

多路径特征融合的 YOLOv8 航拍图像检测算法^{*}王灵超¹ 沈学利² 艾强³ 闫海龙¹

(1. 辽宁理工学院信息工程学院 锦州 121000; 2. 辽宁工程技术大学软件学院 葫芦岛 125105;

3. 青海师范大学计算机学院 西宁 810008)

摘要: 针对无人机航拍图像中目标密集、背景复杂导致小目标检测精度较低的问题,提出了一种改进的航拍目标检测算法(MF-YOLO)。首先,增强 YOLOv8 的多路径特征融合能力,整合不同层次特征以保留浅层细节,提高小目标检测精度;其次,采用 EMA 注意力机制,提高目标区域识别率和目标框定位精度,有效区分目标与背景区域;然后,提出密集注意力层(DAL),通过聚焦密集目标区域和抑制无关特征,提升算法对密集区域的特征提取能力;接着,设计挤压激励检测头,结合 SE 注意力机制抑制冗余特征,进一步提升小目标检测精度;最后,构建视频数据集并设计目标检测系统,以可视化算法检测效果。在 VisDrone2019 数据集上的实验验证表明,MF-YOLO 的 mAP_{0.5} 达到 30.3%,较基线算法 YOLOv8n 提升 3.4%。结果显示,该算法显著提升了无人机图像的目标检测性能,具有广泛的应用前景。

关键词: 无人机检测;多路径特征融合;密集注意力;挤压激励检测头;YOLOv8n

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:** 520.6040

YOLOv8 aerial image detection algorithm with multi-path feature fusion

Wang Lingchao¹ Shen Xueli² Ai Qiang³ Yan Hailong¹

(1. College of Information Engineering, Liaoning Institute of Science and Engineering, Jinzhou 121000, China;

2. College of Software, Liaoning Technical University, Huludao 125105, China;

3. College of Computer, Qinghai Normal University, Xining 810008, China)

Abstract: To address the challenge of low detection accuracy for small objects in drone aerial images due to dense targets and complex backgrounds, we proposed MF-YOLO. First, the multi-path feature fusion capability is enhanced to integrate features from different layers, preserving shallow details and improving small object detection accuracy. Second, the EMA attention mechanism is adopted to improve the recognition rate of target regions and the accuracy, effectively distinguishing targets from background regions. Then, a Dense Attention Layer (DAL) is introduced to enhance the algorithm's feature extraction capability in dense regions by focusing on these areas and suppressing irrelevant features. Next, a Squeeze-and-Excitation detection head is designed, incorporating the SE attention mechanism to suppress redundant features and further improve small object detection accuracy. Finally, a video dataset is constructed, and a target detection system is designed to visualize the algorithm's detection performance. Experimental validation on the VisDrone2019 dataset shows that MF-YOLO achieves a mAP_{0.5} of 30.3%, a 3.4% improvement compared to the YOLOv8n baseline algorithm. The results demonstrate that the algorithm significantly improves object detection performance in drone images and has broad application prospects.

Keywords: UAV detection; multi-path feature fusion; dense attention layer; squeeze excitation detection head; YOLOv8n

0 引言

随着无人机技术的快速发展,其在航拍、应急响应、农业监测和城市管理等领域的应用日益广泛^[1]。然而,无人机航拍图像中的小目标检测仍然面临诸多挑战。首先,目

标尺度变化显著,由于无人机飞行高度和视角的不确定性,同一目标在不同图像中可能呈现出较大尺度差异,使得检测器难以提取稳定的特征。其次,背景复杂,无人机拍摄的场景通常包含大量相似纹理和干扰信息,如建筑、树木和道路等,容易导致误检。再者,目标易受遮挡,特别是在城市

环境或密集目标场景中,目标间的相互遮挡会导致部分关键信息丢失,增加检测难度^[2]。此外,光照变化、运动模糊和成像分辨率受限等因素也对检测精度产生不利影响。

为应对这些挑战,深度学习方法已广泛应用于无人机航拍目标检测任务,主要分为二阶段和一阶段目标检测算法。二阶段检测算法(如Fast R-CNN、Faster R-CNN和RetinaNet)依赖于区域提议网络(region proposal network, RPN)生成候选区域,在一般目标检测任务中表现良好,但由于计算复杂度较高,难以满足无人机航拍的实时性需求。一阶段检测算法以YOLO系列为代表,因其端到端检测模式和较高的检测速度,被广泛应用于无人机场景。然而,这类算法在小目标检测方面仍然存在局限性,主要表现为:1)特征提取能力不足,小目标在下采样过程中容易丢失关键信息,导致检测精度下降;2)目标与背景的区别能力有限,复杂背景容易干扰小目标识别,增加误检率;3)特征融合方式相对简单,难以充分利用不同层级的特征信息,影响模型对小目标的表达能力。

为提升YOLO等系列算法在无人机图像中目标检测的精确度,研究者们从改进特征金字塔网络(feature pyramid networks, FPN)^[3]、多尺度特征融合以及添加注意力机制等角度提出了不同优化方法。对于无人机航拍图像中目标尺寸较小、特征信息有限而导致检测效果不佳的问题,袁玲玲等^[4]提出了一种基于YOLOv7-tiny的航拍图像车辆检测方法,通过设计自适应感知特征提取模块和双分支空间感知下采样模块,混合不同通道的特征图,最大限度地保留小目标特征信息。Li等^[5]提出一种名为平衡空间和语义信息融合金字塔网络的新型颈部架构,通过增加上采样的频次,以融合浅层空间信息与深层语义信息,但显著增加了计算量。Xiong等^[6]提出AS-YOLOV5,通过采用软池化技术优化了特征提取网络,减少了小目标关键边缘信息在下采样过程中的损失,并通过包含可学习参数的特征融合方法,有效平衡了特征层。为了检测密集小目标,Tian等^[7]提出MD-YOLO,将DenseNet块和自适应注意力模块(adaptive attention module, AAM)整合到特征提取部分,有效平衡了特征层,提升了小目标检测精度。为实现复杂背景中的小目标检测,苏雨蕾等^[8]提出YOLO-DA算法,通过小目标和极小目标检测层,提高网络对小目标特征的学习能力;同时引入空间自适应特征融合检测头,通过学习空间过滤冲突信息来抑制不同尺度特征的不一致性,实现多尺度特征的自适应融合。Zhang等^[9]提出一种基于差的局部对比增强技术来提高目标检测性能。

上述改进算法能有效提升下目标检测精度,同时也增加了运算负担或者参数量。因此,需要提高算法的小目标检测性能的前提下,保持算法轻量化,以提高算法在复杂背景下的小目标识别精度,降低误检率。

1 YOLOv8n 算法简介

YOLOv8n^[10]是一阶段目标检测算法,为YOLOv8系

列中深度和特征宽度最小的版本,在检测精度与推理速度上均优于YOLOv5^[11]和YOLOv7^[12],适合无人机航拍、自动驾驶等实时场景。其整体结构由Backbone、Neck、Head构成;Backbone改进自CSPDarknet53^[13],并以梯度分流的C2f模块强化信息流;Neck采用PAN-FPN^[14]进行多尺度双向融合;Head将分类与回归分离,借助损失函数兼顾精度与速度。输入图像经卷积、标准化、激活函数等处理后可显著降低计算开销,助力网络轻量化部署。

2 改进YOLOv8n网络结构

针对无人机的实际部署需求,提出基于YOLOv8n的改进算法MF-YOLO。它在主干网络(backbone)、颈部(neck)和头部(head)方面相比以往版本进行了显著的改进。首先,提出多路径特征融合网络(multi-path feature fusion network, MFFN)来整合不同层次的特征以保留浅层细节。其次,采用EMA(efficient multi-scale attention, EMA)^[15]注意力机制,有效区分目标与背景。此外,提出密集注意力层(dense attention layer, DAL)以聚焦密集目标区域,抑制无关特征,增强特征提取效果。最后,设计挤压激励检测头,结合SE(squeeze-and-excitation network, SENet)注意力^[16]机制抑制冗余特征,进一步提升小目标检测性能。具体的网络结构如图1所示。

2.1 多路径特征融合网络

传统特征融合方法,如FPN采用自顶向下的方法,从最高分辨率的特征图开始,逐步融合来自更低分辨率的特征。然而,FPN在处理多尺度特征时会忽略细粒度信息,导致小目标检测性能不足。为此,路径聚合特征金字塔网络(path aggregation feature pyramid network, PAFPN)通过增加增强双向的信息流,优先保留空间细节特征。然而,其在复杂场景下的信息传递不充分,影响图像的检测精度。

双向特征金字塔网络(bidirectional feature pyramid network, BiFPN)^[17]通过双向集成不同分辨率的特征,利用自底向上和自顶向下的路径,进一步提升了目标检测性能。然而,深度堆叠的BiFPN块在训练过程中可能导致梯度消失,影响整体网络性能,尤其在处理更深层次特征时表现不佳。

为了解决上述特征融合问题,提出MFFN,该改进方法在不显著增加计算复杂度或延迟的情况下,实现跨尺度特征融合,并提升检测性能。具体而言,MFFN通过参数化和消除额外上采样操作,简化了复杂性,降低了计算开销,其网络结构如图2所示。

在改进后的MFFN架构中,特征融合模块采用了C2f-EMA模块,如图3所示,该模块取代了传统的基于 3×3 卷积的特征融合,并引入了注意力机制。C2f-EMA模块通过引入注意力机制,能够更精准地选择和融合高层语义特征与底层空间细节,从而提升了小目标检测精度。

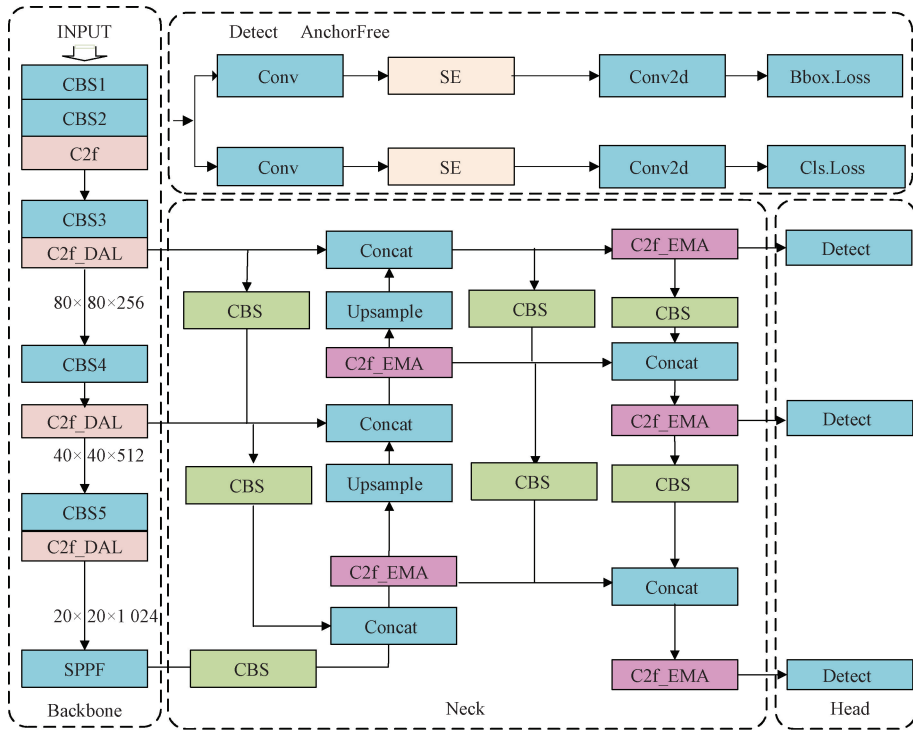


图 1 改进 YOLOv8 网络结构

Fig. 1 Improvement of YOLOv8 network structure

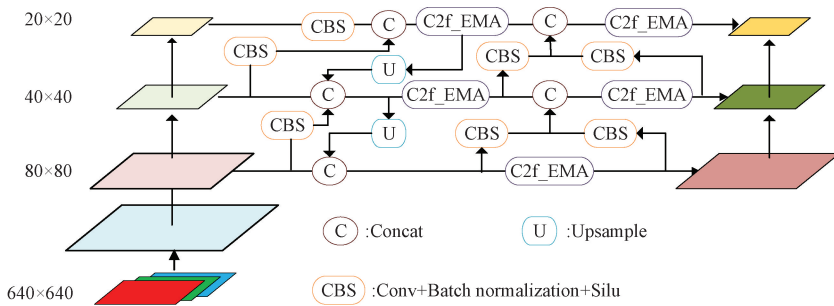


图 2 多路径特征融合网络结构

Fig. 2 Multi-path feature fusion network architecture

改进后的网络不仅保留了多尺度特征融合的优势,还通过减少计算开销和延迟,使其更适合实际应用场景。

聚焦目标区域并抑制冗余特征干扰,本文提出密集注意力层,其目的在于通过多重注意力机制强化网络对目标的专注度,尤其是在拍摄距离远、目标尺度小且背景复杂的情况下,为后续检测和识别提供更加清晰、突出的特征表示,从而显著提升整体检测性能与鲁棒性,其结构如图 4 所示。

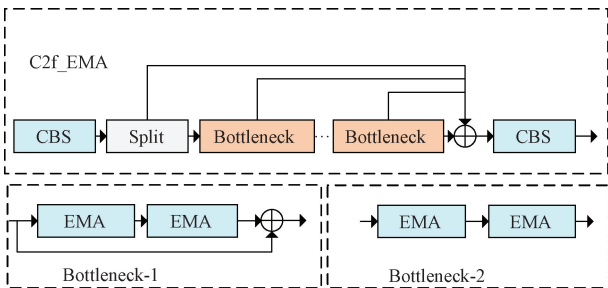


图 3 C2f-EMA 结构

Fig. 3 C2f-EMA structure

2.2 密集注意力层 (DAL)

为了在无人机航拍等高度密集的小目标场景中,充分

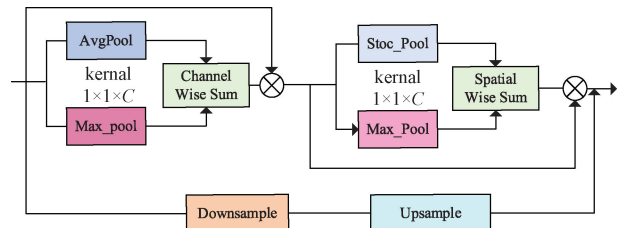


图 4 密集注意力层结构

Fig. 4 Dense attentional layer structure

具体而言,DAL首先在通道维度上对输入特征 D 分别执行平均池化和最大池化操作,得到通道注意力权重,并与原始特征 F 逐元素相乘,生成经过通道校正的特征 D_1 ,如式(1)所示。

$$D_1 = D \odot (\text{AvgPool}(D)_{1 \times 1 \times C} + \text{MaxPool}(D)_{1 \times 1 \times C}) \quad (1)$$

其中, \odot 表示逐元素相乘。这一步旨在突出通道维度更具判别力的特征,并抑制通道间的噪声或冗余信息,让网络在提取与融合小目标关键特征时更为精准。

将经过通道关注的调整特征 F_1 作为输入。为了进一步融合特征并更加关注空间位置,引入空间随机池化和最大池化。随机池化是随机选择池化核区域内的值来计算最大值的一种池化方法。与传统的最大池化方法相比,这种随机池化方法提高了模型的鲁棒性,特别是在对于密集区域的目标,浅层特征可能存在噪声干扰的情况下。在此基础上,DAL在空间维度进一步结合随机最大池化与随机池化操作,用大小为 $H \times W \times 1$ 的池化核强化目标高度集中的区域并抑制背景干扰。对通道校正后的 D_1 施加两种空间注意力权重后,可得到空间校正特征 D_2 。

$$D_2 = D_1 \odot \left(\begin{matrix} \text{StocPool}(D_1)_{H \times W \times 1} + \\ \text{MaxPool}(D_1)_{H \times W \times 1} \end{matrix} \right) \quad (2)$$

经过通道注意和空间注意校正后,将校正后的权重特征与特征提取器的输入特征融合,如式(3)所示。这种融合过程在特征中引入了更多的关注信息,有助于在密集特征中提取更有效的目标。此外,注意信息有助于特征的集中,符合无人机应用中密集物体检测的特点。

$$D_{\text{output}} = \text{Extractor}(D) \odot D_2 \quad (3)$$

2.3 挤压激励检测头

YOLOv8在检测头中采用了相对精简的解耦结构,通过将分类(classification)和回归(regression)分支分开,避免二者在特征学习中互相干扰。由于检测头本身通常只在有限的尺度特征上完成预测,并未在此处引入更多的注意力机制来专门强化对小目标特征的挖掘,导致小目标在与背景或其他干扰相似度较高时,检测头无法充分突出其关键特征通道,从而出现漏检或定位不准的问题。

为改进上述缺陷,在YOLOv8检测头的卷积层之后插入SENet模块,结构如图5所示。SENet的核心原理在于对通道维度进行自适应重标定,具体过程可分为“挤压”与“激励”两个步骤:首先对特征图 X 进行全局平均池化,计算每个通道的全局响应向量。

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_c(i, j) \quad (4)$$

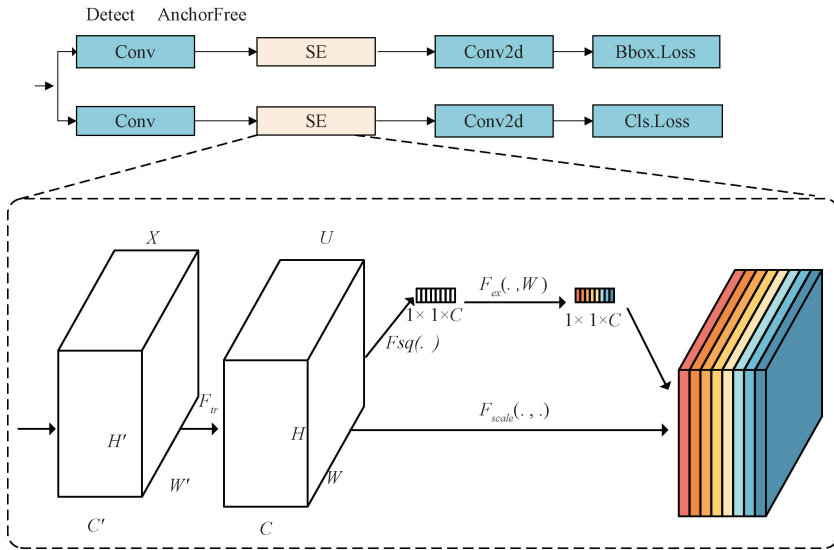


图5 挤压激励检测头

Fig. 5 Squeeze excitation detection heads

这一步称为挤压, $X_c(i, j)$ 代表位于位置 (i, j) 的第 c 个通道的值。随后将该向量输入到由两层全连接(含激活函数)和最终的Sigmoid组成的映射函数,得到通道级别的加权系数向量。

$$s = \sigma(W_2 \cdot \delta(W_1 \cdot z)) \quad (5)$$

式中: s 代表通道权重向量, σ 是Sigmoid激活函数,用于生成权重向量中的每个元素。 W_1 和 W_2 代表全连接层的权重矩阵。 δ 为ReLU激活函数。 z 为Squeeze操作输出

的特征向量。

最后,利用该权重对原特征逐通道相乘,得到重标定后的特征 \tilde{X} 。

$$\tilde{X} = X \odot s \quad (6)$$

通过这一通道加权机制,SENet能够对“有助于检测小目标”的特征通道赋予更高权重,抑制对识别贡献度不大的通道,从而弥补YOLOv8检测头在设计上对于通道

注意力的忽视问题。

SENNet 的优势在于其结构相对简单明了,易于直接插入到检测头的卷积层之后,同时只需要少量的额外参数和计算量即可显著提升通道注意力。对于无人机航拍中多变、复杂的拍摄场景而言,SENNet 能够更好地突出小目标在分类和回归分支上的特征,使检测头即使在有限的尺度输出上,也能对小目标保有更高的敏感度。

3 实验结果分析

3.1 实验数据集

VisDrone2019^[18]数据集由天津大学机器学习与数据挖掘实验室携手 AISKYEYE 数据挖掘团队共同开发,是目前在无人机航拍领域中广受关注且具有广泛影响力的研究基准之一。VisDrone2019 数据集不仅在地理分布、环境条件和目标类型上体现了高度多样性,还特别收录了白天与夜晚等不同光照条件下的拍摄内容;同时,它所关注的目标类型范围广泛,包括行人、人、自行车、汽车、货车、三轮车、遮阳三轮车、公共汽车、卡车和摩托车,其中,处于站立或行走姿态的人类图像分类为“行人”,其他姿态的分类为“人”。本实验使用官方提供的数据集,数据中训练集是 6 471、验证集是 548、测试集 1 610 张。

DOTA-v1.5^[19]是一个专为遥感图像目标检测设计的大规模数据集,它涵盖了 16 个目标类别,包括飞机、轮船、大型车辆、小型车辆、桥梁等,数据丰富且场景复杂。DOTA-v1.5 包含 2 806 张图像,分辨率从 800×800 到 20 000×20 000 不等,标注了超过 40 万个目标实例。数据集中采用定向边界框(OBB)和水平边界框(HBB)两种标注方式,支持不同类型的目标检测任务。DOTA-v1.5 的广泛应用推动了遥感图像领域的研究与发展,是目标检测模型评估的重要基准数据集。论文采用官方提供的 10 595 张训练图像、3 360 张验证图像和 71 888 张测试图像进行实验验证。

3.2 实验环境和参数

本研究以 YOLOv8n 作为主要对照网络,实验环境详如表 1 所示。训练阶段选用 Adam 优化器,初始学习率为 0.01,权重衰减系数设为 0.000 5,批大小(batch_size)取 16,并运行 300 个训练轮次(epoch)。图像输入分辨率统一为 640×640,数据增强操作采用默认配置。为确保一致性,所有消融与对比实验皆基于相同设置,无额外训练,其他参数如表 1 所示。

3.3 评价指标

本文基于精确率(precision)、召回率(recall)、平均精确率均值(mean average precision, mAP)、模型参数量(parameter, Params)以及浮点数运算量(floating point operations, FLOPs)五方面,对模型的检测效能进行衡量。相关公式定义如下:

$$P = \frac{TP}{TP + FP} \quad (7)$$

$$R = \frac{TP}{TP + FN} \quad (8)$$

$$AP = \int_0^1 P(R) dR \quad (9)$$

$$mAP = \frac{\sum_{i=1}^k AP_i}{k} \quad (10)$$

$$FLOPs = C_i \times K^2 \times C_0 \times W \times H \quad (11)$$

式中:TP 表示被正确识别的正样本数量,FP 表示被错误判定为正样本的数量,FN 则指实际为正样本却被错误判定为负样本的数量, k 代表数据类别数。此外,符号 C_i 用于表示网络的输出通道数, C_0 表示输入通道数, W 和 H 分别指代卷积核的宽度与高度。通过这些变量的配合,可进一步描述检测过程中对正负样本的识别性能,以及卷积操作在网络结构中的维度转换情况。

表 1 实验硬件环境

Table 1 Experimental hardware environment

参数	实验环境
CPU	AMD Ryzen 7 2700X
GPU	RTX 3060(12 G)
操作系统	Ubuntu20.04
Python	3.9
深度学习框架	Pytorch1.8.1

3.4 消融实验

为了验证改进算法的有效性,采用 VisDrone2019 数据集对基线模型进行消融实验,结果如表 2 所示。相较于基线模型,改进的模型 F 在实验结果精度上有明显提升,mAP@0.5 提升 3.6%。实验 A 使用 MFFN 代替模型的颈部网络,mAP@0.5 提升 0.9%,可知改进后的 FFMN 能有效融合不同尺度的图像信息。提高网络对位置特征的处理能力。实验 B 在颈部结构嵌入了 DAL,mAP@0.5 提升了 0.6%,说明该模块通过通道与空间两个维度的注意力机制,在高密度场景中突出关键目标并抑制冗余特征,从而增强网络对小目标的检测能力。实验 C 是设计的挤压激励检测头,检测精度提升 1%,且没有明显的参数量和计算量的增长,证明设计的检测头可在保持轻量化的同时显著增强网络对复杂场景和小目标的检测效果,进一步提升小目标检测精度。实验 D 使用 MFFN 替换颈部网络并增加 DAL,检测精确度进一步增加。实验 E 结合了 DAL 和挤压激励检测头,使得算法能有精准定位小目标,促进了模型整体性能的优化。模型 F 为改进后的 MF-YOLO 模型,模型参数量和浮点数运算相较于基线分别增加了 0.8 MB 和 3.2 GB,mAP@0.5 提升了 3.4%,说

明改进的模型在不显著增加参数数量的情况下,能显著提升目标检测精度。

表2 采用不同改进策略后的检测结果

Table 2 Detection results after using different improvement strategies

实验	MFFN	DAL	SE_Head	mAP@0.5/%	参数量/MB	计算量/GB
YOLOv8n				26.9	3.16	8.9
A	✓			27.8	3.45	9.3
B		✓		27.5	3.17	9.0
C			✓	27.9	3.01	8.2
D	✓	✓		28.7	3.46	9.3
E		✓	✓	28.1	3.24	9.5
F	✓	✓	✓	30.3	3.86	11.6

注:✓表示采用了改进后的策略

3.5 对比实验

为了进一步证明 MF-YOLO 算法在无人机航拍小目标中的检测效果,在 VisDrone 数据集上将 MF-YOLO 与经典的主流模型进行对比。如表 3 所示,加粗部分为该类别在所有算法中的最优值。

通过表 3 可以得出, MF-YOLO 算法在 VisDrone 数据集上的综合性能优异,尤其是在行人、人和摩托车等较小物体的检测任务中, AP 值分别达到了 29.4%、18.7% 和

28.9%。此外,对于汽车等较大物体的检测精度也明显优于其他模型。尽管与 YOLOv7-tiny 相比提升幅度不大,但其参数量仅为 YOLOv7-tiny 的 62%,在综合性能上更显优势。在速度方面,本文算法与 YOLOv8 基线基本持平,且优于其他算法。总之, MF-YOLO 在无人机航拍检测任务重表现良好,不仅在较小物体上取得了明显的精度提升,对较大物体也有明显的优势,相较于其他常用的检测方法, MF-YOLO 算法综合检测性能占优。

表3 不同算法在 VisDrone 测试集上的 AP 与 mAP 对比

Table 3 AP vs. mAP for different algorithms on the VisDrone test set

方法	AP50/%										mAP0.5/%	FPS
	行人	人	自行车	汽车	货车	卡车	三轮车	遮阳三轮车	公共汽车	摩托车		
YOLOv8n	22.6	12.0	6.1	66.3	29.3	29.0	12.1	15.1	50.0	24.2	26.7	163
Faster R-CNN ^[20]	21.4	15.6	6.7	51.7	29.5	19.0	13.1	7.7	31.4	20.7	21.7	4.9
RetinaNet ^[21]	13.0	7.9	1.4	45.5	19.9	11.5	6.3	4.2	17.8	11.8	13.9	23.8
YOLOv4-tiny ^[22]	11.4	11.1	3.1	52.5	22.8	20.4	8.2	7.6	41.25	12.45	19.1	66
YOLOv5s	25.2	16.3	8.7	67.4	31.2	33.6	14.0	13.5	55.0	24.1	28.9	116.8
YOLOv7-tiny	27.1	18.3	11.0	68.1	35.6	34.0	14.1	14.7	50.5	29.1	30.2	109.8
MF-YOLO	29.3	18.8	7.4	72.0	33.0	31.6	14.3	16.5	51.4	28.9	30.3	158

为了验证本文所提模型的泛化性,使用 DOTA-v1.5 进行实验,具体结果如表 4 所示,由实验可知,本文算法精度较基线模型提升 0.7%,具有较好的泛化性。

表4 DOTA 数据集上的泛化实验

Table 4 Generalization experiment results on the DOTA dataset

模型	参数量/MB	计算量/GB	精度/%	召回率/%	mAP0.5/%
YOLOv8n	3.16	8.9	65.1	33.8	35.8
MF-YOLO	3.86	11.6	68.6	33.2	36.5

3.6 可视化对比实验

针对 MF-YOLO 算法在实际场景中的检测性能验证,

本文选取 DOTA-v1.5 与 VisDrone2019 测试集中具有多视角、小目标密集特性的样本进行对比实验。实验结果对比分析如图 6 和 7 所示,具体而言:对比图 6(a)与(d)的检测结果可见,改进后的算法有效抑制了误检现象;通过图 6(b)与(e)的对比分析表明,本文算法通过特征增强机制降低算法的漏检率;进一步分析图 6(c)与(f)的遮挡场景检测效果,改进模型在保持高精度的同时,有效缓解小目标遮挡问题。图 7 展示了在 VisDrone2019 中 MF-YOLO 算法的优势。图 7(a)和(b)分别展示了基准方法与 MF-YOLO 算法的检测效果对比。可视化结果表明,改进算法在无人机视角下的小尺寸目标检测任务中展现出更强的特征辨识能力,更适用于航拍小目标检测任务。

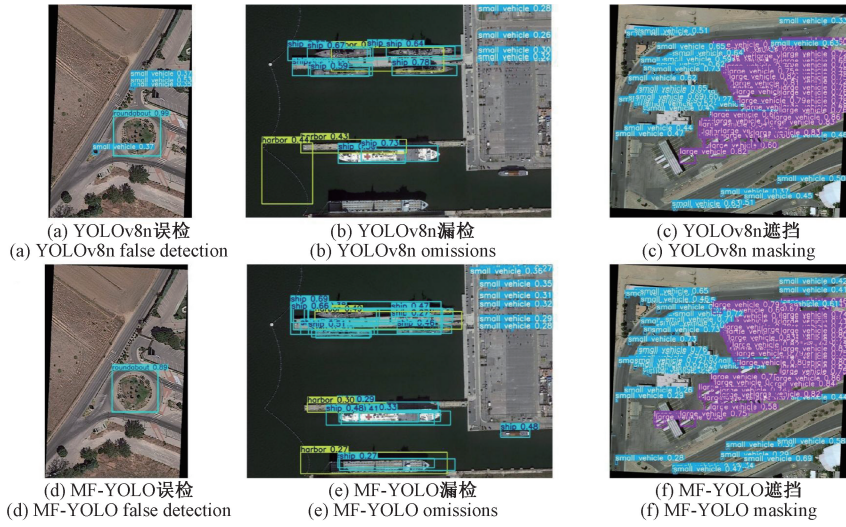


图 6 DOTA v1.5 可视化结果对比

Fig. 6 Comparison of visualisation results on DOTA v1.5

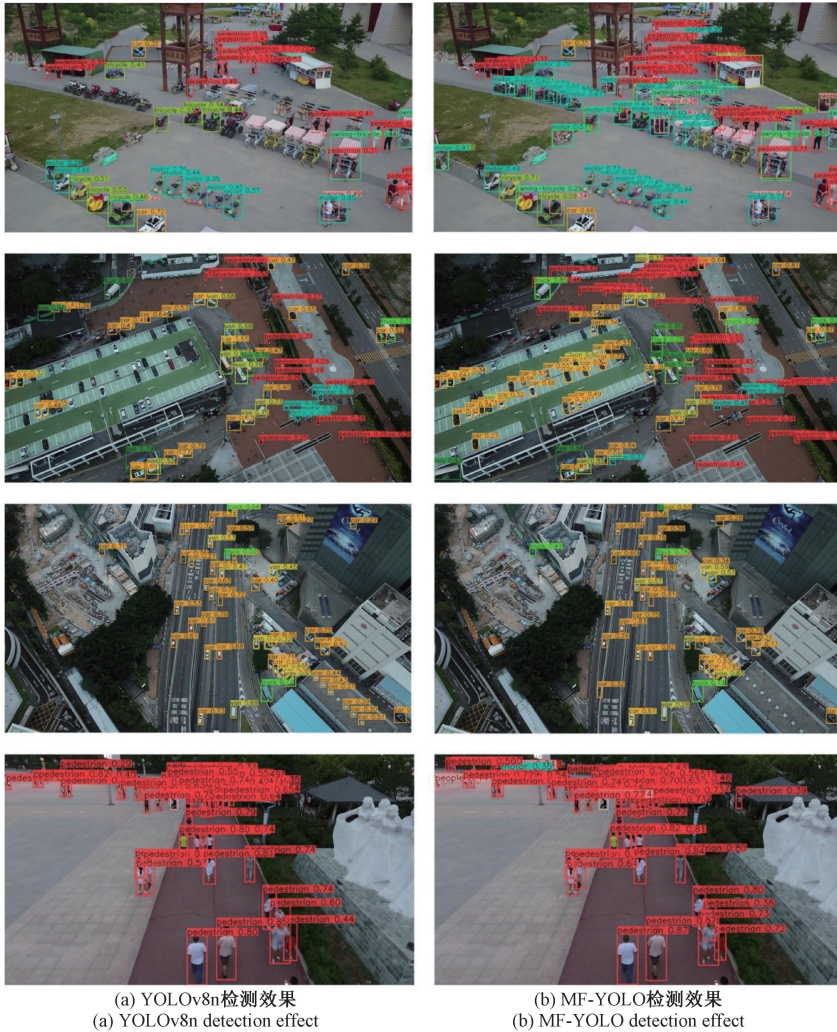


图 7 Visdrone 可视化结果对比

Fig. 7 Comparison of visualisation results on Visdrone

3.7 系统验证

为验证 MF-YOLO 在实际场景下的检测性能,本文构建了一个可加载预训练模型的目标检测系统,该系统可以加载训练模型,检测目标可以是图片和视频。本次验证选取青海省西宁市城西区五四西路与冷湖路交叉路口作为拍摄地点,并自建视频数据集测评算法实时目标检测效果。具体结果如图8所示,本文对比了系统加载的YOLOv8n模型与MF-YOLO在同一真实航拍视频中的表现,结果显

示,基线模型能够识别出大部分目标,但在小目标的检测上存在较大的误差和遗漏,尤其是在低分辨率和远距离目标的情况下,检测精度较低。对比之下,本文提出的模型在处理小目标时展现了明显的优势。综上,本文模型相较于传统次优模型,不仅提升了检测精度,还显著提高了小目标的识别率,为航拍视频中的小目标检测提供了更加可靠的解决方案,并展现了该技术在实际应用中的广泛潜力。



图8 系统检测结果

Fig. 8 Detection results obtained by the system

4 结 论

本文提出了一种改进的航拍目标检测算法(MF-YOLO),针对无人机视角下的小目标密集分布和复杂背景干扰问题,通过强化YOLOv8的多路径特征融合能力,增强浅层细节保留,从而显著提高对小目标的检测精度,并结合EMA注意力机制进一步提升目标区域识别率与精确定位能力;此外,算法在检测过程中引入密集注意层(DAL)以聚焦目标密集区域,抑制无关特征干扰在检测头部分;采用挤压激励检测头(SENet),有效减少冗余特征的影响,整体性能得到大幅提升。在VisDrone2019数据集上的实验结果表明,MF-YOLO在mAP@0.5指标上达到30.3%,较基线算法YOLOv8n提升3.4%,突出体现了算法在行人等小目标检测场景中的优异表现。与其他检测算法相比,MF-YOLO兼顾了较高精度与相对较低的参数量和运算量,适合嵌入到无人机设备中。

参考文献

- [1] ZHAO C, LIU R W, QU J, et al. Deep learning-based object detection in maritime unmanned aerial vehicle imagery: Review and experimental comparisons[J]. Engineering Applications of Artificial Intelligence, 2024, 128: 107513.
- [2] TANG G, NI J, ZHAO Y, et al. A survey of object

- detection for UAVs based on deep learning[J]. Remote Sensing, 2023, 16(1): 149.
- [3] TLIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [4] 袁玲玲,陈春梅,朱天鑫,等.基于航拍图像的自适应感知目标检测网络[J].电子测量技术,2025,48(2): 57-65.
- YUAN L L, CHEN CH M, ZHU T X, et al. Aerial image-based adaptive perception object detection network[J]. Electronic Measurement Technology, 2025, 48(2): 57-65.
- [5] LI Y, LI Q, PAN J, et al. SOD-YOLO: Small-object-detection algorithm based on improved YOLOv8 for UAV images[J]. Remote Sensing, 2024, 16(16): 3057.
- [6] XIONG X, HE M, LI T, et al. Adaptive feature fusion and improved attention mechanism based small object detection for UAV target tracking[J]. IEEE Internet of Things Journal, 2024, 11(12): 21239-21249.
- [7] TIAN Y, WANG S, LI E, et al. MD-YOLO: Multi-scale dense YOLO for small target pest detection[J]. Computers and Electronics in Agriculture, 2023, 213:

- 108233.
- [8] 苏雨蕾, 黄丹丹, 刘智, 等. 无人机对地小目标检测方法研究[J]. 电子测量与仪器学报, 2024, 38(9):144-154. SU Y L, HUANG D D, LIU ZH, et al. Research on small ground target detection method for UAVs[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(9): 144-154.
- [9] ZHANG K, YANG K, LI S, et al. A difference-based local contrast method for infrared small target detection under complex background [J]. IEEE Access, 2019, 7: 105503-105513.
- [10] SOHAN M, SAI RAM T, REDDY R, et al. A review on YOLOv8 and its advancements [C]. International Conference on Data Intelligence and Cognitive Informatics, 2024: 529-545.
- [11] THUAN D. Evolution of YOLO algorithm and YOLOv5: The state-of-the-art object detection algorithm[D]. Oulu: Oulu University of Applied Sciences, 2021.
- [12] 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.
- [13] MAHASIN M, DEWI I A. Comparison of CSPDarkNet53, CSPResNeXt-50, and EfficientNet-B0 backbones on YOLOv4 as object detector [J]. International Journal of Engineering, Science and Information Technology, 2022, 2(3): 64-72.
- [14] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [15] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning[C]. ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). IEEE, 2023: 1-5.
- [16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [17] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
- [18] DU D, ZHU P, WEN L, et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results [C]. IEEE/CVF International Conference on Computer Vision Workshops, 2019.
- [19] XIA G S, BAI X, DING J, et al. DOTA: A large-scale dataset for object detection in aerial images[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3974-3983.
- [20] JIANG H, LEARNED-MILLER E. Face detection with the faster R-CNN [C]. 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition(FG 2017), 2017: 650-657.
- [21] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [22] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. ArXiv preprint arXiv:2004.10934, 2020.

作者简介

王灵超, 硕士, 主要研究方向为图像与视觉信息计算、目标检测等。

E-mail:20241035@lise.edu.com

沈学利, 博士, 教授, 主要研究方向为智能信息处理、目标检测等。

E-mail:Shenxueli@lntu.edu.cn

艾强, 硕士研究生, 主要研究方向为时空数据挖掘、图像处理等。

E-mail:aiqiang@stu.qhnu.edu.cn

闫海龙(通信作者), 本科, 副教授, 主要研究方向为图像处理、目标检测等。

E-mail:77904555@qq.com