

基于YOLOv8改进的无人机视觉小目标检测模型

刘纪红, 时瑞瑞

(东北大学 信息科学与工程学院, 辽宁 沈阳 110819)

摘要: 针对无人机航拍图像中小目标易误检和漏检的问题以及无人机检测对实时性和轻量化的需求, 提出一种基于YOLOv8改进的轻巧高效模型. 首先, 将YOLOv8的Neck部分简化为特征金字塔网络, 使模型有效利用浅层网络提取的细节信息, 并增加特征融合模块为Head层提供更利于小目标检测的特征; 其次, 在Backbone部分集成高效局部注意力机制以实现目标区域的精确定位. 实验结果表明, 与YOLOv8s相比, 改进模型的参数量和模型规模分别降低50%, $mAP_{0.5}$ 和检测速度分别提升4%. 该改进模型为无人机检测领域的部署提供了新思路.

关键词: 小目标检测; YOLOv8; 高效局部注意力机制; 模型轻量化; 无人机航拍

中图分类号: TP 391.4 文献标志码: A 文章编号: 1005-3026(2025)12-0029-09

An Improved Small Object Detection Model Based on YOLOv8 for UAV Vision

LIU Ji-hong, SHI Rui-rui

(School of Information Science & Engineering, Northeastern University, Shenyang 110819, China. Corresponding author: LIU Ji-hong, E-mail: liujihong@ise.neu.edu.cn)

Abstract: In view of easy false detection and missed detection of small objects in unmanned aerial vehicle (UAV) aerial images, as well as the requirements for real-time performance and lightweight design in UAV detection tasks, an improved lightweight and efficient model based on YOLOv8 was proposed. Firstly, the Neck part of YOLOv8 was simplified into a feature pyramid network, enabling the model to effectively utilize the detailed information extracted by shallow networks. Meanwhile, a feature fusion module was added to provide more favorable features for small object detection to the Head layer. Secondly, an efficient local attention (ELA) mechanism was integrated into the Backbone part to achieve accurate localization of target regions. Experimental results show that compared with YOLOv8s, the parameters and model size of the improved model are reduced by 50%, while the $mAP_{0.5}$ and detection speed are improved by 4%. This improved model provides a new idea for the deployment of UAV detection.

Key words: small object detection; YOLOv8; efficient local attention mechanism; lightweight model; UAV aerial photography

无人机体积小、方便灵活、应用领域广泛, 近年来被大量应用于军用或民用领域. 其应用的关键在于目标检测技术, 尤其是小目标检测技术的有力支撑. 目前主流的检测算法对中大目标可以实现高效迅速的检测, 但由于小目标分辨率低且所占区域较小, 现有检测算法对小目标的实时高效检测存在困难. 因此, 提升目标检测模型对小

目标的检测能力, 实现对小目标的实时精准检测具有重要的实际意义.

基于深度学习的目标检测算法分为基于Transformer的算法、基于Anchor Based的算法以及基于Anchor Free的算法.

Transformer^[1]最初主要应用于自然语言处理(natural language processing, NLP)领域. 2020年

Carion 等^[2]提出的 DETR 首次将其应用于端到端的目标检测,然而 DETR 系列检测器在没有相应领域的预训练情况下很难应用于新领域。

基于 Anchor Based 的算法分为两类:一类是“两阶段检测”,其先驱是 Girshick 等^[3]提出的 R-CNN 算法. Ren 等^[4]进一步提出了 Faster R-CNN,这是一种端到端的检测器,基本实现了实时性. 2017 年,文献^[5]提出了基于 Faster R-CNN 的特征金字塔网络(feature pyramid network, FPN). 另一类是“单阶段检测”算法,主要是由 Redmon 等^[6]提出的 YOLOv1 算法. 后来文献^[7]提出了 SSD 算法. 2018 年,出现了 YOLOv3 算法^[8],之后 Glenn^[9]提出了 YOLOv5 算法.

基于 Anchor Free 的算法是为了使检测器更简单高效而提出的. 研究者探索如何去除锚框以设计更为简洁高效的检测器. 2018 年 CornerNet^[10]被提出,该算法基于关键点检测目标的角点,从而确定目标的位置和大小. 2019 年 Tian 等^[11]提出了 FCOS 算法,此算法使用锚点对目标进行分类和回归. 2024 年, YOLOv5 的提出者 Glenn 提出了 YOLOv8 算法^[12],该算法与 YOLO 系列前几个版本的最大区别在于其采用 Anchor Free 检测头,无需预定义锚框即可直接预测目标信息,因而能够更准确地捕捉目标的形状和位置. 目前 YOLO 算法已发展至 YOLOv9^[13],该系列算法检测精度高、速度快,是现今应用最广泛的实时检测器.

为了实现小目标的高效检测,文献^[14]采用基于密度的裁剪方法提取和处理拥挤的小目标区域. 文献^[15]设计了多跳多尺度融合模块并改进损失函数以提升模型对红外微小目标的检测能力. 文献^[16]提出一种基于 YOLOv5x 的改进算法. 文献^[17]提出无需额外模块的基于密度裁剪的高效级联放大目标检测器. 尽管现有改进方法提升了小目标的检测精度,但多数伴随模型复杂度增加或检测时间延长的问题,不利于对无人机拍摄的小目标图像实现快速精准检测. 本文基于 YOLOv8 检测算法,通过改进其特征提取和特征融合模块来提升对小目标的检测性能,构建便于在边缘设备部署的实时高效小目标检测器.

本文的具体研究如下:

1) 将 YOLOv8 的 Neck 部分由路径聚合网络(path aggregation network, PAN)^[18]简化为特征金字塔网络(FPN),并增加一个特征融合模块,在降低模型大小和推理时间的同时提升模型对小目标的召回率;

2) 在 YOLOv8 的 Backbone 中加入高效局部注意力(efficient local attention, ELA)机制^[19],提升模型对关键特征信息的提取能力;

3) 使用 VisDrone-2019 数据集进行实验,验证所提的改进模型的有效性.

1 基于 YOLOv8 改进的检测模型

1.1 改进 YOLOv8 模型 Neck 部分

目标检测任务主要包括特征提取、特征融合以及分类与定位. 检测模型通常由 Backbone, Neck 以及 Head 三部分组成. YOLOv8 模型的 Backbone 采用 YOLOv3 中提出的 Darknet-53 网络, Neck 为 PAN 网络, Head 采用解耦头进行目标的分类和定位. 模型整体结构如图 1a 所示, 模块细节如图 1b 所示, YOLOv8 的参数配置如图 1b 左下角的表格所示. 图 1 中: 参数 k 表示卷积核大小; s 表示步长; p 表示填充大小; n 表示每个模块的数量; H, W, C 分别为输入特征的高度、宽度和通道数; a 表示 YOLOv8 模型的深度倍数; b 表示宽度倍数; r 表示比率. YOLOv8 模型的 Backbone 部分包含多个由卷积和跨阶段部分融合模块(cross stage partial fusion module, C2f 模块)组成的特征提取模块,以及快速空间金字塔池化(spatial pyramid pooling-fast, SPPF)模块,其主要功能是提取输入图像的特征信息. C2f 模块主要由瓶颈模块和卷积块组成,瓶颈模块的 Shortcut 参数用于控制是否启用跳跃连接. Shortcut 为 True 时,瓶颈模块为残差模式; Shortcut 为 False 时,瓶颈模块为普通卷积块模式. Neck 部分采用 PAN 网络结构,由多个特征融合模块 1 和特征融合模块 2 组成,用于对 Backbone 部分提取的特征进行多尺度融合和进一步提取. Head 部分的输入为 P3, P4, P5 层的特征,输入特征的通道数分别为 256, 512, 1 024, 特征尺寸与模型输入图像尺寸相比分别为下采样 8 倍、16 倍和 32 倍.

PAN 是在 FPN 的基础上,使用一个额外的自底向上的路径将底层特征信息传递给高层. 大多数研究者对 Neck 部分的改进都是在 PAN 的基础上进一步复杂化特征融合的方式,如改进为双向特征金字塔网络(bi-directional feature pyramid network, BiFPN)^[20]. 尽管模型检测性能得到了提升,但增加了检测模型的部署难度. 本文在 Neck 部分舍弃路径聚合网络,仅采用特征金字塔网络融合 Backbone 与 Neck 部分的特征以降低网络深

度,减少模型参数量和规模.同时为更有效地利用浅层网络提取的细节信息,增加了一个特征融合模块来整合浅层网络特征,为Head层提供更利于小目标检测的特征表达,从而提升模型对小目标的召回率.改进后的模型将P2,P3,P4层的特

征作为Head层的输入,其特征图相对于输入图像的下采样倍数分别为4倍、8倍、16倍.为在提升小目标检测性能的同时保持模型轻量化,将Head层输入特征的通道数缩减至64,128,256.

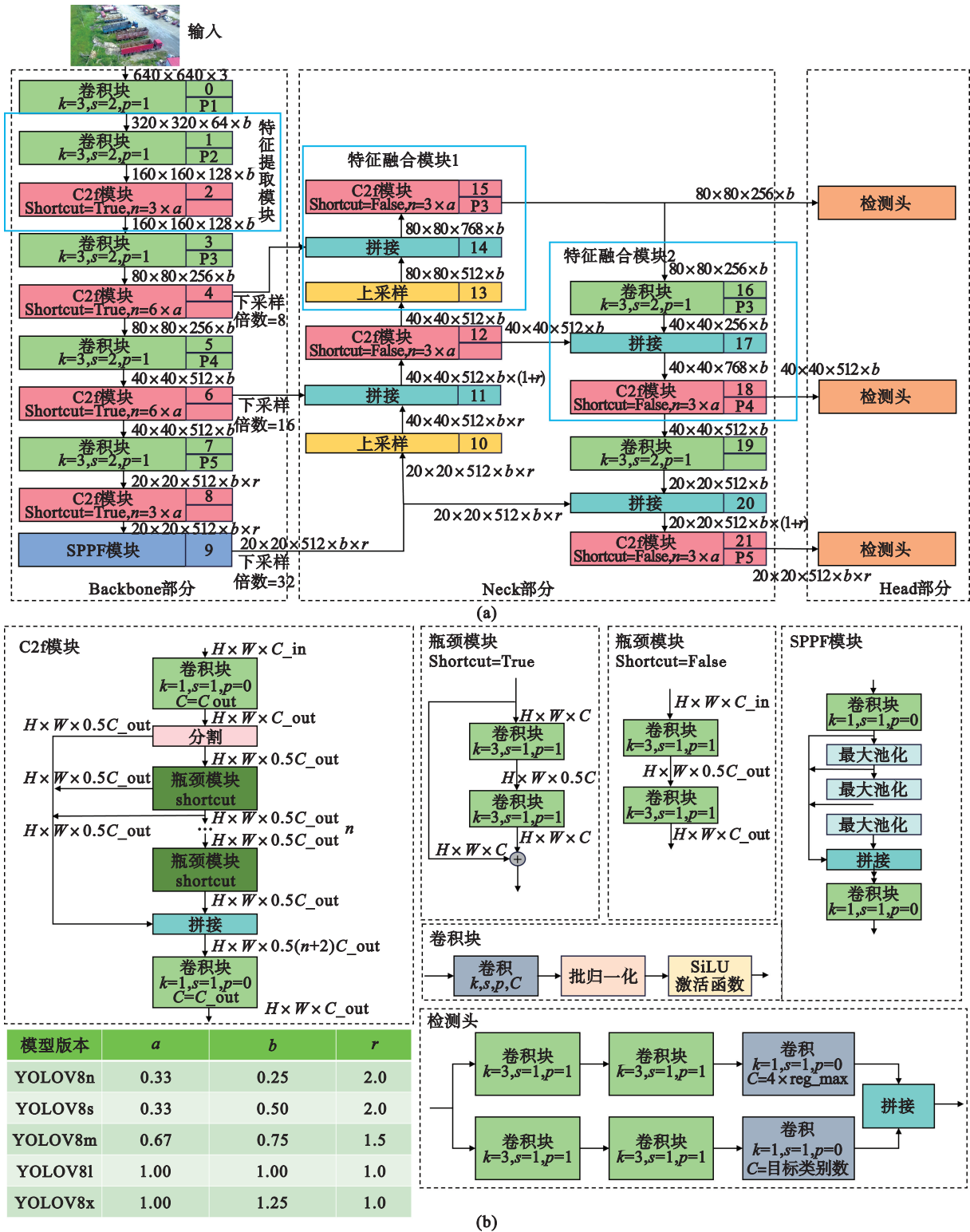


图1 YOLOv8模型结构图

Fig. 1 YOLOv8 model architecture diagram

(a)—YOLOv8模型结构; (b)—模块细节.

YOLOv8 模型中超参数 Reg_max 代表输出特征图中预测框的上下左右 4 个边框到中心点距离的最大预测范围. YOLOv8 模型最大下采样倍数为 32 倍, Reg_max 值为 16, 表示预测框的上下左右 4 个边框到中心点距离的最大预测范围为 512, 即 16 个特征图单元所映射的实际距离. 本文改进模型的 Reg_max 值设置为 8, 原因如下: 一是无人机航拍图像包含的目标多为小目标, 小目标中心点到边界框的距离较小, 改进模型的超参数 Reg_max 无需设置过大; 二是超参数 Reg_max 的值与模型最后的输出通道数相关. 本文改进模型最大下采样倍数为 16 倍, 最小下采样倍数为 4 倍, 对应输出特征图尺寸为 160 像素×160 像素. 模型预测时会增加一定计算量, 为减少改进模型的计算量并提升检测速度, 将超参数 Reg_max 由 16 调整为 8.

1.2 改进 YOLOv8 模型 Backbone 部分

YOLOv8 模型 Backbone 部分的主要功能是从输入的图像中提取丰富的特征信息, 这些信息对于后续的特征融合、分类以及定位任务至关重要. 由于小目标尺寸较小, 关键特征信息较少, 模型提取的特征信息中包含大量背景冗余信息, 难以有效关注到有利于小目标检测识别的关键特征. 因此, 本文在 Backbone 部分引入 ELA 机制, 使模型在进行特征提取时能够以轻量级的方式实现对感兴趣区域的准确定位, 从而更加关注与小目标最相关的关键特征.

ELA 针对协调注意力 (coordinate attention, CA) 机制^[21] 涉及多个特征图在水平和垂直方向上的分离合并的复杂性, 以及 CA 机制中批量归一化影响模型泛化能力的问题, 进行了优化改进. ELA 机制首先在空间维度上采用条带池化^[22] 来获得水平和垂直方向上的特征向量, 使用窄的内核形状来捕获长程依赖性并防止不相关区域影响模型对目标的预测, 从而在各自方向上产生丰富的目标位置特征. 然后在每个方向上独立地处理上述特征向量以获得注意力预测, 再使用乘积运算将其组合, 确保模型对感兴趣区域的准确定位.

ELA 机制的具体实现方式如下: 第一步是通过条带池化来获得水平和垂直方向上的特征向量; 第二步是应用一维卷积分别对两个特征向量进行局部交互, 产生的特征向量经过分组归一化和非线性激活函数处理后, 生成两个方向的位置注意力预测, 并相乘得到最终的位置注意力. 式 (1)~式 (5) 为 ELA 机制的数学表达式.

$$Z_{t,g}(g) = \frac{1}{W} \sum_{0 \leq i < W} x_t(g, i), \quad (1)$$

$$Z_{t,k}(k) = \frac{1}{H} \sum_{0 \leq j < H} x_t(j, k), \quad (2)$$

$$y_g = \sigma \left(G_n \left(F_g(z_g) \right) \right), \quad (3)$$

$$y_k = \sigma \left(G_n \left(F_k(z_k) \right) \right), \quad (4)$$

$$S = x_t \times y_g \times y_k. \quad (5)$$

式中: x_t 为输入特征的第 t 通道的特征表示; $Z_{t,g}$ 为 x_t 中高度为 g 的特征经过水平方向条带池化后的输出特征; $Z_{t,k}$ 为 x_t 中宽度为 k 的特征经过垂直方向的条带池化后的输出特征; σ 为非线性激活函数; F_g 和 F_k 为一维卷积; G_n 为分组归一化; y_g 和 y_k 为水平和垂直方向上的注意力; S 为 ELA 的输出.

图 2 是 CA 和 ELA 的结构图. 图中, X 表示水平方向, 对应特征图的宽度; Y 表示垂直方向, 对应特征图的高度. ELA 在保持输入特征图通道维度不变的前提下, 以轻量级的方式实现了对输入特征图感兴趣区域的准确定位. ELA 有 4 种参数配置方案. 本文为了使改进模型在提升对小目标检测性能的同时保持模型的轻量化和检测速度, 采用如下参数配置: 一维卷积的卷积核大小为 5, 分组数为输入通道数的 1/8, 分组归一化的分组数为 16.

1.3 基于 YOLOv8 改进的检测模型

模型 Backbone 部分的特征提取能力和 Neck 部分的特征融合方式与模型的检测性能密切相关. 本文对 YOLOv8 基础模型的这两部分进行改进以提升模型的检测性能. 改进后的模型总体结构如图 3 所示, 改进 1 为图 3 中红色矩形框和蓝色框标注的部分 (即对 YOLOv8 的 Neck 部分进行改进); 改进 2 为图 3 中橙色矩形框标注的部分 (即对 YOLOv8 的 Backbone 部分的改进).

2 实验设计及结果分析

2.1 实验数据

VisDrone2019 数据集^[23] 是由天津大学收集的开源数据集, 是研究小目标检测问题的理想基准. 该数据集拍摄场景多样, 图像分辨率高, 背景复杂, 目标尺寸小, 共包含行人、人群、自行车、汽车、厢式货车、卡车、三轮车、带棚三轮车、公交车和摩托车 10 个类别. VisDrone2019 数据集的划分如表 1 所示. 本文使用 training, validation 和 test-dev 子集进行实验, 其中 training 子集用于模型训

练, validation子集用于模型训练过程中的验证, test-dev子集用于模型性能的最终测试.

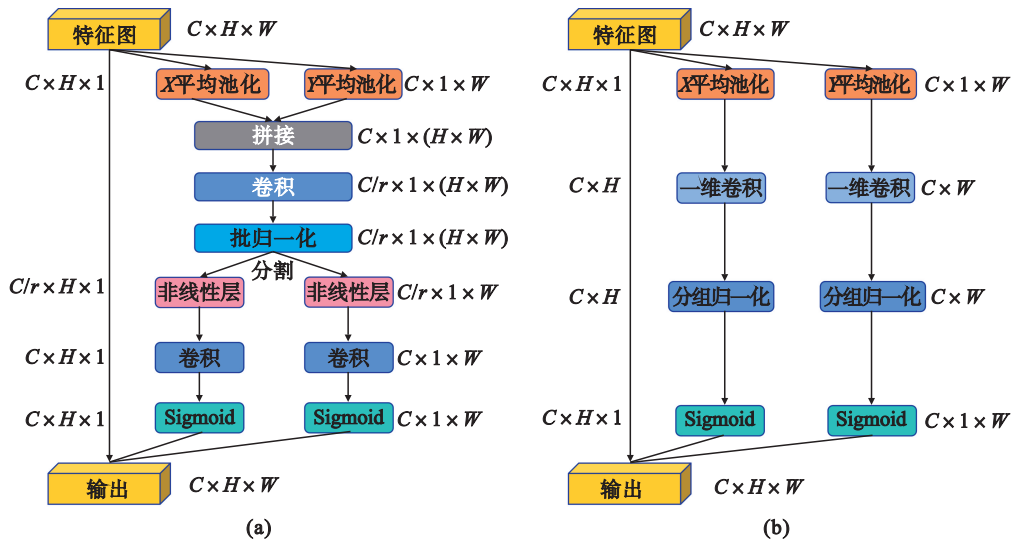


图 2 CA和ELA结构图

Fig. 2 CA and ELA structure diagram

(a)—CA 结构图; (b)—ELA 结构图.

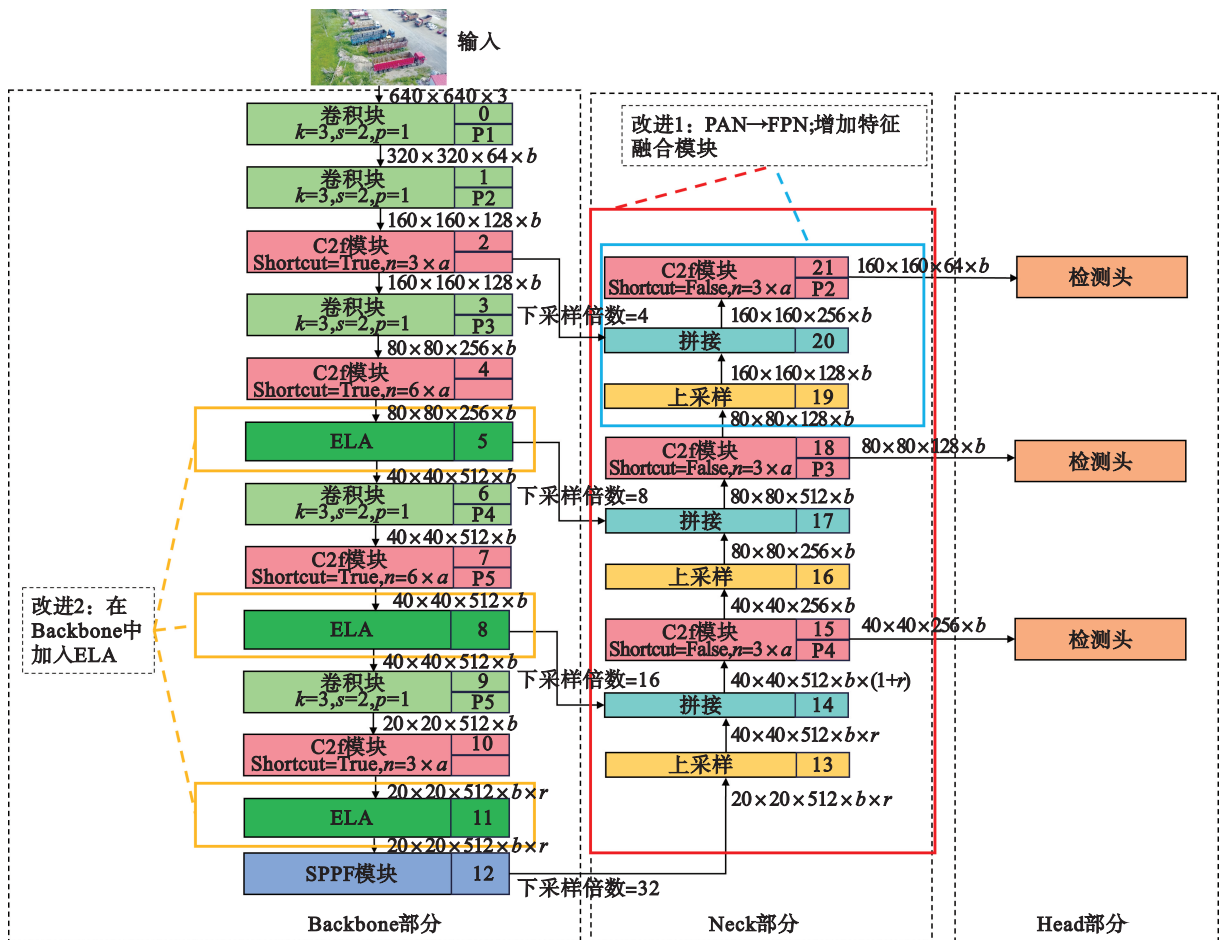


图 3 改进 Backbone和 Neck部分的YOLOv8模型结构图

Fig. 3 YOLOv8 model structure diagram for improved Backbone and Neck sections

2.2 评价指标

在目标检测任务中,真正例(true positives, TP)为模型正确检测到的目标物体数量,假正例

(false positives, FP)为模型错误将背景区域预测为目标物体的数量,真负例(true negatives, TN)为模型正确地将背景(即非目标物体区域)识别

为负样本,假负例(false negatives, FN)为模型未能检测到的目标物体数量. TN 的数量通常非常大,且对于评估模型性能并没有直接意义,因此目标检测中主要关注 TP, FP 和 FN, 这些结果直接反映了目标检测模型对目标物体的检测能力.

表 1 VisDrone2019数据集
Table 1 VisDrone2019 dataset

子集名称	training	validation	test-dev
图像数量	6 471	584	1 610

精确率(precision, P), 又称为“查准率”, 是指模型正确检测到的目标物体数量占模型检测到的所有目标数量的比例. 精确率的计算式为

$$P = \frac{TP}{TP + FP}. \quad (6)$$

召回率(recall, R), 又称为“查全率”, 是指模型正确检测到的目标物体数量占所有目标物体数量的比例, 是反映目标物体漏检情况的指标. 召回率的计算式为

$$R = \frac{TP}{TP + FN}. \quad (7)$$

精确率和召回率在一定程度上是矛盾的, 通常一个值较高时, 另一个值往往偏低. 以 P 为纵坐标、 R 为横坐标绘制 $P-R$ 曲线, 该曲线与坐标轴所围成的区域面积值即为平均精确率(average precision, AP).

平均精确率均值(mean average precision, mAP)是评估所有类别 AP 的指标^[24]. mAP 是将多个类别的 AP 进行加和求平均值, mAP 值越大表明模型性能越好. $mAP_{0.5}$ 是在交并比(intersection over union, IoU) 阈值为 0.5 时预测结果的 mAP 值, 若一个目标检测任务中共有 M 个类别, AP_i 表示第 i 个类别的 AP. mAP 的计算式为

$$mAP = \frac{1}{M} \sum_{i=1}^M AP_i. \quad (8)$$

检测速度用每秒帧率(frames per second, FPS)表示, 反映了模型每秒钟能够处理的图像帧数, 硬件条件对该指标会有较大影响.

浮点运算次数(floating point operations, FLOPs)代表了模型在处理数据时需要进行多少次浮点数的加、减、乘、除等基本运算. GFLOPs 指模型需要进行 10^9 次浮点运算, 是衡量模型计算复杂度的一个重要参考.

2.3 实验环境配置

本文实验环境搭建于高性能计算平台, 表 2 为所有实验的环境配置.

2.4 实验结果与分析

为了满足不同应用场景的需求, YOLOv8 模型提供了 5 种不同版本的配置, 即 tiny 版 YOLOv8n, small 版 YOLOv8s, medium 版 YOLOv8m, large 版 YOLOv8l 和 x-large 版 YOLOv8x. YOLOv8 模型不同版本配置的主要区别在于网络规模、计算复杂度、推理速度以及准确度等方面. 各版本对应的网络宽度和深度等参数设置如表 3 所示.

表 2 环境配置
Table 2 Environment configuration

环境	型号
GPU	NVIDIA GeForce RTX 4090
内存	24 GB
操作系统	Windows
系统位数	64
编译器	Pycharm 2023
CUDA	CUDA11. 8
编译环境	Python3. 9. 19
深度学习框架	PyTorch2. 2. 2

表 3 YOLOv8 不同版本配置参数设置
Table 3 Different version parameters setting of YOLOv8

YOLOv8 版本	深度倍数	宽度倍数	比率
YOLOv8n	0. 33	0. 25	2. 0
YOLOv8s	0. 33	0. 50	2. 0
YOLOv8m	0. 67	0. 75	1. 5
YOLOv8l	1. 00	1. 00	1. 0
YOLOv8x	1. 00	1. 25	1. 0

考虑到只有边缘设备才能在无人机上实现实时目标检测和推理, 这种局限性要求模型参数量少、占用内存少且推理时间短. 因此, YOLOv8s 常被用作改进和推广的基准模型. 本文以 YOLOv8s 模型作为基准模型进行改进. 为保证模型检测性能的公平性和可比性, 本文所有实验在训练时均未使用任何预训练权重.

本文将图像样本大小归一化为 $640 \text{ 像素} \times 640 \text{ 像素}$, 该尺寸可使模型在部署至边缘设备时保留足够的有效信息. 本文改进模型的最小下采样倍数为 4 倍, 其检测头较基础模型会输出更多预测结果, 故需更多的后处理时间. 为平衡检测速度与性能, 模型验证阶段将置信度阈值设为 0.005. 实验训练的关键参数设置见表 4.

2.4.1 消融实验

本文提出的检测模型主要对 YOLOv8s 的

Neck 部分和 Backbone 部分进行改进.为了系统分析各改进模型的性能情况,本文定量分析了基准模型 YOLOv8s、改进模型 1(改进 Neck 部分)、改进模型 2(改进 Neck 和 Backbone 部分)的评价指标变化情况.各模型在 test-dev 子集上的实验结果如表 5 所示.

表 4 训练参数设置

参数	数值
训练轮数	250
批次大小	8
图像尺寸/像素	640×640
交并比	0.7
初始学习率	1×10^{-2}
最终学习率	1×10^{-4}
数据增强	1.0
动量	0.937

表 5 的实验数据表明,相较于 YOLOv8s,本文的改进模型在高效降低模型参数量和规模的同时,检测性能与检测速度均有所提升.改进模

表 5 消融实验结果

模型	P/%	R/%	mAP _{0.5} /%	参数量×10 ⁻⁶	检测速度/FPS	计算量/GFLOPs	模型规模/MB
YOLOv8s	43.9	33.4	32.1	11.13	385	28.5	21.4
改进模型 1	43.8	34.8	33.1(↑3%)	5.51(↓50%)	400(↑4%)	17.1(↓40%)	10.8(↓50%)
改进模型 2	44.6	34.6	33.4(↑4%)	5.55(↓50%)	400(↑4%)	17.2(↓40%)	10.8(↓50%)

注:↑和↓分别表示提高和降低.

树莓派 4B 是典型的低成本、资源受限的边缘设备,NVIDIA Jetson 系列则是具有强大图形处理能力和深度学习加速能力的高性能边缘设备.本文提出的改进模型在规模和参数量方面完全符合树莓派 4B 的存储和内存承载能力,同时能在 NVIDIA Jetson 等高性能边缘人工智能平台上发挥卓越性能,可满足无人机视觉等对实时性和精度要求较高的边缘计算场景需求.

为了更直观地表达本文改进模型对小目标检测性能的提升,图 4a 和图 4b 展示了基础模型 YOLOv8s 和本文改进模型在 test-dev 子集上部分图像的检测结果对比.从图 4 第 1 行中的蓝色方框可以看出,改进模型减少了对小目标的错误检测;黄色方框显示改进模型对小目标的检测性能更好,漏检率更低.图 4 第 2 行和第 3 行的检测结果对比表明,改进模型在多种环境下的图像检测性能均优于 YOLOv8s,具有更好的鲁棒性.图

型 1 在 test-dev 子集上的实验结果显示:模型参数量和规模较 YOLOv8s 降低 50%,计算复杂度降低 40%,mAP_{0.5} 提升 3%,检测速度提升 4%.实验结果表明,对 YOLOv8s 的颈部结构的改进使模型参数量和规模减半的同时,提高了对小目标检测的召回率与检测速度.改进模型 2 是在对 YOLOv8s 的 Neck 部分改进的基础上,进一步对 Backbone 部分改进后的模型.本文将 ELA 集成在模型的 Backbone 部分,以强化模型对特征图中目标区域的精准定位能力,进而提升对小目标的检测性能.改进模型 2 的 mAP_{0.5} 相较于 YOLOv8s 提升了 4%,且模型参数量和计算复杂度与改进模型 1 相比有极少增加.本文提出的改进模型 2 参数量为 5.55×10^6 ,模型规模为 10.8 MB,计算量为 17.2 GFLOPs;模型在 test-dev 子集上的 mAP_{0.5} 为 33.4%,检测速度为 400 FPS.实验验证表明,本文的改进模型在有效提升小目标检测性能和检测速度的同时,能够更方便地部署在边缘设备上,实现对无人机航拍小目标实时高效检测.

4 第 4 行的检测结果对比进一步说明,改进模型对多个类别都具有更好的检测能力.综上,图 4 直观地证明了本文改进模型能够实现对小目标更准确、更全面的检测识别.

2.4.2 与其他模型对比实验

为了验证改进模型综合性能的优势,将本文的改进模型与 YOLOv5s 模型、YOLOv9c 模型以及基础模型 YOLOv8s 进行实验对比,在 test-dev 子集上的结果如表 6 所示.

由表 6 可看出,YOLOv5s 模型与 YOLOv8s 模型相比,模型参数量、规模以及计算复杂度更小,但模型检测精度 mAP_{0.5} 和检测速度 FPS 均有显著降低.2024 年提出的 YOLOv9 模型已开源的配置为 YOLOv9c 和 YOLOv9e,本文对 YOLOv9c 进行实验,结果表明其 mAP_{0.5} 相较 YOLOv8s 提升了 13%,但模型检测速度大幅下降了 34%.YOLOv9c 模型的参数量、计算复杂度以及模型规模分别约为 YOLOv8s 的 2 倍、4 倍以及 2 倍,约为

本文改进模型的4倍、6倍以及5倍,不适合部署在边缘设备上来实现对无人机航拍小目标的检测识别. 综上,实验结果表明本文的改进模型综

合性能较优,在检测精度、速度和模型部署难度方面均满足实际应用场景的需求,具有较高的鲁棒性和实用性.



图4 模型改进前后在test-dev子集的检测结果对比

Fig. 4 Comparison of detection results in test-dev subset before and after model improvement

(a)—YOLOv8s检测结果示例; (b)—改进模型检测结果示例.

表6 不同模型结果比较

Table 6 Comparison of results from different models

模型	P/%	R/%	mAP _{0.5} /%	参数量×10 ⁻⁶	检测速度/FPS	计算量/GFLOPs	模型规模/MB
YOLOv8s	43.9	33.4	32.1	11.13	385	28.5	21.4
YOLOv5s	37.7	30.5	27(↓16%)	7.04(↓37%)	238(↓38%)	15.8(↓45%)	13.7(↓36%)
YOLOv9c	50.3	37.4	36.4(↑13%)	25.33(↑128%)	256(↓34%)	102.4(↑259%)	49.2(↑130%)
本文模型	44.6	34.6	33.4(↑4%)	5.55(↓50%)	400(↑4%)	17.2(↓40%)	10.8(↓50%)

3 结 语

针对无人机航拍图像中小目标检测所面临的误检与漏检问题,以及无人机实时检测应用对

模型高效性和轻量化的需求,本文提出了一种基于YOLOv8改进的小目标检测模型. 通过将Neck部分简化为FPN结构并增加特征融合模块,本文的改进模型在有效降低模型参数量和规模大小的同时,能够更加充分地利用浅层网络提取的细

节信息,为Head层提供有利于小目标检测的特征,从而降低模型的漏检率.同时,在Backbone部分集成ELA,进一步增强了模型对目标区域的精确定位能力,提升了对小目标的检测效果.实验结果表明,本文改进模型在高效降低模型参数数量和规模大小的同时,提升了模型的检测性能和检测速度.未来将继续深入研究小目标检测技术,探索更多有效的模型改进方法,以推动无人机航拍技术的进一步发展和应用.

参考文献:

- [1] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [J]. *Advances in Neural Information Processing Systems*, 2017, 30: 6000–6010.
- [2] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with Transformers [C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2020: 213–229.
- [3] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2014: 580–587.
- [4] Ren S 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*, 2016, 39(6): 1137–1149.
- [5] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, 2017: 936–944.
- [6] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, 2016: 779–788.
- [7] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector [C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2016: 21–37.
- [8] Farhadi A, Redmon J. YOLOv3: an incremental improvement [C]//2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, 2018: 1–6.
- [9] Glenn J. YOLOv5 release v7.0 [EB/OL] (2022–11–22) [2024–03–10]. <https://github.com/ultralytics/YOLOv5/releases/tag/v7.0>.
- [10] Law H, Deng J. CornerNet: detecting objects as paired keypoints [C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018: 765–781.
- [11] Tian Z, Shen C H, Chen H, et al. FCOS: fully convolutional one-stage object detection [C]// International Conference on Computer Vision (ICCV). Seoul, 2019: 9626–9635.
- [12] Glenn J. YOLOv8 release v8.1.0. [EB/OL]. (2024–01–10) [2024–03–10]. <https://github.com/ultralytics/ultralytics/releases/tag/v8.1.0>.
- [13] Wang C Y, Yeh I H, Liao H Y M. YOLOv9: learning what you want to learn using programmable gradient information [C]//Proceeding of the European Conference on Computer Vision (ECCV). Cham: Springer, 2024: 1–21.
- [14] Li C L, Yang T, Zhu S J, et al. Density map guided object detection in aerial images [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, 2020: 737–746.
- [15] 燕舒乐, 陈润宇, 蔡念, 等. 基于多跳深度网络的红外微小目标检测方法 [J]. *激光与光电子学进展*, 2024, 61(22): 2237008.
(Yan Shu-le, Chen Run-yu, Cai Nian, et al. Infrared small target detection method based on multi-hop depth network [J]. *Laser & Optoelectronics Progress*, 2024, 61(22): 2237008.)
- [16] 吴海斌, 张亚, 胡鹏. 面向无人机航拍图像小目标检测方法 [J]. *安徽工业大学学报(自然科学版)*, 2024, 41(1): 65–73.
(Wu Hai-bin, Zhang Ya, Hu Peng. A small target detection method for unmanned aerial vehicle aerial photography images [J]. *Journal of Anhui University of Technology (Natural Science)*, 2024, 41(1): 65–73.)
- [17] Meethal A, Granger E, Pedersoli M. Cascaded zoom-in detector for high resolution aerial images [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Vancouver, 2023: 2046–2055.
- [18] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation [C]//2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, 2018: 8759–8768.
- [19] Xu W, Wan Y. ELA: efficient local attention for deep convolutional neural networks [EB/OL]. (2024–03–02) [2024–04–10]. <https://arxiv.org/abs/2403.01123>.
- [20] Tan M X, Pang R M, Le Q V. EfficientDet: scalable and efficient object detection [C]//2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 10778–10787.
- [21] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design [C]//2021 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, 2021: 13708–13717.
- [22] Hou Q B, Zhang L, Cheng M M, et al. Strip pooling: rethinking spatial pooling for scene parsing [C]//2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 4002–4011.
- [23] Zhu P F, Wen L Y, Du D W, et al. Detection and tracking meet drones challenge [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 44(11): 7380–7399.
- [24] Zou Z X, Chen K Y, Shi Z W, et al. Object detection in 20 years: a survey [J]. *Proceedings of the IEEE*, 2023, 111(3): 257–276.