

DOI:10.19651/j.cnki.emt.2417692

# 基于 LDF-YOLO 的小目标检测方法<sup>\*</sup>

刘洋<sup>1</sup> 任旭虎<sup>1</sup> 刘宝弟<sup>2</sup> 刘伟峰<sup>2</sup>

(1. 中国石油大学华东海洋与空间信息学院 青岛 266580; 2. 中国石油大学华东控制科学与工程学院 青岛 266580)

**摘要:** 小目标检测是计算机视觉中极具挑战性的任务,现有的检测算法复杂度高、计算量大且检测精度低导致了漏检和误检的问题,本文针对小目标的独有特征提出了 LDF-YOLO 算法以提高检测精度并降低漏检率。首先是对 Head 部分的改进,在特征融合网络中引入了特征转换模块,设计了针对微小物体的检测头 LP-Detect;其次,借鉴残差门控机制和局部特征增强机制设计了 LR-C2f 模块,增强模型提取局部特征的能力;最后,融入了局部特征增强模块,以强化骨干网络提取小目标信息的能力。在公开数据集 Tiny Person 上,LDF-YOLO 比原 YOLOv8 在 mAP0.5 上提高了 4.5%,召回率提高了 5.5%,实验结果验证了改进方法的有效性,同时在 NWPU VHR-10 和 VisDrone2019 数据集上做了泛化对比实验,经实验表明各项指标均有提升。

**关键词:** 小目标检测;YOLOv8;残差门控机制;特征转换;特征融合

中图分类号: TP391.4; TN911.73 文献标识码: A 国家标准学科分类代码: 520.6

## Small object detection method based on LDF-YOLO

Liu Yang<sup>1</sup> Ren Xuhu<sup>1</sup> Liu Baodi<sup>2</sup> Liu Weifeng<sup>2</sup>

(1. College of Ocean and Space Information, China University of Petroleum East China, Qingdao 266580, China;

2. College of Control Science and Engineering, China University of Petroleum East China, Qingdao 266580, China)

**Abstract:** Small target detection is an extremely challenging task in computer vision, where existing detection algorithms suffer from high complexity, large computational overhead, and low detection accuracy, leading to issues such as missed detections and false alarms. In this paper, the LDF-YOLO algorithm is proposed to enhance detection accuracy and decrease missed detection rates for small objects. Firstly, improvements are made to the Head section by introducing a feature transformation module in the feature fusion network and designing the LP-Detect detection head tailored for small objects. Secondly, drawing inspiration from residual gated mechanisms and local feature enhancement strategies, the LR-C2f module is devised to bolster the model's capability in extracting local features. Finally, the local feature enhancement module is integrated to enhance backbone's ability to extract information from small objects. On the publicly available Tiny Person dataset, LDF-YOLO outperforms the original YOLOv8 by achieving a 4.5% improvement in mAP0.5 and a 5.5% increase in recall. Experimental results validate the effectiveness of our proposed improvements. Furthermore, generalization comparison experiments on the NWPU VHR-10 and VisDrone2019 datasets demonstrate improvements across all metrics.

**Keywords:** small object detection; YOLOv8; residual gated mechanisms; feature transformation; feature fusion

## 0 引言

在目标检测方向,小目标检测是自动驾驶、无人机侦察和视频监控等领域的关键问题。研究小目标检测算法不仅满足行业需求,且能够推动视觉技术创新,具有实用价值和科研意义。

近年来人们对这一领域进行了大量的研究,实现小目

标检测的模型主要有两类,一类是基于 Transformer<sup>[1]</sup>的,文献[2]提出了一种自上而下的方法,逐部分检测目标信息,随后聚类相联目标点。相较于直接预测目标信息,此方法召回率更高,但模型效率较低;文献[3]提出了细化粗糙特征思想,在 DETR(detection transformer)基础上引入局部信息到全局上下文信息流的方法,精细化目标特征。然而,随着解码器层数增加,导致模型推理速度显著减慢;文

收稿日期:2024-12-21

\*基金项目:国家自然科学基金(62372468)、山东省重点基础研究项目(ZR2023ZD32)、山东省自然科学基金(ZR2023MF008)项目资助

献[4]设计了一种前景增强注意力框架,将上下文信息集中到 Swin Transformer 的原始骨干网络中,并通过跳跃连接操作提出了改进的加权双向特征金字塔网络。另一类是基于 CNN<sup>[5]</sup>的,文献[6]在 YOLOv5s 基础上提出引入了通道注意力机制并使用 $\partial$ -EIOU 替换 GIOU 损失函数,在复杂环境中取得了较好的检测结果;文献[7]设计了跨阶段特征提取模块和跨梯度空间金字塔池化模块,在增强特征提取能力的同时有效融合多尺度细节信息,有效降低误检率,但多次跨层连接导致模型计算效率较低。文献[8]在 YOLOv8n 基础上提出引入了 CBAM(convolutional block attention module)注意力机制并改进了原 C2f 模块以更好地捕捉小目标的细节特征,但是检测精度提升不明显;雷帮军等<sup>[9]</sup>从高效多尺度注意力机制汲取灵感,设计了新的特征提取模块,并在下采样阶段引入无参注意力机制构建了多尺度网络结构,有效提升了模型的特征提取能力,但是对特征不明显的目标仍然存在漏检的情况。

尽管这些方法在一定程度上改善了模型的检测性能,但是改进方法多偏向于引入注意力机制,针对小目标的细节信息提取不充分,在检测分辨率较小的目标时仍然存在漏检或者误检的问题。因此,本文根据前人的改进方法,从骨干网络提取特征时局部信息逐渐丢失和模型输出之前全局信息与局部信息高度综合这两个问题入手,改进模型结构以改善漏检或误检的问题。

本文贡献如下:1)将局部特征增强模块(convolutional feed-forward network,ConvFFN)与骨干网络中的普通卷积相结合,提高骨干网络聚合局部信息的能力;2)设计了使得模型能够学习更深层次和更丰富的特征表示的 LR-C2f 模块;3)改进特征融合网络,并设计了融入特征转换模块的小物体检测层 LP-Detect,扩大输出特征图的感受野以提升模型对小目标的关注度。

## 1 YOLOv8 算法

YOLOv8 算法是目前应用范围最广的目标检测算法之一,基本的组成部分包括 Backbone、Neck 和 Head。其中 Backbone 部分由 CBS、C2f 和 SPPF 模块堆叠而成,C2f 模块的设计参考了 YOLOv5 的 C3 模块和 YOLOv7<sup>[10]</sup>的高效层聚合网络 (efficient layer aggregation networks, ELAN)。Neck 部分的设计思想主要来源于路径聚合结构<sup>[11]</sup>(path aggregation net, PAN),它采用了横向连接与自底向上的信息传递方式,用于融合不同尺度特征图的信息。Head 部分采用了解耦策略,分别计算分类损失与定位损失。

## 2 LDF-YOLO 算法

LDF-YOLO 算法是在 YOLOv8 的基础上针对小目标检测改进的,首先,将能够聚合局部信息的 ConvFFN 融入到骨干网络中的普通下采样操作;然后借鉴残差门控机制

和局部特征增强机制设计了 LR-C2f 模块,提升模型对目标信息的提取能力;最后引入用于从综合信息中捕获局部信息的特征转换模块(layer-wise contextual block hierarchical attention module,LCBHAM)代替检测层之前的常规卷积,并改进特征融合网络,融入针对小物体的检测头。

模型的输入为  $x \in \mathbb{R}^{H \times W \times C_I}$ , 骨干网络中每一部分的输出  $y_{B_i}$  为:

$$y_{B_i} = DCL(y_{B_i}), i = 0, 1, 2, 3, 4 \quad (1)$$

其中,DCL 为常规卷积、ConvFFN 与 LR-C2f 的顺序组合,且  $y_{B_0} = CBS(x)$ ; 特征融合网络中每一部分的输出  $y_{N_h}$  为:

$$y_{N_h} = CL[u(y_{N_{h-1}}), y_{B_{1-h}}], h = 0, 1, 2, 3 \quad (2)$$

其中,CL 表示 Contact 与 LR-C2f 的组合, $u$  表示上采样操作,且  $y_{N_0} = SPPF(y_{B_4})$ ; Head 中 3 个模块的输出  $y_{H_l}$  为:

$$y_{H_l} = \begin{cases} y_{N_3}, l = 0 \\ CL[y_{H_0}, y_{N_2}], l = 1 \\ CL[CBS(y_{H_1}), y_{N_1}], l = 2 \end{cases} \quad (3)$$

最终的输出  $y_{(4,8,16)} \in \mathbb{R}^{H/(4,8,16) \times W/(4,8,16) \times C_O}$  由  $y_{H_l}$  经过 LP-Detect 后得到。LDF-YOLO 模型结构如图 1 所示。

### 2.1 设计新的 PAN 结构与检测层

#### 1) 改进检测头

YOLOv8 沿用了早期 YOLO 版本的检测头设计策略,网络模型最终输出的 3 个特征图分别为输入图像下采样 8 倍、16 倍、32 倍,分别用于检测小、中、大目标。下采样倍数小的特征图感受野较小,能够保留图像更多的局部特征信息,即小目标的信息。基于此,本文在 Head 部分去掉了下采样 32 倍的检测头,额外增加了下采样 4 倍的检测头。本文采取的方法是将 Backbone 的第一个 C2f 模块与扩展后的 PAN 结构进行横向连接,然后连接到检测头,扩展后的 Neck 部分能够融合到更多局部特征信息<sup>[12-13]</sup>。此外,本文进一步引入了部分卷积 (PConv) 和轻量化卷积 (LightConv) 替换检测头中的普通卷积,设计了 LP-Detect 模块。PConv 的输入只有其中一部分进行普通卷积,将卷积后的特征图与另一部分特征图拼接得到 PConv 的输出; LightConv 由深度卷积与普通卷积组成,深度卷积用于提取更完整的特征信息,普通卷积用于得到指定通道数的特征图。PConv 与 LightConv 的结合在增强模型对目标特征的表达能力的同时保证不会降低模型的计算效率。改进后的检测头如图 2 所示。

#### 2) LCBHAM 模块

YOLOv8 中的 PAN 结构使用常规的卷积实现下采样后与特征金字塔结构(feature pyramid networks,FPN)实现跨层链接,特征图进入检测头模块之前,全局信息与局部信息已经高度综合,常规卷积无法捕获重要的局部信息并抑制不重要的特征信息,因此引入 LCBHAM 模块<sup>[14]</sup>以解决此问题。

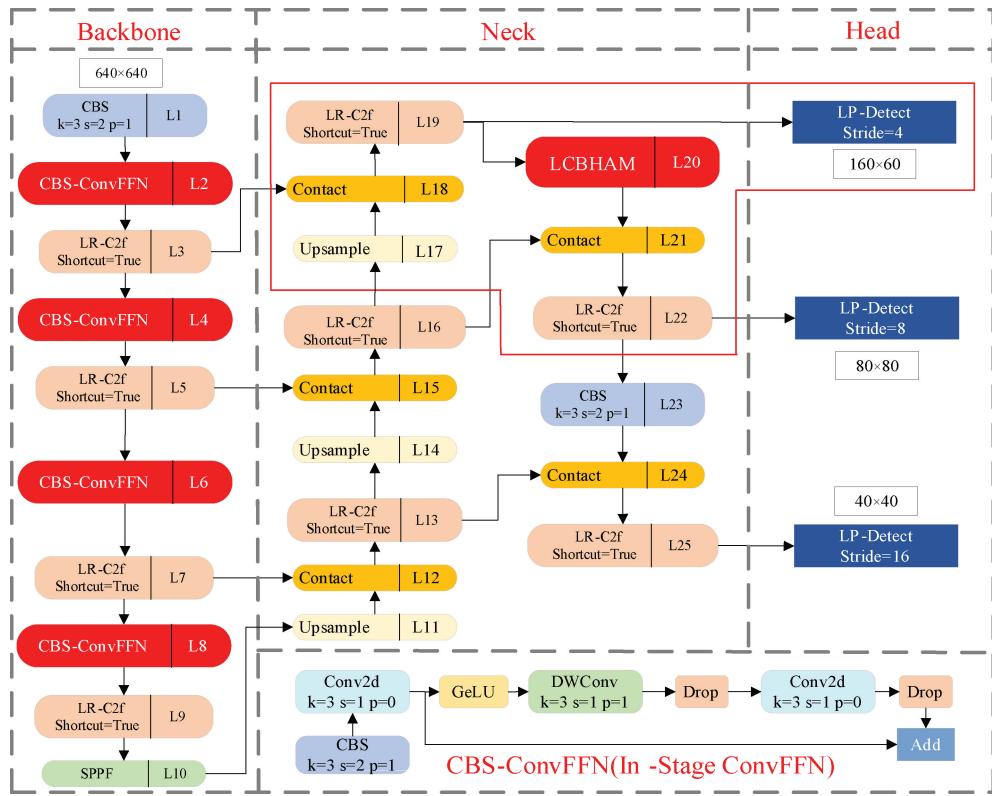


图 1 LDF-YOLO 模型结构

Fig. 1 LDF-YOLOv8 model structure

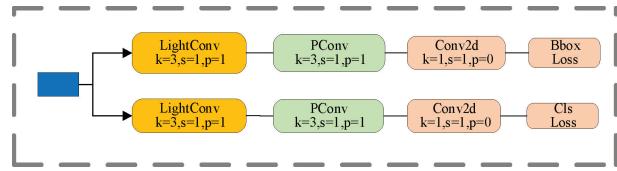


图 2 LP-Detect

Fig. 2 LP-Detect

LCBHAM 模块如图 3 所示, 特征图首先经过一个下采样卷积, 压缩特征图并增加感受野, 同时提取局部特征; 然后在归一化后的特征图上使用具有更高的计算效率 HardSwish 激活函数; 随后特征图与分层上下文注意力模块 (layer-wise contextual attention module, LCAM) 的输出相乘, 其中 LCAM 模块分别使用全局最大池化和全局平均池化将每个特征图转化为  $1 \times 1$  的向量, 将得到两种信道的全局统计特征输入到一个共享卷积中, 将通道数降为输入通道的  $1/16$ , 经过 Relu 激活函数后再通过一个共享卷积将通道数增加到原始通道, 最后在通道维数上将两个分支的特征图相加, 通过 Sigmoid 激活函数得到表示通道权重的特征映射。LCAM 模块主要参考 CBAM 注意力机制的通道域路径设计, 与后者中的全连接层相比, 共享卷积不仅可以自适应调整通道的重要性并捕获通道之间的特征关系, 而且参数数量少、计算复杂度低; 然后特征图输入到 LD-SAM 模块中, 局部空间注意力模块 (local domain

spatial attention module, LD-SAM) 设计思想来源于的 CBAM 注意力机制的空间域路径, 首先将输入特征图分别在通道维度上进行全局最大池化和全局平均池化, 得到两个二维的特征图, 然后将两个特征图在通道维度上进行拼接并输入到卷积核大小为  $3 \times 3$  的卷积中, 得到一个输出通道为 1 的二维向量, 然后通过 Sigmoid 激活函数得到表示空间权重的特征映射。与使用  $7 \times 7$  卷积核的 CBAM 相比, 使用  $3 \times 3$  卷积核更有利于小目标的特征信息捕获并减少计算量; 最后 LD-SAM 模块的输出与该模块的输入相乘得到最终的输出。

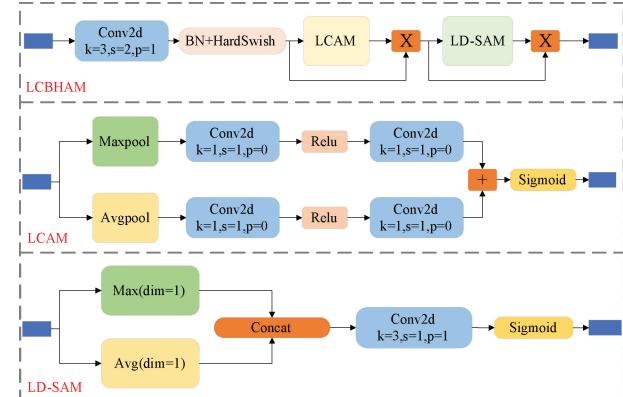


图 3 LCBHAM 模块

Fig. 3 LCBHAM module

PAN 结构的底层已经融合了图像的全部信息, LCBHAM 模块是基于注意力的特征转换而设计的, 该模块能够从综合的特征中提取更丰富的特征信息, 增加感受野, 同时也能突出特征中的重要信息并抑制不重要的信息, 所以将 LCBHAM 模块集成到 PAN 结构的底层与输出部分结合。此外, LCBHAM 模块与 CBS 模块的参数量相近, 不会额外增加模型的参数量。

## 2.2 构建 LR-C2f 模块

对于小目标, 细微的特征差异非常重要, 但在经过多次下采样后, 这些细节信息可能被模糊化或丢失<sup>[15]</sup>, 局部特征增强机制(local spatial block)能够高效地捕获并保留输入特征层的局部空间信息, 并将其与原始输入融合, 以增强特征信息的表示, 而残差门控机制(ResGated Block)旨在利用深度可分离卷积和残差连接的特性提高模型性能并降低计算成本。因此本文借鉴残差门控机制和局部特征增强机制设计了 LR-C2f 模块, 提升小目标信息提取能力的同时能够保留更多的细粒度特征, LR-C2f 模块如图 4 所示。

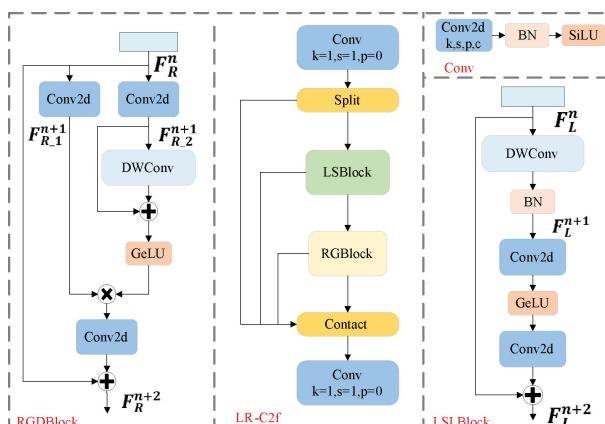


图 4 LR-C2f 模块

Fig. 4 LR-C2f module

在 LocalSpatial Block 中, 对于输入的特征图  $F_L^n$ , 首先经过深度可分离卷积, 该卷积分别在每个输入通道上进行相乘操作, 而不混合通道信息, 有效地提取了输入特征图的局部空间信息, 同时减少了计算成本和参数数量, 然后经过批量归一化以实现一定程度的正则化效果, 同时减少过拟合, 得到的中间层  $F_L^{n+1}$ , 即:

$$F_L^{n+1} = \text{BN} [\text{DWConv}_{3 \times 3} (F_L^n)] \quad (4)$$

中间层通过  $1 \times 1$  卷积聚合通道信息, 并通过非线性激活函数 GeLU, 使模型能够学习到更复杂的特征信息; 最后, 通过残差连接将原始输入与处理后的特征融合得到最终的输出  $F_L^{n+2}$ , 即:

$$F_L^{n+2} = F_L^n \oplus \text{Conv}_{1 \times 1} \{ \text{GeLU} [\text{Conv}_{1 \times 1} (F_L^{n+1})] \} \quad (5)$$

LocalSpatial Block 使得模型能够理解和整合图像中不同维度的特征, 从而增强对图像在下采样过程中尺度变化的鲁棒性。

ResGated Block 的输入特征层分为两个分支, 并以  $1 \times 1$  卷积的形式实现全连接, 得到中间状态  $F_{R,1}^{n+1}$  和  $F_{R,2}^{n+1}$ , 其中:

$$F_{R,1}^{n+1} = \text{Conv}_{1 \times 1} (F_R^n) \quad (6)$$

$$F_{R,2}^{n+1} = \text{Conv}_{1 \times 1} (F_R^n) \quad (7)$$

深度可分离卷积被用作其中一个分支上的位置编码模块, 通过残差连接使得模型在训练过程中更高效地回流梯度, 这降低了计算成本并通过保留和利用图像的空间结构信息以提高性能。而后采用非线性 GeLU 作为激活函数, 在每个级别控制信息流, 然后通过元素乘法与另一个分支合并, 通过  $1 \times 1$  卷积与全局特征融合以混合通道信息, 最后通过残差连接与原始输入特征层相加得到最终输出  $F_R^{n+2}$ , 即:

$$F_R^{n+2} = F_R^n \oplus \text{Conv}_{1 \times 1} \{ F_{R,1}^{n+1} \odot \text{GeLU} [F_{\text{temp}}] \} \quad (8)$$

$$F_{\text{temp}} = F_{R,2}^{n+1} \oplus \text{DWConv}_{3 \times 3} (F_{R,2}^{n+1}) \quad (9)$$

ResGated Block 模块将全局依赖性与全局特征传递给每个像素, 以捕获相邻特征层之间的依赖关系, 从而丰富上下文信息, 进一步增强了模型的表达能力。

## 2.3 引入局部特征增强模块

在骨干网络提取特征的过程中, 输入图像的尺寸会随着模型层数堆叠而减小, 这个过程中图像的局部信息会逐渐丢失, 导致模型对小目标的检测准确率降低<sup>[16]</sup>, 引入 ConvFFN 用于聚合局部信息, 减少局部信息丢失。

FFN(Feed-Forward Network)是一种普通的特征增强模块, 然而传统 FFN 仅用多个全连接层在通道维度上进行增强, 并没有考虑到特征图的二维特征信息, 因此引入深度卷积用于弥补这个缺陷, ConvFFN 是在 FFN 模块中的 GeLU 激活函数后增加深度可分离卷积<sup>[17]</sup>, 这使得 ConvFFN 能够有效聚合局部信息。ConvFFN 有两种类型, 一种是输入端直接与经过 ConvFFN 操作的输出端相加的 In-Stage ConvFFN, 另一种是输入端经过一个深度可分离卷积、归一化和普通卷积后与经过 ConvFFN 操作的输出端相加的 Down-Sample ConvFFN, 如图 5 所示。

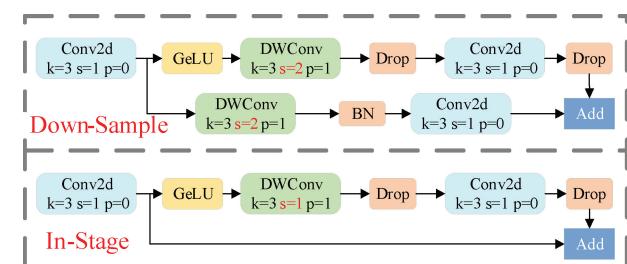


图 5 两种 ConvFFN 结构对比

Fig. 5 Comparison of two ConvFFN structures

骨干网络中的下采样操作由普通的卷积来完成, 因此, 本文将能够聚合局部信息的 ConvFFN 模块与普通卷积组合, 在满足下采样要求的同时使得模型能够学习到更多的局部特征, 增强骨干网络对小目标的特征提取能力,

两种融合方式以及不同尺寸卷积核对检测结果的影响如本文 4.3 节所示。

### 3 数据集与评价指标

#### 3.1 数据集

本文采用 Tiny Person<sup>[18]</sup>公开数据集验证改进算法的有效性,该数据集中标注的最大目标框只占了整幅图像的 2.2%,所以检测 Tiny Person 中的小目标更具挑战性。此外,本文通过水平翻转、垂直翻转等数据增强方法将数据集扩充至 3 786 幅,最后设置训练集与验证集的比例为 4:1。

#### 3.2 评价指标

为了验证本文所提出的 LDF-YOLO 模型的有效性,本研究采用了深度学习领域常用的评估指标来评估模型的检测性能。这些评估指标包括准确率(Precision)、召回率(Recall)、平均精度均值(mAP0.5, mAP0.5 : 0.95)、参数量(parameters)以及浮点运算量(GFLOPs)。以上的评价指标使用了以下参数:TP(预测结果为正样本,实际也为正样本的样本数量)、FP(预测结果为正样本,实际为负样本的样本数量)、FN(预测结果为负样本,实际为正样本的样本数量)、IoU(预测边界框与真实边界框的交集和并集的比率)。各指标的计算如下:

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

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

$$AP = \int_0^1 Precision(Recall) d(Recall) \quad (12)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (13)$$

其中,AP 是 PR 曲线在坐标轴上所围成的面积,表示所有类别的平均值;mAP 是所有样本类别 AP 值的加权平均值;N 表示所有样本类别的数量;mAP0.5 表示将 IoU 设置为 0.5 所得 mAP 值;mAP0.5 : 0.95 表示以间隔 0.05 将 IoU 设置为 0.5~0.95 之间所得 mAP 值。

## 4 实验与结果分析

#### 4.1 实验环境

本实验是基于 Ubuntu20.04.5 操作系统,在一张显存大小为 32 GB 的 NVIDIA V100 显卡上训练、验证模型,CPU 的配置为 10 vCPU Intel Xeon Processor (Skylake, IBRS), 实验环境为 Python3.8、PyTorch2.0.0、CUDA11.8。模型相关参数设置:训练迭代次数为 300;每次迭代的批次大小为 8;动量设置为 0.937;权重衰减系数设置为 0.0005;开启预热和早停机制;采用余弦退火算法更新学习率,初始化学习率与最终学习率为 0.01;使用 SGD 优化器。

#### 4.2 改进模型检测结果

为了体现 LDF-YOLO 算法的有效性,将其与原

YOLOv8 算法进行各指标上的对比,对比结果如表 1 所示。由表 1 可知各项指标涨幅明显,说明 LDF-YOLO 算法对小目标的检测效果显著。

表 1 原算法与改进算法各项指标对比

Table 1 Comparison of various indicators between the original algorithm and the improved algorithm

模型结构	YOLOv8	LDF-YOLO
输入尺寸	640×640	640×640
准确率/%	38.4	42.9(↑4.5)
召回率/%	28.4	33.9(↑5.5)
mAP0.5/%	24.9	29.6(↑4.7)
mAP0.5:0.95/%	7.5	10.3(↑2.8)

图 6 和 7 是原 YOLOv8 算法与 LDF-YOLO 算法的各项指标在 Tiny Person 数据集上随迭代轮数变化的曲线图。可以看出改进后算法比原 YOLOv8 算法,在训练次数一致的情况下,改进后的算法精度更高,并且模型在训练后期更加稳定。

#### 4.3 改进机制对比实验

##### 1) 检测层不同改进机制对比实验

在检测层的改进部分,本文将 LCBHAM 模块添加到特征融合网络中的 C2f 模块之后代替普通的卷积操作,因此将 LCBHAM 模块分别添加到第 19 层和第 22 层之后进行了对比实验,实验结果如表 2 所示,由表可知,将其添加在第 19 层之后效果更好。

##### 2) 引入不同类型 ConvFFN 模块对比实验

在骨干网络的改进部分,本文引入的 ConvFFN 模块有 4 种不同的结构,因此将它们分别与常规卷积结合,实验结果如表 3 所示,“(3×3)”表示 DWConv 的卷积核尺寸,由表可知,不同结构的 ConvFFN 模块提升效果相近,但是 In-Stage ConvFFN(3×3)的计算量更低,因此将其融入骨干网络中。

##### 3) 骨干网络不同优化方法对比实验

为了更好的体现出在骨干网络中融入 ConvFFN 的有效性,引入 SPDConv<sup>[19]</sup>、AKConv<sup>[20]</sup> 和 EfficientNet<sup>[21]</sup> 进行了对比,实验结果如表 4 所示,虽然各种改进方法在准确率上的提升不相上下,但是融入 ConvFFN 使得召回率有明显大于其他改进方法的提升。

##### 4) C2f 模块添加不同机制对比实验

基于骨干网络和特征融合网络中 C2f 模块的改进主要分为两种:一是引入经典的注意力机制,如 CBAM<sup>[22]</sup>、GAM<sup>[23]</sup> 和 Simam<sup>[24]</sup>;二是融合较为新颖的模块化操作,如 STARNet<sup>[25]</sup>、SANet<sup>[26]</sup> 和 RFAConv<sup>[27]</sup>。相关对比实验表 5 所示。表 5 显示,除了 STAR-C2f 之外,其余改进皆有所提升,其中 RFA-C2f 与本文设计的 LR-C2f 提升效果相近,但 RFA-C2f 结构复杂,不满足模型轻量化的要求。

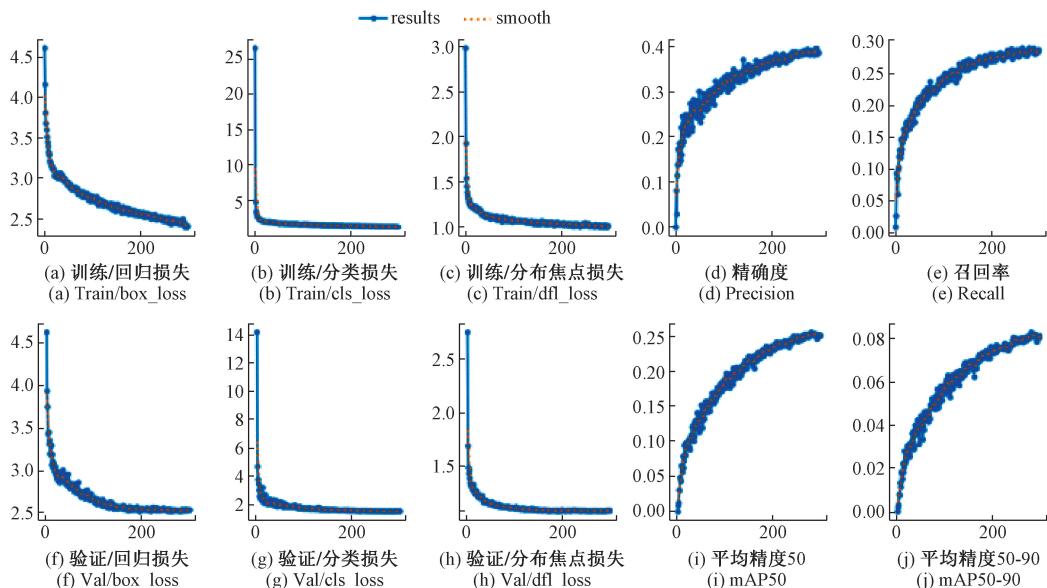


图 6 YOLOv8 各项指标曲线图

Fig. 6 YOLOv8 indicator curve chart

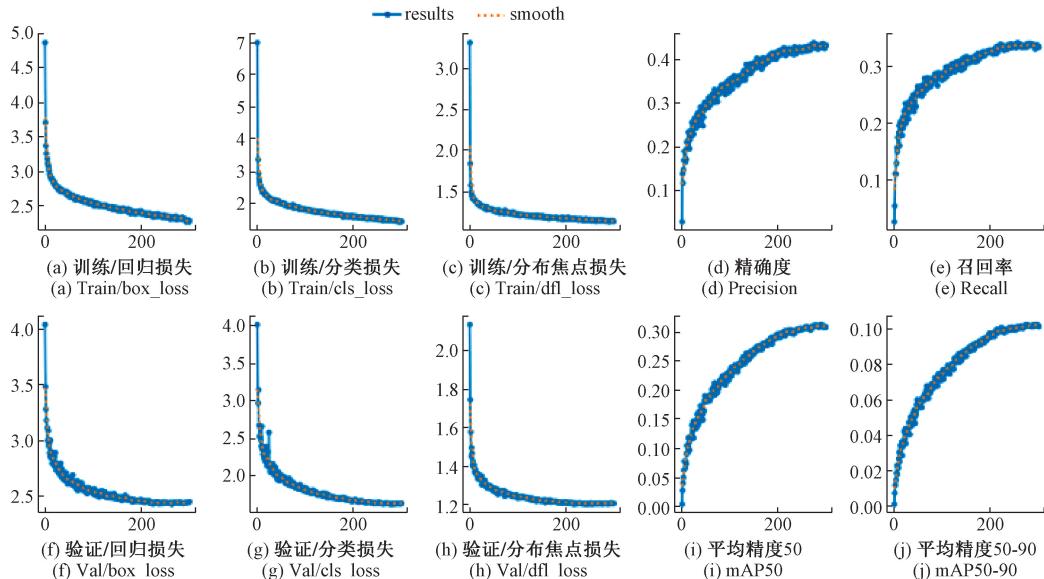


图 7 LDF-YOLOv8 各项指标曲线图

Fig. 7 LDF-YOLOv8 indicator curve chart

表 2 检测层不同改进机制对比实验

Table 2 Comparison experiment of different improvement mechanisms in the detection layer %

基线模型	P	R	mAP0.5	mAP0.5:0.95
L19	<b>41.6</b>	33.1	<b>28.8</b>	<b>9.2</b>
L22	40.5	33.1	28.1	8.8

#### 4.4 消融实验

为了验证每个改进方法对提高 YOLOv8 算法检测性能的有效性, 使用 Tiny Person 数据集并以 YOLOv8n 为基本模型进行了消融实验, 实验结果如表 6 所示, 其中

表 3 不同类型 ConvFFN 模块对比实验

Table 3 Comparative experiments of different types of ConvFFN modules

基线模型	P/%	R/%	mAP0.5/%	FLOPs/G
In-Stage(3×3)	38.9	29.6	25.3	<b>8.5</b>
Down-Samplef(3×3)	38.6	29.7	25.4	8.7
In-Stage(5×5)	39.1	29.5	25.4	8.8
Down-Samplef(5×5)	<b>39.0</b>	<b>29.9</b>	<b>25.6</b>	9.2

“√”表示使用了对应的改进策略, 表 6 的消融实验表明, 不同的改进策略对模型的小目标检测性能有不同的提升

表 4 骨干网络不同优化方法对比实验

Table 4 Comparison experiment of different optimization methods for backbone networks %

基线模型	P	R	mAP0.5	mAP0.5:0.95
SPDConv	39.0	28.6	25.1	7.9
AKConv	38.4	28.5	24.9	7.5
EfficientNet	38.6	28.3	24.8	7.5
ConvFFN	38.9	<b>29.6</b>	<b>25.3</b>	<b>8.1</b>

表 5 C2f 模块添加不同机制对比实验

Table 5 Comparison experiment of adding different mechanisms to C2f module %

基线模型	P	R	mAP0.5	mAP0.5:0.95
C2f	38.4	28.4	24.9	7.5
CBAM-C2f	39.3	28.9	25.4	7.5
STAR-C2f	38.1	28.6	24.1	7.1
SA-C2f	39.6	<b>30.0</b>	25.4	7.8
GAM-C2f	40.1	29.2	26.2	8.2
RFA-C2f	40.5	29.6	26.3	8.7
Simam-C2f	40.1	29.1	26.1	8.2
LR-C2f(ours)	<b>41.1</b>	29.8	<b>26.6</b>	8.7

效果。在骨干网络中融入 ConvFFN 后,增强了骨干网络聚合局部信息的性能,准确率和 mAP 值均有小幅度提升。使用 LR-C2f 模块替换原始的 C2f 模块之后,使得模型能够高效地捕获和表示输入特征图的局部空间信息,并将其与原始输入融合,增强特征的表示,准确率提升了 2.7%。在模型的输出端添加针对小目标的 LPD(LP-Detect)检测头,准确率提升了 3.2%,召回率提升了 5.7%,漏检率大幅度降低。在融入小目标检测头的基础上,分别加入 ConvFFN 与 LR-C2f 模块,准确率与 mAP 值均有接近 4% 的提升,说明在降低漏检率的同时,针对小目标的检测性能进一步提升。最后将 3 种改进方法结合起来,mAP0.5 达到了 29.6%,有 4.7% 的提升。通过消融实验的结果可知,相较于原始 YOLOv8 算法,改进后的算法在小目标的检测中效果更佳,检测准确率大幅上升,证明了改进方法的有效性。

#### 4.5 对比实验

为了进一步体现改进后的模型在小目标检测上的优越性,使用了目标检测领域的经典算法进行对比,包括 Faster R-CNN、YOLOv5n、YOLOv5s、DETA<sup>[28]</sup>、DeoT<sup>[29]</sup>、YOLOv7-tiny、VA-YOLOv5s<sup>[30]</sup>、Drone-YOLO<sup>[31]</sup> 和 LW-YOLOv8<sup>[32]</sup>,在相同的实验环境和模型参数的情况下使用 Tiny Person 数据集进行验证,对比实验结果如表 7 所示。

表 6 消融实验

Table 6 Ablation experiment

YOLOv8n	LPD	LR-C2f	ConvFFN	P/%	R/%	mAP0.5/%	mAP0.5:0.95/%	FLOPs/G
✓				38.4	28.4	24.9	7.5	<b>8.1</b>
✓			✓	38.9	29.6	25.3	8.1	8.5
✓		✓		41.1	29.8	26.6	8.7	11.4
✓	✓			41.6	<b>34.1</b>	28.8	9.2	11.9
✓	✓	✓		42.5	33.4	29.3	10.1	15.4
✓	✓		✓	41.7	32.9	29.1	9.6	12.3
✓	✓	✓	✓	<b>42.9</b>	33.9	<b>29.6</b>	<b>10.3</b>	15.9

表 7 对比实验

Table 7 Comparative experiment

模型	P/%	R/%	mAP0.5/%	mAP 0.5:0.95/%	FLOPs/G
Faster R-CNN	18.5	12.4	7.6	1.0	270.6
YOLOv5n	30.9	25.6	18.2	4.9	<b>4.7</b>
YOLOv5s	32.3	26.8	20.1	5.4	15.3
DETA	38.6	20.5	20.3	5.6	129.1
DeoT	39.3	16.5	19.1	6.0	144.6
VA-YOLOv5s	38.1	29.0	24.1	7.1	16.1
YOLOv7-tiny	35.6	27.0	22.8	6.0	7.6
YOLOv8n	38.4	28.4	24.9	7.5	8.1
YOLOv8s	42.5	30.0	28.1	9.2	28.4
Drone-YOLO	<b>43.1</b>	31.9	28.8	9.2	19.5
LW-YOLOv8	41.8	31.0	28.2	8.7	17.4
本文	42.9	<b>33.9</b>	<b>29.6</b>	<b>10.3</b>	15.9

从表 7 可以看出 Faster R-CNN 精度低且模型过于复杂; 基于 Transformer 的模型 DETA 和 DeoT 在准确度上有一定的优势, 但是计算量过大, 且召回率较低, 存在严重的漏检问题; YOLOv5 和 YOLOv7 的特征提取模块对小目标的细节信息并没有做针对性改进, 且检测头感受野较小, 检测性能明显低于本文算法。VA-YOLOv5s 的参数量与本文算法相近, 但是在漏检率和误检率上高于本文算法。YOLOv8s 相比于 YOLOv8n, 检测性能有所提升, 但是单纯的增加网络深度并没有降低漏检率, 且计算量增加了 3 倍以上。Drone-YOLO 算法在 neck 部分引入“sandwich-fusion”模块并增加相较于原图下采样 4 倍的检测头, 忽略了骨干网络中特征层的原始信息, 增加了计算量且漏检率降低不明显; LW-YOLOv8 算法在骨干网络中增加通道注意力机制以充分捕获全局信息, 并使用 GSConv 替代普通卷积缓解通道注意力机制导致参数量上升的问题, 但没有考虑跨层连接过程中细节信息的丢失, 检测结果不理想。本文算法凭借 ConvFNN 增强对小目标信息的提取能力, 并在检测层之前利用 LR-C2f 模块保留细粒度特征的特点实现细节信息的保留, 最后利用 LCBHAM 模块从全局信息与局部信息高度综合的特征图中捕获重要的局部信息并抑制不重要的特征信息, 因此模型性能均优于上述算法。

综上所述, 本文提出的 LDF-YOLO 算法, 相较于原始 YOLOv8 算法以及其他经典的目标检测算法在降低漏检率和误检率上有明显的优势, 因此进一步证明了本文所做出的改进在小目标检测上是有效的。

#### 4.6 泛化性验证

为了验证改进方法的适用性和泛化性, 在 NWPU VHR-10<sup>[33]</sup> 和 VisDrone2019<sup>[34]</sup> 数据集上进行了对比实验, 实验结果如表 8 所示, 由表 8 实验结果可知, 本文算法在多目标且目标体积小的数据集上各项指标均有大幅度提升, 充分证明了 LDF-YOLO 算法在检测小目标上具有高性能的同时, 也具备良好的适用性。

表 8 不同数据集对比实验

Table 8 Comparison experiment of different datasets

数据集-模型	P	R	mAP0.5	mAP 0.5:0.95	%
NVHR-YOLOv8n	80.5	78.2	83.0	61.3	
NVHR-本文	86.7	83.6	85.6	64.2	
VisD-YOLOv8n	43.6	32.0	31.1	12.2	
VisD-本文	46.3	34.2	31.9	13.0	

#### 4.7 实验结果可视化与分析

在实验过程中分别选取了 3 组不同视角下的图像, 如图 8(a)所示, LDF-YOLO 算法在 3 组不同视角下的图像中表现出了较好的效果。相比于图 8(b)的检测结果, 改进

后的算法成功地检测到了原始模型没有检测到的小目标, 在图 8(c)中用黄色圆形框标注, 有效地降低了漏检率; 此外, 由第 1 组和第 2 组的输出结果可知, 无论是远距离成像还是小目标较为密集的场景, LDF-YOLO 检测结果的置信度更高, 因此改进后的算法不仅能够明显提高召回率, 并且还能够保持较高的识别精度。



图 8 YOLOv8 与 LDF-YOLO 检测结果可视化

Fig. 8 Visualization of YOLOv8 and LDF-YOLO detection results

## 5 结 论

本文针对小目标检测效果不佳的问题, 提出了一种基于 LDF-YOLO 的目标检测算法。首先, 通过在骨干网络融入 ConvFFN, 有效提高了骨干网络聚合局部信息的能力; 其次, 设计了 LR-C2f 模块, 该模块借鉴了残差门控机制与局部特征增强机制, 旨在提升模型提取局部特征的能力; 最后, 引入特征转换模块 LCBHAM, 设计了 LP-Detect 检测层并调整特征融合网络, 使得模型在输出阶段捕获重要的局部信息的同时抑制不重要的特征信息, 有效提升模型小目标检测的能力并降低漏检率。实验结果表明, 本文算法在保证检测速度的前提下, 准确率、召回率以及 mAP 上均高于原始 YOLOv8 算法, 且检测性能优于其他经典的目标检测算法。虽然改进模型的参数量有所上升, 但是仍能满足一定的实时性的要求, 后续将在保持检测准确度的基础上进一步降低模型的参数量, 使其更易于部署在边缘计算设备上。

## 参考文献

- [1] ZHAO W X, ZHOU K, LI J, et al. A survey of large language models [J]. ArXiv preprint arXiv: 2303.18223, 2023.
- [2] DUAN K W, BAI S, XIE L X, et al. Centernet++ for object detection [J]. ArXiv preprint arXiv: 2204.08394, 2022.
- [3] CAO X, YUAN P, FENG B, et al. CF-DETR: Coarse-to-fine transformers for end-to-end object detection[C]. AAAI Conference on Artificial Intelligence, 2022, 36(1): 185-193.
- [4] XU W, ZHANG C, WANG Q, et al. FEA-swin: Foreground enhancement attention swin transformer network for accurate UAV-based dense object detection[J]. Sensors, 2022, 22(18): 6993.
- [5] MOHAMMED A, KORA R. A comprehensive review on ensemble deep learning: Opportunities and challenges [J]. Journal of King Saud University-Computer and Information Sciences, 2023, 35(2): 757-774.
- [6] 王铮帅,邱联奎,李迎港.复杂环境下的YOLOv5s烟火检测方法[J].电子测量技术,2023,46(24):149-156.
- WANG ZH SH, QIU L K, LI Y G. YOLOv5s fireworks detection method in complex environments[J]. Electronic Measurement Technology, 2023, 46(24): 149-156.
- [7] 邵凯,李浩刚,梁燕,等.跨尺度特征融合的遥感微小目标检测算法[J/OL].系统工程与电子技术,1-13[2025-01-23]. <http://kns.cnki.net/kcms/detail/11.2422.TN.20240930.1400.027.html>.
- SHAO K, LI H G, LIANG Y, et al. Remote sensing micro target detection algorithm based on cross scale feature fusion [J/OL]. System Engineering and Electronics, 1-13[2025-01-23]. <http://kns.cnki.net/kcms/detail/11.2422.TN.20240930.1400.027.html>.
- [8] 赵金宪,赵志滢.改进YOLOv8n的小目标检测算法[J].黑龙江科技大学学报,2024,34(4):642-647.
- ZHAO J X, ZHAO ZH Y, Improving YOLOv8n's small object detection algorithm [J]. Journal of Heilongjiang University of Science and Technology, 2024,34(4): 642-647.
- [9] 雷帮军,余鞠,余快.基于YOLOv8s改进的小目标检测算法[J].无线电工程,2024,54(4):857-870.
- LEI B J, YU AO, YU K. Improved small object detection algorithm based on YOLOv8s [J]. Radio Engineering, 2024,54(4):857-870.
- [10] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [11] SHAMSHAD F, KHAN S, ZAMIR S W, et al. Transformers in medical imaging: A survey [J]. Medical Image Analysis, 2023, 88: 102802.
- [12] CHEN J R, KAO S, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021-12031.
- [13] JIANG T, LU P, ZHANG L, et al. Rtmpose: Real-time multi-person pose estimation based on mmpose [J]. ArXiv preprint arXiv:2303.07399, 2023.
- [14] ZHANG Y, ZHANG H, HUANG Q, et al. DsPYOLO: An anchor-free network with DsPAN for small object detection of multiscale defects[J]. Expert Systems with Applications, 2024, 241: 122669.
- [15] WANG Z, LI C, XU H, et al. Mamba YOLO: SSMs-based YOLO for object detection [J]. ArXiv preprint arXiv:2406.05835, 2024.
- [16] FAN Q, HUANG H, GUAN J, et al. Rethinking local perception in lightweight vision transformer[J]. ArXiv preprint arXiv:2303.17803, 2023.
- [17] 王成龙,赵倩,赵琰,等.基于深度可分离卷积的实时遥感目标检测算法[J].电光与控制,2022,29(8):45-49.
- WANG CH L, ZHAO Q, ZHAO Y, et al. Real-time remote sensing target detection algorithm based on depth separable convolution [J]. Electro Optics and Control, 2022,29(8):45-49.
- [18] YU X, GONG Y, JIANG N, et al. Scale match for tiny person detection [C]. IEEE/CVF Winter Conference on Applications of Computer Vision, 2020: 1257-1265.
- [19] HUANG F Z, LI S. Lightweight you only look once v8: An upgraded you only look once v8 algorithm for small object identification in unmanned aerial vehicle images[J]. Applied Sciences, 2023, 13(22): 12369.
- [20] ZHANG X, SONG Y, SONG T, et al. AKConv: Convolutional kernel with arbitrary sampled shapes and arbitrary number of parameters [J]. ArXiv preprint arXiv:2311.11587, 2023.
- [21] WOO S, DEBNATH S, HU R H, et al. Convnext v2: Co-designing and scaling convnets with masked autoencoders [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 16133-16142.
- [22] TERVEN J, CORDOVA E D M, ROMERO G J A. A comprehensive review of YOLO architectures in computer

- vision: From YOLOv1 to YOLOv8 and YOLO-nas[J]. Machine Learning and Knowledge Extraction, 2023, 5 (4): 1680-1716.
- [23] LIU Y, SHAO Z, HOFFMANN N. Global attention mechanism: Retain information to enhance channel-spatial interactions [J]. ArXiv preprint arXiv: 2112.05561, 2021.
- [24] YANG L, ZHANG R Y, LI L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks [C]. International Conference on Machine Learning. PMLR, 2021: 11863-11874.
- [25] MA X, DAI X, BAI Y, et al. Rewrite the stars[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 5694-5703.
- [26] ZHANG Q L, YANG Y B. Sa-net: Shuffle attention for deep convolutional neural networks[C]. ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021: 2235-2239.
- [27] ZHANG X, LIU C, YANG D, et al. RFACConv: Innovating spatial attention and standard convolutional operation[J]. ArXiv preprint arXiv:2304.03198, 2023.
- [28] OUYANG-ZHANG J, CHO J H, ZHOU X, et al. NMS strikes back [J]. ArXiv preprint arXiv: 2212.06137, 2022.
- [29] DING T, FENG K, WEI Y, et al. DeoT: An end-to-end encoder-only Transformer object detector[J]. Journal of Real-Time Image Processing, 2023, 20(1): 1.
- [30] 刘展威,陈慈发,董方敏.基于 YOLOv5s 的航拍小目标检测改进算法研究[J].无线电工程, 2023, 53(10): 2286-2294.
- LIU ZH W, CHEN C F, DONG F M. Improved aerial small object detection algorithm based on YOLOv5s[J]. Radio Engineering, 2023, 53 (10): 2286-2294.
- [31] ZHANG Z. Drone-YOLO: An efficient neural network method for target detection in drone images[J]. Drones, 2023, 7(8):526-549.
- [32] HUANG F Z, LI S. Lightweight you only look once v8: An upgraded you only look once v8 algorithm for small object identification in unmanned aerial vehicle images[J]. Applied Sciences, 2023, 13(22): 12369.
- [33] OSCO L P, WU Q, DE LEMOS E L, et al. The segment anything model (sam) for remote sensing applications: From zero to one shot[J]. International Journal of Applied Earth Observation and Geoinformation, 2023, 124: 103540.
- [34] ZHU P, WEN L, DU D, et al. Detection and tracking meet drones challenge [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(11): 7380-7399.

## 作者简介

**刘洋**,硕士研究生,主要研究方向为目标检测。

E-mail:2292152718@qq.com

**任旭虎**(通信作者),副教授,主要研究方向为智能测控技术、智能信息处理。

E-mail:Z23160105@s.upc.edu.cn

**刘宝弟**,副教授,主要研究方向为机器学习理论及其在计算机视觉领域中的应用。

E-mail:liubaodi@upc.edu.cn

**刘伟峰**,教授,主要研究方向为模式识别、机器学习理论及应用。

E-mail:liuwf@upc.edu.cn