计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
2026,35(1):152−163 [doi: 10.15888/j.cnki.csa.010028] [CSTR: 32024.14.csa.010028]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 YOLO11 的远距复杂场景小目标检测①

熊诗雨[1,2,3], 狄永正[1,2,3], 纪　雯[1], 史红周[1,2]

[1](中国科学院 计算技术研究所, 北京 100190)
[2](移动计算与新型终端北京市重点实验室, 北京 100190)
[3](中国科学院大学 计算机科学与技术学院, 北京 100049)
通信作者: 史红周, E-mail: hzshi@ict.ac.cn

摘　要: 远距复杂场景中的小目标检测任务因目标尺寸小、形态不规则、纹理信息弱且易被背景干扰, 长期面临检测精度低与鲁棒性差的挑战. 针对上述问题, 本文提出一种改进的检测算法 ReF-YOLO (remote-enhanced fusion YOLO), 在 YOLO11 框架基础上从特征提取、特征融合与检测头设计这 3 方面进行系统优化. 具体而言, 引入融合通道注意与空间建模的 C3k2DCASC 模块, 增强主干网络对非规则目标的表达能力; 设计结合主干同尺度特征的 L-Fuse 结构与高效下采样模块 SCDown, 提升语义与细节对齐效果; 并增设高分辨率 P2 检测分支, 有效提升极小目标的感知与定位能力. 在 VisDrone2019 典型小目标数据集上的实验表明, 所提方法的 mAP@0.5 相较于 YOLO11n 提升 4.9%, 在小目标检测任务中表现出更优的准确性与稳定性, 验证了其在远距复杂场景下的实用性与泛化能力.

关键词: 小目标检测; YOLO11; 特征提取; 多尺度融合; 远距复杂场景

引用格式: 熊诗雨,狄永正,纪雯,史红周.基于 YOLO11 的远距复杂场景小目标检测.计算机系统应用,2026,35(1):152–163. http://www.c-s-a.org.cn/1003-3254/10028.html

Small Object Detection in Long-range Complex Scenes Based on YOLO11

XIONG Shi-Yu[1,2,3], DI Yong-Zheng[1,2,3], JI Wen[1], SHI Hong-Zhou[1,2]

[1](Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)
[2](Beijing Key Laboratory of Mobile Computing and Pervasive Device, Beijing 100190, China)
[3](School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Small object detection in remote and complex scenes faces persistent challenges, including low detection accuracy and poor robustness, due to the objects' small size, irregular shape, weak texture, and high susceptibility to background interference. To address these issues, this study proposes an enhanced detection algorithm named remote-enhanced fusion YOLO (ReF-YOLO), which systematically optimizes the YOLO11 framework from three aspects: feature extraction, feature fusion, and detection head design. Specifically, a module called C3k2DCASC is introduced, integrating channel attention and spatial modeling, to enhance the backbone network's representational capacity for irregular objects. The L-Fuse structure, combined with the same-scale features from the backbone and the efficient downsampling module SCDown, is introduced to improve semantic-detail alignment. Additionally, a high-resolution P2 detection branch is added to strengthen the perception and localization capacities of the algorithm for detecting extremely small objects. Experiments on a representative small object detection dataset VisDrone2019, demonstrate that the proposed method improves mAP@0.5 by 4.9% over YOLO11n, along with enhanced accuracy and stability across various small object detection tasks. These results validate the utility and generalization capability of ReF-YOLO in

① 基金项目: 国家重点研发计划 (2023YFB4502805)
收稿时间: 2025-05-14; 修改时间: 2025-06-09, 2025-06-30; 采用时间: 2025-07-14; csa 在线出版时间: 2025-11-17
CNKI 网络首发时间: 2025-11-18

remote and complex scenes.

Key words: small object detection; YOLO11; feature extraction; multi-scale fusion; long-range complex scene

1 引言

随着人工智能与计算机视觉技术的不断发展, 目标检测作为基础性视觉任务, 已广泛应用于无人机监控、远距安防、工业质检和交通巡查等实际场景[1,2]. 在这些应用中, 小目标检测因目标尺寸小、纹理特征弱、易被遮挡或背景干扰而成为研究中的难点问题, 尤其在远距复杂环境下更具挑战性[3]. 以高空航拍或远程监控图像为例, 目标通常仅占据图像极小区域, 不仅尺寸小于 32×32 像素, 甚至低于原图的 0.1%, 且常伴随密集分布、形态不规则、边界模糊、光照不均等问题, 这些因素均显著增加了检测难度[4]. VisDrone2019 和 TinyPerson 数据集正是该类场景的代表, 分别对应于密集城市航拍与远距人体监控环境, 目标均存在尺度小、分布密、遮挡重等特征, 传统检测方法在此类数据上的表现普遍不佳, 漏检和误检问题突出[5].

当前目标检测技术主要划分为两类: 两阶段检测方法与单阶段检测方法. 以 Faster R-CNN[6] 与 Cascade R-CNN[7] 为代表的两阶段方法通常通过候选区域生成与后续分类回归的方式获得较高精度, 但结构复杂、推理速度慢, 不适用于实时场景. 相比之下, 单阶段方法如 SSD[8] 和 YOLO[9] 系列以其结构简洁、速度快、部署效率高的优势广泛应用于工业与边缘设备中. YOLO 系列算法自 YOLOv1 以来不断演进, Ultralytics 官方版本 YOLO11[10] 在主干网络结构、特征融合与检测头设计方面进行了多项优化, 提升了整体检测性能. 然而, 面对远距复杂场景中的极小目标和非规则形态, YOLO11 仍存在特征表达弱化、尺度适应性不足与语义空间失配等问题. 小目标由于在图像中的占比极小, 在下采样过程中其特征极易被压缩甚至消失, 导致深层检测头难以准确定位. 与此同时, 传统卷积结构在建模弱纹理与复杂边界形态时存在适应性差的问题, 造成模型在遮挡或背景干扰条件下感知能力下降. 此外, 检测头结构多集中于中等尺度层级, 对极小目标缺乏足够的分辨率支持, 限制了其在实际检测任务中的边界回归能力.

针对上述问题, 本文在 YOLO11 框架基础上提出一种面向远距复杂场景的小目标检测改进算法 ReF-YOLO (remote-enhanced fusion YOLO), 该算法从特征提取、特征融合与检测头设计这 3 方面系统提升模型对小目标的建模能力. 在特征提取阶段, 引入空间-通道协同结构 C3k2DCASC, 融合通道注意力与动态蛇形卷积, 增强模型对弱纹理与不规则边界的建模能力; 在特征融合阶段, 设计双向信息流融合结构 L-Fuse, 引入主干同尺度特征作为横向补偿通路, 提升浅层语义与空间细节对齐能力; 在结构效率方面, 引入 YOLOv10[11] 提出的高效下采样模块 SCDown, 减少小目标在多层压缩中的信息损失. 在检测头设计上, 新增 P2 分支结构, 基于高分辨率浅层特征图增强对极小目标的响应能力, 有效弥补深层检测头在极小尺度下的性能短板.

2 相关工作

2.1 小目标检测研究进展

小目标检测作为目标检测中的重要子任务, 广泛应用于安防监控、无人机航拍、遥感图像分析、交通监测等实际场景. 小目标通常被定义为其在图像中占比极低 (如小于 32×32 像素) 或在特征图中分辨率较低的目标, 其检测难点主要体现在 3 个方面: 一是下采样过程中目标信息容易丢失, 导致高层语义难以准确表达其位置与类别; 二是小目标具有弱纹理、模糊边缘、形态不规则等特性, 使得特征提取器难以学习稳定的判别特征; 三是背景干扰强、前景信号弱, 极易引发漏检与误检. 为缓解上述问题, 研究者从不同角度提出了多种改进方法, 主要集中于多尺度特征融合、特征信息增强与分辨率感知优化这 3 个方向.

多尺度特征融合是当前最主流的解决策略之一. 通过融合来自浅层的高分辨率特征与深层的高语义信息, 可以实现空间细节与语义信息的互补, 从而更好地定位和识别小目标. FPN[12]是该方向的经典结构, 自顶向下构建融合路径, 提升了低层特征的语义表达能力. 后续如 PANet[13]进一步引入自底向上的强化路径, 增强信息双向流动. YOLOF[14]等方法通过定制的融合单元与轻量感受野扩展模块, 有效提升了感知深度与小目标表达能力, 同时控制了计算复杂度.

特征信息增强方法则聚焦于提升小目标在特征图

中的显著性. 例如, 空洞卷积[15]可在不增加参数的情况下扩展感受野, 增强小目标的上下文理解能力; 注意力机制 (如 SE[16]、CBAM[17]、SimAM[18]等) 引导网络关注目标关键区域, 抑制背景噪声干扰; 此外, 结合监督策略的空间注意机制也被用于突出关键目标区域, 提升网络的聚焦能力[19].

超分辨率感知与分辨率补偿也被证明对小目标检测有效[20,21]. 在遥感图像或视频中, 研究者常将图像重建与目标检测联合训练, 以提升图像细节和目标清晰度, 从而增强小目标的可识别性. 一些工作如 DCSCN[21]、SRD[22]等在低分辨率数据上显著提升了小目标检测精度. 这类方法也可与多尺度建模联合使用, 进一步优化检测结果.

总体来看, 小目标检测技术已在网络结构、特征建模、信息增强等方面取得显著进展. 但在实际部署中仍需面对如下挑战: 其一, 深度融合结构带来的计算成本问题; 其二, 对极端尺度不平衡或非规则目标的适应能力仍有限; 其三, 部分增强策略易引入冗余信息, 造成训练收敛困难. 因此, 在保证效率的前提下实现鲁棒性小目标检测仍是当前研究关注的重点方向.

2.2 YOLO 系列算法演进

YOLO 系列算法因其高效性与部署便捷性, 在实时小目标检测中被广泛应用. 从 YOLOv1 至 YOLOv12, 该系列经历了 4 个阶段的发展.

在尺度感知阶段, YOLOv2[23]引入 Anchor 机制, YOLOv3[24]采用 FPN 结构提升了多尺度预测能力. 特征增强阶段的 YOLOv4[25]与 YOLOv5[26]通过引入 PANet、SPP 模块与轻量卷积, 提高了对小目标的表达能力. 后续的 YOLOv6[27]与 YOLOv7[28]进一步在结构重参数化与跨层融合方面优化, 实现了精度与速度的平衡. 进入几何感知阶段后, YOLOv8[29]引入 Anchor-Free 机制, YOLOv9[30]至 YOLOv12[31]持续强化对目标形状边缘和空间结构的建模能力, 如 PGI 梯度引导、动态感受野、自注意力机制与 Transformer 集成等.

尽管 YOLO 系列不断提升模型的尺度感知与边界表达能力, 但在处理远距、非规则小目标时仍面临 3 方面挑战: 深层语义压缩目标细节、锚框与规则网格难以适应变形边界、多尺度融合对齐误差影响定位精度.

2.3 远距复杂场景下的小目标检测挑战

远距复杂场景常见于无人机航拍、远程监控与遥感应用, 图像中目标通常尺寸更小、边界模糊、遮挡严重、背景复杂, 检测难度显著增加. 小目标在下采样过程中特征易丢失, 传统卷积结构难以适配不规则目标几何形态, Anchor 机制也难以准确拟合其实际边界.

同时, 复杂背景中的高噪声区域、弱对比度和非结构化布局进一步降低了检测器对前景的判别能力. 虽然已有方法尝试增强浅层特征、增加高分辨率检测头等手段改善检测效果, 但常以增加计算代价为代价, 不利于实际部署.

综上所述, 当前小目标检测技术在特征融合、结构优化等方面取得了丰富成果, YOLO 系列算法持续演进, 展现出较强的实用性与扩展潜力. 然而, 在面对远距复杂环境中非规则小目标时, 现有方法在信息保持、几何建模与分支结构设计方面仍存在不足.

为此, 本文在 YOLO11 基础上提出 ReF-YOLO 算法, 结合空间-通道协同模块 (C3k2DCASC)、双向特征融合结构 (L-Fuse)、高效下采样模块 (SCDown) 与极小目标检测分支 (P2), 从特征提取、特征融合与检测精度这 3 方面系统优化检测器结构, 提升其在远距复杂场景中的小目标检测能力.

3 ReF-YOLO 算法设计与实现

本文在 YOLO11 目标检测框架的基础上, 提出一种面向远距复杂场景的小目标检测算法——ReF-YOLO. 该算法从结构设计角度出发, 围绕“增强感知能力、提升信息保真、优化尺度支持”这 3 个目标, 对 YOLO11 的主干网络、颈部融合路径与检测头结构进行多维度优化, 形成系统性的性能提升方案.

ReF-YOLO 的架构体系由 3 个核心组件组成: 包括主干网络、颈部特征融合网络和检测头部分, 如图 1 所示. 在主干部分, 本文设计了空间–通道协同特征提取模块 DCASC, 并嵌入原始 C3k2 模块中构建 C3k2DCASC 模块, 用以增强网络对非规则轮廓与弱纹理目标的早期感知能力. 在颈部网络部分, 设计了 L-Fuse 双向融合模块, 引入主干同尺度特征补偿路径, 有效缓解路径信息流中语义断层问题. 同时, SCDown 模块通过重构压缩路径, 实现对小目标纹理与边界的特征保持, 为下游检测任务提供更具判别性的融合特征. 在检测头部分, 除常规 P3–P5 检测层外, ReF-YOLO 新增了 P2 检测分支, 以更高的空间分辨率 (1/4 特征图) 专注于极小目标区域的检测任务, 提升了网络对低像

素小目标的响应能力. 最终检测头覆盖 P2–P5 这 4 个尺度, 分别对应 160×160、80×80、40×40 和 20×20 的输出分辨率, 可适应从极小至中大目标的多尺度检测需求.

整体而言, ReF-YOLO 通过在不同层级引入结构增强与分辨率补偿机制, 实现了对远距复杂场景中非规则小目标的感知强化与定位优化, 为后续章节中各模块的具体设计与性能分析奠定了结构基础.

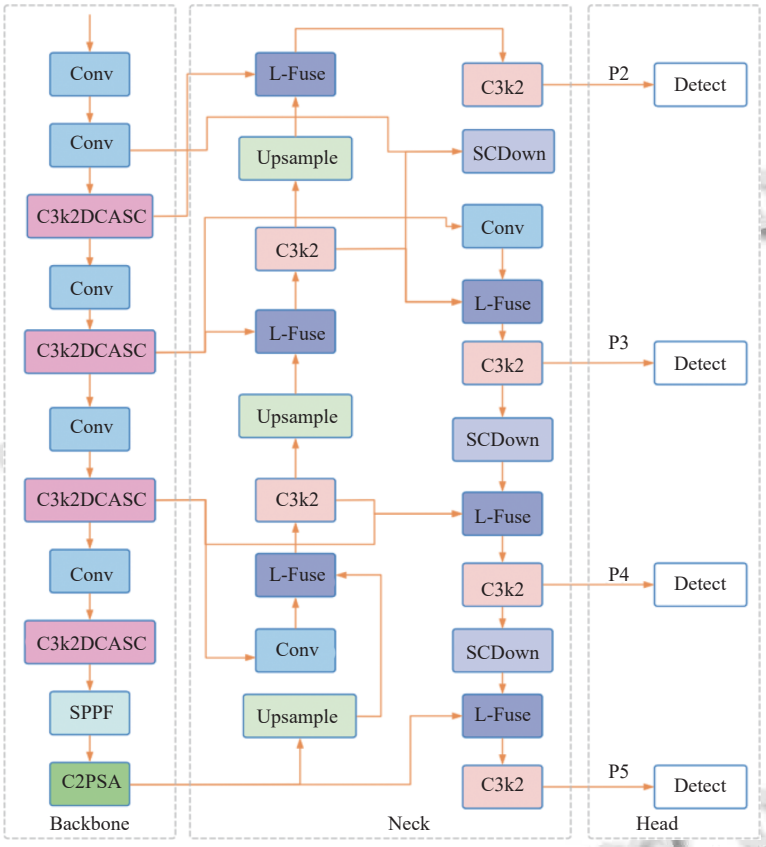


图 1 ReF-YOLO 结构图

3.1 主干网络优化: C3k2DCASC

在目标检测任务中, 主干部分主要承担原始图像的多尺度信息建模任务, 其提取到的特征在很大程度上决定了模型对目标形状与语义语境的表征质量. 然而, 在远距复杂场景中, 小目标往往呈现出尺寸极小、边界不规则、纹理信息弱等特性, 常规卷积结构在提取其有效特征时存在明显不足. 一方面, 小目标在前几层特征图中响应较弱, 在逐层下采样过程中易发生信息消失; 另一方面, 传统卷积核具有固定感受野和对规则结构的强依赖性, 难以适应形态多变的目标的复杂空间结构.

为解决上述问题, 在 YOLO11 主干网络中 C3k2 模块的基础上, 本文引入 DCASC 结构, 形成 C3k2DCASC 模块. 该模块融合了通道维度的注意力调控与空间维度的动态形态建模, 以实现更具自适应性的特征提取能力, 强化主干网络在早期阶段对远距小目标的感知与建模. DCASC 模块主要由两部分组成, 如图 2 所示: (a) 是动态蛇形卷积 (DSC) 分支, 用于建模目标边界的空间结构变化; (b) 是包含自适应通道注意力 (AdaCA) 的分支, 用于增强目标语义通道.

整体处理过程可分为 3 个连续的阶段, 分别对应通道增强、动态采样与特征融合. 在第 1 阶段, 模块对输入特征图 $F \in \mathbb{R}^{C\times H\times W}$ 进行通道维度建模, 分别采用全局平均池化与最大池化提取通道级统计特征, 并通过共享多层感知器 (*MLP*) 进行通道权重预测. 具体地, 通道注意力权重向量 S 的计算如下.

$$F_{\text{avg}} = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}F(:,i,j) \tag{1}$$

$$F_{\max} = \max_{i,j}F(:,i,j) \tag{2}$$

$$S=\sigma(MLP(F_{\text{avg}})+MLP(F_{\max})) \tag{3}$$

其中, σ 表示 Sigmoid 函数, MLP 包括两个线性变换层, 具有共享参数. 该权重向量用于对每个通道特征图进行加权, 得到通道增强后的输出.

$$F_{\text{attn}}(c,:,:)=F_{\max}=\max_{i,j}F(c,:,:)\cdot S(c) \tag{4}$$

进入第 2 阶段, 模块引入空间动态采样机制, 通过轻量卷积层预测每个卷积位置上的偏移量 $\Delta P\in\mathbb{R}^{2K\times H\times W}$, 并采用归一化与正切函数进行约束, 以保持偏移稳定性.

$$\Delta P=\tanh(GN(Conv(F_{\text{attn}}))) \tag{5}$$

采用轴向递推策略生成蛇形采样路径, 使卷积核的采样点可以灵活地贴合目标边界形态, 从而更好地适配不规则结构的空间分布.

$$p'=p+\Delta p \tag{6}$$

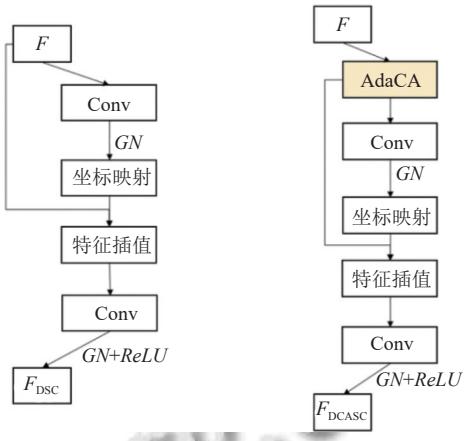
由于偏移点多为非整数位置, 本文采用双线性插值从邻近整数点中获取特征值, 插值表达为:

$$F_{\text{deformed}}(p')=\sum_{q\in N(p')}F_{\text{attn}}(q)\cdot\max(0,1-|p'_x-q_x|)\cdot\max(0,1-|p'_y-q_y|) \tag{7}$$

在第 3 阶段, 模块对所有变形采样点进行特征加权融合, 并通过归一化与非线性激活得到最终的输出特征图. 整体输出表达为:

$$F_{\text{DCASC}}(p_0)=ReLU(GN(\sum_{p_n\in R}W(p_n)\cdot F_{\text{deformed}}(p_0+p_n+\Delta p_n))) \tag{8}$$

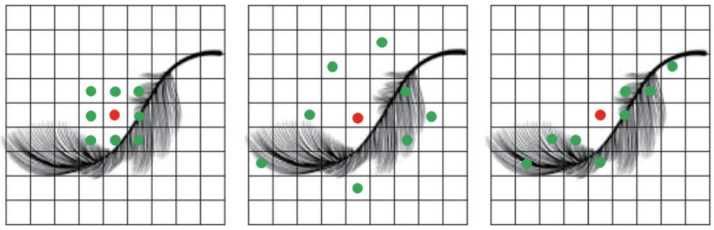
其中, R表示标准卷积核位置集合, GN为 group normalization, $W(p_n)$为可学习的卷积权重. 该模块通过通道与空间两个维度的协同增强, 显著提升了主干网络在早期阶段对远距小目标的感知能力.



(a) 动态蛇形卷积 (DSC) 分支 (b) 包含AdaCA的DCASC分支

图 2 DCASC 结构

为进一步说明动态蛇形卷积 (DSC) 在不规则小目标建模中的优势, 本文以图示方式对比了传统卷积、可变形卷积与本文引入的动态蛇形卷积的采样行为 (见图 3). 图 3 中以一片弯曲羽毛类结构为例, 其边界具有明显的非规则曲率, 红点表示中心采样位置, 绿点为对应卷积核的采样点. 如图 3(a) 是传统卷积采用固定网格, 传统卷积采用固定的 9 点正方形采样区域, 难以适应弯曲边界, 难以覆盖边界的关键区域; 图 3(b) 中的可变形卷积虽能在一定程度上自由采样, 但其偏移量学习受限于卷积核感受野, 采样点仍存在漂移、分散的问题. 相比之下, 图 3(c) 展示的动态蛇形卷积通过引入轴向递推策略, 引导采样点沿边界连续分布, 使得采样路径更紧凑地贴合边界曲线, 增强了空间结构建模能力, 更贴合真实结构. 这一机制可视为对卷积核采样几何结构的引导性建模, 使卷积从“规则网格”拓展为“边界自适应路径”, 从而实现对不规则目标 (如弯曲人影、非对称物体边缘) 的更精准感知与表达. 通过这种动态偏移 ΔP 的引导, DSC 提升了特征提取的几何适配性与感知灵活性, 是提升 C3k2DCASC 模块对非规则小目标表达能力的关键所在.



(a) 传统卷积 (b) 可变形卷积 (c) 动态蛇形卷积

图 3 不同卷积方式在非规则边界上的采样策略对比

为了将 DCASC 模块有效集成至 YOLO11 的主干网络中, 本文将 DCASC 模块嵌入至 YOLO11 主干中的 C3k2 模块中, 形成改进后的 C3k2DCASC 结构. 该结构在保持原有 CSP 分组卷积主干架构 (C3k2) 的基础上, 将标准 Bottleneck 块替换为更具几何感知能力的 Bottleneck_DCASC 模块. 在实际网络实现中, C3k2 结构包含两种不同配置形式, 分别对应 C3k=True 与 C3k=False. 同样的, C3k2DCASC 也包含这两种情况, 如图 4 所示.

当参数 C3k=True 时, 主路径使用一个 C3kDCASC 模块, 该模块由两个串联的 Bottleneck_DCASC 子模块组成, 用于增强深度空间建模能力; 当参数 C3k=False 时, 主路径则直接使用一个 Bottleneck_DCASC 模块, 以实现更高的效率与结构紧凑性.

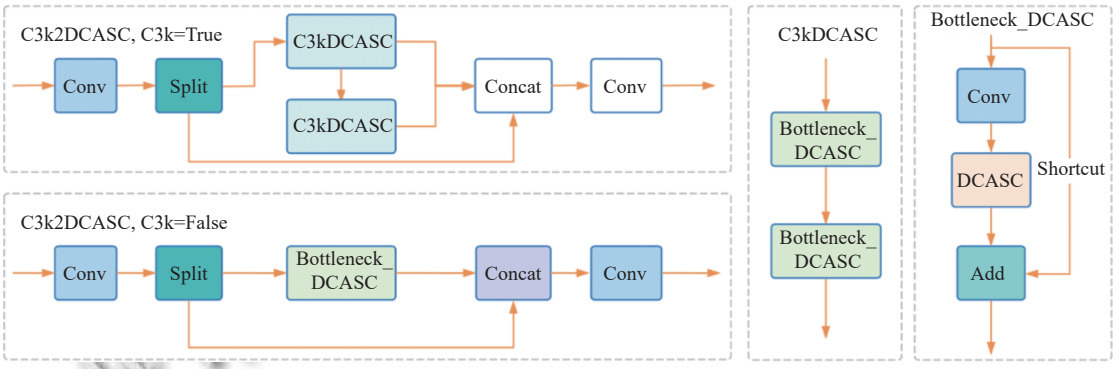


图 4 C3k2DCASC 模块结构图

3.2 颈部网络设计: L-Fuse 与 SCDown

颈部网络是连接主干与检测头的中间模块, 承担着多尺度特征融合与语义信息整合的重要任务. 其结构设计对小目标的可感知性和检测准确性具有关键影响. YOLO11 中采用的标准 PAN 结构具备基本的上下特征流融合能力, 但在远距复杂场景中, 面对纹理弱、边界模糊和尺度不一的小目标时, 仍存在特征细节衰减、结构对齐不充分等问题, 难以有效支撑后续检测头的精确定位与分类.

为了增强对复杂小目标的语义建模能力, 本文在 YOLO11 原有上下双向路径设计的基础上, 借鉴 BiFPN 的多分支融合思想, 提出增强型三路融合模块 L-Fuse. 该结构不仅保留了上下层信息的传递, 还引入主干网络中间层的横向特征作为辅助路径, 进一步缓解浅层细节弱化与语义失衡的问题. 同时, 为提升尺度压缩过程中的信息保持能力, 本文还结合 YOLOv10 中提出的高效下采样模块 SCDown, 构建出面向远距复杂场景小目标检测任务的鲁棒颈部网络结构.

L-Fuse 模块的核心机制是通过可学习权重对多路特征进行自适应融合. 如图 5 所示, 设输入特征集合为$\{F_1, F_2, \cdots, F_n\}$, 对应可学习权重为 $\{w_1, w_2, \cdots, w_n\}$. 首先对权重进行 *Hardswish* 激活与归一化处理, 计算方式为:

$$w_i' = \frac{Hardswish(w_i)}{\sum_{j=1}^{n} Hardswish(w_i) + \varepsilon}, \quad i = 1, 2, \cdots, n \tag{9}$$

然后对各分支特征进行加权求和, 得到融合输出:

$$F_i' = \sum_{i=1}^{n} w_i' \cdot F_i \tag{10}$$

该加权融合方式能够根据不同场景特征响应情况动态分配信息通道比例, 有效提升融合节点对关键特征的选择性与稳定性, 特别适用于尺寸小、形态多变、边界模糊目标的感知.

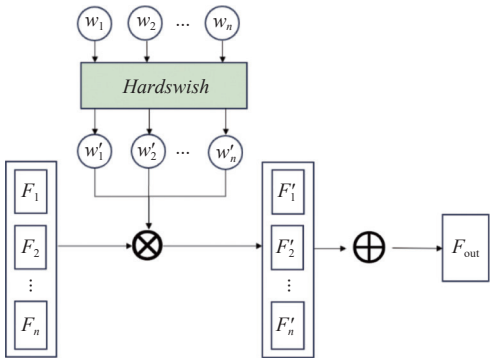


图 5 L-Fuse 结构图

为进一步增强主干至颈部路径中的特征完整性, 本文在下采样阶段引入结构紧凑、表达能力强的 SCDown 模块. 在实现上, SCDown 采用了由 1×1 卷积与

步幅为 2 的深度可分离卷积串联构建的双层结构, 用于增强特征压缩过程中的信息保留能力, 其输出表示为:

$$F_{sc} = Conv_{dw,k=3,s=2}(Conv_{1\times1}(F)) \quad (11)$$

该结构在压缩特征图分辨率的同时最大程度保留通道表达能力, 提升模型在远距视角下对小目标边界的保留效果.

为进一步验证 L-Fuse 结构在缓解语义断层问题方面的有效性, 本文在 VisDrone2019 数据集中选取了典型的远距航拍场景图像, 进行融合特征响应的热力图对比分析, 如图 6 所示. 图 6(a) 为原始图像, 图 6(b) 为 YOLO11n 原始结构下的特征响应, 图 6(c) 为引入 L-Fuse 结构后的响应结果. 图中红黄热区代表特征激活强度, 越亮表示模型越关注该区域. 从图 6(b) 中可以观察到, 在场景中多个车辆分布稀疏、光照逆向、目标尺寸小的区域, 原始 YOLO11n 模型的响应强度极低, 热区极少甚至断裂, 反映出信息在浅层向深层传递过程中存在“语义断层”现象: 浅层细节信息因缺乏足够语义引导而难以保留, 在多尺度融合阶段被进一步削弱, 最终导致检测头难以激活该区域. 相比之下, 图 6(c) 中在相同区域下的特征热力图表现出更密集、更连续的响应模式, 多个目标区域均出现明显激活, 说明融合后的特征表达更均衡、语义与细节信息整合更充分. 究其原因, 一方面是 L-Fuse 在上下路径之间引入主干中间层的横向补偿通路, 从而引导浅层空间细节在融合过程中与深层语义进行更充分对齐, 避免了信息“直接拼接”带来的特征突变; 另一方面, L-Fuse 通过可学习的融合权重机制对不同通路的贡献进行动态调整, 缓解了部分路径中存在的语义过强或细节过弱的偏置, 使得融合节点的整体感知能力更强.

(a) 原始图 (b) YOLO11n特征响应图 (c) 引入L-Fuse后的特征响应图

图 6 L-Fuse 结构引入前后的特征响应热力图对比

从感知效果角度看, L-Fuse 的引入提升了模型对边界模糊、光照复杂和尺寸极小目标的响应能力, 从而在远距复杂场景中建立了更鲁棒的特征表征基础, 也为后续检测分支提供了更稳定的输入支持.

在颈部网络中, L-Fuse 模块通过轻量归一化加权机制构建出鲁棒灵活的特征融合路径, SCDown 模块则以轻量卷积单元提升下采样表达密度, 二者协同强化了 ReF-YOLO 在远距复杂场景中对微弱目标的识别感知能力, 为检测头提供了稳定可靠的语义支撑.

3.3 检测头优化: 引入高分辨率 P2 分支

检测头模块是目标检测框架中完成分类与定位预测的核心部分, 其感受野与分辨率设计直接影响模型对不同尺度目标的检测能力. YOLO11 原始结构采用 3 层输出 (P3–P5) 用于多尺度目标检测, 然而对于远距复杂场景中的小目标, 尤其是边界模糊、纹理弱的极小目标, 其在下采样后的高层特征图中往往仅占据 1–2 个像素, 感知信息极为有限, 难以实现稳定响应与精准定位.

针对上述问题, 本文在保持 YOLO11 原有检测结构的基础上, 引入更高分辨率的 P2 检测头分支, 用于增强模型对极小目标的检测能力. 该分支通过上采样路径从 P3 上采样特征与主干中较浅层的 P2 特征进行 L-Fuse 融合得到, 具有更大的空间分辨率 (160×160), 更有效地维持了图像中的边缘轮廓、纹理结构以及微观细节的完整性.

在检测头的上采样分支中, 模型通过融合低层的细节特征与高层上采样得到的语义特征, 借助 L-Fuse 结构构建空间分辨率最高的特征表示. 该融合结果随后通过卷积层进行特征提炼, 最终输出小目标检测结果分支. 该路径对目标边界变化的响应更加敏锐, 有效增强了模型对小尺寸目标 (如 16 像素以下目标) 的定位与识别能力.

值得注意的是, P2 检测头的引入在一定程度上增加了模型的计算量和显存需求, 但在小目标场景中的精度收益明显优于资源开销. 通过实验证明, P2 分支有效提升了模型在高密度、小尺寸目标场景中的召回率

与 mAP 表现, 为远距复杂条件下的小目标检测任务提供了关键支持.

因此, 新增的 P2 分支构建了更密集的空间感知通路, 使得 ReF-YOLO 在极端距离和复杂环境中具备更强的小目标检测鲁棒性.

4 实验结果与分析

4.1 数据集

为了评估 ReF-YOLO 方法在远程复杂场景中应对小尺度目标检测的有效性, 本文选用了具有代表性的 VisDrone2019 与 TinyPerson 两个小目标数据集开展实验对比分析.

VisDrone2019[32] 是广泛应用于无人机视觉检测任务的代表性数据集之一, 由天津大学的 AISKYEYE 团队构建并公开发布, 涵盖了多种真实场景下的航拍图像. 该数据集共收录了 6 471 张训练图、548 张验证图与 1 610 张测试图, 覆盖行人、自行车、汽车、卡车、公交车等 10 类常见目标. 其图像拍摄角度主要为城市上空, 场景中目标数量密集且尺度变化大, 其中小于 32 像素的目标占比较高, 为远距小目标检测算法提供了具有挑战性的测试平台.

TinyPerson[33]数据集是一个专门针对极小人体目标检测任务设计的数据集. 该数据集收集自多个真实监控场景, 涵盖多种拍摄角度和环境条件, 图像中目标高度大多低于 20 像素, 具备极高密集度与显著遮挡性. 数据集中共包含 8 122 张图像, 并已提供标准化的训练与测试划分. TinyPerson 被广泛用于评估模型在超小尺寸目标、远距模糊目标等极端条件下的检测能力.

上述两个数据集共同构成对所提算法在城市空中视角与远距监控场景中的实用性验证基础, 能够全面反映 ReF-YOLO 在不同小目标检测环境下的鲁棒性与适应性.

4.2 实验设置与评价指标

本文的所有实验工作均基于 YOLO11n 框架统一进行, 并采用一致的训练参数配置, 以保证消融分析与对比评估具备可比性. 具体实验运行环境为 Ubuntu 22.04.5 系统, 使用 NVIDIA RTX 2080 Ti 显卡, 并依托 PyTorch 2.0.1 与 CUDA 11.7 实现模型训练, 详见表 1.

在模型训练阶段, 初始学习率设置为 0.01, 批处理大小为 8, 总共迭代 400 个 epoch. 实验中图像统一输入为 640×640 的尺寸, 并采用 YOLO11 默认的数据增强策略进行预处理.

表 1 实验环境配置

参数	实验环境
CPU	Intel Xeon Silver 4110
GPU	RTX 2080 Ti (62 GB)
操作系统	Ubuntu 20.04
Python	3.8
PyTorch	2.0.1
CUDA	11.7

模型性能评估指标包括 mAP@0.5、召回率 (Recall)、精确度 (Precision)、参数量 (Params) 与 GFLOPs 这 5 项. mAP@0.5 是在 IoU 阈值为 0.5 时的平均精度, 反映模型在目标识别任务中的总体准确性; mAP@0.5:0.95 是更严格的检测指标, 能够考察模型对目标边界的拟合能力. Recall 表示模型对正样本的检出能力, Precision 衡量误检率. 在资源开销方面, 参数量用于表示模型的存储大小, GFLOPs 衡量模型每次前向推理的计算复杂度, 便于在轻量化部署时评估模型的实际适用性.

4.3 消融实验与分析

为评估 ReF-YOLO 各结构模块对小目标检测性能的影响, 本文基于 VisDrone2019 数据集设计了系统的消融实验, 对 4 个关键组件 (C3k2DCASC 主干结构、L-Fuse 融合模块、SCDown 下采样机制和 P2 检测头) 在不同组合下的性能表现进行量化分析, 结果如表 2 所示.

以 YOLO11n 为基础模型, 其 mAP@0.5 为 27.4%. 实验 A 单独引入 C3k2DCASC 模块, mAP@0.5 提升至 28.4%, 仅增加 0.04M 参数和 0.4 GFLOPs, 验证该模块在保持高效率的同时具备良好的边界建模能力. 实验 B 与实验 C 分别引入 L-Fuse 与 BiFPN 作为多尺度融合模块, mAP@0.5 分别达到 28.7% 和 28.8%, 二者在精度上相近, 但 BiFPN 不引入横向补偿路径, 缺乏细粒度信息的自适应调整能力. 实验 D 单独替换为 SCDown 下采样模块后, mAP@0.5 下降至 27.1% (低于基线 0.3 个百分点). 主要原因在于缺乏横向补偿时, SCDown 一次 stride=2 的压缩会过度丢失浅层纹理与边缘信息, 导致极小目标的早期感知不足. 因此 SCDown 更适合与 L-Fuse 或 P2 分支等细粒度语义补偿路径配合使用, 而非孤立部署.

进一步地, 实验 H 和实验 I 分别对比了 L-Fuse 与 BiFPN 在完整模块组合下的效果, 结果显示 L-Fuse 组合 (H) 最终 mAP@0.5 提升至 32.3%, 略高于 BiFPN 组

合 (I) 的 32.1%, 验证了 L-Fuse 在语义对齐与融合鲁棒性上的优势. 虽然二者的计算量均为 13.9 GFLOPs, 但 L-Fuse 通过更简洁的三路结构和可学习加权机制实现更高精度, 具备更优的性能效率比, 说明其设计在复杂场景下对小目标信息整合效果更强, 体现了结构改进的量化价值.

表 2 VisDrone2019 数据集采用不同改进策略后的检测结果

方法	C3k2DCASC	L-Fuse	BiFPN	SCDown	P2	mAP@0.5 (%)	参数量 (M)	GFLOPs
YOLO11n	—	—	—	—	—	27.4	2.58	6.3
A	√	—	—	—	—	28.4	2.62	6.7
B	—	√	—	—	—	28.7	3.11	7.8
C	—	—	√	—	—	28.8	3.11	7.8
D	—	—	—	√	—	27.1	2.42	6.2
E	—	—	—	—	√	28.2	2.66	10.2
F	—	√	—	√	—	28.1	2.77	7.4
G	√	√	—	√	—	29.0	2.71	7.8
H	√	√	—	√	√	**32.3**	2.83	13.9
I	√	—	√	√	√	32.1	2.83	13.9

值得注意的是, P2 分支的引入 (实验 E、H、I) 虽然导致 GFLOPs 增加约 1 倍 (由 6.3 GFLOPs 增至 13.9 GFLOPs), 但其显著提升了对极小目标的感知能力, 使 mAP@0.5 相较基线提升 4.9%. 考虑到实际部署中边缘计算设备普遍具备中等计算资源, ReF-YOLO 的整体参数量控制在 2.8M 以内, 仍符合轻量化检测框架标准. 因此, 通过在结构层面嵌入自适应特征建模机制与多尺度感知补偿路径, ReF-YOLO 在保持部署友好的同时实现了对小目标检测性能的实质性提升.

4.4 对比实验与分析

4.4.1 与 YOLO11n 进行精确度对比

为进一步验证 ReF-YOLO 在小目标检测任务中的实际检测性能, 本文在 VisDrone2019 测试集上统计了模型在各类别上的检测精确度与 mAP@0.5 指标, 结果如表 3 所示. 可以看出, ReF-YOLO 在所有 10 个类别中均实现了精度提升, 尤其在小尺寸目标如行人、自行车、三轮车、摩托车等类别上优势显著.

表 3 VisDrone2019 数据集各类别对比结果 (%)

类别	YOLO11n		Ours	
	精确度	mAP@0.5	精确度	mAP@0.5
行人	41.7	23.0	45.6	31.1
人	43.9	12.7	47.1	19.8
自行车	22.8	6.89	30.4	10.6
汽车	60.6	67.1	64.0	73.0
货车	31.8	31.2	37.9	36.6
卡车	37.7	30.0	38.2	33.0
三轮车	23.2	12.6	26.4	17.1
遮阳三轮车	34.2	14.7	39.9	17.4
公共汽车	61.5	51.6	64.6	53.0
摩托车	37.4	24.7	40.1	31.2

其中, 在“行人”类别上, ReF-YOLO 的 mAP@0.5 从 23.0% 提升至 31.1%, 提升了 8.1%, 精确度也从 41.7% 提升至 45.6%; “人”类别的目标 mAP@0.5 提升了 7.1%; 而“自行车”类目标 mAP@0.5 从 6.9% 提升至 14.8%, 提升幅度最大, 达 7.9%. 此外, “遮阳三轮车”与“摩托车”类别中, ReF-YOLO 提升均超过 6%, 说明改进结构有效增强了模型对极小尺寸目标的感知能力.

中等与大尺寸目标如“汽车”和“卡车”类别上, ReF-YOLO 仍保持了与原模型相近甚至更优的精度, 分别达 73.0% 和 61.0%, 表明 ReF-YOLO 在提升小目标性能的同时未显著削弱对中大目标的识别效果, 体现出良好的模型稳定性与通用性.

4.4.2 VisDrone2019 对比实验

为了进一步验证 ReF-YOLO 在无人机拍摄场景下对小目标检测的有效性, 本文选取 VisDrone2019 数据集, 将所提方法与多种主流模型进行性能对比. 对比结果如表 4 所示, 其中加粗数据代表该类别下的最优检测指标.

在 VisDrone2019 测试集上的对比实验结果显示, ReF-YOLO 在 mAP@0.5 指标上获得了 32.3% 的检测精度, 显著优于 YOLO12n (25.8%)、YOLO11n (27.4%)、YOLOv10n (26.8%)、YOLOv8n (26.7%) 和 Faster R-CNN (21.7%) 等方法, 展现出较强的整体检测性能. 具体类别上, ReF-YOLO 在人 (19.8%)、行人 (31.1%)、自行车 (10.6%) 和摩托车 (31.2%) 等典型小目标上均取得最高 AP50, 与 YOLO11n 相比分别提升 7.1%、8.1%、3.7% 和 6.5%, 有效缓解了目标尺寸小、遮挡严

重等挑战. 同时, 在复杂背景下的“遮挡三轮车”和“三轮车”类别上, ReF-YOLO 亦保持领先, 表明其对非规则结构目标具备更强的鲁棒性. 即便在中等目标类别如“汽车”和“公共汽车”中, ReF-YOLO 仍取得最高检测精度, 进一步验证了其在远距复杂场景下的普适性与有效性.

表 4 多种算法在 VisDrone2019 数据测试集所得的 AP50 及 mAP@0.5 对比结果 (%)

方法	AP50										mAP@0.5
	行人	人	自行车	汽车	货车	卡车	三轮车	遮阳三轮车	公共汽车	摩托车	
YOLOv5n	21.8	11.7	6.48	66.7	30.8	31.1	11.6	14.0	50.6	22.2	26.7
YOLOv8n	22.8	11.9	6.61	66.3	31.0	29.9	12.0	13.6	49.9	22.9	26.7
YOLOv10n	23.3	13.2	6.77	66.1	29.9	28.8	12.4	13.7	50.0	23.5	26.8
YOLO11n	23.0	12.7	6.89	67.1	31.2	30.0	12.6	14.7	51.6	24.7	27.4
YOLOv12n	21.5	11.2	5.57	65.4	28.9	30.5	11.8	12.6	47.8	22.7	25.8
Faster R-CNN[34]	21.4	15.6	6.7	51.7	29.5	19.0	13.1	7.7	31.4	20.7	21.7
RetinaNet[35]	13.0	7.9	1.4	45.5	19.9	11.5	6.3	4.2	17.8	11.8	13.9
ReF-YOLO	**31.1**	**19.8**	**10.6**	**73.0**	**36.6**	**33.0**	**17.1**	**17.4**	**53.0**	**31.2**	**32.3**

4.4.3 通用性对比实验

为验证 ReF-YOLO 模型在不同检测任务下的适应能力与通用性, 本文进一步在与训练数据分布存在差异的场景下评估其泛化表现, 并与 YOLO11n 基线进行直接对比, 实验结果如表 5 所示.

表 5 通用性对比实验

模型	精确度 (%)	召回率 (%)	mAP@0.5 (%)	参数量 (M)	GFLOPs
YOLO11n	38.5	24.3	21.5	2.58	6.3
Ours	39.2	29.3	25.3	2.83	13.9

从表 5 中结果可见, ReF-YOLO 在未重新调参的条件下仍展现出较强的适应性, 整体 mAP@0.5 达到 25.3%, 相比 YOLO11n 提升 3.8%. 与此同时, 其召回率从 24.3% 提升至 29.3%, 说明模型对潜在目标具有更高的识别能力.

4.5 可视化结果与分析

为了进一步评估 ReF-YOLO 在复杂环境中对小尺度目标的感知能力, 本文选取 VisDrone2019 和 TinyPerson 数据集中具有代表性的图像样本, 利用热力图技术进行了可视化分析, 相关结果展示于图 7 和图 8 中, 红框内是差异明显的检测目标区域.

在 VisDrone2019 场景中, ReF-YOLO 能够更准确地定位远处道路上的行人、自行车等小型目标. 在基线模型中, 这些目标常由于纹理弱化和背景干扰而未能有效激活, 而 ReF-YOLO 在对应区域的响应更强, 热力图中显著增强, 体现了主干中 DCASC 模块在特征提取方面的优势.

(a) 原始图 (b) ReF-YOLO结果 (c) YOLO11n结果

图 7 VisDrone2019 数据可视化热力图

(a) 原始图 (b) ReF-YOLO结果 (c) YOLO11n结果

图 8 TinyPerson 数据可视化热力图

在 TinyPerson 数据集被遮挡和分布稠密的小目标场景中 (目标尺寸多小于 20 像素), 热图可视化结果表明 ReF-YOLO 能够在多个复杂区域内持续激活关键目标区域, 特别是在海边及远距离遮挡环境中展现出良好的关注能力, 热区与真实目标分布高度吻合, 体现出该模型在复杂背景与小目标密集情况下的鲁棒性与稳定性.

此外, 通过与 YOLO11n 的可视化结果对比可观察到, ReF-YOLO 的特征响应范围更加聚焦于目标本体, 背景区域抑制明显, 说明 L-Fuse 和 P2 检测分支增强了模型在浅层特征融合与细粒度检测上的表现. 这进一步从视觉层面验证了所提出结构在提升小目标检测准确性的有效性与必要性.

5 结论与展望

本文针对远距复杂场景中小目标存在目标尺寸极小、边界模糊、纹理弱、易被遮挡等检测挑战, 提出了一种基于 YOLO11 的改进检测算法 ReF-YOLO. 该算法围绕“增强特征感知、优化特征融合、扩展尺度支持”这 3 方面开展结构优化, 具体包括引入空间-通道协同特征提取模块 C3k2DCASC, 提升主干网络对非规则小目标的建模能力; 设计融合主干同尺度信息的 L-Fuse 结构与高效下采样策略 SCDown, 缓解语义断层与特征衰减问题; 并在检测头中增设高分辨率 P2 分支, 有效提升对极小目标的感知与定位能力. 实验在 VisDrone2019 与 TinyPerson 两个典型小目标检测数据集上进行, 结果显示 ReF-YOLO 较现有主流方法取得显著性能提升, 尤其在人、行人等高密度遮挡场景下的 mAP@0.5 提升超过 7 个百分点, 验证了所提方法在复杂背景下的实用性与鲁棒性.

未来工作可从以下几个方面进一步拓展: 一是引入跨尺度注意力机制与多模态辅助信息 (如红外、深度图像), 以增强小目标在极端场景下的表征能力; 二是结合自监督预训练或迁移学习策略, 提升模型在低标注或新领域数据下的适应能力; 三是探索基于动态推理或显著性引导的轻量推理策略, 进一步降低部署成本; 最后, 也可尝试将 ReF-YOLO 与检测-跟踪一体化模型结合, 以满足实时远距监控等场景的更高需求. 本研究为复杂小目标检测任务提供了一种兼顾精度与效率的可行路径, 为后续小目标检测算法设计提供了重要参考.

参考文献

1 高新波, 莫梦竟成, 汪海涛, 等. 小目标检测研究进展. 数据采集与处理, 2021, 36(3): 391–417.

2 Shi TJ, Gong JN, Hu JM, *et al*. Feature-enhanced CenterNet for small object detection in remote sensing images. Remote Sensing, 2022, 14(21): 5488. [doi: 10.3390/rs14215488]

3 Liu J, Yang SJ, Tian L, *et al*. Multi-component fusion network for small object detection in remote sensing images. IEEE Access, 2019, 7: 128339–128352. [doi: 10.1109/ACCESS.2019.2939488]

4 Zhong H, Li F, Kuang P, *et al*. Small object detection algorithm based on context information and attention mechanism. Proceedings of the 19th International Computer Conference on Wavelet Active Media Technology and Information Processing. Chengdu: IEEE, 2022. 1–6.

5 Sun Y, Liu WK, Gao YT, *et al*. A dense feature pyramid network for remote sensing object detection. Applied Sciences, 2022, 12(10): 4997. [doi: 10.3390/app12104997]

6 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]

7 Cai ZW, Vasconcelos N. Cascade R-CNN: High quality object detection and instance segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(5): 1483–1498. [doi: 10.1109/TPAMI.2019.2956516]

8 Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37.

9 Redmon J, Divvala S, Girshick R, *et al*. You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779–788.

10 Khanam R, Hussain M. YOLOv11: An overview of the key architectural enhancements. arXiv:2410.17725, 2024.

11 Wang A, Chen H, Liu LH, *et al*. YOLOv10: Real-time end-to-end object detection. Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver: ACM, 2024. 3429.

12 Lin TY, Dollár P, Girshick R, *et al*. Feature pyramid networks for object detection. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 936–944.

13 Liu S, Qi L, Qin HF, *et al*. Path aggregation network for instance segmentation. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Salt Lake City: IEEE, 2018. 8759–8768.
14 Chen Q, Wang YM, Yang T, *et al*. You only look one-level feature. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13034–13043.
15 袁帅, 王康, 单义, 等. 基于多分支并行空洞卷积的多尺度目标检测算法. 计算机辅助设计与图形学学报, 2021, 33 (6): 864–872.
16 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141.
17 Woo S, Park J, Lee JY, *et al*. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. Munich: IEEE, 2018. 3–19.
18 Yang LX, Zhang RY, Li LD, *et al*. SimAM: A simple, parameter-free attention module for convolutional neural networks. Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021. 11863–11874.
19 Sun J, Gao HW, Wang XN, *et al*. Scale enhancement pyramid network for small object detection from UAV images. Entropy, 2022, 24(11): 1699. [doi: 10.3390/e24111699]
20 苏继贤. 基于超分辨率重建的小目标检测算法研究. 建模与仿真, 2023, 12(3): 3224–3237.
21 Yamanaka J, Kuwashima S, Kurita T. Fast and accurate image super resolution by deep CNN with skip connection and network in network. Proceedings of the 24th International Conference on Neural Information Processing. Guangzhou: Springer, 2017. 217–225.
22 Jin Y, Zhang Y, Cen YG, *et al*. Pedestrian detection with super-resolution reconstruction for low-quality image. Pattern Recognition, 2021, 115: 107846.
23 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6517–6525.
24 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv:1804.02767, 2018.
25 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934, 2020.
26 Jocher G, Chaurasia A, Stoken A, *et al*. ultralytics/yolov5: v6.2-YOLOv5 classification models, Apple M1, reproducibility, ClearML and Deci.ai integrations. https://ui.adsabs.harvard.edu/abs/2022zndo…7002879J/abstract. (2022-08-17).
27 Li CY, Li LL, Jiang HL, *et al*. YOLOv6: A single-stage object detection framework for industrial applications. arXiv:2209.02976, 2022.
28 Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 7464–7475.
29 Kukartsev VV, Ageev RA, Borodulin AS, *et al*. Deep learning for object detection in images development and evaluation of the YOLOv8 model using ultralytics and roboflow libraries. Proceedings of the 13th Computer Science Online Conference. Cham: Springer, 2024. 629–637.
30 Wang CY, Yeh IH, Liao HYM. YOLOv9: Learning what you want to learn using programmable gradient information. Proceedings of the 18th European Conference on Computer Vision. Milan: Springer, 2025. 1–21.
31 Tian YJ, Ye QX, Doermann DS. YOLOv12: Attention-centric real-time object detectors. arXiv:2502.12524, 2025.
32 Du DW, Zhu PF, Wen LY, *et al*. VisDrone-DET2019: The vision meets drone object detection in image challenge results. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Seoul, 2019. 213–226.
33 Yu XH, Gong YQ, Jiang N, *et al*. Scale match for tiny person detection. Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision. Snowmass: IEEE, 2020. 1246–1254.
34 Jiang HZ, Learned-Miller E. Face detection with the faster R-CNN. Proceedings of the 12th IEEE International Conference on Automatic Face & Gesture Recognition. Washington: IEEE, 2017. 650–657.
35 Lin TY, Goyal P, Girshick R, *et al*. Focal loss for dense object detection. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2999–3007.

(校对责编: 张重毅)