

基于 AMD-YOLOv8 的无人机图像小目标检测^①



杨树莹, 葛华勇

(东华大学 信息科学与技术学院, 上海 201620)

通信作者: 葛华勇, E-mail: gehuayong@dhu.edu.cn

摘要: 无人机航拍图像目标检测在实时监控、异常检测等领域具有重要应用, 但航拍图像中复杂的背景、目标尺度变化大和小目标比例高等问题增加了检测难度. 为此, 本文提出了一种改进的目标检测算法 AMD-YOLOv8. 首先, 该算法采用微小目标检测头替代大目标检测头, 有效保留小目标特征信息; 引入动态目标检测头 Dyhead, 通过对水平维度、空间维度以及通道维度进行细粒度的注意力调整, 显著提高了模型对目标细节特征的提取能力. 其次, 设计出多尺度全局注意力模块 MSGA, 利用不同卷积核运算和全局上下文布局, 增强对远距离小目标的检测能力. 然后, 提出了双向密集扩展特征金字塔网络 BDEFPN, 该网络通过扩展尺度和密集连接, 实现了高效的多尺度信息融合. 最后, 采用 LAMP 剪枝策略对模型进行轻量化处理, 通过自适应评估层间冗余连接并进行删除, 有效降低计算量并加快推理速度. 实验结果表明, 改进后的 AMD-YOLOv8 在 VisDrone2019 数据集上, 参数量比 YOLOv8n 减少 46.0%, *mAP50* 提升了 8.6%, FPS 达到 98.3 f/s; 在 UAVDT 数据集的测试结果进一步验证了该算法优越的泛化能力, 证明了其在无人机航拍图像检测中的有效性.

关键词: 小目标检测; YOLOv8; 多尺度特征提取; 特征金字塔; 多重注意力

引用格式: 杨树莹, 葛华勇. 基于 AMD-YOLOv8 的无人机图像小目标检测. 计算机系统应用, 2025, 34(10): 173-183. <http://www.c-s-a.org.cn/1003-3254/9994.html>

Small Target Detection in UAV Images Based on AMD-YOLOv8

YANG Shu-Ying, GE Hua-Yong

(College of Information Science and Technology, Donghua University, Shanghai 201620, China)

Abstract: Target detection for UAV aerial images is significant in the applications of real-time monitoring and anomaly detection. However, the detection process is complicated due to challenges such as complex backgrounds, large variations in target scales, and a high proportion of small targets. To address these issues, this study proposes an improved algorithm for small target detection, AMD-YOLOv8. First, the algorithm replaces the large-target detection head with a micro-target detection head to effectively preserve small target features. A dynamic target detection head (Dyhead) is introduced, which applies fine grained attention adjustments across the horizontal, spatial, and channel dimensions, significantly enhancing the model's ability to extract detailed features of the target. Second, a multi-scale global attention (MSGA) module is designed, utilizing various convolutional kernel operations and global context layouts to enhance the detection capability for distant small targets. Third, a bidirectional dense extended feature pyramid network (BDEFPN) is proposed to efficiently integrate multi-scale information through scale expansion and dense connections. Finally, the LAMP pruning strategy is applied for model lightweight by adaptively assessing and removing redundant connections between layers, effectively reducing the computational load and accelerating inference speed. Experimental results demonstrate that the improved AMD-YOLOv8 reduces parameters by 46.0% compared to YOLOv8n on the VisDrone2019 dataset

① 基金项目: 国家自然科学基金 (62372100)

收稿时间: 2025-03-13; 修改时间: 2025-05-07; 采用时间: 2025-05-26; csa 在线出版时间: 2025-09-03

CNKI 网络首发时间: 2025-09-04

while achieving an 8.6% increase in $mAP50$ and an FPS of 98.3 f/s. Additionally, test results of the UAVDT dataset further confirm the algorithm's superior generalization ability, validating its effectiveness in the detection of UAV aerial images.

Key words: small target detection; YOLOv8; multi-scale feature extraction; feature pyramid; multiple attention

随着目标检测技术和无人机 (unmanned aerial vehicle, UAV) 技术的迅速发展, 基于无人机视角的目标检测已成为计算机视觉领域的关键研究方向, 并广泛应用于智能交通^[1]、森林火灾监测^[2]、军事行动^[3]、输电线异常检测^[4]等领域. 然而, 航拍图像受到飞行高度和拍摄角度等因素的影响, 具有复杂背景、小目标占比高以及目标尺度变化显著等特点, 这导致通用目标检测算法在无人机图像中的检测性能显著下降^[5]. 因此, 亟需提出适用于无人机航拍图像的小目标检测算法.

近年来, 研究者们提出了多种提升航拍图像目标检测性能的方法. Wang 等^[6]通过将三预测头改为双预测头并引入瓶颈模块, 增强了小目标特征提取能力, 但在复杂环境下对密集小目标的识别仍有限. Jiang 等^[7]提出多尺度特征提取模块, 通过不同尺度的卷积提取丰富的多尺度特征信息, 显著提高小目标的检测精度, 但网络加深导致部分信息丢失. Li 等^[8]使用 GhostblockV2 替代部分 C2f 模块, 在减小模型参数的同时缓解了长距离信息丢失, 但受限于骨干网络结构, 对密集小目标的提取仍存在不足.

针对当前无人机目标检测模型中小目标检测性能差、多尺度信息提取不足、深浅层信息利用不充分等问题, 本文提出一种基于改进 YOLOv8n 的无人机航拍图像小目标检测算法 AMD-YOLOv8, 主要贡献如下.

(1) 优化检测头设计: 移除大目标检测头以减少参数, 新增微小目标检测头强化小目标特征提取; 同时引入具多重注意力机制的动态检测头 Dyhead (dynamic head)^[9], 提升航拍图像中的特征提取与识别能力, 降低误检率.

(2) 设计多尺度全局注意力模块 (multi-scale global attention module, MSGA): 此模块通过设计 3 个分支卷积操作和全局上下文提取器, 以有效提取多尺度特征信息, 降低漏检率.

(3) 提出双向密集扩展特征金字塔网络 (bidirectional dense extended feature pyramid network, BDEFPN):

通过扩展尺度、创新的密集连接和尺度序列特征融合模块 (scale sequence feature fusion module, SSFF)^[10], 以实现浅层和深层特征信息的全面整合, 有效提升模型对不同尺度目标的检测性能.

(4) 采用 LAMP (layer-adaptive magnitude-based pruning)^[11]剪枝策略对模型进行优化, 通过动态评估并去除冗余连接, 有效减少模型计算量并加快推理速度.

1 YOLOv8 概述

YOLOv8 是 YOLO^[12]系列的代表算法之一, 由输入端、骨干网络、颈部网络和检测头组成. 输入端通过 Mosaic 等方法对数据进行预处理, 以增加数据集的多样性. 骨干网络由 Conv、C2f 和 SPPF 模块组成, Conv 模块集成了卷积、归一化和 SiLU 操作, C2f 模块通过多次 3×3 卷积捕获丰富的梯度信息, SPPF 模块通过多个 5×5 最大池化层扩大感受野并融合多尺度特征. 颈部网络借鉴了路径聚合网络 (path aggregation network, PAN)^[13]和特征金字塔网络 (feature pyramid network, FPN)^[14]的设计, 整合不同层级的特征信息. 检测头采用解耦设计, 分别提取目标的位置信息和类别信息, 以提升检测精度.

相较同系列算法, YOLOv8 在结构设计和检测性能上更具优势. 其中, YOLOv8n 作为最轻量版本, 具有最小的深度与宽度、最低的计算开销和最快的运行速度. 鉴于无人机航拍对检测精度和轻量化的需求, 本文选用 YOLOv8n 作为改进基础.

2 AMD-YOLOv8 检测算法

本文在 YOLOv8n 基础上提出了一种面向无人机航拍图像的小目标检测算法 AMD-YOLOv8, 结构如图 1 所示. 该算法在 4 个方面进行改进: 检测头移除大目标检测头, 新增微小目标检测头, 并引入 Dyhead 以提升检测性能; 骨干网络加入 MSGA 模块, 增强特征提取能力; 颈部结构优化为 BDEFPN, 提升多尺度特征融合效果. 此外, 采用 LAMP 剪枝策略实现模型轻量化, 加快推理速度.

其中, $Conv_{1 \times 1}$ 表示核为 1×1 的卷积模块, 每个卷积模块由卷积层、批处理归一化和 SiLU 激活函数构成, 批

处理归一化用于加速训练过程并增强模型的稳定性, SiLU 激活函数则能够引入非线性变换以提升表达能力。

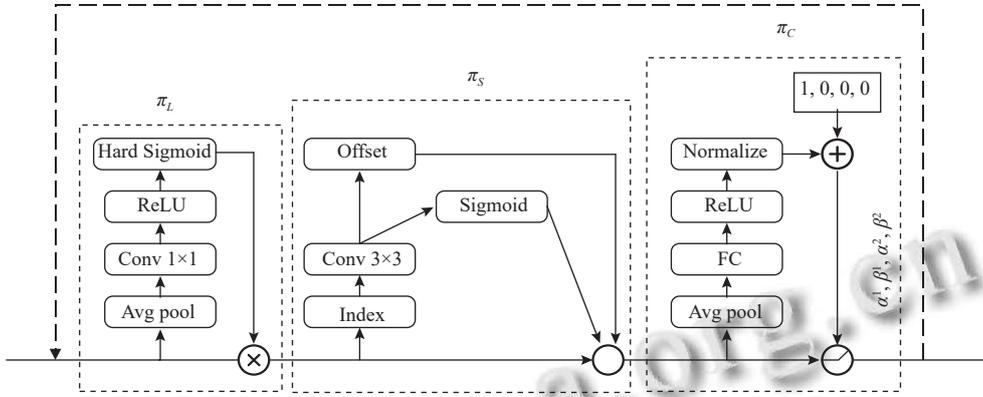


图2 Dyhead 结构

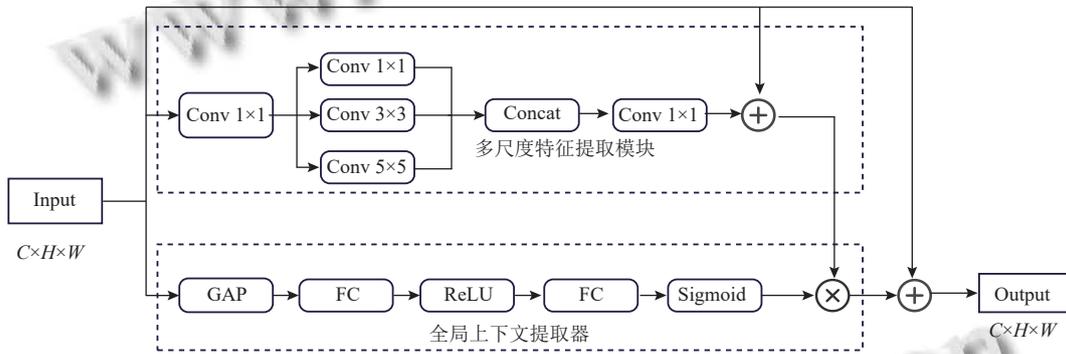


图3 MSGA 结构

随后, 设计 3 个不同尺度的卷积路径 (1×1 、 3×3 、 5×5), 以模拟多种感受野并捕捉局部与中远距离的上下文特征. 不同尺度分支输出后, 在通道维度进行拼接; 考虑到多尺度拼接可能导致信息冗余, 模块在拼接后进一步使用 1×1 卷积进行通道融合和语义统一, 压缩冗余通道、提升判别性, 从而实现高效的信息整合与表达压缩, 如式 (3) 和式 (4) 所示:

$$Y_c = [Conv_{1 \times 1}(X), Conv_{3 \times 3}(X), Conv_{5 \times 5}(X)] \quad (3)$$

$$Y_1 = Conv_{1 \times 1}(Y_c) \quad (4)$$

其中, $Conv_{i \times i}$ 表示核为 $i \times i$ 的卷积模块, $[]$ 表示 Concat 操作, Y_c 表示经过拼接得到的特征图。

为弥补局部卷积感受野的局限, MSGA 引入轻量的全局上下文提取器: 首先, 通过全局平均池化 (global average pooling, GAP) 汇聚整图语义特征, 得到压缩表达; 接着, 通过双层感知机结构建模通道间的依赖关系,

第 1 层使用 ReLU 激活提升非线性建模能力, 第 2 层采用 Sigmoid 激活用于生成注意力权重. 该全局权重随后作用于融合后的多尺度特征图上, 以实现全局语义引导下的通道级特征重标定, 从而强化关键目标区域、抑制冗余与背景干扰, 如式 (5) 所示:

$$Y_2 = Sig(fc(Re(fc(gap(x)))))) \quad (5)$$

其中, gap 表示全局平均池化, fc 表示全连接, $Re()$ 表示 ReLU 函数, $Sig()$ 表示 Sigmoid 函数。

为提升模块内部的信息流动能力, 防止因特征变换过多而导致的语义信息衰减, MSGA 同时引入了局部残差连接 (local residual learning, LRL) 和全局残差连 (global residual learning, GRL). 其中, LRL 将输入特征与多尺度特征提取输出相加, 实现局部范围内的结构优化与特征保留; GRL 则连接输入特征与全局上下文建模输出, 实现更长路径的信息补充, 进一步提升语

义传导的连续性与稳定性,如式(6)所示:

$$Y_{final} = X + Y_2 \otimes (Y_1 + X) \quad (6)$$

其中, \otimes 表示逐元素相乘, X 表示输入特征, Y_1 表示多尺度特征提取模块的输出映射, Y_2 表示全局上下文提取器的输出特征, Y_{final} 表示MSGa模块的输出特征。

2.3 颈部网络的优化

在航拍图像中,受目标距离镜头远近的影响,检测目标尺度变化显著。YOLOv8的颈部网络使用FPN-PAN来提取和聚合特征,但这种简单的上下和横向连接未能充分利用深层与浅层信息,导致特征融合效果不佳。为更有效地整合多层次目标信息,本文提出了一

种改进的颈部网络BDEFPN。

BDEFPN主要在以下3个方面进行了改进,结构如图4所示。首先,在YOLOv8n的基础上,扩展了颈部网络的尺度,在每个尺度的特征层上增加多个CBS和C2f,以获得更丰富的深层语义信息。其次,BDEFPN引入多条跳跃连接和跨层连接:同层跳跃连接减少因扩展颈部网络带来的特征损失,而跨层连接可促进不同尺度特征的信息流动。通过这两种连接方式,BDEFPN可确保不同尺度特征的有效融合和信息传递,充分利用来自不同层次的特征信息。最后,使用SSFF模块以充分整合骨干网络中小目标的浅层位置信息。

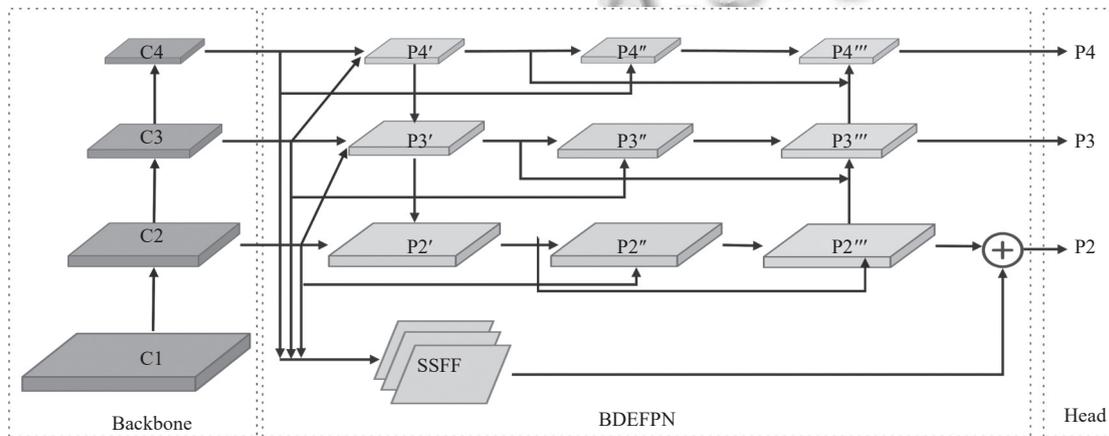


图4 BDEFPN结构

SSFF聚合了骨干网络中不同层级的特征,其结构如图5所示。对于并行输入的P2、P3和P4级特征,SSFF首先通过 1×1 卷积统一各分支的通道数。考虑到P2特征图包含丰富的小目标信息,采用最近邻插值将P3和P4上采样至与P2相同的尺寸。随后,将每个特

征图从3D张量扩展为4D,并在深度维度进行拼接,形成供后续卷积操作的融合特征。最后,SSFF通过3D卷积、3D批量归一化和SiLU激活函数提取尺度序列特征,并将其与P2的深层语义信息融合,传递至检测头进行后续处理。

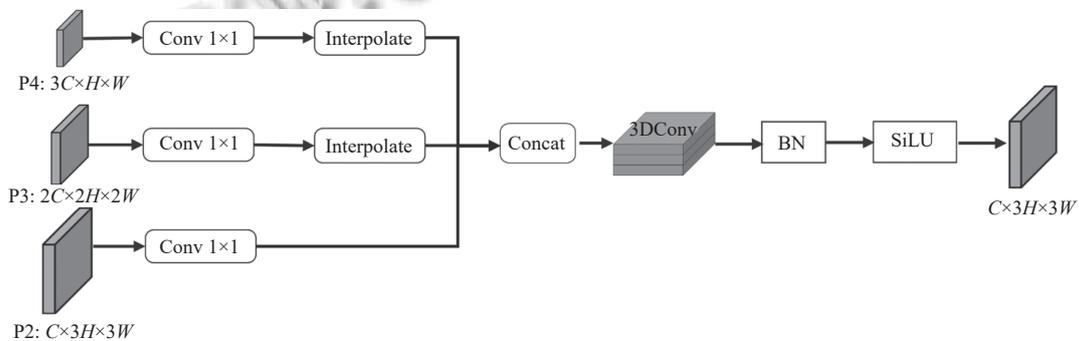


图5 SSFF结构

扩展的颈部结构与SSFF模块协同作用,使BDEFPN能够有效解决航拍图像中目标尺度不一致的问题。

2.4 LAMP Pruning

在无人机航拍的小目标检测任务中,嵌入式设备

通常受限于计算能力和存储空间,对模型的轻量化与实时性提出了较高要求。改进后的YOLOv8网络虽在检测精度方面有所提升,但相应增加了参数量和计算开销。为此,本文引入LAMP模型剪枝方法,通过全局评分机制有效识别冗余连接并进行删除,在保持检测性能的同时显著压缩模型规模,加快推理速度。

LAMP剪枝方法的核心思想是通过评估每个连接的权重幅值,衡量其对模型性能的贡献,从而识别出冗余的连接进行剪除。计算每个连接的LAMP评分公式如下:

$$score(u, W) = \frac{(W[u])^2}{\sum_{v>u} (W[v])^2} \quad (7)$$

其中, $score(u, W)$ 表示LAMP评分, $W[u]$ 表示由索引 u 映射的权重项, $W[v]$ 表示由索引 v 映射的权重项。在剪枝过程中, LAMP方法首先计算所有连接的评分,并按照评分高低对连接进行排序。评分较高的连接对模型的性能贡献较大,因此会被保留;而评分较低的连接则可被剪除。LAMP的实质是在全局范围内对神经网络进行剪枝,综合考虑每个连接的重要性,实现更精确的剪枝,能够在提升推理效率的同时保持较高的检测精度。

3 实验结果与分析

本文实验是基于Windows 11系统,使用PyTorch框架,显卡采用NVIDIA GeForce RTX 4060,部署环境为Python 3.9,加速计算架构是CUDA 11.8。训练参数如表1所示。

表1 训练参数

参数名称	参数信息
图片尺寸 (imgsz)	640×640
批量大小 (batchsize)	2
训练轮次 (epochs)	200
权重衰减系数 (weight_decay)	0.0005
学习率动量 (momentum)	0.937
余弦退火参数 (lrf)	0.01
初始学习率 (lr0)	0.01

3.1 数据集概述

为评估AMD-YOLOv8在无人机航拍图像中小目标检测的有效性,本文采用了两个公开数据集VisDrone和UAVDT进行实验。其中,VisDrone数据集用于模型的改进与训练,而UAVDT数据集则用

于补充实验,以验证模型在不同数据集上的适应性和性能表现。

(1) VisDrone2019数据集是由不同型号的无人机在不同场景、天气和光照条件下拍摄,包含8629幅静态图像,其中6471张用于训练,548张用于验证,1610张用于测试。该数据集提供行人、汽车、自行车和三轮车等10种对象类别标签。数据集中小型车辆和行人等小目标密集,符合本文实验要求。

(2) UAVDT数据集涵盖汽车、卡车、公共汽车3种类别,提供超过8万个注释。因数据集规模庞大,本研究随机选取4000张图像按7:1:2的比例分为训练集、验证集和测试集,以补充验证模型的通用性。

3.2 评价指标

本实验采用 P (准确率)、 R (召回率)、 $mAP50$ 、 $mAP50:95$ 、Params (参数量) 和 GFLOPs (每秒浮点运算次数) 作为主要性能指标。其中, $mAP50$ 为IoU阈值为0.5时的平均精度, $mAP50:95$ 则为IoU在0.5–0.95间、步长为0.05的平均精度。 P 、 R 、 AP 和 mAP 的计算方法如式(8)–式(10)所示。

$$P = \frac{TP}{TP+FP} \quad (8)$$

$$R = \frac{TP}{TP+FN} \quad (9)$$

$$AP = \int_0^1 P(R) dR \quad (10)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP(i) \quad (11)$$

其中, TP 为正确预测的正样本数, FP 为误判为正样本的负样本数, FN 为误判为负样本的正样本数, n 为类别总数; $P(R)$ 表示精度-召回率曲线, mAP 为各类别 AP 的平均值。

3.3 MSGA 组件分析实验

为了验证MSGa模块中上下文建模机制的有效性,我们设计了对比实验,分别评估以下3种模型配置: ①在YOLOv8n基础上引入多尺度特征提取模块,但不包含全局上下文提取器和残差连接的模型(模型C1); ②在C1的基础上添加全局上下文提取器,但仍未引入残差连接(模型C2); ③在C2的基础上进一步加入残差连接,构成完整的MSGa模块。各模型的实验结果如表2所示。

表2 MSGA 组件分析实验 (%)

模型	$mAP50$	$mAP50:95$
YOLOv8n	26.0	14.6
C1	26.7	15.0
C2	26.9	15.1
MSGA	27.1	15.3

实验结果表明,增加多尺度提取模块的模型 C1 相比 YOLOv8n 精度显著提升, $mAP50$ 和 $mAP50:95$ 分别提高了 0.7% 和 0.4%,验证了多尺度特征提取的有效性.在此基础上,加入全局上下文提取器后(模型 C2),精度进一步提升, $mAP50$ 和 $mAP50:95$ 分别提高了 0.2% 和 0.1%,表明全局上下文信息有助于增强对远距离小目标的检测.进一步引入残差连接构成的完整 MSGA 模块,精度再次提升, $mAP50$ 和 $mAP50:95$ 分别达到了 27.1% 和 15.3%,验证了残差连接在上下文信息融合中的关键作用,它有效促进了信息流动,优化了特征表示,进一步增强了模型的检测性能.

3.4 注意力机制对比实验

为评估 MSGA 注意力机制在小目标检测中的有效性,在 VisDrone2019 数据集上将其与不同注意力机制进行对比实验,结果如表 3 所示.

表3 注意力机制对比实验 (%)

模型	$mAP50$	$mAP50:95$
YOLOv8n	26.0	14.6
+CBAM ^[15]	25.9	14.5
+CoordAtt ^[16]	25.7	14.2
+ECA ^[17]	26.2	14.5
+EMA ^[18]	26.0	14.6
+ACmix ^[19]	26.1	14.6
+MSGA	27.1	15.3

实验结果表明,加入 MSGA 注意力机制后,模型的 $mAP50$ 和 $mAP50:95$ 分别达到了 27.1% 和 15.3%.与 CBAM、CoordAtt 和 ECA 注意力机制相比,MSGA

的 $mAP50$ 分别提高了 1.2%、1.4% 和 0.9%,这是由于 MSGA 更加注重全局上下文信息的建模,能够在捕获丰富语义特征的同时,精确提取小目标的局部细节.与 EMA 和 ACmix 相比,MSGA 的 $mAP50$ 分别提高了 1.1% 和 1.0%.MSGA 采用多尺度信息提取与全局交互相结合的方式,精确保留并增强了小目标的空间与语义信息,提高了检测精度.

3.5 颈部结构对比实验

为验证 BDEFPN 在提升航拍图像检测精度方面的优势,在 VisDrone2019 数据集上对其与不同颈部网络进行对比实验,结果如表 4 所示.

表4 颈部结构对比实验 (%)

模型	$mAP50$	$mAP50:95$
YOLOv8n	26.0	14.6
+BiFPN ^[20]	26.0	14.7
+HSFPN ^[21]	24.8	13.7
+Slim-Neck ^[22]	25.2	14.0
+AFPN ^[23]	22.6	12.5
+BDEFPN	27.4	15.5

实验结果表明,将颈部网络替换为 BDEFPN 后,模型的 $mAP50$ 和 $mAP50:95$ 分别达到了 27.4% 和 15.5%.与 BiFPN 和 HSFPN 相比,BDEFPN 通过扩展尺度和创新连接增强了信息融合与细节保留能力, $mAP50$ 分别提高了 1.4% 和 2.6%.此外,BDEFPN 采用尺度序列特征融合模块,深入整合骨干网络的多层次信息,提升了多尺度特征的表达能力,相较 Slim-Neck 和 AFPN, $mAP50$ 分别提高了 2.2% 和 4.8%,改善了目标检测效果.

3.6 消融实验

为验证每个改进模块对于模型检测能力的提升,本文以 YOLOv8n 作为基准模型,分别在 VisDrone2019 数据集上进行消融实验,结果如表 5 所示.

表5 VisDrone2019 数据集上的消融实验

模型	P234+Dyhead	CSEL	BDEFPN	LAMP	P (%)	R (%)	$mAP50$ (%)	$mAP50:95$ (%)	Params (M)	GFLOPs	FPS (f/s)
A	—	—	—	—	37.3	28.4	26.0	14.6	3.01	8.1	137.8
B	√	—	—	—	45.2	34.6	33.1	18.9	1.46	19.5	65
C	—	√	—	—	38.7	29.4	27.1	15.3	3.26	9.9	129.0
D	—	—	√	—	38.4	29.9	27.4	15.5	3.73	10.0	123.3
E	√	√	—	—	46.1	35.0	34.0	19.6	1.78	22.8	57.2
F	√	—	√	—	45.7	35.9	34.6	19.8	1.85	22.6	57.9
G	—	√	√	—	39.7	31.1	28.6	16.3	3.98	11.8	110.5
H	√	√	√	—	46.4	36.5	35.0	20.0	2.03	24.6	49.8
I	√	√	√	√	45	36.0	34.6	19.6	1.62	14.8	98.3

基础模型 YOLOv8n (模型 A) 采用 P3–P5 层级输出特征,具有较高的推理效率.模型 B 将检测头下移至

浅层 (P2–P4) 并引入 Dyhead 模块,提升了小目标的感知能力和多尺度语义融合效果.实验表明,模型 B 显著

提高了小目标检测精度,但由于引入多维注意力与浅层特征,GFLOPs增至19.5,FPS降至65 f/s,适用于对精度要求较高的场景。

模型C在YOLOv8n的基础上加入了MSG模块,进一步提升了远距离目标的检测能力。实验结果表明,模型C在 $mAP50$ 和 $mAP50:95$ 上分别提升了1.1%和0.7%,尽管增加了结构复杂度,但GFLOPs仅为9.9,FPS达123.0 f/s,保持了良好的精度与效率平衡。

模型D通过BDEFPN改进了颈部网络结构,增强了网络信息融合能力。在仅对颈部结构进行调整的前提下,模型D在 $mAP50$ 和 $mAP50:95$ 上分别提升了1.4%和0.9%,GFLOPs为10.0,FPS达到99.6 f/s,验证了BDEFPN在低计算成本下的有效性。

为验证各改进策略的相互作用,本研究设计并测试了模型E、F、G。融合两种改进的模型表现更优,进一步验证了改进方法的协同增益效应。

模型H集成了上述改进,相比YOLOv8n, P 和 R 分别提升了9.1%和8.1%, $mAP50$ 和 $mAP50:95$ 分别

提高了9.0%和5.4%,并将参数量减少约32.5%。尽管GFLOPs增至24.6,但单张图像推理的FPS仍可达到49.8 f/s,满足边缘计算平台对实时性($FPS \geq 30$)和高精度的双重需求。

为进一步提升模型H的计算效率,进行了LAMP剪枝优化。剪枝后的模型I(AMD-YOLOv8)虽然精度略有下降,但 $mAP50$ 和 $mAP50:95$ 仍可达到34.6%和19.6%,同时GFLOPs降低至14.8,FPS提升至98.3 f/s,在保持较高精度的同时显著提高了推理效率,适合在资源受限的实时应用场景中部署。

为直观展示AMD-YOLOv8在航拍图像检测中的性能,本文对改进前后的模型进行了可视化对比。如图6所示,YOLOv8n在低光模糊环境中将垃圾桶误检为行人,并漏检多个行人目标;在高空视角下也遗漏了大量远距离目标。而经过网络架构优化和注意力增强的AMD-YOLOv8能准确地定位并识别了上述错检和漏检的目标,且检测框置信度更高,体现出更强的识别精度。

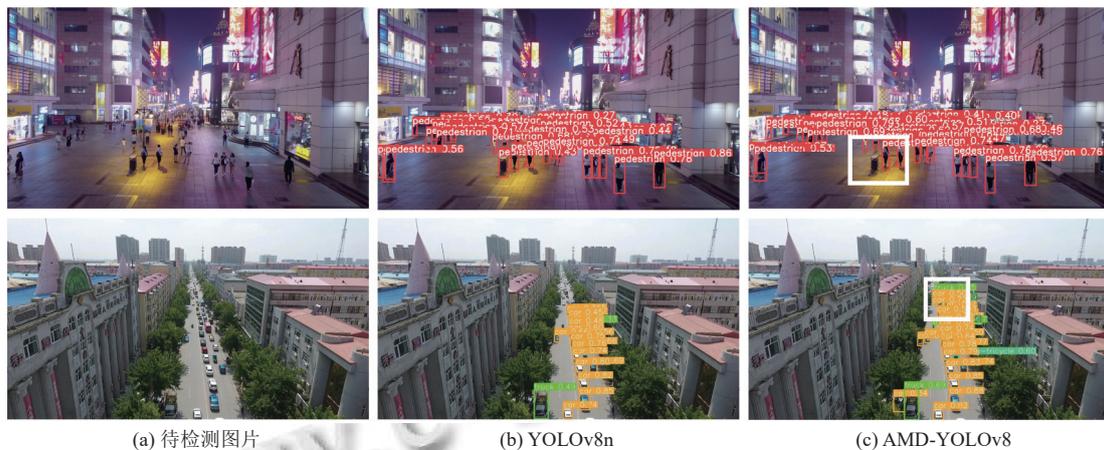


图6 改进前后检测效果对比图(白框标示的区域为改进部分)

为更全面地展示模型对目标区域的关注情况,采用梯度加权类别激活映射法(gradient-weighted class activation mapping, Grad-CAM)可视化改进前后模型的关注差异。Grad-CAM结合卷积层的梯度和特征图权重生成类激活图,能够高亮图像中重要的区域,直观展现模型的关注点,并评估其对目标和背景噪声的响应。如图7所示,YOLOv8n对小目标关注不足,且对远距离目标的敏感度较低。而AMD-YOLOv8则有效抑制了背景噪声,注意力更集中于目标中心,提升了对小目标及远距离目标的检测准确性。

3.7 对比实验

3.7.1 VisDrone2019数据集上的对比实验

为评估AMD-YOLOv8在航拍图像中的检测性能,本文在相同实验环境下,将其与目前的主流模型进行对比实验,结果如表6所示。

结果表明,CenterNet和SSD模型的规模较大,且其检测性能显著低于AMD-YOLOv8。与YOLOv5s、YOLOv6s、YOLOv8s、YOLOv9s和YOLOv10s相比,AMD-YOLOv8和 $mAP50$ 分别提高了3.7%、5.0%、3.0%、2.8%和2.9%, $mAP50:95$ 分别提高了2.0%、

2.6%、1.5%、1.3% 和 1.4%，参数量和计算量均为最小。此外，与 BD-YOLO、SC-YOLO、NDM-YOLOv8 以及 TPH-YOLOv5 这几个同样改进 YOLO 系列的模型相比，AMD-YOLOv8 在检测精度与轻量化上具备更大的优势。这主要源于 AMD-YOLOv8 对特征提取能力的优化。通过改进检测头，AMD-YOLOv8 能够更有

效地处理不同尺寸的目标，增强了对细节的捕捉能力。骨干网络中的 MSGA 模块，突破了传统 YOLO 系列依赖固定卷积感受野的局限性，使模型能够更深层次地挖掘目标特征。此外，BDEFPN 颈部网络优化了 YOLO 系列在特征融合方面的不足，使其更适应多尺度特征提取的需求。



图7 改进前后热力图对比图(白框标示的区域为改进部分)

表6 VisDrone2019 数据集上的对比实验

模型	$mAP50$ (%)	$mAP50:95$ (%)	Params (M)	GFLOPs
CenterNet	27.8	16.1	15.38	39.4
SSD	25.3	14.9	91.04	78.9
YOLOv5s	30.9	17.6	9.1	23.8
YOLOv6s	29.6	17.0	16.3	44.0
YOLOv8s	31.6	18.1	11.1	28.5
YOLOv9s ^[24]	31.8	18.3	7.2	26.7
YOLOv10s ^[25]	31.7	18.2	7.4	21.4
BD-YOLO ^[26]	32.6	17.9	7.9	27.6
SC-YOLO ^[27]	34.8	19.8	12.3	30.1
NDM-YOLOv8 ^[28]	31.9	17.8	3.4	16.4
TPH-YOLOv5 ^[29]	31.2	17.1	8.41	35.1
AMD-YOLOv8	34.6	19.6	1.62	14.8

3.7.2 UAVDT 数据集上的对比实验

为验证模型的泛化性，本文基于 UAVDT 数据集，开展了 AMD-YOLOv8 与 YOLO 系列模型的对比实验，结果如表 7 所示。

实验结果表明，AMD-YOLOv8 表现最佳， $mAP50$ 和 $mAP50:95$ 分别达到了 90.6% 和 70.7%。与 YOLOv8n 相比，AMD-YOLOv8 在 $mAP50$ 和 $mAP50:95$ 方面分别提升了 16.0% 和 14.0%，且参数量减少约 46.0%；相较于 YOLOv5s、YOLOv6s、YOLOv8s、YOLOv9s 和 YOLOv10s，AMD-YOLOv8 不仅在 $mAP50$ 和

$mAP50:95$ 方面取得了更高的检测精度，而且参数量和计算量最低，显著减少了模型冗余。

表7 UAVDT 数据集上的对比实验

模型	$mAP50$	$mAP50:95$	Params (M)	GFLOPs
YOLOv5s	78.2	61.8	9.12	23.8
YOLOv6s	74.7	57.6	16.3	44.0
YOLOv8n	74.6	56.7	3.0	8.1
YOLOv8s	79.5	64.1	11.1	28.5
YOLOv9s	75.3	60.4	7.2	26.7
YOLOv10s	80.6	64.8	7.4	21.4
AMD-YOLOv8	90.6	70.7	1.62	14.8

上述模型的 PR 曲线对比如图 8 所示。与其他 YOLO 系列模型相比，AMD-YOLOv8 在检测头、注意力机制和颈部结构的优化使得其在低召回率时仍能保持较高精确率，曲线面积更大，表明它能够在不同决策阈值下更好地平衡精确率与召回率，展现出更优的检测性能。

综上所述，AMD-YOLOv8 在 UAVDT 数据集上的表现验证了其优异的泛化能力，在无人机航拍图像检测领域具有广泛的应用潜力。

4 结束语

针对无人机航拍图像中因环境复杂、目标尺度变化大及小目标细节模糊导致的漏检与误检问题，本文

提出了一种基于 YOLOv8n 的改进算法——AMD-YOLOv8。该算法采用了 P2、P3 和 P4 这 3 种尺度的检测头,以增强模型对密集小目标的关注;融合 Dyhead 改进检测头结构,提升模型在复杂场景下的信息提取效果,从而有效降低误检与漏检率;在骨干网络中引入 MSGA 模块,以提取更丰富的小目标细节信息,进一步提升检测精度;同时,设计 BDEFPN 颈部结构,强化多尺度特征的融合能力;此外,结合 LAMP 剪枝策略,有效去除冗余连接,降低计算复杂度并提升推理效率。在 VisDrone2019 和 UAVDT 数据集上的实验结果表明,AMD-YOLOv8 显著提高了无人机航拍图像的检测精度,相较于其他主流模型,表现出更优的性能。未来的研究将致力于进一步优化小目标检测算法,并探索其在无人机平台上的实际应用。

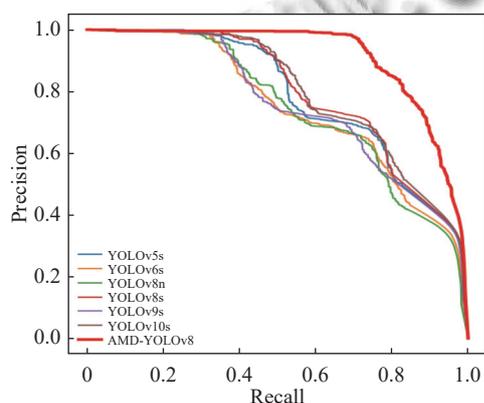


图8 PR曲线对比图

参考文献

- Jiang CC, Ren HZ, Ye X, *et al.* Object detection from UAV thermal infrared images and videos using YOLO models. *International Journal of Applied Earth Observation and Geoinformation*, 2022, 112: 102912. [doi: [10.1016/j.jag.2022.102912](https://doi.org/10.1016/j.jag.2022.102912)]
- Zhang L, Wang MY, Fu YJ, *et al.* A forest fire recognition method using UAV images based on transfer learning. *Forests*, 2022, 13(7): 975. [doi: [10.3390/f13070975](https://doi.org/10.3390/f13070975)]
- Chen CL, Zheng ZY, Xu TY, *et al.* YOLO-based UAV technology: A review of the research and its applications. *Drones*, 2023, 7(3): 190. [doi: [10.3390/drones7030190](https://doi.org/10.3390/drones7030190)]
- Bao WX, Ren YX, Wang N, *et al.* Detection of abnormal vibration dampers on transmission lines in UAV remote sensing images with PMA-YOLO. *Remote Sensing*, 2021, 13(20): 4134. [doi: [10.3390/rs13204134](https://doi.org/10.3390/rs13204134)]
- Cheng N, Wu S, Wang XC, *et al.* AI for UAV-assisted IoT applications: A comprehensive review. *IEEE Internet of Things Journal*, 2023, 10(16): 14438–14461. [doi: [10.1109/JIOT.2023.3268316](https://doi.org/10.1109/JIOT.2023.3268316)]
- Wang J, Li XQ, Chen JF, *et al.* DPH-YOLOv8: Improved YOLOv8 based on double prediction heads for the UAV image object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5647715. [doi: [10.1109/TGRS.2024.3487191](https://doi.org/10.1109/TGRS.2024.3487191)]
- Jiang LJ, Yuan BX, Du JW, *et al.* MFFSODNet: Multiscale feature fusion small object detection network for UAV aerial images. *IEEE Transactions on Instrumentation and Measurement*, 2024, 73: 5015214. [doi: [10.1109/TIM.2024.3381272](https://doi.org/10.1109/TIM.2024.3381272)]
- Li YT, Fan QS, Huang HS, *et al.* A modified YOLOv8 detection network for UAV aerial image recognition. *Drones*, 2023, 7(5): 304. [doi: [10.3390/drones7050304](https://doi.org/10.3390/drones7050304)]
- Dai XY, Chen YP, Xiao B, *et al.* Dynamic head: Unifying object detection heads with attentions. *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021. 7369–7378. [doi: [10.1109/cvpr46437.2021.00729](https://doi.org/10.1109/cvpr46437.2021.00729)]
- Kang M, Ting CM, Ting FF, *et al.* ASF-YOLO: A novel YOLO model with attentional scale sequence fusion for cell instance segmentation. *Image and Vision Computing*, 2024, 147: 105057. [doi: [10.1016/j.imavis.2024.105057](https://doi.org/10.1016/j.imavis.2024.105057)]
- Xu KX, Wang Z, Geng X, *et al.* Efficient joint optimization of layer-adaptive weight pruning in deep neural networks. *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision*. Paris: IEEE, 2023. 17401–17411. [doi: [10.1109/iccv51070.2023.01600](https://doi.org/10.1109/iccv51070.2023.01600)]
- 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. [doi: [10.1109/cvpr.2016.91](https://doi.org/10.1109/cvpr.2016.91)]
- Liu S, Qi L, Qin HF, Shi JP, Jia JY. Path aggregation network for instance segmentation. *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2018. 8759–8768. [doi: [10.1109/cvpr.2018.00913](https://doi.org/10.1109/cvpr.2018.00913)]
- 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*. Honolulu: IEEE, 2017. 936–944. [doi: [10.1109/cvpr.2017.106](https://doi.org/10.1109/cvpr.2017.106)]
- Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block

- attention module. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 3–19. [doi: [10.1007/978-3-030-01234-2_1](https://doi.org/10.1007/978-3-030-01234-2_1)]
- 16 Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13708–13717. [doi: [10.1109/cvpr46437.2021.01350](https://doi.org/10.1109/cvpr46437.2021.01350)]
- 17 Wang Q, Wu BG, Zhu PF, *et al.* ECA-Net: Efficient channel attention for deep convolutional neural networks. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 11531–11539. [doi: [10.1109/cvpr42600.2020.01155](https://doi.org/10.1109/cvpr42600.2020.01155)]
- 18 Ouyang DL, He S, Zhang GZ, *et al.* Efficient multi-scale attention module with cross-spatial learning. Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Rhodes Island: IEEE, 2023. 1–5. [doi: [10.1109/ICASSP49357.2023.10096516](https://doi.org/10.1109/ICASSP49357.2023.10096516)]
- 19 Pan XR, Ge CJ, Lu R, *et al.* On the integration of self-attention and convolution. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 805–815. [doi: [10.1109/cvpr52688.2022.00089](https://doi.org/10.1109/cvpr52688.2022.00089)]
- 20 Tan MX, Pang RM, Le QV. EfficientDet: Scalable and efficient object detection. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 10778–10787. [doi: [10.1109/cvpr42600.2020.01079](https://doi.org/10.1109/cvpr42600.2020.01079)]
- 21 Chen YF, Zhang CY, Chen B, *et al.* Accurate leukocyte detection based on deformable-DETR and multi-level feature fusion for aiding diagnosis of blood diseases. Computers in Biology and Medicine, 2024, 170: 107917. [doi: [10.1016/j.compbiomed.2024.107917](https://doi.org/10.1016/j.compbiomed.2024.107917)]
- 22 Li HL, Li J, Wei HB, *et al.* Slim-neck by GSConv: A lightweight-design for real-time detector architectures. Journal of Real-time Image Processing, 2024, 21(3): 62. [doi: [10.1007/s11554-024-01436-6](https://doi.org/10.1007/s11554-024-01436-6)]
- 23 Yang GY, Lei J, Zhu ZK, *et al.* AFPN: Asymptotic feature pyramid network for object detection. Proceedings of the 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Honolulu: IEEE, 2023. 2184–2189. [doi: [10.1109/SMC53992.2023.10394415](https://doi.org/10.1109/SMC53992.2023.10394415)]
- 24 Wang CY, Yeh IH, Mark Liao HY. YOLOv9: Learning what you want to learn using programmable gradient information. Proceedings of the 18th European Conference on Computer Vision. Milan: Springer, 2024. 1–21. [doi: [10.1007/978-3-031-72751-1_1](https://doi.org/10.1007/978-3-031-72751-1_1)]
- 25 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: Curran Associates Inc., 2024. 3429.
- 26 吴明杰, 云利军, 陈载清, 等. 改进 YOLOv5s 的无人机视角下小目标检测算法. 计算机工程与应用, 2024, 60(2): 191–199. [doi: [10.3778/j.issn.1002-8331.2307-0223](https://doi.org/10.3778/j.issn.1002-8331.2307-0223)]
- 27 李岩超, 史卫亚, 冯灿. 面向无人机航拍小目标检测的轻量级 YOLOv8 检测算法. 计算机工程与应用, 2024, 60(17): 167–178.
- 28 程期浩, 陈东方, 王晓峰. 基于 NDM-YOLOv8 的无人机图像小目标检测. 计算机技术与发展, 2024, 34(9): 63–69. [doi: [10.20165/j.cnki.ISSN1673-629X.2024.0177](https://doi.org/10.20165/j.cnki.ISSN1673-629X.2024.0177)]
- 29 Zhu XK, Lyu SC, Wang X, *et al.* TPH-YOLOv5: Improved YOLOv5 based on Transformer prediction head for object detection on drone-captured scenarios. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 2778–2788. [doi: [10.1109/ICCVW54120.2021.00312](https://doi.org/10.1109/ICCVW54120.2021.00312)]

(校对责编: 张重毅)