

DOI:10.19651/j.cnki.emt.2415282

# 基于位置感知和跨层特征融合的航拍小目标检测算法<sup>\*</sup>

雷帮军<sup>1,2,3</sup> 余 翩<sup>1,2,3</sup> 吴正平<sup>1,2,3</sup> 余 快<sup>1,2</sup>(1. 湖北省水电工程智能视觉监测重点实验室 宜昌 443002; 2. 三峡大学计算机与信息学院 宜昌 443002;  
3. 水电工程视觉监测宜昌市重点实验室 宜昌 443002)

**摘要:** 针对航拍图像目标尺度小、背景复杂、漏检和误检严重,提出了一种基于位置感知和跨层特征融合的航拍小目标检测算法DC-YOLOv8s。DC-YOLOv8s新增小目标检测层,增强对小目标尺度的敏感性,提高检测精度。为了减少特征信息的丢失,设计了跨层特征融合模块,充分融合小目标浅层语义信息和深层语义信息,丰富特征表示。改进C<sub>2</sub>f结构,结合可变性卷积设计了基于位置感知融入残差的感受野注意力模块,适应航拍小目标形状的变化,快速提取感受野特征,降低漏检和误检率。最后使用基于注意力机制的动态检测头在尺度感知、空间感知、任务感知方面提高复杂场景下小目标的定位性能。实验表明,在VisDrone2019数据集上,DC-YOLOv8s在P、R、mAP上相较于YOLOv8s分别提高了7.2%、7.5%、9.1%,显著提高了小目标检测的性能,FPS为71帧,满足实时性要求。在VOC2007+2012上进行模型泛化性实验验证,效果优于其他经典算法。

**关键词:** 小目标检测层;跨层特征融合;位置感知;感受野;动态检测头

中图分类号: TP391.4 文献标识码: A 国家标准学科分类代码: 520.6040

## Small target detection algorithm for aerial photography based on location awareness and cross-layer feature fusion

Lei Bangjun<sup>1,2,3</sup> Yu Ao<sup>1,2,3</sup> Wu Zhengping<sup>1,2,3</sup> Yu Kuai<sup>1,2</sup>(1. Hubei Key Laboratory of Intelligent Visual Monitoring for Hydropower Engineering, Yichang 443002, China;  
2. School of Computer and Information, China Three Gorges University, Yichang 443002, China;  
3. YiChang Key Laboratory of Hydropower Engineering Vision Supervision, Yichang 443002, China)

**Abstract:** Aiming at the small target scale, complex background, and serious leakage and misdetection of aerial images, an aerial small target detection algorithm based on location awareness and cross-layer feature fusion, DC-YOLOv8s, is proposed. DC-YOLOv8s adds a new small target detection layer, which enhances the sensitivity to the small target scale and improves the detection accuracy. In order to reduce the loss of feature information, a cross-layer feature fusion module is designed to fully fuse the small target shallow semantic information and deep semantic information to enrich the feature representation. Improve the C<sub>2</sub>f structure, combined with variability convolution to design a sensory field attention module based on position-aware incorporation of residuals, adapting to the changes in the shape of aerial small targets, quickly extracting sensory field features, and reducing the rate of leakage detection and false detection. Finally, the dynamic detection head based on the attention mechanism is used to improve the localization performance of small targets in complex scenes in terms of scale perception, spatial perception, and task perception. The experiments show that on VisDrone2019 dataset, DC-YOLOv8s improves 7.2%, 7.5%, and 9.1% on P, R, and mAP respectively compared to YOLOv8s, which significantly improves the performance of small target detection, and the FPS is 71 frames, which meets the real-time requirement. Experimental verification of model generalizability is carried out on VOC2007+2012, and the effect is better than other classical algorithms.

**Keywords:** small target detection layer; cross-layer feature fusion; position-aware; sensory field; dynamic detection head

## 0 引言

势被广泛应用于农业作物检测、电力管廊检测<sup>[1]</sup>、军事情报采集、遥感测绘<sup>[2]</sup>、嵌入式视觉目标定位<sup>[3]</sup>等领域。与传统的地面检测不同,无人机往往在高空进行拍摄,面临着拍摄

随着科技的不断进步,无人机航拍技术以其独特的优

视角变化大、背景复杂、目标尺度较小细节特征不明显、部分重叠、相互遮挡、误检和漏检严重等问题,给检测带来了较大的困难。近年来,随着深度学习的飞速发展,目标检测领域取得了显著的进展。基于深度学习的航拍小目标检测算法可以分为两阶段(two-stage)和一阶段(one-stage)和两类。两阶段目标检测算法首先通过候选框生成阶段提取潜在目标框,经过非极大值抑制,最终在目标分类与位置回归阶段输出目标检测结果,这类经典算法以Fast R-CNN<sup>[4]</sup>、Faster R-CNN<sup>[5]</sup>为代表,优点是精度高、尺度适应性好,但也存在着训练和推理耗时长、计算复杂度高、实时性差等缺点。一阶段目标检测算法则省略了候选框生成这个阶段,直接输出图像中目标类别和位置信息,实现端到端的检测,这类算法以SSD<sup>[6]</sup>、RetinaNet<sup>[7]</sup>、YOLO<sup>[8-11]</sup>为代表,在保持较高检测精度的同时,兼具模型复杂度低、检测速度快、实时性高的优点。YOLOv8s是2023年Ultralytics公司推出的一阶段目标检测算法,相较于其他目标检测算法,在目标检测任务上具有精度高、速度快、实时性高、易于部署到移动端设备等优点。但是无人机航拍图像往往分辨率低、小目标居多、背景复杂,原始YOLOv8s算法在航拍检测任务上精度低、误检和漏检较多。具体存在如下劣势:1)YOLOv8s在80×80、40×40、20×20三个尺度的特征层上进行检测,对应着检测小、中、大物体,但是航拍目标大部分像素小,导致80×80尺度的小目标检测头无法有效识别目标,这就造成了误检或漏检。2)YOLOv8s采用FPN-PAN结构进行特征融合,传递特征信息,但是FPN-PAN结构中频繁的上下采样操作导致小目标细节信息丢失严重,而且主干网络不经过FPN-PAN处理的浅层特征图中含有的丰富位置细节信息没有得到有效利用,造成小目标检测精度低。3)YOLOv8s的检测头为解耦头,对比耦合头,在速度和精度上有一定的提升,但是相较于注意力机制的新型检测头,在航拍目标密集的复杂背景下,定位能力较差。因此需要合理改进YOLOv8s网络结构,满足航拍检测任务的需要。

针对小目标检测,吴旭红等<sup>[12]</sup>在YOLOv7网络基础上基于注意力机制构建MP-SimAM模块,提出边界框回归损失函数SCIoU Loss,提高小目标检测精度,但是网络参数量大、计算速度慢。郑有凯等人<sup>[13]</sup>提出基于自适应特征融合和任务对齐的小目标检测算法,为高空或远距离场景小目标检测应用提供了解决方案,但是模型较为复杂。吴明杰等<sup>[14]</sup>提出无人机视角下小目标检测算法BD-YOLO,解决无人机拍摄目标存在大小差异和遮挡等问题,但是检测精度不高。熊娟等<sup>[15]</sup>针对遥感图像方向任意性等问题,提出了一种基于上下文信息增强的特征金字塔网络,减少目标的信息流失,但是各个类别之间的性能难以做到较好的平衡。

鉴于YOLOv8s算法的优势,一些研究人员开始改进YOLOv8s并提出了一些优化算法,以提高航拍小目标检

测的精度。程换新等<sup>[16]</sup>针对航拍图像目标检测算法精度低,提出一种改进YOLOv8s的算法Bi-YOLOv8,在骨干网络中加入多尺度注意力EMA,捕捉细节信息,改进C2f模块,减少模型的计算量,此外使用BiFPN改进颈部,增强模型多尺度特征融合能力,但是实验平均检测精度仅为40.7%,在小目标检测方面仍然有较大的提升空间;邓天民等<sup>[17]</sup>针对无人机小目标检测精度低和模型参数量大的问题,提出一种改进YOLOv8s的算法FS-YOLO,通过降低通道维数和改进网络结构,实现对冗余特征信息的高效复用并减小模型大小,在特征融合阶段引入内容感知特征重组模块,加强小目标语义信息的关注,提升网络对航拍图像的检测性能。但FS-YOLO的检测效果在目标像素点个数极少、特征不明显、背景复杂的情况下仍有改进空间。Wang等<sup>[18]</sup>提出一种基于UAV航拍场景的目标检测算法UAV-YOLOv8,采用WIoUv3作为边界框回归损失,使得模型更加关注普通质量样本,引入BiFormer注意力机制对骨干网络进行改进,提高模型对关键信息的关注,设计特征处理模块FFNB融合浅层特征与深层特征,有效提高了对小目标的检测能力,但是计算资源消耗过大、实时性不足。Sui等<sup>[19]</sup>针对无人机被探测目标的大小差异较大、物体遮挡等问题,提出了一种基于YOLOv8s的无人机高空小目标检测算法BDH-YOLO,将双向加权金字塔特征网络BiFPN与骨干网络相结合,增强多层次特征融合,获得更多语义信息,针对特征图多次下采样导致分辨率降低,丢失大量空间信息的问题,提出结合自注意力的检测头,改善检测性能。但是平均检测精度为42.9%,精度提升幅度有限,算法训练、收敛时间较长。目前对于航拍图像检测仍然存在以下难点:1)目标尺度小含有的特征信息较少。2)目标密集受复杂背景影响较大。3)小目标难以定位,误检和漏检严重。4)特征提取过程中,容易丢失细节信息。

对于上述航拍目标检测中存在的难点以及现有改进YOLOv8s算法存在的问题,本文改进YOLOv8s<sup>[20]</sup>,提出了一种针对航拍小目标检测的改进算法DC-YOLOv8s。与改进YOLOv8s的文献算法<sup>[16-19]</sup>一样,本文算法也是通过改进YOLOv8s,提高航拍小目标检测精度。不同之处在于,本文根据航拍小目标的特性,从多元化角度进行改进,涉及改进特征提取和特征融合网络、优化检测头结构、设计D\_RFCBAM\_C2f模块等。每个改进点各不相同,但都有效提升检测精度。在平均检测精度大幅提升的情况下,召回率、实时性等指标依旧较高,做到了较好的平衡,没有顾此失彼。本文算法设计中的难点在于:1)跨层特征融合模块中不同尺度特征层的选取以及跨层交互融合的实现。2)D\_RFCBAM\_C2f模块的设计过程中如何将DCBS与RFCBAMConv有机结合,最大化提高检测精度。DC-YOLOv8s算法主要创新为:1)针对小目标尺度不敏感,调整特征提取和特征融合网络,合理增加160 pixel×160 pixel的极小目标检测层,提升检测效果。2)利用设计

的跨层特征融合模块(cross-layer feature fusion moudle, CLFFM), 强化浅层特征图局部信息和深层特征图全局信息的融合, 减少小目标位置细节信息的丢失, 提高检测精度。3) 使用可变性卷积联合感受野注意力模块 RFCBAMConv 和  $C_2f$  设计了全新的基于位置感知融入残差的感受野注意力模块  $D\_RFCBAM\_C_2f$ , 引入空间变换, 适应小目标的形状, 快速提取感受野特征, 减少漏检和误检。4) 使用基于注意力机制的动态检测头(dynamic head, DyHead)<sup>[21]</sup> 在尺度感知、空间感知、任务感知 3 个方面提高复杂场景下小目标的定位能力。VisDrone2019 数据集上的实验结果表明, 本文改进算法 DC-YOLOv8s 在精确率、平均检测精度上相较于 YOLOv8s 分别提高了 7.2%、9.1%, FPS 为 71 帧, 实时性较高。对比其他主流以及改进算法, 平均检测精度均有不同程度的提升, 综合性能较好。

## 1 YOLOv8s 算法介绍

YOLOv8 是 Ultralytics 公司于 2023 年发布的一阶段目标检测算法模型。一共有 5 个版本, 分别是 YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l、YOLOv8x。考虑到精度和实时性, 选择 YOLOv8s 作为本次实验的 Baseline。YOLOv8s 整体网络由 Backbone 特征提取、FPN-PAN 特征融合、Head 三部分构成。Backbone 部分对输入图像进行特征提取, FPN-PAN 特征融合网络利用多尺度金字塔, 通过路径聚合融合不同层级的特征信息。Head 部分使用

anchor-free 的检测头生成最终的预测结果。

## 2 DC-YOLOv8s 改进算法介绍

由于无人机航拍图像往往目标尺寸较小、遮挡严重、背景复杂, YOLOv8s 针对此种情况检测精度较低、误检和漏检严重。基于此, 本文设计了基于位置感知和跨层特征融合的目标检测算法 DC-YOLOv8s。DC-YOLOv8s 算法的整体网络结构如图 1 所示, 右上角黑色实线框为本文改进的 4 个模块, 分别为基于位置感知融入残差的感受野注意力模块、跨层特征融合模块、基于注意力机制的动态检测头、增加小目标检测层。针对小目标检测精度低效果不好, 本文在 YOLOv8s 的基础上新增加一个 160 pixel × 160 pixel 的检测头, 增强小目标的捕获能力。为了减少小目标语义信息丢失, 设计 CLFFM 对 Backbone 特征提取部分 80 pixel × 80 pixel 和 40 pixel × 40 pixel 的特征图进行空间和通道维度的交互, 之后跟 PAN 中相同尺度的特征图融合, 丰富特征信息表示, 优化小目标检测效果。设计基于位置感知融入残差的感受野模块  $D\_RFCBAM\_C_2f$ , 更好的适应小目标的形状和尺度, 获得更加灵活的感受野, 减少小目标细节信息的丢失, 降低误检和漏检率。使用基于任务感知、空间感知、尺度感知的动态检测头 DyHead, 在增加极少额外计算量的情况下, 显著增强了检测头对目标的表达能力, 加强密集复杂场景下小目标的特征提取能力。下面将具体介绍这 4 个改进点。

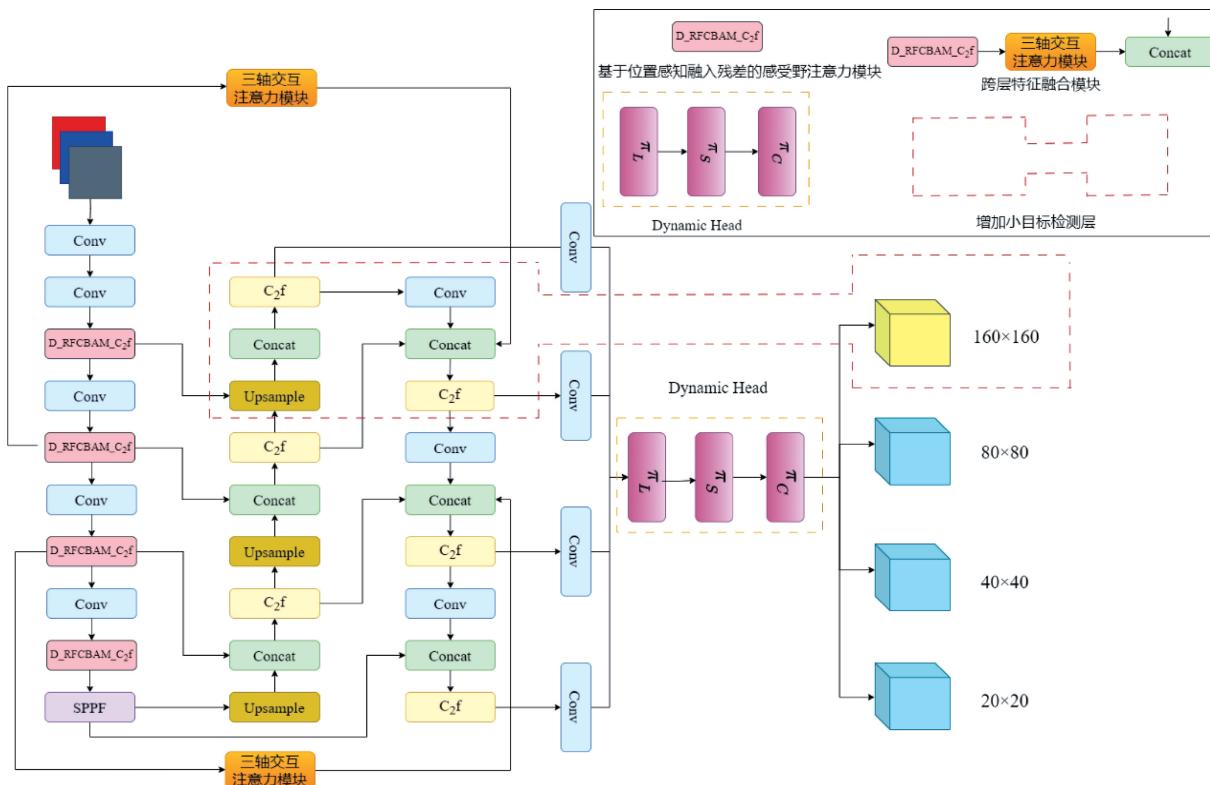
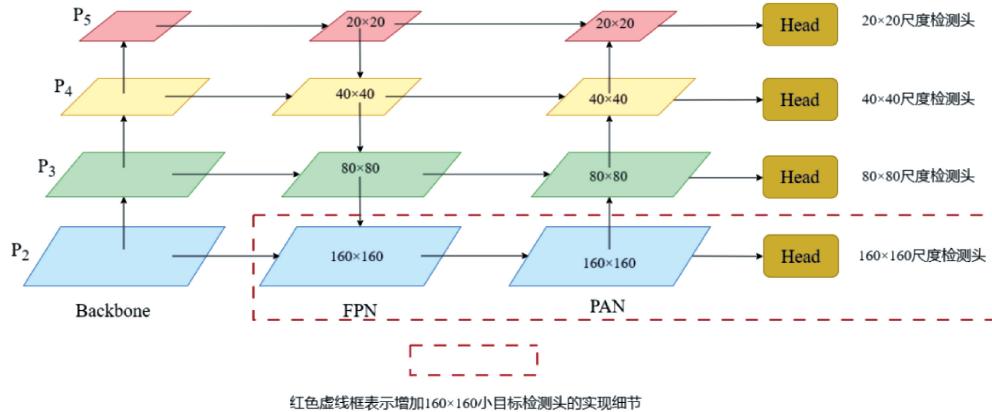


图 1 DC-YOLOv8s 算法整体网络结构图

## 2.1 增加小目标检测层

小目标检测精度不足一直是无人机航拍图像的一个难点,由于航拍图像大部分为小目标,而且较为密集、背景复杂,这给检测带来了较大的困难。在原始 YOLOv8s 网络中,尺寸大小为 640 pixel×640 pixel 的输入图像依次经过 Backbone 特征提取网络、FPN-PAN 特征融合网络,最后在 Head 部分进行检测。Head 部分一共有 3 个检测头,对应的特征图尺寸依次为 80 pixel×80 pixel、40 pixel×40 pixel、20 pixel×20 pixel,代表着检测小、中、大物体。但是小目标的特征信息多存在于浅层特征图中,原始的 80 pixel×80 pixel 的检测头对小目标检测效果并不理想,检测精度低。基于此,DC-YOLOv8s 新增一个 160 pixel×160 pixel

的小目标检测头,使得检测头对小目标更加敏感,提高小目标检测的精度。改进具体过程如图 2 红色虚线框部分所示:在自上向下融合网络(feature pyramid networks, FPN)80 pixel×80 pixel 的特征图后面增加了一个上采样操作,突破感受野极限,将特征图尺寸变为 160 pixel×160 pixel,之后和 Backbone 相同尺寸的特征图进行融合,尽可能多地保留小目标的细节信息。在自下向上特征融合网络(path aggregation network, PAN)中与来自 FPN 的特征图再次进行融合,以达到强化特征信息保留的目的。之后将特征图信息输入到 160 pixel×160 pixel 检测头中。经过改进后的 DC-YOLOv8s 网络包含 4 种不同尺度的检测头,更加适应航拍图像复杂尺度的剧烈变化。



红色虚线框表示增加 $160\times 160$ 小目标检测头的实现细节

图 2 添加小目标检测层细节图

## 2.2 跨层特征融合模块

YOLOv8s 采用 FPN-PAN 金字塔结构进行特征融合,在一定程度上提高了小目标检测的准确性,但是频繁上下采样操作导致特征细节模糊化和丢失。另外,Backbone 部分的特征保留了大量的小目标位置细节信息,单纯经过 FPN-PAN 特征融合会导致小目标特征信息的丢失。基于此,本文在增加小目标检测层的基础上设计 CLFFM,丰富 Backbone 的小目标信息,在增加极少量额外计算开销的情况下,建立空间和通道之间的联系,通过跨层连接,充分融合浅层与深层语义信息,强化小目标位置细节信息的保留。

如图 3 红色虚线框所示,CLFFM 过程如下:对 Backbone 中  $p_3$ 、 $p_4$  层的输出特征图,在 3 个分支上通过池化操作分别将 C、H 和 W 维度缩减为 2 维,捕获对应维度上的平均池化特征和最大池化特征。之后通过卷积层、归一化、激活函数,得到  $(H, W)$ 、 $(C, W)$ 、 $(C, H)$  维度之间的关系权重向量。然后利用权重对原始特征图进行优化。随后将 3 个分支中优化后的特征图进行加权求和,帮助模型关注 Backbone 特征提取网络中小目标关键信息的同时,建立空间和通道之间的联系。最后将细化后的输出特征图以及 PAN 网络中  $p'_3$  和  $p'_4$  的输出特征图进行跨层信息融合,捕获多尺度信息,提高小目标检测的鲁棒性。与

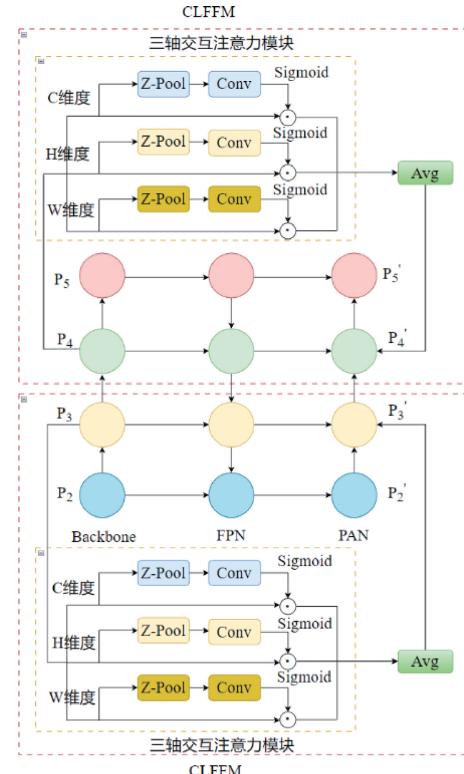


图 3 跨层特征融合模块 CLFFM

FPN-PAN 特征融合网络相比,CLFFM 增强 Backbone 浅层特征图中的小目标细节信息,提高模型对多尺度信息的感知能力,强化小目标细粒度信息的保留。

以  $p_3$  层输出的特征图经过 CLFFM 为例,具体数学计算过程如下:

$$y' = \text{Cat}(y, p'_3) \quad (1)$$

$$y = \frac{1}{3}(\hat{\chi}_1 u \omega_1 + \hat{\chi}_2 u \omega_2 + \hat{\chi}_3 u \omega_3) \quad (2)$$

$$\omega_1 = \sigma(\psi_1(\hat{\chi}_1^* u)) \quad (3)$$

$$\omega_2 = \sigma(\psi_2(\hat{\chi}_2^* u)) \quad (4)$$

$$\omega_3 = \sigma(\psi_3(\hat{\chi}_3^* u)) \quad (5)$$

其中,  $y'$  表示经过跨层特征融合后的输出,  $\text{Cat}$  表示在通道上面进行拼接,  $p'_3$  表示  $p_3$  层的特征图经过 PAN 网络后的输出,  $y$  表示在 C、H、W 三个维度上经过注意力机制作用的加权输出,  $\hat{\chi}_1 u$ 、 $\hat{\chi}_2 u$ 、 $\hat{\chi}_3 u$  表示 C、H、W 三个维度上的输入,  $\omega_1$ 、 $\omega_2$ 、 $\omega_3$  表示 C、H、W 三个维度上生成的注意力机制权重。 $\hat{\chi}_1^* u$ 、 $\hat{\chi}_2^* u$ 、 $\hat{\chi}_3^* u$  表示在 C、H、W 三个维度上面经过 Z-Pool 后的输出。

### 2.3 基于位置感知融入残差的感受野注意力模块

#### 1) 设计可变性卷积 DCBS 模块

传统卷积的卷积核尺寸不变,只能在固定大小的位置上进行采样,对全局信息的感知不足。可变性卷积(deformable convolution, DCN)<sup>[22]</sup>引入位置偏移量,调整卷积核在位置上对输入特征图的采样方式,增大感受野,建立全局上下文信息,更好的适应航拍目标复杂尺度的变化。受 Deformable ConvNets v2<sup>[23]</sup>的启发,本文采用可变性卷积 v2(deformable convolution v2, DCNv2),相较于 DCN,引入调制机制,通过调制标量  $\Delta m_k$  和可学习偏移量  $\Delta p_k$  根据输入特征进行学习调整,具备空间感知自适应能力。结合 DCNv2,本文设计了 DCBS 模块,如图 4 所示,DCBS 采用批量归一化(batch normalization, BN)对每一层的输入进行标准化,降低输入分布的偏移,使得网络学习的特征更加明显,加速模型的收敛能力。同时 BN 减轻了神经网络对训练数据的过拟合,使得网络对于输入数据的微小变化更加鲁棒,提高模型的泛化能力。之后引入 SiLU 激活函数,通过非线性映射,使得网络学习更为复杂的特征和表示,减少梯度消失,增强网络的拟合能力。DCNv2 的计算公式如下:

$$y(p) = \sum_{k=1}^K \omega_k \cdot x(p + p_k + \Delta p_k) \cdot \Delta m_k \quad (6)$$

其中,  $\omega_k$  表示第  $k$  个位置上的权重,  $y(p)$  表示输出特征图在 P 处的特征,  $x$  输入特征图在 P 处的特征,  $p_k$  表示预先指定的偏移量,  $\Delta p_k$  表示第  $k$  个位置的可学习偏移量,  $\Delta m_k$  表示第  $k$  个位置的调制标量。

#### 2) 设计基于残差的感受野注意力模块

标准卷积是神经网络构建的基础模块,使用具有共享

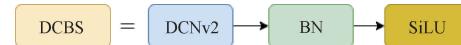


图 4 DCBS 结构图

参数的滑动窗口进行特征信息的提取。虽然在一定程度上减少了计算量和计算开销,但是无法捕捉不同位置带来的信息差异,限制了网络的性能。标准卷积的数学计算过程如式(7)所示。

$$F_N = X_{N1} \times K_1 + X_{N2} \times K_2 + X_{N3} \times K_3 + \dots + X_{NS} \times K_S \quad (7)$$

其中,  $F_N$  表示经过卷积操作输出的值,  $X_N$  表示每个滑块内对应位置的像素值,  $K$  表示卷积核,  $S$  为卷积核中参数的个数,  $N$  表示滑块的总数,  $K_S$  表示滑块内相同位置的特征共享的参数。

空间注意力机制,例如 CBAM<sup>[24]</sup>中的空间注意力部分,在一定程度上可以提高神经网络的性能。假定输入特征图的通道数为 1,在  $3 \times 3$  卷积的前面加入空间注意力机制,则空间注意力机制的数学计算过程如式(8)所示。

$$F_N = X_{Ni} \times A_{Ni} \times K_i \quad (8)$$

其中,  $F_N$  表示加权运算后的值,  $X_i$  表示输入特征图在不同位置的值,  $A_i$  表示注意力机制在不同特征图位置的值,  $N$  表示宽度和高度的乘积,  $K_i$  表示  $3 \times 3$  卷积核。仔细分析发现,  $A_{12} = A_{21}, A_{13} = A_{22} \dots$  等。由此可知,空间注意力机制对于  $3 \times 3$  这种大核卷积依然没能够解决参数共享的问题,没有充分考虑整个感受野的空间特征。

Liu 等<sup>[25]</sup>提出了感受野模块(receptive field block, RFB),从模拟人类视觉感受野出发,加强网络的特征提取能力。受此启发,本文设计了基于残差的感受野注意力模块 RFCBAMConv。将空间注意力机制引导到感受野空间上,强调感受野滑块内不同特征的重要性,解决不同尺寸卷积核参数共享的问题,并对远程信息进行建模。RFCBAMConv 的结构如图 5 所示,分为左右两分支,左支首先进行全局平均池化减少参数量,并聚合感受野的全局特征信息,之后经过两个全连接层,生成通道注意力权重  $w_1$ 。右支首先使用  $3 \times 3$  的分组卷积快速提取感受野特征,利用归一化和 ReLU 激活函数加速收敛速度,提高网络的表达能力。其次,对生成的特征图变换形状后,在通道维度上进行平均池化和最大池化捕获空间分布信息,经过标准卷积和 sigmoid 激活函数得到空间注意力权重  $w_2$ 。然后利用  $w_1$  和  $w_2$  对变换后的特征图进行加权。最后,在经过  $3 \times 3$  卷积提取特征信息后,与原始图像进行残差连接,保留更多细节信息,提高模型的性能。

#### 3) 设计 D\_RFCBAM\_C2f 结构

以 DCBS、RFCBAMConv 与 C2f 联合设计了 D\_RFCBAM\_C2f 模块,适应小目标的形状尺度,充分获取感受野特征,建立全局信息之间的依赖,降低误检率和漏检率。图 6(c)以 3 个 D\_RFCBAM\_C2f 为例,展示了设计的 D\_RFCBAM\_C2f 结构。对输入特征图经过卷积提取图

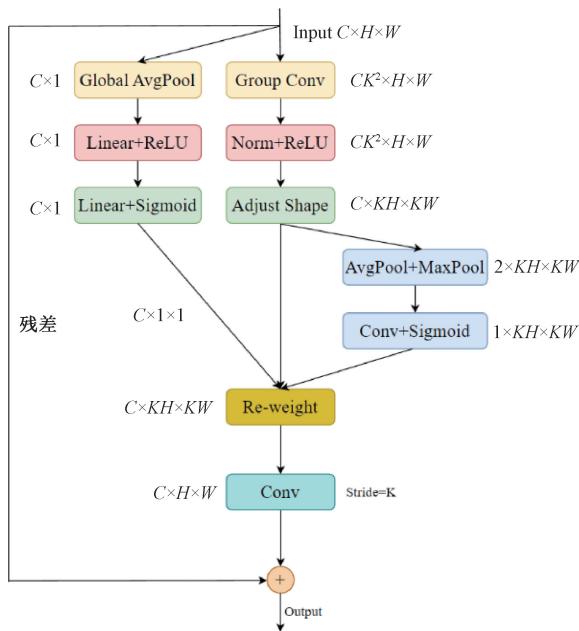
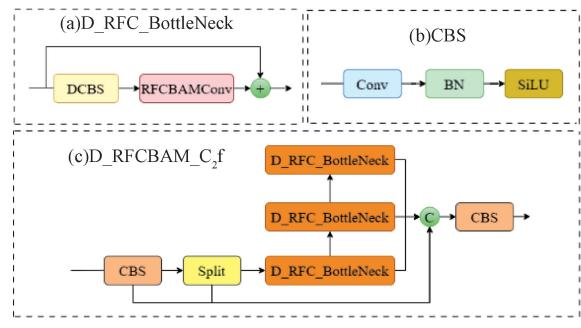


图 5 RFCBAMConv 结构图

像特征信息后,利用 split 对其进行通道切分,并引入多路径信息流。其次,为了避免网络加深带来的特征信息丢失,split 之后串联 3 个基于残差的 D\_RFC\_BottleNeck 结构。最后,将每一层的输出进行跨阶段融合。其中,D\_RFC\_BottleNeck 的结构如图 6(a)所示,下分支经过 DCBS 和 RFCBAMConv 分别达到捕捉位置细节信息和增大感受野目的,之后与捷径分支相加,缓解梯度消失,保留更多底层信息。相较于原始的 C<sub>2</sub>f 模块,D\_RFCBAM\_C<sub>2</sub>f 模块在增加小幅参数量的情况下,获得了更加丰富的梯度信息。

## 2.4 基于注意力机制的动态检测头

检测头起到分类和定位的作用,高效的检测头可以提高复杂场景下小目标的定位性能,提高检测的精确度。随着深度学习的发展,注意力机制被广泛运用在计算机视觉任务中,注意力机制引导模型关注更加重要的部分,从而

图 6 D\_RFCBAM\_C<sub>2</sub>f 模块整体及细节结构图

提高模型的表达能力。基于此,本文改进 YOLOv8s 的检测头,使用基于注意力机制的动态检测头 DyHead,在小幅增加参数量的情况下,有效提升检测的性能。如图 7 所示,DyHead 在尺度感知、空间位置、任务感知 3 个角度运用注意力机制,尺度感知注意力机制学习语义信息的相对重要性,空间位置感知注意力机制学习特征的连贯判别表示,任务感知注意力机制引导不同的特征通道支持不同的任务,例如分类、框回归、关键点学习等任务。3 个注意力机制构成了高性能的目标检测头。对于给定的特征张量  $F \in R^{L \times S \times C}$ , 动态检测头表示如下:

$$W_F = \pi_C(\pi_S(\pi_L(F) \cdot F) \cdot F) \quad (9)$$

其中,  $\pi_C$  表示任务感知注意力机制,  $\pi_S$  表示空间位置感知注意力机制、 $\pi_L$  表示尺度感知注意力机制。 $\pi_L$  的定义为:

$$\pi_L(F) \cdot F = \sigma(f(\frac{1}{SC} \sum_{s,c} F)) \cdot F \quad (10)$$

$$\sigma(x) = \max(0, \min(1, \frac{x+1}{2})) \quad (11)$$

其中,  $f(\cdot)$  是由  $1 \times 1$  卷积层近似的线性函数,  $\sigma$  为 hard-sigmoid 激活函数。 $\pi_S$  的定义为:

$$\pi_S(F) \cdot F = \frac{1}{L} \sum_{l=1}^L \sum_{k=1}^K \omega_{l,k} \cdot F(l; p_k + \Delta p_k; c) \cdot \Delta m_k \quad (12)$$

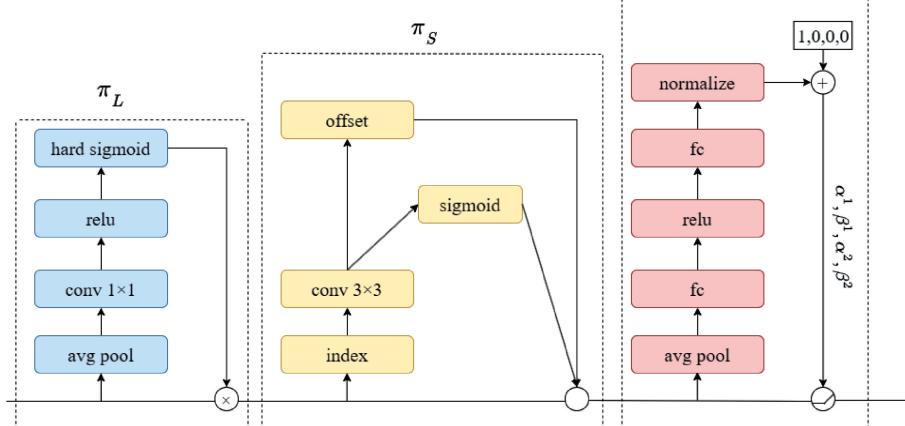


图 7 DyHead 结构图

其中,  $K$  表示稀疏采样位置的数量,  $p_k + \Delta p_k$  是通过自主学习空间偏移了  $\Delta p_k$  的空间位置,  $\Delta m_k$  是位置  $p_k$  处的自主学习重要性标量, 两者都是从  $F$  的输入特征中学习而得到。 $\pi_c$  的定义如下:

$$\pi_c(F) \cdot F = \max(\alpha^1(F) \cdot F_c + \beta^1(F), \alpha^2(F) \cdot F_c + \beta^2(F)) \quad (13)$$

$$\theta(\cdot) = [\alpha^1, \alpha^2, \alpha^3, \alpha^4]^T \quad (14)$$

其中,  $F_c$  是第  $c$  通道的切片,  $\theta(\cdot)$  是一个超函数, 学习控制激活的阈值。它首先在  $L \times S$  的维度上通过平均池化降维, 之后使用两个全连接层和归一化操作, 最后使用移位的 sigmoid 激活函数将输出值映射到  $[-1, 1]$  之间。

### 3 实验结果及分析

#### 3.1 实验环境及参数设置

本实验环境为:NVIDIA GeForce RTX3090, 显存 24 GB, Ubuntu20.04 操作系统。编程语言为 Python, CUDA 为 11.3, 深度学习框架为 PyTorch 为 1.12.1。YOLOv8s 依赖库 Ultralytics 版本为 8.0.202。训练过程中的参数设置如表 1 所示。

表 1 训练参数设置

参数名称	参数设置	参数名称	参数设置
optimizer	SGD	close_mosaic	10
epochs	300	momentum	0.937
batchsize	8	weight_decay	0.0005
workers	8	warmup_epochs	3
imgs	640	warmup_momentum	0.8
lr0	0.01	patience	50
lrf	0.01	warmup_bias_lr	0.1

#### 3.2 实验数据集

为了验证本文改进算法的效果, 使用公开的无人机航拍数据集 VisDrone2019 进行实验验证。VisDrone2019 数据集是由天津大学机器学习与数据挖掘实验室发布的, 数据集包括城镇晴天、多云、雨天等多种天气和光照条件, 检测目标多样, 遮挡严重, 目标密集, 涵盖不同尺度的目标, 给检测带来了较大的困难。该数据集训练集图片数量为 6 471 张, 验证集图片数量为 548 张, 测试集图片数量为 1 610 张, 一共包含 10 个类别, 分别为 Pedestrian、People、Bicycle、Car、Van、Truck、Tricycle、Awning-tricycle、Bus、Motor。

#### 3.3 评价指标

本实验使用精确率(Precision)、召回率(Recall)、平均精度均值(mean average precision)、参数量(Parameters)、每秒帧数(frames per secend, FPS)评价指标对模型的性能进行评价。Precision、Recall、mAP 的计算公式如下:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (15)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (16)$$

$$AP = \int_0^1 P(r) dr \quad (17)$$

$$mAP = \frac{\sum_{i=1}^k AP_i}{k} \quad (18)$$

式中:  $TP$  表示模型正确预测为正类别的正样本数量,  $FP$  表示模型错误预测为正类别的负样本数量,  $FN$  表示模型错误预测为负类别的正样本数量。 $AP$  表示单个类别的精确率,  $mAP$  是以 *Precision* 为横坐标, *Recall* 为纵坐标进行积分所得图像的面积, 代表了所有类别精确率的均值。

#### 3.4 消融实验

为了验证本文提出的 4 个改进方法对无人机航拍图像性能提升的作用, 以 YOLOv8s 为基线网络, 进行了一系列的消融实验, 实验结果如表 2 所示。“√”表示添加了该方法, 表中第一行表示什么方法都不添加的情况下 YOLOv8s 的检测指标。从表 2 可以看出, 添加小目标检测层,  $P$ 、 $R$ 、 $mAP$  相较于 YOLOv8s 分别提高了 5.5%、3.6%、4.8%, 参数量下降了 0.5 M, 说明加入小目标检测层, 不仅提高了航拍小目标检测的精确率, 而且使得模型更加轻量化。在此基础上接着引入跨层特征融合模块 CLFFM, 增强浅层语义信息与深层语义信息的融合, 丰富多尺度特征表示, 保留更多小目标的细节纹理特征,  $R$  和  $mAP$  分别提高了 1.2%、0.4%, 参数量增加了 0.09 M, FPS 小幅下降。使用基于注意力机制的动态检测头,  $P$ 、 $R$ 、 $mAP$  分别提高了 3.2%、2.1%、3.4%, 说明动态检测头的引入显著提升了复杂场景下小目标检测的精确率。最后引入基于位置感知融入残差的感受野注意力模块 D\_RFCBAM\_C2f, 解决参数共享的问题, 快速提取感受野特征, 建立全局上下文信息的关联,  $R$ 、 $mAP$  分别提高了 0.6%、0.5%, 误检和漏检率进一步降低。综合来看, 本文改进模型在  $P$ 、 $R$ 、 $mAP$  上相较于 YOLOv8s, 提升了 7.2%、7.5%、9.1%, 参数量同比下降 0.27 M, 说明本文提出的 4 个改进点显著增强了对航拍小目标的捕获能力, 大幅提升检测性能。虽然 FPS 降至 71 帧, 但是满足实时性(30 FPS)的要求。

#### 3.5 对比实验

为了验证本文提出的算法 DC-YOLOv8s 对航拍小目标检测的优越性, 在 VisDrone2019 数据集上, 与当前先进的航拍小目标检测算法进行对比, 实验环境和参数设置均保持一致。为了实验设计的科学性, 本文选取了经典主流算法、较新现有改进算法、最新改进 YOLOv8 算法等先进算法作为对比算法。其中 BD-YOLO、VA-YOLO、SDS-YOLO、Tph-YOLOv5 为改进 YOLOv5 的较新文献算法; 文献算法 1 为改进 YOLOv7-tiny 的最新文献算法; FS-YOLO、BDH-YOLO、UAV-YOLOv8 为改进 YOLOv8 的最新文献算法。对比实验主要从各个类别 AP 值和模型

表 2 消融实验结果

小目标检测层	CLFFM	DyHead	D_RFCBAM_C <sub>2</sub> f	P/%	R/%	mAP/%	Parameters/M	FPS
✓				50.5	38.7	39.4	11.12	105
	✓			56.0	42.3	44.2	10.62	93
		✓		51.1	38.9	40.1	11.21	100
			✓	54.2	40.1	41.9	11.16	85
✓	✓			51.3	39.4	40.5	11.22	98
✓	✓	✓		54.4	43.5	44.6	10.71	89
✓	✓	✓	✓	57.6	45.6	48.0	10.75	74
			✓	57.7	46.2	48.5	10.85	71

P、R、mAP 等指标这两个方面进行全面综合的性能对比。不同先进算法在 VisDrone2019 数据集上各个类别 AP 值对比实验结果如表 3 所示,“—”表示原文献没有提到此项指标从表 3 可以看出 DC-YOLOv8s 不仅在 People、Bicycle、Van、Truck、Tricycle、Awning-tricycle、Motor 这七个类别上均取得了最佳的检测性能,AP 值分别为 46.1%、23.0%、51.8%、42.5%、35.5%、21.1%、57.5%,而且平均检测精度 mAP 也取得了最好成绩,为 48.5%,对于 Pedestrian、Car、Bus 这种纵横较大的类别,算法也表现出不错的检测效果,AP 值分别为 56.1%、85.7%、65.6%,说明 DC-YOLOv8s 算法在各种目标检测场景下都有着优秀的检测性能,对比其他先进目标检测算法,综合性能最优。DC-YOLOv8s 相较于 YOLOv8s,在较小目标类别 Pedestrian、Bicycle、Tricycle 上 AP 值分别提升了 12.7%、

9.6%、7.2%。相较于 YOLOv8m 这种大一个量级的模型,mAP 提升了 5.3%。和较新现有改进文献算法 BD-YOLO、VA-YOLO、SDS-YOLO、Tph-YOLOv5、文献算法 1 相比,DC-YOLOv8s 在 Pedestrian 等 10 个目标类别上 AP 值均为最高,平均检测精度分别提高 15.9%、9.5%、8.5%、3.6%、7.6%。和改进 YOLOv8 的文献算法 FS-YOLO 对比,DC-YOLOv8s 平均检测精度高出 1.5%,而且在 Pedestrian、People 等 9 个目标类别上 AP 值均处于领先的水平,在 Bus 这个目标类别上 AP 值为 65.6%,仅低 0.4%。和 UAV-YOLOv8 算法相比,DC-YOLOv8s 平均检测精度高出 1.5%,在 People、Bicycle 等 8 个目标类别上均取得最好结果。实验结果充分说明 DC-YOLOv8s 提出的 4 个改进点,能够充分提取小目标的特征信息,大幅提高检测精度,在航拍检测任务上平均检测精度最高、综合性能最佳。

表 3 不同先进算法在 VisDrone2019 数据集上的 AP 与 mAP 对比

算法	mAP	Pedestrian	People	Bicycle	Car	Van	Truck	Tricycle	Awning-tricycle	Bus	Motor
Faster R-CNN <sup>[5]</sup>	21.7	21.4	15.6	6.7	51.7	29.5	19.0	13.1	7.7	31.4	20.7
RetinaNet <sup>[7]</sup>	13.9	13.0	7.9	1.4	45.5	19.9	11.5	6.3	4.2	17.8	11.8
YOLOv4	40.4	43.4	33.3	16.1	78.9	44.2	39.3	28.5	16.3	60.9	44.2
YOLOv5s	33.0	40.2	32.2	10.3	73.3	35.9	27.5	18.1	10.7	41.5	39.7
YOLOv5m	37.3	45.5	35.7	15.0	76.3	40.2	32.7	24.5	11.9	48.7	42.3
YOLOX <sup>[26]</sup>	40.3	46.2	38.5	13.8	80.3	43.5	33.2	28.7	14.8	55.2	49.2
YOLOv7-tiny <sup>[27]</sup>	35.5	39.6	36.2	9.6	77.5	38.3	30.3	19.4	10.2	49.6	44.5
YOLOv8s	39.4	43.4	33.0	13.4	79.8	45.0	36.0	28.3	15.5	54.6	44.6
YOLOv8m	43.2	46.2	36.6	17.8	81.7	48.1	40.3	33.8	15.9	62.1	49.7
BD-YOLO <sup>[14]</sup>	32.6	42.6	34.9	12.6	70.5	34.7	24.9	16.8	9.4	40.4	39.1
VA-YOLO <sup>[28]</sup>	39.0	48.2	37.4	14.8	81.2	40.8	33.0	24.7	12.4	51.4	44.8
SDS-YOLO <sup>[29]</sup>	40.0	46.7	35.4	14.4	82.0	45.1	35.8	26.5	12.7	54.3	47.4
Tph-YOLOv5 <sup>[30]</sup>	44.9	53.3	42.1	21.1	83.7	45.2	42.5	33.0	16.3	61.1	51.0
文献算法 1 <sup>[31]</sup>	40.9	49.0	41.4	13.9	82.4	42.4	35.6	26.1	13.9	52.9	51.0
FS-YOLO	47.0	53.1	43.6	18.9	85.0	51.4	41.3	35.3	21.1	<b>66.0</b>	54.9
BDH-YOLO	42.9	44.9	35.2	—	81.3	—	41.8	33.2	19.4	60.1	48.7
UAV-YOLOv8	47.0	<b>56.8</b>	44.9	18.8	<b>85.8</b>	50.8	39.0	33.3	19.7	64.3	56.2
DC-YOLOv8s	<b>48.5</b>	56.1	<b>46.1</b>	<b>23.0</b>	85.7	<b>51.8</b>	<b>42.5</b>	<b>35.5</b>	<b>21.1</b>	65.6	<b>57.5</b>

注:粗体字表示最优结果

为了更进一步验证 DC-YOLOv8s 算法的优势和有效性,与当前先进算法 Faster-RCNN、RetinaNet、YOLOv5m、YOLOv8s、BD-YOLO、文献算法 1、FS-YOLO、UAV-YOLOv8 等在 P、mAP、FPS 等指标上进行综合对比,实验结果如表 4 所示,“—”表示原文献没有提到此项指标。从表 4 可以看出,DC-YOLOv8s 有着相较于其他先进算法更高的检测精度和更低的漏检率,在精确率、召回率、平均检测精度指标上取得了最优的结果,分别为 57.7%、46.2%、48.5%。对比 YOLOv8s,DC-YOLOv8s 在 P、R、mAP 上分别提高了 7.2%、7.5%、9.1%,参数量下降 2.42%,每秒帧率 FPS 为 71 帧。相较于 YOLOv8m 更大体量的模型,DC-YOLOv8s 在 P、mAP 上提高 3.1%、5.3% 的同时,漏检率降低 4.5%,参数量降低 15.01 M,兼顾精度和模型的轻量化。对比 Faster-RCNN 和 RetinaNet 两阶段算法,DC-YOLOv8s 在 P、R、mAP 上均提升了 20% 以上,每秒帧率 FPS 提升明显。对比基于 Anchor-Based 的 YOLOv5m 算法,R、mAP 提升了 8.7%、11.2%,FPS 提升 16 帧。对比基于 Anchor-Free 的

YOLOX,P、mAP 提升了 9.1%、8.2%,FPS 提升 13 帧。对比较新现有改进文献算法 BD-YOLO、VA-YOLO,DC-YOLOv8s 在参数量小幅增加的情况下,平均检测精度提升幅度较大,均在 9.5% 以上,并且每秒帧数 FPS 较高为 71 帧。对比改进 YOLOv7-tiny 的文献算法 1,在参数量减少 3.65 M 的情况下,精确率和平均检测精度分别提升 6.5%、7.6%,FPS 提升 35 帧。对比最新改进 YOLOv8 的文献算法 FS-YOLO、UAV-YOLOv8,DC-YOLOv8s 在 R、mAP 上均有不同程度的提升,R 最大提高 1.5%,mAP 最大提高 1.5%,而且 FPS 也相对较高。综合表 3 与表 4 的实验结果表明 DC-YOLOv8s 在满足航拍目标检测任务实时性的要求下,显著提高了复杂场景下小目标检测的精度,检测效果优于经典主流算法、较新现有改进算法。并且,对比已有最新改进 YOLOv8 的文献算法,DC-YOLOv8s 在精确率、召回率、平均检测精度指标上处于领先地位,可行性较高,在各个类别检测性能之间做到了较好的平衡,综合性能最优,为无人机航拍目标检测任务面临的难题提出了一个较好的解决方案。

表 4 不同先进算法在 VisDrone2019 数据集上综合指标对比

算法	P/%	R/%	mAP/%	Parameters/M	FPS
Faster R-CNN	31.2	24.3	21.7	136.3	18
RetinaNet	20.5	16.8	13.9	48.60	14
YOLOv4	50.3	41.0	40.4	52.59	54
YOLOv5s	42.4	35.3	33.0	7.03	60
YOLOv5m	49.0	37.5	37.3	20.9	55
YOLOX	48.6	41.5	40.3	8.05	58
YOLOv7-tiny	46.6	37.5	35.5	6.03	73
YOLOv8s	50.5	38.7	39.4	11.12	105
YOLOv8m	54.6	41.7	43.2	25.86	69
BD-YOLO	—	—	32.6	7.86	—
VA-YOLO	51.63	—	39.0	—	75
SDS-YOLO	—	—	40.0	1.8	36
文献算法 1	51.2	41.9	40.9	14.5	36
FS-YOLO	—	44.7	47.0	5.48	68
BDH-YOLO	—	41.6	42.9	9.39	—
UAV-YOLOv8	54.4	45.6	47.0	10.3	51
DC-YOLOv8s	57.7	46.2	48.5	10.85	71

### 3.6 实验效果分析

为了验证 DC-YOLOv8s 算法在实际场景中的效果,选取 VisDrone2019 测试集中难度较大的图片在高空小目标、远距离密集小目标、夜间光照复杂场景、较大目标四个代表性场景进行检测。检测效果对比图如图 8 所示,左边为 YOLOv8s 检测效果图,右边为 DC-YOLOv8s 检测效果图。从图 8 可以看出,DC-YOLOv8s 在高空人群小目标下,对小目标尺度更加敏感,能够检测检测出更多和更远

的行人,有效降低了漏检率。在远距离密集小目标下,DC-YOLOv8s 也能够有效提取小目标感受野特征,将其从拥有多尺度信息区域中分离出来,保留更多细节信息,将更多远距离汽车和微小行人检测出来。在夜间光照复杂场景下,由于目标和背景之间的对比度减弱,导致模型无法正确定位和识别目标,YOLOv8s 出现了对行人和公交车的漏检,DC-YOLOv8s 则正确的将行人和公交车检测。在较大目标场景下,YOLOv8s 错误地将蓝色厂棚房识别为

卡车,而本文算法不仅没有出现这个误检,而且将其他大目标类别也精准识别。与 YOLOv8s 相比,本文算法 DC-YOLOv8s 在平衡大目标检测精度的同时,显著提高了航

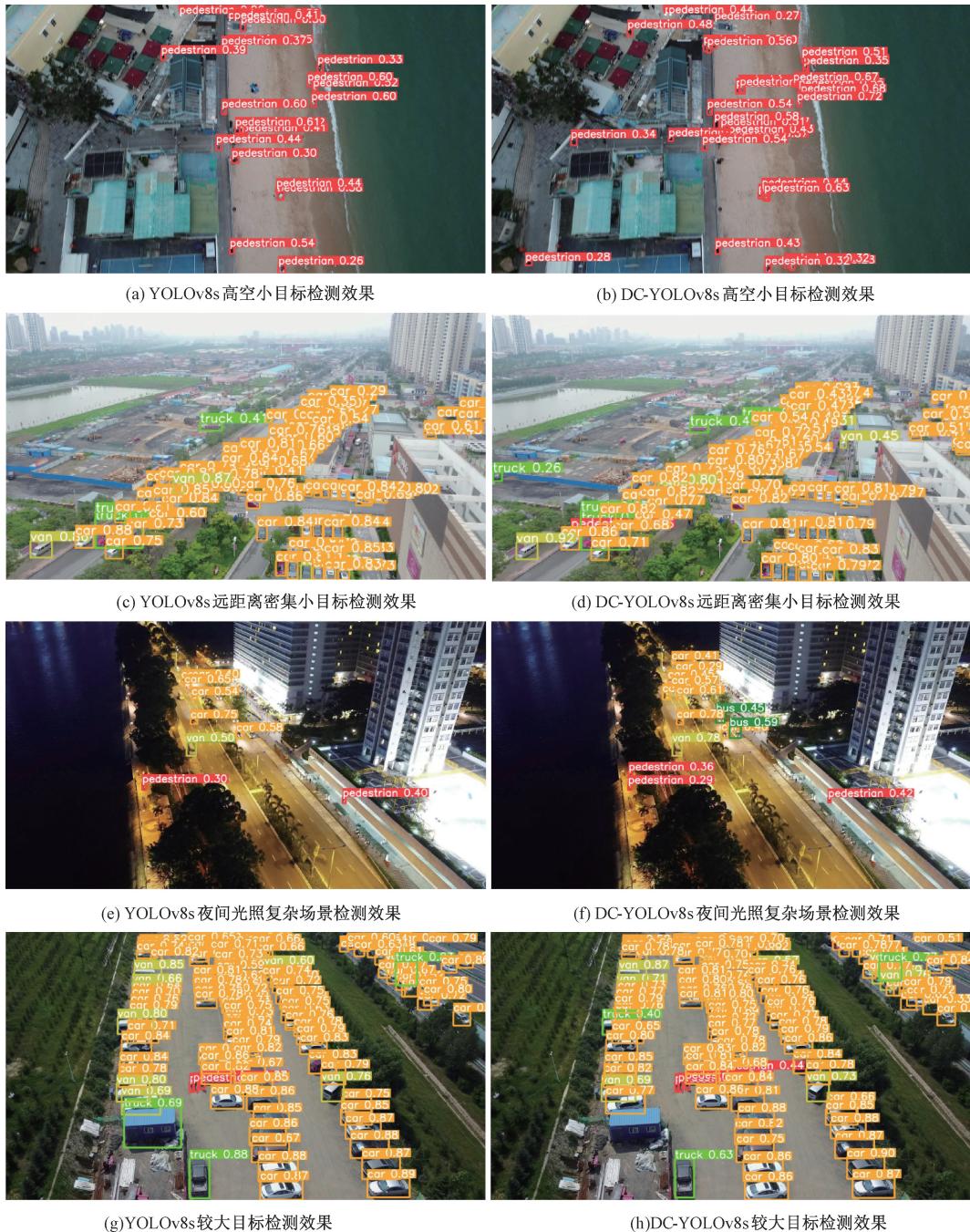


图 8 检测效果对比图

### 3.7 模型泛化性实验验证

为了验证本文提出的算法具有适用性、正确性在其他数据集上表现优异,使用 VOC2007+2012 数据集进行模型泛化性实验验证,实验训练参数设置和消融实验、VisDrone2019 对比实验设置完全一样。将本文算法 DC-YOLOv8s 和 YOLOv8s、YOLOv7-tiny、YOLOv5s、YOLOv4-

拍小目标在密集场所、光照复杂场景,较大目标等场景的特征提取能力和检测性能,漏检和误检进一步减少,在航拍小目标检测任务上表现优越。



tiny、Gold-YOLO 先进算法在 P、R、mAP、Parameters、FPS 五个指标方面进行对比分析。Gold-YOLO 是华为诺亚方舟实验室于 2023 年提出的一阶段目标检测算法,不同于以往许多研究通过修改结构、增加数据和设计新的损失,将基线推到较高的水平,Gold-YOLO 提供了一种先进的 GD 机制,该机制通过卷积和自关注操作实现,提高了模型

多尺度特征融合能力，并在所有模型尺度上实现了延迟和准确性之间的平衡。DC-YOLOv8s 和 Gold-YOLO 都是基于 YOLO 思想的算法，具备一阶段目标检测算法的优势。不同之处在于，Gold-YOLO 适用于通用目标检测，而 DC-YOLOv8s 是专门针对航拍小目标检测任务而设计的算法，针对性更强，效果更好。

模型泛化性实验结果如表 5 所示。从表 5 可以看出，DC-YOLOv8s 在 P、R、mAP 三个重要指标上取得了最好的成绩，分别为 81.3%、73.1%、80.7%。相较于次优的 YOLOv8s，精确率 P 和平均检测精度 mAP 提升了 1.6%、2.6%，参数量下降了 2.51%。对比 Gold-YOLO，在 P、R、mAP 上分别提高了 3.7%、3.9%、4.5%，参数量下降 2.79 M。相较于 YOLOv4-tiny、YOLOv7-tiny 这种轻量级检测算法，P、mAP 分别提高了 17.2%、18.0% 和 8.9%、8.6%，参数量小幅上升。FPS 虽然降至 78 帧，但是满足实时性的要求，以小幅参数量的提升换取检测精度的较大提升，属于合理范围。综合上述实验表格的结果，充分说明本文改进算法不仅适用于航拍小目标检测，也适用于其他场景下的目标检测任务，具有通用性和较好的泛化性。

表 5 模型泛化性对比实验

算法	P/%	R/%	mAP/%	Parameters	FPS
YOLO4-tiny	64.1	59.3	62.7	6.36	161
YOLOv5s	74.2	63.4	68.7	7.06	83
YOLOv7-tiny	72.4	68.2	72.1	6.06	112
YOLOv8s	79.7	70.4	78.1	11.14	103
Gold-YOLO <sup>[32]</sup>	77.6	69.2	76.2	13.65	89
DC-YOLOv8s	81.3	73.1	80.7	10.86	78

## 4 结 论

针对航拍图像小目标背景复杂、误检和漏检率高、难以检测，本文改进 YOLOv8s，提出了一种基于位置感知和跨层特征融合的航拍小目标检测算法 DC-YOLOv8s。针对小目标尺度小，难以定位，新增一个小目标检测层，提高对小目标尺度的敏感性，提高检测精度。为了减少特征信息的丢失，设计跨层特征融合模块 CLFFM，捕捉空间和通道之间的复杂联系，充分利用 Backbone 中小目标丰富的特征信息，强化浅层语义信息和深层语义的融合，最大化保留小目标特征。结合可变性卷积，设计了基于位置感知融入残差的感受野模块，更好的适应小目标的形状，优先提取感受野特征，降低漏检和误检率。使用基于注意力机制的动态检测头，提升小目标边界定位的准确性，提高检测性能。最后，在 VisDrone2019 数据集上，DC-YOLOv8s 在七个类别上取得了最佳的检测性能并且平均检测精度最高。效果优于其他先进算法，显著提高了小目标的检测能力。在 VOC2007+2012 数据集上，平均检测精度相较于

YOLOv8s 提高了 2.6%，模型适用性较好。实验表明改进方法不仅显著提升了航拍小目标检测的精度，而且在其他公开数据集上泛化性较好，具有通用性。

## 参考文献

- [1] 黄郑, 卞尹蕾, 王红星, 等. 无先验地图条件下电力管廊无人机自主巡检方法研究[J]. 电子测量技术, 2022, 45(19): 30-35.
- [2] 刘晓霞, 杨风暴, 卫红, 等. 一种农田作物无人机遥感影像阴影检测算法[J]. 电子测量技术, 2022, 45(2): 135-139.
- [3] 张君轶, 陈仁文, 刘飞, 等. 基于嵌入式视觉的无人机目标定位系统[J]. 国外电子测量技术, 2023, 42(2): 171-176.
- [4] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [5] REN S, HE K, 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.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision Heidelberg: Springer, 2016: 21-37.
- [7] TSUNG Y L, PRIYA G, ROSS G, et al. Focal loss for dense object detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [8] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6517-6525.
- [9] REDMON J, FARHADI A. Yolov3: An incremental improvement [J]. ArXiv Preprint, 2018, ArXiv: 1804. 02767.
- [10] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection [J]. ArXiv Preprint, 2020, ArXiv: 2004. 10934.
- [11] 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, 2016: 779-788.
- [12] 吴旭红, 赵清华. 基于改进 YOLOv7 的无人机航拍图像目标检测[J/OL]. 电光与控制, 1-9 [2023-12-27] http://kns. cnki. net/kcms/detail/41. 1227. TN.

- 20231031.1430.002.html.
- [13] 郑有凯,胡君红,田春欣. 基于自适应特征融合和任务对齐的小目标检测算法[J/OL]. 激光杂志,1-8[2023-12-27] <http://kns.cnki.net/kcms/detail/50.1085.TN.20231012.1338.018.html>.
- [14] 吴明杰,云利军,陈载清,等. 改进YOLOv5s的无人机视角下小目标检测算法[J/OL]. 计算机工程与应用,1-12[2023-12-27] <http://kns.cnki.net/kcms/detail/11.2127.tp.20230920.1219.042.html>.
- [15] 熊娟,张孙杰,阙亚亚,等. 基于CAFPN和细化双头解耦的遥感图像目标检测[J]. 应用科学学报,2023,41(6): 989-1003.
- [16] 程换新,乔庆元,骆晓玲,等. 基于改进YOLOv8的无人机航拍图像目标检测算法[J/OL]. 无线电工程,1-10[2024-02-23]. <http://kns.cnki.net/kcms/detail/13.1097.TN.20231127.1157.002.html>.
- [17] 邓天民,程鑫鑫,刘金凤,等. 基于特征复用机制的航拍图像小目标检测算法[J/OL]. 浙江大学学报(工学版),1-12[2024-02-23]. <http://kns.cnki.net/kcms/detail/33.1245.T.20240219.1245.010.html>.
- [18] WANG G, CHEN Y, AN P, et al. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios[J]. Sensors, 2023, 23(16): 7190.
- [19] SUI J, CHEN D, ZHENG X, et al. A new algorithm for small target detection from the perspective of unmanned aerial vehicles[J]. IEEE Access, 2024.
- [20] PANDEY S, CHEN K F, DAM E B. Comprehensive multimodal segmentation in medical imaging: Combining YOLOv8 with SAM and HQ-SAM models [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 2592-2598.
- [21] DAI X, CHEN Y, XIAO B, et al. Dynamic head: Unifying object detection heads with attentions[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 7369-7378.
- [22] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 764-773.
- [23] ZHU X, HU H, LIN S, et al. Deformable convnets v2: More deformable, better results[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9300-9308.
- [24] WOO S, PARK J, LEE J Y, et al. Cbam:
- Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision, 2018: 3-19.
- [25] LIU S, HUANG D. Receptive field block net for accurate and fast object detection[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 385-400.
- [26] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021 [J]. ArXiv Preprint, 2021, ArXiv:2107.08430.
- [27] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [28] 刘展威,陈慈发,董方敏. 基于YOLOv5s的航拍小目标检测改进算法研究[J]. 无线电工程, 2023, 53(10): 2286-2294.
- [29] 王恒涛,张上,陈想,等. 轻量化无人机航拍目标检测算法[J]. 电子测量技术, 2022, 45(19): 167-174.
- [30] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 2778-2788.
- [31] 张光华,李聪发,李钢硬,等. 基于改进YOLOv7-tiny的无人机航拍图像小目标检测算法[J/OL]. 工程科学与技术, 1-14 [2024-02-24]. <https://doi.org/10.15961/j.jsusee.202300593>.
- [32] WANG C, HE W, NIE Y, et al. Gold-YOLO: Efficient object detector via gather-and-distribute mechanism [J]. ArXiv Preprint, 2023, ArXiv: 2309.11331.

## 作者简介

- 雷帮军**,博士,博士生导师,主要研究方向为计算机视觉、图像处理、人工智能、模式识别。  
E-mail:Bangjun.Lei@ctgu.edu.cn
- 余翱**,硕士研究生,主要研究方向为计算机视觉、图像处理。  
E-mail:1056002405@qq.com
- 吴正平**(通信作者),博士,讲师,主要研究方向为图像处理、人工智能、模式识别。  
E-mail:34673637@qq.com
- 余快**,硕士研究生,主要研究方向为图像处理、遥感影像建筑物提取。  
E-mail:2993846564@qq.com