

Article

DBM-YOLO: A Dual-Branch Model with Feature Sharing for UAV Object Detection in Low-Illumination Environments

Liwen Liu, Huilin Li, Gui Fu , Bo Zhou, You Wang and Rong Fan

College of Aviation Electronic and Electrical Engineering, Civil Aviation Flight University of China, Chengdu 641419, China; llw@cafuc.edu.cn (L.L.); m15823613256@163.com (H.L.); 18881036982@163.com (B.Z.); wangyou@cafuc.edu.cn (Y.W.); fanrong@cafuc.edu.cn (R.F.)

* Correspondence: abyfugui@cafuc.edu.cn

Highlights

What are the main findings?

- We propose a parallel network framework that consists of a Zero-DCE illumination enhancement network and the backbone of a YOLOv11n-based object detection network.
- The DPSA module is introduced to enhance feature representation and multi-scale adaptability through dynamic channel and spatial attention, while the HLSAFM module refines high- and low-frequency features to enable richer feature extraction and improved discriminative capability.

What are the implications of the main findings?

- This parallel architecture establishes a dual-branch framework that enables collaborative feature training and real-time updates, effectively enhancing the feature adaptability and accuracy of object.
- Ablation studies confirm that the synergistic effect of the proposed modules yields more robust and discriminative feature representations, significantly enhancing detection performance in UAV scenarios in low-illumination environments.

Abstract

To resolve the issue of degraded detection accuracy for unmanned aerial vehicle object detection under low-illumination environments, this paper introduces a parallel object detection model. First, a dual-branch architecture is established by parallelly integrating a Zero-Reference Deep Curve Estimation (Zero-DCE) illumination enhancement network with a You Only Look Once (YOLOv11n)-based object detection network, enabling collaborative feature training and real-time updates. Through a feature-sharing mechanism, the two branches are jointly optimized during training, thus enhancing the model's generalization capability in low-illumination environments. Furthermore, to further improve detection accuracy, a Dynamic Pooling Synergy Attention (DPSA) module is introduced into the backbone of YOLOv11n. By integrating dynamic pooling-based channel attention with spatial attention, this module improves feature representation, improves performance under complex environments, and increases adaptability to multi-scale targets. In addition, a High and Low Frequency Spatially-adaptive Feature Modulation (HLSAFM) module is added to the detection network's Neck. Through high- and low-frequency feature refinement, segmented feature processing, and dynamic modulation, the network is able to capture richer feature information, thereby strengthening feature representation and discriminative capability. Extensive experiments on the VisDrone (Night) and DroneVehicle (Night) datasets demonstrate superior performance over multiple existing methods under low-illumination object detection tasks. Compared with the original YOLOv11n model, the



Academic Editor: Anastasios Dimou

Received: 6 January 2026

Revised: 21 February 2026

Accepted: 25 February 2026

Published: 28 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

proposed model $mAP50$ increases by 6.0% and 1.0% and $mAP50 : 95$ increases by 3.1% and 0.8%, respectively. These results confirm the enhanced detection capability achieved by our method in challenging low-illumination unmanned aerial vehicle (UAV) scenarios.

Keywords: unmanned aerial vehicle; parallel network; low-illumination object detection; YOLOv11; image enhancement

1. Introduction

Over recent years, advances in unmanned aerial vehicle (UAV) platforms have significantly expanded their applicability in a wide range of vision-based tasks, including intelligent transportation, nighttime monitoring, emergency response, and public security. Compared with conventional ground-based vision systems, UAV platforms provide a broader field of view and flexible deployment capabilities. However, object detection under low-illumination conditions remains a particularly challenging scenario. In nighttime or dim-light environments, insufficient illumination causes the captured images to have low contrast, severe noise interference, and significant degradation of visual information quality. These factors substantially weaken the discriminative capability of deep neural networks, leading to frequent missed detections and false alarms, especially for small-scale and long-distance targets, thereby limiting the stability and reliability of detection systems [1–3].

In practical UAV missions, such detection failures are not merely numerical degradations but may directly influence operational effectiveness. For example, in nighttime surveillance and urban traffic monitoring, missed detections of pedestrians or vehicles may delay the response or compromise situational awareness. In emergency search-and-rescue scenarios, the inability to reliably detect small or distant targets under poor illumination may significantly reduce the rescue efficiency and increase operational risk. Recent surveys on UAV-based human detection for search-and-rescue missions emphasize that missed detections in low-light, cluttered, or high-altitude conditions can directly delay victim localization and compromise mission success, highlighting the critical role of robust detection performance in real deployments [4]. Therefore, improving quantitative detection metrics such as mAP and recall in low-light conditions is closely related to enhancing the reliability, safety, and responsiveness of UAV-based perception systems in real-world deployments.

Although infrared (IR) imagery has been explored for low-illumination UAV object detection, this work focuses on visible-light imagery for three main reasons. First, visible-light cameras provide rich color, texture, and shape information, which are crucial for fine-grained object recognition and localization in UAV scenarios. Second, visible-light sensors are significantly more affordable and widely available, making them more suitable for practical and large-scale deployment. In contrast, high-resolution infrared cameras are considerably more expensive. Third, visible–infrared fusion involves heterogeneous feature alignment and additional model complexity, which constitutes a separate research direction beyond the scope of this study. Therefore, our work aims to enhance detection robustness under low illumination using visible-light imagery alone.

To address low-illumination UAV object detection, traditional methods typically adopt a two-stage processing pipeline. In this paradigm, illumination enhancement is first applied to the input images to improve the visual quality, followed by object detection using conventional detection networks. Representative image enhancement techniques include Retinex-based models, illumination correction networks, and generative adversarial networks (GANs). Typical works include detection frameworks based on EnlightenGAN [1],

Zero-DCE [5], URetinex [5], SCI [5], and RUAS [5]. These methods can enhance the image brightness and contrast to a certain extent, thereby improving the detection performance.

Nevertheless, two-stage methods still suffer from inherent limitations. First, image enhancement and object detection are usually optimized independently, neglecting the intrinsic coupling between illumination restoration and task-oriented feature learning. Second, improved visual quality does not necessarily guarantee better detection performance, as some enhancement methods may amplify the background noise or suppress critical target features while increasing the brightness [6,7]. Moreover, the serial processing structure introduces additional computational overhead and inference latency, which is unfavorable for real-time application on UAV platforms with limited computational resources [8].

Nowadays, numerous studies have explored low-illumination object detection for UAV applications, with particular emphasis on addressing the challenges posed by severe illumination degradation. Different from the traditional two-stage paradigm that sequentially performs illumination enhancement and then object detection, most recent methods tend to adopt a one-stage (end-to-end) detection framework, where illumination-robust representation learning is integrated into the detector. By embedding illumination-aware feature enhancement, attention mechanisms, or multi-scale feature fusion modules into YOLO-style networks, these methods jointly optimize low-light feature enhancement and detection objectives, thereby alleviating the mismatch between perceptual enhancement and task performance, while also reducing additional computational overhead and inference latency compared with serial enhancement–detection pipelines. Notable methods include the integration of illumination-aware feature enhancement and attention mechanisms into YOLO variants. For example, Reference [2] introduced illumination-aware feature enhancement and attention mechanisms into YOLOv7, achieving improved detection accuracy in dim environments. Reference [7] proposed a fusion-based feature enhancement strategy built on YOLOv5, effectively improving the detection performance in dim environments. Reference [9] developed 3L-YOLO with a lightweight design tailored for low-illumination environments, while Reference [10] proposed LLE-YOLO by modifying the YOLOv5 architecture to improve low-illumination detection and reduce noise interference.

Meanwhile, attention mechanisms and multi-scale feature modeling have been widely adopted in low-illumination object detection. Reference [11] introduced GAME-YOLO, which bolsters UAV detection by incorporating global attention and multi-scale feature enhancement. Reference [6] introduced En-YOLO, utilizing dual attention mechanisms in conjunction with implicit feature learning to strengthen low-illumination feature representations. Edge information at multiple scales along with illumination-aware features were introduced into the YOLOv8 framework in [12], while Reference [13] proposed LLD-YOLO, which employs dynamic weighting of shallow and deep features. Additionally, methods combining Transformer architectures with dynamic feature fusion have demonstrated promising results in low-illumination detection tasks [14,15].

Beyond YOLO-based frameworks, methods such as A-RetinaNet [16], DBLDNet [17], and FCMA-Det [18] have further validated the effectiveness of structured feature enhancement through dual-branch architectures, feature complementarity, and multi-content aggregation strategies for object detection under low-illumination conditions.

Despite these advances, several obstacles continue to be unaddressed. Most existing methods focus primarily on feature enhancement or attention modeling, while lacking a systematic mechanism for collaborative learning between illumination enhancement and object detection. Moreover, UAV-captured targets often exhibit significant scale variations, complex backgrounds, and frequent occlusions, which are further exacerbated under low-

illumination conditions [19–22]. In addition, UAV platforms impose strict constraints on model complexity and inference efficiency, whereas some existing methods remain computationally expensive and less suitable for real-time deployment [23–28].

Our breakthroughs in this paper are outlined as follows:

- (1) This work presents a parallel detection architecture, where illumination enhancement network and object detection network are jointly incorporated to achieve synchronized feature learning. This parallel architecture allows the two networks to adapt to each other during training while sharing feature representations, thereby improving feature adaptability and detection accuracy under low-illumination environments.
- (2) A High and Low Frequency Spatial-Adaptive Feature Modulation (HLSAFM) module is designed to enhance model robustness in extremely low-illumination scenarios by decomposing features into high and low frequency and modulating spatially adaptive modulation.
- (3) A Dynamic Pooling Synergistic Attention (DPSA) module is introduced, which integrates dynamic pooling, channel attention, spatial attention, and multi-scale convolution to improve discriminative capability for objects at different scales.

2. Materials and Methods

2.1. Dataset

The VisDrone2019 dataset [29] is a large-scale publicly available dataset captured by UAV released by Tianjin University and other research teams. The dataset is made up of 6471 training images, 548 validation images, and 1610 test images, with varying spatial resolutions including 1360×765 , 1400×1050 , and 2000×1500 pixels. Inspired by Reference [15], a comprehensive analysis of the VisDrone dataset was conducted, from which nighttime images along with their corresponding annotation files were carefully selected to construct the VisDrone (Night) subset used in this study. The constructed VisDrone (Night) dataset contains 2023 images for training and 56 images for testing. All samples undergo consistent preprocessing and are scaled to a fixed size of 640×640 pixels.

The DroneVehicle dataset [30] includes 56,878 images captured by UAVs, consisting of paired RGB and infrared images covering five vehicle categories: bus, SUV, truck, car, and freight car. During dataset construction, nighttime image pairs were first selected, followed by further filtering to retain only those RGB images with ground-truth annotations. This process yielded the DroneVehicle (Night) subset, which includes 11,406 training images and 880 test images. All images in this subset have a fixed spatial resolution of 640×640 pixels.

2.2. DBM-YOLO

2.2.1. Architecture

To address the performance degradation caused by low illumination, small object scale, dense distribution, and background interference in UAV night scenes, we propose a parallel low-illumination UAV detection framework built upon YOLOv11n. The proposed architecture integrates an illumination enhancement network and an object detection network in a parallel manner, enabling synchronized feature updates and mutual optimization during training. In addition, two plug-and-play modules, namely the Dynamic Pooling Synergistic Attention (DPSA) module and the High and Low Frequency Spatial-Adaptive Feature Modulation (HLSAFM) module, are incorporated to strengthen multi-scale representation and improve robustness to small and densely distributed targets. The overall architecture of the proposed framework is illustrated in Figure 1.

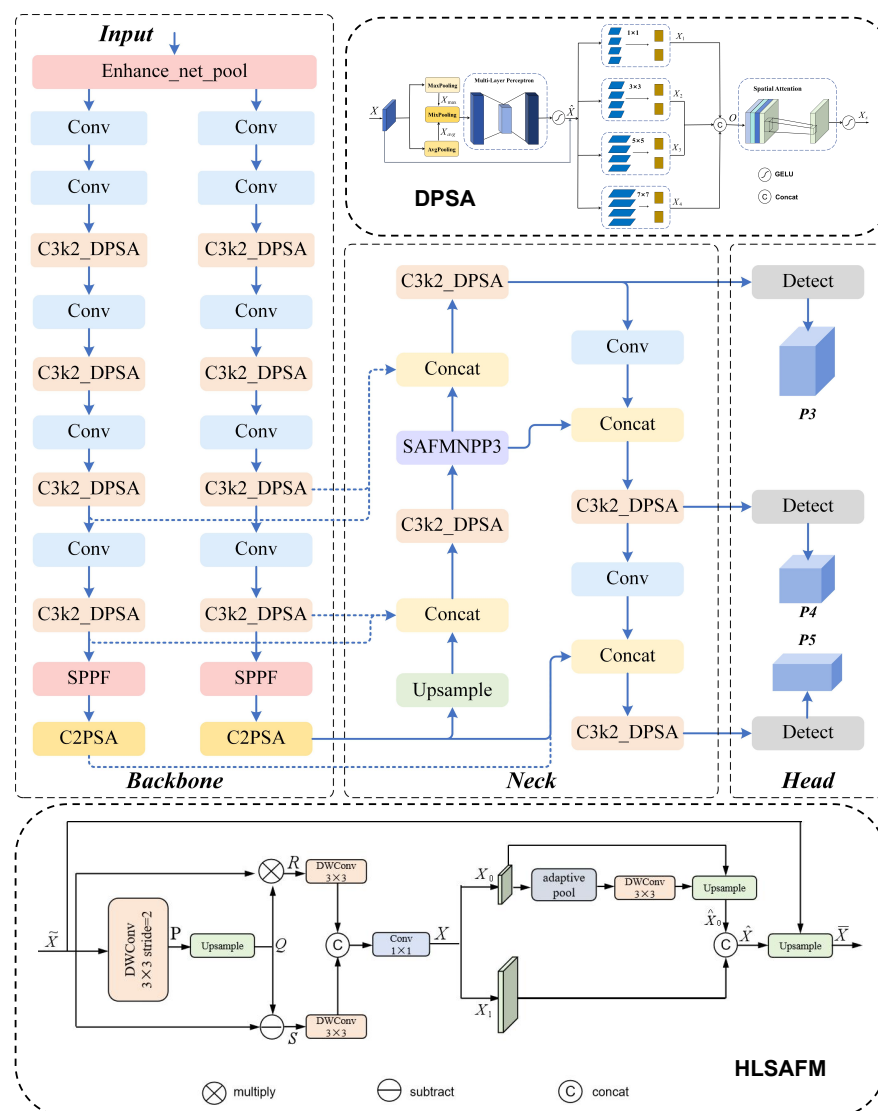


Figure 1. Architecture of the proposed DBM-YOLO.

2.2.2. Parallel Network Framework

The overarching architecture of the introduced parallel network is shown in Figure 1. This framework has been meticulously engineered to tackle the distinctive challenges associated with enhancing illumination and detecting targets in dimly illuminated scenarios, employing a concurrent processing paradigm to substantially elevate the overall efficacy. Images are simultaneously fed into the Zero-DCE illumination enhancement network and the backbone of the YOLOv11n detector, leveraging parallel processing for autonomous illumination adaptation.

The Zero-DCE network employs a zero-reference deep curve estimation approach, overcoming limitations of traditional methods reliant on paired data or external references. It learns directly from image-specific illumination via a non-reference loss, enhancing generalization without paired training data.

In this study, YOLOv11n is deliberately selected as the baseline detection network. Compared with heavier detection models, YOLOv11n provides a lightweight architecture with reduced parameter size and computational complexity while maintaining competitive detection performance. Its compact structure makes it more suitable for edge deployment on embedded UAV systems, where real-time processing (e.g., >30 FPS) is often required to ensure stable flight control and timely decision-making. By adopting YOLOv11n as the

backbone detector, the proposed parallel framework maintains computational feasibility, while allowing additional enhancement modules (DPSA and HLSAFM) to be integrated without causing prohibitive increases in GFLOPs or latency.

More generally, the architectural design of the proposed method follows a lightweight and modular principle. All additional modules are designed as plug-and-play components with controlled computational overhead, ensuring that performance gains do not come at the expense of deployability. This design philosophy enhances the practical applicability of the proposed framework for real-time low-light UAV detection scenarios.

In the detection branch, features from the enhancement and backbone paths are fused using Feature Pyramid Network (FPN) and additional fusion techniques across multiple scales. The proposed DPSA and HLSAFM modules are embedded into the backbone and neck, respectively, to boost accuracy.

A pretrained YOLOv11n network provides auxiliary fixed-parameter features, fused to augment the final detection performance.

2.2.3. HLSAFM

Low-illumination conditions introduce noise and brightness deficiencies, complicating target feature extraction. The HLSAFM module addresses this by refining high- and low-frequency features, separating processing paths, and applying dynamic modulation for improved discriminability and robustness.

High-frequency components capture edges and details critical for small targets, while low-frequency components provide global structure and illumination invariance. Inspired by image sharpening and contrast enhancement, HLSAFM extracts low-frequency information via downsampling and smoothing, with high-frequency details obtained as residuals.

Input features $\tilde{X}$ undergo depthwise convolution to acquire a downsampled feature map P . In the field of image processing, high-frequency details are commonly extracted by computing the difference between an image and its blurred counterpart. Inspired by this principle, the present work adopts a similar approach. By upsampling feature P to capture the spatial resolution of $\tilde{X}$, the resulting feature is denoted as Q . Subsequently, element-wise multiplication between $\tilde{X}$ and Q is performed to effectively emphasize fine details, resulting in the high-frequency component S . In the parallel branch, element-wise subtraction is applied between $\tilde{X}$ and Q to capture the low-frequency component R . The low-frequency feature R and high-frequency feature S are then processed through depthwise convolution, aggregated, and further refined via a 1×1 convolution to obtain an enhanced fused feature X . The fused feature map X , which has undergone high-low frequency processing, is then split channel-wise in a 1:3 ratio into two parts, X_0 and X_1 , as follows:

$$[X_0, X_1] = \text{Split}(X), \quad (1)$$

where $\text{Split}()$ denotes the channel-wise splitting function.

Single-scale feature modulation is performed on X_0 through adaptive pooling, depthwise convolution, and upsampling operations to generate a richer feature representation that promotes non-local feature interactions. The process is formulated as follows:

$$\hat{X} = \uparrow_p ((DW - \text{Conv}_{3 \times 3}(\downarrow_{\frac{p}{8}}(X_0))), X_0) \quad (2)$$

where DW denotes depthwise convolution, $\uparrow$ represents upsampling, and $\downarrow_{\frac{p}{8}}$ indicates adaptive pooling.

The feature $\hat{X}_0$ is element-wise multiplied with X_0 to perform spatial-adaptive feature modulation. The resulting modulated feature $\hat{X}$ is then integrated with the initial input feature $\tilde{X}$, further enhancing the representation of salient features:

$$\bar{X} = \text{Concat}(\hat{X}, \tilde{X}), \quad (3)$$

where $\text{Concat}()$ denotes the concatenation function.

The feature fusion is formulated as follows:

$$X_{Fus} = \text{Concat}(X_{ori}, X_{Low}), \quad (4)$$

where X_{ori} denotes an original feature, and X_{Low} denotes a low-level feature.

In summary, the HLSAFM module dynamically modulates the representation of input features while maintaining computational efficiency, thus enhancing the performance of object recognition in low-illumination settings. The integration of high and low frequency components, allows the model to fully exploit both local and global contextual information, which demonstrates notable benefits for precise object identification.

2.2.4. DPSA

Deep neural networks typically face difficulties in effectively identifying useful information across diverse channels or spatial locations. Furthermore, conventional convolutional operations lack the ability to directly capture feature representations from different image regions or at multiple scales. To address these problems, we propose the Dynamic Pooling Synergistic Attention (DPSA) module, which integrates dynamic pooling-based channel attention with spatial attention mechanisms. The primary goal of DPSA is to strengthen the representation capability of input features, improve the performance of the network in complex scenarios, and increase its adaptability to features of varying scales.

The DPSA dynamically selects salient feature channels and spatial regions through adaptive weighting in both the channel and spatial dimensions. Channel-wise adaptivity assigns differentiated weights to individual channels, thereby reducing the feature redundancy and semantic interference, emphasizing edge-related information, and suppressing less informative features such as background and texture noise. Spatial adaptivity addresses multi-scale object localization conflicts and background interference by generating a spatial weight map through dynamic fusion. This enhances the response within object-related areas while simultaneously reducing the influence of background noise. The synergistic integration of channel and spatial attention enables DPSA to substantially enhance the model's generalization capability, resulting in higher stability and reliability across diverse real-world scenarios.

The DPSA module begins by performing global average pooling and global maximum pooling to the input feature map X to extract global information for each channel:

$$X = \text{Conv}_{k=1}(\text{RELU}(\text{Conv}_{k=1}(\text{avg_pool}(X)))), \quad (5)$$

where $\text{avg_pool}()$ denotes global average pooling, $\text{RELU}()$ denotes RELU activation functions, $\text{Conv}_{k=1}$ denotes 1×1 convolution and X_{avg} is the resulting feature descriptor obtained from average pooling.

$$X_{max} = \text{Conv}_{k=1}(\text{RELU}(\text{Conv}_{k=1}(\text{max_pool}(X)))) \quad (6)$$

In the above equations, $\text{max_pool}()$ denotes maximum pooling, and X_{max} represents the resulting feature descriptor obtained from maximum pooling.

To further enhance the adaptivity of the pooling operation, a balanced pooling module is introduced. By controlling a learnable balancing parameter α , the module adaptively combines global average and max pooling during the pooling process, thereby making the pooling operation more flexible.

$$X_{\alpha} = \alpha * \text{avg_pool}(X) + (1 - \alpha) * \text{max_pool}(X) \quad (7)$$

In the above equation, α represents the balancing parameter, and X_{α} denotes the result of balanced pooling.

$$X_{min} = Conv_{k=1}(RELU(Conv_{k=1}(X_{\alpha}))) \quad (8)$$

In the above equation, X_{min} represents the result of the balanced pooling operation.

The pooled features are aggregated and subsequently normalized using a Sigmoid function to produce the channel attention output feature:

$$X_c = \sigma(X_{avg} + X_{max} + X_{min}) \quad (9)$$

Here, the Sigmoid activation is denoted by σ , and X_c corresponds to the feature output of the channel attention network.

$$\hat{X} = X * X_c \quad (10)$$

In the above equation, X_c represents the feature attained by combining the channel attention map with the initial feature map.

To handle the scale variation of objection in UAV imagery, the DPSA module adopts a multi-scale feature fusion strategy. This is achieved by deploying parallel convolutional kernels with receptive fields of 1×1 , 3×3 , 5×5 , and 7×7 , whose outputs are subsequently aggregated, thereby enhancing the model's capability to handle objects of varying sizes.

$$X_i = Conv_{k=\beta}(\hat{X}_i), i \in (1, 2, 3, 4), \beta \in (1, 3, 5, 7) \quad (11)$$

In the above equation, $Conv_{k=\beta}$ signifies the kernel size of the convolution of $\beta \times \beta$, while X_i denotes the output feature produced by the corresponding convolution branch.

$$O = Concat(X_i), i \in (1, 2, 3, 4) \quad (12)$$

In the above equation, O represents the fused feature map obtained by concatenating the outputs from convolutions of different kernel sizes.

In the spatial attention branch, the feature map is initially processed using average and max pooling across channels to obtain global information at each spatial location. Subsequently, these pooled representations are combined, followed by the application of a convolutional layer to generate the spatial attention map. Spatial attention weights each pixel location adaptively, enabling the model to emphasize important spatial regions while suppressing irrelevant areas, thereby enhancing the representational capability of the spatial features.

$$X_s = \sigma(Conv_{k=7}(avg_pool(O); max_pool(O))), \quad (13)$$

where X_s represents the spatial attention output feature.

The DPSA module simultaneously refines channel-wise and spatial representations of the feature map by allocating adaptive weights according to the relative significance of each channel and spatial position. Additionally, by employing convolutions with multiple kernel sizes, it captures features at different scales, thus strengthening the network's ability to extract both local and global information.

2.3. Experimental Settings

2.3.1. Experimental Environment

All experiments were conducted on a workstation running the Windows operating system (Microsoft Corporation, Redmond, WA, USA). The hardware configuration included an Intel Core i5-10400 processor (Intel Corporation, Santa Clara, CA, USA) and an NVIDIA GeForce RTX 4080 GPU with 24 GB memory (NVIDIA Corporation, Santa Clara, CA, USA). The software environment consisted of Python (v3.10.19, Python Software Foundation, Wilmington, DE, USA), PyTorch (v2.9.1, Meta AI, Menlo Park, CA, USA), and CUDA (v13.0, NVIDIA Corporation, Santa Clara, CA, USA). A batch size of 8 was adopted, and the model was trained for 300 epochs. The learning rate was set to 0.01, and stochastic

gradient descent (SGD) was used as the optimizer due to its stable convergence behavior and strong generalization performance in large-scale object detection tasks, particularly for YOLO-based architectures. The input image size was set to 640×640 pixels.

2.3.2. Evaluation Metrics

To objectively evaluate the detection performance of the proposed model on the VisDrone and DroneVehicle datasets, several widely used evaluation metrics are adopted in this study, including Average Precision (AP) for a single class, Mean Average Precision (mAP) across all classes, Recall (R), and Precision (P). The specific formulations are defined as follows:

$$AP = \frac{TP + TN}{TP + TN + FP} \quad (14)$$

$$APR = \frac{TP}{TP + FN} \quad (15)$$

$$P = \frac{TP}{TP + FP} \quad (16)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (17)$$

In the above equations, TP (True Positive) denotes the number of correctly identified positive samples, FP (False Positive) represents the number of negative samples incorrectly classified as positive samples, FN (False Negative) refers to the number of positive samples that fail to be detected, and TN (True Negative) denotes the number of correctly identified negative samples. Let n denote the total number of object categories in the training dataset, and let P_i represent the Average Precision corresponding to the i -th category.

The Mean Average Precision (mAP) serves as a comprehensive evaluation metric for assessing the overall performance of the detection network in terms of both classification accuracy and localization capability across all object categories. Mathematically, the mAP can be interpreted as the mean value of the Average Precision over all categories and is equivalently represented by the area under the Precision–Recall (PR) curve. A higher mAP value indicates superior detection performance and stronger generalization ability of the network. Furthermore, the detection speed, measured in Frames Per Second (FPS), is adopted as an additional critical performance indicator. FPS quantifies the number of image frames processed by the model per second. A higher FPS value implies enhanced real-time inference capability, which is particularly essential for time-sensitive UAV-based object detection applications.

3. Experimental Results and Analysis

3.1. Comparison Experiments

3.1.1. Comparison of Low-Light Detection Methods on the VisDrone (Night) Dataset

In this paper, object detection methods are categorized into three groups: Category A: Original standalone object detection networks without low-illumination adaptation; Category B: Two-stage serial methods that first apply illumination enhancement followed by object detection; Category C: End-to-end parallel methods that jointly optimize enhancement and detection.

As clearly shown in Table 1, the proposed enhanced algorithm demonstrates superior detection performance on the VisDrone (Night) dataset. Faster R-CNN [31], a classical two-stage detection method, indeed delivers solid performance but is outperformed by the method presented in this section.

Relative to the YOLOv11n algorithm, our approach delivers notable performance gains, leading to gains of 6.0% in $mAP@50$ and 3.1% in $mAP@50 : 95$.

Against representative two-stage serial methods, URetinex [32] + YOLOv11n; SCI [33] + YOLOv11n; Zero-DCE [34] + YOLOv11n; ENGAN [35] + YOLOv11n; and RUAS [36] + YOLOv11n, the proposed algorithm yields the following gains: $mAP50 : 95$ increases by 4.25%, 4.06%, 4.29%, 3.64%, and 4.85%, respectively; $mAP50$ improves by 6.0%, 5.9%, 5.7%, 4.3%, and 7.6%, respectively.

Even against stronger baselines that integrate advanced enhancement techniques (RUAS or ENGAN) with more recent detectors (YOLOv8n and YOLOv11n), the proposed method still demonstrates clear superiority, with improvements in the $mAP50 : 95$ of 2.51%, 2.9%, 4.1%, and 2.5%, and in the $mAP50$ of 6.2%, 4.4%, 7.1%, and 4.1%, respectively.

Furthermore, compared with four representative end-to-end low-illumination object detection models—RetinaNet [37], PG-YOLO [38], GBS-YOLOv11n [39], and Drone-YOLO [40]—the proposed algorithm outperforms them by 8.9%, 5.6%, 6.5%, and 3.8% in the $mAP50$, respectively.

These comprehensive quantitative results strongly confirm that the improved algorithm proposed in this paper surpasses existing low-illumination object detection methods in both overall performance and key evaluation metrics.

Table 1. Experimental results of various methods on the DroneVehicle (Night) dataset.

	Model	$mAP50 : 95\%$	$mAP50\%$	Precision	Recall	F1-Score
A	RT-DETR [41]	13.3	23.2	45.2	20.1	27.8
	YOLOv5n	8.37	17.5	40.8	18.3	25.3
	YOLOv8n	10.1	17.9	42.3	18.2	25.4
	YOLOv11n	10.2	18.1	38.8	19.6	26.0
B	URetinex+YOLOv11n	9.05	18.1	41.9	19.8	26.9
	SCI+YOLOv11n	9.24	18.2	42.0	18.6	25.8
	Zero_DCE+YOLOv11n	9.01	18.4	43.2	18.5	25.9
	ENGAN+YOLOv11n	9.66	19.8	48.3	21.1	29.4
	RUAS+YOLOv11n	8.45	16.5	39.7	20.3	26.9
	ENGAN+YOLOv8n	10.4	19.7	45.9	19.3	27.2
	RUAS+YOLOv8n	9.79	17.9	44.1	17.8	25.0
	RUAS+YOLOv11n	9.2	17.0	44.9	18.4	25.7
	ENGAN+YOLOv11n	10.8	20.0	39.5	22.9	29.0
C	RetinaNet	8.1	15.2	42.3	14.5	21.6
	PG-YOLO	9.36	18.5	38.0	19.6	25.9
	GBS-YOLOv11n	10.0	17.6	41.4	17.7	24.8
	Drone-YOLO	10.5	20.3	40.2	21.3	27.9
	Our Method	13.3	24.1	45.0	25.0	32.1

In addition, to subjectively analyze the feature extraction capability of different models in complex nighttime traffic scenarios, the feature response distributions of each model were visualized, as shown in Figure 2. This visualization intuitively illustrates the key regions attended by the networks during the detection process. Different colors indicate varying degrees of contribution to the detection results, where colors closer to red represent stronger feature responses and higher attention levels.

As shown in the first column of Figure 2, in elevated road scenes with significant variations in vehicle scale, YOLOv8 and YOLOv11 exhibit relatively dispersed attention distributions. Some high-response regions extend into background areas, such as road edges and building structures, indicating insufficient focus on distant small-scale vehicles. In contrast, DBM-YOLO presents more concentrated and continuous high-response feature distributions along the main lanes and vehicle-dense regions, effectively suppressing irrelevant background interference and significantly enhancing multi-scale target perception.

In the second column, which represents scenarios with complex illumination and dense traffic flow, background lights and road reflections tend to cause false activations. YOLOv8 suffers from pronounced background over-response, while YOLOv11 shows partial improvement but still exhibits discontinuous responses in distant small-object regions. By comparison, DBM-YOLO focuses its activation more accurately on real vehicle targets, with a more balanced and structurally coherent heatmap distribution. This demonstrates its superior background suppression and feature discrimination capabilities, mainly attributed to the multi-scale feature fusion mechanism of the DPSA module and the adaptive modulation of high- and low-frequency information by the HLSAFM module, enabling stable target responses under complex lighting conditions.

In the third column, which depicts intersection scenes with local occlusion and densely distributed targets, occlusion and scale variation often lead to missed detections or feature response shifts. YOLOv8 and YOLOv11 show weakened responses or attention drift in partially occluded regions, whereas DBM-YOLO maintains strong responses in overlapping vehicle areas, demonstrating more stable and robust feature representations under complex occlusion conditions.

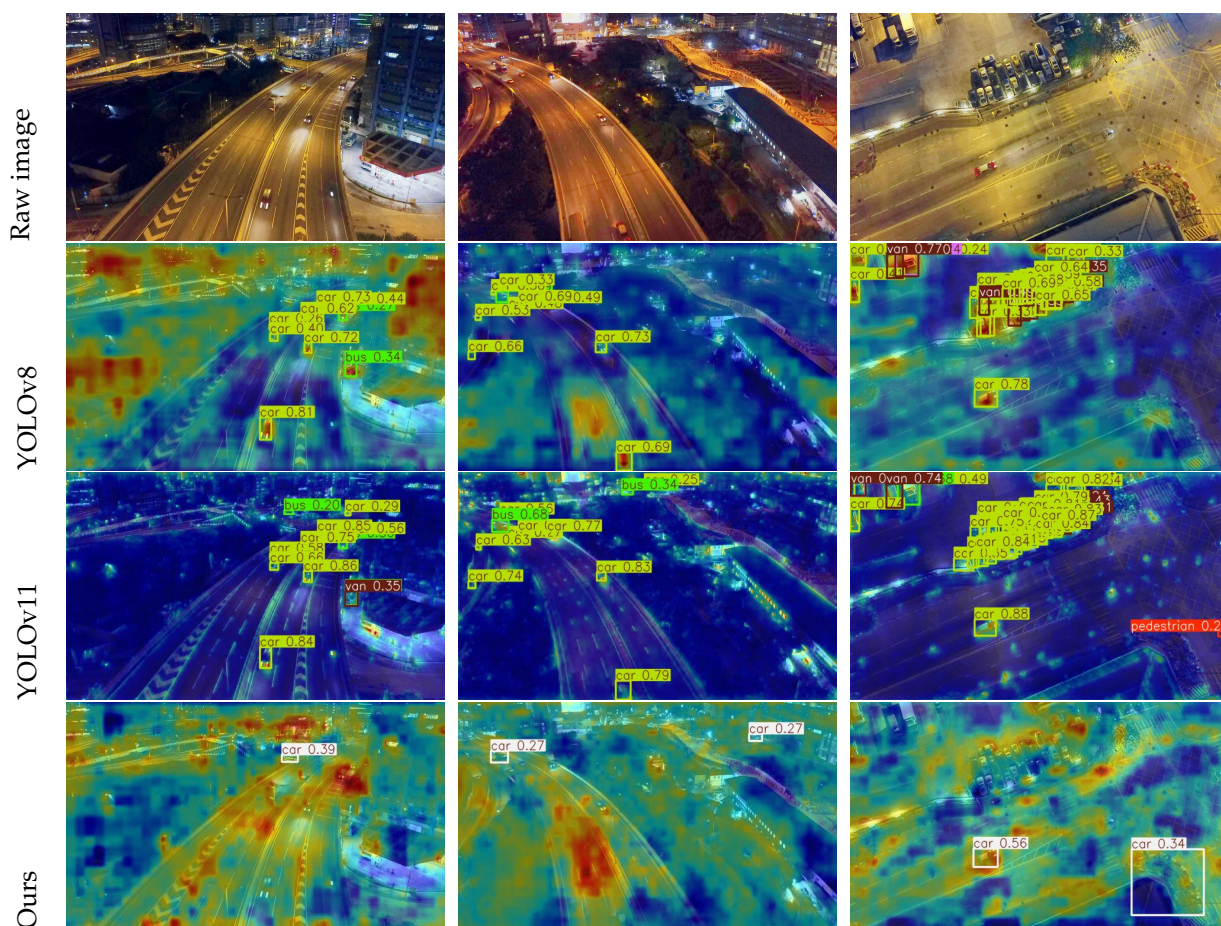


Figure 2. Comparison of heatmaps from different detection methods. Red regions indicate areas with higher attention or stronger responses from the model, whereas blue regions represent lower response intensity.

Overall, benefiting from the parallel enhancement framework and the collaborative optimization of the DPSA and HLSAFM modules, DBM-YOLO exhibits more concentrated, continuous, and discriminative response characteristics under challenging conditions such as low illumination, multi-scale distributions, background interference, and local occlusion. While maintaining high computational efficiency, it further improves the detection accuracy

and localization stability, validating the effectiveness and generalization capability of DBM-YOLO for multi-scale object detection tasks in complex nighttime traffic scenarios.

3.1.2. Comparison of Heatmaps from Different Backbone Networks

As evident from Table 2, the improved algorithm designed in the paper achieves the best detection performance on the DroneVehicle (Night) dataset. When compared with the YOLOv11n baseline, the proposed approach yields improvements of 0.5% in $mAP50$ and 0.8% in $mAP50 : 95$.

When evaluated against representative two-stage serial pipelines, URetinex [32] + YOLOv11n; SCI [33] + YOLOv11n; Zero-DCE [34] + YOLOv11n; ENGAN [35] + YOLOv11n; and RUAS [36] + YOLOv11n, the proposed approach delivers consistent gains: the $mAP50 : 95$ increases by 3.15%, 2.96%, 1.5%, 2.54%, and 3.75%, respectively, and the $mAP50$ improves by 5.7%, 5.6%, 1.8%, 3.0%, and 7.6%, respectively.

Even when compared with stronger baselines that incorporate advanced enhancement techniques (RUAS or ENGAN) into more recent detectors such as YOLOv8n and YOLOv11n, our methodology consistently exhibits pronounced superiority in both the $mAP50 : 95$ and $mAP50$.

Table 2. Experimental results of various methods on the DroneVehicle (Night) dataset.

	Model	$mAP50 : 95\%$	$mAP50\%$	Precision	Recall	F1-Score
A	RT-DETR [41]	41.8	73.1	78.2	64.5	70.7
	YOLOv5n	43.2	72.4	79.0	63.0	70.1
	YOLOv8n	44.0	70.0	75.0	64.0	69.1
	YOLOv11n	46.6	72.4	75.5	68.1	71.6
B	URetinex+YOLOv11n	43.2	71.5	75.4	64.8	69.4
	SCI+YOLOv11n	42.6	71.2	74.3	64.6	69.1
	Zero_DCE+YOLOv11n	43.2	72.5	77.9	65.5	71.2
	ENGAN+YOLOv11n	42.9	71.8	76.5	64.0	69.7
	RUAS+YOLOv11n	41.6	70.2	76.9	62.2	68.8
	ENGAN+YOLOv8n	45.7	72.3	76.4	66.6	71.2
	RUAS+YOLOv8n	41.9	67.5	73.2	61.5	66.9
	RUAS+YOLOv11n	46.3	72.4	73.9	67.5	70.6
	ENGAN+YOLOv11n	45.4	71.8	75.3	65.2	69.9
C	RetinaNet	36.4	61.5	61.5	43.8	51.2
	PG-YOLO	38.2	70.3	75.1	63.8	69.0
	GBS-YOLOv11n	42.3	66.9	74.5	65.3	69.6
	Drone-YOLO	42.5	72.1	72.4	62.6	67.1
	Our Method	47.4	73.4	78.6	67	72.3

3.2. Ablation Study

3.2.1. Ablation Study of the Improved Modules for Object Detection

The previous experiment confirmed the utility of the proposed parallel structure. To further investigate the impact of the two proposed enhancement modules, namely HLSAFM and DPSA, we assess their stepwise integration into configuration (d) in the ablation study.

Ablation studies conducted on the VisDrone (Night) benchmark as presented in Table 3 reveal that integrating the HLSAFM component leads to $mAP50$ and $mAP50 : 95$ improvements of 0.4% and 1.0%, respectively, over the baseline parallel architecture. Independently incorporating the DPSA component results in enhancements of 1.4% in $mAP50$ and 0.7% in $mAP50 : 95$. Notably, the simultaneous deployment of both HLSAFM and DPSA components produces markedly superior outcomes, with substantial relative improvements

of 2.2% in $mAP50$ and 1.1% in $mAP50 : 95$. These empirical findings robustly underscore the individual merits of the introduced components, while highlighting their synergistic complementarity when jointly applied in low-illumination UAV imagery.

Table 3. Ablation study on the VisDrone (Night) dataset.

Model	<i>P</i>	<i>R</i>	<i>mAP50%</i>	<i>mAP50 : 95%</i>	<i>F1-Score</i>	<i>FPS</i>	<i>GFLOPS</i>
d	43.0	23.6	22.1	12.2	30.5	75.2	84.3
d+HLSAFM	43.5	25.6	22.5	13.2	32.2	59.2	102.6
d+DPSA	34.4	24.9	23.5	12.9	28.8	61.6	101.7
d+HLSAFM+DPSA	45.0	25.0	24.1	13.3	32.1	63.3	119.9

Figure 3 presents the normalized confusion matrices of YOLOv11n and DDJK-YOLO11 on the VisDrone dataset. Compared with the baseline model, the improved model exhibits stronger diagonal responses in key categories such as car, bus, motor, and pedestrian, while confusion between foreground targets and background is significantly reduced, leading to lower false detection rates. In scenarios involving small and densely distributed objects, inter-class confusion is further alleviated, indicating that the proposed feature enhancement strategy improves the representation of distant and small-scale targets. Overall, the improved model demonstrates higher detection reliability and lower misclassification risk in complex UAV-view scenarios.

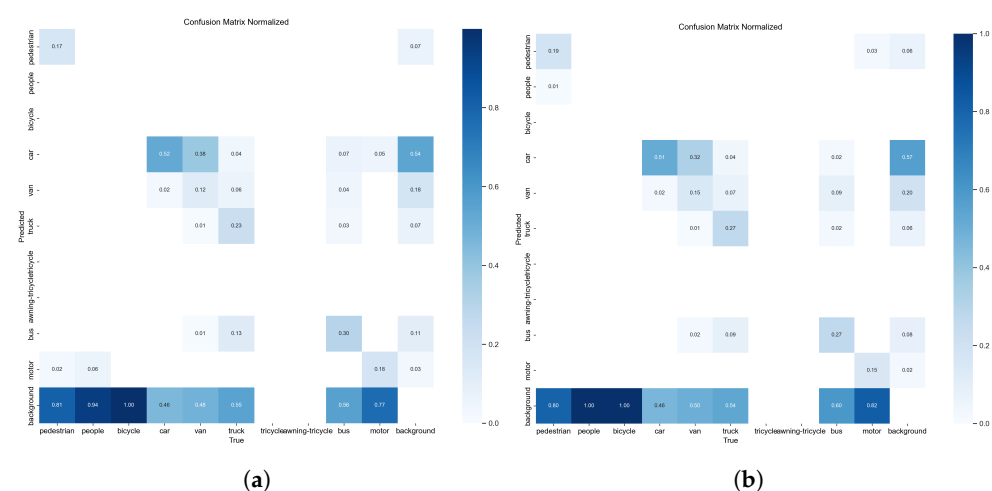


Figure 3. Comparison of confusion matrices for different detection networks on VisDrone dataset. (a) YOLOv11n. (b) Ours.

On the DroneVehicle (Night) dataset (Table 4), the HLSAFM module alone contributes improvements of 1.1% in $mAP50$ and 0.9% in $mAP50 : 95$. When applied alone, the DPSA module alone provides gains of 0.7% in $mAP50$ and provides gains of 0.7% in $mAP50 : 95$. When both modules are used together, the performance further leads to $mAP50$ and $mAP50 : 95$ improvements of 1.0% and 0.6%, respectively.

The normalized confusion matrices of YOLOv11n and DBM-YOLO11 on the DroneVehicle dataset are as shown in Figure 4. Overall, both models exhibit strong diagonal responses, indicating that the dataset has relatively concentrated category distributions and that the baseline already achieves high performance. Compared with YOLOv11n, the improved model shows more concentrated diagonal responses in major categories such as car and van, while confusion between foreground targets and background is further reduced, suggesting enhanced discriminative capability and robustness to interference. However, since the DroneVehicle dataset is relatively vehicle-centric with clearer inter-

class distinctions, the baseline model already performs strongly, leaving limited room for improvement and resulting in comparatively modest overall gains.

Table 4. Ablation study on the DroneVehicle (Night) dataset.

Model	<i>P</i>	<i>R</i>	<i>mAP50%</i>	<i>mAP50 : 95%</i>	<i>F1-Score</i>	<i>FPS</i>	<i>GFLOPS</i>
d	74.8	67.9	72.4	46.8	73.1	118.72	84.3
d+HLSAFM	75.7	68.3	73.5	47.7	71.8	125.49	102.6
d+DPSA	75.9	67.8	73.1	47.5	71.6	148.76	101.7
d+HLSAFM+DPSA	78.6	67	73.4	47.4	72.3	162.31	119.9

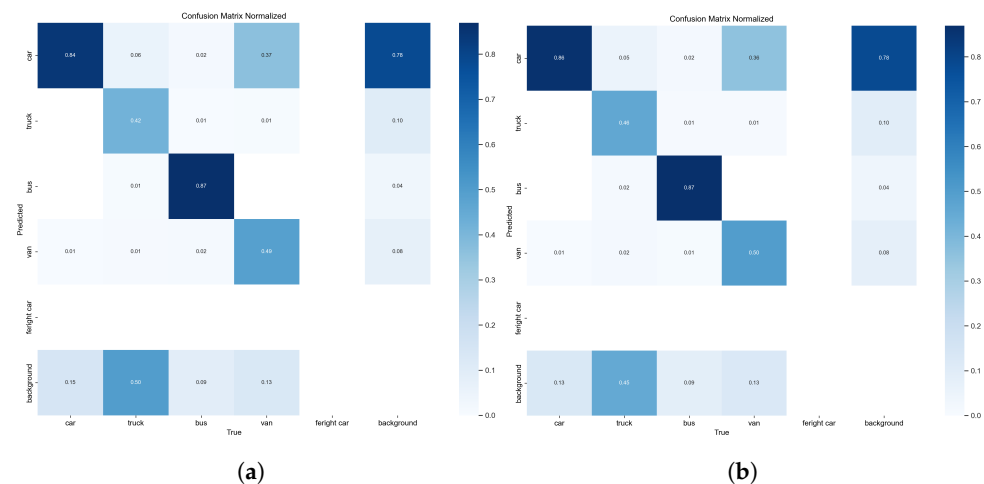


Figure 4. Comparison of confusion matrices for different detection networks on DroneVehicle dataset. (a) YOLOv11n. (b) Ours.

Overall, the ablation studies confirm the individual contributions of the HLSAFM and DPSA modules and highlight their synergistic effect. Compared with the original YOLOv11n model, the proposed model *mAP50* increases by 6.0% and 1.0%, and the *mAP50 : 95* increases by 3.1% and 0.8%, respectively. However, the performance gains differ across datasets mainly due to their distinct characteristics and available room for improvement. VisDrone contains highly challenging UAV scenes with numerous small objects, significant scale variation, frequent occlusion, and complex backgrounds (including more difficult low-illumination conditions in our setting). The proposed design emphasizes stronger multi-scale representation and improved robustness for small and dense targets, thus yielding more pronounced improvements on VisDrone. In contrast, DroneVehicle is more vehicle-centric, with more consistent object appearance and less extreme scale variation and background complexity. Consequently, the baseline already performs strongly on DroneVehicle, leaving limited headroom for further improvement and resulting in a smaller *mAP50* gain.

3.2.2. Experimental Analysis

The module of HLSAFM, through its dynamic high and low frequency feature fusion mechanism, increases the recall by 0.4% compared to the baseline model (from 0.679 to 0.683). This substantial improvement indicates a significantly enhanced capability to perceive features of blurred or low-contrast targets. In contrast, the DPSA module improves the detection precision for fine-grained targets by dynamically selecting channel and spatial features and incorporating multi-scale feature fusion. The Precision–Recall (PR) curves evaluated on the VisDrone (Night) benchmark are illustrated in Figure 5a.

As shown in Figure 5, the average precision (AP) for the category “car” reaches 0.584, while the AP for the category “bike” is 0. This disparity arises because the dataset was constructed solely based on image brightness levels, without considering category balance during selection, resulting in severe class imbalance rather than inherent limitations of the proposed model. As a result, certain categories (e.g., “bicycle”) are severely under-represented, leading to extreme class imbalance. In such cases, the model has insufficient training samples to learn discriminative features for minority classes, which directly results in zero AP for those categories. In real-world UAV surveillance scenarios, object categories often follow long-tailed distributions, where dominant classes such as cars appear far more frequently than rare classes such as bicycles. Therefore, the observed performance gap reflects realistic data conditions to some extent. Despite the severe imbalance, the proposed method achieves strong overall detection performance, demonstrating its robustness under challenging data distributions. The high AP for major categories (e.g., “car”) and competitive mean Average Precision (mAP) indicate that the model effectively captures representative features even when trained on imbalanced data.

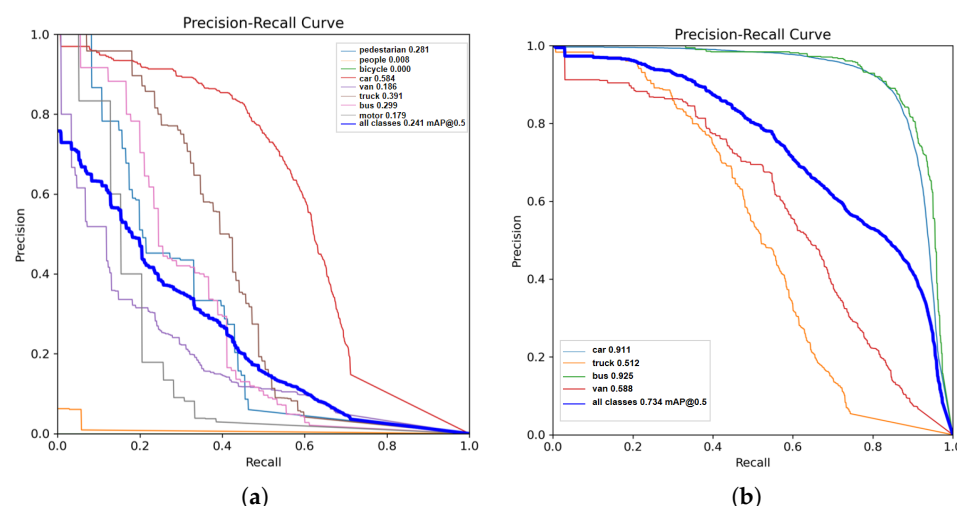


Figure 5. Precision–Recall curves on VisDrone (Night) and DroneVehicle (Night) datasets. (a) The PR curves evaluated on the VisDrone (Night). (b) The PR curves evaluated on the DroneVehicle (Night).

The complete model integrating both DPSA and HLSAFM achieves an mAP_{50} of 73.4% on the DroneVehicle (Night) dataset, marking a 1.0% improvement relative to the original YOLOv11n.

The corresponding Precision–Recall (PR) curves are illustrated in Figure 5b. Notably, the average precision (mAP_{50}) for the “car” and “bus” categories reaches 91.1% and 92.5%, respectively, while the remaining categories also exhibit strong detection performance. These experimental data verify the synergistic enhancement effect of the dual feature optimization mechanisms (HLSAFM and DPSA) in complex nighttime scenarios, enabling robust handling of varying vehicle scales and severe illumination degradation.

As the DroneVehicle (Night) subset was constructed solely based on image brightness levels without explicit consideration of category distribution, certain class imbalances and performance disparities exist across categories. However, these factors do not compromise the validity or superiority of the overall detection performance.

3.3. Visualization

This section presents qualitative detection examples on the VisDrone (Night) and DroneVehicle (Night) datasets to further highlight the performance advantages of our approach. Figure 6 illustrates that the visualization results from nighttime UAV-captured images in the VisDrone (Night) test set demonstrate that the proposed algorithm exhibits

superior object detection performance in diverse scenarios. Baseline YOLO models, when applied directly to low-illumination conditions, suffer from blurred object contours and loss of fine details due to insufficient illumination (e.g., indistinct edges of vehicles and pedestrians), leading to frequent false positives and misdetections (e.g., erroneous detections in the second row of the YOLOv8 results). The overall image darkness, coupled with the prominent background noise, results in high missed detection rates (e.g., incomplete vehicle detections in YOLOv8 and YOLOv11 visualizations).

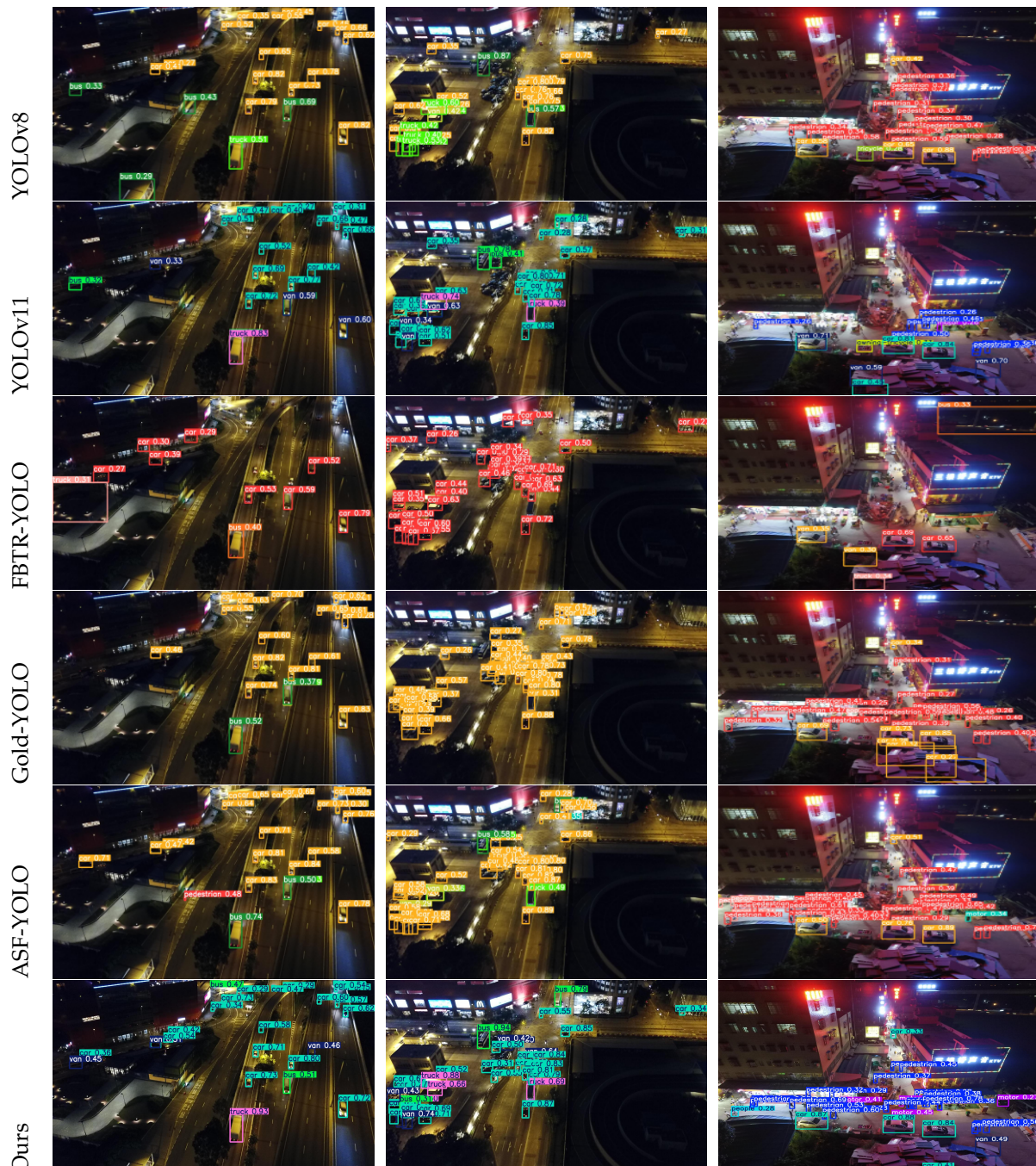


Figure 6. Visual Evaluation of detection results on the DroneVehicle (Night) dataset.

In contrast, the proposed method accurately detects the majority of targets even under complex nighttime constraints, such as strong light interference from street lamps and vehicle headlights, as well as occlusion by road structures. This robust performance fully validates the model's strong environmental adaptability and detection capability.

Compared with other baseline algorithms, the proposed approach, leveraging the multi-scene characteristics of the dataset, consistently achieves effective target detection across various conditions, reflecting its excellent adaptability to complex nighttime environments.

Figure 7 illustrates that YOLOv8 and YOLOv11 exhibit noticeable missed detections under low-illumination conditions owing to limited feature extraction capabilities; for example, the “freight car” is not detected in several scenes. Although methods such as FBRT-YOLO, Gold-YOLO, and ASF-YOLO achieve improved detection accuracy, they still suffer from partial missed detections and false positives, with a large number of erroneous “truck” detections evident in the third column. In contrast, the proposed method demonstrates significantly stronger target capture ability in dark environments. For instance, in the third column, it accurately identifies “truck” targets even in extremely dark conditions without missing “freight car” instances. These results fully validate that the proposed algorithm substantially enhances the detection performance in complex scenarios, including severe low illumination, noise interference, and varying target scales. Qualitative visualizations align closely with the quantitative improvements reported earlier, further confirming the effectiveness of the parallel architecture and the synergistic contributions of the HLSAFM and DPSA modules in handling the problems of low-illumination UAV vehicle detection.

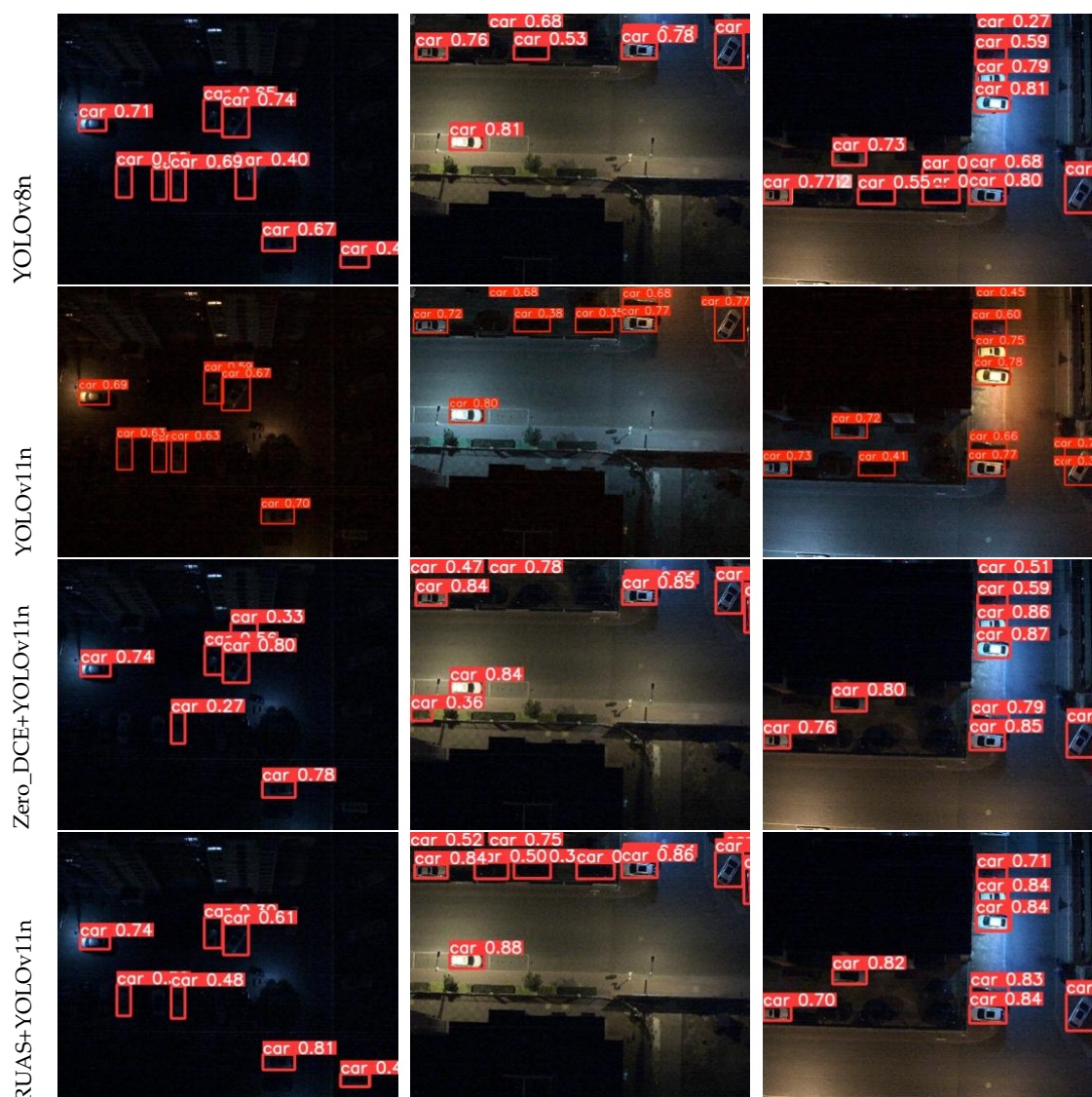


Figure 7. Cont.



Figure 7. Visual evaluation of detection results on the DroneVehicle (Night) dataset. The red bounding boxes represent detections belonging to the car category.

4. Discussion

4.1. Advantages

The proposed parallel network framework exhibits clear and practical advantages for low-illumination UAV object detection tasks. Unlike conventional two-stage pipelines that sequentially perform image enhancement and object detection, the proposed parallel architecture enables synchronous feature updating and joint optimization between enhancement and detection branches. This design allows illumination correction and semantic representation learning to interact during training, thereby reducing information degradation and feature misalignment that commonly arise in decoupled processing pipelines.

From a practical perspective, the framework is particularly effective under challenging night-time conditions where targets often suffer from low contrast, blur, and background interference. The introduction of the HLSAFM module enhances discriminative feature representation by explicitly separating low-frequency structural information from high-frequency detail cues. This decomposition enables the network to preserve global object shapes while refining fine-grained local details, which is critical for reliably detecting weakly illuminated or partially blurred targets.

In addition, the DPSA module further strengthens the framework's robustness in real-world UAV scenarios by enhancing multi-scale perception and suppressing background noise. Through dynamic channel-wise and spatial attention combined with pooling and multi-scale convolutions, the model effectively captures objects of varying sizes caused by changes in flight altitude and viewing perspective. This capability is especially important for UAV-based detection, where scale variation and cluttered backgrounds pose significant challenges to conventional detectors.

Compared with existing low-illumination UAV detection approaches, the proposed parallel framework significantly improves both precision and recall in low-light UAV scenarios, effectively reducing false detections and the risk of target overestimation while mitigating missed detections and enhancing the completeness and stability of spatial target distribution modeling. By leveraging its advantages in complex background suppression, multi-scale representation, and robustness under low-illumination conditions, the framework provides a reliable and scalable technical foundation for the practical deployment of low-light UAV target detection systems.

4.2. Challenges and Limitations

Despite the promising performance achieved by the proposed framework, several challenges and limitations should be acknowledged. First, the constructed nighttime subset was primarily selected based on image brightness without explicitly balancing object categories, which resulted in a long-tailed class distribution. This imbalance is reflected in the zero AP observed for certain underrepresented categories (e.g., "bicycle" in VisDrone (Night)). Although the overall detection performance across major categories remains competitive, the insufficient representation of minority classes may limit the applicability of the model in scenarios where rare-object detection is critical, such as emergency response or abnormal-event monitoring. Future dataset construction strategies could incorporate class-aware sampling, reweighting mechanisms, or targeted data augmentation to alleviate this imbalance and enhance robustness at the category level.

Second, under extremely low-illumination conditions, illumination enhancement operations may inadvertently amplify sensor noise. If noise suppression is insufficient, the amplified noise may interfere with feature extraction and reduce the detection stability. In addition, excessive brightening can cause saturation in already well-exposed regions, leading to the loss of fine texture and edge information. Since object detection models rely heavily on structural and boundary cues for accurate localization, such degradation may negatively affect performance in extreme lighting scenarios. Incorporating noise-aware enhancement strategies or adaptive exposure control mechanisms could help mitigate these effects in future improvements of the framework.

Third, although the proposed parallel architecture maintains real-time inference capability, the introduction of additional modules (DPSA and HLSAFM) increases the computational complexity and GFLOPs compared with the baseline detector. While the trade-off between accuracy and efficiency remains acceptable for typical edge-device deployment, further lightweight optimization may be necessary for ultra-low-power UAV platforms with stricter onboard computational constraints. Model compression techniques, pruning strategies, or hardware-aware optimization could be explored to improve deployment flexibility.

Overall, addressing dataset imbalance, extreme illumination degradation, and computational efficiency under resource-limited deployment conditions remains an important direction for future research.

4.3. Future Perspectives

Future work will focus on enhancing the generalization, robustness, and evaluation rigor of the proposed framework. This includes extending validation to more diverse datasets spanning multiple geographic regions, growth stages, and environmental conditions, as well as incorporating repeated-run experiments and statistical variance analysis for more reliable performance assessment. In addition, class-rebalancing strategies, noise-aware or exposure-aware enhancement mechanisms, and multimodal RGB–infrared fusion may be explored to further improve robustness under long-tailed distributions and extremely low-light conditions. Model compression techniques such as knowledge distillation, structured pruning, and quantization will also be investigated to improve computational efficiency without sacrificing detection accuracy.

From a real-world deployment standpoint, the parallel framework provides a balanced trade-off between detection accuracy and computational cost, making it suitable for UAV platforms with moderate onboard computing resources. Through further optimization of lightweight design and enhanced spatial suppression mechanisms, the framework is expected to maintain stable performance in extreme density or boundary scenarios, thereby supporting reliable and scalable large-scale monitoring applications under complex and low-illumination environments.

5. Conclusions

This paper introduces a parallel framework for low-illumination UAV object detection, which combines illumination enhancement and detection networks to mitigate performance degradation under adverse lighting conditions. By deploying the illumination enhancement network and object detection network in parallel, the proposed framework achieves synchronized feature updates during training, enabling mutual adaptation and joint optimization between the two networks. Furthermore, shared feature representations promote effective information exchange across networks, establishing a strong foundation for enhanced overall performance.

In terms of feature enhancement, two novel plug-and-play modules are presented: the Dynamic Pooling Synergistic Attention (DPSA) and the High and Low Frequency Spatial-Adaptive Feature Modulation (HLSAFM). The DPSA module integrates dynamic pooling-based channel attention with spatial attention mechanisms, which enhances the representational capacity of the input feature maps, improves the network's performance in complex scenarios, and increases its adaptability to multi-scale features. The HLSAFM module refines high- and low-frequency features through separated processing paths and dynamic modulation mechanisms, allowing the network to capture richer representations and improve feature expressiveness and discriminability.

The experimental results demonstrate substantial improvements over the baseline YOLOv11n detector, with $mAP50$ gains of 2.0% on the VisDrone (Night) dataset and 1.0% on the DroneVehicle (Night) dataset. In addition to improvements in academic performance metrics, these gains translate into tangible operational benefits for real-world UAV applications. The proposed method is expected to provide valuable support for safety-critical UAV missions, where enhanced detection robustness can reduce false alarms, decrease the need for manual intervention, and improve mission success rates in complex nighttime environments.

Author Contributions: Conceptualization, L.L. and G.F.; methodology, L.L. and H.L.; software, B.Z.; validation, H.L. and B.Z.; formal analysis, B.Z.; investigation, L.L., H.L. and Y.W.; resources, Y.W.; data curation, L.L., H.L. and Y.W.; writing—original draft preparation, L.L.; writing—review and editing, G.F. and R.F.; visualization, L.L. and R.F.; supervision, G.F.; project administration, G.F.; funding acquisition, G.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fund Project for Basic Scientific Research Expenses of Central Universities (Grant No. 25CAFUC03008) and the Natural Science Foundation of Sichuan Province (Grant No. 2024NSFSC0507).

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

Acknowledgments: The authors would like to thank the institute of Electronic and Electrical Engineering, Civil Aviation Flight University of China, for administrative and technical support.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Ma, J.; Sun, Z.; Wu, Y.; Jin, D. Efficient Detection Model of UAVs under Low-Light Conditions Based on LL-YOLO and EnlightenGAN. *Meas. Control* **2025**. [\[CrossRef\]](#)
2. Zhao, D.; Shao, F.; Zhang, S.; Yang, L.; Zhang, H.; Liu, S.; Liu, Q. Advanced Object Detection in Low-Light Conditions: Enhancements to YOLOv7 Framework. *Remote Sens.* **2024**, *16*, 4493. [\[CrossRef\]](#)
3. Weng, T.; Niu, X. Enhancing UAV Object Detection in Low-Light Conditions with ELS-YOLO: A Lightweight Model Based on Improved YOLOv11. *Sensors* **2025**, *25*, 4463. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Abdelnabi, A.A.B.; Rabadi, G. Human Detection From Unmanned Aerial Vehicles' Images for Search and Rescue Missions: A State-of-the-Art Review. *IEEE Access* **2024**, *12*, 152009–152035. [\[CrossRef\]](#)
5. Jaffri, S.M.A.A.; ul Haq, M.; Farhan, M. Enhancing Object Detection in Low Light Environments Using Image Enhancement Techniques and YOLO Architectures. In *Proceedings of the 26th International Multi-Topic Conference (INMIC), Karachi, Pakistan, 30–31 December 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6.
6. Liu, X.; Liu, C.; Zhou, X.; Fan, G. Enhancing Low-Light Object Detection with En-YOLO: Leveraging Dual Attention and Implicit Feature Learning. *Multimed. Syst.* **2025**, *31*, 249. [\[CrossRef\]](#)
7. Peng, D.; Ding, W.; Zhen, T. A Novel Low-Light Object Detection Method Based on the YOLOv5 Fusion Feature Enhancement. *Sci. Rep.* **2024**, *14*, 4486. [\[CrossRef\]](#)
8. Shovo Abir, S.U.A.; Kabir, M.G.R.; Mridha, M.M.; Mridha, F.M. Advancing Low-Light Object Detection with You Only Look Once Models: An Empirical Study and Performance Evaluation. *Cogn. Comput. Syst.* **2024**, *6*, 119–134. [\[CrossRef\]](#)
9. Han, Z.; Yue, Z.; Liu, L. 3L-YOLO: A Lightweight Low-Light Object Detection Algorithm. *Appl. Sci.* **2025**, *15*, 90. [\[CrossRef\]](#)
10. Li, J.; Qian, L. LLEYOLO: A Target Detection Algorithm Based on Improved YOLOv5 for Low-Light Environments. *J. King Saud Univ. Comput. Inf. Sci.* **2025**, *37*, 284. [\[CrossRef\]](#)
11. Di, R.; Fan, H.; Ma, Y.; Wang, J.; Qian, R. GAME-YOLO: Global Attention and Multi-Scale Enhancement for Low-Visibility UAV Detection with Sub-Pixel Localization. *Entropy* **2025**, *27*, 1263. [\[CrossRef\]](#)
12. Gong, B.; Zhang, H.; Ma, B.; Tao, Z. Enhancing Real-Time Low-Light Object Detection via Multi-Scale Edge and Illumination-Guided Features in YOLOv8. *J. Supercomput.* **2025**, *81*, 1120. [\[CrossRef\]](#)
13. Cai, W.; Chen, Y.; Qiu, X.; Niu, M.; Li, J. LLD-YOLO: A Low-Light Object Detection Algorithm Based on Dynamic Weighted Fusion of Shallow and Deep Features. *IEEE Access* **2025**, *13*, 69967–69979. [\[CrossRef\]](#)
14. Cai, T.; Chen, C.; Dong, F. Low-Light Object Detection Combining Transformer and Dynamic Feature Fusion. *Comput. Eng. Appl.* **2024**, *60*, 135–141.
15. Fu, S.; Zhao, Q.; Liu, H.; Tao, Q.; Liu, D. Low-Light Object Detection via Adaptive Enhancement and Dynamic Feature Fusion. *Alex. Eng. J.* **2025**, *126*, 60–69. [\[CrossRef\]](#)
16. Xu, Z.; Su, J.; Huang, K. A-RetinaNet: A Novel RetinaNet with an Asymmetric Attention Fusion Mechanism for Dim and Small Drone Detection in Infrared Images. *Math. Biosci. Eng.* **2023**, *20*, 6630–6651. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Zhang, X.; Di, X.; Liu, M. DBLDNet: Dual Branch Low Light Object Detector Based on Feature Localization and Multi-Scale Feature Enhancement. *Multimed. Syst.* **2025**, *31*, 288. [\[CrossRef\]](#)

18. Ji, J.; Zhao, Y.; Zhang, Y.; Zuo, X.; Wang, C.; Shi, F. FCMA-Det: Low-Light Image Object Detection Based on Feature Complementarity and Multicontent Aggregation. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5636414. [\[CrossRef\]](#)
19. Jung, M.; Cho, J. Enhancing Detection of Pedestrians in Low-Light Conditions by Accentuating Gaussian–Sobel Edge Features from Depth Maps. *Appl. Sci.* **2024**, *14*, 8326. [\[CrossRef\]](#)
20. Qi, G.; Yu, Z.; Song, J. Multi-Scale Feature Fusion and Context-Enhanced Spatial Sparse Convolution Single-Shot Detector for Unmanned Aerial Vehicle Image Object Detection. *Appl. Sci.* **2025**, *15*, 924. [\[CrossRef\]](#)
21. Wei, Y.; Tao, J.; Wu, W.; Yuan, D.; Hou, S. RHS-YOLOv8: A Lightweight Underwater Small Object Detection Algorithm Based on Improved YOLOv8. *Appl. Sci.* **2025**, *15*, 3778. [\[CrossRef\]](#)
22. Telçeken, M.; Akgun, D.; Kacar, S. An Evaluation of Image Slicing and YOLO Architectures for Object Detection in UAV Images. *Appl. Sci.* **2024**, *14*, 11293. [\[CrossRef\]](#)
23. Choutri, K.; Lagha, M.; Meshoul, S.; Batouche, M.; Bouzidi, F.; Charef, W. Fire Detection and Geo-Localization Using UAV's Aerial Images and YOLO-Based Models. *Appl. Sci.* **2023**, *13*, 11548. [\[CrossRef\]](#)
24. Jung, H.-K.; Choi, G.-S. Improved YOLOv5: Efficient Object Detection Using Drone Images under Various Conditions. *Appl. Sci.* **2022**, *12*, 7255. [\[CrossRef\]](#)
25. Kim, J.; Huh, J.; Park, I.; Bak, J.; Kim, D.; Lee, S. Small Object Detection in Infrared Images: Learning from Imbalanced Cross-Domain Data via Domain Adaptation. *Appl. Sci.* **2022**, *12*, 11201. [\[CrossRef\]](#)
26. Zhang, M.; Wang, Z.; Song, W.; Zhao, D.; Zhao, H. Efficient Small-Object Detection in Underwater Images Using the Enhanced YOLOv8 Network. *Appl. Sci.* **2024**, *14*, 1095. [\[CrossRef\]](#)
27. Deng, H.; Zhang, S.; Wang, X.; Han, T.; Ye, Y. USD-YOLO: An Enhanced YOLO Algorithm for Small Object Detection in Unmanned Systems Perception. *Appl. Sci.* **2025**, *15*, 3795. [\[CrossRef\]](#)
28. Miao, Y.; Wang, X.; Zhang, N.; Wang, K.; Shao, L.; Gao, Q. Research on a UAV-View Object-Detection Method Based on YOLOv7-Tiny. *Appl. Sci.* **2024**, *14*, 11929. [\[CrossRef\]](#)
29. Zhu, P.; Wen, L.; Du, D.; Bian, X.; Fan, H.; Hu, Q.; Ling, H. Detection and Tracking Meet Drones Challenge. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 7380–7399. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Sun, Y.; Cao, B.; Zhu, P.; Hu, Q. Drone-Based RGB-Infrared Cross-Modality Vehicle Detection Via Uncertainty-Aware Learning. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 6700–6713. [\[CrossRef\]](#)
31. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Wu, W.; Weng, J.; Zhang, P.; Wang, X.; Yang, W.; Jiang, J. URetinex-Net: Retinex-Based Deep Unfolding Network for Low-Light Image Enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, June 2022*; IEEE: Piscataway, NJ, USA, 2022; pp. 5901–5910.
33. Ma, L.; Ma, T.; Liu, R.; Fan, X.; Luo, Z. Toward Fast, Flexible, and Robust Low-Light Image Enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, June 2022*; IEEE: Piscataway, NJ, USA, 2022; pp. 5637–5646.
34. Li, C.; Guo, C.; Loy, C.C. Learning to Enhance Low-Light Image via Zero-Reference Deep Curve Estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 4225–4238. [\[CrossRef\]](#)
35. Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; Wang, Z. EnlightenGAN: Deep Light Enhancement Without Paired Supervision. *IEEE Trans. Image Process.* **2021**, *30*, 2340–2349. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Liu, R.; Ma, L.; Zhang, J.; Fan, X.; Luo, Z. Retinex-Inspired Unrolling with Cooperative Prior Architecture Search for Low-Light Image Enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual, 19–25 June 2021*; IEEE: Piscataway, NJ, USA, 2021; pp. 10561–10570.
37. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [\[CrossRef\]](#)
38. Guo, W.; Li, W.; Li, Z.; Gong, W.; Cui, J.; Wang, X. A Slimmer Network with Polymorphic and Group Attention Modules for More Efficient Object Detection in Aerial Images. *Remote Sens.* **2020**, *12*, 3750. [\[CrossRef\]](#)
39. Liu, H.; Duan, X.; Lou, H.; Gu, J.; Chen, H.; Bi, L. Improved GBS-YOLOv5n Algorithm Based on YOLOv5n Applied to UAV Intelligent Traffic. *Sci. Rep.* **2023**, *13*, 9577. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Zhang, Z. Drone-YOLO: An Efficient Neural Network Method for Target Detection in Drone Images. *Drones* **2023**, *7*, 526. [\[CrossRef\]](#)
41. Kong, Y.; Shang, X.; Jia, S. Drone-DETR: Efficient Small Object Detection for Remote Sensing Image Using Enhanced RT-DETR Model. *Sensors* **2024**, *24*, 5496. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.