Membrane Technology
ISSN (online): 1873-4049

An Improved YOLOv8 Network Architecture for Enhancing Detection Accuracy of Bird Nests with Small Size and Intricate Background on Power Transmission Lines

Dan Yang[1], Hailong Yang[1,*], Tao Yang[1,*], Shengjian Zeng[1,*], Xiaobo Wang[1], Wei Liu[1], Liuliang Zhao[1], Shuang Jiang[1] and Yulong Yang[2]

[1]State Grid Guangyuan Power Supply Company, Guangyuan, China

[2]School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China

**Corresponding Author*

Abstract:
The harm caused by birds building nests on power transmission lines is enormous, greatly increasing the possibility of short circuits and damages to the transmission lines. Bird nest detection on transmission lines presents significant challenges due to the small size of nests and complex backgrounds. To address these challenges, this paper proposes an improved YOLOv8 model specifically designed for bird nest detection. The model introduces several key innovations: the inclusion of an exponential moving average mechanism in the C2f module to enhance feature extraction and improve detection performance; the development of a dynamic weighted feature fusion network to more effectively integrate contextual features, significantly reducing both parameter count and computational load, thus achieving network lightweighting; and the introduction of weight-CIOU as a bounding box loss function, which better measures the similarity between targets, accelerates convergence, and enhances detection accuracy. Experimental results demonstrate that the improved YOLOv8 model outperforms the original YOLOv8 by achieving a mAP of 96.3%. Visualization of detection results further confirms the model's robustness in various challenging scenarios, highlighting its potential for real-world deployment in power line monitoring systems.

Keywords: network architecture; YOLOv8; bird nest; transmission lines

INTRODUCTION

The inspection of foreign objects on power transmission lines is one of the primary tasks in power line patrols. Among these foreign objects, bird nests on power towers are of particular concern, as short circuits caused by bird nests can have severe impacts on the power industry and disrupt daily life [1]. Traditional inspection methods involve manually identifying whether bird nests are present on power lines through visual inspection of images, a process that is not only time-consuming but also prone to errors, especially when dealing with large volumes of data.

The use of intelligent automated detection for bird nests is now the mainstream development trend in power line inspections. The intelligent detection of bird nests has undergone several stages of development [2-4]. However, these methods are labour and resource intensive [5]. Intelligent bird nest detection now falls under the broader problem of object detection, which requires the application of object detection algorithms. Region Proposal-based R-CNN series algorithms such as R-CNN, Fast R-CNN, and Faster R-CNN, are two-stage approaches [6,7]. The one-stage algorithms like YOLO and SSD, which use a single CNN network to directly predict the classes and locations of different objects [8,9].

As computer vision technology continues to advance, object detection is increasingly being applied across various aspects of daily life, facilitating the identification of bird nests on transmission line towers. To improve detection accuracy, one approach integrates AKConv within YOLOv8, designing the C2f-BE module to enhance the algorithm's feature processing capabilities [10]. Another method replaces standard convolution layers with depthwise separable convolutions, inverted residuals, and linear bottleneck structures to improve the extraction of small object features, although this increases model complexity [11]. Additional approaches enhance multi-scale structural feature maps by adding extra context-semantic enhancement modules, thereby improving small object recognition, albeit with a higher false positive rate for similar targets [12-14]. While these methods somewhat improve small object detection performance, they also inadvertently amplify noise, resulting in only marginal performance gains. Another approach proposes an edge-guided context module that extracts multi-scale spatial edge information at a finer granularity, improving small object localization [15]. The insufficient utilization of multi-scale features in existing networks is addressed by designing a feature fusion method that comprehensively extracts and utilizes features [16]. Additionally, a multi-scale feature optimization network is proposed, utilizing a feature optimization fusion module, a multi-scale local feature aggregation module, and a feature enhancement module to mitigate issues related to varying target scales,

though the detection performance for small objects remains suboptimal [17]. These methods attempt to improve small object detection performance through multi-scale feature fusion; however, treating all input features equally during the fusion process leads to stacked redundant features and noise as the network deepens, hindering further performance improvements based on multi-scale learning. Finally, the SPM attention based on CA is introduced to learn detailed information at locations of interest [18]. Another study incorporates attention into the backbone network, constructing an efficient layer attention aggregation structure to enhance feature extraction [19]. Furthermore, the integration of attention mechanisms with the detection head from the perspectives of scale, space, and task awareness improves the detection head's representational capacity [20]. These researchers have improved the model's focus on small objects using attention mechanisms, but the computational cost remains high when processing high-resolution images, limiting their application in bird nest detection.

Therefore, YOLO still has some limitations. Specifically, it exhibits lower detection accuracy for small objects and is prone to false positives and missed detections in complex backgrounds. To overcome these issues, this paper proposes an improved YOLOv8 network architecture designed to better meet the practical needs of bird nest detection. The proposed enhancements include the introduction of the exponential moving average (EMA) mechanism into the C2f module to strengthen feature extraction and improve detection performance, particularly for small objects. Additionally, a dynamic weighted feature fusion network (DWFFN) is introduced to integrate contextual features, reducing both the parameter count and computational load, thereby achieving network lightweighting more effectively. Furthermore, a weight-CIOU (W-CIOU) loss function is employed to measure the similarity more accurately between bounding boxes, accelerating convergence and enhancing detection accuracy. These improvements aim to address the challenges faced by the original YOLOv8 model, making it more suitable for the specific requirements of bird nest detection in complex environments.

MATERIALS AND METHODS

2.1 Improvement of YOLOv8 Network

The YOLOv8 is a state-of-the-art (SOTA) object detection model characterized by its Decoupled Head and Anchor-Free strategies. The Decoupled Head removes the traditional objectness branch, retaining only the classification and regression branches, thereby simplifying the model structure. The Anchor-Free strategy allows YOLOv8 to operate without anchor points, directly predicting the center position of objects rather than relying on present anchor boxes. This not only reduces redundant bounding boxes but also accelerates the non-maximum suppression process, improving both detection speed and accuracy.

Additionally, YOLOv8 adopts the ELAN backbone and Neck design from YOLOv7, incorporating the C2f structure and carefully adjusting the number of channels to enhance detection performance. For loss calculation, it combines VFLLoss, CIOULoss, and DFLLoss to optimize classification and bounding box regression. In terms of data augmentation, YOLOv8 borrows from YOLOX, disabling Mosaic augmentation in the later stages of training to allow the model to focus more on learning target features. These improvements enable YOLOv8 to excel in object detection tasks.

The final Head section employs multiple parallel detection heads, each corresponding to a feature map of different scales, ensuring that the model can simultaneously detect and output the positions, classes, and confidence scores of objects of varying sizes within the image. By leveraging feature maps at multiple scales, the model captures diverse features ranging from high-resolution details to low-resolution contextual information, enabling comprehensive detection of objects of different sizes.

This paper introduces an optimized YOLOv8 network architecture, as illustrated in Figure 1, with improvements focused on network lightweighting, feature enhancement, and detection accuracy. The aim is to better meet the practical needs of bird nest detection.

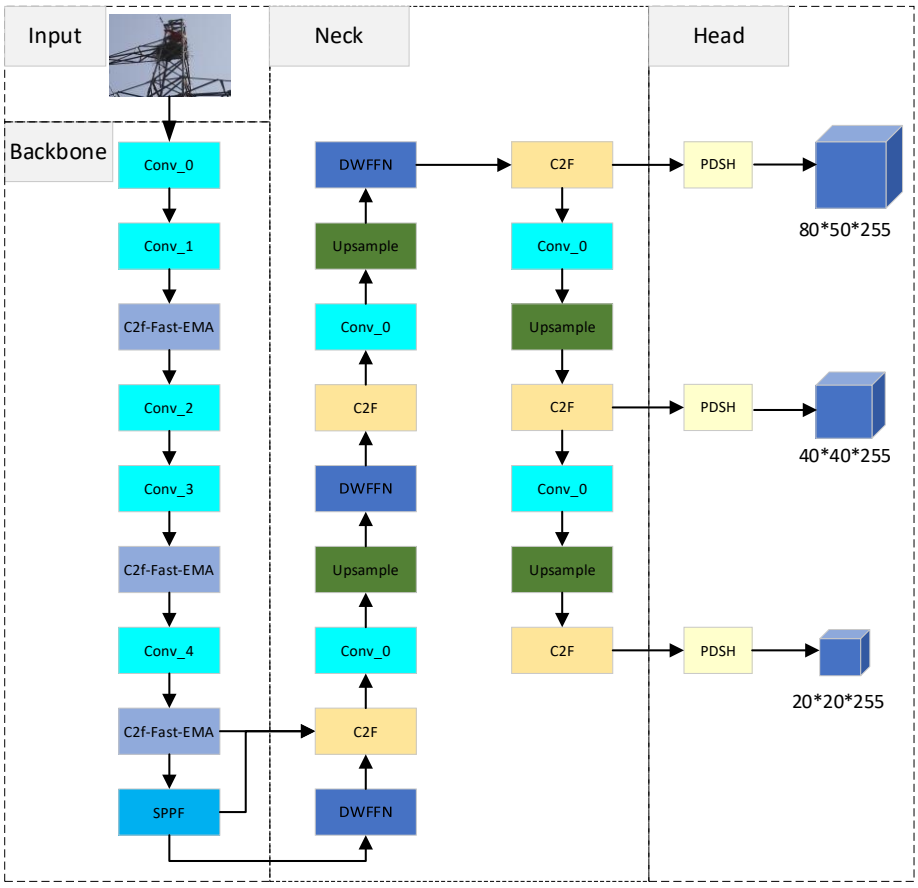

Figure 1. Improved YOLOv8 network structure

2.2 Improvement of Backbone Network

The backbone network of YOLOv8 performs well in feature extraction, but its convolution operations are computationally expensive, particularly during shallow feature extraction, which can lead to redundant computations and resource wastage. Additionally, the original feature extraction module may struggle to adequately capture critical information when dealing with complex backgrounds and small objects, which can impact detection performance. Therefore, this paper proposes an improved C2f-EMA module that combines depthwise separable convolution (DSConv) with the exponential moving average (EMA) attention mechanism to enhance the efficiency and accuracy of feature extraction.

In the backbone network of YOLOv8, the C2f module (cross stage partial networks with fusion) is responsible for the preliminary extraction of features from the input image. The original C2f module uses standard convolution operations to process all channels. Although this design can achieve good feature representation, it is computationally expensive, particularly when extracting shallow features, where redundant computations can occur. Therefore, this paper introduces the DSConv and EMA attention mechanism into the C2f module, forming the C2f-EMA module, which replaces the original C2f module.

The basic idea of DSConv is to decompose the standard convolution operation into two independent steps. Each input channel is convolved with a separate convolution kernel. This means that if the input feature map has C channels, there will be C convolution kernels, each processing only its corresponding single channel. The output feature map still has C channels, but each channel contains only the output of one convolution kernel. Through a 1×1 convolution operation, the output channels of each depthwise convolution are linearly combined to finally obtain the required number of output channels. This step effectively performs a linear combination of the channels at each pixel, achieving inter-channel information fusion. In the improved C2f-EMA module, we replaced the traditional full convolution operation with DSConv, significantly reducing computational overhead while maintaining the original feature extraction capability, as shown in Figure 2. The calculation formula for DSConv is as follows:

$$DepthwiseConv(X)[:,:,c\ i\ ] = K_{ci}\ \ * X[:,:,c\ i\ ] \qquad (1)$$

where K_{ci} represents the depthwise convolution kernel, and * represents the convolution operation.

$$PointwiseConv(X)[:,:,j] = \sum_{i=1} C_{in} W\_(i,j) \cdot DepthwiseConv(X)[:,:,i] \quad (2)$$

where $W_{i,j}$ is the pointwise convolution weight, and the final output feature map size is $H \times W \times C_{out}$. Compared to traditional convolution operations, depthwise separable convolution significantly reduces the computational load. The computational complexity of standard convolution is $H \times W \times C_{in} \times C_{out} \times k \times k$, while that of depthwise separable convolution is $H \times W \times (C_{in} \times k \times k + C_{in} \times Cout)$, reducing computational complexity and making the model more lightweight and suitable for running on resource-constrained devices.

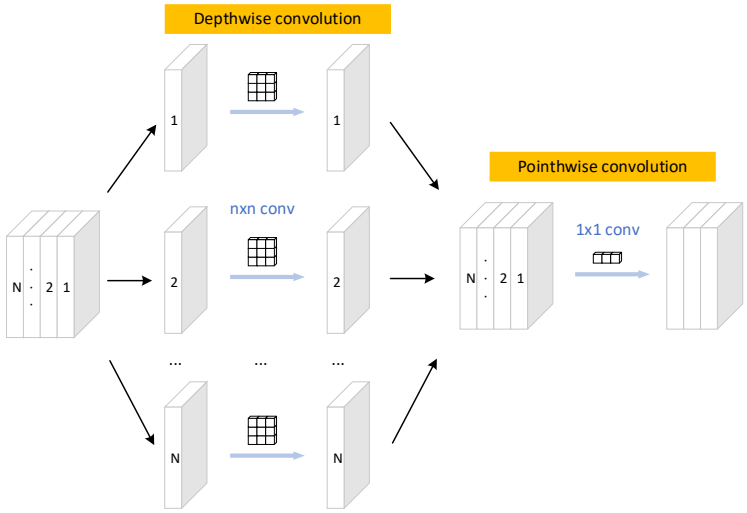

Figure 2. Depthwise separable convolutional network

The EMA attention mechanism is a module designed to enhance the focus of the feature extraction module on critical information, as illustrated in Figure 3. It encodes multi-scale features to generate cross-channel spatial attention maps, thereby highlighting important features while suppressing background noise. Specifically, the EMA module consists of two parallel branches: one branch captures multi-scale features using a 3x3 convolution, while the other branch encodes channel features through global average pooling. Both branches share a 1x1 convolution, and the attention map is generated using the Sigmoid function, achieving spatial and channel interaction to enhance the model's sensitivity to targets. The calculation formula for the EMA attention mechanism is as follows:

$$EMA(X) = \sigma(W_1 * GAP(X) + W_2 * Conv(X)) \quad (3)$$

where GAP(X) represents the global average pooling operation, Conv(X) represents the 3x3 convolution operation, σ is the Sigmoid activation function, and W_1 and W_2 are the corresponding weight matrices. Through the EMA attention mechanism, the network can emphasize important feature regions, particularly those containing small targets, thereby improving detection accuracy.

By introducing the C2f-EMA module, the network not only improves computational efficiency during the feature extraction stage but also enhances the ability to extract critical information in scenarios involving small targets and complex backgrounds.

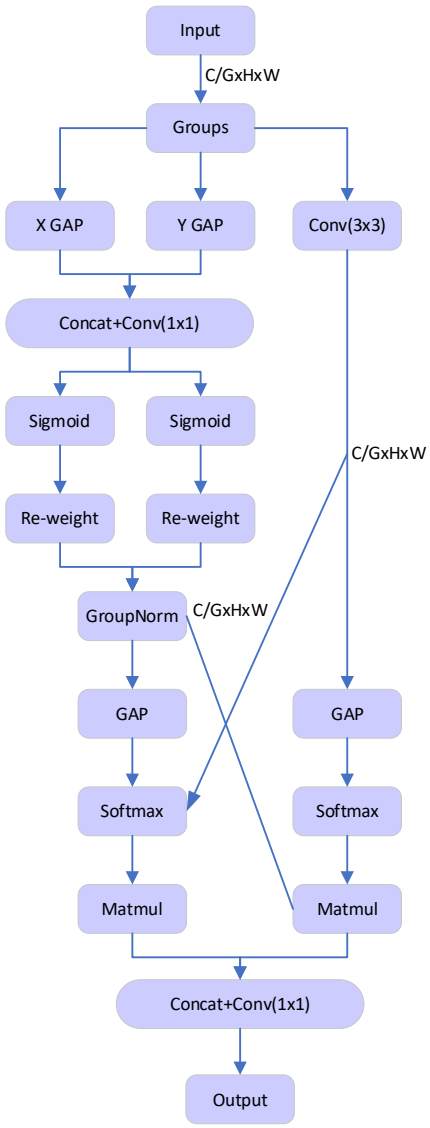

Figure 3. EMA module structure diagram

2.3 Improvement of Feature Fusion Module

The original feature fusion module (Neck) in YOLOv8 utilizes the traditional feature pyramid network (FPN) and path aggregation network (PANet) to achieve feature fusion across different scales. However, this structure may experience issues with feature information loss or insufficient fusion when processing images with complex backgrounds and small objects. In the context of bird nest detection, the scale variation of targets and background complexity demands more from feature fusion. To address these challenges, this paper proposes a new multi-scale feature fusion method named dynamic weighted feature fusion network (DWFFN), aimed at more efficiently fusing multi-scale features and improving the accuracy of small object detection, as shown in Figure 4.

In the DWFFN module, we propose a dynamic weighted upsampling method to replace the traditional fixed upsampling approach. Dynamic weighted upsampling (DWU) introduces dynamic weights and adaptive offsets, allowing the upsampling process to adjust the position and scope of sampling points dynamically based on the content and context of the input features. This approach not only enhances the resolution of the feature map but also better preserves the feature information of small objects, avoiding information loss during the upsampling process. The upsampling process of DWU can be represented by the following formula:

$$DWU(X) = W_d \cdot UpSample(X) + b \quad (4)$$

where UpSample(X) represents the traditional upsampling operation, W_d is the dynamic weighting matrix, and b is the bias term. Through this weighted operation, DWU can dynamically adjust the sampling results based on the content of the input features, effectively enhancing the features of small objects, especially in complex backgrounds.

DWFFN also introduces a bidirectional feature flow mechanism(BFF), which fuses shallow and deep features through bidirectional paths. During the feature fusion process, DWFFN can adaptively adjust the weights of features at each level, ensuring that features of different scales are fully expressed in the final feature map. This design not only enhances the network's ability to detect small objects but also improves the overall quality of the feature map, enabling the model to recognize targets more accurately in complex backgrounds. The process of bidirectional feature flow can be expressed as follows:

$$F_{iout} = Concat(F_{iup}, F_{idown}) \quad (5)$$

where F_{iup} represents the features transmitted from lower levels through the upward path, F_{idown} represents the features transmitted downward from higher levels, and Concat represents the concatenation of features. Through this bidirectional feature flow mechanism, DWFFN can better fuse features of different scales, enhancing the detection capability for small objects.

The DWFFN module achieves more flexible and efficient multi-scale feature fusion through the design of dynamic weighted up-sampling and bidirectional feature flow. Compared to the original PANet structure, DWFFN can better handle the feature fusion challenges in scenarios involving small objects and complex backgrounds, significantly improving the model's detection accuracy and recall rate. This improvement is particularly effective in reducing missed detections and false positives in bird nest detection.

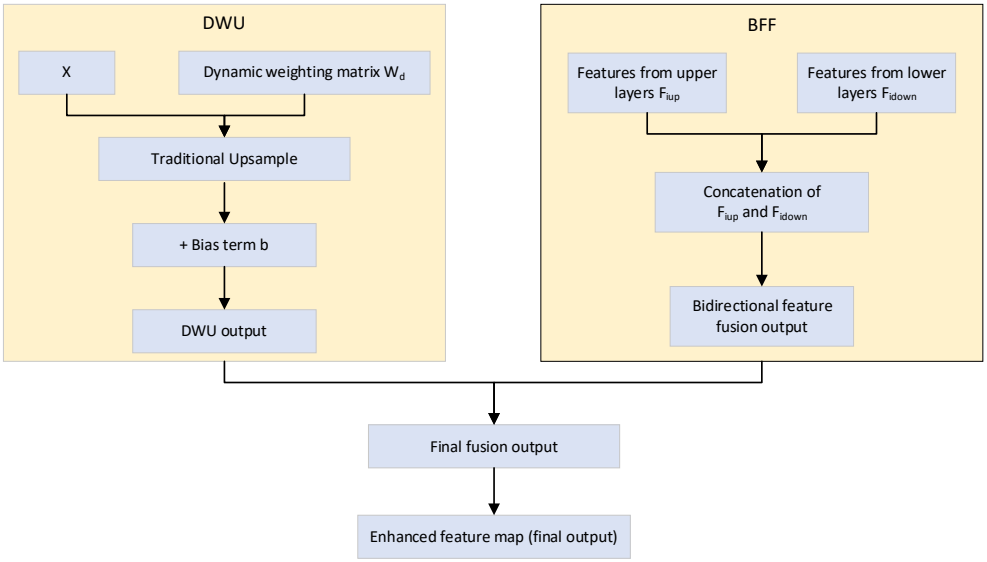

Figure 4. Structure of DWFFN module

2.4 Improvement of Detection Head

The detection head of YOLOv8 is primarily responsible for object classification and localization, and its design directly impacts the overall performance of the model. The traditional YOLOv8 detection head can encounter parameter redundancy when handling small objects, leading to increased model complexity and affecting real-time performance. In small object detection, particularly in resource-constrained environments such as embedded systems or real-time monitoring applications, it is crucial to reduce model complexity and computational resource requirements. Therefore, this paper proposes a parameter-sharing detection head (PSDH) to reduce model complexity while maintaining high detection performance, as shown in Figure 5.

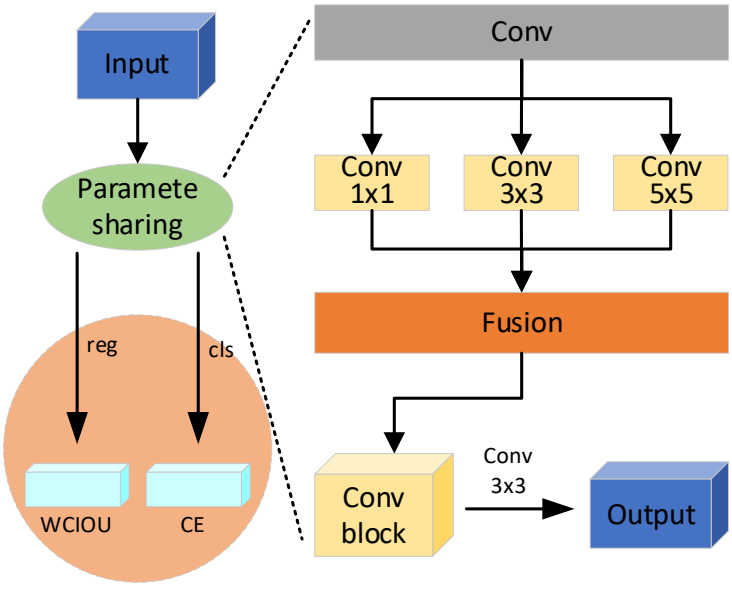

Figure 5. Structure of PSDH module

The core idea of PSDH is to share parameters between the regression and classification tasks within the detection head, thereby reducing the number of parameters in the model. Specifically, PSDH shares parameters in certain convolutional layers (such as the first 3x3 convolutional layer), allowing both the regression and classification tasks to make different predictions within the same feature space. This approach not only reduces the number of parameters but also improves computational efficiency. The computation process of PSDH can be expressed as follows:

$$F_{cls} = W_{cls} * F_{shared} + b_{cls} \quad (6)$$

$$F_{reg} = W_{reg} * F_{shared} + b_{reg} \quad (7)$$

where F_{shared} represents the shared feature representation, W_{cls} and W_{reg} the weight matrices for the classification and regression tasks, respectively, and b_{cls} and b_{reg} are the corresponding bias terms. By sharing parameters in certain convolutional layers, PSDH reduces computational overhead while enhancing parallel processing.

To further improve the model's real-time performance, PSDH incorporates a series of lightweight design measures. The number of 3x3 convolutional filters in the original detection head is reduced from 64 to 32, and the size of some 1x1 convolutional filters is decreased from 256 to 128. By reducing the number and size of convolutional filters, the computational graph is optimized, allowing the model to complete predictions more quickly. In small object detection tasks, this lightweight design not only maintains detection accuracy but also significantly enhances the model's response speed, making it more suitable for running on embedded devices or low-power equipment.

The PSDH module significantly reduces the number of parameters and computational overhead through parameter sharing and lightweight design. While maintaining detection accuracy, PSDH enhances the model's real-time performance, making it particularly suitable for small object detection tasks such as bird nest detection, where quick response is essential.

2.5 Improvement of Loss Function

In object detection models, the loss function is a critical component, and its design directly impacts the model's detection performance. An effective bounding box loss function can significantly enhance model performance. The Intersection over Union (IoU) loss function is a widely used method in the field of object detection. It measures the overlap between the predicted bounding box and the ground truth bounding box, allowing for more accurate localization of the predicted bounding box and improving detection accuracy.

Using IoU directly as a loss function has some issues. When the predicted bounding box does not overlap with the ground truth bounding box, i.e., when IoU is zero, the gradient of the loss function is also zero, which can lead to stagnation in training. YOLOv8, building on CIOU, introduces a weighting factor W to construct W-CIOU. This weighting factor is calculated based on the ratio of the squared distance between the centres of the predicted and ground truth boxes to the squared length of the

diagonal of the smallest enclosing box. This approach allows the loss value to reflect the matching degree more flexibly between the predicted and ground truth boxes. By incorporating this weighting mechanism, W-CIOU can significantly increase the loss value when the distance between the target box and the ground truth box is large, promoting faster convergence of the model and a quicker approach to the ground truth box. The formula for W-CIOU is:

$$W - CIOU = \alpha \cdot CIOU + \beta \cdot d_{iou} \quad (8)$$

where α and β are dynamically adjusted weighting factors, and d_{iou} represents the distance measure between the target box and the ground truth box.

Through this weighting mechanism, W-CIOU retains the advantages of CIOU in measuring the overlap and aspect ratio of the bounding boxes, while the dynamic adjustment of the weighting factors enhances the model's adaptability across different detection tasks. In particular, for bird nest detection, W-CIOU is more effective in handling target localization issues in complex backgrounds, reducing both false positives and missed detections.

RESULTS AND DISCUSSION

3.1 Experimental Environment

This experiment is based on a custom-built bird nest detection dataset and uses the PyTorch 1.12.1 deep learning framework for model training and evaluation. The hardware and software environments used in the experiment are detailed in Table 1. During network training, the optimizer used is SGD with an initial learning rate of 0.05, which is gradually decayed to a final value of 0.0005. The momentum parameter starts at 0.58 and is adjusted to 0.937, with a weight decay coefficient set to 0.0005. The input image size is fixed at 640×640, the batch size is set to 8, and the maximum number of training epochs is 300. An early stopping strategy is employed, where the training is halted if there is no significant improvement in validation accuracy over 25 epochs.

Table 1. Experimental environment

Operating system	Ubuntu 20.04.1
Python	3.8
Cuda	11.3
GPU	NVIDIA RTX 3090
CPU	13th Gen Intel(R) Core(TM) i7-13700KF
Memory	64G

3.2 Experimental Dataset

The dataset used for the experiment is a custom-built bird nest detection dataset containing 5000 annotated images. The images cover bird nest instances from various environments and shapes. The dataset is divided into training, validation, and test sets in a 7:2:1 ratio. The primary challenges of this dataset include the small volume and diverse shapes of the target bird nests, as well as the high similarity in color and shape between the nests and their surrounding environment, which places higher demands on the model's object detection capabilities.

3.3 Evaluation Indicators

To comprehensively evaluate the performance of the improved YOLOv8 model, the following evaluation metrics are used: mRecall, mAP@0.5, GFLOPs, and the number of parameters (Params). Specifically, mRecall represents the proportion of correctly detected objects relative to the total number of actual objects, reflecting the model's detection rate; mAP@0.5 indicates the average precision for each class at an IoU threshold of 0.5; GFLOPs and Params represent the model's computational complexity and resource consumption, respectively.

3.4 Comparative Experiments

In this section, we conduct comparative experiments between the improved YOLOv8 model and other mainstream object detection algorithms to assess the performance enhancements of the model in bird nest detection tasks. The algorithms compared include Faster R-CNN, YOLOv5, YOLOv7, the original YOLOv8 version, and several recently proposed improved models. To ensure fairness, we keep the hyperparameters consistent across all algorithms.

The experiments utilized our custom-built bird nest dataset, which features images of bird nests captured from various angles, distances, and densities, characterized by small targets and complex backgrounds. We trained and tested all algorithms on this

dataset and quantified model performance using evaluation metrics such as mRecall, mAP@0.5, GFLOPs, and the number of parameters (Params).

As shown in Table 2, the improved YOLOv8 model achieved a mAP@0.5 of 96.3%, a significant improvement over the 94.8% of the original YOLOv8. The improved model reached an FPS of 54.7, which, although slightly lower than the 57.8 FPS of YOLOv7-tiny, still demonstrates commendable real-time performance. This indicates that the improved model balances high accuracy with detection speed. The data comparison reveals that the improved YOLOv8 reduced the number of parameters by approximately 15%, increased mAP@0.5 by 1.5 percentage points, and decreased GFLOPs by around 10%. These enhancements are primarily attributed to the lightweight backbone network and optimized feature fusion modules, making the model more efficient in handling small targets. By optimizing the model structure and parameters, the improved YOLOv8 maintains high accuracy while reducing model size and parameter count. This makes the model more suitable for deployment on edge devices or embedded systems, meeting practical requirements for model size and computational resources. The improvements in key modules such as feature extraction, feature fusion, and detection heads have effectively enhanced the model's ability to detect small targets. Specifically, the C2f-Faster-EMA module has significantly improved model accuracy by strengthening the extraction of deep semantic information, while the DWFFN module has balanced computational load and precision in cross-scale feature fusion. These improvements ensure that the model achieves a reasonable control of computational complexity while maintaining high accuracy, providing reliable support for rapid detection in practical applications.

Table 2. Comparison of experimental results

Model	mRecall (%)	mAP@0.5 (%)	GFLOPs	Params (M)	FPS
Faster R-CNN	74.5	79.4	180.7	63.2	36.1
YOLOv5s	78.9	83.7	104.3	7.2	42.7
YOLOv7-tiny	78.1	90.7	88.9	87.8	57.8
YOLOv8	85.0	94.8	87.6	3.9	54.1
Improved YOLOv8	86.7	96.3	78.9	3.3	54.7

3.5 Ablation Experiments

Ablation experiments aim to evaluate the impact of each improved module on the overall model performance, analyzing the contribution of each module to the enhancement of model accuracy and efficiency, thereby validating the effectiveness of the design improvements. By progressively removing individual modules (C2f-Faster-EMA feature extraction module, DWFFN weighted fusion module, PSDH detection head) from the improved YOLOv8 model, several comparative experiments were conducted to observe their effects on model performance.

Table 3 shows the individual effects of each improved module. Adding the C2f-EMA module only led to a 0.55 percentage point drop in mAP@0.5, but the parameter count decreased by about 5%. Adding the DWFFN module resulted in a 2.5 percentage point decrease in mAP@0.5, indicating that this module significantly improves feature extraction accuracy. Adding the PSDH module caused a significant reduction in the model's lightweight advantage, with the parameter count increasing by about 8% and mAP@0.5 dropping by 1.2 percentage points.

Table 3. Experiment results of single module ablation experiment

Module	mAP@0.5 (%)	GFLOPs	Params (M)
Basic model	94.80	87.6	3.9
C2f-EMA	92.30	84.3	3.7
DWFFN	95.75	90.8	4.5
PSDH	93.60	90.2	4.2

Table 4 presents the results of multi-module ablation experiments, validating the effectiveness of various module combinations. The combination of Improvement 1 and Improvement 2 resulted in the highest accuracy, with a 1.4 percentage point increase in mAP@0.5, achieving the best average accuracy but with larger parameter and computational requirements. The combination of Improvement 1 and Improvement 3 achieved the smallest computational and parameter counts and the fastest recognition speed, but the accuracy was the lowest among all improvements. The combination of Improvement 2 and Improvement 3, due to an inability to suppress small target information loss, did not achieve the best detection performance and had the highest parameter and computational costs. Finally, the combination of all three improvements balanced model lightweight and detection accuracy,

reaching a mAP@0.5 of 96.3% and an FPS of 54.7. The integration of Improvements 1, 2, and 3 effectively addressed issues related to small targets, multi-scale detection, and model lightweight.

Table 4. Experimental results of multi-module combination

Combination	mAP@0.5 (%)	GFLOPs	Params (M)	FPS
Improvement 1+Improvement 2	96.4	79.4	3.5	53.8
Improvement 1+Improvement 3	94.5	78.1	3.3	54.9
Improvement 2+Improvement 3	95.0	80.7	3.6	53.2
All Improvements	96.3	78.9	3.3	54.7

The ablation experiments conclude that the C2f-EMA feature extraction module effectively enhances the model's ability to detect small targets by mining deep semantic information; the DWFFN weighted fusion module improves cross-scale feature fusion and makes a significant contribution to overall accuracy; and the PSDH detection head module reduces model complexity through parameter sharing, enhancing model lightweight characteristics while maintaining accuracy.

3.6 Visual Analysis

By visualizing the detection results of the improved YOLOv8 model, the effectiveness of the model improvements and its performance in practical applications are further validated. Figure 6 presents the detection results of both the original and improved YOLOv8 models in different scenarios.

The visual results indicate that when bird nests appear small in the images, the original YOLOv8 fails to detect them. Additionally, in cases where the bird nests are significantly occluded, the original YOLOv8 model also experiences missed detections. The experimental results demonstrate that the improved YOLOv8 model exhibits stronger robustness in handling complex backgrounds and multi-scale targets. Specifically, in scenarios with small targets and substantial occlusion, the improved YOLOv8 model is more accurate in identifying the location of bird nests, with a significantly reduced false detection rate.

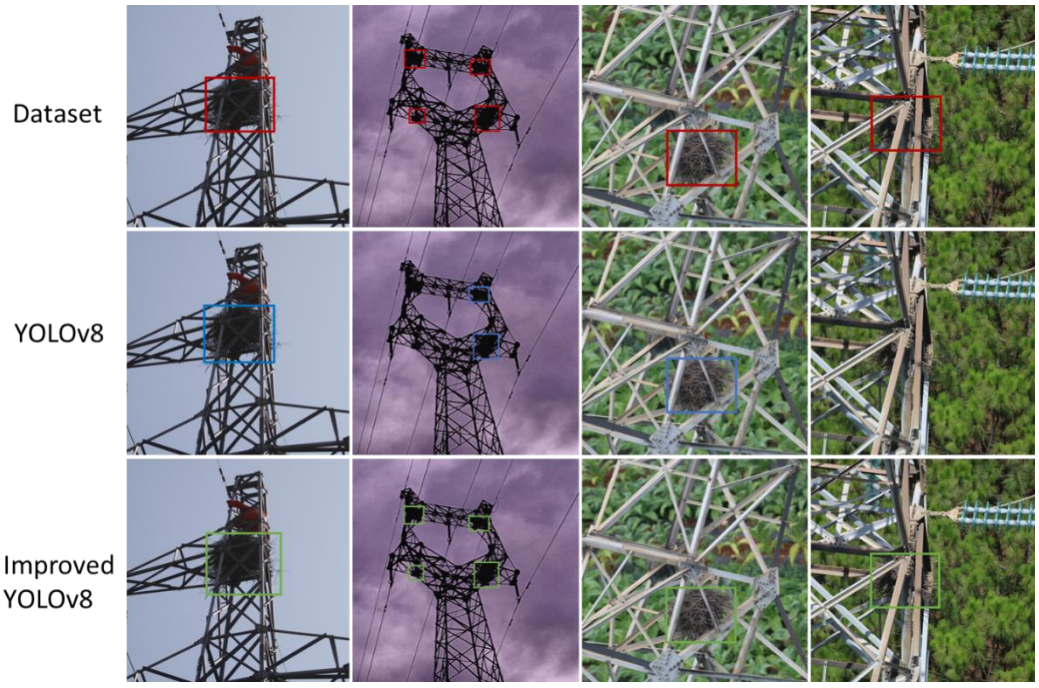

Figure 6. Comparison of detection results in different scenarios

CONCLUSIONS

In this paper, we proposed significant improvements to the YOLOv8 model, aimed specifically at enhancing bird nest detection on power transmission lines. By incorporating the EMA mechanism into the C2f module, we effectively enhanced the model's

feature extraction capabilities, leading to improved detection performance, particularly for small objects in complex environments. The introduction of the DWFFN allowed for more efficient integration of contextual features, enabling the model to achieve lightweighting by significantly reducing parameter count and computational load. Furthermore, the adoption of the W-CIOU provided a more accurate measurement of the similarity between bounding boxes, which accelerated convergence and improved the precision of the detection results. Experimental validation demonstrated that the modified YOLOv8 model achieved notable improvements in detection accuracy, precision, and recall, surpassing the original YOLOv8 model in challenging scenarios typical of bird nest detection tasks. The model not only reduced false positives and negatives but also maintained real-time detection capabilities, making it highly suitable for deployment in power line monitoring systems.

Looking forward, further enhancements could be explored to push the boundaries of bird nest detection on power transmission lines. One potential direction is the integration of advanced attention mechanisms to improve the model's focus on small objects and challenging backgrounds. Additionally, exploring the use of hybrid models that combine the strengths of different architectures could lead to even more robust detection capabilities.

FUNDING STATEMENT

This research was funded by State Grid Guangyuan Power Supply Company Innovation Project, erp number 521907240001.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest to report regarding the present study.

REFERENCES

[1] Manville, A. M. (2005). Bird strikes and electrocutions at power lines, communication towers, and wind turbines: state of the art and state of the science–next steps toward mitigation. USDA Forest Service General Technical Report PSW-GTR-191, 1051-1064.

[2] Li, H., Dong, Y., Liu, Y., Ai, J. (2022). Design and Implementation of UAVs for Bird's Nest inspection on Transmission Lines Based on Deep Leaning. Drones, 6, 252.

[3] Han, G., Wang, R., Yuan, Q., Li, S., Zhao, L., He, M., Yang, S., Qin, L. (2023). Detection of Bird Nests on Transmission Towers in Aerial Images Based on lmproved YOLOv5s. Machines, 11, 257.

[4] Liu, J., Jia, R., Li, W., Ma, F., Abdullah, H. M., Ma, H., Mohamed, M. A. (2020). High precision detection algorithm based on improved RetinaNet for defect recognition of transmission lines. Energy Reports, 6, 2430-2440.

[5] Ren, S., He, K., Girshick, R., Sun, J. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE T. Pattern Anal., 39, 1137-1149.

[6] Chen, Y., Li, W., Sakaridis, C., Dai, D., Gool, L.V. (2018). Domain Adaptive Faster R-CNN for Object Detection in the Wild. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3339-3348.

[7] Terven, J., Córdova-Esparza, D. M., Romero-González, J. A. (2023). A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Mach. Learn. Knowl. Extr., 5, 1680-1716.

[8] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S. (2020). End-to-End Object Detection with Transformers. European conference on computer vision. Cham: Springer International Publishing, 213-229.

[9] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779-788.

[10] Han, Y., Zhang, H., Wan, J., Li, X. (2024). Small target detection algorithm for UAV aerial photography based on attention mechanism. IEEE 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), 672-676.

[11] Peng, G., Yang, Z., Wang, S., Zhou, Y. (2023). AMFLW-YOLO: A Lightweight Network for Remote Sensing Image Detection Based on Attention Mechanism and Multiscale Feature Fusion. IEEE T. Geosci. Remote, 61, 4600916.

[12] Ma, N., Su, Y., Yang, L., Li, Z., Yan, H. (2024). Wheat Seed Detection and Counting Method Based on Improved YOLOv8 Model. Sensors, 24, 1654.

[13] Singh, G., Stefenon, S. F., Yow, K. C. (2023). Interpretable visual transmission lines inspections using pseudo-prototypical part network. Mach. Vision Appl., 34, 41.

[14] Qin, Z., Chen, D., Wang, H. (2024). MCA-YOLOv7: An Improved UAV Target Detection Algorithm Based on YOLOv7. IEEE Access, 12, 42642-42650.

[15] Sorbelli, F. B., Palazzetti, L., Pinotti, C. M. (2023). YOLO-based detection of Halyomorpha halys in orchards using RGB cameras and drones. Comput. Electron. Agr., 213, 108228.

[16] Ouyang, W., Luo, P., Zeng, X., Qiu, S., Tian, Y., Li, H., Yang, S., Wang, Z., Xiong, Y., Qian, C., Zhu, Z., Wang, R., Loy, C., Wang, X., Tang, X. (2014). DeepID-Net: multi-stage and deformable deep convolutional neural networks for object detection. arXiv:1409.3505

[17] Lan, Z., Zhuang, F., Lin, Z., Chen, R., Wei, L., Lai, T., Yang, C. (2024). MFO-Net: A Multiscale Feature Optimization Network for UAV Image Object Detection. IEEE Geosci. Remote S., 21, 6006605.

[18] Zhao, C., Shu, X., Yan, X., Zuo, X., Zhu, F. (2023). RDD-YOLO: A modified YOLO for detection of steel surface defects. Measurement, 214, 112776.

[19] Wang, H., Zhang, J., Tian, Y., Chen, H., Sun, H., Liu, K. (2018). A simple guidance template-based defect detection method for strip steel surfaces. IEEE T. Ind. Inform., 15, 2798-2809.

[20] Dai, X., Chen, Y., Xiao, B., Chen, D., Liu, M., Yuan, L., Zhang, L. (2021). Dynamic head: Unifying object detection heads with attentions. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7373-7382.