Journal of Computer Science and Digital Technology, 2026, 2(1), 37-50
https://doi.org/10.71204/ew979106

Night UAV Vehicle Detection Based on Infrared-Visible Fusion and Improved YOLO11

Hongxia Su [1, *]

[1] School of Computer Science and Technology, Taiyuan Normal University, Shanxi 030619, China

*** Correspondence:**

Hongxia Su

2862201740@qq.com

Received: 25 February 2026 / Accepted: 15 March 2026 / Published online: 16 March 2026

Abstract

To address the challenges of vehicle detection accuracy in nighttime UAV aerial photography scenarios caused by low light conditions, strong noise, and dense small object distribution, this paper proposes an improved YOLO11 vehicle detection method based on multi-modal fusion. First, a collaborative enhancement strategy combining CLAHE and Gamma correction is applied to preprocessing nighttime visible light images, effectively restoring vehicle texture details in dark areas. Second, the enhanced visible light images are fused with infrared images through fixed-weight weighted fusion (infrared weight α=0.7), fully leveraging the complementary advantages of both modalities. Finally, the MANet module is introduced on top of YOLO11n to enhance backbone multi-scale feature extraction capabilities, while the ADFPN-DASI module is designed to collaboratively optimize neck multi-scale feature representation. Experiments conducted on the DroneVehicle dataset demonstrate that the proposed method achieves an mAP50 of 80.53%, representing an improvement of 11.45 percentage points over the YOLO11n baseline. The method maintains lightweight advantages in parameter and computational resources, outperforming mainstream comparison methods such as YOLOv5n.

Keywords: Nighttime Vehicle Detection; Drone Aerial Photography; Multimodal Fusion; YOLO11; Feature Pyramid Network

1. Introduction

In recent years, the widespread adoption of drones (UAVs) and high-resolution aerial imaging technology has demonstrated significant value in target detection based on aerial imagery for urban traffic monitoring, intelligent traffic management, and disaster response (Mittal et al., 2020). However, nighttime drone aerial imaging faces severe challenges such as extreme low-light conditions, strong noise, and loss of target details, making it a major research challenge. Li et al. (2024) also highlighted the prominent issue of accuracy degradation in nighttime scenarios. Sun

et al. (2022) constructed the DroneVehicle dataset, which includes 28,439 pairs of RGB-infrared images covering various urban scenes during both day and night, providing a crucial multimodal benchmark for related research.

In image enhancement, CLAHE effectively improves the visibility of details in visible light images under low-light conditions through local adaptive contrast equalization (Persiya et al., 2025). When used in conjunction with Gamma correction, it further restores the recognizability of textures in dark areas (Ma et al., 2023). Yuan et al. (2023) also validated the practical value of CLAHE in preprocessing nighttime low-light scenarios for drones. To compensate for the limitations of visible light modality at night, infrared-visible fusion technology has become a key breakthrough (Ma et al., 2019). ESM-YOLO (Zhang et al., 2024) achieves high accuracy in small target detection for aerial remote sensing through its bilateral excitation fusion module, while IV-YOLO (Tian et al., 2024) reaches a 74.6% mean absolute precision (mAP) on UAV remote sensing datasets.

In the field of detection networks, first-stage algorithms like the YOLO series (Redmon and Farhadi, 2018)are better suited for nighttime drone scenarios due to their real-time performance advantages. YOLO11 (Khanam and Hussain, 2024)introduced the C3k2 module and C2PSA attention mechanism, with its nano version containing only about 2.6 million parameters, making it suitable for edge deployment. YOLOv12 (Tian et al., 2025)further incorporated area attention mechanism and R-ELAN, surpassing previous CNN-dominated frameworks in balancing speed and accuracy. Wang et al.achieved significant improvements in small object detection accuracy on the VisDrone dataset through enhancements to YOLOv8 (Wang et al., 2023). Luo et al. (2025) proposed NOC-YOLO, which reached a mAP50 of 79.5% on the DroneVehicle dataset. Feng et al. (2024) introduced Hyper-YOLO, integrating supergraph computation and MANet backbone into the detection framework, achieving a 12% mAP improvement over YOLOv8-N on the COCO dataset. However, existing methods still exhibit notable limitations in dense small object detection at night and cross-modal fusion.

To address this, this paper proposes an improved YOLO11 vehicle detection method for nighttime scenes in the DroneVehicle dataset, with the following key contributions:

(1) The CLAHE+Gamma collaborative enhancement strategy is applied to pre-process nighttime visible light images, effectively restoring vehicle texture details in dark areas.

(2) The enhanced visible light image and infrared image are fused by fixed weight weighted fusion (infrared weight 0.7) to give full play to the complementary advantages of the two modes;

(3) The self-developed ADFPN (Aggregated Diffusion Feature Pyramid Network) integrates the HCFNet-based DASI module and MANet backbone, achieving significant improvements in nighttime dense small object detection accuracy while maintaining its lightweight advantages.

2. Methodology Design

The proposed method is structured in three phases as shown in Figure 1: visible light image preprocessing, infrared-visible fusion, and an enhanced YOLO11 detection network.

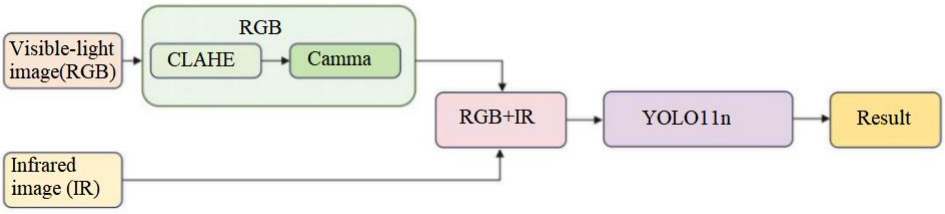

Figure 1. Overall framework of this paper

2.1. CLAHE+Gamma Visible Light Image Enhancement

To address the low contrast and loss of details in dark areas of nighttime visible light images, this paper adopts a collaborative enhancement strategy combining CLAHE and Gamma correction.

The strategy comprises two key processing steps: First, the CLAHE module converts the image to the LAB color space and applies adaptive histogram equalization (LUT) solely to the L channel (8×8 sub-block size, clipLimit=2.0) to suppress noise while enhancing local contrast and restoring dark area texture details. Second, the Gamma correction module (γ=0.7) performs efficient nonlinear brightness mapping through a lookup table (LUT) to specifically brighten dark areas, further improving vehicle contour recognition. Through this two-step collaborative processing, the contours and texture details of vehicles in nighttime dark areas are effectively restored, providing high-quality input for subsequent infrared-visible fusion.

2.2. Infrared-Visible Weighted Image Fusion

To address the complementary limitations of visible light modality in nighttime imaging and infrared modality's insufficient texture representation, this study proposes a fixed-weight fusion strategy. The approach leverages the inherent complementarity between the two modalities: infrared images remain stable in low-light conditions while reliably capturing vehicle thermal radiation characteristics, whereas CLAHE+Gamma-enhanced visible light images preserve rich texture and structural details. The fusion formula is as follows:

$$I_{fused}=\alpha\cdot\ I_{IR}+(1-\alpha)\cdot I_{RGB\_enhanced}$$

Here, IIR denotes infrared images, IRGB_enhanced represents enhanced visible light images, and α is the infrared weight. Ablation experiments on the DroneVehicle validation set demonstrated that the model achieves optimal performance (79.14% mAP50) with α=0.7, indicating that infrared modality should be given higher fusion weight in nighttime scenes to fully leverage its light-insensitivity advantage while retaining 30% of visible light texture information for further precision improvement. The final fused images are input into the subsequent detection network as three-channel inputs.

2.3. The Improved YOLO11 Algorithm

To address challenges in nighttime UAV aerial photography—such as target detail loss due to low light, severe degradation of visible light single-modality performance, and high false-negative rates caused by dense small object distribution—this study proposes an improved algorithm for

nighttime dense small object detection by enhancing YOLO11n through multi-modal fusion. The enhanced network architecture is illustrated in Figure 2.

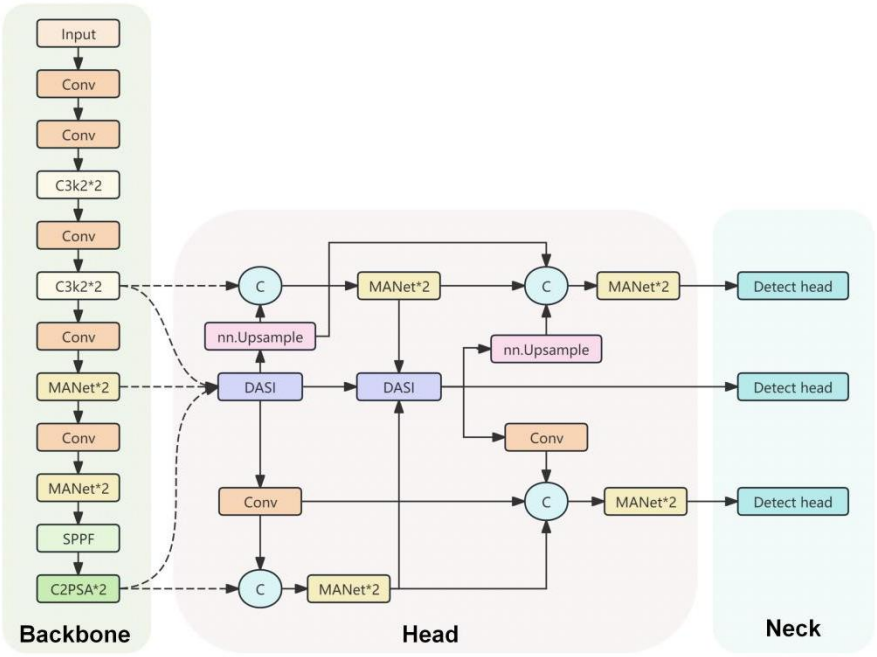

Figure 2. Network Architecture Diagram

The algorithm primarily consists of two enhancement modules:

(1) MANet Module: To address the backbone network's limited capability in extracting multi-scale features of small targets at night, this paper introduces the MANet (Mixed Aggregation Network) module, replacing some C3k2 feature extraction units in the YOLO11n backbone. MANet simultaneously receives multi-scale feature maps as inputs and employs a hybrid aggregation mechanism to achieve cross-scale adaptive fusion of multi-sensory field features. Features at each scale are first segmented through channel splitting and then fed into the C2f Block for local feature extraction. Subsequently, the HyperC2 Block utilizes the Hyperedge mechanism to enable cross-scale semantic transfer, allowing each detection scale to perceive contextual information from other scales. The final output combines features from all scales with the original input through channel dimension concatenation. This cross-scale hybrid aggregation mechanism significantly enhances the network's multi-scale feature representation capability for densely distributed small targets in complex nighttime backgrounds, effectively mitigating target detail loss caused by low illumination.

(2) ADFPN-DASI Module: To address the issues of insufficient cross-scale information transfer and strong background noise interference in traditional feature pyramid networks, this paper designs the ADFPN-DASI module. Based on a bidirectional feature pyramid architecture, this module replaces the fixed fusion structure in traditional FPN with an improved DASI (Dimension-aware Selective Integration) node derived from HCFNet. In the top-down path, the DASI node simultaneously receives three-scale inputs (P3, P4, P5), performing adaptive weighting of multi-scale features in both channel and spatial dimensions to selectively enhance

effective feature responses while suppressing background noise. In the bottom-up path, the second DASI node performs cross-scale aggregation on enhanced features from each scale, completing secondary feature diffusion through symmetric sampling and splicing operations to ensure semantic-rich features are fully distributed across all detection scales. The coordinated operation of the two DASI nodes enables each detection scale to obtain context-rich feature representations through dimension-aware adaptive fusion, maintaining lightweight advantages while collaboratively improving multi-scale detection capabilities for dense small targets in nighttime scenarios.

2.3.1. MANet

To address the limitations of backbone networks in nighttime aerial photography for multi-scale feature extraction of small targets, this study introduces the MANet (Mixed Aggregation Network) module to replace certain C3k2 feature extraction units in the YOLO11n backbone. Unlike the original C3k2 module that extracts features locally at a single scale, MANet employs a cross-scale hybrid aggregation mechanism to simultaneously model semantic relationships across multiple detection scales, making it more suitable for feature enhancement of densely distributed small targets under nighttime low signal-to-noise ratio conditions.

As shown in Figure 3, the MANet's overall processing workflow consists of three distinct phases.

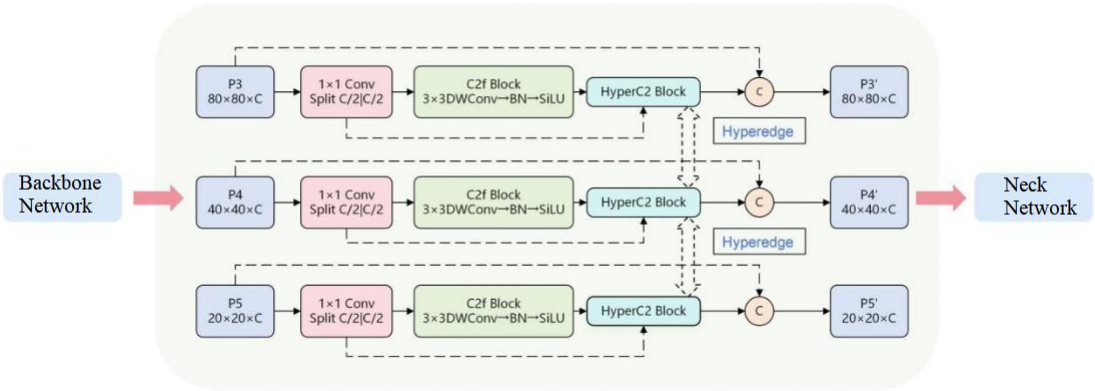

Figure 3. MANet structure diagram

(1) Multi-scale Input and Channel Splitting: MANet processes three-scale feature maps (P3: 80×80×C, P4: 40×40×C, P5: 20×20×C) simultaneously. Each scale's features undergo 1×1 convolution for channel splitting (Split C/2 | C/2), dividing the feature channels into two streams—one for local feature extraction and the other for cross-scale semantic transfer. This approach maintains complete feature information across scales while controlling computational load.

(2) Local Feature Extraction and Cross-Scale Semantic Transfer: The segmented features are fed into the C2f Block for local feature extraction, which employs a 3×3 depthwise separable convolution (3×3DWConv→BN→SiLU) to efficiently capture local details across scales while reducing parameter size and maintaining receptive field coverage. The extracted features are then

processed by the HyperC2 Block, which establishes explicit cross-scale semantic connections between P3, P4, and P5 scales through the Hyperedge mechanism. This enables each scale to perceive and integrate contextual semantic information from other scales. This mechanism is particularly critical in nighttime scenes—when target responses become weak due to low illumination, semantic supplements from other scales effectively compensate for feature incompleteness.

(3) Feature Fusion and Output: After cross-scale aggregation, the enhanced features from each scale are concatenated (concat) with the original input features along the channel dimension, forming an enhanced feature representation that integrates local details and global semantics. The final output consists of three multi-scale feature maps (P3 ', P4', P5') for subsequent use by the neck network and detection head.

Through the cross-scale hybrid aggregation mechanism, MANet enables the model to simultaneously integrate local detail information and global semantic context across three detection scales. This significantly enhances the feature representation capability for densely distributed small targets in complex nighttime backgrounds, thereby reducing false negatives and improving overall detection accuracy.

2.3.2. ADFPN-DASI

The ADFPN-DASI framework employs a bidirectional feature fusion pathway, replacing the fixed fusion structure of traditional FPNs with a DASI (Dimension-Aware Selective Integration) module enhanced by HCFNet (Zhu et al., 2021) as the core aggregation node. Unlike conventional FPNs that rely solely on simple upscaling and concatenation for cross-scale feature fusion, the DASI node performs adaptive weighting of multi-scale features simultaneously across channel and spatial dimensions. It selectively amplifies target-relevant feature responses while actively suppressing noise interference from complex nighttime backgrounds, making it particularly suitable for multi-scale detection of dense small targets in low-light conditions.

The ADFPN-DASI workflow comprises two distinct approaches: a top-down and a bottom-up methodology, as illustrated in Figure 4.

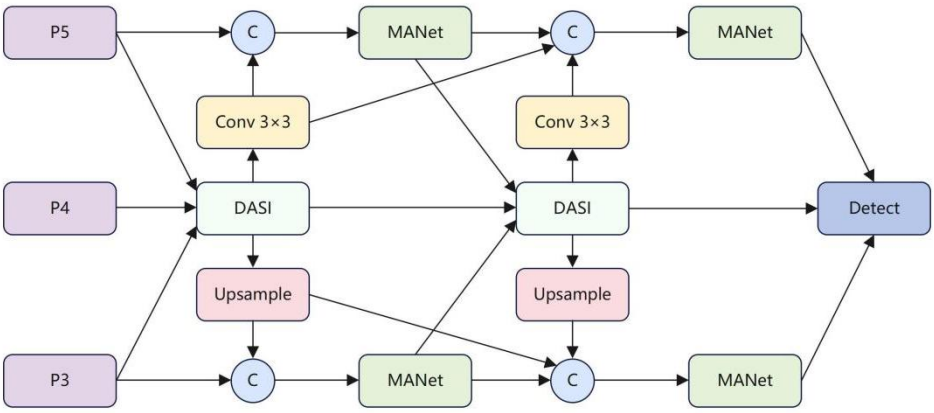

Figure 4. ADFPN-DASI structural diagram

(1) Top-down pathway: The P5 feature map undergoes down-sampling via Convolutional Downsampling to align spatial resolution, while the P3 feature map is upsampled to a uniform scale. The three feature streams (P3, P4, P5) then converge into the first DASI node. The DASI node first evaluates the importance of features across channels and applies adaptive weighting to filter redundant channel responses. It then selectively activates each feature map position in the spatial dimension, highlighting target region features to achieve active background noise suppression during fusion. The aggregated features are fed into the MANet module for further cross-scale local feature extraction, yielding preliminary enhanced feature representations at each scale.

(2) Bottom-up Path: The feature enhancement outputs from the top-down path are reintegrated into the second DASI node for secondary dimension-aware adaptive fusion. Symmetrically aligned with the first DASI node, this path undergoes corresponding sampling and stitching operations to fully integrate low-level detail information with high-level semantic information, completing secondary feature diffusion. The final output consists of three bidirectional path-enhanced feature maps (P3, P4, P5), which are respectively fed into corresponding scale detection heads for target localization and classification.

Two DASI nodes collaboratively handle cross-scale aggregation in bidirectional paths, forming a complete bidirectional enhancement loop that combines "top-down semantic transfer with bottom-up detail supplementation." This ensures each detection scale receives context-rich feature representations through dimension-aware adaptive fusion. The design proves particularly critical in nighttime dense small-object scenarios: the top-down path supplements global semantic information for small-scale targets, while the bottom-up path preserves fine spatial localization details for large-scale features. Their synergy significantly enhances multi-scale detection capabilities for nighttime dense small objects.

3. Experiments and Result Analysis

3.1. Introduction to the Dataset

The experiment was conducted using the DroneVehicle dataset, which contains 28,439 registered RGB-infrared image pairs. These images cover diverse urban scenarios including roads, parking lots, and residential areas, with multiple lighting conditions (daytime and nighttime). The dataset is labeled with five categories: car, truck, bus, van, and freight-car. In this study, we follow the official classification to separate the training and validation sets.

3.2. Experimental Environment and Evaluation Indicators

The hardware and software configurations used in the experiment are shown in Table 1. To comprehensively evaluate model performance, this study employs five evaluation metrics: Precision, Recall, mAP, Parameters (M), and GFLOPs. Precision measures the proportion of predicted positive samples that are actually positive, while Recall indicates the proportion of true positive samples correctly detected. mAP represents the average AP value across all categories. Specifically, we use mAP@0.5 and mAP@0.5:0.95 as metrics, corresponding to average

precision at IoU thresholds of 0.5 and 0.5–0.95 (with a step size of 0.05). Parameters (in megabytes) and GFLOPs (in gigaflops) reflect the model's storage overhead and computational inference load, respectively, serving as key indicators for measuring model lightweighting.

The model was trained for 250 epochs with a batch size of 64 and an input resolution of 640×640 pixels. The SGD optimizer was adopted with an initial learning rate of 0.01, a momentum of 0.937, and a weight decay of 0.0005. Early stopping with a patience of 100 epochs was applied to prevent overfitting.

Table 1. Experimental Environment Configuration

environment	Parameter configuration
operating system	Windows11
CPU	Intel® Xeon® Gold 5418Y processor, 10 cores
GPU	NVIDIA GeForce RTX4090
exploitation environment	PyCharm
compiling environment	Python 3.10
Deep learning framework	Pytorch 2.2.1

3.3. Ablation Experiment

Table 2 presents the detection performance of YOLO11n under various input modalities and fusion weight configurations, demonstrating the effectiveness of preprocessing enhancement and fusion strategies.

Table 2. Experimental results of different modalities and fusion weight ablation

Model configuration	P	R	mAP50	mAP50-95	GFLOPs	Parameters
YOLO11n+RGB	71.51	65.12	69.08	43.70	6.3	2.58 million
YOLO11n+IR	77.56	74.56	78.79	59.05	6.3	2.58 million
YOLO11n+ fusion (α=0.5)	76.73	75.17	78.49	58.25	6.3	2.58 million
YOLO11n+ fusion (α=0.6)	76.96	74.83	78.83	58.73	6.3	2.58 million
YOLO11n+ fusion (α=0.7)	77.41	75.34	79.14	59.15	6.3	2.58 million
YOLO11n+ fusion (α=0.8)	76.99	75.55	78.78	59.17	6.3	2.58 million

Table 2 reveals the following conclusions: The single infrared modality (78.79%) significantly outperforms the single visible RGB modality (69.08%), with a 9.71 percentage point improvement, demonstrating the core advantage of infrared modality in nighttime scenes. Weighted fusion outperforms single RGB modality across all weight configurations, achieving optimal mAP50 (79.14%) when the infrared weight α=0.7. This indicates that infrared information should dominate in nighttime scenes while retaining some visible texture information to further enhance accuracy. The fusion scheme (79.14%) surpasses the single infrared modality (78.79%), proving that the visible images enhanced by CLAHE+Gamma indeed provide incremental information for fusion.

Table 3 presents the ablation results of three enhancements—MANet, FDPN, and DASI—introduced progressively on the optimal fusion input (α=0.7).

Table 3. Results of Ablation Experiment for Network Modules

Model configuration	P	R	mAP50	mAP50-95	GFLOPs	Parameters
YOLO11n+ fusion (α=0.7)	77.41	75.34	79.14	59.15	6.3	2.58 million
+MANet	77.61	77.19	80.33	60.41	8.4	3.78 million
+ADFPN-DASI	78.46	75.39	79.51	59.37	7.4	2.68 million
+MANet+ADFPN-DASI	79.12	76.01	80.53	60.28	8.6	3.28 million

As shown in Table 3, the introduction of MANet increased mAP50 by 1.19 percentage points (79.14%→80.33%) and recall rate by 1.85 percentage points, demonstrating that the hybrid aggregation mechanism effectively enhances the backbone's feature extraction capability for small nighttime targets. The addition of ADFPN-DASI improved mAP50 by 0.37 percentage points (79.14%→79.51%), achieving effective multi-scale feature optimization with only an increase of approximately 100,000 parameters, highlighting its lightweight characteristics. When both modules are used simultaneously, the optimal mAP50 (80.53%) is achieved, representing a 1.39 percentage point improvement over the fusion baseline and an 11.45 percentage point enhancement compared to the YOLO11n+RGB baseline, validating the synergistic effects of each module.

3.4. Comparison of Mainstream Algorithms

Table 4 compares the final method proposed in this paper with mainstream detection models of comparable scale on the DroneVehicle dataset.

Table 4. Comparison of Experimental Results with Mainstream Methods

model	P	R	mAP50	mAP50-95	GFLOPs	Parameters
YOLOv5n	76.71	75.13	78.49	58.39	5.8	2.18 million
YOLOv8n	77.03	74.55	78.21	58.38	6.8	2.69 million
YOLOv12n	76.63	76.25	79.32	59.28	5.8	2.51 million
YOLOv13n	76.79	74.32	78.09	58.72	6.1	2.45 million
YOLO11n (baseline)	71.51	65.12	69.08	43.70	6.3	2.58 million
Methodology of this paper	79.12	76.01	80.53	60.28	8.6	3.28 million

As shown in Table 4, our proposed method outperforms mainstream baseline models (e.g., YOLOv5n, YOLOv8n, YOLOv12n, and YOLOv13n) in both mAP50 and mAP50-95 metrics, achieving a 1.21 percentage point improvement over the closest competitor YOLOv12n. While the proposed method requires more parameters and GFLOPs than MANet before its integration, these metrics remain within a lightweight range, ensuring deployability on edge devices in drones. These results conclusively validate the effectiveness of our multimodal fusion and network optimization strategy.

3.5. Visualization Results

To further validate the effectiveness of our proposed method, Figures 5-8 systematically compare YOLO11n, YOLOv12n, and our method through visual detection in multiple typical application scenarios, clearly demonstrating their performance differences under various challenging conditions.

Figure 5 demonstrates a relatively uniform distribution of scene targets under moderate lighting conditions, which facilitates focused evaluation of the modeling capabilities of various methods in target confidence assessment. Comparative results reveal that our proposed method achieves significantly higher confidence scores across all detection targets compared to YOLO11n and YOLOv12n. This phenomenon indicates that the CLAHE+Gamma collaborative enhancement preprocessing and the integration of the MANet attention mechanism effectively strengthen the model's feature representation capability in target regions, enabling the network to complete target localization and classification with higher confidence.

Figure 6 depicts a scene with numerous targets exhibiting varying degrees of occlusion and scale variations. Experimental results demonstrate that YOLOv12n exhibits significant missed detection in this scenario, failing to effectively retrieve some small and edge targets. While YOLO11n shows some improvement, its detection completeness remains inadequate. The proposed method achieves target detection quantities comparable to or even surpassing YOLO11n, with a markedly reduced missed detection rate. These results conclusively validate the positive

impact of the ADFPN-DASI multi-scale feature fusion module in enhancing small target perception capabilities and suppressing missed detection.

Figure 7 depicts a complex background with coexisting multi-category objects, demanding high classification discrimination capability from the model. Comparative analysis reveals that both YOLO11n and YOLOv12n exhibit certain category misclassifications, misclassifying background interference or adjacent objects. In contrast, our method demonstrates the lowest false detection rate and superior category prediction accuracy. This indicates that through multimodal feature fusion and attention enhancement, the model achieves enhanced inter-category discrimination, enabling more precise differentiation of objects across diverse categories in complex scenarios.

Figure 8 presents a typical nighttime drone aerial photography scenario with high object density and significant overlap, representing one of the most challenging conditions. Experimental results demonstrate that both YOLO11n and YOLOv12n exhibit varying degrees of confidence degradation and missed detection, indicating overall unstable performance. In contrast, our proposed method achieves more accurate object localization and consistently reliable detection results. This demonstrates that the infrared-visible weighted fusion strategy effectively compensates for information loss in low-light nighttime visibility. By leveraging the infrared modality's high sensitivity to thermal radiation targets, the model significantly improves detection performance in dense and complex nighttime environments.

The comprehensive visual comparison results demonstrate that the proposed method outperforms the baseline in multiple metrics including confidence level, recall rate, classification accuracy, and nighttime robustness, further validating the effectiveness of the synergistic effects among the improved modules.

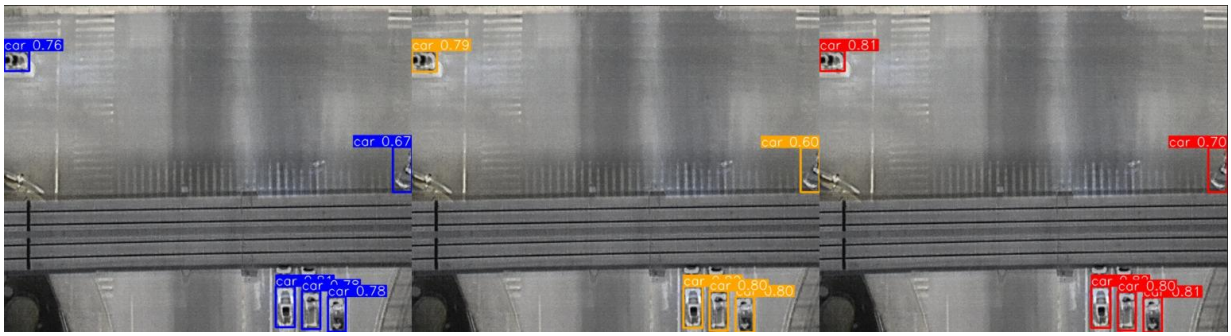

Figure 5. Visualizes Figure

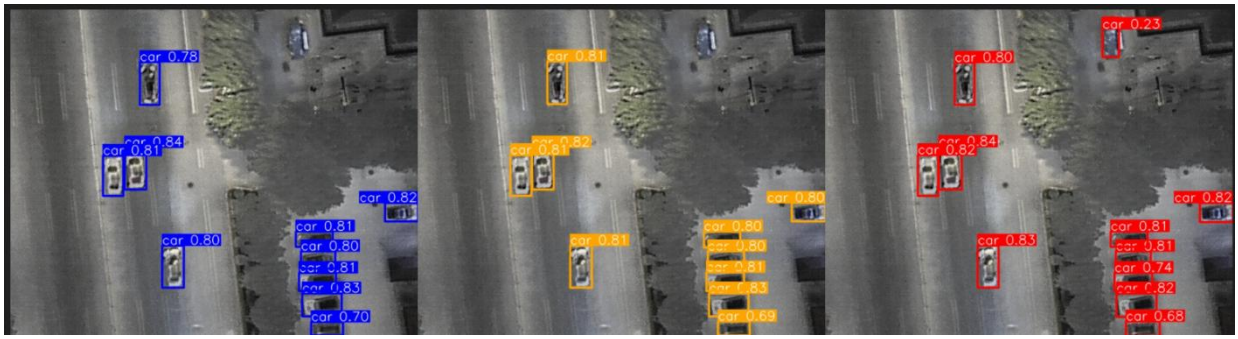

Figure 6. Visualizes Figure

Figure 7. Visualizes Figure

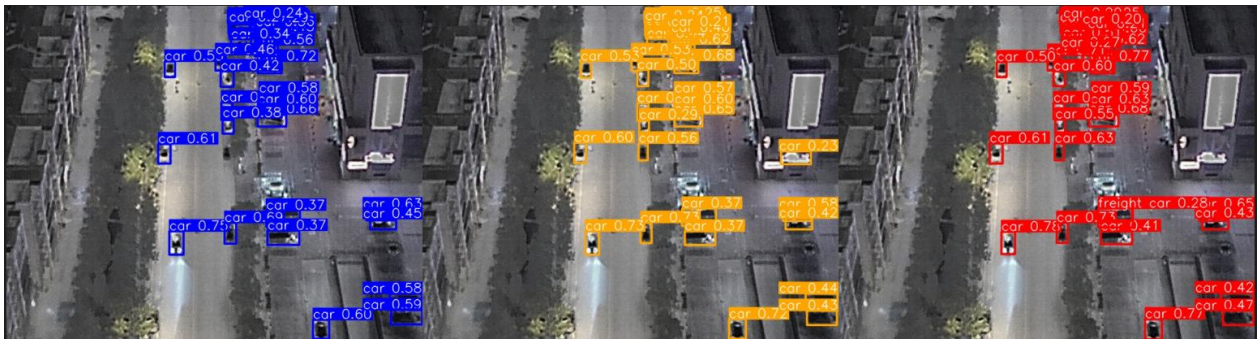

Figure 8. Visualizes Figure

4. Conclusion

This study addresses key challenges in nighttime UAV vehicle detection, including low illumination, severe noise, and multi-modal data fusion difficulties. We propose a multi-modal fusion-enhanced detection method: image preprocessing adopts a CLAHE-Gamma collaborative enhancement strategy to improve low-light contrast and detail visibility; a weighted fusion mechanism fuses infrared (thermal radiation sensitivity) and visible (rich texture-semantic information) images for modal complementarity; the network integrates the MANet attention mechanism and ADFPN-DASI feature pyramid module to enhance multi-scale feature aggregation and cross-layer interaction, boosting small/dense target detection.

Experiments on the DroneVehicle dataset show the method achieves an mAP50 of 80.53%, 11.45 percentage points higher than the baseline, validating its effectiveness. Ablation experiments confirm the independent and synergistic contributions of the enhancement module, fusion strategy, and network improvements.

Limitations include: fixed-weight fusion fails to adapt to dynamic lighting in real aerial scenarios, leading to performance bottlenecks; unoptimized computational complexity restricts deployment on resource-constrained edge platforms. Future work will focus on lightweight deployment and diverse dataset construction to enhance robustness and practicality.

Author Contributions:

All authors have read and agreed to the published version of the manuscript.

Funding:

This research received no external funding.

Institutional Review Board Statement:

Not applicable.

Informed Consent Statement:

Not applicable.

Data Availability Statement:

Not applicable.

Conflict of Interest:

The authors declare no conflict of interest.

References

Feng, Y., Huang, J., Du, S., et al. (2024). Hyper-YOLO: When visual object detection meets hypergraph computation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 47(4), 2388–2401.

Khanam, R., & Hussain, M. (2024). YOLO11: An overview of the key architectural enhancements. arXiv preprint arXiv:2410.17725.

Li, X., Li, X., et al. (2024). A survey of object detection for UAVs based on deep learning. Remote Sensing, 16(1), 149.

Ma, J., Ma, Y., & Li, C. (2019). Infrared and visible image fusion methods and applications: A survey. Information Fusion, 45, 153–178.

Ma, M., Wang, H., & Wang, J. (2023). An underwater image enhancement algorithm based on improved MSRCR–CLAHE fusion. Infrared Technology, 45(1), 23–32.

Mittal, P., Singh, R., & Sharma, A. (2020). Deep learning-based object detection in low-altitude UAV datasets: A survey. Image and Vision Computing, 104, 104046.

Persiya, J., & Sasithradevi, A. (2025). Synergistic fusion: An integrated pipeline of CLAHE, YOLO models, and advanced super-resolution for enhanced thermal eye detection. PLOS ONE, 20(7), e0328227.

Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Sun, Y., Cao, B., Zhu, P., et al. (2022). Drone-based RGB-infrared cross-modality vehicle detection via uncertainty-aware learning. IEEE Transactions on Circuits and Systems for Video Technology, 32(10), 6700–6713.

Tian, D., Yan, X., Zhou, D., et al. (2024). IV-YOLO: A lightweight dual-branch object detection network. Sensors, 24(19), 6181.

Tian, Y., Ye, Q., & Doermann, D. (2025). YOLOv12: Attention-centric real-time object detectors. arXiv preprint arXiv:2502.12524.

Wang, G., Wang, G., et al. (2023). UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. Sensors, 23(16), 7190.

Yuan, Z., Zeng, J., Wei, Z., et al. (2023). CLAHE-based low-light image enhancement for robust object detection in overhead power transmission system. IEEE Transactions on Power Delivery, 38(3), 2240 – 2243.

Zhang, Q., Qiu, L., Zhou, L., et al. (2024). ESM-YOLO: Enhanced small target detection based on visible and infrared multi-modal fusion. In Proceedings of the Asian Conference on Computer Vision (pp. 1454 – 1469).

Zhang, Y., Dai, Z., Pan, C., et al. (2025). NOC-YOLO: An exploration to enhance small-target vehicle detection accuracy in aerial infrared images. Infrared Physics & Technology, 149, 105905.

Zhu, P., Wen, L., Du, D., et al. (2021). Detection and tracking meet drones challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(11), 7380 – 7399.