

UAVAI-YOLO: 无人机航拍图像的小目标检测模型

何植仟¹, 曹立杰^{1,2}

(1. 大连海洋大学信息工程学院, 辽宁 大连 116023;
2. 辽宁省海洋信息技术重点实验室, 辽宁 大连 116023)

摘要: 针对无人机航拍图像目标检测效果差的问题, 提出改进的UAVAI-YOLO模型。首先, 为使模型获得更加丰富的语义信息, 使用改进可变形卷积网络(deformable convolutional networks, DCN)替换原骨干(backbone)网络部分通道到像素(channel-to-pixel, C2f)模块原始卷积。其次, 为增加P2特征层而不增加模型参数量, 提出Conv_C模块将骨干网络输出通道降维, 同时避免通道降维导致的语义信息丢失, 使用改进ODConv卷积替换颈部(neck)部分C2f模块原始卷积。然后, 为充分利用上下文语义信息, 引入双向特征金字塔网络(bi-directional feature pyramid network, BIFPN)。最后使用Wise-IoU替换原始损失函数, 提高模型目标检测框的准确性。在公开的VisDrone2019数据集和UAVDT数据集的实验结果表明, UAVAI-YOLO模型相比于原YOLOv8n模型mAP@0.5分别提升了4.4%和1.1%。与其他主流目标检测模型相比具有较高的检测精度。

关键词: 无人机航拍图像; 小目标检测; YOLOv8; 可变形卷积网络; 注意力机制

中图分类号: TP391

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202422

UAVAI-YOLO: dense small target detection algorithm based on UAV aerial images

HE Zhiqian¹, CAO Lijie^{1,2}

1. School of Information Engineering, Dalian Ocean University, Dalian 116023, China
2. Liaoning Provincial Key Laboratory of Marine Information Technology, Dalian 116023, China

Abstract: An improved UAVAI-YOLO model was proposed to address the problem of poor target detection in UAV aerial images. Firstly, in order to obtain richer semantic information for the model, the original convolution of the C2f module of the original backbone part was replaced with the improved DCN convolution. Secondly, in order to increase the P2 feature layer without increasing the number of model parameters, the Conv_C module was proposed to downscale the output channel of the backbone network, and at the same time, because of avoiding the loss of semantic information due to channel downsizing, the original convolution of the C2f module in the neck part was replaced by the improved ODConv dynamic convolution. Then, the BIFPN module was introduced to make full use of the contextual semantic information. Finally, Wise-IoU was used to replace the original loss function to improve the accuracy of the model target detection frame. Experimental results on the publicly available VisDrone2019 dataset and UAVDT dataset showed that the UAVAI-YOLO model improves 4.4% and 1.1% compared to the original YOLOv8n model mAP0.5, respectively, high detectability accuracy compared to other mainstream object detection models.

Key words: UAV aerial images, small object detection, YOLOv8, DCN, attention mechanisms

收稿日期: 2024-04-10; 修回日期: 2024-05-21

通信作者: 曹立杰, caolijie@dlou.edu.cn

基金项目: 辽宁省教育厅科研项目(No.LJKZ0731)

Foundation Item: Liaoning Province Department of Education Research Fund (No.LJKZ0731)

0 引言

随着无人机技术的迅速发展,其在军工业、商业、农业等多个领域得到了广泛应用^[1-3]。但是,无人机飞行高度较高,拍摄对象比较复杂,导致航拍图像中的目标较小且密集,这对现有算法的目标检测精度提出了挑战。此外,无人机所采用的目标检测算法还受到移动端硬件能力的限制。因此,提升无人机航拍图像的目标检测精度已经成为目标检测领域中一个极具挑战性的研究课题。

自 LeNet-5 问世以来,它作为卷积神经网络(convolutional neural networks, CNN)的奠基石,推动了CNN技术的发展。2012年, AlexNet的诞生进一步加速了基于CNN的目标检测模型的发展。目前,目标检测模型主要分为两阶段和单阶段两种类型。两阶段算法,如区域卷积神经网络(region-CNN, R-CNN)、Fast R-CNN、Mask R-CNN等,首先生成候选区域,然后对这些感兴趣区域进行精确定位和分类。这种方法虽然检测精度较高,但检测速度较慢,不适用于需要实时检测的场景。单阶段算法,如YOLO和SSD,直接在单次检测中预测物体的类别和位置,无须候选区域步骤,因此速度更快,能够满足实时检测的需求。随着YOLO系列的不断改进,YOLOv5、YOLOv7、YOLOv8等模型在检测大目标方面展现了更高的精度。例如, Qu等人^[4]提出的基于YOLOv5的多尺度特征融合网络,在PASCAL VOC和MS COCO数据集上mAP@0.5分别达到89.1%和67.8%。黄路等人^[5]基于改进的YOLOv5实现了基于区域的多目标检测,在PASCAL VOC数据上mAP@0.5达到81.7%。

无人机航拍图像具有检测目标数量大、背景复杂等特点,同时,受无人机硬件限制,部署在无人机的目标检测模型参量需要足够小。大量学者采用YOLO作为无人机目标检测模型,满足无人机目标检测对实时性、准确性、鲁棒性的要求。但是,YOLO目标检测模型在网络中使用较小的特征图进行边界框的预测,这可能导致小目标的细节信息丢失。因此,有学者对YOLO系列模型进行改进以提高无人机图像目标检测准确率。如张朝阳等人^[6]将YOLOv5融合Swin Transformer多头注意力机制并使用完整交并比(complete intersection over union, CIoU)作为边界框回归的损失函数,在遥感目标数据集NWPU VHR-10上mAP@0.5比基线模型提

高了3.05%;刘晋川等人^[7]使用一种多阶段特征融合方法,并将其与注意力机制串联到特征挖掘模块,在小目标检测数据集上的检测精度比原有模型提高了8.5%;Ju等人^[8]使用改进的Darknet53作为骨干网络,利用尺度匹配策略为小目标检测选择合适的尺度和锚点大小,在航空影像中的车辆检测(vehicle detection in aerial imagery, VEDAI)数据集上实现较高的检测精度;胡均平等^[9]基于YOLOv5在颈部特征融合中增设深、浅卷积特征的可学习自适应权重,在TT100K(Tsinghua-Tencent 100K)交通数据集上mAP@0.5达到77.3%。Shao等人^[10]基于YOLOv8提出了一种轻量级的无人机图像目标检测模型Aero_YOLO,在VisDrone2019数据集上mAP@0.5达到28.6%。Cai等人^[11]提出了一种新的损失函数CFIoU,解决小物体检测中边界框收敛速度缓慢和空间信息利用率不高的问题。在VisDrone2019数据集上,应用CFIoU之后,基于YOLOv8s模型的mAP@0.5指标达到33.6%。顾宙瑜等人^[12]设计了一种高效瓶颈模块,提出Rapid-YOLOX模型,在PASCAL VOC 2007数据集mAP@0.5达到77.92%、每秒处理帧数达到45.40。宋华杰等人^[13]对YOLOv3算法的损失函数进行了改进,在不影响车辆检测速度的同时,召回率、精确率分别提升了0.1、0.105。但是,上述学者采用的改进方法没有考虑模型不同层级之间高低特征语义信息缺失的问题,以及无人机硬件对模型参数量的要求。

因此,本文在YOLOv8架构上进行了深入优化,提出了适用于无人机航拍图像的目标检测模型UAVAI-YOLO。首先,针对无人机航拍图像中目标小且特征提取困难的问题,本文在骨干网络中采用改进的C2f_DCN模块,该模块利用改进的可变形卷积网络(deformable convolutional networks, DCN)扩大特征提取时的感受野,从而提高骨干网络的特征提取能力。其次,为了有效利用P2特征层中的小目标语义信息,本文将颈部(neck)部分进行重新构建,将P2、P3、P4、P5特征层输入neck部分。同时,为了抑制引入P2特征层带来的参数量增加问题,在骨干网络与颈部特征链接处插入了Conv_C模块,并在neck部分采用了改进的C2f_ODConv模块,这样在降低模型参数量的同时能够提取更丰富的语义信息。然后,为了更好地结合P2、P3、P4、P5特征层的上下文语义信息,本

文在特征融合采用双向特征金字塔网络 (bidirectional feature pyramid network, BIFPN), 更好地帮助网络模型捕捉上下文依赖关系。最后, 为了提高模型对目标检测框的判断准确性, 本文用具有动态聚焦机制的边界框回归损失 (Wise-IoU) 函数替代了原始损失函数, 避免了因交并比 (intersection over union, IoU) 判别造成的网络模型检测精度较低的问题。在公开的 VisDrone2019 数据集和无人机基准目标检测和跟踪 (UAVDT) 数据集上的实验结果表明, 本文对无人机航拍图像中的目标检测具有优良的检测效果。

1 YOLOv8 基线网络模型

本文采用的基线网络模型为 YOLOv8, 其网络架构如图 1 所示。YOLOv8 是 YOLO 系列中具有更快轻量的模型, 且相比于 YOLOv4、YOLOv5 增加了 focus 模块和跨阶段局部网络 (CSPNet), 在 YOLOv8 网络模型主要分为 backbone、neck、prediction 3 部分。

对输入图像进行数据处理, 将其缩减到 640×640 的尺寸。backbone 由 CBS 模块、通道到像素 (channel-to-pixel, C2f) 模块与空间金字塔池化 (SSPF) 模块组成, 主要作用是对输入的图像进行特征提取。neck 主要采用了路径聚合网络 (path aggregation network, PANet) 将不同特征层和检测

层进行融合。predictions 采用 CIoU_loss 作为 Bounding_box 的损失函数, 使用非极大值抑制 (non-maximum suppression, NMS) 对识别目标进行筛选。

2 改进的 UVAI-YOLO 模型

2.1 UVAI-YOLO 总体架构

改进的 UVAI-YOLO 网络模型如图 2 所示。首先将 YOLOv8 中 backbone 的 C2f 模块替换为改进的 C2f-DCN 模块, 使用可变形卷积在网络和感受野中引入偏移量, 在不同感受野下对特征图进行提取信息, 在 neck 部分使用全维动态卷积 (omni-dimensional dynamic convolution, ODConv) 构成 C2f_ODConv 模块, 同时考虑了空域、输入通道、输出通道等维度上的动态性, 极大提高了模型检测精度。其次对 neck 部分进行重构, 新增一个 P2 特征层, 使其增加对小目标的检测能力。然后对原始 YOLOv8 模型的 IoU 损失函数改进, 引入 Wise-IoU。最后特征融合时为使模型更好地结合上下文语义信息, 引入 BIFPN 模块替换原网络中特征融合网络。

2.2 改进的 C2f_DCN 模块

DCN^[14]通过引入可学习的偏移量来允许卷积核在执行卷积操作时进行形变, 这样可以捕捉到图像中更加灵活和多变的特征, 解决了传统 CNN 模

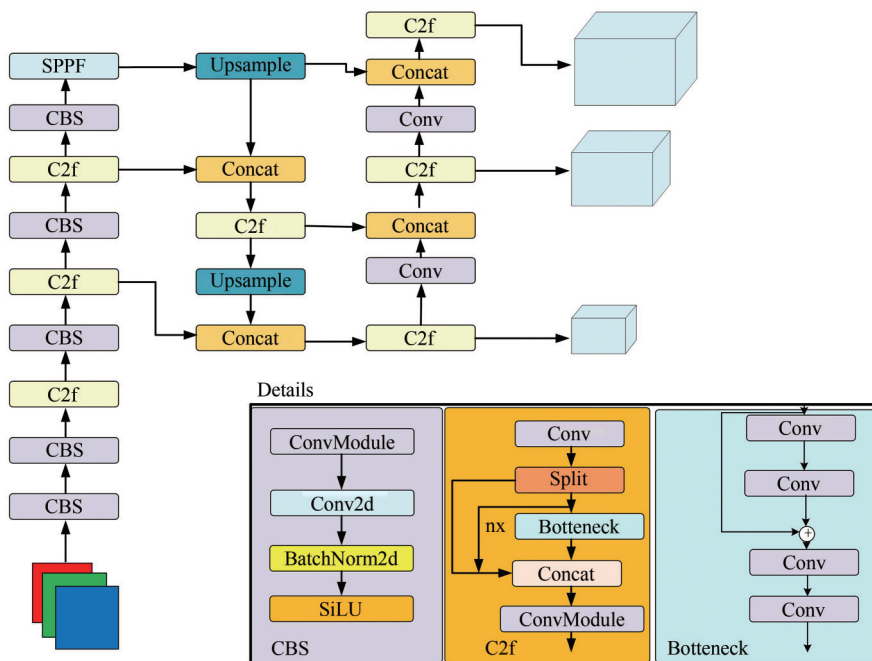


图1 YOLOv8n网络模型

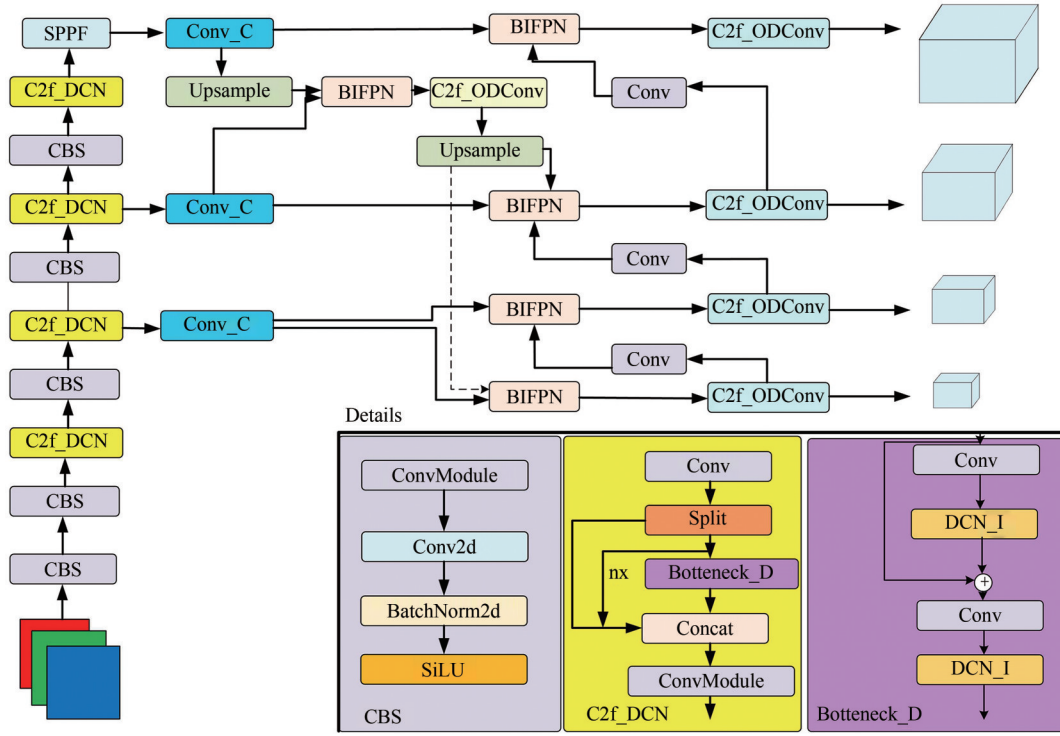


图2 改进的UAAVI-YOLO网络模型

型中空间提取特征不足的问题。通常2D卷积由两个步骤组成：1) 在输入特征图 X 上使用规则网格 R 进行采样，如式(1)所示；2) 对由 w 加权的采样值求和，如式(2)所示。网格 R 定义了感受野的大小和扩张。

$$R = \{(-1, -1), (-1, 0), \dots, (0, -1), (1, 1)\} \quad (1)$$

$$y(P_0) = \sum_{P_n \in R} w(p_n) \times x(P_0 + P_n) \quad (2)$$

其中 w 为权重矩阵， P_n 为卷积核， x 为输入特征图， P_0 为卷积中心点坐标。

相比传统的CNN卷积，DCN卷积引入了由双线性插值得到的 ΔP_n 偏移量，如式(3)所示，以此提升模型的感受野，解决CNN模型面对小目标检测时特征提取困难不足的问题。

$$y(P_0) = \sum_{P_n \in R} w(p_n) \times x(P_0 + P_n + \Delta P_n) \quad (3)$$

但是DCN卷积引入了额外的可学习偏移量，相比于传统的卷积操作需要更多的计算资源。同时当DCN卷积在学习位置偏移量 ΔP_n 时，无法确定其感受野区域的有效性，可能会导致优化过程中梯度消失或爆炸的问题。因此本文对DCN卷积采样点范围进行改进，利用多层网格 R 叠加的方法，确定位置偏移量采样点的感受野，如式(4)所示。

$$R_{\text{Original}} = \{(-1, -1), (-1, 0), \dots, (0, -1), (1, 1)\}$$

$$R_{\text{TL}} = \{(-2, -2), (-2, 0), \dots, (0, 1), (-1, -1)\}$$

$$R_{\text{BL}} = \{(-2, 0), (-1, -1), \dots, (-1, -1), (-2, -2)\}$$

$$R_{\text{TR}} = \{(0, 2), (0, 1), \dots, (2, 1), (2, 0)\}$$

$$R_{\text{BR}} = \{(0, 0), (-1, -1), \dots, (2, 1), (2, 2)\}$$

$$R_{\text{Improved}} = R_{\text{Original}} + R_{\text{TL}} + R_{\text{BL}} + R_{\text{TR}} + R_{\text{BR}} \quad (4)$$

本文提出的改进的DCN卷积在不增加模型参数数量的情况下，实现了对DCN卷积位置偏移量的采样点感受野多层次采样。卷积采样点感受野如图3所示。

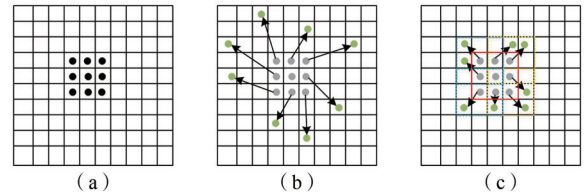


图3 卷积采样点感受野 (a) 传统卷积采样点; (b) DCN卷积采样点; (c) 改进的DCN卷积采样点

2.3 重构YOLOv8的目标检测头

目标检测头是用来对特征金字塔输出特征进行目标检测的，它包括了卷积层、池化层和全连接层

等。在原YOLOv8目标检测层特征图有3层(P3、P4、P5)，分别用于检测8×8、16×16、32×32大小的特征图，但是对于较小目标检测效果较差，因此增加P2层，用于检测160×160大小的特征图，实现较小目标检测。但是增加P2层会导致模型参数量以及计算复杂度的提升，因此，在骨干网络与neck特征链接处添加了额外的Conv_C模块，通过对输出特征通道进行调整，降低了模型参数与计算复杂度，改进的UAVA-YOLO网络模型如图2所示。但是，随着通道数的降低，对于neck解码部分需要更高效的特征提取模块，获取有效的语义信息。所以，在neck部分采用改进的C2f_ODConv模块。动态卷积层将 n 个卷积核的线性组合与注意力机制进行动态加权，使卷积操作依赖于输入，如式(2)所示。ODConv通过引入注意力机制对卷积核空间的4个维度进行学习，提取更加关注的语义信息。

$$y = (a_{w1}W_1 + \dots + a_{wn}W_n) \times x \quad (5)$$

ODConv中使用通道注意力机制，计算 W_i 卷积核的4个维度，但没有关注到特征图中的空间信息。因此改进的C2f_ODConv添加了卷积注意力模块(CBAM)，关注到空间维度特征信息，使C2f_ODConv的 W_i 卷积核提取具有通道和空间信息的语义信息。

2.4 添加BIFPN

传统的特征融合面对多尺度目标检测无法充分地结合上下文语义信息。因此，为解决高低层特征融合时的多尺度问题，本文引入BIFPN加权双向特征金字塔替换YOLOv8的网络模型中特征融合网络。BIFPN相比特征图金字塔网络(feature pyramid network, FPN)的改进之处主要在于其能够更有效地融合不同尺度的特征信息，并且通过双向特征路径增强解决目标检测中上下文信息融合的问题。BIFPN能够更好地应对目标检测中的多尺度问题，提升模型的性能。

FPN和BIFPN加权双向特征金字塔如图4所示，BIFPN相比FPN在P3~P7部分具有自上而下的通路传递着高层的语义信息，同时P3~P6具有跳链通路传递同一层特征图的语义信息。

在目标检测网络中，引入BIFPN模块能够通过自底向上的方式构建多尺度特征图金字塔，有效地实现特征融合，同时捕捉局部细节和全局上下文信息。这一结构设计有助于提升网络对多尺度目标

的处理能力，尤其是在检测小目标和远距离目标时表现出色。

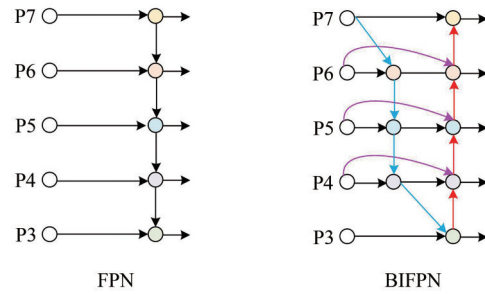


图4 FPN和BIFPN的网络模型

此外，BIFPN通过共享卷积核和更加高效的设计，能够有效减少计算量，这使得网络在保持高性能的同时更加高效。同时，BIFPN还能够增强网络对上下文信息的整合能力，为后续的目标检测任务提供更丰富的特征表示。因此，本文引入BIFPN模块解决多尺度目标处理、捕捉上下文信息等方面的问题。

2.5 损失函数的改进

YOLOv8的损失函数更加侧重于检测大物体或者具有较高置信度的物体，当面对小物体和遮挡或者重叠的物体时，检测的准确性较低。为解决这个问题，对损失函数的优化方向有分类和回归两部分。分类损失函数仍采用交叉熵损失函数(BCELoss)，回归部分采用分布焦点损失函数(DFL)和目标边界框损失函数(BBRL)，其完全损失函数表示为：

$$f_{\text{loss}} = \lambda_1 f_{\text{BCE}} + \lambda_2 f_{\text{DFL}} + \lambda_3 f_{\text{BBRL}} \quad (6)$$

为了提高模型的泛化能力和类别性能，使用Wise-IoU损失函数^[15]作为边界框的损失函数。当训练数据标注质量较低时，损失函数结合动态非单调聚焦机制，利用“离群值”对锚框质量进行评估，避免对模型几何因素(如距离、宽高比)的过度惩罚。

同时，当预测框与目标框重合程度较高时，损失函数通过弱化几何因子的惩罚，使模型在训练干预较少的情况下获得更好的泛化能力。基于此，本文采用了具有两层注意机制和动态非单调FM机制的Wise-IoU。其表达式如式(7)和式(8)所示。

$$f_{\text{BBRL}} = \left(1 - \frac{W_i H_i}{S_u}\right) \exp \left(\frac{\left(x_p - x_{gt}\right)^2 + \left(y_p - y_{gt}\right)^2}{W_\xi^2 + H_\xi^2} \right) \gamma \quad (7)$$

$$\gamma = \frac{\beta}{\delta \alpha^{\beta - \sigma}} \quad (8)$$

其中, W_i 和 H_i 是预测框与实际框的交集的长度和宽度, S_u 是预测框与实际框的联合集, α 和 δ 是超参数, β 表示预测框的差异程度。因此, 利用 β 构造非单调焦点数, 可以为异常值较大的预测框分配较小的梯度增益, 有效减少低质量训练样本的有害梯度。其预测框与 GT 框示意图如图 5 所示。 X_{gt} 和 Y_{gt} 表示预测框坐标值, x 和 y 代表 GT 框的坐标值, H 和 W 以及 H_i 和 W_i 分别代表两个方框的宽度和高度。

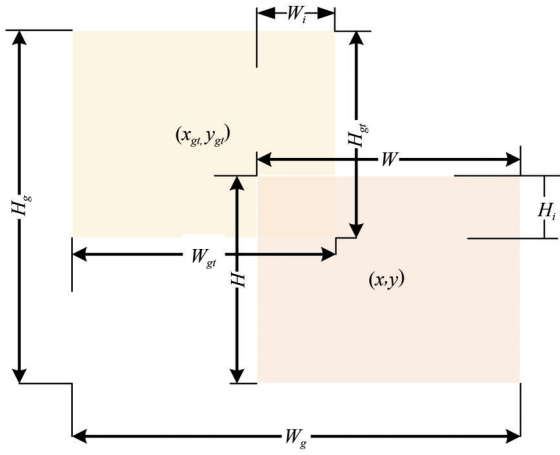


图5 Wise-IoU 预测框示意

3 实验结果与分析

3.1 实验数据集

实验数据集采用中国天津大学机器学习与数据挖掘实验室 AI SKYEYE 团队收集的 VisDrone2019 数据集^[16]和 UAVDT 数据集^[17]。VisDrone2019 是由无人机在不同场景和天气下航拍得到的, 一共收集了 10 个类别、8 599 张图片、540 000 个标签信息。

VisDrone2019 数据集中存在大量的密集小目标, 其目标尺寸如图 6 所示。从图 7 中可以得出该数据集的 GT 框长宽比都在 0.1 以下。

AI SKYEYE 团队将 VisDrone2019 数据集进行划分, 其中 6 471 张图片做训练集, 548 张图片做验证集, 1 580 张图片做测试集。图 7 展示了 VisDrone2019 数据集中部分图片的标注信息。

UAVDT 数据集是关于无人机目标检测追踪的大规模视频数据集, 包含 10 小时的原始视频, 约 8 000 个具有代表性的视频帧, 这些视频帧带有手动标注边界框和一些有用的标签, 例如车辆类别和遮挡。该数据集是由无人机在各种复杂场景下捕获的, 主要用于执行 3 个基本任务: 目标检测 (DET)、

单目标跟踪 (SOT) 和多目标跟踪 (MOT), UAVDT 数据集部分展示如图 8 所示。该实验采用的目标检测任务数据集包括 31 353 张训练图片、9 382 张验证图像。

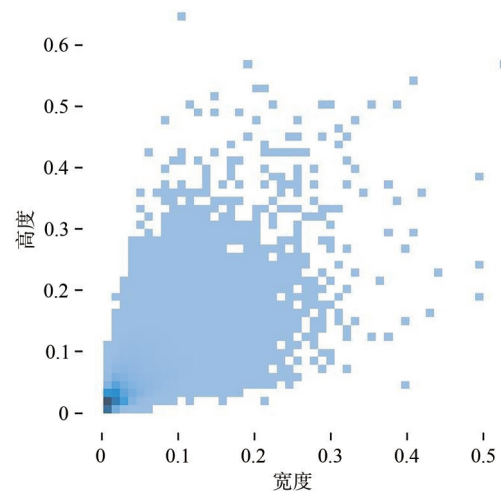


图6 VisDrone2019 数据集宽高分布



图7 VisDrone2019 数据集部分展示



图8 UAVDT 数据集部分展示

3.2 实验环境

模型的实验环境: CPU 为 Intel®Core (TM) i5-12 400F, 运行内存 32 GB, GPU 为 GeForce RTX 4 080, 显存 16 GB, 操作系统为 64 位 Windows 11。采用 Pytorch 1.8 深度学习框架, CUDA 版本为 10.2, epochs 为 200。

3.3 实验评价指标

本文采用 $\text{mAP}@0.5$ 、精确率 P 、召回率 R 与参数量作为评价指标。

$$P = \frac{TP}{TP + FP} \quad (9)$$

$$R = \frac{TP}{TP + FN} \quad (10)$$

其中, TP 为正确样本数, FP 为误报样本数, FN 为漏检的正样本数。平均精度 (AP) 是衡量模型检测效果的常用标准, AP 是召回率取值 0 到 1 时准确率的平均值, mAP 综合考虑了准确率和召回率, 能够很好地衡量目标检测效果。

$$\text{mAP} = \frac{1}{n} \sum_{k=1}^n J(P, R)_k \quad (11)$$

其中, $J(P, R)_k$ 为平均精度函数, 即类别编号 k 时的准确率 P 与召回率 R 构成的 P - R 曲线下面积。IoU 的阈值为 0.5 时的 mAP 称为 $\text{mAP}@0.5$, 将其作为目标检测模型检测精度的评价指标。

3.4 实验结果分析

在 YOLOv8n 模型中验证原始模块与改进模块的有效性。实验结果如表 1~表 4 所示。

表 1 改进的 C2f_DCN 对比实验

模型	精确率	召回率	$\text{mAP}@0.5$	参数量/MB
YOLOv8n	44.9%	33.0%	33.0%	3.0
+DCN	44.7%	33.3%	33.6%	3.1
+C2f_DCN	45.0%	33.5%	34.0%	3.2

表 2 重构的检测层对比实验

模型	精确率	召回率	$\text{mAP}@0.5$	参数量/MB
YOLOv8	44.9%	33.0%	33.0%	3.0
+ODConv	43.9%	32.8%	32.7%	3.1
Our	46.2%	35.4%	35.4%	3.9

表 3 BIFPN 对比实验

模型	精确率	召回率	$\text{mAP}@0.5$	参数量/MB
YOLOv8	44.9%	33.0%	33.0%	3.0
+BIFPN	44.5%	33.1%	33.2%	3.2

表 4 Wise-IoU 损失函数对比实验

模型	精确率	召回率	$\text{mAP}@0.5$	参数量/MB
YOLOv8	44.9%	33.0%	33.0%	3.0
Wise-iou	45.0%	33.9%	33.6%	3.0

从表 1 可知, 在 YOLOv8 的 backbone 使用 C2f_DCN 卷积, $\text{mAP}@0.5$ 达到 34.0%, 相比原始

的 DCN 卷积, $\text{mAP}@0.5$ 提高了 0.4%。

为了不增加模型的参数量, 在 neck 部分添加 P2 检测层。本文使用 Conv_C 模块对检测特征层进行降维, 同时为避免通道压缩造成的语义信息丢失, 使用改进的 ODConv 卷积, 加强模型的特征提取能力。

从表 2 可知, 在 YOLOv8 的 neck 部分使用改进的 C2f_ODConv 并添加 P2 特征层, 在参数量仅比基线模型 YOLOv8 增加 0.9 MB 的情况下, $\text{mAP}@0.5$ 达到 35.4%, 相比使用原始的 ODConv 卷积, $\text{mAP}@0.5$ 提高了 2.7%。

为了使模型在特征融合时能够充分结合高层语义信息与底层语义信息, 使用 BIFPN 特征融合模块。

从表 3 可知, 使用 BIFPN 特征融合模块后, $\text{mAP}@0.5$ 达到 33.2%。

为提高模型的收敛速度与准确性, 本文使用 Wise-IoU 损失函数。从表 4 可知, 使用 Wise-IoU 损失函数后, $\text{mAP}@0.5$ 提高了 0.6%。

以 YOLOv8n 为基线模型, 将所改进的模块逐步添加到基线模型, 实验结果如表 5 所示。首先在骨干网络中使用改进的 C2f_DCN 模块, 提高模型的特征提取能力。其次, 对 neck 部分进行重构, 添加 P2 特征层, 并使用改进的 C2f_ODConv 卷积替换原模型的 C2f 模块 (AD_P2)。然后, 使用 BIFPN 模块结合上下文语义信息。最后使用 Wise-IoU 损失函数。

表 5 在 VisDrone 数据集消融实验

序号	C2f_DCN	AD_P2	BIFPN	Wise-IoU	$\text{mAP}@0.5$	参数量/MB
1					33.0%	3.0
2	√				34.0%	3.2
3	√	√			35.8%	3.9
4	√	√	√		36.7%	3.9
5	√	√	√	√	37.4%	3.9

表 5 中序号 1 为原始的 YOLOv8n 基线模型, 序号 2~4 为分别添加不同模块得到的改进模型。从表 5 可以看出, 当在基线模型基础上添加 C2f_DCN 模块时, $\text{mAP}@0.5$ 达到 34.0%, 相比基线模型提高 1%。在序号 3 中对 neck 部分进行重构, 使用 Conv_C 模块将骨干网络输出通道调整为 512, 该方法不仅增加了 P2 特征检测层, 而且模型参数量仅有 3.9 MB, 其 $\text{mAP}@0.5$ 达到 35.8%。在序号 4

中使用 BIFPN 进行上下文语义信息融合, mAP@0.5 达到 36.7%。在序号 5 中使用 Wise-IoU 损失函数, 其 mAP@0.5 达到 37.4%。实验结果证明改进的 UAVAI-YOLO 模型相比基线模型, mAP@0.5 提升了 4.4%。

将 UAVAI-YOLO 模型与目前主流目标检测模型在 VisDrone2019 数据集和 UAVDT 数据集进行对比, 实验结果如表 6 和表 7 所示。

表 6 在 VisDrone2019 数据集与其他主流目标检测模型对比

模型	Input	backbone	mAP@0.5	参数量/MB
YOLOv8n	640×640	CSPDarkNet50	33.0%	3.0
YOLOv7tiny	640×640	CSPDarkNet50	26.8%	6.0
YOLOv5s	640×640	CSPDarkNet50	34.9%	9.2
YOLOv4s	640×640	CSPDarkNet50	26.5%	9.2
YOLOxs	640×640	CSPDarkNet50	26.3%	8.9
UCHNet ^[18]	640×640	ResNet50	32.8%	17.0
MobilNet-SDD ^[19]	512×512	MobilNet	34.2%	3.9
KPE-YOLOv5 ^[20]	640×640	CSPDarkNet50	39.2%	8.2
YOLO-UAVlite ^[21]	640×640	CSPDarkNet50	36.6%	3.0
UAVAI-YOLO	640×640	CSPDarkNet50	37.4%	3.9

从表 6 和表 7 可以看出, 相比 YOLOv8n, UAVAI-YOLO 在 VisDrone2019 数据集上的 mAP@0.5 提高了 4.4%, 在 UAVDT 数据集上, mAP@0.5 提高了 1.1%。与其他网络模型相比, UAVAI-YOLO 模型的检测精度都具有优势, 虽然在 VisDrone2019 数据集中检测精度低于 KPE-YOLOv5 模型, 但是

UAVAI-YOLO 模型参数量比 KPE-YOLOv5 模型低 4.3 MB。整体而言, UAVAI-YOLO 模型对于小目标的检测效果优于其他模型。

表 7 在 UAVDT 数据集与其他主流目标检测模型对比

模型	Input	backbone	mAP@0.5
YOLOv8n	640×640	CSPDarkNet50	33.8%
YOLOv7tiny	640×640	ResNet50	26.8%
YOLOv5s	640×640	CSPDarkNet50	34.9%
SSD	512×512	VGG16	21.4%
DMNet ^[22]	600×1 000	ResNet50	24.6%
R-FCN ^[23]	600×100	ResNet-101	17.5%
UAVAI-YOLO	640×640	CSPDarkNet50	34.9%

3.5 可视化结果分析

为验证 UAVAI-YOLO 模型的检测效果, 本文使用 YOLOv8 模型与 UAVAI-YOLO 模型在 VisDrone2019 数据集测试集进行推理, 结果如图 9 所示。从 UAVAI-YOLO 模型预测图标红区域可以看出, 本文模型在无人机目标检测精度方面具有较高的性能。

4 结论

针对目前主流检测模型对密集小目标检测存在误检、漏检等问题, 基于 YOLOv8 提出一种航拍下密集小目标检测 UAVAI-YOLO 模型。首先, 引入了改进的 C2f_DCN 模块到骨干网络中, 这显著增强了网络的特征提取性能。接着, 为了更有效地利用 P2 特征层中的细小目标语义信息, 本文重新

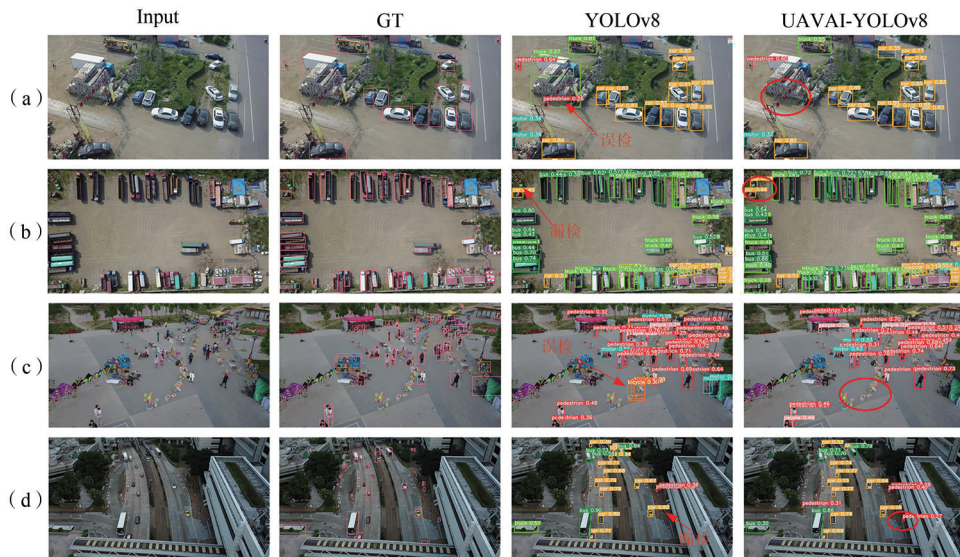


图9 在 VisDrone2019 测试集的结果

设计了neck部分。此外,考虑到P2特征层引入的参数量增加问题,本文在骨干网络与neck特征层连接处引入了Conv_C模块,这不仅减少了模型参数,还能提取更丰富的语义信息。随后,为了更紧密地整合P2、P3、P4、P5特征层中的上下文语义信息,本文特征融合采用BIFPN架构,以更好地辅助网络模型捕捉上下文依赖关系。最终,为了提高模型对目标检测框的判断准确性,本文采用Wise-IoU损失函数替代原有的损失函数。在实验结果中验证了本文改进的DCN卷积和重构的neck部分,在仅增加较少的参数量的情况下,能够极大地提升模型性能。UAVAI-YOLO模型在VisDrone2019数据集和UAVDT数据集的mAP@0.5检测精度分别达到37.4%和34.9%,比基线模型分别提升了4.4%和1.1%,这证明UAVAI-YOLO模型具有更高的检测精度和鲁棒性。

参考文献:

- [1] 宗有帅, 俞明, 张红顺, 等. 美军大中型察打一体无人机发展特点及趋势[J]. 战术导弹技术, 2022(6): 32-41.
ZONG Y S, YU M, ZHANG H S, et al. Developing characteristics and trend of US military's large and medium-sized reconnaissance/strike UAV[J]. Tactical Missile Technology, 2022(6): 32-41.
- [2] 欧阳桃花, 郑舒文, 张凤, 等. 研究型大学服务国家战略工程研究: 以长鹰无人机原始创新为例[J]. 科学学与科学技术管理, 2023, 44(1): 20-37.
OUYANG T H, ZHENG S W, ZHANG F, et al. Research on national strategic engineering of research universities: a case study of the original innovation from changying UAV[J]. Science of Science and Management of S & T, 2023, 44(1): 20-37.
- [3] LUCAS Simon, 沈甜雨, 王晓, 等. 基于统计前向规划算法的游戏通用人工智能[J]. 智能科学与技术学报, 2019, 1(3): 219-227.
SIMON L, SHEN T Y, WANG X, et al. General game AI with statistical forward planning algorithms[J]. Chinese Journal of Intelligent Science and Technology, 2019, 1(3): 219-227.
- [4] QU Z, GAO L Y, WANG S Y, et al. An improved YOLOv5 method for large objects detection with multi-scale feature cross-layer fusion network[J]. Image and Vision Computing, 2022, 125: 104518.
- [5] 黄路, 李泽平, 杨文帮, 等. 基于区域感知的多尺度目标检测算法[J]. 计算机工程, 2023, 49(12): 178-185.
HUANG L, LI Z P, YANG W B, et al. Multi-scale target detection algorithm based on area awareness[J]. Computer Engineering, 2023, 49(12): 178-185.
- [6] 张朝阳, 张上, 王恒涛, 等. 多尺度下遥感小目标多头注意力检测[J]. 计算机工程与应用, 2023, 59(8): 227-238.
ZHANG Z Y, ZHANG S, WANG H T, et al. Multi-head attention detection of small targets in remote sensing at multiple scales[J]. Computer Engineering and Applications, 2023, 59(8): 227-238.
- [7] 刘晋川, 黎向峰, 刘安旭, 等. 改进RetinaNet的无人机小目标检测[J]. 科学技术与工程, 2023, 23(1): 274-282.
LIU J C, LI X F, LIU A X, et al. Improved RetinaNet for unmanned aerial vehicle small target detection[J]. Science Technology and Engineering, 2023, 23(1): 274-282.
- [8] JU M R, LUO J N, LIU G Q, et al. A real-time small target detection network[J]. Signal, Image and Video Processing, 2021, 15(6): 1265-1273.
- [9] 胡均平, 王鸿树, 戴小标, 等. 改进YOLOv5的小目标交通标志实时检测算法[J]. 计算机工程与应用, 2023, 59(2): 185-193.
HU J P, WANG H S, DAI X B, et al. Real-time detection algorithm for small-target traffic signs based on improved YOLOv5[J]. Computer Engineering and Applications, 2023, 59(2): 185-193.
- [10] SHAO Y F, YANG Z X, LI Z H, et al. Aero-YOLO: an efficient vehicle and pedestrian detection algorithm based on unmanned aerial imagery[J]. Electronics, 2024, 13(7): 1190.
- [11] CAI D L, ZHANG Z Y, ZHANG Z. Corner-point and foreground-area IoU loss: better localization of small objects in bounding box regression[J]. Sensors, 2023, 23(10): 4961.
- [12] 顾宙瑜, 於跃成, 者甜甜. Rapider-YOLOX: 高效的轻量级目标检测网络[J]. 智能科学与技术学报, 2023, 5(1): 92-103.
GU Z Y, YU Y C, ZHE T T. Rapider-YOLOX: lightweight object detection network with high precision[J]. Chinese Journal of Intelligent Science and Technology, 2023, 5(1): 92-103.
- [13] 宋华杰, 周磊. 基于函数改进的YOLOv3车辆检测与识别算法[J]. 智能科学与技术学报, 2023, 5(4): 535-542.
SONG H J, ZHOU L. Vehicle detection and recognition algorithm based on function improvement of YOLOv3[J]. Chinese Journal of Intelligent Science and Technology, 2023, 5(4): 535-542.
- [14] DAI J F, QI H Z, XIONG Y W, et al. Deformable convolutional networks[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 764-773.
- [15] TONG Z J, CHEN Y H, XU Z W, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism[EB]. arXiv preprint, 2023, arXiv: 2301.10051v3.
- [16] DU D W, ZHU P F, WEN L Y, et al. VisDrone-DET2019: the vision meets drone object detection in image challenge results[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Piscataway: IEEE Press, 2019: 213-226.
- [17] YU H Y, LI G R, ZHANG W G, et al. The unmanned aerial vehicle benchmark: Object detection, tracking and baseline[J]. International Journal of Computer Vision, 2020, 128(5): 1141-1159.
- [18] LIAO J J, PIAO Y C, SU J H, et al. Unsupervised cluster guided object detection in aerial images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 11204-11216.
- [19] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 10778-10787.
- [20] YANG R J, LI W F, SHANG X N, et al. KPE-YOLOv5: an improved small target detection algorithm based on YOLOv5[J]. Electronics, 2023, 12(4): 817.
- [21] LIU C, YANG D G, TANG L, et al. A lightweight object detector based on spatial-coordinate self-attention for UAV aerial images[J]. Remote Sensing, 2022, 15(1): 83.
- [22] LI C L, YANG T, ZHU S J, et al. Density map guided object detection in aerial images[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2020: 737-746.
- [23] DAI J F, LI Y, HE K M, et al. R-FCN: object detection via region-based fully convolutional networks[C]// Proceedings of the 29th Annual Conference on Neural Information Processing Systems, [S.l.:s.n.], 2016: 379-387.

[作者简介]



何植仟 (1997-), 男, 大连海洋大学信息工程学院硕士生, 主要研究方向为计算机视觉、图像处理。



曹立杰 (1978-), 女, 博士, 大连海洋大学信息工程学院副教授、副院长, 主要研究方向为计算机视觉、无线传感器网络。