2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

LADF-YOLO: A Highly Accurate Low-Light Target Detection Algorithm

Songyang Li[1][0009-0005-3690-9608] and Jianping Shuai[1][0009-0009-5698-3732]* and Ya Zhou[1][0009-0007-9488-7712] and Yaoyang Zhang[1][0009-0008-5380-3057] and Yingying Chen[1][0009-0002-6828-4436]

[1] Guilin University of Electronic Technology, Guilin 541004, China
*shuaijp@guet.edu.cn

Abstract. Images captured in complex low-light environments often exhibit weak contrast, high noise, and blurred edges. Directly applying existing target detection models to low-light images can lead to missing details and inaccurate localization, resulting in poor detection accuracy. To address these issues, this paper presents a low-light target detection method based on LADF-YOLO. The method first introduces a ReS Feature Pyramid Network (ReSFPN) integrated with a backbone network to capture more effective image features in low-light conditions. The method then designs a detection head that eliminates the need for non-maximum suppression (NMS-Free), utilizing a dual-label assignment strategy and a consistent matching metric to align the optimization direction of the head, thereby enhancing the model's overall performance. Finally, experiments on the real low-light image dataset DarkFace demonstrate that the proposed LADF-YOLO outperforms other leading target detection algorithms in low-light conditions. Compared to the benchmark model YOLOv8, LADF-YOLO achieves a 10.8% improvement in mAP@0.5 and a 9.9% improvement in Recall.

Keywords: Low-light images, Targeted Detection, YOLO, Feature Pyramid Network.

1 Introdution

Low-light target detection, a key research area in computer vision, holds significant application value in fields such as nighttime security [1], autonomous driving [2], and agricultural production [3]. However, its technical implementation faces substantial challenges. Due to issues like low image brightness, noise interference, and color distortion from insufficient lighting, traditional target detection methods struggle to effectively extract feature information, leading to frequent false detections and missed detections. Overcoming the bottleneck of feature representation in low-light environments and enhancing the robustness of target recognition are key breakthroughs for advancing the practical application of target detection technology.

Current mainstream methods for low-light target detection typically follow a two-stage framework: first, the low-light image undergoes pre-processing with brightness enhancement and noise suppression using image enhancement techniques, and then it is input into the target detection network to perform the recognition task. Image enhancement methods are mainly classified into two types: traditional algorithm-based strategies, such as histogram equalization, and deep learning-based end-to-end enhancement models. Both approaches require independent enhancement modules that run sequentially with the detection network, leading to computational redundancy and error accumulation within the system.

Traditional image enhancement methods improve low-light image quality through various techniques. Histogram equalization-based algorithms enhance visibility by adjusting global or local contrast, such as adaptive HE [4] and JHE, which incorporates pixel neighborhood information [5].The STAR model, based on Retinex theory [6], optimizes image structure and texture by decomposing light and reflection. There are also the Progressive Recursive Network (PRIEN) [7] and fused gamma correction automatic conversion technology [8], which enhance brightness from the perspectives of deep learning and probability distribution, respectively. However, these methods often lead to issues like local overexposure, noise amplification, and color distortion when enhancing dim areas. Balancing background suppression with target feature preservation is challenging, resulting in key information being obscured by noise or halo effects during contrast adjustment, which significantly impacts subsequent detection accuracy.

Current deep learning-based low-light detection methods still predominantly rely on a discrete enhancement-detection architecture. For example, the IAT network [9] enhances the image using an illumination adaptive transformer before applying YOLOv3 for detection, while Vinoth et al. [10] build a serial system using a lightweight YOLOv8. Although these methods outperform traditional algorithms in terms of enhancement, their discrete architecture requires separate training of the enhancement module and detection network, leading to issues such as computational redundancy and error propagation. Additionally, achieving end-to-end joint optimization is challenging, which limits model deployment efficiency and hinders performance breakthroughs.
The single-stage object detection algorithm directly performs object localization and classification using dense anchor boxes or feature points. Its core architecture is exemplified by the YOLO [11] and SSD [12] series: YOLO uses a joint regression strategy with global image input and output layers, while SSD predicts both target category and position through multi-scale convolutional layers. Both achieve efficient detection via end-to-end single forward inference. Studies have shown that YOLO outperforms SSD and two-stage models in detection speed and low-light adaptability [13]. ViT-YOLO [14] further incorporates the MHSA-Darknet backbone network to enhance global feature extraction and introduces a weighted BiFPN to optimize cross-scale feature fusion, significantly improving detection robustness in complex scenarios. With its real-time performance and accuracy, the YOLO series has become the mainstream solution for industrial-grade low-light detection.

To overcome the structural limitations of existing two-stage methods, this paper proposes an end-to-end low-light object detection algorithm, LADF-YOLO. The model innovatively builds a light-adaptive deep network based on the YOLOv8 architecture, enabling direct target localization and classification on the original low-light images through multi-scale feature fusion and noise robustness optimization. Compared to the traditional cascade enhancement-detection paradigm, LADF-YOLO eliminates the need for an independent image enhancement preprocessing module, effectively avoiding issues like error accumulation. This significantly improves detection accuracy and recall rate in low-light scenes.

The core innovation of the LADF-YOLO algorithm proposed in this paper consists of two key modules: First, we build a ReS feature pyramid network (ReSFPN), create a bidirectional cross-level feature interaction channel using deformable convolution, and dynamically adjust multi-scale feature weights by combining spatial-channel dual attention mechanisms, This effectively addresses the issues of information redundancy and semantic conflict when shallow detail features are fused with deep semantic features. Secondly, we design the NMS-Free [15] detection head, which utilizes a decoupled dual-label allocation strategy, The coordinated optimization of classification confidence and localization accuracy is achieved through task-consistent matching metrics, significantly improving recall rate and bounding box regression accuracy for target detection in low-light images, while avoiding the computational overhead and parameter sensitivity issues caused by traditional NMS post-processing.

In summary, the main contributions of this paper are as follows.

- This article overcomes the bottleneck of low-light detection through the innovative dual-path architecture: The ReS feature pyramid network (ReSFPN) introduces a bidirectional cross-level feature interaction mechanism and uses channel-spatial attention to dynamically fuse multi-scale features, significantly enhancing target detection accuracy and model inference efficiency in low-light scenes.
- A low-light target detection algorithm, LADF-YOLO, is proposed, with a dual-path optimization framework built on the YOLOv8 [16] architecture: By integrating the ReS feature pyramid network (ReSFPN), spatial attention-guided multi-scale feature interaction is achieved. Deformable convolution is combined to enhance the multi-level semantic representation of dark area targets, significantly improving the multi-scale feature fusion effectiveness. Additionally, an NMS-Free detection head is designed, utilizing a decoupled dual-label allocation strategy. It achieves coordinated optimization of confidence and localization accuracy through task-consistent matching metrics, offering an efficient solution for accurate real-time detection in low-light environments.
- LADF-YOLO obtains reliable results through rigorous experiments on the publicly available real low-light image dataset, DarkFace. Compared to the baseline YOLOv8 model, LADF-YOLO improves mAP@0.5 by 10.8% and Recall by 9.9%.

This study conducted a systematic evaluation on the low-light benchmark dataset, DarkFace. The experimental results demonstrate that: Compared to the baseline YOLOv8 model, LADF-YOLO's mAP@0.5 increased by 10.8 percentage points, reaching 58.6%, and its recall rate improved by 9.9%, reaching 52.3%. The model demonstrates significant performance advantages in ultra-low-light conditions (illuminance <1 lux), highlighting its effective collaboration between feature extraction and target positioning through synergistic optimization.

In summary, experimental verification shows that LADF-YOLO achieves breakthrough innovative advantages in low-light scenes through the collaborative optimization mechanism of the ReS feature pyramid network and the NMS-Free detection head: Compared to mainstream detection models, its detection accuracy is improved by 10.8%, with a particularly significant improvement in target recall rate in extremely dark areas (illuminance <1 lux). This algorithm effectively addresses the issues of feature degradation in traditional two-stage methods, providing a robust and efficient solution for low-light real-time detection scenarios, such as nighttime security and autonomous driving.

2 Related Work

2.1 The YOLOv8 object detection model

YOLOv8 (You Only Look Once Version 8) is a single-stage, real-time object detection framework released by Ultralytics in 2023. It strikes an excellent balance between detection accuracy and real-time performance under standard lighting conditions. Its architecture consists of four core modules: the input preprocessing module, which enhances model robustness through adaptive image scaling. The backbone network utilizes CSPDarkNet for efficient feature extraction. The neck structure is configured with a Path Aggregation Feature Pyramid (PAFPN) for multi-scale feature fusion. The detection head simplifies the positioning process using an anchor-free mechanism. By designing dual-dimensional parameterization for width (number of channels) and depth (number of layers), five progressively enhanced model architectures—n, s, m, l, and x—are formed, The model complexity and detection performance exhibit a positively correlated gradient, offering a flexible model selection space for deployment from mobile to cloud.

Since YOLOv8 requires fewer parameters, it offers better detection performance compared to YOLOv5 and YOLOv7. Additionally, YOLOv8 requires relatively low computational cost. Therefore, this paper chooses YOLOv8 as the foundational framework.

2.2 Feature Pyramid Network (FPN)

Feature Pyramid Networks (FPN) [17] are a fundamental architecture in object detection. They integrate deep semantic information into shallow features through a top-down feature propagation mechanism, This effectively addresses the semantic gap problem in cross-scale object detection. This paradigm has been thoroughly validated in classic models such as Faster R-CNN [18] and Mask R-CNN [19]. However, traditional FPN uses a homogenized feature fusion strategy and does not differentiate between the semantic value and spatial detail contribution of features at different levels. To overcome this bottleneck, this paper introduces the ReS Feature Pyramid Network (ReSFPN), through the spatial-channel dual attention mechanism, multi-scale feature weights are dynamically calibrated. A bidirectional interactive channel is established between shallow detail features and deep semantic features, enabling adaptive fusion of cross-level heterogeneous features and significantly enhancing multi-scale target detection accuracy.

2.3 Non-maximum suppression (NMS)

Non-Maximum Suppression (NMS) is a core technology in target detection post-processing. It removes redundant detection boxes through confidence sorting and intersection-over-union (IoU) screening. However, its threshold sensitivity often results in missed detections or the incorrect removal of key targets. To address this issue, researchers have proposed various improvement solutions. For example, Zhao et al. [20] developed a dynamic NMS network (D-NMS Net) to achieve image-adaptive threshold prediction. Kumar et al. [21] designed a differentiable grouping NMS based on geometric constraints to improve the consistency of box regression in 3D detection tasks. However, traditional NMS and its derivative methods still introduce significant inference delay, and the parameter tuning process often disrupts the end-to-end nature of the detector. To address this, Wang et al. proposed the NMS-Free detection paradigm. Using a dual-label assignment strategy and task-alignment measurement method, they established a strong correlation between classification confidence and positioning accuracy during training, enabling the model to directly output non-redundant, high-precision detection boxes.

3 The proposed LADF-YOLO model

3.1 Overview of the LADF-YOLO Model

This paper presents the LADF-YOLO model for low-light target detection, which is composed of a CSPDarknet backbone network, a ReS feature pyramid network (ReSFPN), and an NMS-Free detection head (as shown in Figure 1). The core innovations are reflected in three aspects: (1) The use of the CSPDarknet backbone network to construct a multi-scale gradient flow, enhancing the feature representation capability of low-light images through a cross-stage partial connection strategy. (2) The design of the ReS feature pyramid network, which dynamically calibrates the fusion weights of P3-P5 multi-scale features through the bidirectional cross-level feature interaction channel and the spatial-channel dual attention mechanism, improving the accuracy of cross-level feature fusion. (3) A novel NMS-Free detection head is proposed, employing a decoupled dual label assignment strategy and achieving coordinated optimization of classification confidence and positioning accuracy through task-consistent matching metrics, thus completely eliminating the need for NMS post-processing.

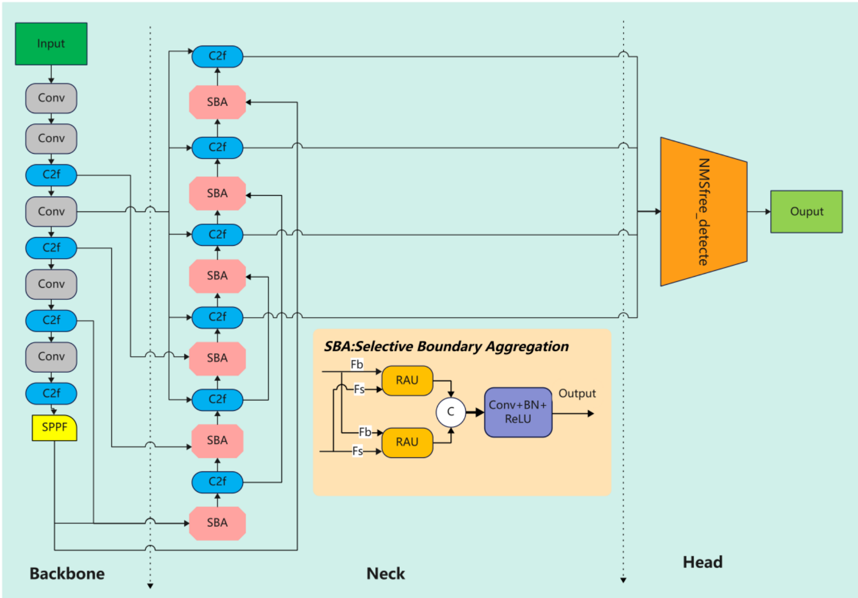

Figure 1 shows the structure of the LADF-YOLO model.

3.2 ReS Feature Pyramid Network (ReSFPN)

To tackle the challenge of target feature extraction in low-light scenes, this paper proposes the ReS Feature Pyramid Network (ReSFPN), as shown in Figure 2. The ReSFPN incorporates the Selective Boundary Aggregation (SBA) module [22], illustrated in Figure 3. This module creates a dynamic feature interaction mechanism using deformable convolution, which can be mathematically expressed as:

$$PAU(T_1, T_2) = (W_\theta(T_1) \odot T_1) + (W_\phi(T_2) \odot T_2 \odot (\ominus\, (W_\theta(T_1)))) + T_1 \qquad (1)$$

Where $\mathrm{W}_\theta(\cdot) = \mathrm{Sigmoid}(\mathrm{Conv}^{\theta}_{1\times1}(\cdot))$, $\mathrm{W}_\phi(\cdot) = \mathrm{Sigmoid}(\mathrm{Conv}^{\phi}_{1\times1}(\cdot))$.

The input features T_1 and T_2 are processed by two linear transformations $W_\theta(\cdot)$ and $W_\phi(\cdot)$ with sigmoid activation, reducing their channel dimensions to 32 to generate refined feature maps $\mathrm{W}_\theta(\mathrm{T}_1)$ and $W_\phi(T_2)$.$\odot$ is Point-wise multiplication.$\ominus$ is the reverse operation by subtracting the feature $W_\theta(T_1)$. The deep semantic features $\mathrm{F}^{\mathrm{s}}$and shallow boundary features F^{b}as T_1, T_2 input dual path processing: The main path is compressed to 32 channels using 1×1 convolution and sigmoid activation to generate a spatial attention mask $\mathrm{W}_\theta(\mathrm{T}_1)$, semantic weights are used to calibrate deep features. The compensation path operates through the inverse mask ($\ominus\, (W_\theta(T_1))$) guides the refinement of shallow boundary features F^{b}. The two outputs are merged across levels and fused using a 3×3 convolution to generate the final feature map:

$$Z = \mathrm{C}_{3\times3}(Concat[PAU(F^s, F^b), PAU(F^b, F^s)]) \qquad (2)$$

The SBA module integrates multi-level features through a $C_{3\times3}$ block (3×3 convolution with batch normalization and ReLU), combining F^s(H/8×W/8×32) from encoder layers 3-4 for semantics and F^b(H/4×W/4×32) from the backbone's initial layer for boundary details. Leveraging a channel-space dual attention mechanism, the shallow path employs deformable convolution to dynamically correct deep feature positioning via pixel-level boundary responses, while the deep path utilizes channel attention to suppress low-light background noise. Concatenation of these enhanced features yields the refined output Z (H/4×W/4×32).

Compared to traditional FPN, it significantly mitigates feature degradation by introducing a mathematically modeled cross-level dynamic calibration and a parameterized feature compensation mechanism.

Furthermore, this paper systematically improves the architectural limitations of YOLOv8's Path Aggregation Feature Pyramid (PAFPN) in low-light scenes. The original PAFPN enhances global semantic perception through a bottom-up path, but its four-level fixed structure (P3-P6) struggles to effectively capture the edge features of the P2/P3 levels in low-light images. ReSFPN innovatively constructs a six-level feature pyramid (P2-P6), significantly expanding the coverage of the multi-scale receptive field by incorporating the P2/P3 levels. In terms of technical implementation, the SBA module replaces the traditional upsample-concat structure, the deformable convolution kernel is used to adaptively capture the target contour, while the spatial attention mechanism suppresses background noise interference. ReSFPN input and output sizes and key parameters are (160,160,128), (80,80,256), (40,40,512), (20,20,512). The design of ReSFPN significantly enhances edge feature response strength, greatly improves multi-scale target detection accuracy, and effectively addresses the edge information degradation issue in feature pyramid networks under low-light conditions.

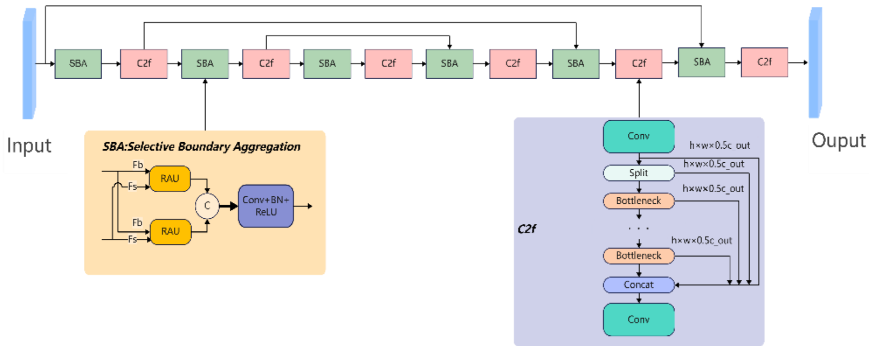

Figure 2 illustrates the diagram of the ReS feature pyramid network.

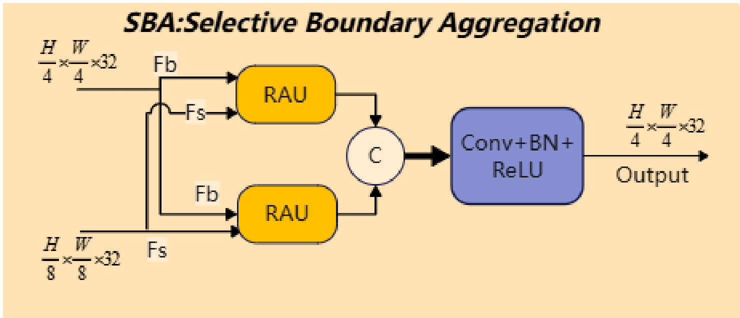

Figure 3 shows the framework of the SBA module.

3.3 No need for non-maximum suppression detector head

To address the NMS dependency issue caused by the one-to-many label assignment strategy in the traditional YOLO detection model, this paper proposes a dual-path collaborative optimization mechanism. Traditional methods enhance supervision signals by matching multiple prediction boxes to a single target, but this leads to a large number of redundant detection boxes during inference, necessitating reliance on non-maximum suppression (NMS) for post-processing. This introduces computational bottlenecks and hyperparameter sensitivity during end-to-end deployment. This method innovatively introduces a dual-branch detection head architecture. During the training phase, a joint supervision paradigm is adopted, simultaneously performing one-to-many label assignment (dense positive sample supervision) and one-to-one label assignment (precise positioning learning). In the inference stage, the one-to-one branch is activated through the task alignment strategy, directly outputting high-confidence detection results (as shown in Figure 4). This dual-label assignment strategy employs a dynamic feature decoupling mechanism to fully eliminate the NMS post-processing step, while preserving the benefits of strong supervision signals, thereby enabling true end-to-end low-light target detection.

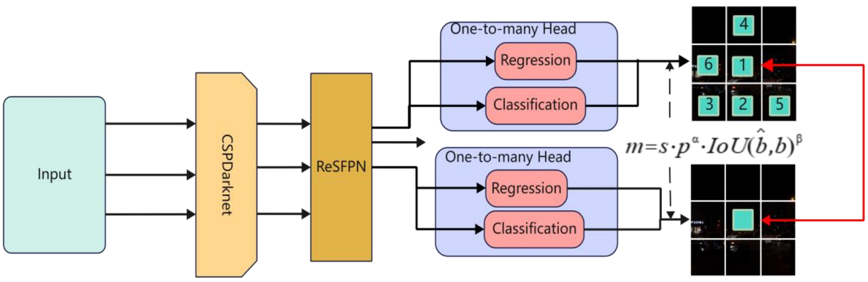

Figure 4 Structure of the Detection Head without Non-Maximum Suppression

To address the supervision mismatch issue between the one-to-many (o2m) and one-to-one (o2o) branches in the dual-label assignment strategy, this paper proposes a unified matching metric framework based on dynamic balancing. By constructing a composite metric function,

$$m(\alpha,\beta) = s \cdot p_{\alpha} \cdot IoU(\hat{b}, b)^{\beta} \tag{3}$$

Classification-localization collaborative optimization is achieved, where p is the classification score, $\hat{b}$ and b denote the bounding box of prediction and instance, the spatial prior $s \in \{0,1\}$ represents the spatial association of the anchor point, and the hyperparameters α and β control the weights of classification confidence p and bounding box regression accuracy IoU, respectively. Given the maximum IoU $u*$ of an instance, the two-branch matching metrics are $m_{o2m} = m(\alpha_{o2m}, \beta_{o2m})$ and $m_{o2o} = m(\alpha_{o2o}, \beta_{o2o})$. The supervision gap can be quantified using the 1-Wasserstein distance of the classification target difference,

$$A = t_{o2o,i} - I(i \in \Omega) t_{o2m,i} + \sum_{k \in \Omega \setminus \{i\}} t_{o2m,k} \tag{4}$$

Where $t_{o2m,j} = u^* \cdot \frac{m_{o2m,j}}{m_{o2m}^*}$ and $t_{o2o,i} = u^*$. The analysis shows that when $t_{o2m,i} \to u^*$ (i.e., when i becomes the best sample in the positive sample set Ω), the gap A is minimized. To address this, parameter ratio constraints $\alpha_{o2o} = r \cdot \alpha_{o2m}$ and $\beta_{o2o} = r \cdot \beta_{o2m}$ with $m_{o2o} = m_{o2m} \cdot r$, are established. This mechanism ensures that the optimal sample from the o2m branch also achieves the highest matching degree in the o2o branch, enabling alignment optimization of the dual-branch target space and effectively mitigating the semantic misalignment issue caused by independent measurements in traditional methods. Where the hyperparameter r serves as a scaling factor that ensures consistency between the one-to-one and one-to-many detection heads.

4 Experiment

4.1 Dataset Processing and Evaluation Metrics

This study conducts a comprehensive experimental evaluation using the DarkFace dataset [23], which consists of 6,000 low-light images collected from real-night scenes. The dataset covers a variety of complex lighting scenarios, including urban roads, building complexes, and transportation hubs, with all images providing detailed annotated face bounding box ground truth. Following standard machine learning practices, the dataset is split into a training set (4,800 images), a validation set (600 images), and a test set (600 images) in an 8:1:1 ratio. These sets are used for model parameter optimization, hyperparameter tuning, and evaluation of generalization ability, ensuring the statistical significance and reproducibility of the experimental results.

According to standard evaluation metrics, this paper uses precision (P), recall (R), and mean average precision (mAP) to assess the performance of the object detection model.

Precision refers to the proportion of samples predicted as positive by the model that are actually positive. In target detection, it refers to the proportion of all detected target frames that are correctly identified. Here, TP represents the number of true positive samples, FP represents the number of false positive samples, and FN represents the number of false negative samples. The calculation formula is as follows,

$$P = \frac{TP}{TP+FP} \tag{5}$$

Recall refers to the proportion of actual positive samples that are correctly identified as positive by the model. In target detection, recall refers to the proportion of all actual targets that are successfully detected by the model. The calculation formula is as follows,

$$R = \frac{TP}{TP+FN} \tag{6}$$

mAP is a commonly used metric for evaluating multi-class classification problems. It is calculated by determining the average precision (AP) for each category and then averaging the results. The value of mAP ranges from 0 to 1, with a value closer to 1 indicating better model performance. The calculation formula is as follows,

$$mAP = \frac{\sum_{i=1}^{N} \int_{0}^{1} P(R)dR}{N} \tag{7}$$

4.2 Implementation Details

This study developed a complete training process using the PyTorch framework [24] and employed the stochastic gradient descent (SGD) optimizer to optimize the model parameters. The experimental setup follows a rigorous deep learning training paradigm, input images are uniformly resized to a 640×640 resolution, and a random initialization strategy is employed to prevent domain adaptation bias from pre-trained weights. The optimizer configuration includes a momentum coefficient and an L2 regularization term to ensure stable parameter updates. The training process employs a cosine annealing learning rate scheduling strategy, completing 300 training epochs of iterative optimization on a standard GPU computing cluster, ultimately achieving stable convergence of the model parameters. The key hyperparameter configurations are provided in Table 1.

Table 1 lists the processors used in the experiment and the parameters employed during training.

Catalogory	Item	Params
Hardware	GPU	NVIDIA A100 80GB PCIe
	CPU	-
Training	Optimizer	SGD
	Learning rate	0.01
	Weight decay	0.0005
	Momentun cofficient	0.937
	Batch_size	16

4.3 Ablation Experiment

This study evaluates the effectiveness of the LADF-YOLO model's innovations through controlled ablation experiments. Building on the YOLOv8 benchmark model, a progressive module stacking strategy is employed. First, the ReS Feature Pyramid Network (ReSFPN) is integrated to enhance multi-scale feature interaction, followed by the design of the NMS-Free detection head to eliminate the reliance on post-processing. The experiment established a multi-dimensional evaluation system based on four aspects: detection accuracy (mAP@0.5/mAP@0.5:0.95), recall rate (Recall), computational complexity (GFLOPs), and real-time performance (FPS). Visual analysis was also conducted through loss curves and indicator change trends.

Table 2 Ablation Experiment Data Summary. The symbol "✓" indicates that the corresponding improved module has been integrated into the model.

YOLOv8	ReSFPN	NMS-Free	mAP@0.5(%)	P(%)	R(%)	mAP@0.5:0.95(%)	GFLOPS	FPS
✓			47.8	71.2	42.4	20.5	8.1	116.3
✓	✓		57.9	75.8	50.8	24.8	14.9	87.1
✓		✓	49.0	69.8	43.7	21.5	8.1	107.5
✓	✓	✓	58.6	72.2	52.3	24.7	14.9	84.4

The results of the ablation experiment (Table 2) demonstrate that the proposed ReS feature pyramid network (ReSFPN) and NMS-Free detection head significantly enhance the low-light detection performance. Compared to the baseline YOLOv8 model (mAP@0.5 = 47.8%), the introduction of the ReSFPN module alone boosts the detection accuracy by 10.1 percentage points, reaching 57.9%, by enhancing the multi-scale feature interaction capability. The NMS-Free detection head provides a 1.2% accuracy improvement by eliminating post-processing redundancy. After the collaborative optimization of both modules, the model's mAP@0.5 on the DarkFace dataset reached 58.6%, resulting in an overall performance improvement of 10.8%. The complementary advantages of ReSFPN in optimizing dark area feature representation and NMS-Free in improving inference efficiency were demonstrated, effectively addressing the dual challenges of feature degradation and post-processing bottlenecks in low-light scenes.

The evaluation index curve (e.g., Fig. 5) shows that the four core metrics—mAP@0.5, mAP@0.5:0.95, precision, and recall—improve simultaneously as the number of training rounds increases. Notably, LADF-YOLO improves the mAP@0.5 metric by over 10 percentage points compared to the original model, consistently outperforming other variants (such as models with only the NMS-Free or ReS feature pyramid modules added). This cross-indicator consistency optimization demonstrates that the ReS feature pyramid network enhances the semantic representation of dark-area targets through multi-level feature interaction, while the NMS-Free detection head reduces redundant detection boxes via an end-to-end optimization mechanism. The synergistic effect of both components effectively addresses the dual challenges of feature degradation and positioning ambiguity in low-light scenes, validating the collaborative innovation of LADF-YOLO in both feature extraction and reasoning efficiency.

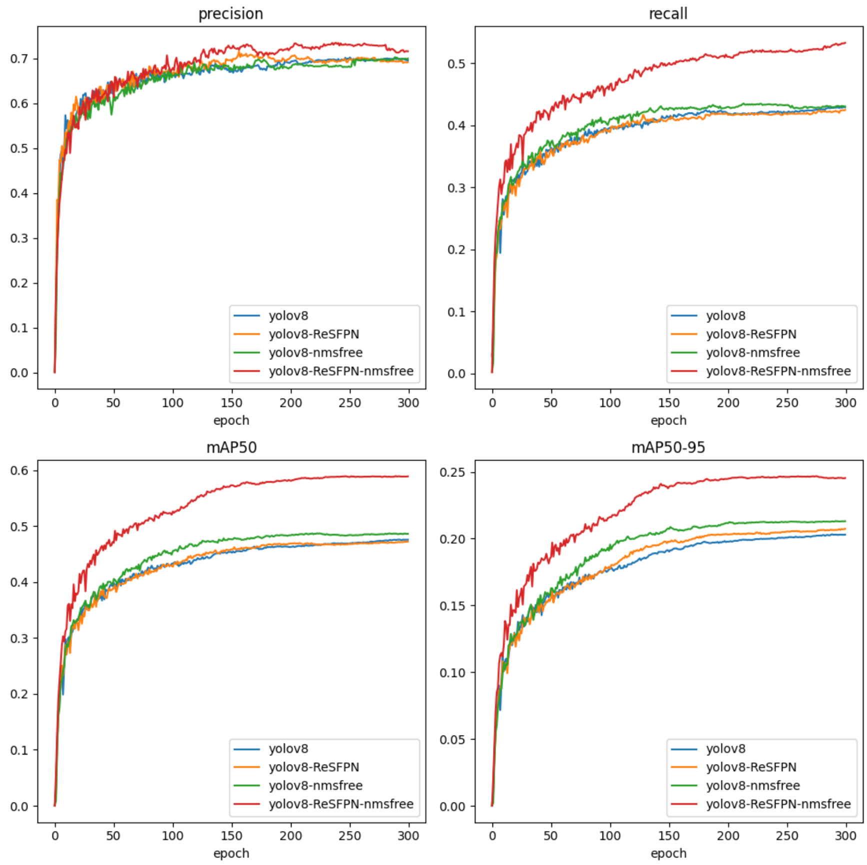

Figure 5 metrics_curve information diagram

4.4 Comparative Experiments on the DarkFace Dataset

This study conducted a systematic comparative experiment on mainstream YOLO series models using the DarkFace dataset (as shown in Table 3), including classic models such as YOLOv5, YOLOv8, and YOLOv10. All models are optimized end-to-end using a unified training strategy, with precision, recall, mAP@0.5, mAP@0.5:0.95, computational complexity (GFLOPs), and real-time performance (FPS) serving as multi-dimensional evaluation benchmarks. Bold indicates the best result for each metric among the low-light object detection algorithms.

This study verified the technological breakthrough of LADF-YOLO in low-light object detection through comparative experiments. Compared to mainstream YOLO series models, this algorithm leverages the cross-level feature interaction mechanism of the ReS feature pyramid network and the end-to-end optimization paradigm of the NMS-Free detection head, a significant leap in detection accuracy was achieved, with a breakthrough improvement of 10.8% in mAP@0.5 and 9.9% in Recall, while the mAP@0.5:0.95 indicator also demonstrated steady growth. Although the model's computational complexity has increased, its innovative design in feature representation effectively addresses core challenges such as blurred target edges and missing semantic information in low-light environments. This verifies the necessity of an accuracy-first algorithm optimization approach in extreme lighting conditions and offers a highly robust detection solution for real-world industrial deployment.

Table 3 Comparison of detection performance between LADF-YOLO and classic YOLO methods on the DarkFace dataset. Bold indicates the best result for each metric among the low-light object detection algorithms.

Method	Backbone	mAP@0.5(%)	mAP@0.5:0.95(%)	P(%)	R(%)	GFlOPS	FPS
YOLOv8	CSPDarknet	47.8	20.5	71.2	42.4	8.1	116.3
YOLOv5 [25]	CSPNet	46.2	19.5	69.5	41.2	7.8	113.8
YOLOv11 [26]	-	42.6	18.2	68.5	38.5	**6.5**	**123.1**
YOLOX [27]	CSPDarknet	56.5	-	-	-	-	-
YOLOv10	CSPNet	41.7	17.6	65.2	37.8	8.6	101.7
YOLOv9m [28]	-	54.5	25.0	**74.0**	47.6	77.0	85.0
LADF-YOLO(Ours)	CSPDarknet	**58.6**	**24.7**	72.2	**52.3**	14.9	84.4

This study systematically compares the performance of LADF-YOLO with mainstream detection models on the DarkFace low-light dataset, as shown in Table 4. Experimental results show that traditional detection methods, such as FCOS [29], YOLOX, and CenterNet [30], generally exhibit low detection accuracy due to uneven illumination and feature degradation in low-light scenes. For instance, the FCOS model achieves only 30.3% mAP@0.5. In contrast, the proposed LADF-YOLO enhances multi-scale semantic fusion through the ReS feature pyramid network and achieves end-to-end optimization with the NMS-Free detection head. As a result, it surpasses all compared models, achieving a mAP@0.5 of 58.6%—a 10.8 percentage point improvement over the YOLOv8 baseline (47.8%). This result validates the synergistic enhancement of the cross-level feature interaction mechanism and post-processing optimization strategy under extreme lighting conditions, offering a novel technical paradigm for low-light object detection.

The visual comparison of detection results (Figure 6) demonstrates that LADF-YOLO exhibits significant advantages in low-light environments, Compared with classic detection methods like FCOS and YOLOX, it is less prone to false detections (e.g., misclassifying light and shadow noise as targets) and missed detections (e.g., failing to recognize certain visible targets) caused by interference from dark backgrounds. Compared to the latest YOLOv12 algorithm (mAP@0.5 = 40.6%), LADF-YOLO achieves a significant improvement, boosting mAP@0.5 to 58.6%. This method precisely delineates target boundaries in low-contrast areas by leveraging the multi-scale semantic fusion of the ReS feature pyramid network and the task alignment strategy of the NMS-Free detection head. Experimental cases demonstrate that in extremely dark areas (illuminance <1 lux) where target visibility falls below 15%, LADF-YOLO can still reconstruct local contour features and capture global context. Compared to the baseline model, it significantly reduces bounding box positioning errors, confirming its strong and robust capability in parsing partially visible targets in low-light environments.

Table 4 Comparison of Detection Results Between This Method and Other Mainstream Object Detection Methods on the DarkFace Dataset.

Method	Backbone	Size	mAP@0.5(%)
YOLOv8	CSPDarknet	640×640	47.8
YOLOX	CSPDarknet	640×640	56.5
YOLOv5	CSPNet	640×640	46.2
CenterNet	ResNet	640×640	49.1
RetinaNet [31]	ResNet	640×640	34.4
Faster R_CNN	ResNet	640×640	38.6
Fcos	ResNet	640×640	30.3
YOLOv12 [32]	-	640×640	40.6
Dynamic R-CNN [33]	ResNet	640×640	38.0
LADF-YOLO(Ours)	**CSPDarknet**	**640×640**	**58.6**

Figure 6 Visualization of detection results for LADF-YOLO and classic object detection methods on the DarkFace dataset.

5 Conclusion

It aims to address the issues of uneven illumination, blurred boundaries, difficulty in feature extraction in low-light target detection. This paper proposes an innovative solution, LADF-YOLO, which constructs a bidirectional cross-level feature interaction ReS feature pyramid network (ReSFPN). By employing deformable convolution and a spatial-channel dual attention mechanism, it enables multi-scale semantic enhancement and boundary feature reconstruction of targets in dark areas. Additionally, an NMS-Free detection head is designed to achieve collaborative optimization of classification and positioning tasks through a decoupled dual-label allocation strategy. This approach completely eliminates the reliance on non-maximum suppression (NMS) post-processing, improves positioning accuracy, and enhances the model's performance in low-light image detection. Compared to the base model YOLOv8, the LADF-YOLO algorithm achieved an average detection accuracy of 58.6% on the low-light target detection DarkFace dataset, marking an increase of 10.8%. Compared to mainstream general target detection algorithms on the low-light target detection DarkFace dataset, the LADF-YOLO model demonstrates higher detection accuracy. However, its detection speed needs improvement. The next step will focus on developing a lightweight network architecture and exploring data transfer between modules to reduce parameter calculations and ultimately enhance detection speed.

Acknowledgments. This work was supported by the Guangxi Key Research and Development Program (No. Guike AB22080047), State-level Innovation and Entrepreneurship Training Program for College Students (No. 202410595012).

Disclosure of Interests. The authors declare that there is no conflict of interest regarding the publication of this manuscript. All authors have contributed to the research and preparation of the manuscript, and there are no financial, professional, or personal relationships that could be perceived as influencing the research outcomes.

References

1. Abba, S., Bizi, A.M., Lee, J.-A., Bakouri, S., Crespo, M.L.: Real-time object detection, tracking, and monitoring framework for security surveillance systems. Heliyon 10(15), e34922 (2024). https://doi.org/10.1016/j.heliyon.2024.e34922.
2. Yue, S., Zhang, Z., Shi, Y., Cai, Y.: WGS-YOLO: A real-time object detector based on YOLO framework for autonomous driving. Comput. Vis. Image Underst. 249, 104200 (2024). https://doi.org/10.1016/j.cviu.2024.104200.
3. Badgujar, C.M., Poulose, A., Gan, H.: Agricultural object detection with You Only Look Once (YOLO) Algorithm: A bibliometric and systematic literature review. Comput. Electron. Agric. 223, 109090 (2024). https://doi.org/10.1016/j.compag.2024.109090.
4. Stark, J.A.: Adaptive image contrast enhancement using generalizations of histogram equalization. IEEE Trans. Image Process. 9(5), 889–896 (2000). https://doi.org/10.1109/83.841534.

5. Agrawal, S., Panda, R., Mishro, P.K., Abraham, A.: A novel joint histogram equalization based image contrast enhancement. J. King Saud Univ. Comput. Inf. Sci. 34(4), 1172–1182 (2022). https://doi.org/10.1016/j.jksuci.2019.05.010.
6. Xu, J., Hou, Y., Ren, D., Liu, L., Zhu, F., Yu, M., Wang, H., Shao, L.: STAR: A structure and texture aware Retinex model. IEEE Trans. Image Process. 29, 5022–5037 (2020). https://doi.org/10.1109/TIP.2020.2974060.
7. J. Li, X. Feng and Z. Hua, "Low-Light Image Enhancement via Progressive-Recursive Network," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 11, pp. 4227-4240, Nov. 2021, doi: 10.1109/TCSVT.2021.3049940.
8. Huang, S.-C., Cheng, F.-C., Chiu, Y.-S.: Efficient contrast enhancement using adaptive gamma correction with weighting distribution. IEEE Trans. Image Process. 22(3), 1032–1041 (2013). https://doi.org/10.1109/TIP.2012.2226047.
9. Cui, Z., Li, K., Gu, L., Su, S., Gao, P., Jiang, Z., Qiao, Y., Harada, T.: You only need 90K parameters to adapt light: A light weight transformer for image enhancement and exposure correction. In: Proc. British Machine Vision Conference (BMVC 2022), pp. 1–13. BMVA Press, London, UK (2022). https://doi.org/10.48550/arXiv.2205.14871.
10. Vinoth, K., Sasikumar, P.: Lightweight object detection in low light: Pixel-wise depth refinement and TensorRT optimization. Results Eng. 23, 102510 (2024). https://doi.org/10.1016/j.rineng.2024.102510
11. Wang, C.-Y., Liao, H.-Y.M.: YOLOv9: Learning what you want to learn using programmable gradient information. arXiv preprint arXiv:2402.13616 (2024). https://doi.org/10.48550/arXiv.2402.13616.
12. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C.: SSD: Single shot MultiBox detector. In: Computer Vision – ECCV 2016 (LNCS, vol. 9905), pp. 21–37. Springer (2016). https://doi.org/10.1007/978-3-319-46448-0_2.
13. Srivastava, S., Divekar, A.V., Anilkumar, C., et al.: Comparative analysis of deep learning image detection algorithms. J. Big Data 8(1), 66 (2021). https://doi.org/10.1186/s40537-021-00446-6.
14. Zhang, Z., Lu, X., Cao, G., Yang, Y., Jiao, L., Liu, F.: ViT-YOLO: Transformer-based YOLO for object detection. In: Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW 2021), pp. 2799–2808. IEEE (2021). https://doi.org/10.1109/ICCVW54120.2021.00314.
15. Wang, A.: YOLOv10: Real-time end-to-end object detection. arXiv preprint arXiv:2405.14458 (2024). https://doi.org/10.48550/arXiv.2405.14458.
16. Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics YOLOv8. https://github.com/ultralytics/ultralytics (2023).
17. Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. arXiv preprint arXiv:1612.03144 (2016). https://doi.org/10.48550/arXiv.1612.03144.
18. 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. 39(6), 1137–1149 (2017). https://doi.org/10.1109/TPAMI.2016.2577031.
19. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: Proc. IEEE Int. Conf. Comput. Vis. (ICCV 2017), pp. 2980–2988. IEEE, Venice, Italy (2017). https://doi.org/10.1109/ICCV.2017.322.
20. Zhao, H., Wang, J., Dai, D., Lin, S., Chen, Z.: D-NMS: A dynamic NMS network for general object detection. Neurocomputing 512, 225–234 (2022). https://doi.org/10.1016/j.neucom.2022.09.080.

21. Kumar, A., Brazil, G., Liu, X.: GrooMeD-NMS: Grouped mathematically differentiable NMS for monocular 3D object detection. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR 2021), pp. 8973–8983. IEEE (2021).
22. Tang, F., Huang, Q., Wang, J., Hou, X., Su, J., Liu, J.: DuAT: Dual-aggregation transformer network for medical image segmentation. arXiv preprint arXiv:2212.11677 (2022). https://doi.org/10.48550/arXiv.2212.11677.
23. Yang, W., Yuan, Y., Ren, W., Liu, J., Scheirer, W.J., Wang, Z., et al.: Advancing image understanding in poor visibility environments: A collective benchmark study. IEEE Trans. Image Process. 29, 5737–5752 (2020). https://doi.org/10.1109/TIP.2020.2981903.
24. Paszke, A., Gross, S., Massa, F., et al.: PyTorch: An imperative style, high-performance deep learning library. In: Proc. 33rd Int. Conf. Neural Inf. Process. Syst. (NeurIPS 2019), p. 721. Curran Associates, Vancouver, BC, Canada (2019).
25. Jocher, G.: Ultralytics YOLOv5. https://github.com/ultralytics/yolov5 (2020).
26. Jocher, G., Qiu, J., Chaurasia, A.: Ultralytics YOLOv11. https://github.com/ultralytics/ultralytics (2024).
27. Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J.: YOLOX: Exceeding YOLO series in 2021. arXiv preprint arXiv:2107.08430 (2021). https://doi.org/10.48550/arXiv.2107.08430.
28. Wang, C.-Y., Liao, H.-Y.M.: YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. arXiv:2402.13616 [cs.CV] (2024).
29. Tian, Z., Shen, C., Chen, H., He, T.: FCOS: Fully convolutional one-stage object detection. arXiv preprint arXiv:1904.01355 (2019). https://doi.org/10.48550/arXiv.1904.01355.
30. Zhou, X., Wang, D., Krähenbühl, P.: Objects as points. arXiv preprint arXiv:1904.07850 (2019). https://doi.org/10.48550/arXiv.1904.07850.
31. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. arXiv preprint arXiv:1708.02002 (2017). https://doi.org/10.48550/arXiv.1708.02002.
32. Tian, Y., Ye, Q., Doermann, D.: YOLOv12: Attention-Centric Real-Time Object Detectors. arXiv:2502.12524 [cs.CV] (2025).
33. Zhang, H., Chang, H., Ma, B., et al.: Dynamic R-CNN: Towards High Quality Object Detection via Dynamic Training. In: Eur. Conf. Comput. Vis. (ECCV) (2020).