

改进 YOLOv5s 的轻量化遥感图像小目标检测方法

张 瑶,王军号

安徽理工大学 计算机科学与工程学院,安徽 淮南 232001

摘要:目的 针对无人机遥感图像检测中复杂场景下小目标精度低和模型复杂的问题,对基准 YOLOv5s 算法进行优化改进。方法 引入轻量化卷积注意力模块 CBAM(Convolutional Block Attention Module)和堆叠融合策略来重构主干 C3 模板,在减少参数数量的同时,提升了网络的特征提取能力;在特征融合层,引入跨层级特征融合来避免特征损失,并在融合过程中引入上采样算子,降低模型复杂度的同时融合了更多的特征信息;使用集成损失函数 ILF(Integrated Loss Function)作为边界框损失函数,增强对目标的定位能力。结果 在 VisDrone2019 数据集测试中,改进后的算法与原始算法相比,平均精度均值提升了 5.8%;模型容量压缩至 6.2 MB,相较原模型大幅下降;参数量相比原模型下降 44.7%。同时,与主流检测方法相比,也取得了更高的检测精度。结论 提出的改进方法在保持模型检测精度的同时,显著降低了模型的复杂度和计算成本,实现了遥感图像中小目标的快速且精准识别。对于小目标检测任务,该方法具有较好的性能,且对于实时性和资源受限的应用场景具有重要意义。

关键词:小目标检测;跨层特征融合;模型轻量化;YOLOv5s;注意力机制

中图分类号:TP751 **文献标识码:**A **doi:**10.16055/j.issn.1672-058X.2026.0003.007

Improved Small Target Detection Method for Lightweight Remote Sensing Images Based on YOLOv5s

ZHANG Yao, WANG Junhao

School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232001, Anhui, China

Abstract: Objective This study aims to optimize and improve the benchmark YOLOv5s algorithm to solve the problems of low accuracy for small targets and complex models in complex scenarios during unmanned aerial vehicle (UAV) remote sensing image detection. **Methods** The lightweight convolutional block attention module (CBAM) and a stacking fusion strategy were introduced to reconstruct the backbone C3 template. This approach enhanced feature extraction ability of the network while reducing the number of parameters. In the feature fusion layer, cross-level feature fusion was introduced to avoid feature loss, and an upsampling operator was incorporated during the fusion process to reduce the complexity of the model and fuse more feature information. Additionally, the integrated loss function (ILF) was adopted as the bounding box loss function to enhance localization accuracy for targets. **Results** In the VisDrone2019 dataset test, the improved algorithm achieved a 5.8% increase in mean average precision (mAP) compared to the original algorithm. The model size was reduced to 6.2 MB, a significant reduction from the original model, while the number of parameters decreased by 44.7%. Moreover, the proposed method achieved higher detection accuracy compared to mainstream detection approaches. **Conclusion** The proposed improved method significantly reduces the complexity and computational cost of the

收稿日期:2024-05-30 **修回日期:**2024-09-24 **文章编号:**1672-058X(2026)03-0053-08

基金项目:国家自然科学基金项目资助(61300001)。

作者简介:张瑶(1998—),男,安徽合肥人,硕士研究生,从事目标检测研究。

通信作者:王军号(1970—),男,安徽淮南人,高级实验师,从事图像处理研究。Email: 173234893@qq.com。

引用格式:张瑶,王军号.改进 YOLOv5s 的轻量化遥感图像小目标检测方法[J].重庆工商大学学报(自然科学版),2026,43(3):53-60.

Zhang Yao, Wang Junhao. Improved small target detection method for lightweight remote sensing images based on YOLOv5s[J].

Journal of Chongqing Technology and Business University (Natural Science Edition), 2026, 43(3): 53-60.

model while maintaining the detection accuracy, and achieves fast and accurate recognition of small targets in remote sensing images. For small target detection tasks, this method has good performance and is of great significance for real-time and resource-constrained application scenarios.

Keywords: small target detection; cross-layer feature fusion; model lightweighting; YOLOv5s; attention mechanism

科技的日新月异使得无人机遥感技术可应用于应急救援^[1]、军事领域^[2]等实际场景。现阶段目标检测算法主要有以 Faster R-CNN^[3] 为代表的两阶段算法和以 YOLO^[4-7]、RetinaNet^[8] 为代表的单阶段算法。无人机遥感图像具有目标密集且小、目标多尺度的特征,因此将传统算法直接应用于其上,检测性能易下降。随着神经网络的不断发展,卷积层的不断叠加,类似 CSPDarkNet53^[9]、ResNet101^[10] 等大型骨干网络被提出,特征提取能力增强,但深度学习模型参数数量和体积也随之增大,所需要的计算资源也不断增加。因此,对主流目标检测模型的轻量化改进有着重要的应用价值。SINGH 等^[11] 提出图像金字塔方法,对原始图像进行多次缩放,生成不同尺度图像金字塔。这种方法虽然能够提高小目标检测精度,但将不同尺度的图像分别输入到网络中进行特征提取,增加了预处理阶段的计算量,从而增加了网络前向传播的计算成本。HAN 等^[12] 提出 ReDet 网络结果,通过预测目标框的旋转角度和引入旋转不变的特征对齐方式,提高了对小目标的检测精度。然而,这种设计考虑到旋转不变性,通常会增加网络的复杂性,并且优化特征对齐操作需要额外的网络参数。WANG 等^[13] 提出 YOLOv7 网络结构,采用更复杂的网络结构和更先进的特征提取方法,能够更准确地识别物体并进行定位。然而,在处理图像时,由于下采样倍数较大,难以学习到小目标的特征信息,这直接影响小目标的检测效果。此外,该模型相对较大,需要更多的计算资源和存储空间。张学锋等^[14] 采用注意力机制增强从图片中提取出的空间信息和语义信息,减少了图像细节的丢失,并通过融合浅层特征,提升了小目标的识别率,同时在实际应用中适应性较好。WANG 等^[15] 提出了 UAV-YOLOv8 模型,引入 BiFormer 注意力机制来优化骨干网络,从而提高了模型对关键信息的关注,有助于更准确地检测和识别小目标。此外,还设计了 Focal FasterNet 的特征处理模块,并在此基础上提出两种新的检测尺度,使得浅层特征和深层特征完全融合,从而提高了模型的检测性能,但需要更深的网络结构和更多的参数,训练时间会更长。

尽管上述算法在检测精度方面都有所提升,但在检测精度和轻量化之间没有达成理想的平衡。

虽然 YOLOv7 和 YOLOv8 等在性能上有所提升,但它们在轻量化方面可能并未做出特别的优化。相比之下,YOLOv5s 作为一个相对较小的模型,在保持一定性能的同时,更容易进行轻量化改进,以满足特定应用场景的需求。此外,对于实时性要求较高,或者需要在有限计算资源上运行模型的场景,YOLOv5s 更加有优势。因此,对 YOLOv5s 基准模型进行轻量化改进,提出了一种改进后的轻量化遥感图像小目标检测算法,在保持一定精度的前提下,实现了模型的轻量化。首先,引入轻量化卷积注意力模块和堆叠融合策略来重构主干网络的 C3 模块。该模块使模型能够聚焦于图像中的特定区域,尤其是小目标区域,从而使模型更加关注与目标相关的特征。这一改进显著减少了模型所需要的参数量,降低了模型的复杂度。其次,提出跨层级特征融合方法,以增加特征信息传播路径。这种设计同时融合了低层特征的高分辨率信息和高层特征的语义信息,有助于提高对小目标的检测能力。结合上采样算子 CARAFE,进一步强化了对小目标的内容感知能力。最后,采用集成损失函数,全面考虑预测框和真实框之间的差异,优化正负样本分配,从而提高模型对小目标形状的识别精度。在 VisDrone2019 数据集上的实验结果表明,改进后的算法在维持较高检测精度的同时,实现了模型的轻量化。

1 模型建立与算法设计

YOLOv5 是一阶段目标检测算法,其中 YOLOv5s 模型小巧、计算轻量且检测迅速。本文在其基础上进行优化,进一步缩减模型体积并提高检测精度。YOLOv5s 模型主要由以下 4 个部分组成:第一部分是输入端,负责处理输入数据并起到数据增强作用;第二部分是 Backbone,用于提取图像中的特征;第三部是 Neck,它将不同级别的特征组合在一起进行处理,从而提高模型对目标的识别精度;最后一部分是 Head,负责对图像特征进行预测,生成边界框并预测类别。

YOLOv5s 网络结构如图 1 所示,各组件详细结构如图 2 所示。

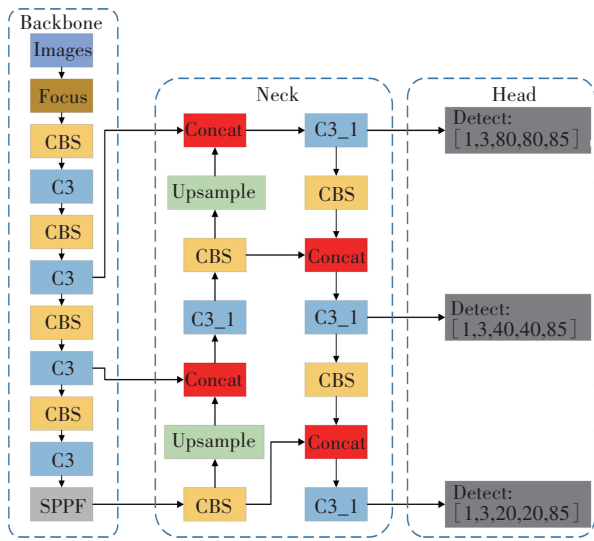


图 1 YOLOv5s 网络结构

Fig. 1 YOLOv5s network structure

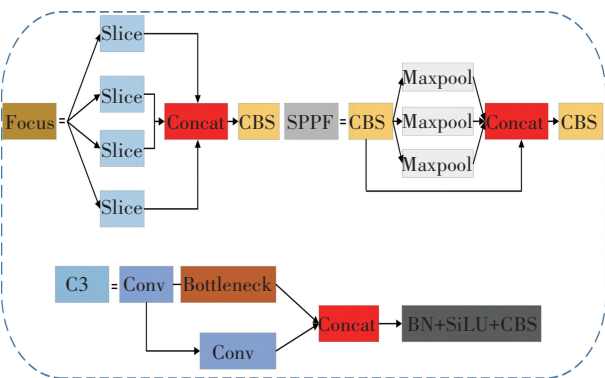


图 2 YOLOv5s 各组件详细结构

Fig. 2 Detailed structure of YOLOv5s components

1.1 重构主干 C3 模块

主干网络的 C3 模块是关键的卷积神经网络模块,能显著提升模型的性能和准确性。C3 模块由 3 个 Conv 模块和一个 Bottleneck 模块组成,每个 Conv 模块都使用 1×1 卷积进行降维或升维操作。然而, Bottleneck 模块虽能避免梯度消失问题,但其较大的参数和计算量不利于设备部署。为了解决以上问题,首先在 Bottleneck 模块拼接后引入轻量级 CBAM^[16] 注意力机制。CBAM 结合了通道注意力(Channel Attention Module, CAM)和空间注意力(Spatial Attention Module, SAM)。当接收特征 F 后,经过通道与空间维度的注意力权重整合,并与原特征相乘,最后输出增强特征 F_1 。这样能够自适应地增强关键特征,提升模型性能,同时保持轻量级和灵活性。CBAM 基本结构如图 3 所示。

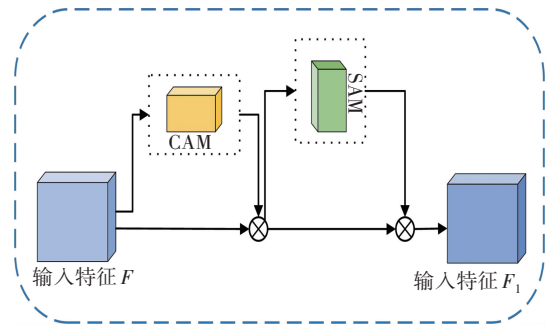


图 3 CBAM 结构

Fig. 3 CBAM structure

其次,使用堆叠融合策略来替换 C3 模块中的普通 1×1 卷积。堆叠融合策略由逐通道卷积和逐点卷积堆叠组成。具体计算过程如下:

当输入特征图的通道数为 C_i 、图片的尺寸为 $h \times w$, 通过 1×1 的常规卷积操作后,输出特征图的通道数为 C_0 、整个过程的计算量为

$$F = C_i \times h \times w \times C_0 \quad (1)$$

假定堆叠融合策略是由卷积核大小为 5×5 的逐通道卷积组成。逐通道卷积的计算量为

$$F_d = C_i \times 5 \times 5 \times h \times w \quad (2)$$

卷积核为 1×1 逐点卷积的计算量为

$$F_p = C_i \times 1 \times 1 \times h \times w \times C_0 \quad (3)$$

综上所述,堆叠融合策略与传统卷积策略在计算量的对比可表述为

$$R = \frac{F_d + F_p}{F} = \frac{C_i \times 5 \times 5 \times h \times w + C_i \times 1 \times 1 \times h \times w \times C_0}{C_i \times 1 \times 1 \times h \times w \times C_0} \quad (4)$$

由式(4)可知,堆叠融合策略通过分解标准卷积操作,显著减少了模型所需要的参数量。这有助于降低模型的存储需求和计算成本,使得模型更加轻量化和高效。

1.2 跨层级特征融合

随着网络深度的不断加深,一方面,模型能够捕获到更复杂、更丰富的特征;另一方面,特征融合层在下采样以及网络向前传播过程中会造成特征信息丢失以及小目标位置信息丢失。为了保持网络性能的同时尽可能减少小目标特征信息的丢失,在特征融合层额外增加了两条链路,将不同层的特征进行融合。一方面,通过整合编码器生成的特征图,可以增强图像中的语义信息和细节信息,有助于模型更全面地理解图像内容,从而提高目标检测的准确性;另一方面,特征融合结合不同尺度的特征信息,使模型能够同时捕获到目标的大致位置和精确边界,有助于提升检测性能。结构如图 4 所示。

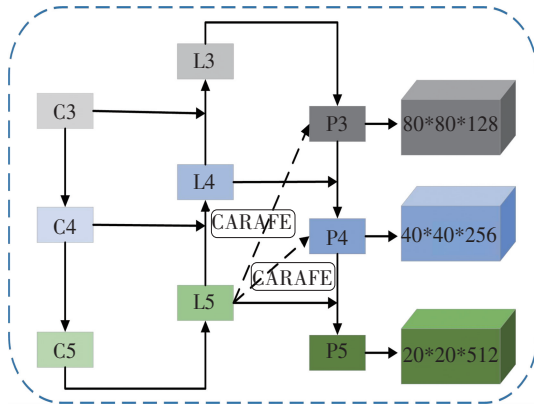


图 4 跨层级特征融合

Fig. 4 Cross-level feature fusion

此外,在特征上采样过程中,使用轻量化上采样算子 CARAFE^[17]。如图 5 所示,该模块是由两个主要部分组成:一个是上采样核预测模块,另一个是内容感知重组模块。

在上采样核预测模块中,首先对特征图的通道数进行压缩,然后利用 $K \times K$ 大小的卷积层对压缩后的特征图进行编码,并将编码后的特征图进行 α 倍上采样预测并扩展得到 $\alpha H \times \alpha W \times K^2$ 大小的特征图,最后归一化所有通道上的上采样核。内容感知重组模块中,在原始输入的特征图中取出 $K \times K$ 的区域,与通道的上采样核进行点积操作,最后得到上采样特征图。

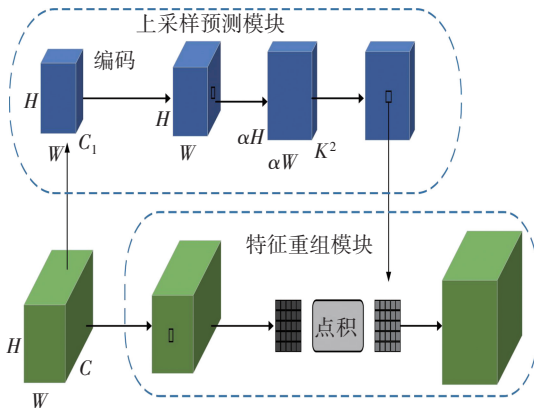


图 5 CARAFE 模块结构

Fig. 5 CARAFE module structure

1.3 集成损失函数

在遥感目标检测中,检测头的损失函数由 3 部分组成:置信度损失,负责衡量模型对目标存在与否的预测准确度;分类损失,用于评估模型对目标类别的分类性能;定位损失,用于量化模型预测的目标位置与实际位置之间的差异。传统的 Yolov5s 中,定位损失由 CIoU^[18] (Generalized Intersection over Union) 来计算,考虑预测框与真实框之间的重叠区域,并引入宽高比等因素,CIoU

损失函数的计算公式如下:

$$L_{CIoU} = 1 - L_{IoU} + \frac{\rho^2(a, a^{gt})}{C^2} + \lambda \nu \quad (5)$$

式(5)中, L_{IoU} 为两个目标框交集与并集的值, $\rho^2(a, a^{gt})$ 为预测框中心点与真实框中心点之间的欧氏距离, λ 为平衡指标参数, ν 为宽高比一致性参数。

针对以上问题,在遥感图像小目标检测中,需要更全面地反映预测框和真实框在尺度上的差异,并考虑重叠区域和中心点距离,从而更好地指导模型学习目标框的准确位置。此外,实际数据集中高质量锚框的数量很少,而低质量锚框的数量很多,通常会影响模型的训练结果。为此,决定使用 EIoU^[19] 和 Focal Loss^[20] 相结合作为正负样本分配的度量值,提出一个适合遥感图像小目标检测的集成损失函数 ILF (Integrated Loss Function)。集成损失函数的计算公式如下:

$$L_{ILF} = \lambda_1 L_{EIoU} + \lambda_2 L_{Focal-Loss} \quad (6)$$

式(6)中: λ_1 、 λ_2 是比例融合系数,与遥感数据集图像特点以及小目标与中大型目标数目相关。

2 实验及结果分析

2.1 实验数据集

选用 VisDrone2019 数据集进行算法验证,该数据集既蕴含了城市的繁华景象,又涵盖了乡村的自然风光。此外,数据集中的图像拍摄于多种天气和光照条件下,有助于在各种复杂环境中进行测试。VisDrone2019 数据集包含 10 个不同的类别,且数据集上的物体尺寸较小、目标数量多,使得目标检测和分类难度较大。数据集由 7 019 张训练集图像,548 张验证集图像与 1 610 张测试集图像组成。为了增强数据集的大小和多样性,采用图像切割方法进行数据扩充。具体来说,将原始数据集中的图片按照相邻块之间间隔 200 像素的方式裁剪成多个 $1\ 024 \times 1\ 024$ 像素的子图像。裁剪之后的图片总共为 23 750 张,从中随机选择 19 000 张图片作为训练集,2 375 张图片作为新的验证集,以及 2 375 张图片作为测试集。

2.2 实验环境及参数设置

实验环境为 Windows10 操作系统,虚拟环境选择 Python3.8 和 CUDA11.2。实验设备方面,采用 i5-12490F 处理器和 NVIDIA RTX 3090 (24GB 显存)。在实验设置中,初始学习率(lr)设置为 0.03,使用随机梯度下降法更新权值,权重衰变(decay)设置为 0.0005。输入图像的分辨率为 640×640 ,批处理大小(batchsize)设置为 16,迭代训练(epoch)统一设置为 200 代。

2.3 评价指标

在实验中,模型的检测性能通过平均精度(Average Precision, AP)和平均精度均值(mean Average Precision, mAP)这两个指标来综合评估。具体而言,AP 用于衡量单一类别在检测任务中的精度,而 mAP 则是将所有类别的 AP 求平均值,从而提供一个全局的评估指标。在评估模型的轻量化方面,采用参数量、运算量(GFLOPs)和模型体积等评价指标。参数量反映模型在训练过程中所需的参数总数,是评估模型复杂度的

重要依据。GFLOPs 用于衡量模型的计算复杂性。模型体积表示的是训练完成后模型权重文件的大小,直接关系到模型在部署时的存储要求。

2.4 消融实验

为了验证无人机遥感数据集中各模块改动点对小目标检测的实际效果,在相同的实验环境下,使用 VisDrone2019 数据集对不同模块进行算法检测性能的对比。在此过程中,选取 YOLOv5s 作为基准模型,并输出训练 200 个 epoch 后的结果。具体对比结果如表 1 所示。

表 1 改进后的算法在 VisDrone2019 数据集的消融结果

Table 1 Ablation results of the improved algorithm in the VisDrone2019 dataset

算 法	重构 C3	特征融合	ILF	精确率/%	召回率/%	mAP/%	GFLOPs	参数量/MB
M				46.8	34.7	33.4	16.0	7.06
M-1	✓			48.2	35.2	37.8	16.8	3.28
M-2		✓		48.5	33.6	35.6	19.6	8.12
M-3			✓	47.1	34.9	34.8	16.0	7.06
M-4	✓	✓		49.3	37.5	38.7	21.3	3.90
M-5	✓	✓	✓	49.8	39.5	39.2	21.3	3.90

M 算法以 YOLOv5s 为基准模型,经过消融实验验证,其网络结构重构后展现出卓越的性能提升。具体来说,平均检测精度显著提高至 39.2%,同时模型的参数量相比未重构模型大幅度减少,降幅达到 44.7%。这一结果充分证明了网络结构轻量化改进具有显著效果,不仅提升了检测精度,还优化了模型的复杂度。模型 M-1 在重构 C3 模块后,其检测精度提升明显。与基准模型 M 相比,M-1 的 mAP 提升了 4.4%,同时参数量下降了 53%,这进一步证明了 C3 模块重构后的有效性。而在模型 M-2 中,通过增加跨层级特征融合,检测精度得到一定的提升,相较于

基础模型提升了 2.2%。不过,这种特征融合过程需要将不同层的特征图进行结合,因此会引入额外计算量和参数量。至于模型 M-3,使得 mAP 进一步提升 1.4%。这一结果表明,ILF 损失函数能够有效提高检测精度。综上所述,改进后的算法在遥感小目标检测中实现了性能与效率的双重提升。

2.5 算法评估

为了更全面估改进后算法的通用性能,进行了一系列对比实验,将优化后的算法与当前主流的目标检测算法进行对比分析。在实验中,主要评估检测精度(mAP)这一指标。结果如表 2 所示。

表 2 改进后模型在 VisDrone2019 数据集中实验结果

Table 2 Experimental results of the improved model in the VisDrone2019 dataset

Model	AP/%										mAP/%
	ped	people	bicycle	car	van	truck	tricycle	awning	bus	motor	
RetinaNet	12.3	4.3	3.1	36.7	19.1	11.7	8.6	6.2	20.1	9.7	13.2
Faster R-CNN	19.8	14.4	7.8	42.3	27.3	19.5	13.8	7.1	29.7	14.5	19.6
YOLOv3	33.2	24.1	8.1	69.6	32.2	19.1	14.7	5.8	42.4	19.9	26.9
YOLOv4	36.5	33.6	14.8	76.5	41.6	41.5	20.9	13.6	40.4	37.1	35.7
YOLOv5s	32.3	28.3	9.5	74.2	39.7	42.3	22.1	9.3	41.6	34.8	33.4
YOLOv7 ^[13]	44.3	35.2	14.7	80.9	45.7	42.5	23.1	12.1	51.6	37.9	38.3
UAV-YOLOv8 ^[15]	47.6	42.4	16.8	81.5	45.9	41.8	23.4	14.2	48.5	44.9	40.7
本文算法	46.8	38.4	15.1	79.3	44.9	40.8	22.6	13.9	48.8	41.4	39.2

通过表 2 可知,改进后的算法与主流算法相比在平均检测精度方面表现卓越,达到 39.2%,相比于其他主流算法如 RetinaNet、Faster R-CNN、YOLOv3、YOLOv4、YOLOv5s、和 YOLOv7,其提升幅度分别为 26%、19.6%、12.3%、3.5%、5.8% 和 0.9%。对于 pedestrian、people、bicycle、awning 和 motor 5 个类别,该算法取得了 46.8%、38.4%、15.1%、13.9% 和 41.4% 的 AP 值。改进后的算法在各类小目标检测中表现出色,证实了所提出方法在小目标检测领域的有效性。表 3 为改进后模型在 VisDrone2019 数据集中的参数对比。

表 3 改进后模型在 VisDrone2019 数据集中参数对比

Table 3 Parameter comparison of the improved model in the VisDrone2019 dataset

算 法	FPS	容量/MB
RetinaNet	15	80
Faster R-CNN	21	136.7
YOLOv3	35	37
YOLOv4	42	244
YOLOv5s	102	13.7
YOLOv7 ^[13]	68	156
UAV-YOLOv8 ^[15]	84	23.4
本文算法	124	6.2

根据表 2 和表 3 的数据可以得出,虽然 UAV-YOLOv8 在检测精度上最为卓越,但在综合性能上相对较弱。改进后的算法在接近 UAV-YOLOv8 检测精度的同时,其检测速度每秒多了 51 帧,模型容量减少了 17.2 MB,展现出多个方面的优势。相比之下,YOLOv4 和 YOLOv7 在检测精度上较低,且模型复杂度较高、检测速度较慢。RetinaNet 和 YOLOv3 算法由于特定的网络结构,对小目标的检测受到了限制。通过与常见算法的性能对比分析,改进后的算法在轻量化、实时性能和检测精度上取得了较好的平衡,在实际工业级别轻量化应用以及部署中更有优势。

2.6 参数设置对实验结果的影响

为了验证不同参数设定对改进后模型性能的影响,在 VisDrone2019 数据集上进行了实验验证。表 4 展示在训练时不同学习率和权重衰减率对平均精度均值(mAP)的影响,其中 lr 为初始学习率,d 为权重衰减率(decay)。

通过表 4 数据分析,当初始学习率(lr)设置为 0.03 时,模型的平均精度均值达到最高。相比之下。当学

习率提高到 0.04 时,平均精度均值呈现出下降趋势。在权重衰减率(decay)方面,当设置为 0.000 5 时,平均精度均值也达到了峰值。综合考虑,当初始学习率(lr)设置为 0.03,并将权重衰减率(decay)设置为 0.000 5 时,模型的检测精度达到最高。

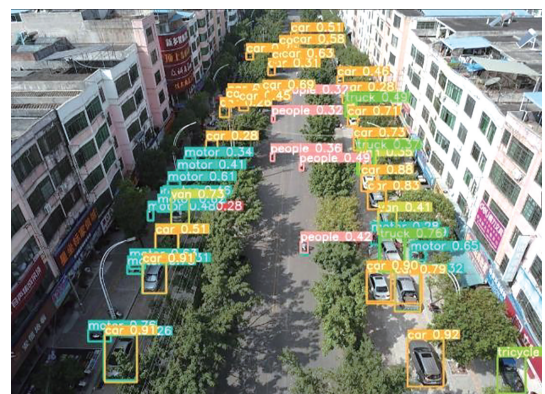
表 4 在 VisDrone2019 数据集上验证不同参数对检测精度结果的影响

Table 4 Validation of the effect of different parameters on detection accuracy on the VisDrone2019 dataset

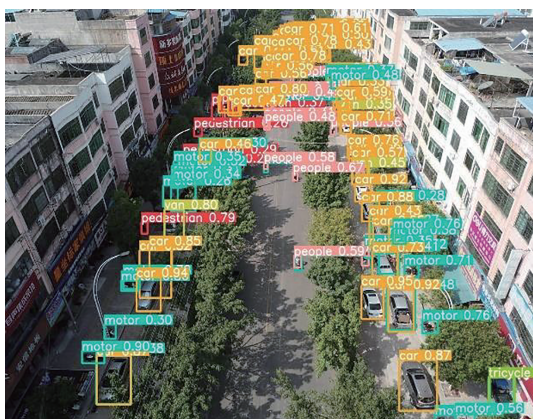
参 数	mAP/%
lr 为 0.02	38.1
lr 为 0.03	38.7
lr 为 0.04	37.4
decay 为 0.000 4	38.3
decay 为 0.000 5	38.9
decay 为 0.000 6	38.5
本文算法	39.2

2.7 实验效果与分析

遥感图像中有着背景复杂多样、目标聚集以及目标较小等检测难点,这些因素导致传统目标检测算法常出现误检和漏检情况。为了验证改进后算法在不同场景中的检测效果,使用 VisDrone2019 测试图片进行检测。图 6 所示的是密集场景小目标检测情况。实验结果表明,相较于基准算法,改进后算法在面临检测目标密集与复杂背景的场景下,能够更准确地检测出更多的小目标,从而展现出更高的检测性能。图 7 所示的是在昏暗光线环境下的检测效果。经过算法改进后,检测准确率获得了显著提升,同时误检率与漏检率均有所降低,从而证明了算法在复杂光线条件下检测性能的优越性。



(a) 基准模型密集场景检测结果



(b) 改进模型密集算法检测结果

图 6 密集场景检测情况对比

Fig. 6 Comparison of detection performance in a dense scene



(a) 基准模型夜间场景检测结果



(b) 改进模型夜间场景检测结果

图 7 昏暗光线场景情况对比

Fig. 7 Comparison of detection performance in a low-light scene

总体而言,相较于基准算法,改进后的算法展现了更为突出的优势,特别是面对小目标密集且背景复杂的场景,其检测能力得到了显著增强,表现出更高的性能和准确性。

3 结论与展望

提出一种改进 YOLOv5s 的轻量化遥感图像小目标检测方法。该方法通过重构主干网络 C3 模块,在减少模型参数数量的同时,将更多的注意力聚焦于密集小目标区域,以提高小目标特征的提取能力。此外,提出跨层级特征融合并引入上采样算子,充分利用浅层信息进行特征融合,并提升了计算效率。同时,设计了集成损失函数,解决了预测框模糊性和难易样本预测局限性的问题,并有助于模型更快地收敛。在 VisDrone2019 数据集上的实验结果显示,改进后的模型相较于基准模型在检测精度上实现了显著提升,并成功降低了模型的参数量。改进后的算法不仅提升了检测精度,还大幅度减少了算法复杂度,提高了模型检测速度,从而展现出更高的综合性能。然而,尽管改进后的算法在准确性和轻量化之间实现了较好的平衡,但与最新的 YOLOv8 和 YOLOv10 等更复杂的模型相比,平均精度仍存在一定差距。此外,在极端天气条件下,改进后模型在小目标检测方面的表现仍需提升。

未来,为了进一步提高检测精度和鲁棒性,需要探索更为先进的神经网络架构和学习策略,以增强在极端天气下的小目标检测能力。此外,计划引入更多类型的数据增强技术和半监督方法,以克服训练数据量有限和标注成本昂贵的问题。

参考文献(References):

[1] 苏立晨, 赵浩然, 郭通, 等. 基于动态分治的大规模多场站无人机应急救援优化方法[J]. 北京邮电大学学报, 2024, 47(1): 65-71.
 SU Li-chen, ZHAO Hao-ran, GUO Tong, et al. Optimization method for large-scale multi-site unmanned aerial vehicle emergency rescue based on dynamic divide-and-conquer strategy [J]. Journal of Beijing University of Posts and Telecommunications, 2024, 47(1): 65-71.

[2] 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.
- [3] GIRSHICK R. Fast R-CNN [C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2015: 1440–1448.
- [4] JIANG P, ERGU D, LIU F, et al. A review of YOLO algorithm developments [J]. Procedia Computer Science, 2022, 199: 1066–1073.
- [5] 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. Piscataway: IEEE Press, 2016: 779–788.
- [6] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Piscataway: IEEE Press, 2021: 2778–2788.
- [7] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6517–6525.
- [8] WANG Y, WANG C, ZHANG H, et al. Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery[J]. Remote Sensing, 2019, 11(5): 531.
- [9] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2020: 1571–1580.
- [10] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770–778.
- [11] SINGH B, DAVIS L S. An analysis of scale invariance in object detection-SNIP [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 3578–3587.
- [12] HAN J, DING J, XUE N, et al. ReDet: a rotation-equivariant detector for aerial object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 2786–2794.
- [13] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 7464–7475.
- [14] 张学锋, 王子琦, 汤亚玲. 基于 YOLO-CDF 神经网络的安全帽检测[J]. 重庆工商大学学报(自然科学版), 2022, 39(4): 32–41.
ZHANG Xue-feng, WANG Zi-qi, TANG Ya-ling. Helmet detection based on YOLO-CDF neural network[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2022, 39(4): 32–41.
- [15] 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.
- [16] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3–19.
- [17] WANG J, CHEN K, XU R, et al. CARAFE: content-aware ReAssembly of FEatures [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 3007–3016.
- [18] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993–13000.
- [19] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146–157.
- [20] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 2980–2988.

责任编辑:李翠薇