Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

LLDFP-YOLO: A lightweight improved YOLO26 model for enhanced pedestrian detection of UAVs under low-light conditions

Xiaonan Shi
Xi'an University of Science and Technology
Qifeng Su
1443031792@qq.com
Xi'an University of Science and Technology
Limei Fu
Xi'an University of Science and Technology

Research Article

Keywords: Drone Pedestrian Detection, Low Light, YOLO26, Dual-Branch Feature Fusion, Structured Pruning

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11656524/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

LLDFP-YOLO: A lightweight improved YOLO26 model for enhanced pedestrian detection of UAVs under low-light conditions

Author： Xiaonan Shi, Qifeng Su, Limei Fu

Corresponding Author： Qifeng Su

Corresponding author's email： 23408010423@stu.xust.edu.cn

Organization: Xi'an University of Science and Technology

Abstract

Nighttime drone patrols and search-and-rescue missions require pedestrian detection in low-light aerial images. Dim lighting conditions result in low image contrast, high noise, and blurred details of pedestrian targets. The traditional two-stage “enhance-then-detect” paradigm suffers from inconsistent optimization objectives, computational redundancy, and difficulties in end-to-end joint training; simultaneously, drone-mounted embedded platforms impose strict constraints on model parameter counts and inference speed. To address these challenges, this paper proposes LLDFP-YOLO, a lightweight model for low-light pedestrian detection on UAVs, using YOLO26s as the baseline. The input layer incorporates CurveEnhance, a learnable low-light enhancement module that achieves adaptive brightness and contrast optimization through exposure loss and spatial smoothing loss; For the shallow, high-resolution P2 and P3 branches, we construct the Dual-Fusion CoordAttn module, which fuses texture and semantic features while incorporating coordinate attention to enhance the localization of small pedestrians in low-light conditions; we also employ SP-LAMP structured pruning to reduce redundant channels based on module channel dependencies, thereby compressing the model. Experiments on the VisDrone dataset show that, compared to the baseline YOLO26s, LLDFP-YOLO achieves a 1.47% increase in precision and a 2.32% increase in mAP, with only a 0.46M increase in parameters; the ExDark generalization test validates its performance in low-light conditions. This model improves the accuracy of low-light small pedestrian detection by drones and is suitable for resource-constrained airborne embedded platforms.

Keywords:Drone Pedestrian Detection; Low Light; YOLO26; Dual-Branch Feature Fusion; Structured Pruning

1.Introduction

Drone-View Object Detection (DVOD) aims to locate and classify objects in images and videos captured by drones [1]. With advancements in computer vision and drone technology, DVOD has been widely adopted in scenarios such as security surveillance,

intelligent transportation, and environmental monitoring. Wang et al. [2]'s YOLO-ADD leverages feature enhancement and context-aware adaptation for complex urban scene detection; Qiu et al. [3]'s AeroLight lightweight network achieves high-quality detection of small objects under resource-constrained conditions; Peng et al.'s [4] YOLOv8s-DroneNet relies on feature selection and the ISIoU loss to optimize the localization of small objects; Lang et al. [5] optimized YOLOv11n at the network, module, and loss function levels to address the detection of small ground targets in aerial imagery; Yan et al. [6]'s ST-YOLO utilizes multi-modules and SI-IoU to improve the detection performance of blurry small targets; Jing et al. [7]'s MCFNet enhances detection robustness in low-light conditions through multimodal fusion; Li et al. [8]'s MFA-YOLO relies on multiple modules to improve the accuracy of small-object detection in complex backgrounds.

Nighttime patrols and search-and-rescue operations require real-time monitoring over large areas. Manual inspections are inefficient and prone to oversights, while traditional surveillance systems suffer from poor low-light imaging and lack automated analysis capabilities. Drones equipped with detection algorithms offer flexible deployment and enable wide-area monitoring [9]; consequently, high-performance object detection in low-light conditions has become an important research direction [10].

Low light conditions result in low image contrast, high noise levels, and blurred details, causing a significant drop in performance when conventional detectors such as YOLO and Faster R-CNN are used directly. The current mainstream approach is a two-stage "enhance-then-detect" method, which first improves image quality before performing detection.

There has been extensive research on low-light enhancement: Edwin H. Land [11] proposed the Retinex theory, which decomposes image illumination and reflection components and is widely used for low-light enhancement and defogging; Cui et al. [12] developed a two-stage perceptual enhancement Transformer that restores brightness textures step-by-step and suppresses noise artifacts; Zhou et al. [13] proposed a lightweight local reference feature transfer method to restore details, correct colors, and adapt to various enhancement networks.

However, this two-stage architecture follows a non-end-to-end, separated paradigm: the enhancement and detection modules lack a unified optimization objective, and the enhancement tends toward general visual effects that do not align with the requirements of the detection task; the modules cannot be jointly optimized, limiting their ability to adapt features; the computational process is redundant, resulting in low training and inference efficiency and poor practical applicability.

To address the above issues, this paper proposes LLDFP-YOLO, a lightweight model for low-light scenarios based on YOLO26s, which balances accuracy and efficiency. Its main contributions are as follows:

(1) To address issues such as uneven illumination, detail degradation, and noise interference in VisDrone aerial images, the

front end incorporates the CurveEnhance learnable low-light enhancement module, which adaptively adjusts brightness and contrast end-to-end. It uses exposure and spatial smoothing losses to suppress overexposure and pseudo-noise, thereby providing high-quality feature inputs for subsequent networks.

(2) The Dual-Fusion CoordAttn module is deployed in the shallow detection heads P2 and P3. By embedding CoordAttn coordinate attention, it enables bidirectional feature encoding and preserves positional dependencies; By concatenating the enhanced image features with the backbone features and using attention-based adaptive weighting, we inject low-light texture information into the shallow layers. This enhances the network's ability to localize dense small objects (such as pedestrians) in low-light conditions and improves its detection robustness against interference, addressing issues such as the loss of shallow-layer textures and the blurring of positional features of small objects in low light.

(3) To mitigate the increase in parameter count and inference overhead caused by the multi-module architecture, we introduce SP-LAMP structured pruning. By evaluating channel importance to prune redundant channels and fine-tuning to restore accuracy, we achieve a balance between detection performance and lightweight embedded deployment.

2.Background

2.1 Object Detection from a Drone's Perspective

Object detection from a drone's perspective is a challenging research problem at the intersection of computer vision and remote sensing, influenced by unique imaging geometry, dynamic environments, and real-time constraints. Top-down imaging results in objects occupying a small proportion of the image, being densely distributed, exhibiting drastic scale variations, and appearing against a complex background, which limits the performance of general-purpose detectors. Research in this area primarily follows two technical approaches: improving the YOLO single-stage detector and exploring the Transformer detection paradigm.

Optimization of YOLO focuses on three levels: network architecture, feature fusion, and attention mechanisms. At the network architecture level, CMS-YOLOv7 introduced a P2 detection layer and removed the deep P5 heads to mitigate the degradation of small-object features caused by downsampling; however, this weakened the detection capability for large objects and resulted in insufficient generalization in extreme-scale scenarios [14]. SOD-YOLO and AFE-YOLOv8, respectively, added a small-object detection head and adopted a weighted downsampling fusion (WDF) module to enhance small-object feature extraction; however, due to limitations imposed by preset anchor boxes, they are prone to localization errors and missed detections when dealing with irregular or densely overlapping objects [15][16].

In terms of feature fusion, HCA-YOLO's hierarchical cross-scale attention framework and BDRNet's background-aware dynamic scale routing mechanism optimize cross-layer information integration and address issues

such as semantic conflicts and loss of detail. However, HCA-YOLO's computational complexity increases nonlinearly with network depth, which is not conducive to real-time inference, while BDRNet suffers from poor interpretability and is prone to routing conflicts and feature confusion in extremely dense scenes [17] [18].

In terms of attention mechanisms, DB-YOLOv5 employs a dual-trunk network combined with coordinate attention to enhance the robustness of small-object detection; however, the increase in parameters and computational overhead makes it difficult to adapt to the low-power embedded deployment requirements of drones [19][20].

Transformer-based detectors possess global modeling capabilities and strong robustness against complex backgrounds, representing a new research direction in this field. LT-DETR employs a lightweight Transformer combined with dual knowledge distillation to balance small-object detection accuracy with the resource constraints of onboard devices; however, when processing high-resolution drone images, it faces significant self-attention memory and computational pressures, and converges slowly on small datasets [21]. GHF-DETR uses a heterogeneous multi-path convolutional network (HMC) backbone and a dual-domain downsampling module to enhance weak target signals and reduce information loss; however, its structure is complex, making it difficult to implement, and the effectiveness of its modules depends on specific data distributions, so its generalization ability remains to be verified [22]. Drone-DETR is an improvement based on RT-DETR that balances accuracy and speed in remote sensing image detection on embedded drone platforms; however, it has not fundamentally resolved the performance bottleneck for small targets caused by query initialization defects .

In addition, environmental factors also pose challenges: high-speed drone flight can cause image blurring and motion blur; weather conditions such as fog and rain can reduce image quality and interfere with flight; and electromagnetic interference can cause malfunctions in motors, sensors, and communication links. All of these issues can compromise the detection system's accuracy and real-time performance.

2.2 Object Detection in Low-Light Conditions

In recent years, a large number of innovative algorithms have emerged in the field of low-light object detection, including improvements to the YOLO series, hybrid CNN-Transformer models, and methods based on Retinex theory.

3L-YOLO introduces Switchable Hole Convolution (SAConv) to eliminate the need for a separate image enhancement module, thereby reducing computational overhead. However, the lightweight architecture based on YOLOv8n limits the model's expressive power [23], resulting in insufficient feature extraction for faint objects under extreme low-light conditions and degraded detection performance for small and distant objects. While SAConv expands the field of view, complex backgrounds can introduce contextual noise, reducing localization accuracy.

GOI-YOLOv8 utilizes a batch-offset mechanism to optimize target localization in low-light conditions, but this increases network complexity and can lead to training instability [24]. This model has limited adaptability to variations in target scale, resulting in uneven detection performance in scenes where large and small targets coexist; the effectiveness of the batch-offset mechanism depends on the richness of the training data. If the dataset fails to cover various low-light degradation scenarios, the model's generalization ability will decline significantly.

NID-DETR relies on attention mechanisms to achieve precise detection, but it retains the high computational complexity inherent in Transformer models [25]. The time complexity of self-attention scales quadratically with image resolution, resulting in high computational costs when processing high-resolution, low-light images, which makes it difficult to adapt to scenarios requiring high real-time performance. Additionally, DETR has a long training convergence period and places high demands on computational resources.

Joint networks based on Retinex theory, such as N-LoLiGan, can separate reflection and illumination components; however, the physical assumptions underlying this theory do not fully align with the complexity of real-world scenarios [26][27]. In environments with strong lighting or complex reflections, decomposition is prone to artifacts and excessive smoothing, which degrade detection performance; furthermore, such methods require additional network modules to perform the decomposition, thereby increasing system complexity.

At the architectural level, research on low-light object detection is moving toward end-to-end optimization. The traditional two-stage "enhancement-then-detection" paradigm suffers from conflicting objectives: image enhancement prioritizes visual quality, while the detection task emphasizes feature representations conducive to classification and localization, resulting in inconsistent goals [28]. Consequently, integrated frameworks have become the mainstream approach. For example, DarkCORE collaboratively performs reflection denoising and target-oriented feature enhancement, balancing detection accuracy with computational efficiency [29]. End-to-end architectures enable all network modules to be optimally coordinated toward a unified objective.

3.Baseline Model

YOLO26 is the latest version in the YOLO series, offering five model variants—n, s, m, l, and x—ranging from lightweight to high-precision, and covering multi-tier deployment environments such as edge devices, embedded platforms, and high-performance computing servers. It features an overall two-stage modular architecture (see Figure 1), divided into two major modules: the backbone network and the detection head. The backbone network is responsible for extracting deep features from raw images and fusing features across different levels, while the detection head uses the fused, effective features to produce the final object detection output.

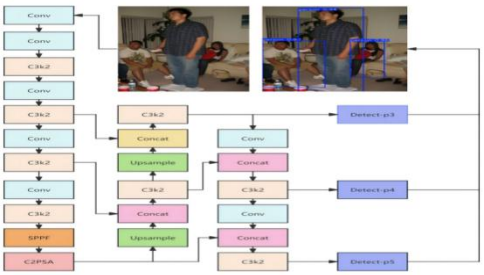

Figure 1: Overall Architecture of the YOLO26 Model

YOLO26 focuses primarily on improvements in end-to-end inference, detection head lightweighting, training strategy optimization, and multi-task adaptation. By default, the model outputs predictions directly without requiring NMS, reducing post-processing overhead and simplifying deployment. Additionally, the DFL structure has been removed to simplify the detection head while maintaining the unconstrained regression range. During the training phase, the model incorporates MuSGD, Progressive Loss, and STAL strategies to optimize convergence and shift the focus of supervision toward the inference-side detection heads, ensuring positive label coverage for small objects. Additionally, dedicated detection heads and loss functions have been designed for instance segmentation, semantic segmentation, pose estimation, and oriented object detection tasks.

We selected YOLO26s as our baseline model; compared to the “n” variant, the “s” version offers higher detection accuracy with only a slight increase in computational overhead, making it particularly well-suited for deployment on lightweight drones.

4.Methodology

4.1 Enhancement Module

To address the significant drop in object detection accuracy of the YOLO model in complex low-light scenarios, this paper designs CurveEnhance—a lightweight, pixel-level low-light enhancement module based on the Zero-DCE curve optimization approach—and embeds it into the front end of the YOLO backbone network to achieve adaptive dynamic brightness optimization of input images.

This module overcomes the limitations of traditional image enhancement algorithms with fixed parameters. Based on the Zero-DCE unsupervised learning paradigm, it does not require paired labeled data; instead, the network autonomously learns pixel-level curve adjustment parameters and iteratively optimizes image brightness and contrast features, effectively enhancing image detail representation in low-light environments. Embedding this module into the YOLO detection framework enables complete end-to-end joint training. The CurveEnhance low-light enhancement module provides high-quality input features for subsequent object feature extraction, detection, and recognition, while balancing module lightweight design with inference efficiency to meet the application requirements of real-time detection in the YOLO series.

The overall structure of the CurveEnhance module consists of two core components: feature encoding and iterative curve optimization (see Figure 2). The module has a simple structure and a very low number of parameters, allowing it to be seamlessly integrated into the YOLO model without causing a significant increase in inference latency.

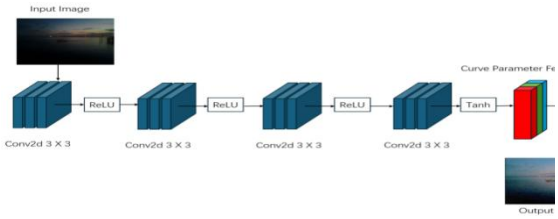

Figure 2: CurveEnhance Module Architecture Diagram

During the feature encoding stage, this paper constructs a lightweight multi-layer convolutional encoding network. Taking a 3-channel RGB raw image as input, the network extracts shallow-level texture, luminance, and contrast features through three layers of 3×3 convolutions combined with the ReLU activation function, gradually distilling effective feature information from the image. Finally, a single-layer convolution maps the extracted features to a curve parameter feature map of a dimension corresponding to the preset number of iterations. The Tanh activation function is then used to constrain all curve parameters within the interval [-1, 1], thereby avoiding issues such as overexposure, underexposure, and distortion caused by parameter extremes, and ensuring the stability and controllability of the image enhancement process.

Iterative curve optimization is the core innovative mechanism of this module. Unlike global, uniform image enhancement methods, this module implements dynamic, adaptive enhancement at the pixel level. The module is set by default to 8 optimization iterations, with each iteration corresponding to a set of independent pixel-level curve parameters. The core iterative optimization formula is:

$$LE_n(x) = LE_{n-1}(x) + \alpha_i \cdot (LE_{n-1}(x) - LE_{n-1}(x)^2) \quad (1)$$

In this equation, $LE_{n-1}(x)$ represents the enhanced image from the previous iteration, α_i represents the pixel-level curve adjustment parameters predicted by the network for the i-th iteration, and $LE_{n-1}(x) - LE_{n-1}(x)^2$ represents the quadratic curve transformation term, which remains non-negative when the image pixels are normalized to the interval [0, 1] and whose range is stable within [0, 0.25].

Based on this core formula, the module can achieve differentiated image enhancement: positive parameters brighten dark areas of the image to highlight object contours and details in low-light environments, while negative parameters suppress overexposed areas in bright regions, balancing the overall light and shadow distribution. This effectively addresses common issues in low-light scenes, such as blur caused by low light, localized overexposure, and loss of detail. After multiple rounds of iterative optimization, the final enhanced image is clipped to the [0,1] interval to ensure pixel values are within the valid range and to prevent color distortion in the image.

4.2 Dual-Branch Feature Fusion Module

To address the issues of missed and false detections in YOLO caused by weak features, missing textures, and insufficient semantic representation, this paper proposes the Dual-Fusion CoordAttn dual-branch feature fusion enhancement module. Optimized for shallow-layer small-object detection, this module incorporates CoordAttn (coordinate attention) to achieve complementary fusion of

dual-branch features and adaptive enhancement based on spatial coordinates. It addresses the issues of one-dimensional feature extraction and the disconnect between lighting and semantic features in single-branch approaches, thereby enhancing the feature representation of small objects and objects with weak textures while maintaining a lightweight architecture and ensuring real-time performance.

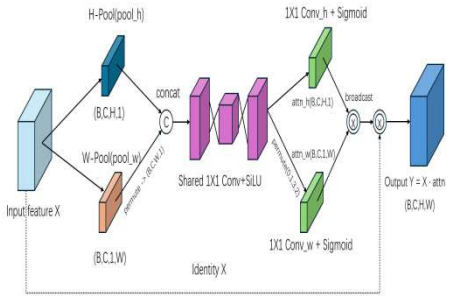

Figure 3: CoordAttn Module Architecture Diagram

The spatial attention module performs pooling on the input features along the height and width dimensions, respectively, to encode positional information. The concatenated features undergo information exchange via shared convolutions, after which the process splits into two branches to predict attention weights in the H and W directions. These weights are then broadcast and multiplied by the original features, adaptively enhancing features in the target region while suppressing irrelevant background noise, thereby efficiently modeling the spatial positional dependencies of the features (see Figure 3).

Traditional YOLO shallow features rely solely on backbone convolutions to extract semantic information, resulting in the loss of fine textures and light-dark boundaries in low-light enhanced images; small objects in complex low-light backgrounds are prone to misclassification due to feature blurring. To address this, this module employs a parallel fusion architecture with two branches—one for detection features and the other for enhanced image textures—to integrate deep semantic information from the main branch with the texture features output by CurveEnhance. This cross-dimensional approach balances semantics and details, compensating for the expressive limitations of shallow features.

Dual-Fusion CoordAttn consists of four stages: dual-branch extraction, scale alignment, feature fusion, and attention enhancement. The architecture is streamlined and free of redundancy; it can adaptively adjust the number of input channels and the downsampling ratio, dynamically generate intermediate channel dimensions, set a lower bound on the number of channels to prevent feature dimension collapse, and control computational complexity while maintaining representation performance (see Figure 4).

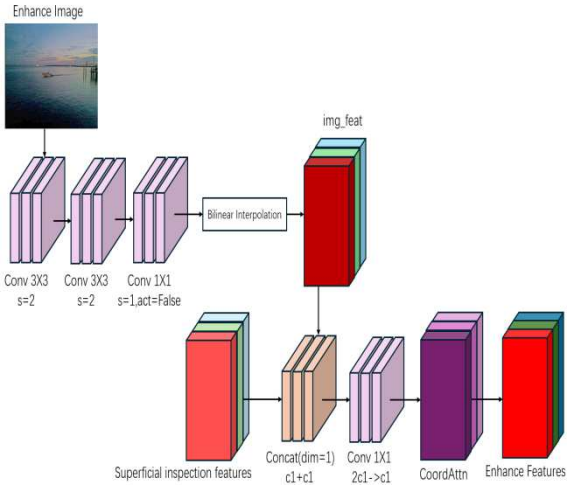

Figure 4: Architecture Diagram of the Dual-Fusion CoordAttn Dual-Branch Feature Fusion Enhancement Module

During the dual-branch feature extraction stage, the module is divided into a main branch for detecting features and a branch for enhancing textures in the original image. The backbone branch directly receives shallow detection features from layers P2 and P3 of the YOLO network, preserving the target semantic and coarse-grained positional features learned by the network; the texture branch takes the enhanced image output by the CurveEnhance module as input and performs step-by-step downsampling and feature extraction through a custom multi-layer lightweight convolutional sequence. The feature extraction process can be represented as:

$$X_{txt} = \mathrm{F}_{orig}(I_{enh}) \qquad (2)$$

In this equation, $\mathrm{F}_{orig}(\cdot)$ represents the multi-level convolutional downsampling feature extraction network defined in this paper, which performs scale compression through multiple convolutions with a stride of 2. It ultimately outputs texture features $X_{txt} \in \mathbb{R}^{B\times C_1\times H\times W}$ that perfectly match the feature channels and dimensions of the main branch, thereby efficiently condensing fine-grained texture features in low-light images and providing detailed texture information to support subsequent dual-branch fusion.

To completely avoid pixel-scale deviations and feature-size mismatches that occur during downsampling, the module incorporates a scale-adaptive correction mechanism. Before merging features from the two branches, the module automatically verifies the height and width dimensions of the two feature streams. If a size error of 1 to 2 pixels is detected, scale alignment is achieved through bilinear interpolation. The scale correction formula is as follows:

$$X_{txt} = \mathrm{Interpolate}(X_{txt}, \mathrm{size} = (H, W), \mathrm{mode} = 'bilinear') \qquad (3)$$

This operation ensures that the dimensions of the two feature spaces match perfectly, completely eliminating issues such as information misalignment and feature loss during feature stitching. It guarantees the integrity and validity of the fused features, laying the foundation for subsequent high-precision feature fusion.

During the feature fusion and enhancement stage, the module first fuses the semantic features of the main branch with the textural detail features through channel concatenation, integrating the complementary information from the two branches. The core formula for channel concatenation and convolutional dimension reduction fusion is:

$$X_{cat} = \mathrm{Concat}(X_{det}, X_{txt}), \quad X_{fuse} = \mathrm{C}_{1\times1}(X_{cat}) \qquad (4)$$

In the equation, Concat represents a channel-dimension concatenation operation, and $\mathrm{C}_{1\times1}$ represents a 1×1 convolutional layer, which enables channel dimension reduction, cross-branch feature coupling, and filtering of redundant noise, thereby integrating dual-channel features into a single composite feature. Unlike conventional feature fusion modules, this module embeds a CoordAttn coordinate attention module in the final stage of fusion to perform dual calibration—both in global space and along the channel dimension—on the fused

composite features. The formula for the attention enhancement process is as follows:

$$X_{out} = \mathrm{A}_{Coord}(X_{fuse}) \otimes X_{fuse} \quad (5)$$

In this equation, $\mathrm{A}_{Coord}(\cdot)$ represents the CoordAttn attention weight mapping function, and $\otimes$ represents the pixel-wise multiplication operation. Leveraging the attention mechanism's ability to allocate weights at the pixel level, this approach precisely enhances the weights of fusion features in small target regions, suppresses redundant background information, and compensates for the loss of spatial position information resulting from the fusion of the two branches, thereby further improving the quality and distinguishability of the fused features.

In addition, the module optimizes the model's deep copy mechanism, implements a custom deep copy function, and uses a memo mechanism to resolve training anomalies caused by nested modules and parameter references. It seamlessly integrates with the model's EMA updates and iterative training process, ensuring the stability and robustness of the training process. This design resolves parameter copying conflicts that arise when multiple innovative modules are nested, enabling the three core modules—CurveEnhance, Dual-Fusion CoordAttn, and CoordAttn—to work together stably.

In summary, the Dual-Fusion CoordAttn module achieves deep fusion of semantic features and textural details through a dual-branch structure. At the same time, it relies on an embedded coordinate attention mechanism to finely enhance the fused features, thereby specifically strengthening the small-object representation capabilities of YOLO's shallow P2 and P3 feature layers. This module is deeply integrated with and works in concert with the low-light enhancement module and the coordinate attention module described earlier, forming a complete chain of image enhancement, feature fusion, and feature optimization. It effectively improves the model's accuracy in detecting small objects under low-light conditions and in complex backgrounds, while its lightweight structure avoids imposing excessive inference overhead, thereby balancing detection performance with real-time capabilities.

4.3 SP-LAMP Pruning and Lightweighting

Deploying an object detection system on a drone platform requires real-time processing of aerial images to accurately identify key objects such as vehicles and pedestrians. However, drones are typically constrained by limited computational resources, memory, and battery capacity. Therefore, detection models must be carefully optimized to reduce computational complexity and memory consumption. Currently, model pruning has emerged as an effective structural optimization technique that can significantly reduce model size while maintaining performance close to that of the original model.

The LAMP unstructured pruning algorithm, proposed in 2021, eliminates differences in weight magnitude distributions across different network layers through intra-layer normalization, enabling global comparability of weight importance without

the need for manual adjustment of pruning ratios for each layer. However, the original LAMP uses individual weights as the pruning granularity, generating an unstructured sparse matrix that fails to achieve practical inference acceleration in general-purpose GPU environments and is thus suitable only for theoretical analysis. To further enhance the real-time inference capabilities of the proposed model on edge devices, this paper employs SP-LAMP, a structured layer-adaptive pruning method, to eliminate redundant model parameters and reduce model complexity.

4.3.1 Channel Importance Score

Let the weights of a convolutional layer be denoted by $W \in \mathbb{R}^{C_{\text{out}} \times C_{\text{in}} \times k \times k}$, where C_{out}、C_{in} represent the number of output and input channels, respectively, and k is the size of the convolutional kernel. For the i-th output channel, compute the square of the L2 norm of the convolutional kernel for that channel:

$$n_i = \|W_{i,\ldots,i}\|_2^2 = \sum c_{\text{in}} \sum_h \sum_w W_{i,c_{\text{in}},h,w}^2 \qquad (6)$$

Sort all channels in this layer in descending order by $\mathbf{n}_i$ to obtain the sequence $\{\mathbf{n}_{(1)}, \mathbf{n}_{(2)}, \ldots, \mathbf{n}_{(C_{\text{out}})}\}$, and calculate the suffix sum of the norm:

$$S_u = \sum_{t=u}^{C_{\text{out}}} n_t \qquad (7)$$

The SP-LAMP importance score for the u-th channel in the sorted list is defined as:

$$\text{SP} - \text{LAMP} - \text{Score}_u = \frac{\mathbf{n}(u)}{\mathbf{s}_u} \qquad (8)$$

Through intra-layer suffix sum normalization, SP-LAMP-Score resolves the issue of inconsistent weight scales across different convolutional layers, enabling the channel scores from different layers to provide a global comparison. By mapping the scores back to the original channel indices, we can obtain the importance scores for each output channel of the convolutional layer.

5. Experimental Results

5.1 Dataset

To evaluate the performance of the LLDFP-YOLO model in detecting human targets under low-light conditions, this paper uses the VisDrone drone aerial photography benchmark dataset [30] for the main experiments. The dataset was augmented by setting the hue shift range to 0.015, the saturation scaling range to 0.7, and the brightness scaling range to 0.4, and extensive experiments were conducted.

To further evaluate the model's generalization performance in real-world low-light environments, this paper uses the ExDark (Exclusively Dark) dataset for generalization validation [31]. This evaluates the cross-dataset generalization capability of LLDFP-YOLO for low-light human detection.

5.2 Experimental Environment

This experiment was conducted using Python 3.11.8 and PyTorch 2.8.0 on an NVIDIA GeForce RTX 5090 GPU. Each model was trained for 100 epochs with a batch size of 16 and without pre-trained weights. All comparative experiments used the same settings.

5.3 Analysis of Experimental Results

To comprehensively evaluate the overall detection performance of LLDFP-YOLO, this paper conducts side-by-side comparison experiments with various mainstream detection algorithms to validate the algorithm's performance advantages over existing networks. Additionally, through ablation experiments, we quantitatively analyze the effectiveness of each improved module and the benefits of their synergistic optimization.

All experiments were conducted using the VisDrone aerial photography dataset, and the models were comprehensively and quantitatively evaluated across multiple dimensions, including precision, recall, mAP detection accuracy, and the number of parameters.This paper specifically establishes multiple sets of progressive ablation control groups:

(1) Baseline model: The original, unmodified YOLO26s model;

(2) Single-module validation: Testing the performance of models that incorporate only the CurveEnhance low-light enhancement module, the Dual-Fusion Attn feature fusion module, or the Dual-Fusion CoordAttn feature fusion module, respectively;

(3) Multi-Module Combination Validation: Testing the detection performance of the CurveEnhance and Dual-Fusion-CoordAttn modules when integrated together, to verify the synergistic optimization capabilities of the two core modules;

(4) Complete Model: The final LLDFP-YOLO model, which integrates all improved modules and has undergone SP LAMP pruning and lightweight optimization.

The quantitative results of the cross-comparison of different algorithms are shown in Table 3, and the detailed quantitative results for each ablation experiment group are shown in Table 4.

Table 3: Results of the LLDFP-YOLO vs. other models in the VisDrone benchmark

model	Precision/%	Recall/%	mAP/%	mAP50/%	Params/M
Faster-R-CNN	43.67	32.41	22.32	38.41	41.76
EfficientDet-D1	44.32	32.61	21.44	36.18	6.62
RT-DETR-r50	43.21	31.23	24.76	43.22	42.14
YOLOX-S	46.51	33.72	23.43	38.51	9.02
YOLO26s	64.88	51.56	25.59	58.73	9.95
LLDFP-YOLO	66.35	51.50	27.91	57.58	10.41

Table 4: Results of the LLDFP-YOLO Ablation Experiment

model	Precision/%	Recall/%	mAP/%	mAP50/%	Params/M

YOLO26s	64.88	51.56	25.59	58.73	9.95
YOLO26-CE	65.36	51.61	26.46	56.34	10.22
YOLO26-DA	65.24	51.56	26.41	55.97	10.43
YOLO26-DC	65.33	51.58	26.43	57.21	10.39
LLDF-YOLO	67.48	51.53	28.03	57.75	10.67
LLDFP-YOLO	66.35	51.50	27.91	57.58	10.41

As shown in Table 3, which presents a side-by-side comparison, Faster-R-CNN, EfficientDet-D1, RT-DETR-r50, YOLOX-S, and the baseline YOLO26s were selected for comparison. In terms of accuracy, the two-stage detectors—Faster-R-CNN, EfficientDet-D1,and RT-DETR-r50—perform poorly in scenarios involving small targets against complex backgrounds in aerial imagery; YOLOX-S also underperforms compared to the YOLO26s baseline (Precision 64.88%, mAP 25.59%). In this paper, LLDFP-YOLO further improves accuracy, achieving a precision of 66.35% and an mAP of 27.91%; with only 10.41M parameters—far fewer than Faster-R-CNN (41.76M) and RT-DETR-r50 (42.14M)—and an increase of just 0.46M compared to the baseline YOLO26s (9.95M); mAP50 decreased slightly to 57.58%. LLDFP-YOLO achieves an increase in mAP with only a modest increase in the number of parameters, striking a good balance between accuracy and parameter count in complex aerial photography scenarios.

The ablation experiments were designed as a progressive series: baseline YOLO26s; sequentially adding the CurveEnhance low-light enhancement module, Dual-Fusion Attn, and Dual-Fusion CoordAttn individually; combining the two modules to obtain LLDF-YOLO; and then applying SP LAMP structured pruning to obtain the final LLDFP-YOLO, thereby validating the independent gains and synergistic effects of each module. Ablation results for individual modules:

1.CurveEnhance (YOLO26-CE) only: Precision 65.36%, mAP 26.46% (+0.87%), which improves performance on low-light aerial images; mAP50 drops to 56.34%, and the number of parameters increases slightly to 10.22M.

2. Dual-Fusion Attn only (YOLO26-DA): mAP 26.41% (+0.82%); bidirectional feature fusion benefits multi-scale aerial object detection.

3.Dual-Fusion CoordAttn only (YOLO26-DC): mAP 26.43% (+0.84%), where the coordinate attention enhancement network improves the ability to capture positional information for small aerial targets.

The combination of CurveEnhance and Dual-Fusion-CoordAttn into LLDF-YOLO yields significant synergistic gains: Precision 67.48%, mAP 28.03% (2.44% higher than the baseline), and mAP50 recovering to 57.75%, with 10.67M parameters. The synergistic gains from combining these two modules exceed those achieved by simply stacking them, confirming the synergistic optimization effect between the modules.

Introducing SP LAMP pruning yields LLDFP-YOLO, with the number of parameters reduced from 10.67M to 10.41M; mAP decreases only slightly to 27.91%, while precision drops to 66.35%. Pruning removes redundant parameters with virtually no loss in accuracy, achieving a balance between detection performance and model lightweighting. Ablation experiments show that both CurveEnhance and Dual-Fusion-CoordAttn can independently improve performance and exhibit synergistic gains; SP LAMP pruning effectively compresses the model with minimal accuracy loss. Overall, these results validate the effectiveness of all the improvement strategies proposed in this paper.

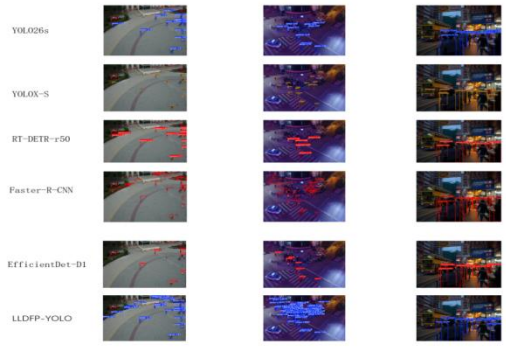

Figure 5: Visual Comparison of Detection Performance Across Different Algorithms in Low-Light Aerial Photography Scenes from the VisDrone Dataset

Figure 5 shows the visualization results of aerial photography samples under low-light conditions. In dimly lit scenes with dense crowds of pedestrians, comparison algorithms generally fail to detect small pedestrians; LLDFP-YOLO can restore details of low-light targets and detect more small pedestrians in dense crowds, intuitively validating the effectiveness of the improvements.

After validating the effectiveness of each module through comparative and ablation experiments on the VisDrone dataset, we conducted generalization comparison experiments on the ExDark dataset—a low-light dataset—to further verify the cross-dataset generalization performance of the LLDFP-YOLO model under low-light conditions and rule out performance gains attributable to dataset-specific characteristics. LLDFP-YOLO was compared with Faster-R-CNN,EfficientDet-D1,RT-DETR-r50, YOLOX-S, and the baseline YOLO26s. Precision, recall, mAP, mAP50, and the number of parameters were again used as evaluation metrics. The experimental results are shown in Table 5.

Table 5: Generalization Experiment ExDark

model	Precision/%	Recall/%	mAP/%	mAP50/%	Params/M
Faster-R-CNN	67.40	52.60	35.22	58.93	41.76
EfficientDet-D1	62.42	54.73	35.21	58.46	6.62
RT-DETR-r50	66.23	56.81	37.72	64.33	42.14
YOLOX-S	63.71	55.20	37.13	60.74	9.02
YOLO26s	60.17	22.57	13.53	24.62	9.95
LLDFP-YOLO	78.39	66.18	27.91	57.58	10.41

As shown in Table 5, the baseline YOLO26s exhibits a significant drop in performance on the ExDark low-light dataset, with a recall of only 22.57% and an mAP of only 13.53%, indicating that the original

network is not sufficiently adapted to low-light images. In this paper, LLDFP-YOLO, under the lightweight constraint of only 10.41M parameters, achieves a precision of 78.39% and a recall of 66.18%,representing significant improvements over the baseline. Although the mAP50 is slightly lower than that of RT-DETR-r50 and YOLOX-S, the precision and recall have improved significantly. These results demonstrate that the low-light enhancement and feature fusion strategies proposed in this paper are not limited to the VisDrone aerial photography dataset but are equally effective on general low-light images, indicating that the model possesses strong cross-dataset generalization capabilities.

6.Conclusion

This study proposes LLDFP-YOLO, an enhanced model based on YOLO26s. The model integrates the learnable low-light enhancement module CurveEnhance, the Dual-Fusion CoordAttn two-branch feature fusion module with embedded coordinate attention, and the SP-LAMP structured pruning strategy, effectively improving the model's performance in detecting small pedestrian targets and its feature perception capabilities in low-light aerial photography scenarios captured by drones.

First, this paper proposes a low-light enhancement module called CurveEnhance. Drawing inspiration from the unsupervised learning approach of Zero-DCE, it uses a convolutional network to learn pixel-level brightness adjustment curves and combines exposure loss with spatial smoothing loss constraints to adaptively improve the quality of low-light images. Second, the Dual-Fusion CoordAttn module uses a dual-branch structure to fuse enhanced image texture features with the main semantic features, and introduces the CoordAttn coordinate attention unit to address the limitations of convolutions in modeling spatial position information; the SP-LAMP structured pruning strategy prunes and reconstructs redundant network channels, alleviating the increased computational cost caused by the multi-module architecture. Extensive experiments conducted on the VisDrone aerial photography dataset and the ExDark low-light dataset fully validate the effectiveness of the proposed algorithm, and ablation studies similarly confirm the positive contributions of each improved module and the synergistic benefits among them. Although the proposed method improves the detection performance of dense small pedestrian targets in highly challenging low-light environments such as nighttime aerial photography, the model still has some limitations. On the one hand, the addition of improved modules increases the complexity of the network architecture, leading to longer inference times for the unpruned model and placing certain demands on the computational power of onboard devices. On the other hand, generalization validation was conducted only on the ExDark dataset; the model's generalization capabilities in complex and adverse imaging scenarios—such as rain, snow, backlighting, and dense fog—as well as on other types of datasets, still require further verification.

Looking ahead, several areas warrant further research. First, the model could be optimized for more challenging imaging conditions—such as backlighting, rain, and fog—which introduce additional interference

in drone-based pedestrian detection. Second, research could focus on enhancing the model's adaptability to scenes with drastic changes in lighting, thereby improving detection stability under complex lighting fluctuations. Additionally, exploring a multi-task collaborative framework that integrates object detection and tracking with trajectory prediction represents another important research direction. Finally, iteratively optimizing pruning and module computation logic to further reduce inference latency and build lightweight models better suited for resource-constrained onboard embedded devices remains a key focus for future work.

Acknowledgments: Part of this research was supported by the Shaanxi Provincial Natural Science Basic Research Program (2026JC-YBMS-0650).

All authors participated in the review process to ensure the quality and completeness of the manuscript.

Data Availability: No datasets were generated or analyzed in this study.

Statement

Conflicts of Interest: The author declares that there are no conflicts of interest.

References

[1] Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q., & Ling, H. (2022). Detection and Tracking Meet Drones Challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(11), 7380 – 7399. https://doi.org/10.1109/tpami.2021.3119563

[2] Wang, R., Yang, C., Teng, X., Chen, Y., Zhang, G., & Mou, X. (2026). YOLO-ADD: A high-precision detection model for small low-altitude drones in complex scenarios. Applied Soft Computing, 199, 115347. https://doi.org/10.1016/j.asoc.2026.115347

[3] Qiu, H., Meng, X., Zhao, Y., Yu, L., & Yin, S. (2025). AeroLight: A Lightweight Architecture with Dynamic Feature Fusion for High-Fidelity Small-Target Detection in Aerial Imagery. Sensors, 25(17), 5369. https://doi.org/10.3390/s25175369

[4] Peng, J., He, H., & Zhang, D. (2025). YOLOv8s-DroneNet: Small Object Detection Algorithm Based on Feature Selection and ISIoU. Computers, Materials & Continua, 84(3), 5047–5061. https://doi.org/10.32604/cmc.2025.066368

[5] Lang, B., Yang, H., Xu, R., & Li, H. (2026). YOLOSO: An Improved YOLO-Based Algorithm for UAV to Detect Small Ground Targets. Drones, 10(7), 484. https://doi.org/10.3390/drones10070484

[6] Yan, H., Kong, X., Wang, J., & Tomiyama, H. (2025). ST-YOLO: An Enhanced Detector of Small Objects in Unmanned Aerial Vehicle Imagery. Drones, 9(5), 338. https://doi.org/10.3390/drones9050338

[7] Jing, J., Hu, J. F., Zhao, Z. P., & Liu, Y. (2025). MCFNet: Research on small target detection of RGB-infrared under UAV perspective with multi-scale complementary feature fusion. IET Image Processing, 19(1). https://doi.org/10.1049/ipr2.13320

[8] Li, S., & Chen, C. (2025). MFA-YOLO: a multi-feature aggregation approach for small-object detection method in drone imagery. Scientific Reports, 16(1). https://doi.org/10.1038/s41598-025-32247-9

[9] Zhou, L., Dong, Y., Ma, B., Yin, Z., & Lu, F. (2024). Object detection in low-light conditions based on DBS-YOLOv8. Cluster Computing, 28(1). https://doi.org/10.1007/s10586-024-04829-1

[10] Kaur, R., & Singh, S. (2023). A comprehensive review of object detection with deep learning. Digital Signal Processing, 132, 103812. https://doi.org/10.1016/j.dsp.2022.103812

[11] Chen, F., Zhu, Y., & Muratova, G. V. (2021). Two-step modulus-based matrix splitting iteration methods for retinex problem. Numerical Algorithms, 88(4), 1989–2005. https://doi.org/10.1007/s11075-021-01103-y

[12] Cui, H., Li, J., Hua, Z., & Fan, L. (2022). TPET: Two-stage Perceptual Enhancement Transformer Network for Low-light Image Enhancement. Engineering Applications of Artificial Intelligence, 116, 105411. https://doi.org/10.1016/j.engappai.2022.105411

[13] Zhou, L., Zhang, W., Zheng, Y., Wang, J., & Zhao, W. (2024). Local Reference Feature Transfer (LRFT): A simple pre-processing step for image enhancement. Pattern Recognition Letters, 186, 330–336. https://doi.org/10.1016/j.patrec.2024.10.013

[14] Qin, J., Yu, W., Feng, X., Meng, Z., & Tan, C. (2024). A UAV Aerial Image Target Detection Algorithm Based on YOLOv7 Improved Model. Electronics, 13(16), 3277. https://doi.org/10.3390/electronics13163277

[15] Sun, K., Li, D., & Song, Y. (2024). Improved multi-scale small target detection by UAV. Multimedia Tools and Applications, 84(22), 24789–24803. https://doi.org/10.1007/s11042-024-20076-3

[16] Li, Y., Li, Q., Pan, J., Zhou, Y., Zhu, H., Wei, H., & Liu, C. (2024). SOD-YOLO: Small-Object-Detection Algorithm Based on Improved YOLOv8 for UAV Images. Remote Sensing, 16(16), 3057. https://doi.org/10.3390/rs16163057

[17] Liu, Y., Li, W., Tan, L., Huang, X., Zhang, H., & Jiang, X. (2023). DB-YOLOv5: A UAV Object Detection Model Based on Dual Backbone Network for Security Surveillance. Electronics, 12(15), 3296. https://doi.org/10.3390/electronics12153296

[18] Zheng, Y., Jing, Y., Zhao, J., & Cui, G. (2024). LAM-YOLO: Drones-based Small Object Detection on Lighting-Occlusion Attention Mechanism YOLO. In Computer Vision and Pattern Recognition.

[19] Zheng, X., Shao, F., Liu, Q., Dai, J., Yue, Y., Zhang, T., & Chen, C. (2026). BDRNet: Background-Aware Dynamic-Scale Routing Network for UAV Remote Sensing Object Detection. Remote Sensing, 18(12), 1987. https://doi.org/10.3390/rs18121987

[20] Han, Z., Jia, D., & Zhang, L. (2025). LT-DETR: lightweight UAV object detection and dual knowledge distillation for remote sensing scenarios. Measurement Science and Technology, 36(3), 036005. https://doi.org/10.1088/1361-6501/adb326

[21] Hu, L., Huang, Q., Liu, Z., & Ye, H. (2026). GHF-DETR: An Improved DETR Framework with a Multi-Path Backbone and Dual-Domain Downsampling for UAV Object Detection. Remote Sensing, 18(13), 2239. https://doi.org/10.3390/rs18132239

[22] Kong, Y., Shang, X., & Jia, S. (2024). Drone-DETR: Efficient Small Object Detection for Remote Sensing Image Using Enhanced RT-DETR Model. Sensors, 24(17), 5496. https://doi.org/10.3390/s24175496

[23] Han, Z., Yue, Z., & Liu, L. (2024). 3L-YOLO: A Lightweight Low-Light Object Detection Algorithm. Applied Sciences, 15(1), 90. https://doi.org/10.3390/app15010090

[24] Mei, M., Zhou, Z., Liu, W., & Ye, Z. (2024). GOI-YOLOv8 Grouping Offset and Isolated GiraffeDet Low-Light Target Detection. Sensors, 24(17), 5787. https://doi.org/10.3390/s24175787

[25] Pan, Q., Liu, Q., & Huang, W. (2025). NID-DETR: A novel model for accurate target detection in dark environments. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-98173-y

[26] Wang, J., Xiao, G., Zhu, H., Li, W., Cui, J., Wan, Y., Wang, Z., & Sui, Q. (2023). N-LoLiGan: Unsupervised low-light enhancement GAN with an N-Net for low-light tunnel images. Digital Signal Processing, 143, 104259. https://doi.org/10.1016/j.dsp.2023.104259

[27] Jiang, Y., Zhu, J., Li, L., & Ma, H. (2024). A Joint Network for Low-Light Image Enhancement Based on Retinex. Cognitive Computation, 16(6), 3241–3259. https://doi.org/10.1007/s12559-024-10347-4

[28] Yuan, J., Hu, Y., Sun, Y., Wang, B., & Yin, B. (2024). A plug-and-play image enhancement model for end-to-end object detection in low-light condition. Multimedia Systems, 30(1). https://doi.org/10.1007/s00530-023-01228-1

[29] Feng, X., Wang, J., Wang, Y., Zeng, J., & Ming, D. (2026). DarkCORE: Efficient low-light object detection via collaborative reflectance denoising and object-oriented feature enhancement. Expert Systems with Applications, 323, 132424. https://doi.org/10.1016/j.eswa.2026.132424

[30] Zhu, P., Wen, L., Bian, X., Ling, H., & Hu, Q. (2018). Vision Meets Drones: A Challenge (Version 2). arXiv. https://doi.org/10.48550/ARXIV.1804.07437

[31] Loh, Y. P., & Chan, C. S. (2018). Getting to Know Low-light Images with The Exclusively Dark Dataset. In Computer Vision and Pattern Recognition.