计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2024,33(7):139−148 [doi: 10.15888/j.cnki.csa.009567]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 YOLOv8n 的无人机航拍目标检测①

沈学利, 王灵超

(辽宁工程技术大学 软件学院, 葫芦岛 125105)
通信作者: 王灵超, E-mail: LingChao_Wang@163.com

摘　要: 针对无人机航拍检测任务中小目标检测精度低的问题, 提出一种基于 YOLOv8n 的目标检测算法 (SFE-YOLO). 首先, 嵌入浅层特征增强模块, 将输入特征的浅层空间信息与颈部获取的深层语义信息融合, 以增强小目标特征表示能力, 并使用全局上下文块 (GC-Block) 对融合信息进行重校准, 抑制背景噪声. 其次, 引入可变形卷积来代替 C2F 中的部分标准卷积, 提高网络对几何变化的适应性. 再次, 引入 ASPPF 模块, 融合平均池化技术, 增强模型对多尺度特征的表达并降低漏检率. 最后, 在颈部网络的基础上嵌入中尺度特征合成层, 融合主干网络中更多的中间特征, 使不同尺度的特征过渡更平滑, 并通过跳跃连接增强特征重用性. 该模型在数据集 VisDrone2019 和 VOC2012 上进行验证, *mAP*@0.5 值达到 30.5% 和 67.3%, 相较于基线算法 YOLOv8n 提升了 3.6% 和 0.8%, 能够提升无人机图像目标检测性能, 同时具有较好的泛化性.

关键词: 无人机检测; 浅层特征融合; ASPPF; YOLOv8n; 中尺度特征融合

引用格式: 沈学利,王灵超.基于 YOLOv8n 的无人机航拍目标检测.计算机系统应用,2024,33(7):139–148. http://www.c-s-a.org.cn/1003-3254/9567.html

UAV Aerial Photography Target Detection Based on YOLOv8n

SHEN Xue-Li, WANG Ling-Chao

(Software College, Liaoning Technical University, Huludao 125105, China)

Abstract: An enhanced YOLOv8n-based object detection algorithm, SFE-YOLO, is developed to tackle the issues of low detection precision for small targets in UAV aerial photography. Initially, a shallow feature enhancement module is embedded to integrate the shallow spatial details of input features with deep semantic information obtained from the neck section. This fusion strengthens the representation capability for small target features. Additionally, a global context block (GC-Block) is utilized to recalibrate this merged information, effectively suppressing background noise. Subsequently, the network's adaptability to geometric changes is increased by substituting deformable convolutions for some standard convolutions in the C2F layer. Furthermore, the ASPPF module, incorporating average pooling technology, is integrated to augment the model's expression of multi-scale features and to decrease miss rates. Finally, a novel weighted feature fusion method is designed. This method blends more intermediate features from the main network, enabling smoother transitions among different scale features and augmenting feature reusability through skip connections. The model's performance is validated on VisDrone2019 and VOC2012 datasets, achieving *mAP*@0.5 values of 30.5% and 67.3%, respectively. These results mark improvements of 3.6% and 0.8% over the baseline YOLOv8n algorithm, demonstrating enhanced performance in UAV image target detection and notable generalization capabilities.

Key words: UAV detection; shallow feature fusion; ASPPF; YOLOv8n; mesoscale feature fusion

① 基金项目: 国家自然科学基金 (62173171)
收稿时间: 2024-01-17; 修改时间: 2024-02-26; 采用时间: 2024-03-11; csa 在线出版时间: 2024-06-05
CNKI 网络首发时间: 2024-06-07

无人机的技术迅速发展提高了目标检测技术在航拍任务中的应用价值, 然而传统的目标检测方法如Faster-RCNN[1]、RetinaNet[2]、YOLO[3]等算法框架在以往的目标检测任务中表现出色, 但它们在无人机航拍视角下, 面对复杂场景和光线不均匀等挑战时, 对于目标的检测能力仍显不足. 因此, 需要开发一种无人机小目标检测模型, 从而有效应对这些应用场景的特殊要求.

当前基于深度学习的无人机图像目标检测方法可以分为双阶段与单阶段两类, 双阶段方法以 Faster-RCNN 为代表, 先通过区域建议网络 (RPN) 产生候选区域, 再进行分类, 此算法对单张特征图进行处理, 导致图像中的空间信息被高度压缩, 不利于小目标的识别. 对此, Yin 等[4]基于 Faster-RCNN 算法增加了相应的域自适应分量, 对 RPN 网络进行了改进以提高小目标检测性能, 但模型对训练样本的质量要求较高, 不具备一定的泛化能力.

与此同时, 单阶段方法如 RetinaNet 和 YOLO 系列也在快速发展. YOLO 系列算法作为单阶段目标检测领域的先驱, 将目标检测任务从传统的分类问题转变为回归问题. 然而, 这类算法在多次下采样的过程中会导致输入图像的分辨率降低, 进而造成特征信息的丢失, 从而影响检测准确度, 尤其对小目标或者密集场景检测具有局限性[5]. RetinaNet 采用单阶段架构, 并通过 Focal Loss 解决类别不均衡问题, 提高了密集目标检测的效率. 这种架构的关键在于能够将深层的强语义信息传递至浅层, 进而增强检测性能. 然而它也存在两大挑战: 特征图在自上而下传播时分辨率降低, 导致小目标信息丢失; 横向连接的通道压缩削弱了小目标特征. 这些限制是当前单阶段检测系统需要优化的关键点. 为了提高这些方法的效率和准确性, 一些研究者进行了进一步的优化. 例如, Zhao 等[6]基于 RetinaNet, 使用系统聚类方法重新选择锚点的规模和数量, 并优化基于 RetinaNet 的焦点损失计算来提升无人机检测精度, 但是推理速度有待进一步提升. 黄海生等[7]在 YOLOv5 基础上引入 MobileNetv3 骨干网络、CBAM 注意力模块和 SiLU 激活函数, 构建了轻量化且高效的 YOLOv5-tiny 网络, 有效平衡了航拍场景目标检测的速度与精度. Tan 等[8]针对特征金字塔结构进行创新性优化, 提出双向特征金字塔结构以融合不同尺度的特征. Wang 等[9]在 YOLOv7-tiny 基础上引入全局注意力机制 (GAM) 和 Context Transformer 模块等方法来提高无人机航拍图像中的检测精度, 同时增加了额外的计算量. Zhang 等[10]在 YOLOv8 中引入 sandwich fusion 模块, 增强了无人机小目标的检测能力, 但同时也增加了计算负担. Du 等[11]设计的稀疏卷积, 在优化检测头的同时减轻了计算量, 但影响模型精度.

基于以上, 针对无人机视角下航拍图像目标检测困难的问题, 本文在 YOLOv8n 模型的基础上进行改进, 具体如下.

(1) 提出浅层特征增强模块 (SFEM), 融合浅层空间信息与颈部自顶向下传播的深层语义信息, 使网络更加关注图像的细节信息, 然后利用全局上下文模块 (GC-Block) 对融合后的特征图进行重校准, 并在此基础上增添小目标检测头, 提升小目标检测性能.

(2) 将骨干网中特定的 C2F 层替换为 DCN_C2F 模块, DCN_C2F 模块引入了可变形卷积 (deformable convolutional networks v2, DCNv2)[12]来自适应学习每个卷积点在感受野内的空间偏移, 有效聚焦于目标特征, 克服因下采样引起的特征失真问题.

(3) 改进空间金字塔池化 (spatial pyramid pooling-fast, SPPF)[13]结构, 结合全局池化和平均池化的双金字塔池化层来保留高层次的语义信息, 之后通过上采样将低层特征结合来减少空间细节信息的丢失, 提高对小目标的检测灵敏度, 抑制复杂背景的干扰, 从而降低误检与漏检的风险, 确保模型的精确性.

(4) 为应对目标特征提取中的尺度下降问题, 引入了中尺度特征合成层 (MFSL) 让特征过度更平滑, 并在此基础上采用跳跃连接和颈部网络的特征融合层来强化对目标的检测能力, 提高特征融合的效率.

(5) 在模型颈部增添简单无参注意力模块 (simple, parameter-free attention module, SimAM)[14], 对特征图中的每个位置赋予不同的权重, 突出重要的特征并抑制不重要的部分, 有助于优化特征传递和融合, 提高检测精度并减少误检率, 进而改善模型的总体性能.

1 网络改进

1.1 改进 YOLOv8n 网络结构

YOLOv8 是 YOLO 系列中最新的目标检测模型, 它在主干网络 (Backbone)、颈部 (Neck) 和头部 (Head) 方面相比以往版本进行了显著的改进.

(1) 主干网络: YOLOv8 主干网络使用 C2F 模块, 与之前的 C3 (CSPDarkNet53) 相比, C2F 模块通过合并

所有瓶颈模块的输出来提高训练速度和改善梯度流, 更快地处理图像并提取重要特征.

(2) 颈部: 颈部的设计对于在不同层级之间传递和融合特征至关重要. YOLOv8 在颈部结构中结合了特征金字塔网络 (FPN) 和路径聚合网络 (PAN)[15], 增强对不同尺度目标的检测能力, 有助于提高模型的整体性能.

(3) 头部: YOLOv8 的头部结构实现了从基于锚点的方法向无锚点的转变. 它采用了解耦式设计, 使得分类和回归任务分开执行, 从而提升了模型性能. 此外, YOLOv8 在损失函数上也做了改进, 引入了任务对齐分数, 这个分数将分类分数与交并比 (intersection over union, IoU) 分数相结合, 有助于模型在分类和定位上的精确性.

此外, YOLOv8 根据网络的深度和宽度分为不同的型号 (nano, small, medium, large, extra large), 以适应从轻量级移动设备到高精度需求的各种应用场景, 提供了显著的性能改进和更高的检测精度.

综合考虑算法的检测精度和实际应用, 本文基于 YOLOv8n 提出了一种改进的目标检测算法 (SFE-YOLO), 结构如图 1 所示, 使用浅层特征增强模块 SFEM 与 C2F_DCN 模块优化主干网络, 保持空间细节并减少形变. 改进快速空间金字塔池化结构, 以减少模型的误检率并提高检测精度. 结合中尺度特征合成层和跳跃连接以增强特征重用性, 融入 SimAM 强化无人机目标检测精度.

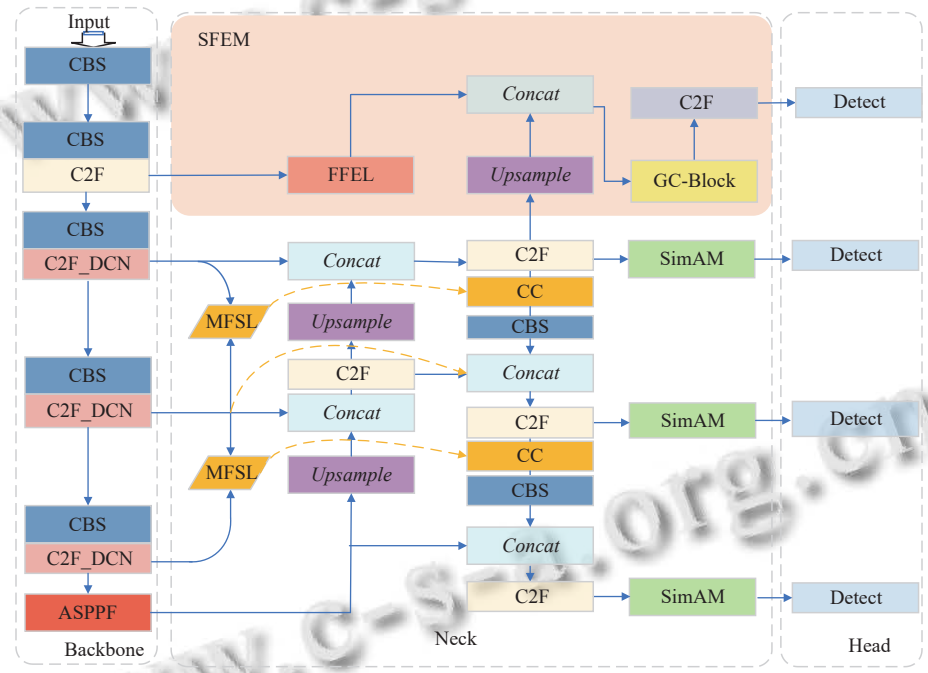


图 1 改进 YOLOv8 网络结构

1.2 浅层特征增强模块

常见的目标检测网络在深层特征提取过程中往往会丢失空间细节和边缘位置等浅层特征信息, 这对小目标的检测性能影响较大. 因此, 本文提出浅层特征增强模块 (shallow feature enhancement module, SFEM). 如图 2 所示, SFEM 首先通过快速特征提取层 (fast feature extraction layer, FFEL) 提取浅层空间信息. 然后将其与主干网络通过颈部自顶向下传播获得的深层语义信息进行融合, 得到更丰富的特征图, 接着嵌入 GC-Block, 通过其捕获全局上下文关系, 对特征图进行重校准, 增强网络去噪能力和细节信息表征能力.

快速特征提取层 (FFEL) 通过减少计算冗余和内存访问, 提高计算效率, 尤其是通过引入部分卷积 (PConv)[16]来减少浮点运算量和参数数量, 同时有效提取特征空间信息. 为解决连续卷积和池化层可能导致关键信息丢失导致模型对小目标检测不利的问题, 嵌入快速特征提取层 (FFEL) 初步提取浅层信息. 如图 2(a) 所示, 它由一个 Conv1×1 层、两个 PConv3×3 层和一个 Conv3×3 组成, 并且加入了残差连接. 其中, Conv1×1

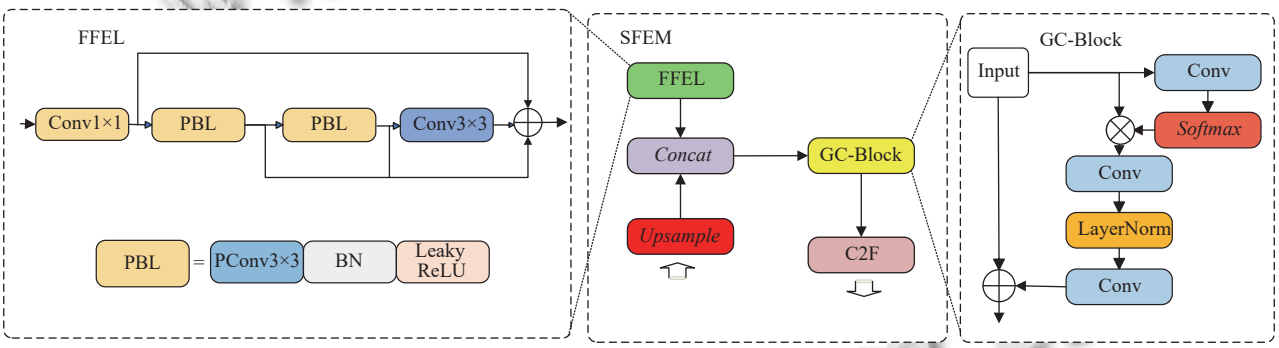
层负责初步特征提取, 随后的 PConv3×3 层则通过减少计算冗余和内存访问来提升效率. Conv3×3 层负责进一步的特征提取, 残差连接增强了输出特征的维度, 减少有用特征的丢失, 同时缓解网络加深导致的特征衰减问题, 有助于保持网络性能并提高检测精度.

全局上下文块 (GC-Block) 结合简化非局部 (SNL) 模块和挤压激励 (SE) 模块的优点, 可以提供长距离依赖信息的有效建模, 同时保持轻量级的计算量, 通过使用瓶颈结构的转换模块, 从而将参数量大幅降低. 通过 GC-Block 实现对全局上下文信息的有效捕捉并对特征进行重校准, 增强对小目标特征的捕获能力, 并通过注意力机制的重分配进一步提升了网络对小尺度目标的敏感度, 如式 (1) 所示, GC-Block 的详细体系结构如图 2(c) 所示.

$$Z_i = X_i + W_{v2} ReLU\left(LN\left(W_{v1}\sum_{j=1}^{Np}\varphi_j\right)\right) \tag{1}$$

$$\varphi_j = \sum_{j=1}^{Np}\frac{e^{W_k X_j}}{\sum_{m=1}^{Np} e^{W_k X_m}} X_j \tag{2}$$

在 GC-Block 中的输入特征图 X_i, 首先经过一个卷积层 W_k 提取关键信息以计算全局注意力权重. 这些权重通过 W_k 与特征 X_j 的线性变换和随后的 *Softmax* 正规化获得, 反映了每个特征在全局上下文中的重要性. 接着, 输入特征图 X_i 通过这些权重进行加权, 强调了重要特征. 加权特征图经由第 2 个卷积层 W_{v1} 进一步变换, 并通过层归一化和 *ReLU* 激活函数处理, 以增强特征的非线性表示和训练稳定性. 最终, 通过与原始输入特征图 X_i 相加, 结合了局部细节和全局上下文信息, 并通过第 3 个卷积层 W_{v2} 的最终处理, 形成了融合后的输出特征图 Z_i, 为网络的下一步操作或作为最终输出提供了综合丰富的特征表达.



(a) FFEL (b) SFEM (c) GC-Block

图 2 SFEM 详细结构

1.3 C2F_DCN

在无人机目标检测领域, 面对多尺度及复杂背景小目标的检测任务, 常规卷积神经网络由于固定的感受野和空间几何结构, 在下采样过程中易于引起信息丢失, 尤其是当目标尺寸小或存在变形时. 为此, 在 YOLOv8 骨干网络中嵌入 C2F_DCN 模块, 以增强网络对变化目标形态的适应性. 该模块结合了可变形卷积网络 (deformable convolutional networks v2, DCNv2), 它能够学习每个卷积点的空间偏移量, 使得卷积核的感受野动态适应目标的实际形态, 缓解传统卷积结构的局限性.

如图 3 所示, 每个 C2F_DCN 模块包含 2 个 DCNv2 和 n 个 BottleNeck. C2F_DCN 模块通过 1×1 卷积调整输入特征的通道数, 并使用 Split 操作来划分特征, 不仅降低了参数的数量, 而且通过增强梯度流使得网络结构更加丰富, 有效提取更多样化和多尺度的特征, 适用于无人机目标检测中的多尺度目标.

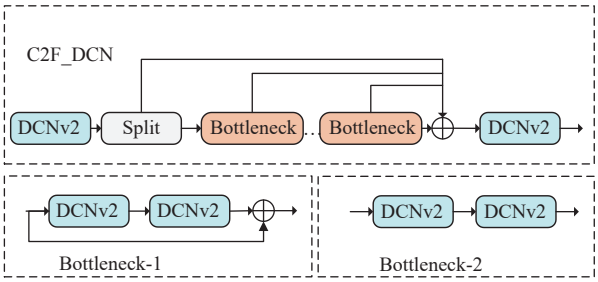


图 3 C2F_DCN 结构

根据是否需要残差连接, 将 Bottleneck 结构分为 Bottleneck-1 结构和 Bottleneck-2 结构. 在 Bottleneck-1

中, 输入输出之间采用残差连接, 用于优化骨干网络, 可以缓解梯度爆炸问题. Bottleneck-2 采用串行连接的方式进行信息传输, 由于特征融合结构作用在颈部, 直接输出可以在更大程度上保留融合后的特征信息. 由于特征融合结构作用在颈部, 直接输出可以在更大程度上保留融合后的特征信息. Backbone 中使用了 3 个 C2F_DCN 模块, 分别用于主干特征提取网络的第 4 层、第 6 层和第 8 层.

综上, C2F_DCN 模块通过集成 DCNv2 及其可变形卷积操作和 Bottleneck 单元, 以处理多尺度和几何变形目标检测任务, 促进特征的有效表示.

1.4 ASPPF

在 YOLOv8 的架构中, 颈部采用的 SPPF (spatial pyramid pooling fast) 结构为多尺度特征提取和目标检测性能的增强提供了基础. 该模块通过连续的最大池化提取特征, 加快了处理速度. 然而, 在无人机航拍的应用场景中, 由于其图像特征的复杂性和小目标的存在, 原模型会导致部分空间信息的丢失, 同时利用最大池化会将噪声当做显著特征保留, 导致模型产生误检. 因此, 本文对 SPPF 模块进行改进, 改进后的结构如图 4 所示. ASPPF 通过以下 4 个步骤实现对无人机航拍图像中目标的高效准确检测.

首先, 特征降维与尺度分解优化: 输入特征图通过一个 1×1 的卷积层进行初步降维处理, 将通道数减半, 以便减轻计算负担同时保留重要特征信息.

$$F' = Conv_{1\times1}(F) \tag{3}$$

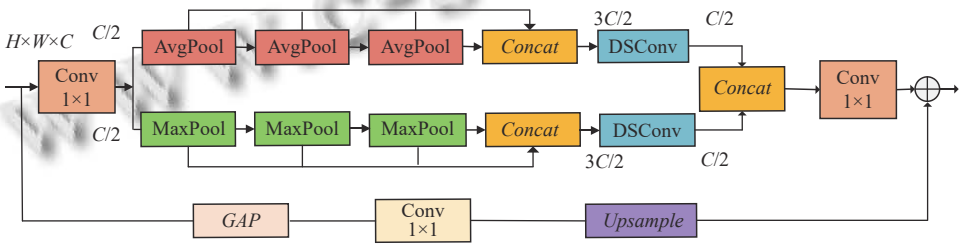


图 4 ASPPF 结构

其次, 双重金字塔池化. 平均池化分支: 对降维后的特征图进行 3 个不同尺度的平均池化操作, 以捕获背景和上下文信息. 最大池化分支: 同时对相同的降维特征图进行 3 个不同尺度的最大池化操作, 以突出显著特征.

$$F_{\text{avg}} = Concat\left(AP_{s_1}(F'), AP_{s_2}(F'), AP_{s_3}(F')\right) \tag{4}$$

$$F_{\max} = Concat(MP_{s_1}(F'), MP_{s_2}(F'), MP_{s_3}(F')) \tag{5}$$

其中, AP 代表平均池化, MP 代表最大池化, s_i (i=3, 5, 7) 代表不同的池化核.

然后, 深度可分离卷积与特征融合. 对两个池化分支的输出分别应用深度可分离卷积, 这种卷积可以进一步提取特征的同时显著降低参数数量. 接着将两个分支的深度可分离卷积输出通过 *Concat* 操作合并, 再通过一个 1×1 卷积层进行融合, 以获得丰富的特征表示.

$$F_{\text{merged}} = Conv_{1\times1}(Concat(DSConv(F_{\text{avg}}, F_{\max})) \tag{6}$$

最后, 全局与局部特征的特征融合. 经过全局平均池化, 1×1 卷积、批量归一化和非线性处理的高级特征生成全局上下文, 然后将这些全局上下文特征通过上采样与低级特征相结合. 从而实现高阶特征对低阶特征图的补充.

$$F_g = Upsample(Conv_{1\times1}(GAP(F))) \tag{7}$$

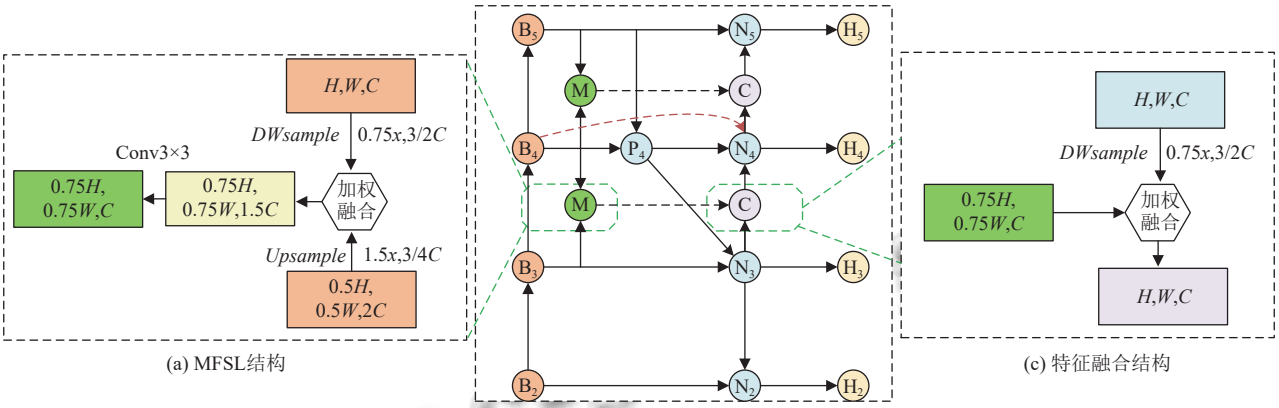
$$F_{\text{final}} = F_{\text{merged}} + F_g \tag{8}$$

ASPPF 优化了无人机航拍图像中微小目标的识别. 池化层与 1×1 卷积的融合不仅锚定了高层次的语义信息, 而且通过上采样与低层特征的结合, 减少空间细节信息丢失. 这种特征融合策略显著提高对小目标的检测灵敏度, 抑制复杂背景中的干扰, 降低误检与漏检的风险, 确保模型的精确性.

1.5 中尺度特征合成层

在 PAN-FPN 架构中, 通过将 PAN 与 FPN 相结合, 构建了一个自上而下及自下而上的网络结构. 这种结构通过特征融合有效地实现了浅层位置信息与深层语义信息的互补. 然而, 由于主干网络中每一层卷积步长为 2, 导致输出特征图的尺寸仅为输入层大小的 1/4. 这种显著的尺度降低会导致相邻层间特征过渡不够平滑, 这在无人机航拍图像的目标特征提取中尤为不利. 针对此问题, 提出了一种新的特征融合策略——中尺度

特征合成层 (mesoscale feature synthesis layer, MFSL), 如图 5(a) 所示. 该层旨在解决主干网络中由于较大步长导致的特征图尺寸快速缩减问题. 具体来说, 对于主干网络中的输出特征图, 通过下采样或上采样操作, 将输出特征图调整至中间特征图的统一尺寸. 随后, 通过平均加权融合操作将两层特征图融合, 避免 *Concat* 操作导致的通道数倍增问题, 并通过 3×3 卷积层进一步提取合成尺度特征.



(a) MFSL结构 (b) SPAN-FPN结构 (c) 特征融合结构

图 5 SPAN-FPN 详细结构

为提升小型无人机目标的检测能力, SFE-YOLO 引入跳跃连接 (skip connections, SK), 将原始输入与同级输出连接, 以丰富深层特征的语义信息. 为充分利用浅层的空间位置信息, 在颈部网络中增加特征融合层, 调整特征图尺度, 并与 MFSL 提取的中尺度特征信息进行加权融合, 确保了特征融合的高效性和动态性.

最后, 添加一层更适合小目标检测的尺度 H_2, 分辨率为 160×160 像素, 相当于在骨干网络中进行了仅两次下采样操作, 包含了更丰富的底层特征信息. H_2 检测头与原检测头结合使用, 可以有效降低尺度方差带来的负面影响, 提高模型对小目标的检测能力. 虽然增加检测头带来了额外的计算量和内存开销, 但显著提升了对小目标的检测性能.

1.6 SimAM 注意力机制

SFE-YOLO 结合 SimAM, 进一步提升无人机目标检测的性能. 如图 6 所示, 该模块借助一个无需参数的能量函数, 自动调节注意力分布, 从而有效地捕获通道和空间层面的关键特征, 并消除了手动调节超参数的需要. 在颈部末尾添加 SimAM 增强 YOLOv8 算法在处理无人机拍摄图像时的目标检测能力, 提高检测的准确性. SimAM 源于神经科学理论, 并基于能量函数提取基本特征. 每个神经元的能量函数如下:

$$e_t = \frac{4\left(\hat{\sigma}^2+\lambda\right)}{(t-\hat{u})^2+2\hat{\sigma}^2+2\lambda} \tag{9}$$

$$\hat{u} = \frac{1}{M}\sum_{i=1}^{M} x_i \tag{10}$$

$$\hat{\sigma}^2 = \frac{1}{M}\sum_{i=1}^{M}(x_i-\hat{\mu})^2 \tag{11}$$

其中, λ是正则项, M 是神经元的个数, $M=H\times W$, i 表示神经元的索引号, t 表示目标神经元, $\hat{u}$和$\hat{\sigma}^2$表示所有神经元在单个通道上的均值和方差. e_t的高值指示了对应位置的特征对于目标的识别非常重要, 应该被增强. $\hat{u}$表示所有M个特征点的平均响应. $\hat{\sigma}^2$是输入特征图的方差, 衡量特征点响应的变异度. 最后, 按照注意力机制的定义, 对特征进行增强处理, SimAM 注意模块可以描述如式 (8) 所示, 其中 E 组合了e_t^{-1}在通道和空间维度上的关注度. 使用 *Sigmoid* 函数是为了避免权重值过大.

$$\tilde{X} = Sigmoid\left(\frac{1}{E}\right)\odot X \tag{12}$$

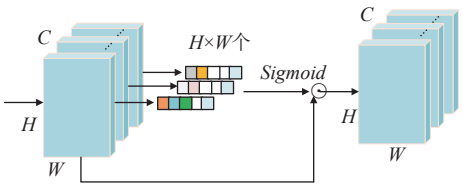


图 6 SimAM 注意力结构

2 实验分析

2.1 实验数据集

VisDrone2019 数据集[17]是主流的无人机航拍数

据集之一, 由天津大学机器学习和数据挖掘团队 AISKYEYE 收集. VisDrone2019 包含 6 471 个训练图像, 548 个验证图像和 1 610 个测试图像. 这个数据集有 10 类, 包括行人、人、自行车、汽车、货车、三轮车、遮阳三轮车、公共汽车、卡车和摩托车. 其中, 处于站立或行走姿态的人类图像分类为“行人”, 其他姿态的分类为“人”.

PASCAL VOC2012 数据集[18]包含来自各种场景的图像, 例如室内环境、城市街景和自然景观. 共有 20 个类别的物体, 包括人、动物 (如猫、狗)、车辆 (如汽车、自行车) 和室内物品 (如瓶子、椅子), 包含数 17 125 张图像, 每张图像都经过了人工标注, 确保了标注的质量和准确性.

2.2 实验环境和参数

实验主要以 YOLOv8n 网络作为基本对照, 具体实验环境如表 1 所示. 网络训练使用 Adam 优化器, 初始学习率为 0.01, 权重值衰减为 0.000 5, 批大小 (batch_size) 为 16, 批次 (epoch) 为 300. 图片输入尺寸为 640×640. 此外, 使用默认的数据增强方法, 并且所有消融实验、对比实验均采用相同设置, 无额外的训练.

表 1 实验硬件环境

参数	实验环境
CPU	AMD Ryzen 7 2700X
GPU	RTX 3060 (12 GB)
操作系统	Ubuntu 20.04
Python	3.9
深度学习框架	PyTorch 1.8.1

2.3 评价指标

本文从精确率 (precision, P)、召回率 (recall, R)、平均精确率均值 (mean average precision, mAP)、模型参数量 (parameter, Params)、浮点数运算量 (floating point operations, $FLOPs$) 这 5 个方面来衡量模型的检测性能. 其中, mAP 用于网络模型评价的整体性能, mAP@0.5 代表 IoU 阈值为 0.5 时的平均 AP (average precision), $FLOPs$ 指每秒浮点运算次数, 理解为计算速度, 可以用来衡量硬件的性能, 相关公式如下:

$$P=\frac{TP}{TP+FP} \tag{13}$$

$$R=\frac{TP}{TP+FN} \tag{14}$$

$$AP=\int_0^1 P(R)\mathrm{d}R \tag{15}$$

$$mAP=\frac{\sum_{i=1}^{k}AP_i}{k} \tag{16}$$

$$FLOPs=C_i\times K^2\times C_0\times W\times H \tag{17}$$

其中, TP 表示准确判定正样本; FP 表示错误判定为正样本; FN 表示错误判定负样本; K 表示样本数据类别. 其中表示 C_0 输出通道数, C_i 表示输入通道数, W 表示卷积核宽, H 表示卷积核高.

2.4 消融实验

为了验证改进算法的有效性, 采用 VisDrone2019 数据集对基线模型进行消融实验, 结果如表 2 所示.

表 2 采用不同改进策略后的检测结果

方法	C2F_DCN	SFEM	ASPPF	EFFL+SK	SimAM	mAP@0.5 (%)	参数量 (M)	计算量 (G)
YOLOv8n	—	—	—	—	—	26.9	3.16	8.9
A	√	—	—	—	—	27.7	3.31	9.1
B	—	√	—	—	—	28.2	3.54	13.9
C	—	—	√	—	—	27.9	3.12	8.8
D	—	—	—	√	—	27.7	3.34	10.3
E	—	—	—	—	√	27.3	3.16	8.9
F	—	√	√	√	—	28.7	3.87	15.9
G	√	√	√	√	√	30.5	3.96	17.1

注: √表示采用了改进后的策略

相较于基线模型, 改进的模型 F 在实验结果精度上有明显提升, mAP@0.5 提升 3.6%. 模型 A 使用 C2F_DCN 代替 YOLOv8n 主干网络中的部分 C2F, mAP@0.5 提升 0.8%, 可变形卷积的应用扩大网络的接受野, 提高网络对不同形状和位置特征的处理能力. 实验 B 在颈部结构嵌入了针对小目标的浅层特征增强模块 SFEM, 同时结合小目标检测头能够有效检测到更多的小目标. mAP@0.5 提升 1.3%, 但是由于增加一个检测头以及 SFEM 模块, 参数量和浮点数运算也有所增加. 实验 C 改进原有的 SPPF 结构, 融合全局语义与局部细节特征强化了模型对小目标的识别能力, 并利用深度可分离卷积保证参数量和计算量微小增加, 精度提升 1%.

模型 D 通过在颈部增加平滑尺度信息并融合不同尺度的特征, *mAP* 提升了 0.8%, 促进了模型整体性能的优化. 模型 E 引入了 SimAM 注意力机制, 能够增强特征表示的而不增加网络的参数量和计算量, *mAP*@0.5 提升了 0.7%, 说明在颈部增加 SimAM 注意力机制能使模型具有更好的特征提取能力. 模型 F 为改进后的 SFE-YOLO 模型, 模型参数量和浮点数运算相较于基线分别增加了 0.8M 和 8.2G, *mAP*@0.5 提升了 3.5%, 说明改进的模型在不显著增加参数量的情况下, 实现了精度更高的检测效果.

2.5 对比实验

2.5.1 与 YOLOv8n 进行准确度对比

为了证明改进模型对检测性能的提高效果, 将 SFE-YOLO 与基线进行对比实验, 如表 3 所示.

表 3 VisDrone2019 数据集各类别对比结果 (%)

类别	YOLOv8n		Ours	
	精确度	*mAP*@0.5	精确度	*mAP*@0.5
行人	43.2	22.6	50.5	29.4
人	45.5	11.9	49.5	18.7
自行车	21.8	6.05	19.4	7.3
汽车	58.4	66.2	61.7	72.0
货车	34.5	29.1	37	34.0
卡车	30.4	28.8	30.7	32.3
三轮车	20.7	12.1	30.5	14.5
遮阳三轮车	34	15.1	41.6	16.6
公共汽车	57.4	49.9	58.6	51.4
摩托车	37.9	24.3	46.5	28.9

根据表 3 的对比结果, 改进方法与基线相比, 各个类别的 *mAP* 均有提升, 尤其是人、行人和摩托车 3 个类别的 *mAP* 值提升幅度均在 5% 以上, 可见模型在对小目标类别具有较好的检测效果, 同时对卡车等相对较大目标检测效果依然明显, 这表明改进后的模型适用于无人机航拍任务. 然而模型对三轮车和遮阳三轮车等特征点不明显的类别检测效果还是较低, 在未来的研究中, 仍需针对这些特殊类别进行更为深入的优化和调整, 以期进一步提高模型的综合检测能力.

2.5.2 主流模型对比对比实验

为了进一步证明 SFE-YOLO 算法在无人机航拍小目标中的检测效果, 在 VisDrone2019 数据集上将 SFE-YOLO 与经典的主流模型进行对比. 如表 4 所示, 加粗部分为该类别在所有算法中的最优值. 通过表 4 可以得出, SFE-YOLO 算法在 VisDrone2019 数据集上的综合性能优异, 尤其是在行人、人和摩托车等较小物体的检测任务中, *AP* 值分别达到了 29.4%、18.7% 和 28.9%. 此外, 对于汽车等较大物体的检测精度也明显优于其他模型. 尽管与 YOLOv7-tiny 相比提升幅度不大, 但其参数量仅为 YOLOv7-tiny 的 62%, 在综合性能上更显优势. 总之, SFE-YOLO 在无人机航拍检测任务重表现良好, 不仅在较小物体上取得了明显的精度提升, 对较大物体也有明显的优势, 相较于其他常用的检测方法, SFE-YOLO 算法综合检测性能占优.

表 4 不同算法在 VisDrone 测试集上的 *AP* 与 *mAP* 对比 (%)

方法	*AP*50										*mAP*@0.5
	行人	人	自行车	汽车	货车	卡车	三轮车	遮阳三轮车	公共汽车	摩托车	
YOLOv8n	22.6	12.0	6.1	66.3	29.3	29.0	12.1	15.1	50.0	24.2	26.7
Faster R-CNN[1]	21.4	15.6	6.7	51.7	29.5	19.0	13.1	7.7	31.4	20.7	21.7
RetinaNet[2]	13.0	7.9	1.4	45.5	19.9	11.5	6.3	4.2	17.8	11.8	13.9
YOLOv4-tiny[19]	11.4	11.1	3.1	52.5	22.8	20.4	8.2	7.6	41.25	12.45	19.1
YOLOv5s	25.2	16.3	8.7	67.4	31.2	33.6	14.0	13.5	**55.0**	24.1	28.9
YOLOv7-tiny[20]	27.1	18.3	**11.0**	68.1	**35.6**	**34.0**	14.1	14.7	50.5	29.1	30.2
SFE-YOLO	**29.4**	**18.7**	7.3	**72.0**	34.0	32.3	**14.5**	**16.6**	51.4	28.9	**30.5**

注: 加粗字体为最优结果

2.5.3 通用性对比实验

为了验证所提模型的泛化能力和准确性, 使用数据集 PASCLA VOC 2012 数据集上设计模型的通用性对比实验. 实验结果如表 5 所示. SFE-YOLO 与基线相比, 精度提升 1.3%, *mAP* 提升 0.8%, 说明改进的算法具有一定的通用性.

表 5 通用性对比实验

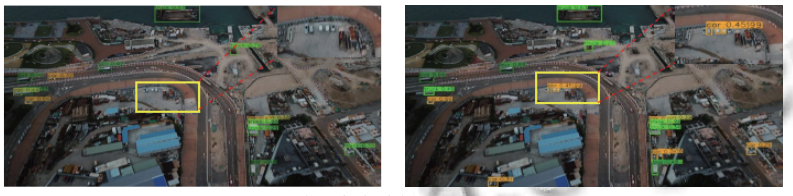
模型	精度	*mAP*@0.5 (%)	参数 (M)	浮点数运算 (G)
YOLOv8n	68.8	66.8	3.0	8.1
Ours	70.1	67.6	3.96	17.1

2.5.4 可视化对比实验

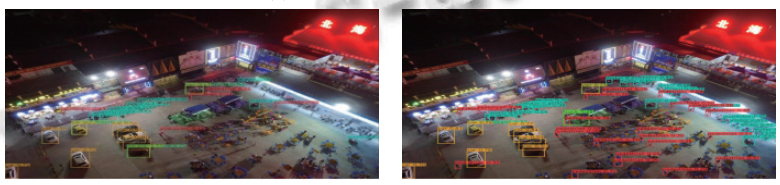
为了验证 SFE-YOLO 算法在实际场景中的检测效果, 使用 VisDrone2019 测试集中白天和夜晚的不同

场景, 包括夜间低光照城市交通场景、夜间小目标密集场景、白天复杂建筑施工场景和白天城市道路交通场景. 最终测试效果如图 7 所示, 左侧为原方法, 右侧为本文的改进方法.

(a) 夜间低光照城市交通场景



(b) 白天复杂建筑工业场景



(c) 夜间小目标密集场景

(d) 城市道路交通场景

图 7 在 VisDrone2019 数据集上的可视化对比图

由图 7(a) 可以发现, 基线算法对公共汽车存在漏检, 而 SFE-YOLO 算法能够准确进行识别. 由图 7(b) 和图 7(c) 可以发现, 在光线较暗的夜间小目标场景和白天复杂工业建筑场景中, 基线算法受到背景噪声的干扰, 从而产生大量漏检, 而 SFE-YOLO 算法通过浅层特征增强模块和颈部特征融合模块, 强化网络对感兴趣区域的关注度, 弱化背景噪声, 有效改善了小目标漏检情况. 在图 7(d) 中, 基线算法将小型建筑物误检为轿车, 而 SFE-YOLO 算法通过改进空间特征金字塔池化模块, 融合多尺度池化特征, 提高模型对小目标的检测灵敏度, 抑制复杂背景中的干扰, 降低误检, 总体而言, 改进后的算法在处理无人机视角下的小尺寸、背景复杂的图像时具有更强的辨识度, 而且有效减少了误检和漏检等现象.

3 结论与展望

本文提出了一种基于 YOLOv8n 的无人机航拍目标检测算法, 通过引入浅层特征增强模块, 结合浅层空间和深层语义信息, 显著提升模型对小目标的检测精度. 此外, 算法在传统特征金字塔结构的基础上, 通过嵌入中尺度特征合成层和跳跃连接, 使得特征过渡更平滑并减少特征丢失. 嵌入 ASPPF 模块, 通过双重金字塔池化技术 (包含平均池化和最大池化), 有效提取和融合特征, 同时降低模型的漏检率.

在效率方面, 采用部分卷积和深度可分离卷积减少了计算复杂度, 确保了高速推理, 同时通过 SimAM 注意力机制, 进一步精确地聚焦于关键信息. 在 VisDrone2019 数据集上的测试结果表明, 该算法在精度方面显著超越基线方法, 特别是在行人等小目标检测上提升明显. 与其他经典算法相比, SFE-YOLO 算法具有较低参数量和较高检测精度, 同时, 有广泛的应用潜力和实际价值. 未来研究将进一步优化模型的计算效率和推理时间, 以便更好地适应边缘设备的部署需求.

参考文献

1 Jiang HZ, Learned-Miller E. Face detection with the faster R-CNN. Proceedings of the 12th IEEE International Conference on Automatic Face & Gesture Recognition. Washington: IEEE, 2017. 650–657.

2 Lin TY, Goyal P, Girshick R, *et al*. Focal loss for dense object detection. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2999–3007.

3 Redmon J, Divvala S, Girshick R, *et al*. You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779–788.

4 Yin GX, Yu M, Wang M, *et al*. Research on highway vehicle detection based on faster R-CNN and domain adaptation. Applied Intelligence, 2022, 52(4): 3483–3498. [doi: 10.1007/s10489-021-02552-7]

5 欧阳权, 张怡, 马延, 等. 基于深度学习的无人机航拍目标检测与跟踪方法综述. 电光与控制, 2024, 31(3): 1–7.

6 Zhao T, Liu JY, Duan ZQ. UAV target detection based on RetinaNet. Proceedings of the 2019 Chinese Control and Decision Conference. Nanchang: IEEE, 2019. 3342–3346. [doi: 10.1109/CCDC.2019.8832517]

7 黄海生, 饶雪峰. 面向无人机航拍场景的轻量化目标检测. 计算机系统应用, 2022, 31(12): 159–168. [doi: 10.15888/j.cnki.csa.008866]

8 Tan MX, Pang RM, Le QV. EfficientDet: Scalable and efficient object detection. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 10778–10787. [doi: 10.1109/CVPR42600.2020.01079]

9 Wang ZY, Liu ZW, Xu G, *et al*. Object detection in UAV aerial images based on improved YOLOv7-tiny. Proceedings of the 4th International Conference on Computer Vision, Image and Deep Learning. Zhuhai: IEEE, 2023. 370–374. [doi: 10.1109/CVIDL58838.2023.10166362]

10 Zhang ZX. Drone-YOLO: An efficient neural network method for target detection in drone images. Drones, 2023, 7(8): 526. [doi: 10.3390/drones7080526]

11 Du BW, Huang YC, Chen JX, *et al*. Adaptive sparse convolutional networks with global context enhancement for faster object detection on drone images. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 13435–13444. [doi: 10.1109/CVPR52729.2023.01291]

12 Wang RX, Shivanna R, Cheng D, *et al*. DCN V2: Improved deep & cross network and practical lessons for Web-scale learning to rank systems. Proceedings of Web Conference 2021. 1785–1797.

13 He KM, Zhang XY, Ren SQ, *et al*. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904–1916. [doi: 10.1109/TPAMI.2015.2389824]

14 Yang LX, Zhang RY, Li LD, *et al*. SimAM: A simple, parameter-free attention module for convolutional neural networks. Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021. 11863–11874.

15 Liu S, Qi L, Qin HF, *et al*. Path aggregation network for instance segmentation. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8759–8768.

16 Chen JR, Kao SH, He H, *et al*. Run, don't walk: Chasing higher FLOPS for faster neural networks. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 12021–12031.

17 Du D, Zhu P, Wen L, *et al*. VisDrone-DET2019: The vision meets drone object detection in image challenge results. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshops. 2019.

18 Everingham M, Van Gool L, Williams CKI, *et al*. The pascal visual object classes (VOC) challenge. International Journal of Computer Vision, 2010, 88(2): 303–338. [doi: 10.1007/s11263-009-0275-4]

19 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934, 2020.

20 Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 7464–7475.

(校对责编: 张重毅)