A Lightweight Small Object Detection Method for Birds Nest on Electric Tower Based on Attention Mechanism

*Jian Xu, Lei Lv, Ying Zhang, Xu Dai, Yaoran Huo**

Information and Communication Branch, State Grid Sichuan Electric Power Company, China
xuj2511@sc.sgcc.com.cn, 464640288@qq.com, zhangy0741a@sc.sgcc.com.cn,
daix0018@sc.sgcc.com.cn, 295579921@qq.com

Abstract

Foreign objects such as bird nests pose significant risks to the safety of transmission lines. Effective inspection and removal of these nests using aerial photography are essential. However, current target detection models face challenges in deployment on embedded devices and in small target detection. This study introduces a lightweight foreign object detection method for transmission lines based on the LW-YOLOv7 algorithm. Enhancements include a small target detection layer in the neck region and a lightweight CBAM module to improve feature extraction. The Ghost module replaces the ELAN and RepConv modules, reducing the model's weight and easing deployment. Additionally, the WIoU loss function is employed to enhance detection accuracy. Using drone-collected aerial images for training and testing, the proposed method achieves an AP of 89.31% and an accuracy of 96.28% in detecting bird nests. The method outperforms advanced target detection models in terms of model size and detection speed, providing a practical solution for ensuring the safe operation of power systems.

Keywords: Bird's nest, YOLOv7 algorithm, Ghost conv, Lightweight model

1 Introduction

As a crucial system to achieve carbon neutrality, the power system is expanding rapidly, which imposes increasing demands on the safe operation and power supply reliability of transmission lines [1]. Bird nests can easily cause line trips, short circuits, and other malfunctions [2]. These can affect the power supply security of a large area, and even cause incalculable losses, so it is necessary to detect and remove bird nests on transmission lines [3].

The advent of target detection methods based on deep learning provides a novel solution for the bird nest detection problem [4]. This scheme uses drones to capture high-definition images of transmission lines and then uses intelligent algorithms to detect whether there are bird nests in the images [5]. Deep learning-based approaches for target detection are typically segmented into two distinct groups: algorithms that operate in two stages for target detection, and those that function in a single stage. One-stage models do not have feature redundancy extraction and information transmission bottlenecks, and process faster than two-stage target detection models [6]. They can satisfy the requirements of online real-time detection of bird nests by drones. Representative one-stage target detection algorithms include single shot multi-box detectors (SSD) [7-8], RetinaNet [9], YOLO series [10-11], and so on. A lot of research indicates that One-stage methods have the advantage of high efficiency and can balance real-time performance while ensuring high accuracy. They are extensively used by researchers [12-13]. Li et al. [14] proposed an optimized YOLOv5-s model for the real-time detection of bird nests, which was implemented on an onboard computer for in-flight verification. Ju et al. [15] suggested a bird's nest detection model based on SSD to detect a bird's nest in real-time. Yang et al. combined YOLOv5 and detection transformer (DETR) to provide a method to pinpoint bird nests located on power lines, utilizing everyday scene data even with limited samples [16]. Ou et al. [17] proposed a bird's nest recognition method based on the RetinaNet deep learning model. They constructed a feature pyramid image to meet the detection of bird's nest targets at different scales. Based on the framework and the YOLOv5 algorithm, Kang et al. [18] developed a bird's nest detection software. In order to perform target detection on bird nests on transmission towers, Jie et al. replaced the pre-network VGGNet with ResNet101 to improve the feature extraction ability of the SSD algorithm. At the same time, they replaced Softmax loss with Focal loss to improve the sample imbalance problem in the SSD algorithm [19]. Compared with some methods mentioned above, the YOLOv7 model proposed by Wang et al. [20] is one of the newest algorithms in the YOLO series, with faster speed and higher accuracy.

However, the original YOLOv7 model generates a large number of parameters while enhancing the model's accuracy, making it challenging to deploy on drones. Moreover, the construction of high-performance target detection networks, including YOLO series models, focuses more on the effects of the detection network on some large-scale datasets in the general domain. When applied to specific scenario domains, they disregard the domain-specific characteristics, wasting the optimization space that the dataset features may offer. They also pay less attention to the ease of deployment and feasibility of the

*Corresponding Author: Yaoran Huo; E-mail: 295579921@qq.com
DOI: https://doi.org/10.70003/160792642025052603008

model, making the model accurate but not practical. Some research pursue the lightweight of the neural network model excessively, but this cause the model accuracy to decline drastically, especially the YOLOv7-tiny and other similar models. Generally speaking, the existing models have achieved good results in the laboratory environment. But there is still a long way to go in engineering applications. Different from the instance segmentation tasks commonly seen on the MS COCO [21] and Pascal VOC [22] datasets, the pixels of bird nests in aerial images account for a very small proportion. This requires the detector to have the ability to detect small target objects, which is a formidable task in computer vision. Therefore, the main motivation of this study is to develop an easy-to-deploy lightweight small target detection method, focusing on detecting bird nests in aerial images.

In light of the above research, we propose a lightweight small object detection method for bird nests on electric towers based on an attention mechanism. The rationale for choosing YOLOv7 as the baseline is that the YOLOv7 model employs an efficient layer aggregation network (ELAN) as the feature extraction unit, which can effectively enhance the feature expression ability and receptive field. This implies that YOLOv7 has more application potential in small object detection than other mainstream detectors. In this study, we use the K-means++ algorithm [23] to cluster the prior boxes, which reduces the clustering error while accelerating the convergence. A small target detection layer is added to the neck area to augment the detection capability of small targets. A lightweight convolution block attention module (CBAM) module [24] is introduced to enhance the feature extraction ability of the convolution module. Moreover, using the lightweight technology of the Ghost module, the ELAN module and RepConv module are replaced with the Ghost module, reducing the parameters and computation of the network model. In order to further improve the accuracy of small target detection, the WIoU loss function [25] is employed to replace the loss function in the original network, which improves the network's positioning ability for the detection target. Utilizing drone-captured imagery, we have assembled a dataset upon which our LW-YOLOv7 demonstrates superior detection capabilities when benchmarked against established object detection algorithms. In essence, the primary contributions of our research can be summarized as follows:

(1) We introduce what is, to our knowledge, the inaugural lightweight detection network tailored specifically for the identification of bird nests.

(2) A smaller scale detection layer is added on the basis of the original three scale detection layers. We demonstrated that this design is especially beneficial for detecting small target bird nests, and does not affect the detection speed.

(3) In order to make it easier to deploy to mobile devices, the lightweight CBAM module and Ghost module are introduced respectively, which reduce the model parameters and computation while enhancing the feature extraction ability.

(4) The experimental results on the self-built dataset show that the proposed LW-YOLOv7 performs well in terms of speed and accuracy in detecting bird nests.

The subsequent chapters of this paper are as follows: Section 2 delineates the proposed LW-YOLOv7 framework. Section 3 details the dataset employed for this study, along with a presentation of the precise experimental outcomes and their analysis. Section 4 encapsulates the concluding remarks.

2 The Proposed Approach

In previous studies, the YOLOv7 network has been widely used for various object detection tasks. Li et al. [26] used YOLOv7 to detect traffic signs. Jiang and colleagues [27] utilized YOLOv7 for the identification of hemp ducks, which tend to be concealed by their counterparts and present as tinier subjects (akin to the task we are tackling). Consequently, YOLOv7 was selected as the foundational network for our investigation. This model is structured with a backbone network, a connecting neck, and a terminal head. This segment will initially outline the comprehensive structure of the LW-YOLOv7 model, proceed to elucidate on each constituent module, and culminate with a discourse on the enhanced loss function.

2.1 Our Framework

As shown in Figure 1, the proposed LW-YOLOv7 aims to perform bird's nest detection tasks on transmission lines in aerial images. LW-YOLOv7 originates from YOLOv7. In previous research, Yolov7 has demonstrated excellent performance in object detection. It is why we chose it as the baseline model.

2.2 Ghost Convolution

At present, conventional detection networks predominantly employ standard convolution techniques for feature extraction. Such network architectures are characterized by extensive convolutional layering, leading to an escalation in both the model's parameters and computational demands. Feature extraction in ordinary convolution does not need to be obtained by convolution and can be replaced by cheaper operations. Ghost module is a lightweight neural network unit, whose main idea is to generate more feature maps by using a small number of convolution operations and some simple linear transformations, thereby reducing the model parameters and computational complexity [28]. The ghost convolution structure is shown in Figure 2.

In this research, we replace the ELAN module with the Ghost module and add a 1×1 convolution layer after each Ghost module to keep the output channel number consistent with the original ELAN module. We also replace the RepConv module with the Ghost module and add a 1×1 convolution layer before each Ghost module to reduce the input channel number. This can reduce the model parameters and computation, enhance the detection speed, and maintain the quality and multi-scale ability of the features.

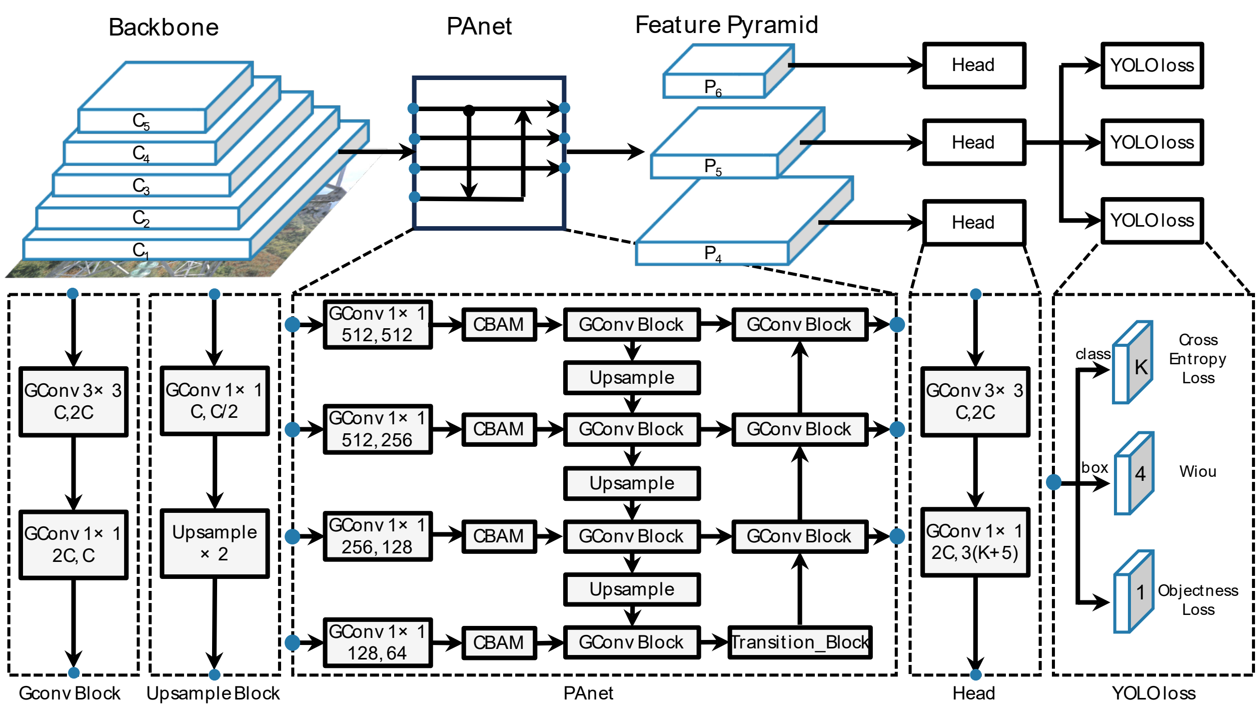

Figure 1. The LW-YOLOv7 framework diagram

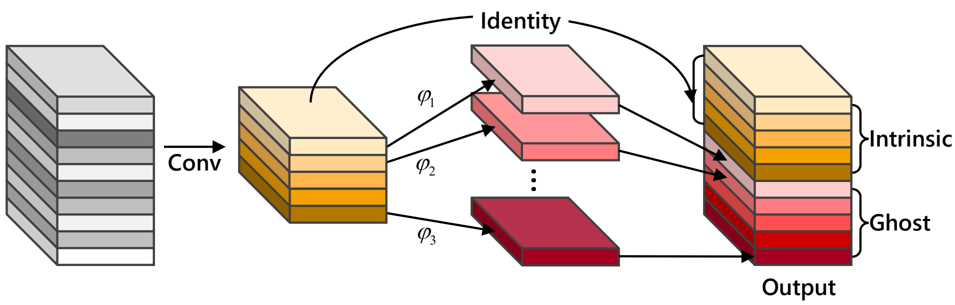

Figure 2. Ghost convolution structure diagram

2.3 K-means++ Algorithm

YOLOv7 initially clusters anchor boxes from the COCO dataset using the K-means algorithm and further refines these clusters with a genetic algorithm throughout the training phase. Yet, the effectiveness of K-means clustering is significantly influenced by the initial placement of cluster centers [29]. To address this, we implement the K-means++ algorithm, which improves detection precision and efficiency. Unlike the traditional K-means that selects several cluster centers simultaneously, K-means++ chooses a single center at a time, steering the random selection towards a more locally optimal solution that approximates the global optimum. The specific steps are as follows:

(1) A random target box is chosen from the dataset to serve as the initial clustering nucleus, followed by the computation of the smallest intersection ratio distance $A(x)$ between the residual sample boxes and the newly established cluster center.

$$A(x) = 1 - IoU_{(x,c)} \quad \textbf{(1)}$$

Where IoU signifies the intersection over union between two anchor boxes. The sub-target label sample frame is denoted as x, while the cluster center is represented by c.

(2) Calculate the probability $O(x)$ for each nest sample box to be selected as the next cluster center, employing the roulette method for the selection process.

$$O(x) = \frac{A(x)^2}{\sum_{x \in X} A(x)^2} \quad \textbf{(2)}$$

where X represents the entire sample of the target marker frame.

(3) Repeat steps 1 and 2 until all cluster centers are determined.

(4) Measure the distance from each sample in the dataset to the cluster center, assign each sample to the

class of the nearest cluster center, and recalculate the cluster center for each category c_i. Continue updating the classification and cluster centers until the size of the anchor box stabilizes.

$$c_i = \frac{1}{c_i}\sum_{x \in c_i} x \tag{3}$$

where i = 1, ..., K, K denotes the number of anchor boxes of varying sizes, determined by the detection model's anchor box count. In this study's detection model, there are three detection feature maps, each corresponding to three anchor boxes, making $k = 9$. The dimensions of each feature map and their respective optimized anchor boxes are presented in Table 1.

Table 1. The dimensions of each feature map and the corresponding anchor boxes

Feature map size	80×80	40×40	20×20
	(14,16)	(41,51)	(112,98)
YOLOv7	(9,67)	(87,37)	(66,293)
	(14,102)	(23,156)	(186,232)

2.4 CBAM Attention Mechanism

Computer vision attention mechanism is a dynamic selection process of vital information in image input, widely employed in machine learning tasks in various fields. The attention mechanism's fundamental concept is to identify the relationships within the original data and emphasize the important features, such as channel attention, pixel attention, and multi-order attention, among others. In order to better identify the tiny bird nest features in large-scale aerial images, we introduce a lightweight CBAM for YOLOv7, which comprises two different parts: channel attention module (CAM) and spatial attention module (SAM). CAM emphasizes the foreground and meaningful areas of the image, while SAM focuses on the location that contains the whole image context information. To prevent the attention mechanism module from disrupting the original weights of the backbone network, we incorporated the attention mechanism into the feature extraction part of the YOLOv7 model.

The working principle of CBAM attention mechanism is shown in Figure 3. Specifically, further, its working principle is as follows:

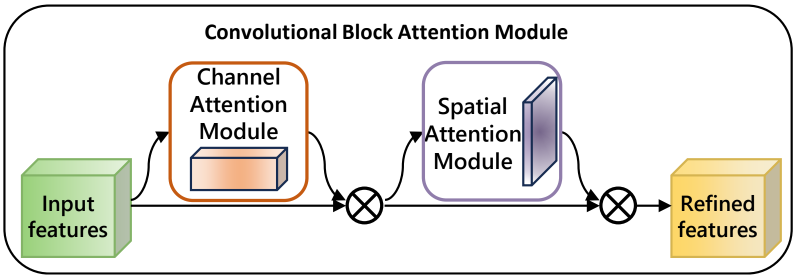

Figure 3. Convolution block attention module structure diagram

In the CAM of CBAM, the input feature map is $F \in \mathbb{R}^{C \times H \times W}$, where C is the number of channels, H is the height, and W is the width. Apply global average pooling (GAP) and global max pooling (GMP) to the feature map, resulting in two feature vectors F_{aug} and F_{max}, each of size $1 \times 1 \times C$. Pass these feature vectors through a shared MLP. The MLP has one hidden layer with C/r neurons, where r is the reduction rate (typically 8). Then add the two feature vectors and apply a Sigmoid function to obtain the channel attention map $M_c \in \mathbb{R}^{C \times 1 \times 1}$. These features are subsequently reweighting the input feature map by the channel attention map. resulting in the modified feature map, which serves as the input for the SAM. This process is illustrated by equation (4).

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \tag{4}$$

The SAM focuses on the importance of different spatial locations. In the SAM, the input feature map is $F' \in \mathbb{R}^{C \times H \times W}$, which is the output from the CAM. Apply GAP and GMP along the channel dimension, resulting in two feature maps $F'_{aug} \in \mathbb{R}^{1 \times H \times W}$ and $F'_{max} \in \mathbb{R}^{1 \times H \times W}$. These maps are then concatenated along the channel dimension to get $F'' \in \mathbb{R}^{2 \times H \times W}$. Passing F'' through a 7×7 dimensionality reduction, convolution layer to generate a single-channel feature map, then apply a Sigmoid function to obtain the spatial attention map $M_s \in \mathbb{R}^{1 \times H \times W}$. Finally, this spatial attention feature is multiplied by the input feature map, resulting in the final feature map. Equation (5) is the model of this operation.

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \tag{5}$$

The CBAM attention mechanism, as a plug and play module, is embedded in the underlying convolution of the detection network neck.

2.5 Small Object Detection Structure

The YOLOv7 model utilizes different scale feature maps to detect objects of various sizes. During the aerial inspection of transmission lines, maintaining a consistent image shooting distance is challenging, leading to variations in the sizes of bird nest targets across different images. Facing samples with both large and small targets, pruning methods should not be used. There are also bird nests that are occluded, and the exposed size is less than 8×8 pixels.

To solve the defects of YOLOv7 in detecting small target objects, we introduce a dedicated small target feature extraction layer in Backbone to augment the performance of the model. We use the 160x160 sized feature layer obtained from the second downsampling in the YOLOv7 backbone network as a supplementary feature layer introduced into the neck. After determining the focus of the feature layer through the CBAM mechanism, we fuse the upsampled feature map with an 80x80 sized feature layer to supplement the position information of the features for the model. Finally, we merge the feature map after downsampling with the original 80×80 feature map as the small target detection head. The introduction of the 160×160 feature layer compensates for the disadvantage of losing the information of small objects during the downsampling process of YOLOv7 model. Enhanced the detection effect of small objects.

2.6 WIoU Loss Function

In YOLOv7 model, the difference between the prediction result and the actual value is usually calculated by the loss function L_{CIoU}.

However, research has found L_{CIoU} has a series of defects [29]. Therefore, we replace the loss function of YOLOv7 with WioU, which has a dynamic non-monotonic focusing mechanism. The calculation formula of L_{WIoU} is as follows:

$$L_{WIoU} = rR_{WIoU}L_{IoU}, r = \frac{\beta}{\delta\alpha^{\beta-\delta}} \tag{6}$$

$$R_{WIoU} = \exp\left(\frac{(x-x_{gt})^2+(y-y_{gt})^2}{(W_g^2+H_g^2)^*}\right) \tag{7}$$

$$L_{IoU} = 1 - IoU = 1 - \frac{W_iH_i}{S_u} \tag{8}$$

$$\beta = \frac{L_{IoU}}{\overline{L}_{IoU}} \in \left[0, +\infty\right) \tag{9}$$

where $\overline{L}_{IoU}$ represents the exponential running average with momentum m. The intersection over Union L_{IoU} is employed to gauge the overlap between the anchor box and the target box; δ and α are hyperparameters; β indicates the outlier of the anchor box; x and y are the horizontal and vertical coordinates of the center point of the predicted anchor box, respectively; x_{gt} and y_{gt} are the coordinates of the real anchor box's center point. W_g and H_g are the dimensions of the smallest enclosing box. L_{IoU} is the intersection ratio of the overlapping area between the predicted box and the real box.

3 Experiments

This section presents several comparative experiments to assess the proposed method for detecting bird nests on power transmission lines. First, the composition of the dataset is described. Next, the detection criteria for bird nests are introduced. Following that, LW-YOLOv7 model is compared with several other methods. Besides, ablation experiments are conducted to demonstrate the effectiveness of the proposed LW-YOLOv7 method.

3.1 Experimental Setup

We performed our experiments on a computer system equipped with 4 NVIDIA GeForce RTX 2080Ti graphics cards. Developed the LW-YOLOv7 model in Python 3.8 and PyTorch 1.8.2. The laboratory setup environment is detailed in Table 2.

Table 2. Configuration

Configuration	Name	Type
Hardware	CPU	Intel® X®(R) Silver 4110
	GPU	NVIDIA GeForce RTX 2080Ti*4
	Disk	4T
	Running memory	192 GB
Software	CUDA	11.3
	Pytorch	1.8.2
	Python	3.8

During the LW-YOLOv7 training process, we set the batch size to 16. Using momentum-based Adam algorithm to update gradients. The initial learning rate of the Adam optimizer is 0.0015. The training includes 250 epochs. In order to accelerate training efficiency and prevent weight damage, the backbone weights were frozen in the first 50 epochs. For testing set, both the IoU threshold and the confidence threshold are set to 0.5. Table 3 summarizes the hyperparameter settings used in our experiments.

Table 3. Hyperparameters

Configuration	Component name/Value
Epoch	250
Input image size	640 × 640
Batch size	16
Initial learning rate	0.0015
Optimizer	Adam
IoU threshold	0.5
Confidence threshold	0.5

3.2 Dataset

Due to the lack of public aerial image datasets of transmission lines, the images used in this paper are all from the image dataset collected by Sichuan Electric Power Company Information Communication Co., Ltd.

using unmanned aerial vehicles. In the actual unmanned aerial vehicle (UAV) inspection process, the dataset is accumulated by aerial videos captured by airborne HDR cameras. The resolution of all aerial videos is 3648 × 5472 at 30 FPS. We extracted 1600 suitable frames from these aerial videos at the same time interval as the dataset. To evaluate the performance of the model, we divided the dataset into training and testing sets in an 8:2 ratio. We manually annotate the bird's nest using Label Studio software to obtain high-quality labels. Figure 4 provides detailed instances and their corresponding basic facts. Please note that the images in the dataset are captured by airborne HDR cameras, which are positioned far from the transmission line corridor, resulting in wide-angle photos of the power towers. Therefore, the background occupies most of the image space. As shown in the figure, the bird nests are very small targets relative to the entire image, and some nests are difficult to identify with the naked eye. It can be observed that there is a significant imbalance between the pixels occupied by backgrounds such as the sky and those occupied by bird nests. In addition, we must note that power towers are also one of the main obstacles to detecting bird nests. Therefore, developing lightweight bird nest detection on this dataset is a highly challenging task.

Figure 4. Some typical samples

3.3 Evaluation Indicators

The widely used evaluation indicators in object detection research include precision (P), recall (R), average precision (AP), and F_1. The P, R, F_1 and AP can be measured using Equations (10)–(13).

TP represents true positives (correct detection). FN represents false negatives (missed detection). And FP represents false positives (incorrect detection). F_1 is the trade-off between P and R, frequently used to comprehensively evaluate the performance of the model. $P_{(R)}$ is a function of R. AP is the overall detection performance. The larger AP value represent the better the performance of the model.

$$P = \frac{TP}{TP + FP} \tag{10}$$

$$R = \frac{TP}{TP + FN} \tag{11}$$

$$F_1 = \frac{2 \times R \times P}{R + P} \tag{12}$$

$$AP = \int_0^1 P_{(R)} dR \tag{13}$$

3.4 Superiority of LW-YOLOv7

To demonstrate the superiority of the LW-YOLOv7 model, we compared EfficientDet [30]. YOLOX [31], YOLOv5 [32], YOLOv7, SSD and FCOS [33] with LW-YOLOv7 model. Table 4 shows the comparison results. We will comprehensively compare these models through multiple dimensions such as accuracy, recall, and average accuracy. Table 4 shows the results of the comparative experiment. From the experiment results, it can be seen that the LW-YOLOv7 model outperforms other methods in terms of accuracy, recall, and average accuracy. The average precision of LW-YOLOv7 model for insulators is 89.31%. The recall rate of this method for insulators has been significantly increased, indicating that this method can reduce the number of missed detections of bird nests.

3.5 Effectiveness of the Improvements

In order to verify the effectiveness of the LW-YOLOv7 model, we evaluate the individual contribution of each improvement through ablation experiments. Each group of experiments is repeated 3 times. The average values of the indicators obtained from the experiments are calculated and filled in Table 5. After adding CBAM, the precision and recall of bird nests are enhanced. The ability of this module to detect small targets is verified. After replacing the loss function with WIoU, the recall of the model detection is significantly increased, which can reduce the problem of missing small targets.

As shown in Table 5, the YOLOv7 method without improvement modules has a precision of 93.75%, but the recall rate is only 71.09%. The second group of experiments added CBAM to the YOLOv7 model. The precision and recall for bird nests detection were improved, and the F_1 indicator increased by 1%. The third group of experiments with the WIoU loss function, and the recall of the model detection was significantly improved, which can reduce the problem of missing small targets. Then, on the basis of the third group of experiments, the CBAM module was added, which increased the AP by 3.56% and the F_1 by 3%. In addition, the fifth group added a dedicated small target feature extraction layer, which achieved the best performance; the *AP*, *P*, and *R* of the improved model reached 89.31%, 96.28%, and 82.31%, respectively.

Table 4. Nested detection performance of different object detection methods

Detection performance	AP	F1	P	FLOPs	FPS	Recall	Parameters
EfficientDet	59.72	0.64	90.52	410B	40.8	49.76	33.42M
FCOS	69.22	0.72	84.18	1.2M	49.8	63.03	32.11M
SSD	74.01	0.71	94.44	47B	50.0	56.4	23.62M
YOLOV5	54.68	0.53	88.89	0.37G	47.1	37.91	46.63M
YOLOV7	83.95	0.81	93.75	0.28G	52.7	71.09	37.01M
YOLOX	85.78	0.85	92.7	0.82G	58.9	78.2	54.14M
Ours	89.31	0.89	96.28	0.24G	61.9	82.31	37.51M

Table 5. Performance of the proposed methods

ID	Method	AP	F_1	P	FLOPs	FPS	Recall
1	YOLOv7	83.95	0.81	93.75	0.28G	52.7	71.09
2	YOLOv7+CBAM	85.95	0.82	95.03	0.29G	51.2	72.51
3	YOLOv7+ WIoU	84.26	0.84	90.11	0.28G	52.1	79.12
4	YOLOv7+CBAM+WIOU	87.82	0.87	94.89	0.28G	53.4	80.1
5	Ours	89.31	0.89	96.28	0.24G	61.9	82.31

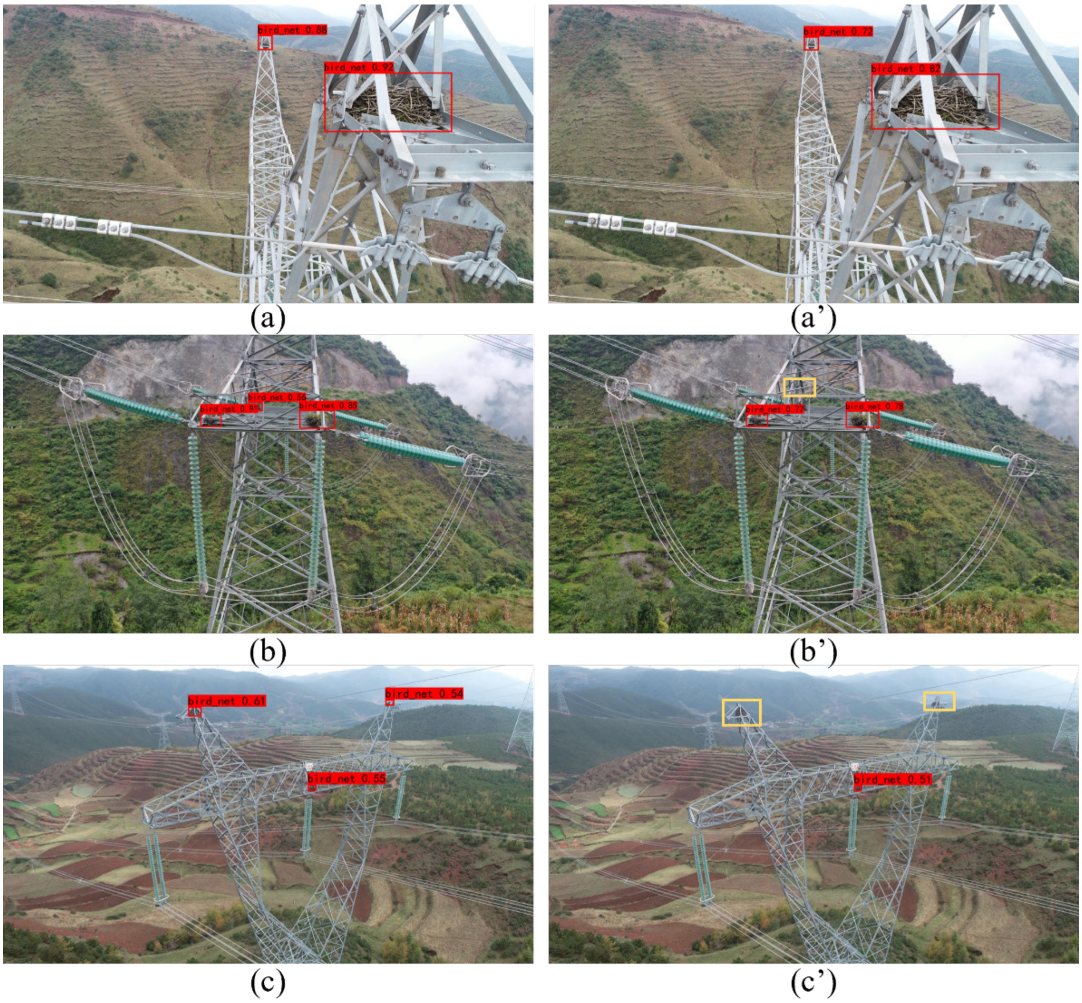

Figure 5. Comparison of different methods
(We obtained the results in Figures a-c using our method, where Figures a '- c' were obtained from the original YOLOv7.)

Figure 5 displays the bird nest detection results of YOLOv7 model and our LW-YOLOv7 model. The red box indicates the bird nest, and the undetected ones are marked with a yellow box. In Figure 5(a), our method detects both the bird nests in the foreground and the background and has higher accuracy than the YOLOv7 model. In the middle of Figure 5(b'), there is a bird nest that is almost entirely occluded, which is detected by LW-YOLOv7 in Figure 5(b). As shown in Figure 5(c'), there are two tiny bird nests at the top of the tower that are severely obscured, which are difficult to find by the naked eye. The YOLOv7 method also did not detect the target, but our method did.

4 Conclusion

This work proposed a lightweight small object detection method for bird nests detection. The ablation experiments show that the AP, P, and R of the LW-YOLOv7 reach 89.31%, 96.28%, and 82.31%, respectively, which are 4.36%, 2.53%, and 11.22% higher than the original YOLOv7 algorithm. The comparative experiments show that this algorithm is better than the existing bird's nest detection network in terms of accuracy and efficiency. Besides, the parameters of LW-YOLOv7 model are less than other YOLO series networks, makes it easier to deploy. The final improved model is more lightweight and effectively solves the problem of difficult balance between detection efficiency and confidence level. Providing a reference scheme for the actual detection of bird nests on transmission lines. To further advance the practicality of related technologies in the scene of small target recognition, we will continue to study how to accelerate the calculation speed of the LW-YOLOv7 model.

Acknowledgment

This work is supported by the Foundation of State Grid Sichuan Electric Power Company under the project 'The Key Technologies and Applications of Defect Recognition with Few-Shot Samples for Electric Transmission Lines' (521947230002).

References

[1] J. Liao, H. Xu, X. Fang, Q. Miao, G. Zhu, Quantitative Assessment Framework for Non-Structural Bird's Nest Risk Information of Transmission Tower in High-Resolution UAV Images, IEEE Transactions on Instrumentation and Measurement, Vol. 72, pp. 1-12, April, 2023.

[2] J. Li, D. Yan, K. Luan, Z. Li, H. Liang, Deep Learning-Based Bird's Nest Detection on Transmission Lines Using UAV Imagery, Applied Sciences-Basel, Vol. 10, No. 18, Article No. 6147, September, 2020.

[3] S. Wang, X. Wang, F. Guo, B. Han, X. Wang, Study on Nest Material Flashover Risk Level Assessment of AC Overhead Transmission Line, *IEEE Electrical Insulation Conference*, Denver, CO, USA, 2021, pp. 656-659.

[4] L. Shi, Y. Chen, G. Fang, K. Chen, H. Zhang, Comprehensive identification method of bird's nest on transmission line, *Energy Reports*, Vol. 8, No. Supplement 6, pp. 742-753. September, 2022.

[5] Z. Ge, H. Li, R. Yang, H. Liu, S. Pei, Z. Jia, Z. Ma, Bird's Nest Detection Algorithm for Transmission Lines Based on Deep Learning, *2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA)*, Changchun, China, 2022, pp. 417-420.

[6] J. Rabbi, N. Ray, M. Schubert, S. Chowdhury, D. Chao, Small-Object Detection in Remote Sensing Images with End-to-End Edge-Enhanced GAN and Object Detector Network, *Remote Sensing*, Vol. 12, No. 9, Article No. 1432, May, 2020.

[7] K. Lavanya, S. Prakash, Y. Gedam, A. Aijaz, L. Ramanathan, Real time digital face mask detection using MobileNet-V2 and SSD with apache spark. International *Journal of Performability Engineering*, Vol. 18, No. 8, pp. 598-604, August, 2022.

[8] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, SSD: Single Shot MultiBox Detector, *14th European Conference on Computer Vision*, Amsterdam, Netherlands, 2016, pp. 21-37.

[9] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, Focal Loss for Dense Object Detection, *16th IEEE International Conference on Computer Vision*, Venice, Italy, 2017, pp. 2999-3007.

[10] A. Kumar, G. S. Lehal, Layout Detection of Punjabi Newspapers using the YOLOv8 Model, *International Journal of Performability Engineering*, Vol. 20, No. 3, pp. 186-193, March, 2024.

[11] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You Only Look Once: Unified, Real-Time Object Detection, *2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 779-788.

[12] J. Chen, Z. An, Z. Tang, Z. Lu, Research on tower defect detection method of UAV patrol inspection image based on improved YOLOv4 model, *Electrical Measurement and Instrumentation*, Vol. 60, No. 10, pp. 155-160, October, 2023.

[13] D. Satheeswari, L. Shanmugam, N. M. J. Swaroopan, Recognition of Bird's Nest in High Voltage Power Line using SSD, *First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, Trichy, India, 2022, pp. 1-7.

[14] H. Li, Y. Dong, Y. Liu, J. Ai, Design and Implementation of UAVs for Bird's Nest Inspection on Transmission Lines Based on Deep Learning, *Drones*, Vol. 6, No. 9, Article No. 252, September, 2022.

[15] M. Ju, C. D. Yoo, Detection of Bird's Nest in Real Time Based on Relation with Electric Pole Using Deep Neural Network, *34th International Technical Conference on Circuits/Systems, Computers and Communications*, JeJu, South Korea, 2019, pp. 389-392.

[16] Y. Yang, X. Wang, Recognition of bird nests on transmission lines based on YOLOv5 and DETR using small samples, *Energy Reports*, Vol. 9, pp. 6219-6226, December, 2023.

[17] J. Ou, Y. Yang, L. Shi, Z. Zhou, S. Qiu, Research on Bird's Nest Recognition Method for Transmission Line Towers Based on Deep Learning, *Mechanical and Electrical Information*, No. 24, pp. 22-23, August, 2020.

[18] Y. Kang, G. Xie, H. Rong, R. Lan, An intelligent detection system for power inspection bird nests based on YOLOv5, *Electronic Technology and Software Engineering*, Vol. 8, pp. 259-262, May, 2022.

[19] Q. Jie, L. Jiao, Bird Nest Detection on Transmission Tower Based on Improved SSD Algorithm, *Computer System Applications*, Vol. 29, No. 5, pp. 202-208, May, 2020.

[20] C.-Y. Wang, A. Bochkovskiy, H.-Y. M. Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, Canada, 2023, pp. 7464-7475.

[21] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, C. L. Zitnick, Microsoft COCO: Common Objects in Context, *13th European Conference on Computer Vision*, Zurich, Switzerland, 2014, pp. 740-755.

[22] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, A. Zisserman, The Pascal Visual Object Classes (VOC) Challenge, *International Journal of Computer Vision*, Vol. 88, No. 2, pp. 303-338, June, 2010.

[23] D. Arthur, S. Vassilvitskii, k-means++: The Advantages of Careful Seeding, *18th ACM-SIAM Symposium on Discrete Algorithms*, New Orleans, LA, 2007, pp. 1027-1035.

[24] S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, CBAM: Convolutional Block Attention Module, *15th European Conference on Computer Vision*, Munich, Germany, 2018, pp. 3-19.

[25] Z. Tong, Y. Chen, Z. Xu, R. Yu, Wise-IoU: Bounding *Box Regression Loss with Dynamic Focusing Mechanism*, arXiv: 2301.10051, April, 2023. https://arxiv.org/abs/2301.10051

[26] S. Li, S. Wang, P. Wang, A Small Object Detection Algorithm for Traffic Signs Based on Improved YOLOv7, *Sensors*, Vol. 23, No. 16, Article No. 7145, August, 2023.

[27] K. Jiang, T. Xie, R. Yan, X. Wen, D. Li, H. Jiang, N. Jiang, L. Feng, X. Duan, J. Wang, An Attention Mechanism-Improved YOLOv7 Object Detection Algorithm for Hemp Duck Count Estimation, *Agriculture*, Vol. 12, No. 10, Article No. 1659, October, 2022.

[28] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, C. Xu, GhostNet: More Features from Cheap Operations, *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 2020, pp. 1577-1586.

[29] L. Bai, J. Liang, C. Dang, F. Cao, A cluster centers initialization method for clustering categorical data, *Expert Systems with Applications*, Vol. 39, No. 9, pp. 8022-8029, July, 2012.

[30] M. Tan, R. Pang, Q. V. Le, EfficientDet: Scalable and Efficient Object Detection, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 10778-10787.

[31] Z. Ge, S. Liu, F. Wang, Z. Li, J. Sun, YOLOX: Exceeding YOLO Series in 2021, arXiv e-prints 2107.08430, August, 20221. https://arxiv.org/abs/2107.08430

[32] J. Yuan, X. Zheng, L. Peng, K. Qu, H. Luo, L. Wei, J. Jin, F. Tan, Identification method of typical defects in transmission lines based on YOLOv5 object detection algorithm, *Energy Reports*, Vol. 9, No. Supplement 7, pp. 323-332, September, 2023.

[33] Z. Tian, C. Shen, H. Chen, T. He, FCOS: Fully Convolutional One-Stage Object Detection, *IEEE/CVF International Conference on Computer Vision*, Seoul, South Korea, 2019, pp. 9626-9635.

Biographies

Jian Xu is currently working in State Grid Sichuan Information & Communication Company. His research interests include smart Grid, Internet of Things.

Lei Lv is currently working in State Grid Sichuan Information & Communication Company. His research interests include smart Grid, Internet of Things.

Ying Zhang is currently working in State Grid Sichuan Information & Communication Company. Her research interests include power communication, and digitalization of new electric power systems.

Xu Dai is currently working in State Grid Sichuan Information & Communication Company. His research interests include AI, Electric Power BDS.

Yaoran Huo is currently working in State Grid Sichuan Information & Communication Company. His research interests include AI, Pre-trained large AI model, Internet of Things.