



计算机科学

COMPUTER SCIENCE

基于特征增强与上下文融合的无人机小目标检测算法

陈崇杨, 彭力, 杨杰龙

引用本文

陈崇杨, 彭力, 杨杰龙. 基于特征增强与上下文融合的无人机小目标检测算法[J]. 计算机科学, 2025, 52(11): 131-140.

CHEN Chongyang, PENG Li, YANG Jielong. UAV Small Object Detection Algorithm Based on Feature Enhancement and Context Fusion [J]. Computer Science, 2025, 52(11): 131-140.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于声纹特征的无人机个体识别技术研究](#)

Research on Individual Unmanned Aerial Vehicles Identification Technology Based on Voiceprint Characteristics

计算机科学, 2025, 52(11): 444-451. <https://doi.org/10.11896/jsjcx.250300079>

[基于多粒度统计特征的僵尸网络流量智能检测方法](#)

Intelligent Botnet Traffic Detection Method Based on Multi-granularity Statistical Features

计算机科学, 2025, 52(11): 373-381. <https://doi.org/10.11896/jsjcx.241100019>

[SCDDA:基于SCA和Dinkelbach的空-天-地一体化网络无人机轨迹与计算卸载优化方法](#)

SCDDA:SCA and Dinkelbach-based Approach for UAV Trajectory and Computation Offloading in Space-Air-Ground Integrated Networks

计算机科学, 2025, 52(11): 270-279. <https://doi.org/10.11896/jsjcx.241100163>

[基于变体注意力的关系与属性感知实体对齐](#)

Relationship and Attribute Aware Entity Alignment Based on Variant-attention

计算机科学, 2025, 52(11): 230-236. <https://doi.org/10.11896/jsjcx.240800140>

[基于联合注意力机制与多阶段特征提取的图像去雨](#)

Image Deraining Based on Union Attention Mechanism and Multi-stage Feature Extraction

计算机科学, 2025, 52(11): 206-212. <https://doi.org/10.11896/jsjcx.240900013>

基于特征增强与上下文融合的无人机小目标检测算法

陈崇杨 彭力 杨杰龙

江南大学物联网工程学院物联网技术应用教育部工程研究中心 江苏 无锡 214122

(chenchongyang22@163.com)

摘要 针对无人机航拍视角下目标尺寸小、特征信息不足、分布密集以及因遮挡导致的检测精度低的问题,提出一种基于特征增强与上下文融合的无人机小目标检测算法。首先,构建增强特征提取的轻量化主干网络,利用特征提取轻量块高效提取特征信息,并设计细粒度通道融合块有效地避免目标细粒度特征的丢失,该主干网络提高了模型的特征提取能力和推理速度;其次,构建小目标检测头,充分提取小目标的位置信息和细节特征;然后,利用自适应选择空间注意力模块,自适应地调整不同目标所需的感受野,以充分利用航拍小目标周围丰富的上下文信息;最后,引入基于最小点距离的边界框回归损失函数 MPDIoU,进一步提高密集小目标检测的精度。所提算法在 VisDrone2019 数据集上的 $mAP_{0.5}$ 和 $mAP_{0.5,0.95}$ 达到了 46.7% 和 28.6%,较基准网络 YOLOv8s 分别提高了 8.5% 和 5.9%;同时算法的参数数量较 YOLOv8s 减少了 23.4%,可高效适用于无人机航拍密集小目标检测场景。

关键词: 无人机;小目标检测;轻量级网络;上下文信息;注意力机制

中图分类号 TP391.4

UAV Small Object Detection Algorithm Based on Feature Enhancement and Context Fusion

CHEN Chongyang, PENG Li and YANG Jielong

Ministry of Education Engineering Research Center for the Application of Internet of Things Technology, School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China

Abstract Aiming at the problems of low detection accuracy caused by small object sizes, insufficient feature information, dense distribution, and occlusion in UAV aerial photography, this paper proposes a UAV small object detection algorithm based on feature enhancement and context fusion. Firstly, a lightweight backbone network for enhanced feature extraction is constructed, utilizing lightweight feature extraction blocks to efficiently extract feature information, and a fine-grained channel fusion block is designed to effectively prevent the loss of fine-grained features. The backbone network improves the feature extraction capability and inference speed of the model. Secondly, a small object detection head is constructed to fully extract the position information and detailed features of small objects. Then, the adaptive spatial attention module is used to adaptively adjust the receptive fields required for different objects, making full use of the rich context information around the aerial small objects. Finally, a minimum point distance-based bounding box regression loss function (MPDIoU) is introduced to further improve the precision of dense small object detection. The proposed algorithm achieves $mAP_{0.5}$ and $mAP_{0.5,0.95}$ of 46.7% and 28.6% on the VisDrone2019 dataset, respectively, representing an improvement of 8.5% and 5.9% over the baseline network YOLOv8s. Moreover, the algorithm reduces parameters by 23.4% compared to YOLOv8s, making it efficient for dense small object detection in UAV aerial photography scenarios.

Keywords Unmanned Aerial Vehicle, Small object detection, Lightweight networks, Context information, Attention mechanism

1 引言

随着无人机(Unmanned Aerial Vehicle, UAV)技术的不断发展,无人机航拍已广泛应用于城市管理、环境监测、航空巡检等领域。无人机航拍^[1]的特殊性质,如目标尺寸小、特征

信息不足、目标密集且遮挡等复杂情况,以及无人机硬件设备对模型规模的限制,使得小目标检测任务更加具有挑战性和需要满足实用性的需求。因此,如何有效应对无人机航拍小目标检测的挑战,具有重要的研究意义和实用价值。

在小目标检测领域,基于深度学习的目标检测算法^[2]逐

到稿日期:2024-10-08 返修日期:2024-12-17

基金项目:国家自然科学基金(62106082,61873112);第九届中国科协青托工程项目(2023QNRC001)

This work was supported by the National Natural Science Foundation of China(62106082,61873112) and 9th China Association for Science and Technology Youth Support Project(2023QNRC001).

通信作者:彭力(penglimail2002@163.com)

渐成为主导趋势。在目标检测的发展历程中,一阶段和二阶段目标检测算法分别代表了两种不同的思路。一阶段检测算法通过卷积神经网络直接回归目标的位置和类别,如 SSD^[3], RetinaNet^[4] 和 YOLO^[5-12] 系列等,都具有较好的实时性,适合无人机航拍等实时性要求较高的场景,但它们的检测精度相对较低,特别是对于无人机视角下的小目标检测需要进一步优化。而二阶段检测算法则将目标检测任务划分为两个阶段,首先提取候选区域,然后对这些区域进行分类和定位,如 R-CNN^[13], Faster R-CNN^[14] 等,这些模型具有较高的检测精度,但推理速度相对较慢,无法满足无人机视角下实时性检测的需求。一阶段检测算法中的 YOLO 系列由于具有较好的实时性和较小的模型规模,在处理无人机航拍小目标检测任务上具有一定的优势和适用性。Zhu 等^[15] 提出了 TPH-YOLOv5 小目标检测网络,其先增加了一个更高分辨率的检测头,然后将原始检测头替换为所提出的 Transform Prediction Head,结合自注意力机制,提升了无人机航拍小目标检测的精度,但其推理计算速度较慢。Feng 等^[16] 提出了注意力特征融合模块来改进 YOLOv5 网络的特征融合方式,提升了特征融合的效果,但其多次使用跨步卷积层进行下采样,导致无人机航拍目标细粒度特征丢失。Yang 等^[17] 提出的 QueryDet 网络利用级联稀疏查询机制,先在低分辨率特征图上预测小目标的大概位置,随后利用这些位置引导高分辨率特征图的检测,避免了对无人机航拍图像背景区域的有效计算。Xie 等^[18] 提出了无人机影像小目标检测算法 Drone-YOLO,改进了多层次特征信息有效融合的特征金字塔结构,但未能充分考虑小目标周围的上下文信息,使得小目标检测的精度提升不大。Zhang 等^[19] 提出一种全局-局部特征引导模块,以有效结合局部信息和全局信息,使模型聚焦于提取目标特征,减少无人机航拍复杂背景的干扰。Dou 等^[20] 引入多头自注意力机制并构建多尺度特征融合网络,提高了模型对图像特征的利用率,但其特征提取网络多次使用最大池化层下采样操作,造成了目标细粒度特征信息的丢失。Zhang 等^[21] 提出了一种遥感图像小目标检测算法 SuperYOLO,通过融合多模态数据和引入超分辨率学习,对多尺度目标进行高分辨率检测,但其推理检测速度相对较慢。

综上,现有小目标检测算法虽然在一定程度上提高了无人机航拍小目标的检测精度,但仍存在一些关键问题:首先,较大的模型虽然提高了检测精度,但模型参数量、计算量和内存访问量过大,导致算法的推理速度较慢,不适合实时应用;其次,航拍目标尺寸小,在网络中多次使用跨步卷积层或池化层进行下采样,易造成目标细粒度特征信息丢失,在低分辨率图像或密集小目标场景下尤为严重;此外,小目标本身特征信息不足,与航拍复杂背景难以区分,现有小目标检测网络的感受野尺度有限且固定,未能充分利用不同类型小目标周围丰富的上下文信息,易导致航拍的误检和漏检。

针对以上关键问题,为了提高无人机小目标检测算法的推理速度,并有效解决航拍目标尺寸小、特征信息不足、目标密集遮挡导致的检测精度低的问题,本文提出了一种基于特征增强与上下文融合的无人机小目标检测算法 FC-YOLO。

本文的主要工作和创新点总结如下:

1) 构建了增强特征提取的轻量化主干网络,利用提出的特征提取轻量块高效地提取特征信息,并构建新的下采样层以有效保留小目标特征。该主干网络降低了参数量、计算量与内存访问量,增强了网络的特征提取能力,提高了模型的检测精度和推理速度。

2) 设计了一种细粒度通道融合块,有效避免了跨步卷积层造成的细粒度特征丢失,并构建了小目标检测头充分提取小目标的位置信息和细节特征,提升了模型对密集小目标的检测性能。

3) 构建了一种自适应选择空间注意力模块,动态调整不同目标所需的合理感受野,以充分利用航拍小目标周围丰富的上下文信息,提高了网络在复杂场景中的检测能力。

4) 引入了基于最小点距离的边界框回归损失函数 MP-DIoU^[22],提高了航拍密集小目标检测中边界框回归的速度和精度。

2 相关工作

2.1 无人机航拍小目标检测

无人机航拍图像具有目标尺寸小、分布密集和复杂背景等特点,并且无人机硬件平台对模型大小有较高的限制,因此,兼顾无人机航拍小目标检测的精度和速度尤为重要。Zeng 等^[23] 提出了一种混合注意力模块来增强航拍小目标的特征提取,并设计多层特征融合结构丰富了浅层特征的语义信息,但改进后的模型计算量较大,检测的实时性较低。Zhu 等^[24] 提出了基于双层路由的动态稀疏注意力模块,通过粗略区域级过滤和细粒度令牌-令牌注意力的路由区域策略,为无人机小目标检测任务提供了合理的计算分配思路。Jiang 等^[25] 设计了一种多尺度特征提取模块,并构建双向密集特征金字塔,提升了航拍小目标检测的精度,但检测速度相对较慢,难以满足无人机航拍实时检测的需求。Zhu 等^[26] 构建高效多尺度特征提取网络,以更好地提取不同尺度航拍目标的特征,但其在网络中多次使用跨步卷积进行下采样操作,导致航拍目标细粒度特征信息丢失。尽管这些方法在一定程度上提升了无人机航拍小目标检测的精度,但仍面临推理检测速度慢、目标细粒度特征丢失的问题。针对上述问题,本文提出一种增强特征提取的轻量化主干网络,在提高检测精度的同时加快推理速度,使模型能够在资源受限的无人机平台上高效运行,提升了无人机航拍小目标检测的性能。

2.2 基于注意力机制的小目标检测

在小目标检测领域,注意力机制是一种提升模型检测性能的重要方法。通过设计注意力机制,能使模型更好地关注图像中的关键区域和重要特征,尤其是在复杂场景和无人机航拍小目标检测任务中,能够有效提升检测精度和鲁棒性。Ouyang 等^[27] 提出致力于保留每个通道信息的多尺度注意力模块,通过对部分通道进行重塑和跨维度交互,提高了无人机图像特征表示的可辨别性。Ma 等^[28] 提出一种多尺度融合注意力机制,强化对小目标关键特征的捕获,并抑制非关键信息,增强了网络的特征学习能力。Zhou 等^[29] 设计了一种轻量级自相关模块,用于增强特征表示,并在颈部网络中扩展为

轻量级互相关模块,以捕获相邻尺度特征之间的依赖关系。然而,这些方法未能充分考虑不同类型小目标周围丰富的上下文信息。为了克服现有注意力机制在感受野调整方面的不足,本文提出一种自适应选择空间注意力模块,该模块能够根据不同类型目标的需求,自适应地调整感受野的合理尺度,从而充分利用航拍小目标周围的上下文信息,有效提高模型在复杂背景下对小目标的检测能力。

3 FC-YOLO 无人机小目标检测算法

本文提出的 FC-YOLO 无人机航拍小目标检测算法是在 YOLOv8s^[12] 基准网络上进行改进的, YOLOv8s 网络包括 3 个组成部分:主干网络使用 CSPDarkNet-53 结构,颈部网络使用 PANet 结构,头部网络使用 Decoupled Head 解耦检测头结构。YOLOv8s 可以输出 3 种不同尺度的预测特征图,这种多尺度的设计实现了较好的检测性能。FC-YOLO 小目标检测算法的整体结构如图 1 所示。首先,构建增强特征提取的轻量化主干网络,利用特征提取轻量块(Feature Extraction Lightweight Block, C2f_FELB)高效提取特征信息,并设计一

种细粒度通道融合块(Fine-grained Channel Fusion Block, FCFB)作为该主干网络的下采样模块,有效避免目标细粒度特征信息丢失,改进后的主干网络有效提高了模型的特征提取能力和推理计算效率,能更好地满足实时性准确检测的应用需求;其次,构建小目标检测头(Small Object Detection Head, SODH),将改进主干网络中的浅层特征图 P2 融入颈部网络中的特征融合过程,充分提取航拍小目标的位置信息和细节特征,弥补特征信息的丢失,有效提高了密集小目标检测的精度;然后,利用所设计的自适应选择空间注意力模块(Adaptive Selective Spatial Attention Module, ASSA),动态调整空间感受野的尺度,有效融合不同类型小目标周围丰富的上下文信息,在复杂背景下减少目标的误检和漏检,增强了模型的鲁棒性和适应性;最后,将原始网络采用的边界框回归损失函数 CIoU 替换为 MPDIoU (Minimum Point Distance Based IoU),充分考虑预测框和真实框之间的重叠或非重叠区域、中心点距离、宽和高的实际偏差,提高无人机航拍小目标检测中边界框回归的效率和准确性,从而进一步提升了模型对密集小目标的检测精度。

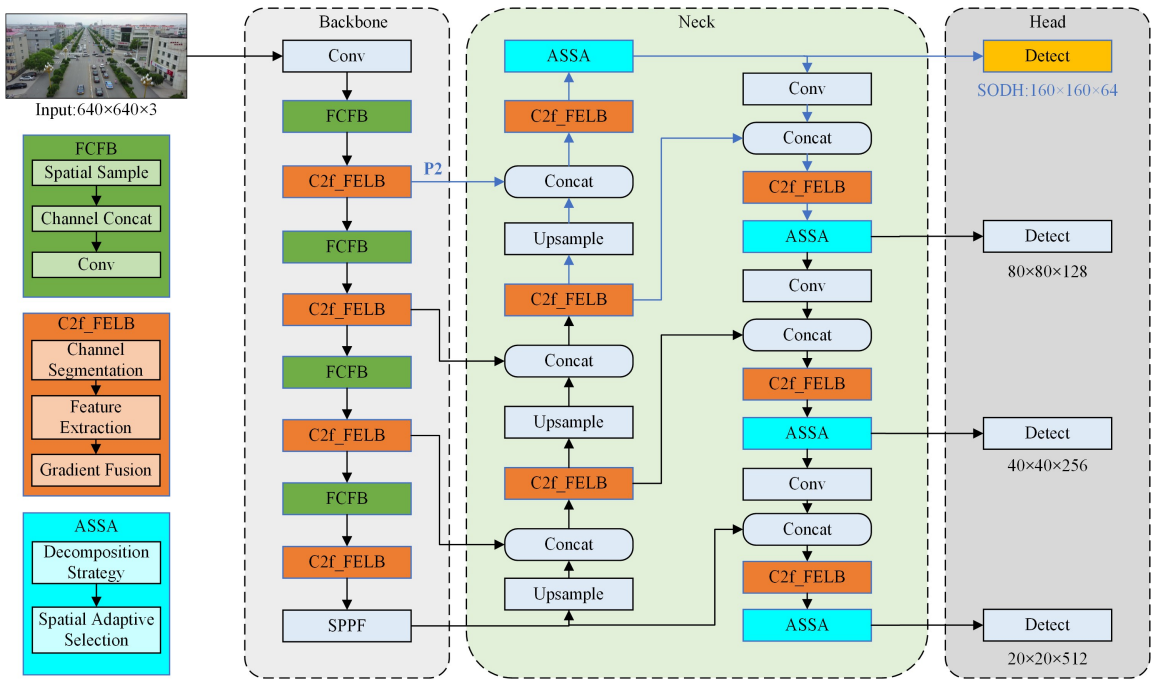


图 1 FC-YOLO 模型整体结构
Fig. 1 Overall architecture of FC-YOLO model

3.1 增强特征提取的轻量化主干网络

为了降低模型的冗余计算量、参数量与内存访问量,提高网络的特征提取能力,构建高性能、轻量化的无人机航拍小目标检测网络,本文提出了一种增强特征提取的轻量化主干网络,利用 C2f_FELB 快速提取关键特征信息,并构建 FCFB 有效避免目标细粒度特征丢失,使其高效适用于无人机航拍小目标检测任务。

具体来说,通过引入部分卷积^[30](Partial Convolution, PConv),充分利用特征图不同通道之间的特征高度相似性,仅在特征图的第一个或最后一个连续的四分之一输入通道上,利用常规卷积提取特征,其余输入通道则保持不变,其结构如图 2 所示。在 PConv 之后引入逐点卷积 PWConv,充分

挖掘特征图所有输入通道的信息,促进通道之间的信息交互,进一步增强特征的表达能力。

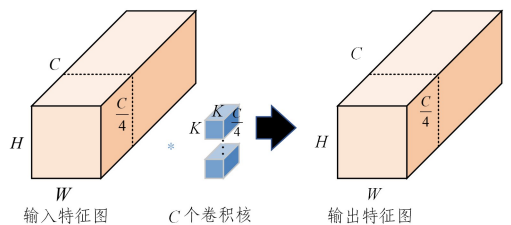


图 2 部分卷积
Fig. 2 Partial convolution

基于 PConv 和 PWConv 提出一种特征提取瓶颈(Feature

Extraction Bottleneck, FEB), 其具体结构如图 3 所示。FEB 由两个 PConv 和两个 PWConv 结合成一个反向残差块, 该模块的中间层具有扩展数量的通道, 有助于梯度的有效流动, 同时通过快捷连接实现输入特征的重用。在 FEB 中的第二个 PConv 后插入归一化层 (Batch Norm, BN) 和 SiLU 激活函数, 以充分保留特征的多样性并加快推理速度。FEB 通过将 PConv 和 PWConv 相结合, 在提取足够特征信息的同时减少了参数量和计算量, 使网络高效提取关键特征, 提高了推理计算效率。

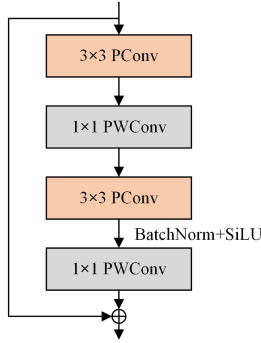


图 3 特征提取瓶颈

Fig. 3 Feature extraction bottleneck

利用 FEB 替换整体网络中 C2f 模块的 Bottleneck 结构, 构建 C2f_FELB, 该模块的具体结构如图 4 所示。

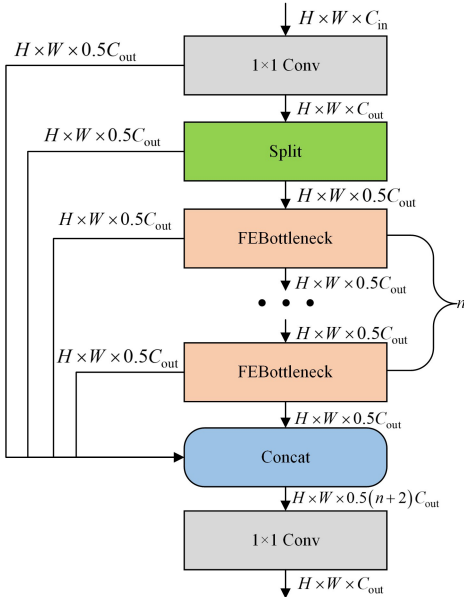


图 4 特征提取轻量化块

Fig. 4 Feature extraction lightweight block

首先, 通过 1×1 的卷积调整特征图的通道数, 增强特征的表达力。然后, 使用 Split 操作将特征图划分成两部分: 一部分用于残差连接, 以保留输入重要特征, 避免特征信息丢失; 另一部分进入主分支, 进行更深层次的特征提取, 以增强模型的表示和非线性能力。主分支通过 n 个串联的 FEB 进行特征提取, 并保持残差连接, 在减少参数量和计算量的同时获得更丰富的梯度流。随后, 将多个并行的梯度流分支通过 Concat 操作进行拼接, 从而整合更丰富的梯度信息, 进一步强化特征表达能力。最后, 通过 1×1 的卷积对拼接后的特征

图进行处理, 得到 C2f_FELB 的最终输出。与原始网络中的 C2f 模块相比, 所构建的 C2f_FELB 在保证检测精度的同时, 有效减少了参数量、计算量和内存访问量, 提高了推理计算速度, 有利于构建轻量化、高性能的无人机航拍小目标检测网络。

无人机航拍目标尺寸小, 原始主干网络使用跨步卷积层进行下采样操作, 容易造成航拍目标细粒度特征信息丢失, 进而降低无人机航拍小目标检测的性能。这一问题在输入图像分辨率较低和密集小目标检测任务中尤为明显。针对该问题, 受 YOLOv5 网络中 Focus 模块结构的启发, 设计 FCFB 作为改进主干网络中的每个下采样层, 以有效提高无人机航拍小目标的检测精度, 该模块的结构如图 5 所示。

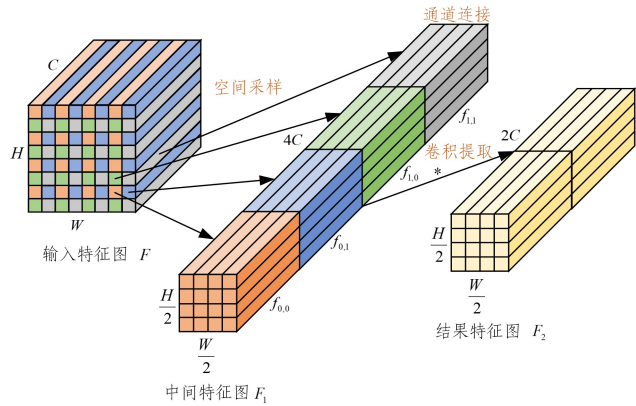


图 5 细粒度通道融合块

Fig. 5 Fine-grained channel fusion block

FCFB 首先对输入特征图 F 进行空间采样和通道连接, 实现对目标细粒度特征信息的保留和融合, 避免细节特征信息丢失。具体而言, 对于给定大小为 $H \times W \times C$ 的输入特征图 F , 按比例因子 2 对 F 进行交错下采样, 得到 4 个子特征图 $f_{0,0}, f_{0,1}, f_{1,0}, f_{1,1}$, 每个子特征图的大小为 $(\frac{H}{2}, \frac{W}{2}, C)$, 表示如下:

$$\begin{aligned} f_{0,0} &= F[0:H:2, 0:W:2] \\ f_{0,1} &= F[0:H:2, 1:W:2] \\ f_{1,0} &= F[1:H:2, 0:W:2] \\ f_{1,1} &= F[1:H:2, 1:W:2] \end{aligned} \quad (1)$$

交错下采样方法旨在保留空间维度中不同位置的细节信息, 确保在特征图的每个区域都能捕捉到目标的关键特征。然后, 将得到的 4 个子特征图沿着通道维度进行拼接, 得到中间特征图 F_1 , F_1 的尺寸为 $(\frac{H}{2}, \frac{W}{2}, 4C)$ 。最后, 利用 $2C$ 个卷积核大小为 3×3 的非跨步卷积层进行特征提取, 把 F_1 转换成大小为 $(\frac{H}{2}, \frac{W}{2}, 2C)$ 的结果特征图 F_2 。使用非跨步卷积层的目的是为了有效保留特征图中所有关键特征信息, 避免跨步卷积导致的特征信息丢失。FCFB 有效解决了航拍目标细粒度特征丢失的问题, 特别适用于检测低分辨率图像和无人机航拍密集小目标场景。

增强特征提取的轻量化主干网络包括 4 个阶段, 每个阶段的 FCFB 作为下采样层, 用于空间下采样和扩展特征图的

通道数,每个阶段 C2f_FELB 的作用是高效提取特征信息。该主干网络充分利用 C2f_FELB 和 FCFB 的结合,实现模型结构的轻量化,同时提高网络的特征提取能力,较好地平衡了精度和速度,使模型能够在资源受限的无人机硬件设备上高效运行,并提升航拍小目标检测的性能。

3.2 构建小目标检测头

航拍小目标在图像中像素占比较小,且特征表示有限,原始网络难以充分学习其位置信息和细节特征。为解决该问题,与 Zhang 等^[19]提出的方法类似,在原始网络的基础上构建专门用于小目标的检测头 SODH,即将改进主干网络中的浅层特征图 P2 融入颈部网络中的特征融合过程,通过浅层特征融合,生成空间分辨率为 160×160 的预测特征图。通过构建 SODH,使模型更好地提取小目标的位置和细节信息,有效降低了密集小目标漏检、误检的概率。

3.3 自适应选择空间注意力模块

航拍小目标的特征信息少,仅根据其本身特征难以实现准确的检测,而目标附近的上下文信息可以提供丰富的背景特征,这些特征对于目标的定位和识别至关重要,所以充分融合小目标周围的上下文信息是提升检测性能的关键。同时,不同种类的小目标需要不同范围的上下文信息,因此需要利用合理尺度的感受野实现准确的检测^[31]。原始网络的感受野有限且固定,未能充分考虑目标周围广泛的上下文信息,易导致无人机航拍小目标的漏检和误检。

上下文信息不仅包括航拍目标的环境背景,还反映了与周围物体的相对位置和关系。为了充分利用目标周围丰富的上下文信息,提高检测性能,设计了一种自适应选择空间注意力模块 ASSA。该模块能够动态调整感受野的不同尺度,更有效地融合小目标周围的上下文信息,从而增强网络对目标的理解,使模型更好地区分目标特征和复杂背景,有效降低了无人机航拍目标漏检和误检的概率。ASSA 由分解的两个具有不同尺度感受野的大型内核卷积和空间自适应选择机制组成,具体结构如图 6 所示。

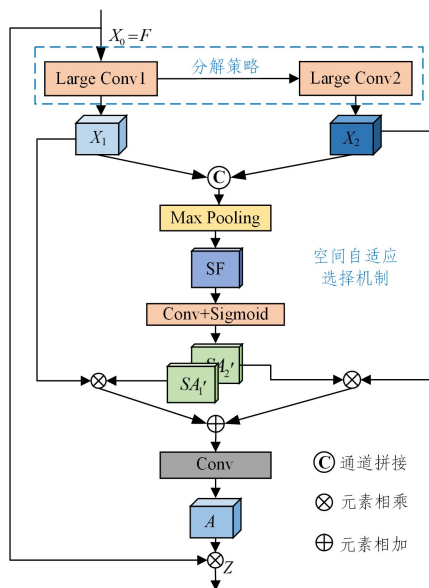


图 6 自适应选择空间注意力模块

Fig. 6 Adaptive selective spatial attention module

首先,利用显式分解策略,将所构造的大型内核卷积分解为串联的两个具有不同尺度感受野的常规卷积,进而生成具有合理感受野的两个特征图。分解后的两个常规卷积的卷积核大小 k 、扩张率 d 以及感受野 RF 的定义式为:

$$k_1 \leq k_2, d_1 = 1, d_1 < d_2 \leq RF_1 \quad (2)$$

$$RF_1 = k_1, RF_2 = d_2(k_2 - 1) + RF_1 \quad (3)$$

式(2)保证了 RF 可以通过 k 和 d 的增大,实现快速扩展。同时,为了避免扩张卷积在特征图之间引入空隙,设定扩张率的上界为分解后第一个卷积的感受野大小,这确保了在不失真特征的前提下,能够有效捕获更广泛的上下文信息。接着,使用分解后的两个具有不同尺度感受野的常规卷积,以充分提取特征信息。

$$X_0 = F, X_1 = f_1(X_0), X_2 = f_2(X_1) \quad (4)$$

其中, f_1, f_2 表示第一、第二个卷积, X_1, X_2 为对应的卷积得到的特征图。然后,利用空间自适应选择机制实现对内核的动态选择,以更好地适应不同目标的尺度和满足所需上下文信息融合的需求,增强模型对不同类型小目标的感知能力。具体来说,先将 X_1, X_2 在通道维度上进行拼接,形成特征集合 X :

$$X = [X_1; X_2] \quad (5)$$

接着对 X 利用基于通道维度的最大池化操作,从多个通道中抽取最显著的特征,以充分提取空间特征关系,增强上下文信息的表达能力,得到最大池化空间特征图 SF :

$$SF = P_{\max}(X) \quad (6)$$

其中, $P_{\max}(\cdot)$ 表示基于通道的最大池化。然后使用卷积层进行非线性变换和特征整合,通过非线性映射来强化上下文特征的表达能力,将池化特征转换为两个空间注意力图:

$$SA = f(SF) \quad (7)$$

对于得到的两个空间注意力图,使用 Sigmoid 激活函数生成空间选择掩码 SA_1' 和 SA_2' ,用于指示在特定上下文信息中更为重要的特征,从而动态调整特征图的权重:

$$SA_1' = \sigma(SA_1), SA_2' = \sigma(SA_2) \quad (8)$$

其中, $\sigma(\cdot)$ 为 Sigmoid 激活函数。将 X_1 和 X_2 与其相应的 SA_1' 和 SA_2' 进行加权操作,然后通过卷积层 $f(\cdot)$ 将加权后的特征充分融合,得到注意特征图 A :

$$A = f(SA_1' \cdot X_1 + SA_2' \cdot X_2) \quad (9)$$

这一加权融合过程使得模型能够有效地融合不同尺度的上下文信息,从而增强对小目标的识别能力。最后,通过 F 与 A 之间的逐元素乘积计算 ASSA 的输出结果 Z :

$$Z = F \cdot A \quad (10)$$

ASSA 可以自适应地调整不同目标所需的感受野,以有效融合小目标周围的上下文信息,使模型更好地区分航拍小目标与复杂背景,提高网络在多种复杂场景中的检测能力和鲁棒性,从而进一步提高无人机航拍小目标检测的性能。

3.4 优化损失函数

基准网络 YOLOv8s 采用 CIoU 作为边界框回归损失函数,CIoU 在 DIoU 的基础上额外考虑了预测框和真实框的宽高比,但该惩罚项描述的是预测框与真实框的宽和高的相对

值,不能反映二者之间的实际误差,并且在预测框和真实框具有相同的宽高比但宽度和高度值完全不同时,CIoU将失效,导致无人机小目标检测的收敛速度和检测精度下降。为了解决该问题,引入基于最小点距离的边界框回归损失函数 MPDIoU,其充分考虑了预测框和真实框之间的重叠或非重叠区域、中心点距离、宽和高的实际偏差,通过最小化预测框和真实框之间的左上角点和右下角点的距离,有效提高了航拍密集小目标检测中边界框回归的收敛速度和准确性,从而提升了无人机航拍密集小目标的检测精度。

假设输入图片的高度和宽度分别为 H 和 W ,真实边界框和预测边界框用左上点和右下点的坐标分别表示为 $B^{gt} = (x_1^{gt}, y_1^{gt}, x_2^{gt}, y_2^{gt})$, $B = (x_1, y_1, x_2, y_2)$, L_{MPDIoU} 具体表示如下:

$$\rho_1 = \sqrt{(x_1 - x_1^{gt})^2 + (y_1 - y_1^{gt})^2} \quad (11)$$

$$\rho_2 = \sqrt{(x_2 - x_2^{gt})^2 + (y_2 - y_2^{gt})^2}$$

$$MPDIoU = \frac{B^{gt} \cap B}{B^{gt} \cup B} - \frac{\rho_1^2}{W^2 + H^2} - \frac{\rho_2^2}{W^2 + H^2} \quad (12)$$

$$L_{MPDIoU} = 1 - MPDIoU \\ = 1 - \frac{B^{gt} \cap B}{B^{gt} \cup B} + \frac{\rho_1^2}{W^2 + H^2} + \frac{\rho_2^2}{W^2 + H^2} \quad (13)$$

式(11)中, ρ_1 为预测框和真实框的左上角的欧氏距离, ρ_2 为预测框和真实框的右下角的欧氏距离。

4 实验结果与分析

4.1 实验平台和数据集

实验平台选用 Ubuntu16.04 操作系统, CPU 是 Intel^(R) Core i5-8500, GPU 使用 NVIDIA GTX 1080Ti, 11 GB 显存。本文算法主要在 VisDrone2019 数据集上进行实验,并在 PASCAL VOC 数据集上进一步验证其有效性,实验中使用了不同的输入图片分辨率,以充分验证本文算法的检测性能。

VisDrone2019^[32] 数据集是由天津大学机器学习与数据挖掘实验室构建的无人机航拍小目标检测数据集,包含 10 种类别,分别为 pedestrian, people, bicycle, car, van, truck, tricycle, awning-tricycle, bus, motor, 共有 260 万个目标标注,其中小目标占比 60% 左右,最小目标的像素宽高为 3×1 。其拍摄场景复杂多变,且存在大量的密集、遮挡小目标。在实验中使用 6471 张训练集图片训练模型,使用 1610 张测试集图片评估模型的检测性能。

PASCAL VOC^[33] 数据集是计算机视觉领域中用于目标检测的常用数据集,包含 20 种目标类别,涵盖大、中、小 3 种不同尺寸的目标。在实验中使用 VOC2007 和 VOC2012 的 train+val 组合训练模型,共计 16551 张图像;利用 VOC2007 的 test 验证集评估模型的性能,包含 4952 张图像。

4.2 实验细节和评价指标

在模型训练阶段,针对不同的输入图片分辨率,使用不同的参数设置,具体参数的选择经过多次实验验证,旨在优化模型的检测性能和提高训练效率。表 1 详细列出了在 Vis-Drone2019 数据集和 PASCAL VOC 数据集上的实验参数配置。

表 1 实验参数设置

Table 1 Experimental parameter settings		
实验参数	VisDrone2019	PASCAL VOC
输入分辨率	640×640	480×480
批量大小	16	32
初始学习率	0.01	0.02
训练轮数	200	200
warmup 轮数	3	3
优化器	SGD	SGD
动量因子	0.937	0.937
权重衰减系数	0.0005	0.0005
翻转系数	0.5	0.5
Mosaic 增强概率	1.0	1.0

本文使用的评价指标包括精确度(Precision, P)、召回率(Recall, R)、单类别平均精度(Average Precision, AP)、所有类别平均精度均值(Mean Average Precision, mAP)、参数量(Parameters, Params)、计算量(Floating Point Operations, FLOPs)及每秒帧数(Frames Per Second, FPS),以全面衡量本文算法的综合性能。AP 和 mAP 的具体计算方法为:

$$AP = \int_0^1 P(R) dR \quad (14)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i = \frac{1}{N} \sum_{i=1}^N \int_0^1 P(R) dR \quad (15)$$

式(15)中, N 为目标类别数。

4.3 消融实验

为了充分验证本文所提改进点的有效性,在 Vis-Drone2019 测试集上进行消融实验,本节详细分析了各改进点对本文算法综合性能的影响,消融实验结果如表 2 所列。相较于实验 A 基准模型,实验 B 利用 C2f_FELB 作为整体网络的特征提取模块,Params 和 FLOPs 分别从 1.11×10^7 , 2.85×10^{10} 降低至 7.9×10^6 , 1.72×10^{10} , 且 FPS 从 79.3 提升至 108.6,在保证检测精度的前提下实现了轻量化,提高了网络的推理计算速度。基于实验模型 B,实验 C 利用 FCFB 作为改进主干网络中的下采样层,Params 和 FLOPs 略微增加,但 $mAP_{0.5}$ 和 $mAP_{0.5,0.95}$ 分别提升了 3.2 个百分点和 2.2 个百分点。实验 C 的结果表明,构建增强特征提取的轻量化主干网络可以有效提高模型的检测精度和推理速度。实验 D 是在实验 C 的基础上构建 SODH,以充分提取无人机航拍小目标的位置信息和细节特征,有效弥补了小目标特征信息的丢失, $mAP_{0.5}$ 和 $mAP_{0.5,0.95}$ 分别提高了 4.1 个百分点和 2.9 个百分点。实验 E 是在实验 D 模型的基础上加入 ASSA,其 $mAP_{0.5}$ 和 $mAP_{0.5,0.95}$ 指标分别提升了 1.0 个百分点和 0.7 个百分点,证明了该模块可以充分融合航拍小目标周围丰富的上下文信息,有效提高无人机航拍小目标检测的精度。相较于实验 E,实验 F 的结果指标 $mAP_{0.5}$ 和 $mAP_{0.5,0.95}$ 进一步提高至 46.7% 和 28.6%,证明了 MPDIoU 损失函数可以提高无人机航拍密集小目标检测的性能。最终,本文算法 FC-YOLO 与基准网络 YOLOv8s 相比,Params 下降了 23.4%, FLOPs 下降了 11.6%, FPS 提升了 10.3%, 精度指标 $mAP_{0.5}$ 提升了 8.5 个百分点, $mAP_{0.5,0.95}$ 提升了 5.9 个百分点。充分证明了本文所提算法 FC-YOLO 可以有效提升无人机航拍小目标检测的精度,并保证检测的实时性。

表2 消融实验结果

Table 2 Ablation experiment results

Method	C2f_FELB	FCFB	SODH	ASSA	MPDIoU	Params	FLOPs	FPS	$mAP_{0.5}/\%$	$mAP_{0.5:0.95}/\%$
A	—	—	—	—	—	1.11×10^7	2.85×10^{10}	79.3	38.2	22.7
B	✓	—	—	—	—	7.9×10^6	1.72×10^{10}	108.6	38.1	22.6
C	✓	✓	—	—	—	8.7×10^6	2.01×10^{10}	102.8	41.3	24.8
D	✓	✓	✓	—	—	8.1×10^6	2.44×10^{10}	92.4	45.4	27.7
E	✓	✓	✓	✓	—	8.5×10^6	2.52×10^{10}	87.5	46.4	28.4
F	✓	✓	✓	✓	✓	8.5×10^6	2.52×10^{10}	87.5	46.7	28.6

此外,基于实验 D,在 VisDrone2019 测试集上对 ASSA 中的大型内核卷积分解数量、感受野 RF 大小进行实验验证,以研究其在无人机航拍小目标检测任务中的最优结果。表 3 是 RF 的大小为 29 时,大型内核卷积分解数量的实验结果,可以看出,在保持相同 RF 大小的前提下,相较于单个大型内核及分解的 3 个内核序列,将大型内核卷积分解为串联的两个常规卷积可以达到最优的检测结果。

表3 大型内核卷积分解数量的实验结果

Table 3 Experimental results for the decomposition of the number of large kernel convolution

(k, d)	RF	$mAP_{0.5}/\%$	$mAP_{0.5:0.95}/\%$
(29,1)	29	45.8	28.0
(5,1)→(7,4)	29	46.2	28.3
(3,1)→(5,2)→(7,3)	29	46.0	28.2

不同 RF 大小的实验结果如表 4 所列。由实验结果可知,当 RF 的大小约为 23 时,模型达到最优的检测性能, $mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别达到了 46.4% 和 28.4%,比 RF 大

小为 11 和 31 时的检测精度更高。表 4 中的结果表明,在无人机航拍小目标检测任务中,采用适当大小的 RF,并进行大型内核卷积分解,可以取得较大程度的检测性能提升。

表4 合理感受野的实验结果

Table 4 Experimental results for reasonable receptive field

(k_1, d_1)	(k_2, d_2)	RF	$mAP_{0.5}/\%$	$mAP_{0.5:0.95}/\%$
(3,1)	(5,2)	11	45.7	27.9
(5,1)	(7,3)	23	46.4	28.4
(7,1)	(9,3)	31	46.1	28.2

4.4 对比实验及结果分析

4.4.1 VisDrone2019 数据集对比实验

VisDrone2019 数据集包含大量的密集小目标,并且背景较为复杂、存在严重的遮挡问题,因此检测难度较大。对比算法包括近几年提出的 YOLO 系列算法、针对小目标检测的 QueryDet 算法、VisDrone 数据集竞赛的获奖算法 TPH-YOLOv5,以及经典的 SSD,RetinaNet,Faster-RCNN 等算法,表 5 为不同算法在 VisDrone2019 数据集上的对比实验结果。

表5 VisDrone2019 数据集上的对比实验

Table 5 Comparative experiments on VisDrone2019 dataset

Model	Params	FLOPs	P/%	R/%	$mAP_{0.5}/\%$	$mAP_{0.5:0.95}/\%$
SSD ^[3]	2.45×10^7	8.79×10^{10}	21.0	35.5	23.9	10.2
RetinaNet ^[4]	1.98×10^7	9.37×10^{10}	23.5	37.9	26.5	12.4
Faster-RCNN ^[14]	4.12×10^7	2.067×10^{11}	45.3	33.8	33.2	17.0
YOLOv5s(6,0) ^[8]	7.0×10^6	1.58×10^{10}	42.5	34.5	32.8	17.9
YOLOX-s ^[9]	8.9×10^6	2.67×10^{10}	43.2	34.8	33.1	18.2
YOLOv6-s ^[10]	1.84×10^7	4.52×10^{10}	45.3	38.2	34.9	19.1
YOLOv7-tiny-SiLU ^[11]	6.0×10^6	1.31×10^{10}	47.3	38.6	36.3	18.8
TPH-YOLOv5 ^[15]	8.4×10^6	2.25×10^{10}	52.6	40.9	41.2	25.8
QueryDet ^[17]	1.89×10^7	4.43×10^{10}	41.1	33.4	31.6	17.4
Proposed-s ^[16]	8.4×10^6	1.92×10^{10}	—	—	39.4	24.6
Drone-YOLO ^[18]	8.3×10^6	3.94×10^{10}	—	—	39.1	24.0
YOLOv8s ^[12]	1.11×10^7	2.85×10^{10}	50.4	37.1	38.2	22.7
FC-YOLO	8.5×10^6	2.52×10^{10}	57.3	45.7	46.7	28.6

相较于表 5 中其他性能优异的多种主流算法,本文算法 FC-YOLO 在 VisDrone2019 数据集上取得了较高的性能优势,P 和 R 指标分别为 57.3% 和 45.7%,精度指标 $mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别达到了 46.7% 和 28.6%,明显优于基准算法 YOLOv8s 和其他对比算法。在保持优秀检测性能的同时,FC-YOLO 还具备较轻量的模型结构,Params 和 FLOPs 分别为 8.5×10^6 和 2.52×10^{10} ,不仅减少了计算资源的消耗,还提高了无人机航拍小目标检测的推理速度。与针对小目标检测的 QueryDet 算法相比,FC-YOLO 算法的精度指标 $mAP_{0.5}$ 提升了 15.1 个百分点、 $mAP_{0.5:0.95}$ 提升了 11.2 个百分点,且模型结构更轻量化,更适合准确性和实时性要求较高的无人机航拍检测场景。相较于 VisDrone 数据集竞赛的获奖算法

TPH-YOLOv5,本文算法具有更加优秀的检测性能, $mAP_{0.5}$ 提升了 5.5 个百分点、 $mAP_{0.5:0.95}$ 提升了 2.8 个百分点,但模型 FLOPs 略高出 2.7×10^9 。与 YOLOv5s 和 YOLOv7-tiny-SiLU 检测算法相比,FC-YOLO 模型的 Params 和 FLOPs 更高,但在检测精度上有明显提升,其更适用于检测难度较大的无人机航拍小目标检测场景。与一众较新提出的 YOLO 系列算法相比,本文算法 FC-YOLO 在 P, R, $mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 性能指标上的表现更为出色。相较于经典的 SSD,RetinaNet,Faster-RCNN 检测算法,FC-YOLO 的精度指标 $mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 提升更为显著,并且模型的 Params 和 FLOPs 相对更低,可以很好地满足无人机设备实际部署的应用需求。对比实验的结果表明,本文算法 FC-YOLO 具有

较高的检测精度和较轻量化的模型结构,与多种主流检测算法相比,其在无人机航拍小目标检测任务中具有明显的性能优势。

4.4.2 PASCAL VOC 数据集对比实验

为了验证所提算法 FC-YOLO 在其他数据集上的有效性,将其与性能优异的多种轻量级算法在 PASCAL VOC 数据集上进行实验对比,具体实验结果如表 6 所列。

表 6 PASCAL VOC 数据集上的结果对比

Table 6 Comparison of results on PASCAL VOC dataset

Model	Params	FLOPs	P/%	R/%	mAP _{0.5} /%
YOLOv5s(6.0)	7.10×10^6	9.10×10^9	79.4	70.7	78.9
YOLOX-s	8.90×10^6	1.50×10^{10}	80.1	71.9	80.2
YOLOv6-s	1.85×10^7	2.54×10^{10}	81.3	74.3	81.7
YOLOv7-tiny-SiLU	6.10×10^6	7.50×10^9	81.2	70.2	79.0
TPH-YOLOv5	8.40×10^6	1.26×10^{10}	80.6	72.2	80.5
QueryDet	1.89×10^7	2.48×10^{10}	79.3	70.4	78.7
YOLOv8s	1.11×10^7	1.60×10^{10}	80.8	75.2	81.9
FC-YOLO	8.50×10^6	1.41×10^{10}	83.4	76.8	84.3

与表 6 中的其他对比算法相比,FC-YOLO 在 P 和 R 指标上均取得了较高的提升,分别为 83.4% 和 76.8%,表现出更高的目标检测精准率和召回率。FC-YOLO 的精度 mAP_{0.5} 达到了 84.3%,超过了基准网络 YOLOv8s 和其他对比算法,进一步证明了 FC-YOLO 具有较高的性能优势。同时,FC-YOLO 还具备较轻量化的模型结构,Params 和 FLOPs 分别为 8.5×10^6 和 1.41×10^{10} ,这表明 FC-YOLO 适用于实时小目标检测任务。

为了更详细地分析本文算法对小目标的检测性能,从 PASCAL VOC 数据集中选取 5 种具有代表性的小目标类别,

不同算法在这 5 种类别上的检测结果如表 7 所列。

表 7 小目标类别检测结果

Table 7 Detection results of small object categories

Model	mAP _{0.5}	AP				
		bird	bottle	plant	chair	boat
YOLOv5s(6.0)	64.2	73.5	66.4	50.6	61.9	68.7
YOLOX-s	65.1	74.3	67.2	51.5	62.7	69.9
YOLOv6-s	66.7	77.6	66.1	53.8	63.2	72.8
YOLOv7-tiny-SiLU	64.5	73.8	65.4	52.0	60.8	70.2
TPH-YOLOv5	68.6	79.3	66.8	56.7	64.9	75.1
QueryDet	64.1	73.2	66.3	50.5	61.6	69.1
YOLOv8s	67.8	78.3	66.3	55.9	64.7	73.9
FC-YOLO	72.2	82.1	75.6	57.6	67.8	78.1

FC-YOLO 在这 5 种小目标类别上均取得了最高的检测精度,特别是 bird, bottle 和 boat 这 3 个类别,AP 指标分别达到了 82.1%, 75.6% 和 78.1%,相对于其他算法的性能提升更高。FC-YOLO 在这 5 类小目标的总体 mAP_{0.5} 达到了 72.2%,明显优于其他对比算法,进一步表明了本文算法在小目标检测任务上具有较好的检测性能。

4.5 可视化结果分析

为了更加直观地验证本文算法在无人机航拍小目标检测场景下的优秀性能,图 7 展示了 VisDrone2019 数据集上典型场景的可视化结果图,并进行详细的对比分析。其中包含本文算法 FC-YOLO 和多种主流算法在目标遮挡、目标密集、夜间黑暗场景和复杂背景下的可视化检测结果,主流算法包括基准算法 YOLOv8s、针对小目标检测的 QueryDet 算法、VisDrone 数据集竞赛的获奖算法 TPH-YOLOv5,以及轻量级检测算法 YOLOv7-tiny-SiLU。

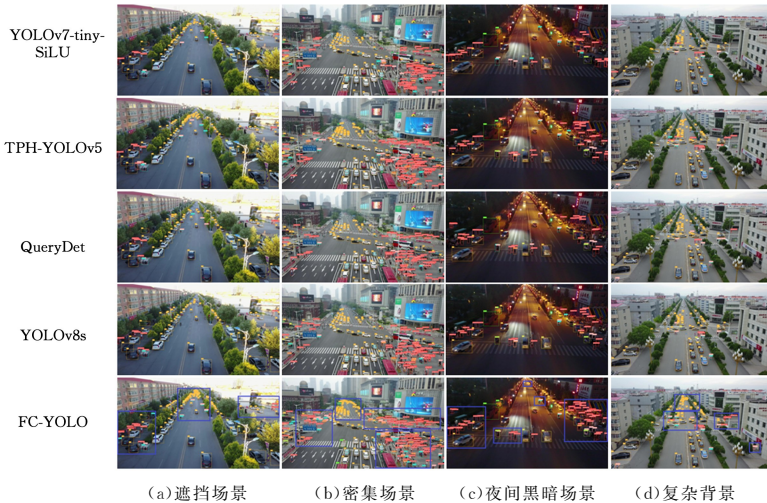


图 7 可视化检测结果对比

Fig. 7 Comparison of visual detection results

图 7(a) 是目标遮挡场景中的检测结果,在无人机航拍过程中,由于拍摄角度的限制,航拍小目标容易被场景中的树木或其他物体遮挡,在这种情况下,基准算法 YOLOv8s 和其他算法如 TPH-YOLOv5, QueryDet 及 YOLOv7-tiny-SiLU 均出现了不同程度的漏检问题,检测效果不理想,而 FC-YOLO 可以更准确地检测出被遮挡的汽车、摩托车和行人,表明其在目标遮挡的情况下具有更强的目标检测能力和鲁棒性。在图 7(b) 的密集目标场景下,基准算法和其他主流算法存在明显的

误检和漏检问题,相比之下,FC-YOLO 对航拍密集小目标的定位和分类更为准确,可以有效避免目标的漏检和误检,提升了无人机航拍密集小目标检测的准确性,在目标密集场景下具有更优秀的检测性能。图 7(c) 展示的是夜间黑暗场景下的检测结果,夜间环境下存在光线不足和对比度降低的问题,而 FC-YOLO 仍然可以准确识别路上的行人与摩托车,成功检测到了更多的小目标,与多种主流算法相比,其在夜间黑暗环境下的检测性能更为出色。从图 7(d) 中可见,基准模型和

其他算法难以区分航拍小目标和复杂背景,对十字路口及街道远处的汽车和行人存在较多漏检的情况,而 FC-YOLO 可以准确检测出远处的小目标汽车、路口的摩托车及行人,有效减少了航拍小目标的误检和漏检。通过可视化结果的对比分析,进一步验证了 FC-YOLO 在处理复杂场景无人机航拍小目标检测任务时的出色表现,具有优秀的检测性能。

结束语 本文提出一种基于特征增强与上下文融合的无人机小目标检测算法 FC-YOLO,旨在有效解决无人机航拍目标尺寸小、特征信息不足、分布密集遮挡导致的检测精度低的问题,并提高模型的推理速度。首先,构建增强特征提取的轻量化主干网络,利用特征提取轻量块高效地提取关键特征,并构建细粒度通道融合块有效地避免目标细粒度特征的丢失,该主干网络有效提高了模型的检测精度和推理速度;其次,构建小目标检测头,充分提取小目标的位置信息和细节特征;然后,利用自适应选择空间注意力模块,充分融合小目标周围丰富的上下文信息;最后,引入基于最小点距离的边界框回归损失函数 MPDIoU,提高航拍密集小目标检测的精度。将 FC-YOLO 与多种主流算法进行了充分的对比,结果表明,本文算法较好地平衡了精度和速度,适用于无人机航拍小目标检测场景。然而,在无人机设备资源受限的环境中,本文算法在轻量化方面仍有提升空间。未来将继续对模型进行优化,在保证检测精度的前提下,利用剪枝或量化技术进一步提高模型的实时性和部署灵活性。

参 考 文 献

- [1] MITTAL P, SINGH R, SHARMA A. Deep learning-based object detection in low-altitude UAV datasets: A survey[J]. *Image and Vision computing*, 2020, 104: 104046.
- [2] LIU Y, SUN P, WERGELES N, et al. A survey and performance evaluation of deep learning methods for small object detection [J]. *Expert Systems with Applications*, 2021, 172: 114602.
- [3] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//*Proceedings of the 14th European Conference on Computer Vision*. Cham: Springer, 2016: 21-37.
- [4] 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.
- [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] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017: 7263-7271.
- [7] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection [J]. *arXiv*: 2004. 10934, 2020.
- [8] GLENN-JOCHER. YOLOv5[EB/OL]. (2021-10-12)[2024-09-12]. <https://github.com/ultralytics/yolov5>.
- [9] GE Z, LIU S, WANG F, et al. Yolox: exceeding yolo series in 2021[J]. *arXiv*: 2107. 08430, 2021.
- [10] LI C, LI L, JIANG H, et al. YOLOv6: a single-stage object detection framework for industrial applications[J]. *arXiv*: 2209. 02976, 2022.
- [11] WANG C Y, BOCHKOVSKIY A, LIAO H Y. 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.
- [12] GLENN-JOCHER. YOLOv8[EB/OL]. (2023-09-27)[2024-09-12]. <https://github.com/ultralytics/ultralytics>.
- [13] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014: 580-587.
- [14] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: towards real-time object detection with region proposal networks[C]//*Advances in Neural Information Processing Systems*. 2015.
- [15] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021: 2778-2788.
- [16] FENG Z Q, XIE Z J, BAO Z W, et al. Real-time dense small object detection algorithm for UAV based on improved YOLOv5 [J]. *Acta Aeronautica et Astronautica Sinica*, 2023, 44(7): 251-265.
- [17] YANG C, HUANG Z, WANG N. QueryDet: cascaded sparse query for accelerating high-resolution small object detection [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022: 13668-13677.
- [18] XIE C H, WU J M, XU H Y. Small object detection algorithm based on improved YOLOv5 in UAV image[J]. *Computer Engineering and Applications*, 2023, 59(9): 198-206.
- [19] ZHANG Y, WU C, ZHANG T, et al. Self-attention guidance and multiscale feature fusion-based UAV image object detection[J]. *IEEE Geoscience and Remote Sensing Letters*, 2023, 20: 1-5.
- [20] DOU Z, HU C G, LIANG J Y, et al. Lightweight target detection algorithm based on improved Yolov4-tiny [J]. *Computer Science*, 2023, 50(S1): 484-490.
- [21] ZHANG J, LEI J, XIE W, et al. SuperYOLO: super resolution assisted object detection in multimodal remote sensing imagery [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-15.
- [22] SILIANG M, YONG X. MPDIoU: a loss for efficient and accurate bounding box regression[J]. *arXiv*: 2307. 07662, 2023.
- [23] ZENG S, YANG W, JIAO Y, et al. SCA-YOLO: A new small object detection model for UAV images[J]. *The Visual Computer*, 2024, 40(3): 1787-1803.
- [24] ZHU L, WANG X, KE Z, et al. BiFormer: vision transformer with bi-level routing attention[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023: 10323-10333.
- [25] JIANG L, YUAN B, DU J, et al. MFFSODNet: Multi-Scale Feature Fusion Small Object Detection Network for UAV Aerial Images[J]. *IEEE Transactions on Instrumentation and Measurement*, 2024, 73: 5015214.

- [26] ZHU J, WU Y, MA T. Multi-Object Detection for Daily Road Maintenance Inspection With UAV Based on Improved YOLOv8 [J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(11):16548-16560.
- [27] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning[C]//ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). IEEE, 2023:1-5.
- [28] MA S, LU H, LIU J, et al. LAYN: Lightweight multi-scale attention yolov8 network for small object detection[J]. IEEE Access, 2024, 12:292924-29307.
- [29] ZHOU Q, SHI H, XIANG W, et al. DPNet: Dual-path network for real-time object detection with lightweight attention[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(3):4504-4518.
- [30] CHEN J, KAO S, HE H, et al. Run, Don't Walk: chasing higher FLOPS for faster neural networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023:12021-12031.
- [31] LIU Y, ZHANG K H, FAN J Q, et al. Progressively aggregating multi-scale scene context features for camouflaged object detection[J]. Journal of Computer Science and Technology, 2022, 45(12):2637-2651.
- [32] DU D, ZHU P, WEN L, et al. VisDrone-DET2019: the vision meets drone object detection in image challenge results [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019.
- [33] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes(voc) challenge[J]. International Journal of Computer Vision, 2010, 88:303-338.



CHEN Chongyang, born in 2000, post-graduate. His main research interests include object detection and computer vision.



PENG Li, born in 1967, Ph.D, professor, Ph.D supervisor. His main research interests include computer vision, deep learning and visual Internet of Things.

(责任编辑:喻黎)