

基于改进 YOLOv8s 的无人机目标检测算法

曲晨阳 吕进 卫策

(长安大学信息工程学院 西安 710000)

摘要:针对目前无人机航拍图像目标尺寸较小,图像背景复杂,导致现有的无人机目标检测算法检测精度较低的问题,提出一种改进 YOLOv8s 的无人机目标检测算法。首先使用可变形卷积替换标准卷积,以增强网络对不规则形状目标的特征提取能力;然后使用可分离大核注意力机制(LSKA)改进快速空间金字塔池化(SPPF)模块,改善因目标尺度差异较大导致检测精度较低的问题。在网络颈部结合双向特征金字塔网络(Bi-FPN)实现多尺度特征融合,改善网络对小目标的漏检和错检问题。在网络头部,使用自注意力机制动态检测头(DyHead)替换原检测头,增强对遮挡物体和小目标的检测能力。最后,针对数据集中存在大量低质量样本对训练过程产生负面影响的问题,使用 Wise-IOU 损失函数,提升模型收敛速度和检测精度。实验结果表明,改进后的方法在 VisDrone2019 数据集上获得了 41.7% 的平均精度均值(mAP),与原 YOLOv8s 算法相比,mAP@0.5 提升了 3.0%,mAP@0.5 : 0.95 提升了 1.9%,参数量下降了 17.5%,计算量下降了 12.63%。实现了模型轻量化和检测精度双重提升。

关键词:无人机;目标检测;可变形卷积;YOLOv8s;注意力机制

中图分类号: TN29 文献标识码:A 国家标准学科分类代码: 520.6040

UAV target detection algorithm based on improved YOLOv8s

Qu Chenyang Lyu Jin Wei Ce

(School of Information Engineering, Chang'an University, Xi'an 710000, China)

Abstract: Aiming at the problem that the target size of the current UAV aerial images is small and the image background is complex, which leads to low detection accuracy of existing UAV target detection algorithms, this article proposes an improved YOLOv8s UAV target detection algorithm. First, deformable convolution is used to replace standard convolution to enhance the network's feature extraction ability for irregularly shaped targets. Then the separable large-kernel attention mechanism (LSKA) is used to improve the SPPF module to improve the problem of low detection accuracy due to large differences in target scales. The weighted bi-directional feature pyramid network (Bi-FPN) is combined at the neck of the network to achieve multi-scale feature fusion and improve the network's missed detection and false detection of small targets. At the head of the network, the dynamic detection head (DyHead) is used to replace the original detection head to enhance the detection ability of occluded objects and small targets. Finally, in order to solve the problem that a large number of low-quality samples in the dataset have a negative impact on the training process, the Wise-IOU loss function was used to improve the model convergence speed and detection accuracy. Experimental results show that the improved method achieved 41.7% mAP on the VisDrone2019 dataset. Compared with the original YOLOv8s algorithm, mAP@0.5 increased by 3.0%, mAP@0.5 : 0.95 increased by 1.9%, the number of parameters decreased by 17.5%, and the amount of calculation dropped by 12.63%. It achieves both model lightweight and detection accuracy.

Keywords: UAV; target detection; deformable convolution; YOLOv8s; attention mechanism

0 引言

随着无人机技术的蓬勃发展以及制造成本的降低,无人机开始从军用领域进入日常生活,给人们的日常生活和工作提供了巨大的便利。无人机凭借其体型小巧、动作灵活性高、易于操控等特点,现已在各种多元化航拍场景中都得到了应用。无人机航拍目标检测算法的研究对于实现智能化、高效化的无人机航拍具有重要意义。

无人机由于其特殊复杂的作业场景,基于无人机的目标检测会面临诸多难点。首先,无人机航拍图像背景复杂,容易有目标遮挡。无人机高空飞行视野广阔,目标检测干扰物较多,使得在复杂背景下对航拍图像中的目标进行准确定位和分类成为难题^[1]。其次,无人机航拍的高度较高,视野较广,使得图像中的待检测对象大多为小目标,相比于中大型目标,小目标的尺寸小,分辨率低,导致在图像中的信息有限,在下采样过程中容易丢失关键特征,需要对算法进行改进,以提升在小目标检测方面的性能。另外,无人机由于轻量化设计,所能搭载的计算资源有限,且无人机目标检测对于实时性有一定要求,因此减少网络参数数量和降低计算复杂程度是十分有必要的。

目前基于深度学习的目标检测主要分为两类。一种是基于区域推荐的目标检测算法,也称两阶段(two-stage)目标检测算法。这类检测算法首先使用一种区域生成方法生成可能包含目标的候选区域,对于每个候选区域,提取特征并进行分类和边界框回归,以确定是否存在目标以及目标的位置和类别。最后通过后处理步骤来消除重叠的检测结果,得到最终的目标检测结果。常见的两阶段目标检测算法包括区域卷积神经网络(region-based convolutional neural network, R-CNN)^[2]、Fast R-CNN^[3]、Faster R-CNN^[4]等。另一种是基于回归的目标检测算法,也称为单阶段(one-stage)目标检测算法。此类算法一般先将整个图像输入网络,通过卷积和其他操作提取特征。然后在特征图上直接进行分类和边界框回归,以确定图像中的目标位置和类别。最后通过阈值筛选和非极大值抑制等后处理步骤,得到最终的目标检测结果。常见的单阶段目标检测算法包括 YOLO^[5]、SSD^[6]、RetinaNet^[7]等。

伴随着无人机目标检测对于实时性和效率的要求不断提高,目前无人机目标检测算法主要以单阶段的目标检测算法为主。杨辉羽等^[8]将改进后的单聚合模块嵌入YOLOv5 网络中,解决因为网络深度造成的梯度衰减问题。裴伟等^[9]在 SSD 的基础上进行改进,通过使用空洞卷积和反卷积扩大感受野,融合了不同层的特征信息,增强了模型的泛化性能。闫钧华等^[10]使用深度卷积和跨层级通道特征融合,添加位置注意力机制,有效提升了遥感图像中弱小目标的检测精度。Yang 等^[11]针对无人机电力线巡检中绝缘子缺陷检测准确率低的问题,提出了改进的YOLOv3^[12]网络。他们使用双向融合网络来提高对小目标的检测精度。同时引入了 EIOU 损失函数^[13],使预测

框更接近真实框。钱承山等^[14]基于 Transformer 改进 YOLOv5 模型,提出了一种应用于无人机设备的轻量山火检测网络 YOLO_MC 与 YOLO_MCLite.,他们使用分组计算和通道注意力机制,结合知识蒸馏算法显著降低了网络的计算复杂度。陈朋磊等^[15]设计出一种高效的小目标特征聚合网络(SFANet),使其能够充分整合浅层特征图中小目标的细节信息。Hou 等^[16]提出了一种名为 YOLOX-Pro 的无人机航拍影像山体滑坡检测方法,基于 YOLOX^[17]进行改进,以提高在不同地貌环境下的检测精度。该方法采用 Focal Loss 损失函数^[18]来应对大小样本分布不均的问题,使用位置注意力机制(CA)提高模型的检测精度和特定区域的识别能力,提升了无人机航拍影像中山体滑坡检测的准确性和鲁棒性。

为了进一步提高小目标检测的精度,同时优化网络结构,实现网络的轻量化。本文在 YOLOv8s 网络基础上进行如下。

1) 针对无人机拍摄目标大多形状不规则,导致标准卷积对其特征提取能力较弱的问题,使用可变形卷积代替标准卷积,使用多路坐标注意力机制改善可变形卷积的学习能力,增强网络对于复杂目标的特征提取能力。

2) 针对待检测目标尺度差异较大,导致检测精度低的问题,使用可分离大核注意力机制(large separable kernel attention, LSKA)改进快速空间金字塔池化(spatial pyramid pooling-fast, SPPF)模块,提高网络的多尺度特征提取能力。

3) 由于小目标特征会随着网络层数加深而丢失,本文结合双向特征金字塔网络(bi-directional feature pyramid network, Bi-FPN)的思想,优化原网络的特征融合模块,设计一个新的特征融合模块 YOLOv8-Bi 增强网络对不同尺度目标的检测能力,同时降低计算复杂度。

4) 针对小目标和遮挡物体容易漏检或误检的问题,本文使用一个更加高效的自注意力机制动态检测头(dynamic head, DyHead)提高网络检测性能。

5) 针对数据集中的低质量样本会对训练效果产生负面影响的问题,使用 Wise-IOU 损失函数替换原来的损失函数,提升训练速度和精度。

1 YOLOv8 网络结构

YOLOv8 是目前最先进的目标检测算法之一,它在速度和准确度之间取得了良好的平衡,能够快速而准确地识别目标,适合部署于移动设备。最新版本的 YOLOv8 具有多个规格大小的模型,如 YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l 和 YOLOv8x。其总体结构包括主干网络(Backbone)、颈部网络(Neck)和检测头(Head)3 部分。网络结构如图 1 所示。

主干网络采用了 Darknet53 结构,其中包括卷积(convolution, Conv)、SPPF 和 C2f(CSPLayer with 2Conv)模块。Conv 模块用于提取和整理特征图,SPPF 模块实现了

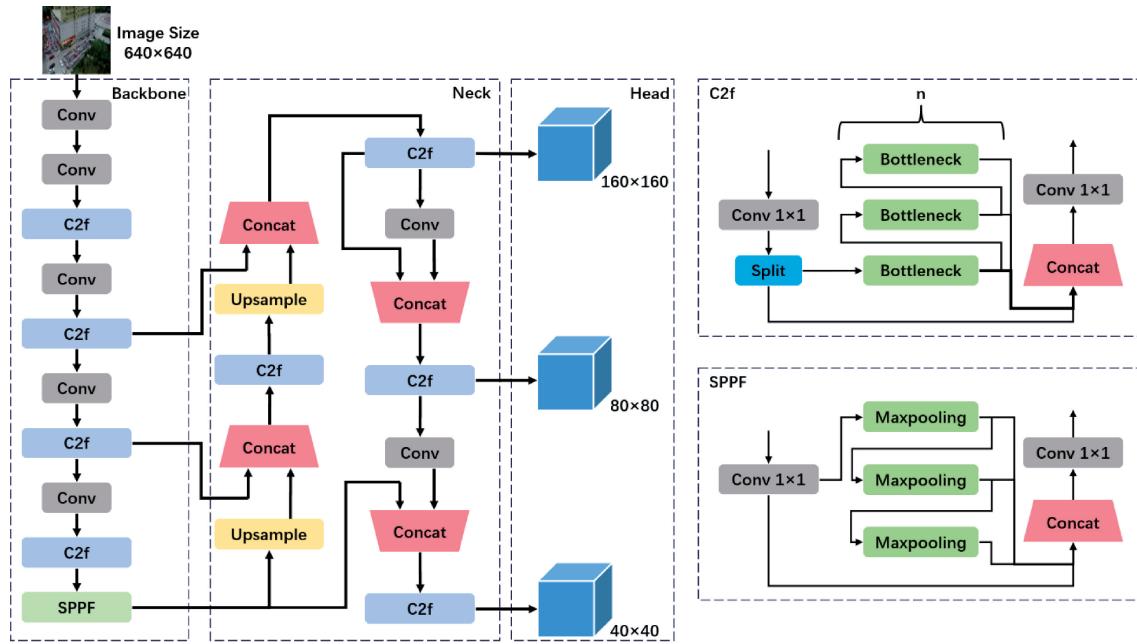


图 1 YOLOv8s 网络结构
Fig. 1 YOLOv8s network structure

空间金字塔池化,用来融合不同尺度的特征,而 C2f 模块融合了 C3 模块的残差结构和 YOLOv7^[19]的泛化高效层聚合网络(efficient layer aggregation network, ELAN)思想,使其能够提供更丰富的梯度流信息。颈部网络使用特征金字塔网络(feature pyramid network, FPN)和路径聚合网络(path aggregation network, PAN)实现了多个尺度特征图的融合,进一步增强了语义表达和定位能力。检测头部部分采用了 YOLOX 的解耦头结构,将分类和检测分离,同时采用了无锚框(anchor-free)的目标检测方法,提高了检测速度和准确度。在损失计算方面,YOLOv8 采用了正负样本动态分配策略,结合了分类损失(varifocal loss, VFL)和回归损失 CIOU+分布焦点损失(distribution focal loss, DFL),以提高模型的性能。

2 本文方法

2.1 主干网络改进

1) C2f_DcnA 模块

在无人机航拍图像目标检测场景中,待检测的目标往往没有规则的几何形状,且由于无人机是在运动中拍摄图像,拍摄视角的改变导致相同的目标也会具有不同的表现形状。标准的卷积操作往往使用固定的矩形卷积核,使得标准卷积在此任务下无法更好的提取到待检测物的特征,从而导致漏检错检。因此,本文使用可变形卷积网络(deformable convolutional networks version 2, DCNv2)^[20]代替标准卷积,增强网络对于复杂目标的特征提取能力。可变形卷积不再使用规则的矩形卷积核,而是在每一个采样位置上增加一个偏移量(offset),以求在不

同的阶段和目标上都使用最优的卷积核结构,增强不规则目标的特征提取。同时在每个采样点添加一个权重,增强网络的形变建模能力。可变形卷积核的示意图如图 2 所示。

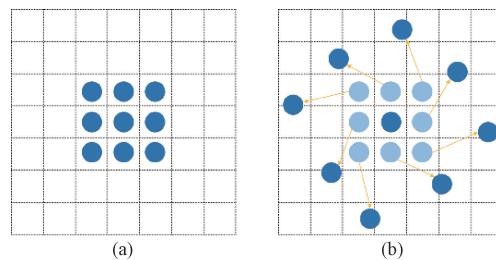


图 2 可变形卷积核
Fig. 2 Deformable convolution kernel

可变形卷积核计算公式如下:

$$y(P_0) = \sum_{P_n \in R} w(P_n) \cdot x(P_0 + P_n + \Delta P_n) \cdot \Delta m_n \quad (1)$$

式中: y 为输出特征图; n 为采样点数, $w(p_n)$ 为第 n 个采样点投影权重, p_n 为预定义卷积网格采样的第 n 个位置;可变形卷积操作在 R 上进行,每个点增加了一个可学习的偏移量 Δp_n 。 Δm_n 是第 n 个位置的调制标量,取值为 $[0, 1]$,若 Δm_n 取值为 0,则表示此部分区域的特征对网络输出无关。

在 DCNv2 中,偏移量和权重是由一个简单的卷积的学习获得,提取各路特征的能力相对较弱。因此,本文设计一个多路坐标注意力机制(multipath coordinate atten-

tion, MPCA)添加在该卷积中,增强对于输入图像特征的学习,生成更契合待检测物体形状的卷积核,增强原网络的特征提取能力,MPCA 的结构如图 3 所示。MPCA 将输入特征分解为两个一维的注意力特征编码:水平方向(x)和垂直方向(y),分别进行平均池化(average pooling, AP)操作,然后连接起来进行通道信息交换,通过一组卷积

积和 sigmoid 操作使特征信息更加精细,重新切分后再融合得到坐标注意力(coordinate attention, CA);在输入特征处添加一个通道分支,对其使用全局平均池化操作(global average pooling, GAP)和卷积操作,并将 sigmoid 操作后的特征求平均后与之融合,调整原通道的权重,最后得到多路坐标注意力机制。

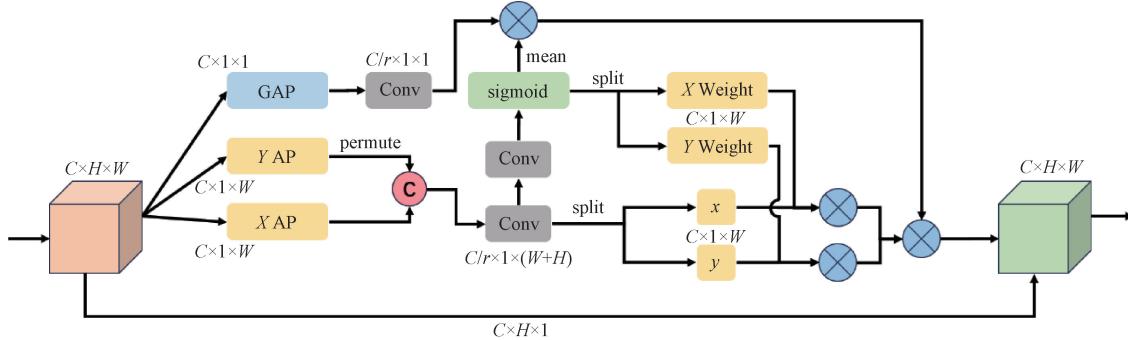


图 3 MPCA 结构

Fig. 3 Multipath coordinate attention structure

2) SPPF_LKSA 模块

无人机目标检测场景中,由于无人机飞行高度变化,待检测目标的尺寸差异较大,图像中的同一物体往往呈现不同的尺寸。原有的 SPPF 模块使用 3 个连续的池化层,将每层的输出结合起来,实现多尺度融合,显著降低了计算复杂度。但 SPPF 模块的多层次池化操作使其容易忽略小目标的特征信息,在无人机目标检测任务上的效果不佳。因此,本文引入 LKSA 注意力机制改进 SPPF 模块,提高网络的多尺度特征提取能力。

LKSA 使用深度卷积(DW-Conv)和空洞深度卷积(DW-D-Conv)来建模大卷积核,使其能够增强空间感知能力并减少大卷积核带来的计算量和参数量。对于一个 $k \times k$ 的大卷积核,LKSA 将其分解为 3 部分:1)一个 $1 \times (2d-1)$ 和一个 $(2d-1) \times 1$ 的深度卷积, d 为膨胀系数的大小;2)一个 $1 \times (k/d)$ 和一个 $(k/d) \times 1$ 的空洞深度卷积;3)一个 1×1 卷积。LKSA 结构如图 4 所示。

特征图输入后,LKSA 使用深度卷积对输入的每个通道都执行单独的卷积操作,每个卷积核处理来自一个输入通道的信息;然后使用空洞卷积在输入通道之间共享参数,从而减少参数量;最后使用一个 1×1 卷积整合信息,形成最终的特征图。LKSA 的输入如下:

$$Z^c = \sum_{H,W} W_{(2d-1) \times 1}^c * (\sum_{H,W} W_{1 \times (2d-1)}^c * F^c) \quad (2)$$

$$Z^c = \sum_{H,W} W_{\lfloor \frac{k}{d} \rfloor \times 1}^c * (\sum_{H,W} W_{1 \times \lfloor \frac{k}{d} \rfloor}^c * \bar{Z}^c) \quad (3)$$

$$A^c = W_{1 \times 1} * Z^c \quad (4)$$

$$\bar{F}^c = A^c \otimes F^c \quad (5)$$

其中,* 和 \otimes 分别代表卷积和 Hadamard 积。

2.2 多尺度特征融合

由于无人机飞行高度较高,拍摄图像中的物体多数为

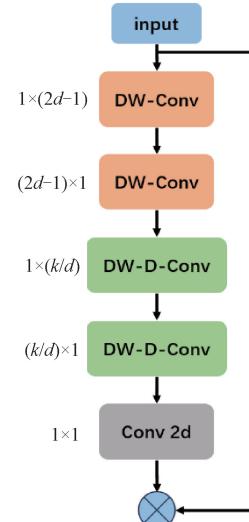


图 4 LSKA 结构

Fig. 4 Large separable kernel attention structure

小目标,其特征信息主要存在于浅层网络,随着网络层数的加深,小目标的特征信息会丢失,导致小目标的检测精度下降。在 YOLOv8 网络中,为了提升对小目标的检测能力,添加了新的模型输出层 P2,P2 层的卷积次数少,特征图尺寸较大,使其能够保留更多的小目标特征。为了进一步提升网络的检测性能,本文结合 Bi-FPN 的思想,优化原网络的特征融合模块,增加不同尺度之间的连接通道,使得特征信息能够跨层级流动,增强网络对不同尺度目标的检测能力。

3 种特征融合结构如图 5 所示,其中图 5(a)为 YOLOv8 中的特征融合方式,虽然添加了小目标检测层 P2,但并没有直接连接到输出层,而是输入到 P3 层中再

输出,在特征提取过程中,小目标的特征容易受到较大目标的干扰,网络层数加深也意味着小目标特征信息的缺失,这使得原网络在小目标检测上不能满足实际要求。Bi-FPN 在同一层的输入节点和输出节点之间跳跃连接,如图 5(b)所示,在尺度相同的情况下,跳跃连接能够更好

地提取和传递特征信息。因此,本文在浅层网络 P2 和 P3 中添加跳跃连接,使其保留更多的浅层特征,同时针对无人机平台有限的计算资源,优化网络结构,减少网络参数量和计算量,改进后的特征融合结构 YOLOv8-Bi 如图 5(c)所示。

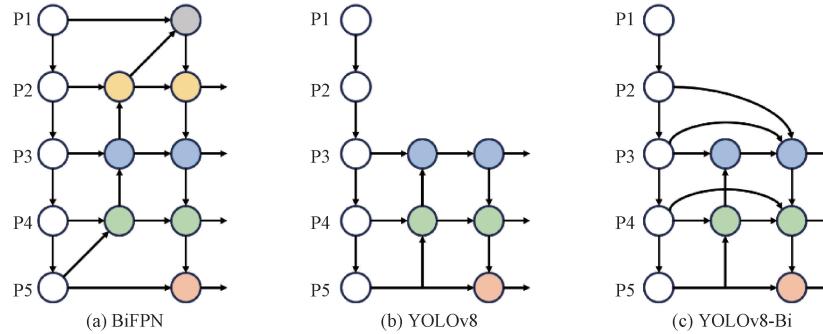


图 5 3 种特征融合模块结构

Fig. 5 Three feature fusion module structures

2.3 自注意力机制动态检测头

YOLOv8 网络使用了 3 个检测头,采用了 Anchor-Free 和 Decoupled-Head 的思想,在多数目标检测任务中表现出色,但在检测遮挡物体和小目标时容易漏检或误检,导致在无人机目标检测任务上效果不佳。因此,本文使用一个更加高效的动态检测头 DyHead 提高网络检测性能。

YOLOv8 主干网络输出的是一个三维张量,level × spatial × channels, 分别为特征层数,空间位置和输出通道,在这 3 个维度上分别使用尺度感知注意力 L ,空间感知注意力 S 和任务感知注意力 C ,3 种注意力结合组成了动态检测头 DyHead,结构如图 6 所示。

1) 尺度感知注意力根据语义重要性融合不同尺度的特征,其表达式如下:

$$\pi_L(F) \cdot F = \sigma\left(f\left(\frac{1}{SC} \sum_{s,c} F\right)\right) \cdot F$$

$$\sigma(x) = \max(0, \min(1, \frac{x+1}{2})) \quad (6)$$

式中: $f(\cdot)$ 是一个 1×1 卷积的线性变换; σ 为 Hard sigmoid 函数。

2) 空间感知注意力首先使用可变形卷积在注意力学习中添加稀疏性,然后在同一空间位置聚合不同层次的特征,其表达式如下:

$$\pi_S(F) \cdot F = \frac{1}{L} \sum_{l=1}^L \sum_{k=1}^K w_{l,k} \cdot F(l; p_k + \Delta p_k; c) \cdot \Delta m_k \quad (7)$$

式中: K 为稀疏采样位置的个数; Δp_k 和 Δm_k 为可变形卷积中的偏移量和偏移权重,两者都由中间层 f 的输入特征学习得到。

3) 任务感知注意力通过动态地打开或关闭特征通道,以支持不同的任务。

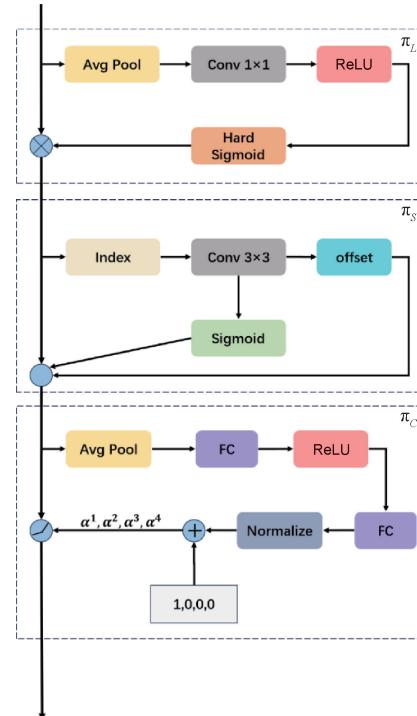


图 6 DyHead 结构

Fig. 6 Dynamic head structures

$$\pi_C(F) \cdot F = \max(\alpha^1(F) \cdot F_c + \beta^1(F), \alpha^2(F) \cdot F_c + \beta^2(F)) \quad (8)$$

式中: F_c 为第 c 个通道的特征切片, $[\alpha^1, \alpha^2, \beta^1, \beta^2]^T = \theta$ 为学习控制激活阈值的超函数。 θ 首先在 $L \times S$ 维上进行全局平均池化以降低维数,然后使用两个完全连接层和一个归一化层,最后使用移位的 sigmoid 函数将输出归一化为 $[-1, 1]$ 。

上述 3 种注意机制是顺序应用的,通过多次嵌套

式(7),可以有效地将多个 π_L 、 π_S 和 π_C 块叠加在一起,组成动态检测头DYHead。

2.4 损失函数改进

在目标检测算法中,通常使用边界框(bounding box)实现精确的目标定位,采用IoU损失函数来衡量网络预测结果与实际结果的相似度,进一步优化边界框的定位。YOLOv8s算法的损失函数由分类损失VFL Loss和回归损失CIOU Loss + DFL两部分组成。相比于IOU函数,CIOU损失函数将预测边界框和真实边界框的重叠面积,中心点距离,以及边界框的纵横比纳入损失计算中,提高了边界框的检测能力和收敛速度。但当预测边界框和真实边界框的纵横比相同时,CIOU会忽略该边界框的优化,使模型收敛过程产生波动。同时,训练数据中不可避免会存在低质量的样本,尤其是在充斥着密集小目标的无人机拍摄图像中,使用CIOU函数计算低质量样本的中心距离和纵横比等几何因素会加重模型对低质量样本的惩罚,从而导致模型的泛化能力下降,影响检测性能^[21]。因此,本文引入Wise-IOU损失函数,提升边界框预测回归效果。

Wise-IOU采用动态非单调的聚焦机制,该机制使用离群度代替IOU对锚框进行质量评估提供了明智的梯度

增益分配策略。该策略在降低高质量锚框的竞争的同时,也减小了低质量示例产生的有害梯度。这使得Wise-IOU可以聚焦于普通质量的锚框,并提高检测器的整体性能。公式定义如下:

$$L_{WIoU_1} = R_{WIoU} L_{IoU}$$

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \quad (9)$$

式中:(x, y)和(x_{gt}, y_{gt})为锚框和目标框的中心点坐标, L_{IoU} 属于[0, 1],能够改进质量较低的锚框,通过在距离度量中引入 R_{WIoU} 属于[1, e],可以减少对高质量锚框和中心距离的关注, $(W_g^2 + H_g^2)^*$ 将 W_g 和 H_g 分开处理,可以防止 R_{WIoU} 产生阻碍收敛的梯度,有效消除了阻碍收敛的因素,因此该方法没有将纵横比纳入损失计算中。通过引入一个单调聚焦系数 L_{IoU}^* ,能够使得模型聚焦于困难样本,同时引入均值 L_{IoU} 作为归一化因子,得到了Wise-IOUv2,公式如下:

$$L_{WIoU_2} = \left(\frac{L_{IoU}^*}{L_{IoU}}\right)^r L_{WIoU_1} \quad (10)$$

2.5 改进的YOLOv8s网络模型

通过以上对原始YOLOv8s模型的改进,得到最终的无人机目标检测网络模型,网络结构如图7所示。

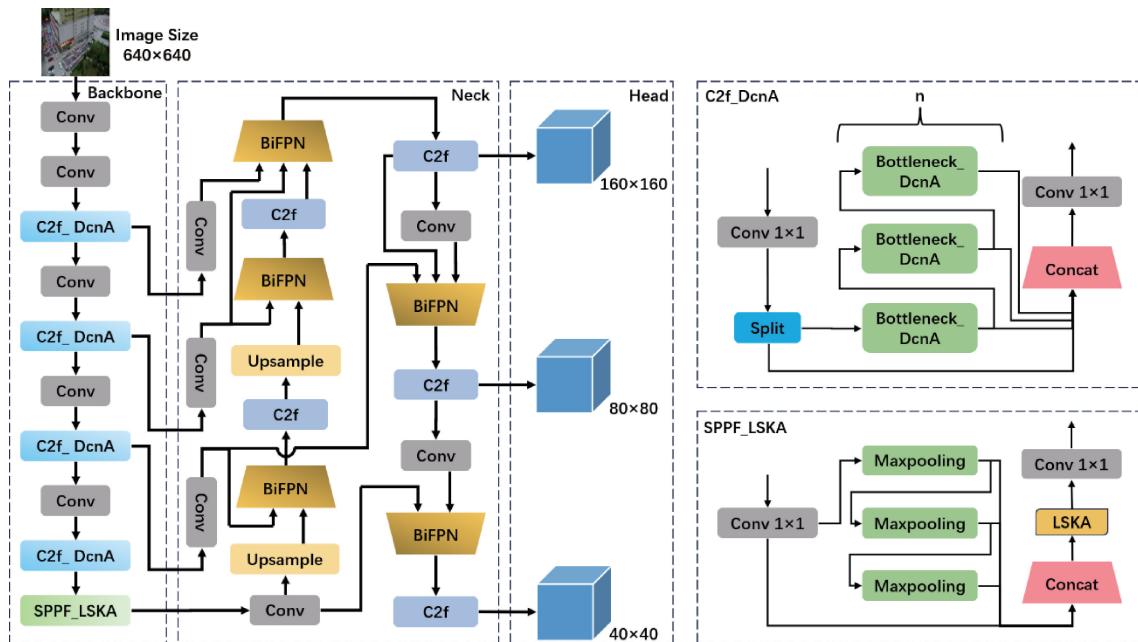


图7 改进的YOLOv8s网络结构
Fig. 7 Improved YOLOv8s network structure

3 实验结果与分析

3.1 实验数据集

本文实验数据集采用无人机目标检测数据集VisDrone2019,该数据集由6 471张训练图片,548张验证图片和1 610张测试图片组成,包括了pedestrian、people、bi-

cycle、car、van、truck、tricycle、awning-tricycle、bus、motor共10个类别,几乎包含了城市场景下无人机航拍图像中的全部常见目标种类。实验数据集的部分图片如图8所示,从图8(a)可以看出图像中待检测目标容易受到楼房,树木等遮挡;图8(b)为由于无人机飞行倾角的变化也会改变图像拍摄的角度,这些因素都给准确检测目标带来了

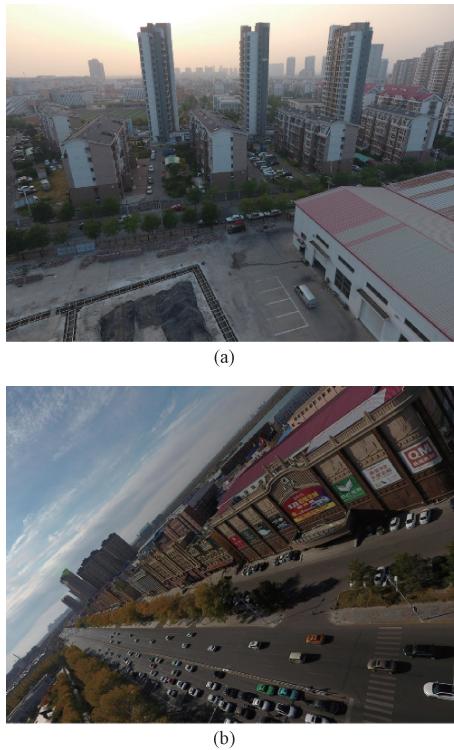


图 8 VisDrone2019 数据集

Fig. 8 VisDrone2019 dataset

挑战。

3.2 实验环境

本文实验硬件环境为 Nvidia RTX3080ti GPU, Xeon (R) Silver 4214R CPU; 软件环境为 Python 3.8.10, Cuda 11.8, pytorch 2.0.0。

训练时输入图像大小为 640, 初始学习率为 0.01, 使用 SGD 优化器, 动量值为 0.937, 在 RTX3080ti GPU 上训练 200 轮, 最后 10 轮关闭 Mosaic 数据增强。

3.3 评价指标

本文使用准确率(precision, P)、召回率(recall, R)、平均精确度均值(mean average precision, mAP)衡量目标检测算法的准确性。

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

$$R = \frac{TP}{TP + FN}$$
(11)

式中: TP(true positive) 表示实际为正类且预测也为正类; TN(true negative) 表示实际为负类且预测也为负类; FP(false positive) 表示实际为负类但预测为正类; FN(false negative) 表示实际为正类但预测为负类; P 表示正确预测的正类个数占总预测正类个数的比例; R 表示正确预测的正类个数占总预测个数的比例。

本文使用 mAP@0.5 和 mAP@0.5 : 0.95 衡量算法的准确度。mAP@0.5 表示 IOU 阈值为 0.5 时目标类别

的评价检测精度, 能够反映网络对于不同类别目标的综合分类能力; mAP@0.5 : 0.95 表示以步长为 0.05, IOU 阈值从 0.5 ~ 0.95 的全部 10 个阈值下的平均检测精度。(mean average, AP) 平均精度为:

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

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i$$
(12)

式中: n 表示数据集的类别数。

除了衡量算法的准确性以外, 本文使用参数量(parameters)和计算量(FLOPs)来衡量网络的计算复杂程度, 衡量算法的轻量化程度。

3.4 损失函数对比实验

为了验证本文所改进的损失函数的有效性和不同版本的 Wise-IoU 损失函数的性能差异, 以 YOLOv8s 为基准模型进行实验, 使用 CIoU、DIoU^[22]、MPDIoU^[23] 和两个版本的 Wise-IoU 损失函数进行对比实验, 实验结果如表 1 所示。从表 1 可以得出, 本文使用的 Wise-IoU 损失函数比其他损失函数的精度更高, 实验结果表明, Wise-IoUv2 版本比 v1 版本的平均精确度高出 0.6%, 更适合无人机小目标检测任务。

表 1 损失函数对比实验

Table 1 IoU loss comparison experiment

损失函数	P/%	R/%	mAP@0.5/%
CIoU	48.4	38.7	38.7
DIoU	48.7	38.2	38.7
MPDIoU	50.9	38.0	38.8
Wise-IoUv1	49.5	39.1	39.0
Wise-IoUv2	51.1	38.5	39.6

3.5 消融实验

为了验证本文所提出的无人机小目标检测算法各个改进模块的有效性, 设计进行消融实验。实验 A 使用 YOLOv8s 算法作为基准模型; 实验 B 在实验 A 基础上将原有的 C2f 模块改进为 C2f_DcnA; 实验 C 将实验 B 算法模型中的 SPPF 模块替换为本文的 SPPF_LKSA; 实验 D 在实验 C 的基础上使用 YOLOv8_Bi 改进其特征融合模块; 实验 E 在实验 D 的基础上将损失函数替换为 Wise-IoUv2; 最后, 实验 F 将实验 E 算法模型的检测头替换为 DyHead。实验结果如表 2 所示。

本文将原始 YOLOv8s 算法作为基准模型, 在此基础上逐个添加改进模块进行消融实验。从实验结果可以看出, 在网络主干部分, C2f_DcnA 和 SPPF_LKSA 改进模块增强了网络的特征提取能力, mAP@0.5 和 mAP@0.5 : 0.95 均提升了 0.4%, 但由于添加了注意力机制, 网络计算复杂度略有上升。在此基础上, 添加多尺度特征融合模块 YOLOv8_Bi 后, 增强了算法对不同尺度目标的特征

表2 消融实验
Table 2 Ablation experiment

实验	模型	mAP@0.5/%	mAP@0.5 : 0.95/%	参数量/($\times 10^6$)	计算量/GFLOPs
A	YOLOv8s	38.7	23.0	11.12	28.5
B	A+C2f_DcnA	39.0	23.3	11.31	26.7
C	B+SPPF_LKSA	39.1	23.4	12.38	27.6
D	C+YOLOv8_Bi	40.2	24.3	8.63	24.3
E	D+Wise-IoUv2	40.6	24.3	8.63	24.3
F	E+DyHead	41.7	24.9	9.17	24.9

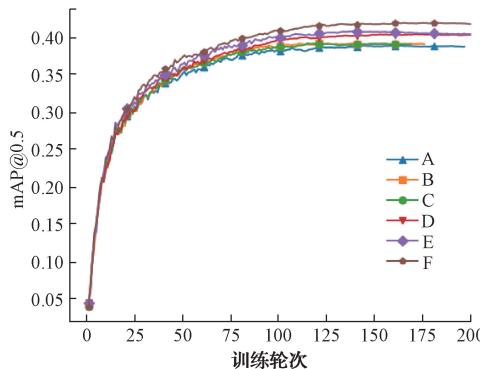


图9 消融实验的 mAP@0.5 曲线
Fig. 9 mAP@0.5 curve of ablation experiment

提取,改善了模型对小目标的检测能力,同时由于弱化了大目标检测层,使模型的参数量和计算量大幅降低,实现了模型的轻量化。实验 E 使用 Wise-IoUv2 损失函数改善数据集中低质量样本过多影响训练效果的问题,加快模型收敛速度,进一步提升检测精度。最后添加了自注意力机制动态检测头 DyHead,虽然参数量和计算量小幅上升,但增强了模型对小目标,遮挡目标的检测能力,mAP@0.5 和 mAP@0.5 : 0.95 相较之前分别提升了 1.1% 和 0.6%。消融实验中各模型在训练中的 mAP@0.5 曲线变化如图 9 所示,可知,模型在 150 轮后趋于收敛,改进后模型的检测精度明显低于改进前的检测精度。

综上所述,本文改进后的模型相比于基准模型 YOLOv8s 在 mAP@0.5 和 mAP@0.5 : 0.95 上分别提升了 3.0% 和 1.9%,参数量和计算量分别下降了 17.5% 和 12.63%,由此可以看出本文改进的模型检测精度显著提升,模型计算复杂程度显著降低,实现模型的轻量化。

3.6 对比实验

为了验证本文提出的无人机小目标检测算法的性能优势,选取目前热门的目标检测算法在 VisDrone2019 数据集上进行对比实验,其中包括两阶段目标检测算法如 Faster-RCNN、FPN^[24],以及单阶段目标检测算法如 SSD、RetinaNet 和 YOLO 系列算法,实验结果如表 3 所示。

从实验结果可以得出,单阶段的 YOLO 系列目标检测算法在检测精度上显著高于 SSD、RetinaNet 和两阶段

表3 对比实验

Table 3 Comparison experiment

模型	mAP@0.5/%	参数量/ $(\times 10^6)$	计算量/GFLOPs
Faster-RCNN	33.3	41.2	206.7
FPN	29.2	11.3	26.7
SSD	23.8	24.5	87.9
RetinaNet	21.2	19.8	93.7
YOLOv5s	33.3	7.2	16.5
YOLOX-s	33.9	9.0	26.8
YOLOv7	41.8	36.9	104.7
YOLOv8s	38.7	11.1	28.5
本文	41.7	9.2	24.9

目标检测算法 FPN, Faster-RCNN 在检测精度上虽然不输 YOLO 系列,但参数量和计算量较大,因此不适合在无人机平台使用。在 YOLO 系列算法中,本文方法相比于 YOLOv7 算法,平均精确度只下降了 0.1%,但参数量下降了 75%,计算量下降了 76%,说明本文方法相比 YOLOv7 更适合在无人机平台上部署。相比于计算复杂度较低的 YOLOv5s 和 YOLOX-s 算法,本文方法的检测精度分别提升了 8.7% 和 7.9%,说明本文方法在无人机小目标检测任务的检测精度更好,同时具有相对较小的参数量和计算量。

4 结 论

本文针对当前无人机目标检测算法存在小目标检测精度不足,实时性差等问题,对 YOLOv8s 算法为基础进行改进。通过使用可变形卷积替换原来的标准卷积增强网络对形状不规则目标的特征提取能力,使网络提取到更加丰富的目标特征信息。无人机拍摄图像中目标尺度差异较大,本文使用可分离大核注意力机制 LSKA 提高网络的多尺度特征提取能力,使网络能够获得不同层次的尺度信息。然后针对无人机拍摄图像中小目标数量较多,将 Bi-FPN 和 YOLOv8s 的特征融合模块结合起来,添加小目标特征融合路线,增强网络对不同尺度特征层的信息进行融合的能力,提升小目标的检测精度。在检测头部分,

使用基于自注意力机制的动态检测头 DyHead 改善网络对小目标的漏检、错检问题,进一步提升检测精度。针对数据集中的低质量样本会对训练效果产生负面影响的问题,使用 Wise-IoU 损失函数替换原来的损失函数,提升训练速度和精度。实验结果表明,在检测精度方面,本文方法相比原始 YOLOv8s 算法,mAP@0.5 提升了 3.0%,mAP@0.5 : 0.95 提升了 1.9%;在计算复杂度方面,参数量下降了 17.5%,计算量下降了 12.63%。相比于其他热门目标检测算法均有一定优势。

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):028822.
- [2] LUO X D, WU Y Q, CHEN J L. Research progress on deep learning methods for object detection and semantic segmentation in UAV aerial images [J]. Acta Aeronautica et Astronautica Sinica,2024,45(6): 028822.
- [3] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [4] 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): 1137-1149.
- [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, 2016: 779-788.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. 2016 European Conference on Computer Vision. Springer, 2016: 21-37.
- [7] 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, 2017: 2980-2988.
- [8] 杨辉羽,李海明.单聚合 YOLO 航拍小目标检测算法[J].国外电子测量技术,2023,42(4):131-140.
- [9] PEI W, XU Y M, ZHU Y Y, et al. The target detection method of aerial photography images with improved SSD[J]. Journal of Software, 2019,30(3): 738-758.
- [10] YAN J H, ZHANG K, SHI T J, et al. Multi-level feature fusion based dim small ground target detection in remote sensing images [J]. Chinese Journal of Scientific Instrument, 2022,43(3):221-229.
- [11] YANG Z, XU Z, WANG Y. Bidirection-fusion-YOLOv3: An improved method for insulator defect detection using UAV image[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-8.
- [12] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. ArXiv preprint arXiv: 1804. 02767, 2018.
- [13] 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.
- [14] 钱承山,沈有为,孙宁,等.基于 Transformer 改进 YOLOv5 的山火检测方法研究[J].电子测量技术,2023,46(16):46-56.
- [15] QIAN CH SH, SHEN Y W, SUN N, et al. Research on improved YOLOv5 forest fire detection method based on Transformer [J]. Electronic Measure Technology, 2023,46(16):46-56.
- [16] 陈朋磊,王江涛,张志伟,等.基于特征聚合与多元协同特征交互的航拍图像小目标检测[J].电子测量与仪器学报,2023,37(10):183-192.
- [17] CHEN P L, WANG J T, ZHANG ZH W, et al. Small object detection in aerial images based on feature aggregation and multiple cooperative features interaction [J]. Journal of Electronic Measurement and Instrumentation, 2023,37(10):183-192.
- [18] HOU H, CHEN M, TIE Y, et al. A universal landslide detection method in optical remote sensing images based on improved YOLOX [J]. Remote Sensing, 2022, 14(19): 4939.
- [19] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021 [J]. ArXiv preprint arXiv: 2107. 08430, 2021.
- [20] Foreign Electronic Measurement Technology, 2023, 42(4):131-140.

- [18] 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.
- [19] 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]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [20] ZHU X, HU H, LIN S, et al. Deformable convnets v2: More deformable, better results[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9308-9316.
- [21] TONG Z, CHEN Y, XU Z, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism[J]. ArXiv preprint arXiv: 2301. 10051, 2023.
- [22] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
- [23] SILIANG M, YONG X. MpIoU: A loss for efficient and accurate bounding box regression[J]. ArXiv preprint arXiv: 2307. 07662, 2023.
- [24] 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, 2017: 2117-2125.

作者简介

曲晨阳(通信作者),硕士研究生,主要研究方向为深度学习、目标检测。

E-mail: q18411012150423@163.com

吕进,博士,副教授,主要研究方向为交通信息工程及控制、人工智能及智能优化与控制、人工神经网络理论及应用、智能交通系统。

E-mail: lvjin@chd.edu.cn

卫策,硕士研究生,主要研究方向为深度学习。

E-mail: weice_5003_0@163.com