

An Infrared Small Target Detection Method for Unmanned Aerial Vehicles Integrating Adaptive Feature Focusing Diffusion and Edge Enhancement

Jiale Wang*

School of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, Shanxi, China

**Author to whom correspondence should be addressed.*

Copyright: © 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: In the context of target detection under infrared conditions for drones, the common issues of high missed detection rates, low signal-to-noise ratio, and blurred edge features for small targets are prevalent. To address these challenges, this paper proposes an improved detection algorithm based on YOLOv11n. First, a Dynamic Multi-Scale Feature Fusion and Adaptive Weighting approach is employed to design an Adaptive Focused Diffusion Pyramid Network (AFDPN), which enhances the feature expression and transmission capability of shallow small targets, thereby reducing the loss of detailed information. Then, combined with an Edge Enhancement (EE) module, the model improves the extraction of infrared small target edge features through low-frequency suppression and high-frequency enhancement strategies. Experimental results on the publicly available HIT-UAV dataset show that the improved model achieves a 3.8% increase in average detection accuracy and a 3.0% improvement in recall rate compared to YOLOv11n, with a computational cost of only 9.1 GFLOPS. In comparison experiments, the detection accuracy and model size balance achieved the optimal solution, meeting the lightweight deployment requirements for drone-based systems. This method provides a high-precision, lightweight solution for small target detection in drone-based infrared imagery.

Keywords: Infrared detection of unmanned aerial vehicles; YOLOv11; Adaptive feature fusion; Edge enhancement; Small target detection

Online publication: October 31, 2025

1. Introduction

With the increasing application of UAV (Unmanned Aerial Vehicle) technology in military reconnaissance, urban inspection, and other fields^[1,2], the infrared thermal sensors equipped on drones have gained prominence due to their advantages, such as all-weather capabilities and penetration visibility. These sensors have become one of the primary methods for target detection. However, during small target detection in infrared imaging under UAV conditions, challenges arise due to the tiny size of the targets (ranging from a few pixels to tens of

pixels), limited feature information such as target color, shape, and texture, as well as the inherent characteristics of infrared imaging, which result in a low signal-to-noise ratio (SNR), blurred target edges, and susceptibility to background interference. Additionally, the features of targets in infrared imaging are affected by material properties, environmental temperature, lighting, and other factors, leading to low detection accuracy for small targets using infrared thermal sensors. In the field of deep learning, the emergence of YOLO series algorithms has significantly improved target detection efficiency^[3,4]. Therefore, this paper selects YOLOv11 as the base model for research. This algorithm, while retaining the first-stage high speed of the YOLO series, incorporates C3k2 and C2PSA spatial attention learning models for feature representation, FPN+PAN feature fusion network structures, and decoupled detection head designs to enhance target detection efficiency. However, challenges such as limited feature extraction capability, high missed detection rates for small targets in complex background environments, and difficulty in balancing detection accuracy with lightweight performance still persist.

2. Proposed method

In response to the challenges of small target detection in drone-based infrared images, such as small target size, low signal-to-noise ratio, and loss of edge details, an improved algorithm based on YOLOv11n is proposed (Figure 1).

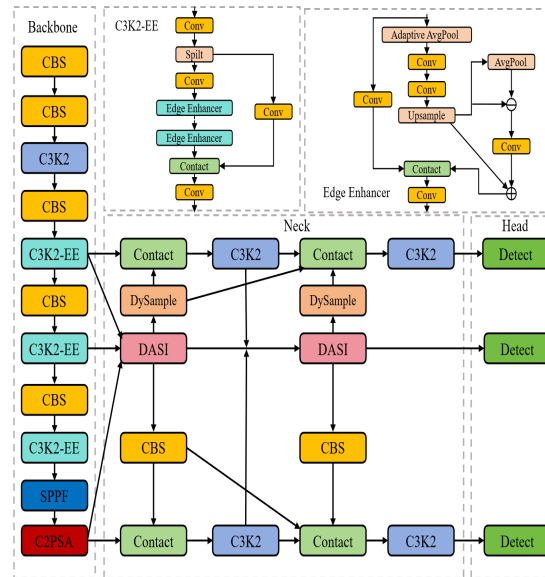


Figure 1. Improved YOLOv11 network architecture.

The improvements are made in both the feature extraction and feature fusion stages. First, an Adaptive Focused Diffusion Pyramid Network (AFDPN) is designed, based on the concept of feature-focused diffusion^[5], to dynamically integrate multi-scale features using a Dimension-Aware Selective Integration Module (DASI)^[6]. The feature fusion process employs a four-step mechanism of feature alignment, channel division, adaptive weight calculation, and fusion. Additionally, the fusion weights for high and low-dimensional features are

adaptively adjusted, with dynamic upsampling (DySample) ^[7] applied to improve sampling efficiency via point sampling. Compared to the FPN-PAN feature fusion network structure in YOLOv11, AFDPN enhances the contextual information expression of shallow small targets, mitigating the issue of context information loss caused by deep network feature sampling. It also reduces feature redundancy. In the P3-P5 layers of the feature extraction network, the concept of high-frequency edge enhancement is introduced ^[8], and an Edge Enhancer (EE) module is added to C3k2 to form C3k2-EE. Through low-frequency suppression, high-frequency extraction, edge enhancement, and fusion, the module strengthens the edge information of infrared small targets, addressing the issue of edge information loss caused by upsampling. By increasing shallow feature expression, cross-scale contextual information fusion, and high-frequency edge feature extraction, the algorithm compensates for the semantic deficiencies between deep and shallow networks. This enhances the ability to capture edge information and transfer semantic information in infrared imaging, improving detection accuracy and robustness.

3. Experiment and result analysis

3.1. Experimental setup

The experiment was conducted using an NVIDIA GeForce RTX 4090 24GB GPU, with the operating system Ubuntu 22.04, based on Python 3.10.14. The system has 60 GB of RAM, and the development framework used is PyTorch 2.2.0-Ubuntu 22.04 (Table 1).

Table 1. Experimental hyperparameter settings

Method	Configuration
Learning rate	0.01
Momentum	0.937
Batch size	32
Optimizer	SGD
Image size	640×640
Epochs	200

3.2. Experimental dataset and evaluation metrics

The dataset used in this experiment is the HIT-UAV ^[9] infrared thermal image dataset, which is specifically designed for high-altitude UAVs. The dataset contains 2,898 infrared images with a resolution of 640×512 pixels, and a total of 24,898 annotated target classes. The dataset includes scenes such as schools, parking lots, roads, and playgrounds, with the main targets being Person, Car, Bike, and OtherVehicle. A “DontCare” label is used to annotate ambiguous targets. The dataset is split into 2,029 training images, 290 validation images, and 579 test images.

The experimental evaluation metrics include the number of parameters (Para) and the number of floating-point operations per second (GFLOPS) to assess the model size and computational complexity. Precision (P), recall (R), and mean average precision (mAP) are used to evaluate the model’s detection accuracy.

3.3. Ablation experiments

To validate the effectiveness of the proposed modules, a progressive ablation experiment was designed (Table 2).

Table 2. Ablation experiments

Methods	P(%)	R(%)	mAP50(%)	mAP50-95(%)	Params(M)	GFLOPS
YOLOv11n	83.5	72.7	78.5	50.8	2.58	6.3
YOLOv11n +AFDPN	82.2	76.7	81.0	53.5	2.86	9.2
YOLOv11n+EE	86.9	71.7	79.3	51.7	2.53	6.3
YOLOv11n+AFDPN+EE	88.2	75.7	82.3	55.5	2.81	9.1

The combination of AFDPN with YOLOv11 resulted in a 2.7% increase in average detection accuracy (mAP50) and a 4% improvement in recall rate. This demonstrates that adaptive bidirectional diffusion can effectively improve small target detection accuracy in infrared scenarios, reduce shallow feature loss, and only adds 0.28M parameters. The C3k2-EE module, which enhances edge features, improves precision by 3.4%, while recall rate decreased by 1%. The parameter count and computational complexity are both lower than the baseline. When combining both modules to optimize the baseline model, the average detection accuracy (mAP50) increased by 3.8%, reaching 82.3%, and mAP50-95 improved by 4.7%. The parameter count increased by 8.9%. The adaptive focused diffusion pyramid network (AFDPN) compensates for the recall rate decrease introduced by the edge enhancement module, achieving a 3.0% improvement over the baseline model, indicating the complementary effect between adaptive multi-scale feature diffusion and edge enhancement.

3.4. Comparison experiments of different detection models

To further validate the performance of the improved model, a comparison with other mainstream YOLO series models was conducted on the HIT-UAV test set (Table 3).

Table 3. Comparison of different detection models

Methods	P(%)	R(%)	mAP50(%)	mAP50-95(%)	Params(M)	GFLOPS
Hyper-YOLOs	84.7	77.9	82.1	55.3	13.51	33.8
YOLOv5s	85.5	75.9	80.9	52.5	9.11	23.8
YOLOv6s	87.1	75.9	81.4	53	16.29	44
YOLOv8s	84.4	77.7	83.7	54.4	11.13	28.4
YOLOv9s	85.1	77.1	81.2	55.0	7.17	26.7
YOLOv10n	79.5	76.8	80.6	51.9	2.26	6.5
YOLOv11s	88.6	74.2	81.5	54.5	9.41	21.3
Ours	88.2	75.7	82.3	55.5	2.81	9.1

It can be observed that the proposed improved model achieves the lowest parameter count and computational complexity compared to most mainstream models, with a parameter count of only 2.81M. Additionally, the average detection accuracy has reached the optimal level. Although the parameter count and computational complexity are higher than YOLOv10n, the average detection accuracy (mAP50) exceeds it by 1.7%. Considering both detection accuracy and model size, the deployment difficulty of the model meets the

Disclosure statement

The author declares no conflict of interest.

References

- [1] Wu X, Li W, Hong D, et al., 2021, Deep Learning for Unmanned Aerial Vehicle-Based Object Detection and Tracking: A Survey. *IEEE Geoscience and Remote Sensing Magazine*, 10(1): 91–124.
- [2] Qiu Z, Bai H, Chen T, 2023, Special Vehicle Detection from UAV Perspective via YOLO-GNS Based Deep Learning Network. *Drones*, 7(2) 117.
- [3] Terven J, Córdova-Esparza DM, Romero-González JA, 2023, A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*, 5(4): 1680–1716.
- [4] Khanam R, Hussain M, 2024, YOLOv11: An Overview of the Key Architectural Enhancements. *arXiv preprint arXiv:2410.17725*.
- [5] Xiangming Q, Yiran Z, 2024, Fire Detection Algorithm with Multi-Scale Feature Focus and Diffusion. 2024 International Conference on Electronics and Devices, Computational Science (ICEDCS), IEEE: 71–76.
- [6] Xu S, Zheng S, Xu W, et al., 2024, HCF-Net: Hierarchical Context Fusion Network for Infrared Small Object Detection. 2024 IEEE International Conference on Multimedia and Expo (ICME), IEEE: 1–6.
- [7] Liu W, Lu H, Fu H, et al., 2023, Learning to Upsample by Learning to Sample. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6027–6037.
- [8] Li S, 2023, Salient Object Detection via High-Frequency Edge Detail Enhancement. *International Conference on Electronic Information Engineering and Computer Science (EIECS 2022)*, SPIE, 12602: 293–300.
- [9] Suo J, Wang T, Zhang X, et al., 2023, HIT-UAV: A High-Altitude Infrared Thermal Dataset for Unmanned Aerial Vehicle-Based Object Detection. *Scientific Data*, 10(1): 227.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.