doi:10.3772/j.issn.1002-0470.2024.07.010

基于改进 YOLOv8 的无人机图像小目标检测算法^①

李云文②* 冯宇平③* 安文志* 赵 军* 聂国潇**

(*青岛科技大学自动化与电子工程学院 青岛 266061) (**烟台东方威思顿电气有限公司 烟台 264000)

摘要 针对无人机(UAV)图像中目标尺寸小及特征信息少导致检测精度低的问题,提出改进YOLOv8n的小目标检测算法。首先,引入Wise-IoU损失函数,通过动态非单调聚焦机制增强网络对普通质量锚框的关注,提高泛化能力。然后,添加小目标检测层(SODL)和双向特征金字塔网络(BiFPN)构建新的特征融合结构——Bi-SODL结构。SODL使网络更充分地捕捉小目标的浅层特征信息,BiFPN可以实现不同尺度特征层之间的信息交流和融合,提高小目标检测的准确性。最后,添加大型可选择模块(LSKBlock)注意力机制,通过空间选择机制和加权方式对输入特征进行处理,进一步提高小目标检测的性能和鲁棒性。实验结果表明,在VisDrone2019数据集上的检测精度指标P、mAP_0.50 和 mAP_0.50:0.95分别提升6.4%、8.3%和5.2%,并且参数量降低25.78%。改进措施使得检测性能优于众多主流算法,证明了改进算法的有效性。

关键词 小目标检测: YOLOv8n: 特征融合: 注意力机制

由于无人机具有操作简单和灵活机动等优势^[1],无人机航拍被广泛应用于城市交通等领域^[2]。但无人机图像中目标尺寸较小、特征信息不易提取会导致检测精度低下,因此提高无人机图像小目标检测的准确率成为研究热点。

卷积神经网络(convolutional neural network, CNN)的出现使基于深度学习的目标检测算法成为主流算法^[3],此类算法分为两阶段和单阶段。例如快速区域卷积神经网络(fast region-based CNN, Fast R-CNN)^[4]和 Faster R-CNN^[5]的两阶段算法先生成候选区域,再对候选区域进行分类和回归。例如单次多框检测器(single shot multibox detector, SSD)^[6]和只看一次(you only look once, YOLO)^[7]的单阶段算法将检测视为回归问题,直接输出目标的类别和位置信息。相较而言,单阶段算法具有更快的速度,但在精度上稍有牺牲。然而随着算法不断的改进,

YOLO 算法兼顾了精度和速度,能够有效应对无人机图像小目标检测任务,因此本文采用 YOLO 算法进行小目标检测。

为了提高对无人机图像小目标的检测性能,国内外学者做了一系列研究。Tang等人^[8]提出改进的 YOLOv5 算法,在主干和颈部之间添加对合块并在主干网络末端引入卷积块注意模块(convolutional block attention module, CBAM),同时添加小目标检测头,提高对小目标的检测能力。Zhu等人^[9]提出改进的飞桨 YOLO(paddle paddle YOLO,PP-YOLO)算法,在颈部网络添加额外的通路,加强不同层级的特征信息交流,同时使用 CutMix 数据增强算法替换Mixup,提高网络的泛化能力。Wang等人^[10]提出YOLOX-w算法,使用切片辅助超推理(slicing aided hyper inference,SAHI)算法进行数据增强,并引入小目标检测头和超轻量子空间注意力模块,最后优化

① 国家自然科学基金(61971253)和国家级大学生创新创业训练项目(202310426296,202310426356)资助。

② 男,2000 年生,硕士生;研究方向:模式识别与机器视觉;E-mail: 2682555143@qq.com。

③ 通信作者,E-mail: qustfyp@126.com。 (收稿日期:2024-01-26)

损失函数,提高检测精度。Liu 等人^[11]提出 Bi-YO-LO 算法,引入 BiFormer 注意力机制,同时使用分组随机卷积(group shuffle convolution, GSConv) 替换Conv,并将跨阶段局部模块(cross stage partial, CSP) 替换为 VoV-GSCSP,达到平衡检测精度和速度的效果。

上述方法对小目标检测的性能有一定改善,但依旧存在对小目标特征信息利用不充分的问题。因此本文提出改进 YOLOv8n 的小目标检测算法,主要工作如下:

首先,使用序列证据加权插值交并比(weighted inter polation of saquential evidence for intersection over union, Wise-IoU),将小梯度增益分配给低质量锚框,削弱低质量样本的影响;其次,添加小目标检测层(small object detection layer, SODL)获取更多小目标的特征信息来提高对小目标的检测精度;然后,引入双向特征金字塔网络(bidirectional feature pyramid network, BiFPN)构成新的特征融合结构 Bi-SODL,使特征信息在不同层级之间双向传播;最后,添加大型可选择模块(large selective kernel block, LSKBlock)注意力机制,使网络自适应地调整感受野大小而后对输入特征进行处理。

经实验验证,本文算法在 VisDrone2019 数据集上的精度指标有明显改善,虽然模型复杂度稍有增加,但满足实时性需求,实现检测精度和速度的平衡。

1 YOLOv8 网络结构

YOLOv8 是 Ultralytics 发布的最新版本的 YOLO 算法,按网络深度与宽度划分为 n,s,m,l 和 x 共 5 个版本。它主要包含主干网络(Backbone)、颈部网络(Neck)和检测头(Head)3部分。在 Backbone中,沿用 CSP 思想,但使用梯度流更丰富的 C2f 模块替换 C3 模块,并将每个阶段的模块数从[3,6,9,3] 修改为[3,6,6,3]。在 Neck 中采用特征金字塔(feature pyramid network, FPN)和路径聚合网络(path aggregation network, PAN)结构并将 C3 模块替换为 C2f 模块。在 Head 中使用解耦头结构将分类

和检测头分离,并换成无锚框检测。损失函数 (loss)方面,抛弃交并比(intersection for union, IoU) 匹配或者单边比例的分配方式,使用对齐分配器 (Task-Aligned Assigner)正负样本匹配方式,并引入分布聚焦损失(distribution focal loss, DFL)。YOLOv8 网络结构如图 1 所示。

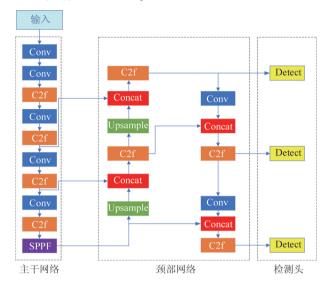


图 1 YOLOv8 网络结构

2 改进的 YOLOv8 网络结构

本文改进的 YOLOv8 网络结构如图 2 所示。首先,使用 Wise-IoU 替换完全交并比(complete IoU, CIoU),降低低质量样本的影响。然后,在 Backbone的第一个C2f块后添加SODL以增强网络提取浅层

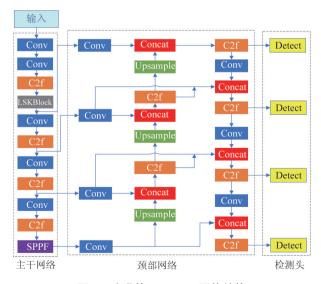


图 2 改进的 YOLOv8 网络结构

特征信息的能力,并引入 BiFPN 到添加了 SODL 的 Neck 部分,构成新的特征融合结构 Bi-SODL,通过 多尺度特征融合来充分利用小目标特征信息。最后,在 Backbone 的第一个 C2f 块后加入 LSKBlock 注意力机制以增强网络的感受野,让网络更好地学习特征信息增强检测的准确性。

2.1 损失函数

YOLOv8 默认使用 CIoU^[12]作为损失函数,其原理如图 3 所示,记真实框为 $b_{gt} = [x_{gt}, y_{gt}, w_{gt}, h_{gt}]$, 预测框为 b = [x, y, w, h], 2 个框的最小包裹框(图 3 黑色虚线框区域)为 B。CIoU 计算公式如式(1)~(4)所示。

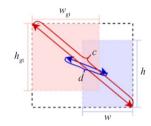


图 3 CloU 计算原理

$$CIoU = IoU - \frac{d^2}{c^2} - \alpha v = IoU - \frac{\rho^2(b, b_{gt})}{c^2} - \alpha v$$

(1)

$$\nu = \frac{4}{\pi^2} \left(\arctan\left(\frac{w_{\text{gt}}}{h_{\text{eq}}}\right) - \arctan\left(\frac{w}{h}\right) \right)^2$$
 (2)

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{3}$$

$$LOSS_{CIoU} = 1 - IoU + \frac{\rho^2(b, b_{gt})}{c^2} + \alpha v$$
 (4)

式中, IoU 为真实框与预测框的重叠程度, d^2 和 $\rho^2(b,b_{\rm gt})$ 为真实框和预测框两中心点的欧氏距离, c^2 为最小包裹框的对角线距离, α 和 v 分别为权重

系数和长宽比的相似性。

CIoU 将真实框和预测框的长宽比考虑在内,解决了由于预测框面积相同但长宽比不同而导致 IoU 值相同的问题,但忽略了低质量样本引起的负面影响。因此本文使用 Wise-IoU 替换 CIoU。Wise-IoU 是一种动态非单调调频的损失函数,在降低高质量锚框竞争力的同时,也减少了低质量样本产生的有害梯度,这使得 Wise-IoU 专注于普通质量的锚框,从而提高检测器的整体性能^[13]。Wise-IoU 的计算公式如式(5)、(6)所示。

$$\beta = \frac{L_{\text{loU}}^*}{\overline{L_{\text{loU}}}} \in [0, + \infty)$$
 (5)

$$L_{\text{WIoUv3}} = \frac{\beta}{\delta r^{\beta - \delta}} \exp\left(\frac{(x - x_{\text{gt}})^2 + (y - y_{\text{gt}})^2}{(W_g^2 + H_g^2)^*}\right) L_{\text{IoU}}$$
(6)

式中, β 代表离群度, L_{loU}^* 代表变量 L_{loU} 转换成的常量, $\overline{L_{\text{loU}}}$ 是动量 m 的运行平均值,r 和 δ 代表学习参数, $(x-x_{\text{gt}})^2+(y-y_{\text{gt}})^2$ 代表真实框和预测框两中心点的欧氏距离, $(W_g^2+H_g^2)^*$ 为最小包裹框的对角线距离。

2.2 Bi-SODL 结构

YOLOv8 的 Neck 部分结合 FPN 和 PAN 进行设计,先下采样操作,然后上采样操作,最后进行特征融合完成特征提取。然而,由于 YOLOv8 的下采样倍率较高,导致浅层特征图中的细节信息容易丢失,这使得 YOLOv8 对小目标检测的准确率较低。针对这一问题,本文设计一种新的特征融合结构 Bi-SODL,该结构添加 SODL 并结合 BiFPN^[14],实现小目标信息的跨尺度连接和特征融合。上述网络结构简图如图 4 所示。

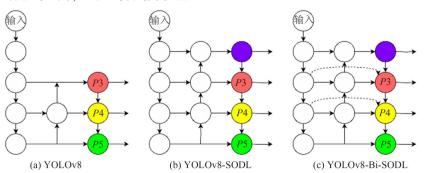


图 4 YOLOv8、YOLOv8-SODL、YOLOv8-Bi-SODL 网络结构简图

图 4 中 *P*3、*P*4 和 *P*5 分别代表小型、中型和大型的目标检测器,它们的检测层输出特征图尺度分别为 80×80、40×40和 20×20。 *P*2则代表 SODL 中的小目标检测头,其特征图尺度为 160×160。

YOLOv8 网络结构简图如图 4(a) 所示。YOLOv8 使用下采样、上采样、FPN 和 PAN 等操作来实现目标检测任务。首先通过下采样将输入图像尺寸缩小,得到高分辨率的特征图,再通过上采样将特征图恢复到原始大小,以便进行目标位置的精确定位。同时引入 FPN 和 PAN,通过融合不同尺度的特征来提供语义信息。

由于 P3、P4 和 P5 输出的特征图尺度过小,导致小目标信息丢失严重。因此,添加 SODL 来捕获小目标特征信息,YOLOv8-SODL 网络结构简图如图 4(b)所示。YOLOv8-SODL 将 160 × 160 的特征图融入特征融合阶段,捕获最浅层的特征信息,以此来增加网络的感受野,捕捉更多周围像素的上下文信息,从而提高对小目标的理解能力。

添加 SODL 可以扩展网络的感受野,但 YOLOv8 网络中的 FPN 只存在单向的信息流动,会出现特征融合不充分和特征丢失等问题。因此,结合 BiFPN 进行二次改进,提出 YOLOv8-Bi-SODL 网络结构,其结构简图如图 4(c)所示。在 YOLOv8-Bi-SODL 的

SODL、P3、P4 和 P5 的特征输入节点和中间特征提取节点添加通道数 256 的 1 × 1 的卷积核来降低特征图的维度,并在 P3 和 P4 层的原始输入和输出节点之间增加额外的连接,以保证特征融合的有效性和双向跨尺度连接。该结构采取双向特征融合的机制,允许特征在自顶向下和自底向上 2 个方向上进行融合,能够在不同层级之间更高效地融合特征。然后使用加权融合机制为每个输入特征添加权重用于优化特征融合效果,计算公式如式(7)所示。

$$O = \sum_{i} w_i \cdot I_i \tag{7}$$

式中,0是输出结果;i表示第i个特征; I_i 是输入特征; w_i 是一个可学习的权重,可以是标量(每个特征)、向量(每个通道)或多维张量(每个像素)。

2.3 LSKBlock 注意力机制

LSKBlock 是源于 LSKNet 的注意力机制^[15],通过空间选择机制来选择不同尺度的大型卷积核,以增强网络对空间上下文区域的关注能力,从而提高小目标检测的准确率。LSKBlock 注意力机制主要由大核卷积(Large Kernel Convolutions)和空间核选择(Spatial Kernel Selection)2部分构成,结构图如图 5所示。

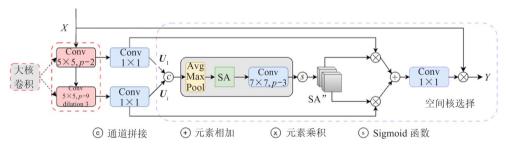


图 5 LSKBlock 结构

Large Kernel Convolutions 中引入分组卷积和空洞卷积对输入进行处理。输入 X 首先会经过大小为 5×5 、填充 (padding) 为 2、分组 (groups) 为维度 (dim) 的卷积核,这种分组卷积能够分别对输入的每个通道进行卷积,增加模型的非线性能力和灵活性。然后再经过大小为 7×7 、padding 为 2、groups 为 dim、扩张 (dilation) 为 3 的卷积核,在分组卷积的同时进行扩张率为 3 的空洞卷积,增加卷积核的有

效尺寸,提高网络的感受野。Large Kernel Convolutions 可以快速增大网络的感受野,使网络能够更好地理解图像的上下文信息,同时分组卷积和空洞卷积也能够减少参数量和计算量,提高网络的效率。

Spatial Kernel Selection 使用空间选择机制从不同尺度的大卷积核中对特征图进行空间选择。首先,将来自于不同感受野卷积核的特征 U_i 进行拼接:

$$U = [U_1, \cdots, U_i] \tag{8}$$

然后,通过平均池化和最大池化(记为 $P_{\text{avg}}(\cdot)$ 和 $P_{\text{max}}(\cdot)$) 操作对特征 U 进行空间关系提取,得到特征描述符 SA_{avg} 和 SA_{max} :

$$SA_{\text{avg}} = P_{\text{avg}}(U)$$
, $SA_{\text{max}} = P_{\text{max}}(U)$ (9)
其中, SA_{avg} 表示平均池化空间特征符, SA_{max} 表示最大池化空间特征符。

为了允许不同空间描述符之间的信息交互,将之前得到的空间特征拼接起来,并使用卷积层 $F^{2-N}(\cdot)$ 将拼接的特征(2个通道)变换为 N 个空间注意力图 SA':

$$SA' = F^{2 \to N}([SA_{\text{avg}}; SA_{\text{max}}])$$
 (10)

将 Sigmoid 激活函数应用到每个空间注意力图,可获得每个解耦的大卷积核所对应的独立的空间选择掩膜:

$$SA''_{i} = \sigma(SA'_{i}) \tag{11}$$

式中, $\sigma(\cdot)$ 为 Simoid 函数。然后,将解耦后的大卷 积核序列的特征与对应的空间选择掩膜进行加权处理,并通过卷积层 $F(\cdot)$ 进行融合获得注意力特征 S:

$$S = F(\sum_{i=1}^{N} (SA''_{i} \cdot U_{i}))$$
 (12)

最终的输出Y可通过输入特征X与注意力特征S的逐元素点乘获得,即:

$$Y = X \cdot S \tag{13}$$

3 实验结果与分析

3.1 实验环境与参数配置

本文实验环境如表 1 所示。相关参数设置如下:输入图像尺寸为 640×640,训练轮数为 300,批大小(batchsize)为4,初始学习率为0.01,优化器选

表1 实验环境

环境	训练、测试环境配置					
CPU	$Intel(R)Core(TM)\ i7\text{-}11700\ CPU\ 2.\ 50\ GHz$					
GPU	NVIDIA GeForce RTX2080Ti(11 GB)					
操作系统	Ubuntu18.04					
PyTorch	PyTorch1.13.1					
CUDA	CUDA11.6					
Python	Python3.8					

用随机梯度下降(stochastic gradient descent,SGD)。

3.2 数据集

本文在 VisDrone2019^[16] 数据集上进行实验验证,该数据集共包含 8 629 张图像,用于训练、验证和测试的图像分别为 6 471 张、548 张和 1 610 张,包含日常生活中的行人、汽车和摩托车等 10 个类别。该数据集类别实例分布情况和各标注目标的长宽分布如图 6 和 7 所示,可以看出数据集中各个类别比例不均衡且物体大多数为小目标。

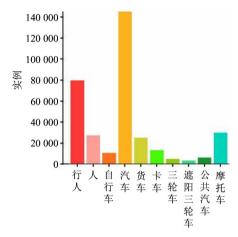


图 6 类别实例分布

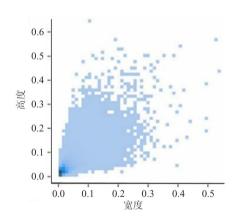


图 7 标注目标的长宽分布

3.3 评价指标

本文实验用到的评价指标主要包括:精准率 (Precision, P)、召回率(Recall, R)、平均精度均值 (mean average precision, mAP)、模型参数量(Params)、每秒检测帧数(FPS)和浮点运算次数 (FLOPs)。P是预测为正样本中实际正样本的比例,计算公式如式(14)所示; R是在所有实际为正类的

样本中,被模型正确预测为正类的比例,计算公式如式(15)所示; mAP用于评估模型在不同类别上的检测准确度,计算公式如式(16)和式(17)所示,本次实验以 mAP_0.50和 mAP_0.50:0.95 作为性能评价指标,分别代表 IoU 阈值为 0.50 时的 mAP 值和 IoU 阈值在 0.50 到 0.95 范围内变化时计算的 mAP值; Params 是模型中参数的数量; FPS 是模型每秒处理的图片数量,用来衡量检测速度,FLOPs 用来衡量模型的复杂度。

$$P = \frac{TP}{TP + FP} \tag{14}$$

$$R = \frac{TP}{TP + FN} \tag{15}$$

$$mAP = \frac{1}{n} \sum_{i=1}^{n} AP(i)$$
 (16)

$$AP = \int_0^1 P(R) \, \mathrm{d}R \tag{17}$$

式中,TP 为正确预测出正样本的数量,FP 为负样本被预测成正类的数量,FN 为被预测成负类的正样本数,n 为检测的类别数量,AP 为 PR 曲线(precision-recall curve)所包含的面积。

3.4 损失函数选择实验

YOLOv8 默认使用 CloU 损失函数,为确定改进的损失函数,分别使用 EloU、SloU 以及 Wise-IoU 进行对比实验,实验结果如表 2 所示。

表 2 YOLOv8 不同损失函数的实验结果

算法	P/%	<i>mAP</i> _ 0.50/%	mAP = 0.50 : 0.95 / %
CIoU	44.1	31.9	18.4
EIoU	43.5	32.9	19.2
SIoU	44.2	33.0	19.3
Wise-I	oU 44.6	33.4	19.6

由表 2 结果可知,使用 EloU、SloU 以及 Wise-IoU 替换 CloU 都可以在一定程度上提升性能。相较而言,Wise-IoU 在各项指标上的提升效果更为显著,因此选用 Wise-IoU 对 YOLOv8n 进行改进。

3.5 消融实验

为验证本文算法对小目标检测的提升效果,以YOLOv8n 算法为基础,依次添加改进模块,在 Vis-Drone2019 上进行消融实验。实验结果如表 3 所示。

表 3 消融实验结果

算法	Wise-IoU	SODL	Bi-SODL	LSKBlock	P/%	mAP = 0.50%	<i>mAP</i> _ 0.50: 0.95/%	Params / M	FLOPs/G	FPS
YOLOv8n					44.1	31.9	18.4	3.01	8.1	174.1
A	$\sqrt{}$				44.6	33.4	19.6	3.01	8.1	176.3
В	$\sqrt{}$	$\sqrt{}$			48.0	36.1	21.4	2.92	12.2	163.8
C	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$		49.3	38.9	23.3	2.23	17.1	160.4
D	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	50.5	40.2	23.6	2.23	17.4	158.9

对表 3 实验结果分析可知,通过逐步添加改进模块,如使用 Wise-IoU、添加 SODL、引入 BiFPN 以及加入 LSKBlock 注意力机制,YOLOv8n 在小目标检测性能方面有显著的提升。改进 A 使用 Wise-IoU 损失函数,使网络更专注普通质量的锚框,有效地降低低质量样本产生的有害梯度,在不增加 Params 和 FLOPs 的情况下,提升了 P、 $mAP_0.50:0.95$ 3 项指标。改进 B 继续添加 SODL,解决下采样过程中浅层信息丢失严重的问题,尽管FLOPs 和 FPS 分别变化到 12.2 和 163.8,但是 P、 $mAP_0.50$ 和 $mAP_0.50:0.95$ 分别提升了 3.4%、

2.7%和1.8%,精度有了明显提升。改进 C 继续引入 BiFPN,通过跨尺度连接和加权特征融合的方式,保证不同层级的特征能够高效融合,这使得 P、mAP_0.50 和 mAP_0.50 : 0.95 分别达到49.3%、38.9%和23.3%,3 项指标都有较为明显的提升,但也导致了 FLOPs 的增加和 FPS 的降低。改进 D 继续加入LSKBlock 注意力机制,通过使用空间选择机制来选择不同尺度的大型卷积核,以增强网络对空间上下文区域的关注能力。相较于 YOLOv8n,改进 D 的 Params 从 3.01 M 减少到 2.23 M, P、mAP_0.50 和 mAP_0.50 : 0.95 分别提升6.4%、8.3%和5.2%,

达到 50.5%、40.2% 和23.6%, 虽然 FLOPs 从 8.1 G 增加到 17.4 G, FPS 从 174.1 下降到 158.9, 但是检测精度提升明显, 能够满足实时性和准确性的需求。

为了更直观地体现本文算法的有效性,绘制出2种算法的热力图如图8(左为YOLOv8n,右为本文算法)所示,并且作图比较YOLOv8n和依次添加改进模块的改进算法在训练过程中的4项指标,结果

如图 9 所示。由图 8 可知, YOLOv8n 更关注近处物体并且聚焦于很多不相关物体,本文算法对于不相关物体的关注度显著降低并且对远处小目标的关注度更高更准确。由图 9 可知,随着依次添加改进模块,各指标都呈现出明显上升的趋势,添加所有改进模块的算法的各指标达到最高,进一步证明本文算法的有效性。

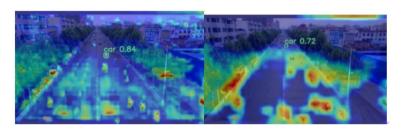


图 8 YOLOv8n 和改进算法的热力图

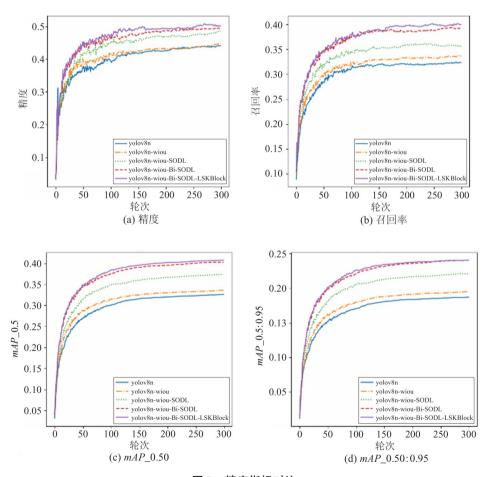


图 9 精度指标对比

为了进一步验证本文算法在 VisDrone2019 数据集上的有效性,将该数据集的训练集、验证集和测试集继续划分为5:3:2和3:4:3两种情况进行实

验,实验结果如表 4 和 5 所示。由结果可知,相较于 默认的数据集划分,因训练集比例变小,修改划分比 例后的精度指标有所下降,但随着改进模块的添加, 所有精度指标仍呈现出上升趋势,证明了本文算法 在该数据集上的有效性。

3.6 与其他算法的对比实验

为进一步验证本文算法对无人机图像小目标的

检测性能,将本文算法与其他主流算法(如 Faster-RCNN、RetinaNet^[17]等)在相同数据集、实验环境及参数配置下进行对比,实验结果如表 6 所示。

表 4 5:3:2 数据集划分实验结果

算法	Wise-IoU	SODL	Bi-SODL	LSKBlock	P/%	mAP _ 0.50/%	mAP _ 0.50:0.95/%
YOLOv8n					37.9	25.7	14.7
A	\checkmark				38.1	26.1	14.9
В	$\sqrt{}$	\checkmark			38.9	28.3	16.1
C	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$		39.8	30.3	17.3
D	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	41.3	30.9	17.6

表 5 3:4:3 数据集划分实验结果

算法	Wise-IoU	SODL	Bi-SODL	LSKBlock	P/%	mAP _ 0.50/%	mAP _ 0.50:0.95/%
YOLOv8n					32.2	21.5	11.8
A	$\sqrt{}$				32.9	22.1	12.2
В	$\sqrt{}$	\checkmark			34.5	23.7	13.1
C	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$		35.9	25.7	14.1
D	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	\checkmark	36.7	26.0	14.3

由表 6 可知,相较于 Faster-RCNN、RetinaNet 和 SSD 算法,本文算法参数量少且准确率高。相较于模型复杂、参数量大的 YOLOv5s 和 YOLOv8s,本文算法在 $mAP_0.50$ 、 $mAP_0.50$:0.952 项指标上表现优异,与 YOLOv8s 在 P 值上相比仅有 0.8% 的差距,并在检测速度上展现出明显的优势。相较于轻

量化算法 YOLOv7-tiny,本文算法的 FPS 降低了2.8,但 P、mAP_0.50 和 mAP_0.50 : 0.95 分别提升 3.2%、5.0% 和 5.4%。结合各项指标综合来看,本文算法对无人机图像小目标具有更好的检测性能,能够较好地平衡检测的精度和速度,证明本文算法 具有优越性。

表 6 对比实验结果

算法	P/%	mAP _ 0.50/%	mAP = 0.50 : 0.95 / %	Params/M	FLOPs/G	FPS
YOLOv8n	44.10	31.90	18.40	3.00	8. 10	174. 10
Faster-RCNN	45.60	33.50	19.30	41.19	206.73	23.40
RetinaNet	41.30	22.10	16.60	19.80	93.70	43.10
SSD	42.90	24.10	17.20	24.50	87.90	49.60
YOLOv5s	50.00	37.90	22.70	9.12	23.80	151.00
YOLOv7-tiny	47.30	35.20	18.20	6.03	13.30	161.70
YOLOv8s	51.30	39.00	23.50	11.13	28.50	140.30
文献[18]	49.00	38.20	20.60	55.89	60.00	
文献[19]	47.10	38.20		11.40		90.91
本文算法	50.50	40.20	23.60	2. 23	17.40	158.90

3.7 结果可视化

为验证本文算法在实际场景中的检测效果,选取 VisDrone2019 测试集中密集、多尺度、黑夜以及复杂背景等场景下的图片进行可视化,检测效果如图 10 所示(左为 YOLOv8n,右为改进算法),图中圆框区域表示检测效果有明显改善的区域。

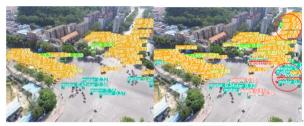










图 10 不同场景的检测效果

由图 10 可知,第 1 组图像中目标密集且尺寸小,YOLOv8n 检测到图像中有 2 个人、99 辆汽车和 7 辆摩托车,存在很多漏检目标,本文算法检测到人、汽车和摩托车分别增至 9、111 和 18,虽然依旧存在漏检问题,但检测效果有明显的改善。第 2 组图像中近远处的目标尺度差异较大,YOLOv8n 对近远处目标存在漏检严重的问题,本文算法能准确地检测到近处的摩托车和行人等目标,还能检测到更

多远处的目标。第3组图像是处于黑夜环境下,物体的很多特征信息会有所变化,对比可知,本文算法不仅能够更准确地检测到光照下的物体,而且对于黑暗环境下目标的检测也有明显改进。第4组和第5组图像受到复杂背景因素的影响,YOLOv8n检测结果中存在较多漏检和错检的情况,然而本文算法明显减少了目标漏检和错检的情况,并且能够更准确地检测到远处以及遮挡的目标。综上分析得知,改进算法对小目标检测效果更佳。

3.8 鲁棒性实验

为了充分验证本文算法的鲁棒性,继续在DOTA1.0 数据集上进行实验。DOTA1.0 包含 2 806 张 4 000 × 4 000 的图像,共 15 个类别 188 282 个目标。为了便于实验,将 DOTA1.0 中的图像裁剪扩充成 21 046 张 1 024 × 1 024 的图像,随机选择 15 749 张图片作为训练集,5 297 张图片作为测试集。Batchsize 调节为 2,其余参数设置和实验环境与 3.1 节保持一致,实验结果如表 7 所示。

表 7 DOTA 数据集实验结果

算法	P/%	mAP = 0.50%	mAP = 0.50 : 0.95 / %
YOLOv8n	72.4	63.7	41.2
本文算法	74.1	66.3	42.5

由实验结果可知,本文算法在 DOTA1.0 数据集的精度指标 $P_{\text{v}} MAP_{\text{o}} D.50$ 和 $mAP_{\text{o}} D.50$: 0.95 较改进前分别提升 1.7%、2.6% 和 1.3%,有效证明了本文算法具有良好的鲁棒性。

4 结论

针对无人机图像小目标检测漏检误检严重导致准确率低下的问题,提出改进YOLOv8n的小目标检测算法。首先使用Wise-IoU损失函数,更准确地评估目标检测框与真实目标框的匹配程度;然后在颈部网络使用Bi-SODL特征融合结构,使网络在充分捕捉浅层特征信息的同时实现多尺度特征的交流和

融合;最后添加 LSKBlock 注意力机制,通过空间选择机制和加权方式提取上下文信息并进行融合,使模型更好地理解图像中的目标并准确地进行检测。实验结果表明,与相关领域其他代表算法相比,本文算法能够有效地平衡检测的精度和速度问题,在检测性能上有优异的表现,并且具有良好的鲁棒性。

参考文献

- [1] 罗旭东,吴一全,陈金林. 无人机航拍影像目标检测与 语义分割的深度学习方法研究进展[J]. 航空学报, 2024,45(6):241-270.
- [2] SURMANN H, WORST R, BUSCHMANN T, et al. Integration of UAVs in urban search and rescue missions [C] //2019 IEEE International Symposium on Safety, Security, and Rescue Robotics. Würzburg, Germany: IEEE, 2019: 203-209.
- [3] ZOU Z, CHEN K, SHI Z, et al. Object detection in 20 years: a survey [EB/OL]. (2023-01-08) [2024-01-06]. https://arxiv.org/pdf/1905.05055.pdf.
- [4] GIRSHICK R. Fast R-CNN [C] // Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015:1440-1448.
- [5] REN S, HE K, 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, 2017,39(6):91-99.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] // Proceedings of the 14th European Conference on Computer Vision. Amsterdam, Netherlands: Springer, 2016:21-37.
- [7] 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. Las Vegas, USA: IEEE, 2016: 779-788.
- [8] TANG S, FANG Y, ZHANG S. HIC-YOLOv5: improved YOLOv5 for small object detection [EB/OL]. (2023-01-09) [2024-01-06]. https://arxiv.org/pdf/2309.16393. pdf.

- [9] 朱福珍,王帅,巫红. 改进的 PP-YOLO 网络遥感图像 目标检测[J]. 高技术通讯, 2022, 32(5):528-534.
- [10] WANG X, HE N, HONG C, et al. Improved YOLOX-X based UAV aerial photography objectdetection algorithm
 [J]. Image and Vision Computing, 2023,135:104697.
- [11] 刘子洋, 徐慧英, 朱信忠,等. Bi-YOLO: 一种基于YOLOv8 改进的轻量化目标检测算法[EB/OL]. (2023-11-08)[2024-01-08]. http://kns.cnki.net/kc-ms/detail/43.1258. TP. 20231107. 1657. 002. html.
- [12] ZHENG Z, WANG P, REN D, et al. Enhancing geometric factors in model learningand inference for object detection and instance segmentation [J]. IEEE Transactions on Cybernetics, 2022,52(8):8574-8586.
- [13] TONG Z, CHEN Y, XU Z, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism [EB/OL]. (2023-04-08)[2024-01-07]. https://arxiv.org/pdf/2301.10051.pdf.
- [14] TAN M, PANG R, LE Q V. EfficientDet: scalable and efficient object detection [C] // Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020:10781-10790.
- [15] LI Y, HOU Q, ZHENG Z, et al. Large selective kernel network for remote sensing object detection [EB/OL]. (2023-03-06) [2024-01-07]. https://arxiv.org/pdf/ 2303.09030. pdf.
- [16] YU W, YANG T, CHEN C. Towards resolving the challenge of long-tail distribution in UAV images for object detection [C] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, USA; IEEE, 2021;3258-3267.
- [17] 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. Venice, Italy: IEEE, 2017;2980-2988.
- [18] 杨秀娟, 曾智勇. 基于 YOLOv5 的无人机航拍改进目标检测算法 Dy-YOLO[J]. 福建师范大学学报(自然科学版), 2024,40(1):76-86.
- [19] 苏佳, 秦一畅, 贾泽,等. 基于 ATO-YOLO 的小目标检测算法[J]. 计算机工程与应用, 2024,60(6)68-77.

Small object detection algorithm for unmanned aerial vehicle image based on improved YOLOv8

LI Yunwen*, FENG Yuping*, AN Wenzhi*, ZHAO Jun*, NIE Guoxiao**

(*College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061)

(*Yantai Dongfang Wisdom Electric Co, Ltd, Yantai 264000)

Abstract

Aiming at the problem of low detection accuracy caused by small object size and little feature information in unmanned aerial vehicle images, an improved small object detection algorithm of YOLOv8n is proposed. Firstly, the Wise-IoU loss function is introduced, which enhances the network's focus on ordinary quality anchor frames through a dynamic non-monotonic focusing mechanism, improving the generalization ability of the algorithm. Secondly, in order to improve the accuracy of small object detection, a new feature fusion structure, the Bi-SODL structure, is constructed by adding a small object detection layer (SODL) and bidirectional feature pyramid network (BiFPN). SODL enables the network to capture the shallow feature information of the small object more adequately. BiFPN can achieve the information exchange and fusion between feature layers of different scales. Finally, LSKBlock attention mechanism is introduced, which processes the input features through spatial selection mechanism and weighting, further improving the performance and robustness of small object detection. The experimental results show that the detection accuracy metrics P, mAP_{-0} . 50 and mAP_{-0} . 50: 0.95 on the VisDrone2019 dataset are increased by 6.4%, 8.3% and 5.2% respectively, and the number of parameters is reduced by 25.78%. The improved measures make the detection performance better than many mainstream algorithms, which proves the effectiveness of the improved algorithm.

Key words: small object detection, YOLOv8n, feature fusion, attention mechanism