ISSN 2616-5775 Vol. 8, Issue 2: 83-89, DOI: 10.25236/AJCIS.2025.080211

Research on Remote Sensing Small Target Detection Method Based on Lightweight YOLOv8

Jingtao Zhu*

School of Computer Science and Technology, Zhengzhou University of Light Industry, Zhengzhou, 450000, China
**Corresponding author: 542207280229@zzuli.edu.cn*

Abstract: *With the evolution of remote sensing technology, accurately detecting small targets in remote sensing photos has become a critical challenge in domains such as economics and defense. This study addresses the difficulty of finding small targets in remote sensing photos by offering a lightweight YOLOv8-based detection algorithm. By integrating GhostNet as the backbone network of YOLOv8, the model has been lightweighted, enhancing its operational efficiency on resource-constrained devices. Experiments with the NWPU VHR-10 dataset demonstrate that YOLOv8s-ghost has a 9% decrease in mAP but a 37.5% loss in GFLOPS when compared to YOLOv8s. This indicates that the lightweight YOLOv8s ghost maintains detection accuracy while reducing computational costs, achieving the goal of model lightweighting. This research provides an efficient solution for small target detection in remote sensing photo, with significant practical value and broad application prospects.*

Keywords: *Remote Sensing Image, Small Object Detection, YOLOv8, Lightweight Network, GhostNet, Deep Learning*

1. Introduction

With the rapid development of remote sensing technology, the application of remote sensing images has become increasingly widespread, including but not limited to disaster casualty detection, urban and rural planning, traffic management, and military reconnaissance[1]. For these applications, the precision and real-time performance of analysis depend on the ability to accurately identify small targets in remote sensing images. However, because of their small size, low contrast, and vulnerability to background interference, small targets are challenging to detect in remote sensing photos. To increase the precision and real-time performance of remote sensing image processing, it is crucial to investigate effective small target recognition techniques.

In recent years, deep learning based object detection algorithms have demonstrated excellent performance[2], gradually replacing traditional methods. Detection algorithms such as YOLO, SSD, RetinaNet, and Faster RCNN typically have high detection accuracy, improving detection efficiency. The You Only Look Once algorithm (YOLO) has become the mainstream method for remote sensing image object detection due to its fast and accurate object detection capability. Weng Junhui et al.[3] introduced a small target detector and replaced the downsampling convolution module with the SPDConv module to achieve efficient detection of dense targets in aerial images, with an mAP0.5 improvement of 9.8% compared to the baseline model. Liang Tiantian et al.[4] optimized the C2F module in the backbone network using an expandable residual structure and redesigned the detection head of the YOLOv8s algorithm. The enhanced algorithm outperformed the original method by an average of 2.56% in terms of accuracy. Lei Bangjun et al.[5] added a new small target layer to enhance the model's sensitivity to small target scales, and added an attention mechanism to the detection head to improve the model's localization performance for small targets in various environments. Wang Xuanhui et al.[6] used the Faster module of the FasterNet network to replace the C2f modules of the backbone and neck in YOLOv8s, and proposed the Inner MPDIOU loss function to enhance the model's ability to handle details and increase the accuracy of its detection. Although the above algorithms have improved detection accuracy to some extent, they have also led to an increase in overall model parameters and computational complexity, which cannot solve the problems of limited performance stability and computing resources in embedded devices[7].

To overcome these challenges, this study introduced GhostNet proposed by Huawei Noah's Ark Laboratory as the backbone network of YOLOv8[8]. At the CVPR conference, Noah's Ark Laboratory showcased GhostNet, a new Ghost module that compresses network parameters and speeds up detection speed while producing more feature maps through low-cost operations. This scheme will introduce GhostNet as the backbone network of YOLOv8 to achieve model lightweighting and improve the running efficiency of the model on resource limited devices. Experiment to confirm the suggested method's efficacy and contrast it with other cutting-edge techniques to show how well this scheme performs in remote sensing small target detection.

This paper is organized as follows: In Chapter 2, the pertinent work and research background are introduced; in Chapter 3, the design and implementation of the lightweight YOLOv8 algorithm based on GhostNet are thoroughly described; in Chapter 4, the experimental results and analysis are presented; and in Chapter 5, the entire paper is summarized and future research directions are suggested.

2. YOLOv8 Network Architecture

YOLOv8 is a YOLO model released by Ultralytics, proposed based on YOLOv5 and incorporating the advantages of other YOLO frameworks[9]. The input layer, the backbone network, the neck network, and the detecting head are the four main parts of the YOLOv8 model's design, as seen in figure 1. To improve the model's ability to adjust to changes in the image, the input layer initially performs data augmentation to the input remote sensing image. The backbone network, which uses the CSPDarkNet structure and is in charge of extracting important information from the image, receives the improved image after that. Deep fusion of features taken from the backbone network is the responsibility of the neck network. Through feature fusion, this procedure improves the model's recognition capabilities for targets with varying scales and sends the fused features to the detecting head. The model's detection head serves as its decision center and is built with two parallel branches: one for category recognition and the other for target location prediction.

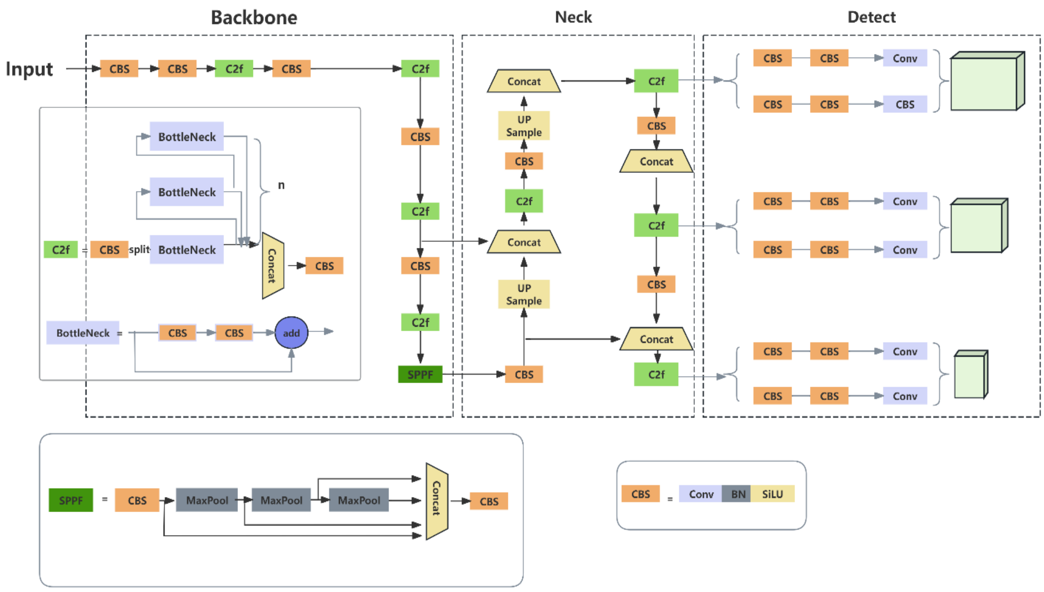

Figure 1 YOLOv8 Network Structure

2.1 The backbone network

YOLOv8 employs the C2f module in place of the C3 module used in YOLOv5. The C2f module utilizes the ELAN module to replace the CSP module of C3. The ELAN[10] module, introduced in YOLOv7, is a design concept based on the CSP module, equipped with more residual connections, offering richer gradient flow, and achieving further lightweighting of the network. The specific network structure diagrams of the C3 and C2f modules are shown in figure 2 and figure 3. C3 uses the CSPNet structure, dividing the input into two branches for operations, and finally performing a Concat operation. Before the branches of C3, there is no dimension reduction operation, ensuring that the input dimensions of each branch are equal to the initial input dimensions. C2f also adopts the CSPNet structure and uses more BottleNeck structures, followed by a Concat operation. Before branching in C2f, a dimension reduction operation is first conducted through convolution, and then the multi-branch structure of

ISSN 2616-5775 Vol. 8, Issue 2: 83-89, DOI: 10.25236/AJCIS.2025.080211

CSPNet is utilized, along with more residual structures for feature extraction. Although C2f has more branches, dimension reduction operations are performed before entering the branches, significantly reducing the computational load in subsequent calculations. This makes C2f not only computationally lighter than C3 but also enhances its feature extraction capabilities.

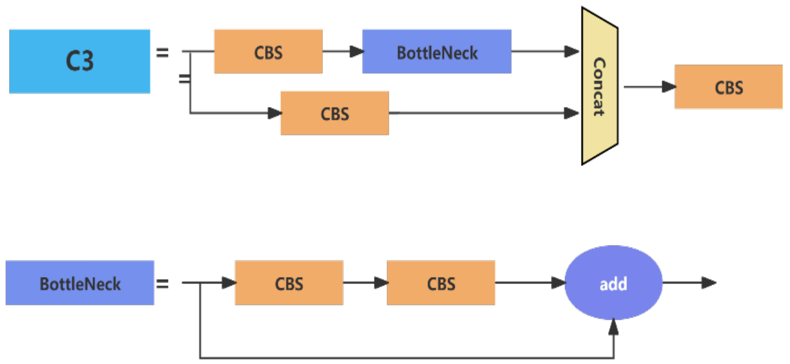

Figure 2 C3 module

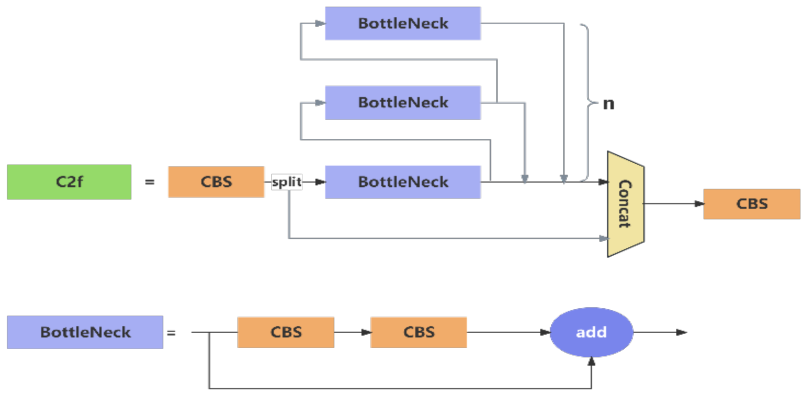

Figure 3 C2f module

2.2 The Detection Head

The detection head part of the YOLOv8 design has seen notable advancements. YOLOv8 separates the duties of object categorization and location using the decoupled head structure that is now popular. This structure allows the model to focus more on each sub-task, thereby enhancing the accuracy of classification and localization. Through decoupling, the model can more flexibly adjust the parameters for classification and localization to accommodate various detection scenarios and requirements. The conventional anchor-based detection technique has been dropped by YOLOv8 in favor of an anchor-free detection approach. This change implies that the model now forecasts the center points and dimensions of the objects directly rather than using predefined anchor boxes to determine their locations. This method improves the model's generalization and adaptation to objects of various scales by streamlining the prediction process and lowering reliance on anchor selection, as shown in Figure 4.

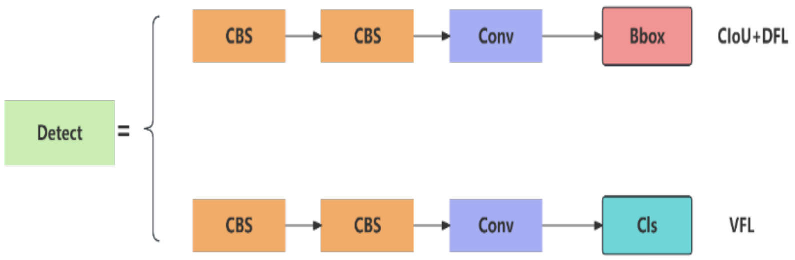

Figure 4 YOLOv8 detection head

3. Remote sensing image horizontal target detection algorithm combining YOLOv8 and GhostNet

3.1 Optimization of the Backbone Network: Incorporating GhostNet to Enhance YOLOv8 Performance

The backbone network of the YOLOv8 model employs CSPDarknet, an architecture that has demonstrated exceptional feature extraction performance and superior results on the ImageNet dataset. Nevertheless, even while CSPDarknet excels at feature extraction, the YOLOv8 model's overall detection performance is somewhat impacted by its computational efficiency. This is because, in the process of extracting deep-level features, CSPDarknet may involve more computational resources and time consumption, especially when dealing with remote sensing images. This computational burden could become a factor limiting the performance of YOLOv8 in real-time applications. Based on this, this chapter introduces the lightweight network GhostNet[11] to replace the original backbone network in the YOLOv8 network structure. The core design of GhostNet lies in identifying and utilizing redundancy and shared information between input feature maps, employing an efficient convolutional strategy: it first performs standard convolution operations on selected channels, then cleverly derives features from these convolutional results through linear transformations for other channels. This approach significantly reduces the computational burden of the backbone network. Through this optimization, the YOLOv8 model maintains its consistent high-precision detection capability while achieving faster detection speeds. By using a more effective feature fusion strategy, GhostNet not only lowers the computational cost of the model but also improves its capacity to extract important features. This enables YOLOv8 to more rapidly identify and locate targets when processing remote sensing images, thereby demonstrating outstanding performance in practical applications.

3.2 The Model Overall Structure

In this chapter, we propose innovative enhancements to the YOLOv8 model aimed at achieving a more efficient lightweight design. Our designed network retains the core architecture of YOLOv8, which includes the input layer, backbone network, feature extraction network, and detection head. For the construction of the backbone network, we employ the CSPDarknet structure, which serves as the cornerstone of YOLOv8. We replace conventional convolution procedures with Ghost convolution to lower the model's computational complexity. This method improves feature extraction effectiveness while drastically lowering the number of model parameters. Specifically, we replaced all modules after the first convolutional layer in the backbone network, transforming the standard convolutional module CBS (Conv-BatchNorm-SiLU) into a more efficient GBS (GhostConv-BatchNorm-SiLU). Furthermore, within the C2f module, we upgraded the BottleNeck structure to GBottleNeck, further enhancing the network's lightweight characteristics. These improvements collectively enable our model to maintain high-precision detection capabilities while achieving faster detection speeds. Our model offers an effective solution for horizontal target detection in remote sensing images, particularly in application scenarios requiring real-time feedback, where it exhibits superior performance. This not only increases the model's operational efficiency but also guarantees its accuracy and dependability in object detection tasks. Through these targeted optimizations, our YOLOv8 model has made significant progress in lightweight and efficiency, offering possibilities for deployment and use in a broader range of application fields.

4. Results

4.1 Experimental environment and data processing

To assess the enhanced lightweight algorithm's detection capabilities for tiny target remote sensing, the hardware environment employed in the experiment includes 12th Generation Intel Core ™ I7 processors, 16GB of memory, Nvidia GeForce RTX 3060 (6GB) graphics card, software environment for Windows 11 Professional operating system, and deep learning framework for Pytorch 2.5.1+cu124. The detection of the remote sensing picture target dataset used in the experiment is the publicly available horizontal detection dataset NWPU VHR-10[12] released by Northwestern Polytechnical University. The NWPU VHR-10 dataset is characterized by its rich category coverage, large area distribution, and a large number of annotated instances, making it highly valuable and challenging for testing and improving remote sensing image target detection algorithms. It functions as a sizable benchmark dataset for the detection of targets in remote sensing images. The NWPU VHR-10 dataset consists of ground truth, a collection of negative images, and a set of positive images. This experiment utilized 650 images from the positive image set, which contains 10 object categories with a total of 3,651 targets. These ten categories

ISSN 2616-5775 Vol. 8, Issue 2: 83-89, DOI: 10.25236/AJCIS.2025.080211

are airplanes, ships, tanks, baseball diamonds, tennis courts, basketball courts, ground runways, ports, bridges, and vehicles. These categories are quite common in remote sensing photos and have a broad variety of uses. The dataset was preprocessed to convert each annotated object in the images into the format required by the YOLOv8 model. This formatting operation ensures that the model can directly accept and process the preprocessed data. Subsequently, the NWPU VHR-10 dataset was divided into a training set(70% of the total)and a test set(30% of the total)ensuring that model training and evaluation are fair and effective.

4.2 Evaluation Metrics

To scientifically and thoroughly analyze the performance and efficiency of the proposed method on remote sensing picture target recognition tasks, two widely applied evaluation metrics were selected: Mean Average Precision (mAP) and Giga Floating-point Operations Per Second (GFLOPS). Mean Average Precision is an important metric for evaluating the performance of object identification algorithms. It is determined as the area under the curve based on the variance of precision and recall rates, using the precision and recall formulas described in Equations.

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} \tag{1}$$

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \tag{2}$$

In this case, TP stands for correctly detected true positives; FP for false positives, which are negative samples that are mistakenly labeled as positive; and FN for false negatives, which are positive samples that are mistakenly identified as negative. Precision on the P-R curve represents the model's ability to identify targets, or the proportion of targets properly classified as belonging to a specific category. Recall measures the comprehensiveness of the model in finding targets, that is, the proportion of all true targets found by the model to all actual existing targets. These two indicators contribute to the AP value, and the greater the AP value, the better the model's detection performance at various thresholds. The mean average precision is the average of AP values across all categories, representing the algorithm's overall detection performance across all categories. It is a well-known comprehensive evaluation statistic in the field of object detection. The greater the mAP number, the better the algorithm's overall performance at detecting targets across all categories. Considering the practical application requirements of the algorithm, the hardware performance metric of Giga Floating-point Operations Per Second (GFLOPS) was also introduced to gauge the computational requirements of the algorithm. The lower the GFLOPS value, the less computational resources the algorithm requires to complete the same task, and the faster the running speed, which is crucial for remote sensing image target detection tasks with high real-time requirements.

Strong data support for the algorithm's use in the field of remote sensing image target detection is provided by the use of mean average precision and floating-point operations as evaluation metrics. This allows for a thorough examination of the algorithm's performance in terms of recognition accuracy as well as an assessment of its viability for real deployment.

4.3 Improved Results Showcase

The YOLOv8s model serves as the foundation for the improvement experiment. Figure 5 displays the PR curves for each category of remote sensing picture for the enhanced YOLOv8s-ghost model. The precision rate is represented by the vertical axis, while the recall rate is represented by the horizontal axis.

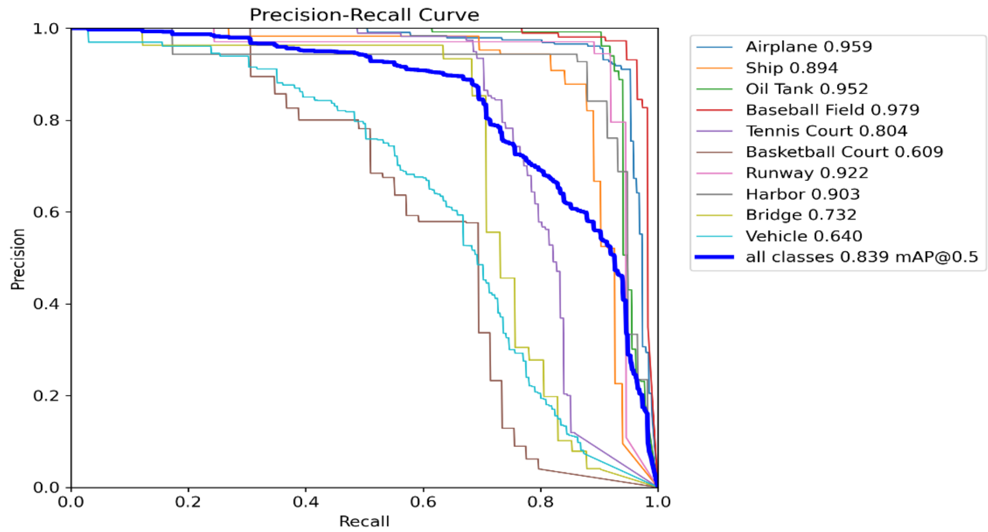

Figure 5 PR curve of YOLOv8-ghost

Figure 6 illustrates the YOLOv8s-ghost model's unique performance during training. Observing the mAP50(B) and mAP50-95(B) metrics in figure 6, by the 200th epoch of training iterations, the average precision values of the model become stable and no longer show significant improvement, indicating that the model has been trained to its optimal state and will not continue to improve with additional training epochs. At the same time, the precision and recall, two evaluation metrics of the model, also exhibit a stable trend, implying that the model has found the best balance in its current state, and there is no need to further increase the number of training iterations.

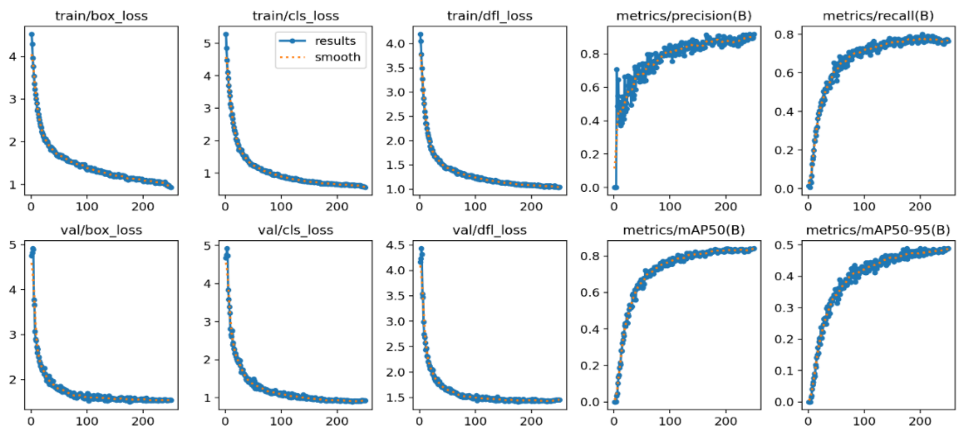

Figure 6 Training process diagram of YOLOv8-ghost

4.4 Comparative experiment

In order to further illustrate the effectiveness of the improved algorithm, experimental comparison is made with YOLOv8s.

Table 1 Comparison of Experimental Results

METHOD	Airplane	Ship	Oil Tank	Baseball Field	Tennis Court	Basketball Court	Runway	Harbor	Bridge	Vehicle	mAP@50	GFLOPS
Yolov8s	**1**	**0.93**	0.90	**0.96**	**0.81**	**0.76**	0.89	**0.98**	**0.73**	**0.79**	**0.89**	28.8
Yolov8s-ghost	0.95	0.89	**0.93**	**0.96**	0.75	0.49	**0.92**	0.91	**0.73**	0.60	0.83	**18.0**

From table 1, it is evident that the lightweight YOLOv8s-ghost has achieved a significant improvement in computational efficiency, with GFLOPS reduced by 37.5%, making the model more appropriate for deployment in resource-constrained environments. However, this lightweight design also entails some performance trade-offs, particularly in the recognition of basketball courts and vehicles. Despite this, the accuracy of oil tank and runway category recognition has improved, indicating that lightweight networks may be more effective in certain categories. Overall, while maintaining excellent detection accuracy,

ISSN 2616-5775 Vol. 8, Issue 2: 83-89, DOI: 10.25236/AJCIS.2025.080211

YOLOv8s-ghost greatly decreases computing costs, which is particularly useful in application scenarios needing quick response and limited resources.

5. Conclusion

This study addresses the difficulty of finding small targets in remote sensing photos by presenting a detection algorithm based on the lightweight YOLOv8. By integrating GhostNet as the backbone network, we have achieved significant lightweighting of the model while maintaining high detection accuracy. The following are the study's primary findings and contributions:

1) Lightweight Design: We successfully decreased the number of model parameters and processing needs by including GhostNet into the YOLOv8 backbone network, which improved the model's suitability for operation on devices with limited resources.

2) Performance Optimization: The improved YOLOv8s-ghost model has achieved a significant enhancement in computational efficiency while maintaining high precision, with GFLOPS reduced by 37.5%, which is particularly important for application scenarios requiring rapid response.

3) Experimental Validation: We validated the proposed method's effectiveness by conducting tests on the NWPU VHR-10 dataset. The results showed that the lightweight YOLOv8s-ghost significantly reduced computational costs while maintaining high detection accuracy.

Future study will concentrate on further improving the lightweight network structure to improve the accuracy of small target recognition and investigating other deep learning models appropriate for remote sensing photos. Additionally, we will consider applying this method to other types of remote sensing data to verify its versatility and scalability. We expect that our work will result in more efficient and accurate target detection tools for remote sensing picture processing.

References

[1] Chen Tianpeng, Hu Jianwen. A Survey on Deep Learning-Based Rotational Object Detection in Remote Sensing Images[J]. Application Research of Computers, 2024, 41(02): 329-340.

[2] Liu L, Ouyang W, Wang X, et al. Deep learning for generic object detection: A survey[J]. International journal of computer vision, 2020, 128: 261-318.

[3] Weng Junhui, Cheng Le, Huang Manli, et al. Small Object Detection in UAV Aerial Photography Images Based on CS-YOLOv5s[J]. Electronic Measurement Technology, 2024, 47(07): 157-162.

[4] Liang Tiantian, Yang Songqi, Qian Zhenming. An Improved YOLOv8s Method for Vehicle and Pedestrian Detection in Adverse Weather Conditions[J]. Electronic Measurement Technology, 2024, 47(09): 112-119.

[5] Lei Bangjun, Yu Ao, et al. Small Target Detection Algorithm for Aerial Photography Based on Location Awareness and Cross-Layer Feature Fusion[J]. Electric Measurement Technology, 2024, 47(5): 112-123.

[6] Wang Xuanhui, Wu Ying, Shao Kaiyang, et al. Research on Multi-Object Tracking and Detection for Autonomous Driving Based on Improved YOLOv8s[J]. Automotive Technology, 2024, (12): 1-7.

[7] Song L ,Tao S ,Fangke J , et al.Infrared road object detection algorithm based on spatial depth channel attention network and improved YOLOv8[J/OL].Optoelectronics Letters,1-8[2024-12-20].

[8] Han K, Wang Y, Tian Q, et al. Ghostnet: More features from cheap operations[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 1580-1589.

[9] 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.

[10] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 7464-7475.

[11] Han K, Wang Y, Tian Q, et al. Ghostnet: More features from cheap operations[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 1580-1589.

[12] Su H, Wei S, Yan M, et al. Object detection and instance segmentation in remote sensing imagery based on precise mask R-CNN[C]//IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019: 1454-1457.