

DOI:10.13232/j.cnki.jnju.2025.02.003

Transformer 结构与注意力机制融合的 无人机图像小目标识别方法

王国帅¹, 黄贻望^{1,3*}, 刘建晓², 毛志⁴, 刘声¹

(1. 铜仁学院大数据学院/经济管理学院, 铜仁, 554300; 2. 华中农业大学信息学院, 武汉, 430070;

3. 民族教育信息化教育部重点实验室, 云南师范大学, 昆明, 650000;

4. 智能计算与信息处理教育部重点实验室, 湘潭大学, 湘潭, 411105)

摘要: 目前无人机(Unmanned Aerial Vehicle, UAV)航拍图像中存在较多密集和尺寸较小的实例, 识别效果往往较差。针对此问题, 在 YOLOv5 的骨干网络中使用带有 Transformer 结构的 C3TR 模块替换原有的 C3 (CSP Bottleneck with 3 Convolutions) 模块来增强骨干网络的特征提取能力, 之后, 在 SPPF 层后加入 CA (Coordinate Attention) 模块来增强模型对小目标区域的关注度。其次, 在颈部网络中使用 ConvNeXtBlock 模块替换 C3 模块, ConvNeXtBlock 的深层卷积能够进一步增强对小目标细节的识别, 提高目标检测的准确性。最后, 使用 ECIoU 损失函数替换 CIoU 损失函数来进一步提升模型的收敛速度和精度。实验结果表明, 在 VisDrone2019 公开数据集上, 与基线模型 YOLOv5l 相比, 改进模型的 $mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别提升 9.5% 和 6.3%, 和其他模型相比也有比较好的效果。

关键词: 无人机, YOLOv5, 小目标, Transformer, 注意力机制

中图分类号: P391.41

文献标志码: A

Small target recognition method integrating Transformer structure and attention mechanism for UAV images

Wang Guoshuai¹, Huang Yiwan^{1,3*}, Liu Jianxiao², Mao Zhi⁴, Liu Sheng¹

(1. School of Data Science/School of Economics and Management, Tongren University, Tongren, 554300, China;

2. College of Informatics, Huazhong Agricultural University, Wuhan, 430070, China; 3. Key Laboratory of

Education Informatization for Nationalities, Ministry of Education, Yunnan Normal University,

Kunming, 650000, China; 4. Key Laboratory of Intelligent Computing and Information Processing,

Ministry of Education, Xiangtan University, Xiangtan, 411105, China)

Abstract: In UAV (Unmanned Aerial Vehicle) aerial images, there are numerous dense and small instances, often resulting in suboptimal detection performance. To address this issue, this paper replaces the original C3 (CSP bottleneck with 3 convolutions) module in the backbone of YOLOv5 with a C3TR module incorporating a Transformer structure to enhance the backbone's feature extraction capability. Then, a CA (Coordinate Attention) module is added after the SPPF layer to improve the model's focus on small-object regions. Additionally, the ConvNeXtBlock module is employed in the neck network to

基金项目: 国家自然科学基金(62066040), 民族教育信息化教育部重点实验室开放课题(EIN2024B003), 2023 年度智能计算与信息处理教育部重点实验室开放课题(2023ICIP05), 国家留学基金委西部地区人才培养特别项目(202108525007), 铜仁市大数据智能计算与应用重点实验(铜仁市科研[2022]5 号), 贵州省科技厅基础研究计划(黔科合基础-zk[2025]面上 096, ZK[2022]557), 贵州省基础研究(自然科学)(MS[2025]096)

收稿日期: 2024-12-24

* 通信联系人, E-mail: yjsyhyw@gztrc.edu.cn

replace the C3 module, where the deep convolutions in ConvNeXtBlock further enhance the recognition of small-object details, thereby improving the detection accuracy. Finally, the ECIoU loss function is introduced in place of the CIoU loss function to further improve the model's convergence speed and accuracy. Experimental results indicate that the modified model achieves increases of 9.5% and 6.3% in $mAP_{0.5}$ and $mAP_{0.5:0.95}$, respectively, on the VisDrone2019 public dataset compared to the baseline YOLOv5l model, also outperforming other models.

Keywords: UAV, YOLOv5, small object, Transformer, attention mechanism

近年来,随着低空经济的兴起,无人机被广泛应用于城市安全监控^[1]、紧急救援^[2]、交通管理^[3]等新兴领域,极大地提高了作业效率。新技术的发展不仅推动了经济增长,还为应对各种社会和环境问题提供了新的视角和解决方案。在这些领域中,无人机航拍图像识别的准确性尤其重要,它直接影响无人机执行任务的效率和可靠性。和一般的图像目标识别相比,无人机航拍图像识别面临更复杂的挑战,因为无人机的飞行高度相对较高,拍摄的目标物体通常较小,这极大地增加了识别难度。此外,无人机常常需要在环境复杂且多变的条件下工作,例如在多变的光照、天气和背景干扰中进行准确的识别任务,这些因素都可能影响目标识别的准确性和实时性。因此,设计一种对小目标物体具有较强识别准确性和鲁棒性的目标检测算法非常重要。

目前的目标检测算法分两类:第一类是一阶段目标检测算法,包括YOLO系列^[4-10]、RetinaNet^[11]等;第二类是两阶段目标检测算法,如RCNN^[12]、Fast-RCNN^[13]等。这两类算法在目标识别策略和识别效率上存在显著差异。和两阶段目标检测算法相比,一阶段目标检测算法的主要优势在于其简洁性和速度,如YOLO系列可直接在网络的输出层预测目标的类别和位置,无须在单独的区域提议步骤^[14]。因此,一阶段目标检测算法在实时性检测应用中非常有效,能够在较高的帧率下进行检测,非常适合部署在无人机上进行航拍图像的目标识别任务。但对于准确性要求较高的场景,以上的原始模型和方法均存在精度不足、鲁棒性不强的缺点。针对此问题,国内外很多学者在原有目标检测模型的基础上提出了诸多改进。李姝等^[15]在YOLOv8的基础上通过改进特征提取网络和引入非线性平滑处理等方法来提升对航拍图像识别的精确率和鲁棒性。Xu and Wu^[16]

使用EfficientNet作为主要骨干网络,替换YOLOX的骨干网络来提升航拍图像目标识别的准确性。潘玮等^[17]改进YOLOv8的空间金字塔池化层部分来增强特征的融合能力,检测效果较好。蒋伟等^[18]改进YOLOv8的C2f模块来提高算法对小目标检测的性能。Zhang^[19]对YOLOv8的颈部网络进行改进,将RepVGG模块作为下采样层,以增强网络学习多尺度特征的能力。

综上,研究者对YOLO网络进行了各种改进以提升无人机航拍图像的识别能力,但在一些复杂场景下依旧存在检测不准确、误检等问题。为了提升模型对无人机航拍图像小目标的检测效果,本研究在YOLOv5的基础上提出三点改进。

(1)在骨干网络中使用带有Transformer结构的C3TR模块来替换原始C3模块,并在SPPF层后引入坐标注意力(Coordinate Attention, CA)模块,以增强骨干网络对小尺度目标特征的提取能力。

(2)在颈部网络中使用ConvNeXtBlock模块替换C3模块,增强模型对特征的表达能力,强化模型对于关键空间位置的响应速度。

(3)使用ECIoU损失函数替换原始的CIoU损失函数,进一步提升模型的收敛速度。

1 相关工作

1.1 YOLOv5 概述 首先,在进行特征提取之前的图像预处理部分,YOLOv5采用Mosaic数据增强^[20]、锚框自适应调整和图像尺寸动态缩放等技术对图像数据进行优化^[21]。Mosaic数据增强通过随机组合不同图像片段来丰富训练数据的变化性,增强模型对不同图像特征的学习能力。自适应锚框算法针对输入图像的真实框和预测框之间的差别不断地进行迭代计算,以更准确地预测不同尺寸的目标。

YOLOv5 的主要网络结构包括骨干网络、颈部网络、头部网络三部分。

(1) 骨干网络负责对输入图像进行特征的提取。它由多个不同的结构组成,包括 Focus, CBS, CSP 和 SPPF 模块,负责执行卷积、批量归一化和非线性激活函数,为颈部网络提取特征图。

(2) 颈部网络作为连接骨干网络和头部网络的中间层,负责对骨干网络提取的特征进行融合,主要包括特征金字塔网络 (Feature Pyramid Networks, FPN) 和路径聚合网络 (Path Aggregation Network, PAN)。FPN 通过自顶向下的结构增强高层语义特征的传递, PAN 通过自底向上的结构增强低层细节特征的传递,这样的组合使模型能够在不同尺度上有效地检测目标。

(3) 头部网络作为模型的最后一部分,负责最终的目标检测任务,包括对象分类、位置定位以及置信度评分等。

YOLOv5 还采用了分类损失、定位损失和置信度损失三种损失函数^[22]同时作用,使模型能够准确地分类、定位目标,并评估模型检测的置信度。

1.2 小目标检测定义 对小目标的定义是根据被检测目标的大小和被检测目标占整个被检测图像的比例来进行的。在第一类小目标的定义中,被检测目标的分辨率小于 32×32 像素即为小目标,这是在 MS COCO 数据集^[23]中被定义的。而对于被检测目标占整个被检测图像的比例进行定义时,一般认为目标的边界框占整个图像的 $0.08\% \sim 0.58\%$ 即为小目标。

无论对哪种定义的小目标物体的检测都是一种挑战,原因在于小目标尺寸较小,缺乏相应的纹理、颜色、形状^[24]等外观信息,经过多次下采样后,特征信息减少,检测难度变大。

2 本文算法

本文对 YOLOv5 模型的具体改进如下。

(1) 在骨干网络中使用带有 Transformer 结构的 C3TR 模块替换原始的 C3 模块,因为小目标在图像中的像素数量少,纹理颜色等信息有限,传统的卷积神经网络 (Convolutional Neural Network, CNN) 在处理这些特征时往往存在局限性。和

CNN 相比,Transformer 结构的自注意力机制能够捕获全局依赖关系,可以提供更丰富的上下文信息,这对于小目标特征的提取非常重要。再将 CA 模块加入 SPPF 层后,这一设计显著增强了模型对特征图中空间信息的处理能力。CA 模块增加了特征图的每个位置的水平和垂直方向的注意力,以提高模型对空间结构的敏感性,使模型能更有效地捕捉和理解目标的空间布局。

(2) 在颈部网络中,使用 ConvNeXtBlock 模块替换原有的 C3 模块。ConvNeXtBlock 是一种改进后的卷积模块,通过更深层的卷积操作优化了对多尺度特征的捕获。这一模块结构不仅包括深层卷积和跳跃连接,还融入了归一化层,使其能够更有效地捕获和整合不同尺度上的特征信息,增强对小目标特征细节的识别,提高了目标检测的准确性和鲁棒性。

(3) 使用 ECIoU 损失函数替换原始的 CIoU 损失函数。ECIoU 损失函数使模型在训练期间能更快地优化边界框的预测,尤其在处理复杂场景的小目标检测时可以确保更快的误差反馈和更准确的梯度更新,显著提高了模型的训练效率和最终性能。

改进的 YOLOv5 模型网络结构如图 1 所示,图中黑色加粗框为改进模块。

2.1 Transformer 结构的 C3 模块 YOLOv5 中的 C3 模块是改进的 CSP 结构,是为了增强特征提取能力而设计的。原始的 CSP 结构中,残差块通过堆叠形成来增加网络的深度,而 C3 模块进一步优化了这一设计,采用双分支结构,其中一个分支包含多个堆叠的 Bottleneck 残差块和一个标准的卷积层,另一个分支包含单个基本卷积模块。最后,将这两个分支的输出通过 Concat 操作合并,以融合各自提取的特征信息,增强模型对输入图像的特征理解。本文采用带有 Transformer 结构的 C3TR 模块,使用 Transformer 编码块替换 C3 模块中的 Bottleneck 残差块,其结构如图 2 所示。Transformer 编码块的核心组件包括多头自注意力机制 (Multi-Head Self-Attention, MSA)、多层感知机 (Multi-Layer Perceptron, MLP)、层标准化 (Layer Normalization, LN),其中,MSA 使模块能够在不同位置的特征间建立直接的依赖关系,

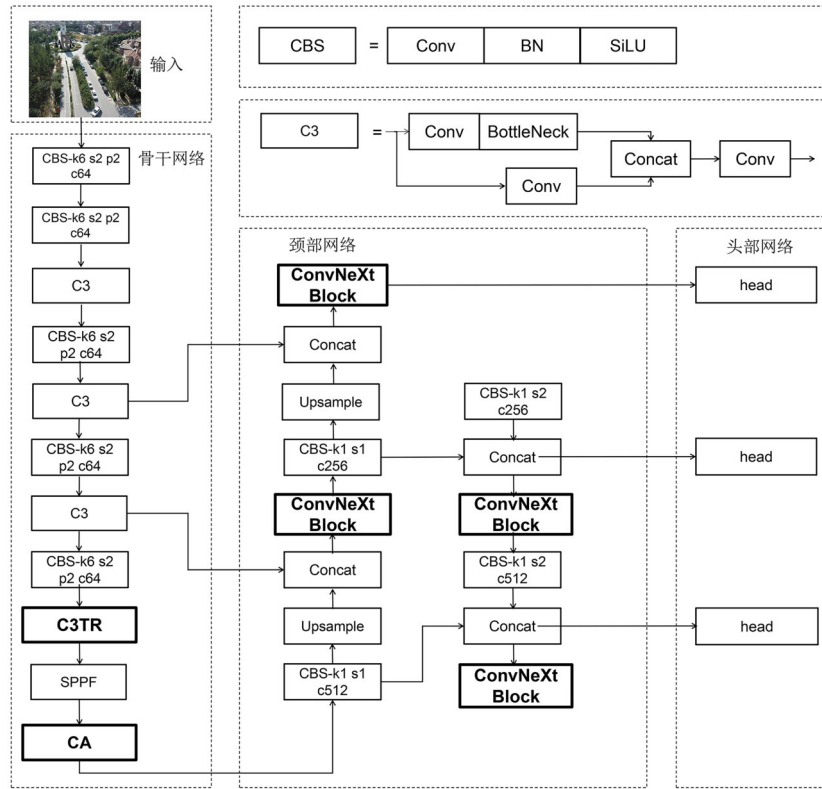


图1 改进后的YOLOv5网络的结构图

Fig. 1 The architecture of improved YOLOv5 network

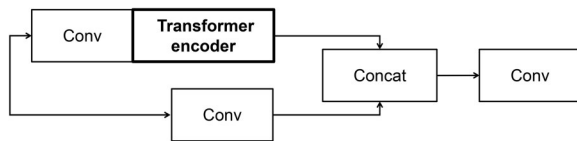


图2 C3TR模块的结构

Fig. 2 The architecture of C3TR module

捕捉更广泛的全局信息和上下文信息. 这对于处理小目标特征非常有利.

2.2 CA注意力机制 在航拍图像识别任务中, 小目标容易受到复杂背景的干扰, 所以在特征提取的最后阶段不容易聚合这些特征. 在骨干网络的最后加入CA模块, 可以对特征图中的每个位置分别进行水平和垂直方向的特征聚合, 不仅增强了模型对特定区域(如小目标所在位置)的关注, 还帮助模型在复杂的背景中区分和识别小目标. CA模块的网络结构如图3所示.

CA模块接收上层骨干网络提取的特征图, 这些特征图包含了通过深层网络和多尺度处理获得的丰富信息. 对输入的特征图进行X和Y两个

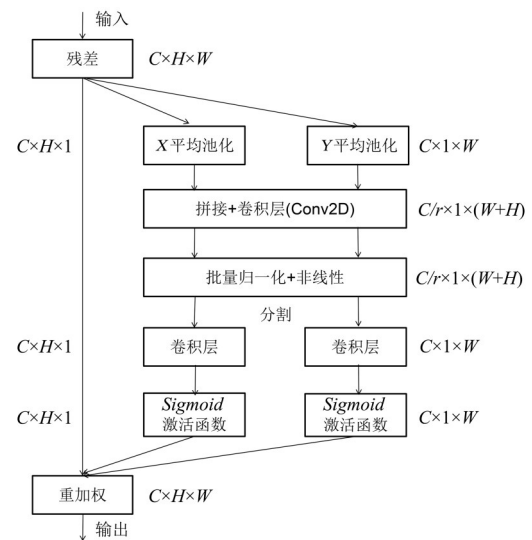


图3 CA模块的结构

Fig. 3 The architecture of CA module

方向的平均池化操作之后, 将平均池化的结果在通道维度上进行合并, 再通过一个2D卷积层进一步处理. 这一步骤不仅融合了来自不同空间方向的特征, 卷积运算也增强了特征的表达力. 对卷

积后的特征图进行批量归一化和非线性激活函数处理,以进一步增强模型的输出.处理后的特征图被分为两部分,每个部分分别经过一个独立的2D卷积层,每个卷积的输出均通过 *Sigmoid* 激活函数生成一个0~1的权重图.最后,对得到的注意力图与输入特征图进行逐元素的乘法操作,这一步精确调整了特征图每个位置的重要性,尤其增强了那些小目标识别中重要区域的特征.

2.3 ConvNeXtBlock 结构

当特征图被骨干网

络提取后,在颈部网络中进行特征的融合操作.在实际的无人机航拍图像目标检测任务中,不同尺寸目标的准确定位依然是不小的挑战.因此,本文选择在颈部网络中使用 ConvNeXtBlock 模块来替换原始的 C3 模块. ConvNeXtBlock 模块通过带有更大感受野的深度卷积来提升对小目标的识别能力,可以捕捉更加细微的特征,这对于在目标较密集的航拍图像中定位小尺寸目标非常有效. ConvNeXtBlock 模块的结构如图 4 所示.

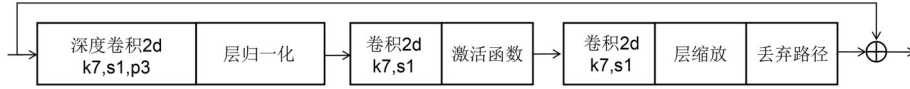


图 4 ConvNeXtBlock 模块的结构

Fig. 4 The architecture of ConvNeXtBlock module

首先使用深度分离卷积来减少参数数量,同时保持卷积网络的性能.分别处理输入特征图的每个通道,然后通过 1×1 的卷积合并这些通道的信息,再使用大卷积核或多尺度卷积核来扩大感受野,捕捉更广泛的上下文信息,这些操作可以增强网络对航拍图像局部信息的捕获能力.最后,经过层归一化和非线性激活函数来提高模型的收敛速度和性能.与传统的 C3 模块相似,ConvNeXtBlock 也包含残差连接,能够在增加网络深度的同时避免梯度消失.

2.4 损失函数改进 YOLOv5 的损失函数采用的是 CIOU 损失, CIOU 损失考虑了几个关键因素,包括边框的重叠面积、框中心点之间的距离以及长宽比.这种设计通过引入中心距离,提供了 IoU 相同的情况下更具区分的梯度,增加了损失函数的精确性. CIOU 的计算如式(1)所示:

$$CIOU = IoU - \frac{\rho^2(b, b^{gt})}{c^2} \alpha \nu \quad (1)$$

其中, α, ν 分别为权重函数和宽高比的度量值,其计算如式(2)和式(3)所示:

$$\alpha = \frac{\nu}{(1 - IoU) + \nu} \quad (2)$$

$$\nu = \frac{4}{\pi} \left(\arctan \frac{\omega^{gt}}{h^{gt}} - \arctan \frac{\omega}{h} \right)^2 \quad (3)$$

其中, ω 和 h 是真实框的宽和高, ω^{gt} 和 h^{gt} 是预测框的宽和高.这些计算帮助损失函数更精确地进行边框的调整,但它主要关注边框尺寸的差异,没有

考虑长宽比对置信度的影响.当 $\omega = k\omega^{gt}, h = kh^{gt}$ 时,惩罚项将失去效果.

EIoU 弥补了 CIOU 的缺点,它将长宽比的影响因子拆分成独立的长宽比损失,以独立地计算真实框和预测框的长宽. EIoU 损失函数考虑了重叠损失、中心距离损失和宽高损失三个部分,在训练中提供了更加全面的梯度信息. EIoU 损失函数的计算如式(4)所示:

$$Loss_{EIoU} = L_{IoU} + L_{dis} + L_{asp} = 1 - IoU + \frac{\rho^2(b^{gt}, b)}{c^2} + \frac{\rho^2(h^{gt}, h)}{c_h^2} + \frac{\rho^2(\omega^{gt}, \omega)}{c_w^2} \quad (4)$$

本文算法结合了 CIOU 和 EIoU 两种损失函数的特点,使用 ECIoU 作为模型的损失函数.此前,曹超和顾幸生^[25]在 YOLOv8 模型中使用 ECIoU 损失函数,有效地加快了模型的收敛速度,还提高了目标检测的整体精度. ECIoU 损失函数的计算如式(5)所示:

$$Loss_{ECIoU} = 1 - IoU + \alpha \nu + \frac{\rho^2(b^{gt}, b)}{c^2} + \frac{\rho^2(h^{gt}, h)}{c_h^2} + \frac{\rho^2(\omega^{gt}, \omega)}{c_w^2} \quad (5)$$

3 实验结果与分析

3.1 数据集处理 公开数据集 VisDrone2019^[26]由天津大学机器学习与数据挖掘实验室团队采集,包括在中国 14 个不同地区于不同时间、不同天气下采集的 10209 幅图像,包含行人、人、自行车、小轿车、货车、卡车、三轮车、雨篷三轮车、公交

车和摩托车在内的10个类别.数据集内的标注实例绝大多数属于小目标范围.从中选取训练集6471张,验证集548张.

3.2 实验环境及参数设置 实验在浪潮高性能人工智能平台上运行,硬件系统为Intel® Xeon® Platinum 8358P CPU@2.60 GHz, GPU 为NVIDIA A40×4, 256 GB内存,单张加速卡显存为45 GB,共160 GB.

代码编译环境为Python 3.8.7,加速卡的CUDA版本为11.2,深度学习框架选择PyTorch 2.2.1.

训练参数设置:训练轮数为100次,批处理大小为256,初始学习率为0.01,优化器为SGD,动量为0.937,权重衰减系数为0.005.

3.3 评价指标 使用的评价指标包括精确率(Precision, P)、召回率(Recall, R)和平均精准均值(Mean Average Precision, mAP).

(1)精确率(P)是一个用于评估分类模型性能的重要指标,是模型预测正确的正例占有被模型预测为正例的比例.其计算如式(6)所示:

$$P = \frac{TP}{TP + FP} \quad (6)$$

其中, TP 是模型正确预测为正类的样本数, FP 是模型正确预测为负类的样本数.

(2)召回率(R)也称查全率,在目标检测或其他分类任务中,主要衡量模型识别出的真正正例占有所有实际正例的比例.其计算如式(7)所示:

$$R = \frac{TP}{TP + FN} \quad (7)$$

其中, FN 是模型错误预测为正例的样本数.

(3)平均精准均值(mAP)是评估目标检测模型性能的一个核心指标,是对模型在不同阈值下精确率(P)和召回率(R)表现的整体评估. $mAP_{0.5}$

是 $IoU = 50\%$ 时的平均精度, $mAP_{0.5:0.95}$ 是 $IoU = 50\% \sim 95\%$,步长为5%的平均精度.其计算如式(8)所示,其中, N 为总样本数.

$$mAP = \frac{\sum_i^N AP}{N} \quad (8)$$

3.4 损失函数对比 对损失函数的改进可以进一步提升模型的收敛速度,提高目标检测的效率.

本文算法使用ECIoU损失函数来替换CIoU损失函数,实验结果证明,ECIoU损失函数的加入可以显著提高模型收敛速度.训练结果如图5所示.

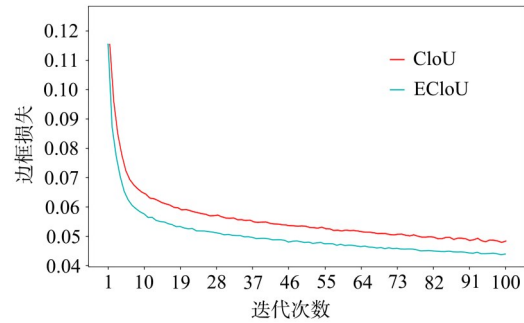


图5 ECIoU与CIoU的回归对比

Fig. 5 Regression of ECIoU and CIoU

3.5 消融实验 进行三组消融实验来验证本文算法三种改进方法的有效性,实验结果如表1所示.

改进1是使用带有Transformer结构的C3TR模块来替换骨干网络中的C3模块,并在SPPF层后加入CA模块,实验证明这项改进是有效的, $mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别提升了3.9%和4.2%.改进2是使用ConvNeXtBlock模块来替换颈部网络中的C3模块,使模型在特征融合阶段能更好地捕获全局信息,也能更准确地定位小目标,效果较

表1 消融实验的结果

Table 1 Results of ablation experiments

	C3TR+CA	ConvNeXt Block	ECIoU	P	R	$mAP_{0.5}$	$mAP_{0.5:0.95}$
YOLOv5	—	—	—	47.3%	34.8%	32.9%	17.9%
改进1	✓	—	—	49.4%	39.9%	36.8%	22.1%
改进2	—	✓	—	49.6%	40.1%	38.1%	22.6%
改进3	—	—	✓	48.4%	35.5%	34.2%	18.7%
本文算法	✓	✓	✓	52.8%	41.5%	42.4%	24.2%

好, $mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别提升了 5.2% 和 4.7%。改进 3 是使用 ECIoU 来替换原始的 CIoU 损失函数。3.4 的对比实验证明对损失函数的改进可以提升模型的收敛速度, 消融实验的结果也证明 ECIoU 损失函数可以有效地提升模型的精度, 其 $mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别提升了 1.3% 和 0.8%。

本文算法拼合了三种改进方法。实验结果表明, 和基线模型 YOLOv5 相比, 本文算法的精确率、召回率、 $mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别提升了 5.5%、6.7%、9.5% 和 6.3%。

3.6 对比实验 为了验证本文算法的性能, 与一

些算法进行了对比, 实验结果如表 2 所示, 表中黑体字表示最优的性能。由表可见, 本文算法在小轿车和公交车的检测中表现优异, 和最新的 YOLOv10n 模型相比, 精确率和 $mAP_{0.5}$ 分别提升了 3.6% 和 3.8%。另外, 对行人等小目标的检测, 本文算法的性能也占优, 和 Faster-RCNN, YOLOv3, YOLOv4, YOLOv5, YOLOv7-tiny, YOLOv8n, YOLOv10n, TPH-YOLOv5^[27] 相比, 其 $mAP_{0.5}$ 提高了 12.3%, 32%, 14.6%, 9.5%, 3.5%, 10.2%, 2.8% 和 3.8%。

综上, 本文算法和其他算法相比, 在小目标识别中具有优势。

表 2 本文算法和其他检测模型的性能对比

Table 2 Performance of our algorithm and other detection models

模型	精确率										$mAP_{0.5}$
	行人	人	自行车	小轿车	货车	卡车	三轮车	雨篷三轮车	公交车	摩托车	
Faster-RCNN	28.2%	16.3%	8.9%	76.9%	33.5%	39.1%	15.4%	28.2%	56.6%	26.8%	30.1%
YOLOv3	12.3%	12.6%	1.3%	41.4%	6.2%	5.6%	4.7%	1.5%	5.7%	12.2%	10.4%
YOLOv4	28.9%	17.2%	8.8%	75.5%	30.8%	38.0%	16.4%	28.2%	55.2%	24.5%	27.8%
YOLOv5	30.4%	17.7%	9.6%	77.2%	34.9%	39.7%	16.3%	30.0%	58.1%	28.0%	32.9%
YOLOv7-tiny	38.0%	25.8%	17.4%	78.2%	41.0%	40.9%	25.2%	21.7%	55.4%	40.0%	38.9%
YOLOv8n	33.6%	26.5%	8.2%	74.3%	36.4%	30.0%	22.3%	10.9%	44.5%	35.1%	32.2%
YOLOv10n	42.2%	33.1%	15.0%	78.5%	41.3%	38.1%	29.8%	17.2%	55.7%	44.8%	39.6%
TPH-YOLOv5	45.3%	36.7%	15.7%	77.9%	40.7%	37.6%	25.7%	12.1%	49.3%	44.7%	38.6%
本文算法	51.2%	38.7%	17.8%	82.1%	43.4%	38.7%	29.5%	15.2%	59.5%	47.3%	42.4%

3.7 对比分析 为了更好地观察本文算法的检测效果, 对模型的检测结果进行了对比。如图 6 所示, 图 6a 为 YOLOv5 的检测结果, 图 6b 为本文算法的检测结果。可以看出, 在目标较稀疏的前两幅图像中, 两种算法的检测效果不分伯仲, 但在目标实例存在遮挡的环境中, 本文算法的检测效果更优。在第三幅图像中, 目标实例非常密集, YOLOv5 在图片右侧多处均出现漏检的情况(红色圆圈为标出的漏检区域), 而本文算法依旧发挥出色, 进一步证明了本文算法在密集场景下检测小目标的可用性。

4 结论

本文针对无人机航拍图像中小目标识别不准确的问题, 在 YOLOv5 模型的基础上进行了三点

改进。首先使用带有 Transformer 结构的 C3TR 模块替换骨干网络中的 C3 模块, 利用 Transformer 的多头自注意力机制, 捕捉了更广泛的全局信息和上下文信息, 更有利于处理小目标特征, CA 模块的加入也增强了骨干网络中对小目标所在位置的关注。之后, 在颈部网络使用 ConvNeXt-Block 模块替换原有的 C3 模块, ConvNeXt-Block 模块的大卷积可以在特征提取阶段拥有更大的感受野, 有效提升了对小目标的识别能力。最后使用 ECIoU 损失函数替换 CIoU 损失函数, 在加快模型收敛速度的同时也提升了精度。在 Vis-Drone2019 数据集上的对比实验也证明本文算法在小目标识别中具有优势。

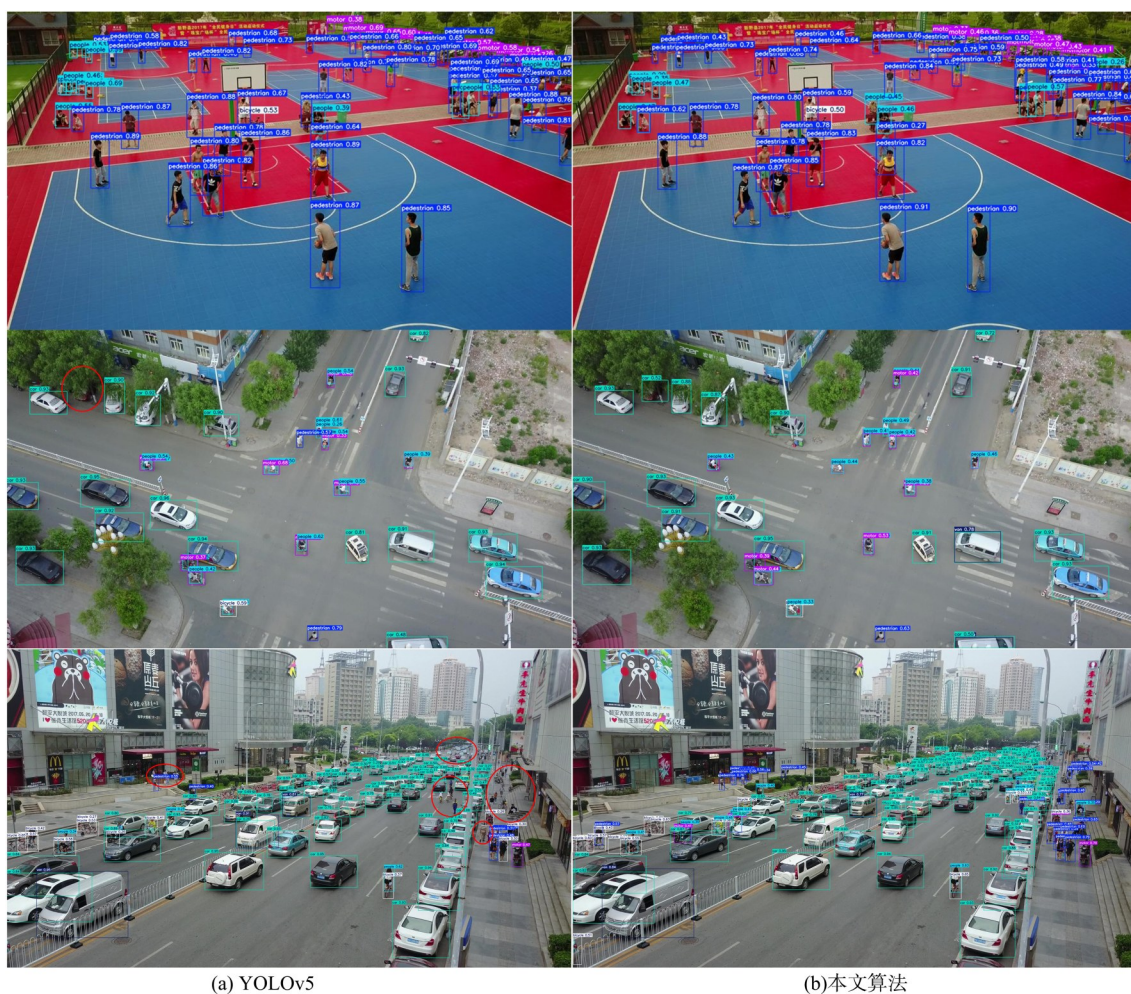


图6 检测效果的对比:(a) YOLOv5;(b)本文算法

Fig. 6 Detection performance:(a) YOLOv5,(b) our method

参考文献

- [1] Muchiri G N, Kimathi S. A review of applications and potential applications of UAV//Proceedings of the Sustainable Research and Innovation Conference. Pretoria, South Africa: Academic Press, 2022: 280—283.
- [2] Mohsan S A H, Othman N Q H, Li Y L, et al. Unmanned aerial vehicles (UAVs): Practical aspects, applications, open challenges, security issues, and future trends. Intelligent Service Robotics, 2023, 16 (1): 109—137.
- [3] Cheng N, Wu S, Wang X C, et al. AI for UAV - assisted IoT applications: A comprehensive review. IEEE Internet of Things Journal, 2023, 10(16): 14438—14461.
- [4] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 779—788.
- [5] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 6517—6525.
- [6] Redmon J, Farhadi A. YOLOv3: An incremental improvement. 2018, arXiv:1804.02767.
- [7] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection. 2020, arXiv:2004.10934.
- [8] Ge Z, Liu S T, Wang F, et al. YOLOX: Exceeding YOLO series in 2021. 2021, arXiv:1606.08415.

- [9] Li C Y, Li L L, Jiang H L, et al. YOLOv6: A single-stage object detection framework for industrial applications. 2022, arXiv:2209.02976.
- [10] 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//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE, 2023: 7464—7475.
- [11] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection//2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2999—3007.
- [12] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014: 580—587.
- [13] Girshick R. Fast R-CNN//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 1440—1448.
- [14] Wu T Y, Dong Y K. YOLO-SE: Improved YOLOv8 for remote sensing object detection and recognition. Applied Sciences, 2023, 13(24): 12977.
- [15] 李姝, 李思远, 刘国庆. 基于YOLOv8无人机航拍图像的小目标检测算法研究. 小型微型计算机系统, 2024, 45(9): 2165—2174. (Li S, Li S Y, Liu G Q. Research on small target detection algorithm based on YOLOv8 UAV aerial images. Journal of Chinese Computer Systems, 2024, 45(9): 2165—2174.)
- [16] Xu X J, Wu X C. Target recognition algorithm for UAV aerial images based on improved YOLO-X//2023 IEEE 5th International Conference on Civil Aviation Safety and Information Technology. Dali, China: IEEE, 2023: 83—87.
- [17] 潘玮, 韦超, 钱春雨, 等. 面向无人机视角下小目标检测的YOLOv8s改进模型. 计算机工程与应用, 2024, 60(9): 142—150. (Pan W, Wei C, Qian C Y, et al. Improved YOLOv8s model for small object detection from perspective of drones. Computer Engineering and Applications, 2024, 60(9): 142—150.)
- [18] 蒋伟, 王万虎, 杨俊杰. AEM-YOLOv8s: 无人机航拍图像的小目标检测. 计算机工程与应用, 2024, 60(17): 191—202. (Jiang W, Wang W H, Yang J J. AEM-YOLOv8s: Small target detection algorithm for UAV aerial images. Computer Engineering and Applications, 2024, 60(17): 191—202.)
- [19] Zhang Z X. Drone-YOLO: An efficient neural network method for target detection in drone images. Drones, 2023, 7(8): 526.
- [20] Fan J H, Cui L Z, Fei S M. Waste detection system based on data augmentation and YOLO_EC. Sensors, 2023, 23(7): 3646.
- [21] Liu Z G, Gao Y, Du Q Q, et al. YOLO-extract: Improved YOLOv5 for aircraft object detection in remote sensing images. IEEE Access, 2023, 11: 1742—1751.
- [22] Zhang J R, Chen Z H, Yan G X, et al. Faster and lightweight: An improved YOLOv5 object detector for remote sensing images. Remote Sensing, 2023, 15(20): 4974.
- [23] Chen X L, Fang H, Lin T Y, et al. Microsoft COCO captions: Data collection and evaluation server. 2015, arXiv:1504.00325.
- [24] 李红光, 于若男, 丁文锐. 基于深度学习的小目标检测研究进展. 航空学报, 2021, 42(7): 024691. (Li H G, Yu R N, Ding W R. Research development of small object tracking based on deep learning. Acta Aeronautica et Astronautica Sinica, 2021, 42(7): 024691.)
- [25] 曹超, 顾幸生. 基于改进YOLO v8的行李追踪技术. 计算机工程与应用, 2024, 60(9): 151—158. (Cao C, Gu X S. Baggage Tracking Technology Based on Improved YOLO v8. Computer Engineering and Applications, 2024, 60(9): 151—158.)
- [26] Du D W, Zhu P F, Wen L Y, et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop. Seoul, Korea (South): IEEE, 2019: 213—226.
- [27] Zhu X K, Lyu S C, Wang X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Montreal, Canada: IEEE, 2021: 2778—2788.

(责任编辑 杨可盛)