DI-YOLO: 一种面向无人机航拍图像的高效小目标检测框架

丁浩晗[1,2†]，贺万程[2]，万 俊[2]，沈易航[2]，崔晓晖[3]

(1. 江南大学 未来食品科学中心，江苏 无锡 214122；
2. 江南大学 人工智能与计算机学院，江苏 无锡 214122；
3. 武汉大学 国家网络安全学院，武汉 430072)

摘 要: 针对航拍图像中目标尺寸微小、纹理特征模糊以及分布密集带来的检测难题, 提出一种基于改进YOLO 架构的 DI-YOLO 检测模型. 当前主流检测方法在微小目标结构信息保留以及多尺度特征提取和融合方面存在明显不足. 鉴于此, 构建内容感知特征增强模块 (CARAFE), 通过动态特征选择机制实现跨层级特征的自适应融合; 同时, 设计并行异构特征调制模块 (PHFM), 有效协调全局上下文建模与局部细节特征的关联性; 并引入形状感知交并比损失函数 (Shape-IoU) 和微小目标检测头, 进一步提升边界框回归精度和微小目标检测能力. 在 VisDrone2019 和 DOTAv1.5 基准数据集上的对比实验结果表明, 所提出模型较基准模型 YOLOv10 取得显著提升: 在 VisDrone2019 数据集上, mAP@0.5 和 mAP@0.5 : 0.95 指标分别提升了 12.7% 和 13.7%; 在 DOTAv1.5 数据集上, 对应提升了 12.1% 和 10.2%, 且在计算效率方面保持优势. 消融实验进一步验证了各模块的有效性, 为航拍场景下的高精度目标检测提供了新的解决方案.

关键词: 深度学习；异构调制器；跨层级特征融合；YOLO；计算机视觉；航拍微小目标检测

中图分类号: TP391.41 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2025.0425

引用格式: 丁浩晗, 贺万程, 万俊, 等. DI-YOLO: 一种面向无人机航拍图像的高效小目标检测框架 [J]. 控制与决策, 2025, 40(10): 3106-3116.

DI-YOLO: An efficient small object detection framework for UAV aerial imagery

DING Hao-han[1,2†]，*HE Wan-cheng*[2]，*WAN Jun*[2]，*SHEN Yi-hang*[2]，*CUI Xiao-hui*[3]

(1. Science Center for Future Food，Jiangnan University，Wuxi 214122，China；2. School of Artificial Intelligence and Computer Science，Jiangnan University，Wuxi 214122，China；3. School of Cyber Science and Engineering，Wuhan University，Wuhan 430072，China)

Abstract: Addressing the detection challenges posed by small target sizes, blurred texture features, and dense distributions in aerial imagery, this research proposes a detection model based on an improved YOLO architecture, named drone imagery YOLO (DI-YOLO). Current mainstream detection methods show significant deficiencies in preserving structural information of small targets and multi-scale feature extraction and fusion. Therefore, we innovatively construct a content-aware reassembly of features (CARAFE) module, achieving adaptive fusion of cross-level features through a dynamic feature selection mechanism; simultaneously design a parallel heterogeneous feature modulator (PHFM) that effectively coordinates the relationship between global context modeling and local detail features; and introduce a shape-aware intersection over union (Shape-IoU) loss function and a tiny object detection he ad to further enhance bounding box regression accuracy and small target detection capabilities. Through comparative experiments on the VisDrone2019 and DOTAv1.5 benchmark datasets, the proposed model achieves significant improvements over the baseline YOLOv10 model: on the VisDrone2019 dataset, mAP@0.5 and mAP@0.5 : 0.95 metrics improve by 12.7% and 13.7%, respectively, while on the DOTAv1.5 dataset, corresponding improvements of

收稿日期: 2025-04-22；**录用日期:** 2025-06-17.
基金项目: “十四五”国家重点研发计划重点专项项目 (2024YFE0199500, 2022YFF1101100).
责任编委: 林志赟.
†通信作者. E-mail: dinghaohan@jiangnan.edu.cn.
本文附带电子附录文件, 可登录本刊官网该文“资源附件”区自行下载阅览.

12.1% and 10.2% are achieved, with maintained advantages in computational efficiency. Ablation experiments further verify the effectiveness of each module, providing a new solution for high-precision object detection in aerial scenes.
Keywords: deep learning; heterogeneous modulation; cross-level feature fusion; YOLO; computer vision; aerial small object detection

0 引 言

航空图像目标检测作为计算机视觉领域的关键研究方向, 在环境监控[1]、智慧城市管理[2]、农业灌溉[3]、灾害评估和应急救援等多领域展现出重要的应用价值和学术研究意义. 近年来, 随着航空遥感技术的迅速迭代和普及, 对空中获取图像的信息提取需求呈现指数级增长. 高精度目标检测算法不仅是实现航拍图像精确定位和追踪的技术基础, 更是推动无人航空系统智能化发展的核心支撑. 相较于常规自然图像处理, 无人机航拍图像的目标检测面临多重独特挑战: 高空拍摄的图像可利用的识别特征相对有限, 存在目标尺寸小、纹理模糊、分布密集、背景复杂多变等突出问题, 这些特性对实际应用环境下的检测算法设计和优化提出了严峻考验[4].

深度学习技术的突破性进展推动了目标检测领域的范式革新. 当前基于深度学习的检测算法根据检测阶段可分为两阶段和单阶段两类架构[5]: 两阶段检测算法采用“区域提议-分类回归”的递进策略, 以 Fast R-CNN、Faster R-CNN、Mask R-CNN 等为代表, 通过区域提议网络生成候选框后再进行精细化处理[6]; 而以 YOLO、SSD 为代表的单阶段算法[7] 采用端到端架构, 直接通过卷积特征层实现目标定位与分类的同步预测, 在实时性要求高的监测系统中展现出显著优势.

YOLO 目标检测模型凭借其高精度和高效率的显著优势, 已成为该领域最具代表性的算法体系. 自 2015 年首次提出初代架构[8] 以来, 该系列算法历经多次重大革新: YOLOv2-v8 通过引入批量归一化、锚框机制、多尺度预测等技术, 在精度与速度间实现了良好平衡[9]; YOLOv10 通过无 NMS 检测架构[10], 首次实现了真正端到端的目标检测范式; YOLOv11 进一步提出了动态稀疏计算框架和多模态预训练技术, 显著提升了开放场景泛化能力[11]. 而在航空图像目标检测领域, 多位研究者针对 YOLO 架构进行了多项适应性改进: 张攀峰等[12] 提出的 DD-YOLO 通过跨阶段四分支模块构建了轻量化双主干架构, 融合训练策略显著提升了无人机视角下小目标检测性能; 武腾辉等[13] 研发的 LS-YOLO 设计了双路径特征增强机制, 通过改进的跨层特征金字塔与空间语义融合模块捕获多尺度上下文信息, 有效缓解了复杂背景干扰; 范博淦等[14] 基于 YOLOv8 框架采用了自适应特征校准机制动态调整通道权重, 并通过 Inner-EIoU 损失函数优化了边界框回归. 然而, 这些方法在特征上采样的内容感知能力和多尺度特征的并行异构融合方面仍然存在显著局限.

首先, 传统上采样方法 (最近邻插值、双线性插值) 仅依赖像素间的空间距离关系, 完全忽视了像素间的语义关联[8,15-16], 这在需要精细结构保持的小尺度目标检测任务中尤为关键. 尽管转置卷积提供了可学习的上采样方式, 但是, 其均匀卷积核设计难以有效应对特征图不同区域的局部内容差异, 导致复杂背景中小目标检测的细节信息损失[17]. 航空图像中的小目标往往具有复杂的局部结构特征, 需要上采样过程能够根据内容自适应地保持和增强这些关键信息. 其次, 是多尺度特征融合的异构处理不足的问题: 在特征融合层面, 现有注意力机制虽然在改进特征表示方面取得了进展, 但是, 在航空小目标检测中仍然面临显著局限. 自 SENet 首次建立通道注意力机制以来, CBAM 融合了通道与空间注意力, GAM 通过跨维度交互突破了传统局限, SimAM 探索了无参注意力方向[18]. 然而, 这些机制多采用串行处理结构或单一注意力形式, 难以同时有效处理航空图像中的全局上下文感知和局部细节捕获[19]. 航空小目标检测恰恰需要全局战场态势理解与局部精细特征的有机结合: 全局信息用于理解目标所处的复杂环境背景, 而局部细节则决定了小目标的准确识别. 更为关键的是, 传统注意力机制在处理尺寸悬殊、数量众多的多尺度目标时, 注意力分配策略往往偏向于大尺寸或高对比度目标, 导致对微小目标的特征表示严重不足. 这种偏向性问题在航空图像的密集小目标场景中表现得尤为突出, 亟需一种能够平衡不同尺度目标重要性的并行异构特征调制机制.

为优化这些核心技术的挑战, 本文提出 DI-YOLO, 这是一种专门为航空影像的目标识别任务设计的增强型算法. 本文主要内容包括以下 3 个方面:

1) 使用内容感知特征增强上采样模块 (CARAFE), 该模块能够根据输入内容动态调整特征融合策略, 显著提升模型保留小尺度目标结构信息

的能力, 有效应对真实环境下目标识别的鲁棒性问题.

2) 设计一种新颖的并行异构特征调制模块(PHFM). 该模块能够有效平衡全局战场态势感知与局部特征表示, 在捕获长程依赖关系和细粒度目标细节方面均能够取得优异的表现, 为高精度信息提取提供技术支持. 同时, 本文重新设计颈部结构, 加入形状感知交并比损失函数和超小目标检测头, 在超小分辨率目标的捕获场景上可获得重大提升.

3) 在 VisDrone2019 和 DOTAv1.5 数据集上的大量实验验证了所提出方法在航拍目标检测方面具有显著效果. 具体而言, 与基准 YOLOv10 相比, DI-YOLO 表现出显著提升: 在 VisDrone2019 数据集上, mAP@0.5 和 mAP@0.5 : 0.95 指标分别提升了 12.7% 和 13.7%; 在 DOTAv1.5 数据集上, 对应提升了 12.1% 和 10.2%. 这些提升对于航空遥感领域的智能信息处理具有重要学术意义和应用价值.

1 DI-YOLO 无人机航拍图像识别模型

针对前文分析的现有方法在特征融合和上采样方面的关键局限, 本文提出 DI-YOLO 模型作为系统解决方案. 如图 1 所示, 该模型通过两个核心创新模块直接应对这些技术挑战: 内容感知特征增强上采样模块 (CARAFE) 专门改善传统上采样方法缺乏语义感知的问题, 并行异构特征调制模块 (PHFM) 则优化了现有特征融合机制在全局-局部特征表示平衡方面的固有局限. 这两个模块共同构成了 DI-YOLO 的技术核心, 为航空小目标检测提供了更优的特征表示和融合机制.

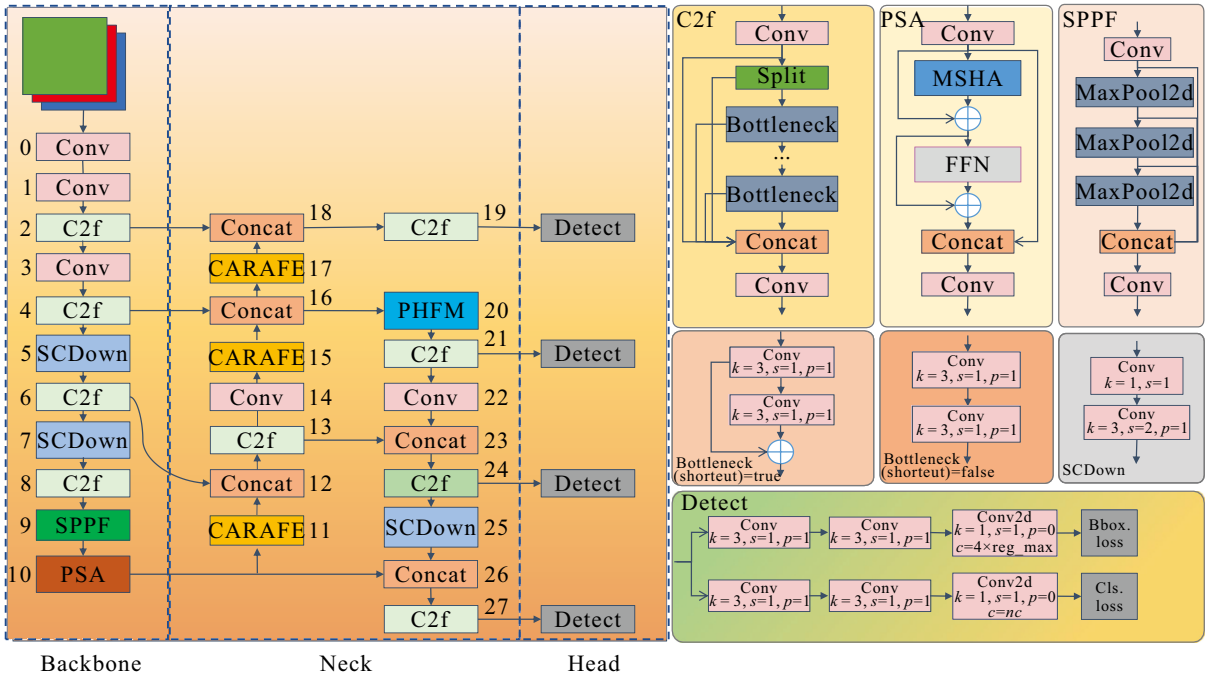

图1 DI-YOLO 模型结构

为平衡检测精度与实时性能, 本文选择 YOLOv10 作为 DI-YOLO 的基础架构. DI-YOLO 由 3 个关键组件组成: 特征提取骨干网络、增强型颈部网络和检测头. 在骨干网络中, 本文采用 YOLO 的原始架构: 利用 C2f 和 SCDown 模块进行特征提取, 并在末端集成 SPPF 和 PSA 模块以增强特征表示.

对于颈部网络, 本文对原始结构进行了重大改进. 原始 YOLOv10 采用标准 FPN 结构, 使用简单的最近邻上采样 (nn.Upsample) 和 $P_3/P_4/P_5$ 三尺度检测. DI-YOLO 对颈部进行了系统性重构: 首先, 用内容感知特征增强上采样模块替代了传统的简单上采样操作, 通过自适应特征重组改进了特征表示; 然后, 在 P_3 特征融合后的关键位置插入所设计并行异构特征调制模块; 最后, 新增 P_2 检测头形成四尺度检测体系.

在损失函数设计方面, 本文引入了形状感知交并比 (Shape-IoU) 机制, 对传统边界框回归损失进行了改进. Shape-IoU 在标准 IoU 的基础上增加了中心距离惩罚和形状差异惩罚, 使得模型能够更精确地捕捉航拍图像中形状各异的目标, 有效提升了边界框拟合精度.

检测头采用经典的解耦设计结构并添加了一个微小目标检测头, 负责将增强的多尺度特征转换为目标位置、类别和边界框坐标的预测, 并结合 Shape-

IoU 损失进行优化, 实现了高效准确的目标检测.

1.1　内容感知特征增强上采样模块

为改善传统上采样方法缺乏语义内容感知能力的关键局限, 本文引入了内容感知特征增强上采样模块替换上采样结构. 内容感知特征增强上采样模块结构如图 2 所示, 是一种新型的上采样模块, 实现了特征的自适应上采样和增强. 给定输入特征图 $\boldsymbol{X} \in \mathbb{R}^{C\times H\times W}$, 该模块通过两个关键的功能子模块输出上采样后的特征图$\boldsymbol{Y} \in \mathbb{R}^{C\times \sigma H\times \sigma W}$, 其中$\sigma$为上采样系数. 需要说明的是, CARAFE 涉及两个关键参数: k_{encoder}为内核预测模块的编码器卷积核大小 (设置为 3), 用于提取局部特征信息: 选择3×3卷积核基于其在特征编码中的最优平衡性, 既能提供足够的局部空间感受野以捕获小目标的边缘和纹理特征, 又能保持较低的计算复杂度; k_{up}为重组邻域大小 (设置为 5), 定义了每个输出位置聚合特征的输入邻域范围. 这种设计使得 CARAFE 能够在保持计算效率的同时, 可实现比传统插值方法更大的感受野.

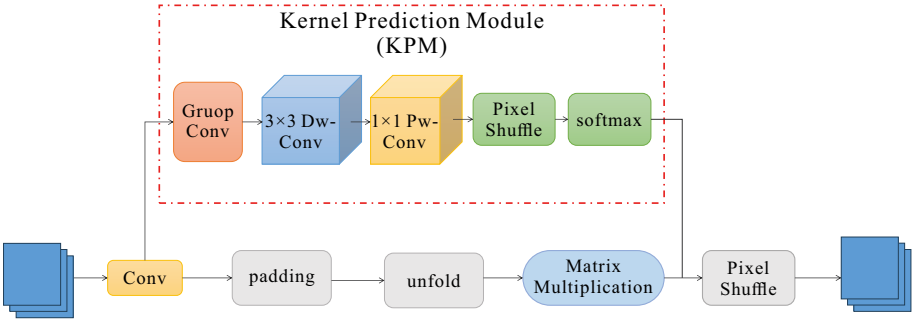

图2　内容感知特征增强上采样模块结构示意图

CARAFE 模块的核心设计包含以下两个互补的功能组件.

1.1.1　内核预测模块

该模块首先通过通道压缩器将输入特征的通道数从C压缩至C_m(在本文的实现中设置为$C/4$), 以提高计算效率; 然后, 内容编码器利用大小为k_{encoder}的卷积核对压缩后的特征进行编码, 生成用于特征重组的自适应内核; 最后, 通过内核归一化器对预测的内核进行 softmax 归一化处理, 以确保重组权重的有效性. 整个预测过程可表示为

$$\boldsymbol{W} = \text{Normalize}(\mathcal{P}(\boldsymbol{X})). \tag{1}$$

其中: $\mathcal{P}$为内核预测网络, $\boldsymbol{W}$为归一化后的重组内核.

在本文的具体实现中, 内核预测模块采用了组卷积和深度可分离卷积的优化策略, 将预测过程分为以下步骤.

step 1: 通过组卷积进行高效的通道降维, 如下所示:

$$\boldsymbol{X}_c = \text{GroupConv}_{1\times 1}(\boldsymbol{X}). \tag{2}$$

step 2: 使用深度可分离卷积提取空间特征, 如下所示:

$$\boldsymbol{X}_d = \text{DWConv}_{k\times k}(\boldsymbol{X}_c). \tag{3}$$

step 3: 通过点卷积生成上采样核, 即

$$\boldsymbol{K} = \text{Conv}_{1\times 1}(\boldsymbol{X}_d). \tag{4}$$

其中: $k = k_{\text{encoder}}$为编码器卷积核大小, 输出通道数为$\sigma^2 \times k_{\text{up}}^2$, 这里$\sigma$为上采样系数.

step 4: 通过像素重排和归一化操作生成重组权重, 如下所示:

$$\boldsymbol{W} = \text{Softmax}(\text{PixelShuffle}(\boldsymbol{K})). \tag{5}$$

这种设计使得模块能够根据输入特征的语义内容动态生成最适合当前区域特征重组的权重, 对于不同特征图区域中的不同目标模块会自适应地生成不同的重组策略, 以最大程度保留目标的结构细节.

1.1.2　内容感知重组模块

基于预测的自适应内核, 该模块在预定义的局部区域内对特征进行重组. 对于目标位置p, 其特征重组过程可表示为

$$\boldsymbol{Y}(p) = \sum_{q\in \Omega_{k_{\text{up}}}(p)} \boldsymbol{W}(p,q)\cdot \boldsymbol{X}(q). \tag{6}$$

其中: $\Omega_{k_{\text{up}}}(p)$是以$p$为中心的$k_{\text{up}}\times k_{\text{up}}$邻域, $\boldsymbol{W}(p,q)$为对应的重组权重.

重组模块的核心在于其自适应特征融合策略, 具体表现在以下几个方面: 首先, 不同于固定权重的传统上采样方法, CARAFE 根据输入特征的语义内容动态调整每个位置的融合权重. 如: 对于航拍图像中的小目标边缘区域, 模块会赋予更高的权重以保留结构细节; 而对于语义不重要的背景区域, 则可能采用更平滑的权重分布. 然后, 通过矩阵乘法实现的特征重组操作允许模块考虑邻域信息的空间分布. 对于航拍图像中的密集小目标区域, 重组模块能够根据目标的空间关系自适应地调整特征融合策略,

有效区分相邻的小目标实例. 最后, 结合像素重排技术, 重组模块能够在保留细节的同时有效整合多尺度特征信息. 这对于处理航拍图像中不同尺度的目标尤为重要.

为优化计算效率, 本文在实现上采用了多项优化策略: 在特征调整层使用点卷积以减少计算量; 在核预测模块中采用深度可分离卷积替代标准卷积; 在特征重组过程中, 通过等效卷积操作来实现特征展开和重组, 有效降低了内存开销. 这些优化确保了模块在提供强大功能的同时保持较低的计算复杂度.

1.2 并行异构特征调制模块

近年来, 计算机视觉研究表明, 全局-局部特征表示的权衡优化仍然是一个关键挑战. 传统卷积神经网络 (CNNs) 虽然在局部特征提取方面表现出色, 但是, 其固有的感受野限制制约了对长程依赖关系的建模. 相反, 注意力机制虽然能够有效捕获全局上下文信息, 却往往在保持局部精细特征方面存在不足. 为突破这一瓶颈并改善现有注意力机制在航空小目标检测中的局限, 本文提出并行异构特征调制模块. 该模块通过并行多流架构和自适应特征调制机制, 实现了对全局-局部空间特征的高效整合和优化表征, 在抑制冗余信息的同时突出任务相关特征, 从而在跨尺度目标定位任务中取得了显著性能提升.

如图 3 所示, 并行异构特征调制模块采用创新性的三流并行架构设计, 包含 (a) 局部注意力单元、(b) 全局注意力单元、门控单元. 局部单元通过软池化和多层卷积提取局部特征, 全局单元利用自适应注意力机制捕获长程依赖, 门控流基于输入的首个通道生成调制信号. 这种多流架构通过特征融合模块来实现局部-全局特征的自适应整合, 并通过门控机制和残差连接动态调节信息流动. 该结构不仅能够同时捕获全局和局部空间特征, 还能进一步优化特征表示和信息流动, 为特征的有效融合提供了保障. 通过巧妙结合自适应注意力机制和深度卷积特

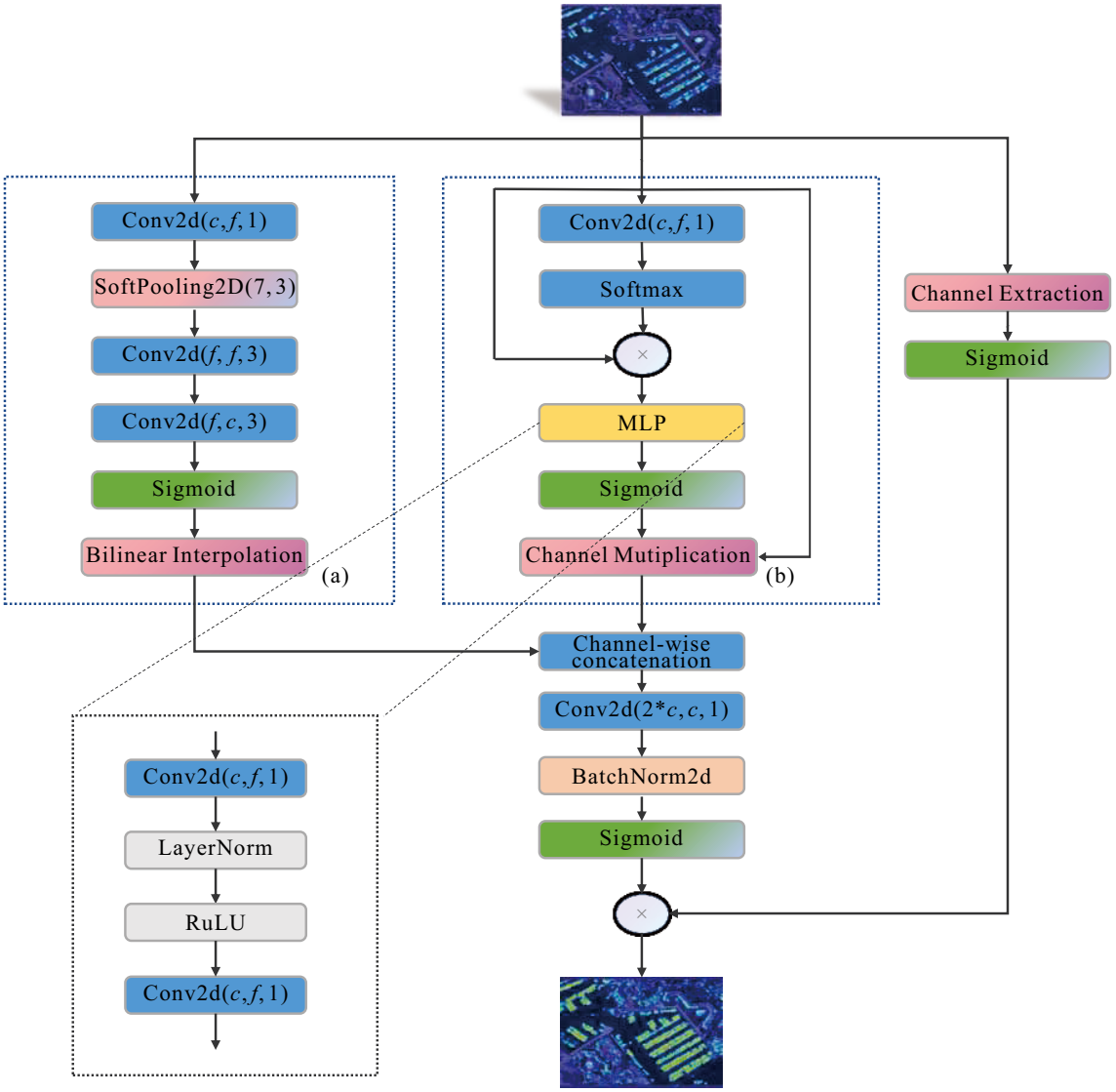

图3 并行异构特征调制模块 (PHFM) 模块结构示意图

征提取, 该架构在模型性能与计算效率间实现了最优平衡. 整体结构可形式化表示为

$$\mathrm{PHFM}(X) = X \cdot F(L(X), G(X)) \cdot T(X). \quad (7)$$

其中: 局部空间特征$L(X)$主要通过局部流进行建模, 全局空间特征$G(X)$则由全局流负责捕获. 经非线性变换后, 这些特征被融合为特征$F(L(X), G(X))$. 此外, 本文还引入了作用于第 1 通道的轻量级通道门控机制, 用于调制注意力响应, 有效防止对局部模式的过度强调. 具体而言, 该门控机制通过对输入特征的首个通道应用 sigmoid 激活函数来生成调制信号. 这种轻量级设计不仅能够有效调节特征响应强度, 还能保持较低的计算开销. 最后, 模块采用残差连接结构, 将输入特征与调制后的特征进行融合, 既保证了信息的充分传递, 又提升了模型的训练稳定性.

1.2.1　局部注意力单元

局部注意力单元专注于捕获精细的局部空间关系, 采用基于 SoftPooling 的加权特征聚合机制. 传统的局部注意力方法往往依赖简单的卷积操作或平均池化, 难以有效区分局部区域内不同特征的重要性. 为突破这一限制, 本文提出了基于 SoftPooling 操作的加权特征聚合局部流机制. 局部注意力单元的核心在于 SoftPooling 操作, 该操作通过指数加权方案实现特征重要性感知的下采样, 如下所示:

$$\mathrm{SoftPool}(x) = \frac{\sum_{i \in R} x_i \mathrm{e}^{x_i}}{\sum_{i \in R} \mathrm{e}^{x_i}}. \quad (8)$$

其中: x_i为位置i处的输入特征, R为局部池化区域. 式 (8) 通过指数项e^{x_i}自然地强化显著特征响应, 同时抑制微弱特征响应. 为提升计算效率, 加权特征在降低空间分辨率的条件下经由两个卷积层处理, 随后恢复至原始通道维度. 最后通过 sigmoid 激活将注意力权重规范化至 [0, 1] 区间.

1.2.2　全局注意力单元

全局注意力单元是 PHFM 模块捕获长程依赖关系的核心组件, 通过空间注意力与通道调制的协同作用优化特征表达. 该组件采用 AdaptiveAttentionBlock 结构, 能够在保持计算效率的同时有效建立远距离像素间的空间关联.

全局分支的长程依赖捕获能力主要体现在其自适应空间注意力池化过程. 给定输入特征$X \in \mathbb{R}^{B \times C \times H \times W}$, 该机制首先生成空间注意力掩码, 有

$$A = \mathrm{Softmax}(\mathrm{Conv}_{1\times1}(X)), \quad (9)$$

其中$\mathrm{Conv}_{1\times1}$将通道数从C降为 1, 生成注意力图$A \in \mathbb{R}^{B \times 1 \times H \times W}$. 然后, 输入特征$X$被重塑并增加维度为$X_{\mathrm{reshape}} \in \mathbb{R}^{B \times 1 \times C \times (H \times W)}$, 注意力掩码$A$被重塑为$A_{\mathrm{reshape}} \in \mathbb{R}^{B \times 1 \times (H \times W)}$. 接着, 通过矩阵乘法实现全局上下文聚合, 如下所示:

$$\mathrm{GlobalContext} = \mathrm{MatMul}(\mathrm{X}_{\mathrm{reshape}}, \mathrm{A}^{\mathrm{T}}_{\mathrm{reshape}}). \quad (10)$$

该操作将所有空间位置的特征按照注意力权重进行加权聚合, 得到$\mathrm{GlobalContext} \in \mathbb{R}^{\mathrm{B} \times \mathrm{C} \times 1 \times 1}$.

获得全局上下文信息后, 模块通过多层感知器实现通道级特征调制, 即

$$\mathrm{G(X)} = \mathrm{X} \cdot \sigma(\mathrm{MLP}(\mathrm{GlobalContext})), \quad (11)$$

其中 MLP 由两个1×1卷积层构成, 中间包含层归一化和 ReLU 激活函数. 为确保训练稳定性, MLP 的最后一层采用零初始化策略, 使得模块在训练初期保持近似恒等映射特性. 这种渐近式学习策略有助于模型在训练过程中平稳收敛.

通过这种全局注意力机制, PHFM 模块能够在航拍图像中建立起道路网络的连通性、建筑群的整体布局以及车辆集群的分布模式等长程空间关系, 为小目标检测提供了丰富的全局上下文信息.

1.3　形状感知交并比损失函数

传统 IoU 损失主要关注边界框的重叠区域, 但是, 对形状差异和中心点偏移等因素考虑不足. 为解决这一问题, 本研究选择使用形状感知交并比 (Shape-IoU) 损失函数作为检测头优化的关键组件.

Shape-IoU 在标准 IoU 的基础上, 引入了中心距离惩罚和形状差异惩罚两个重要惩罚项, 有

$$\mathrm{Shape-IoU} = \mathrm{IoU} - \mathrm{Distance} - 0.5 \cdot \mathrm{ShapeCost}, \quad (12)$$

其中中心距离惩罚项考虑边界框中心点距离, 并通过形状权重系数进行自适应调整, 如下所示:

$$\mathrm{Distance} = \frac{hh \cdot d_x^2 + ww \cdot d_y^2}{c^2}. \quad (13)$$

这里: d_x和d_y分别为两个边界框中心在水平和垂直方向上的距离, c^2为包围两个边界框最小矩形的对角线长度平方, ww和hh为根据目标框宽高比自适应计算的权重系数.

形状差异惩罚项通过非线性变换量化边界框在宽度和高度上的相对差异, 如下所示:

$$\mathrm{ShapeCost} = (1 - \mathrm{e}^{-\omega_w})^4 + (1 - \mathrm{e}^{-\omega_h})^4, \quad (14)$$

其中 $\omega_w = hh \cdot \dfrac{|w_1 - w_2|}{\max(w_1, w_2)}$和 $\omega_h = ww \times \dfrac{|h_1 - h_2|}{\max(h_1, h_2)}$分别为宽度和高度的归一化差异.

与传统 IoU、GIoU 和 CIoU 等损失函数相比,

Shape-IoU 具有显著优势: 显式考虑边界框的形状差异, 对航拍图像中形状各异的目标检测尤为重要; 通过自适应权重系数, 能够根据目标的宽高比特性动态调整惩罚强度; 非线性惩罚机制使得模型能够稳定收敛.

2 实 验

2.1 数据集与评估指标

在本研究中, 在两个基准数据集上进行了实验: VisDrone2019 和 DOTAv1.5. VisDrone2019 数据集包含 7019 张高质量的无人机拍摄图像 (6471 张用于训练, 548 张用于验证). 目标被分为 10 个类别: 行人、人群、自行车、汽车、面包车、卡车、三轮车、遮阳三轮车、公交车和摩托车. 为进一步评估所提出模型对密集分布小目标的检测能力, 本文还在 DOTAv1.5 数据集上进行了实验, 该数据集包含超过 400 000 张带有实例标注的航拍图像, 涵盖 16 个目标类别 (包括飞机、船只、储罐和各类设施). 这些数据集的组合为目标检测算法提供了严格的测试场景, 具有航空图像分析中典型的多样化视角、目标尺度和场景复杂性.

本文使用多个指标评估模型, 主要准确率指标是在两个 IoU 阈值 (0.5 和 0.5 : 0.95) 下的 mAP (平均精确度均值). mAP 的计算公式为

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{AP}_i, \tag{15}$$

$$\mathrm{AP} = \int_0^1 P(R)\mathrm{d}R. \tag{16}$$

其中: N为类别数量, P为精确度, R为召回率. 为评估模型效率, 本文还测量了计算成本 (GFLOPs)、参数数量 (Params) 和推理速度 (FPS).

2.2 实现细节

所有实验均在 Linux 操作系统环境下进行, 使用配备 CUDA 12.2 的 NVIDIA A100 GPU. 软件环境基于 Python 3.10 和 PyTorch 2.0.1 框架, 以实现高效的模型训练和推理. 在实验期间, 批量大小设为 8. 最大训练轮数设置为 250, 并实施了早停策略: 当验证指标连续 50 轮没有改善时终止训练. 为促进模型在接近实际应用场景的数据分布下实现更精确的参数优化和稳定收敛, 本研究采用了渐近式训练策略. 具体而言, 在训练的最终阶段, 有计划地停用了马赛克增强技术, 使得模型能够直接暴露于未经几何变换的原始数据分布中. 这种从强增强到弱增强的训练范式转变, 有助于模型参数从粗粒度优化逐步过渡到精细化调整, 从而提高模型在真实场景中的泛化性能和预测精度. 其他大多数实验参数遵循默认参数设置以保持实验的可重复性和可比性, 针对特定场景进行了少量调整.

2.3 消融实验

为评估所提出改进策略在无人机航拍图像检测任务中的有效性, 在 VisDrone2019 数据集上, 以 YOLOv10m 为基准模型开展一系列消融实验. 实验过程中依次在 YOLOv10m 模型的颈部加入 PHFM 模块、上采样新引入 CARAFE 模块、形状感知交并比损失函数和小目标检测头以优化模型对微小目标对象的精准定位能力. 在此基础上, 分别比较引入不同模块对于基准模型在检测精度上所产生的具体影响, 实验结果如表 1 所示.

表1 不同组件在 DI-YOLO 中的消融研究

YOLO	CARAFE	PHFM	Shape-IoU	Head	mAP@0.5	mAP@0.5 : 0.95
√					0.418	0.254
√	√				0.424	0.259
√		√			0.432	0.264
√	√	√			0.439	0.268
√	√	√	√		0.441	0.269
√	√	√	√	√	0.471	0.29

从性能数据分析, CARAFE 替代传统最近邻上采样后, mAP@0.5 从 0.418 提升至 0.424, mAP@0.5 : 0.95 从 0.254 提升至 0.259. 这种提升源于 CARAFE 的内容感知机制 —— 不同于固定权重的插值方法, CARAFE 为每个位置动态生成上采样核, 能够更好地保留航拍图像中小目标的边缘和纹理细节. 在颈部引入 PHFM 模块使得 mAP@0.5 进一步提升至 0.432. PHFM 插入在P_3层 (原图 1/8 分辨率) 这一关键位置, 既保留了足够的空间细节, 又具有适当的感受野进行全局建模. 值得注意的是, 当 CARAFE 与 PHFM 联合使用时, mAP@0.5 达到了 0.439, 超过单独使用的线性叠加效果, 表明两个模块存在协同增强效应. 最显著的性能提升来自P_2检测头的引入: mAP@0.5 跃升至 0.471, mAP@0.5 : 0.95 提升至 0.290. P_2层保持原图 1/4 分辨率, 包含了丰富的细节信息, 对于检测航拍图像中的微小目标至关重要. 这一大幅提升充分表明了高分辨率特征在小目标检测中的关键作用.

综合来看, 这种系统性提升验证了 DI-YOLO 颈部设计的先进性: CARAFE 提供高质量特征上采样, PHFM 实现有效的多尺度特征融合, P_2检测头充分利用高分辨率信息, 三者相互配合构建了一个针对航拍小目标检测高度优化的特征处理架构.

2.4　对比实验

如表 2 所示, 实验结果有力地验证了所提出 DI-YOLO 模型在航空小目标检测领域的显著优势. 在挑战性极高的 VisDrone2019 数据集上, DI-YOLO 模型在 mAP@0.5 和 mAP@0.5 : 0.95 两项关键指标上分别达到了 0.471 和 0.290, 展现出卓越的检测性能. 尽管最新的 YOLOv11、YOLOv12 以及 RTDETRv2 等模型在检测性能上有了显著提升, DI-YOLO 仍然保持了性能领先优势. 相较于 RTDETRv2 (0.460 和 0.269), DI-YOLO 分别提升了 2.4% 和 7.8%; 与 YOLOv11 (0.439 和 0.271) 相比, mAP@0.5 提升了 7.3%, 而在 mAP@0.5 : 0.95 上提升了 7.0%; 与 YOLOv12 (0.435 和 0.266) 相比, 两项指标分别提升了 8.3% 和 9.0%.

表2　不同模型在 VisDrone2019 和 DOTAv1.5 数据集上的性能对比

方法	参数量/M	GLOPs	VisDrone2019			DOTAv1.5		
			mAP@0.5	mAP@0.5 : 0.95	FPS	mAP@0.5	mAP@0.5 : 0.95	FPS
Faster-RCNN	41.39	208.0	0.329	0.194	22.6	0.310	0.180	17.4
RetinaNet	36.59	210.0	0.275	0.164	35.2	0.260	0.155	28.8
RTDETRv2	42.42	112.1	0.460	0.269	88.4	0.421	0.258	64.5
YOLOv12	19.62	60.2	0.435	0.266	223.4	0.393	0.242	116.8
YOLOv11	20.04	67.7	0.439	0.271	212.7	0.410	0.262	114.9
YOLOv10	16.50	64.0	0.418	0.255	181.3	0.381	0.236	99.0
YOLOv9	20.02	77.6	0.439	0.268	147.0	0.408	0.256	62.1
YOLOv8	25.86	79.1	0.422	0.256	161.3	0.400	0.252	80.6
YOLOv6	52.00	161.6	0.411	0.249	153.8	0.375	0.233	61.7
YOLOv5	25.07	64.4	0.422	0.255	166.7	0.391	0.244	66.7
DI-YOLO (本文)	26.14	96.3	**0.471**	**0.290**	113.4	**0.427**	**0.260**	64.5

在 DOTAv1.5 数据集的评估中, 观察到了类似的性能优势模式. DI-YOLO 分别在 mAP@0.5 和 mAP@0.5 : 0.95 指标上表现优异 (0.427 和 0.260), 与最具竞争力的 RTDETRv2 相比仍然保持微弱领先. 这一致性的性能表现有力地验证了所提出方法在不同航空影像场景下的泛化能力和鲁棒性.

从计算效率和实时性角度分析, 虽然 YOLOv12 在 VisDrone2019 上达到了 223.4 FPS 的最高推理速度, 但是其检测精度显著低于 DI-YOLO. 所提出模型成功地在保持可接受的实时处理能力 (113.4 FPS) 的同时, 实现了更高的检测精度, 这种精度与效率的平衡对于航空小目标检测这类对精度要求极高的实际应用场景尤为重要.

如图 4 所示, DI-YOLO 在检测性能上实现了全面提升, 尤其是在处理具有挑战性的小目标检测任务时表现突出. 以行人和摩托车类别为例, 相较于现有 YOLO 系列模型, 所提出方法取得了显著的性能改进. 在中等尺度目标 (如汽车和面包车) 的检测中也展现出明显优势, 同时, 对大目标的检测能力也有所增强. 这种跨尺度的性能提升主要得益于两个方

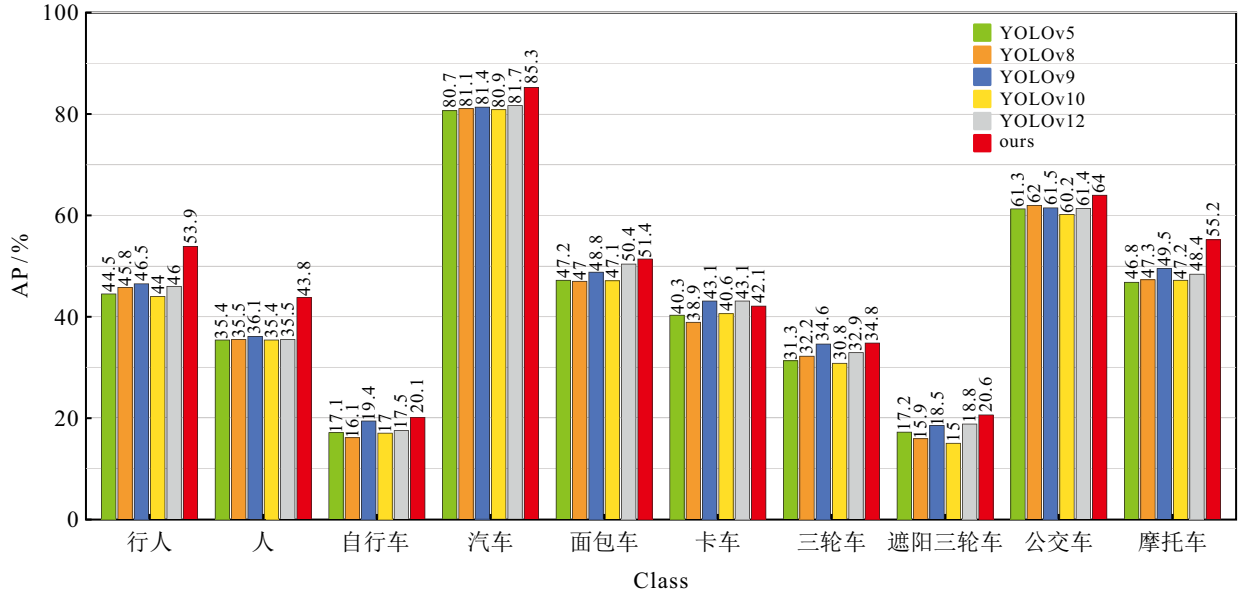

图4　不同 YOLO 系列模型在 VisDrone2019 数据集上各类别的 AP 性能 (%) 对比

面: 首先, PHFM 模块通过并行异构特征处理机制, 有效地平衡了全局上下文信息与局部细节特征的捕获, 使得模型能够同时关注目标的整体语义信息和精细结构特征; 其次, 改进的特征融合策略增强了模型对于不同尺度目标的适应性, 使其在处理各类复杂场景时均能够保持稳定的检测性能. 实验结果表明, 这种多尺度特征学习和融合的策略在提升检测准确率方面发挥了关键作用.

2.5 与基线模型的可视化对比

为进一步验证 PHFM 模块在特征表示增强方面的有效性, 通过热力图可视化的方式直观展示了模块对特征激活模式的改进效果. 图 5 为典型港口航拍场景在 PHFM 模块处理前后的特征响应对比.

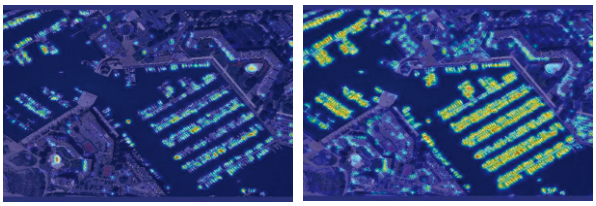

图5 PHFM 模块特征激活热力图对比分析

由图 5 对比结果可以清晰地观察到, 原始输入图像主要呈现基础的空间信息, 而经 PHFM 模块处理后的输出特征图展现出显著的增强效果. 输出热力图中的高激活区域 (绿色和黄色区域) 精确地对应了场景中的关键目标位置, 特别是密集停靠的船只区域. 这种激活模式的改变充分体现了 PHFM 模块通过并行异构特征调制机制实现的特征增强能力.

为验证 DI-YOLO 中 CARAFE 模块相对于传统上采样方法的优势, 使用 VisDrone2019 数据集的部分图片在相同的实验条件下对 YOLOv10 原始模型 (使用 nn.Upsample) 与 DI-YOLO 改进模型 (使用 CARAFE) 进行了定量对比分析. 本文选用同一位置的两个上采样模块的特征变化图进行比较, 图 6 的对比清晰地展示了两种上采样方法的差异. 传统最近邻上采样的 Enhancement Map 接近零值 (浅色), 表明其仅进行了简单的像素复制, 无法对特征进行优化; 而 CARAFE 则显示出显著的特征调制区域 (红色), 这些区域主要集中于语义关键位置, 体现了模块的内容感知能力. 输出特征的增强 (绿色高激活区域) 进一步验证了 CARAFE 在特征质量提升方面的优势.

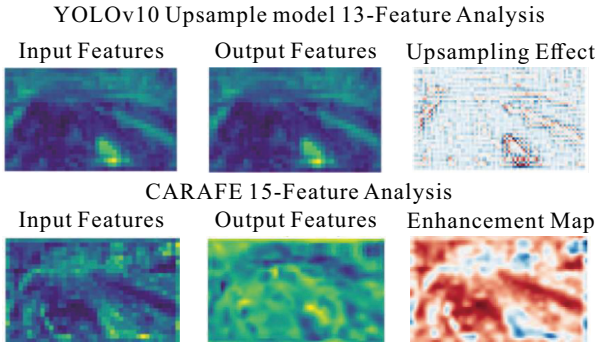

图6 CARAFE 与传统上采样在特征变化图的可视化结果

为了直观地展示 DI-YOLO 的检测性能, 本文从 VisDrone2019 测试集中选取了一组具有代表性的图像, 并以 YOLOv10m 作为基线模型进行对比分析. 图 7 为在密集低空场景和稀疏低空场景中的检

图7 DI-YOLO 与基线模型在 VisDrone2019 数据集低空场景上的可视化结果比较

测结果. 图 7 按照 3 列组织: 左列显示使用 YOLOv12 的检测结果图像, 中间列展示基线模型的检测结果, 右列呈现所提出 DI-YOLO 模型的输出. 这些结果生动地展示了所提出模型在处理各种类型和尺度目标方面的能力. 从零星的行人到密集停放的小型车辆, DI-YOLO 始终表现出精确的定位和分类能力, 在不同场景复杂度下均显著优于基线模型.

3 结 论

本文提出了 DI-YOLO, 一种面向航拍图像的增强型目标检测框架. 本文结合了内容感知特征增强模块和形状感知交并比损失函数的优势, 并设计了并行异构特征调制模块. 该模块通过独特的三流并行架构实现了全局上下文信息与局部细节特征的有效融合, 显著增强了模型的特征表示能力. 通过这些模块的协同作用, DI-YOLO 在保留小目标结构信息和多尺度特征融合方面取得了突破性进展. 在 VisDrone2019 和 DOTAv1.5 两个基准数据集上的实验验证了所提出方法的有效性. DI-YOLO 相比于基准模型 YOLOv10 取得了显著提升, 尤其是在小目标检测方面表现突出. 虽然引入了少量计算开销, 但是所提出模型仍然保持了良好的实时性能, 满足了实际航拍应用的需求.

未来工作将进一步优化模型结构, 降低计算复杂度, 并探索将所提出方法扩展到目标跟踪和实例分割等相关任务中的可能性. 特别是计划研究旋转目标框的检测方法, 进一步提升检测精度和实用性.

参考文献 (References)

[1] 张志豪, 杜丽霞, 郝紫微, 等. 多核上下文特征引导下的无人机航拍图像可信检测算法[J]. 北京航空航天大学学报, DOI: 10.13700/j.bh.1001-5965.2024.0548.
(Zhang Z H, Du L X, Hao Z W, et al. Multi-core contextual feature-guided algorithm for trusted detection of UAV aerial images[J]. Journal of Beijing University of Aeronautics and Astronautics, DOI: 10.13700/j.bh.1001-5965.2024.0548.)

[2] 陈志旺, 肖迪创, 吕昌昊, 等. 基于多尺度融合和高分辨特征增强的无人机航拍目标检测[J]. 控制与决策, 2025, 40(7): 2290-2299.
(Chen Z W, Xiao D C, Lv C H, et al. UAV aerial target detection based on multi-scale fusion and high-resolution feature enhancement[J]. Control and Decision, 2025, 40(7): 2290-2299.)

[3] 袁婷婷, 赖惠成, 汤静雯, 等. LMFI-YOLO: 复杂场景下的轻量化行人检测算法[J]. 计算机工程与应用, 2025.
(Yuan T T, Lai H C, Tang J W, et al. LMFI-YOLO: Light weight pedestrian detection algorithm in complex scenes[J]. Computer Engineering and Applications, 2025.)

[4] 卢迪, 赵庆. 空间分组内卷积轻量级目标检测算法[J]. 控制与决策, DOI: 10.13195/j.kzyjc.2025.0035.
(Lu D, Zhao Q. Lightweight object detection algorithm based on SGWInvo[J]. Control and Decision, DOI: 10.13195/j.kzyjc.2025.0035.)

[5] 刘振江, 张会娟, 姬淼鑫, 等. 轻量级多尺度特征融合增强的空间非合作小目标检测算法[J]. 北京航空航天大学学报, DOI: 10.13700/j.bh.1001-5965.2024.0509.
(Liu Z J, Zhang H J, Ji M X, et al. Lightweight multi-scale feature fusion enhanced spatial non-cooperative small target detection algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, DOI: 10.13700/j.bh.1001-5965.2024.0509.)

[6] 梁礼明, 冯耀, 龙鹏威, 等. 融合岛式双向特征金字塔的遥感图像目标检测[J]. 计算机工程与应用, 2024.
(Liang L M, Feng Y, Long P W, et al. Target detection in remote sensing image based on island bidirectional feature pyramid[J]. Computer Engineering and Applications, 2024.)

[7] 王宁, 智敏. 深度学习下的单阶段通用目标检测算法研究综述[J]. 计算机科学与探索, 2025, 19(5): 1115-1140.
(Wang N, Zhi M. Review of one-stage universal object detection algorithms in deep learning[J]. Journal of Frontiers of Computer Science and Technology, 2025, 19(5): 1115-1140.)

[8] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 779-788.

[9] Jocher G, Chaurasia A, Qiu J. Ultralytics YOLOv8[EB/OL]. (2023-01-10)[2024-05-20]. https://github.com/ultralytics/ultralytics.

[10] Wang A, Chen H, Liu L H, et al. YOLOv10: Real-time end-to-end object detection[J/OL]. 2024, arXiv: 2405.14458.

[11] Ultralytics. YOLO11[EB/OL]. (2024-09-30)[2024-09-30]. https://github.com/ultralytics/ultralytics.

[12] 张攀峰, 陈文强, 神显豪, 等. DD-YOLO, 一种面向无人机的小目标检测算法[J]. 电光与控制, 2025, 32(5): 20-26.
(Zhang P F, Chen W Q, Shen X H, et al. DD-YOLO, a small target detection algorithm for UAVs[J]. Electronics Optics & Control, 2025, 32(5): 20-26.)

[13] 武腾辉, 邓炳光. LS-YOLO: 基于改进 YOLOv8n 的航拍小目标检测算法[J]. 电讯技术, DOI: 10.20079/j.issn.1001-893x.241024003.
(Wu T H, Deng B G. LS-YOLO: A small target detection algorithm based on improved YOLOv8n[J]. Telecommunication Engineering, DOI: 10.20079/j.issn.1001-893x.241024003.)

[14] 范博淦, 王淑青, 陈开元. 基于改进 YOLOv8 的航拍无人机小目标检测模型[J]. 计算机应用, DOI: 10.11772/j.issn.1001-9081.2024070946.
(Fan B G, Wang S Q, Chen K Y. Small target detection

model for aerial photography UAV based on improved YOLOv8[J]. Journal of Computer Applications, DOI: 10.11772/j.issn.1001-9081.2024070946.)

[15] 杨永刚, 姜文韬, 高志云. 低空无人机实时目标检测算法[J]. 航空学报, DOI: 10.7527/S1000-6893.2025.31619.
(Yang Y G, Jiang W T, Gao Z Y. Algorithm for real-time target detection in low-altitude UAVs[J]. Acta Aeronautica et Astronautica Sinica, DOI: 10.7527/S1000-6893.2025.31619.)

[16] 陈崇杨, 彭力, 杨杰龙. 基于特征增强与上下文融合的无人机小目标检测算法[J]. 计算机科学, 2024, 51(12): 312-325.
(Chen C Y, Peng L, Yang J L. UAV small target detection algorithm based on feature enhancement and context fusion[J]. Computer Science, 2024, 51(12): 312-325.)

[17] 高卫峰, 易宇轩, 黄玲玲, 等. 一种高效的无人机航拍小目标检测算法[J]. 控制与决策, 2025, 40(8): 2525-2533.
(Gao W F, Yi Y X, Huang L L, et al. An efficient algorithm for small object detection in unmanned aerial vehicle images[J]. Control and Decision, 2025, 40(8): 2525-2533.)

[18] 高志霖, 王劲滔, 孟琪翔, 等. 基于双流特征增强的轻量级图像超分辨率重建[J]. 激光与光电子学进展, 2025, 62(16): 2.
(Gao Z L, Wang J T, Meng Q X, et al. Lightweight image super-resolution reconstruction based on dual-stream feature enhancement[J]. Laser & Optoelectronics Progress, 2025, 62(16): 2.)

[19] 张轩宇, 周思航, 黄健, 等. 基于高阶空间特征提取的无人机航拍小目标检测算法[J]. 计算机工程与应用, 2025, 61(12): 210-221.
(Zhang X Y, Zhou S H, Huang J, et al. High-order spatial feature extraction based small target detection for UAV aerial photographs[J]. Computer Engineering and Applications, 2025, 61(12): 210-221.)

作者简介

丁浩晗 (1992−), 男, 讲师, 博士, 主要研究方向为联邦学习、工业智能化, E-mail: dinghaohan@jiangnan.edu.cn;

贺万程 (2001−), 男, 硕士生, 主要研究方向为伪造图像识别、目标检测, E-mail: 6233115010@stu.jiangnan.edu.cn;

万俊 (2003−), 女, 本科, 主要研究方向为图像处理、多元时序预测, E-mail: 1091210408@stu.jiangnan.edu.cn;

沈易航 (2003−), 男, 本科, 主要研究方向为强化学习、图神经网络, E-mail: 1193220320@stu.jiangnan.edu.cn;

崔晓晖 (1971−), 男, 教授, 博士, 博士生导师, 主要研究方向为大数据、区块链, E-mail: xcui@whu.edu.cn.