrounds, wire occlusions, and illumination changes, and thus offers a solid basis for validating the performance of tower detection models.

2) Implementation Details

All experiments were implemented in

PyTorch and executed on a Windows system equipped with an NVIDIA GeForce real-time ray tracing (RTX) 3090 (24 GB) graphics processing unit (GPU) and the 12th Gen Intel Core i5-12600KF central processing unit (CPU). During training, we used an initial learning rate of 0.01, a batch size of 16, an input resolution of 640 pixel \times 640 pixel, and 300 epochs. Optimization was performed with stochastic gradient descent (SGD), and Mosaic data augmentation was enabled throughout training. In addition, to assess the feasibility of embedded deployment, we conducted integration tests in a UAV onboard computing environment. The results indicated that the key component detections produced by our method can be directly used in a perspective- n point (P n P) based pose solver, providing the visual input required for estimating the UAV's relative distance and pose with respect to transmission towers.

3.2 Evaluation Metrics

To objectively evaluate detection performance, we focused on detection-quality measures and reported component detection precision (P_{CD}), component detection recall (R_{CD}), component detection accuracy at intersection over union (IoU) = 0.5 (CDA_{50}), and component detection robustness across IoU thresholds ($CDR_{50:95}$); their computational definitions are provided as follows

$$P_{CD} = \frac{N_{TP}^{CD}}{N_{TP}^{CD} + N_{FP}^{CD}} \quad (1)$$

$$R_{CD} = \frac{N_{TP}^{CD}}{N_{TP}^{CD} + N_{FN}^{CD}} \quad (2)$$

$$CDA_{50} = \frac{1}{n} \sum_{i=1}^n ACDP_i \quad (3)$$

$$CDR_{50:95} = \frac{1}{10} \sum_{t=0.50}^{0.95} CDA_t \quad (4)$$

where N_{TP}^{CD} represents the number of power pole tower components correctly detected by the model, N_{FP}^{CD} represents the number of non-targets incorrectly detected as power pole tower components, N_{FN}^{CD} represents the number of actual power pole tower components that were not detected, $ACDP_i$ represents the average component detection precision for the i -th category, i represents the index of the i -th component category, n represents the total number of component categories, CDA_t represents the component detection accuracy at the IoU threshold t , and t represents the IoU threshold.

3.3 Ablation Study

To verify the effectiveness and synergistic effect of the proposed structural improvements, this study conducted seven groups of ablation experiments based on the YOLOv11 model using the InsPLAD power tower dataset. Taking YOLOv11 as the baseline, the C3k2-AMRB, HWIU, and DACR-Head were successively introduced to analyze the contribution and synergistic effect of each module on detection accuracy and stability. The experimental results are shown in [Tab. 1](#), and the evaluation metrics include P_{CD} , R_{CD} , CDA_{50} , and $CDR_{50:95}$.

Based on the baseline model, the introduction of C3k2-AMRB increased CDA_{50} from 84.6%

Tab. 1 Ablation results

Method	YOLOv11	C3k2-ARMB	HWIU	DACR-HEAD	P_{CD} (%)	R_{CD} (%)	CDA_{50} (%)	$CDR_{50:95}$ (%)
Group 1	√	—	—	—	85.9	83.8	84.6	65.2
Group 2	√	—	—	√	88.2	82.0	86.2	67.9
Group 3	√	—	√	—	89.3	83.8	86.2	68.6
Group 4	√	√	—	—	85.8	84.4	87.2	68.3
Group 5	√	√	—	√	85.9	83.8	86.6	69.1
Group 6	√	√	√	—	87.2	83.9	87.4	69.6
Group 7	√	√	—	√	90.4	83.4	87.5	69.1
Group 8	√	√	√	√	88.2	84.8	88.3	69.8

to 86.2% (+1.6%) and $CDR_{50:95}$ from 65.2% to 67.9% (+2.7%), indicating that the combination of multi-dilated convolutions and re-parameterization effectively expanded the receptive field and enhanced the model’s representation capability for multi-scale structures. After incorporating the HWIU, P_{cd} increased to 89.3% (+3.4%) and $CDR_{50:95}$ rose to 68.6% (+3.4%), demonstrating that the interaction between high- and low-frequency features improved semantic consistency in complex backgrounds. When the DACR-Head was introduced individually, $CDR_{50:95}$ increased to 67.9% (+2.7%), suggesting that the distribution-aware confidence optimization mechanism alleviated the mismatch between classification confidence and localization accuracy.

In the dual-module combination experiments, the synergistic effects among the modules were further demonstrated. The combination of C3k2-AMRB and HWIU increased CDA_{50} to 87.4% (+2.8%) and $CDR_{50:95}$ to 69.6% (+4.4%), reflecting the semantic complementarity between the backbone layer and the feature fusion layer. The combination of C3k2-AMRB and DACR-Head raised the R_{cd} to 84.4% (+0.6%), indicating the cooperation between spatial feature modeling and confidence optimization. For the combination of HWIU and DACR-Head, $CDR_{50:95}$ reached 69.1% (+3.9%), suggesting that the fusion of high- and low-frequency features provided more discriminative semantic support for confidence estimation. When all three modules were applied simultaneously, the model achieved the best overall performance, with CDA_{50} reaching 88.3% (+3.7%) and $CDR_{50:95}$ reaching 69.8% (+4.6%). The continuous improvement in these metrics indicates that the multi-level enhancements—from feature extraction and feature fusion to confidence calibration—formed a comprehensive optimization pipeline. This systematic refinement strengthened the model in terms of feature diversity, semantic consistency, and confidence stability, thereby providing a solid

foundation for achieving high-precision detection and stable localization of UAVs in complex environments.

3.4 Experimental Results and Analysis

This section presents the comparative experimental results on the InsPLAD power transmission line component detection dataset. Fig. 5 illustrates the detection performance of the original YOLOv11 and the improved algorithm across twelve representative samples. Overall, the improved algorithm outperforms the baseline model in terms of target completeness, boundary precision, and detection stability in complex backgrounds.

In the experiments of group 2, group 3, and group 12, the improved algorithm successfully detected components that the original YOLOv11 model failed to recognize. In group 2, the bolt connectors were completely missed by the baseline model, whereas the improved algorithm accurately identified all targets. In group 3, the detection of fittings partially occluded by conductors was more complete, and in group 12, the small upper connection components were also successfully recognized. These results demonstrated that the model’s capability to detect small objects and complex structures was enhanced through multi-scale feature fusion. In the experiments of group 1, group 4, group 6, and group 10, boxes exhibited poor alignment with object contours and contained redundant regions whereas the improved algorithm achieved complete coverage and more conditions, and the model maintained stable detection performance.

Overall, the improved algorithm outperformed the original YOLOv11 model in terms of boundary precision, target coverage completeness, and small-object recognition. According to the quantitative results presented earlier, the improved algorithm achieves increases of 4.4% and 7.0% in CDA_{50} and $CDR_{50:95}$, respectively, with the P_{cd} rising to 90.7%. By incorporating the multi-scale feature expansion of C3k2-AMRB, the high- and low-frequency feature

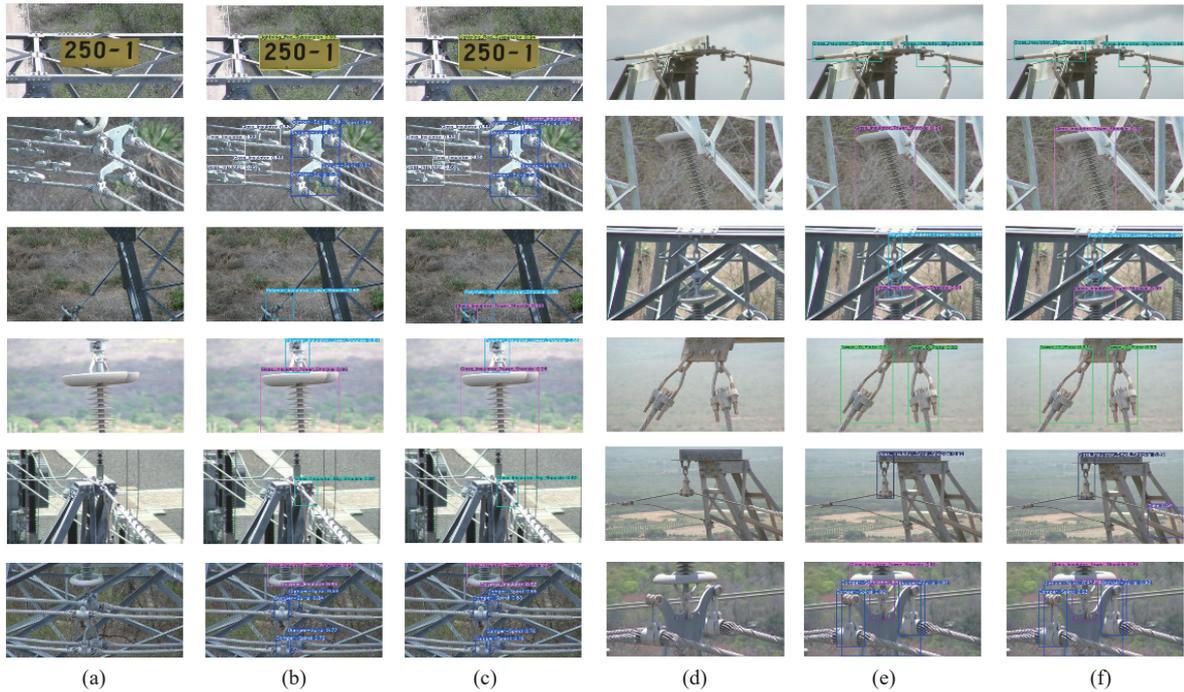


Fig. 5 The training results of the proposed algorithm model: (a), (d) original image; (b), (e) YOLOv11 detection result; (c), (f) improved detection result

fusions of HWIU, and the confidence optimization of DACR-Head, the model achieves enhanced performance in multi-scale, occlusion, and complex background conditions. The improved model demonstrates superior performance in detection completeness, boundary precision, and classification stability, validating the effectiveness and robustness of the proposed algorithmic design.

3.5 Comparison Experiments

This paper proposes an innovative algorithm based on the YOLOv11 framework, aiming to improve the accuracy and computational efficiency of power transmission tower detection. To verify the effectiveness of the proposed method, rigorous experimental evaluations were conducted, and the improved model was compared with several popular object detection algorithms in this field. All experiments followed the principle of controlled variables to ensure fairness, maintaining consistent hardware and software environments to achieve unbiased evaluation. The experiments were conducted on the InsPLAD power transmission tower dataset,

which contains multiple tower components characterized by multi-scale variations, complex backgrounds, and occlusion challenges. In terms of the CDA_{50} and $CDR_{50:95}$ metrics, the improved algorithm demonstrated significant performance gains. Compared with baseline models such as YOLOv3-Tiny, YOLOv5n, YOLOv6n, YOLOv7-Tiny, YOLOv8n, YOLOv9, YOLOv10n, and real-time detection Transformer-large (RT-DETR-L), the proposed method achieved improvements of 5.1%, 1.0%, 1.9%, 0.7%, 0.3%, 1.0%, 1.6%, and 3.9%, respectively, as shown in Tab. 2. In particular, the improved model outperformed the baseline YOLOv11n by 3.7% in CDA_{50} and 4.6% in $CDR_{50:95}$, demonstrating the superiority of the proposed approach in complex power tower detection tasks.

Tab. 2 also compares the complexity and efficiency of different models while achieving high accuracy. Specifically, the parameter count is only 2.5×10^6 , and the computational cost is as low as 6.4×10^9 , which is comparable to the most lightweight models (e.g., YOLOv5n). Furthermore, an inference latency of 1.0 ms demon-

Tab. 2 Comparison of detection results of different algorithms on the InsPLAD Dataset

Model	P_{cd} (%)	R_{cd} (%)	CDA_{50} (%)	$CDR_{90.95}$ (%)	Parameter (10^6)	GFLOP	Size (MB)	Inference (ms)
YOLOv3	89.9	78.5	83.2	64.1	12.1	18.9	46.5	1.7
YOLOv5n	89.8	83.7	87.6	69.5	2.5	7.1	5.0	1.0
YOLOv6n	85.7	84.2	86.4	68.1	4.2	11.8	8.3	0.9
YOLOv7	91.5	83.3	87.6	67.7	6.0	13.2	11.7	1.4
YOLOv8n	88.2	84.9	88.0	70.2	3.0	8.1	6.0	1.1
YOLOv8s	89.8	82.3	87.5	70.5	11.3	28.5	21.4	2.5
YOLOv9	88.9	81.9	87.3	67.3	2.6	10.7	5.8	2.2
YOLOv10n	83.7	81.3	86.7	67.9	2.7	8.3	5.5	1.3
RT-DETR-L	85.9	83.8	84.4	68.4	32.0	103.5	63.1	17.0
Ours	88.2	84.8	88.3	69.8	2.5	6.4	5.3	1.0

Note: GFLOP represents Giga floating-point operation.

strates its capability to meet the real-time requirements of UAV platforms. Overall, our model achieves the best balance between accuracy and efficiency.

4 Conclusion

This study has systematically constructed and validated a precision detection scheme for key components of power transmission towers, tailored for UAV autonomous localization. The scheme introduces three core improvements within the YOLOv11 framework, and experimental results confirm its effectiveness in addressing the fundamental challenges prevalent in power tower inspection, such as multi-scale variations, complex background interference, and unstable small-target localization. Specifically, the C3k2-AMRB module expands the receptive field through multi-dilated convolutions and re-parameterization, enhancing multi-scale target representation capability; the HWIU module mitigates semantic mismatch in complex backgrounds through high- and low-frequency feature fusion, significantly improving the model’s overall robustness; the DACR-Head module adaptively optimizes confidence based on the distribution characteristics of bounding boxes, thereby markedly enhancing the localization accuracy and detection stability of small targets. Collectively, these modules form a comprehensive optimiza-

tion pipeline, whose final output of high-precision and highly stable detection results provides crucial and reliable visual input for subsequent PnP -based pose estimation. This research demonstrates that the proposed method not only significantly improves various performance metrics for component detection but also, more importantly, offers a solid and viable technical pathway to overcome the challenge of global positioning system (GPS) positioning drift for UAVs during power inspections, enabling stable and autonomous close-range localization.

Despite the significant achievements, there remains room for further improvement. Future research can be carried out in two main directions. First, component detection accuracy will be enhanced by focusing on particularly challenging scenarios, such as improving the recognition robustness for small components of severe occlusion or in extreme scales. This will involve exploring advanced attention mechanisms and targeted data augmentation strategies to strengthen model generalization in these complex conditions. Second, future work will integrate millimeter-wave radar and inertial navigation data, employing radar-vision fusion and unsupervised learning strategies to further enhance the UAV’s autonomous localization accuracy and improve the multi-sensor fusion perception system for power inspection in complex environments.

References:

- [1] L. Zuo, Y. Yan, J. Wang, X. Sang, and Y. Wang, "Detection of UAV target based on continuous radon transform and matched filtering process for passive bistatic radar," *Journal of Beijing Institute of Technology*, vol. 33, no. 1, pp. 9-18, 2024.
- [2] X. Zhang, Y. Zhang, K. Shen, Q. Fu, and H. Shen, "FAFNet: An overhead transmission line component detection method based on feature alignment and fusion," *IEEE Sensors Journal*, vol. 25, no. 15, pp. 30197-30206, 2025.
- [3] Y. Zhang, C. Wu, W. Guo, T. Zhang, and W. Li, "CFANet: Efficient detection of UAV image based on cross-layer feature aggregation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1-11, 2023.
- [4] Y. Zhang, C. Wu, T. Zhang, and Y. Zheng, "Full-scale feature aggregation and grouping feature reconstruction-based UAV image target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1-11, 2024.
- [5] Y. Fu, J. Meng, and H. Li, "Unmanned aerial vehicles and edge AI for power line inspection," in *2025 IEEE International Conference on Mechatronics and Automation (ICMA)*, Beijing, China, pp. 801-808, 2025.
- [6] Y. Li, M. Liu, Z. Li, and X. Jiang, "CSSAdet: Real-time end-to-end small object detection for power transmission line inspection," *IEEE Transactions on Power Delivery*, vol. 38, no. 6, pp. 4432-4442, 2023.
- [7] X. Tian, M. Zhang, and G. Lu, "Power line insulator defect detection using CNN with dense connectivity and efficient attention mechanism," *Multi-media Tools and Applications*, vol. 83, no. 10, pp. 28305-28322, 2024.
- [8] Y. Tan, F. Jiao, W. Mo, H. Liu, X. Bai, and J. Ma, "Detection in optical remote sensing images of transmission tower based on oriented object detection," *CSEE Journal of Power and Energy Systems*, vol. 11, no. 1, pp. 217-226, 2025.
- [9] W. Ma, J. Xiao, G. Zhu, J. Wang, D. Zhang, X. Fang, and Q. Miao, "Transmission tower and power line detection based on improved Solov2," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1-11, 2024.
- [10] H. Cheng, Y. Gu, M. Xi, Q. Zhong, and W. Liu, "A few-shot collapsed transmission tower detection method combining large and small models in remote sensing image," *IEEE Access*, vol. 13, pp. 41670-41681, 2025.
- [11] Y. Xu, C. Song, Y. Sun, and S. Yu, "Improved detection of transmission tower equipment using YOLOv7 with saliency data augmentation," in *2024 5th International Conference on Clean Energy and Electric Power Engineering (ICCEPE)*, Yangzhou, China, pp. 254-260, 2024.
- [12] X. Zhao, H. Zhang, C. Song, H. Li, and H. Guo, "Multi-scale component detection in high-resolution power transmission tower images," *IEEE Transactions on Power Delivery*, vol. 40, no. 4, pp. 2391-2401, 2025.
- [13] A. J. Peterlevitz, M. A. Chinelatto, A. G. Menezes, C. A. M. Motta, G. A. B. Pereira, and G. L. Lopes, "Sim-to-real transfer for object detection in aerial inspections of transmission towers," *IEEE Access*, vol. 11, pp. 110312-110327, 2023.
- [14] F. Liu, L. Yao, C. Zhang, T. Wu, X. Zhang, X. Jiang, and J. Zhou, "Boost UAV-based object detection via scale-invariant feature disentanglement and adversarial learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1-13, 2025.
- [15] Y. Bazi and F. Melgani, "Convolutional SVM networks for object detection in UAV imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3107-3118, 2018.
- [16] J. Wang, X. Li, J. Chen, L. Zhou, L. Guo, Z. He, H. Zhou, and Z. Zhang, "DPH-YOLOv8: Improved YOLOv8 based on double prediction heads for the UAV image object detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1-15, 2024.
- [17] A. Munir, A. J. Siddiqui, M. S. Hossain, and A. El-Maleh, "YOLO-RAW: Advancing UAV detection with robustness to adverse weather conditions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 6, pp. 7857-7873, 2025.
- [18] H. Wang, C. Wang, and Q. Fu, "MINIAOD: Lightweight aerial image object detection," *IEEE Sensors Journal*, vol. 25, no. 5, pp. 9167-9184, 2025.
- [19] S. Yang, B. Hu, B. Zhou, F. Liu, X. Wu, X. Zhang, J. Gu, and J. Zhou, "Power line aerial image restoration under adverse weather: Datasets and baselines," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 10105-10119, 2025.
- [20] D. Wu, W. Yang, J. Li, K. Du, L. Li, and Z. Yang, "CRL-YOLO: A comprehensive recalibration and lightweight detection model for aav power line inspections," *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1-21, 2025.



Luqi Zhang received the B.S. degree from Hebei University of Water Resources and Electric Power, Cangzhou, China, in 2024. He is currently pursuing the master's degree with School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang, China. His research interests include image

processing and object detection.



Yunzuo Zhang received the B.S. degree from School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, China in 2007, the M.S. degree from School of Information and Communication, Guilin University of Electronic Science and Technology, Guilin, China, in 2011 and the Ph.D. degree from School of Information and Electronics, Beijing Institute of Technology, Beijing, China in 2016. In 2018, he was a visiting scholar in California State University, California, United States. He is a Professor with School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang, China. He is currently a senior member of China Computer Federation (CCF) and a member of Chinese-American Engineers and Scientists Association of Southern California (CESASC). His research interests include image processing, video intelligent analysis and big data processing.



Song Tang is a senior engineer of Institute of Applied Mathematics, Hebei Academy of Sciences, Shijiazhuang, China and a doctoral candidate of Tianjin University, Tianjin, China. He is mainly responsible for the research and application of privacy-preserving computation and artificial intelligence algorithms. To date, he has published more than 10 academic papers as the first

author and obtained 5 invention patents.



Wei He received his Ph.D. degree in mechanical engineering from Xinjiang University, Urumqi, China, in 2023. He has been a research associate of Institute of Applied Mathematics at Hebei Academy of Sciences, Shijiazhuang, China since 2024. His current research interests include visual servoing, aerial robotics and robot control.



Tianliang Zhang received his Ph.D. degree in agricultural mechanization engineering from China Agricultural University, Beijing, China, in 2022. He has been an assistant researcher of Institute of Applied Mathematics at Hebei Academy of Sciences, Shijiazhuang, China since 2023. His current research interests include image and radar processing technology, 3D reconstruction and drone application technology.



Yubo Hu received the B.S. degree from Hebei Agricultural University, Baoding, China, in 2024. He is currently studying for a master's degree at School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang, China. His research interests include image processing and lane detection.