



Object Detection Techniques in UAV Aerial Imaging Scenarios

Zhaojie Yao

School of Computer Science, Civil Aviation Flight University of China, Guanghan, Sichuan
618307, China

Yaozhaojie319@outlook.com

Abstract. With the rapid development of computer vision technology, object detection in drone-captured imagery has become a highly important research topic in recent years. This technology is now applied in a wide range of fields, including urban planning, agricultural monitoring, disaster response, traffic management, and military reconnaissance. However, using drones for object detection still presents several major challenges. Targets in these images are often very small, backgrounds are typically complex and contain many distracting elements, and the sizes of objects can vary significantly. Additionally, lighting and weather conditions frequently change, and occlusion is a common issue. All of these factors make it difficult to achieve accurate and efficient object detection. This paper begins by introducing two primary categories of object detection models based on deep learning: single-stage and two-stage models. It then examines the main technical challenges encountered in real-world drone-based detection and proposes strategies for improvement. Finally, drawing on recent research, the paper offers recommendations for further enhancing object detection methods.

Keywords: Computer Vision Technology, Deep Learning, Object Detection.

1 Introduction

In the context of rapidly evolving technology, unmanned aerial vehicles (UAVs) are increasingly being utilized in a wide range of important fields, including military reconnaissance, disaster monitoring, environmental protection, and intelligent transportation. Simultaneously, the integration of deep learning-based object detection technology with UAV systems has emerged as a major research focus in recent years [1]. However, several technical bottlenecks persist when drones undertake object recognition tasks. Altitude constraints often result in ground objects appearing at small scales, and increasing altitude further reduces image resolution, leading to insufficient detection accuracy. Furthermore, the high flight speed of drones and their wide-area imaging capabilities often result in images containing numerous targets within highly complex scenes. This leads to practical challenges such as frequent occlusions and motion blur. Furthermore, changes in lighting and extreme weather conditions—including strong winds, rain, snow, hail, and dense fog—can degrade

image quality and increase uncertainty in detection tasks. Additionally, significant variations in the size of detection targets make it more challenging for models to adapt. Therefore, enhancing the reliability and accuracy of UAV-based object detection under such complex conditions remains a challenging problem that requires innovative solutions.

Numerous researchers have systematically studied UAV-based object detection. Doll and Loos compared the performance of various target detectors using a dataset of drone-captured sheep images [2]. Zhao et al. examined the challenges, methodologies, and aerial datasets associated with UAV-based maritime object detection [3]. Li et al. provided a comprehensive summary of traditional object detection methods, deep learning-based approaches, enhanced detection techniques, and relevant datasets, along with performance evaluations [4]. However, current research on UAV object detection does not yet provide comprehensive coverage of the latest advancements in the field.

This paper begins by presenting two main categories of object detection models based on deep learning: two-stage models and single-stage models. It then examines the major challenges faced in drone-based object detection, such as the presence of very small targets, significant variations in object scale, and cluttered backgrounds. The study also reviews the limitations of existing methods and proposes strategies for improving detection performance. Finally, drawing on recent research findings, the paper offers several recommendations to further enhance object detection using UAVs, aiming to provide both theoretical support and practical guidance for future studies.

2 UAV-Based Object Detection Algorithms

2.1 Fundamental Principles

With rapid advances in deep learning technology, deep learning-based object detection algorithms have become the dominant approach in this field. Forming the basis for UAV object detection, these algorithms fall into two-stage and single-stage models, distinguished by their structural design. Two-stage object detection algorithms begin by conducting a rapid scan of the entire input image to identify candidate regions that may contain potential targets. These candidate regions are then subjected to more precise processing to achieve high detection accuracy. Representative algorithms include R-CNN, Faster R-CNN and Cascade R-CNN. By contrast, single-stage algorithms predict directly from the image's feature maps, enabling simultaneous target localization and classification with increased detection speed. Examples include the SSD and YOLO series. Although these algorithms differ structurally, their core object detection workflows are generally similar. Fig.1 illustrates the overall workflow.

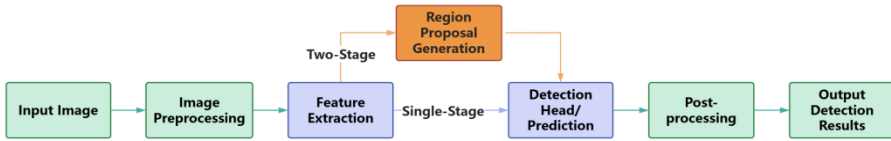


Fig. 1. Deep learning-based object detection algorithms. (Picture credit: Original)

2.2 Optimization of the Model

The application of deep learning-based object detection algorithms in UAV-specific scenarios amplifies their inherent limitations and introduces additional challenges. These challenges include small target sizes, complex background environments, and significant variations in target scale. Such issues arise because drones often operate at high altitudes and high speeds, causing ground objects to occupy only a small portion of the image. This results in fragmented feature representations and pronounced differences in object scale, thereby increasing the difficulty of detection. Furthermore, the wide-angle imaging capability of UAVs provides a broad field of view, which further complicates the background and increases the likelihood of interference from irrelevant objects. In cases of occlusion caused by extreme weather conditions, target features are further weakened, significantly exacerbating the difficulty of object recognition and detection. In response to these issues, many researchers have actively conducted relevant studies to address these unique challenges.

Models for Small Object Detection. LIANG et al. proposed a Feature Fusion and Scaling-based Single Shot Detector (FS-SSD) based on FSSD. By introducing an additional deconvolution scaling branch into the deconvolution module and combining it with average pooling operations, they constructed a feature pyramid. Predictions were made using two feature pyramids derived from the feature fusion module and the deconvolution module. Additionally, a spatial context analysis method was proposed for target re-detection [5].

LI et al., building upon YOLOv5, enhanced global feature extraction by incorporating a multi-head attention mechanism. They employed shallow feature fusion to preserve critical detailed information and designed a dynamic multi-level feature fusion strategy to optimize detection weights for targets of different scales [6].

ZHAO et al., based on YOLOv7, integrated a stride-free convolution module, a deformable attention module, and an efficient multi-scale attention module to optimize the network architecture [7]. Whether on the VisDrone2019 or the NWPU VHR-10 dataset, the new model outperformed other models in terms of mAP50, demonstrating an ability to detect more small objects with a lower missed detection rate. However, the FPS of this model was slightly lower than that of YOLOv8, indicating that further optimization of computational efficiency is needed in future work.

ZHU et al., building upon YOLOv8n, designed a multi-scale feature fusion layer to achieve effective fusion of feature maps at different scales. They then introduced the TFE and SSFF modules to fully leverage multi-scale information and reduce the loss

of detailed information. Subsequently, ConvNeXt v2 was adopted in the backbone network to improve the localization accuracy of small objects. Finally, the LAMP pruning algorithm was employed to reduce the number of parameters and computational cost (FLOPs) while maintaining high precision [8].

Models for Addressing Challenges in Complex Environments. Xi et al. proposed a Fine-grained Object Focus Network (FiFoNet), which adaptively fuses multi-scale features through the FIFA module to aggregate information from sub-regions within objects. This approach enhances the model's recognition capability by providing more refined feature representations. Furthermore, they designed a TFB module that utilizes masks to distinguish between foreground targets and background regions in images, thereby suppressing the impact of background noise. Additionally, a GLCC module was introduced to capture both global and local contextual information, further improving overall detection performance [9].

Zhang et al. introduced SSG-YOLOv7, an enhancement of the YOLOv7 framework. Initially, they expanded the VisDrone2019 and RSOD datasets by simulating five types of challenging environments. The K-means++ algorithm was then applied to generate four anchor box sizes better suited for UAV targets. Subsequently, a parameter-free SimAM 3D attention module was integrated into the backbone network to improve the model's capacity to identify salient features. The SPPCSPC module was restructured to better fuse multi-scale pooling information, and standard convolutions were replaced with GhostConv. Finally, Soft NMS was employed instead of traditional NMS to optimize post-processing [10].

Luo et al. designed three improvement strategies based on the YOLOv5s model. By pairing these strategies, three multi-strategy models were constructed and evaluated through comparative experiments in complex environments. The results showed that Solution 2, which combines the Multi-Head Self-Attention (MHSA) mechanism and the SimAM attention mechanism, achieved the best overall performance [11].

3 Analysis and Discussion

3.1 Limitations of Small Object Detection

Although existing research has made notable progress in UAV object detection, several significant limitations and shortcomings persist. In optimization studies targeting small object detection models, Liang et al. employed VGG-16 as the backbone network. However, its feature extraction capability is inferior to that of modern architectures such as ResNet and EfficientNet, leading to relatively low computational efficiency. The use of deconvolution modules may introduce noise, and upsampling low-level features often fails to sufficiently restore details. Moreover, validation was conducted solely on the CARPK and SDD datasets, without encompassing a broader range of small object categories or testing robustness under extreme conditions.

The study by Li et al. resulted in a modified model with an increase of 2.795 million parameters compared to YOLOv5s, while the frames per second (FPS) decreased from 56 to 41. This elevated computational burden may impose limitations on deployment for drone platforms. Furthermore, the experimental validation was conducted solely on the VisDrone2021 dataset, lacking evaluation across diverse datasets or real-world scenarios. Although the dataset contains numerous small objects, it does not further categorize targets of different scales.

In Zhao et al.'s research, the VisDrone dataset was utilized, with small objects comprising 88.12% of the data. Nevertheless, visualization results revealed instances of missed detections, particularly for densely distributed small objects. The study only incorporated Mosaic data augmentation and did not introduce enhancement techniques specifically tailored for small object detection. Furthermore, experiments were limited to two UAV datasets, without validation on general-purpose datasets.

In Zhu et al.'s study, comparisons were made exclusively with YOLO series models, omitting both models specifically designed for small objects and state-of-the-art UAV detection models. After LAMP pruning, the mAP50 decreased from 32.2% to 30.2%, indicating a clear loss in accuracy. The improvement in detection accuracy for occluded objects was limited, with mAP50 reaching only 17.8%. Additionally, the study did not propose dedicated optimization strategies for extremely small objects.

3.2 Limitations Analysis in Complex Environments

In the optimization research of models designed for complex environments, several limitations remain evident. In Xi's work, the model's performance deteriorates significantly under adverse weather conditions such as fog and rain. Furthermore, the inference time of the model on the NVIDIA Jetson Xavier NX edge device is 509.8 ms, which does not satisfy real-time processing requirements. More importantly, the experiments relied on the VisDrone_Foggy dataset, which is synthetically generated and differs significantly from real-world foggy images, thereby limiting the reliability of the results. In addition, model validation was restricted to standard datasets (e.g., VisDrone, UAVDT) without evaluation in extreme scenarios, and no comparative analysis with state-of-the-art models was conducted.

Zhang's study simulated only five weather conditions (light fog, rain, dense fog, motion blur, and fog-rain) using the imgauge library, failing to cover a broader range of complex meteorological conditions. Moreover, only the SimAM attention mechanism was employed, without assessing the applicability of other mechanisms in complex environments. The experiments did not report AP values for small objects or the false detection rate, and lacked real-world testing in ultra-low visibility scenarios such as heavy rain and sandstorms, creating a gap between simulated results and practical applications. Additionally, the model's parameter count is high (28.5M), which hinders direct deployment on UAV platforms.

In Luo's study, only mAP50 was used as the accuracy metric, without evaluation under higher thresholds. The methodology involved only pairwise combinations of strategies, omitting a holistic integration of all three strategies, which may have resulted in missing the optimal solution. To present the specific challenges faced by

different models in various application scenarios more clearly and intuitively, the main limitations are summarized in Table 1.

Table 1. A Critical Analysis of the Limitations in UAV Object Detection Algorithm Models (Data from: this study)

Problem	Model	Shortcomings
Small objects	FS-SSD	The backbone network exhibits certain limitations; insufficient detail restoration capability; the experimental validation is incomplete.
	BA-YOLOv5s	An imbalance between parameter quantity and computational efficiency; the experimental validation is insufficient.
	Improved YOLOv7	The model demonstrates insufficient robustness in small object detection and exhibits limitations in its training strategy; the experimental validation is inadequate.
	Improved YOLOv8n	Experimental comparisons are incomplete, and the pruning strategy presents certain limitations; persistent missed detections of small objects remain an unresolved issue.
	FifoNet	The model exhibits poor robustness under adverse weather conditions, and its computational efficiency may be insufficient for real-time applications; the evaluation scenarios are limited, restricting the assessment of its generalizability.
Complex environments	SSG-YOLOv7	There are limitations in data augmentation, and only a single attention mechanism has been utilized; the experimental validation is insufficient, and the model faces challenges in deployment.
	Improved YOLOv5s	There are limitations in the combination of strategies, and only a single evaluation metric has been employed.

3.3 Recommendations for Model Enhancement

For small object detection, the backbone network of the FS-SSD model can be replaced with higher-performance architectures such as ResNet, and an attention mechanism can be introduced into the feature pyramid to enhance detail retention and reduce background interference. It is also recommended to validate the model on large-scale UAV datasets such as VisDrone and UAVDT, covering complex scenarios including rain, fog, low illumination, and motion blur. For BA-YOLOv5s, channel pruning can be applied to the MF module to reduce the number of parameters, and experiments can be conducted on remote sensing datasets such as DIOR and DOTA to improve generalization performance. For the improved YOLOv7 model, targeted adversarial training for small object detection can be implemented to enhance robustness. Additionally, data augmentation techniques specifically tailored for small object detection can be introduced, and the loss function can be optimized. It is recommended to evaluate mAP on general datasets while optimizing efficiency

without compromising accuracy. For the improved YOLOv8n model, comparative experiments can be conducted with models such as TPH-YOLO and Gold-YOLO, and tests can be performed on datasets like SeaDronesSee, with particular attention to performance in foggy and dynamic blur scenarios. A hierarchical pruning strategy combined with knowledge distillation is recommended to minimize accuracy loss during pruning.

For addressing complex environments, the FifoNet model can incorporate multi-source data such as infrared or radar to compensate for the lack of visible-spectrum information in adverse weather conditions. It is also recommended to evaluate the model on complex datasets such as NightOwls and DAIR-V2X to enhance its applicability in real-world UAV scenarios. For the SSG-YOLOv7 model, additional complex and harsh scenarios can be included to optimize the training data. A hybrid attention mechanism—such as combining SimAM with CBAM and embedding it into the backbone network—can be used to enhance the robustness of feature selection. Key performance metrics should also be supplemented in the experimental evaluation. For the improved YOLOv5s model, in addition to the existing three optimization strategies, further experiments should be designed to combine MHSA, BiFPN, and SimAM to verify whether their synergistic effect outperforms pairwise combinations. Ablation studies should be conducted to quantify the contribution of each module, and efficiency metrics such as FPS should be included to achieve a balanced evaluation of accuracy and speed. The above recommendations aim to improve the models in key aspects such as network architecture, training strategies, and data validation. For clarity, the specific improvement plans for each model are summarized in Table 2.

Table 2. Recommendations for the Enhancement of UAV Object Detection Algorithm Models (Data from: this study)

Problem	Model	Recommendations
Small objects	FS-SSD	Replace the backbone network with a higher-performing architecture, incorporate an attention mechanism, and expand validation to include diverse datasets and scenarios.
	BA-YOLOv5s	Adopt a lightweight design and conduct validation across multiple datasets.
	Improved YOLOv7	Enhance the capability for small object detection, optimize training strategies, and validate performance on general datasets.
	Improved YOLOv8n	Add comparative experiments, refine pruning operations, and enhance the detection performance for small objects.
Complex environments	FifoNet	Employ multimodal data fusion and incorporate more complex testing scenarios.
	SSG-YOLOv7	Optimize data augmentation techniques, adopt hybrid attention mechanisms, and include real-world data in the evaluation process.

Improved YOLOv5s	Design comprehensive ablation studies and incorporate additional evaluation metrics to ensure a thorough assessment of model performance.
------------------	---

4 Conclusions

This paper systematically reviews the research progress in object detection algorithms for UAVs. First, it provides an overview of two categories of deep learning-based object detection models: single-stage models and two-stage models. Next, it examines the key challenges faced in UAV object detection, including difficulties in detecting small objects, significant interference from complex backgrounds, stringent real-time requirements, and substantial variations in object scales. In response to these challenges, targeted optimization strategies—such as multi-scale feature fusion and the incorporation of attention mechanisms—are discussed. Based on the analysis presented above, this study proposes several directions for improving UAV object detection models. Key recommendations include designing lightweight network architectures and developing high-quality, task-specific datasets. These results provide theoretical support and technical guidance for improving object detection algorithms in UAV applications. They also help establish a practical foundation for deploying such systems in real-world scenarios, including smart city management, disaster emergency monitoring, and intelligent agricultural inspection.

References

1. Srivastava, S., Narayan, S., Mittal, S.: A survey of deep learning techniques for vehicle detection from UAV images. *Journal of Systems Architecture* 117, 102152 (2021).
2. Doll, O., Loos, A.: Comparison of object detection algorithms for livestock monitoring of sheep in UAV images. In: *Camera Traps, AI, and Ecology—3rd International Workshop*. LNCS, pp. 1–7. Springer, Heidelberg (2023).
3. Zhao, C.J., Liu, R.W., Qu, J.X., et al.: Deep learning-based object detection in maritime unmanned aerial vehicle imagery: review and experimental comparisons. *Engineering Applications of Artificial Intelligence* 128, 107513 (2024).
4. Li, Q., et al.: A Review of Object Detection Research in UAV Aerial Images. *Journal of Graphics* 45(6), 1145–1164 (2024).
5. Liang, X., Zhang, J., Zhuo, L., et al.: Small object detection in unmanned aerial vehicle images using feature fusion and scaling-based single shot detector with spatial context analysis. *IEEE Transactions on Circuits and Systems for Video Technology* 30(6), 1758–1770 (2019).
6. Li, L.X., Wang, X., Wang, J., Zhang, Y.Y.: Small Object Detection Algorithm for UAV Images Based on Feature Fusion and Attention Mechanism. *Journal of Graphics* 44(4), 659–666 (2023).
7. Zhao, D.W., Shao, F.M., Liu, Q., Yang, L., Zhang, H., Zhang, Z.H.: A small object detection method for drone-captured images based on improved YOLOv7. *Remote Sensing* 16(6), 1002 (2024).

8. Zhu, L.Z., Wei, H.: UAV Small Object Detection Algorithm Based on YOLOv8n. *Laser & Optoelectronics Progress* 62(10),1037009–1037009 (2025).
9. Xi, Y., Jia, W.J., Miao, Q.G., et al.: FiFoNet: fine-grained target focusing network for object detection in UAV images. *Remote Sensing* 14(16), 3919 (2022).
10. Zhang, R.M., Xiao, Y.F., Jia, Z.N., Chen, Z., Chen, Z.H., Yuan, B., Cao, W.W., Song, W.W.: Improved YOLOv7-based object detection algorithm in complex environments from UAV perspective. *Opto-Electronic Engineering* 51(5), 240051–240051(2024).
11. Luo, Q., Lu, R.C., Wang, Y.Z., Wu, X.M., Ge, J.X., Wang, F.: UAV Small Object Detection Based on Multi-Strategy YOLOv5s in Complex Environments. *Journal of Artificial Intelligence and Robotics Research* 14 (3), 590–604 (2025).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

