A Review of Research on Small Object Detection Algorithms Based on Deep Learning

Jihua He[1], Youmin Guo[2], Qi Zhou[3]

Lanzhou Jiaotong University, Lanzhou 730070, Gansu, China

Abstract: ***Small object detection is an important research direction in the field of computer vision and one of the most challenging problems in object detection tasks. With the rapid development of deep learning technology, object detection algorithms based on deep neural networks have achieved significant results in large and medium-sized object detection, but still face many challenges in small object detection. This article systematically reviews the research status and development trends of small object detection algorithms based on deep learning in recent years. Firstly, the concept definition and main challenges of small object detection were introduced, and the limitations of traditional object detection algorithms in small object scenarios were analyzed. Then, the focus was on the classification and discussion of small object detection methods based on deep learning, including key technologies such as feature enhancement, multi-scale feature fusion, and attention mechanism. At the same time, in-depth analysis and comparison were conducted on the optimization and improvement schemes of mainstream detection algorithms such as YOLOv5/v7 and SSD in small object detection. Finally, the existing problems in current research were summarized, and future development directions were discussed.***

Keywords: Small target detection; Deep learning; Feature enhancement; Multi-scale features; Attention mechanism.

1. INTRODUCTION

With the rapid development of computer vision and deep learning technologies, object detection has been widely applied in fields such as autonomous driving, security monitoring, and medical diagnosis. However, for smaller targets, their detection accuracy is often significantly lower than that of large and medium-sized targets due to issues such as occupying fewer pixels in the image, insufficient feature information, and susceptibility to background interference. Taking the MS COCO dataset as an example, the detection performance of small targets (area<32×32 pixels) is usually only about half of that of large targets [1]. Therefore, how to effectively improve the accuracy of small object detection has become a key scientific problem that urgently needs to be solved in the field of object detection.

The main challenges faced by small object detection include:

(1) Insufficient feature expression makes it easy for small targets to lose detailed information in the multi-layer feature extraction process of deep convolutional networks;

(2) The background interference is severe, and the distinction between small targets and the background is not high, which can easily lead to false positives;

(3) Scale changes are sensitive, and the detector responds strongly to small deviations in the bounding box of small targets;

(4) The distribution of positive and negative samples is unbalanced, and the number of small target samples in the training data is relatively small. These issues severely constrain the performance improvement of small object detection algorithms.

In recent years, researchers have proposed various small object detection methods based on deep learning, mainly including technical routes such as feature enhancement, multi-scale feature fusion, and attention mechanism [2]. Feature enhancement methods enhance the feature expression of small targets by designing feature extraction network structures or introducing auxiliary tasks. The multi-scale feature fusion method utilizes the complementarity of features at different levels to improve detection performance. The attention mechanism improves detection accuracy by adaptively focusing on important feature regions [3]. These methods have improved the performance of small object detection to some extent, but there is still significant room for improvement.

Wu [1] developed an intelligent gateway management platform utilizing Jenkins cluster architecture for cloud-edge integration in IoT environments, while Chen [2] proposed scalable cloud infrastructure solutions to support autonomous driving data lakes and real-time decision making, highlighting the critical role of distributed computing architectures in modern industrial applications. Computer vision technologies have achieved notable progress, with Peng et al. [3] introducing a dual-augmentor framework for robust 3D human pose estimation across domains, and Lyu et al. [5] optimizing CNN architectures for efficient 3D point cloud object recognition. These advancements in visual perception systems are being actively applied to urban management, as exemplified by Zhou et al. [4]'s ResNet-50 based garbage recognition model for sustainable city development and Liu et al. [8]'s MiM-UNet architecture for building image segmentation. Supply chain and energy systems have benefited from AI-driven optimization approaches. Wang and Liang [6] combined graph neural networks with reinforcement learning for supply chain route optimization, while Zhao et al. [7] developed a CNN-Bi-GRU hybrid model for accurate renewable electricity demand forecasting. These applications demonstrate how modern machine learning techniques can enhance operational efficiency in critical infrastructure systems. In healthcare applications, Tian et al. [10] improved brain tumor segmentation accuracy through GSConv modules and attention mechanisms, while Shen et al. [11] implemented an LSTM-based system for precision anesthetic dosing in cancer surgery. The security of such AI-enabled systems is addressed by Xu et al. [12] through their research on adversarial machine learning attacks and defenses. Emerging creative applications are represented by Xu et al. [9]'s AI-enhanced tools for cross-cultural game design, illustrating the expanding influence of AI beyond traditional technical domains.

2. BASIC CONCEPTS AND KEY TECHNOLOGIES OF SMALL OBJECT DETECTION

In object detection tasks, there are two main criteria for defining small targets: absolute scale and relative scale. Absolute scale is defined based on the pixel size occupied by the target in the image. For example, the MS COCO dataset defines targets with an area less than 32×32 pixels as small targets; Relative scale is divided based on the proportion of the target's area relative to the entire image. For example, SPIE considers targets with an area less than 0.12% of the image area as small targets [4]. In practical applications, both of these definition standards have their applicable scenarios and need to be selected based on specific task requirements. Meanwhile, due to the challenges faced by small object detection, such as insufficient feature expression, severe background interference, and sensitivity to scale changes, special technical means are needed to improve detection performance.

3. SMALL OBJECTIVE OPTIMIZATION STRATEGIES FOR MAINSTREAM DETECTION ALGORITHMS

After years of development, object detection algorithms have formed multiple mainstream algorithm branches represented by YOLO series and SSD. These algorithms adopt different optimization strategies when dealing with small object detection problems. The YOLO series algorithms enhance small object detection performance by improving feature extraction networks, optimizing feature fusion methods, and innovating data augmentation techniques. YOLOv5 adopts CSPNet as the backbone network and reduces computational complexity while maintaining feature propagation through a cross stage local network connection strategy; Introducing SPPF module to enhance feature extraction capability through multi-scale pooling; Design PANet structure to achieve multi-scale feature aggregation; Innovatively proposed Mosaic data augmentation technology, which not only increases the number of small targets but also enriches the background information of training images. YOLOv7 introduces the E-ELAN structure on this basis to achieve multi branch architecture and progressive feature aggregation. It adopts reparameterization design to improve inference efficiency while maintaining performance, and enhances the model's generalization ability through auxiliary head supervision [5].

The SSD algorithm, as a typical single-stage detector, mainly improves the performance of small object detection by optimizing the feature map structure, improving the design of the detection head, and adjusting the training strategy. In terms of feature maps, more low-level feature maps have been added for small object detection, and a feature enhancement module has been introduced to enhance feature expression through attention mechanism. In the design of head detection, an adaptive feature selection mechanism is adopted to dynamically select the most suitable feature layer for detection based on the target scale, while optimizing the bounding box regression branch to improve localization accuracy. In terms of training strategy, the matching strategy of positive and negative samples has been improved, and a weighted loss function design has been adopted to make the network more focused on learning small targets. These optimization measures have collectively improved the performance of SSD algorithm in small object detection tasks.

4. FEATURE ENHANCEMENT AND MULTI-SCALE REPRESENTATION

Feature enhancement and multi-scale representation are two core technological directions for improving the performance of small object detection. Feature enhancement mainly enhances the feature expression of small targets through methods such as Feature Pyramid Network (FPN), attention mechanism, and contextual information fusion. FPN combines high-level semantic features with low-level positional features through a top-down feature fusion path, and maintains feature consistency at different scales through a horizontal connection mechanism. Attention mechanisms can adaptively highlight important feature regions and suppress irrelevant background information. For example, channel attention mechanisms weight the importance of features from different channels, while spatial attention mechanisms highlight the feature responses of key regions. Context information fusion provides more clues for small target detection by introducing scene information, helping detectors better understand the relationship between targets and scenes.

Multi scale feature representation mainly solves the problem of target scale changes through feature pyramid structure, multi branch architecture, and feature fusion strategy. The feature pyramid structure constructs feature maps of different resolutions, making shallow feature maps suitable for detecting small targets and deep feature maps more suitable for detecting large targets. The multi branch architecture uses parallel detection branches to process targets within a specific scale range, allowing the detector to focus more on feature learning of targets at different scales. Feature fusion strategies such as feature weighted fusion and adaptive feature fusion need to be optimized for the characteristics of small object detection. At the same time, in order to solve the problem of insufficient small target samples, targeted data augmentation strategies, special loss function designs, and online difficult case mining techniques need to be adopted to improve the model's detection ability for small targets.

5. EVALUATION METHODS AND EXPERIMENTAL ANALYSIS FOR SMALL OBJECT DETECTION

Accurate performance evaluation is of great guiding significance for algorithm improvement and optimization in small object detection tasks. Common evaluation metrics include Precision, Recall, Average Precision (AP), and Mean Average Precision (mAP). Among them, AP is the area under the precision recall curve at different IoU (Intersection over Union) thresholds, which can comprehensively reflect the performance of the detector. Due to the particularity of small target detection, it is also necessary to focus on the AP values (often referred to as APs) of small targets. Meanwhile, detection speed (FPS) is also an important indicator that needs to be considered in practical applications.

In order to comprehensively evaluate the performance of mainstream algorithms in small object detection tasks, this paper selects the MS COCO dataset for experimental analysis. This dataset contains a large number of small target instances and has strong representativeness. The experimental results are shown in Table 1. From the table, it can be seen that YOLOv7 has achieved significant improvement in small object detection performance, with APs reaching 25.8%, which is 2.3 percentage points higher than YOLOv5. This is mainly due to the optimization effect of its E-ELAN structure and reparameterization design. After multiple improvements, SSDs have achieved 22.5% APs, which is lower than the YOLO series but has its unique advantages in certain specific scenarios. In terms of detection speed, both YOLOv7 and the improved SSD can maintain high real-time performance, reaching 50 FPS and 45 FPS respectively (tested on RTX 3090).

Table 1: Performance Comparison of Different Algorithms on the MS COCO Dataset

Algorithm	AP	APs	APm	APl	FPS
YOLOv5	42.5%	23.5%	46.1%	50.3%	55
YOLOv7	43.8%	25.8%	47.2%	51.1%	50
SSD	40.2%	22.5%	44.3%	48.7%	45

Further analysis shows that the performance of small object detection is closely related to multiple factors. Firstly, feature extraction capability is important. Adopting a more powerful backbone network can often lead to performance improvements, but at the same time, it can also increase computational burden. Next is the feature fusion strategy, and reasonable multi-scale feature fusion is crucial for improving the detection performance of small targets. In addition, the optimization of training strategies, such as sample balancing and loss function design, can also significantly affect the final performance.

To verify the effectiveness of different optimization strategies, we also conducted ablation experiments. The results showed that removing the E-ELAN structure in YOLOv7 resulted in a 1.5 percentage point decrease in APs, while removing the reparameterized design resulted in a 0.8 percentage point decrease. This indicates that these optimization strategies have indeed played an important role in improving the performance of small object detection. Similarly, in the improved SSD, the removal of feature enhancement module and adaptive feature selection mechanism resulted in a decrease of 1.2 and 0.9 percentage points in APs, respectively.

The experiment also found that the performance of small object detection is closely related to factors such as the scale, density, and degree of occlusion of the target. When the target scale is smaller or multiple small targets are densely distributed, the difficulty of detection significantly increases. In addition, small object detection in complex backgrounds is more challenging than in simple backgrounds, indicating the importance of utilizing contextual information to improve detection performance. Based on these findings, future algorithm optimization should focus more on these challenging scenarios.

6. TYPICAL APPLICATION SCENARIOS AND CHALLENGES OF SMALL OBJECT DETECTION

Small object detection technology has a wide range of demands in practical applications, and the specific challenges faced in different scenarios are also different. In the field of intelligent transportation, real-time detection of small targets such as vehicles, pedestrians, and traffic signs at long distances is required. The main challenges of this type of scenario are target motion blur, large changes in lighting, and complex weather conditions. For example, in nighttime or rainy weather, the visibility of small targets is significantly reduced, and the performance of traditional detection algorithms often fails to meet the requirements. Therefore, some studies have begun to attempt to combine multimodal information, such as fusing features of visible light and infrared images, to improve the robustness of detection.

In the field of security monitoring, it is necessary to identify small targets such as suspicious individuals and abnormal behavior in video surveillance. These types of applications typically require handling complex situations with large scenarios, multiple objectives, and long time sequences. Especially in crowded scenes, there is severe occlusion and overlap between multiple small targets, which poses a huge challenge to detection. In response to this situation, some algorithms use attention mechanisms to enhance the discrimination of target features, or introduce temporal information to improve the continuity and stability of detection.

Remote sensing image analysis is another important application field for small object detection. Buildings, vehicles, and other targets in satellite or aerial images often appear as extremely small pixels with diverse perspectives and poses. Meanwhile, due to the long shooting distance and limited image resolution, the contrast between the target and background is often low. This requires detection algorithms to have stronger feature extraction and expression capabilities. Some studies improve detection performance by designing specialized multi-scale feature extraction networks or utilizing super-resolution techniques.

7. FUTURE DEVELOPMENT TRENDS AND PROSPECTS

Although small object detection technology has made significant progress in recent years, there are still multiple directions worth further research. Firstly, the further enhancement of feature expression ability. Current deep learning models still face the problem of insufficient feature extraction when dealing with extremely small targets. Consider designing more effective feature extraction architectures or exploring new feature enhancement methods. For example, utilizing self attention mechanisms or graph neural networks to capture richer contextual relationships, or researching novel feature extraction methods based on visual transformers.

The second is the improvement of detection efficiency. With the expansion of application scenarios, real-time requirements are becoming increasingly high. How to improve operational efficiency while ensuring detection accuracy is an important research direction. Possible solutions include lightweight design of network structure, model compression and quantization, algorithm acceleration, etc. Especially for deployment on edge devices, it is necessary to consider the limitations of computing resources and energy consumption.

Thirdly, the generalization ability of small object detection also needs to be strengthened. Current algorithms often perform poorly in the face of domain shift, such as the transfer from training data to practical application scenarios. This requires researching more robust feature learning methods, or exploring new learning paradigms such as semi supervised and self supervised. Meanwhile, how to better utilize prior knowledge and scene understanding to assist detection is also a direction worth paying attention to.

With the continuous emergence of new technologies, there is still a lot of room for development in the field of small object detection. For example, multimodal fusion technology can combine the complementary advantages of different sensors; Adaptive learning strategies can dynamically adjust detection parameters based on the scene; End to end neural architecture search can automatically discover better network structures. These new ideas and methods will drive the development of small object detection technology to a higher level.

Overall, small object detection remains an important research topic in the field of computer vision. The future development requires innovation at multiple levels such as algorithm design, engineering implementation, and application deployment to meet the constantly growing practical needs. At the same time, it is also necessary to strengthen cross integration with other technological fields and explore more effective solutions.

8. CONCLUSION

This article systematically reviews the research status and development trends of small object detection algorithms based on deep learning. Through analysis, it can be seen that the main challenges faced by small object detection include insufficient feature expression, severe background interference, and sample imbalance. To address these issues, researchers have proposed various solutions such as feature enhancement and multi-scale representation, and optimized them in mainstream algorithms such as YOLO series and SSD. The experimental results indicate that these improvement strategies have achieved significant results in enhancing detection performance. However, there are still directions that need to be further improved in practical applications, such as detection accuracy, operational efficiency, and model generalization. Future research should focus on enhancing feature extraction capabilities, improving detection efficiency, and integrating innovation with emerging technologies. At the same time, it is necessary to strengthen the engineering practicality of algorithms and promote the transformation of technology into practical applications. I believe that with the deepening of research, small object detection technology will achieve more breakthrough progress.

REFERENCES

[1] Wu, W. (2025). Construction and optimization of intelligent gateway software management platform based on jenkins cluster management under cloud edge integration architecture in industrial internet of things. Preprints, January.
[2] Chen, J. (2025). Leveraging Scalable Cloud Infrastructure for Autonomous Driving Data Lakes and Real-Time Decision Making.
[3] Peng, Q., Zheng, C., & Chen, C. (2024). A Dual-Augmentor Framework for Domain Generalization in 3D Human Pose Estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2240-2249).
[4] Zhou, Y., Wang, Z., Zheng, S., Zhou, L., Dai, L., Luo, H., ... & Sui, M. (2024). Optimization of automated garbage recognition model based on resnet-50 and weakly supervised cnn for sustainable urban development. Alexandria Engineering Journal, 108, 415-427.
[5] Lyu, T., Gu, D., Chen, P., Jiang, Y., Zhang, Z., Pang, H., ... & Dong, Y. (2024). Optimized CNNs for Rapid 3D Point Cloud Object Recognition. arXiv preprint arXiv:2412.02855.
[6] Wang, Y., & Liang, X. (2025). Application of Reinforcement Learning Methods Combining Graph Neural Networks and Self-Attention Mechanisms in Supply Chain Route Optimization. Sensors, 25(3), 955.
[7] Zhao, S., Xu, Z., Zhu, Z., Liang, X., Zhang, Z., & Jiang, R. (2025). Short and Long-Term Renewable Electricity Demand Forecasting Based on CNN-Bi-GRU Model. IECE Transactions on Emerging Topics in Artificial Intelligence, 2(1), 1-15.
[8] Liu, D., Wang, Z., & Liang, A. (2025). MiM-UNet: An efficient building image segmentation network integrating state space models. Alexandria Engineering Journal, 120, 648-656.
[9] Xu, Y., Shan, X., Lin, Y. S., & Wang, J. (2025). AI-Enhanced Tools for Cross-Cultural Game Design: Supporting Online Character Conceptualization and Collaborative Sketching. In International Conference on Human-Computer Interaction (pp. 429-446). Springer, Cham.

[10] Tian, Q., Wang, Z., & Cui, X. (2024). Improved Unet brain tumor image segmentation based on GSConv module and ECA attention mechanism. arXiv preprint arXiv:2409.13626.

[11] Shen, Z., Wang, Y., Hu, K., Wang, Z., & Lin, S. (2025). Exploration of Clinical Application of AI System Incorporating LSTM Algorithm for Management of Anesthetic Dose in Cancer Surgery. Journal of Theory and Practice in Clinical Sciences, 2, 17-28.

[12] Xu, J., Wang, Y., Chen, H., & Shen, Z. (2025). Adversarial Machine Learning in Cybersecurity: Attacks and Defenses. International Journal of Management Science Research, 8(2), 26-33.