基于全局空间感知的轻量级无人机航拍图像目标检测

吴家骏 朱松豪
(南京邮电大学自动化学院 南京 210023)

摘 要 无人机航拍图像目标检测因目标尺度小(面积<1%或<32×32 像素)、背景复杂及实时性要求而极具挑战性。传统方法在特征提取、特征融合及定位优化方面存在精度不足或计算复杂度高的局限性。本文提出基于全局空间感知的轻量级无人机航拍图像目标检测网络(Lightweight Global-Spatial Perception Network, LGSP-Y10),通过协同优化实现高精度与实时性。首先,提出跨通道-分组空间注意力(Cross-Channel Group-Wise Spatial Attention, CGSA)模块,通过动态跨通道上下文聚合和分组空间注意力增强小目标稀疏纹理特征提取能力,参数量减少 10.7%。其次,构建跨尺度-多注意力协同双卷积瓶颈(Cross-Scale Multi-Attention Dual-Convolution Bottleneck, CMDCB)模块,通过动态通道剪枝(7%剪枝率)和多注意力协同融合优化多尺度特征。最后,提出归一化交并比-距离损失(L_{NIoU}),通过融合尺度归一化交并比损失与高效 Wasserstein 距离损失,有效提升了小目标的定位精度。大量实验结果表明,相较于现有方法,LGSP-Y10 在 VisDrone2019 和 TinyPerson 数据集上实现了最佳性能,mAP 从 31.41%提升至 33.60%,提升了 2.19 个百分点,且在 NVIDIA GeForce RTX 4060 Ti 上达 115.73 FPS,在 NVIDIA Jetson Orin Nano 上达 27.45 FPS,满足无人机实时检测需求。

关键词 无人机航拍图像目标检测;跨通道-分组空间注意力;跨尺度-多注意力协同双卷积瓶颈;归一化交并比-距离损失
中图法分类号 TP391 **DOI 号** 10.11897/SP.J.1016.2026.01620

Global-Spatial Perception For Lightweight Unmanned Drone Aerial Image Small Object Detection

WU Jia-Jun ZHU Song-Hao
(School of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210023)

Abstract Unmanned drone aerial image small object detection is highly challenging due to the small object scale (area<1% or<32×32 pixels), complex background, and real-time requirement. Traditional methods suffer from limitations in feature extraction, feature fusion, and localization optimization, resulting in suboptimal detection accuracy or high computational complexity. This paper proposes a lightweight unmanned drone aerial image small object detection based on global-spatial perception (Lightweight Global-Spatial Perception Network, LGSP-Y10), which achieves high precision and real-time performance through collaborative optimization. Firstly, a Cross-Channel-Group-Wise Spatial Attention module (CGSA) is proposed, which enhances the small object sparse texture feature extraction capability through dynamic cross-channel contextual information aggregation and group-wise spatial attention, with a 10.7% parameter reduction. Secondly, a Cross-Scale Multi-Attention Dual-Convolution Bottleneck (CMDCB) is proposed, which optimizes multi-scale features through dynamic channel pruning (7% pruning rate) and multi-attention collaborative fusion. Finally, a Normalized Intersection-over-

收稿日期:2025-07-15;在线发布日期:2026-03-13。本课题得到国家自然科学基金(No. 52405065)资助。**吴家骏**,硕士研究生,中国计算机学会(CCF)会员,主要研究领域为图像处理、人工智能。E-mail:1224056235@njupt.edu.cn。**朱松豪(通信作者)**,博士,副教授,主要研究领域为图像处理、智能感知。E-mail:zhush@njupt.edu.cn。

Union-Distance Loss (L_{NIoU}) is proposed, which through scale-normalized Intersection-over-Union loss with efficient Wasserstein distance loss. Extensive experimental results demonstrate that, compared to existing methods, LGSP-Y10 achieves state-of-the-art (SOTA) performance on the VisDrone2019 and TinyPerson datasets. It improves mAP by 6.97% (31.41%→33.60%), reduces parameters by 18.11% (2.71M→2.21M), while attaining 115.73 FPS on an NVIDIA GeForce RTX 4060 Ti and 27.45 FPS on an NVIDIA Jetson Orin Nano, meeting the real-time requirements for UAV applications.

Keywords　unmanned drone aerial image small object detection; cross-channel group-wise spatial attention; cross-scale multi-attention coordinate dual-convolution bottleneck; normalized intersection over union-distance loss

1　引　言

近年来,无人机技术因其灵活性和高效性,在交通监控、灾害救援、农业监测、城市规划等领域得到广泛应用,尤其在需要高精度目标检测的任务中表现出巨大潜力。例如,在农业领域,Pretto 等人利用无人机的定向干预功能,开发了多光谱感知算法的空地系统,实现了对杂草的准确定位与精细分类[1];在城市规划领域,Hague 等人设计了一种小型无人机,用于城市街景的三维建模与环境感知,实现目标的自主跟踪[2]。然而,无人机航拍图像目标检测仍然面临巨大挑战,这主要源于图像覆盖范围广、背景复杂、目标尺度小、像素占比低、光照条件变化等问题[3-4]。此外,无人机端侧硬件(如 NVIDIA Jetson Orin Nano)的计算资源和功耗限制对模型的实时性和轻量化提出了更高要求。因此,设计一种兼顾高精度和低延时的轻量级小目标检测模型具有重要的研究意义和应用价值。

传统目标检测方法在无人机航拍小目标检测领域存在显著局限性。基于卷积神经网络的模型依赖深层网络提取特征,但对小目标的稀疏纹理特征捕获不足,导致检测精度较低,尤其在复杂背景下易出现漏检。近年来,基于 Transformer 的检测模型通过全局建模提升特征表征能力,但其高计算复杂度(FLOPs 高达数百 G)使其难以在资源受限的无人机平台实现实时推理[5]。注意力机制的引入增强了特征表征能力,但传统通道注意力或空间注意力对小目标长程语义信息建模不足,且参数量较大,不适合轻量化需求[6]。

针对小目标区域像素占比小引起的传统注意力易受复杂背景干扰、特征金字塔信息传递过程中高频细节易丢失、错误目标边界框回归引起检测性能降低三重挑战,本文提出基于全局空间感知的轻量级无人机航拍目标检测网络(Lightweight Global-Spatial Perception Network,LGSP-Y10),通过协同优化实现实时小目标精准检测,主要贡献如下:

(1)针对无人机航拍图像中小目标稀疏纹理特征难以捕获、易受复杂背景干扰的问题,本文提出跨通道-分组空间注意力模块,通过动态跨通道上下文聚合和分组空间注意力机制,增强小目标显著性特征提取,同时抑制背景噪声,为后续特征融合提供高质量输入。

(2)为解决传统特征金字塔网络在低分辨率小目标空间信息丢失和参数量过大的问题,本文提出跨尺度多注意力协同双卷积瓶颈模块。通过动态通道剪枝(7%剪枝率)和多注意力协同融合机制,优化多尺度特征融合,增强小目标空间细节感知能力。该模块有效平衡了检测精度与计算效率,适合包括无人机平台等在内的资源受限场景。

(3)针对传统交并比损失导致小目标边界框偏移敏感、归一化 Wasserstein 距离损失计算复杂度高的问题,本文提出归一化交并比-距离损失。该损失通过尺度归一化缓解小目标偏移,并通过坐标归一化将 Wasserstein 距离计算复杂度从 $O(n^2)$ 降至 $O(n)$,训练 FLOPs 降低 20%。此外,利用动态加权机制根据目标尺寸自适应调整 L_{IoU} 和 L_{NWD} 权重,优化密集小目标定位,满足无人机实时检测需求。

2　相关工作

目标检测方法广泛应用于多个领域,但传统方法在处理多目标或重叠目标时鲁棒性较差。相比之下,基于深度学习的检测方法显著提升了复杂环境下的性能,尤其在资源受限的无人机平台上成为研

究热点，研究者们提出了一系列针对无人机小目标检测的创新方法，进一步提升了检测精度和效率。

注意力机制广泛应用于目标检测任务，以增强关键特征的提取能力。Quan 等人[7]提出融合远距离注意力机制的多级特征融合策略，提高小目标特征提取能力。Yang 等人[8]通过引入注意力机制的水平集模型，提高滑坡区域目标检测能力。Cheng 等人[9]引入区域交叉自注意力机制，提升小目标特征表达能力。

在多尺度特征融合方面，特征金字塔成为研究重点。Yu 等人[10]提出基于改进的特征金字塔网络，增强小水域特征提取能力法。Pan 等人[11]提出基于 Transformer 的多尺度特征融合检测网络，通过层次化窗口机制增强多尺度特征表征能力。

IoU 损失及其变体[12]在小目标检测中对边界框偏移高度敏感，尤其在目标像素占比低时易导致训练不稳定，因此针对小目标检测的损失函数设计成为研究热点。Zhu 等人[13]提出归一化 Wasserstein 距离损失函数，利用目标定位与语义标签一致性损失、轻量级目标性损失、低秩定位约束三种损失函数进行训练，以优化小目标检测精度。Cai 等人[14]提出角点与前景结合的损失函数，优化边界框回归过程。

在模型轻量化优化上，Li 等人[15]提出 EMFE-YOLO 框架，通过引入高效多级特征增强模块，利用动态卷积替换传统卷积，实现参数量减少。Chen 等人[16]开发 YOLO-LE 算法，整合动态通道注意力变体模块和动态下采样策略，减少参数量，提高推理速度。

尽管上述研究在提升小目标检测性能方面取得一定成效，但仍存在以下问题：(1)当前方法大多专注于检测精度的提升，但对计算效率的优化较少；(2)一些模型在引入注意力机制或复杂特征融合模块后，计算复杂度大幅提高。(3)目前方法虽在参数压缩和推理速度上取得进展，却多局限于模块级替换，缺乏多组件间协同，难以全面平衡小目标检测精度与实时性。因此，设计一种既能提升检测精度，又能优化计算效率的轻量级目标检测模型仍是一个重要研究方向。本文提出的基于全局空间感知的轻量级无人机航拍图像目标检测模型通过整合跨通道-分组空间注意力模块、跨尺度自适应上下文特征金字塔网络模块和归一化交并比-距离损失模块，不仅实现了对小目标的高效检测，还显著降低了模型的计算复杂度。

3 本文方法

本文提出一种基于全局空间感知的轻量级无人机航拍图像目标检测网络，针对无人机航拍图像小尺度目标(面积<1%或<32×32 像素)、背景复杂、密集遮挡的挑战，通过协同优化特征提取、特征融合和目标定位实现小目标检测的高精度与实时性。与图 1(a)所示的 YOLOv10 [17]相比，图 1(b)所示的本文提出的 LGSP-Y10 网络引入三个关键模块：CGSA、CMDCB 和 L_{NIoU}。这三个引入的模块分别对应图 1(b)中的多尺度特征提取层(浅层 P3、中层 P4、深层 P5，分别表示下采样 8 倍、16 倍、32 倍的特征图)、特征融合层(FPN 结构)和损失计算模块，协同工作以提升小目标检测性能并降低计算开销。以下详细阐述各模块的设计思路与实现过程。

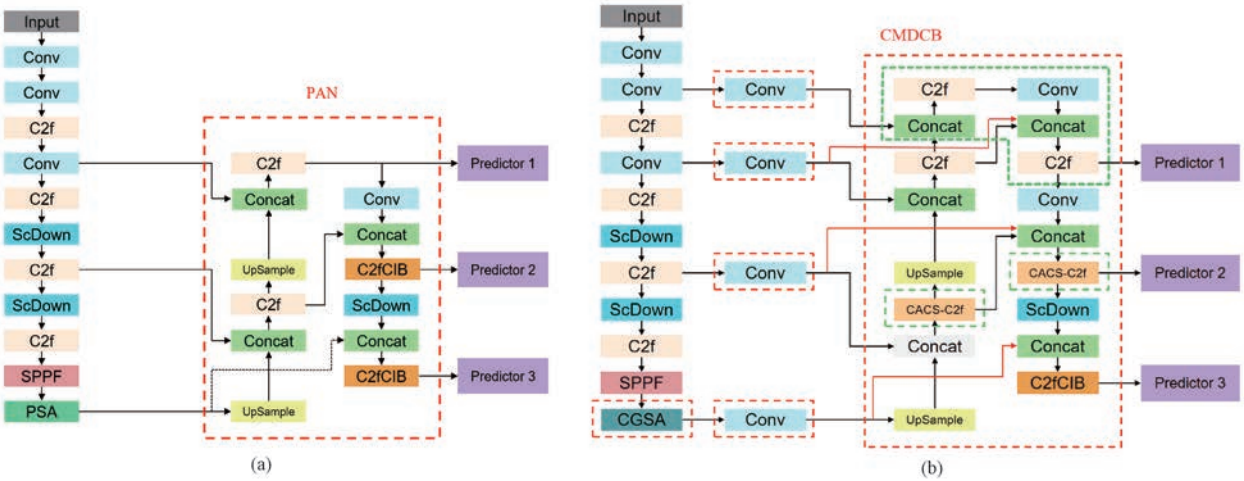


图 1 模型结构图((a)为 YOLOv10 模型结构图；(b)为本文提出的 LGSP-Y10 模型结构示意，包含跨通道-分组空间注意力模块(CGSA，特征提取层 P3、P4、P5，分别对应下采样 8 倍、16 倍、32 倍的特征图，用于捕获不同尺度目标)、跨尺度多注意力协同双卷积瓶颈(CMDCB，特征融合层)和 L_{NIoU}(损失计算))

3.1　基于跨通道—分组空间注意力的特征提取

在无人机航拍图像目标检测中,YOLOv10 通过像素级空间注意力(Pixel-Level Spatial Attention, PSA)增强目标区域关注能力,但其对小目标稀疏纹理特征提取能力不足,尤其在高空视角下复杂背景干扰显著。传统注意力机制依赖静态通道或空间加权,难以有效捕捉航拍图像目标长程语义关系与稀疏特征,且计算复杂度高,不适合轻量化需求。为此,本文提出如图 2 所示的 CGSA。针对航拍图像中小尺度目标、背景复杂、密集遮挡的特性,CGSA 设计了动态跨通道上下文聚合、分组空间注意力、动态跨组特征交互机制,显著提升小目标特征提取能力,同时降低参数量。

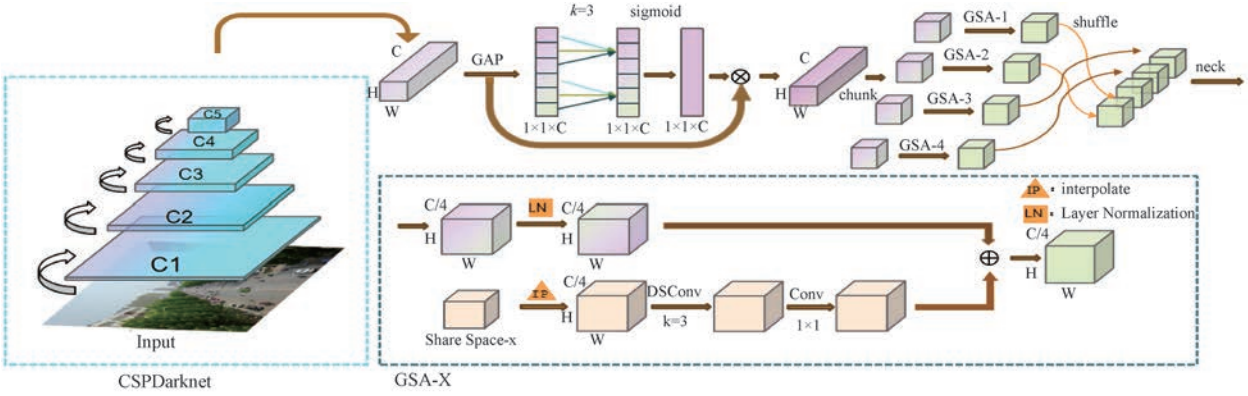


图 2　本文提出的跨通道-分组空间注意力模块的结构示意图

提出的跨通道—分组空间注意力模块包含以下三个创新。

(1)动态跨通道上下文聚合。为精准捕捉航拍图像目标的稀疏纹理特征,CGSA 首先通过全局平均池化与一维卷积生成通道特征表征,并引入动态通道选择策略增强与小目标相关的高频特征通道。具体实现过程可用如下公式表示:

$$z_c=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}x_c(i,j) \tag{1}$$

其中,z_c 表示输入特征第 c 个通道的特征表征,$x_c(i,j)$表示输入特征第 c 个通道的第(i,j)个位置的值;然后,利用一维卷积计算通道注意力;最后,对输入特征 $\boldsymbol{X}$ 的每个通道进行加权:

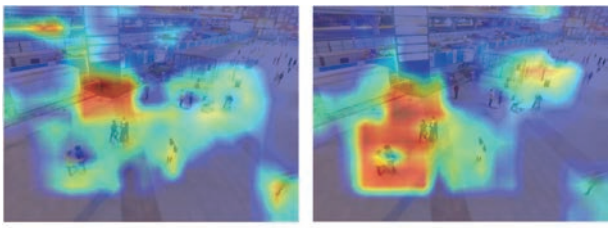
$$\boldsymbol{X}'=\boldsymbol{X}\oplus\alpha_c,\alpha_c=\delta(Conv_{1D}(z_c)) \tag{2}$$

其中,δ 表示 Sigmoid 激活函数,Conv_{1D} 表示一维卷积,$\boldsymbol{X}'$表示加权的通道特征表征。与 SE 的静态通道加权不同,SE 依赖静态全局平均池化,仅生成固定通道权重(忽略目标尺度变异,导致背景噪声易干扰小目标高频特征)。相比之下,CGSA 引入目标尺度感知分支,首先,利用如下式所示的特征图的统计方差

$$\sigma_c=\sqrt{\frac{1}{H\times W}\sum(x_c(i,j)-z_c)^2} \tag{3}$$

估计通道纹理稀疏度;其次,动态计算目标尺度因子,自适应调整权重,优先提高高频通道权重,抑制背景低频通道权重。这种自适应机制能够根据航拍图像目标尺度实时调整通道权重,显著抑制背景噪声干扰。

本文采用以下过程实现小目标稀疏纹理提取:首先,采用动态 σ_c 估计(式(3))自适应提升高频通道权重;其次,优先激活小目标稀疏纹理($\sigma_c>0.05$)相关通道,避免受到传统静态加权低频背景干扰;最后,结合后续分组空间注意力捕捉长程空间依赖,强化稀疏区域语义建模。由图 3 所示的可视化实验结果可知,在小目标场景下,CGSA 的注意力热力图更集中于稀疏纹理区域,而 SE 的注意力均匀分布导致漏检风险增加。



(a) SE　　(b) CGSA

图 3　SE 与 CGSA 的通道注意力热力图对比(小目标区域:CGSA 高频通道激活提升约 20%,背景低频通道抑制约 15%)

(2)分组空间注意力。为解决小目标与背景间的空间关系和长程依赖问题,CGSA 模块引入分组空间注意力机制(GSA)。首先,将加权的通道特征表征 $\boldsymbol{X}'\in R^{C\times H\times W}$ 按通道维度分为 G 组 $\boldsymbol{X}'_1,\boldsymbol{X}'_2,\cdots,\boldsymbol{X}'_G$(每组通道数为 C/G),对于每一组特征 $\boldsymbol{X}'_g$,在空间维度上分别生成查询(Query)、键(Key)和值(Value)矩阵:

$$\boldsymbol{Q}_g=\Phi_q(\boldsymbol{X}'_g),\boldsymbol{K}_g=\Phi_k(\boldsymbol{X}'_g),\boldsymbol{V}_g=\Phi_v(\boldsymbol{X}'_g) \tag{4}$$

其中，Φ_q、Φ_k、Φ_v 分别表示利用 1×1 卷积获得的线性映射，其中查询 $\boldsymbol{Q}$ 表示当前空间位置的特征信息，键 $\boldsymbol{K}$ 表示全局空间中各位置的特征索引，值 $\boldsymbol{V}$ 表示该位置对应的语义信息。通过计算 $\boldsymbol{Q}$ 与 $\boldsymbol{K}$ 的相似度获得注意力权重，并以此加权 $\boldsymbol{V}$，实现空间维度的特征聚合：

$$\boldsymbol{A}_g=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}_g\boldsymbol{K}_g^T}{\sqrt{d}}\right),\boldsymbol{Y}_g=\boldsymbol{A}_g\boldsymbol{V}_g \tag{5}$$

其中，$\boldsymbol{A}_g$ 为第 g 组空间注意力矩阵，d 表示嵌入程度。最终，将所有组输出在通道维度上拼接并通过轻量卷积进行融合，以增强不同组间的特征交互。

与传统全局空间注意力机制直接对全部 C 个通道进行联合计算（计算复杂度为 $O(C^2)$）相比，GSA 通过分组计算显著降低了计算复杂度。对于每组通道数为 C/G 的情况，注意力计算复杂度可降至 $O(C^2/G)$，此时整体浮点运算量（FLOPs）下降约 $1/G$（例如，当 $G=4$ 时，理论计算量和参数量均可减少约 75%）。这种复杂度降低首先来源于每组通道的独立投影减少了乘加次数，其次源于组内注意力计算范围的限制，减少了键—值匹配的搜索域。

实验结果表明，引入 GSA 后，模型参数从 2.71M 降至 2.42M（降低约 10.7%），FLOPs 下降约 8.3%，推理速度得到明显提升，同时保持对小目标特征的高敏感性和空间结构建模能力。这说明 GSA 在平衡精度与计算效率方面具有良好的优势。

（3）动态跨组特征交互机制。为进一步增强不同通道组别间特征的全局交互，CGSA 引入动态跨组特征交互机制。首先，通过通道混洗（Channel Shuffle）操作[18]随机调整各组顺序，实现不同组间的特征交互，提高全局注意力；然后，利用深度可分离卷积[19]融合通道特征：

$$\boldsymbol{Y}=DSConv(S(Concat(\boldsymbol{X}''_1,\boldsymbol{X}''_2,\cdots,\boldsymbol{X}''_G))) \tag{6}$$

其中，S 表示通道混洗操作。

3.2 基于跨尺度—多注意力协同的特征融合

在小目标检测中，高效融合多尺度特征是关键挑战。路径聚合网络（Path Aggregation Network，PAN）[20]通过连接高层特征和低层特征增强特征表征能力，但其对低分辨率小目标的空间信息捕获不足，且参数量较大。为此，本文提出跨尺度-多注意力协同机制，通过动态加权和多注意力协同优化特征融合，同时降低模型复杂度。

提出的跨尺度—多注意力协同机制包含以下三个创新：

（1）引入动态通道剪枝策略

传统的通道剪枝方法通过手动设定阈值，剪掉不重要的通道，从而减少计算量和参数量；而动态通道剪枝策略依据每个通道的重要性自适应调整通道权重，剪除冗余通道，实现更高效的通道压缩。具体地，动态剪枝策略会在每一层的特征图生成后，依据通道的重要性评分（基于特征图的激活度或梯度信息）动态判断可以剪除哪些通道。

在网络结构层面，动态通道剪枝策略通过调整每层特征图的权重，有效减少了冗余通道计算，从而一定程度上降低了网络的参数量和计算复杂度。动态通道剪枝的具体实现过程可表示为

$$\dot{\boldsymbol{F}}_l=\alpha_l\boldsymbol{F}_l+\boldsymbol{F}_{l+1} \tag{7}$$

其中，$\dot{\boldsymbol{F}}_l$ 表示特征映射融合后的特征图，$\boldsymbol{F}_l$ 和 $\boldsymbol{F}_{l+1}$ 分别表示来自第 l 层和第（$l+1$）层的特征映射，α_l 是自适应加权系数。例如，由表 5 所示的实验结果可知，通过动态剪枝，通道数从原始的 C 减少到 C'（其中 $C'<C$），最终参数量减少约 7%；同时，动态通道剪枝策略还显著提升了推理效率。

在动态通道剪枝策略的基础上，L1 正则化进一步对通道权重进行压缩，保留对小目标检测有较大贡献的通道。这两个策略相辅相成，动态通道剪枝策略去除冗余通道，稀疏化对剩余通道进一步细化压缩，从而有效提升模型的计算效率。

（2）引入自适应加权机制，依据特征重要性动态调整特征权重，优化特征融合效果。为进一步降低计算复杂度，采用稀疏特征选择策略，通过 L1 正则化约束权重参数矩阵 $\boldsymbol{W}_l$，优先保留对小目标检测贡献较大的特征通道，抑制冗余通道。具体实现过程可用如下公式表示：

$$\alpha_l=Sigmoid(Conv1\times1(Norm(\boldsymbol{F}_l)))+\lambda\cdot\|\boldsymbol{W}_l\|\_1 \tag{8}$$

其中，Conv1x1 表示 1×1 卷积，Norm 表示归一化操作，λ 表示 L1 正则化系数，$\|\boldsymbol{W}_l\|\_1$ 表示权重参数矩阵的 L1 范数。表 3 所示的实验结果表明，通过 L1 正则化，实现约 15%权重参数的稀疏化（参数从 2.71M 进一步优化至 2.56M），参数量显著减少；同时，稀疏化降低冗余特征的计算开销，FLOPs 额外减少约 5%，推理速度进一步提升（FPS 从 94.62 增至 101.47）。图 4(a)和图 4(c)所示的实验结果表明，基准模型的注意力分布偏向大目标，而图 4(b)和图 4(d)所示的实验结果表明，引入动态加权和稀

疏化策略后,小目标区域的关注度显著提高,验证了稀疏化对检测精度的贡献。

(3)引入融合多注意力协同机制的跨尺度双卷积瓶颈模块[21],增强小目标位置和形态感知能力。基于跨阶段部分瓶颈的双卷积(Cross-Stage Partial Bottleneck with 2 Convolutions, C2f)[22]模块因未将局部空间信息关联,难以捕捉小目标空间细节。为此,本文在 C2f 基础上引入多注意力协同机制,提出如图 4 所示的跨尺度协同注意力双卷积瓶颈模块,通过精确的空间坐标加权提升小目标特征提取能力。

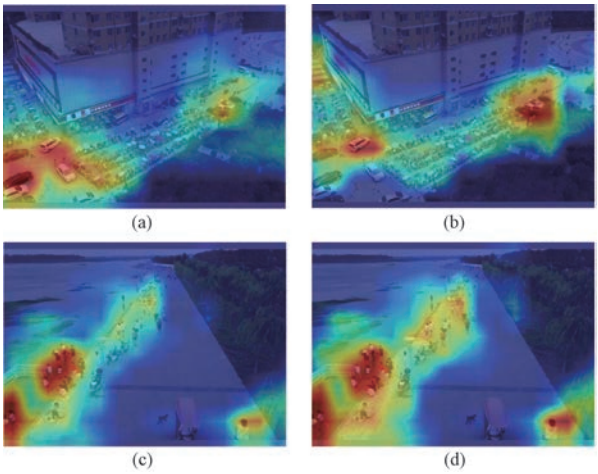


图 4　不同方法获得的注意力分布的热力图((a)和(c)分别表示利用基准模型获得的注意力分布的热力图,(b)和(d)分别表示利用本文提出的动态加权策略获得的注意力分布的热力图)

如图 5 所示,将 C2f 双卷积瓶颈结构改进为融合多注意力协同的跨尺度协同注意力双卷积瓶颈模块结构,有利于每个位置的特征映射都能依据其空间坐标进行相应加权。

与传统通道注意力不同,多注意力协同机制聚焦空间维度,通过空间位置感知精准加权,突出小目标特征。由图 6 所示可知,多注意力协同机制将二维全局池化拆为水平、垂直一维池化,同时捕捉空间位置信息和上下文信息。

多注意力协同机制的实现过程如下所述:①对输入特征 $\boldsymbol{X}\in\boldsymbol{R}^{C\times H\times W}$ 进行水平以及垂直方向的 1 维全局池化,生成特征映射 $\boldsymbol{z}_h\in\boldsymbol{R}^{C\times H\times 1}$ 以及 $\boldsymbol{z}_w\in\boldsymbol{R}^{C\times 1\times W}$:

$$\begin{cases}\boldsymbol{z}_h(h)=\dfrac{1}{W}\displaystyle\sum_{i=0}^{W-1}x_c(h,i)\\ \boldsymbol{z}_w(w)=\dfrac{1}{H}\displaystyle\sum_{j=0}^{H-1}x_c(j,w)\end{cases}\tag{9}$$

其中,$\boldsymbol{z}_h(h)$和 $\boldsymbol{z}_w(w)$表示池化后的特征映射中的一个元素,$x_c(h,i)$和 $x_c(j,w)$表示输入特征映射中特定位置的值。需要注意的是,在进行水平池化时,针对每个高宽位置 H/W,计算该位置上所有高宽方向的特征的平均值。②对 $\boldsymbol{z}_h$ 和 $\boldsymbol{z}_w$ 依次进行拼接、1×1 卷积和非线性化,生成特征图 $\boldsymbol{f}\in\boldsymbol{R}^{C/r\times(H+W)\times 1}$:

$$\boldsymbol{f}=\delta(F_1([\boldsymbol{z}_h,\boldsymbol{z}_w]))\tag{10}$$

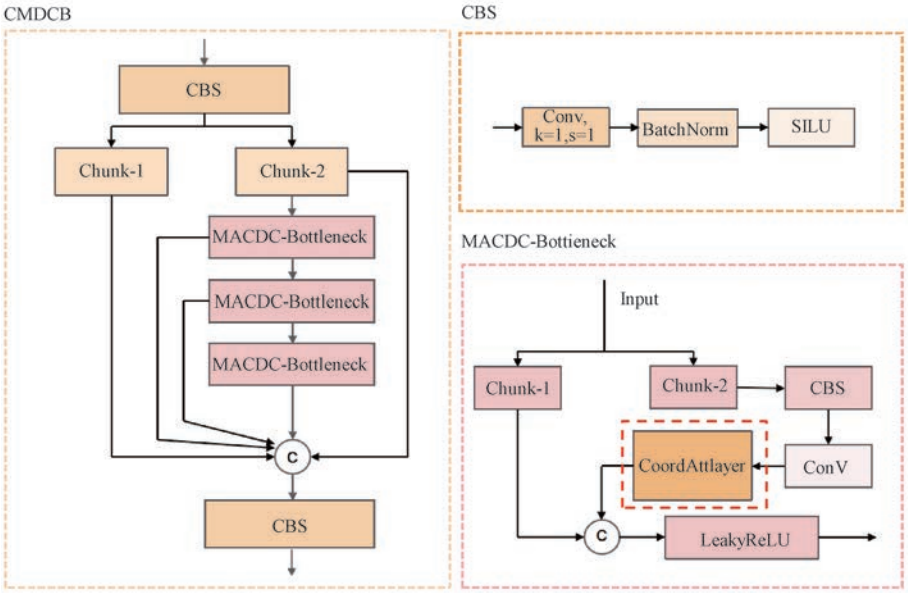


图 5　提出的跨尺度协同注意力双卷积瓶颈模块的结构图

其中,F_1 表示 1×1 卷积,δ 表示 Sigmoid 激活函数。③沿着空间维度,将 $\boldsymbol{f}$ 进行分割为 $\boldsymbol{f}^h\in\boldsymbol{R}^{C/r\times H\times 1}$ 和 $\boldsymbol{f}^w\in\boldsymbol{R}^{C/r\times 1\times W}$,并再次利用 1×1 卷积和非线性化获得注意力向量 $\boldsymbol{g}^h\in\boldsymbol{R}^{C\times H\times 1}$ 和 $\boldsymbol{g}^w\in\boldsymbol{R}^{C\times 1\times W}$。协同注意力的特征映射可用下式表示:

$$y_c(i,j)=x_c(i,j)\times g_c^h(i)\times g_c^w(j)\tag{11}$$

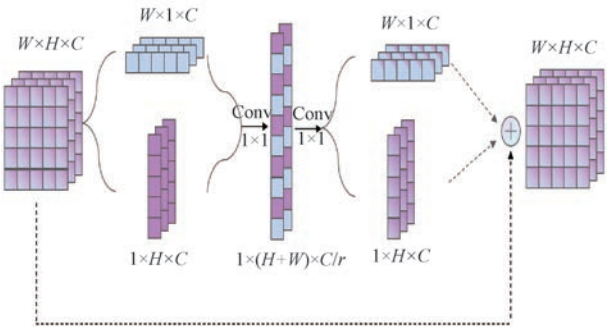


图 6 多注意力协同模块的结构示意图

其中，$y_c(i,j)$为位置(i,j)处的加权特征值，$x_c(i,j)$为输入特征图中处于相同位置的特征值，$g_c^h(i)$和$g_c^w(j)$为水平和垂直方向注意力图在位置(i,j)的加权值。

3.3 归一化的交并比—距离损失函数

在小目标检测中，交并比损失(L_{IoU})通过衡量预测框与真实框的重叠程度引导网络优化目标定位。然而，L_{IoU}对小目标的边界框偏移高度敏感，尤其当目标与背景重叠小或无重叠时，损失值变化剧烈，导致训练和优化困难。归一化 Wasserstein 距离损失(Loss Normalized Wasserstein Distance，L_{NWD})[23]通过建模边界框的概率分布，能有效捕捉小目标与背景的空间关系，但其计算复杂度较高，且与L_{IoU}的融合缺乏自适应性。为此，本文提出归一化的交并比—距离损失函数(L_{NIoU})，通过引入尺度归一化和动态加权机制，优化L_{IoU}和L_{NWD}的融合，显著提升小目标检测精度，同时降低训练计算开销，支持模型的轻量化目标。

归一化的交并比—距离损失的具体内容如下所述。

(1)L_{NIoU}改进了交并比损失(L_{IoU})，通过引入尺度归一化因子，缓解小目标偏移敏感性。传统L_{IoU}直接计算 1-IoU，容易放大小目标的边界框偏差。本文提出如下公式所示的尺度归一化的L_{IoU}：

$$L_{IoU}=1\text{-}IoU+\lambda_s\cdot\frac{1\text{-}IoU}{Smax/Sref} \tag{12}$$

其中的 IoU 为预测框与真实框间的交并比，λ_s为尺度加权系数(通常设为 0.5)，$Smax$为预测框和真实框中面积较大的值，$Sref$为参考面积(设为图像面积的 1%，即小目标阈值)。尺度归一化因子($Smax/Sref$)减小了小目标偏移对损失的过度惩罚，提升了优化稳定性，尤其在小目标重叠较小时效果显著。

(2)进一步优化归一化 Wasserstein 距离损失(L_{NWD})，降低计算复杂度并增强小目标适配性。L_{NWD}将预测框和真实框分别建模为二维高斯分布 Np 和 Ng，并计算分布差异：

$$L_{NWD}=W_2(Norm(Bp)),(Norm(Bg))/\sigma \tag{13}$$

其中，W_2表示二阶 Wasserstein 距离，$Norm(Bp)$和$Norm(Bg)$分别表示归一化的预测框坐标和归一化的真实框坐标，σ表示归一化因子(基于图像尺寸)。传统二维 Wasserstein 距离通过比较高斯分布Bp和Bg的相似度，寻找所有点的最小“移动成本”，需要遍历成千上万对点，从而导致巨大计算量($O(n^2)$)。与此相比，归一化 Wasserstein 距离损失可显著降低计算复杂度：通过坐标归一化，可将框坐标“缩放”至[0,1]×[0,1]范围，这将原本复杂的二维分布简化为两个一维分布：$W_1(Bp\_1,\ Bg\_1)+W_1(Bp\_2,\ Bg\_2)$，此时计算复杂度从$O(n^2)$降至$O(n)$。

(3)通过动态加权机制融合L_{IoU}和L_{NWD}，生成归一化的交并比—距离损失L_{NIoU}：

$$\begin{cases}L_{NIoU}=\alpha\cdot L_{IoU}+(1-\alpha)\cdot L_{NWD}\\ \alpha=Sigmoid(Conv1\times1(Norm(Sg/Simg)))\end{cases} \tag{14}$$

其中，α为动态加权系数，Sg为真实框面积，$Simg$为图像面积，$Conv1\times1$为 1×1 卷积，$Norm$为归一化操作。根据目标尺寸动态调整L_{IoU}和L_{NWD}的权重：当$Sg/Simg<1\%$(小目标)时，增加L_{NWD}权重($\alpha>0.6$)以强化分布差异优化——传统L_{IoU}对密集遮挡(重叠>50%)高度敏感，小偏移易导致损失剧增和训练不稳定，而L_{NWD}通过量化预测框与真实框的概率分布差异，即使在无重叠情况下也能有效优化边界定位；当目标较大时，增加L_{IoU}权重($\alpha<0.4$)以提升边界框重叠精度。结合尺度/坐标归一化(公式(12)～(13))，进一步平滑小偏移下的梯度变化。这种自适应机制使L_{NIoU}在多尺度目标检测中表现更优，尤其缓解密集遮挡场景下的定位偏差。由图 7 所示的不同损失函数得到的小目标检测结果可知，利用本文提出的归一化交并比-距离损失有效提高小目标检测精度的同时，提高了小目标检测的置信度。

4 实 验

4.1 实验设置

本文选择 VisDrone2019[24]和 TinyPerson[25]作为基准数据集。VisDrone2019 数据集包含城市、乡村等多场景高分辨率航拍图像，涵盖行人、车辆等

(a)　　(b)

图 7　不同损失得到的小目标检测结果,(a)为交并比损失,(b)为提出的归一化交并比-距离损失 L_{NIoU}

10 类目标,适用于交通监控等任务。其主要问题包括:目标尺寸偏小且多为小目标;类别分布不均,易使模型特征表征受样本数量多的类别主导;存在密集遮挡,增加检测难度。

为解决上述问题,对 VisDrone2019 数据集进行如下划分:包含 6471 个样本的训练集,包含 548 个样本的验证集,包含 1610 个样本的测试集,图像尺寸统一处理为 640×640 像素。为确保性能对比实验的公平性,所有对比算法均在如表 1 所示的相同硬件环境下进行推理测试。推理实验统一采用单精度浮点数(FP32)格式,以确保推理指标的可比性。训练过程采用表 2 列出的关键参数。针对无人机端侧硬件的推理性能测试,单独在 NVIDIA GeForce 4060Ti 上进行。

表 1　训练/测试实验环境的硬件配置

实验环境	参数
CPU	13th Gen Intel(R) Core(TM) i5-13490F
GPU	NVIDIA GeForce RTX 4060 Ti
RAM	32 GB
Operating System	Microsoft Windows 11
Language	Python 3.9.20
Frame	Pytorch 2.0.1
CUDA Version	11.8

表 2　模型训练参数

轮数	每批次样本数	初始学习率	最终学习率
150	16	1×10^{-2}	1×10^{-4}
优化器		动量	损失权重
SGD		0.937	0.5

4.2　消融实验

为评估本文所提各个模块对模型性能提升的效果,在 YOLOv10 的基础上,逐步引入 CGSA、CMDCB 和 L_{NIoU},并对其性能进行消融实验。这些实验采用包括精确率(Precision)、召回率(Recall)、平均精度(mean Average Precision, mAP)、每秒处理帧数(Frames Per Second, FPS)、模型参数(Parameter)和模型规模(Model Size)等评估指标。实验在 NVIDIA GeForce RTX 4060Ti (4060Ti) 和 NVIDIA Jetson Orin Nano (Nano)上进行,以验证模型在高性能 GPU 和无人机端的表现。此外,若非特殊说明,本实验中所有的 mAP 均为 mAP@50。

由表 3 所示的消融实验结果可得以下结论。(1) 引入 CGSA 模块后,验证集的 mAP 从 31.41% 提升至 31.93%,提升了 0.52 个百分点。(2) 在 CGSA 的基础上引入 CMDCB 模块,验证集的 mAP 从 31.93%提升至 32.82%,提升了 0.89 个百分点,这表明:①该模块通过分析卷积层参数权重,选择一定比例的权重参数进行剪枝,有效减少模型规模(5.01M→4.69M)、提高模型效率;②通过多级跳跃连接及多层 C2f 块融合信息,优化了网络结构,使得推理速度提升 20.39% (22.80→27.45 FPS)。(3) 引入 L_{NIoU} 后,验证集的 mAP 从 31.41%提升至 31.83%,提升了 0.42 个百分点,有效缓解了目标间的相互遮挡问题。(4) 当 CGSA、CMDCB 与 L_{NIoU} 联合作用时,模型性能达到最优:验证集的 mAP 从 31.41%提升至 33.60%,提升了 2.19 个百分点,同时模型规模降低 18.11%(5.20→4.69MB)、在 NVIDIA GeForce RTX 4060 Ti 的 FPS 提升 8.98%(94.62→103.12FPS)。(5) 消融实验结果表明,CGSA 模块增强复杂背景下的特征选择能力;CMDCB 提升重叠场景下的目标检测精度;L_{NIoU} 损失函数提升了密集场景中的目标检测精度。(6) 在无人机端性能上,相较于 YOLOv10-N 的 24.12 FPS,本文提出的 LGSP-Y10N 以 7%剪枝率实现 27.45

FPS，提升 13.81%的同时满足无人机实时检测需求（＞20 FPS）；此外，相较于 YOLOv10-N，LGSP-Y10N 模型规模缩小 9.81%（5.20MB→4.69MB），验证了其在资源受限的无人机端侧硬件上的高效性。

由表 4 的实验结果，可得类似于由表 3 所得的结论。

表 3 本文基于 YOLOv10-N 的优化模型在 VisDrone2019 数据集的消融实验

Data	CGSA	CMDCB	L_{NIoU}	Precision/%	Recall/%	mAP/%	Parameter/M	Size/MB	FPS(4060Ti)	FPS(Nano)
Val	×	×	×	42.61	31.92	31.41	2.71	5.20	94.62	21.35
	√	×	×	43.72	32.47	31.93	2.42	5.01	96.20	22.10
	×	√	×	43.92	32.61	32.09	2.56	4.77	101.47	23.87
	×	×	√	42.94	32.25	31.83	2.71	5.20	94.62	21.35
	√	√	×	44.13	32.90	32.82	2.21	4.69	103.12	24.83
	√	×	√	43.79	32.81	32.54	2.42	5.01	96.20	22.10
	×	√	√	44.59	32.89	32.92	2.56	4.77	101.47	23.87
	√	√	√	**44.51**	**33.32**	**33.60**	**2.21**	**4.69**	**103.12**	**24.83**
Test	×	×	×	36.42	28.04	25.32	2.71	5.20	107.31	24.12
	√	×	×	37.41	28.37	25.73	2.42	5.01	109.10	24.85
	×	√	×	37.69	28.81	26.07	2.56	4.77	110.18	25.63
	×	×	√	37.60	28.50	26.04	2.71	5.20	107.31	24.12
	√	√	×	37.81	28.93	26.61	2.21	4.69	115.73	27.45
	√	×	√	37.68	28.61	26.39	2.42	5.01	109.10	24.85
	×	√	√	38.49	28.73	26.58	2.56	4.77	110.18	25.63
	√	√	√	**38.05**	**29.43**	**27.20**	**2.21**	**4.69**	**115.73**	**27.45**

表 4 本文基于 YOLOv10-S 的优化模型在 VisDrone2019 数据集的消融实验

Data	CGSA	CMDCB	L_{NIoU}	Precision/%	Recall/%	mAP/%	Parameter/M	Size/MB	FPS(4060Ti)	FPS(Nano)
Val	×	×	×	49.82	37.28	38.18	9.07	14.02	81.20	18.50
	√	×	×	50.51	38.04	39.03	8.85	13.69	83.49	19.12
	×	√	×	50.80	38.20	39.21	8.70	13.47	85.02	19.75
	×	×	√	49.46	37.47	38.45	9.07	14.02	81.20	18.50
	√	√	×	51.42	38.51	39.38	7.59	13.32	95.91	22.30
	√	×	√	50.73	38.00	39.04	8.85	13.69	83.49	19.12
	×	√	√	51.56	38.48	39.51	8.70	13.47	85.02	19.75
	√	√	√	**52.02**	**38.82**	**40.21**	**7.59**	**13.32**	**95.91**	**22.30**
Test	×	×	×	43.02	32.04	30.33	9.21	14.24	90.14	20.50
	√	×	×	43.51	32.51	30.54	8.42	13.91	92.02	21.00
	×	√	×	43.76	32.77	31.04	8.65	13.40	93.02	21.50
	×	×	√	42.81	32.20	30.21	9.21	14.24	90.14	20.50
	√	√	×	43.74	32.87	30.86	7.59	13.32	100.10	23.25
	√	×	√	43.82	32.59	30.72	8.42	13.91	92.02	21.00
	×	√	√	44.04	33.20	31.19	8.65	13.40	93.02	21.50
	√	√	√	**44.48**	**33.37**	**32.03**	**7.59**	**13.32**	**100.10**	**23.25**

由表 5 所示实验结果可知，剪枝率为 7%时模型性能达到最优。

表 5 对 LGSP-Y10N 进行不同剪枝比例的性能对比实验

剪枝率	mAP/%	FPS (Jetson Orin Nano)	模型规模(M)
0	27.51	22.80	5.01
5%	27.35	24.20	4.75
7%	**27.20**	**27.45**	**4.69**
10%	26.04	28.10	4.52

由图 8 可知，对比基线方法，在引入 CMDCB 后，重叠行人及车辆的检测精度得到提高，而在引入 CGSA 及 L_{NIoU} 后明显降低误检率以提高置信度。

4.3 超参数敏感性分析

为进一步验证关键超参数对模型鲁棒性和性能的影响，本文对 CGSA 的分组数 G（调控空间注意力复杂度 $O(C^2/G)$）、CMDCB 的注意力融合方式（包括水平-垂直池化及动态加权系数 α）和通道压缩率进行了网格搜索（步长 0.1/2%）敏感性实验。在 VisDrone Val 数据集上评估 mAP、FLOPs 和 FPS(NVIDIA Jetson Orin Nano)。

表 6 揭示了在最优参数配置下（分组数 $G=4$，水平-垂直，融合方式 $\alpha=0.5$，压缩率＝7%）实现了最佳平衡效果。由表 6 所示实验结果可得以下结论。

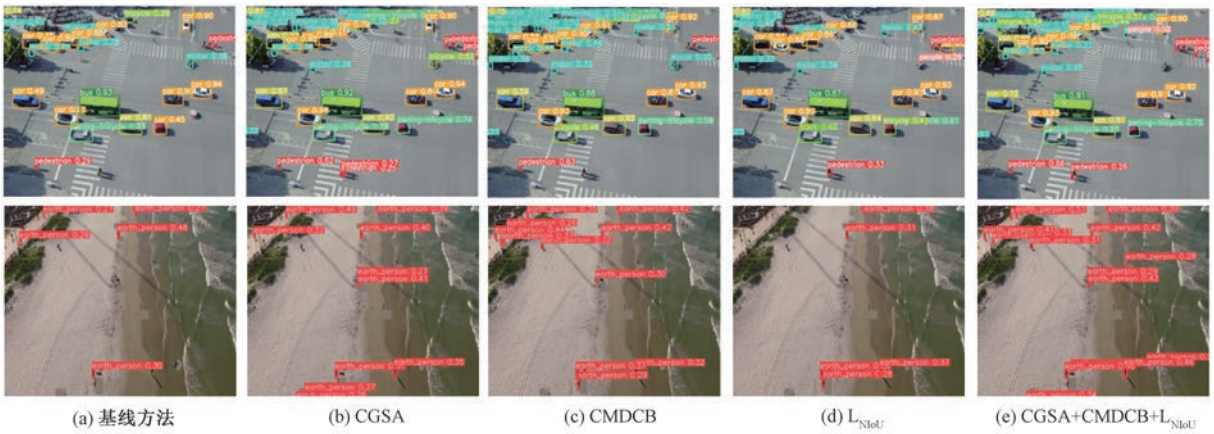


图 8　消融实验结果的可视化

(1) 分组数 G 对计算量-精度平衡至关重要:G=2 时计算量较大,mAP 达 33.45%,但 FLOPs 增加 8%;G=8 时 mAP 下降 0.48%,但 FLOPs 下降 7%;G=4 (最优)时 FLOPs 下降至 8.4G,mAP 达 33.60%,证明分组机制有效缓解 $O(C^2)$ 瓶颈而不丢失长程语义。

(2) 融合方式方面,全局池化简单但忽略空间坐标,mAP 仅为 33.21%;水平-垂直池化通过一维分解捕捉位置语义,mAP 达到 33.60%;α=0.5 时自适应小目标权重最优,mAP 达到 33.60%。

(3) 压缩率在 7%达到最优效果,FLOPs 达到 8.4G,FPS 达到 27.45。

表 6　超参数敏感性分析

超参数	值	mAP/%	FLOPs/G	FPS(Nano)
分组数 G (CGSA)	2	33.45	9.2	25.12
	4(最优)	**33.60**	**8.4**	**27.45**
	8	33.12	7.8	29.10
融合方式 (CMDCB)	全局池化	33.21	8.7	26.20
	水平-垂直(**最优**)	**33.60**	**8.4**	**27.45**
	α=0.3	33.28	8.5	26.80
	α=0.5(最优)	**33.60**	**8.4**	**27.45**
	α=0.7	33.41	8.3	27.90
压缩率 (CMDCB)	0%	33.64	9.1	22.80
	5%	33.44	8.7	24.20
	7%(最优)	**33.60**	**8.4**	**27.45**
	10%	32.92	8.0	28.10

4.4　性能对比

本文旨在研究适用于无人机航拍图像目标检测,重点考虑低硬件配置和良好适用性。因此,选择一阶段目标检测算法进行性能对比,如 YOLOv8[26] 系列模型、YOLOv10[17] 系列模型,其他轻量级目标检测模型,如 MobileNetV4-SSD[27]、GhostNetV3[28]、CSPDNetV2[29] 等。

表 7 和表 8 给出对比实验结果,所有模型均在 VisDrone2019 数据集上进行训练,参数量和平均精度作为性能指标,且均未采用预训练权重进行训练。由实验结果,可以得出以下结论。(1) LGSP-Y10N 表现卓越。以 2.21M 的参数量实现 33.60% 的 mAP,较 YOLOv10-N,从 31.41%提升至 33.60%,提升了 2.19 个百分点,在行人和自行车的检测精度上分别提升了 6.07% (32.60% → 34.58%) 与 1.12%(8.04%→8.13%)。显示出在小目标检测上的显著优势。(2) LGSP-Y10S 在不同场景下的性能都很突出。①检测精度。相较于 YOLOv10-S,提升了 5.31%(38.18%→40.21%),特别是在卡车检测任务中,LGSP-Y10-S 的检测精度(38.27%)超越了参数量约为其 1.5 倍(7.59M→11.12M)的 YOLOv8-S(35.28%)。② 模型轻量化。相比 GhostNetV3 和 CSPDNetV2,LGSP-Y10-S 在参数量上仅增加了 24.63%(6.09→7.59M)以及 45.40%(5.22→7.59M),但检测精度却提升了 11.54%(36.05%→40.21%)和 13.17%(35.53%→40.21%)。(3) LGSP-Y10 在检测精度和模型规模间实现了良好的平衡。相较于 CSPDNetV2 和 GhostNetV3 等轻量级模型,提供了更高的检测精度和推理性能,适合高精度要求较高的应用场景。

表 9 给出不同模型在 TinyPerson 数据集的对比实验结果。TinyPerson 数据集包含多种复杂背景、密集小尺寸行人的实际场景,适合测试目标检测算法在小目标、密集场景、遮挡环境下的性能。实验过程中,将 TinyPerson 数据集按 7∶3 比例随机划分为训练集和验证集,训练参数与评估指标设置与此前实验相同。表 9 所示的实验结果表明:LGSP-Y10 模型在更少参数量的情况下,取得比基准模型 YOLOv10 更好的检测性能,且推理速度更快,具有

表 7 不同轻量级模型在 VisDrone2019 数据集的性能对比实验

Model	Precision/%	Recall/%	mAP/%	FPS(4060 Ti)	Parameter/M	GFLOPs/G	Model Size/MB
Faster-rcnn	37.25	28.91	27.62	104.73	\	7.51	2.34
YOLOv8-S	48.80	36.49	37.87	81.19	11.12	28.60	10.75
YOLOv8-N	43.00	32.88	31.02	93.20	3.01	8.70	2.94
YOLOv8-N-ghost	38.01	30.03	28.64	111.88	1.72	5.80	2.12
YOLOv8-S-ghost	46.20	36.47	35.22	96.44	5.93	16.40	7.65
MobileNetv4-SSD	41.52	31.33	30.47	91.31	5.25	10.35	5.79
GhostNetV3	45.74	36.40	36.05	101.28	6.09	25.90	12.34
CSPDNetV2	44.59	35.88	35.53	100.42	5.22	7.10	11.01
YOLOv10-S	49.82	37.28	38.18	81.20	9.07	25.50	14.02
LGSP-Y10S(Ours)	**52.02**	**38.82**	**40.21**	95.91	7.59	26.10	13.32
YOLOv10-N	42.61	31.92	31.41	94.62	2.71	8.20	5.20
LGSP-Y10N(Ours)	44.51	33.32	33.60	**103.12**	**2.21**	**8.40**	**4.69**

表 8 不同轻量级模型在 VisDrone2019 数据集的各类检测性能对比实验

Model	mAP/%										
	All	Pedestrian	People	Bicycle	Car	Van	Truck	Tricycle	Awning Tricycle	Bus	Motor
Faster-RCNN	27.62	27.71	24.53	6.12	70.11	29.83	21.03	13.91	8.65	37.22	28.78
YOLOv8-S	37.87	40.29	31.58	11.30	78.73	43.38	35.28	25.62	15.32	54.42	42.80
YOLOv8-N	31.02	32.82	26.22	6.51	74.20	36.21	24.40	18.89	10.62	43.31	35.61
YOLOv8-N-ghost	28.64	29.48	25.00	6.74	72.19	32.88	22.21	16.56	9.59	39.40	31.43
YOLOv8-S-ghost	35.22	37.86	29.79	9.37	76.82	41.01	29.88	23.78	13.17	51.01	39.00
MobileNetv4-SSD	30.47	31.91	27.02	7.89	73.33	36.10	26.09	17.91	10.30	44.04	35.22
GhostNetV3	36.05	38.42	30.18	9.41	77.31	42.03	30.62	24.13	14.45	51.11	39.45
CSPDNetV2	35.53	38.01	29.61	8.90	76.88	41.29	30.04	23.12	14.01	48.03	38.80
YOLOv10-S	38.18	39.74	31.49	12.33	78.67	44.73	36.68	27.43	15.53	52.41	42.78
LGSP-Y10S(Ours)	**40.21**	**42.17**	**34.21**	**13.62**	**79.52**	**46.00**	**38.27**	**28.92**	**17.73**	**56.89**	**44.61**
YOLOv10-N	31.41	32.60	0.274	8.04	74.32	36.27	26.62	18.34	10.52	45.04	35.10
LGSP-Y10N(Ours)	33.60	34.58	29.28	8.13	75.79	40.61	29.39	20.91	10.71	48.33	37.89

表 9 不同模型在 TinyPerson 数据集的性能对比实验

Model	P/%	R/%	mAP/%			FPS (4060 Ti)	Parameter /M	GFLOPs /G	Model size /MB
			All	earth_person	sea_person				
Faster-RCNN	31.94	24.00	16.14	12.08	20.06	111.72	\	7.51	2.33
YOLOv8-S	37.21	27.79	21.73	16.28	26.91	97.17	11.12	28.60	10.73
YOLOv8-N	34.93	26.01	19.13	14.73	23.51	107.19	3.01	8.70	2.93
YOLOv8-N-ghost	33.00	24.88	16.90	12.68	20.96	111.90	1.72	5.80	2.11
YOLOv8-S-ghost	35.31	26.43	19.57	14.63	24.18	95.42	5.93	16.40	7.64
MobileNetv4-SSD	35.08	26.81	21.16	15.83	26.16	96.31	5.25	10.35	5.79
GhostNetV3	35.04	26.21	19.34	14.48	23.93	105.28	6.09	25.90	12.33
CSPDNetV2	34.81	26.00	19.00	14.25	23.56	104.40	5.22	7.10	11.01
YOLOv10-S	36.49	27.35	20.51	15.38	25.42	94.54	9.07	25.50	13.70
LGSP-Y10S(Ours)	**42.22**	**28.01**	**24.55**	**18.38**	**30.38**	99.15	7.59	26.10	12.92
YOLOv10-N	32.53	24.82	17.21	13.53	20.80	103.11	2.71	8.20	5.14
LGSP-Y10N(Ours)	35.57	27.31	21.72	16.01	27.53	**110.32**	**2.21**	**8.40**	**4.61**

较高的准确性和实时性，适合在边缘设备中进行部署。与其他轻量级模型（如 GhostNet、CSPDNet）相比，LGSP-Y10 在检测性能上也有明显优势。

4.5 可视化

为直观展现本文方法的优势，选取类别完整及识别难度大的图像进行可视化分析。由图 9 所示的 VisDrone 数据集样本的检测结果可知，相较于其他方法，本文提出的 LGSP-Y10 能够准确检测绝大多数行人目标，且具有较高的置信度。

由图 10 所示的实验结果，可以得到类似于由图 8 所示实验结果得到的相关结论：(1) 当小目标高度密集及相互重叠而导致信息缺失时，对比方法会出现大量漏检及置信度较低的问题，而在小目标边界不明显时，对比方法会出现误判的问题，这是因为全局上下文信息的缺乏导致边界特征信息表征的错误引导；(2) 本文方法能够正确检测小目标边界区域，

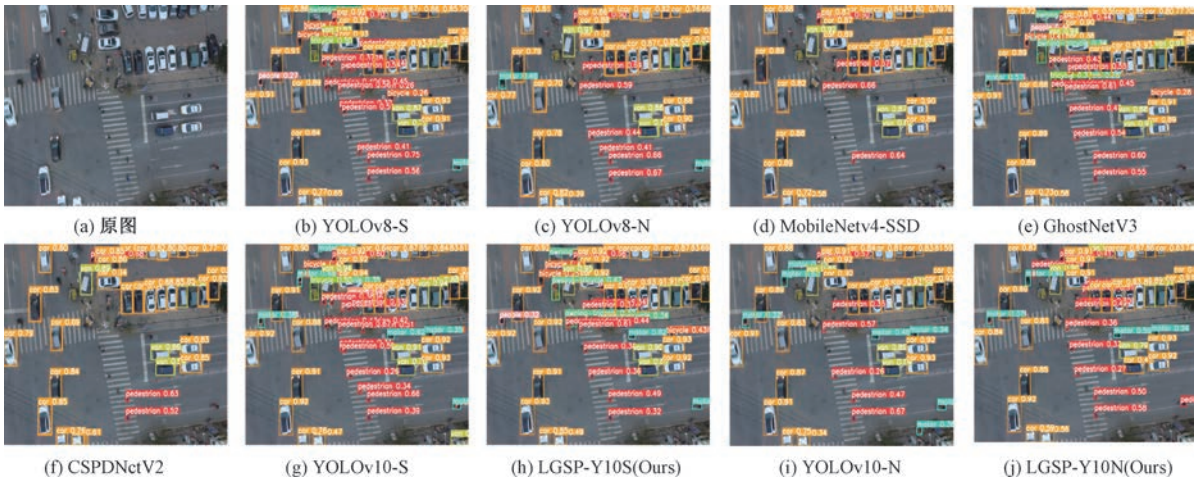


图 9　不同方法在 VisDrone 数据集检测结果的示例图

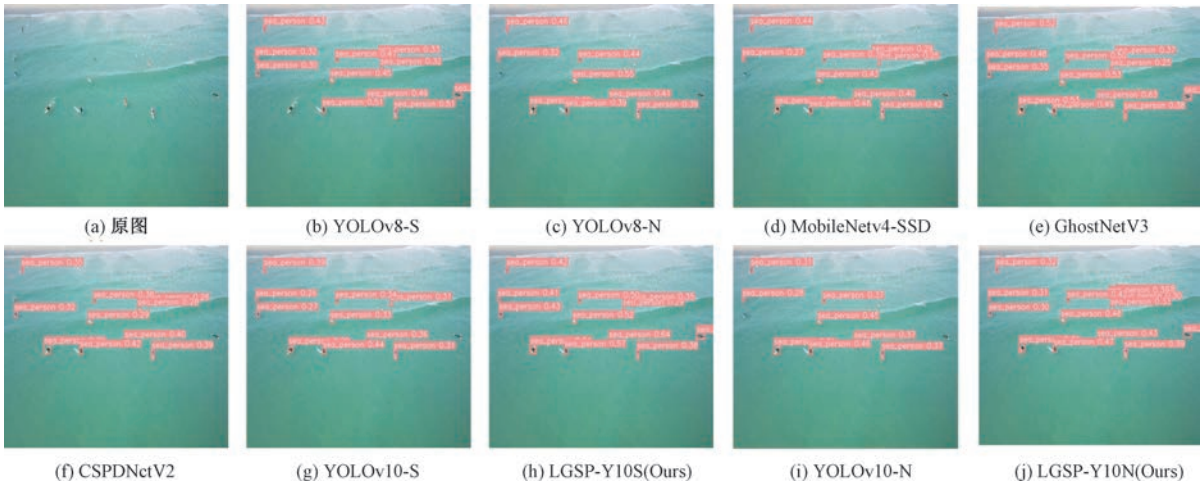


图 10　不同方法在 TinyPerson 数据集检测结果的示例图

验证本文方法能够平衡全局及局部信息,从而更好判别目标类别。

由图 9 和图 10 所示实验结果可知:本文提出的 LGSP-Y10 模型能够准确识别不同行人小目标的位置和姿态,即使在人群密集情况下也能保持较高的区分度,这表明 LGSP-Y10 模型具有很高的行人小目标检测的召回率,即使在高度密集的极端条件下,LGSP-Y10 模型依然能够保持较高的检测精度。

4.6　失败案例分析

虽然本文提出的 LGSP-Y10 模型在无人机航拍图像目标检测中取得不错的效果,但也存在一定的局限性。本小节对 LGSP-Y10 模型在 VisDrone2019 和 TinyPerson 数据集上的失败案例进行了分析,以明确模型的不足并为后续研究提供指导方向。

通过对无人机航拍图像目标检测结果的深入分析,发现主要存在以下两类失败案例。

(1)复杂背景下的漏检。在极低对比度场景下,LGSP-Y10 模型对稀疏纹理特征的捕获能力受限,可能忽略与背景高度相似的微弱高频特征,从而导致漏检。这是因为在高分辨率航拍图像的背景中存在大量纹理噪声或与目标特征相似的干扰,导致 LGSP-Y10 模型偶尔漏检小目标(如行人或自行车)。如图 8 所示,靠近建筑物边缘的行人,因纹理对比度低(目标面积$<0.5\%$)而未被检测。

(2)密集遮挡下的误检。在密集遮挡场景中,注意力机制可能过分强调局部空间细节,从而导致边界框划分错误。如 TinyPerson 数据集中的人群密集区域,LGSP-Y10 模型会将背景区域遮挡目标误判为完整目标,从而出现误检。

5　结　论

在无人机航拍图像目标检测任务中,准确、高效地检测小目标依然是一项挑战。传统目标检测方法

在小目标检测上面临精度不足、计算复杂度高和实时性差的问题。为解决这些问题，本文提出了一种基于 YOLOv10 改进的轻量级目标检测框架—LGSP-Y10，重点优化了小目标检测性能，增强了在无人机航拍图像中的应用价值。首先，提出了跨通道-分组空间注意力机制，提高了小目标特征提取能力并有效抑制了复杂背景噪声；其次，设计了跨尺度多注意力协同双卷积瓶颈网络，改进了特征传播与交互方式，提升了不同尺度目标的检测能力，同时优化了计算复杂度；最后，引入了归一化交并比-距离损失机制，优化了目标定位精度。大量实验结果表明，LGSP-Y10 模型在 VisDrone2019 和 TinyPerson 数据集中取得了检测精度与计算复杂度间的良好平衡，具备在嵌入式设备和无人机平台的应用潜力。

尽管 LGSP-Y10 模型在检测精度和计算效率上取得显著突破，但仍存在复杂背景下的漏检问题和密集遮挡下的误检问题。未来可通过引入更鲁棒的背景噪声抑制机制（如基于对比学习的特征分离）或优化注意力机制的局部-全局平衡，进一步提升模型性能。此外，可探索更轻量化的网络结构，如通过神经架构搜索自动优化模型结构，或结合高效 Transformer 结构以增强小目标特征表征能力。

参 考 文 献

[1] Pretto A, Aravecchia S, Burgard W, Chebrolu N, Dornhege C, Falck T, Fleckenstein F, Fontenla A, Imperoli M, Khanna R. Building an aerial-ground robotics system for precision farming: An adaptable solution. IEEE Robotics and Automation Magazine, 2021,28(3):29-49

[2] Hague C, Kakavitsas N, Zhang J, Christopher B, Willis A, Wolek A. Design and flight demonstration of a quadrotor for urban mapping and target tracking research. 2024, arXiv preprint: 2402.13195

[3] Cai X, Lai Q, Wang Y, Wang W, Sun Z, Yao Y. Poly kernel inception network for remote sensing detection//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2024:27706-27716

[4] Liang X, Zhang J, Zhuo L, Li Y, Tian Q. Small object detection in unmanned aerial vehicle images using feature fusion and scaling-based single shot detector with spatial context analysis. IEEE Transactions on Circuits and Systems for Video Technology, 2020,30(6):1758-1770

[5] Jiang L, Yuan B, Du J, Chen B, Xie H, Tian J, Yuan Z. MffsodNet: Multiscale feature fusion small object detection network for UAV aerial images. IEEE Transactions on Instrumentation and Measurement, 2024,73:1-14

[6] Zhang Y, Gao T, Xie H, Liu H, Ge M, Xu B, Zhu N, Lu Z. Narrowband radar micromotion targets recognition strategy based on graph fusion network constructed by cross-modal attention mechanism. Remote Sensing, 2025,17:64101-64122

[7] Quan Z, Sun J. A feature-enhanced small object detection algorithm based on attention mechanism. Sensors, 2025,25(2):58901-589923

[8] Yang Y, Miao Z, Zhang H, Wang B, Wu L. Lightweight attention-guided yolo with level set layer for landslide detection from optical satellite images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024,17:3543-3559

[9] Cheng K, Cui H, Ghafoor H, Wan H, Mao Q, Zhan Y. Tiny object detection via regional cross self-attention network. IEEE Transactions on Circuits and Systems for Video Technology, 2024,34:8984-8996

[10] Yu Y, Wang M, Zhao Q, Wang Y, Chen Q. A method for extracting small water areas based on an improved FPN network//Proceedings of the International Conference on Ubiquitous Information Management and Communication. Bangkok, Thailand, 2025:1-6

[11] Pan J, Bai Y, Shu Q, Zhang Z, Hu J, Wang M. M-Swim: Transformer-based multiscale feature fusion change detection network within cropland for remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 2024,62:1-16

[12] Zhang S, Li C, Jia Z, Liu L, Zhang Z, Wang L. Diag-Iou loss for object detection. IEEE Transactions on Circuits and Systems for Video Technology, 2023,33(12):7671-7683

[13] Yang C, Shen Y, Wang L. EMFE-YOLO: A lightweight small object detection model for UAVs. Sensors, 2025,25(16):520001-520020

[14] Chen Z, Zhang Y, Xing S. YOLO-LE: A lightweight and efficient UAV aerial image target detection algorithm. Computers, Materials and Continua, 2025,80(2):238-258

[15] Zhu Z, Wang S, Gu S, Li Y, Li J, Shuai L, Qi G. Driver distraction detection based on lightweight networks and tiny object detection. Mathematical Biosciences and Engineering, 2023,20(10):18248-18266

[16] Cai D, Zhang Z, Zhang Z. Corner-point and foreground-area IoU loss: Better localization of small objects in bounding box regression. Sensors, 2023,23(10):496101-496117

[17] Wang A, Chen H, Liu L, Chen K, Lin Z, Han J, Ding G. Yolov10: Real-time end-to-end object detection//Proceedings of the Annual Conference on Neural Information Processing Systems, 2024:1-28

[18] Huang K, Xu Z. Lightweight video salient object detection via channel-shuffle enhanced multi-modal fusion network. Multimedia Tools and Applications, 2024,83(1):1025-1039

[19] Jiang Y, Zhou X, Wang M, Chen T, Chen L. Global convolutional network and sample weighted loss for the robust gear fault diagnosis of reciprocating manipulator. IEEE Transactions on Instrumentation and Measurement, 2025,74:1-10

[20] Ying Z, Zhou J, Zhai Y, Quan H, Li W, Genovese A, Vincenzo Piuri, Scotti F. Large-scale high-altitude uav-based vehicle detection via pyramid dual pooling attention path aggregation network. IEEE Transactions on Intelligent Transportation Systems, 2024,25(10):14426-14444

[21] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Online, 2021: 13708-13717

[22] Pan J, Xu S, Cheng Z, Lian S. C2F-YOLO: A coarse-to-fine object detection framework based on YOLO//Proceedings of the Asia Conference on Algorithms, Computing and Machine Learning. Shanghai , China, 2024:150-157

[23] Liu Y, Luo P. Yolo-Ts: A lightweight yolo model for traffic sign detection. IEEE Access, 2024,12:169013-169023

[24] Du D, Zhang Y, Wang Z, Wang Z, Song Z, Liu Z, Bo L. Visdrone-Det2019: The vision meets drone object detection in image challenge results//Proceedings of the IEEE Conference on Computer Vision. California, USA, 2019:213-226

[25] Yu X, Chen P, Wang K, Han X, Li G, Han Z, Ye Q, Jiao J. CPR++: Object localization via single coarse point supervision. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024,46(7):4908-4925

[26] Varghese R, Sambath M. Yolov8: A novel object detection algorithm with enhanced performance and robustness//Proceedings of the International Conference on Advances in Data Engineering and Intelligent Computing Systems, 2024:1-6

[27] Qin D, Leichner C, Delakis M, Fornoni M, Luo S, Yang F, Wang W, Banbury C, Ye C, Akin B, Aggarwal V, Zhu T, Moro D, Howard A. Mobilenetv4-universal models for the mobile ecosystem//Proceedings of the European Conference on Computer Vision. Milan, Italy, 2024:78-96

[28] Liu Z, Hao Z, Han K, Tang Y, Wang Y. GhostNetV3: Exploring the training strategies for compact models. 2024, arXiv preprint: 2404.11202

[29] Hua W, Chen Q, Chen W. A new lightweight network for efficient UAV object detection. Scientific Reports, 2024,14: 1328801-1328815

WU Jia-Jun, M. S. candidate. His primary research interests include deep learning and image processing.

ZHU Song-hao, Ph. D. , associate professor. His main interests include image processing and artificial intelligence.

Background

Small-object detection in unmanned-aerial-vehicle (UAV) imagery is a central issue in computer vision and intelligent aerial surveillance, falling within the scope of lightweight deep learning for edge deployment. The task demands that average precision (AP) for objects smaller than 32×32 pixels be improved while keeping the model under 2—3 M parameters and 10 G FLOPs—strict envelopes set by NVIDIA Jetson-class hardware. To date, the best-published ultra-light detectors (YOLOv10-N, YOLOv11-N) achieve 28%—31% mAP@0.5 on the VisDrone2019 test set at 90—110 FPS on an RTX-4060 Ti; heavier models exceed 35 % mAP but require>8 M parameters and >25 G FLOPs, precluding real-time inference on edge devices. A clear performance gap therefore persists between heavy and ultra-light architectures, and closing it without extra latency remains an open challenge.

This paper reduces that gap to within 1 mAP of the heavyweights while staying fully ultra-light: LGSP-Y10-N attains 33.60% mAP (+2.5 over the strongest lightweight rival) with only 2.21 M parameters and 8.4 G FLOPs, delivering 103 FPS on RTX-4060 Ti and 27.5 FPS on Jetson Orin Nano. The work is supported by the National Natural Science Foundation of China project "UAV Visual Navigation and Safety Monitoring for Urban Air Mobility" (No. 52405065). The long-term goal of the programme is to equip micro-UAVs with real-time vision capable of locating a 10×10 cm marker within 1 km^2 in 5 min, enabling rapid search-and-rescue and precision agriculture.

Our research group has iteratively released aerial detectors; the latest generation, LGSP-Y10 (2024), is presented herein.