Research on Small Object Detection in UAV Aerial Imagery Based on Improved YOLOv11

Bitong Liu
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: liubitong@st.xatu.edu.cn

Yaoxuan Yuan
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: smartyuan@126.com

***Abstract*—To enhance the robustness of small object detection in UAV aerial imagery, this paper proposes an improved detection method based on the YOLOv11 architecture. Initially, to mitigate the occlusion of small-target features by surrounding background clutter, a Multi-scale Edge Information Enhancement module is introduced to amplify fine-grained local details. Subsequently, an efficient feature fusion architecture is constructed to achieve comprehensive integration of global contextual semantics with small-object feature representations. Furthermore, a Task-Dynamic Aligned Head (TDAH) based on shared convolutions is proposed to mitigate the inconsistency between the classification and regression tasks. Finally, a loss function named WSIoU, which incorporates dynamic focusing and shape-aware constraints, is introduced to reduce the interference of low-quality samples during model optimization. Experimental results on the VisDrone2019 dataset confirm that the proposed method not only achieves a 4.72% improvement in mAP@0.5 but also provides a practical and viable enhancement strategy for small object detection from a drone perspective.**

Keywords-Small Object Detection; Feature Enhancement; Feature Pyramid; Task Aligned

I. INTRODUCTION

Owing to their maneuverability, expansive aerial coverage, and capacity for on-board analysis, Unmanned Aerial Vehicles (UAVs) offer distinct benefits across domains ranging from security surveillance to emergency disaster relief [1]. Consequently, vision-based object detection for UAVs has gained widespread application. However, small object detection from a drone perspective remains a formidable challenge. Firstly, due to the long shooting distance, small objects occupy minimal pixels in images, resulting in an extremely low signal-to-noise ratio. Their discriminative features, such as texture and shape, are inherently weak, making them difficult to extract effectively by conventional convolutional networks. Secondly, UAV aerial scenes often contain complex and diverse backgrounds. Small objects are prone to feature confusion with background textures and are susceptible to environmental interference like illumination changes and shadows. Thirdly, targets in practical scenarios are often densely distributed with severe occlusion and overlap. Meanwhile, objects exhibit a significant scale variance, with small, medium, and large targets coexisting, posing great difficulty for detectors to maintain accuracy across all scales. These factors collectively undermine the detection model's accuracy and robustness. Consequently, there is an urgent need for a detection method capable of achieving high accuracy in such complex environments to enhance the reliability of UAV vision systems.

The proportion of small objects in images was improved through image partitioning [2], which enhanced the detection performance to some extent, but also inevitably introduced additional noise, thereby compromising the overall performance. Zhai et al. [3] proposed the GAM attention mechanism, which effectively enhances feature fusion for small objects. However, this method introduces substantial computational overhead, hindering its deployment on resource-constrained UAV platforms. Despite advances in the performance of small object detection, most existing studies overlook the resource constraints in computational power and memory inherent to UAV platforms. The challenge of improving accuracy while ensuring model lightweight design persists.

To tackle the aforementioned challenge, this paper introduces an enhanced detection framework tailored for small object perception in UAV imagery. The proposed approach achieves robust accuracy under intricate environmental conditions while incurring modest computational overhead, thereby offering considerable practical utility. The principal contributions of this work are outlined below:

(1) This study devises a specialized Edge Information Enhancement unit spanning multiple resolutions—termed MSEIE—to redirect the network's attention toward the elusive signatures of tiny objects. This design markedly improves the model's responsiveness to minute targets, boundary details, and cluttered backgrounds.

(2) Down-sampling operations inevitably erode the subtle signatures of tiny objects. To counteract this effect, we construct an efficient feature pyramid that merges abstract semantic context with granular spatial cues, preserving low-level fidelity without incurring prohibitive computational costs.

(3) To mitigate the risk of overlooking or misidentifying small targets against cluttered scenes, a detection head featuring dynamic alignment of regression and classification objectives is introduced. This design helps alleviate the performance degradation stemming from conflicting optimization goals in decoupled detection architectures.

(4) Aiming to alleviate the problem of loss function over-penalization due to the inherent scale constraints of small targets, this study introduces a composite loss function that combines a dynamic focusing mechanism with shape-aware constraints, thus reducing the disruptive influence of low-quality samples during training.

II. Related Work

A. Object detection network

Object detection approaches typically split into two types: traditional and deep learning based. The former heavily use handcrafted features. Their main advantage is a lower need for computing power. However, their representational ability is restricted, and they are easily disturbed by different lighting conditions, object poses, and occlusions. In contrast, Convolutional Neural Network (CNN)-based object detection methods [4] achieve joint optimization of feature extraction and classification. CNNs possess a powerful capability for automatic feature extraction, enabling them to learn richer shape, texture, and semantic information directly from input images, without dependence on manual feature engineering. Currently, mainstream CNN-based object detection algorithms can be divided into two categories: two-stage algorithms and one-stage algorithms. Two-stage object detection algorithms first generate region proposals on the original image and then perform classification and regression on these regions [5]. These algorithms often demonstrate high accuracy in small object detection tasks. However, they suffer from large computational overhead, slow inference speed, and strong dependence on hardware resources, making them difficult to deploy on resource-constrained UAV systems. One-stage object detection algorithms, such as the YOLO series, perform classification and regression directly on feature maps, bypassing the region proposal stage. This design grants them significant speed advantages, making them highly suitable for real-time [6] UAV applications and earning them widespread attention for their excellent speed-accuracy trade-off.

B. UAV Aerial Photography Small Target Detection Method

Existing networks primarily rely on convolutions for feature extraction. However, this approach often leads to insufficient extraction of global information [7] and an increased loss of fine-grained details for small objects as the network[6] deepens. In response, researchers have proposed various improvement strategies, such as refining network architectures, optimizing loss functions, and enhancing feature fusion mechanisms.

The High-Resolution Feature Pyramid Network [8] is proposed for UAV-view object detection. It is designed to effectively handle scale variations and minimize feature redundancy. While Deeper and Wider YOLO [9] improves feature extraction for aerial images through stage-wise residual block optimization and expanded convolutional kernels, its performance is still constrained by complex environmental conditions. A shallow feature fusion mechanism was specifically constructed in

DBF-YOLO for drone-based target detection, contributing to markedly enhanced recall and precision metrics [10]. Despite these improvements, the model exhibits limited robustness in complex scenes characterized by dense target distributions. Improved YOLOX-X Network [11] improves small target detection performance through a combination of attention mechanisms, data augmentation, shallow feature maps, and loss function optimization. The trade-off for this combined strategy is an increase in computational demand and a reduction in processing frame rate.

The detection of dense small objects in UAV-based aerial images remains a non-trivial problem [12]. To tackle these long-standing difficulties, this paper puts forward a novel framework for UAV detection, featuring an algorithm that is specially adapted to dense small targets in aerial views.

III. Proposed Model

A. Overview of Proposed Model

The comprehensive structure of the introduced framework is depicted Figure 1. It is built upon the YOLOv11n framework, which has been further optimized and extended. YOLOv11 adopts an anchor-free architecture, inheriting the design philosophy of YOLOv8. Within this framework, the detector formulates predictions via direct regression of bounding box parameters—center coordinates, dimensions, and class likelihoods—thus dispensing with the requirement for predefined anchor templates. This anchor-agnostic paradigm is especially vital when localizing small objects. The reason is twofold: such targets cover extremely few pixels, and their scale statistics cluster around minuscule values. This makes it difficult for anchor boxes to achieve precise matching, thereby exacerbating the imbalance in positive and negative sample assignment and increasing regression errors. The anchor-free architecture of YOLOv11 reduces dependence on anchor box hyperparameters, demonstrating higher adaptability and robustness when processing small objects and high-resolution images. Consequently, it possesses inherent suitability and application potential for drone-based small object detection scenarios. YOLOv11n is the most lightweight version within the YOLOv11 series, featuring the fewest parameters, the smallest computational load, the most compact model size, and the fastest inference speed. Its efficiency profile renders it particularly compatible with hardware of limited computational capacity and scenarios requiring rapid, high-frequency processing. For these reasons, YOLOv11n is selected as the foundational reference in this study.

B. Multi-Scale Edge Information Enhancement (MSEIE) Module

Small objects occupy only a minimal number of pixels in an image, resulting in insufficient feature representation. Furthermore, they are often occluded by complex backgrounds or larger objects, making it difficult for the network to attend to their subtle and weak feature signals. To mitigate the issue of inadequate feature representation for small objects and to better capture their fine-grained details, a Multi-Scale Edge Information Enhancement (MSEIE) module is proposed. This module is integrated into the C3K2 block of the backbone network, forming a new composite block named C3K2-MSEIE. The MSEIE module enhances the model's perceptual capacity for small objects, edge details, and complex scenes through multi-scale edge information enhancement and strategic feature fusion.

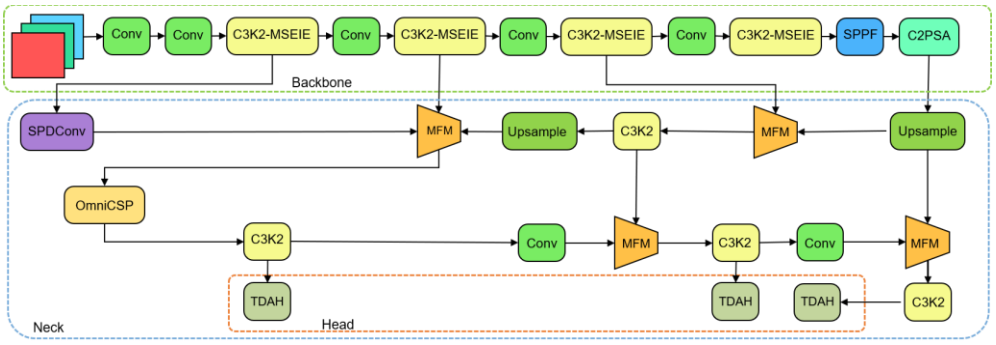

Figure 1. Network architecture of Proposed Model

The core component of the MSEIE module is the EdgeEnhancer (EE), designed based on high-pass filtering and edge detection principles to amplify edge information, which is crucial for small object detection. For an input feature map $X \in \mathbb{R}^{C\times H\times W}$, where C, H and W represent the channels, height, and width of the feature map, respectively. The EE can be summarized as follows:

$$X_{smooth} = AvgPool2d_{3\times3}(X) \quad (1)$$

$$X_{edge} = X - X_{smooth} \quad (2)$$

$$F_{out} = X + \sigma(Conv(X_{edge})) \quad (3)$$

As shown in Figure 2. , this module extracts enhanced edges by subtracting a smoothed feature map from the original, further refines them via convolution, and finally fuses them with the original features through addition.

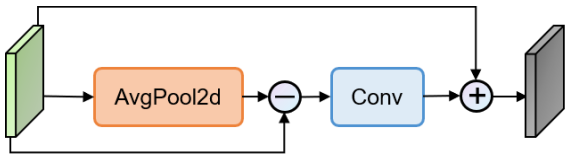

Figure 2. EdgeEnhancer

The complete architecture of the MSEIE module is shown in Figure 3. We design four scale branches of 3×3, 6×6, 9×9, and 12×12 to extract contextual information with different receptive fields. Each scale independently passes through the EE module for enhanced discrimination, and finally, the feature information from all scales is concatenated and compressed through a 1×1 convolution for output feature.

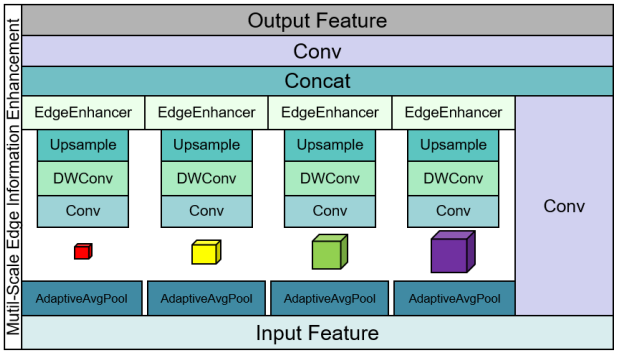

Figure 3. Multi-Scale Edge Information Enhancement Module

MSEIE's main innovation is the joint design of multi-scale context modeling and edge enhancement. This design helps the model better preserve features of small objects, reduce information loss during down sampling, and improve robustness against complex backgrounds. As a result, it achieves stronger performance in edge sensitive detection tasks.

C. Detail Enhancement Feature Pyramid Network (DEFPN)

Small objects retain high spatial resolution in shallow features but lack sufficient semantics. In deeper features, while semantic expression is stronger, the features of small targets are often severely compressed or even vanish due to their limited size. For small object detection, relying solely on three detection layers, P3, P4, and P5, is often insufficient for meeting feature resolution requirements. During progressive down-sampling at these layers, part of the detailed information is lost, limiting the network's representation capacity for small objects and consequently leading to suboptimal detection performance. Recent research adds a P2 detection layer to improve small object detection using shallow features, but this increases computational cost. To address this, we propose DEFPN, a neck network that enhances small object detection while controlling complexity, as shown in Figure 4.

DEFPN incorporates the P2 layer's shallow features while achieving an accuracy-speed balance via structural optimization. To avoid spatial information loss from conventional down-sampling, DEFPN employs a Space-to-Depth Convolution (SPDConv) module, which redistributes spatial information into channels, preserving detail crucial for small objects. Instead of traditional concatenation, DEFPN uses a Modulation Fusion Module (MFM) that weights channels adaptively before fusion. This produces feature maps with higher information density and fewer channels, reducing computational overhead. Let $\{X_i\}_{i=1}^{H}$ denote H input feature maps, where $X \in \mathbb{R}^{C_i\times H\times W}$, we first unify their channel numbers to C using a 1×1 convolution, obtaining $\hat{X}_i \in \mathbb{R}^{C\times H\times W}$. We

compute the element-wise sum across the H feature maps: $S = \sum_{i=1}^{H} \hat{X}_i$, $S \in \mathbb{R}^{C \times H \times W}$. The vector S is fed into a multi-layer perceptron (MLP) to obtain attention weights $\alpha \in \mathbb{R}^{H \times C \times 1 \times 1}$. Finally, the output feature map $Y \in \mathbb{R}^{C \times H \times W}$ is computed as a weighted sum: $Y = \sum_{i=1}^{H} \alpha_i \odot \hat{X}_i$.

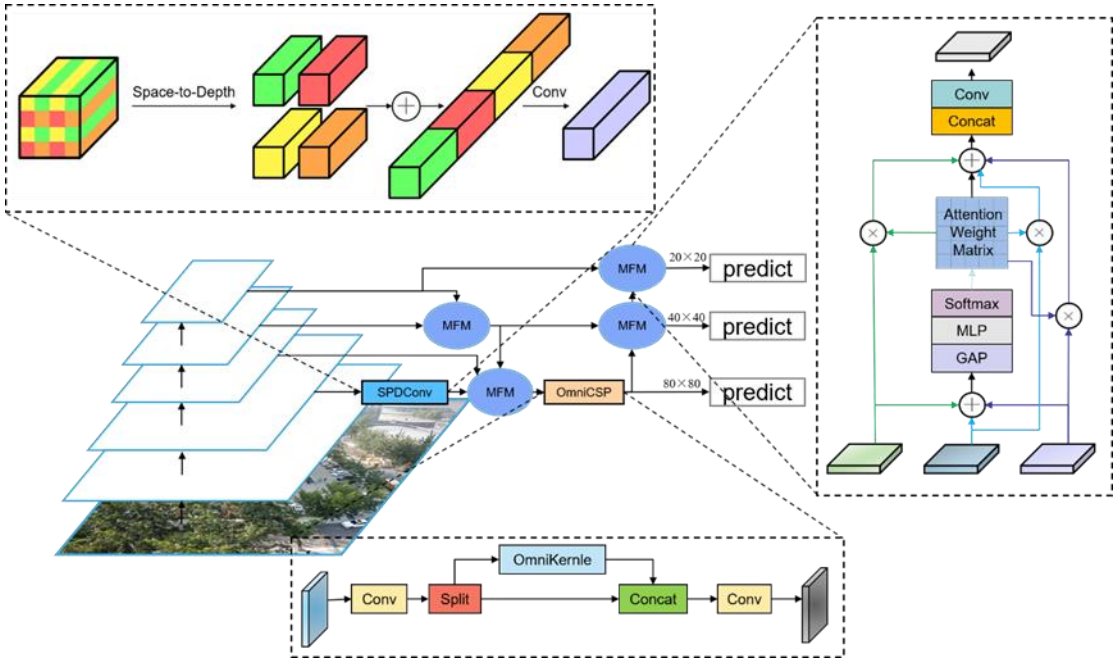

Figure 4. Detail Enhancement Feature Pyramid Network (DEFPN)

Due to their limited size, small objects often lack sufficient contextual features and can become submerged in complex background information during down-sampling. The OmniKernel module [13] addresses this by incorporating frequency-domain convolution and large-kernel convolution, which extend the receptive field far beyond that of traditional convolutions. This enables the network to better distinguish small targets from cluttered backgrounds, though at the cost of increased computational overhead. Inspired by Cross Stage Partial (CSP), we design OmniCSP, which splits input channels such that one branch preserves features directly while the other processes them through OmniKernel. This design maintains information flow while reducing computational cost and improving gradient flow.

Figure 5. shows DEFPN achieves a larger receptive field than PAN-FPN at P3, capturing richer context.

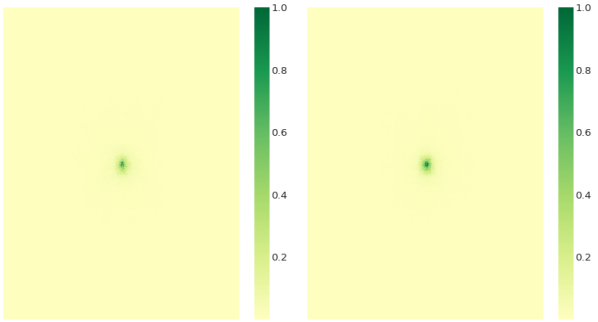

Figure 5. Receptive fields of the left PAN-FPN and the right DEFPN

D. Task Dynamic Aligned Head (TDAH)

YOLOv11 adopts a decoupled detection head design. However, this separation can lead to inconsistency between the predictions, negatively impacting small object detection performance. The underlying cause lies in the inherent differences in objective properties and feature requirements between the two tasks. To address this, we propose a Task-Dynamic Aligned Head (TDAH). We also incorporate shared convolutions to keep the

parameter count low, leading to the final improved TDAH structure shown in Figure 6.

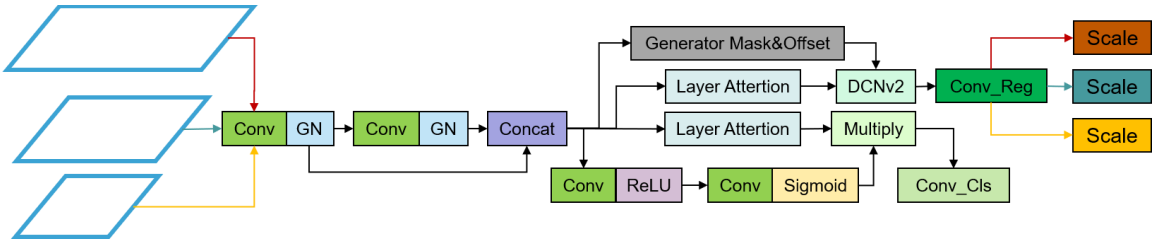

Figure 6. Task Dynamic Aligned Head (TDAH)

The feature outputs from DEFPN at the P3, P4, and P5 levels undergo a series of convolutions and normalization operations to form interaction features. As shown in Figure 7. for each scale of the input feature map X_i, we first apply two consecutive 3×3 group normalization convolutions to obtain intermediate features, which are then concatenated along the channel dimension to form a shared representation F_i. A global average pooling operation extracts a channel-wise descriptor from F_i, which, together with F_i itself, is fed into two separate task decomposition modules to generate classification-specific features F_i^{cls} and regression-specific features F_i^{reg}. To spatially align the regression features, we learn a set of offsets and modulation masks from F_i via a standard convolution, and then apply a deformable convolution (DCNv2) to F_i^{reg} produce aligned regression features. For the classification branch, we compute a spatial attention map from F_i. For the classification branch, we compute a spatial attention map from F_i using a small two-convolution subnetwork followed by a sigmoid function, and multiply it with F_i^{cls} to obtain aligned classification features. Finally, a 1×1 convolution maps the aligned regression features to channels, anther 1×1 convolution maps the aligned classification features to N_c class scores, and the two outputs are concatenated to form the final prediction tensor for that scale. Through this adaptive sampling mechanism, the model's receptive field can automatically adjust according to the target structure, thereby enhancing localization accuracy.

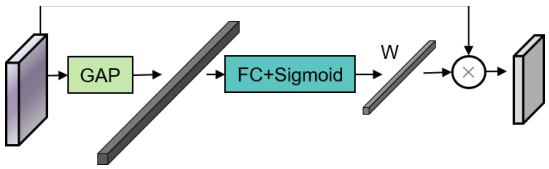

Figure 7. Layer Attention

TDAH incorporates shared convolutions between its classification and regression branches, replacing the separate convolutional layers used in traditional decoupled heads. This design reduces parameter redundancy and computational cost while promoting collaborative use of low-level features, enhancing feature compactness and task consistency. To address output scale inconsistencies across detection heads, a learnable scale factor is added to the regression branch for adaptive feature adjustment, ensuring prediction stability. The classification branch utilizes an adaptive weighting mechanism to focus on discriminative information, improving classification accuracy. By balancing lightweight design with task-specific adaptation, TDAH offers an effective structure for UAV-based small object detection.

E. Proposed Loss Function

In UAV-based small object detection, the minimal pixel scale of bounding boxes means even slight center or aspect-ratio deviations can cause CIoU to over-penalize predictions, destabilizing training. To mitigate this, WIoUv3 introduces a dynamic non-monotonic focusing mechanism that reduces excessive penalties from IoU drops in small objects. This approach suppresses gradients from low-quality samples while emphasizing those from

moderate-quality ones, promoting more stable and efficient training. Additionally, ShapeIoU incorporates shape and scale awareness into IoU calculation, offering consistent quality assessment even under low overlap, which helps alleviate over-penalization due to small target scales. By integrating WIoUv3 and ShapeIoU, this paper proposes a WSIoU regression loss, formulated as follows:

$$L_{ShapeIoU} = (1 - IoU) + \frac{d_c}{c^2} + \frac{\left(1 - e^{-\omega_w}\right)^4 + \left(1 - e^{-\omega_h}\right)^4}{2} \tag{4}$$

$$\beta = \frac{1 - IoU}{\overline{(1 - IoU)}} \tag{5}$$

$$scale = \frac{\beta}{\delta \cdot \alpha^{(\beta - \delta)}} \tag{6}$$

$$L_{WSIoU} = scale \cdot L_{ShapeIoU} \tag{7}$$

In the formula, d_c represents the weighted center distance between the predicted box and the ground-truth box. The symbols ω_w and ω_h indicate the difference in width-height ratios. Several hyperparameters are used for tuning. The computational graph of WSIoU is illustrated in Figure 8.

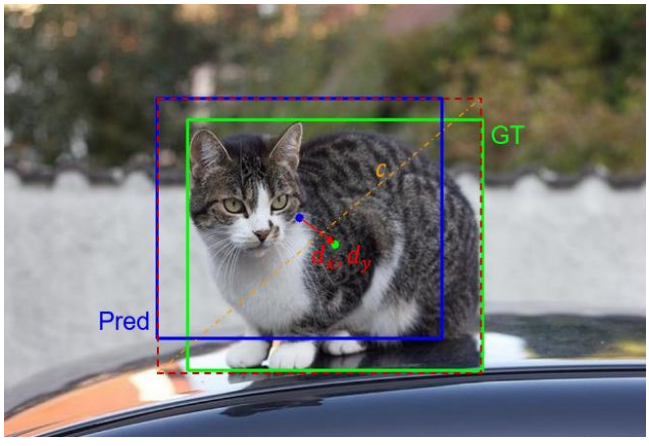

Figure 8. Computation graph of WSIoU

WSIoU integrates a dynamic focusing mechanism with shape-aware constraints, effectively mitigating the insufficient localization accuracy of traditional loss functions in small object scenarios. This combination not only enhances the sensitivity of the loss function to fine-grained features of small targets but also improves the stability of the training optimization process. Consequently, WSIoU demonstrates superior localization accuracy and convergence behavior in small object detection.

IV. Experimental Design and Result Analysis

A. Experimental Environment and Dataset

Our experimental setup used Ubuntu 20.04, a 16-core Intel Xeon Gold 6430 CPU with 120 GB RAM, an NVIDIA RTX 4090 GPU (24 GB), PyTorch 2.0.1+cu118, and Python 3.9.23. The baseline model for the experiments is YOLOv11n, the detailed parameter configuration is listed in TABLE I. To guarantee impartiality, we applied identical experimental setups and hyperparameter values during training and testing.

TABLE I. EXPERIMENTAL PLATFORM

Hyper-parameters	Value
Inputs	640x640
Epochs	300
Batchsize	32
Lr0	0.01
Lrf	0.0001
Amp	False
Momentum	0.937
Optimizer	SGD

This experiment employs the VisDrone2019 dataset to evaluate the performance of the improved YOLO algorithm. VisDrone2019 is a comprehensive multi-task dataset for UAV perspectives. The dataset was gathered using various UAV platforms across diverse scenarios, weather conditions, and lighting environments. It includes both urban and rural areas, as well as scenes with high and low object densities. The data split contains 6,471 images for training, 548 for validation, and 1,610 for testing. With numerous small objects and cluttered scenes, this dataset

effectively captures the typical difficulties of small-object detection and directly corresponds to the main research focus of this paper.

B. Evaluation

This study primarily employs Precision (P), Recall (R), mAP@0.5, GFLOPs to evaluate model performance.

C. Ablation Study

We conducted a comprehensive ablation study, and the experimental results are shown in TABLE II.

TABLE II. RESULTS OF ABLATION EXPERIMENT

Model	A	B	C	D	mAP@0.5	GFLOPs
1					0.3326	6.3
2	√				0.3414	6.3
3		√			0.3514	9.1
4			√		0.3535	8.1
5				√	0.3385	6.3
6	√	√	√	√	0.3798	10.8

Note: A: Multi-Scale Edge Information Enhancement (MSEIE); B: Detail Enhancement FPN (DEFPN); C: Task-Dynamic Aligned Head (TDAH); D: Proposed WSIoU Loss.

According to the experimental outcomes, every proposed improvement module yields a certain degree of performance gain for the baseline model, and their joint use produces synergistic benefits. Specifically, the baseline model achieves an mAP@0.5 of 33.26%. Replacing the C3K2 block in the backbone with C3K2-MSEIE increases mAP@0.5 by 0.88%. Incorporating the DEFPN module further raises mAP@0.5 by 1.88%. Subsequently, adding the TDAH contributes an additional 1.09% gain in mAP@0.5, indicating that TDAH enhances collaboration between the classification and regression tasks through task-dynamic alignment, effectively mitigating the prediction inconsistency inherent in decoupled detection heads. Finally, replacing CIoU with WSIoU improves the model's sensitivity to small object features. The complete improved model achieves a 4.72% improvement in mAP@0.5 over the baseline model. To more intuitively represent the overall detection performance, we also plot the comprehensive Precision-Recall (P-R) curves for all categories, as shown in Figure 9.

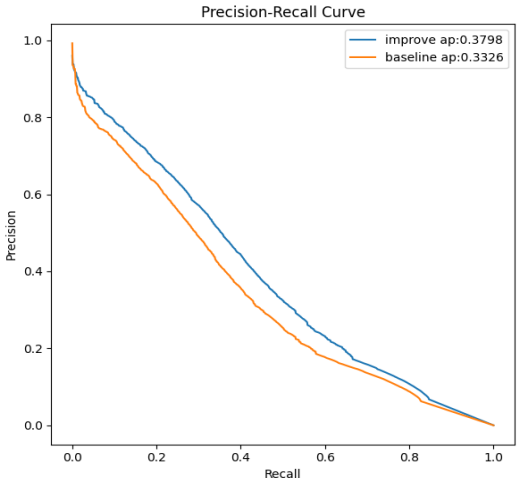

Figure 9. Precision-Recall curves for all categories

D. Comparative Experiments with Different Detection Algorithms

We compare our enhanced model with several leading lightweight approaches on the VisDrone dataset to further verify its advantage in detecting small objects under UAV conditions. As shown in TABLE III. , Ours achieves competitive detection accuracy while maintaining high efficiency—its computational cost is only 59% of that of YOLOv11s. The results confirm that improved model effectively balances accuracy and efficiency, underscoring its practical value and competitiveness for real-time UAV applications.

TABLE III. THE RESULTS OF COMPARATIVE EXPERIMENTS ON DIFFERENT DETECTION ALGORITHMS

Model	P	R	mAP@0.5	GFLOPs
SSD	0.211	0.358	0.24	63.2
YOLOv8n	0.431	0.334	0.33	8.1
YOLOv9t	0.455	0.333	0.34	7.6
YOLOv10n	0.435	0.326	0.328	6.5
YOLOv11n	0.445	0.332	0.3326	6.3
YOLOv11s	0.499	0.384	0.391	18.3
Ours	0.484	0.375	0.3798	10.8

E. Visualization results

We used heatmap visualization to compare the proposed model against the baseline, offering a more intuitive validation of our method's detection

performance. As shown in Figure 10. , it can be observed that the original YOLOv11n exhibits certain limitations in detecting small objects, particularly under conditions of dense target distribution and complex backgrounds where distant targets are easily missed. In contrast, ours focuses more precisely on small targets, enhancing target awareness and effectively suppressing background interference, leading to significant overall performance improvement.

Image YOLOv11n Ours

Figure 10. Heatmap visualization

Furthermore, we selected several practical cases to demonstrate how our model performs when dealing with challenges such as large variations in target scale, complex backgrounds, dense small objects, and severe occlusion. As shown in Figure 11. , the proposed model successfully detected a number of occluded targets with high confidence. This visually demonstrates that the introduced improvements effectively reduce missed detections caused by occlusion or extremely small target sizes, thereby validating that our enhancements strengthen the model's capability to recognize true small objects, lower both omission and false alarms, and enhance overall robustness.

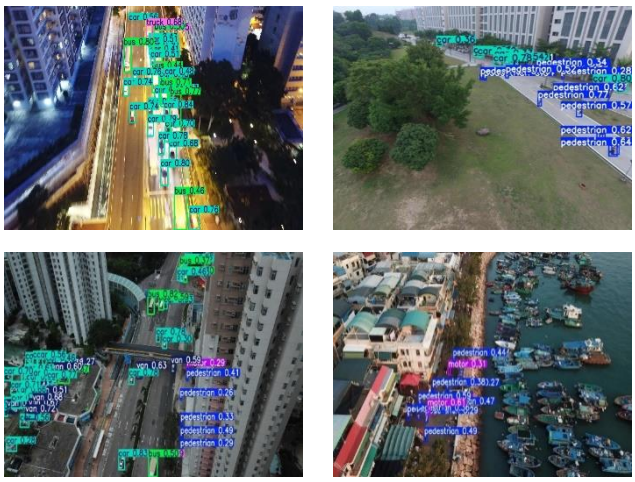

Figure 11. Visualization results

In sharp contrast, the improved model effectively suppresses these false alarms in the same complex environments while detecting more genuine small objects. This fully confirms that the proposed enhancements strengthen the model's ability to perceive true small targets, reduce both missed and false detections, and improve overall robustness.

V. Conclusions

To improve the detection accuracy of small objects in UAV scenarios, this study proposes the improved model. Four key improvements are integrated into the baseline: a C3K2-MSEIE module to enhance perception in complex scenes; a DEFPN structure to preserve fine-grained details and context; a TDAH mechanism to align classification and regression tasks; and a WSIoU loss to stabilize training with low-quality samples. Experimental results demonstrate that ours achieves significant accuracy gains across multiple benchmarks. Future work will explore multimodal fusion and more efficient lightweight designs to further advance practical deployment in resource-constrained scenarios.

References

[1] F. Y. Shih, AI Deep Learning in Image Processing. Boca Raton: CRC Press, 2025. doi: 10.1201/9781003474395.

[2] G. Cheng et al., “Towards Large-Scale Small Object Detection: Survey and Benchmarks,” IEEE Trans. Pattern Anal. Mach. Intell., pp. 1–20, 2023, doi: 10.1109/TPAMI.2023.3290594.

[3] “YOLO-Drone: An Optimized YOLOv8 Network for Tiny UAV Object Detection.” [Online]. Available: https://www.mdpi.com/2079-9292/12/17/3664 Accessed: Apr. 18, 2026.

[4] S. Zhou, H. Zhou, and L. Qian, “A multi-scale small object detection algorithm SMA-YOLO for UAV remote sensing images,” Sci Rep, vol. 15, no. 1, p. 9255, Mar. 2025, doi: 10.1038/s41598-025-92344-7.

[5] C. Wang, Z. Li, Q. Gao, T. Cui, D. Sun, and W. Jiang, “Lightweight and Efficient Air-to-Air Unmanned Aerial Vehicle Detection Neural Networks,” in 2023 IEEE International Conference on Unmanned Systems (ICUS), Oct. 2023, pp. 1575–1580. doi: 10.1109/ICUS58632.2023.10318340.

[6] X. Feng et al., “Research on Small Target Detection Method for Poppy Plants in UAV Aerial Photography Based on Improved YOLOv8,” Agronomy, vol. 15, no. 12, p. 2868, Dec. 2025, doi: 10.3390/agronomy15122868.

[7] Z. Liu, G. He, and Y. Hu, “Multi-Level Contextual and Semantic Information Aggregation Network for Small Object Detection in UAV Aerial Images,” Drones, vol. 9, no. 9, p. 610, Sep. 2025, doi: 10.3390/drones9090610.

[8] L. Lu, D. He, C. Liu, and Z. Deng, "MASF-YOLO: An Improved YOLOv11 Network for Small Object Detection on Drone View," Apr. 25, 2025, arXiv: 2504.18136. doi: 10.48550/arXiv.2504.18136.

[9] Y. Chen, W. Zheng, Y. Zhao, T. H. Song, and H. Shin, "DW-YOLO: An Efficient Object Detector for Drones and Self-driving Vehicles," Arab J Sci Eng, vol. 48, no. 2, pp. 1427–1436, Feb. 2023, doi: 10.1007/s13369-022-06874-7.

[10] S. Lu, Y. Guo, J. Long, Z. Liu, Z. Wang, and Y. Li, "Dense small target detection algorithm for UAV aerial imagery," Image and Vision Computing, vol. 156, p. 105485, Apr. 2025, doi: 10.1016/j.imavis.2025.105485.

[11] X. Wang, N. He, C. Hong, Q. Wang, and M. Chen, "Improved YOLOX-X based UAV aerial photography object detection algorithm," Image and Vision Computing, vol. 135, p. 104697, Jul. 2023, doi: 10.1016/j.imavis.2023.104697.

[12] H. Zhang and J. Tang, "SFFN-YOLO for small object detection in aerial images," Multimedia Systems, vol. 31, no. 5, p. 346, Aug. 2025, doi: 10.1007/s00530-025-01922-2.

[13] Y. Cui, W. Ren, and A. Knoll, "Omni-Kernel Network for Image Restoration," in Proceedings of the AAAI Conference on Artificial Intelligence, Mar. 2024, pp. 1426–1434. doi: 10.1609/aaai. v38i2.27907.