doi:10.3772/j.issn.1002 - 0470.2026.06.005

基于 YOLOv8n 的航拍地面目标检测改进算法①

黄 成② 林生佐③ 吴建东

(广东环境保护工程职业学院实训中心 佛山 528216)

摘 要 针对传统目标检测算法在航拍图像小目标检测中效果不佳的情况,本文基于单次目标检测器第8代最小(you only look once version 8 nano,YOLOv8n)模型进行了改进。在C2f(Conv 2 fusion)模块中引入滑动窗口变换器(Swin Transformer)模块和多尺度特征融合技术,提高了模块对不同尺寸目标的特征提取能力;提出了平方距离交并比(squared distance intersection over union,SDIoU)损失函数,通过对角距离平方与平均面积比的优化策略,使定位预测更为准确。设计了多维小目标检测头,从多个维度对目标进行检测,大幅减少小目标的漏检。在VisDrone 2019数据集上进行实验,改进后的模型在关键性能指标mAP@50和mAP@50:95上,较原模型分别提高了5.8%和3.7%。

关键词 YOLOv8n;目标检测;航拍

近年来,无人机的应用领域日益拓宽,现已深度融入航拍摄影、环境监测、灾难救援、电力巡检等诸多场景[1]。作为无人机核心应用之一,航拍图像的目标检测在诸多领域发挥了重要作用。然而,航拍图像因其复杂的背景、多样的内容以及目标小、分辨率低等特点,为精确检测带来了诸多挑战。尤其是在航拍图像中,目标容易相互遮挡,同时受光照条件和天气因素影响显著,这些因素进一步增加了检测的难度。因此,航拍图像的小目标检测已演变为计算机视觉的关键研究要点之一[2]。

向征等人[3]提出的小目标多尺度轻量化第8代单阶段检测器(small-target multi-scale you only look once version 8 nano,SM-YOLOv8n)新增了小目标增强金字塔(small object enhance pyramid,SOEP)模块并引人双卷积融合模块调制可变形卷积网络第2版(Conv 2 fusion-modulated deformable convolution network version 2,C2f-MDCNv2),但依旧使用完整交并比(complete intersection over union,CIoU)损失函数[4],导致小目标边界框回归精度不足。韦柳梅等人[5]提出的多尺度动态混合第8代单阶段检测器(multi-scale dynamic hybrid YOLOv8,MDH-YOLOv8)用焦点高效交并比(focal-efficient intersection over union,Focal-EloU)代替CIoU,引入了小目标注意力增强(small-object attention enhancement,SAE)和嵌入可变形卷积的双卷积融合(Conv 2 fusion-deformable convolution network,C2f_DCN)模块,还增加了针对小目标检测的小目标专用检测头(small-target detection head,STDH)模块,但模块对小目标的检测能力提高有限。Nie等人[6]以YOLOv8n为基础,新增小目标检测层以保留细节特征,引入尺度敏感特征融合(scale-sensitive feature fusion,SSFF)模块并设计高效路径聚合网络(high-efficiency path aggregation network,HPANet)替代路径聚合网络(path aggregation network,PANet),有效降低了参数量,但HPANet的移除降低了模型对多尺度目标共存的适配性。钟帅和王丽萍[7]提出双向动态改进单阶段检测器(bi-directional dynamic improved YOLO,BDI-YOLO)模型,通过加权双向特征金字塔网络(bi-di-

① 广东省2024年度普通高校重点科研平台(2024ZDZX4152)资助项目。
② 男,1996年生,硕士,高级工程师;研究方向:计算机视觉;E-mail:309709075@qq.com。
③ 通信作者,E-mail:atchessgd@163.com。
(收稿日期:2025 - 09 - 01)

rectional feature pyramid network, BiFPN)和动态检测头(dynamic detection head, DyHead)增强特征提取能力,并优化损失函数为内部约束 CIoU,使模型的平均精度均值(mean average precision, mAP)@50 和 mAP@50∶95 分别提升3.8%和2.7%。王殿伟等人[8]改进双头区域卷积神经网络(region-convolutional neural network, RCNN),引入 Transformer 和可变形卷积模块,结合集成内容感知特征重组上采样的特征金字塔网络(feature pyramid network, FPN)提升小目标特征提取能力。实验结果显示,平均精度(average precision, AP)、AP50 和 AP75 分别提升 2.6%、6.2%和2.1%,但参数量和每秒帧率(frame per second, FPS)不太理想。Wang 等人[9]介绍了一种基于高效多尺度注意模块(efficient multi-scale attention, EMA)的 C2f-E 结构,以提高网络对遥感图像中不同大小物体的检测能力,但实验结果表明提升的幅度十分有限。朱彦等人[10]在 YOLOv8n 中使用空间转深度卷积(space-to-depth convolution, SPDConv)模块来提高主干网络对各尺度特征的提取能力,并引入多尺度注意力 EMA 和高效交并比(efficient intersection over union, EIOU)损失函数,改进后模型的 mAP@50 有所提高。崔家礼等人[11]对 YOLOv8 采用可见光和红外线双分支结构,引入 BiFPN 融合模块和点交并比(point IoU, PIoU)第 2 版损失函数以提高网络的检测能力,但实验结果表明提升的幅度比较有限。Sharshar 等人[12]结合大尺度选择性卷积核网络和基于扩散模型的目标检测框架,针对航拍图像中的小目标检测问题,通过消融实验优化模型。实验结果显示,平均精度达到 45.7%,较 RCNN 提升4.7%。

为了让模型在航拍图像检测中实现更高的精度和更低的误检率,本文进行了 3 项改进,主要贡献如下所述。

(1)提出了滑动窗口变换器-双卷积融合(Swin Transformer-C2-fusion, ST-C2f)模块,引入了异构卷积组合机制和 Swin Transformer[13]模块,显著提高了对航拍图像中小目标的特征提取能力。

(2)提出了距离平方交并比(square distance intersection over union, SDIoU),通过引入对角距离平方与平均面积比的优化策略,提升了模型的定位精度和鲁棒性,可以更精确地定位航拍图像中小目标。

(3)设计了多维小目标检测头(multi-dimensional small object detection head, MD-Head),并融入了语义特征模块,提高模型对特征的识别能力,减少误检。

1 算法改进

1.1 C2f 模块改进

YOLOv8n 的 C2f 模块在航拍图像的小目标处理中表现不理想,随着下采样次数的增加,小目标的特征信息容易在传递过程中丢失,使模型对小目标的识别能力不足。C2f 的特征整合策略在处理不同尺度目标时整合不够充分,未能有效结合大目标、背景以及上下文信息,缺乏全局特征的提取能力,从而限制了模型的精确度和适应性。另外,在背景复杂的场景中 C2f 层容易受到噪声的干扰,无法专注于目标的关键特征区域,导致模型的检测精度下降。因此,本文提出了 ST-C2f 模块,引入了异构卷积组合机制和 Swin Transformer 模块,显著提高了对小目标的特征提取能力。

由于全局自注意力的计算量非常巨大,Swin Transformer 模块基于窗口自注意力[14],在局部窗口内计算特征的自注意力,并通过窗口平移机制整合全局的上下文信息。这样一来,Swin Transformer 模块就可以有效地弥补原始 C2f 模块在细节捕捉以及全局信息整合上的不足。

具体实现上,对于一个输入特征图 $\boldsymbol{X}=\boldsymbol{R}^{H\times W\times C}$,其中 H 为高、W 为宽、C 为通道数,先进行通道压缩得到 $\boldsymbol{X}'$,减少后续计算量,同时保留核心特征,然后进行通道分割,将其分为局部特征分支 $\boldsymbol{X}'_{\mathrm{L}}$ 和全局特征分支得到 $\boldsymbol{X}'_{\mathrm{G}}$。

$$\boldsymbol{X}'=\mathrm{Conv}(\boldsymbol{X}) \tag{1}$$

$$\boldsymbol{X}'_{\mathrm{L}}=\boldsymbol{R}^{H\times W\times C/2} \tag{2}$$

$$\boldsymbol{X}'_{\mathrm{G}}=\boldsymbol{R}^{H\times W\times C/4} \tag{3}$$

局部分支采用双流异构卷积,3×3 深度可分离卷积负责提取像素级细节,可以有效捕捉更多的细节特征及其他局部特征,大幅增强了模块对小目标局部特征的提取能力。7×7 空洞卷积主要负责扩大感受野,提高在复杂背景中的特征整合能力。

全局分支引入 Swin Transformer 模块,其窗口自注意力将特征图划分为不重叠的局部窗口,在每个

窗口内独立计算自注意力,从而显著降低计算复杂度。将输入特征图 $\boldsymbol{X'}_G$ 划分为不重叠的局部窗口 $\boldsymbol{X}_i = \boldsymbol{R}^{M\times M\times C'/2}$。

$$\boldsymbol{X} = \text{SplitWin}(\boldsymbol{X'}) = [\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n] \tag{4}$$

式中:SplitWin( )为划分窗口函数,$n = HW/M^2$。

在每个窗口内,注意力 WinA 计算方法为

$$\text{WinA}(\boldsymbol{X}_i) = \text{SelfA}(\boldsymbol{X}_i\boldsymbol{P}_Q, \boldsymbol{X}_i\boldsymbol{P}_K, \boldsymbol{X}_i\boldsymbol{P}_V) \tag{5}$$

其中窗口自注意力机制 SelfA 为:

$$\text{SelfA}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d}}\right)\boldsymbol{V} \tag{6}$$

式中:$\boldsymbol{P}_Q$、$\boldsymbol{P}_K$、$\boldsymbol{P}_V$ 是投影矩阵矩阵;$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 是输入特征通过投影矩阵生成的 3 个向量,分别对应查询矩阵、键矩阵和值矩阵;d 为通道数;Softmax( )是一个函数,用于将 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 的点积结果转化为概率分布,使注意力权重归一化。

为了捕获全局信息,Swin Transformer 引入窗口平移机制,周期性地移动窗口位置,使相邻窗口之间的信息能够交互。

$$\text{SW}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{WinA}(S(\boldsymbol{Q}), S(\boldsymbol{K}), S(\boldsymbol{V})) \tag{7}$$

式中:SW( )表示窗口平移机制,S( )表示平移变换。最后通过动态门控融合模块,将全局分支和局部分支的输出进行融合,并用卷积恢复通道数,得到输出特征图,如图 1 所示。

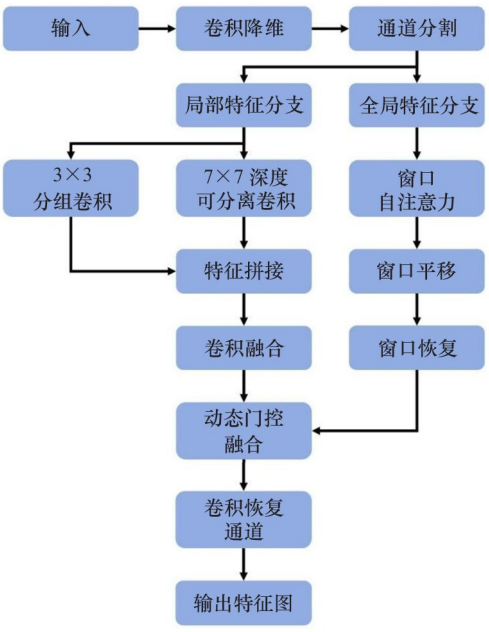

图 1　ST-C2f 结构

1.2　改进损失函数

在模型的训练过程中,损失函数可以直接影响模型对目标的定位精度和分类准确性。YOLOv8n 的回归损失结合了交并比(intersection over union, IoU)损失和分布焦点损失(distribution focal loss, DFL)2 种损失函数,其中,IoU 损失以 CIoU 作为损失函数[15]。CIOU 损失函数在考虑了交并比的基础上,还加入了预测框与真实框的中心点距离以及宽高比的信息,能够更全面地衡量两者之间的差异,从而在目标定位中更加准确。然而,CIoU 在航拍图像这类复杂图像上的表现并不理想。因为 CIoU 主要考虑的是中心点距离和宽高比等因素,而航拍图像中存在大量分辨率低的小目标,模型难以精准预测中心点。为了提高模型对小目标定位和修正的准确度,本文提出了 SDIoU,通过引入对角距离平方与平均面积比的优化策略,提升了模型的定位精度和鲁棒性。它利用预测框和真实框在 2 个对角点上的距离差异,构建更为精准的约束机制。其主要原理如下:

$$\text{SDIoU} = 1 - \text{IoU} + \frac{d_1^2 + d_2^2}{S_t + S_p} + S/4 + \alpha \times v \tag{8}$$

$$\text{IoU} = \frac{A \cap B}{A \cup B} \tag{9}$$

$$d_1^2 = (x_1^B - x_1^A)^2 + (y_1^B - y_1^A)^2 \tag{10}$$

$$d_2^2 = (x_2^B - x_2^A)^2 + (y_2^B - y_2^A)^2 \tag{11}$$

$$S = \left(\frac{x_1^B - x_1^A}{x_1^A - x_2^A}\right)^2 + \left(\frac{y_1^B - y_1^A}{y_1^A - y_2^A}\right)^2 + \left(\frac{x_2^B - x_2^A}{x_1^A - x_2^A}\right)^2 + \left(\frac{y_2^B - y_2^A}{y_1^A - y_2^A}\right)^2 \tag{12}$$

$$\alpha = v/(1 - \text{IoU} + v + \text{eps}) \tag{13}$$

$$v = \left(\frac{4}{\pi^2}\right) \cdot \left(\arctan\left(\frac{w_A}{h_A}\right) - \arctan\left(\frac{w_B}{h_B}\right)\right)^2 \tag{14}$$

式中:(x_1^A, y_1^A),(x_2^A, y_2^A),(x_1^B, y_1^B),(x_2^B, y_2^B) 分别为真实框 A 和预测框 B 左上角和右下角点的坐标;S_t 和 S_p 分别为真实框和预测框的面积;esp 为一个非常小的非零常数值,用于避免出现除零错误。

便于理解公式,如图 2 所示的 IoU 示意图。

在对损失函数进行改进时,为了验证改进损失函数的有效性,本文将多个常见损失函数应用到模型进行对比实验,包括距离交并比(distance IoU, DIoU)[16]、

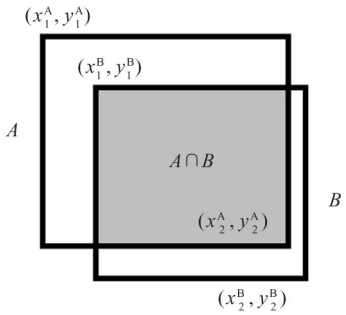

图 2　IoU 示意图

加权交并比(weighted IoU, WIoU)[17]、广义交并比(generalized IoU, GIoU)[4]、多路径距离交并比(multi-path DIoU, MPDIoU)[18],实验结果如表 1 所示。

表 1　损失函数对比实验

损失函数	mAP@50	mAP@50∶95
DIoU	0.353	0.204
WIoU	0.354	0.205
GIoU	0.353	0.205
MPDIoU	0.356	0.206
SDIoU	0.360	0.210

从实验结果可知,SDIoU 损失函数在 mAP@50 和 mAP@50∶95 2 个指标上均明显优于其他损失函数。具体来看,大部分损失函数的 mAP@50 相近,在 0.353 到 0.354 范围内波动,MPDIoU 的表现较好,但仍落后于 SDIoU 的 0.360;至于更严格的 mAP@50∶95指标,SDIoU 以 0.210 的表现,明显超越其他损失函数,充分证明了 SDIoU 改进的有效性。

1.3　多维小目标检测头

因为原 YOLOv8 主要是面向一般图像的目标检测,其检测头自然也是针对一般尺寸目标进行设计,而航拍图像的内容通常比普通图像更复杂多样,原检测头难以充分挖掘和利用多尺度特征信息,导致模型的检测效果不理想。针对这一挑战,本文参考基于注意力机制统一目标检测头的动态检测头[19]设计了一个多维小目标检测头,并融入了语义特征模块,见图 3。

常见的航拍场景可能既包括大尺度的建筑、巴士等目标,也包含尺度较小的自行车、行人等微小目标。针对这一特性,本文在检测头中引入了多尺度感知注意力机制,通过多尺度特征提取模块,充分捕捉小目标的细节信息并保留大目标的全局信息。考

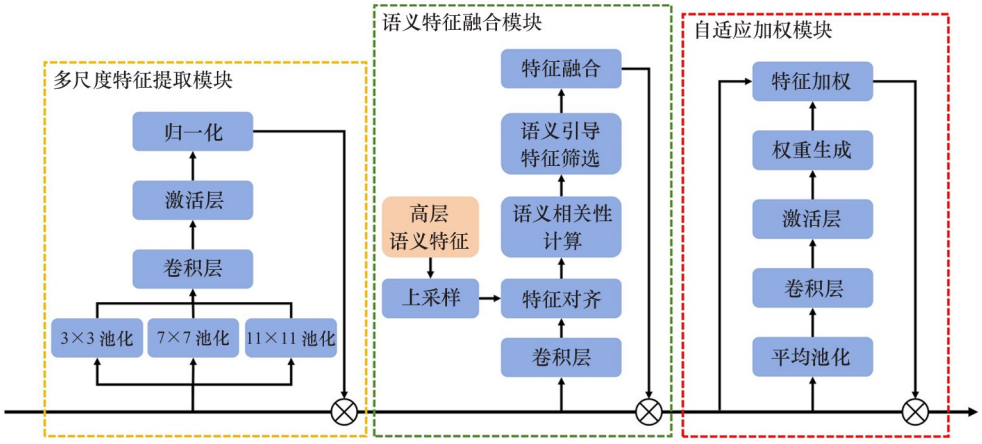

图 3　MD-Head 结构图

虑到小目标细节容易被高层语义特征覆盖,该检测头采用语义引导门控融合机制,对高层语义特征进行上采样重构并激活生成权重矩阵,用于标记高层特征中与小目标相关的语义信息(如轮廓、移动性特征等)。最后,对特征进行自适应加权,通过注意力机制实现对特征图中各通道权重的动态调整,使模型聚焦于目标物体的关键特征,从而提升检测的准确性与可靠性。

对于特征图 $\boldsymbol{X} = R^{H \times W \times C}$,在传统方案中,大多数时候都是直接使用一个全连接层来实现这个注意力函数。然而,鉴于张量具有较高维度,若直接进行全维度注意力机制的学习,会导致巨量的计算开销。

鉴于此,本文采用分阶段学习策略,将注意力函数拆解为 3 个串行的模块,最后同样得到多维的特征金字塔[20],从而有效降低计算成本。

$$F_{\text{out}}(\boldsymbol{X}) = F_3(F_2(F_1(\boldsymbol{X}))) \tag{15}$$

式中:$F_{\text{out}}(\ )$表示检测头的输出,$F_1(\ )$、$F_2(\ )$、$F_3(\ )$分别表示多尺度特征提取模块函数、语意特征提取模块函数以及自适应加权模块函数。

$F_1(\ )$中引入了多尺度感知注意力,可以充分提取不同尺度的特征。

$$F_1(\boldsymbol{X}) = \text{Nor}(\text{ReLU}(F_{\text{Conv}}(P_3(\boldsymbol{X}) \oplus P_7(\boldsymbol{X}) \oplus P_{11}(\boldsymbol{X})))) \tag{16}$$

式中:$P(\ )$表示平均池化,$F_{\text{Conv}}(\ )$表示卷积层运算,ReLU(\ )表示激活层,Nor(\ )表示归一化处理。

语意特征融合模块先通过预处理,将高层语义特征 $\boldsymbol{S}$ 与输入特征 $\boldsymbol{X}$ 进行对齐,通过矩阵乘法计算 $\boldsymbol{S}$ 和 $\boldsymbol{X}$ 的语义相似度,生成筛选权重 $\boldsymbol{Sim}$,接着对筛选权重归一化,进行语义引导特征筛选与融合。

$$\boldsymbol{Sim}(x, y, c) = \boldsymbol{X}(x, y, c) \cdot \boldsymbol{S}(x, y, c) \tag{17}$$

$$F_2(\boldsymbol{X}) = \text{BN}(\text{ReLU}(F_{\text{Conv}}(\boldsymbol{X} \odot \text{Nor}(\boldsymbol{Sim})))) \tag{18}$$

式中:BN(\ )表示归一化,$\boldsymbol{X} \odot \text{Nor}(\boldsymbol{Sim})$表示将归一化的筛选矩阵应用于 $\boldsymbol{X}$。

最后一个模块 $F_3(\ )$,主要对经过前序处理的特征进行动态权重分配,通过聚焦关键语义区域(如小目标所在位置)并抑制背景噪声,提升特征的判别性。该模块针对航拍场景中小目标易被复杂背景淹没的问题,设计了基于全局上下文的权重生成机制,使模型能自适应调整不同特征重要性。对于输入特征 $\boldsymbol{X}$,通过全局平均池化压缩特征的空间维度,提取每个通道的全局统计信息。

$$F_{\text{全局平均池化}}(\boldsymbol{X}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x(i, j, :) \tag{19}$$

式中:$(i, j, :)$ 表示在 (i, j) 位置上,所有通道的特征向量。

接着,对全局特征进行非线性变换,生成与输入特征尺寸匹配的空间权重矩阵 $\boldsymbol{W}$。最后将生成的权重矩阵与输入特征进行逐元素相乘,实现对关键区域的增强和噪声区域的抑制。

$$F_3(\boldsymbol{X}) = \boldsymbol{X} \odot F_{\text{激活}}(F_{\text{Conv}}(F_{\text{全局平均池化}})) \tag{20}$$

式中:$F_{\text{全局平均池化}}$、F_{Conv}、$F_{\text{激活}}(\ )$分别表示自适应加权模块的全局平均池化运算、卷积层运算和激活层运算。

2　实验结果

2.1　实验数据集

本研究采用 VisDrone 2019[21]数据集,该数据集专为无人机视觉任务构建,包含训练集、验证集与测试集 3 部分。数据集中涵盖了行人、自行车、汽车、货车、公交车等 10 类不同目标,每张图像均附带详细标注,涵盖了目标的位置坐标、尺寸大小等特征,为算法检测结果的量化评估提供了标准化依据。

此外,VisDrone 2019 数据集具备显著的多样性与复杂性,其图像样本覆盖了不同天气条件、光照强度、拍摄角度及飞行高度下的场景,能够精准模拟各类复杂实际环境,从而为验证算法的有效性与鲁棒性提供了全面支撑。

2.2　实验环境

本研究的实验环境如表 2 所示。

表 2　实验环境

环境	型号
操作系统	Windows10
CPU	Intel Core i5-13600KF
GPU	NVIDIA GeForce RTX 4060 Ti
RAM	16 GB
深度学习框架	PyTorch(2.1.1)
编译环境	Python(3.9.18)
CUDA	CUDA(12.1)

2.3　消融实验

为了验证本文提出的 3 个改进的有效性,本文选择 YOLOv8n 作为基线模型,使用 VisDrone 2019 数据集,逐一增加 3 个改进,从而评估每个改进对模型检测小目标性能的影响。

表 3　模型改进的消融实验

组别	模型	mAP@50	mAP@50:95	GFLOPS
A	YOLOv8n	0.348	0.204	8.1
B	A + ST-C2f	0.359	0.211	9.7
C	A + SDIoU	0.360	0.210	8.1
D	A + 小目标检测头	0.391	0.233	14.7
E	D + ST-C2f + SDIoU	0.406	0.241	16.1

注:GFLOPS 表示每秒十亿次浮点运算。

实验结果表明,应用 ST-C2f 改进后模型的 mAP@50 和 mAP@50∶95 达到了 0.359 和 0.211,2 个指标均优于原模型的 0.348 和 0.204。当 SDIoU 应用到模型时,其相关指标也有所提高,而且模型的复杂度并没有增加。将小目标检测头应用到模型时,改进效果最为明显,mAP@50 提升至 0.391,提升了 4.3%,mAP@50∶95 提升至 0.233,提升了 2.9%。当 3 个模块同时使用时,模型的 mAP@50 和 mAP@50∶95 分别达到了 0.406 和0.241,提升了 5.8% 和 3.7%,模型也表现出了最佳的检测性能,不足的是计算量有所增加。

2.4　性能曲线对比

图 4 和图 5 分别展示了 YOLOv8n 模型和本研究改进后模型的性能评估曲线。精确率-召回率(precision-recall,PR)曲线层面,改进后模型的每一条曲线都优于原模型。该结果由于 ST-C2f 模块深度可分离卷积的细节捕捉以及 MD-Head 的语义引导筛选的相互协同作用,有效增强了模型对小目标的辨识度。召回率-置信度(recall-confidence,RC)曲线则反映了模型的召回率随置信度的变化情况,SDIoU 通过对角距离平方与平均面积比约束,使小目标的边界框预测更精准,配合MD-Head的多尺

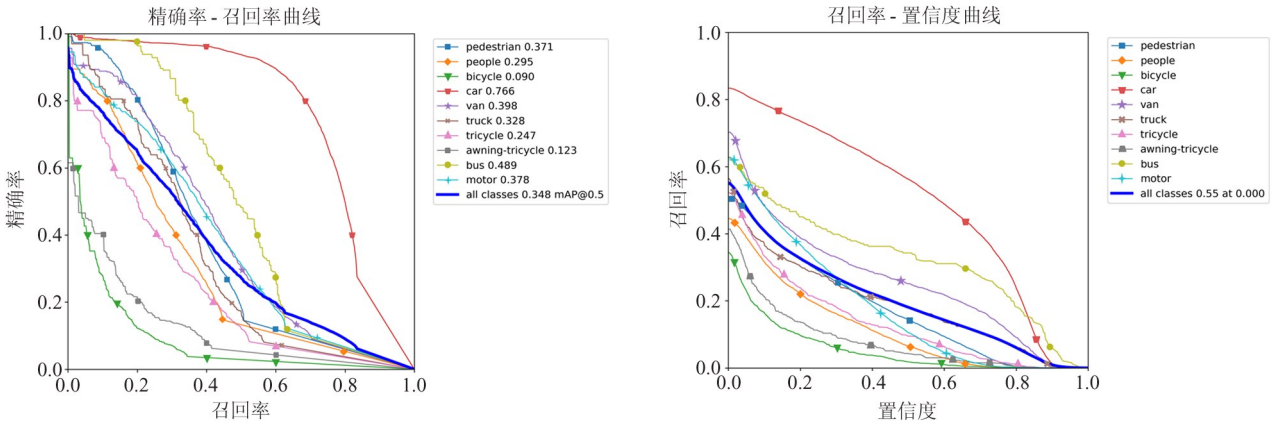

图 4　YOLOv8n 性能曲线

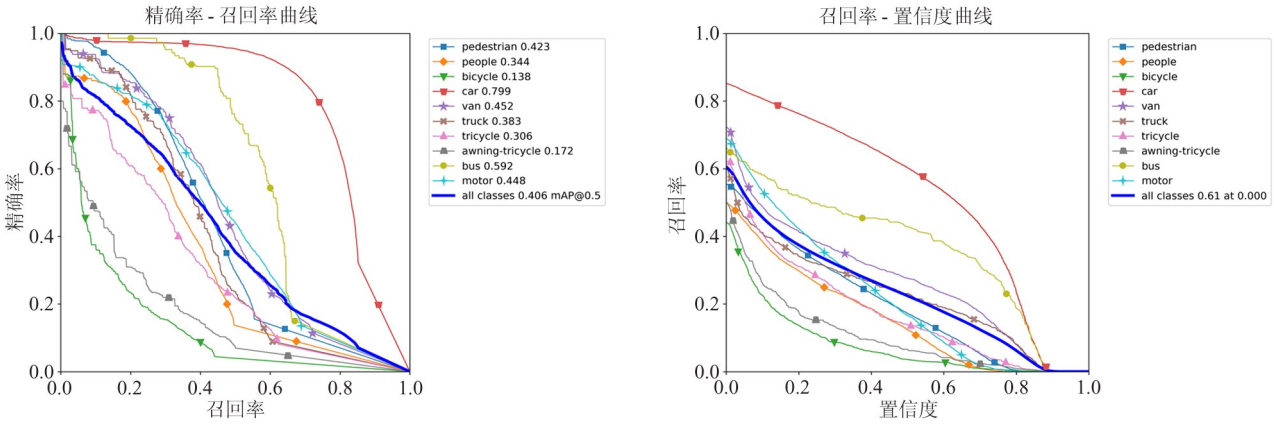

图 5　改进模型的性能曲线

度感知注意力,使小目标在低置信度下仍能有效识别,所以改进模型整体召回率在全置信度区间均高于原模型。通过对比可知,改进后模型在精确率与召回率上实现双重提升,可有效降低航拍小目标检测中的误检与漏检数量。

2.5　可视化对比

图 6～图 8 分别展示了 YOLOv8n 模型和改进后模型的检测结果图以及标签图。通过对比可以发现,改进后模型在多个目标预测上给出了更高的置信度,而且准确率也更高,如左上角场景下方,图 6

中检测为“truck0.5”的目标实际标签为“car”,图 7 中检测正确;同样,右下角场景的下方,图 6 的标签虽然有遮挡,但可以看到是 2 个“motor”标签,而通过图 8 标签可知,应为一个“motor”标签和一个“bicycle”标签,图 7 则将这 2 个目标正确检出。通过结果图的对比,进一步证明了性能曲线对比的结论,改进后模型有更高的精度并可以有效降低误检及漏检数量。

图 6　YOLOv8n 检测结果

图 7　改进模型检测结果

图 8　图片标签

为更客观地体现模型性能,将本文模型与其他研究改进的 YOLOv8n 模型进行对比,各模型在 VisDrone 2019 数据集上的表现如表 4 所示。实验结果表明,本文改进模型的 mAP@50 和 mAP@50∶95

达到了 0.406 和 0.241,在这 2 项指标上都大幅优于其他模型,但由于在模型的检测头增加了多尺度特征提取模块、语义特征融合模块及自适应加权模块,导致参数增加约 1.6×10^6。

表 4 对比实验

模型	mAP@50	mAP@50:95	参数/$\times10^6$
YOLOv8n	0.348	0.204	3.00
SM-YOLOv8n[3]	0.374	0.225	3.30
MDH-YOLOv8[5]	0.375	0.227	3.00
Nie 等人[6]	0.383	0.226	1.97
本文模型	0.406	0.241	4.70

3 结 论

针对 YOLOv8n 模型在航拍图像检测中存在的性能瓶颈,本文从 3 个维度提出改进方案:通过优化 ST-C2f 模块增强小目标特征表征能力,提出 SDIoU 损失函数提升边界框回归精度,并设计了多维小目标检测头强化对低像素目标的捕捉能力。实验结果表明,改进后的模型在 VisDrone 2019 数据集上的性能显著提升。其中,mAP@50 达到 0.406,较原模型提升 5.8%,mAP@50∶95 达到 0.241,提升 3.7%。说明上述改进有效解决了航拍场景中小目标特征稀疏、检测精度不足的问题,为无人机视觉任务提供了更优的技术方案。

需客观指出的是,当前改进模型的参数量较原模型有所增加,不利于模型推理效率。因此,后续研究将围绕“精度-效率平衡”展开,重点探索轻量化网络结构设计。

参考文献

[1] 沈海军. 无人机的发展与应用[J]. 科学新闻,2023,25(1):42 - 45.

[2] 钟帅,王丽萍. 无人机航拍图像目标检测技术研究综述[J]. 激光与光电子学进展,2025,62(10):71 - 89.

[3] 向征,张佳浩. 基于 SM-YOLOv8n 的无人机航拍目标检测[J]. 海军航空大学学报,2025,40(2):321 - 328.

[4] Rezatofighi H, Tsoi N, Gwak J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE,2019:658 - 666.

[5] 韦柳梅,罗雪梅,康健. 改进 YOLOv8 的航拍图像小目标检测算法[J]. 计算机工程与科学,2025,47(5):882 - 890.

[6] Nie H, Pang H, Ma M, et al. A lightweight remote sensing small target image detection algorithm based on improved YOLOv8[J]. Sensors, 2024,24(9):2951.

[7] 钟帅,王丽萍. 改进 YOLOv8 的无人机航拍图像目标检测方法[J]. 航空兵器,2025,32(3):45 - 52.

[8] 王殿伟,胡里晨,房杰,等. 基于改进 Double-Head RCNN 的无人机航拍图像小目标检测算法[J]. 北京航空航天大学学报,2024,50(7):2141 - 2149.

[9] Wang H,Yang H,Chen H,et al. A remote sensing image target detection algorithm based on improved YOLOv8[J]. Applied Sciences, 2024,14(4):1557.

[10] 朱彦,张月霞. SEP-YOLO: 基于 YOLOv8 改进的道路目标检测算法[J]. 计算机应用与软件,2025,42(6):78 - 85.

[11] 崔家礼,王涵,郑瀚,等. 基于 BPP-YOLOv8 的多模态目标检测算法[J]. 红外技术,2025,47(9):1021 - 1029.

[12] Sharshar A, Matsun A. Innovative horizons in aerial imagery: LSKNet meets DiffusionDet for advanced object detection[EB/OL]. (2023 - 11 - 21)[2025 - 09 - 11]. https://arxiv.org/pdf/2311.12956.

[13] Liu Z,Lin Y,Cao Y,et al. Swin Transformer: hierarchical vision transformer using shifted windows[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada:IEEE,2021:10012 - 10022.

[14] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]// Advances in Neural Information Processing Systems. Long Beach, USA:IEEE,2017:5998 - 6008.

[15] Zheng Z, Wang P, Liu W, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA:IEEE,2020:9937 - 9946.

[16] Zheng Z,Yan S,Yang Y. Distance-IoU loss: faster and better training of object detection models[EB/OL]. (2019 - 11 - 19)[2025 - 09 - 11]. https://arxiv.org/pdf/1911.08287.

[17] 殷波. 基于 ASF-WIoU-YOLOv8 的无人机航拍图像多目标检测算法[J]. 计算机科学与应用, 2024,14(5): 137 - 146.

[18] Siliang M, Yong X. MPDIoU: a loss for efficient and accurate bounding box regression[EB/OL]. (2023 - 07 - 14)[2025 - 09 - 11]. https://arxiv.org/pdf/2307.07662.

[19] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE, 2017:2117 - 2125.

[20] Dai X, Chen Y, Xiao B, et al. Dynamic head: unifying object detection heads with attentions[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Honolulu,USA:IEEE,2021:7373 - 738.

[21] Zhu Z, Wen L, Du D, et al. Detection and tracking meet drones challenge[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Seoul, Korea:IEEE,2019:1901 - 1910.

Improved algorithm for aerial ground target detection based on YOLOv8n

Huang Cheng, Lin Shengzuo, Wu Jiandong

(Training Center, Guangdong Environmental Protection Engineering Vocational College, Foshan 510000)

Abstract

To address the poor performance of traditional object detection algorithms in small object detection for aerial images—attributed to challenges like low target resolution and complex backgrounds—this study proposes three key improvements based on the YOLOv8n (you only look once version 8 nano) model. First, the Swin Transformer module and multi-scale feature fusion technology are embedded into the original C2f (conv 2 fusion) module, forming an enhanced feature extraction unit that strengthens the model's capability to capture fine-grained features of objects across different scales, particularly small ones. Second, a square distance intersection over union (SDIoU) loss function is developed; by integrating an optimization strategy that combines squared diagonal distance and average area ratio, this function effectively reduces bounding box regression errors, enabling more accurate localization of small targets. Third, a multi-dimensional small object detection head is designed to analyze targets from semantic, spatial, and scale dimensions, which significantly mitigates the missed detection of small objects in aerial scenes. Comprehensive experiments were conducted on VisDrone 2019, a widely recognized benchmark dataset for aerial object detection. The results demonstrate that the improved model achieves mAP@50 of 0.406 and mAP@50:95 of 0.241, representing increases of 5.8% and 3.7%, respectively, compared to the original YOLOv8n model. This outcome validates that the proposed improvements effectively balance detection accuracy and efficiency for small object detection in practical aerial scenarios.

Key words: YOLOv8n, object detection, aerial photography