ISSN: 2832-790X
Online Journal of
Robotics & Automation Technology
DOI: 10.33552/OJRAT.2026.04.000594
Iris Publishers

Mini Review

Research Summary of Small Target Detection Technology

Xuejiao Bai[1,2], Luyao Tao[1,2], Rui Fan[1,2] and Xiaoli Luo[1,2]*

[1]School of Mathematics and Computer Science, Ningxia Normal University, Guyuan, Ningxia 756099, China

[2]Artificial Intelligence and Intelligent Medical Engineering Technology Research-Centere, China

***Corresponding author:** Xiaoli Luo, School of Mathematics and Computer Science, Ningxia Normal University, Guyuan, Ningxia 756099, China

Received Date: January 26, 2026

Published Date: April 07, 2026

Abstract

Small target detection is an important research direction in the field of computer vision. Owing to the small size of targets, weak discriminative features, and strong interference from complex backgrounds, small target detection remains a challenging task in applications such as remote sensing monitoring, intelligent transportation, and medical imaging. This paper systematically reviews recent advances in small target detection, with a particular focus on key techniques including multi-scale feature fusion, context modeling, attention mechanisms, loss function optimization, and super-resolution enhancement. In addition, the development potential of emerging research directions, such as neural architecture search, cross-modal fusion, and weakly supervised learning, is discussed. By summarizing representative methods and practical applications, this paper aims to provide a comprehensive technical reference and theoretical insights for small target detection research, thereby facilitating further progress in complex real-world scenarios.

Keywords: Small Target Detection; Deep Learning; Multi-scale Features; Attention Mechanism; Context Information

Introduction

As an important branch of target detection [1], small target detection plays a critical role in many key application scenarios, including remote sensing monitoring, autonomous driving, medical image analysis, and intelligent surveillance. In remote sensing applications, targets such as ships and vehicles often occupy only a very small number of pixels, and their detection accuracy directly affects the reliability of remote sensing interpretation and situational awareness [2]. In the field of autonomous driving, the low-resolution characteristics of distant pedestrians and small traffic signs make missed detections or false alarms more likely, which can pose serious threats to system safety [3]. In medical imaging, small lesions and early-stage nodules are typically characterized by small size and low contrast, making high-precision detection a crucial technical support for early disease screening [4].

In industrial visual inspection, accurate detection of small defects is directly related to production quality and operational safety. In addition, emerging applications such as intelligent surveillance and UAV-based inspection impose higher requirements on the accurate detection of densely distributed small targets and long-range anomalous objects. From the perspective of technological evolution, small target detection has gradually expanded from early handcrafted feature-based methods to complex real-world scenarios across multiple domains. With the advancement of UAV platforms, high-resolution imaging, and infrared sensing technologies, small target detection continues to evolve toward more realistic and challenging application environments, and its performance has become an important indicator for evaluating the practicality and robustness of detection systems [5-6].

Compared with conventional targets, small targets typically occupy only a limited number of pixels (e.g., less than 32×32 pixels in the COCO dataset), and are characterized by blurred structures and weak discriminative features. As a result, they are highly susceptible to background clutter and occlusion in complex scenes, leading to a significant degradation in the performance of traditional detection methods. The core challenges of small target detection manifest at multiple levels [7]. First, drastic scale variations in feature maps pose fundamental difficulties. While shallow features preserve rich spatial details, they lack semantic representation, whereas deep features contain strong semantic information but tend to lose fine-grained details of small targets. This imbalance makes multi-scale feature fusion still suffer from issues such as insufficient feature alignment and information redundancy [8].

Second, due to the limited intrinsic discriminative information of small targets, detection performance heavily relies on contextual cues. However, existing context modeling methods often struggle to suppress irrelevant background information and to impose effective semantic constraints. In addition, although attention mechanisms are capable of enhancing salient regions, they are prone to noise and complex backgrounds in small target scenarios, which may result in attention drift or imbalanced feature suppression. Meanwhile, the scarcity of small target samples and their high sensitivity to localization errors further exacerbate the limitations of traditional loss functions in handling class imbalance and bounding box regression accuracy [9]. Finally, super-resolution enhancement techniques attempt to recover fine-grained details of small targets, but their collaborative optimization with detection tasks, computational cost, and generalization ability remain major challenges in practical applications. Collectively, these issues significantly constrain the robustness and practicality of small target detection in complex real-world environments [10].

Technologic progress for small target detection challenges

Multi-scale feature fusion

Multi-scale feature fusion is one of the core approaches for alleviating the insufficient feature representation of small targets. By jointly exploiting features at different scales, it provides more comprehensive representations for small targets and effectively enhances detection performance. Ye et al. [11] proposed MFF-Net for mechanical component fault detection, demonstrating that multi-feature fusion can effectively overcome the limitations imposed by small target size. Building upon this idea, Huang et al. [12] integrated image enhancement and feature fusion within the SSD framework, significantly improving the detection accuracy of small targets in complex indoor environments. Subsequently, in the context of UAV imagery, Ma et al. [13] introduced YOLO-UAV, which mitigates the imbalance between positive and negative samples through multi-scale fusion and accelerates model convergence. Tan et al. [14] further enhanced aerial image detection performance by increasing small-target detection scales and incorporating adaptive fusion mechanisms. For remote sensing imagery, Li et al. [15] proposed RSI-YOLO, an improved YOLOv5-based model, while Xu et al. [16] developed the progressive fusion network D-FPN. Both approaches strengthen multi-scale information integration and effectively improve the recognition performance of long-range small targets.

In terms of lightweight design, Ma et al. [17] introduced the LAYN network, which incorporates a multi-scale attention mechanism to enhance small-target feature extraction while maintaining computational efficiency. Similarly, Li et al. [18] proposed SOD-YOLO by integrating the RFCBAM module, alleviating spatial information loss caused by down sampling. In addition, multi-scale feature fusion techniques have been successfully extended to other application scenarios. Wang et al. [19] applied them to head detection for classroom teaching quality assessment, while Liu et al. [20] proposed a recursive fusion architecture, EMB-RFNet, which significantly improves the detection accuracy of small targets such as traffic signs. In summary, by integrating features across multiple levels, multi-scale feature fusion effectively addresses key challenges in small target detection, including limited target size and complex backgrounds, and demonstrates substantial application value and broad potential across diverse domains.

Context information modeling

Context information modeling plays a critical role in improving the performance of small target detection networks, as small targets often lack sufficient discriminative features and rely heavily on surrounding contextual cues. Guan et al. [21] constructed a large-scale, semantically rich feature map through a top-down pathway, and combined multi-level feature fusion with pyramid pooling to generate context-aware representations. Building upon this idea, Xiao et al. [22] introduced skip pooling and context fusion mechanisms into the Faster R-CNN framework, demonstrating that multi-stage context fusion at the feature level can significantly enhance detection performance when coupled with multi-scale features.In the domain of remote sensing imagery, Feng et al. [23] proposed TCANet, which embeds context reasoning into the candidate region generation stage by explicitly modeling semantic and spatial relationships between targets, thereby improving detection accuracy.

Min et al. [24] further leveraged a Transformer-based CoT framework to integrate both global and local contextual information throughout the entire detection pipeline, from input feature extraction to output prediction. For lightweight detection models, Yu et al. [25] incorporated adaptive context fusion and spatial attention mechanisms into the detection head of YOLOv5s, strengthening context modeling during the feature aggregation stage without introducing excessive computational overhead. Meanwhile, Song et al. [26] proposed BAFNet, which enhances small target representation and localization accuracy through boundary-aware multi-level feature fusion. Overall, the paradigm of context information modeling has evolved from localized context extraction to cross-layer collaborative modeling. The dominant technical pathway emphasizes feature fusion based on top-down architectures, complemented by multi-scale perception and attention mechanisms, effectively alleviating missed detections and false alarms caused by the inherent information scarcity of small targets.

Attention mechanism

To enhance small target detection performance, attention mechanisms have been extensively integrated into feature pyramid networks and multi-scale fusion frameworks. These mechanisms improve the model's ability to extract and discriminate small targets in complex scenes by enhancing scale awareness and performing feature recalibration. Li et al. [27,28] proposed a pyramid feature reconstruction approach that combines neighborhood erasure and transfer (NET), effectively strengthening scale perception while suppressing interference from large targets. Building upon this foundation, subsequent studies further incorporated attention mechanisms into network architectures. For example, Han et al. [29] integrated coordinate attention, a Transformer encoder, and a Swin Transformer module into YOLOv5, significantly improving detection accuracy in specific tasks such as wood defect detection. Similarly, Zheng et al. [30] embedded attention mechanisms into the multi-scale steel bar detection network RebarNet, mitigating missed detections and false alarms of densely distributed small targets through optimized feature interactions.

In recent years, self-attention mechanisms have attracted increasing interest due to their flexibility in feature enhancement. Wu et al. [31] proposed OE-YOLO, which employs attention mechanisms to preserve small-target features without substantially increasing network depth. Zhang et al. [32] introduced hR-YOLOv8, where self-attention enhances high-resolution feature extraction, leading to improved detection performance for small targets such as crop growth indicators. Furthermore, Wang et al. [33] demonstrated that combining multi-scale feature fusion pyramids with attention mechanisms can significantly optimize the feature aggregation process. Overall, attention mechanisms have evolved from auxiliary components to core elements of modern detection networks; when embedded within feature pyramids or multi-scale fusion layers, they effectively enhance feature representation and precise localization of small targets.

Loss function optimization

Loss functions specifically designed to address the characteristics of small targets play a critical role in guiding deep learning models toward more effective optimization. Early studies, such as the work of Deng et al. [34], enhanced multi-scale feature representation by expanding the feature pyramid network, laying the groundwork for subsequent task-oriented loss function design. Building on this, Xiao et al. [35] explicitly proposed a difference maximization loss to strengthen the model's ability to distinguish salient small-target features from complex backgrounds. With the growing adoption of anchor-free detection paradigms, including the mask-guided feature method introduced by Qu et al. [36], loss functions have been required to exhibit greater flexibility in adapting to variations in target scale and morphology, thereby motivating the development of diverse novel loss formulations. For instance, the VIoU loss proposed by Lu et al. [37] improves localization accuracy through vectorized IoU computation, while the SCIoU loss introduced by Wang et al. [38] enhances detection robustness by incorporating shape-aware constraints.

Meanwhile, lightweight architectures designed for specific application scenarios-such as EL-YOLO by Hu et al. [39] and HP-YOLOv8 by Yao et al. [40]-also rely heavily on carefully balanced loss functions to jointly optimize detection accuracy and computational efficiency. In addition, the effectiveness of attention mechanisms integrated into SOD-YOLOv10 by Sun et al. [41] is closely tied to loss function components capable of accurately regulating feature weighting. In summary, specialized loss functions-including difference maximization loss, VIoU loss, and SCIoU loss-have continuously driven advances in small target detection by improving localization precision, enhancing adaptability to multi-scale scenes, and synergistically complementing innovations in network architecture.

Super-resolution enhancement

Limited by imaging conditions and sensor resolution, small targets often suffer from blurred appearances and missing fine details, which severely constrains the discriminative capability of subsequent detection models. To address this issue, super-resolution enhancement techniques have been introduced into small target detection frameworks. By reconstructing and enhancing low-resolution features, these methods provide clearer fine-grained information for detection networks. In practice, super-resolution modules are mainly integrated into the feature extraction stage or the multi-scale fusion layer to enhance the saliency and structural details of small targets. Liu et al. [42] designed a dual-CNN architecture for bacterial colony detection, embedding a super-resolution enhancement mechanism into the network to improve feature clarity while maintaining real-time performance. By fusing multi-scale features with image enhancement techniques at the detection layer and adjacent feature layers, Huang et al. [43] enabled more effective capture of small-target details in complex indoor environments.

The dense detection network proposed by Song et al. [44] further deepened the understanding of multi-scale semantics by integrating multi-scale information fusion with hierarchical feature enhancement modules. LDHD-Net, introduced by Zhang et al. [45], employs a lightweight dual-branch head and incorporates super-resolution enhancement into the feature extraction module, achieving a favorable balance between detection accuracy and computational efficiency. In addition, MSMA-Net developed by Ma et al. [46] adopts a coarse-to-fine multi-scale super-resolution strategy, distributing enhancement components across multiple network levels, which significantly improves robustness when detecting small, blurry targets in complex scenes. Overall, these studies demonstrate that super-resolution enhancement effectively amplifies salient features of both large and small targets, thereby systematically improving the accuracy and robustness of small target detection across diverse and challenging environments.

The future development direction of small target detection

In the future, the development of small target detection technology is expected to further deepen and integrate advances in multi-scale feature fusion, context modeling, attention mechanisms, loss function optimization, and super-resolution enhancement. Multi-scale feature learning will place greater emphasis on cross-level dynamic and adaptive fusion, combined with neural architecture search to construct efficient and flexible fusion structures tailored to targets of varying scales. Context modeling will evolve toward collaborative reasoning over global and local semantics, strengthening spatial topological relationships and scene-level semantic constraints among objects, while exploring graph neural networks and Transformer-based modeling paradigms.

Attention mechanisms are likely to be more tightly coupled with feature extraction and fusion processes, leading to the design of attention modules with enhanced scale awareness and anti-interference capabilities, thereby enabling more precise feature focusing and background suppression. Loss function design will shift toward multi-task coordinated optimization and fine-grained boundary regression. By considering the extreme sample imbalance and scale sensitivity inherent to small targets, more robust loss combinations that balance localization accuracy and stability will be developed, alongside emerging optimization strategies integrating knowledge distillation and self-supervised learning. Meanwhile, super-resolution enhancement is expected to form a more tightly coupled end-to-end collaborative optimization framework with detection tasks. Future research will focus on designing small-target-oriented reconstruction networks that enhance fine details under strict computational constraints, while also exploring their potential in multi-modal scenarios and cross-domain detection tasks.

Conclusion

Small target detection plays an important role in applications such as remote sensing monitoring, intelligent transportation, medical imaging, and industrial inspection. However, due to the small object size, weak discriminative features, complex backgrounds, and highly imbalanced sample distributions, its detection performance still faces significant challenges. To address these issues, recent research has mainly focused on multi-scale feature fusion, context modeling, attention mechanisms, loss function optimization, and super-resolution enhancement. Existing studies demonstrate that multi-scale learning and context modeling can effectively alleviate scale variation and feature information loss. Improvements in attention mechanisms and loss functions further enhance the model's ability to focus on small target features and improve training stability, while super-resolution techniques provide new perspectives for fine-detail enhancement. Overall, small target detection is evolving toward deeper integration and collaborative optimization of multiple complementary strategies.

Acknowledgements and Funding

The authors gratefully acknowledge financial support from Artificial Intelligence and Intelligent Medical Engineering Technology Research Center, the National Natural Science Foundation of China under Grant 62566048, Key R&D Program Project of Ningxia Hui Autonomous Region (Priority) under Grant 2023BEG02072, National Natural Science Foundation of Ningxia (Outstanding Youth Program) under Grant 2025AAC050086, and Ningxia Hui Autonomous Region Education Department Higher School Scientific Research Project (Young Teacher Support Project) under Grant NYG202420.

Conflict of interest

No conflict of interest.

References

1. Yue T, Lu X, Cai J, Chen Y, Chu S (2025) YOLO-MST: Multiscale deep learning method for infrared small target detection based on super-resolution and YOLO. Optics & Laser Technology 187: 112835.
2. Zhang M, Yue K, Guo J, Zhang Q, Zhang J, et al. (2025) Computational fluid dynamic network for infrared small target detection. IEEE Transactions on Neural Networks and Learning Systems.
3. Li Q, Zhang W, Lu W, Wang Q (2025) Multi-branch mutual-guiding learning for infrared small target detection. IEEE Transactions on Geoscience and Remote Sensing.
4. Shen Y, Li Q, Xu C, Chang C, Yin Q (2025) Graph-based context learning network for infrared small target detection. Neurocomputing 616: 128949.
5. Yan X, Du J, Li X, Wang X, Sun X, et al. (2025) A Hierarchical Feature Fusion and Dynamic Collaboration Framework for Robust Small Target Detection. IEEE Access.
6. Ye B, Tong G, Li Q (2025) Lightweight Mamba residual network integrating center-focused strategy for infrared small target detection. Applied Soft Computing, pp. 114322.
7. Ying X, Liu L, Lin Z, Shi Y, Wang Y, et al. (2025) Infrared small target detection in satellite videos: a new dataset and a novel recurrent feature refinement framework. IEEE Transactions on Geoscience and Remote Sensing.
8. Zhang H, Liang M, Wang Y (2025) YOLO-BS: a traffic sign detection algorithm based on YOLOv8. Scientific Reports 15(1): 7558.
9. Guan T, Chang S, Wang C, Jia X (2025) SAR Small Ship Detection Based on Enhanced YOLO Network. Remote Sensing 17(5): 839.
10. Ye T, Zhang Z, Zhang X, Chen Y, Zhou F (2021) Fault detection of railway freight cars mechanical components based on multi-feature fusion convolutional neural network. International Journal of Machine Learning and Cybernetics 12(6): 1789-1801.
11. Huang L, Chen C, Yun J, Sun Y, Tian J, et al. (2022) Multi-scale feature fusion convolutional neural network for indoor small target detection. Frontiers in Neurorobotics 16: 881021.
12. Ma C, Fu Y, Wang D, Guo R, Zhao X, et al. (2023) YOLO-UAV: Object detection method of unmanned aerial vehicle imagery based on efficient multi-scale feature fusion. IEEE Access 11: 126857-126878.
13. Li Z, Yuan J, Li G, Wang H, Li X, et al. (2023) RSI-YOLO: Object detection method for remote sensing images based on improved YOLO. Sensors 23(14): 64
14. Xu Y, Ji H, Zhang Y, Zhang W, Wang P, et al. (2024) Detail feature pyramid network for small object detection in aerial images. In International Conference on Computer Application and Information Security (ICCAIS 2023) 13090: 905-910.
15. Ma S, Lu H, Liu J, Zhu Y, Sang P (2024) Layn: Lightweight multi-scale attention yolov8 network for small object detection. IEEE Access 12: 29294-29307.
16. Li Y, Li Q, Pan J, Zhou Y, Zhu H, Wei H, Liu C (2024) Sod-yolo: Small-object-detection algorithm based on improved yolov8 for UAV images. Remote Sensing 16(16): 3057.
17. Wang R, Chen S, Tian G, Wang P, Ying S (2024) Post-secondary classroom teaching quality evaluation using small object detection model. Scientific reports 14(1): 5816.
18. Liu X, Wang X, Li F, Duan X, Yin S, et al. (2025) A feature cascade and recursive fusion architecture for traffic sign detection in vehicle perception. Scientific Reports 15(1): 41953.
19. Tan S, Duan Z, Pu L (2024) Multi-scale object detection in UAV images based on adaptive feature fusion. Plos one 19(3): e0300120.
20. Guan L, Wu Y, Zhao J (2018) Scan: Semantic context aware network for accurate small object detection. International Journal of Computational Intelligence Systems 11(1): 951-961.
21. Xiao Y, Wang X, Zhang P, Meng F, Shao F (2020) Object detection based on faster R-CNN algorithm with skip pooling and fusion of contextual information. Sensors 20(19): 5490.
22. Feng X, Han J, Yao X, Cheng G (2020) TCANet: Triple context-aware network for weakly supervised object detection in remote sensing images. IEEE Transactions on Geoscience and Remote Sensing 59(8): 6946-6955.
23. Min L, Fan Z, Lv Q, Reda M, Shen L, et al. (2023) YOLO-DCTI: small object detection in remote sensing base on contextual transformer enhancement. Remote Sensing 15(16): 3970.
24. Yu G, Wang T, Guo G, Liu H (2023) SFHG-YOLO: A simple real-time small-object-detection method for estimating pineapple yield from unmanned aerial vehicles. Sensors 23(22): 9242.
25. Song J, Zhou M, Luo J, Pu H, Feng Y, et al. (2024) Boundary-aware feature fusion with dual-stream attention for remote sensing small object detection. IEEE Transactions on Geoscience and Remote Sensing.
26. Li Y, Pang Y, Shen J, Cao J, Shao L (2020) NETNet: Neighbor erasing and transferring network for better single shot object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 13349-13358.

27. Li Y, Pang Y, Cao J, Shen J, Shao L (2021) Improving single shot object detection with feature scale unmixing. IEEE Transactions on Image Processing 30: 2708-2721.

28. Han S, Jiang X, Wu Z (2023) An improved YOLOv5 algorithm for wood defect detection based on attention. IEEE Access 11: 71800-71810.

29. Yanmei Zheng, Guanghui Zhou, B Lu, (2023) A Multi-Scale Rebar Detection Network with An Embedded Attention Mechanism, APPLIED SCIENCES.

30. Kun Wu, Zhijian Zhang, Zeyu Chen, Guohua Liu (2024) Object-Enhanced YOLO Networks for Synthetic Aperture Radar Ship Detection, REMOTE. SENS.

31. Zhang J, Yang W, Lu Z, Chen D (2024) HR-YOLOv8: a crop growth status object detection method based on YOLOv8. Electronics 13(9): 1620.

32. Wang X, Liu Y, Xu H, Xue C (2024) Spatial small target detection method based on multi-scale feature fusion pyramid. Applied Sciences 14(13): 5673.

33. Deng C, Wang M, Liu L, Liu Y, Jiang Y (2021) Extended feature pyramid network for small object detection. IEEE Transactions on Multimedia 24: 1968-1979.

34. Xiao X, Wang B, Miao L, Li L, Zhou Z, et al. (2021) Infrared and visible image object detection via focused feature enhancement and cascaded semantic extension. Remote Sensing 13(13): 2538.

35. Qu H, Shen L, Guo W, Wang J (2021) Ships detection in SAR images based on anchor-free model with mask guidance features. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 15: 666-675.

36. Lu S, Lu H, Dong J, Wu S (2023) Object detection for UAV aerial scenarios based on vectorized IOU. Sensors 23(6): 3061.

37. Hu M, Li Z, Yu J, Wan X, Tan H, Lin Z (2023) Efficient-lightweight YOLO: improving small object detection in YOLO for aerial images. Sensors 23(14): 6423.

38. Yao G, Zhu S, Zhang L, Qi M (2024) Hp-yolov8: high-precision small object detection algorithm for remote sensing images. Sensors 240(15): 4858.

39. Wang L, Shi Y, Mao G, Dharejo FA, Javed S, et al. (2023) Consumer-centric insights into resilient small object detection: Sciou loss and recursive transformer network. IEEE Transactions on Consumer Electronics 70(1): 2178-2187.

40. Yao G, Zhu S, Zhang L, Qi M (2024) Hp-yolov8: high-precision small object detection algorithm for remote sensing images. Sensors 240(15): 4858.

41. Sun H, Yao G, Zhu S, Zhang L, Xu H, et al. (2025) SOD-YOLOv10: Small object detection in remote sensing images based on YOLOv10. IEEE Geoscience and Remote Sensing Letters.

42. Liu S, Gai Z, Chai X, Guo F, Zhang M, et al. (2021) Bacterial colonies detecting and counting based on enhanced CNN detection method. In E3S Web of Conferences 233: 02012

43. Huang L, Chen C, Yun J, Sun Y, Tian J, et al. (2022) Multi-scale feature fusion convolutional neural network for indoor small target detection. Frontiers in Neurorobotics 16: 881021.

44. Song Z, Huang X, Ji C, Zhang Y (2023) Deformable YOLOX: Detection and rust warning method of transmission line connection fittings based on image processing technology. IEEE Transactions on Instrumentation and Measurement 72: 1-21.

45. Zhang C, Gao Q, Shi R, Yue M (2024) LDHD-Net: A Lightweight Network with Double Branch Head for Feature Enhancement of UAV Targets in Complex Scenes. International Journal of Intelligent Systems, 2024(1): 7259029.

46. Ma T, Wang H, Liang J, Peng J, Ma Q, et al. (2023) Msma-net: An infrared small target detection network by multiscale super-resolution enhancement and multilevel attention fusion. IEEE Transactions on Geoscience and Remote Sensing 62: 1-20.