Advances in Computer, Signals and Systems (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2026.100201
ISSN 2371-8838 Vol. 10 Num. 2

Research on Foreign Object Detection Algorithm for High-Speed Railway Catenary Based on Improved YOLOv8n

Jian Sun, Zhenhua Wang, Tianijng Zhang, Siqi Chen, Rui Wang

China Railway Xi'an Group Co., Ltd., Xi'an, 710000, China

Keywords: Foreign Objects in High-speed Railway Catenary, Deep Learning, YOLOV8n, Characteristic Pyramids, Multi-Scale Attention

Abstract: With the rapid development of industrialization, the demand for railway transportation has been increasing. In this context, the detection of foreign objects in railway catenary systems remains a significant challenge to ensuring the safe operation of railways. To address the issues of low accuracy and poor real-time performance in detecting foreign objects in railway catenary systems, this paper proposes a foreign object detection algorithm based on the AFPN-YOLOv8n deep learning model. To fundamentally improve detection accuracy, the algorithm introduces a feature pyramid network (AFPN) in the Head module, effectively integrating low-level detail features with high-level semantic features, thereby enhancing the network's ability to detect foreign objects in high-speed railway catenary images. Additionally, an efficient multi-scale attention (EMA) module is added after the C3 layer in the Backbone module, further improving the network's ability to extract features of foreign objects. Experimental results show that compared to the original YOLOv8n model, the proposed model achieves an increase of 8.3% in mean average precision (mAP) at IOU=0.5, reaching 0.957, with a detection speed of 1.5 FPS. This provides a new approach and method for detecting foreign objects in railway catenary systems.

1. Introduction

In recent years, with the rapid development of China's railway network and the continuous expansion of its operational mileage, safety concerns regarding high-speed railway catenary systems have become increasingly prominent. Among these issues, frequent safety incidents are caused by avian activities near railway tracks and lightweight foreign objects (such as kites, plastic bags, agricultural film, and balloons) interfering with the contact lines. To mitigate the adverse effects caused by bird nests, avian perching, and the accumulation of lightweight debris on catenary wires, current detection and identification of foreign objects in contact line areas primarily rely on manual analysis of surveillance video images for judgment and annotation. This human-dependent image screening approach not only involves heavy workloads and high rates of missed detection but also proves inefficient. Therefore, improving the detection efficiency of avian nests on catenary systems and timely hazard elimination hold significant importance for ensuring railway operational safety.

With the continuous advancement of deep learning technologies, numerous object detection algorithms have been successively proposed, accompanied by significant improvements in image-based detection performance. These developments have enabled widespread applications in overhead catenary foreign object detection. This non-contact automated detection methodology demonstrates particular efficacy in identifying avian nests and lightweight debris (e.g., plastic films, balloons), thereby substantially enhancing inspection efficiency and presenting considerable application potential. Current deep learning-based object detection algorithms can be categorized into two paradigms: two-stage and one-stage detection models. Two-stage architectures (exemplified by Mask R-CNN and Faster R-CNN) operate through region proposal generation followed by classification and bounding box refinement, achieving superior detection accuracy with reduced false positives and false negatives, albeit at the cost of computational efficiency due to their inherent complexity. In contrast, one-stage models (such as YOLO and SSD variants) integrate feature extraction with bounding box regression in an end-to-end manner, enabling real-time inference speeds while typically exhibiting marginally lower precision metrics.

Ren Zhijun et al. [1] proposed an enhanced Mask R-CNN framework incorporating modified feature pyramid networks (FPN), which demonstrated improved mean average precision (mAP) for both target boundary segmentation and bounding box detection. Yin et al. [2] addressed the prevalent issues of incomplete feature extraction and low classification accuracy in existing image detection methods by developing an optimized Faster R-CNN variant. Their approach minimizes pooling artifacts while enhancing classifier performance through diversified feature representations, resulting in more discriminative and robust network characteristics.

Sun et al. [3] proposed Multi-YOLOv8, an efficient multi-input approach that enhances detection performance through improved optical flow computation methods, optimized bounding box regression loss functions, and the incorporation of BiFormer modules with lightweight GSConv convolutional techniques. Zhou et al. [4] improved the generalization capability and stability of bird's nest recognition by fine-tuning a pre-trained DSSD network. Wang Xiaohong et al. [5] developed the ESA4-YOLOV5s network model for monitoring foreign objects in bird nests, which utilizes EfficientNet-B4 as the feature extraction backbone and employs SUBA architecture for feature fusion, thereby improving both detection speed and accuracy.Jiang Xinlan et al. [6] first identified potential regions of interest (ROIs) where bird nests might appear using the LSD line segment detection algorithm, then employed YOLOv3 to detect nests within these ROIs. Wang Qian et al. [7] presented an enhanced YOLOv5-based method for detecting foreign objects in railway catenary systems, incorporating convolutional attention modules, depthwise separable convolutions, and an improved SPPF module to accelerate processing. Wang Keli et al. [8] proposed a deep learning-based approach for bird's nest detection, conducting a comprehensive comparative analysis between YOLOv3 and Faster R-CNN, evaluating metrics including accuracy, false-positive rate, false-negative rate, and detection speed. Experimental results demonstrated that the Faster R-CNN model achieved superior performance in nest detection accuracy, lower false-positive and false-negative rates, though with slower detection speed compared to YOLOv3 algorithm.

He Deqiang et al. [9] developed an automated bird's nest detection system for catenary systems using a deep convolutional neural network based on Faster R-CNN. Their approach involved training a Fast R-CNN network followed by joint training of the Region Proposal Network (RPN) and Fast R-CNN, establishing an optimized Faster R-CNN architecture specifically for avian nest detection. Wang Jiwu et al. [10] proposed an enhanced Faster R-CNN method for railway catenary bird nest identification, employing VGG16 as the backbone network and utilizing sliding windows across multi-scale convolutional feature maps to capture targets while improving feature map resolution. Yang Pei et al. [11] introduced a Dual Discriminator Generative Adversarial Network (DDGANs) that achieved satisfactory classification performance in catenary bird nest detection. Wu et al. [12] presented an improved detection and extraction method based on Faster R-CNN, which incorporated three key enhancements: (1) anchor box parameter optimization using K-means clustering, (2) implementation of Scaled Exponential Linear Unit (SELU) activation functions, and (3) refinement of the ResNet-34 backbone network, collectively significantly boosting the model's recognition capability.

To address the challenges of low detection accuracy in catenary foreign object detection while meeting the stringent real-time requirements of railway operations, this study selects the relatively lightweight YOLOv8n model from the YOLOv8 series - known for its favorable balance between detection speed and accuracy - as the base network for improvement, subsequently applying it to catenary foreign object detection. The proposed AFPN-YOLOv8n algorithm achieves an enhanced trade-off between real-time performance and detection accuracy, enabling more precise feature extraction of foreign objects while maintaining satisfactory real-time detection capabilities.

2. Methodology

The YOLOv8n architecture primarily consists of four components: Input, Backbone, Neck, and Head networks. The Backbone module serves for feature extraction while the Head network performs detection. Recognized for its high detection accuracy and fast inference speed, this architecture has found widespread applications across various scenarios. This paper proposes an improved AFPN-YOLOv8n algorithm with modified Backbone and Head structures as illustrated in Figure 1. The enhancements include: (1) incorporation of an Asymptotic Feature Pyramid Network (AFPN) in the Head module to boost detection capability, and (2) addition of an Efficient Multiscale Attention (EMA) module after the C3 layer in the Backbone to emphasize target-specific key information in images and strengthen feature extraction capacity.

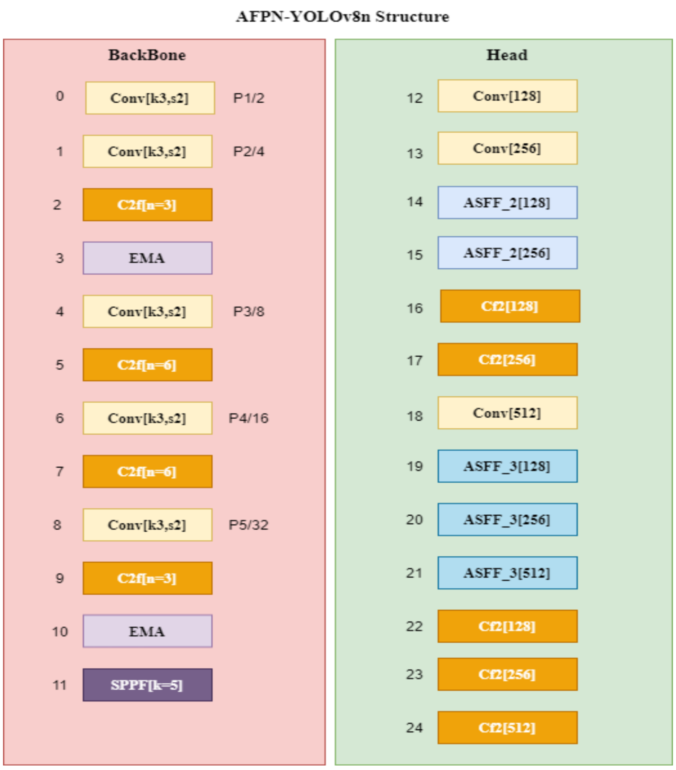

Figure 1. Architecture of the AFPN-YOLOv8n Model

2.1 Cross-space Efficient Multi-scale Attention Mechanism (EMA)

The Efficient Multiscale Attention (EMA) mechanism represents a novel and effective approach for enhanced feature extraction in visual data. This module demonstrates compatibility with various object detection architectures, including YOLOv5 and YOLOv8. Comparative studies reveal that the multiscale attention mechanism yields superior performance improvements over conventional attention methods. Specifically, EMA incorporates cross-space learning to integrate attention maps from parallel subnetworks. Through matrix dot-product operations, it effectively captures pixel-level pairwise relationships while emphasizing global contextual information across all pixels, as illustrated in Figure 2.

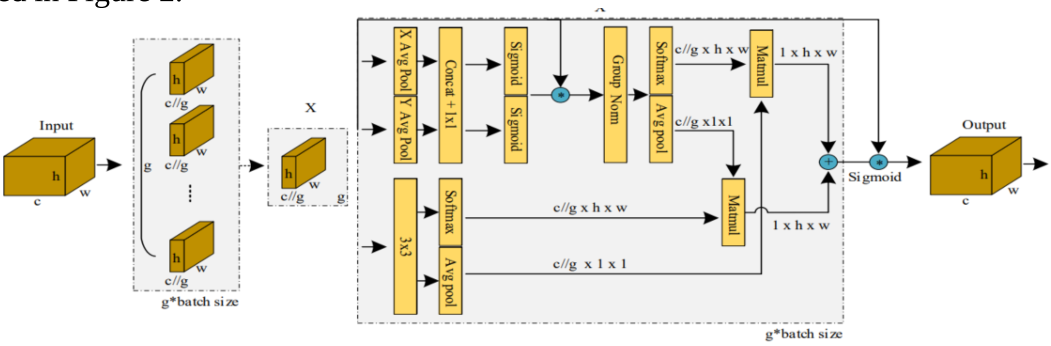

Figure 2. Architecture of the EMA

In this study, we integrate the EMA (Efficient Multiscale Attention) mechanism after the C3 layer in the Backbone network of YOLOv8n. This enhancement establishes feature mapping relationships for target detection and performs attention-based feature map reconstruction. By assigning higher feature weights to targets, the modified architecture significantly improves both detection accuracy and recognition performance for catenary foreign objects.

2.2 Asymptotic Feature Pyramid Network (AFPN)

YOLO (You Only Look Once) is a deep learning-based object detection algorithm that performs detection by dividing images into grids and predicting bounding boxes with class probabilities within each grid. However, since targets of different scales may exhibit distinct features across various hierarchical levels of the feature maps, relying solely on single-level predictions may compromise detection accuracy. To address this limitation, this study introduces an Asymptotic Feature Pyramid Network (AFPN) architecture into YOLOv8, as illustrated in Figure 3.Specifically, the AFPN initiates feature fusion by combining two adjacent low-level features and progressively incorporates high-level features into the fusion process. This hierarchical approach effectively mitigates the substantial semantic gaps that typically exist between non-adjacent levels. Furthermore, to resolve potential multi-target information conflicts during feature fusion at each spatial location, an adaptive spatial fusion operation is implemented to alleviate these inconsistencies. Bird nests and lightweight foreign objects present unique challenges, including indistinct contours, variable sizes, and potential information loss during dimensionality reduction - particularly for nascent bird nests which are often small in size. Therefore, we replace the original network structure of YOLOv8 with the proposed AFPN architecture to better accommodate these characteristics.

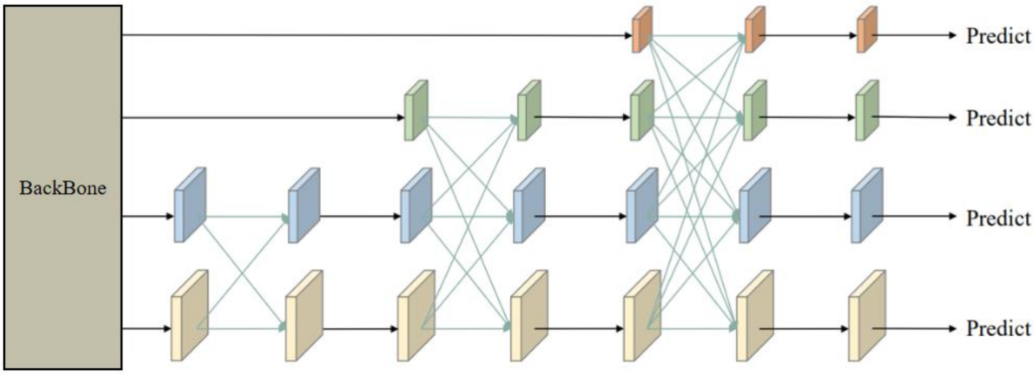

Figure 3. Architecture of the AFPN

2.3 Model Evaluation Metrics

For model performance evaluation, this study employs three key metrics: precision, recall, and mean average precision (mAP). Precision, also referred to as positive predictive value, represents the percentage of correctly predicted positive samples among all detected samples (true positives + false positives), reflecting the accuracy of prediction results. Recall, alternatively called sensitivity, measures the proportion of correctly identified positive samples relative to all actual positive samples (true positives + false negatives), indicating the model's coverage capability of the original dataset. The mean average precision (mAP) integrates both precision and recall to provide a comprehensive evaluation of global performance. The computational formulas for these metrics are detailed in Table 1.

Table 1. Calculation Methods of Model Evaluation Metrics

Evaluation Metrics	Expression	Description
Precision	$p = \dfrac{TP}{TP + FP}$	P denotes Precision. TP stands for True Positives, which are the samples that are correctly predicted as positive instances. FP stands for False Positives, which are the samples that are incorrectly predicted as positive.
Recall	$R = \dfrac{TP}{TP + FN}$	R denotes Recall. FN stands for False Negatives, which are the samples that are incorrectly predicted as negative instances.
mean Average Precision	$mAP = \dfrac{1}{N}\sum_{i=1}^{N}\int_0^1 P(R)d(R)$	N represents the number of object classes. mAP@0.5 indicates the mean average precision when the IoU threshold is set to 0.5, while mAP@0.5:0.95 denotes the mean average precision averaged over multiple IoU thresholds ranging from 0.5 to 0.95 with a step size of 0.05. A higher mAP value indicates more accurate bounding box localization.

3. Network Training and Testing

The proposed AFPN-YOLOv8n model for catenary foreign object detection was implemented using the PyTorch 1.10 deep learning framework. All experiments were conducted on a hardware platform equipped with an Intel® Core™ i7-1165G7 CPU (2.8 GHz) and an NVIDIA GeForce RTX 3090 GPU.

3.1 Dataset Collection and Preprocessing Methods

Due to the scarcity of public datasets for catenary foreign objects, we established a dedicated image dataset through three approaches: (1) Web crawling of high-speed railway catenary images followed by manual curation, (2) Collection of on-site photographs taken by railway maintenance personnel, and (3) Frame extraction from inspection vehicle videos. The compiled dataset comprises 1,442 high-quality images of catenary foreign objects, with representative samples shown in Figure 4.

To prevent overfitting and enhance model robustness and generalization capability, we employed multiple data augmentation techniques including random flipping, cropping, translation, scaling, Gaussian noise injection, and illumination correction. These operations expanded the dataset to 7,210 images. All augmented images were then annotated using LabelImg software with two defined classes: 0 for bird nests and 1 for lightweight foreign objects, formatted in YOLO-compatible label files.

Figure 4. Representative images of foreign objects in high-speed railway catenary systems

3.2 Model Training and Testing

The augmented dataset comprising 7,210 annotated images was randomly partitioned into training, validation, and test sets at an 8:1:1 ratio. To address the limitations of fixed learning rates in traditional stochastic gradient descent, we employed the Adam optimizer [13] for model training. Initial weights were pretrained on the COCO dataset (source: https://github.com/ultralytics/ultralytics) to accelerate convergence. The model was trained for 150 epochs with a batch size of 64 and a learning rate of 0.0001.Figure 5 illustrates the training dynamics, where the x-axis represents epoch numbers and the y-axis shows metric values. The first row displays the training set trends for box loss, classification loss (cls_loss), distribution focal loss (dfl_loss), precision (P), and recall (R). The second row presents validation set metrics including box loss, cls_loss, dfl_loss, mAP@50, and mAP@50:95. The curves demonstrate consistent reduction in loss values and improvement in evaluation metrics with increasing epochs, ultimately converging at P=0.956, R=0.912, mAP@50=0.957, and mAP@50:95=0.651.Figure 6 showcases the model's detection performance on the test set, demonstrating robust capability in both single-object and multi-object scenarios. Quantitative results indicate superior detection accuracy across all test cases.

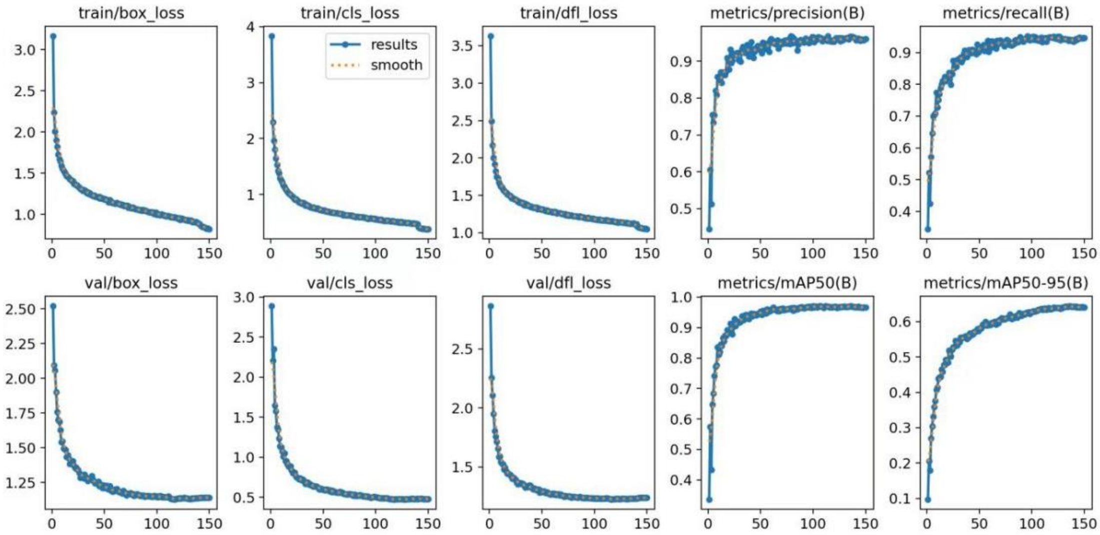

Figure 5. Training Process Metrics

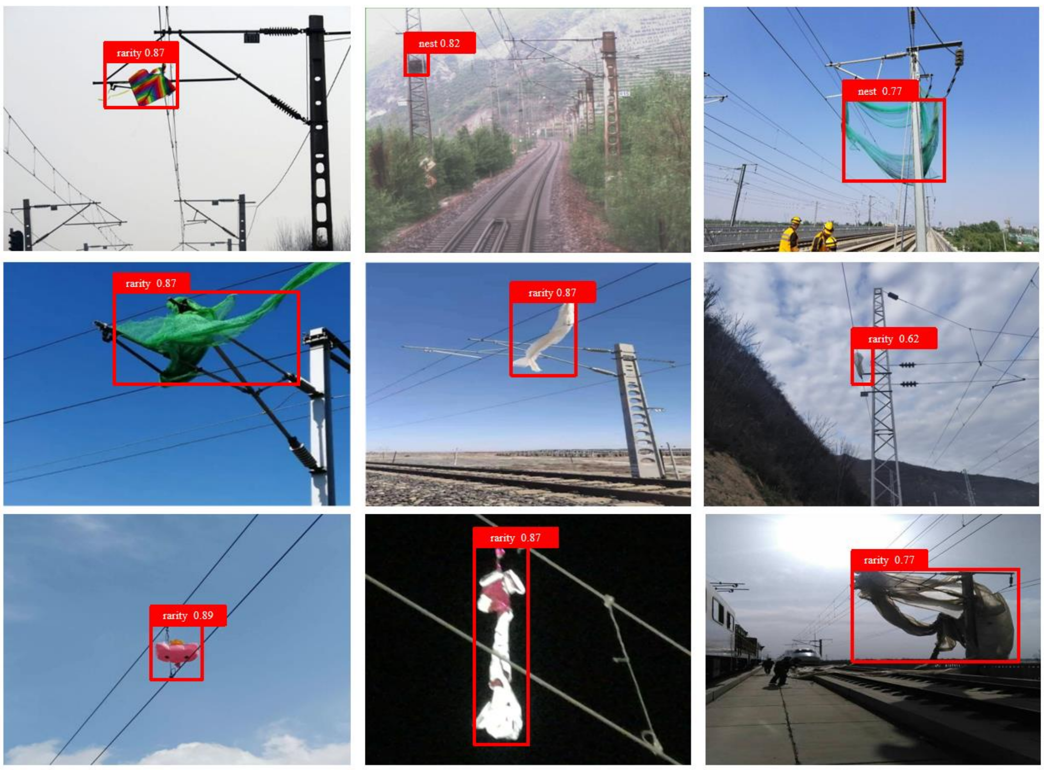

Figure 6. Foreign Object Detection Performance on Test Data

4. Comparative Experiments

This study conducts comprehensive comparisons between the improved YOLOv8n model and its conventional counterpart, followed by systematic ablation experiments to validate the efficacy of the proposed enhancements.

4.1 Comparison with Baseline YOLOv8n

This study validates the efficacy of the proposed AFPN-YOLOv8n model through both visual comparisons and quantitative metrics against the original YOLOv8n architecture, with both models trained under identical experimental conditions.

(1) Visual Comparison

Figure 7 presents side-by-side detection results on identical test images, contrasting the performance between the baseline YOLOv8n and our improved model. The comparative visualization demonstrates enhanced detection accuracy and reduced false positives in challenging scenarios.

Figure 7. Qualitative Results: Original YOLOv8n (Left) vs. Our Method (Right)

As evidenced in Figure 7, the baseline YOLOv8n fails to detect small-scale foreign objects, whereas our proposed model demonstrates superior capability in both small-target identification and precise localization. Quantitative confidence scores further confirm the enhanced reliability of our approach compared to the original architecture.

(2)Quantitative Metrics

Table 2 presents the comprehensive performance metrics, revealing statistically significant improvements across all evaluation criteria.

Table 2. Algorithm Performance Metrics

Model	Precision	Recall	mAP@0.5	mAP@0.5:0.95
YOLOv8	0.863	0.844	0.892	0.568
Our model	0.956	0.912	0.957	0.651

As quantitatively demonstrated in Table 2, for foreign object detection in railway catenary systems, the detection precision improved from the original 86.3% to 95.6% in our modified model - a significant 9.3 percentage point increase. The mean average precision (mAP) also showed an 8.3% improvement over the baseline model. These results confirm that our model enhancements effectively achieve higher detection accuracy, successfully fulfilling the primary objective of precision improvement.

Compared to the typical YOLOv8n model, the improved YOLOv8n model demonstrates a significant enhancement in detection accuracy. Additionally, during the detection process, both the missed detection rate and false detection rate of the model are notably reduced. This indicates that, in various complex scenarios, the improved YOLOv8n model can significantly enhance the accuracy of target detection for classifying foreign objects.

4.2 Ablation Experiment

To further verify the accuracy of the proposed algorithm for detecting foreign objects in railway catenary systems, a comparison of the improved model's performance was conducted to evaluate whether the detection accuracy has been enhanced. An ablation experiment was designed and performed on the catenary foreign object dataset created in this study. The "+" sign indicates that the corresponding improvement method was applied, and each experiment was conducted three times to obtain the respective results. The experimental results are shown in Table 3.

Table 3.Algorithm Performance Metrics

Model	AFPN	EMA	Precision	Recall	mAP@0.5	mAP@0.5:0.95
YOLOv8			0.863	0.844	0.892	0.568
Improvement 1	+		0.904	0.887	0.915	0.601
Improvement 2		+	0.913	0.896	0.926	0.634
Our model	+	+	0.956	0.912	0.957	0.651

As shown in Table 3, the original YOLOv8 model achieves an mAP of 89.2%. Based on the original YOLOv8 model, Improvement 1 introduces an Advanced Feature Pyramid Network (AFPN), resulting in a 2.3% increase in mAP. This indicates that the addition of the pyramid module enables feature fusion, enriching the local feature information and thereby improving the accuracy of foreign object detection in the catenary system. On the basis of the original YOLOv8 model, Improvement 2 incorporates a cross-space efficient multi-scale attention mechanism (EMA), leading to a 3.4% increase in mAP. This demonstrates that the EMA module enhances the network's ability to extract features of foreign objects from images, thereby improving detection accuracy.

The model in this paper, based on the original YOLOv8 model, simultaneously incorporates both Improvement 1 and Improvement 2, resulting in a 6.5% increase in mAP. This significant enhancement indicates that during the improvement of the YOLOv8 model, the combination of AFPN and EMA markedly boosts the network's capability to detect foreign objects.

From the above ablation experiments, it can be seen that in the detection process of foreign objects in railway catenary systems, the improved YOLOv8 model used in this paper achieves an average detection accuracy of 95.7%, which represents a significant improvement compared to the original YOLOv8 model. Therefore, the proposed model meets the practical requirements for detecting foreign objects in railway catenary systems.

5. Conclusion

To address the issues of low detection accuracy and weak real-time performance in foreign object detection for high-speed railway catenary systems, this paper proposes the AFPN-YOLOv8n algorithm. This algorithm employs YOLOv8n, the lightweight version of the YOLOv8 series, as the base network. A feature pyramid network (AFPN) is introduced into the Head module to enhance the network's ability to detect foreign objects in high-speed railway catenary images. To further improve detection accuracy, an efficient multi-scale attention mechanism (EMA) is added after the C3 layer in the Backbone module. Through comparative experiments and ablation studies, the effectiveness of the proposed algorithm has been verified. This method provides a reliable basis for the automatic detection of bird nests on railway catenary systems.

References

[1] Ren Zhijun, Lin Suzhen, Li Dawei, et al. Mask R-CNN Object Detection Method Based on Improved Feature Pyramid[J].Laser & Optoelectronics Progress, 2019, 56(4):041502.DOI:10.3788/LOP56.041502.

[2] Yin X , Chen L .Image Object Detection Method Based on Improved Faster R-CNN[J].Journal of Circuits, Systems and Computers,2024, 33(07).DOI:10.1142/S0218126624501305.

[3] Sun S , Mo B , Xu J ,et al.Multi-YOLOv8: An infrared moving small object detection model based on YOLOv8 for air vehicle[J].Neurocomputing,2024, 588.DOI:10.1016/j.neucom.2024.127685.

[4] Zhou Jun, Chen Jianyun. Research on Catenary Bird's Nest Recognition and Detection Based on DSSD [J]. Journal of East China Jiaotong University, 2019, 36(6):9.DOI:CNKI:SUN:HDJT.0.2019-06-011.

[5] Wang Xiaohong, Du Yunfei, Liu Chang. Detection of Bird Nests and Foreign Objects in Catenary Systems Based on YOLOv5s [J]. Journal of Yangtze Information and Communication, 2023, 36(6):51-54.

[6] Jiang Xinlan, Jia Wenbo. Machine Vision Detection Methods for Foreign Object Intrusion in High-Speed Railway Catenary Systems [J]. Computer Engineering and Applications, 2019.

[7] Wang Qian. Research on Foreign Object Detection in Railway Catenary Systems Based on Improved YOLOv5 [J]. Academic Research of Xi'an Jiaotong University Institute of Technology, 2023, 8(3):49-53.

[8] Wang Keli, Gao Fulai, Yang Peng, et al. Research on Bird Nest and Foreign Object Recognition in Catenary Systems Based on Deep Learning [J]. Railway Locomotives and Vehicles, 2022(002):042.

[9] He Deqiang, Jiang Zhou, Chen Jiyong, et al. Research on Detection Methods for Bird Nests in Railway Catenary Systems Based on Deep Convolutional Neural Networks [J]. Electric Drive for Locomotives, 2019(4):5.DOI: CNKI:SUN:JCDC.0.2019-04-030.

[10] Wang Jiwu, Luo Haibao, Yu Pengfei, et al. Detection of Bird Nests in Railway Catenary Systems Based on Faster R-CNN [J]. Railway Locomotives and Vehicles, 2020, 40(2):5.DOI:CNKI:SUN:TDJC.0.2020-02-019.

[11] Yang Pei. Research on Dual-Discriminator Generative Adversarial Networks and Their Application in Detecting Bird Nests in Catenary Systems [D]. Southwest Jiaotong University,[2024-07-20].

[12] Wu C , He X , Wu Y .An object detection method for catenary component images based on improved Faster R-CNN[J].IOP Publishing Ltd, 2024.DOI:10.1088/1361-6501/ad4c01.

[13] Kingma D , Ba J .Adam: A Method for Stochastic Optimization[J].Computer Science, 2014.DOI:10.48550/arXiv. 1412.6980.