



# 计算机科学

COMPUTER SCIENCE

## 面向无人机航拍图像的YOLOv8目标检测改进算法

胡惠娟, 秦一锋, 徐鹤, 李鹏

引用本文

胡惠娟, 秦一锋, 徐鹤, 李鹏. 面向无人机航拍图像的YOLOv8目标检测改进算法[J]. 计算机科学, 2025, 52(4): 202-211.

HU Huijuan, QIN Yifeng, XU Heand LI Peng. [An Improved YOLOv8 Object Detection Algorithm for UAV Aerial Images](#) [J]. Computer Science, 2025, 52(4): 202-211.

---

## 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)

### [基于虹膜纹理感知的自适应关联学习方法](#)

Adaptive Contextual Learning Method Based on Iris Texture Perception

计算机科学, 2025, 52(4): 185-193. <https://doi.org/10.11896/jsjx.250100022>

### [基于边缘约束和改进Swin Unetr的复杂器官分割方法](#)

Complex Organ Segmentation Based on Edge Constraints and Enhanced Swin Unetr

计算机科学, 2025, 52(4): 177-184. <https://doi.org/10.11896/jsjx.240600007>

### [SCFNet:一种面向时空预测的外部空间特征融合框架](#)

SCFNet:Fusion Framework of External Spatial Features for Spatio-temporal Prediction

计算机科学, 2025, 52(4): 110-118. <https://doi.org/10.11896/jsjx.241000094>

### [空天地一体化网络的无人机轨迹和计算卸载联合优化](#)

Joint Optimization of UAV Trajectories and Computational Offloading for Space-Air-GroundIntegrated Networks

计算机科学, 2025, 52(4): 74-84. <https://doi.org/10.11896/jsjx.241000098>

### [渐进自适应特征融合的轻量化火焰检测算法研究](#)

Study on Lightweight Flame Detection Algorithm with Progressive Adaptive Feature Fusion

计算机科学, 2025, 52(4): 64-73. <https://doi.org/10.11896/jsjx.241000093>

# 面向无人机航拍图像的 YOLOv8 目标检测改进算法

胡惠娟<sup>1</sup> 秦一锋<sup>1</sup> 徐鹤<sup>1,2</sup> 李鹏<sup>1,2</sup>

1 南京邮电大学计算机学院、软件学院、网络空间安全学院 南京 210023

2 江苏省高性能计算与智能处理工程研究中心 南京 210023

**摘要** 针对无人机视角下航拍图像目标检测中存在的目标尺度变化多样、背景复杂、小目标聚集以及无人机平台计算资源受限等问题,提出了一种改进 YOLOv8 目标检测算法 YOLOv8-CEBI。首先,在骨干网络引入轻量级 Context Guided 模块,显著降低模型参数量与计算量,同时引入多尺度注意力机制 EMA,捕获细粒度空间信息,提升对小目标和在复杂背景下的检测能力。其次,引入加权双向特征金字塔网络 BiFPN,对颈部进行改造,在保持参数成本的前提下,增强多尺度特征融合能力。最后利用 Inner-CIOU 损失函数生成辅助边框以更精准地计算损失并加速边界框回归过程。在 VisDrone 数据集上进行实验,结果表明,与原始 YOLOv8s 算法相比,改进方法参数量减少 51.3%,运算量减少 28.5%,mAP50 提升 1.6%。所提模型在轻量化的同时提升了精度,取得了在减少计算资源与保证精度之间的平衡。

**关键词:** 无人机;航拍图像;注意力机制;损失函数;轻量化

**中图分类号** TP391

## An Improved YOLOv8 Object Detection Algorithm for UAV Aerial Images

HU Huijuan<sup>1</sup>, QIN Yifeng<sup>1</sup>, XU He<sup>1,2</sup> and LI Peng<sup>1,2</sup>

1 School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

2 Jiangsu HPC and Intelligent Processing Engineer Research Center, Nanjing 210023, China

**Abstract** Aiming at the problems of diverse target scales, complex backgrounds, small target aggregation, and limited computing resources of drone platforms target detection of aerial images, an improved YOLOv8 target detection algorithm YOLOv8-CEBI is proposed. Firstly, a lightweight Context Guided module is introduced into the backbone network to significantly reduce the number of model parameters and computation. At the same time, a multi-scale attention mechanism EMA is introduced to capture fine-grained spatial information and improve the detection ability for small targets and complex backgrounds. Secondly, the weighted bidirectional feature pyramid network BIFPN is introduced to transform the neck, and the multi-scale feature fusion ability is enhanced under the premise of maintaining the parameter cost. Finally, the Inner-CIOU loss function is used to generate the auxiliary bounding box to calculate the loss more accurately and accelerate the bounding box regression process. Experiments on the VisDrone dataset show that compared with the original YOLOv8s algorithm, the proposed method parameter amount is reduced by 51.3%, the computation amount is reduced by 28.5%, and the mAP50 is increased by 1.6%. The proposed model ensures the improvement of accuracy and achieves a balance between reducing computing resources and ensuring accuracy.

**Keywords** Drones, Aerial images, Attention mechanism, Loss function, Lightweighting

计算机视觉技术为无人机在智能化和信息化方面带来了革命性的改变,无人机获得了自主感知、分析和决策的能力,极大扩展了其应用范围和功能性。通过应用目标检测技术,无人机获得实时感知目标的能力,充分发挥其广阔的鸟瞰视角,从空中俯视目标,并利用灵活机动性进行调整,在信息探索和数据收集方面更加高效灵活<sup>[1]</sup>。

目前,在深度学习和人工智能的推动下,目标检测正处于快速发展阶段<sup>[2]</sup>。深度学习目标检测算法可以分为单阶段和双阶段两种。这两种方法都是通过预定义一系列锚框来进行目标检测。双阶段算法首先生成一系列目标潜在的区域,

然后对这些区域进行分类和边界框回归,代表性算法有 R-CNN<sup>[3]</sup>, Fast R-CNN<sup>[4]</sup>, Faster R-CNN<sup>[5]</sup>, Mask R-CNN<sup>[6]</sup>等。单阶段算法直接从原始图像中提取特征信息,无需生成区域建议,就可以直接预测目标类别和边界位置,经典单阶段代表性算法有 SSD<sup>[7]</sup>, YOLO<sup>[8-12]</sup>, RetinaNet<sup>[13]</sup>等。单阶段目标算法由于处理速度较快且资源需求低,因此更适用于实时响应的应用场景。

尽管目标检测研究不断深入,但在无人机平台部署仍然面临重大挑战。一方面,无人机属于典型的边缘计算设备,限制了在其上运行的算法的复杂度。另一方面,无人机通常处

到稿日期:2024-05-10 返修日期:2025-02-17

基金项目:国家重点研发计划(2019YFB2103003)

This work was supported by the National Key Research and Development Program of China(2019YFB2103003).

通信作者:徐鹤(xuhe@njupt.edu.cn)

于动态变化的高空环境之中,在复杂的背景中准确识别地面的微小目标面临着巨大的挑战。这要求算法在保证效能的同时,对资源消耗严格控制,以适应在边缘环境中进行数据处理与决策分析。

为了应对这一挑战,国内外学者开展了一系列的研究。针对无人机边缘计算轻量化需求,Wang 等<sup>[14]</sup>通过设计多路径反向残差块并结合 CBAM 注意力机制取代 C3 模块,有效减少了参数及运算量。Niu 等<sup>[15]</sup>通过将 YOLOv8 与上下文指导网和具有多分支结构的残余网络合并,增加了模型学习特征的能力,而没有增加计算需求。为了解决无人机视角观察时目标微小且分布密集导致的难题,Pan 等<sup>[16]</sup>在特征融合网络中新增一个层级,利用浅层特征图所包含的位置信息和轮廓信息,来处理微小目标的特征融合。Gong 等<sup>[17]</sup>提出了一种针对微小目标特征的多尺度卷积模块特征融合技术,设计融合因子用于调节浅层和深层的融合程度,并探索如何针对不同的数据集确定最佳值。Li 等<sup>[18]</sup>从关注大尺度特征映射角度出发,引入 Bi-PAN-FPN 思想,增强了模型对小目标的关注度,并提升了多尺度特征融合的有效性。

尽管上述方法取得了一定效果,但模型往往在轻量化后面临精度下降的问题,而在增强特征融合等方面的改进又导致模型复杂度上升,增加了计算量,导致无人机视角下的目标检测难以取得低计算量和高准确性之间的平衡。基于这些问题,本文选择 YOLOv8s 作为基准模型进行改进研究。YOLOv8 作为单阶段实时目标检测算法,在之前 YOLO 系列成功的基础上进行多项关键核心改进,较之前版本运行速度更快,检测精度更高。small 版本计算规模适中,整体性能更适用于无人机平台资源受限的环境。本文针对无人机航拍图像,对 YOLOv8s 进行了以下改进研究。

1)在骨干网络引入低计算成本的上下文特征提取 CG 模块,有效降低了模型的计算成本,并结合多尺度注意力机制

EMA,提高了复杂背景中对小目标特征的关注,增强了模型对复杂背景中相似目标区分的能力。

2)引入 BiFPN 的思想改进颈部特征融合方式,在增强多尺度特征融合的基础上精细考虑通道的数量,解决了因目标尺度差异大导致小目标特征丢失的问题,同时精简了网络结构。

3)利用 Inner-CIoU 替换 CIoU,通过缩放因子控制辅助边框的大小,更好地平衡不同目标尺度的损失,在不增加新的损失项的情况下,加速低 IOU 目标边框回归过程,提高收敛速度和精度。

## 1 YOLOv8 模型概述

YOLOv8,作为 ultralytics 公司于 2023 年 1 月 10 日发布的 YOLOv5 后续版本,是一个全新 SOTA 模型,支持图像分类、目标检测等多种计算机视觉任务。

在 YOLOv8 的设计中,骨干网络采用改进的 CSPDarknet53,对输入特征进行 5 次下采样,得到 5 个不同尺度的特征。CSP 模块将预处理中的 3 次卷积简化为两次卷积,同时吸收了 YOLOv7 ELAN 的多重堆叠设计理念。此外,YOLOv8 网络中使用 C2f 模块代替 YOLOv5 的 C3 模块。C2f 模块运用了梯度并联连接方式,结构中具有更多的跳层连接和额外的 Split 操作,在维持轻量级参数的同时丰富了特征提取网络的信息流。

相较于传统基于锚框的方法,YOLOv8 的检测头部分采用无锚分体式结构,有助于提升检测的准确性和效率。解耦头部使用两个独立的分支进行目标分类和预测边界框回归。对于分类任务,使用二进制交叉熵损失函数(BCE loss)。预测边界框回归任务使用分布式聚焦损失函数(Distribution Focal Loss DFL)和 CLOU 损失函数。这种检测结构可以提高检测精度,加快模型收敛速度。YOLOv8 的网络结构如图 1 所示。

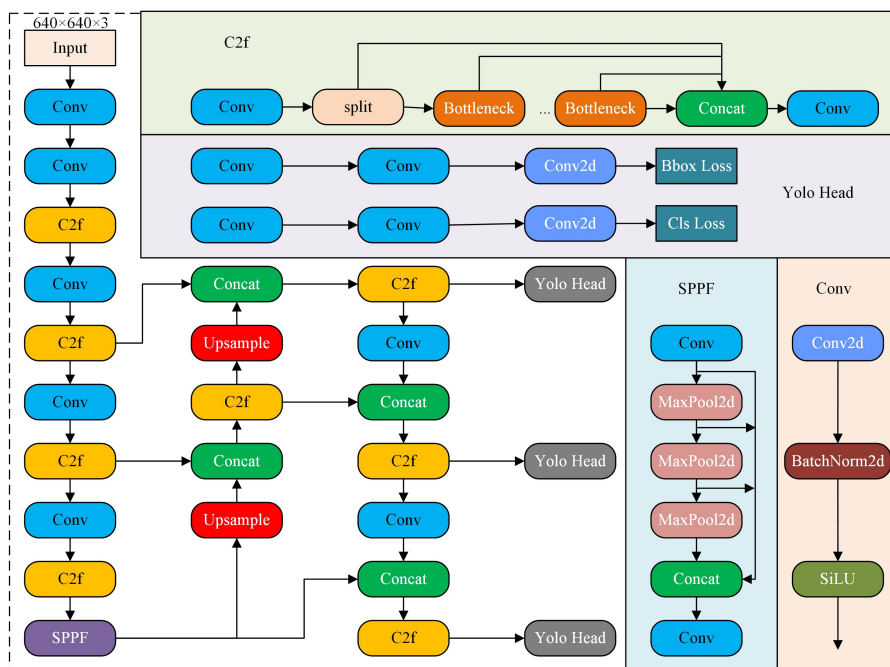


图 1 YOLOv8 网络结构

Fig. 1 YOLOv8 network structure

## 2 YOLOv8s 的改进

### 2.1 骨干网络 C2f 模块的改进

#### 2.1.1 上下文引导模块

上下文引导模块<sup>[19]</sup> (Context Guided Block), 最初被设计用于 CGNet, 它是一种专注于移动设备上语义分割任务的轻量级卷积神经网络。CG 模块旨在捕捉空间依赖和语义上下文信息, 实现像素级分类和目标定位。其总体结构图如图 2 所示。

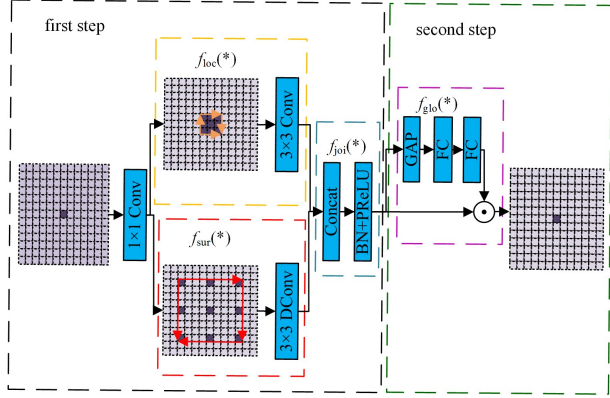


图 2 CG 模块的总体结构

Fig. 2 Overall structure of CG module

C2f 模块虽然在 C3 的基础上进行了改进, 但对于边缘部署的无人机环境来说仍存在参数多、计算量大的问题。相比之下, CG 模块采用轻量、高效的设计, 针对尺度各异或被遮挡的物体, 引入全局上下文, 综合考虑物体周围区域环境来提升识别的准确性, 更适合动态的空中无人机视角。CG 模块共分为两步, 第一步学习目标位置的局部信息和周边上下文, 生成其联合特征; 第二步利用全局上下文信息得到权重向量来对第一步中提取的联合特征进行通道级加权。

在 CG 模块的第一步中, 首先采用一个  $1 \times 1$  卷积来处理输入, 而后通过两个并行卷积来获得多尺度信息, 其中  $f_{loc}()$  提取器负责提取局部特征, 采用  $3 \times 3$  的标准卷积实现, 用于从 8 个相邻的特征向量中学习局部特征; 而  $f_{sur}()$  提取器则用于提取周围的上下文特征, 使用  $3 \times 3$  的扩张卷积进行处理。 $f_{joi}()$  则为联合特征提取器, 负责整合这两种特征, 以提取联合特征, 其定义如式(1)所示:

$$f_{joi}^* = f_{joi}(f_{loc}^*, f_{sur}^*) = BN(PReLU(f_{loc}^*, f_{sur}^*)) \quad (1)$$

其中, PReLU 表示参数化线性整流单元, BN 表示批量归一化。 $f_{loc}^*$  和  $f_{sur}^*$  分别表示局部特征和周围的上下文特征,  $f_{joi}^*$  则表示整合后的联合特征。

其次, 在 CG 模块的第二步中,  $f_{glo}()$  提取器用于提取全局上下文特征, 通过全局池化层和两个全连接层抽取特征, 得到一个权重向量。权重向量用于指导联合特征融合, 将得到的权重和输入按元素相乘, 用于逐通道微调联合特征, 强调有用元素, 压缩无用元素。其定义如式(2)和式(3)所示:

$$f_{glo}^* = f_{glo}(f_{joi}^*) = FC(FC(GAP(f_{joi}^*))) \quad (2)$$

$$f_{out}^* = f_{glo}^* \odot f_{joi}^* \quad (3)$$

其中,  $f_{glo}^*$  表示全局上下文特征,  $f_{out}^*$  表示输出特征, FC 表示完全连接层, GAP 表示全局平均池化,  $\odot$  表示逐元素相乘。

#### 2.1.2 多尺度注意力机制模块

注意力机制作为一种在神经网络领域广泛应用的增强策略, 常用于目标检测任务中关键信息的捕捉和处理。EMA 模块<sup>[20]</sup> 通过编码全局信息来重新调整每个并行分支的权重, 突出强调小目标的特征, 再通过跨维度交互进一步集成并行分支的输出, 跨空间处理长短期依赖关系, 增强通道间信息的融合。

如图 3 所示, EMA 模块采用并行子结构, 包括两个  $1 \times 1$  分支和一个  $3 \times 3$  分支, 分别对应于图中的 X Avg Pool, Y Avg Pool 和  $3 \times 3$  卷积分支。此外, EMA 模块还引入一个跨空间信息处理学习模块。在  $1 \times 1$  分支中, 首先对输入的特征图进行处理, 分别针对图像的宽度(x 方向)和高度(y 方向)进行一维全局平均池化操作, 编码通道信息, 从而实现跨通道信息的交互建模。接着, 将编码后的通道特征在高度方向上拼接, 形成一个综合的特征表示, 并共享一个  $1 \times 1$  卷积, 对融合后的特征进行整合输出。最后通过 Sigmoid 激活函数处理  $1 \times 1$  卷积的输出, 再经自适应重新加权后产生  $1 \times 1$  分支的最终输出。 $3 \times 3$  分支使用  $3 \times 3$  卷积核捕获多尺度特征并作为输出。跨空间学习模块通过二维全局平均池化和 Softmax 函数处理  $1 \times 1$  分支和  $3 \times 3$  分支的输出, 并分别进行点积操作生成两个空间注意力图。最终经 Sigmoid 激活函数和自适应重新加权后为模型提供了丰富的全局上下文信息, 使 EMA 能够在深层特征图上产生更好的像素级注意力。

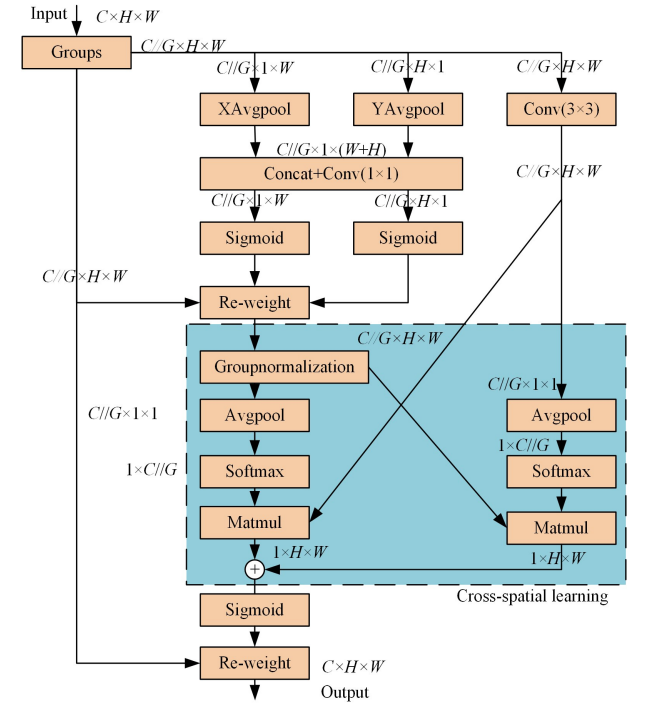


图 3 EMA 注意力机制结构

Fig. 3 EMA attention mechanism structure

## 2.2 基于 BiFPN 的多尺度特征融合网络

在航拍图像目标检测领域, 小目标检测的精度不高一直是一个突出的问题, 其根源在于深层网络旨在提取更抽象和高级的特征, 而在这个过程中浅层网络捕捉的细节信息可能会逐渐减少或被认为是冗余的。

在 YOLOv8 模型中, 特征图被细分为 5 个不同的尺度,

分别标记为 B1—B5(骨干网络)、P3—P4(特征金字塔网络 FPN)和 N4—N5(路径聚合网络 PAN),其具体结构如图 4 所示。YOLOv8 采用 PAN-FPN 结构,在 FPN 顶层加入自下而上的 PAN 结构,解决了 FPN 网络受到单向信息流限制的问题,并采用 P4—N4 和 B5—N5 来增强定位特征的学习,但仍存在小目标特征丢失和利用不够充分的情况。

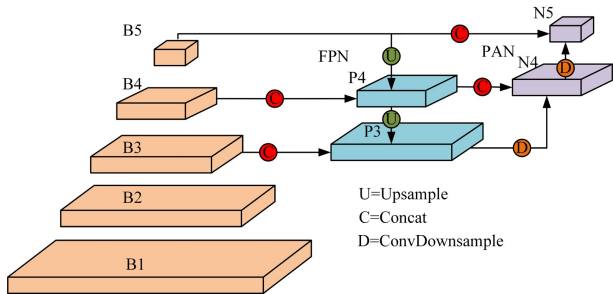


图 4 YOLOv8 中 PAN-FPN 结构示意图

Fig. 4 PAN-FPN structure diagram in YOLOv8

为了有效解决这一问题,本文在 YOLOv8 的基础上,采用了加权双向特征金字塔网络(bidirectional feature Pyramid Network, BiFPN)的架构,对其特征融合部分进行改进。BiFPN<sup>[21]</sup>最初是由 Google 在 EfficientDet 检测算法中提出的全新特征融合方法,通过跨尺度连接和加权特征融合来增强特征的语义信息。BiFPN 主要有 3 种优化思路:1)如果一个节点没有进行特征融合,仅具有单一的输入边,那么它对整合多样特征的网络贡献就会减少,可以将其移除;2)对于处于同一层级的原始输入和输出节点,在它们之间添加一条额外的边,以便在不增加太多成本的前提下实现更广泛的特征融合;3)将双向路径当作一个单独层级,并通过多次重复这样的层级,来实现更高层次的特征融合。

如图 6 所示,本文提出的方法增加一个连接层 N3,同时在 B5 和 N5、B4 和 N4、B3 和 N3 之间添加新的跨尺度连接方法,保留更多的原始信息特征。接着,将 B2 与 N3 相连接,

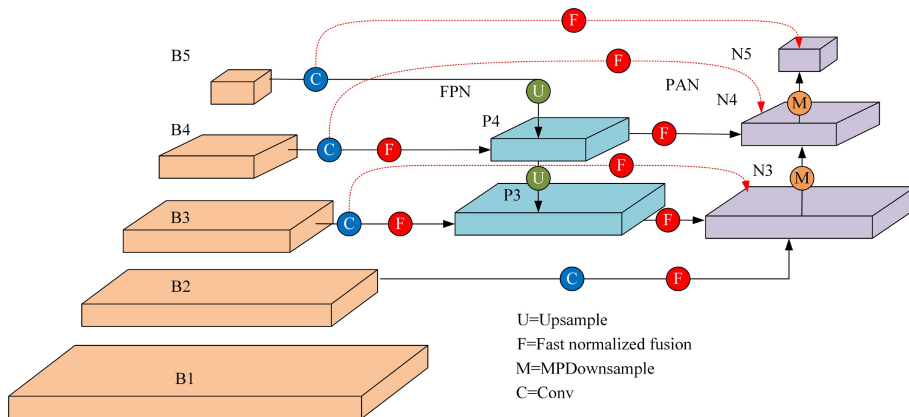


图 6 改进 BiFPN 结构图

Fig. 6 Improved BiFPN structure diagram

### 2.3 损失函数

传统的目标检测损失函数侧重于通过引入新的损失项来加速模型收敛,而忽略了 IoU 损失项本身的局限性。在实际的应用中,IoU 损失无法根据不同的检测器和检测任务进行自适应调整,缺乏泛化能力。为了解决这个问题,本文引用

用于捕获骨干网络中更高层次的特征信息。在特征融合前,本研究对原始卷积层的参数进行调整,将通道数降低到 256,且降低参数数量和计算复杂度,大幅减小模型规模,使其更适合在小型硬件设备上部署。

不同于 YOLOv8 在 PAN 结构中简单使用两个步长为 2 的 3×3 卷积进行下采样操作,本文采用 YOLOv7<sup>[22]</sup>中设计更为精细的 MP 下采样模块,其结构如图 5 所示。第一个分支首先经过一个最大池化层,通过取每个池化窗口内的最大值进行下采样,然后经过一个 1×1 卷积改变通道数。第二条分支先进行 1×1 卷积,改变特征图的通道数,接着使用 3×3 卷积实现进一步下采样,捕获更多的空间上下文。最后,将两个分支的结果合并,得到空间分辨率缩小两倍的特征图。相比普通卷积,MP 下采样模块通过结合最大池化和 3×3 卷积操作,充分保留局部区域最显著特征,并捕获更大的感受野,以包含更多的上下文信息,在多尺度下采样过程中更好地保留小目标的信息。

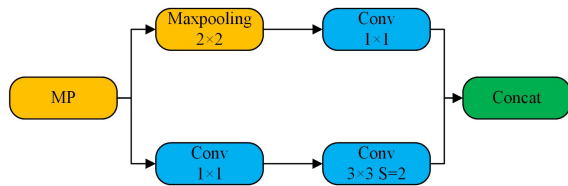


图 5 MP 模块结构图

Fig. 5 MP module structure diagram

其中,为了融合不同分辨率的特征,BiFPN 引入了额外的权重参数,采用快速化归一融合(Fast normalized fusion),使权值保持在 0~1,提高模型在 GPU 上的运行速度。快速化归一融合的计算式如下:

$$O = \sum_{i \in \epsilon} \frac{W_j}{\sum_j W_j} \cdot I_i \quad (4)$$

其中, $W_j$ 表示可以学习的权重大小; $I_i$ 为输入特征的大小; $\epsilon$ 设定为一个很小的数,以防止数值不稳定。

Inner-CIoU<sup>[23]</sup>损失函数,在原 Ciou 损失函数的基础上,引入尺度因子 ratio 来控制辅助边框的尺寸大小,根据检测目标的 IoU 和尺度的不同,采用不同大小的辅助框用于计算损失,从而加速收敛速度,以更高效地产生回归结果,提高预测框的准确度。

图 7 分别给出了 Inner-IoU 在检测目标为大尺度和小尺度情况下辅助框的生成方式。对于高 IoU 的大尺度目标, Inner-IoU 会生成比真实框更小的辅助框。通过更小的辅助框计算 IoU 损失, 可以加强梯度值, 从而加速高 IoU 样本回归。与之相反, 对于低 IoU 的小尺度目标, 则使用比真实框更大的辅助框, 扩大回归的有效范围, 增加计算低 IoU 损失的灵敏度, 从而加强低 IoU 样本回归。将 Inner-IoU 应用至现有基于 CIoU 的边框回归损失函数中,  $L_{\text{Inner-CIoU}}$  的定义如下:

$$L_{\text{Inner-CIoU}} = L_{\text{CIoU}} + \text{IoU} - \text{IoU}^{\text{inner}} \quad (5)$$

$$\text{IoU}^{\text{inner}} = \frac{\text{inter}}{\text{union}} \quad (6)$$

$$\text{union} = (\omega^{\text{gt}} * h^{\text{gt}}) * (\text{ratio})^2 + (\omega * h) * (\text{ratio})^2 - \text{inter} \quad (7)$$

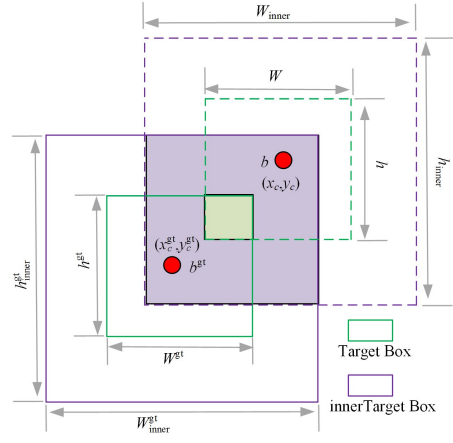
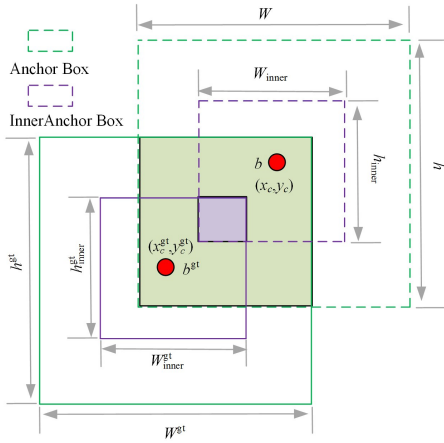


图 7 Inner-IoU 描述图

Fig. 7 Description of Inner-IoU

## 2.4 改进后的网络结构

改进后的算法 YOLOv8s-CEBI 模型主要用于应对大范围目标、复杂背景和边缘部署所带来的挑战。本文重新设计了网络体系结构, 简化了网络模型。在原始 YOLOv8 中, 骨干网络中所有 C2f 模块都被 CG 模块和 EMA 注意力机制模块取代, 以减少参数计算量的同时增强深度特征提取。

此外, 基于 BiFPN 思想提出一种新的跨尺度的特征融合网络, 在控制计算成本的同时, 保留更多原始特征, 从而提升网络在小目标方面的准确性。最后通过 Inner-CIoU 损失函数加速模型对低 IoU 样本的收敛速度, 提升模型的整体性能。

改进后的 YOLOv8s-CEBI 网络结构如图 8 所示。

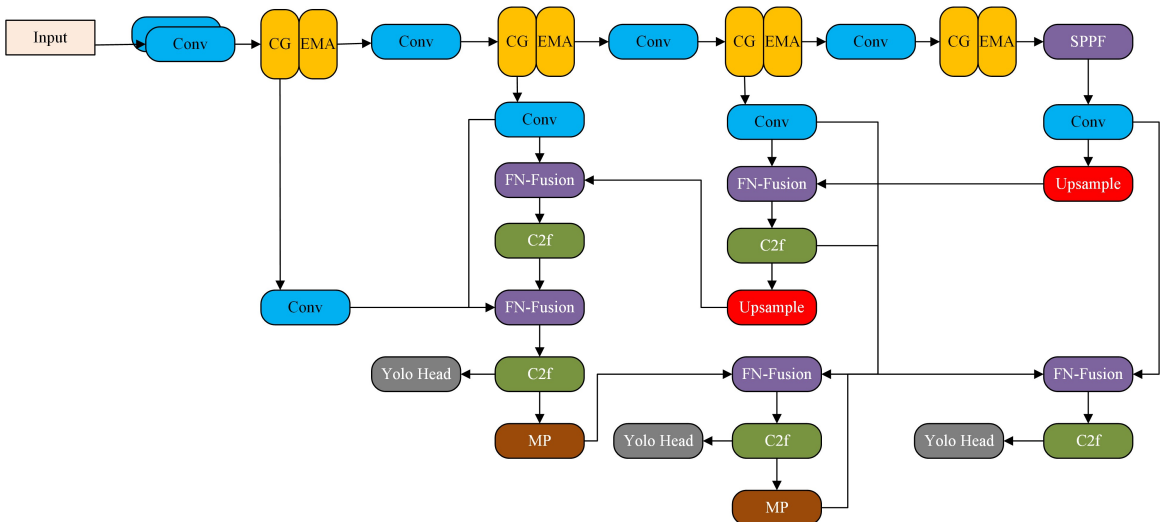


图 8 YOLOv8s-CEBI 结构图

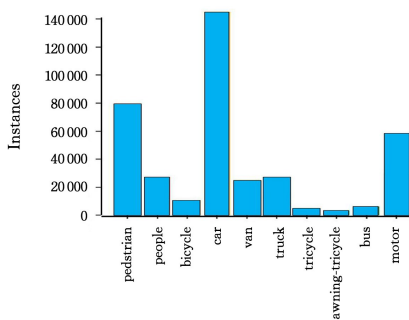
Fig. 8 YOLOv8s-CEBI structure diagram

### 3 实验结果与分析

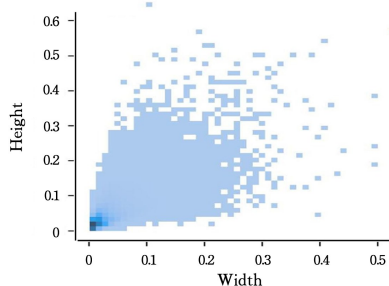
#### 3.1 实验数据集

本研究采用天津大学 AISKEYEYE 团队发布的 Vis-Drone2019<sup>[24]</sup>数据集进行训练和性能评估。该数据集中包含不同型号的无人机在中国 14 个不同城市、不同场景和高度、不同天气和照明条件下拍摄的共 10 209 幅静态图像,划分为训练集(6 471 张)、验证集(578 张)和测试集(3 190 张),包含约 260 万个目标实例,涵盖了 10 种不同类别的目标。

在目标检测任务中,将原图像尺寸下长宽比小于 0.1 的目标定义为小目标。如图 9 所示,在 VisDrone2019 数据集,大多数目标长宽比在 0.1 以内,符合小目标的相对尺寸,且样本分布不均匀,汽车和行人占据 10 种类别中的大部分,而自行车、卡车等出现较少。与其他数据集相比,VisDrone2019 数据集小目标数量更多,种类尺度差异大,更适用于无人机视角下的目标检测情景。



(a) 相对大小分布统计



(b) 类样本统计

图 9 VisDrone2019 数据集上相对大小分布统计和类样本统计

Fig. 9 Relative size distribution statistics and class sample statistics on VisDrone 2019 dataset

#### 3.2 实验环境及参数配置

本文采用 Windows11 操作系统作为测试平台,实验采用 Python3.8 作为语言环境、Pytorch2.0 为深度学习框架。实验所用计算机 GPU 为 NVIDIA GeForce RTX 3070ti, CUDA 版本为 11.7, CPU 为 Intel(R)Core(TM)i7-12700H, 32 GB 运行内存。实验参数的批量大小为 8, 训练轮数为 200, 学习率为 0.01, 权重衰减系数为 0.0005, 随机种子固定为 1, 输入图

片尺寸为  $640 \times 640$ 。训练从 0 开始, 不利用任何预训练权重, 其余参数保持默认。

#### 3.3 实验评价指标

为了有效评估改进模型的性能, 采用准确率(Precision)、召回率(Recall)、平均精度均值(mAP)、模型参数量(Params)、总浮点运算量(FLOPs)作为评价指标。

准确率用于衡量在所有预测为正样本的样本中, 实际为正样本的比例。召回率用于评估模型能够成功预测所有实际正样本的能力。mAP 表示平均精度, 计算所有类别的 AP 值, 然后取这些 AP 值的平均值, 综合评估算法在不同类别上的性能。mAP50 表示 IOU 取 0.5 时所有类别的平均精度, 用于衡量模型整体的好坏程度。准确率、召回率和 mAP 的定义分别如式(13)~式(15)所示:

$$P = \frac{T_p}{T_p + F_p} \quad (13)$$

$$R = \frac{T_p}{T_p + F_n} \quad (14)$$

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

其中,  $T_p$  表示正确预测的正样本个数;  $F_n$  表示实际是正样本但错误预测为负样本的个数;  $F_p$  表示实际是负样本但错误预测为正样本的个数;  $T_n$  表示正确将负样本判断为负样本的个数。

模型参数量反映模型的大小, 与模型结构规模成正比。总浮点运算量统计的是模型计算过程中浮点运算操作次数, 用于衡量模型的复杂度和资源消耗, 总浮点运算量越小, 表示模型所需要的计算能力就越小。

#### 3.4 BiFPN 的对比实验

为了验证本文提出的基于 BiFPN 的多尺度特征融合网络的有效性, 本文进行了一系列对比分析实验。通过对 BiFPN 特征融合结构与 YOLOv8 中的传统的 PAN+FPN 结构进行比较, 分析改进后网络结构的优越性, 实验结果如表 1 所列。

从表 1 中可以看出, 相较于 YOLOv8 中的 PAN+FPN 结构, 基于 BiFPN 的多尺度特征融合网络新增了一个连接层和新的跨尺度连接方法, 导致推理时间变长, FPS 相较于基准模型略有下降, 但该结构在其他多个指标上均有较大的提升, 成功提高了对小目标的检测精度, 并减轻了模型负担, 在参数量和计算量上展现出明显优势。

在颈部多尺度的特征融合中, 采用不同的融合方式会对模型产生不同的效果。在 BiFPN 特征融合结构中对比 5 种不同的融合方式, 分别是加权融合(Weighted Fusion)、自适应融合(Adaptive Fusion)、串联融合(Concatenation Fusion)、快速归一化融合(Fast normalized fusion)和语义信息和细节融合<sup>[25]</sup>(Semantics and Detail Infusion), 旨在探索适合无人机视角的最佳融合方式。

表 1 PAN+FPN 结构与 BiFPN 的对比

Table 1 Comparison of PAN+FPN structure with BiFPN

Methods	P/%	R/%	mAP50/%	mAP50;95/%	Params	GFLOPs	FPS
YOLOv8s	50.4	37.9	39.1	23.6	$11.12 \times 10^6$	28.8	165
YOLOv8s+BiFPN	52.8	39.5	40.7	24.4	$7.22 \times 10^6$	25.1	162

实验结果如表 2 所列,快速归一化融合相较于其他 4 种方法,在兼顾模型计算开销和参数数量的同时,也给予检测精度足够重视,因而在综合各项指标的情况下性能最佳。

表 2 不同融合方式的对比

Table 2 Comparison of different fusion methods

Methods	mAP50/%	Params	GFLOPs	FPS
Weighted Fusion	40.4	$7.60 \times 10^6$	26.8	132
Adaptive Fusion	40.3	$7.60 \times 10^6$	26.8	126
Concatenation Fusion	40.7	$7.72 \times 10^6$	26.2	155
Semantics and Detail Infusion	40.5	$7.52 \times 10^6$	26.2	145
Fast normalized fusion	40.6	$7.40 \times 10^6$	25.4	156

在无人机小目标检测任务中,由于进行了多次下采样操作,小目标的关键信息容易丢失。在颈部特征融合的 PAN 结构中对比 6 种不同的下采样方式,分别是步长为 2 的  $3 \times 3$  标准卷积(Conv)、空间深度转换卷积模块<sup>[26]</sup>(SPD-Conv)、小波下采样<sup>[27]</sup>(Haar Wavelet Downsampling)、YOLOv7 中的 MP 下采样、YOLOv9<sup>[28]</sup>中的 ADown 下采样和鲁棒特征降采样模块<sup>[29]</sup>(Robust Feature Downsampling),旨在探索更有效保留小目标信息的下采样方式。

实验结果如表 3 所列,通过对比 YOLO 系列常用的下

采样方法和几种新型下采样方式可知,YOLOv7 中 MP 下采样有效缓解了小目标信息丢失的问题,且 FPS 有一定的提升,说明该方法更适用于无人机视角下的检测任务。

表 3 不同下采样方式的对比

Table 3 Comparison of different downsampling methods

Methods	mAP50/%	Params	GFLOPs	FPS
Conv	40.6	$7.40 \times 10^6$	25.4	156
SPDConv	40.3	$8.29 \times 10^6$	27.2	158
Haar Wavelet Downsampling	40.5	$7.24 \times 10^6$	25.1	150
MP	40.7	$7.22 \times 10^6$	25.1	162
ADown	40.6	$7.19 \times 10^6$	25.0	151
Robust Feature Downsampling	40.2	$73.5 \times 10^6$	25.3	149

### 3.5 消融实验结果与分析

为了严格评估本文提出的改进方法的效果,在 YOLOv8s 基准模型上进行消融实验,以验证每项改进对算法性能的影响。具体实验步骤如下:首先,将 C2f 模块替换为 CG 模块+EMA 注意力机制,减少模块参数和计算量;随后,将颈部结构替换为 BiFPN 特征金字塔网络结构;最后,将原有的 CIOU 损失函数替换为 Inner-CIoU。为了确保实验的公平性,在不同的实验中只增加相应的模块,保持各实验的优化方法和超参数一致。

表 4 消融实验指标的比较

Table 4 Comparison of ablation experimental indexes

Base-line	CG	EMA	BiFPN	Inner-CIoU	P/%	R/%	mAP50/%	mAP50:95/%	Params	GFLOPs	FPS
✓					50.4	37.9	39.1	23.6	$11.12 \times 10^6$	28.8	165
✓	✓				49.2	37.2	38.6	23.3	$9.30 \times 10^6$	23.3	169
✓		✓			50.5	38.7	39.3	23.7	$11.21 \times 10^6$	29.8	155
✓			✓		52.8	39.5	40.7	24.4	$7.22 \times 10^6$	25.1	162
✓				✓	50.3	38.4	39.3	23.7	$11.12 \times 10^6$	28.8	165
✓	✓	✓			49.5	37.9	39.0	23.5	$9.36 \times 10^6$	24.3	155
✓	✓		✓		50.8	38.6	40.0	24.0	$5.35 \times 10^6$	19.6	163
✓	✓	✓	✓		51.5	39.0	40.4	24.2	$5.41 \times 10^6$	20.6	151
✓	✓	✓	✓	✓	52.3	39.4	40.7	24.4	$5.41 \times 10^6$	20.6	151

分析表 4 中的实验结果可知:从 YOLOv8s 基准模型开始,通过在骨干网络中使用 CG 模块替换 C2f 模块,模型参数量和计算量分别下降 16.3% 和 19.0%,FPS 上升 4,而 mAP50 却下降了 0.5%,可见单使用 CG 模块取代原有型号的 C2f 模块改进效果一般。CG 模块利用空间上下文信息关联的机制,能有效解决无人机视角下目标尺度变化多样的问题,但会略降低小目标的精度。在每个 CG 模块之后加上 EMA 注意力机制,能提高对特征图中关键信息的捕获,减少背景噪声对特征提取过程的干扰,使得 mAP50 提高了 0.4%,与基准模型基本持平。这表明使用 CG 模块+EMA 注意力机制的组合替换 C2f 模块能够在有效控制模型参数和计算量的同时增强模型对图像关键区域的注意能力。然而,添加 EMA 注意力机制,会导致 FPS 出现略微下降。EMA 通过多尺度卷积和全局平均池化来建模长程依赖关系,尽管在特征提取能力方面有显著效果,但频繁的数据访问带来了额外的计算延迟。随后,引入 BiFPN 特征金字塔网络结构,

有效降低了小目标的漏检率,使 mAP50 提升 1.4%,同时在特征融合之前应用卷积层调整通道数,使得模型参数和计算量大幅下降,减轻了特征融合的计算和内存负担,该改进模块提升效果显著。新的颈部特征融合网络在深层网络中有效缓解了小目标信息的丢失问题,降低了小目标的漏检率。CG 模块和基于 BiFPN 的多尺度特征融合网络的联合应用有效缓解了无人机平台计算资源受限的问题。然后,应用 Inner-CIoU 损失函数,mAP50 增加了 0.3%,但模型的计算量并未增加,实现了无损提升,这突出了 Inner-CIoU 损失函数在处理不同数据集时使用不同大小的辅助框的有效性。

改进后的模型平均检测精度提升了 1.6%,模型的参数量和计算量远小于基线模型,分别下降了 51.3% 和 28.5%,FPS 略微下降,但算法仍能保持较高的帧率,满足无人机实时目标检测的需求。如图 10 所示,该算法在轻量化的前提下,提升了精度,其查准率、召回率和 mAP50 均有提升,达到了在减少计算资源与保证精度之间的平衡。

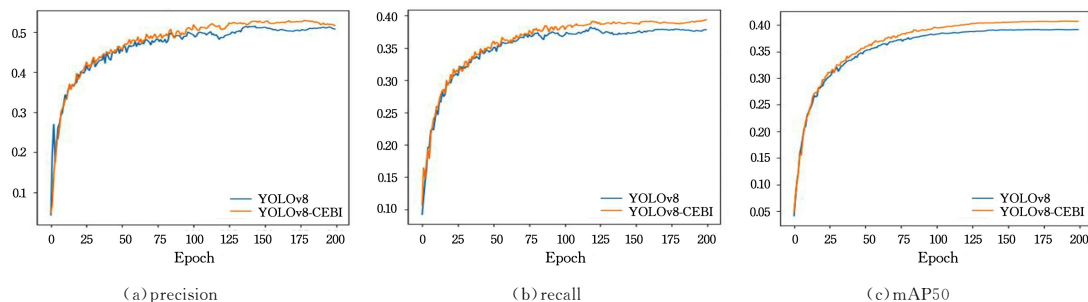


图 10 查准率、召回率、mAP50 的对比

Fig. 10 Comparison of precision, recall and mAP50

### 3.6 改进算法可视化分析

不同场景下 YOLOv8s 与 YOLOv8s-CEBI 的检测效果如图 11 所示。通过模型推理结果对比可以看出,在不同的场景下,改进 YOLOv8s 算法的性能明显优于基准模型 YOLOv8。

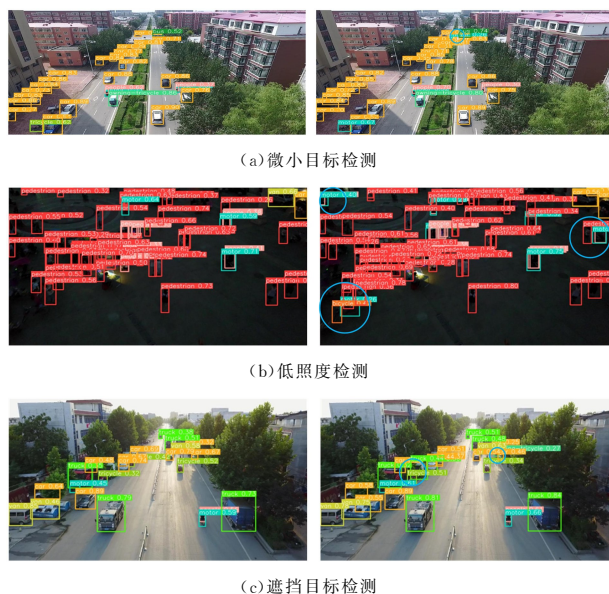


图 11 不同场景下 YOLOv8s 与 YOLOv8s-CEBI 的检测效果

Fig. 11 Detection effect of YOLOv8s and YOLOv8s-CEBI in different scenarios

在图 11(a)中,由于拍摄距离较远,YOLOv8s 无法检测到远处微小的摩托车,存在漏检的情况,改进后的 YOLOv8s 算法能检测到更多小目标。在图 11(b)中,在夜间和弱光条件下,由于光照不足,目标物体的特征变得模糊和难以辨别,YOLOv8s 无法检测到目标的存在,而改进算法在光线不足的

环境下也能保持较高的检测精度。如图 11(c)所示,无人机视角下拍摄的场景中,车辆由于位置重叠,部分车身外露,原始算法难以检测,而改进后的算法综合了上下文特征学习,利用未遮挡区域线索和相邻目标信息准确识别出了车辆,取得了良好的互遮条件下的检测性能。

### 3.7 对比实验

为了证明本文提出的改进算法的性能,将本文方法与 YOLO 系列算法进行了对比实验,实验结果如表 5 所列。最初版本的 YOLO 检测算法,如 YOLOv3 和 YOLOv4,模型结构繁琐复杂,计算资源需求高,难以在计算能力有限的无人机设备上实时运行和部署,目标检测能力相对较低。YOLOv3-tiny, YOLOv7-tiny 和 YOLOv8n 实现了轻量级模型,但是精度损失严重,模型识别能力不能满足检测任务需求。YOLOv7 属于后续推出的版本,但是模型尺寸相对较大,识别准确度一般。在 YOLO 系列算法中,YOLOv5s 和 YOLOv8s 则在目标检测领域引起了广泛关注,YOLOv5s 的突出特点为其轻量级网络和高效性能,而 YOLOv8s 则通过进一步优化提高了检测精度和鲁棒性。但 YOLOv8-CEBI 相较于 YOLOv5s 和 YOLOv8s 都具有一定的优势,其在任务效果和模型效率之间取得了平衡。与 YOLOv5s 相比,本文改进的算法计算开销上升,但通过模型优化降低了参数规模,且检测精度和 FPS 大幅提升;与 YOLOv8s 相比,则具有更少的参数和更高的检测精度,且仍能保持较快的检测速度。此外,与华为诺亚的 YOLO-gold<sup>[30]</sup>相比,本文所改进的基于 BiFPN 的多尺度特征融合网络更适用于小目标场景。结合各项指标来看,本文算法在无人机视角下的目标检测任务中具有良好的性能,显现出本文算法的优越性。

表 5 VisDrone2019 数据集上 YOLO 系列算法的指标比较

Table 5 Comparison of indicators of YOLO series algorithms on VisDrone2019 dataset

Models	P/%	R/%	mAP50/%	mAP50;95/%	Params	GFLOPs	FPS
YOLOv3	49.2	38.3	38.3	23.1	$61.55 \times 10^6$	154.90	69
YOLOv3-tiny	38.2	24.8	23.8	14.2	$12.10 \times 10^6$	8.68	175
YOLOv4	36.0	48.6	42.1	25.6	$60.99 \times 10^6$	111.35	66
YOLOv5s	44.8	34.1	33.3	33.2	$7.20 \times 10^6$	16.50	146
YOLOv7	51.3	41.9	39.6	22.6	$37.20 \times 10^6$	104.80	96
YOLOv7-tiny	47.5	36.2	35.3	19.6	$6.02 \times 10^6$	13.20	130
YOLOv8s	50.4	37.9	39.1	23.6	$11.12 \times 10^6$	28.80	165
YOLOv8n	42.8	32.9	32.3	18.8	$3.15 \times 10^6$	8.90	186
YOLOv8s-CEBI	51.3	39.2	40.6	24.4	$5.60 \times 10^6$	20.90	151
YOLO-gold	50.6	38.6	39.5	23.8	$21.50 \times 10^6$	46.00	155

将 YOLOv8s-CEBI 与目标检测领域其他主流的算法进行比较,实验结果如表 6 所列,结果进一步突出了改进方法在无人机航拍图像方面的优势,证明本文模型相对于其他优秀模型具有最好的检测性能。

表 6 VisDrone2019 数据集上主流算法的指标比较

Table 6 Comparison of indicators of mainstream algorithms on VisDrone2019 dataset

Models	mAP50/%	Params	GFLOPs	FPS
Faster R-CNN	37.50	$41.19 \times 10^6$	134.38	21
RetinaNet	20.12	$36.10 \times 10^6$	127.82	32
CenterNet <sup>[31]</sup>	33.50	$32.67 \times 10^6$	109.34	61
SSD	23.50	$24.50 \times 10^6$	87.90	50
Cascade R-CNN <sup>[32]</sup>	39.50	$69.17 \times 10^6$	52.64	17
YOLOv8s-CEBI	40.60	$5.60 \times 10^6$	20.90	151
RT-DETR <sup>[33]</sup>	44.60	$32.00 \times 10^6$	110.00	182

**结束语** 本文提出了一种面向无人机的航拍改进算法 YOLOv8s-CEBI。该算法考虑了轻量级模块、注意力机制、多尺度特征融合、损失函数等策略,旨在充分利用边缘计算资源的情况下,增强复杂场景下如尺度变化、光线减弱、目标遮挡等负面影响的鲁棒性。在特征提取的骨干网络中,采用考虑空间上下文的轻量级特征提取 CG 模块,并利用 EMA 注意力机制增强特征选择性抑制,能够显著减少计算量,同时保持较强的特征提取能力。在颈部特征融合网络中,针对无人机视角下小目标漏检的普遍问题,引入 BiFPN 的思想,在尽可能保持参数成本不增加的前提下实现更先进、更完整的特征融合。最后,采用 Inner-CIoU 损失函数,通过辅助边框,有效推进小目标的回归过程,提高模型的检测效果。在 Vis-Drone2019 数据集上进行实验,结果表明,相比基线模型,改进后的模型在平均检测精度上提升了 1.6%,参数量和计算量分别下降了 51.3% 和 28.5%,在保持较高检测精度的同时,成功地降低了模型的复杂性和计算负载,虽然 FPS 有一定下降,但仍高于 YOLOv5s,能满足无人机实时检测任务的需求。下一步研究方向为:1)考虑到实际应用场景光照条件较差的情况,进一步优化预处理流程,引入图像增强和图像曝光等技术,增强算法在光照复杂环境中的鲁棒性;2)将算法进一步轻量化,通过模型压缩技术如参数剪枝和知识蒸馏等手段,不断简化模型结构,从而在保持精度的同时实现模型规模的进一步缩减。

## 参考文献

[1] LENG J X, MO M C, ZHOU Y H, et al. Research progress on target detection from the perspective of UAV [J]. Chinese Journal of Image and Graphics, 2023, 28(9): 2563-2586.

[2] DUAN Z J, LI S B, HU J J, et al. Review of Deep Learning Based Object Detection Methods and Their Mainstream Frameworks[J]. Laser & Optoelectronics Progress, 2020(12): 59-74.

[3] 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 (CVPR). Columbus: IEEE, 2014: 580-587.

[4] GIRSHICK R. Fast R-CNN[C]// Proceedings of the IEEE In-

ternational Conference on Computer Vision, 2015: 1440-1448.

[5] REN S Q, HE K M, 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.

[6] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. arXiv: 1703. 06870, 2017.

[7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]// Proceedings of the 14th European Conference on Computer Vision (ECCV). Amsterdam: Springer, 2016: 21-37.

[8] 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 (CVPR). Las Vegas: IEEE, 2016: 779-788.

[9] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger [C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2017: 6517-6525.

[10] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[J]. arXiv: 1804. 02767, 2018.

[11] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. arXiv: 2004. 10934, 2020.

[12] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios [J]. arXiv: 2108. 11539, 2021.

[13] 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 (ICCV). Venice: IEEE, 2017: 2980-2988.

[14] WANG J, ZHANG F, ZHANG Y, et al. Lightweight Object Detection Algorithm for UAV Aerial Imagery [J]. Sensors, 2023, 23: 5786.

[15] NIU Y, CHENG W, SHI C, et al. YOLOv8-CGRNet: A Lightweight Object Detection Network Leveraging Context Guidance and Deep Residual Learning [J]. Electronics, 2024, 13(1): 43.

[16] PAN X, CHEN Q B, HUANG A, et al. Small target detection algorithm for UAV aerial images based on improved YOLOX [J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2024, 44(1): 90-100.

[17] GONG Y, YU X, DING Y, et al. Effective Fusion Factor in FPN for Tiny Object Detection [C]// Workshop on Applications of Computer Vision. IEEE, 2021.

[18] LI Y, FAN Q, HUANG H, et al. A Modified YOLOv8 Detection Network for UAV Aerial Image Recognition [J]. Drones, 2023, 7(5): 304.

[19] WU T, TANG S, ZHANG R, et al. CGNet: A Light-Weight Context Guided Network for Semantic Segmentation [J]. IEEE Transactions on Image Processing, 2021, 30: 1169-1179.

[20] OUYANG D, HE S, ZHANG G Z, et al. Efficient Multi-Scale Attention Module with Cross-Spatial Learning [C]// 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2023). Rhodes Island, Greece, 2023: 1-5.

- [21] TAN M X, PANG R M, LE Q V. EfficientDet: Scalable and Efficient Object Detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020; 10781-10790.
- [22] 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.
- [23] ZHANG H, XU C, ZHANG S. Inner-IoU: More effective intersection over union loss with auxiliary bounding box [J]. arXiv: 2311.02877, 2024.
- [24] 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 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Seoul: IEEE, 2019; 213-226.
- [25] YAO P, SONKA M, CHEN D Z. U-Net v2: Rethinking the Skip Connections of U-Net for Medical Image Segmentation [J]. arXiv: 2311.17791, 2023.
- [26] SUNKARA R, LUO T. No More Strided Convolutions or Pooling: A New CNN Building Block for Low-Resolution Images and Small Objects [J]. Machine Learning and Knowledge Discovery in Databases, 2023, 13715: 27.
- [27] XU G P, LIAO W T, ZHANG X, et al. Haar wavelet downsampling: A simple but effective downsampling module for semantic segmentation [J]. Pattern Recognition, 2023, 143: 109819.
- [28] WANG C Y, YE H I H, LIAO H Y M. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information [J]. arXiv: 2402.13616, 2024.
- [29] LU W, CHEN S B, TANG J, et al. A Robust Feature Downsampling Module for Remote-Sensing Visual Tasks [J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-12.
- [30] WANG C C, HE W, NIE Y, et al. Gold-yolo: Efficient object detector via gather-and-distribute mechanism [J]. arXiv: 2309.11331, 2023.
- [31] DUAN K, BAI S, XIE L, et al. Centernet: Keypoint triplets for object detection [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019; 6569-6578.
- [32] CAI Z, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018; 6154-6162.
- [33] LV W, XU S, ZHAO Y, et al. Detsr beat yolos on real-time object detection [J]. arXiv: 2304.08069, 2023.



**HU Huijuan**, born in 1975, master, lecturer. Her main research interests include artificial intelligence and Internet of Things technology.



**XU He**, born in 1985, Ph.D, professor, master supervisor, is a senior member of CCF (No. 19957S). His main research interests include big data and Internet of Things technology.

(责任编辑:何杨)