ISSN 2616-5767 Vol.8, Issue 1: 8-17, DOI: 10.25236/AJETS.2025.080102

FMH-YOLO: Detecting Foreign Objects on Transmission Lines via Enhanced Yolov8

Guangyao Wang*

College of Information Engineering, Nanjing University of Finance and Economics, Nanjing, 210023, China
**Corresponding author:884255402@qq.com*

Abstract: *Power transmission lines are a critical component of the power system, and due to their widespread presence in outdoor environments, they are highly susceptible to the attachment of foreign objects. These foreign objects pose a risk of causing accidents in power transmission lines. To ensure the stability and safety of the power supply, regular inspections of transmission lines are necessary. Aiming at the precise and efficient detection of foreign objects along power transmission lines, this research puts forward a new FMH-YOLO model which is founded on the YOLOv8 network. Firstly, we introduce Partial Convolution and the EMA to design a new Faster-EMA Block (FEB) structure. Based on this, we construct the FE-C2F module to enhance the model's feature extraction capabilities. Secondly, we propose a multi-scale lightweight convolution module based on GhostConv. Building on this, we design the ML-C2F module to improve the model's ability of feature integration. Additionally, a detection head specifically for small objects is added. The experimental results demonstrate that FMH-YOLO outperforms the baseline YOLOv8 in multiple evaluation metrics, achieving improvements of 2.7% in Precision, 2.2% in Recall, 2.9% in AP50, and 1.8% in mAP@0.5. These results evidently signify its enhanced efficacy and outstanding capabilities in the mission of identifying foreign objects on power transmission lines. Moreover, the graphical representations also attest to the remarkable proficiency of FMH - YOLO. Therefore, this method provides technical support for foreign object detection on power transmission lines, and when deployed on drones, it can effectively carry out inspection tasks.*

Keywords: *YOLOv8, object detection, transmission line, Faster-EMA Block, multi-scale lightweight convolution*

1. Introduction

Due to human activities and the influence of the natural environment, transmission lines are highly vulnerable to the accumulation of foreign objects including bird nests, beehives, branches, kites, balloons, and trash[1]. These foreign objects can pose potential hazards to transmission lines and even trigger accidents, making regular inspections of the transmission lines extremely necessary. In the past, transmission lines were primarily inspected manually, but this method had several disadvantages, including low efficiency, high risk, and limited scope of inspection. Subsequently, helicopter inspections were adopted, which improved efficiency but came with high maintenance costs and poor applicability. Today, with the advancement of technology and continuous improvement in GPU performance, researchers are combining machine vision technology with deep learning algorithms to conduct research on object detection algorithms[2]. By utilizing drones equipped with object detection algorithms, unmanned aerial vehicle (UAV) inspections have been implemented, quickly becoming a hot topic in the industry. Drones offer several advantages, such as low operational costs, high efficiency, and simple operation, enabling them to effectively complete the task of inspecting transmission lines.

There are two broad categories into which object detection algorithms can be classified: two-stage and single-stage ones. Two-stage algorithms generate target candidate regions and extract features from these regions for classification, providing an advantage in accuracy and offering higher detection precision. However, the processing flow of two-stage algorithms is relatively complex, and their operational speed is slow. In scenarios with limited resources or where high-speed real-time detection is required, they are less applicable than single-stage algorithms, resulting in lower practicality. Representative algorithms include R-CNN, Faster R-CNN, and Mask R-CNN.Single-stage object

ISSN 2616-5767 Vol.8, Issue 1: 8-17, DOI: 10.25236/AJETS.2025.080102

detection algorithms eliminate the generation of candidate regions. By utilizing classification branches and bounding box regression branches, they can simultaneously perform classification and localization tasks. Although this reduces accuracy to some extent, the simplified processing flow significantly enhances processing speed. This enables single-stage algorithms to be more applicable to real-time detection scenarios. Representative algorithms include the YOLO series, SSD, and RetinaNet[3].

However, many earlier deep convolutional neural networks were mainly developed for detecting objects in natural scenes. When applied directly to mission of identifying foreign objects in power transmission lines, these models yield suboptimal results. With the aim of overcoming the drawbacks of current approaches, this paper presents a modified YOLOv8-based algorithm, FMH-YOLO, specifically designed for foreign object detection in power transmission lines. As follows are the key innovations that this method encompasses:

(1) In the backbone section, Partial Convolution (PConv) and Efficient Multi-scale Attention (EMA) are used to introduce a novel Faster-EMA Block (FEB) structure. Based on this, the new FE-C2f module is constructed to reduce redundant computations, enhance operational speed, and improve detection performance by enhancing the model's feature retrieval proficiency.

(2) In the neck section, by introducing GhostConv, a Multi-scale Lightweight Convolution module (MLConv) is proposed. Based on this, the novel ML-C2f module is designed, which reduces both false negatives and false positives while being lightweight, effectively enhancing accuracy.

(3) An additional detection head is incorporated to enhance the detection effectiveness for small targets.

2. YOLOv8

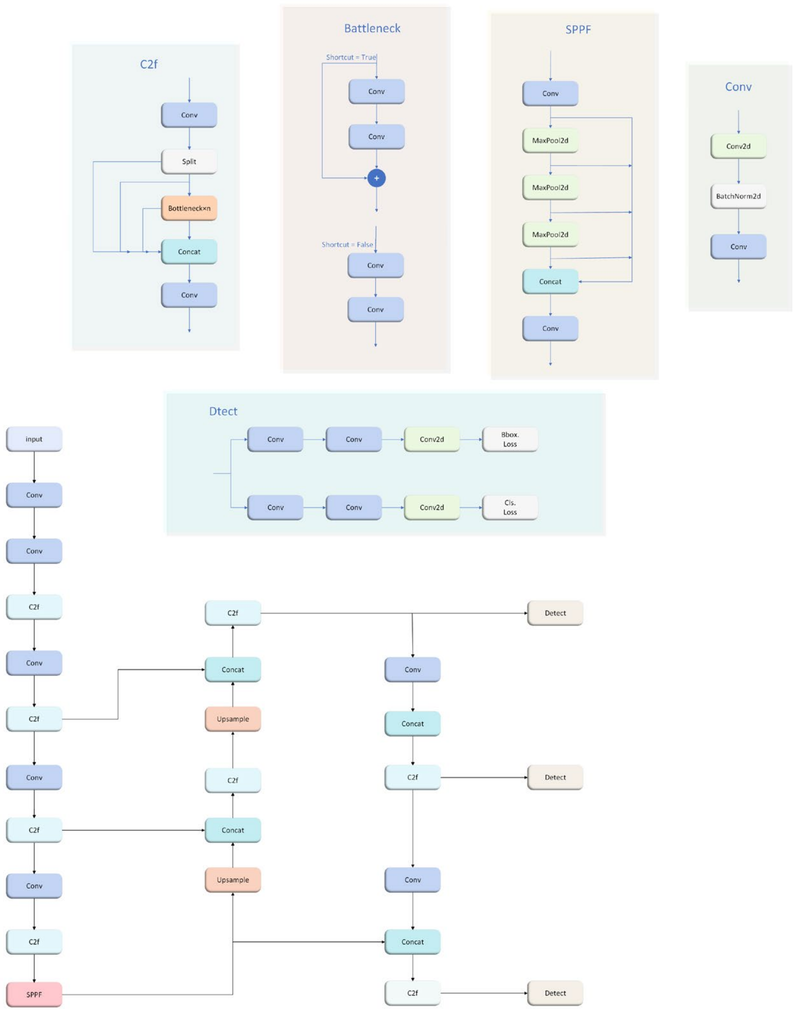


Figure 1. The network structure of YOLOv8.

ISSN 2616-5767 Vol.8, Issue 1: 8-17, DOI: 10.25236/AJETS.2025.080102

On January 10, 2023, Ultralytics introduced and publicized YOLOv8, which represents an object detection algorithm. YOLOv8 builds on the foundations of its predecessor models, incorporating multiple optimizations and enhancements, which enhance its applicability in the field of object detection.Based on varying model sizes, the YOLOv8 algorithm is available in five versions: YOLOv8-n, YOLOv8-s, YOLOv8-m, YOLOv8-l, and YOLOv8-x. In this research, the YOLOv8-n is adopted as the fundamental model, upon which improvements are made.

The network architecture of YOLOv8 primarily consists of three components: the Backbone, the Neck , and the Head. In the architecture of YOLOv8, the backbone network holds a pivotal position. It undertakes the task of extracting the characteristic details from the input image, thereby laying the groundwork essential for the ensuing detection operations. It primarily consists of the convolutional module (Conv), the C2f module, and the SPPF module used in YOLOv5. The neck part is mainly responsible for integrating feature maps derived from various layers of the backbone, aiming to obtain more comprehensive and accurate feature representations. The head is mainly responsible for further processing the multi-scale feature maps output by the neck, to achieve classification and localization. For the loss function utilized in YOLOv8, the Binary Cross-Entropy Loss (BCE Loss) is selected as the means to calculate the classification loss. BCE Loss is particularly effective for handling multi-class classification problems, as it adjusts the weight of each class, ensuring accurate classification of targets from different categories[4].As depicted in Figure 1, it is the network structure of YOLOv8.

3. FMH-YOLO

With regard to the traditional algorithms applied to detect foreign objects within power transmission lines, problems like elevated false-negative ratios, pronounced false-positive ratios, and low detection accuracy are common. In an attempt to surmount these obstacles, the current research presents an enhanced model founded on the basis of YOLOv8-n, named FMH-YOLO, aimed at overcoming the aforementioned problems. In the FMH-YOLO model, the Backbone section incorporates the Pconv[5] and EMA[6] modules to construct the Faster-EMA Block (FEB), and introduces a novel FE-C2f module by combining C2F. This approach succeeds in diminishing the computation and network parameters while enhancing the feature extraction capability for foreign bodies in power transmission lines. In the Neck section, a new multi-scale lightweight module, MLConv, is designed based on GhostConv, and the C2F module is reconstructed to create a new ML-C2f module, which not only ensures lightweight design but also improves detection performance. In the Head section, to address the challenges of detecting small objects, a small-object detection head is added, improving the detection accuracy for small targets. The detailed network architecture of FMH-YOLO is shown in Figure 2.

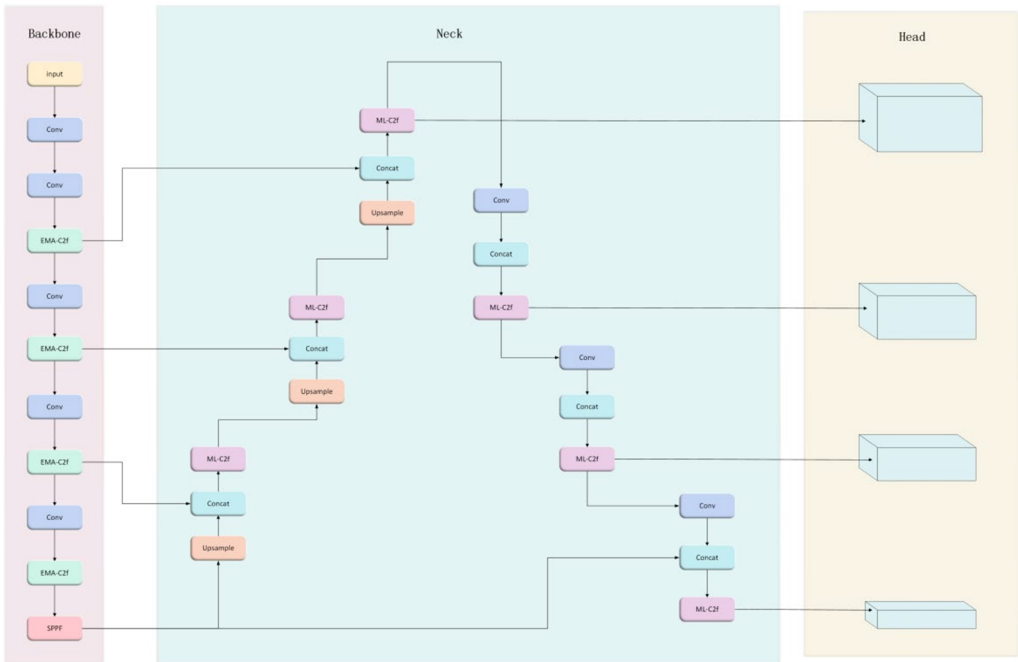


Figure 2. The network structure of FMH-YOLO.

ISSN 2616-5767 Vol.8, Issue 1: 8-17, DOI: 10.25236/AJETS.2025.080102

3.1. FE-C2f

This paper employs Partial Convolution(PConv) and Efficient Multi-scale Attention (EMA) module to construct the Faster-EMA Block(FEB) structure, and based on this, the C2f module is reconstructed to design the novel FE-C2f module.This significantly enhances the network's capability for feature extraction, computational efficiency, robustness, and small object detection, while also reducing memory usage and energy consumption.FLOPS, an abbreviation for floating-point operations per second, measures effective computational speed[7]. The FE-C2f module achieves lower FLOPS by utilizing partial convolutions, which reduce redundant computations and memory accesses, thereby efficiently extracting spatial features. For PConv, the FLOPs are determined through the formula:

$$h\times w\times k^2\times c_p^2, \tag{1}$$

In this regard, the feature map's height and width are signified by h and w respectively, the convolution kernel's size is signified by k, and the number of channels involved in the computation is signified by c_p. Upon the condition that cp/c is equal to 1/4, the number of FLOPs for PConv turns out to be only 1/16 of that for a standard Conv. At this ratio, PConv demonstrates a reduced memory footprint, calculated as

$$h\times w\times 2c_p+k^2\times c_p^2\approx h\times w\times 2c_p, \tag{2}$$

which is only 1/4 of that required for conventional convolution.Through this method, PConv not only minimizes computational load but also significantly decreases memory access, thereby enhancing overall computational efficiency, making it particularly suitable for resource-constrained environments. The structure of Partial Convolution is depicted in Figure 3.

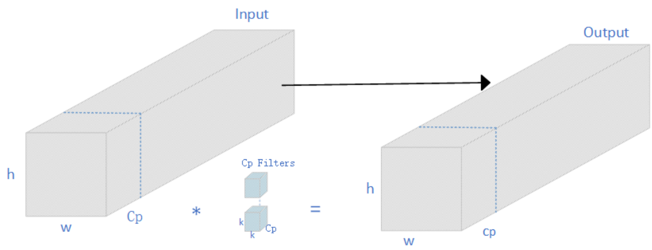


Figure 3. The structure of Partial Convolution.

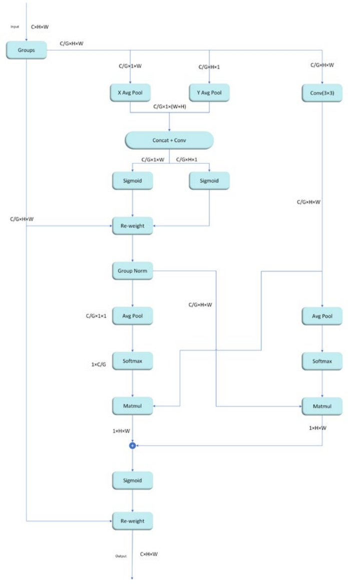


Figure 4.The structure of EMA.

By smoothing feature variations through the EMA module, network performance is enhanced when faced with unstable inputs or noisy data, improving the model's robustness and its applicability in intricate situations. The design of the EMA module is depicted in Figure 4.

By employing Partial Convolution and the EMA module, the FasterNet-EMA Block (FEB) is constructed, balancing performance and efficiency. While ensuring the module's lightweight nature, it

ISSN 2616-5767 Vol.8, Issue 1: 8-17, DOI: 10.25236/AJETS.2025.080102

delivers superior performance and is easily integrated into existing networks as a crucial component of more complex networks, enhancing the comprehensive efficiency of the network model. Figure 5 illustrates the architecture of the FasterNet-EMA Block (FEB).

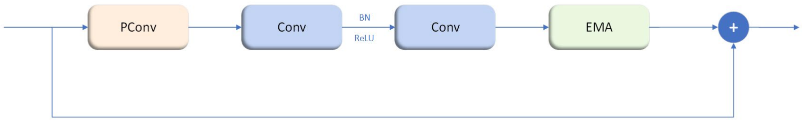


Figure 5.The structure of FasterNet-EMA Block (FEB).

In the C2f module, the Bottleneck is primarily used to optimize the feature representation ability and computational performance of the network. However, during the process of compressing feature dimensions, issues such as information loss and limited feature fusion capability arise. To address these challenges, we introduce a more efficient FasterNet-EMA Block (FEB) for substituting the Bottleneck within C2f, thereby designing the novel FE-C2f module. By employing FE-C2f in the Backbone section of YOLOv8, replacing the traditional C2f, the model's detection accuracy for multiple targets is significantly improved, and its performance in challenging detection scenarios is enhanced. In Figure 6, the new FE-C2f module is presented.

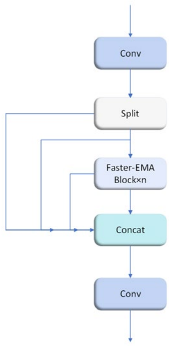


Figure 6.The The FE-C2f Module.

3.2. ML-C2f

This paper introduces a multi-scale lightweight convolution module, MLConv, designed using GhostConv. Based on this, the C2f module is restructured to create the ML-C2f module.In the neck portion of YOLOv8, the ML-C2f replaces the traditional C2f, resulting in a lighter-weight model and a significant boost to its detection performance.

GhostConv is a lightweight convolutional structure that is specifically developed for reducing the computational and parameter burden, all the while keeping or enhancing the model performance. Figure 7 illustrates the architecture of GhostConv.

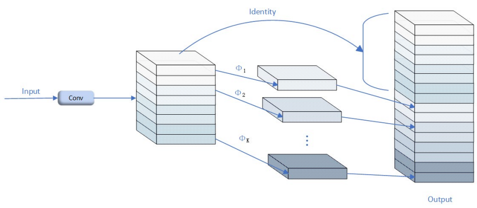


Figure 7.The structure of GhostConv.

The design methodology of MLConv consists of the following steps: Initially, the input channels are grouped into four separate branches, each receiving an equal division of the channels. Each branch employs a lightweight convolution, GhostConv, with varying kernel sizes—3x3, 5x5, 7x7, and 9x9—to extract features of foreign objects at different scales. The number of channels allocated to each branch is a quarter of the total input channels, ensuring that there is no interference between the branches.Subsequently, the diverse scale information extracted from the four channels is concatenated using the Concat operation, allowing all sub-features to be re-aggregated into the original number of

ISSN 2616-5767 Vol.8, Issue 1: 8-17, DOI: 10.25236/AJETS.2025.080102

input channels.Finally, the processed Group branches are added to the original input information and subjected to a 1x1 convolution operation. This integration enhances the fusion of processed feature information with the original data, augmenting the model's feature extraction capability across various scales, concurrently decreasing the number of the model's parameters. Figure 8 shows the MLConv structure.

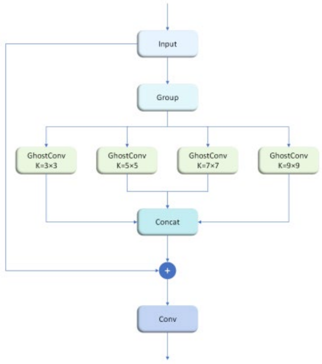


Figure 8.The structure of MLConv.

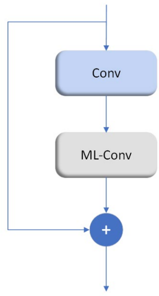


Figure 9.The structure of ML-Bottleneck.

Then, utilizing the MLConv design, a novel ML-Bottleneck is developed to boost the network's ability in feature extraction for multi-scale tasks. Figure 9 illustrates the structure of the ML-Bottleneck.

Next, by replacing the Bottleneck with ML-Bottleneck, the ML-C2f is obtained.In the neck portion of YOLOv8, the traditional C2f is replaced with ML-C2f to enhance the performance of YOLOv8 in object detection tasks. ML-C2f not only retains the advantages of C2f, such as feature branching and feature fusion, but also enhances the multi-scale feature extraction capabilities and feature fusion abilities through the introduction of multi-scale features. It demonstrates excellent performance in real-time detection and processing tasks involving targets of varying scales. Figure 10 illustrates the architecture of ML-C2f.

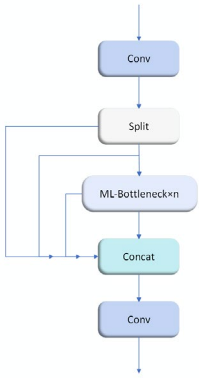


Figure 10.The structure of ML-C2f.

3.3. Small Object Detection Head

Owing to the considerable variation in shooting angles and distances during drone operations, the size of target objects can change dramatically, leading to many small objects that are prone to being missed during detection. In order to tackle this challenge, our research introduces an additional detection head specifically designed for small object detection, which works in conjunction with the other three detection heads to improve overall detection performance.

4. Experimental Results and Discussion

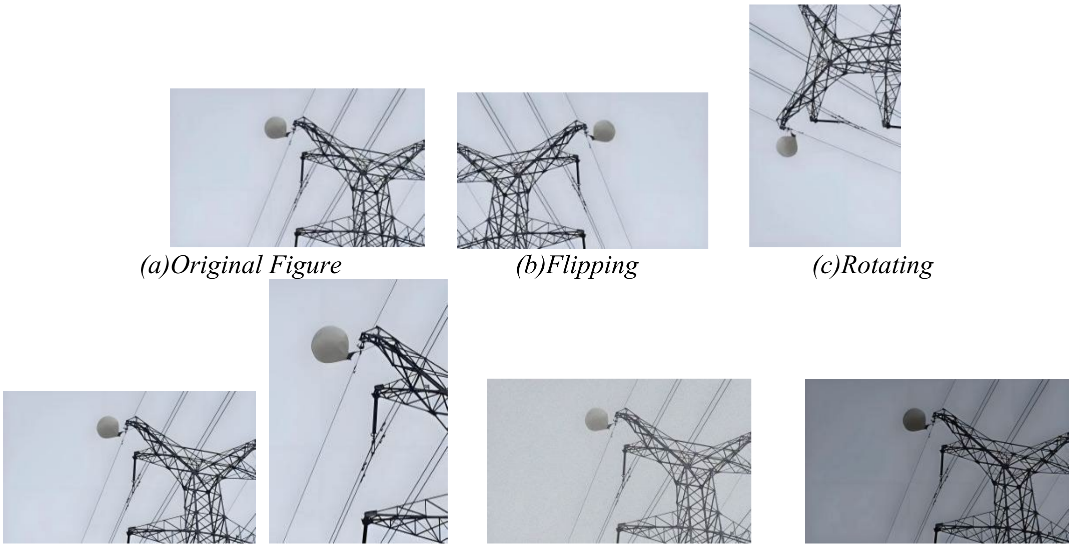
4.1. Dataset Construction

Given the unique characteristics of the power industry, acquiring data samples presents substantial challenges, and the quality of these samples varies significantly. At the current stage, there are few

ISSN 2616-5767 Vol.8, Issue 1: 8-17, DOI: 10.25236/AJETS.2025.080102

open-source datasets available for foreign bodies on transmission lines. Our research has compiled an open-source dataset by collecting 3,537 images as the initial dataset, which includes 886 images of bird nests, 409 of bee hives, 357 of tree branches, 813 of kites, 783 of balloons, and 289 of trash.

To avoid the issues of underfitting or overfitting due to insufficient sample sizes, which could impair the model's generalizability and subsequently affect detection performance.Therefore, we have implemented data augmentation techniques to expand the dataset, including flipping, rotating, scaling, cropping, introducing Gaussian noise, and adjusting brightness.The various data augmentation techniques are depicted in Figure 11.



(a)Original Figure *(b)Flipping* *(c)Rotating*

(d)Scaling *(e)Cropping* *(f)Introducing Gaussian noise* *(g)Adjusting brightness*

Figure 11.Data Augmentation Techniques.

Based on the initial dataset, a range of data augmentation techniques were performed, expanding the original 3,537 images to a final dataset of 7,812 images. This includes 1,963 images of bird nests, 859 of bee hives, 721 of tree branches, 1,705 of kites, 1,926 of balloons, and 638 of trash. This collection has been named the Transmission Line Foreign Objects Dataset,as depicted in Figure 13.

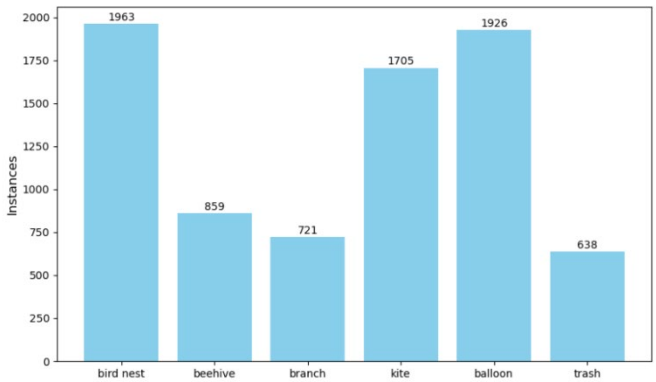


Figure 12.Transmission Line Foreign Objects Dataset

4.2. Evaluation Index

For the purpose of evaluating the effectiveness of FMH-YOLO, the present study employs the following metrics as evaluation criteria.

P: P stands for the proportion of true positive samples within the collection of all samples predicted as positive by the model, demonstrating the model's precision in predicting positive classes. The formula is as follows:

ISSN 2616-5767 Vol.8, Issue 1: 8-17, DOI: 10.25236/AJETS.2025.080102

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

R: The fraction of actual positive samples true predicted as positive by the model is represented by R, which reflects the model's ability to precisely recognize positive instances. The formula is as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

TP: It stands for the quantity of samples which the model exactly classifies as positive, indicating that the prediction results are consistent with the true labels, both being positive. FP: It symbolizes the count of samples that the model erroneously classifies as positive, where these samples are truly negative but are erroneously labeled as positive. FN: It indicates the count of samples that the model wrongly classifies as negative. These samples are actually positive but are erroneously labeled as negative[8].

AP: The value of AP is computed based on the area beneath the Precision-Recall curve, namely AUC. This metric offers a thorough evaluation of the model's proficiency in classification. The formula is as follows:

$$\text{AP} = \int_0^1 \text{P(R)dR} \tag{5}$$

mAP: mAP is obtained by averaging the AP values corresponding to every category. It functions as a means to assess the holistic efficacy of the model throughout the detection assignment. The following is the formula:

$$\text{mAP} = \frac{1}{\text{N}} \sum_{\text{i}=1}^{N} AP_i \tag{6}$$

4.3. Experimental Results

For the purpose of confirming the effectiveness of the algorithm enhancements put forward within this paper, we conducted experiments using FMH-YOLO on the Foreign Objects Dataset of Transmission Line. In Table 1 below, the detection outcomes of all categories are presented.

Table 1: Experimental results.

Categories	P(%)	R(%)	AP50(%)	mAP@0.5(%)
bird nest	93.6	89.4	92.6	87.1
beehive	87.5	84.1	89.2	82.3
branch	88.1	88.5	93.4	86.2
kite	91.8	92.8	95.6	90.3
balloon	89.8	86.7	90.8	85.4
trash	78.5	78.6	85.5	68.5
All	87.4	85.8	90.6	81.4

4.4. Ablation Experiments

With the aim of gauging the effect of each improvement module within FMH-YOLO, we carried out several ablation experiments on the Foreign Objects Dataset of Transmission Line. Table 2 displays the findings:

Table 2: Ablation experiments.

Experiment number	FE-C2F	ML-C2F	Head	AP50(%)
1				87.7
2	✓			88.5
3	✓	✓		89.1
4	✓	✓	✓	90.6

YOLOv8-n was selected as the baseline model. In the first experiment, the baseline YOLOv8-n model was used. In the second experiment, the FE-C2F module was added to the baseline model. The third experiment introduced the ML-C2F module, while the fourth experiment integrated the small

ISSN 2616-5767 Vol.8, Issue 1: 8-17, DOI: 10.25236/AJETS.2025.080102

object detection head. The ablation study results indicate that the baseline model achieved a mean Pixel Accuracy (mPA) of 87.7%. After incorporating the FE-C2F, ML-C2F, and the small object detection head, the mPA of the improved FMH-YOLO reached 90.6%, representing an improvement of 2.9% over the baseline model. This demonstrates the enhanced detection capability of the proposed model.

4.5. Comparative Experiments

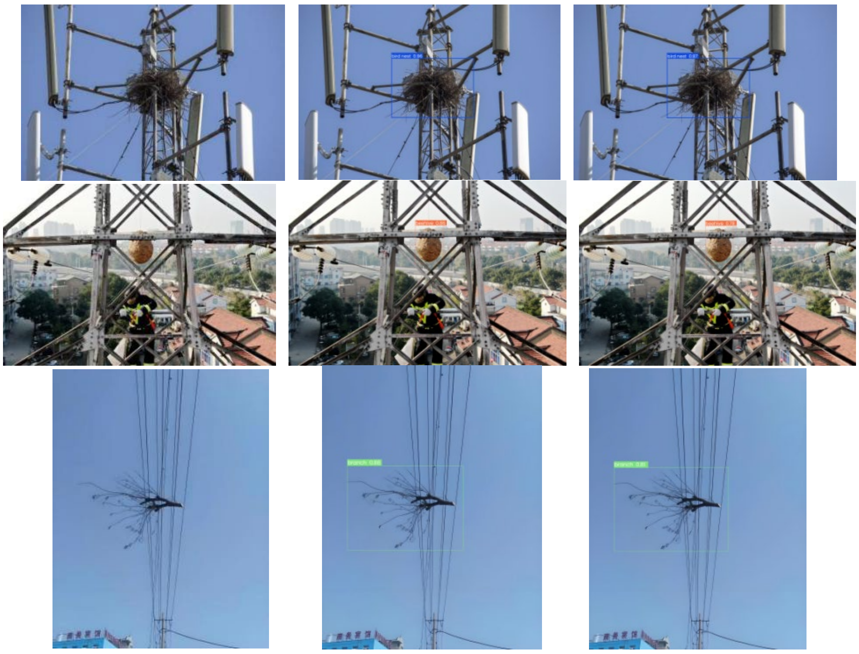
In order to verify the performance of FMH-YOLO, this research carried out comparative experiments with several classical algorithm models, such as Faster R-CNN, SSD, YOLOv3, YOLOv5-n, YOLOv7-tiny and YOLOv8-n. Table 3 displays the results of these experiments.

Table 3: Comparative experiments.

Modles	Precision/%	Recall/%	AP50/%	mAP@0.5/%
Faster R-CNN	76.7	71.1	81.9	65.4
SSD	72.4	68.7	77.3	59.7
YOLOv3	78.8	70.2	80.6	68.4
YOLOv5-n	83.2	80.8	85.6	72.7
YOLOv7-tiny	78.6	75.7	82.1	73.6
YOLOv8-n	84.7	83.6	87.7	79.6
FMH-YOLO	87.4	85.8	90.6	81.4

As shown in the table above, compared to other classical models, our improved model, FMH-YOLO, achieves the highest values in evaluation metrics such as Precision, Recall, AP50, and mAP@0.5.Faster R-CNN, as a two-stage object detection algorithm, is characterized by its complex network architecture, slow inference speed, and high consumption of memory and computational resources. Due to these factors and its inferior detection performance, it is not suitable for real-time object detection.Given its relatively outdated architecture, the SSD exhibits low detection performance.Compared to YOLOv3, the FMH-YOLO shows significant improvements in all metrics.Due to its lightweight architecture, YOLOv7-Tiny does not achieve good detection performance.Compared to YOLOv5, FMH-YOLO demonstrates a significant enhancement in detection capabilities.Compared to the baseline model YOLOv8-n, FMH-YOLO exhibits significant improvements across all metrics, demonstrating the effectiveness of our method.

4.6. Model Inference Outcomes



(a) Original Images (b) FMH-YOLOv8 (c) YOLOv8-n

Figure 13. Inference outcomes of diverse models regarding the dataset. (a) Original images ; (b) Inference outcomes of FMH-YOLOv8; (c) Inference outcomes of YOLOv8-n.

ISSN 2616-5767 Vol.8, Issue 1: 8-17, DOI: 10.25236/AJETS.2025.080102

For the purpose of visually presenting the detection capabilities of our approach, we conducted comparative experiments using YOLOv8-n and FMH-YOLO to validate the detection performance for foreign objects on transmission lines. The results obtained from the experiment are shown within Figure 15. The figure clearly demonstrates that, compared to YOLOv8-n, FMH-YOLO exhibits superior detection performance.

5. Conclusions

For the task of foreign object detection on power transmission lines in drone application scenarios, challenges such as complex backgrounds, target occlusion, significant variations in target scale, and poor performance of existing models persist. This paper proposes FMH-YOLO to address these issues. Firstly, by introducing Partial Convolution and the EMA module, the Faster-EMA Block (FEB) is constructed as a novel lightweight module. Based on FEB, the C2f module is restructured into FE-C2f, which is applied in the Backbone to improve performance while reducing resource consumption. Secondly, through utilizing GhostConv to build a multi-scale lightweight convolution module, the C2f module is further reorganized into ML-C2f. Subsequently, ML-C2f is integrated into the Neck part, thereby remarkably strengthening the model's ability to extract multi-scale features. Lastly, an additional detection head is incorporated into the Head section to detect small objects. Experimental results demonstrate that compared to YOLOv8-n, FMH-YOLO achieves improvements of 2.7%, 2.2%, 2.9%, and 1.8% in Precision, Recall, AP50, and mAP@0.5, respectively, highlighting its superior performance in the task of foreign object detection for power transmission lines.

In future work, we aim to balance detection performance with resource consumption to further optimize the network model. Additionally, we anticipate extending the proposed network model to other object detection applications, thereby addressing challenges in various industries.

References

[1] Zhang W, Liu X, Yuan J, et al. RCNN-based foreign object detection for securing power transmission lines (RCNN4SPTL)[J]. Procedia computer science, 2019, 147: 331-337.

[2] Wang G, Chen Y, An P, et al. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios[J]. Sensors, 2023, 23(16): 7190.

[3] Hao J ,Yan G ,Wang L , et al.A Lightweight Transmission Line Foreign Object Detection Algorithm Incorporating Adaptive Weight Pooling[J].Electronics,2024,13(23):4645-4645.

[4] Yang G, Wang J, Nie Z, et al. A lightweight YOLOv8 tomato detection algorithm combining feature enhancement and attention[J]. Agronomy, 2023, 13(7): 1824.

[5] Liu G, Reda F A, Shih K J, et al. Image inpainting for irregular holes using partial convolutions[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 85-100.

[6] Ouyang D, He S, Zhang G, et al. Efficient multi-scale attention module with cross-spatial learning[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.

[7] Chen J, Kao S, He H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 12021-12031.

[8] Terven J, Córdova-Esparza D M, Romero-González J A. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas[J]. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680-1716.