

## 基于 YOLOv8-FRX 的无人机小目标检测改进算法<sup>①</sup>

林 波<sup>②</sup> 黄洪琼<sup>③</sup>

(上海海事大学信息工程学院 上海 201306)

**摘 要** 针对无人机拍摄图像中小目标分布密集、背景复杂以及传统只看一次(you only look once, YOLO)算法对小目标的误检和漏检率较高的问题,提出了一种改进的目标检测算法 YOLOv8-FRX(feature refinement and eXtended detection, FRX)。该方法通过以下 3 方面改进模型性能:第一,在骨干网络中引入重参数化跨阶段部分网络(reparameterized cross stage partial network, RepCSP),并在梯度流通分支上使用重参数化卷积(reparameterized convolution, RepConv),增强特征提取能力的同时减少参数量;第二,设计小目标增强金字塔(small object enhance pyramid, SOEP),提升小目标特征的捕获效率,同时保证计算效率;第三,采用加权交并比(wise intersection over union v3, Wise-IoU v3)策略,优化梯度增益分配,提高边界框回归的准确性。在 VisDrone 数据集上的实验结果显示,该方法的 mAP50 值相比基线模型提高了 4.0%,参数量减少了 3.6%。在无人机目标检测和跟踪数据集(unmanned aerial vehicle detection and tracking, UAVDT)上的泛化实验中,检测精度提升了 1.7%,验证了该方法的通用性与有效性。上述改进不仅提升了模型对小目标的检测能力,还在一定程度上平衡了精度与效率,为无人机图像中的目标检测提供了更优解决方案。**关键词** 无人机;小目标检测;重参数化跨阶段部分网络;小目标增强金字塔;加权交并比

随着无人机技术的迅速进步,其在军事侦察、环境监测、农业管理、灾难救援和城市管理等多个领域的应用日益广泛。特别是在动态环境中,基于无人机航拍图像的目标检测已经成为一项关键技术需求<sup>[1-3]</sup>。传统的目标检测方法严重依赖人工标注和视觉分析,不仅成本高且效率低,而且容易受到人为偏差的影响。相比之下,基于无人机的目标检测<sup>[4-6]</sup>技术将高分辨率传感器(如相机和红外摄像机)与计算机视觉和深度学习相结合,实现了目标检测和识别的高效自动化。

尽管无人机视角提供了大量数据,航拍图像中的目标检测仍然具有挑战性,特别是对于小目标。尺度变化大、背景复杂、目标之间严重遮挡和重叠等

问题限制了检测精度。因此,优化小型航拍目标的检测性能已成为一个重要的研究方向。

目前,主流的基于深度学习的目标检测算法主要分为单阶段和双阶段方法。双阶段方法,如区域卷积神经网络(region-based convolutional neural network, R-CNN)<sup>[7]</sup>、快速区域卷积神经网络(fast region-based convolutional neural network, Fast R-CNN)、Faster R-CNN 和 Mask R-CNN,通常采用 2 步流程:首先生成区域提议,然后进行精确的分类和定位。这些方法能提供较高的精度,但冗余的特征提取过程使得检测速度较慢。另一方面,单阶段方法,如单步多框检测器(single shot multibox detector, SSD)、RetinaNet 和只看一次(you only look once, YOLO)系

① 国家自然科学基金(62271303)资助项目。

② 男,2001 年生,硕士生;研究方向:深度学习目标检测;E-mail: 18912652240@163.com。

③ 通信作者,E-mail: hqhuang@shmtu.edu.cn。

(收稿日期:2025-01-20)

列<sup>[8-12]</sup>,在单一的特征提取步骤中同时进行目标分类和边界框回归,大大提升了实时性能。这些方法广泛应用于动态场景中,尽管其精度通常略低于双阶段方法。

为应对无人机图像中小目标检测的难题,研究人员提出了多种改进方法。例如,如 Liu 等人<sup>[13]</sup>提出跨阶段局部注意力模块,通过通道重加权增强小目标响应,但未解决深层特征分辨率衰减问题。Girshick 等人<sup>[14]</sup>构建级联空洞卷积金字塔,虽提升了多尺度特征覆盖度,却显著增加了内存开销。Lyu 等人<sup>[15]</sup>在 YOLOv7 中嵌入像素级注意力融合模块,有效抑制背景噪声,但对密集目标的区分能力不足。Cao 等人<sup>[16]</sup>通过引入多尺度特征处理模块增强了 YOLOv5,改善了小目标的检测精度,但增加了计算复杂度,影响了实时性能。Zhang 等人<sup>[17]</sup>通过引入三层路径聚合特征金字塔网络(path aggregation feature pyramid network, PAFPN)增强了 YOLOv8 的特征融合,取得了显著的性能提升,尽管其针对特定任务的适用性有限。Cui 等人<sup>[18]</sup>提出了优化感知网络(optimized perception network, OKNet),能够捕捉多尺度感受野,在细节提取上取得了突破,但其对局部特征的依赖可能会限制其有效处理全局特征的能力。Han 等人<sup>[19]</sup>引入了 Ghost 卷积以减少计算成本,同时提取更多特征,但深度网络训练不稳定以及在不同数据集和任务中的泛化验证不足等问题仍未解决。

尽管上述方法在不同维度改进了小目标检测性能,但仍存在 3 方面核心缺陷:其一,复杂网络模块(如密集注意力机制、多分支结构)的引入导致模型复杂度急剧上升,难以满足无人机平台的实时处理需求;其二,现有特征金字塔多采用层级间单向融合,低层细粒度特征在传递过程中易被高层语义信息稀释,造成小目标细节丢失;其三,传统损失函数未考虑无人机影像中大量低质量锚框(如严重遮挡、模糊目标)对梯度更新的干扰,导致边界框回归精度受限。

针对上述挑战,本文提出 YOLOv8-FRX 算法,在提升 YOLOv8 对小目标的检测精度的同时,优化了模型的参数量和检测速度,从而实现了模型的轻

量化<sup>[20-22]</sup>,具体表现为以下 3 个方面的改进。

(1) 在骨干网络中加入重参数化跨阶段部分网络(reparameterized cross stage partial network, RepC-SP)模块,舍弃 YOLOv8 中常用的残差模块,在梯度流通分支上使用重参数化卷积,以此来增强特征提取和梯度流通的能力,减少模型参数量的同时保持较高的检测性能。

(2) 设计小目标增强金字塔(small object enhance pyramid, SOEP),使用 P2 特征层经过结构重参数化可分离卷积(structural reparameterization depthwise convolution, SPDCConv)得到富含小目标信息的特征给到 P3 进行融合,加强小目标的特征捕获,保证计算效率。

(3) 引入加权交并比(wise intersection over union, Wise-IoU v3)作为损失函数,提供梯度增益分配策略,使模型更加关注低质量样本,提升边界框回归的准确性。

## 1 YOLOv8 算法简介

YOLOv8 是单阶段检测算法,经过精心设计,它不仅继承了 YOLO 系列快速检测和良好实时性的传统优点,还在准确性上取得了显著提升。YOLOv8 的架构由几个关键部分组成:主干网络、颈部网络和检测头。

主干网络使用了创新的跨阶段部分融合(cross-stage partial with fusion, C2f)、卷积块(Conv-BatchNorm-SiLU, CBS)和空间金字塔池化模块(spatial pyramid pooling fast, SPPF),这些模块协同作用,高效提取图像中的多种特征。颈部网络则负责融合主干网络输出的特征。YOLOv8 采用特征金字塔网络(feature pyramid network, FPN)和路径聚合网络(path aggregation network, PAN),通过多尺度特征融合技术,增强了模型对不同尺寸物体的检测能力。检测头在融合后的特征图上执行检测,并输出物体的位置信息和类别。

此外,YOLOv8 是一种无锚点的检测模型,通过 Task-Aligned 分配器动态分配训练样本,从而进一步提高检测精度。YOLOv8 提供了  $n, s, m, l$  和  $x$  这 5 种不同规模的模型,以满足多样化的计算资源和实

时性需求。其中, YOLOv8n 版本模型最轻量, 同时在检测速度上具有显著优势。本文对 YOLOv8n 进

行了改进, 旨在保持高检测精度的同时满足轻量化设计。YOLOv8 网络框架如图 1 所示。

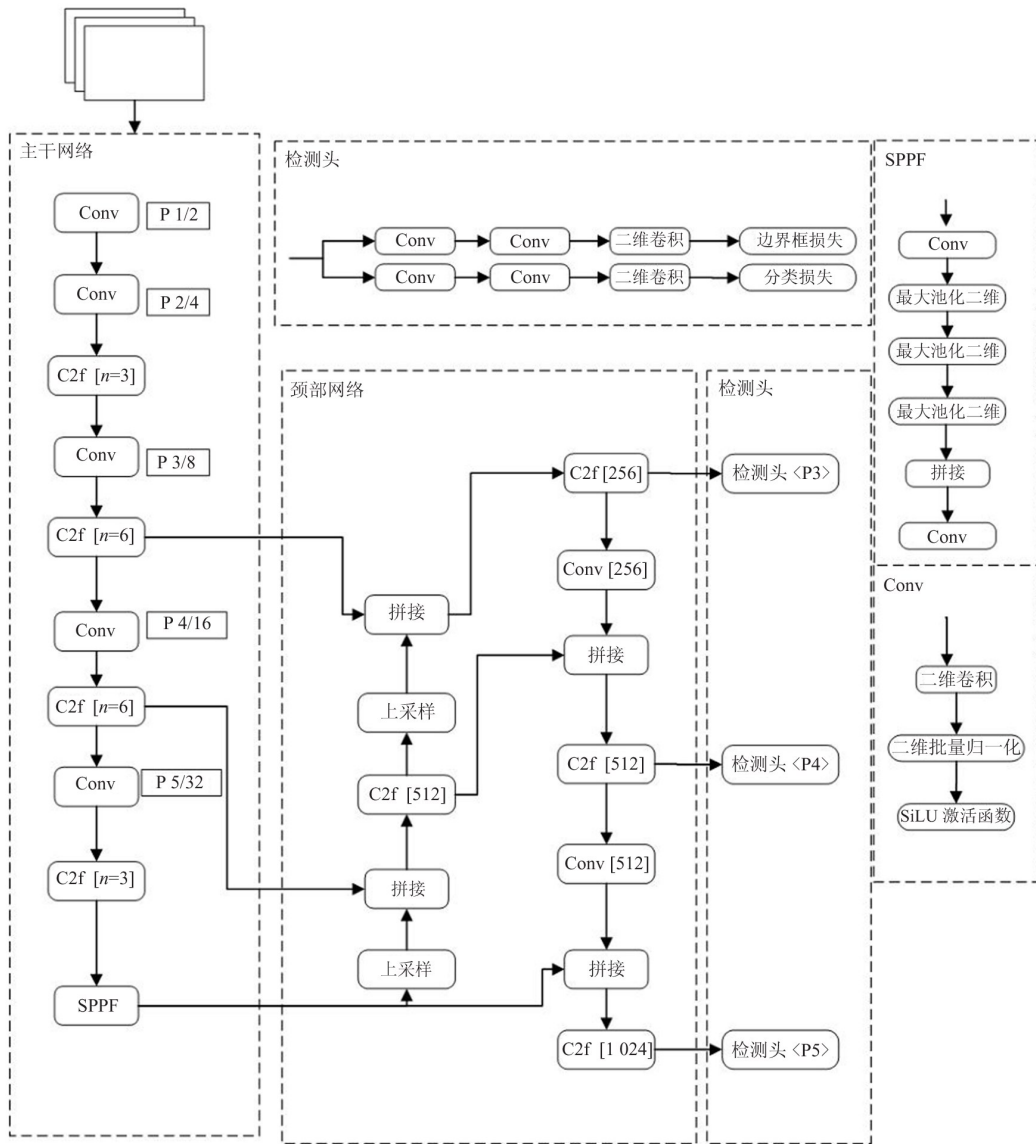


图 1 YOLOv8 网络框架图

## 2 YOLOv8-FRX 算法

针对无人机航拍图像中目标尺寸较小、尺度变化显著、背景复杂以及数据集分布不均的问题, 本文基于 YOLOv8n 进行了以下 3 方面的改进。(1) 在骨干网络中加入 RepCSP 模块, 使用 RepConv 替代传统的残差模块, 增强特征提取和梯度流通的能力; 通过有效的参数优化和特征融合, 优化了模型整体性能, 降低了复杂性, 在保证性能的同时提升了计算

效率。(2) 设计小目标增强金字塔, 不同于传统的方法(如直接增加 P2 检测层), 该方法使用重参数化可分离卷积对 P2 特征层进行处理, 提取出包含丰富小目标信息的特征, 然后与 P3 层融合。这种方法有效增强了对小目标的关注而不增加额外的计算负担。(3) 引入 Wise-IoU v3 作为损失函数, 通过 Wise-IoU v3 的梯度增益分配策略, 模型能够自动调整对不同样本的关注度, 从而增强了检测性能, 尤其是在处理低质量样本时表现更为出色。改进的网络结构如图 2 所示。

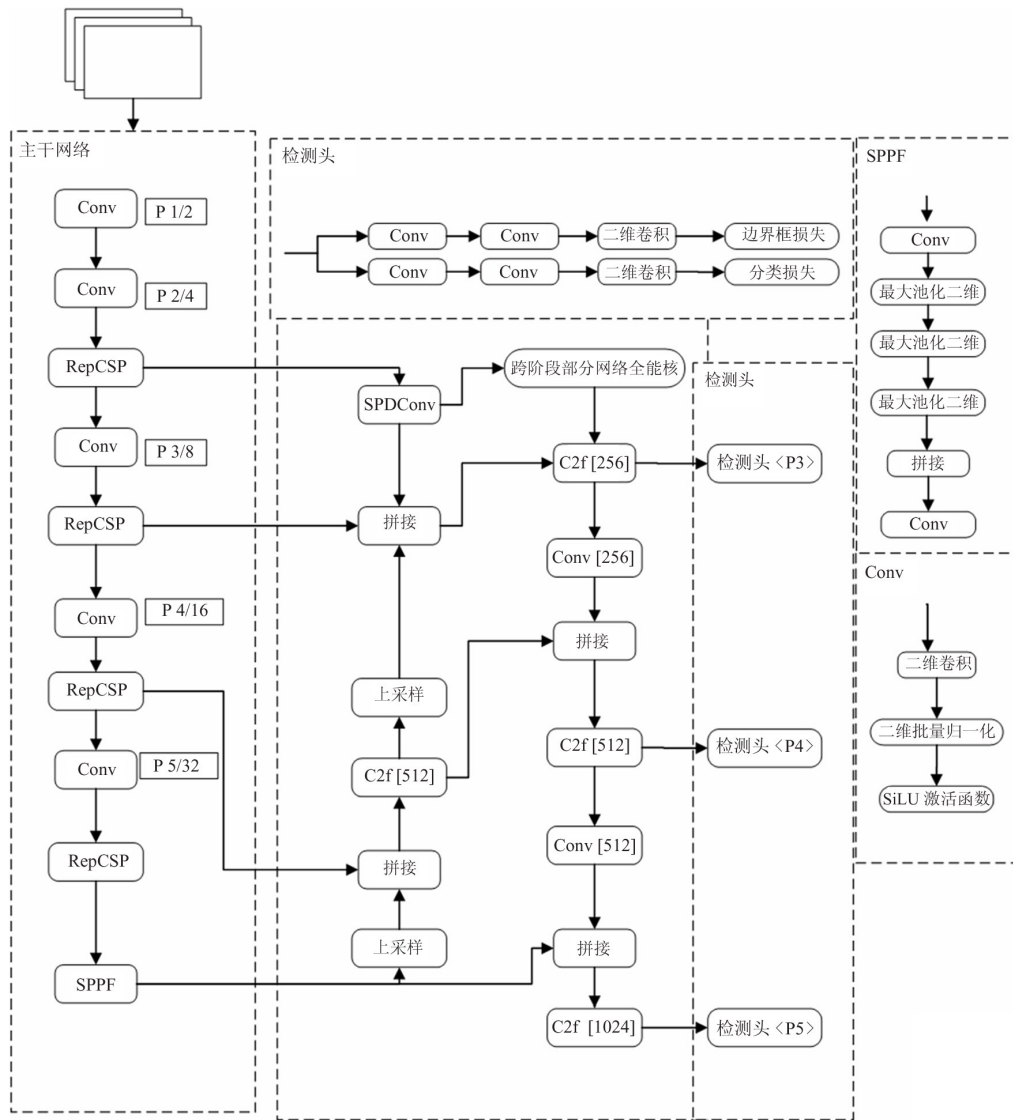


图 2 YOLOv8-FRX 网络结构图

## 2.1 RepCSP 模块

在 YOLOv8n 的骨干网络中的 C2f 结构处引入 RepCSP 模块,以替代部分卷积层,从而在减少模型参数数量的同时保持较高的检测性能。

RepCSP 模块通过在传统卷积结构中引入可学习的特征通道,增强了模型的特征能力。通过灵活的通道分组和分层特征提取,它能有效地适应不同尺度和形状的目标。该模块采用 RepConv 卷积操作,将一个大的卷积操作拆分为多个小的卷积操作,从而减少冗余计算,降低参数量。具体来说,该模块通过以下方式实现这一目标。

(1)特征通道增强:RepCSP 模块在传统卷积结构中引入可学习的特征通道,增强了模型的特征能

力。通过灵活的通道分组和分层特征提取,它能有效地适应不同尺度和形状的目标。

(2)RepConv 卷积操作:RepCSP 模块采用 RepConv 卷积操作,将一个大的卷积操作拆分为多个小的卷积操作,从而减少冗余计算,降低参数量。这种设计允许网络在推理阶段进行特征图的融合,增强了梯度流动和特征提取的能力,进一步提升效率和速度。

(3)梯度流通分支:在梯度流通分支上使用 RepConv,以此来增强特征提取和梯度流通的能力。

RepCSP 模块显著提升了计算效率和检测精度,尤其在复杂场景中的目标识别中表现优异。该模块的具体操作流程如图 3 所示。

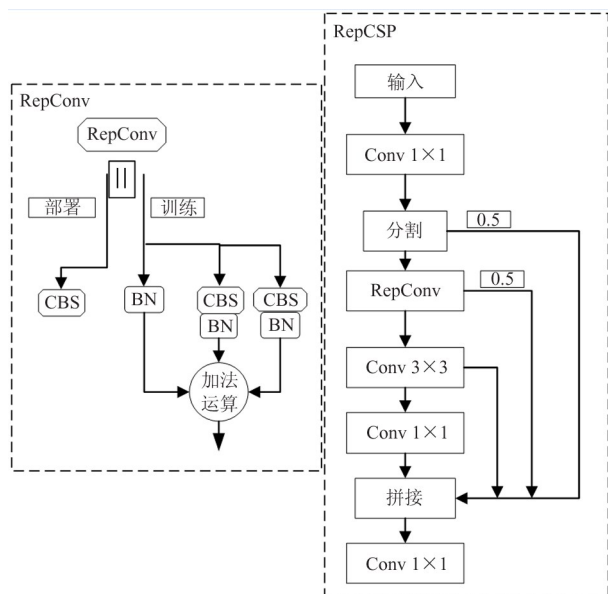


图3 RepCSP 模块和 RepConv 卷积操作示意图

RepCSP 模块在特征提取过程中,通过引入通道分组和层次化卷积操作,增强了对不同目标尺度和形状的适应能力。其输出特征值  $Y$  为

$$Y = \sum_{k=1}^K W_k \cdot F(P_k) \quad (1)$$

式中: $Y$  表示输出特征图,  $F(P_k)$  为特征图中像素位置,  $W_k$  为对应位置的卷积核权重,  $K$  为 RepCSP 模块将输入特征图划分的通道分组数。为了进一步提升模型对复杂目标的拟合能力,RepCSP 模块通过引用额外的隐含通道和动态权重调整,输出特征值  $Y'$  为

$$Y' = \sum_{k=1}^K (W_k + \Delta W_k) \cdot F(p_k + \Delta p_k) \quad (2)$$

式中: $\Delta W_k$  和  $\Delta p_k$  分别表示第  $k$  个位置的动态权重调整和可学习偏移量。该设计增强了模型在应对不同目标形状时的鲁棒性和泛化能力。

RepConv 卷积在推理阶段将多个卷积模块整合为一个,从而提升了模型计算的效率。卷积层  $Conv(x)$  的计算公式为

$$Conv(x) = W \cdot x + b \quad (3)$$

式中: $x$  是输入特征图, $W$  是卷积核权重, $b$  为偏置项。

批归一化(batch normalization, BN)通过计算小批量数据的均值和方差,将输入特征标准化为零均值和单位方差,从而加快收敛速度并提高模型稳定性。其计算公式为

$$BN(x) = \gamma \cdot \frac{x - mean}{\sqrt{var}} + \beta \quad (4)$$

式中: $\gamma$  是缩放因子,调整特征的分布; $\beta$  是偏移项,进一步调整特征; $mean$  和  $var$  分别是输入特征图的均值和方差。

## 2.2 SOEP 特征金字塔

SOEP 是一种专门设计用于提高小目标检测性能的特征金字塔结构。传统的目标检测方法通常在 P3、P4、P5 层上进行小目标的检测,但这些层对小目标的检测能力有限。为了解决这个问题,常见的做法是添加 P2 层以增强小目标检测能力,但这会导致计算量大幅增加和后处理时间延长等一系列问题。

SOEP 通过在原有 PAFPN 的基础上进行改进,提出了更高效的特征提取和融合策略。相较于 PAFPN 特征金字塔,SOEP 可以增强小目标检测能力,减少计算复杂度,并且它的特征整合效率高,拥有更好的特征流动性。它的核心思想是利用经过 SPDCConv 处理的 P2 特征层,提取富含小目标信息的特征,并将其与 P3 层进行融合。这一过程使用了跨阶段部分连接(cross stage partial, CSP)思想和 OmniKernel 模块,后者由全局、大分支和局部分支组成,有效地从全局到局部学习特征表征。SOEP 的特征金字塔架构图如图 4 所示。

以下是对 SOEP 特征金字塔公式各个部分的解释:

$$x \in R^{C \times H \times W} \quad (5)$$

这是 SOEP 的输入特征,其中  $C$  是通道数(特征维度), $H$  和  $W$  是特征图的高度和宽度。输入特征是经过前几层网络处理后的输出。

SPDCConv 模块通过重排输入特征来获取更丰富的小目标信息,并对其进行卷积操作。其中, $x'$  是将 4 个子图 ( $x_{00}, x_{10}, x_{01}, x_{11}$ ) 在通道维度上拼接后的结果, $y_{SPD}$  是经过 SPDCConv 模块最终输出的特征图, $ouc$  是输出通道数;式(8a)~(8d)分别对应于输入特征图的偶数行和偶数列、奇数行和偶数列、偶数行和奇数列以及奇数行和奇数列,其中, $\dots$ 代表选取所有通道, $:2$  指的是步长为 2。SPDCConv 模块的具体计算公式如下:

$$x' = Concat(x_{00}, x_{10}, x_{01}, x_{11}) \quad (6)$$

$$y_{SPD} = Conv(x', ouc) \quad (7)$$

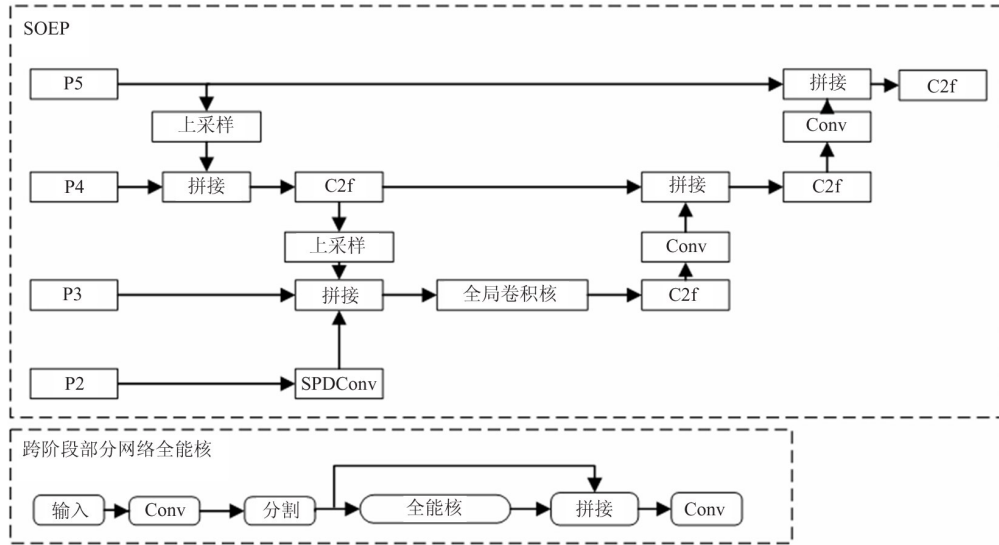


图4 SOEP 特征金字塔结构图

$$x_{00} = x[\dots, :, :2, :, :2] \quad (8a)$$

$$x_{10} = x[\dots, 1::2, :, :2] \quad (8b)$$

$$x_{01} = x[\dots, :, :2, 1::2] \quad (8c)$$

$$x_{11} = x[\dots, 1::2, 1::2] \quad (8d)$$

OmniKernel 模块通过特征通道注意力和空间通道注意力的结合,从全局和局部两个方面增强特征表达,提高了小目标检测的性能。首先,输入特征  $x$  通过初始卷积层  $in\_conv$  处理后,计算特征通道注意力;随后,对输入特征进行快速傅里叶变换 (fast Fourier transform, FFT),在频域中应用通道注意力权重后,再通过逆傅里叶变换 (inverse Fourier transform, IFT) 还原特征;然后,通过卷积和池化操作计算空间通道注意力,并将其应用于特征,得到新的表示。最后将原始输入、深度卷积输出与空间通道注意力的结果进行相加,得到经过整流线性单元 (rectified linear unit, ReLU) 激活后,得到最终输出  $y_{OK}$ 。

对 CSPOmniKernel 模块的输入分支处理时,首先输入特征  $x$  通过卷积层  $cv1$  提取特征。然后经过卷积后的特征被拆分为 2 个部分: $a$  为小目标特征,专注于提取小目标的特征; $b$  为残余特征,用于保留输入特征中未经过滤的重要信息。最后,通过卷积融合小目标特征和残差特征,得到最终输出  $y$ 。而 SOEP 输出是通过将 SPDConv、OmniKernel 和 CSPOmniKernel 的输出相加得到的,最大化利用小目标特征的信息。

CSPOmniKernel 模块计算如式(9)~(10)所示, SOEP 模块计算如式(11)所示。

$$a, b = Split(cv1(x)) \quad (9)$$

$$y = cv2(Concat(m(a), b)) \quad (10)$$

$$y_{SOEP} = y_{SPD} + y_{OK} + y \quad (11)$$

### 2.3 对于边界框损失函数的改进

YOLOv8 采用完全交并比 (complete intersection over union, CIoU) 作为边界框回归损失函数,然而 CIoU 在处理小目标时可能表现不佳,因为它主要关注的是框的重叠度和相对位置,而小目标可能无法提供足够的信息。在某些情况下,框的形状变化可能导致评估不准确,不能有效捕捉目标的真实几何特征。因此,选择 Wise-IoU v3 能够更好地处理目标的长宽比,提升对于各种形状目标的检测性能;Wise-IoU v3 能更有效地评估小目标的检测效果,改善检测精度;并且它的计算效率较高,可以更快地处理大规模数据集,提高训练和推理速度。

Wise-IoU v3 以 Wise-IoU v1 和 Wise-IoU v2 为基础额外增加了聚焦系数提出动态非单调聚焦机制,使预测框更加贴合数据集中的目标。Wise-IoU v1 将距离作为注意力的度量标准,公式如下所示:

$$L_{WIoUv1} = R_{WIoU} L_{IoU} \quad (12)$$

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \quad (13)$$

式中:  $R_{WIoU}$  是基于中心点距离的注意力权重系数,

$L_{IoU}$  是衡量预测框与真实框的重叠程度,取值范围为  $[0,1]$ ;  $(x_{gt}, y_{gt})$  为真实框的中心点坐标,  $(x, y)$  为预测框的中心点坐标,  $W_g, H_g$  分别为包围预测框和真实框的最小外接矩形的宽和高。  $R_{WIoU} \in [1, e)$ , 会放大普通质量的锚框的  $L_{IoU}$ ;  $L_{IoU} \in [0,1]$ , 会降低高质量锚框的  $R_{WIoU}$ , 并在锚框与目标框重合较好的情况下显著降低其对中心点距离的关注。

较低的离群度表明锚框质量较高, Wise-IoU v3 通过为这些锚框分配较小的梯度增益, 使边界框回归更专注于普通质量的锚框。对于离群度极高的极低质量锚框, Wise-IoU v3 同样会分配较低的梯度增益, 从而有效避免低质量示例产生过大的负面梯度。通过  $\beta$  构造非单调聚焦系数, 并将其应用于 Wise-IoU v1:

$$\beta = \frac{L_{IoU}^*}{L_{IoU}} \in [0, +\infty) \quad (14)$$

$$L_{WIoUv3} = rL_{WIoUv1}, r = \frac{\beta}{\delta\alpha^{\beta-\delta}} \quad (15)$$

当  $\beta = \delta$  时,  $r = 1$ 。当锚框的离群程度达到  $\beta = C$  ( $C$  为常数) 时, 锚框将获得最大梯度增益。 Wise-IoU v3 作为损失函数, 通过动态非单调机制评估锚框质量, 使模型更加关注普通质量的锚框, 进而提升目标定位能力。在无人机航拍场景中, 由于小物体的比例较大, 检测具有一定挑战性。 Wise-IoU v3 能够动态调整小物体的损失权重, 从而提高模型的检测性能。

### 3 实验设计与结果分析

#### 3.1 实验数据集

VisDrone 2019 数据集<sup>[23]</sup> 由天津大学机器学习与数据挖掘实验室收集。这些数据来自各类无人机摄像, 是在不同城市、环境、物体、密度、场景、天气和光照条件下使用不同型号的无人机平台收集的。 VisDrone 2019 数据集包含 10 209 张静态图像, 其中 6 471 张用于训练, 548 张用于验证, 3 190 张用于测试, 涵盖了 10 个目标类别。

无人机目标检测和跟踪数据集 (unmanned aerial vehicle detection and tracking, UAVDT) 是一个针对无人机场景中的目标检测和跟踪任务设计的大规模数据集。该数据集包含约 8 万帧图像, 主要针对

车辆目标, 每帧图像均带有手动标注的边界框, 标明了车辆类别以及遮挡情况。本文对 UAVDT 数据集进行处理, 将其转化为分辨率为  $640 \times 640$  像素的图片, 并将其分为训练集与验证集, 比例为 8:2。

#### 3.2 实验环境

本实验在操作系统 Windows 11 下进行, GPU 为 NVIDIA GeForce RTX 4070 Laptop, CPU 为 Intel(R) Core i9-13900HX, 16 GHz 运行内存。深度学习框架为 Pytorch 2.2.1, CUDA 12.1, Python 3.8.18。模型输入图像大小为  $640 \times 640$ , 优化器选用随机梯度下降 (stochastic gradient descent, SGD), 动量设置为 0.937, 训练周期为 300 轮, 批量大小设为 16, 其他训练参数采用官方推荐参数。

#### 3.3 评价指标

本文实验在检测性能和模型参数 2 个方面进行评价。实验指标包括精确率 ( $P$ ), 召回率 ( $R$ ), 平均精确率均值 (mAP), 参数量 ( $Para$ ) 和十亿次浮点运算 (Giga floating point operations per second, GFLOPs), 计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (16)$$

$$R = \frac{TP}{TP + FN} \quad (17)$$

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

式中: TP 表示正样本被识别正确的数量, FP 与 FN 分别表示正样本与负样本被识别错误的数量。以  $P, R$  为纵坐标、横坐标得到模型性能的 PR (Precision Recall) 曲线,  $AP_i$  表示第  $i$  个类别的 AP,  $c$  表示数据集中的标签数。本文使用 mAP50 与 mAP50:95 分别表示交并比 (intersection over union, IoU) 阈值取 0.5 时的 mAP 值和 IoU 阈值取 0.5~0.95 内等间隔选取 10 个 IoU 阈值时的 mAP 平均值。精确率 (precision,  $P$ ) 是用来预测在正样本的结果中正确的比例, 召回率 (recall,  $R$ ) 是用来衡量模型在真实为正的样本中被正确预测为正确的比例,  $P$  和  $R$  值越高证明模型的检测性能越好,  $Para$  是用来衡量模型的规模和空间复杂度的指标, GFLOPs 是用来对模型的计算复杂度和性能进行评估的指标, 表示一秒内

模型进行浮点运算的次数。

### 3.4 消融实验分析

为了进一步分析和验证 YOLOv8-FRX 算法模型在无人机图像检测性能上的有效性,本文设计了

一系列消融实验,以 YOLOv8n 作为基线模型,每个实验均未使用预训练权重,并在相同的实验环境和参数下,分别添加了 RepCSP 模块、小目标增强金字塔和 Wise-IoU v3 损失函数。实验结果如表 1 所示。

表 1 总体消融实验

	RepCSP	SOEP	Wise-IoU v3	P/%	R/%	mAP <sub>50</sub> /%	mAP <sub>50,95</sub> /%	Para	GFLOPs	FPS
1	×	×	×	43.9	32.5	32.6	18.9	$3.0 \times 10^6$	8.1	174.1
2	×	×	✓	43.4	33.1	32.9	19.2	$2.9 \times 10^6$	10.6	169.5
3	✓	×	×	40.3	31.2	30.1	17.0	$2.2 \times 10^6$	6.5	151.5
4	×	✓	×	47.0	35.0	35.7	21.0	$3.3 \times 10^6$	11.8	109.9
5	✓	✓	×	43.5	33.9	33.7	19.3	$3.0 \times 10^6$	8.1	156.3
6	✓	×	✓	42.2	31.9	31.5	17.9	$2.2 \times 10^6$	6.1	120.5
7	×	✓	✓	47.3	35.7	36.3	21.1	$2.9 \times 10^6$	10.6	121.9
8	✓	✓	✓	47.6	35.8	36.6	21.4	$2.9 \times 10^6$	10.6	125.0

消融实验表格系统地评估了 YOLOv8-FRX 算法中每个模块的影响,包括 RepCSP、SOEP 和 Wise-IoU v3。基线模型的 mAP50 为 32.6%,参数量(*Para*)为  $3.0 \times 10^6$ ,GFLOPs 为 8.1,作为对比参考。引入 Wise-IoU v3 模块提高了边界框回归精度,mAP50 提高到 32.9%,但参数量略微减少至  $2.9 \times 10^6$ 。RepCSP 模块显著减少了参数量和 GFLOPs,分别降至  $2.2 \times 10^6$  和 6.5;但性能有所下降,显示了效率和检测能力之间的权衡。SOEP 模块增强了小目标的特征捕捉能力,其 mAP50 提高到 35.7%,尽管计算成本较高且每秒帧数(frames per second,FPS)有所下降。结合 RepCSP 和 SOEP 实现了效率与精度的平衡,mAP50 为 33.7%。SOEP 和 Wise-IoU v3 的协同作用进一步提高了检测精度,mAP50 达到了 36.3%,突出了它们的互补优势。三者——RepCSP、SOEP 和 Wise-IoU v3 的完全整合实现了最佳性能,mAP50 为 36.6%,mAP50 : 95 为 21.4%,并在参数、计算复杂度和速度之间取得了最佳的平衡。这些结果验证了所提设计的有效性,并证明其在解决小目标检测挑战方面的优势。

### 3.5 热力图对比实验

为了更直观地体现本文算法的有效性,绘制出 2 种算法的 3 个场景热力图,如图 5(左边为本文算法,右边为 YOLOv8n)所示。由图 5 可知,YOLOv8n

更关注近处物体,本文算法对于远处小目标的关注度更高、更准确。

### 3.6 损失函数对比实验

YOLOv8 默认使用完全交并比(complete intersection over union,CIoU)损失函数。为了确定改进的损失函数,分别使用欧几里得交并比(Euclidean intersection over union,EIoU)、分离交并比(separation intersection over union,SIoU)、距离交并比(distance intersection over union,DIoU)以及 Wise-IoU v3 进行对比,实验结果如表 2 所示。

从表 2 可以看出,采用 EIoU、SIoU、DIoU 和 Wise-IoU v3 替代 CIoU 均能在不同程度上提升系统性能。尤其是 Wise-IoU v3,在各项指标上表现优于其他损失函数,因此选择它来改进 YOLOv8n。

表 2 不同损失函数的对比实验分析

损失函数	P	R	mAP <sub>50</sub> /%	mAP <sub>50,95</sub> /%
CIoU	43.9	32.5	32.6	18.9
EIoU	41.4	32.1	31.5	18.1
SIoU	43.8	32.8	33.1	19.2
DIoU	43.1	32.5	32.2	18.4
Wise-IoU v3	43.5	33.9	33.7	19.3

### 3.7 对比实验设计

为进一步验证本文算法在无人机航拍图像上的

有效性,将其与当前主流的优秀目标检测算法进行对比,结果如表 3 所示。

从表 3 可以看出, YOLOv8-FRX 集成了 RepCSP 模块、小目标增强金字塔(SOEP)和 Wise-IoUv3,有



图 5 YOLOv8n 和改进算法的热力图

表 3 对比实验

算法	参数量	mAP <sub>50</sub> /%	GFLOPs
YOLOv8n	$3.0 \times 10^6$	32.6	8.1
YOLOv8s	$11.1 \times 10^6$	37.9	28.5
YOLOv8n + BiFPN	$2.0 \times 10^6$	33.1	7.1
YOLOv8n + goldYOLO	$6.0 \times 10^6$	34.1	10.3
YOLOv8n + SE	$3.0 \times 10^6$	32.9	10.6
YOLOv8n + SPPELAN	$3.5 \times 10^6$	33.1	8.5
YOLOv8n + C2f-SCConv + EMA	$2.2 \times 10^6$	33.2	6.3
YOLOv8n + convnextv2	$5.7 \times 10^6$	33.9	14.1
YOLOv8n + Focal-Modulation	$3.1 \times 10^6$	33.1	8.2
YOLOv8n + CBAM	$3.1 \times 10^6$	32.6	8.1

(表 3 续)

YOLOv7-tiny	$6.0 \times 10^6$	35.1	13.3
YOLOv10l	$25.7 \times 10^6$	37.2	126.4
YOLOv9-tiny	$3.0 \times 10^6$	31.6	10.7
F-RCNN	$41.2 \times 10^6$	32.4	206.7
YOLOv8n + GhostHGNetV2	$2.3 \times 10^6$	31.1	6.9
本文算法	$2.9 \times 10^6$	36.6	10.6

注: BiFPN (bi-directional feature pyramid network, 双向特征金字塔网络); SE (squeeze-and-excitation, 压缩与激励网络); C2f-SCConv + EMA (cross-to-fine spatial and channel reconstruction convolution + efficient multi-dimensional attention, 交叉到精细-空间和通道重建卷积-高效多尺度注意力); SPPELAN (spatial pyramid pooling enhanced layer aggregation network, 空间金字塔多尺度特征融合); CBAM (convolutional block attention module, 卷积块注意力模块); FRCNN (faster region-based convolutional neural network, 更快区域卷积神经网络)

效地解决了无人机图像中小目标检测的挑战。根据实验结果,该方法相比其他方法具有明显优势。

首先, YOLOv8-FRX 在参数减少方面表现突出。与其他轻量级模型相比,如 YOLOv8n + Focal-Modulation (参数量为  $3.1 \times 10^6$ , mAP50 为 32.6%) 和 YOLOv8nC2f-SCConv + EMA (参数量为  $2.2 \times 10^6$ , mAP50 为 33.2%), YOLOv8-FRX 仅使用  $2.9 \times 10^6$  参数量, mAP50 达到了更高的 36.6%。这一提升得益于 RepCSP 模块的引入,该模块在减少参数量的同时保持了强大的特征提取能力,使得 YOLOv8-FRX 非常适合计算受限的无人机应用。

其次, YOLOv8-FRX 在检测精度上优于基线模型 YOLOv8n,其 mAP50 提高了 4.0%。它还超越了其他方法,如 YOLOv8n + SE (32.9%) 和 YOLOv8n + CBAM (32.6%)。这一提升主要归因于 SOEP 模块,专门针对小目标检测设计。通过集成多尺度特征融合,SOEP 显著提高了模型捕捉小目标的能力。相比之下,像 CBAM 和 SE 这样的通用注意力机制虽然提供了一定的改进,但在密集和复杂的小目标场景下其效果较差。

在计算效率方面, YOLOv8-FRX 的 GFLOPs 为 10.6,与 YOLOv8n + SE (10.6) 和 YOLOv8n + Focal-Modulation (8.2) 相当,显示出在效率和性能之间的优秀平衡。SOEP 模块在增强小目标检测能力的同时没有引入过多的计算开销,使得 YOLOv8-FRX 在实时无人机监控任务中具有高度的实用性。

此外, YOLOv8-FRX 的泛化能力得益于 Wise-IoU v3 的引入,该方法采用了梯度增益分配策略来优化边界框回归。该创新增强了检测形状不规则或密集分布的小目标的鲁棒性。相比之下,像 YOLOv8n + GhostHGNetV2 这样极轻量化的模型,过度强调参数减少,可能牺牲了模型的泛化性能。

总体而言, YOLOv8-FRX 在参数效率和检测精度之间达成了出色的平衡。借助专为小目标检测优化的模块和创新的 IoU 回归策略,该方法在性能上超越了其他算法。它不仅显著提高了检测精度,还保持了优异的计算效率和强大的泛化能力,为无人机场景中的小目标检测提供了一种高效的解决方案。

### 3.8 泛化实验

将本文算法与 YOLOv8n 在 UAVDT 数据集上

进行对比实验,以证明本文算法的通用性。实验结果如表 4 所示。

表 4 泛化实验

算法	<i>P</i>	<i>R</i>	mAP <sub>50</sub>	mAP <sub>50,95</sub>
YOLOv8n	40.8	28.6	31.9	19.0
本文方法	42.5	34.1	33.6	19.7

由表 4 可知,本文算法在 UAVDT 数据集上的精度指标 *P*、mAP<sub>50</sub> 和 mAP<sub>50,95</sub> 提升了 1.7%、1.7% 和 0.7%,有效证明了本文算法有较好的泛化性能,具有良好的鲁棒性。

### 3.9 可视化结果分析

为更直观地评估本文改进算法的有效性,在 VisDrone 2019 的测试结果中,分别选取了遮挡、高空、密集和夜间场景的典型图片进行可视化对比,结果如图 6 所示。

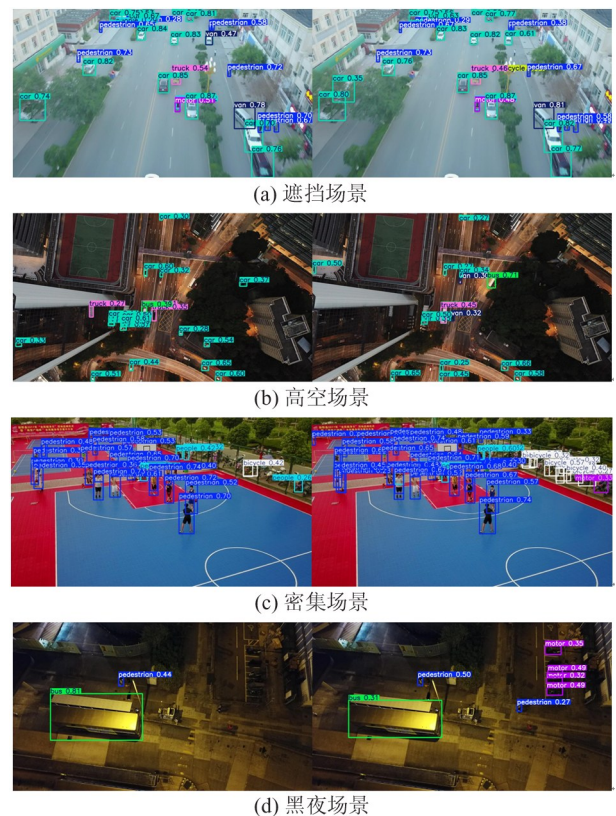


图 6 检测结果对比

图 6 左列小图为 YOLOv8n 的检测结果;右侧小图为本文算法的检测结果。图 6(a) 为遮挡场景,路边的树木挡住了部分目标,使目标边界变得模糊不

清,但是改进后的算法适应能力较强,可以准确检测出被树叶遮挡住的目标;图 6(b)为高空场景,在高空场景中,原模型出现错检漏检的情况,而本文算法不仅能成功识别出原模型所漏检的汽车,还能纠正原模型的误检目标;图 6(c)为密集场景,两者对比,很容易看出本算法比原模型能更精确地识别出右侧停放的自行车与摩托车;图 6(d)为黑夜场景,本文算法可以检测出原模型漏检的行人与路边停放的摩托车。综上所述,本文算法具有良好的检测性能。

## 4 结 论

航拍图像通常存在目标尺寸小、背景复杂、目标类别多且尺度变化大的问题,传统目标检测算法在这些情况下表现不佳。因此,本文提出了一种改进的 YOLOv8n 算法——YOLOv8-FRX,用于无人机小目标的检测。改进后的算法包括 4 个步骤:首先,舍弃 YOLOv8 中常用的残差模块,在梯度流通分支上使用 RepConv,以此来增强特征提取和梯度流通的能力;然后,在骨干网络中加入 RepCSP 模块,减少模型参数量的同时保持较高的检测性能;接着,设计小目标增强金字塔(SOEP),使用 P2 特征层经过 SPDCConv 得到富含小目标信息的特征给到 P3 进行融合,在不增加 P2 检测层的情况下,增强了小目标特征的捕获,同时保持计算效率;最后,采用 Wise-IoU v3 作为损失函数,引入梯度增益分配策略,帮助模型更好地处理低质量样本,提高边界框回归精度。通过消融实验验证了改进策略的有效性,并与其他模型进行了对比分析。实验结果表明,YOLOv8-FRX 算法模型对比 YOLOv8n 算法在 VisDrone 2019 数据集上的 mAP50 值提升了 4.0%,参数量减少了 3.6%。在 UAVDT 数据集上进行泛化实验,mAP50 值提升了 1.7%,以此验证了改进后算法的通用性。

尽管 YOLOv8-FRX 算法在 VisDrone 2019 和 UAVDT 数据集上取得了显著的性能提升,但在更为复杂的场景和更高分辨率的图像下,仍存在进一步优化的潜力。具体来说,算法在处理极小目标和遮挡目标时的检测精度仍有提升空间。此外,模型的计算量和实时检测效率在某些情况下仍需改进,特别是在资源受限的设备上进行高效部署时,模型的轻

量化和实时性需要进一步优化。

未来的研究将聚焦于进一步减少模型的计算量,提高实时检测的效率,引入自适应特征融合和动态注意力机制等先进模块,增强算法在极小目标和遮挡目标检测中的能力。通过进一步优化模型结构,减少冗余计算,降低模型的参数量和计算复杂度。并探索模型在资源受限设备上的高效部署方法,如模型压缩、量化和剪枝等技术,提高模型的实时性和应用范围。通过上述改进,YOLOv8-FRX 算法有望在无人机小目标检测领域取得更大的突破,为智能交通、环境监测等领域的实际应用提供更强大的技术支持。

## 参考文献

- [ 1 ] 宋晓茹,刘康,高嵩,等. 复杂战场环境下改进 YOLOv5 军事目标识别算法研究[J]. 兵工学报, 2024, 45(3):934-947.
- [ 2 ] Mittal P, Singh R, Sharma A. Deep learning based object detection in low-altitude UAV datasets: a survey[J]. Image and Vision Computing, 2020, 104:104046.
- [ 3 ] 杨锦辉,李鸿,杜芸彦,等. 基于改进 YOLOv5s 的轻量化目标检测算法[J]. 光电与控制, 2023, 30(2): 24-30.
- [ 4 ] 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. Las Vegas, USA: IEEE, 2016:779-788.
- [ 5 ] 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.
- [ 6 ] 闫建红,冉同霄. 基于 YOLOv8 的轻量化无人机图像目标检测算法[J]. 图学学报, 2024, 45(6):1328-1337.
- [ 7 ] 刘谱,张兴会,张志利,等. 从 RCNN 到 YOLO 的目标检测综述[C] // 第 16 届全国信号和智能信息处理与应用学术会议论文集. 北京:中国高科技产业化研究会, 2022:16-23.
- [ 8 ] 陈金林,吴一全,苑玉彬. 无人机视角下目标检测的 YOLO 系列算法研究进展[J]. 北京航空航天大学学报, 2024, 50(5):1-33.
- [ 9 ] Xie C H, Wu J M, Xu H Y. Improved small target detection algorithm for UAV imagery with YOLOv5[J]. Computer Engineering and Applications, 2023, 59(9):198-206.
- [ 10 ] Zhang R M, Xiao Y F, Jia Z N, et al. Improved YOLOv7 algorithm for target detection in complex environment from UAV viewpoint[J]. Optoelectronic Engineering, 2024, 51(5):92-103.
- [ 11 ] Li Y C, Shi W Y, Feng C. A lightweight YOLOv8 detec-

- tion algorithm for small target detection in UAV aerial photography [J]. *Computer Engineering and Applications*, 2024,60(17):167–178.
- [12] Zhu E W, Liang J, Xiao J W, et al. Research on small target detection problem based on MDS-YOLO model [J]. *Journal of Hunan University (Natural Science Edition)*, 2024,51(12):78–86.
- [13] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector [C] // *The 14th European Conference on Computer Vision*. Amsterdam, Netherlands: Springer, 2016:21–37.
- [14] Girshick I R. Fast R-CNN [C] // *Proceedings of the IEEE International Conference on Computer Vision*. Santiago, Chile: IEEE, 2015:1440–1448.
- [15] Lyu Z, Jin H, Zhen T, et al. Small object recognition algorithm of grain pests based on SSD feature fusion [J]. *IEEE Access*, 2021,9:43202–43213.
- [16] Cao S, Wang T, Li T, et al. UAV small target detection algorithm based on an improved YOLOv5s model [J]. *Journal of Visual Communication and Image Representation*, 2023,97:103936.
- [17] Zhang Z. Drone-YOLO: an efficient neural network method for target detection in drone images [J]. *Drones*, 2023,7(8):526.
- [18] Cui Y, Ren W, Knoll A. Omni-kernel network for image restoration [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024,8(2):1426–1434.
- [19] Han K, Wang Y, Tian Q J, et al. GhostNet: more features from cheap operations [C] // *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Washington, USA: IEEE, 2020:1577–1586.
- [20] Wu Y Q, Tong K. Research progress on small target detection in UAV aerial images based on deep learning [EB/OL]. [2025–01–02]. <https://hkxb.buaa.edu.cn/CN/10.7527/S1000-6893.2024.30848>
- [21] Li Y C, Shi W Y, Feng C. Lightweight YOLOv8 detection algorithm for small target detection in UAV aerial photography [J]. *Computer Engineering and Applications*, 2024,60(17):167–178.
- [22] Zhu F Y, Peng D. YOLO-SP: a local multiscale target detection network for industrial safety [EB/OL]. (2025–01–02) [2025–03–04]. <https://ieeexplore.ieee.org/document/10637416>.
- [23] Du D W, Zhu P F, Wen L Y, et al. VisDrone-DET 2019: the vision meets drone object detection in image challenge results [C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop*. Seoul, Korea: IEEE, 2019:213–226.

## Research on an improved YOLOv8-FRX algorithm for UAV small object detection

Lin Bo, Huang Hongqiong

(School of Information Engineering, Shanghai Maritime University, Shanghai 201306)

### Abstract

To address the issues of dense small object distribution, complex backgrounds, and high false detection and missed detection rates of small objects in traditional YOLO (you only look once) algorithms for drone-captured images, an improved object detection algorithm, YOLOv8-FRX (feature refinement and eXtended detection), is proposed. This method enhances model performance through three main improvements: first, introducing reparameterised cross stage partial network (RepCSP) in the backbone network and using reparameterized convolution (RepConv) on the gradient circulation branch to enhance the feature extraction capability while reducing the number of parameters; second, designing a small object enhanced pyramid (SOEP) to improve the efficiency of small object feature capture while maintaining computational efficiency; and third, adopting the Wise-IoU v3 (wise intersection over union v3) strategy to optimize gradient gain allocation and enhance the accuracy of bounding box regression. Experimental results on the VisDrone dataset demonstrate that the proposed method improves the mAP50 by 4.0% compared to the baseline model while reducing parameters by 3.6%. In generalization experiments on the UAVDT (unmanned aerial vehicle detection and tracking) dataset, detection accuracy increased by 1.7%, validating the method's generality and effectiveness. These improvements not only enhance the detection capability for small objects but also balance accuracy and efficiency, providing a superior solution for object detection in drone-captured images.

**Key words:** unmanned aerial vehicle, small object detection, reparameterized cross stage partial network, small object enhanced pyramid, wise intersection over union