Design and Development of a Lightweight YOLOv11-Based Model for UAV Image Recognition

Yan YAN[1], Qi Li[2] and Lin Meng[2,*]

[1]Graduate School of Science and Engineering, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga, 525-8577, Japan
[2]College of Science and Engineering, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga, 525-8577, Japan

Abstract

As drone numbers rise and illegal flights become more common, security and privacy issues grow more serious. To monitor and manage drone flights effectively, this paper proposes YOLOv11-mini, a lightweight model improved from YOLOv11. By using GhostConv, C3Ghost, and pruning, YOLOv11-mini keeps high detection accuracy while cutting model size by 87%, making it fit for edge devices. This paper tests the model on a small custom nano-size drone dataset and applies data augmentation to boost performance. Results show that with augmentation, YOLOv11-mini increases mAP50 by about 4% over the unaugmented model, with accuracy only 2% lower than the original YOLOv11. This shows the model's strong advantages and potential in resource-limited settings.

Keywords

Nano-sized UVA, YOLOv11, Small object detection, Lightweight detection

1. Introduction

With the rapid increase in the number of drones, illegal drone flights continue despite repeated bans, frequently leading to incidents such as flight disruptions, public disturbances, and personal injuries. In addition to issuing relevant laws and regulations, it is also urgent to strengthen comprehensive supervision of drones [1]. However, achieving around-the-clock real-time monitoring of drones through manpower alone is difficult. Furthermore, due to their small size and high speed, drones can easily be misidentified or completely overlooked by human observers, especially under low-light conditions. Therefore, it is imperative to leverage more advanced and efficient technological solutions to achieve real-time monitoring and precise control of low-altitude drone activities, such as deploying edge devices for real-time target detection and identification.

In recent years, the rapid development of deep learning has driven continuous breakthroughs in object detection algorithms [2, 3, 4, 5, 6]. From the initial Region-CNN(R-CNN) [7], through Single Shot MultiBox Detector(SSD) [8], to the rapidly evolving YOLO series [9], each technological iteration brings new vitality and expands the possibilities of object detection.

The 7th International Symposium on Advanced Technologies and Applications in the Internet of Things (ATAIT 2025), August 10-11, 2025, Kusatsu, Shiga, Japan
*Corresponding author.
gr0662pf@ed.ritsumei.ac.jp (Y. YAN); liqi24@fc.ritsumei.ac.jp (Q. Li); menglin@fc.ritsumei.ac.jp (L. Meng)
0000-0003-4351-6923 (L. Meng)

Although R-CNN achieves high detection accuracy, its processing speed is slow, so it is not suitable for real-time applications [10]. SSD usually provides higher accuracy than YOLO. However, its large model size and high computational cost make it hard to meet the demands of scenarios that require strong real-time performance. In contrast, YOLO achieves an effective balance between accuracy and speed, making it especially suitable for real-time detection. In summary, lightweight improvements based on YOLO significantly enhance inference speed while maintaining detection accuracy, which makes it an ideal choice for latency-sensitive low-altitude applications such as real-time drone monitoring. This article adopts the YOLOv11 model proposed by Rahima Khanam and Muhammad Hussain in 2024 [10] as the baseline architecture. Although the original YOLOv11 achieves excellent detection accuracy and inference speed, its model size remains relatively large. For edge devices with limited computing resources, such as drones, this makes smooth deployment challenging and may lead to inference delays, falling short of real-time monitoring requirements. However, existing research mostly focuses on introducing attention mechanisms, improving feature extraction modules, or optimizing loss functions to enhance detection accuracy, while paying little attention to the deployment and adaptation of the YOLO model on edge devices with limited computing resources. This is particularly critical in anti-drone scenarios, where edge devices such as surveillance cameras and drones have constrained computing power, creating an urgent need to significantly reduce model size and computational overhead while maintaining detection performance.

Therefore, to achieve efficient monitoring and precise control of low-altitude drone activities, this paper proposes a lightweight object detection model based on YOLOv11, named YOLOv11-mini. By redesigning the backbone network and introducing lightweight convolution modules, the model size is reduced by approximately 87% compared to the original YOLOv11, enabling efficient deployment on resource-constrained edge devices. With the help of edge devices equipped with lightweight detection models, small drones in the monitoring area can be automatically detected and identified around the clock. This not only improves the efficiency and accuracy of monitoring but also significantly reduces the workload of manual patrols.

The main contributions of this paper are as follows:

- In response to the detection requirements for nano-sized drones in low-altitude flight scenarios, this paper independently collects high-resolution images and complete detailed annotations to construct a dedicated dataset. This fills the gap of insufficient publicly available data and lays the foundation for subsequent model training and evaluation.
- In the YOLOv11 framework, we remove the redundant detection layer and replace several standard convolutions with GhostConv and C3Ghost, thereby significantly reducing the number of parameters and computational cost. As a result, the model size reduce by approximately 87% compared to the original version, making it more suitable for deployment on edge devices with limited computing power.
- Systematically compares various data augmentation strategies (such as Mosaic and CutMix), selects the optimal combination, and effectively improves the model's robustness and generalization ability. Experiments show that the enhanced YOLOv11-mini achieves a 4% improvement in mAP50 compared to the unenhanced version, with merely a 2% decrease in accuracy relative to the original YOLOv11, thereby significantly reducing the computational burden while maintaining high detection performance.

In summary, through comprehensive improvements in dataset construction, lightweight network design, and data augmentation optimization, this paper enables the model to maintain high detection accuracy while reducing its model size by 87%, thereby meeting the requirements of real-time drone monitoring in resource-constrained devices.

The remainder of this paper is organized as follows. Section 2 reviews related work.Section 3 introduces the dataset and the YOLO models relevant to this study. Section 4 provides a detailed description of the data augmentation methods and evaluation metrics used. Section 5 presents the experimental results along with an analysis. Finally, Section 6 concludes the paper.

2. Related work

The rapid development of drone technology has driven in-depth research into low-altitude monitoring and counter-drone systems. The YOLO series of models has become a hot topic in research and applications in this field due to its effective balance between detection accuracy and real-time performance. To meet the needs of real-time monitoring in complex environments, many scholars have made various improvements to the YOLO architecture, proposing more adaptive detection methods.

2.1. Improvements for the YOLO series

Ghazlane Yasmine et al. propose a series of improvements based on the YOLOv7 model, incorporating the CSPResNeXt module into the backbone, a transformer block with the C3TR attention mechanism, and a decoupled head structure to enhance the model's performance. While ensuring an accuracy of 0.97, their model achieves an inference speed of 0.02 milliseconds per image, successfully achieving an optimal balance between inference speed and detection performance [11]. Xueqi Cheng et al. propose an IRWT-YOLO model based on YOLOv8 that integrates object detection and image segmentation, incorporates BiFormer into the backbone network, and introduces the RCSCAA and DCPPA modules to improve the detection of weak objects. The proposed model improves the robustness and effectiveness of the original model in detecting weak objects under complex infrared conditions, thereby addressing the problems of low object visibility and background interference in infrared UAV image detection [12]. Ruixi Liu et al. propose a distributed anti-drone system based on YOLOv5, which achieves automatic target locking through a mechanical structure, effectively improves detection accuracy, and adopts distributed cluster deployment to overcome the shortcomings of detection blind spots and target loss. This provides a deployment concept for airport countermeasures against lightweight UAVs and offers theoretical guidance for future anti-UAV strategies [13]. Juanqin Liu et al. propose a detection method called GL-YOMO, which combines the traditional YOLOv5 framework with multi-frame motion detection technology. It enhances the recognition of small drone targets by fusing features of different scales and introducing an attention module. In addition, they integrate the Ghost module into the network to further reduce computational cost and improve inference efficiency, thus achieving a better balance between accuracy and real-time performance and underscoring its potential in UAV detection applications [14].

3. Main work

3.1. Dataset creation

The image acquisition device is a Nikon Z50II mirrorless camera (APS-C DX format) equipped with a NIKKOR Z DX 18-140mm f/3.5-6.3 VR lens kit. The target is a nano-sized low-altitude UAV, namely the Bitcraze Crazyflie 2.1. The standard on nano-sized drone is shown in Figure 1. Images are captured from different heights and angles under bright indoor, dark indoor, and bright outdoor conditions. The self-constructed nano-sized drone dataset shows in Figure 2.

DRONES TAXONOMY BY VEHICLE CLASS SIZE

Class	Size:Weight [cm:kg]	Power [W]	Device
Std-size [*]	$\sim 50 : \geq 1$	≥ 100	Desktop
Micro-size [*]	$\sim 25 : \sim 0.5$	~ 50	Embedded
Nano-size [*]	$\sim 10 : \sim 0.01$	~ 5	MCU
Pico-size [*]	$\sim 2 : \leq 0.001$	~ 0.1	ULP

Figure 1: Drones taxonomy by vehicle class size

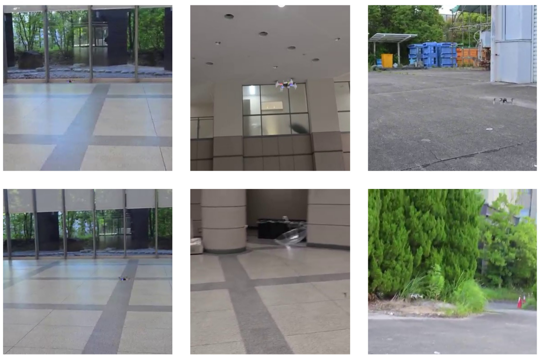

Figure 2: Presentation of a self-made nano-size drone dataset

The initial dataset consists of a total of 413 images of nano-sized drones. Python scripts are used to randomly sample 10% of the images as a validation set, 10% as a test set, and the remaining 334 images as the training set for model development and evaluation. In the training set, 173 shots are taken indoors and the remaining 139 shots are taken outdoors. All images are annotated using LabelImg to generate YOLO-format label files for nano-sized drones.

3.2. Yolov11 network structure

YOLOv1, proposed by Joseph Redmon in 2015, treats object detection as a regression problem, allowing the detection performance to be directly optimized end-to-end [15]. YOLOv11 is officially released on October 1, 2024, and is developed based on the YOLO series framework, leading to significant improvements in detection accuracy and efficiency.

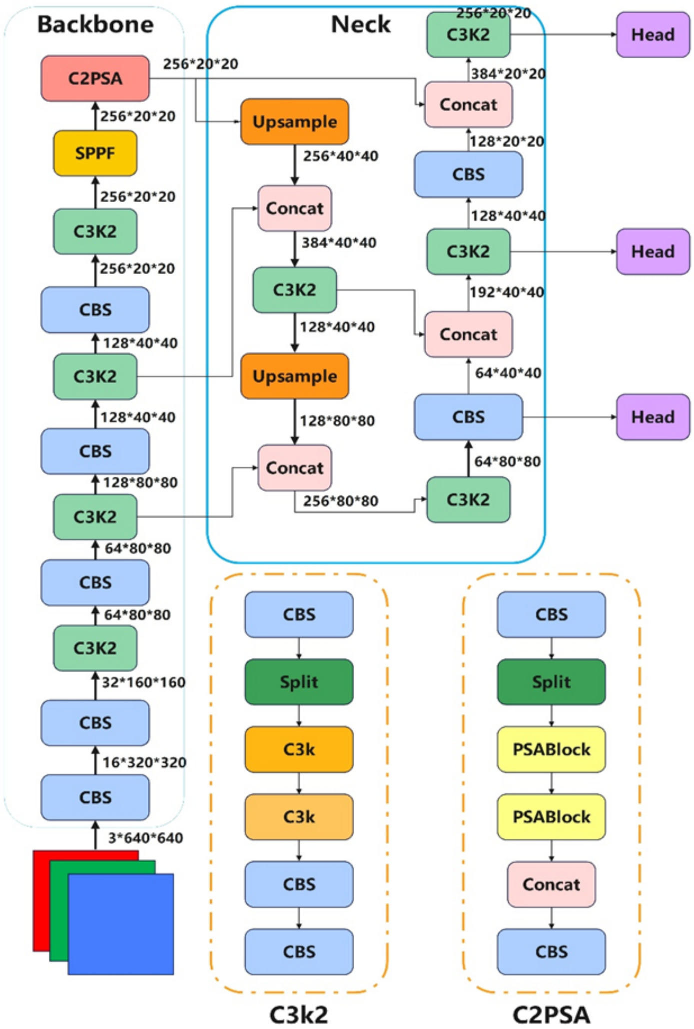

Figure 3: Overall architecture of the YOLOv11 network

Figure 3 shows that YOLOv11 introduces the C3k2 block and the convolutional block with parallel spatial attention(C2PSA) attention mechanism [10].

Among these components, C3k2 plays a key role in enhancing the feature extraction capability. It is an optimized version of the traditional CSP bottleneck structure in YOLOv11 [16], with its core feature being the use of two parallel convolutional layers. This design enables the extraction of features across different channels, thereby improving the model's adaptability to complex scenes. This makes data processing more efficient while maintaining high accuracy. C2PSA enhances multi-scale feature extraction by combining the Cross Stage Partial(CSP) structure and the Pyramid Squeeze Attention(PSA) mechanism. Additionally, it dynamically weights channel features through the Squeeze-and-Excitation(SE) mechanism, thereby strengthening the responses of important channels.

In summary, YOLOv11 achieves significant improvements in detection accuracy while maintaining real-time performance by integrating advanced convolution techniques and innovative attention mechanisms. This structural simplification not only significantly reduces the number of network parameters and computational overhead, thereby improving inference speed, but also prevents excessive aggregation of fine-grained small target information under a large receptive field. This helps to maintain the detection sensitivity and discrimination capability for small targets.

3.3. Lightweight improvement of the YOLOv11 model: YOLOv11-mini

We propose a lightweight model YOLOv11-mini based on YOLOv11 with its architecture shown in Figure 4.

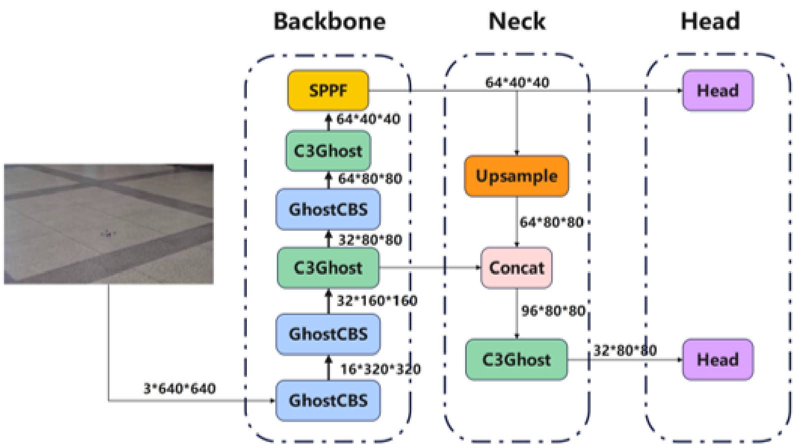

Figure 4: Overall network architecture of the improved YOLOv11mini

YOLOv11 performs well in object detection and other visual tasks due to its outstanding inference speed and high accuracy. However, when deployed on edge devices with limited computation resources, YOLOv11 still suffers from certain limitations. This paper draws on the Ghost module and feature layer pruning concepts to systematically prune and modify YOLOv11 [17], designing YOLOv11-mini, which retains only two output layers.

Based on YOLOv11, we set both the depth and width multipliers to 0.25 and reduce the number of layers in the neck and head, significantly decreasing the network's depth and width. To further reduce the computational burden, we replace the standard convolution with the lightweight GhostConv module and substitute the original C3k2 structure with C3Ghost, which effectively eliminates redundant computations, reducing the number of parameters and FLOPs.

In the standard YOLOv11 architecture, there are typically three detection heads, namely p3, p4, and p5, each responsible for detecting objects at different scales. In particular, the p5 output feature map, with its larger receptive field, is mainly used for detecting larger objects in images. However, in our target application scenarios, the drones are generally small and rely less on large-scale feature layers. Therefore, in designing YOLOv11-mini, we remove the p5/32 output, which is more sensitive to large object detection, and retain only the p3/8 and p4/16 output branches, which are better suited for detecting small and medium-sized objects.

This structural simplification not only significantly reduces the number of network parameters and computational overhead, thereby improving inference speed, but also avoids the excessive aggregation of fine-grained small target information under a large receptive field. This helps maintain the detection sensitivity and discriminative capability for small targets.

In addition, we retain the lightweight SPPF module to enhance the receptive field and feature aggregation capability while further reducing the model size. After these improvements, the lightweight YOLOv11-mini has a simpler network structure and a smaller model size, making it highly suitable for deployment on edge devices with limited computing resources.

4. Experimental data and evaluation metrics

In this section, we use a self-constructed drone dataset and evaluate multiple data augmentation methods to determine the optimal solution. In addition, we introduce the performance evaluation metrics employed in this study.

4.1. Experimental data and data augmentation methods

The computer operating system used in this experiment is Windows 11 Professional Edition. VSCode is employed to remotely connect to the server for training and testing. Python 3.10 is used as the primary programming language, and PyTorch 2.7.0 (CUDA 11.8) serves as the deep learning framework. The dataset is a nano-sized drone dataset that is self-collected and annotated, comprising a total of 413 images, with 334 used for training, 40 for testing, and 39 for validation.

This experiment uses a total of seven data augmentation methods, which can be divided into two categories: optical content transformation and geometric texture transformation. The visual effect is shown in Figure 5 and is described in detail as follows.

1. Optical content transformation

- Color-light: Slight adjustments are made to the hue, saturation, and brightness of the image. The amplitude is small, slightly altering the visual characteristics while maintaining the fundamental features of the original image, thereby improving the model's robustness to minor lighting changes.
- Color-medium: Building on Color-Light, it applies a wider range of brightness and contrast adjustments, yielding more noticeable visual effects and further enhancing the model's generalization ability.
- Color-medium-noise: Building on Color-Medium, Gaussian noise is superimposed to simulate sensor interference in real-world scenarios, enhancing adaptability to noisy environments.
- Color-medium-clahe: Building on Color-Medium, CLAHE is applied to improve local contrast, making it suitable for scenes with uneven lighting or unclear details.

2. Geometry Texture Transformation

- Mosaic: Randomly splicing 4 images into one to increase the number of target in each batch and to enrich the combination diversity, which is especially beneficial for small object detection.
- Mixup: The two images and their labels are linearly mixed in proportion to improve the model's robustness to partial occlusion and enhance its adaptability to complex scenes.
- Cutmix: Rectangular regions are randomly cropped and pasted between two images, and labels are mixed according to the area ratio to further improve the robustness and generalization ability of the model in complex scenes.

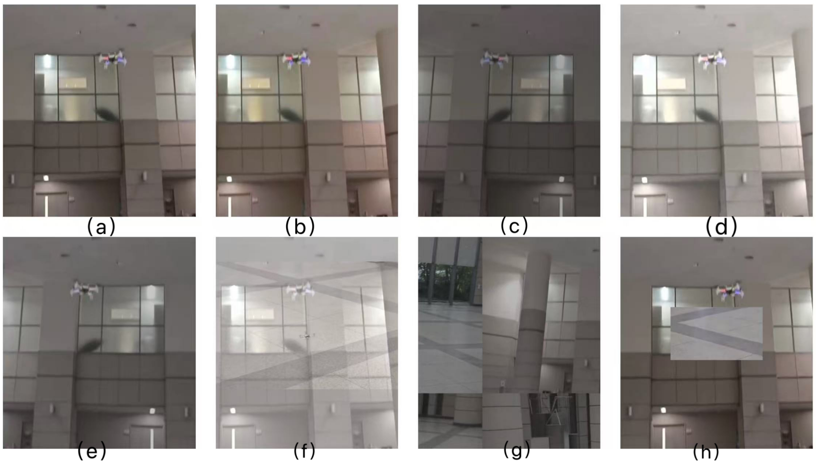

Figure 5: Examples of image data augmentation: (a) Original image; (b) Color-light; (c) Color-medium; (d) Medium-clahe; (e) Medium-noise; (f) Mixup; (g) Mosaic; (h) Cutmix.

4.2. Evaluation metrics

To evaluate the accuracy of the YOLOv11-mini model in recognizing nano-sized drones, this paper adopts performance metrics commonly used in object detection, including Precision, Recall, mAP50, mAP50-95, as well as the number of parameters and model size. The definitions of these metrics are provided below.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{1}$$

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{2}$$

$$AP = \sum_{n=1}^{N} (R_{n+1} - R_n) Precision_{max}(R_{n+1}) \tag{3}$$

$$mAP = \frac{1}{C} \sum_{j}^{C} AP_j \tag{4}$$

The Average Precision (AP) of all classes is the area of the region below the precision-recall curve. R_n represents the recall of the nth value, and $Precision_{max}(R_{n+1})$ represents the highest precision value in the range R_n to R_{n+1}. The mAP is calculated by averaging the AP of each class in the dataset. Specifically, mAP50 refers to the mean AP when the IoU threshold is fixed at 0.5, whereas mAP50-95 is calculated by averaging the mAP values over IoU thresholds ranging from 0.5 to 0.95.

5. Experimental results and analysis

In this section, we compare and analyze the performance of the YOLOv11 model before and after lightweight optimization, and evaluate the performance of the YOLOv11-mini model using various data augmentation strategies.

5.1. Model performance comparison

We use Parameters, Model Size, Precision, Recall, mAP50, and mAP50-95 as performance evaluation metrics to compare the YOLOv11 model before and after lightweight optimization. The results are shown in Table 1.

Table 1
Evaluation of model performance before and after lightweighting

Model	Parameters(M)	Size(MB)	Precision	Recall	mAP50	mAP50-95
YOLOv11	2.5	5.5	0.897	0.894	0.927	0.473
YOLOv11-mini	0.22	0.7	0.87	0.795	0.867	0.436

Compared to YOLOv11, the YOLOv11-mini model has significantly fewer parameters and a smaller model size. However, its performance metrics such as Recall and mAP50 are slightly lower than those of YOLOv11. To improve these metrics, we further apply data augmentation techniques to enhance the model.

We use Precision, Recall, mAP50, and mAP50-95 as performance evaluation metrics to assess the YOLOv11-mini model under eight different data augmentation strategies: base, light, medium, noise, clahe, mosaic, mixup, and cutmix. The results are shown in Table 2.As shown in Table 2, after applying Clahe data augmentation, the model achieves a precision of 1.0, a recall of 0.837, and an mAP50 of 0.905, demonstrating the best overall performance among the seven data augmentation methods and significantly improving detection accuracy. Notably, the Mosaic method yields the highest mAP50-95 value (0.49), while the Medium and Cutmix strategies also exhibit strong overall performance.

5.2. Ablation experiments and results analysis

After comparing the training effects of seven data augmentation methods on the YOLOv11-mini model, four methods that demonstrated better performance—medium color, clahe, mosaic, and cutmix—are initially selected. Subsequently, ablation experiments are conducted to compare the

Table 2
Model performance evaluation of different data augmentation methods

Data augmentation methods	Precision	Recall	mAP50	mAP50-95
Base	0.915	0.795	0.867	0.436
Light	0.938	0.777	0.867	0.45
Medium	0.915	0.821	0.884	0.446
Noise	0.939	0.821	0.868	0.472
Clahe	1	0.837	0.905	0.48
Mosaic	0.933	0.821	0.891	0.49
Mixup	0.894	0.795	0.861	0.434
Cutmix	0.921	0.795	0.886	0.476

training performance of YOLOv11-mini with different combinations of these data augmentation methods, aiming to determine the optimal augmentation strategy. The experimental results are shown in Table 3.

Table 3
Ablation experimental results

Data augmentation methods	Precision	Recall	mAP50	mAP50-95	fps(ms)
Clahe + Mosaic	0.91	0.795	0.869	0.491	0.6
Clahe + Mosaic + Cutmix	0.865	0.819	0.848	0.463	5.2
Clahe + Mosaic + Cutmix0.8	0.858	0.777	0.885	0.495	5.3
Medium + Mosaic	0.878	0.795	0.882	0.507	4.3
Medium + Cutmix	0.88	0.795	0.873	0.476	4.7
Medium + Mosaic + Cutmix	0.963	0.821	0.907	0.493	5.7
Medium + Mosaic + Cutmix0.8	0.906	0.795	0.881	0.478	5.7

The ablation experiments on various data augmentation methods and their combinations indicate that the combination of medium color augmentation, mosaic, and cutmix (1.0) achieves outstanding performance across all metrics. Specifically, it achieves a precision of 0.963, recall of 0.821, mAP50 of 0.907, and mAP50-95 of 0.493, significantly surpassing the baseline model (mAP50 = 0.876). Moreover, the combination of medium and mosaic yields the highest mAP50-95 (0.507), demonstrating more stable detection performance across different IoU thresholds. Overall, integrating medium, mosaic, and cutmix augmentation strategies effectively improves the detection accuracy and generalization ability of the YOLOv11-mini model for nano-sized UAV target detection. Future research can further optimize this augmentation strategy to fully exploit the model's potential.

5.3. Visualization of detection resultss

After adopting the combined data augmentation strategy of medium, mosaic, and cutmix (1.0), the YOLOv11-mini model proposed in this paper demonstrates excellent detection performance on the self-constructed nano-sized low-altitude drone dataset.

As illustrated in Figure 6, the model accurately locates drone targets across various low-altitude scenarios. Figure 7 presents the Precision–Recall curve on the test set, with an mAP0.5 of 0.906, indicating that the lightweight YOLOv11-mini still achieves high detection accuracy and recall while significantly reducing the model's complexity, making it suitable for deployment on edge devices with limited computation resources.

Figure 6: Visualization results of the proposed method on the nano-size uva dataset

6. Conclusion

Aiming at the real-time monitoring needs of nano-sized civilian drones in low-altitude scenarios, this paper proposes a lightweight improved YOLOv11-mini model based on YOLOv11, and constructs a dedicated dataset to support subsequent training and testing. Based on the original YOLOv11 architecture, GhostConv, C3Ghost, and a pruning strategy are introduced to reduce the model size by approximately 87%, significantly lowering the computational burden and making it more suitable for deployment on edge devices. Furthermore, a systematic comparison of multiple data augmentation methods and their combinations shows that the augmentation strategy combining Medium, Mosaic, and CutMix (1.0) effectively improves detection performance, with mAP50 increasing to 0.907 and Precision reaching 0.963. It also performs well on mAP50-95, thus verifying the effectiveness of the proposed approach. In summary, this study achieves

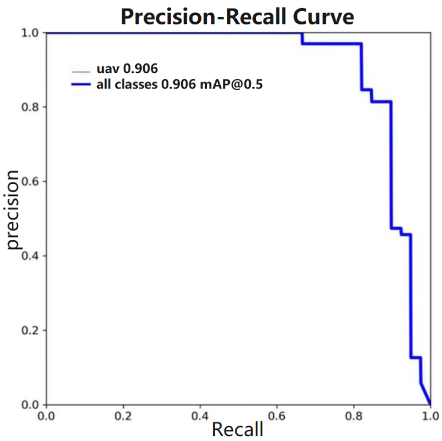

Figure 7: Test results of YOLOv11-mini on the test dataset after using the optimal data augmentation strategy

significant improvements in dataset construction, network lightweight optimization, and data augmentation, enabling YOLOv11-mini to greatly reduce model parameters and computational overhead while maintaining excellent detection accuracy and recall. This meets the requirements of real-time, resource-constrained low-altitude drone monitoring. Future work focuses on exploring more efficient attention mechanisms and small object detection strategies to further enhance the model's robustness and generalization in complex scenarios.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] S. Park, H. T. Kim, S. Lee, H. Joo, H. Kim, Survey on anti-drone systems: Components, designs, and challenges, IEEE access 9 (2021) 42635–42659.

[2] Y. Ge, Z. Li, X. Yue, H. Li, L. Meng, Dataset purification-driven lightweight deep learning model construction for empty-dish recycling robot, IEEE Transactions on Emerging Topics in Computational Intelligence (2025).

[3] Z. Li, Y. Yan, X. Wang, Y. Ge, L. Meng, A survey of deep learning for industrial visual anomaly detection, Artificial Intelligence Review 58 (2025) 1–82.

[4] Z. Li, Y. Ge, L. Meng, A multi-scale information fusion framework with interaction-aware global attention for industrial vision anomaly detection and localization, Information Fusion (2025) 103356.

[5] Z. Li, Y. Ge, X. Wang, L. Meng, 3d industrial anomaly detection via dual reconstruction network, Applied Intelligence 54 (2024) 9956–9970.

[6] Z. Li, Y. Ge, X. Yue, L. Meng, Mcad: Multi-classification anomaly detection with relational knowledge distillation, Neural Computing and Applications 36 (2024) 14543–14557.

[7] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961–2969.
[8] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, Ssd: Single shot multibox detector, in: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, Springer, 2016, pp. 21–37.
[9] N. Jegham, C. Y. Koh, M. Abdelatti, A. Hendawi, Evaluating the evolution of yolo (you only look once) models: A comprehensive benchmark study of yolo11 and its predecessors, arXiv preprint arXiv:2411.00201 (2024).
[10] R. Khanam, M. Hussain, Yolov11: An overview of the key architectural enhancements, arXiv preprint arXiv:2410.17725 (2024).
[11] G. Yasmine, G. Maha, M. Hicham, Anti-drone systems: An attention based improved yolov7 model for a real-time detection and identification of multi-airborne target, Intelligent Systems with Applications 20 (2023) 200296.
[12] X. Cheng, F. Wang, X. Hu, X. Wu, M. Nuo, Irwt-yolo: A background subtraction-based method for anti-drone detection, Drones 9 (2025) 297.
[13] R. Liu, Y. Xiao, Z. Li, H. Cao, Research on the anti-uav distributed system for airports: Yolov5-based auto-targeting device, in: 2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA), IEEE, 2022, pp. 864–867.
[14] J. Liu, L. Plotegher, E. Roura, C. de Souza Junior, S. He, Real-time detection for small uavs: Combining yolo and multi-frame motion analysis, IEEE Transactions on Aerospace and Electronic Systems (2025).
[15] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.
[16] J. P. Polinar, N. M. Al Jastin, S. J. A. Daño, A. M. Aparicio, Deep learning approach for weed detection to determine soil condition, in: 2025 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream), IEEE, 2025, pp. 1–5.
[17] R. Xiao, H. Wang, L. Wang, H. Yuan, C3ghost and c3k2: performance study of feature extraction module for small target detection in yolov11 remote sensing images, in: Second International Conference on Big Data, Computational Intelligence, and Applications (BDCIA 2024), volume 13550, SPIE, 2025, pp. 464–470.