

# 结合特征融合增强和细节特征的低照度小目标检测方法

江泽涛,程留明\*,杨建琛

(桂林电子科技大学广西图像图形处理与智能处理重点实验室,广西桂林 541004)

**摘要:** 正常照度环境下的图像小目标检测具有挑战性,而低照度环境下的图像因低亮度、低对比度、低信噪比等,信息丢失严重,使得小目标特征信息更加弱化,更加难以获取特征等信息,从而导致低照度环境下小目标检测的研究极少. 本文针对这一问题,提出一个低照度小目标检测方法 MC-YOLO. 该方法包含多尺度特征融合增强(Multi-Scale Feature Fusion Enhancement, MFFE)模块、细节特征提取(Detail Feature Extraction, DFE)模块、Neck 模块和 Head 模块. 该方法的第一部分由 MFFE 模块采用可变形卷积和多尺度特征融合来提取低照度环境下的小目标特征,再通过全局平均池化进行全局特征增强,从而使小目标特征信息更加显著;第二部分由 DFE 模块在充分利用上下文信息提取小目标特征的同时保留小目标的位置信息,解决低照度小目标物体细节特征信息易丢失的问题;第三部分由 Neck 模块进行特征提取和多尺度特征融合;最后,在 Head 模块的大分辨率特征图上引入小目标检测层,实现低照度小目标的检测. 实验结果表明,本文方法在低照度小目标检测精度方面具有良好的表现,在自制低照度小目标数据集 LLSOD 上的 mAP 为 83.2%,比目前先进的目标检测方法 YOLOv11 高出 3.6%.

**关键词:** 低照度小目标;多尺度;特征融合;特征增强;细节特征

**基金项目:** 国家自然科学基金(No.62473105);广西自然科学基金重点项目(No.2021GXNSFDA196002);广西图像图形智能处理重点实验室项目(No.GIIP2302, No.GIIP2303, No.GIIP2304);桂林电子科技大学研究生教育创新计划项目(No.2024YCXB09, No.2024YCXs039, No.2024YCXs035)

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2025)07-2229-12

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20241155

## Low-Light Small Target Detection Method Combining Feature Fusion Enhancement and Detail Features

JIANG Ze-tao, CHENG Liu-ming\*, YANG Jian-chen

(Guangxi Key Laboratory of Image and Graphics Processing and Intelligent Processing, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China)

**Abstract:** The detection of small targets in normal illumination conditions is challenging. In low-light environments, images suffer from severe information loss due to low brightness, low contrast, and low signal-to-noise ratio, which further weakens the feature information of small targets, making feature extraction more difficult. As a result, research on small target detection in low-light environments is scarce. To address this issue, this paper proposes a low-light small target detection method, MC-YOLO. The MC-YOLO includes four modules: multi-scale feature fusion enhancement (MFFE) module, detail feature extraction (DFE) module, Neck module and Head module. Firstly, the method uses the MFFE module to extract small target features in low-light environment through deformable convolutions and multi-scale feature fusion and to enhance global features through global average pooling making small target feature information more salient. Next, the DFE module fully utilizes contextual information to extract small target features while preserving the positional information of the small targets, which solves the problem of easy loss of detail feature information of small targets in low-light environments. Then, the neck module performs feature extraction and multi-scale feature fusion. Finally, the head module introduces a small target detection layer on the high-resolution feature map to detect small targets in low-light environments. Experi-

mental results show that this method performs well in the accuracy of low-light small target detection, with the mAP of 83.2% on the self-made low-light small target dataset LLSOD, which is 3.6% higher than the current advanced target detection method YOLOv11.

**Key words:** low-light small targets; multi-scale; feature fusion; feature enhancement; detail features

**Foundation Item(s):** National Nature Science Foundation of China (No.62473105); Nature Science Key Foundation of Guangxi (No.2021GXNSFDA196002); Guangxi Key Laboratory of Image and Graphic Intelligent Processing under Grants (No.GIIP2302, No.GIIP2303, No.GIIP2304); Innovation Project of GUET Graduate Education (No.2024YCXB09, No.2024YCXs039, No.2024YCXs035)

## 1 引言

低照度条件下,图像小目标往往会丧失细节信息,对比度降低,甚至与背景融为一体,致使小目标特征更加弱化,更加难以提取.现有正常照度环境下的小目标检测方法无法很好地检测低照度环境下的小目标物体,且目前低照度小目标检测的研究极少.因此,如何有效地在低照度环境下检测小目标成为需要解决的重点问题之一.

早期的研究主要集中在正常照度条件下的小目标检测方法.Kang等人<sup>[1]</sup>发现SSD小目标检测中传统的IoU算法存在匹配机制的问题,提出对齐匹配算法,通过算法反复过滤非目标框,得到最匹配的默认框以确保小对象检测的准确度,但是会增加训练和推理的时间和计算资源消耗.Liu等人<sup>[2]</sup>在主干网中加入注意力模块,在通道注意模块与空间注意模块的顺序连接中减少非目标特征的通道数放大剩余神经元的占比,以此放大目标特征的权重,将BiFPN(Bidirectional Feature Pyramid Network)运用在特征融合部分,使得浅层特征能融合高层的语义信息,捕捉更多细节信息,在小目标检测中效果更好,但该方法只适用于机场工作人员检测,无法泛化到其他复杂的环境中.Betti等人<sup>[3]</sup>提出YOLO-S,通过一个重塑穿透层促进跨网络的特征重用和联合低级别的位置信息与更有意义的高级别信息,但是该方法的主干网络采用的是Darknet20,并未考虑信息传递过程中小目标信息丢失的问题.Hu等人<sup>[4]</sup>提出Efficient-Lightweight YOLO,通过设计有效的空间金字塔池,增加图像中小目标特征表示,但是需要大量的高质量标注数据,如何在低照度小目标检测场景下获取大量高质量标注数据是一个难题.Qi等人<sup>[5]</sup>提出SODNet,通过多尺度感受野自适应地获取相应的空间信息,从而提高对小目标特征的提取能力,但其针对小目标优化的设计导致对大目标的检测能力下降,无法适用于更加广泛的场景.Ji等人<sup>[6]</sup>提出YOLO-TLA,通过设计全局关注注意力,将通道信息与全局信息相结合以创建加权特征图,突出小目标对象的属性,同时有效地忽略不相关的细节,但带来了更大的参数量和计算成本.

随着无人机、智能监控、智能驾驶等领域的快速发展,低照度环境下的小目标物体检测的需求日益增加,这使得探索在低照度条件下提高小目标检测效果的方法变得尤为紧迫和重要.而正常照度下的小目标检测方法在检测低照度图像小目标时会产生以下问题:(1)现有模型需要大量图像数据支持,而低照度环境下的图像难以对小目标进行高质量的标注,导致目前没有低照度小目标数据集,使得模型缺少数据支持;(2)正常照度下的小目标检测方法无法解决低照度环境造成图像小目标信息不完整、特征不显著的问题;(3)现有小目标检测方法在提取小目标特征时会造成小目标细节信息丢失的问题,使得小目标信息更加稀少.而现有的低照度小目标检测方法极少.Wang等人<sup>[7]</sup>利用合成数据集提出一种结合低照度图像增强和小目标检测的交通标志检测方法,但该方法无法满足正常生活场景检测的需求.Dai等人<sup>[8]</sup>提出SwinVision,使用低照度图像增强数据集LOL训练增强模块STLE,然后从VisDrone2019-DET数据集中提取夜间图像进行整个网络的训练和测试,但是提取的夜间图像数据较少,无法支撑端到端检测网络的训练.针对上述问题,本文提出MC-YOLO方法.该方法包含多尺度特征融合增强(MFFE)模块、细节特征提取(DFE)模块、Neck模块和检测头Head模块.MFFE模块首先全局增强低照度目标特征,初步增强后可获得包含更加显著小目标特征信息的特征图 $P_1$ ;然后DFE模块对 $P_1$ 进行特征提取并保留小目标的位置等细节信息,获得具有深层语义信息和细节信息的低照度小目标特征图 $P_2$ ;再将特征图 $P_2$ 输入Neck模块进行特征提取和多尺度特征融合;最后将Neck模块融合后的特征图送入Head模块检测,预测回归目标检测框的位置、大小和目标类别.本方法实现了低照度环境下图像小目标检测,其创新及贡献如下.

(1)提出了低照度小目标检测方法MC-YOLO,该方法包含多尺度特征融合增强(MFFE)模块、细节特征提取(DFE)模块、Neck模块和Head模块,并针对小目标物体引入了一个小目标检测头S-head,实现了低照度环境下小目标物体的检测.

(2)提出多尺度特征融合增强MFFE模块,解决了

低照度图像小目标物体特征不显著,难以提取的问题,通过拼接融合多级特征,避免小目标信息丢失,并引入全局上下文增强 GCE 块对融合特征进行全局增强,使得低照度图像特征更加显著,易于提取。

(3)提出了细节特征提取 DFE 模块,该模块充分利用上下文信息,使模型更加专注于小目标区域,解决了低照度图像小目标物体细节特征信息易丢失的问题。

(4)为了解决低照度小目标检测领域缺乏数据集的问题,本文制作了 LLSOD 低照度小目标数据集,提供数据支持。

## 2 相关工作

本文针对低照度小目标检测开展研究,涉及低照度图像增强和注意力机制。

### 2.1 低照度图像增强

低照度图像普遍存在整体亮度不足、对比度较低等问题,人眼难以获取图像信息,低照度图像增强算法可以有效解决这些视觉难题。早期的尝试方法主要是使用基于直方图均衡化<sup>[9]</sup>的方法或者是基于 Retinex 理论的方法<sup>[10]</sup>来调整图像的照明程度。Arici 等人<sup>[11]</sup>提出了一种全局直方图均衡化算法,通过低照度图像的直方图来扩展像素强度的动态范围,可以有效地提高图像的对比度。Guo 等人<sup>[12]</sup>提出了 LIME,选择各像素通道中的极大值来初始化输入图像的照明图,并施加一个先验结构对初始照明图进行细化,然后通过 Retinex 原理合成增强图像。现代基于深度学习的图像增强方法则是构建卷积神经网络,从特征层面对低照度图像进行增强操作。Yang 等人<sup>[13]</sup>设计了一种用于协作低光图像增强的隐式神经表示,从语义、文本和颜色这 3 个角度对两种模态监督训练,以增加约束条件,加速寻找最优解并得到感知更佳的结果。这些增强方法虽然有效地提高了低照度图像的视觉效果,但是并未充分考虑增强过程中小目标信息丢失的问题,因此这些方法对于低照度小目标的增强效果并不好,不能直接用于低照度小目标的增强。

### 2.2 注意力机制

注意力机制根据输入图像中各部分区域对输出结果的影响程度,分配不同的权重,以突出需要强调的特征信息,并与原始输入的特征图进行相乘,完成自适应特征加权<sup>[14]</sup>,有助于计算机视觉的各项任务,如图像分类、图像分割、目标检测和图像增强等的开展。在过去几年中,研究者们提出了许多新颖有效的注意力模块。Hou 等人<sup>[15]</sup>提出了一种协调注意力模块,结合了通道注意力的优势,并在通道维度中嵌入位置信息,利用精准的坐标信息来捕捉特征图中对象之间长距离依赖的关系。Park 等人<sup>[16]</sup>提出的瓶颈注意力模块,将通道维度与空间维度上的注意力相结合,以增强网络的特征学

习能力。Azad 等人<sup>[17]</sup>提出了一种可变形大核注意力,该模块采用大卷积核来充分理解上下文,灵活地扭曲采样网格,使模型能够适应不同的数据模式。上述部分注意力机制虽然同时考虑了空间和通道信息,但是都把通道注意力和空间注意力简单地串联使用,这会导致在使用空间注意力时丢失一部分相关通道信息。

## 3 本文方法

### 3.1 设计思想和总体框架

本文提出了一种低照度小目标检测方法 MC-YOLO,该方法包含多尺度特征融合增强(MFFE)模块、细节特征提取(DFE)模块和基于 YOLOv5s 的颈部和改进的头部,网络结构如图 1 所示。针对低照度环境带来的低亮度、低对比度、暗区域特征不显著以及小目标物体信息易丢失的问题,为了使模型更容易提取到小目标特征,同时保证小目标信息不丢失,本文的设计思想是:首先通过 MFFE 模块进行全局特征增强,针对小目标物体,利用可变形卷积 TDS 块对低照度图像进行 3 次下采样,形成多尺度特征,并使用 SPL 和 Concat 操作把浅层特征信息和深层语义信息融合,再使用 GCE 块为每个图像通道乘以分配的亮度值进行全局特征增强,获得包含显著小目标特征信息的特征图  $P_1$ ;其次由细节特征提取模块 DFE 通过堆 3C\_3、3C\_6、3C\_9 特征提取块对特征图  $P_1$  提取特征,使用通道和空间并行注意力功能块 CSCE 更加关注小目标区域,在提取不同尺度特征信息的同时,充分利用上下文信息来保留小目标的位置等细节信息,从而获得具有丰富小目标特征信息的特征图  $P_2$ ;然后通过 Neck 模块,对特征图  $P_2$  进行深层特征提取以及多尺度特征融合,保证浅层特征能与深层特征更好地融合;最后为了提高对小目标的检测能力,在原有 3 个检测头的基础上,在浅层特征层添加了 1 个小目标检测头 S-head,分别对 Neck 模块输出的 4 种大小的特征图进行预测目标框的位置、宽高和对应的类别,并对这些预测结果通过非极大值抑制(NMS)算法消除冗余边界框,选择最具置信度的边界框,得到检测结果。本文提出的模型具备直接检测低照度图像中小目标物体的能力,并输出带有类别和目标框信息的图像。

### 3.2 MFFE 模块

多尺度特征融合增强 MFFE 模块是针对低照度环境中图像亮度低、对比度低以及暗区域造成小目标物体的特征信息不足,从而导致小目标特征难以提取的情况进行特征增强。该模块首先通过可变形卷积实现 3 次下采样形成多尺度特征信息,然后利用上采样和一个向下分裂层(SPL)融合不同尺度的特征信息,最后对融合后的特征信息进行全局特征增强。本文设计的 MFFE 模块结构细节如图 2 所示。

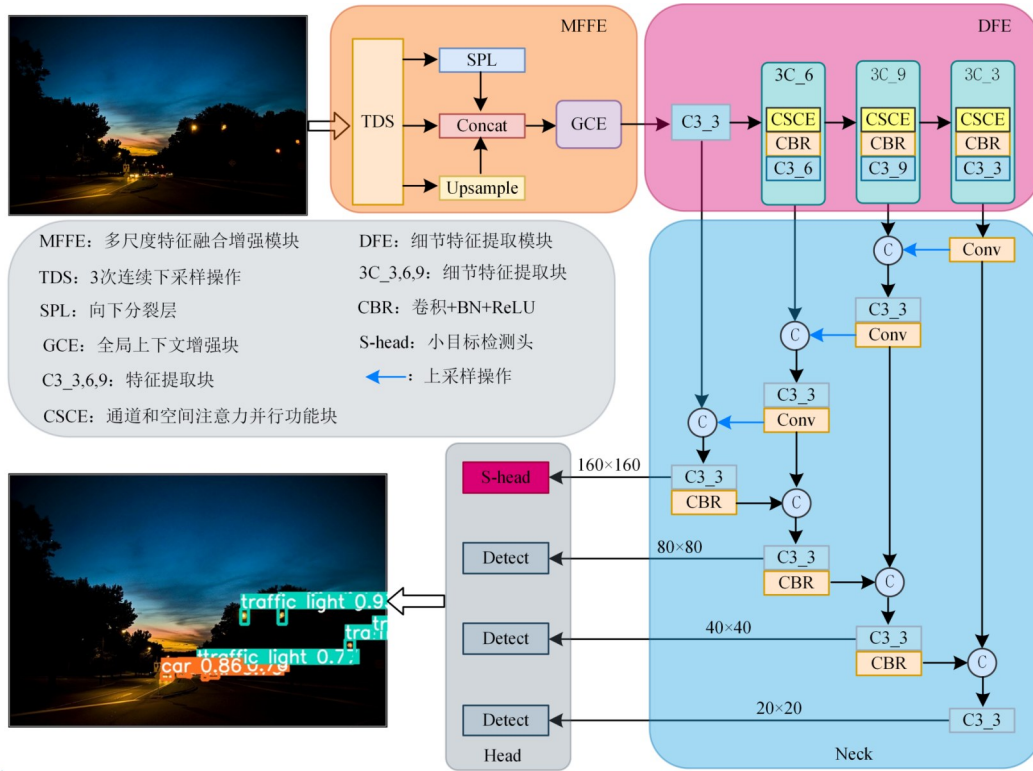


图1 MC-YOLO低亮度小目标检测方法结构图

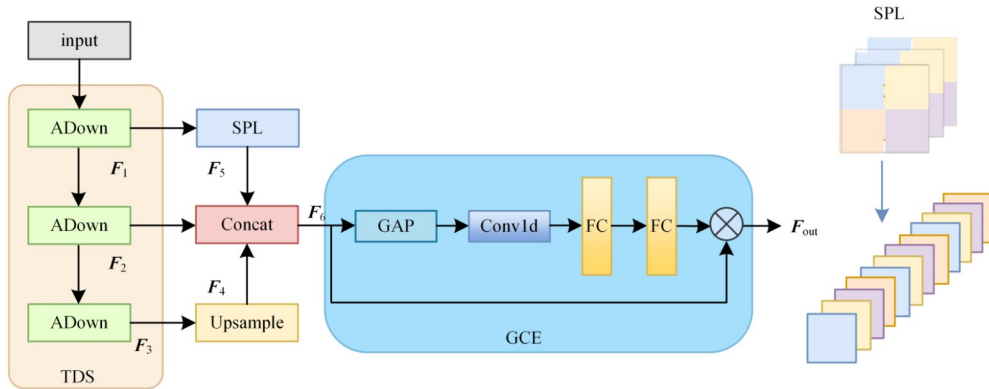


图2 多尺度特征融合增强MFEE模块结构图

首先通过3次 ADown 模块进行下采样操作<sup>[18]</sup>分别得到具有深层语义特征的特征图  $F_i (i=1, 2, 3)$ . ADown 模块采用可变形卷积,具有一定的可学习能力,能根据目标的形状动态调整卷积核大小,适应不同形状的小目标,可以更好地提取小目标特征,公式表示如下:

$$\begin{aligned} F_1 &= \text{ADown}(x) \\ F_2 &= \text{ADown}(F_1) \\ F_3 &= \text{ADown}(F_2) \end{aligned} \quad (1)$$

其中,  $x$  是原始输入图像, ADown 是下采样模块,  $F_i$  是下采样得到的特征图.

然后  $F_3$  通过一次上采样操作得到具有深层语义特征的多尺度特征图  $F_4$ , 再将第一次下采样后的特征图

$F_1$  经过一个向下分裂层可以得到具有浅层位置特征的多尺度特征图  $F_5$ , 并将多尺度特征图  $F_2, F_4, F_5$  进行拼接. 向下分裂层 SPL 可以将浅层具有精确位置信息的细粒度特征图与来自上采样操作具有更深层语义信息的深层特征图更好地结合起来, 使得融合后的特征图能具有丰富的语义信息, 又带有精确的位置信息. 公式表示如下:

$$\begin{aligned} F_4 &= \text{Upsample}(F_3) \\ F_5 &= \text{SPL}(F_1) \\ F_6 &= \text{Concat}(F_2, F_4, F_5) \end{aligned} \quad (2)$$

其中, Upsample 是上采样操作, SPL 是向下分裂操作, Concat 是通道方向串接操作.

最后引入 GCE 块进一步调整融合特征的全局亮度. GCE 块首先通过全局平均池化提取融合特征的全局上下文信息,然后经过一维卷积和两层全连接层生成对应通道的亮度值,每个通道乘以对应的亮度值以实现全局特征进行增强,公式表示如下:

$$F_{\text{out}} = F_6 \otimes \text{MLP}(\text{Conv1d}(\text{GAP}(F_6))) \quad (3)$$

其中,  $\otimes$  是逐元素相乘操作, MLP 是两层全连接操作, Conv1d 是一维卷积, GAP 是全局平均池化操作.

### 3.3 DFE 模块

低照度小目标物体的细节特征信息在特征提取与下采样操作后易丢失,针对这一问题,本文设计了细节特征提取模块 DFE,使模型能够充分利用上下文信息,更加关注小目标区域,同时保留小目标的位置等细节信息,提高小目标的检测能力. DFE 模块由 1 个 C3\_3 特征提取块和 3 个细节特征提取块(3C\_6、3C\_9 和 3C\_3)

组成. 3C\_6、3C\_9 和 3C\_3 的区别是特征提取块中使用的残差块 ResX 的数量不同,可以统称为 3C\_X,其结构如图 3 所示. 3C\_X 块由 CSCE、CBR 和 C3\_X 块构成. 首先经过通道和空间注意力并行功能块 CSCE 充分利用上下文信息进行空间细节信息保留,并使模型更加关注小目标区域. 该功能块有 3 个分支,最上层分支分别采用自适应平均池化和自适应最大池化处理输入的特征图,再通过两层带有激活函数的全连接层,把两个结果相加. 该分支目的是使模型能够捕捉到不同尺度上的特征,两种池化特征的融合,有助于模型整合来自不同池化策略的信息,使模型能更好地理解 and 识别小目标物体. 该分支的公式可以表示如下:

$$C(F) = \sigma(\text{MLP}(\text{AAP}(F))) + \sigma(\text{MLP}(\text{AMP}(F))) \quad (4)$$

其中,  $\sigma$  是 Sigmoid, AAP 是自适应平均池化, AMP 是自适应最大池化.

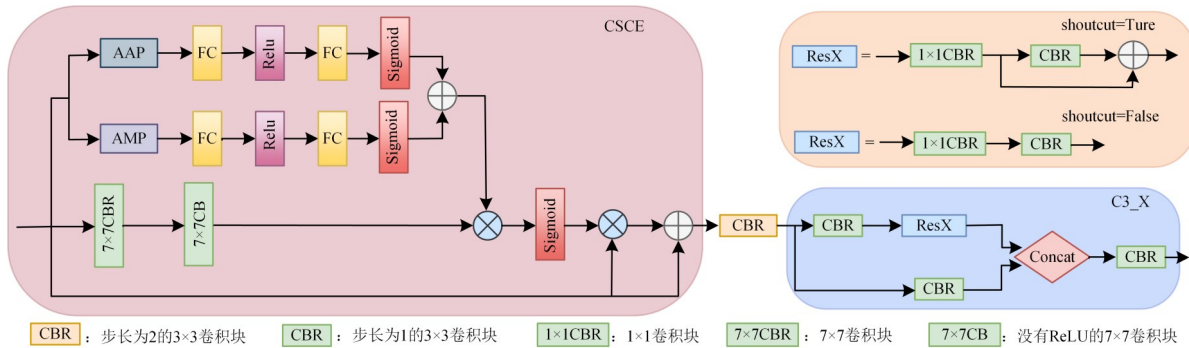


图3 3C\_X块结构图

中层分支采用 2 个  $7 \times 7$  卷积块,目的是通过大感受野捕获大范围的上下文信息,使模型更关注小目标区域,从而提高模型对小目标物体的检测和定位准确性,降低误检率. 该分支的公式可以表示如下:

$$S(F) = \text{BN}(f_1^{7 \times 7}(R(\text{BN}(f_0^{7 \times 7}(F)))))) \quad (5)$$

其中, BN 是 Batch Normalization,  $7 \times 7$  是卷积核大小, R 是 ReLU 激活函数.

下层分支暂不做处理. 中层分支与最上层分支相乘,通过 Sigmoid 后与下层分支相乘,相乘结果再和下层分支相加,目的是在提取小目标特征的同时保留更多的位置信息,提高检测效果和定位精度. 对于输入的特征图  $F$ , CSCE 模块可以得到一个 3D 的增强特征图  $\text{CSCE}(F)$ ,具体公式表示如下:

$$\text{CSCE}(F) = F \oplus (F \otimes (\sigma(C(F) \otimes S(F)))) \quad (6)$$

其中,  $\oplus$  是逐元素相加,  $\otimes$  是逐元素相乘,  $\sigma$  是 Sigmoid.

其次经过卷积块 CBR 进行下采样,形成多尺度特征信息. 最后经过特征提取块 C3\_X,得到具有丰富细节信息且更加关注小目标区域的特征图,提高

模型的小目标检测能力. 3C\_X 功能块的计算公式表示如下:

$$3C_X(F) = C3_X(\text{CBR}(\text{CSCE}(F))) \quad (7)$$

其中,  $3C_X(\cdot)$  是特征提取功能块,  $\text{CBR}(\cdot)$  是卷积块,  $\text{CSCE}(\cdot)$  是通道与空间并行注意力功能块.

### 3.4 Neck 模块与 Head 模块

本文的 Neck 模块采用了 YOLOv5s 的 Neck 结构. 为了提升本文模型在小目标检测方面的性能,本文对 YOLOv5s 的 Head 模块进行改进. 具体来说,本文对特征图的尺度和锚框尺寸进行了精细调整. 首先应用 K-means 聚类算法,对预定义的锚框尺寸进行重新校准,确保锚框能够更好地匹配 LLSOD 数据集中实际目标的尺寸;然后在 Head 模块引入了 1 个小目标检测层 S-head,在 YOLOv5s 的 Neck 网络中执行上采样操作,生成了 1 个分辨率为  $160 \text{ 像素} \times 160 \text{ 像素}$ 、拥有 256 个通道的特征图,再将这个特征图与 DFE 模块的第一层输出进行拼接,在拼接后的特征图上引入 1 个小目标检测层 S-head,用于小目标的分类和检测. 本文的 Neck 和 Head 模块设计巧妙地在不同尺度下生成了几个特征图,每个特征图对应 Head 模

块里的一个检测头,即一组特定的锚框尺寸.这种设计策略使得模型能够在较大的特征图上有效地检测到小目标物体,在较小的特征图上准确地识别出较大物体.通过这种方式,显著增强了模型对图像中小目标物体的表示和检测能力.由于Head模块需要对图像目标进行分类和定位,故使用了3种损失函数:分类损失、定位损失和置信度损失.分类损失函数的具体公式表示如下:

$$L_{\text{class}} = -(1 - p_i)^{\gamma} \log(p_i) \quad (8)$$

真实目标预测概率公式 $p_i$ 定义如下:

$$p_i = \begin{cases} p, & y = 1 \\ 1 - p, & y \neq 1 \end{cases} \quad (9)$$

定位损失函数公式表示如下:

$$L_{\text{local}} = 1 - \left( \text{IoU} - \frac{\rho^2}{c^2} - \alpha v \right) \quad (10)$$

$$\alpha = \frac{v}{(1 - \text{IoU}) + v} \quad (11)$$

其中, IoU 是预测框和真实框的交并比,  $\rho$  为框 A 和框 B 的中心点距离,  $c$  为框 A 和框 B 的最小包围矩形的对角线长度,  $v$  为框 A、框 B 的宽高比相似度,  $\alpha$  为  $v$  的影响因子.

置信度损失函数公式表示如下:

$$L_{\text{obj}} = -(y \log(x) + (1 - y) \log(1 - x)) \quad (12)$$

其中,  $x$  是模型预测结果,  $y$  是真实标签.

在式(8)中,当 $p_i$ 趋于1时,分类正确且是易分类样本,调制系数 $(1 - p_i)^{\gamma}$ 趋于0,对于总的损失贡献很小.当 $p_i < 0.5$ 时,为困难样本,即难分类样本, $(1 - p_i)^{\gamma}$ 趋于1,对于总的损失贡献大,这样的话,困难样本的权重就相对提升了很多.对于检测模型来说,大尺寸的目标预测结果 $p_i$ 一般都比较大,但小目标的预测结果 $p_i$ 一般都比较小.因此,通过Focal Loss损失函数的计算,小目标可以获得较大的损失值,提高整个模型对于小目标的预测能力.

## 4 实验结果与分析

本节主要在LLSOD低照度小目标图像数据集上进行实验.本节重点讨论5个部分:自制低照度小目标数据集LLSOD、评价指标、实现细节、对比实验、消融实验.本文提出的算法主要针对低照度环境下的小目标物体检测,因此以LLSOD数据集的实验结果作为主要的评价标准.

### 4.1 LLSOD数据集

低照度小目标检测领域目前存在的研究不多,也不存在已经公开的公共数据集,这给本领域的研究带来了不便,因此本文制作了低照度小目标数据集LLSOD.该数据集的原始数据图像来自MS COCO2017数据集和低照度图像数据集ExDark,本文从上述2个数据集中挑选出1 640张低照度图像,然后对这1 640张

低照度图像中的小目标物体通过Labelimg工具进行手工标注,形成初始的低照度小目标图像数据集.然后对上述初始数据集通过Gamma变换、旋转、镜像、二次采样等数据扩充方法进行处理,最终形成拥有6 658张图像数据的LLSOD数据集.

该数据集有部分低照度图像选自MS COCO2017数据集,因此数据集的类别和MS COCO2017相同,共有80个类别.数据集的亮度分布和目标尺寸分布如图4和图5所示.图4是该数据集的亮度分布图,横坐标是数据集中每张图像的平均灰度值,纵坐标是该灰度值出现的次数.从图4中可以看出,LLSOD数据集中的图像平均灰度值几乎都在100以下,大部分集中在50左右,可以说明数据集中的图像亮度是比较低的.图5则是该数据集中目标尺寸分布图,横坐标表示目标边界框面积的开方,在0.05以下表示是小目标,纵坐标是目标出现次数.从图5中可以看出,LLSOD数据集中小目标物体的数量占比最高,大物体数量占比很小,足以支持本文方法的训练和验证.

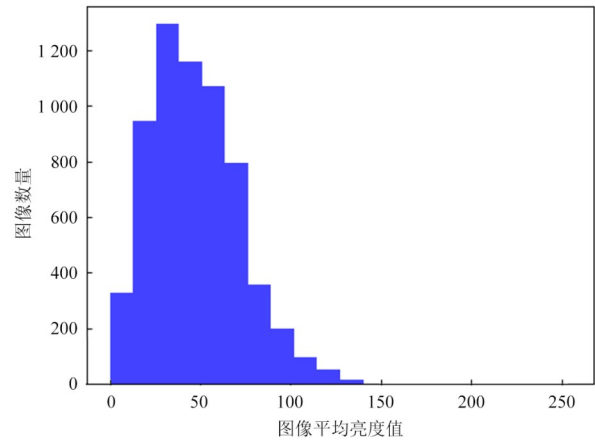


图4 LLSOD数据集亮度分布图

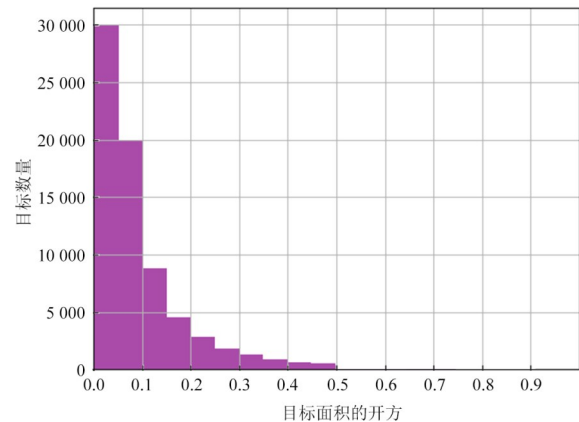


图5 LLSOD数据集目标尺寸分布图

实验开始之前,本文按照 7:2:1 的比例对 LLSOD 数据集进行了划分,形成训练集 4 649 张、测试集 1 492 张和验证集 517 张.

#### 4.2 评价指标

本文采用平均精度 (mean Average Precision, mAP)、查准率 (Precision)、召回率 (Recall) 作为衡量本文方法在低照度小目标检测性能上的评价指标. mAP50 表示当 IoU 阈值为 0.50 时所有目标类别 AP 值的平均值. mAP50~95 表示 IoU 阈值由 0.50~0.95, 以 0.05 为步长计算的 AP 值的平均值,可以提供关于模型在不同 IoU 范围内表现得更全面的信息. 上述评价指标的计算公式表示如下:

$$P = \frac{TP}{TP + FP} \quad (13)$$

$$R = \frac{TP}{TP + FN} \quad (14)$$

$$AP = \int_0^1 P dR \quad (15)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (16)$$

其中, TP、FP 和 FN 分别表示正确检测框、误检框和漏检框的数量,  $P$  表示查准率,  $R$  表示召回率, AP 值表示一个类别的检测平均精度.

#### 4.3 实现细节

本文实验是在 LLSOD 数据集的训练集上训练 300 轮, 在其验证集上评估检测性能, 并在测试集上验证了本文模型的泛化能力. 本文的模型采用随机梯度下降优化器以 0.01 的初始学习率训练, 并使用余弦退火方式将学习率衰减到 0.000 1. 本文的实验设备为 Tesla P40 GPU, 运行环境为 Python3.7 和 PyTorch1.8.

#### 4.4 对比实验

本节在 LLSOD 数据集上对各算法性能进行对比, 结果如表 1 所示, 其中最优结果以加粗字体表示. 本文算法在测试集上取得了 83.2% 的 mAP50 和 58.4% 的 mAP50~95, 在检测精度 mAP 上比目前先进的 YOLOv11 等算法表现更好. 图 6 展示了本文提出的低照度小目标检测算法与其他目标检测算法的比较, 其中第 1 行是标签对应的真实框, 最后一行是本文方法的检测结果. 与其他算法相比, 本文算法使用了多尺度特征融合增强以及带有更加关注小目标区域的注意力机制的细节特征提取, 在低照度物体的特征提取和处理上对小目标特征更加友好, 实现了对低照度区域小目标物体的检测任务, 从而达到了更高的 mAP. 从图 6 第 2 行中可以看出, YOLOv5s 检测小目标的能力较强, 能够检测一些较小的目标, 如第 1 张和第 2 张图像中的交通灯、第 3 张图像中远处的部分绵羊以及第 5 张图像中部分海面上的人, 这也是本方法选择 YOLOv5s

作为基准的原因. 但是, 第 1 张图像中被遮挡的汽车和第 5 张图像中海面上更小的船却无法被检测到, 第 4 张图像中低照度区域里的人也无法被检测到. 第 3 行是 Faster RCNN 的检测结果, 可以看出该方法对低照度和较远的小目标的检测能力一般, 第 1 张图像中被交通灯遮挡的汽车被误检, 第 3 张和第 5 张图像中较远的小目标都未能被检测, 第 4 张图像低照度区域里的人也未能全部被检测. 第 4 行是 RetinaNet 的检测结果, 可以看出该方法对低照度和小目标的检测能力较差, 第 1 张图像中被遮挡的汽车和第 3 张图像的绵羊被误检, 第 2 张、第 4 张、第 5 张图像中的小目标被大量漏检. 第 5 行是 FCOS 的检测结果, 可以看出该方法对低照度和小目标的检测能力很差, 基本检测不到低照度下的目标和小目标. 第 6 行展示了 PP-YOLOE 的检测结果, 可以看出 PP-YOLOE 对低照度和较远的小目标的检测能力较弱, 比如第 1 张、第 2 张、第 3 张和第 5 张图像中远处的小目标以及第 4 张图像中低照度区域里的人都无法被检测. 第 7 行结果表明 HDETR 对低照度区域的目标检测能力还不足, 会产生大量漏检和误检的现象, 比如第 1 张图像中的汽车、第 2 张图像中的交通灯和船、第 3 张图像中远处的绵羊以及第 5 张图像中较远的船都没能被有效检测. 从第 8 行可以看出, YOLOv8 对低照度和小目标的检测能力较差, 第 1 张和第 2 张图像中的交通灯、第 3 张图像中的羊群以及第 5 张图像中海面上的船未能被完全检测, 第 4 张图像中低照度区域里的人也未能被检测. 从第 9 行可以看出, YOLOv9 对低照度区域的小目标检测能力较差, 比如第 1 张图像中的汽车、第 3 张图像远处的绵羊以及第 4 张图像中低照度区域里的人无法被很好地检测. YOLOv10 对低照度区域的目标检测能力较差, 第 1 张图像中的汽车和交通灯均未能被检测, 第 5 张图像中海面上的人和船也未能被检测. YOLOv11 对低照度区域的小目标检测能力较差, 比如第 3 张图像中的羊群、第 4 张图像低照度区域里的人以及第

表 1 LLSOD 数据集上各算法性能对比结果

方法	$P$	$R$	mAP50/%	mAP50~95/%	Parameters/ $10^6$
YOLOv5s (2020)	0.828	0.64	70.9	47	7.2
Faster RCNN (2017) <sup>[19]</sup>	0.619	0.499	60.1	42.1	41.2
RetinaNet (2020) <sup>[20]</sup>	0.711	0.576	61.9	54.1	35.7
FCOS (2022) <sup>[21]</sup>	0.466	0.337	40.8	26.4	31.9
PP-YOLOE (2022) <sup>[22]</sup>	0.862	0.683	77.3	55.9	52.2
HDETR (2023) <sup>[23]</sup>	0.836	0.673	74.4	54.1	46.1
YOLOv8(2023)	0.858	0.681	73.8	55.4	25.9
YOLOv9(2024) <sup>[18]</sup>	0.897	0.729	79.7	56.3	25.5
YOLOv10(2024) <sup>[24]</sup>	0.867	0.72	79.1	55.3	24.4
YOLOv11(2024)	<b>0.907</b>	0.74	79.6	56.4	25.3
MC-YOLO (ours)	0.892	<b>0.790</b>	<b>83.2</b>	<b>58.4</b>	22.1

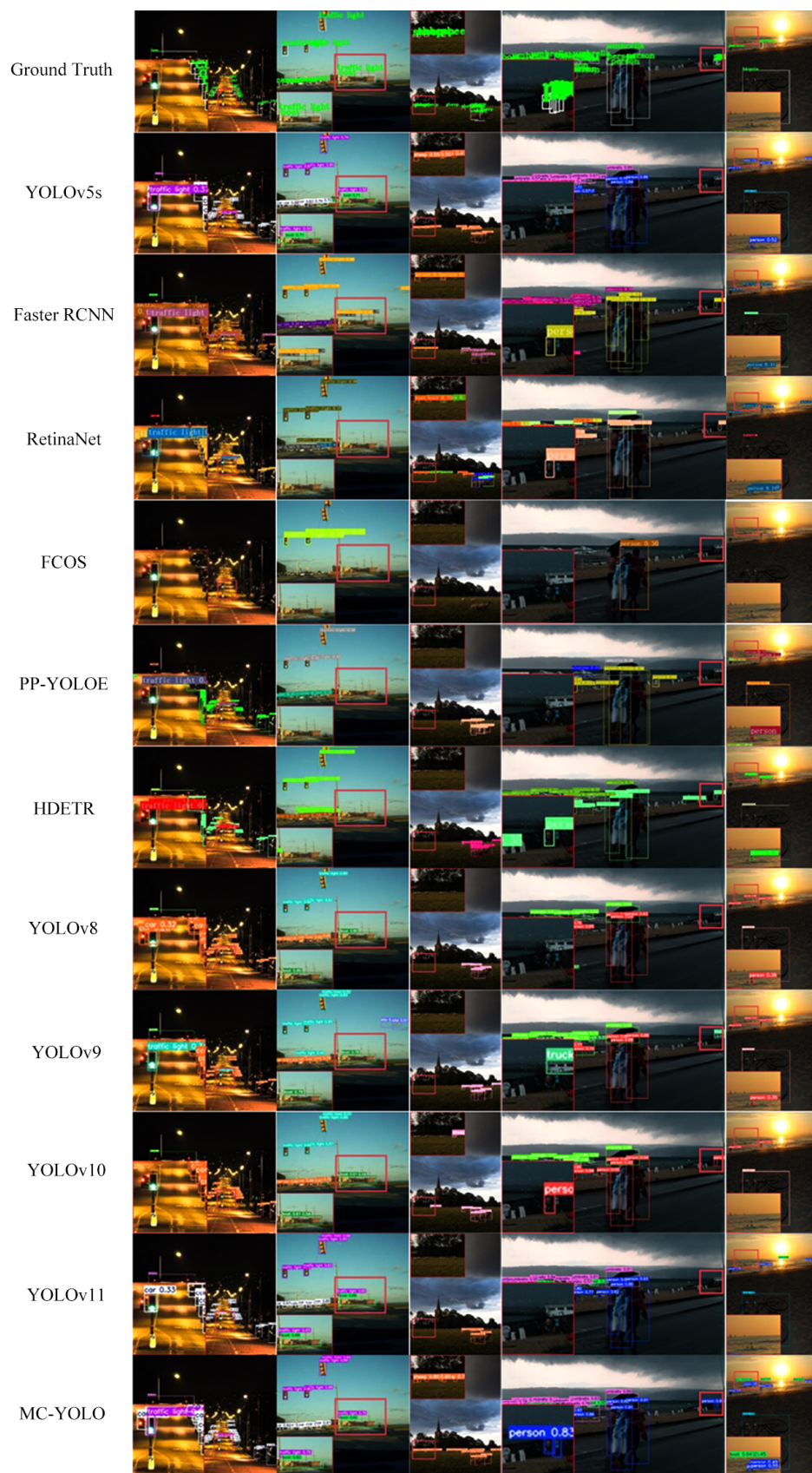


图6 各算法可视化结果对比图

5张图像中海面上的船和人均无法被很好地检测.最后一行是本文方法的检测结果,可以看出本文方法能很好地检测到低照度区域里的小目标物体,这也是检测精度提升的直接原因.综上所述,对低照度图像中小目标物体的漏检和误检,导致现有目标检测算法无法较好地应对低照度区域里小目标的检测任务,而本文提出的方法能够较好地检测到低照度图像中的小目标物体,实现了对低照度环境下的小目标物体的检测,从而实现了检测精度的提升.

#### 4.5 消融实验

为了进一步探讨本文提出算法的有效性,本节对多尺度特征融合增强 MFFE 模块、细节特征提取 DFE 模块以及小目标检测层 S-head 进行了消融实验,并对各个模块对实验结果的影响进行了分析.通过删减本文所提出的模

块,得到 MC-YOLO\_del\_DFE\_S-head、MC-YOLO\_del\_MFFE\_S-head、MC-YOLO\_del\_DFE\_MFFE、MC-YOLO\_del\_S-head、MC-YOLO\_del\_DFE 和 MC-YOLO\_del\_MFFE 这 6 种算法,去掉 MFFE 和 DFE 模块后仅有 Neck 和 Head 模块无法完成检测任务,因此本文设计的 MC-YOLO\_del\_DFE\_MFFE 算法是在 YOLOv5s 网络的基础上添加小目标检测层 S-head 以验证 S-head 的效果,6 种算法的实验结果如表 2 所示,其中最优结果以加粗字体表示.从表 2 中可以看出,3 个模块在单独使用时都能实现不同程度的提升效果,其中单独使用 MFFE 模块的效果最好,印证了特征增强对提升低照度小目标检测能力的帮助.通过对任意两个模块组合的实验结果对比可以发现,MFFE 模块对模型性能的影响较大,且当同时使用 MFFE 模块和 DFE 模块时模型的性能更强.从最后一行可以看出同时使用 3 个模块时模型的性能达到最佳.

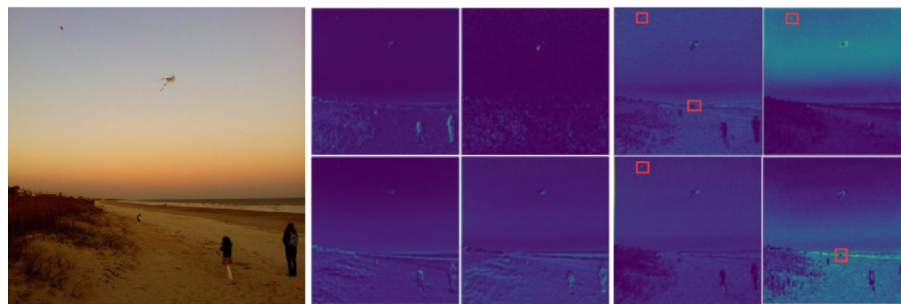
表 2 本文提出模块的消融实验结果

策略	MFFE	DFE	S-head	<i>P</i>	<i>R</i>	mAP50/%	mAP50~95/%
Yolov5s	×	×	×	0.828	0.64	70.9	47
MC-YOLO_del_DFE_S-head	√	×	×	0.873	0.752	80.3	54.2
MC-YOLO_del_MFFE_S-head	×	√	×	0.868	0.653	74.4	50
MC-YOLO_del_DFE_MFFE	×	×	√	0.840	0.685	76.6	47.9
MC-YOLO_del_S-head	√	√	×	0.836	0.714	81.1	<b>58.4</b>
MC-YOLO_del_DFE	√	×	√	0.831	<b>0.791</b>	81.5	53.7
MC-YOLO_del_MFFE	×	√	√	0.838	0.689	77.1	50.9
MC-YOLO	√	√	√	<b>0.892</b>	0.790	<b>83.2</b>	<b>58.4</b>

##### 4.5.1 MFFE 模块有效性

本节添加多尺度特征融合增强模块后的模型与基线模型相比,mAP50提高了9.4%,mAP50~95提高了7.2%.该模块采用了可变形卷积代替传统的卷积进行下采样操作,更加适合小目标,且使用向下分裂层保证不损失位置信息的情况下与深层语义信息进行融合,再经过增强全局特征使小目标特征更加显著,从而提高了模型的检测能力.经过多尺度特征融合增强模块与不添加多尺度特

征融合增强模块的特征图可视化结果如图 7 所示.图 7(a)是输入图像,为了便于观察特征图可视化的结果,选取的是亮度较高的图像作为输入图像.图 7(b)是添加 MFFE 模块前的特征图可视化结果,与输入图像相比,远处的人和风筝的特征基本丢失.而图 7(c)是添加 MFFE 模块后的特征图可视化结果,该模块采用了自适应可变形卷积和向下分裂层以及特征增强,因此小目标特征保留得比较完整,这也直接证明了 MFFE 模块的有效性.



(a) 原图

(b) 未添加 MFFE 模块的可视化图 (c) 添加 MFFE 模块后的可视化图

图 7 不添加与添加 MFFE 模块后的特征图可视化

本节对 MFFE 模块有效的原因进行了探究,对 MFFE 模块中的多尺度特征融合和全局上下文增强 GCE

块进行了消融实验,实验结果如表 3 所示.从表 3 中可以看出,多尺度特征融合和 GCE 块都能给模型带来不同程

度的涨点,而由于多尺度特征融合能够减少小目标信息的丢失,对提升召回率 $R$ 和 $mAP50$ 的效果更好.

#### 4.5.2 DFE 模块有效性

本节添加细节特征提取模块后的模型与基线模型相

表3 MFFE 模块的消融实验结果

策略	$P$	$R$	$mAP50/\%$	$mAP50\sim95/\%$
MC-YOLO_del_MFFE	0.838	0.689	77.1	50.9
MC-YOLO_del_MFFE + 多尺度特征融合	0.883	0.787	83.1	57.6
MC-YOLO_del_MFFE + GCE	0.879	0.769	82.4	57.1
MC-YOLO	0.892	0.790	83.2	58.4

比, $mAP50$ 提高了3.5%, $mAP50\sim95$ 提高了3.0%. 该模块通过并行使用通道注意力和空间注意力,在关注小目标区域的同时,保留更多的位置信息,提高低照度小目标的检测精度. 添加细节特征提取模块与不添加细节特征提取模块的特征图可视化结果如图8所示. 图8展示了在采用MFFE模块的基础上对是否添加DFE模块的特征图可视化结果,图8(b)是不采用DFE模块的特征图可视化结果,从结果可以看出有一部分小目标特征已经丢失,而图8(c)是采用DFE模块后的特征图可视化结果,从可视化结果可以看出添加DFE模块后小目标特征保留得比较完整. 因此,本文提出的DFE模块对提升低照度小目标物体的检测能力是有效的.



(a) 原图

(b) 未添加DFE模块的可视化图

(c) 添加DFE模块后的可视化图

图8 不添加与添加DFE模块后的特征图可视化

#### 4.6 方法性能验证

本节在公开的VisDrone2019-DET无人机视角小目标数据集上验证了本文方法的性能,与目前最先进的小目标检测方法相比具有竞争力,实验结果如表4所示.

本文方法在该数据集上取得了49.1%的 $mAP50$ 和29%的 $mAP50\sim95$ . 与NE-CDMNet方法相比,本文方法的 $mAP50$ 存在2.3%的差异,但 $mAP50\sim59$ 高出1.7%. 与 $\alpha$ S-YOLO方法相比,本文方法的 $mAP50$ 高出0.8%. 本文方法与VisDrone2019-DET数据集上最先进的方法性能相当,这表明本文方法具有很强的泛化能力,且本文方法具有更少的参数量. 图9展示了各检测方法的可视化结果. 从可视化结果图像最后3行可以看出,虽然本文所提方法在正常照度环境下检测小目标的能力与 $\alpha$ S-YOLO、NE-CDMNet方法相当,但在低照度环境下本文所提方法的检测性能更佳,足以验证本文所提模型的泛化性.

## 5 结论

本文针对低照度小目标检测提出了一种结合特征融合增强和细节特征的低照度小目标检测方法MC-YOLO. 该方法为了解决低照度图像中小目标特征不显著的问题,提出了多尺度特征融合增强模块,采用多尺度特征信息融合增强低照度图像中小目标特征. 此外,提出了细节特征提取模块,增强对小目标区域的关注

表4 VisDrone2019-DET数据集上各算法性能对比结果

方法	$mAP50/\%$	$mAP50\sim95/\%$	Parameters/ $10^6$
Yolov5s(2020)	34.2	18.9	7.2
Faster RCNN (2017) <sup>[19]</sup>	33.2	12.0	41.2
Cascade R-CNN (2018) <sup>[25]</sup>	31.9	19.5	69.3
RetinaNet (2020) <sup>[20]</sup>	28.6	18.8	36.1
TOOD (2021) <sup>[26]</sup>	34.2	20.4	32
FCOS (2022) <sup>[21]</sup>	15.5	8.7	31.9
PP-YOLOE (2022) <sup>[22]</sup>	35.6	19.6	52.2
HDETR (2023) <sup>[23]</sup>	35.2	20.8	46.1
YOLOv8(2023)	35.3	20.5	25.8
YOLOv9(2024) <sup>[18]</sup>	37.8	21.7	25.5
YOLOv10(2024) <sup>[24]</sup>	36.3	20.6	24.4
YOLOv11(2024)	38.5	21.3	25.3
ESO-DETR (2025) <sup>[27]</sup>	41	24	14.9
EM-YOLO (2024) <sup>[28]</sup>	43.5	25.1	8.88
PC-YOLO11s (2025) <sup>[29]</sup>	43.8	26.3	7.1
$\alpha$ S-YOLO(2025) <sup>[30]</sup>	48.3	30.4	43.6
NE-CDMNet (2024) <sup>[31]</sup>	51.4	27.3	76
MC-YOLO (ours)	49.1	29	21.7

能力,并在充分利用上下文信息提取目标特征的同时保留更多的小目标细节信息,以此来共同提高模型的

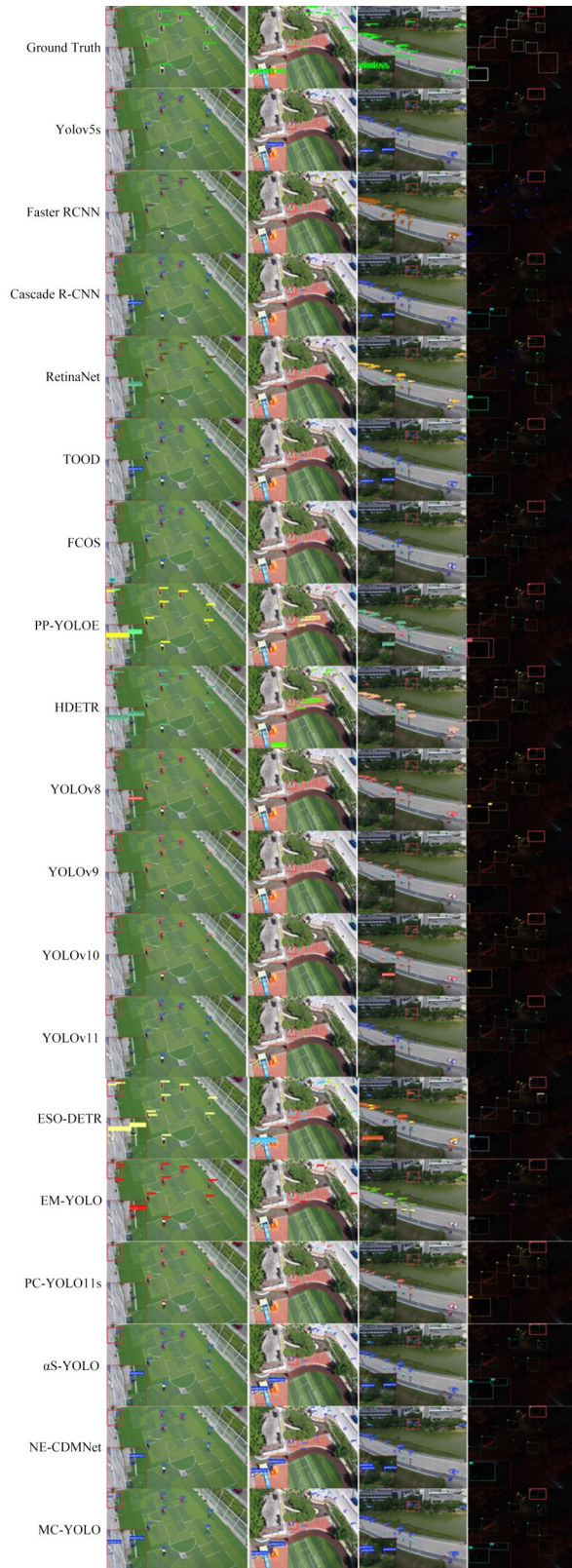


图9 VisDrone数据集上各算法可视化结果对比图

检测能力. 与其他先进的目标检测算法在LLSOD数据集上的检测结果相比, 本文方法能够检测到其他方法

检测不到的低照度区域的小目标物体, 因此本文方法能较好地实现低照度环境下的小目标检测. 下一步将在提升低照度环境下的小目标检测速度及抑制噪声干扰方面开展进一步的探索研究.

#### 参考文献

- [1] KANG S H, PARK J S. Aligned matching: Improving small object detection in SSD[J]. *Sensors*, 2023, 23(5): 2589.
- [2] LIU H, DING M, LI S, et al. Small-target detection based on an attention mechanism for apron-monitoring systems[J]. *Applied Sciences*, 2023, 13(9): 5231.
- [3] BETTI A, TUCCI M. YOLO-S: A lightweight and accurate YOLO-like network for small target selection in aerial imagery[J]. *Sensors*, 2023, 23(4): 1865.
- [4] HU M Z, LI Z Y, YU J, et al. Efficient-lightweight YOLO: Improving small object detection in YOLO for aerial images[J]. *Sensors*, 2023, 23(14): 6423.
- [5] QI G Q, ZHANG Y C, WANG K P, et al. Small object detection method based on adaptive spatial parallel convolution and fast multi-scale fusion[J]. *Remote Sensing*, 2022, 14(2): 420.
- [6] JI C L, YU T, GAO P, et al. Yolo-tla: An efficient and lightweight small object detection model based on YOLOv5[J]. *Journal of Real-Time Image Processing*, 2024, 21(4): 141.1-141.16.
- [7] WANG T Q, QU H Q, LIU C A, et al. LLE-STD: Traffic sign detection method based on low-light image enhancement and small target detection[J]. *Mathematics*, 2024, 12(19): 3125.
- [8] DAI T, WANG Q, SHEN Y C, et al. SwinVision: Detecting small objects in low-light environments[J]. *IEEE Access*, 2025, 13: 42797-42812.
- [9] DONG X, WANG G, PANG Y, et al. Fast efficient algorithm for enhancement of low lighting video[C]//2011 IEEE International Conference on Multimedia and Expo. Piscataway: IEEE, 2011: 1-6.
- [10] PARTHASARATHY S, SANKARAN P. Fusion based multi scale RETINEX with color restoration for image enhancement[C]//2012 International Conference on Computer Communication and Informatics. Piscataway: IEEE, 2012: 1-7.
- [11] ARICI T, DIKBAS S, ALTUNBASAK Y. A histogram modification framework and its application for image contrast enhancement[J]. *IEEE Transactions on Image Processing*, 2009, 18(9): 1921-1935.
- [12] GUO X J, LI Y, LING H B. LIME: Low-light image enhancement via illumination map estimation[J]. *IEEE Transactions on Image Processing*, 2017, 26(2): 982-993.
- [13] YANG S Z, DING M X, WU Y M, et al. Implicit neural representation for cooperative low-light image enhancement[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2023: 12872-12881.
- [14] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-

- attention generative adversarial networks[EB/OL]. (2018-05-21)[2025-05-05]. <https://arxiv.org/abs/1805.08318v2>.
- [15] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 13708-13717.
- [16] PARK J, WOO S, LEE J Y, et al. Bam: Bottleneck attention module[C]//British Machine Vision Conference. Newcastle: BMVC, 2018: 147-160.
- [17] AZAD R, NIGGEMEIER L, HÜTTEMANN M, et al. Beyond self-attention: Deformable large kernel attention for medical image segmentation[C]//2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2024: 1276-1286.
- [18] WANG C Y, YE H I, MARK LIAO H Y. YOLOv9: Learning what you want to learn using programmable gradient information[M]//Computer Vision - ECCV 2024. Cham: Springer Nature Switzerland, 2024: 1-21.
- [19] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [20] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 2999-3007.
- [21] TIAN Z, SHEN C H, CHEN H, et al. FCOS: A simple and strong anchor-free object detector[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(4): 1922-1933.
- [22] XU S L, WANG X X, LV W Y, et al. PP-YOLOE: An evolved version of YOLO[EB/OL]. (2022-12-12)[2025-05-05]. <https://arxiv.org/abs/2203.16250v3>.
- [23] JIA D, YUAN Y H, HE H D, et al. DETRs with hybrid matching[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023: 19702-19712.
- [24] WANG A, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection[EB/OL]. (2024-12-30)[2025-05-05]. <https://arxiv.org/abs/2405.14458v2>.
- [25] CAI Z W, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 6154-6162.
- [26] FENG C J, ZHONG Y J, GAO Y, et al. TOOD: Task-aligned one-stage object detection[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2021: 3490-3499.
- [27] LIU Y F, HE M, HUI B. ESO-DETR: An improved real-time detection transformer model for enhanced small object detection in UAV imagery[J]. Drones, 2025, 9(2): 143.
- [28] WU M J, YUN L J, WANG Y B, et al. Detection algorithm for dense small objects in high altitude image[J]. Digital Signal Processing, 2024, 146: 104390.
- [29] WANG Z, SU Y T, KANG F, et al. PC-YOLO11s: A lightweight and effective feature extraction method for small target image detection[J]. Sensors, 2025, 25(2): 348.
- [30] HOU W, WU H M, WU D, et al. Small object detection method for UAV remote sensing images based on  $\alpha$ S-YOLO[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2025, 18: 8984-8994.
- [31] GUO Z, BI G L, LV H Y, et al. No-extra components density map cropping guided object detection in aerial images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5644813.

## 作者简介



**江泽涛** 男, 1961年3月出生于江西省九江市. 现为桂林电子科技大学教授、博士生导师. 主要研究方向为图像处理、计算机视觉、人工智能. 中国知网高被引学者TOP1%.  
E-mail: 394503704@qq.com



**程留明** 男, 1999年1月出生于河南省周口市. 现为桂林电子科技大学硕士研究生. 主要研究方向为计算机视觉、低照度小目标检测.  
E-mail: 22032201007@mails.guet.edu.cn



**杨建琛** 男, 1998年5月出生于河北省张家口市. 现为桂林电子科技大学硕士研究生. 主要研究方向为计算机视觉、低照度图像增强.  
E-mail: 22032202038@mails.guet.edu.cn