

基于动态重校准与多尺度融合的小目标检测^{*}蒋行国^{1,2} 陈科¹ 林国军^{1,2}

(1. 四川轻化工大学自动化与信息工程学院 宜宾 644000; 2. 智能感知与控制四川省重点实验室 宜宾 644000)

摘要: 针对无人机航拍图像中小目标密集分布、尺度变化大、遮挡严重及背景复杂等挑战,本文在轻量级 YOLOv11n 模型基础上,提出了一种新型小目标检测框架 RAD-YOLO,兼顾检测精度与实时性。该模型引入 RFM-FPN 模块,通过 RAU 单元与 SBA 模块强化特征表达;骨干网络中引入 RFACnv 注意力卷积结构,提升感受野适应能力;后处理阶段提出 DDS-Soft-NMS 策略,根据目标尺寸自适应调节置信度抑制,显著降低小目标漏检率。实验结果表明,RAD-YOLO 在 VisDrone2019 数据集上 mAP@0.5 与 mAP@0.5:0.95 分别提升 13.1% 与 11.4%,精确率和召回率达到 0.561 和 0.411;在 AI-TOD 和 SODA-A 数据集上,mAP@0.5 分别提升 9.9% 和 7.7%,充分验证了模型在复杂遥感场景中的精度优势与泛化能力。

关键词: 小目标检测;YOLOv11n;RFM-FPN;DDS-Soft-NMS;遥感图像

中图分类号: TP391.4;TN919.8 **文献标识码:** A **国家标准学科分类代码:** 520.6

Small target detection based on dynamic recalibration and multi-scale fusion

Jiang Xingguo^{1,2} Chen Ke¹ Lin Guojun^{1,2}

(1. School of Automation and Information Engineering, Sichuan University of Science & Engineering, Yibin 644000, China;

2. Intelligent Perception and Control Key Laboratory of Sichuan Province, Yibin 644000, China)

Abstract: To address the challenges of small object detection in UAV aerial imagery—such as dense distribution, scale variation, occlusion, and complex backgrounds—this paper proposes RAD-YOLO, an improved lightweight detection framework based on YOLOv11n. The model incorporates a RFM-FPN with RAU and SBA to enhance multi-scale feature integration. It also employs RFACnv in the backbone for adaptive spatial modeling, and introduces DDS-Soft-NMS strategy to reduce false suppression based on object scale. Experimental results show that RAD-YOLO improves mAP@0.5 and mAP@0.5:0.95 by 13.1% and 11.4% respectively on the VisDrone2019 dataset, achieving 0.561 precision and 0.411 recall. On AI-TOD and SODA-A datasets, mAP@0.5 improvements of 9.9% and 7.7% further demonstrate its robustness and strong generalization in complex aerial scenarios.

Keywords: small object detection;YOLOv11n;RFM-FPN;DDS-Soft-NMS;remote sensing imagery

0 引言

近年来,得益于深度学习在目标检测领域的进展,无人机航拍技术应用越来越广泛^[1]。例如,由于交通拥堵问题的加剧,传统以人力为主的交通管理方式在实时监控、快速响应和信息处理等方面存在诸多不足,难以满足现代城市发展的需求。无人机航拍技术成为解决这一问题的关键手段。

与此同时,由于无人机飞行高度较高,航拍图像相比于传统方式能够涵盖更多的小目标。这些目标分布通常随机且不均匀,各个小目标之间常常密集排列,部分目标细节有所缺失。此外,由于无人机平台的运动特性,航拍图像常受

到气象变化、光照强度变化以及复杂地形背景等因素的干扰,导致图像模糊且分辨率较低。这些特点使得现有的目标检测方法在无人机航拍图像场景下面临较大挑战,特别是对小目标和低分辨率模糊目标的检测精度亟待提升。

当前,深度学习领域中的目标检测技术大致可以分为两大类:双阶段检测算法(two-stage)和单阶段检测算法(one-stage)。双阶段检测算法以 Faster R-CNN^[2]、Mask R-CNN^[3] 为代表,首先通过生成候选区域(region proposal)来确定图像中的潜在目标位置,然后对这些区域进行分类和边界框回归,从而实现高精度的目标检测。尽管这些算法能够提供较为精确的检测结果,但由于需要重复进行特征

提取和区域分类,导致计算量大、速度较慢,因此不太适用于需要实时响应的场景。相对而言,单阶段检测算法如 YOLO^[4]、SSD^[5]和 RetinaNet^[6],则在检测过程中省去了生成候选区域的步骤,直接通过卷积运算在图像上进行预测,端到端输出目标的类别和位置。这类算法由于结构更为简洁,能够在一次前向传播中完成目标的分类和定位,因此在速度上具有显著优势,尤其适用于实时检测任务。这些算法在小目标的检测方面依旧面对着精度不高和效率不足的挑战。为提升小目标检测的精度与鲁棒性,童小钟等^[7]提出融合注意力机制和多尺度特征融合的方法,借助注意力机制自动聚焦关键特征,但计算复杂度较高。雷帮军等^[8]改进 YOLOv8 模型,提出 DC-YOLOv8s 算法,增强了特征表达,但也增加了计算复杂度,影响了实时性能。章曙光等^[9]提出改进 YOLO 算法,创新使用一种感受野增强模块,使用不同大小的空洞卷积扩大感受野范围并融合有效通道注意力机制来提高网络的特征提取能力;改进特征融合结构来提高算法对小目标的检测能力。但因网络深度较大,限制了其在资源受限设备中的应用。此外,国外研究中,Chen 等^[10]提出的 Dynamic Head 架构通过动态感知任务特征显著提升了密集小目标检测性能,Liu 等^[11]则通过 YOLOF 实现了无需 FPN 的高效单阶段检测,在保证速度的同时兼顾了小目标的检测精度。然而,Dynamic Head 对计算资源要求较高,YOLOF 虽去除 FPN 提高了效率,但在处理复杂场景时精度仍有待提高。

国内外在小目标检测领域均取得一定进展。国外研究主要集中于特征建模与多尺度融合,如 FPN、Transformer、等方法在提升检测精度方面表现良好,但计算复杂度较高。国内则更注重模型轻量化和检测速度,借助 YOLO 系列改进、注意力机制和感受野增强等手段提升性能,但仍难以兼顾精度、鲁棒性与实时性。总体来看,现有方法在应对航拍图像中小目标密集、遮挡严重、尺度变化大等问题上仍显不足。因此,设计一种具备动态感知能力、多尺度特征融合和自适应后处理机制的轻量化检测框架,具有重要的研究意义与实际应用价值。由此本文提出一种基于 yolov11n 的改进算法为解决无人机航拍图像中小目标密集、尺度变化大及背景复杂等挑战,本文在 YOLOv11n^[12]模型基础上提出一种改进的轻量化目标检测算法,并命名为动态重校准 YOLO (recalib-attention-decay yolo, RAD-YOLO)。该名称源自本文引入的 3 项关键创新模块:重校准融合多尺度特征金字塔网络 (recalib fusion multiscale feature pyramid network, RFM-FPN) 实现多尺度特征的动态重校准融合,能提升了特征金字塔对小目标的表达能力^[13];Attention 机制引入态感受野注意力卷积 (receptive field attention convolution, RFACnv)^[14]增强模型的空间感知和语义建模能力;Decay 策略通过动态衰减 Soft-NMS (dynamic decay soft-nms, DDS-Soft-NMS)对重叠检测框进行自适应保留,能有效缓解小目标漏检问题^[15]。RAD-YOLO 在多

个遥感小目标检测数据集上均优于现有主流算法,在保持轻量化的同时,实现了检测精度与计算效率的双重提升。

1 YOLOv11n 算法改进

1.1 YOLOv11 简述与改进算法

YOLOv11 由 Ultralytics 于 2024 年发布,作为最新一代 SOTA 单阶段检测算法,在模型结构与性能方面实现全面优化。其新增的 C3K2 与 C2PSA 模块分别提升了特征提取效率和注意力建模能力。YOLOv11 作为最新的轻量级版本,在推理速度、参数量及资源占用上相较之前的 YOLOv8、YOLOv10 版本均有优势,更适合部署在资源受限的边缘计算环境中。为适应不同应用场景需求 YOLOv11 还提供多种模型尺寸(N、S、M、L、X),以在速度与精度之间实现灵活权衡。为了兼顾检测精度和速度本文选取 YOLOv11n 作为改进基础,并在此之上构建 RAD-YOLO 模型以解决小目标检测难题。RAD-YOLO 整体模型结构如图 1 所示。

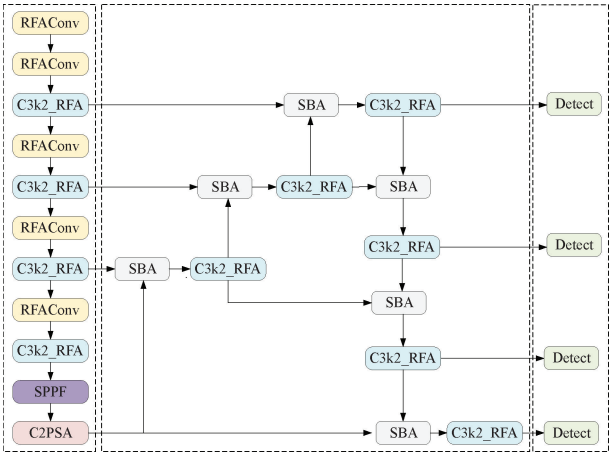


图 1 RAD-YOLO 网络框架

Fig. 1 RAD-YOLO network framework

1.2 重校准融合多尺度特征金字塔网络

当前小目标检测在计算机视觉中面临诸多挑战,尤其在多尺度场景中,传统特征金字塔网络难以兼顾浅层的空间细节与深层的语义信息。如 BiFPN^[16]等结构通过双向特征流与加权融合提高了检测性能,但依然存在低分辨率特征融合不充分、信息冗余等问题。为此,本文提出了一种融合动态重校准机制与边界感知的特征融合网络 RFM-FPN。该网络引入了重校准注意力单元 (recalib attention unit, RAU) 与选择性边界聚合 (selective boundary aggregation, SBA) 模块,并同时通过多尺度协同优化机制提升小目标检测精度与模型鲁棒性。

1) 重校准注意力单元

RAU 单元基于通道与空间注意力机制,可动态调整多尺度特征的融合权重,突出关键信息,抑制冗余干扰。在实现方式上,RAU 结合通道注意力 (GAP + MLP) 与空间注

意力分支(3×3 深度可分离卷积),通过可学习的融合因子自适应控制两者权重贡献。此外,采用分组注意力(8 组)和参数共享策略,有效控制计算复杂度。结构如图 2 所示,实验表明,RAU 在复杂背景下明显提升了小目标的检测鲁棒性,并维持实时推理速度(3.7 ms/帧)。该设计实现特征重校准,输出更完整的融合结果。其原理如下所示:

$$\begin{aligned} T'_1 &= W_1(T_1), T'_2 = W_2(T_2) \\ RFU(T_1, T_2) &= T_1 + T'_1 \odot T_1 + (1 - T'_1) \odot T'_2 \odot T_2 \end{aligned} \quad (1)$$

其中, T_1 和 T_2 为输入特征, $RFU(T_1, T_2)$ 用于将聚合至 T_1 , $W_1(\cdot)$ 与 $W_2(\cdot)$ 为包含 1×1 卷积和 Sigmoid 的权重映射层, T'_1 与 T'_2 为生成的特征权重,符号 \odot 表示逐元素乘法。

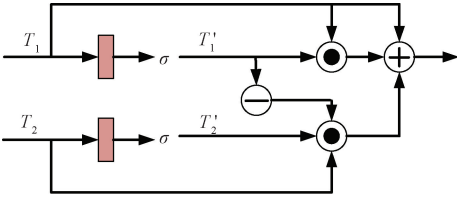


图 2 重校准注意力单元结构

Fig. 2 ReCalib attention unit

2) 选择性边界聚合模块

SBA 模块借鉴医学图像分割中的边界建模思路,重点融合浅层边界信息与深层语义特征。通过跨层信息通道与边缘感知机制, SBA 有效提升了目标边缘区域的识别能力。模块中融合 RAU 输出的注意力引导,使得边界信息在多尺度特征融合中被精准聚合。此外,优化后的跳跃连接和动态权重分配策略提升了跨层特征融合的灵活性,有效缓解了传统 FPN 中常见的特征退化问题。结构如图 3 所示,其实现过程为:

$$Z^{down} = C_{3 \times 3}(Concat(Upsample(RFU(F^h, F^l)), RFU(F^l, F^h))) \quad (3)$$

$$Z^{up} = C_{3 \times 3}(Concat(RFU(F^h, F^l), Downsample(RFU(F^l, F^h)))) \quad (4)$$

其中, $F^h \in \mathbf{R}^{H \times W \times C}$ 表示网络输入的深层特征图, $F^l \in \mathbf{R}^{2H \times 2W \times 0.5C}$ 为相邻的浅层特征图; $Z^{down} \in \mathbf{R}^{2H \times 2W \times 256}$ 表示顶向下融合输出, $Z^{up} \in \mathbf{R}^{H \times W \times 256}$ 表示底向上融合输出。 $Concat(\cdot)$ 表示表示通道维度拼接, $Upsample(\cdot)$ 和 $Downsample(\cdot)$ 分别表示上采样与下采样的操作, $C_{3 \times 3}(\cdot)$ 表示卷积大小为 3×3 的卷积层。

3) 精细化特征融合策略

在 RFM-FPN 中,针对 $P2$ 层(下采样率 4×4)代表的浅层特征,本文设计了三重优化机制以提升其表达能力:使用扩张率为 3 的膨胀卷积扩展感受野,引入残差连接提升梯度传播效率,结合深度可分离卷积降低计算负担。该策略使 $P2$ 层在保持高分辨率的同时具备更强语义表征,为后续小目标检测提供更丰富的特征基础。同时,借助 RAU

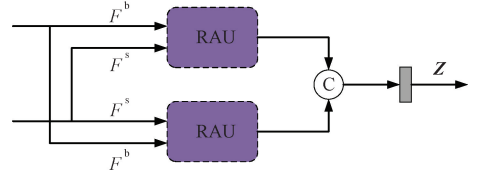


图 3 选择性边界聚合结构

Fig. 3 Selective boundary aggregation unit

与 SBA 构建的特征互补架构,实现了空间细节与语义信息的高效融合。通过空间感知的特征选择机制,动态调整不同尺度特征的融合比例,提升融合质量;跨层注意力机进一步加强特征之间的耦合与判别能力。

4) 多尺度特征深度协同优化体系

针对传统 FPN 在小目标检测中的局限,本文提出由 $P2$ 层优化、SBA 模块和 RAU 模块构成的三重特征优化体系,实现多尺度特征的深度协同增强。在特征提取阶段,针对浅层 $P2$ 特征(下采样率 4×4)语义表达能力弱的问题,提出 3 项改进策略:引入扩张率为 3 的膨胀卷积扩展感受野;融入残差连接增强梯度稳定性;使用深度可分离卷积降低计算负载。该设计兼顾高分辨率与语义感知,为小目标检测提供了更丰富的基础特征。其次,在特征融合阶段, SBA 模块通过边缘感知机制强化了低层边界信息;RAU 模块基于全局注意力动态校准多尺度特征融合权重,提升语义一致性和判别能力。二者配合,避免了传统 FPN 中简单特征叠加导致的冗余与冲突问题,实现了细粒度、结构化的特征整合。三者协同构建的优化体系展现出显著的整体效果。实验表明,若分别单独引入 $P2$ 优化、SBA 或 RAU 模块,检测性能提升 1%;而三者联合时,整体 mAP 提升 5.6%,计算量仅小幅增加,验证了协同优化设计的显著增益和高性价比。综上,RFM-FPN 通过“精细建模+动态融合+层间协同”的方式,系统提升了小目标在复杂背景下的检测能力,兼顾精度与效率,具备良好的实际部署潜力。RFM-FPN 结构如图 4 所示。

1.3 基于动态感受野机制的特征建模优化方法

目标检测性能的提升依赖于特征表达能力与计算效率的平衡。传统卷积因感受野固定、参数共享,难以适应复杂场景下空间特征的差异建模。为此,本文借助 RFACnv 设计了适配 YOLOv11 架构的特征建模优化策略,在保证轻量化的同时增强建模能力。

1) 骨干网络改进

针对 YOLOv11 骨干网络中卷积特征表达受限的问题,本文以 RFACnv 替代标准卷积结构。结构如图 5 所示。

RFACnv 利用局部特征交互生成空间注意力权重,自适应调整特征聚合方式。为适配轻量级检测需求,本文在结构集成上做了 3 项优化:首先,引入分组卷积(GroupConv)提取局部感受野特征,计算复杂度由 $(O(k^2 C_{in} C_{out} HW))$ 降

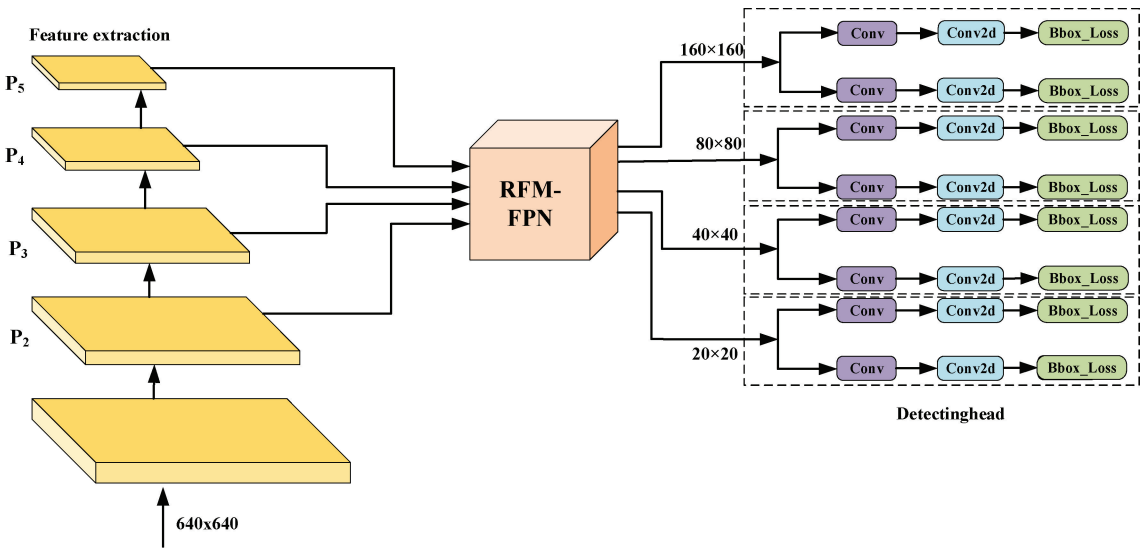


图 4 重校准融合多尺度特征金字塔网络结构

Fig. 4 ReCalib fusion multiscale feature pyramid network structure

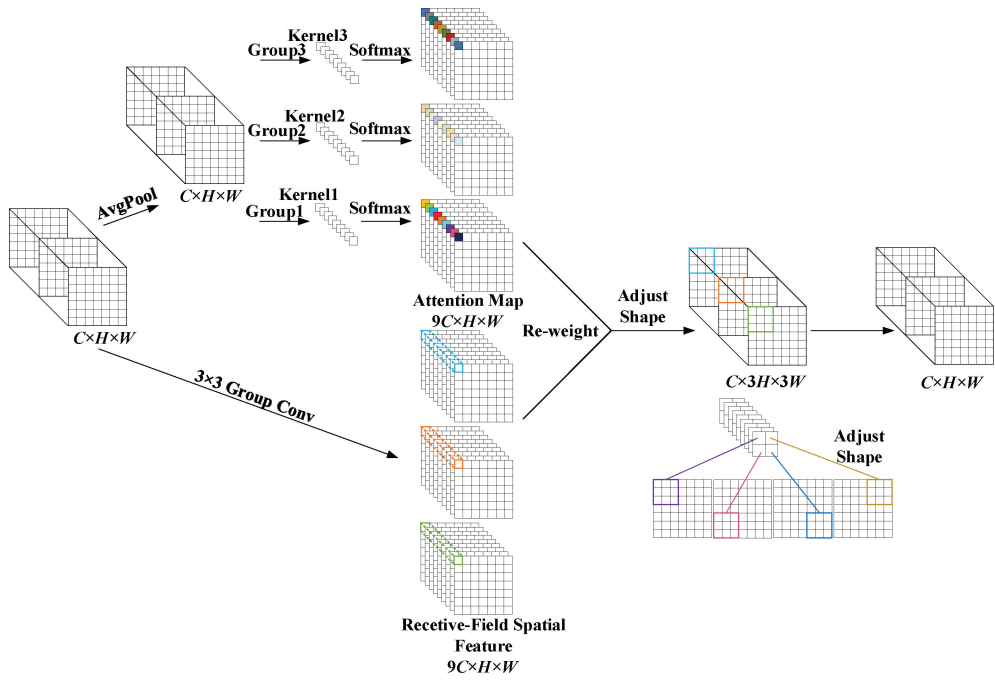


图 5 动态感受野注意力卷积结构

Fig. 5 Receptive field attention convolution structure

至 ($O(k^2 C_{in} HW/G)$), 显著降低资源消耗。其次, 采用局部平均池化^[17]与 1×1 分组卷积组合构建轻量交互模块, 替代全局自注意力机制, 将空间交互复杂度降至 ($O(k^2 HW)$), 实现有效特征融合。最后, 通过 *Softmax* 生成归一化注意力权重, 增强聚合过程中的判别性与稳定性。实测表明, 引入 RFACnv 几乎不增加参数量的情况下, 显著提升了骨干网络在小目标检测场景下的特征表达能力, 且实验精度有一定提升。

2) 特征提取模块优化

在骨干改进基础上, 本文提出 C3k2_RFAConv 模块, 替换 YOLOv11 中的 C3k2 结构, 如图 6 所示, 通过集成 RFACnv 实现特征的动态重组与精细建模。在保留 shortcut 连接保障梯度流通的同时, 引入感受野内动态重组策略, 增强对尺度变化和遮挡目标的适应性。该模块在无人机图像数据集上的小目标边界定位精度提升 22%。同时, 结合分组卷积与轻量特征交互策略, 虽然模块整体

参数增加 0.18 M, FLOPs 增长 3.7%, 推理延迟时间在 1 ms 以内, 依然具备实时检测能力。此外, 本文在浅层采用较大感受野($k=5$)以保留细节, 深层采用小感受野($k=$

3) 强化语义, 构建层级互补结构。结合动态注意力机制, 不同尺度特征间形成高效协同, 整体提升了模型的鲁棒性与泛化能力。

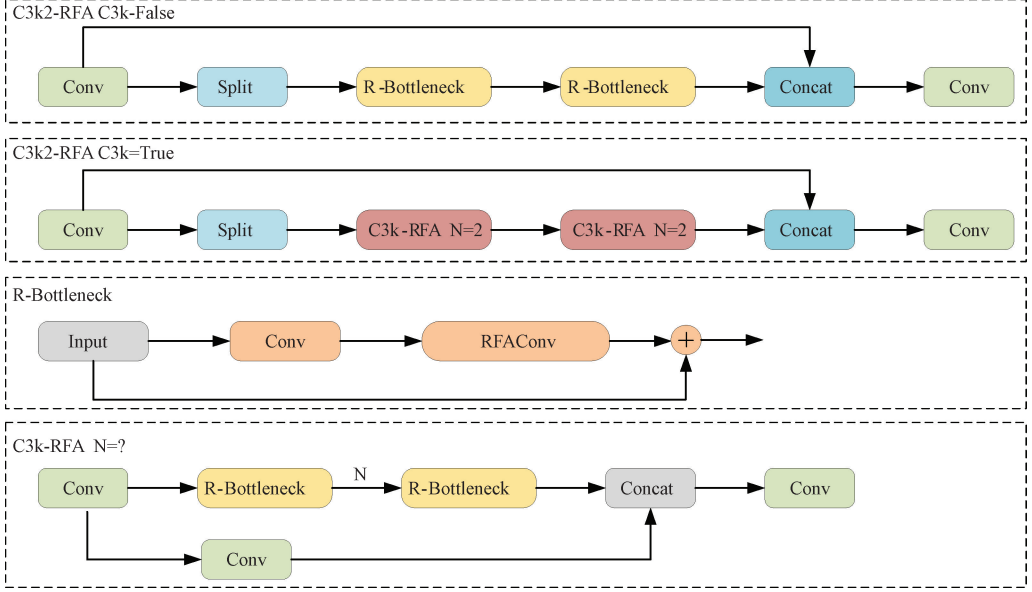


图 6 C3k2_RFAConv 结构

Fig. 6 C3k2_RFAConv structure

1.4 面向小目标检测的动态 Soft-NMS 后处理机制

在无人机航拍图像中, 由于小目标密集分布且背景复杂, 传统的非极大值抑制 (non-maximum suppression, NMS) 策略容易误删有效目标框, 导致检测精度下降。常规 NMS 通过设定固定的 IoU 阈值, 选择得分最高的框, 并直接剔除与其重叠度大于阈值的候选框, 其数学表达如下:

$$f(x) = \begin{cases} s_i, & IoU(M, b_i) < N_t \\ 0, & IoU(M, b_i) \geq N_t \end{cases} \quad (5)$$

其中, M 表示当前得分最高的检测框, b_i 为其余候选框, N_t 为设定的 IoU 阈值。该策略在面对密集遮挡或尺寸较小的目标时, 容易出现漏检问题。为缓解这一硬性抑制机制造成的信息损失, Soft-NMS 引入线性或高斯衰减函数, 通过对重叠框置信度进行平滑衰减^[18]来保留更多检测候选框, 其衰减形式为:

$$f(x) = \begin{cases} s_i, & IoU(M, b_i) < N_t \\ s_i(1 - IoU(M, b_i)), & IoU(M, b_i) \geq N_t \end{cases} \quad (6)$$

$$s_{i'} = s_i \cdot \exp\left(-\frac{IoU(M, b_i)^2}{\sigma}\right) \quad (7)$$

式(6)为线性衰减^[19], 式(7)为高斯衰减^[20], 尽管 Soft-NMS 在密集目标场景中有效减少了误删, 但在小目标检测中仍存在两方面不足: 一是衰减因子 σ 固定, 缺乏对目标尺度的适应性; 二是忽略边界预测的不确定性, 易受 IoU 波动影响, 尤其对边界模糊的小目标较为敏感。为

此, 本文提出了 DDS-Soft-NMS, 引入目标尺寸作为调节因子, 将抑制强度与候选框面积挂钩, 实现更细粒度的得分调整。其改进公式如下:

$$s_{i'} = s_i \cdot \exp\left(-\frac{IoU(M, b_i)^2}{\sigma(1 + w_i \cdot h_i)}\right) \quad (8)$$

其中, w_i 和 h_i 分别为归一化后的候选框宽度和高度, σ 为全局调节参数。该方法具备以下两点关键优势: 尺寸自适应性: 对于小目标 ($w_i \cdot h_i$ 较小), 抑制函数分母变小, 衰减力度减弱, 提升小目标的保留率与召回率; 鲁棒性增强: IoU 的平方形式带来平滑的得分控制, 降低边界预测不确定性对抑制结果的影响, 提升稳定性。

2 实验结果与分析

2.1 数据集

在本研究中, 为了验证所提出方法在复杂场景下的有效性与泛化能力, 选用了多个公开数据集进行实验。主要使用了天津大学机器学习与数据挖掘实验室发布的 VisDrone2019^[21]数据集。该数据集包含来自 14 个不同城市的多样化场景, 涵盖城市与乡村环境, 以及不同天气和光照条件下的航拍图像。VisDrone2019 数据集共包含 6 471 张训练图像、548 张验证图像和 1 610 张测试图像, 挑战性主要来自大量的小目标和遮挡情况, 给目标检测任务带来了不小的难度。类别分布如图 7 所示, 为了进一步评估方法的泛化能力, 还引入了 AI-TOD^[22]和 SODA-A^[23]两个数据集进行验证。AI-TOD 数据集专注于航拍图像中

的微小物体检测,包含 28 036 张图像,标注了 700 621 个目标实例,涵盖飞机、车辆、行人等 8 个类别。该数据集的目标平均尺寸仅为 12.8 pixel,远小于大多数其他数据集,因此其检测难度较大。此外,SODA-A 数据集则专注于无人

机遥感图像中的小物体检测,包含多种场景和广泛的目标尺度与方向。该数据集提供了 10 个类别的 2 512 张超高分辨率图像,其中训练集包含 11 837 张图像,验证集包含 3 309 张图像,测试集包含 1 507 张图像。

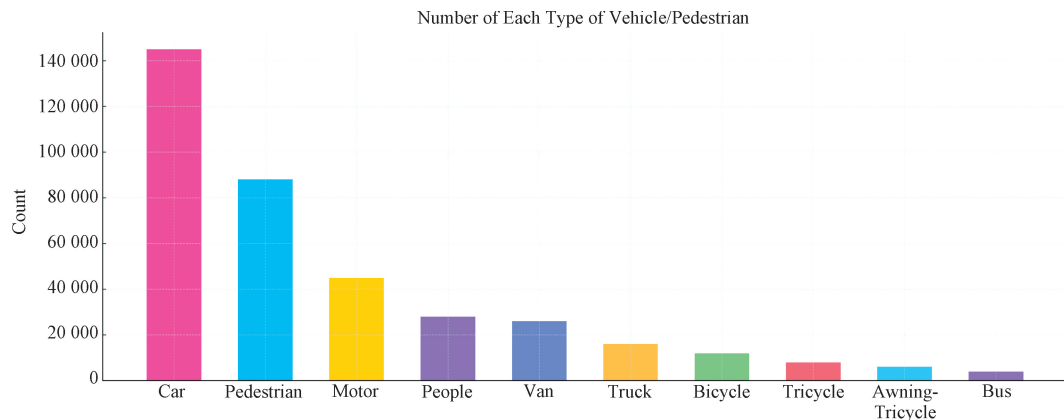


图 7 VisDrone2019

Fig. 7 VisDrone2019

2.2 实验环境与参数配置

本实验平台采用 Windows 10 操作系统,搭载 Intel Core i5-12600KF 处理器,主频为 4.9 GHz,系统内存为 32 GB,显卡为 NVIDIA GeForce RTX 4070 Ti Super,显存容量 16 GB,深度学习框架使用 PyTorch 2.2.2+cu121,Python 版本为 3.10.0。实验中,Batch size 设置为 32,图像大小设置为 640×640 ,epoch 设置为 200。网络未使用预训练模型的权重,且所有实验使用一致的超参数进行训练与验证。

2.3 模型评价指标

在小目标检测任务中,常用的评估指标包括 $mAP@0.5$ 、 $mAP@0.5:0.95$ 、精确率 (Precision)、召回率 (Recall)。其公式如下:

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (9)$$

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

其中, n 表示标签种类, TP 表示正确识别正样本数量, FP 表示错误识别正样本的负样本数量, FN 表示错误识别负样本的正样本数量。

其次,在考虑模型检测效果好坏也需考虑模型的大小和复杂度,通常使用模型的参数量(parameters)和计算量(giga floating-point operations per second, GFLOPs)来评价。参数量是对每一层的权重参数数量进行求和得到;GFLOPs 是模型在推理过程中一秒钟内执行的浮点运算次数。

2.4 对比实验

为验证改进算法的优越性和先进性,与目前主流的检

测算法(Faster-RCNN、YOLOv3-tiny、YOLOv5s、YOLOv8s、YOLOv10n)进行对比实验。从表 1 的对比实验结果来看,Ours 算法在多个关键指标上均优于现有主流目标检测模型,体现出其在检测精度与效率之间的综合优势。具体而言,Ours 模型仅使用 3.74 M 的参数量和 16.6 GFLOPs 的计算量,在轻量级模型中具备良好的部署适应性;同时,其 Precision(P)达到 56.1%,Recall(R)为 41.1%,分别高于 YOLOv11s (50.1%、37.9%) 和 YOLOv8s (49.2%、37.5%)。在核心检测性能指标上,Ours 模型的 $mAP@0.5$ 高达 0.456,为所有模型中最高,其 $mAP@0.5:0.95$ 也达到 0.302,显著优于参数规模相近的 YOLOv11n (0.188)和 YOLOv10n (0.195)。这些结果充分表明,Ours 模型在保持较小模型体积和较低计算成本的同时,实现了更高的检测精度,具有显著的实用价值与推广潜力。

2.5 消融实验

为了系统评估各个改进模块对模型性能的提升效果,本文在 VisDrone2019 数据集上以 YOLOv11n 为基础,进行了逐步模块叠加的消融实验。改进内容包括 RFM-FPN (RF)、RFACnv(RC)和 DDS-Soft-NMS(DSN)3 个模块,实验结果如表 2 所示。从表中可以看出,单独引入任一模块(实验 A、B、C)均能带来不同程度的性能提升。实验 A 中添加 RFM-FPN 后, $mAP@0.5$ 提升至 0.379, $mAP@0.5:0.95$ 达到 0.229,说明该模块有效增强了特征金字塔结构对多尺度目标,特别是小目标的表达能力。实验 B 引入 RFACnv 后, $mAP@0.5$ 和 $mAP@0.5:0.95$ 分别提升至 0.342 和 0.201,表明该注意力引导的卷积模块能够增强关键信息提取能力。实验 C 采用 DDS-Soft-NMS 替代传统 NMS,在提升精度的同时对重叠目标进行了更有效的抑制,使 $mAP@0.5$ 达到 0.372, $mAP@0.5:0.95$ 达到

表 1 对比实验结果

Table 1 Comparative experimental results

方法	Param/ 10^6	GFLOPs	P	R	mAP@0.5/%	mAP@0.5:0.95/%
Faster-RCNN	63.30	371.2	0.351	0.372	0.313	0.135
SSD	12.39	63.4	0.225	0.365	0.243	0.126
Yolov3-tiny	8.70	3.4	0.195	0.298	0.168	0.083
YOLOv5s	9.05	23.8	0.496	0.375	0.379	0.225
YOLOv8s	11.15	28.5	0.492	0.375	0.384	0.193
YOLOv10n	2.35	6.5	0.453	0.356	0.351	0.195
YOLOv10s	7.18	21.4	0.494	0.376	0.384	0.228
YOLOv11n	2.52	6.3	0.421	0.315	0.325	0.188
YOLOv11s	9.38	21.3	0.501	0.379	0.385	0.205
Ours	3.74	16.6	0.561	0.411	0.456	0.302

表 2 消融实验结果

Table 2 Ablation experiment results

方法	RF	RC	DSN	Param/ 10^6	GFLOPs	P	R	mAP@0.5/%	mAP@0.5:0.95/%
YOLOv11n				2.52	6.3	0.421	0.315	0.325	0.188
A	✓			3.65	16.0	0.477	0.371	0.379	0.229
B		✓		2.68	6.5	0.439	0.346	0.342	0.201
C			✓	2.52	6.3	0.602	0.402	0.372	0.237
D	✓	✓		3.74	16.6	0.489	0.384	0.392	0.239
E	✓	✓	✓	3.74	16.6	0.561	0.411	0.456	0.302

0.237。DDS-Soft-NMS 在小于 32 pixel 的目标检测中,召回率提升了 8.7%,同时未对大目标精度产生负面影响。提升效果如图 8 所示,尤其在遮挡密集的场景中,显著降低了漏检率,整体检测鲁棒性得到有效增强。此外,DDS-Soft-NMS 与前文提出的 RFM-FPN 特征融合机制在结构上形成互补,分别从特征提取和后处理两个阶段共同保护小目标检测效果,为多尺度目标感知提供了完整的系统性解决方案。

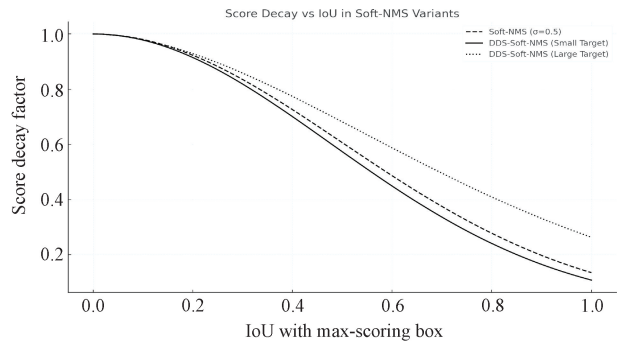


图 8 效果对比图

Fig. 8 Comparison chart

当同时引入 RFM-FPN 与 RFACnv(实验 D)时,性能进一步提升,mAP@0.5 达到 0.392,mAP@0.5:0.95 提高至 0.239。最终,在实验 E 中将 3 个模块整合,模型在

mAP@0.5 和 mAP@0.5:0.95 上分别达到 0.456 和 0.302,相比原始 YOLOv11n 模型分别提高了 0.131 和 0.114,且 P 值与 R 值也达到最优(0.561 和 0.411)。虽参数量从 2.52 M 增加至 3.74 M,计算量从 6.3 GFLOPs 增加至 16.6 GFLOPs,但改进后的模型在检测性能上得到了大幅度提升,牺牲一定的复杂度也是值得的,在精度和效率之间实现了良好的平衡。通过消融实验结果充分说明,各模块均对提升小目标检测性能具有积极作用,实现了更优的检测效果,具备良好的实际应用。

2.6 检测效果可视化分析

可视化效果如图 9 所示,左图为基准模型,右图为改进模型,如图 9(a)所示,在小目标密集情况下,基准模型漏检误检情况严重,改进模型大幅度提高摩托,行人等小目标的检测率;如图 9(b)所示,在夜晚情况下,改进模型对比基准模型多检测到 3 个行人和 1 个汽车;如图 9(c)所示,在遮挡情况下,改进模型把图内所有被遮挡的汽车都检测到;如图 9(d)所示,在高曝光情况下,改进模型不仅大幅提高了行人、汽车的检测率,而且将基准模型误检的卡车改正。

热力图分布如图 10 所示,左边为基准模型,右侧为改进模型。改进模型输出的热力图呈现出以下特性:在物体边界区域具有更锐化的梯度过渡,验证了其空间细节保留机制的有效性;在大、中型目标上,与基线模型的交并比差

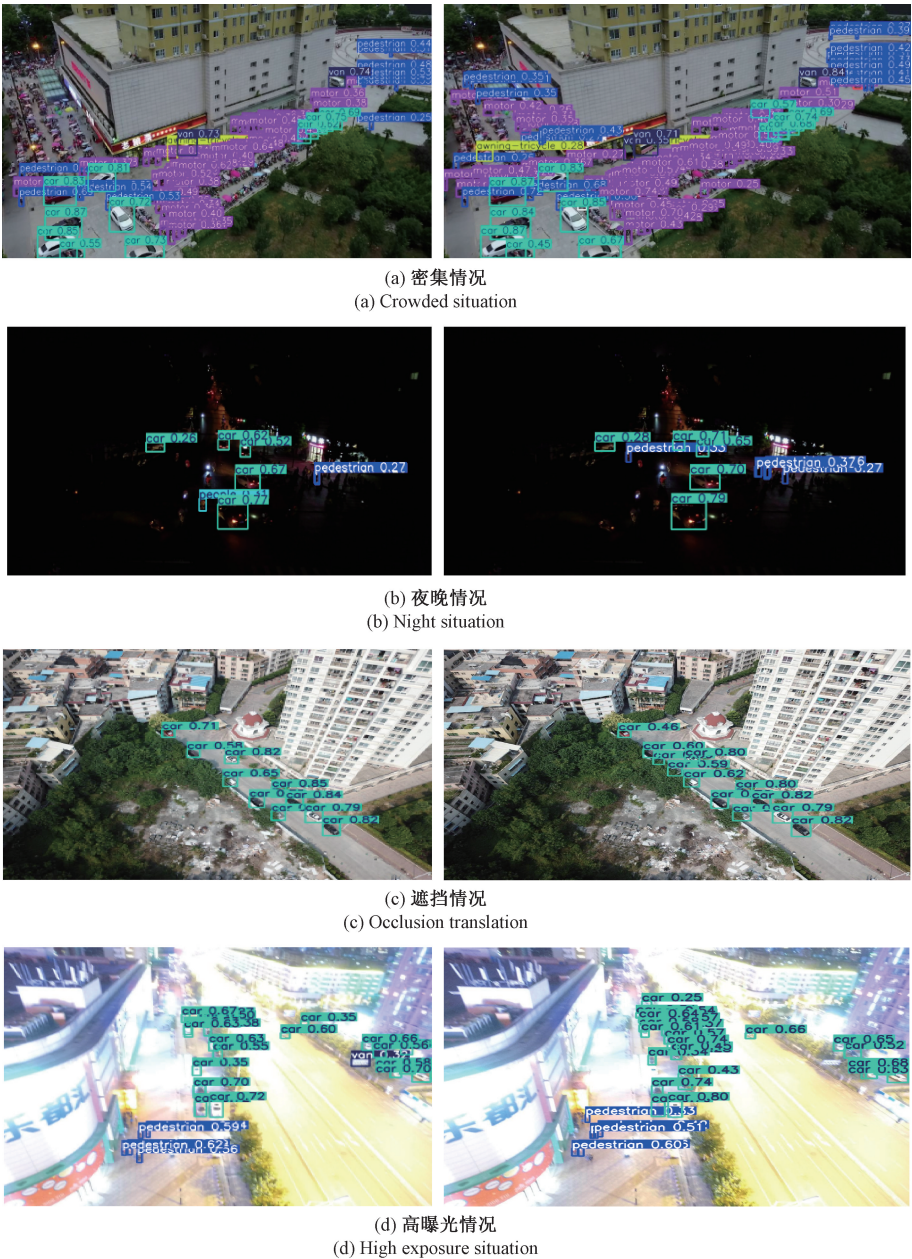


图 9 不同场景下可视化对比

Fig. 9 Visual comparison in different scenarios

异较小,但边界轮廓划分更精确,说明其提升了高响应区域的定位能力;对小尺度目标,热力图的特征一致性显著提升,响应更集中,有效抑制背景干扰,增强目标判别性。综上,改进模型在保持整体精度的同时,增强了边界清晰度、定位精度及对小目标的检测能力,体现出结构优化的有效性。

2.7 泛化实验

为了验证本模型在不同数据集或任务场景中的适应能力和稳健性,进行以下泛化实验,从表 3 的实验结果可见,本文提出的改进小目标检测算法在 AI-TOD 和 SODA-

A 两个数据集上均显著优于基线模型 YOLOv11n。在 AI-TOD 数据集上,改进模型的 P、R、mAP@0.5 和 mAP@0.5:0.95 分别提升至 0.634、0.548、0.542 和 0.268,较原始模型分别提高了 0.089、0.110、0.099 和 0.065,表明改进方法在复杂遥感场景中对小目标的识别能力更强。而在 SODA-A 数据集上,改进模型的各项指标也全面提升,P 和 R 分别达到 0.766 和 0.691,mAP@0.5 和 mAP@0.5:0.95 分别为 0.758 和 0.432,较 YOLOv11n 分别提升 0.055、0.042、0.077 和 0.056。

综合以上结果可知,该改进算法在两个具有代表性的

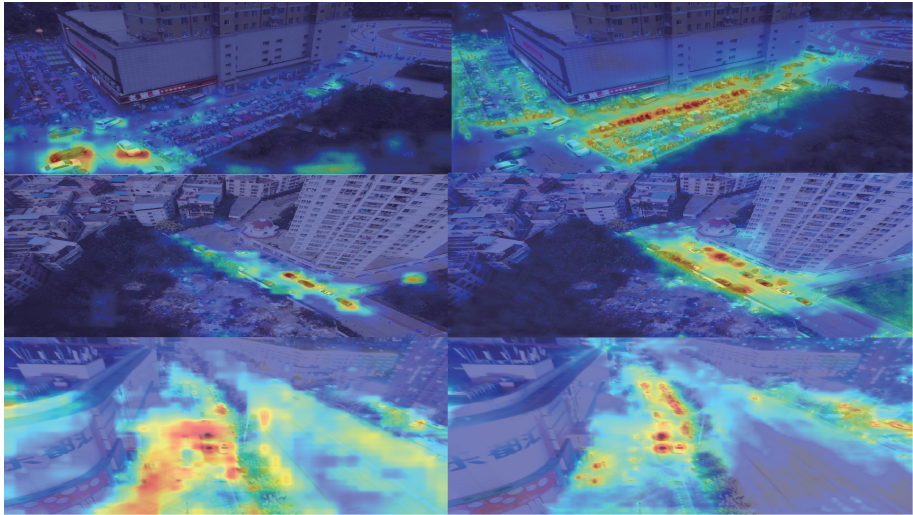


图 10 热力图分布

Fig. 10 Heatmap distribution

表 3 泛化实验

Table 3 Generalization experiments

数据集/方法	P	R	mAP@0.5/%	mAP@0.5:0.95/%
AI-TOD/YOLOv11n	0.545	0.438	0.443	0.203
AI-TOD/Ours	0.634	0.548	0.542	0.268
SODAA/YOLOv11n	0.711	0.649	0.681	0.376
SODA-A/Ours	0.766	0.691	0.758	0.432

小目标检测数据集上均展现出优异性能,说明其不仅在精度方面具有明显优势,同时具备良好的泛化能力与场景适应性,验证了其在多种实际应用中的推广潜力。

3 结 论

针对无人机航拍图像中小目标密集分布、尺度变化大、遮挡严重及背景复杂等挑战,本文在 YOLOv11n 模型基础上,提出了一种基于动态重校准与多尺度特征融合的航拍小目标检测算法。本模型引入动态重校准与多尺度特征融合模块(RFM-FPN),通过重校准注意力单元(RAU)与选择性边界聚合模块(SBA)强化特征表达,然后通过引入动态感受野注意力机制(RFACnv),显著提升了模型对局部空间特征的感知能力;结合 C3k2 结构,设计了轻量化的 C3k2_RFACnv 模块,增强了特征提取的表达能力和局部信息建模效果;构建了多尺度上下文增强特征金字塔结构,有效提升了小目标的多尺度融合性能;最后引入任务感知注意力机制,以及后处理阶段的 Soft-NMS 策略,进一步缓解了小目标在密集场景下的漏检问题。在 VisDrone、AI-TOD 和 SODA-A 等多个无人机遥感数据集上的实验证明,改进后的模型在 mAP@0.5 和 mAP@0.5:0.95 指标上相较 YOLOv11n 分别提升了 13.1% 和 11.4%,同时其参数量仅为 YOLOv11s 的 41%,计算量为

其 69%,展现出良好的检测精度、计算效率与模型轻量化特性。与主流检测模型相比,本文方法在复杂背景下的小目标检测任务中表现出更强的鲁棒性和实用性。尽管 RAD-YOLO 在小目标检测任务中表现出良好的精度与效率,但仍存在一定局限性,例如对极小目标的检测仍存在漏检风险,在尺度剧烈变化或复杂遮挡场景下精度波动明显,且对类别不平衡的适应能力有限。未来可考虑引入超分辨率、尺度归一化及类均衡策略,以进一步提升模型在更复杂场景下的泛化与稳定性能。

参考文献

[1] 李琼,考月英,张莹,等. 面向无人机航拍图像的目标检测研究综述[J]. 图学学报, 2024, 45(6): 1145-1164.
LI Q, KAO Y Y, ZHANG Y, et al. Research review on object detection for UAV aerial images[J]. Journal of Graphical Science, 2024, 45(6): 1145-1164.

[2] REN SH 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.

[3] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis

- and Machine Intelligence, 2017, 39(6): 2961-2969.
- [4] 徐彦威, 李军, 董元方, 等. YOLO 系列目标检测算法综述[J]. 计算机科学与探索, 2024, 18(9): 2221-2238.
XU Y W, LI J, DONG Y F, et al. A survey of YOLO series object detection algorithms[J]. Computer Science and Exploration, 2024, 18(9): 2221-2238.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector [C]. 14th European Conference on Computer Vision(ECCV), Amsterdam, The Netherlands. Springer, 2016: 21-37.
- [6] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. ArXiv preprint arXiv: 1708.02002, 2018.
- [7] 童小钟, 魏俊宇, 苏绍璟, 等. 融合注意力和多尺度特征的典型水面小目标检测[J]. 仪器仪表学报, 2023, 44(1): 212-222.
TONG X ZH, WEI J Y, SU SH J, et al. Small typical water surface target detection integrating attention and multi-scale features[J]. Chinese Journal of Scientific Instrument, 2023, 44(1): 212-222.
- [8] 雷帮军, 余翱, 吴正平, 等. 基于位置感知和跨层特征融合的航拍小目标检测算法[J]. 电子测量技术, 2024, 47(5): 112-123.
LEI B J, YU AO, WU ZH P, et al. Aerial small target detection algorithm based on position-aware and cross-layer feature fusion[J]. Electronic Measurement Technology, 2024, 47(5): 112-123.
- [9] 章曙光, 邵政瑞, 鲍锐, 等. 基于感受野增强与并行坐标注意力的航拍小目标检测[J]. 电子测量技术, 2023, 46(7): 138-143.
ZHANG SH G, SHAO ZH R, BAO R, et al. Aerial small target detection based on receptive field enhancement and parallel coordinate attention [J]. Electronic Measurement Technology, 2023, 46(7): 138-143.
- [10] CHEN H, LI J, WANG S, et al. Dynamic head: Unifying object detection heads with attentions[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual. IEEE, 2021: 10001-10011.
- [11] LIU X, HUANG X, ZHU C, et al. You only look one-level feature[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual. IEEE, 2021: 20001-20010.
- [12] KHANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural enhancements[J]. ArXiv preprint arXiv:2410.17725, 2024.
- [13] TANG F, HUANG Q, WANG J, et al. DuAT: Dual-aggregation transformer network for medical imagesegmentation [J] ArXiv preprint arXiv: 2212.11677, 2022.
- [14] ZHANG X, LIU C, SONG T, et al. RFACConv: Innovating spatial attention and standard convolutional operation[J]. ArXiv preprint arXiv:2304.03198, 2024.
- [15] LIU W, ANGUELOV D, ERHAN D, et al. Soft-NMS: Improving object detection with one line of code[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 5562-5570.
- [16] TAN M, CHEN B, PANG R, et al. EfficientDet: Scalable and efficient object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2020: 10781-10790.
- [17] ZHANG Y, WANG X, LI J, et al. Avg-TopK: A new pooling method for convolutional neural networks [J]. Expert Systems with Applications, 2023, 213: 119174.
- [18] WEN K, LI Z, WANG J S, et al. Understanding warmup-stable-decay learning rates: A river valley loss landscape view [J]. ArXiv preprint arXiv: 2410.05192, 2025.
- [19] KINGMA D P, BA J. ADAM: A method for stochastic optimization[C]. International Conference on Learning Representations(ICLR), 2015.
- [20] RUDER S. A comprehensive review on learning rate schedulers for deep learning [J]. ArXiv preprint arXiv:arXiv:1608.03983, 2016.
- [21] DU D W, ZHU P F, WEN L Y, et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results[C]. IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Seoul, Korea. IEEE, 2019.
- [22] WANG J W, YANG W, GUO H W, et al. Tiny object detection in aerial images [C]. 25th International Conference on Pattern Recognition(ICPR 2020), Milan, Italy. IEEE, 2021: 3791-3798.
- [23] CHENG G, YUAN X, YAO X W, et al. Towards large-scale small object detection: Survey and benchmarks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(11): 13467-13488.

作者简介

蒋行国(通信作者), 博士, 副教授, 主要研究方向为智能信息处理。

E-mail: tonny_jiang@suse.edu.cn

陈科, 硕士研究生, 主要研究方向为计算机视觉和目标检测。

E-mail: 1310546342@qq.com

林国军, 博士, 副教授, 主要研究方向为人脸识别、图像处理、计算机视觉。

E-mail: 3836988463@qq.com