OEE | 光电工程 Opto-Electronic Engineering

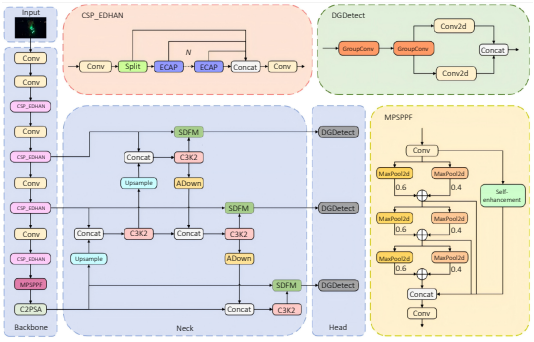
Article

2026 年 3 月, 第 53 卷, 第 3 期

融合多尺度特征与局部细节强化的低照度目标检测

林珊玲[1,2], 刘　梅[1,2], 林坚普[1,2*], 林志贤[1,2,3], 郭太良[2,3], 孙　磊[4]

[1]福州大学先进制造学院，福建 泉州 362000；
[2]中国福建光电信息科学与技术实验室，福建 福州 350116；
[3]福州大学物理与信息工程学院，福建 福州 350116；
[4]福州大学至诚学院，福建 福州 350116



摘要：暗夜环境中曝光不足导致所得图像受噪声影响严重，同时存在图像细节、纹理、轮廓信息丢失等问题，使得后续目标检测算法对低照度图像的检测效果不理想。为此，本文基于 YOLOv11n 模型，提出一种融合多尺度特征与局部细节强化的低照度目标检测算法 LMD-YOLO。首先，提出 MPSPPF 模块，借助多级混合并行池化与特征强化机制，有效抑制背景与随机噪声的干扰，同时缓解低光照场景下图像对噪声的高敏感度，进一步抑制细节信息的丢失。其次，设计 CSP-EDHAN 模块实现多粒度特征融合以及高效特征提取，以减少低照度图像中纹理、边缘和轮廓等低级特征的丢失。接着，本文提出动态细节语义融合金字塔网络，在双向金字塔路径中实现多尺度特征的高效对齐与增强，提升低照度场景下的目标检测精度。最后，设计双组检测头，利用分组卷积与共享特征提取干实现轻量化，解耦分支设计提升检测精度。实验结果表明：LMD-YOLO 在精确度、召回率、mAP50 和 mAP50-95 等关键指标上均优于原始 YOLOv11n，在 Exdark 数据集上分别提升了 1.3%、2.3%、3.1% 和 2.2%；此外，与当前主流算法及针对低照度目标检测的改进模型相比，LMD-YOLO 在 mAP50 和 mAP50-95 指标上分别实现了 3% 和 3.4% 的平均提升，证明了其在低照度图像检测任务中的优越性能。

关键词：低照度目标检测；YOLOv11；多尺度特征融合；细节增强

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250241 | CSTR: 32245.14.oee.2026.250241 | 中图分类号：TP391.4 | 文献标识码：A

林珊玲，刘梅，林坚普等. 融合多尺度特征与局部细节强化的低照度目标检测 [J]. 光电工程，2026，**53**(3): 250241

Lin S L, Liu M, Lin J P et al. Low-light object detection with multi-scale feature fusion and local detail enhancement[J]. *Opto-Electron Eng*, 2026, **53**(3): 250241

1　引　言

目标检测是多项视觉任务的核心基础，已被广泛应用于如智能分拣[1]、工业检测[2]、表面缺陷检测[3]、舰船检测[4] 以及火灾检测[5] 等各类实际场景。目标检测旨在对输入图像中的所有实例进行定位，通过生成边界框标定目标，并判定其类别。近年来，目标检测技术发展迅速，但在低照度环境下采集的图像常存在亮度不足、颜色失真、噪声较多以及细节纹理丢失等问题，给后续任务的开展带来较大挑战[6]。

收稿日期：2025-08-16；修回日期：2025-11-03；录用日期：2025-11-06；网络出版日期：2026-03-19
基金项目：国家自然科学基金青年基金 (62101132)
*通信作者：林坚普，ljp@fzu.edu.cn
版权所有©2026 中国科学院光电技术研究所

随着深度学习的快速演进，基于深度神经网络的目标检测已成为计算机视觉领域的研究前沿。现有方法大致可以分为两类：其一是以 RCNN[7]、Fast RCNN[8] 和 Faster RCNN[9] 为代表的两阶段检测框架；其二是以 YOLO[10]、SSD[11] 为代表的单阶段检测框架。然而，这些主流算法大多在高质量图像的设定下开发，对图像质量下降的影响考虑不足，因而在低光照场景中检测性能显著衰减。近年来，提升低照度图像目标检测性能成为了研究的热点，针对低照度目标检测任务，现有的研究方法主要分为以下两类：

一类是采用如 MBLLEN[12]、RetinexNet[13] 和 ZeroDCE[14] 等图像增强算法作为检测前的预处理环节，从而改善低照度图像的质量，降低光照不足对检测效果的影响。例如，Yuan 等[15] 提出了一种面向目标检测的即插即用低光图像增强方法，在低照度场景下能一定程度提升检测性能，但其增强方式较为简单，未考虑复杂环境光照，因此提升效果有限；Lin 等[16] 提出了一种频率引导的双折叠 Transformer 网络，改善了图像质量，并将该网络应用于暗检测的预处理步骤，但该方法侧重低光图像增强，未与检测网络联合优化，对目标检测性能提升有限。这些方法的共同点是将图像增强作为预处理环节，优点是能在一定程度上改善暗光图像的可见性，但缺点是增强算法往往难以区分真实信号与噪声，容易将背景噪声一同放大，导致增强结果并不一定能有效提升后续的目标检测性能。

另一类是从网络架构层面进行改进，使模型能够在暗光图像中挖掘更丰富的特征信息，从而提升检测性能。但是在低光照环境中，图像信噪比低，目标与背景的区分度显著降低，导致目标的关键纹理、边缘等细节信息严重模糊甚至丢失。这使得传统的特征提取网络难以捕捉到足够的判别性特征，即使是引入多尺度特征提取、注意力机制等改进方法，如史丽晨等[17] 通过多尺度空间特征提取同时结合坐标注意力机制、江泽涛等[18] 在常规照度目标检测框架中加入注意力机制与多尺度特征感知、蒋畅江等[19] 提出一种融合多注意力机制的低照度目标检测模型，也难以彻底解决这一根本性问题，检测精度的提升受阻。此外，现有改进模型往往计算开销较大，例如 DBS-YOLOv8[20]、MPE-DETR[21]、PE-YOLO[22]、Dark-YOLO[23]、MLFE-YOLOX[24] 等方法虽然在一定程度上提升了检测性能，但往往增加了模型深度、引入了复杂模块或多级联结构。这导致模型参数量和计算量显著增加，使得算法难以部署到算力受限的边缘设备或实时性要求高的应用场景中。

总体来看，虽然低照度目标检测已取得一定进展，但仍存在以下挑战：1) 现有方法多通过图像增强来提升低照度图像质量，容易将噪声一同放大，导致检测模型出现误检、漏检；2) 对图像细节信息识别和整合不足，纹理、边缘和轮廓等低级特征丢失严重；3) 尽管部分方法在提升检测精度方面有所突破，但计算量和参数量较高，导致算法在实际部署时面临过高的计算资源消耗。基于上述问题，本文在 YOLOv11n 的基础上，提出了一种低照度目标检测算法 LMD-YOLO，在未大量增加模型复杂度的基础上，强化模型对低照度图像的多尺度特征融合以及局部细节特征提取能力，同时缓解低光照场景下图像对噪声的高敏感度，从而提高低照度图像检测的准确性。

2 融合多尺度特征与局部细节强化的低照度目标检测算法

2.1 网络整体结构

YOLOv11[25] 是 Ultralytics 于 2024 年 9 月 30 日发布的，在检测精度和推理速度方面相比以往 YOLO 系列均有显著提升。YOLOv11 的模型架构由骨干网络、颈部架构和头部网络 3 部分组成，共同实现了高效的目标检测。相比较于 YOLOv8[26] 模型，YOLOv11 将 CF2 模块改成 C3K2，同时在骨干网络最后添加了一个 C2PSA 模块，且将 YOLOv10[27] 的 head 思想引入到 YOLOv11 的 head 中，使用深度可分离的方法，减少冗余计算，提高效率。然而，尽管 YOLOv11 在常规照明条件下的目标检测任务中表现出色，但在低照度图像的检测中仍面临图像模糊、背景噪声与检测目标难以区分等问题，导致检测困难。因此，本文选择 YOLOv11 作为基准模型，旨在提升模型在低照度图像检测中的性能，同时确保其在常规照明条件下依然保持较高的检测精度。此外，为满足低照度图像检测在计算量、实时响应和轻量化部署方面的需求，本文基于 YOLOv11 的最小体量版本 YOLOv11n 进行了改进。首先设计了多池空间金字塔快速池化模块 MPSPPF (multi-pool spatial pyramid fast pooling module)，通过多级混合池化以及特征增强抑制细节信息的丢失。其次提出了双混合注意力跨阶段增强模块 CSP-EDHAN (cross stage partial - enhanced dual hybrid attention network)，这是一种结合多路径特征融

合、深度可分离卷积和混合注意力机制的结构，旨在在低光照条件下高效提取并增强微弱目标的细粒度特征，同时抑制噪声、提升检测精度。然后设计了动态细节语义融合金字塔网络 DDS-FPN (dynamic detail-semantic fusion pyramid network)，通过表层细节融合模块 SDFM (superficial detail fusion module) 的通道-空间注意力机制，将低照度图像中的噪声抑制与关键细节增强相结合，动态提升目标区域的特征响应；同时，通过自适应下采样模块 ADown (adaptive downsampling)，将深层语义信息与浅层细节跨层级融合，补偿暗光导致的细节丢失，在双向金字塔路径中实现多尺度特征的高效对齐与增强，提升了低照度场景下的目标检测精度与鲁棒性。最后提出了双组检测头 DGDetect (DualGroupDetect)，通过引入分组卷积提取更丰富的特征表示，并采用共享特征提取干结合 1×1 卷积的方式减少重复计算与卷积层数量，实现了轻量化；同时，解耦分支设计有效避免了任务间干扰，在保持轻量化的同时显著提升检测精度。LMD-YOLO 模型结构图如图 1 所示。

2.2 MPSPPF 模块

YOLOv11 延用了快速空间金字塔池化模块[28] SPPF (spatial pyramid pooling fast)，但由于 SPPF 模块只使用单一的最大池化操作，而最大池化操作会丢弃一些非最大值的特征信息[29]，这使得 SPPF 模块对特征提取不够全面，从而导致细节信息丢失。为了解决低照度图像噪声敏感和细节丢失的问题，本文提出了多池空间金字塔快速池化模块 MPSPPF。其结构图如图 2 所示。MPSPPF 模块通过引入并行池化分支与自增强特征路径，显著提升了低照度图像的特征提取能力。具体而言，并行池化结合了最大池化与平均池化：前者用于强化低对比度下的显著目标特征，后者则提供平滑表示以抑制噪声。两者加权融合后，能够兼顾全局结构与丰富的细节。同时，该模块通过多次顺序池化构建更深的金字塔结构，以捕获多尺度的局部与全局上下文信息，有效应对低照度场景下目标模糊与遮挡的难题。在特征拼接方面，不同于传统 SPPF 的直接拼接，MPSPPF 设计了自增强特征路径。它将初始特征与其通道维度的均值相加，把全局语义信息融合回原始特征。该机制实现了特征的自增强，有效抑制了低照度、低对比度环境下的随机噪声，保障了检测精度的稳定性。最后，该模块保留了首尾两层卷积层以降低计算量并整合通道信息。多尺度池化特

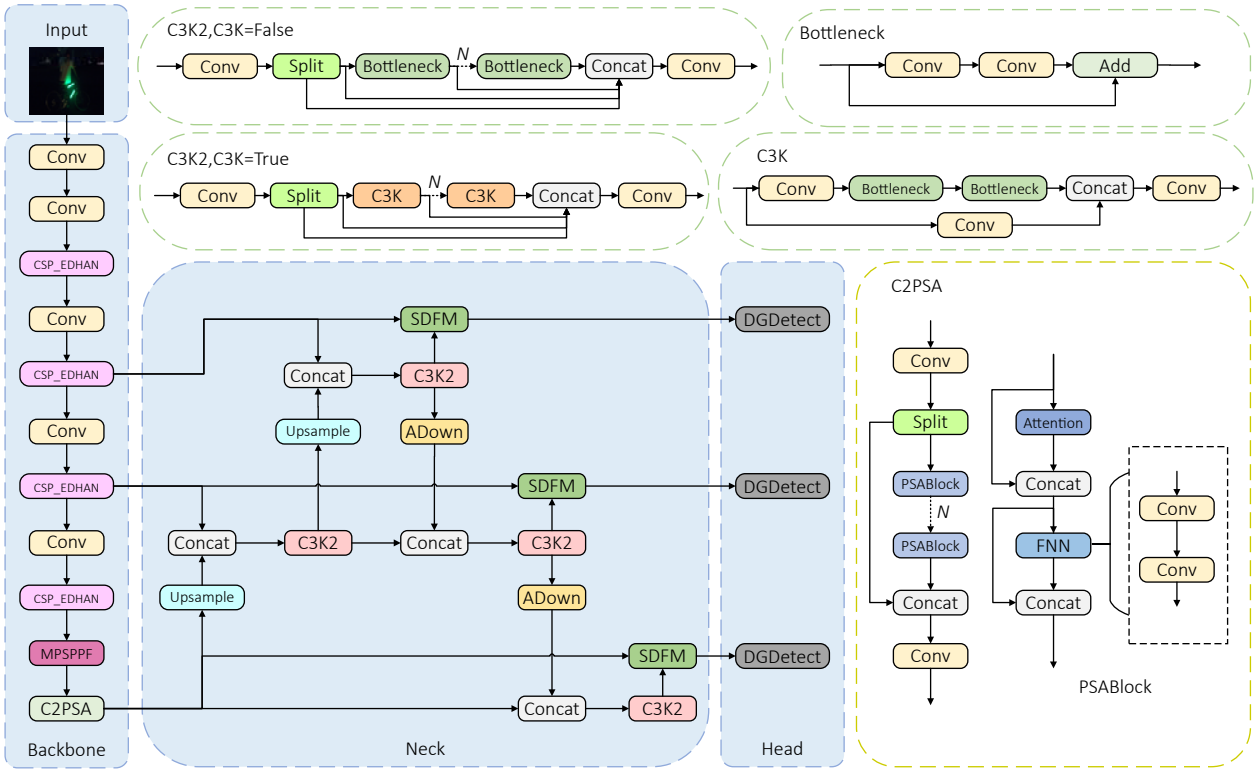


图 1 LMD-YOLO 模型结构图

Fig. 1 LMD-YOLO model architecture diagram

征与自增强特征拼接后，经由末端卷积进行深度融合，在保证计算效率的同时，全面提升了目标检测的准确性与鲁棒性。

2.3 CSP-EDHAN 模块

由于低光照图像存在噪声多、目标的边缘与纹理信息微弱等问题，现有模型难以有效放大与增强细节特征提取关键特征，同时噪声与模糊效应容易让网络混淆真实目标与背景。因此，为了增强模型捕获图像中的细节特征信息和区分真实目标与背景的能力，本文提出了双混合注意力跨阶段增强模块 CSP-EDHAN，如图 3 所示。该模块通过高效的深度可分离卷积提取特征，利用残差连接确保信息的稳定传递，并通过多路径特征融合增强低照度下微弱目标的信息表达，同时结合混合注意力机制精准聚焦关键区域，从而提升整体性能。

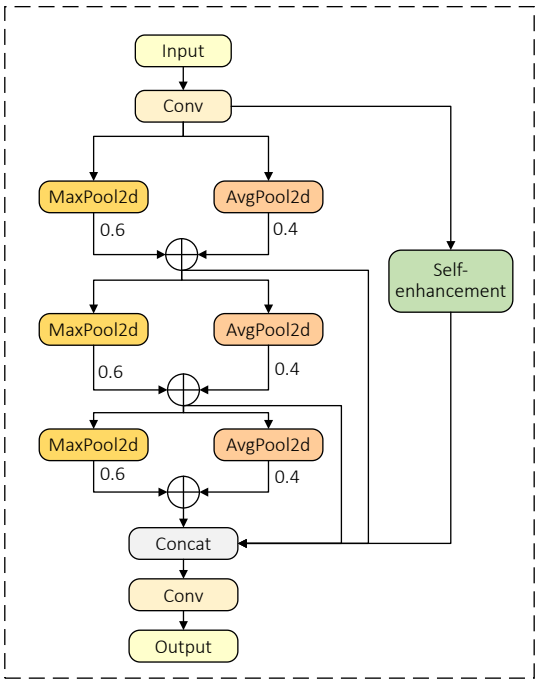


图 2 MPSPPF 模块结构图

Fig. 2 Structure of MPSPPF module

具体而言，该模块通过 1×1 卷积后将输入特征展开为两路：一路直接向前传递以保留原始信息；另一路则依次穿过 N 个 ECAP 模块逐步提取更丰富的语义特征。随后，将两路特征进行拼接并经由卷积层完成多路径特征的整合。这种设计既保留了初始细节，又融合了深层细粒度特征，显著增强了模型在低照度条件下区分目标与噪声的能力。作为核心组件，每个 ECAP 模块内部的具体特征处理流程为：首先，输入特征一方面直接跨层传递，另一方面依次穿过三个级联的 EDConv 层，以逐步提取多层次的深层空间信息；接着，将初始输入与这三个 EDConv 层的输出特征进行拼接，并送入由卷积、批量归一化和 SiLU 激活函数组成的复合层中，进行多尺度特征的降维融合与重校准；随后，校准后的特征进入混合注意力机制（HybridAttention）中，通过通道与空间的联合加权突出关键特征与显著区域；最后，特征经过 Dropout2d 空间正则化处理，并通过残差连接与模块的初始输入相加，在保留基础语义信息的同时，有效缓解了深层网络的梯度消失问题。为了在限制计算量

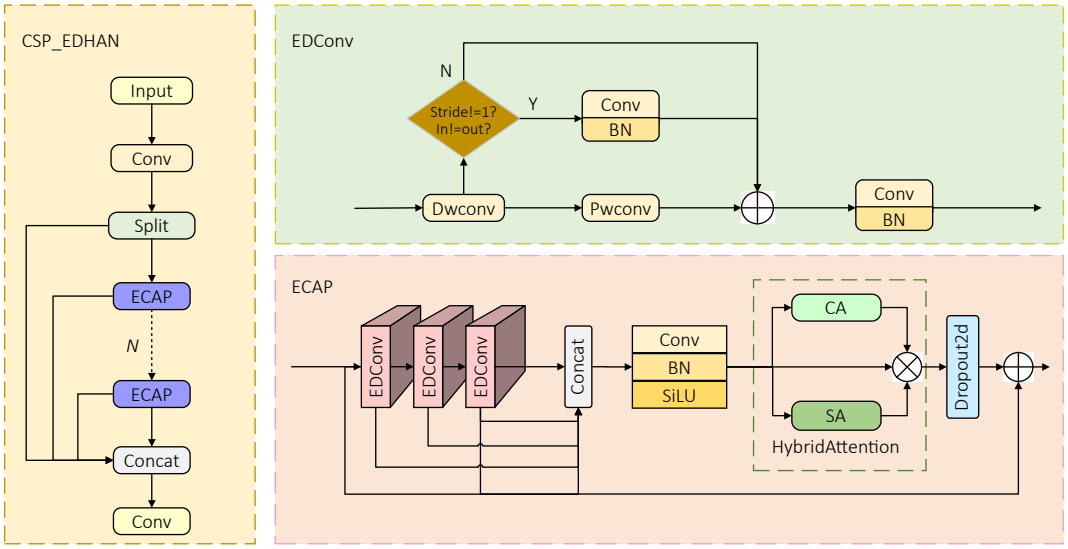


图 3 CSP-EDHAN 模块结构图

Fig. 3 Structure of CSP-EDHAN module

的同时高效捕捉局部细节，上述 ECAP 内部的 EDConv 模块通过深度卷积提取空间特征、逐点卷积融合通道信息，并结合条件残差连接及后续的卷积归一化增强特征表达。此外，ECAP 中的混合注意力机制由通道注意力与空间注意力并行构成。两者接收相同输入，前者通过分析全局响应突出高贡献通道并抑制噪声，后者则聚焦于目标边缘等显著空间区域。两者计算出的权重通过逐元素相乘融合后作用于原始特征，实现双重特征强化。这种机制使网络能够同时在通道选择与空间定位上实现精准聚焦，在低光照及噪声干扰的复杂场景下大幅提升了检测性能。

2.4 动态细节语义融合金字塔网络

原网络颈部采用经典的 FPN+PAN 结构，通过自顶向下的上采样与自底向上的下采样构建 P3、P4、P5 多尺度特征金字塔，并使用 C3K2 模块整合跨层信息。然而，该结构在低照度场景下存在明显不足：一是浅层 P3 仅有 256 个通道，且上下采样过程缺乏抗混叠与特征保护机制，导致微弱的暗光纹理严重衰减并易被噪声淹没；二是跨层融合仅采用简单的特征拼接，缺乏自适应权重调节，难以动态增强暗弱目标特征并抑制背景噪声，从而导致低照度条件下漏检与误检频发。针对这个问题，本文设计了动态细节语义融合金字塔网络 DDS-FPN，如图 4 所示。

动态细节语义融合金字塔网络 DDS-FPN 在传统 FPN + PAN 的主链上引入三处关键改动： 1) 将最浅层 P3、P4 融合后的通道数统一扩充到 512，增强低照度下的高分辨率纹理表征，从而充分保留低照度场景中极度微弱的边缘信息； 2) 采用自适应下采样模块 ADown 替代普通步长卷积，在下采样前先进行低通平滑处理，再结合深度可分离卷积操作，这种方法能够有效抑制低照度环境下的噪声混叠，同时更好地保留图像的边缘信息；3) 在 P3/8、P4/16、P5/32 三个输出尺度分别插入 SDFM 模块，将浅层的高频细节与颈部的深层语义信息进行融合，并将融合后的特征送入检测头。整体设计在保持轻量高效的同时，显著强化了细节恢复能力和噪声鲁棒性，可以有效的提升小目标和弱对比目标的检测精度。其中 SDFM 模块是 PSFusion[30] 提出的，该模块通过全局平均池化、逐点卷积和注意力机制对两路特征进行加权融合。该设计不仅能够有效抑制背景噪声与光照干扰，还能实现低层细节的精准对齐和互补增强，为后续检测阶段提供高保真度的底层特征表示，从而显著提升模型在复杂场景中的目标检测鲁棒性与定位准确性。表层细节融合模块结构如图 5 所示。计算式如式 (1)~(6) 所示。

$$\hat{F}_{\mathrm{A}}^{i}=F_{\mathrm{A}}^{i}\oplus\left(F_{\mathrm{B}}^{i}\otimes\delta\left(PWConv^{n}\left(GAP\left(C\left(F_{\mathrm{A}}^{i},F_{\mathrm{B}}^{i}\right)\right)\right)\right)\right), \tag{1}$$

$$\hat{F}_{\mathrm{B}}^{i}=F_{\mathrm{B}}^{i}\oplus\left(F_{\mathrm{A}}^{i}\otimes\delta\left(PWConv^{n}\left(GAP\left(C\left(F_{\mathrm{A}}^{i},F_{\mathrm{B}}^{i}\right)\right)\right)\right)\right), \tag{2}$$

$$A_{\mathrm{C}}^{i}=PWConv^{n}\left(GAP\left(PWConv\left(C\left(F_{\mathrm{A}}^{i},F_{\mathrm{B}}^{i}\right)\right)\right)\right), \tag{3}$$

$$A_{\mathrm{S}}^{i}=PWConv^{n}\left(PWConv\left(C\left(F_{\mathrm{A}}^{i},F_{\mathrm{B}}^{i}\right)\right)\right), \tag{4}$$

$$W^{i}=\delta\left(A_{\mathrm{C}}^{i}\otimes A_{\mathrm{S}}^{i}\right), \tag{5}$$

$$\hat{F}_{\mathrm{C}}^{i}=\left(W^{i}\otimes\hat{F}_{\mathrm{A}}^{i}\right)\oplus\left(\left(1-W^{i}\right)\otimes\hat{F}_{\mathrm{B}}^{i}\right), \tag{6}$$

式中：i 表示 SDFM 模块所在的特征层级的索引，不同

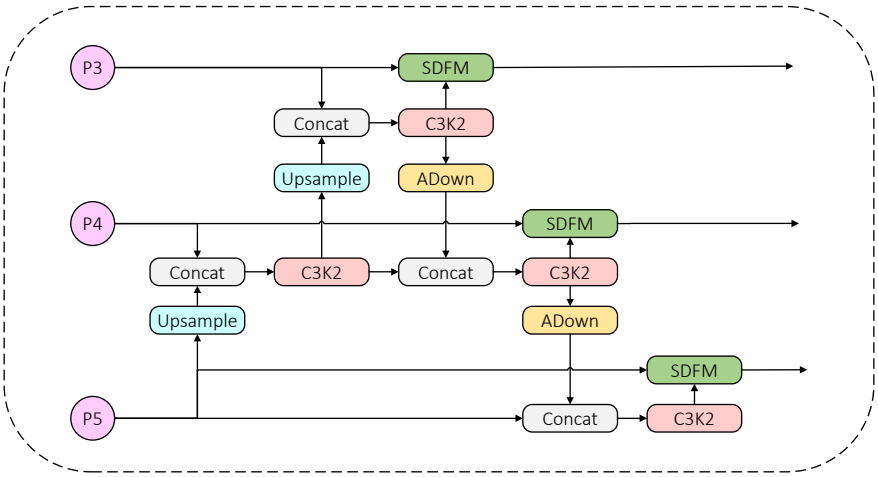


图 4 动态细节语义融合金字塔网络示意图

Fig. 4 Schematic diagram of the dynamic detail-semantic fusion pyramid network

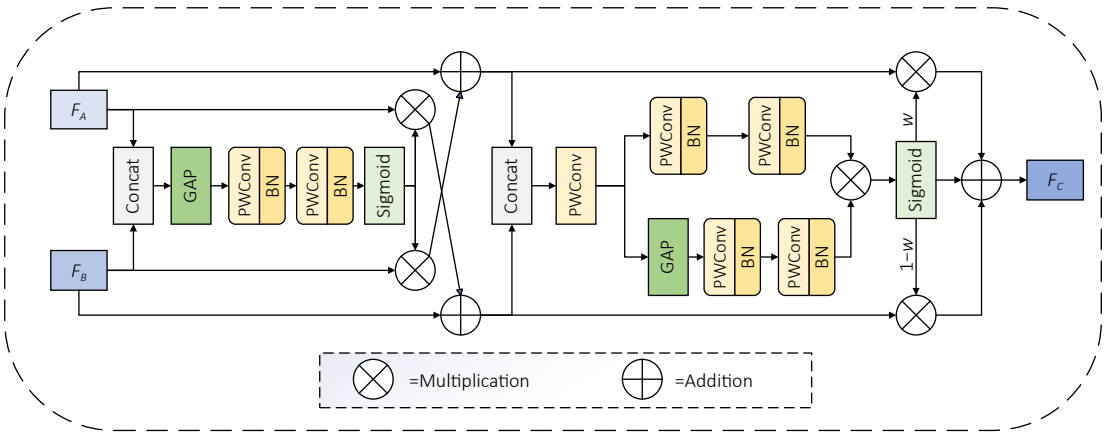


图 5 表层细节融合模块结构示意图

Fig. 5 Schematic diagram of the superficial detail fusion module

i 对应着不同分辨率和语义层次的特征融合；F_{A}^{i} 表示来自主干网络的浅层特征；F_{B}^{i} 表示经过上采样/下采样与多层融合后的深层特征；A_{C}^{i} 表示第 i 层的通道注意力权重特征；A_{S}^{i} 表示第 i 层的空间注意力权重特征；W^{i} 表示第 i 层计算所得的混合注意力自适应融合权重；$\hat{F}_{\mathrm{C}}^{i}$ 表示第 i 层最终输出的融合特征；$PWConv^{n}(\cdot)$表示 n 个级联的逐点卷积层；$C(\cdot)$表示在通道维度上的拼接操作；$\delta(\cdot)$表示 sigmoid 激活函数；而 $GAP(\cdot)$ 表示全局平均池化操作。

动态细节语义融合金字塔网络引入的 ADown 模块是 YOLOv9[31] 中一种用于目标检测任务中的下采样操作的卷积块。ADown 模块首先通过对特征图进行平均池化操作，减少特征图中的噪声和细节，然后分为两条支路：一条采用 3×3 卷积以提取局部空间特征，另一条使用 Max-Pool 以保留显著结构信息，再接 1×1 卷积，最后在通道维度上进行融合，形成多尺度特征表示，有助于模型同时感知细节与全局结构。ADown 模块的整体结构如图 6 所示。

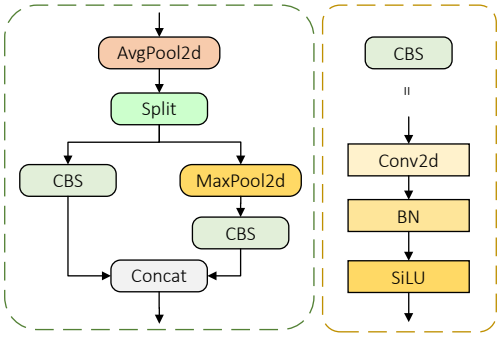


图 6 ADown 模块结构图

Fig. 6 Structure of ADown module

2.5 改进的检测头

由于在低照度环境下，通常存在目标的纹理细节不清晰、边缘模糊、信息丢失严重的问题，导致特征表达较弱、误检率高。针对以上问题，本文提出了双组检测头 DGDetect，其结构如图 7 所示。输入的多尺度特征图首先经过一个共享的特征提取干，由两层分组卷积构成，通过引入分组卷积来提取更丰富的特征表示，帮助检测头更准确地定位和识别目标，回归分支与分类分支在该阶段使用相同的卷积层和参数；随后，特征被送入两条独立的输出分支，每条分支由一个 1×1 卷积层组成，分别预测边界框分布和类别概率，两者互不共享权重，实现解耦检测分支设计。具体而言，分类分支专注于提炼光照不变的语义表征，对背景噪声或轻微位移更具鲁棒性；而回归分支则独立学习目标的几何形变与边缘细节，对模糊或不规则的目标边界更加敏感。这种解耦与轻量化策略不仅提升了整体检测精度，还在低照度、强噪声等极端条件下显著增强了模型的鲁棒性与推理效率。这样做的作用在于：一方面，采用共享的特征提取干结合 1×1 卷积的方式，减少了重复计算和卷积层数量，显著降低了高通道数检测层的计算量与参数量；另一方面，解耦的输出分支使分类与回归任务各自专注于最相关的特征学习，避免了任务间的干扰，从而在保持轻量化的同时提升了检测精度，尤其在复杂背景和极端环境下表现更优。其数学表达式如式 (7) 所示：

$$X_{l}^{\mathrm{out}} = SiLU\left(BN\left(W_{2\cdot G}^{l} SiLU\left(BN\left(W_{1\cdot G}^{l} X_{l}^{\mathrm{in}}\right)\right)\right)\right), \tag{7}$$

$$Reg_{l} = W_{\mathrm{reg}}^{l} * X_{l}^{\mathrm{out}}, \tag{8}$$

$$Cls_{l} = W_{\mathrm{cls}}^{l} * X_{l}^{\mathrm{out}}, \tag{9}$$

$$Y_l = Concat(Reg_l, Cls_l), \tag{10}$$

式中：l为多尺度特征的层级索引，代表不同的特征图尺度；X_l^{in}为第l个尺度层级的输入特征；X_l^{out}为经过共享特征提取干后的输出特征；$*G$ 表示分组卷积操作；W_{1*G}^l、W_{2*G}^l表示对应于第l 尺度层级的卷积核；BN 表示批归一化；$SiLU$ 表示 Sigmoid 线性单元激活；Reg_l 和 Cls_l 为分别表示第l尺度下回归分支和分类分支的预测输出；Y_l 为输出拼接；W_{reg}^l与W_{cls}^l为相应独立分支的1×1卷积核。

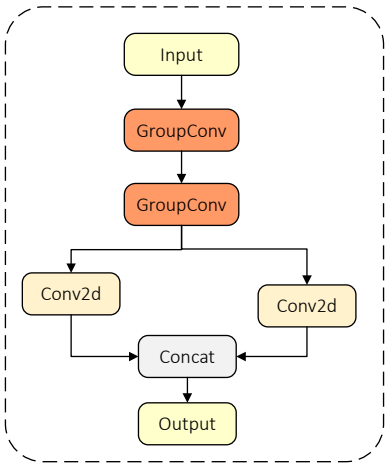


图 7 双组检测头结构图

Fig. 7 Structure of DualGroupDetect

通过将普通检测头替换为双组检测头，显著减少了模型整体的参数量和乘加运算次数实现了轻量化，同时充分利用了 DGDetect 更丰富的特征提取能力，强化了低照度环境下模型对图像细节及目标位置的感知效果，使模型在处理复杂场景时能够更准确地提取和识别细节特征并定位目标，从而有效地解决低照度环境下特征表达不足的问题。

3 实验结果

3.1 数据集

为评估改进模型在低照度环境中的检测性能，本文选用专为低照度视觉任务构建的 ExDark 数据集[32] 进行对比实验。ExDark 共收录 7363 张低光照图像，覆盖 12 类目标，场景包含城市夜景、昏暗室内、交通工具及行人等，可真实反映日常弱光条件。数据集按 8∶1∶1 的比例，随机划分为训练集、验证集与测试集，分别对应 5890、736 与 737 张图像。

3.2 实验设置

实验环境依托 Windows 11 平台，硬件配置为 NVIDIA RTX3080 Ti (显存为 12 GB)。代码采用 Python 3.8 编写，并以 PyTorch 2.0.0 作为深度学习框架。基准模型 YOLOv11n 和本文提出的方法均遵循统一的训练策略。具体而言，模型首先加载了在标准大规模数据集上预训练的官方权重作为初始参数，以利用其在大规模数据上学习到的通用特征。随后，所有模型均在目标低照度数据集上进行了完整的微调。微调过程持续 300 个 epoch，采用随机梯度下降 (SGD) 优化器，batch size 为 8，初始学习率为 0.015，动量参数为 0.937，并应用了余弦学习率调度器。此外，我们还引入了包括马赛克增强、图像翻转、平移、缩放和剪切等一系列数据增强技术。所有输入图像在训练前均统一缩放到 640 pixel×640 pixel。

3.3 模型评价指标

在低照度目标检测任务中，漏检和误检尤为普遍，因此通常将 mAP50 与 mAP50-95 作为主要评估指标。mAP 同时兼顾精确率和召回率，因而被普遍采用为评估目标检测算法整体表现的关键指标，同时，模型的体积与推理成本也是重要指标，常用参数量 (Params) 和浮点运算次数 (FLOPs) 对二者进行衡量。指标计算式如式 (11)~(14) 所示。

$$AP = \int_0^1 P(y)\mathrm{d}y, \tag{11}$$

$$mAP = \frac{1}{n}\sum_{i=0}^{n} AP_i, \tag{12}$$

$$P = \frac{TP}{TP+FP}, \tag{13}$$

$$R = \frac{TP}{TP+FN}, \tag{14}$$

式中：TP 为被正确识别的正样本数；FP 为被错误的检测为正样本的数量；FN 为未被检测出的目标数；AP 表示平均精度，将全部 n 个类别的 AP 取算术平均即可得到 mAP，mAP50 和 mAP50-95 分别表示不同 IoU 阈值下的平均精度均值。

3.4 不同空间金字塔池化模块的实验

为了验证本文提出的 MPSPPF 在低照度目标检测任务中的有效性以及对比的公平性，本文基于

YOLOv11n 模型，将 MPSPPF 与 SPP[33]、SPPF、ASPP-lite 以及 SimCSPSPPF-lite 在 ExDark 数据集上进行了实验对比。在不改变模块核心结构思想的前提下通过调整分支通道数和分支数量匹配复杂度，使其参数量和 FLOPs 与所提出的 MPSPPF 模块差异控制在±5%，排除模型规模差异对性能的影响，从而更准确地评估不同模块在相同复杂度下的表现。实验结果如表 1 所示，使用-lite 后缀标识 ASPP[34] 和 SimCSPSPPF[35] 的等复杂度版本。

由表 1 可知，在保证参数量和计算量尽可能相等的条件下，MPSPPF 模块在主要评估指标 mAP50 和 mAP50-95 上均表现优异。与原始模型使用的 SPPF 模块相比，MPSPPF 模块在 mAP50 上提升了 0.8%，在 mAP50-95 上提升了 0.3%，显著提高了模型的检测精度。这表明，MPSPPF 模块在资源消耗较低的情况下，成功增强了模型的检测性能。同时，对比其他主流空间金字塔池化模块，MPSPPF 模块也表现出更高的性能。这些结果表明了 MPSPPF 模块的有效性，它在保持了模型低复杂度的同时，提升了检测的准确性，为低照度目标检测任务提供了一个有效的选择。

3.5 不同注意力机制在 CSP-EDHAN 模块中的对比实验

为了验证本文提出的混合注意力机制在低照度目标检测任务中的有效性，本文基于 YOLOv11n 模型，在 ExDark 数据集上进行了实验对比。在保证参数量和计算量尽量相等的条件下，替换 CSP-EDHAN 模块中的注意力机制，来比较 CSP-EDHAN 模块中的混合注意力机制与其他主流空间注意力机制 (包括 coordinate attention[36] (CA)、triplet attention[37] (TA)、efficient multi-scale attention[38] (EMA) 和 shuffle attention[39] (SA)) 对模型性能的影响。实验结果如表 2 所示。

从表 2 可以看出，在相同网络结构下，采用混合注意力机制的 CSP-EDHAN 模块在 mAP50 和 mAP50-95 两项关键指标上均取得最高成绩。尽管 EMA 在精确率方面表现最优，但混合注意力在召回率上的显著提升以及关键指标 mAP 的整体增长，表明其在复杂环境中更注重捕捉更多目标，即便牺牲少量精度，但整体性能依然更优。与 SA 相比，混合注意力在召回率、精确率和 mAP 上均有明显优势，体现了更均衡的检测能力。综合来看，混合注意力机制在低照度目标检测任务中展现出更优的综合性能。

3.6 消融实验

为验证所提改进策略的有效性与可靠性，本文在同一超参数设置下设计了多组消融实验，并在数据集上逐一测试，其中“√”表示采用该改进模块，“×”表示不采用该改进模块。结果汇总于表 3。

根据消融结果，可以得出以下结论。首先，引入

表 1　不同空间金字塔池化模块的性能对比

Table 1　Comparison of performance of different spatial pyramid pooling module

Model	mAP50/%	mAP50-95/%	Precision/%	Recall/%	Params/M	FLOPs/G
SPP	65.7	40.2	70.9	60.5	2.58	6.3
SPPF	65.3	40.5	71.4	59.3	2.58	6.3
ASPP-lite	65.2	40.0	68.6	**60.8**	2.65	6.3
SimCSPSPPF-lite	64.8	40.2	75.1	57.0	2.62	6.3
MPSPPF	**66.1**	**40.8**	**74.4**	58.3	2.58	6.3

表 2　不同注意力机制在 CSP-EDHAN 模块中的对比实验

Table 2　Comparative experiments of different attention mechanisms in the CSP-EDHAN module

Model	mAP50/%	mAP50-95/%	Precision/%	Recall/%	Params/M	FLOPs/G
CA	65.6	40.5	72.2	59.8	2.63	7.6
TA	65.5	40.2	72.5	58.8	2.62	7.6
EMA	65.4	40.5	**72.9**	58.3	2.63	7.6
SA	65.7	40.6	71.0	60.2	2.62	7.6
Hybrid attention	**66.2**	**40.8**	71.4	**60.5**	2.62	7.6

表 3 消融实验

Table 3 Ablation experiment

MPSPPF	CSP-EDHAN	DDS-FPN	DGDetect	mAP50/%	mAP50-95/%	Precision/%	Recall/%	Params/M	FLOPs/G
×	×	×	×	65.3	40.5	71.4	59.3	2.58	6.3
√	×	×	×	66.1	40.8	74.4	58.3	2.58	6.3
×	√	×	×	66.2	40.8	71.4	60.5	2.62	7.6
×	×	√	×	66.5	41.0	73.7	59.7	3.13	8.7
×	×	×	√	65.7	40.8	74.5	60.3	**2.32**	**5.1**
√	√	×	×	66.6	41.8	**76.7**	58.4	2.62	7.6
√	√	√	×	67.6	42.2	74.3	60.5	3.17	10.0
√	√	√	√	**68.4**	**42.7**	72.7	**61.6**	2.81	8.2

MPSPPF 模块，mAP50 和 mAP50-95 分别提高了 0.8% 和 0.3%，说明 MPSPPF 模块通过捕获局部细粒度特征与全局上下文信息，显著增强了模型的检测准确性。接着引入 CSP-EDHAN 模块，虽然单独引入 CSP-EDHAN 模块也能提升模型效果，但是当其与 MPSPPF 模块协同工作时，网络性能得到更大幅度的提升，与基础模型相比，mAP50 和 mAP50-95 均提高了 1.3%，这说明 MPSPPF 模块和 CSP-EDHAN 模块具有互补优势，两者的协同作用在于，MPSPPF 为 CSP-EDHAN 提供了丰富且稳定的多尺度上下文信息和梯度增强，促使 CSP-EDHAN 通过更精准的注意力机制和高效的特征融合，共同提升了模型的特征表达能力和精度。然后引入 DDS-FPN，单独引入 DDS-FPN 时，mAP50 和 mAP50-95 较基础模型分别提高了 1.2% 和 0.5%，这说明 DDS-FPN 能使高低维度的特征信息更好地融合在一起。同时，虽然单独引入 DDS-FPN、CSP-EDHAN 或 MPSPPF 均能提升检测精度，但三者协同工作时网络性能得到更大幅度的提升。MPSPPF 为 CSP-EDHAN 提供稳定的多尺度上下文与梯度增强，促使其以更精准的注意力和高效融合提升特征表达与精度。在此基础上，DDS-FPN 将经过优化的特征进行细节恢复与多尺度融合，实现语义与细节的双重强化，从而在低照度、强噪声等复杂场景下显著提升特征表达能力与目标检测精度。最后使用双组检测头代替原始检测头，保持轻量化的同时，mAP50 进一步提高了 0.8%。

综上所述，LMD-YOLO 模型的 mAP50 可达 68.4%，相对于基线模型提升了 3.1%，每个改进点都能够有效提升模型的检测性能，且在提升平均精度均值的同时，依然保持较低的计算量和参数量，确保了模型的运行效率。

3.7 对比实验

为了进一步验证本文改进模型在同类方法中的优势，将 LMD-YOLO 与现有主流目标检测模型以及最新的低照度目标检测算法进行比较，测试结果如表 4 所示，其中加粗表示最优值。

实验结果表明，LMD-YOLO 在 mAP50 和 mAP50-95 两项关键评估指标上的表现均优于目前主流目标检测算法和最新的低照度目标检测算法。相较于主流目标检测算法 YOLOv5n、YOLOv5s、YOLOv8n、YOLOv10n、YOLOv11n、Faster R-CNN、RT-DETR，LMD-YOLO 在 mAP50 指标上分别提高了 4.6%、0.6%、1.7%、4.2%、3.1%、6.1%、4%。与近年来最新的低照度目标检测算法 LOL-YOLO、MLFE-YOLOX、EnlightenGAN+YOLOv6s 相比，LMD-YOLO 的 mAP50 分别提升了 0.3%、2.2%、6.1%。同时，现有目标检测算法的参数量和计算量均值在 5.3 M 和 13.5 G 左右，高于本文模型 LMD-YOLO。上述方法由于对低光照图像细节信息识别和整合不足、忽略了背景因素的影响以及缺乏对上下文的自适应学习能力导致较低的指标。与之不同的是，LMD-YOLO 能更充分地提取出低照度图像的特征信息，并通过动态细节语义融合金字塔网络的引入以及检测头的改进等让算法可以在保持边缘细节的同时，动态整合多尺度上下文信息。实验数据也证明了 LMD-YOLO 的有效性，在计算资源消耗保持较低水平的同时 mAP 值达到 68.4% 和 42.7%，有效提高了低照度环境下的检测精度。

为了进一步验证 LMD-YOLO 不仅在低照度场景下优于当前主流模型，在常规照度场景下也能具有适用性，采用 PASCAL VOC2012[41] 进行对比实验，PASCAL VOC2012 数据集包含了 20 个不同的类别，涵盖了人、动物、车辆和室内物品等。在保证模型参数一致的前提

表 4 ExDark 数据集对比实验

Table 4 Comparative experiment on ExDark

Model	Precision/%	Recall/%	Params/M	FLOPs/G	mAP50/%	mAP50-95/%
YOLOv5n	73.1	57.9	**2.18**	**5.8**	63.8	38.8
YOLOv5s	**75.4**	61.5	7.82	18.8	67.8	41.4
YOLOv8n	72.5	60.2	2.69	6.8	66.7	41.1
YOLOv10n	72.5	55.6	2.72	8.2	64.2	40.1
YOLOv11n	71.4	59.3	2.58	6.3	65.3	40.5
Faster R-CNN	37.4	**75.8**	137.12	370.2	62.3	31.3
RT-DETR	71.1	57.1	35.63	118.8	64.4	39.6
LOL-YOLO[19]	70.9	62.5	/	/	68.1	42.3
MLFE-YOLOX[24]	/	/	10.23	29.7	66.2	39.0
EnlightenGAN+YOLOv6s[40]	66.1	53.7	/	/	62.3	33.6
LMD-YOLO	72.7	61.6	2.81	8.2	**68.4**	**42.7**

下，对 LMD-YOLO 与 YOLOv11n 进行了实验比较，对比实验结果如表 5 所示。

在 PASCAL VOC2012 测试中，LMD-YOLO 较 YOLOv11n 的 mAP50 与 mAP50-95 分别提升 1.8% 和 1.3%，精度和召回率同步增长 1.4% 和 1.5%。实验表明 LMD-YOLO漏检率和误检率均有所下降，证明了 LMD-YOLO 不仅能在低照度环境中保持较高性能，在常规照明条件下也能够可靠的发挥作用。

同时，为了进一步验证 LMD-YOLO 在复杂低光场景下的优越性，本文利用 Albumentations 图像增强库，在 Roboflow 平台提供的 Low-light-v2 数据集上模拟了复杂低光环境，生成了包含复杂环境特征的低照度数据集，并在此低照度数据集上对模型进行了对比实验，对比实验结果如表 6 所示。

在复杂低光场景数据集上的测试结果表明，改进模型相较于基础模型展现出更强的泛化能力和鲁棒性。改进模型在 mAP50% 和 mAP50-95% 上分别取得了 2.2% 和 1.3% 地提升，并且在召回率上实现了显著增长，这表明它在复杂低光环境中能够检测出更多目标，有效减少漏检。改进模型在复杂环境中为了捕捉更多目标而牺牲了一小部分精度，但整体性能更优越，这证明了其在复杂低光环境下的有效性。

3.8 低照度图像检测可视化及分析

为进一步验证 LMD-YOLO 在低光照检测中的优势，对 YOLOv11n 与 LMD-YOLO 生成的热力图进行可视化对比 (见图 8)，图中亮区为模型关注的关键位置。

实验数据显示：在相同低照度测试样本下，原始模型的热力响应呈弥散分布，注意力易受背景噪声干扰。相比之下，LMD-YOLO 通过引入 DDS-FPN 与 MPSPPF 等结构，使优化后模型在低照度场景下对目标的关注更加集中且显著，同时优化模型在暗区边缘轮廓展现出更精细的响应特性。该现象验证了改进结构通过增强局部特征表征力与跨层级信息聚合，显著提升了模

表 5 PASCAL VOC2012 数据集对比实验

Table 5 Comparative experiment on PASCAL VOC2012

Model	Precision/%	Recall/%	mAP50/%	mAP50-95/%
YOLOv11n	70.5	61.2	67.3	50.4
LMD-YOLO	71.9	62.7	69.1	51.7

表 6 复杂低光环境下的对比实验

Table 6 Comparative experiments in complex low-light environments

Model	Precision/%	Recall/%	mAP50/%	mAP50-95/%
YOLOv11n	80.1	66.1	72.3	35.8
LMD-YOLO	76.4	74.6	74.5	37.1

型在低信噪比环境下的注意力鲁棒性。此外，为了更直接地检验 LMD-YOLO 的检测效果，本文在低光照场景下进行了可视化实验，具体的对比检测结果如图 9 所示。

由图 9 可知，我们可以发现，RT-DETR、YOLOv8n 和 YOLOv10n 都未能准确定位和区分左侧模糊的自行车，这反映出模型对模糊场景下的检测能力有限，容易出现漏检；Faster R-CNN、YOLOv5s 不仅将无关的背景噪声错误判断为椅子和摩托车，而且还出现了出现漏检；虽然 YOLOv5n 成功地检测出了目标，但由于对局部细节过度敏感，它将部分无关的背景噪声与纹理误认为目标，因而出现了误检；尽管 YOLOv11n 能够检测出左侧模糊自行车，但是由于图片背景较暗，右边的自行车虽然清晰，但与背景的对比度较低，模型难以区分目标和背景，导致模型出现漏检。在相同测试场景下，相比于其他模型，LMD-YOLO 不仅能够准确定位低照度条件下目标的位置，同时实现了更准确的分类识别，使其在低照度环境下的目标检测任务中具有更高的实用价值和可靠性。

综上，在处理低照度环境下图像模糊以及背景噪声与检测目标难以区分等问题时，相比于其他模型，LMD-YOLO 在低光环境下不仅能够精确锁定目标真实位置，并以较高置信度输出正确类别标签，在位置定位与类别判别两项评估指标上表现出显著优势，展现出更高的检测性能和精确度。

4 结 论

本文针对低照度环境下检测任务中夜间模糊、边界不明显和明暗差距大等问题，提出了一种融合多尺度特征与局部细节强化的低照度目标检测算法 LMD-YOLO。首先通过设计 MPSPPF 模块，来应对低照度图像噪声敏感和细节丢失的问题。然后提出了 CSP-EDHAN 模块，结合双路径特征交互以及注意力增强，减少低照度图像中低级特征的丢失。接着设计动态细节语义融合金字塔网络，在双向金字塔路径中实现多尺度特征高效对齐与增强，进一步提升了目标检测的精度和鲁棒性。最后设计双组检测头，更好地整合上下文信息，同时强化了低照度下的特征提取，以适应光线不足的环境，提高对低照度环境下目标的识别能力。基于 EXDark 与 PASCAL VOC 2012 两个公开数据集的对比试验显示，与基线模型 YOLOv11n 相比，LMD-YOLO 的精确度、召回率、mAP50 和 mAP50-95 等关键指标，在 ExDark 数据集上分别提升了 1.3%、2.3%、3.1% 和 2.2%，在 PASCAL VOC2012 数据集上分别获得了 1.4%、1.5%、1.8% 和 1.3% 的增幅。此外，在复杂低照度场景数据集上，LMD-YOLO 与基线模型 YOLOv11n 相比，mAP50 和 mAP50-95 分别提升了 2.2% 和 1.3%。实验结果证明了本文算法在低照度图像目标检测任务上的优越性和泛化性，并展示了本文算法在常规光照场景下的适用性。为

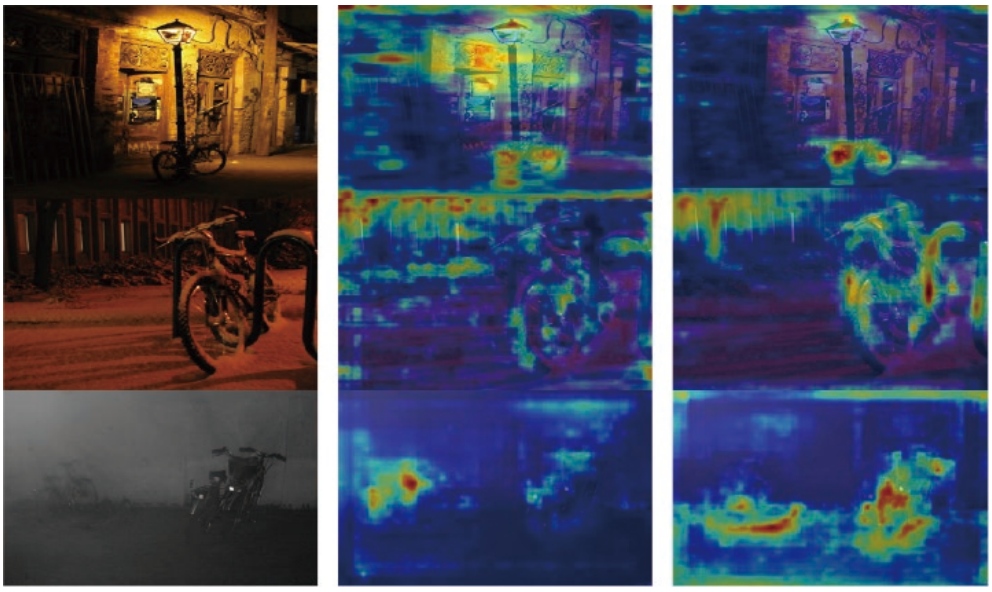


图 8 热力图对比

Fig. 8 Heatmap comparison

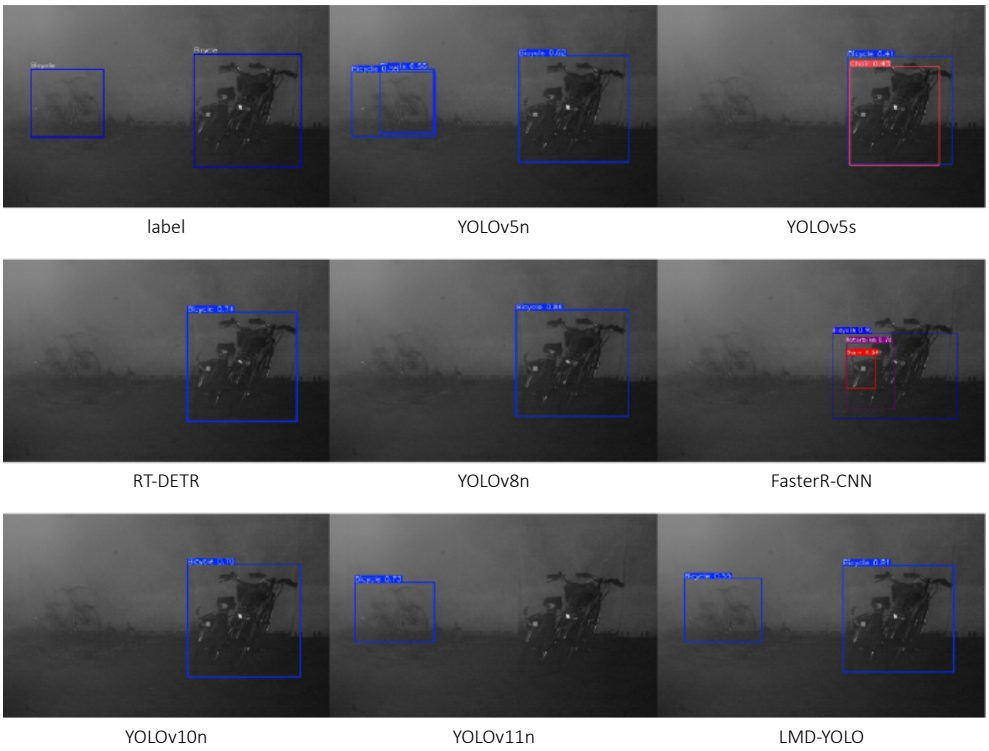


图 9 不同模型检测结果对比

Fig. 9 Comparison of detection results among different model

促进计算机视觉技术在低照度图像检测领域的应用提供了可靠方案。后续工作将围绕更为多变的照明场景展开深入的目标检测研究，并计划面向自然场景设计具有更高鲁棒性与泛化能力的检测框架。

作者贡献

林珊玲：提出改进方法和实验设计，论文撰写与修改；刘梅：进行实验设计及数据整理和分析；林坚普：论文审核与修改；林志贤：监督和领导研究活动的策划和执行；郭太良：为研究活动的策划和执行进行管理和协调；孙磊：获取资助。

利益冲突

所有作者声明无利益冲突

参考文献

1. Xu X, Xue Z P, Zhao Y. Research on an algorithm of express parcel sorting based on deeper learning and multi-information recognition[J]. *Sensors*, 2022, **22**(17): 6705.
2. Gorschlüter F, Rojtberg P, Pöllabauer T. A survey of 6D object detection based on 3D models for industrial applications[J]. *J Imaging*, 2022, **8**(3): 53.
3. Liu Y, Zhang C S, Dong X J. A survey of real-time surface defect inspection methods based on deep learning[J]. *Artif Intell Rev*, 2023, **56**(10): 12131−12170.
4. Li S B, Xiao Z J, Qu H C et al. Multi-granularity feature and shape-position similarity metric method for ship detection in SAR images[J]. *Opto-Electron Eng*, 2025, **52**(2): 240254.
李士博, 肖振久, 曲海成, 等. 面向 SAR 图像舰船检测的多粒度特征与形位相似度量方法[J]. 光电工程, 2025, **52**(2): 240254.
5. Zhang P, Ren H Y, Tian J Q et al. Image-based aerial fire detector based on cross-scale fusion[J]. *Opto-Electron Eng*, 2025, **52**(1): 240253.
张沛, 任恒英, 田佳麒, 等. 基于跨尺度融合的图像型航空火灾探测器[J]. 光电工程, 2025, **52**(1): 240253.
6. Liu G H, Yang Q, Meng Y B et al. A progressive fusion image enhancement method with parallel hybrid attention[J]. *Opto-Electron Eng*, 2023, **50**(4): 220231.
刘光辉, 杨琦, 孟月波, 等. 一种并行混合注意力的渐进融合图像增强方法[J]. 光电工程, 2023, **50**(4): 220231.
7. Girshick R, Donahue J, Darrell T et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, USA, 2014: 580–587. https://doi.org/10.1109/CVPR.2014.81.
8. Girshick R. Fast R-CNN[C]//*Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 2015: 1440–1448. https://doi.org/10.1109/ICCV.2015.169.
9. Ren S Q, He K M, Girshick R et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, **39**(6): 1137−1149.
10. Redmon J, Divvala S, Girshick R et al. You only look once: unified, real-time object detection[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, 2016: 779–788.

https://doi.org/10.1109/CVPR.2016.91.
11. Liu W, Anguelov D, Erhan D et al. SSD: single shot MultiBox detector[C]//*Proceedings of the 14th European Conference on Computer Vision (ECCV 2016)*, Amsterdam, The Netherlands, 2016: 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.
12. Lv F F, Lu F, Wu J H et al. MBLLEN: low-light image/video enhancement using CNNs[C]//*Proceedings of the British Machine Vision Conference (BMVC 2018)*, Newcastle, UK, 2018: 220.
13. Wei C, Wang W J, Yang W H et al. Deep Retinex decomposition for low-light enhancement[C]//*Proceedings of the British Machine Vision Conference (BMVC 2018)*, Newcastle, UK, 2018: 155.
14. Guo C L, Li C Y, Guo J C et al. Zero-reference deep curve estimation for low-light image enhancement[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020)*, Seattle, WA, USA, 2020: 1777–1786. https://doi.org/10.1109/CVPR42600.2020.00185.
15. Yuan J J, Hu Y L, Sun Y F et al. A plug-and-play image enhancement model for end-to-end object detection in low-light condition[J]. *Multimed Syst*, 2024, **30**(1): 27.
16. Lin J P, Lai F W, Lin S L et al. Frequency-guided dual-collapse transformer for low-light image enhancement[J]. *Eng Appl Artif Intell*, 2025, **142**: 109906.
17. Shi L C, Yang C, Liu X C et al. Lightweight low-light object detection algorithm based on CDD-YOLO[J]. *Comput Eng Appl*, 2025, **61**(6): 106−117.
史丽晨, 杨超, 刘雪超, 等. 基于 CDD-YOLO 的轻量级低光照目标检测算法[J]. 计算机工程与应用, 2025, **61**(6): 106−117.
18. Jiang Z T, Zhai F S, Qian Y et al. Low illumination object detection combined with feature enhancement and multi-scale receptive field[J]. *J Comput Res Dev*, 2023, **60**(4): 903−915.
江泽涛, 翟丰硕, 钱艺, 等. 结合特征增强和多尺度感受野的低照度目标检测[J]. 计算机研究与发展, 2023, **60**(4): 903−915.
19. Jiang C J, He X Y, Xiang J et al. LOL-YOLO: low-light object detection incorporating multiple attention mechanisms[J]. *Comput Eng Appl*, 2024, **60**(24): 177−187.
蒋畅江, 何旭颖, 向杰, 等. LOL-YOLO: 融合多注意力机制的低照度目标检测[J]. 计算机工程与应用, 2024, **60**(24): 177−187.
20. Zhou L, Dong Y Y, Ma B Y et al. Object detection in low-light conditions based on DBS-YOLOv8[J]. *Cluster Comput*, 2025, **28**(1): 55.
21. Xue R, Duan J L, Du Z W. MPE-DETR: a multiscale pyramid enhancement network for object detection in low-light images[J]. *Image Vis Comput*, 2024, **150**: 105202.
22. Yin X C, Yu Z D, Fei Z T et al. PE-YOLO: pyramid enhancement network for dark object detection[C]//*Proceedings of the 32nd International Conference on Artificial Neural Networks and Machine Learning - ICANN 2023*, Heraklion, Crete, Greece, 2023: 163–174. https://doi.org/10.1007/978-3-031-44195-0_14.
23. Jiang Z T, Xiao Y, Zhang S Q et al. Low-illumination object detection method based on dark-YOLO[J]. *J Comput-Aided Des Comput Graphics*, 2023, **35**(3): 441−451.
江泽涛, 肖芸, 张少钦, 等. 基于 Dark-YOLO 的低照度目标检测方法[J]. 计算机辅助设计与图形学学报, 2023, **35**(3): 441−451.
24. Tan H, Zhang J L, Jia X. Low-light target detection algorithm based on multi-level feature extraction[J]. *Comput Eng Appl*, 2024, **60**(24): 235−242.
谭豪, 张惊雷, 贾鑫. 基于多级特征提取的低光照目标检测算法[J]. 计算机工程与应用, 2024, **60**(24): 235−242.
25. Khanam R, Hussain M. YOLOv11: an overview of the key architectural enhancements[Z]. arXiv: 2410.17725, 2024. https://doi.org/10.48550/arXiv.2410.17725.
26. Jocher G, Chaurasia A, Qiu J. Ultralytics YOLOv8 [CP/OL]. (2023-01-10) [2024-09-20]. https://github.com/ultralytics/ultralytics.
27. Wang A, Chen H, Liu L H et al. YOLOv10: real-time end-to-end object detection[C]//*Proceedings of the 38th International Conference on Neural Information Processing Systems*, Vancouver, Canada, 2024: 3429.
28. Jocher G, Stoken A, Borovec J et al. Ultralytics YOLOv5 [CP/OL]. (2020-05-29) [2024-10-31]. https://github.com/ultralytics/yolov5.
29. Springenberg J T, Dosovitskiy A, Brox T et al. Striving for simplicity: the all convolutional net[C]//*Proceedings of the 3rd International Conference on Learning Representations*, San Diego, CA, USA, 2015.
30. Tang L F, Zhang H, Xu H et al. Rethinking the necessity of image fusion in high-level vision tasks: a practical infrared and visible image fusion network based on progressive semantic injection and scene fidelity[J]. *Inf Fusion*, 2023, **99**: 101870
31. Wang C Y, Yeh I H, Liao H Y M. YOLOv9: learning what you want to learn using programmable gradient information[C]//*Proceedings of the 18th European Conference on Computer Vision*, Milan, Italy, 2025: 1–21. https://doi.org/10.1007/978-3-031-72751-1_1.
32. Loh Y P, Chan C S. Getting to know low-light images with the Exclusively Dark dataset[J]. *Comput Vis Image Underst*, 2019, **178**: 30−42.
33. He K M, Zhang X Y, Ren S Q et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Trans Pattern Anal Mach Intell*, 2015, **37**(9): 1904−1916.
34. Chen L C, Papandreou G, Kokkinos I et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Trans Pattern Anal Mach Intell*, 2018, **40**(4): 834−848.
35. Li C Y, Li L L, Geng Y F et al. YOLOv6 v3.0: a full-scale reloading[Z]. arXiv: 2301.05586, 2023. https://doi.org/10.48550/arXiv.2301.05586.
36. Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2021: 13708–13717. https://doi.org/10.1109/CVPR46437.2021.01350.
37. Misra D, Nalamada T, Arasanipalai A U et al. Rotate to attend: convolutional triplet attention module[C]//*Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, 2021: 3138–3147. https://doi.org/10.1109/WACV48630.2021.00318.
38. Ouyang D L, He S, Zhang G Z et al. Efficient multi-scale attention module with cross-spatial learning[C]//*Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece, 2023: 1–5. https://doi.org/10.1109/ICASSP49357.2023.10096516.
39. Zhang Q L, Yang Y B. SA-Net: shuffle attention for deep convolutional neural networks[C]//*Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, ON, Canada, 2021: 2235–2239. https://doi.org/10.1109/ICASSP39728.2021.9414568.
40. Wang J, Yang P, Liu Y S et al. Research on improved YOLOv5 for low-light environment object detection[J]. *Electronics*, 2023, **12**(14): 3089.
41. Everingham M, Van Gool L, Williams C K I et al. The PASCAL visual object classes (VOC) challenge[J]. *Int J Comput Vis*, 2010, **88**(2): 303−338.

作者简介



林珊玲 (1991-)，女，福建泉州人，博士，副教授，硕士研究生导师，2014 年、2020 年于福州大学分别获得学士、博士学位，主要从事显示驱动、图像处理等方面的研究。

E-mail：sllin@fzu.edu.cn



【通信作者】林坚普 (1989-)，男，福建泉州人，博士，副教授，硕士研究生导师，2012 年、2019 年于福州大学分别获得学士、博士学位，主要从事图像处理、深度学习、液晶透镜器件、裸眼 3D 显示技术等方面的研究。

E-mail：ljp@fzu.edu.cn

Low-light object detection with multi-scale feature fusion and local detail enhancement

Lin Shanling[1,2], Liu Mei[1,2], Lin Jianpu[1,2*], Lin Zhixian[1,2,3], Guo Tailiang[2,3], Sun Lei[4]

[1] School of Advanced Manufacturing, Fuzhou University, Quanzhou, Fujian 362000, China;
[2] Fujian Science and Technology Innovation Laboratory for Photoelectric Information, Fuzhou, Fujian 350116, China;
[3] College of Physics and Information Engineering, Fuzhou University, Fuzhou, Fujian 350116, China;
[4] Zhicheng College, Fuzhou University, Fuzhou, Fujian 350116, China

Abstract

Objective Nighttime low-light imaging is often affected by underexposure, limited dynamic range, and amplified sensor noise. These factors weaken edge, texture, and contour cues, and they increase background interference and feature ambiguity. Such degradation reduces detector reliability and causes missed detections, false positives, and localization bias, with more severe impact on small objects, distant targets, and dim instances. Based on YOLOv11n, LMD-YOLO was proposed as a low-light object detection method that integrates multi-scale features and local detail enhancement. Without substantially increasing model complexity, it strengthens multi-scale feature fusion and local detail extraction for low-light images and reduces high noise sensitivity in low-illumination scenes, thereby improving low-light detection accuracy.

Methods LMD-YOLO combined multi-scale feature fusion and local detail enhancement through four components. 1) A Multi-Pool Spatial Pyramid Fast Pooling module (MPSPPF) was inserted into the backbone. It contained parallel pooling branches and a gradient-enhanced concatenation part. The parallel pooling used MaxPool2d and AvgPool2d. Max pooling strengthened salient responses and benefited key-object extraction in low-contrast images. Average pooling produced smoother representations and reduced noise sensitivity. The pooled outputs were fused with weighted aggregation to preserve global structure while retaining local detail cues. MPSPPF also used repeated sequential pooling. The same pooling operation was applied stage by stage to the previous pooled output to form a deeper pyramid. This design enlarged the effective receptive field with limited cost and built a richer feature hierarchy. It captured local and global context that supported recognition of blurred, partially occluded, or weakly illuminated objects. The gradient-enhanced concatenation improved information flow across pooling stages and reduced detail attenuation in early feature extraction. 2) A Cross Stage Partial–Enhanced Dual Hybrid Attention Network module (CSP-EDHAN) was constructed for fine-grained low-light features. Depthwise separable convolution reduced computation while maintaining spatial sensitivity. Residual connections stabilized feature propagation under severe noise. A multi-path fusion structure increased cross-stage information flow and strengthened weak-target cues. A dual hybrid attention mechanism combined channel emphasis and spatial selection. Channel emphasis highlighted informative channels under illumination degradation. Spatial selection focused responses on object regions and suppressed random-noise activations. This module improved feature contrast for dim targets and reduced false activations caused by background clutter. 3) A Dynamic Detail–Semantic Fusion Pyramid Network (DDS-FPN) was designed for neck fusion. A channel–spatial attention unit from the SDFM module guided feature selection during cross-scale aggregation. Shallow detail cues and deep semantic cues were fused with dynamic weights, which raised responses in target regions and lowered responses in noisy background regions. An adaptive downsampling module (ADown) aligned deep semantics with shallow details during scale transitions. It reduced feature misalignment, improved feature consistency across

Foundation item: National Natural Science Foundation of China (NSFC) Youth Fund (62101132)

*Correspondence: Lin J P, E-mail: ljp@fzu.edu.cn

pyramid levels, and compensated for detail loss caused by low illumination. A bidirectional pyramid path strengthened bottom-up detail propagation and top-down semantic guidance. Multi-scale features were aligned and enhanced for small, medium, and large objects under low-light conditions. 4) A dual-group detection head was proposed to balance accuracy and efficiency. Input multi-scale feature maps first passed through a shared feature-extraction stem composed of two grouped convolution layers. Grouped convolution increased feature diversity at low cost and improved localization and recognition. The classification and regression branches shared the stem parameters to reduce redundant computation. Two independent output branches were then applied. Each branch used a 1×1 convolution layer to predict class probabilities and bounding-box distributions. The decoupled branch design reduced task interference and improved training stability for noisy low-light inputs.

Results and Discussions Comparative experiments were conducted on EXDark and PASCAL VOC 2012 with YOLOv11n as the baseline. On EXDark, LMD-YOLO improved precision, recall, mAP50, and mAP50-95 by 1.3%, 2.3%, 3.1%, and 2.2%. On PASCAL VOC 2012, the same metrics increased by 1.4%, 1.5%, 1.8%, and 1.3%. Additional evaluation on a complex low-light scene dataset showed gains of 2.2% in mAP50 and 1.3% in mAP50-95 over the baseline. The results indicated that MPSPPF reduced early noise sensitivity through multi-branch pooling and context aggregation. Weighted fusion preserved structure cues while limiting noise amplification. Sequential pooling improved long-range context modeling, which benefited targets with weak boundaries. CSP-EDHAN strengthened weak-target representation through cross-stage fusion and attention-guided enhancement. Depthwise separable convolution preserved efficiency and improved feature robustness. DDS-FPN improved multi-scale alignment through dynamic detail–semantic fusion and attention-based selection. ADown reduced information loss during downsampling and improved semantic–detail consistency across pyramid levels. The dual-group head maintained lightweight computation and improved optimization through decoupled branches. Performance gains on PASCAL VOC 2012 suggested improved general feature quality rather than low-light domain bias. The method also showed stable behavior under mixed illumination and cluttered backgrounds, which are common in real nighttime scenes.

Conclusions LMD-YOLO improves low-light object detection through noise-robust feature extraction, weak-target enhancement, and dynamic multi-scale fusion while preserving a lightweight design. The method achieves consistent accuracy gains on low-light benchmarks and remains effective under normal illumination. Future work targets broader illumination distributions, stronger robustness to extreme noise, glare, and motion blur, and improved generalization in natural scenes.

Keywords: low-light object detection; YOLOv11; feature fusion; detail enhancement

Lin S L, Liu M, Lin J P et al. Low-light object detection with multi-scale feature fusion and local detail enhancement[J]. *Opto-Electron Eng*, 2026, **53**(3): 250241; DOI: 10.12086/oee.2026.250241

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文