http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2025.0417

基于轻量化 YOLOv8n 的无人机视角目标检测算法

宋骏宇[1,2]，庞岩[1,2,*]

(1. 大连理工大学 工业装备结构分析优化与 CAE 软件全国重点实验室，大连 116024；
2. 大连理工大学 辽宁省空天飞行器前沿技术重点实验室，大连 116024)

摘 要：在无人机 (UAV) 视角下的目标检测任务中，为使其在有限资源内保证目标检测精度，提出一种轻量化 YOLOv8n 的无人机视角下目标检测算法 YOLOv8n-WMCBS。采用 Wise-IoU v3 为网络整体损失函数，动态聚焦低质量预测框；引入 MobileNetV3 网络与坐标注意力 (CA) 机制骨干网络，降低参数量与计算复杂度；引入小目标检测层与检测头，并去除大目标检测层，保留小目标细节特征和位置信息，减少模型计算负担；设计残差幻影卷积 (RGConv)，进一步增强骨干网络的特征丰富性，与双向特征金字塔网络 (BiFPN) 模块结合作为新颈部网络模块，增强颈部网络的动态特征融合；提出改进的非跨步空深卷积 (Conv-SPD) 取代原跨步下采样，充分保留高分辨率小目标信息。实验结果表明：在 VisDrone 2019 数据集中，改进算法在参数量降低 40.5% 和模型文件大小减小 35.5% 的同时，mAP0.5 和 mAP0.5:0.95 较原算法分别提升 4.6% 和 2.9%，表明模型有着良好的检测效果和更好的轻量化性能。

关 键 词：YOLOv8；无人机；损失函数；小目标检测层；轻量级网络
中图分类号：V19; TP183; TP391.41
文献标志码：A **文章编号**：1001-5965(2026)11-4122-13

无人机因其灵活、成本低且轻便易携带等优点，可在各种危险或复杂的场景中执行任务，使其自主巡检技术得到了迅速发展。而 YOLO[1] 系列的单阶段目标检测算法简洁的网络结构有着更快的推理速度，降低了对设备硬件和计算资源的需求，因此，该算法成为目标检测领域内的一大研究热点。然而，无人机视角下图像目标常呈现小尺寸、低分辨率及分布密集等特点，将目标检测模型部署在无人机等边缘设备并提高图像小目标检测精度仍极具挑战性，因此，无人机视角下目标检测算法研究广受关注。Tan 等[2] 首次提出一种加权双向特征金字塔网络 (bidirectional feature pyramid network, BiFPN)，改进 YOLOv8 算法颈部路径聚合网络-特征金字塔网络 (path aggregation network and feature pyramid network, PAN-FPN)，增强了模型多尺度特征融合能力。张瑞芳等[3] 在 BiFPN 基础下设计双向加权多尺度动态颈部网络，实现了浅层与深层特征深度融合，显著提升小目标检测精度。李金凤等[4] 在无人机红外小目标检测中，针对小目标检测问题在原颈部网络插入坐标注意力 (coordinate attention, CA)，有效增强了红外小目标的边缘响应。潘玮等[5] 融合感受野注意力卷积 (receptive-field attention convolution, RFAConv)[6] 与卷积块注意力模块 (convolutional block attention module, CBAM)[7]，并在空间池化金字塔模块中融入大核可分离卷积注意力 (large separable kernel attention, LSKA)[8]，增强骨干网络特征提取的同时，提升了不同层间的语义融合。Yuan 等[9] 为解决面部穴位检测表面无明显物理特征与几何形状问题，采用广义交并比 (generalized intersection over union, GIoU)[10] 损失函数，既关注检

收稿日期：2025-06-30；录用日期：2025-11-14；网络出版时间：2025-12-30 16:10
网络出版地址：link.cnki.net/urlid/11.2625.V.20251230.1323.004
*通信作者. E-mail：ypang@dlut.edu.cn

引用格式：宋骏宇，庞岩. 基于轻量化 YOLOv8n 的无人机视角目标检测算法 [J]. 北京航空航天大学学报，2026，52（11）：4122-4134.
Song J Y，Pang Y. UAV perspective target detection algorithm based on lightweight YOLOv8n[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（11）：4122-4134（in Chinese）.

测帧的重叠区域, 也关注非重叠区域, 极大程度增强了模型面部物体检测能力和面部穴位关键点检测。为解决 YOLOv8 回归损失函数无法优化宽高值不同、纵横比相同的预测框与真实框问题, Ma 等[11]提出一种基于最小点距离的交并比 (minimum point distance intersection over union, MPDIoU), 包含了现有损失函数的所有相关因素。

尽管以上改进方法在目标检测任务中取得了不错效果, 但也存在模型过大、计算复杂度较高及在复杂场景中难以定位分布密集的目标等问题。即目前无人机视角下的目标检测任务仍存在以下难点: ①模型性能与计算资源需求的矛盾; ②航拍目标尺度多变且目标分布密集; ③背景复杂、小目标定位困难。

对于以上无人机视角下目标检测存在的问题, 本文提出一种基于 YOLOv8n 的改进算法——YOLOv8n-WMCBS。针对无人机等平台资源受限问题, 为平衡模型效率与精度, 改进算法引入 MobileNetV3-Small 网络, 嵌入 CA 形成轻量化骨干网络, 降低网络复杂度与计算量, 有效保留了目标细粒度坐标信息。同时, 特征传递路径引入残差幻影卷积 (residual ghost convolution, RGConv), 对冗余特征进行压缩与重建, 进一步降低了模型的参数量与计算复杂度; 针对目标尺度多变、目标分布密集问题, 颈部网络采用 BiFPN, 构建了多路径特征融合网络, 强化网络对小目标特征的表达能力, 提升多尺度检测性能; 新增 P2 小目标检测层和检测头替代原 P5 检测层, 以降低模型复杂度, 提升模型小目标感知能力; 改进 Wise-IoU v3 损失函数以降低模型小目标的尺度敏感性。

1　改进算法结构

在无人机航拍目标检测任务中, 存在小目标居多、复杂遮挡、目标尺度跨度较大、背景复杂等问题, 而原 YOLOv8n[12]算法采用跨阶段局部网络 (cross stage partial net, CSPNet)[13]的骨干网络, 全局感受野过大, 难以聚焦小目标局部细节, 其中, 三阶段检测头 (P3～P5) 缺少高分辨率浅层目标信息, 且在多次上下采样后易导致特征丢失, 使该算法对无人机视角下中小目标检测精度不高, 且容易造成误检和漏检。因此, 本文提出一种改进的轻量化 YOLOv8n 无人机视角下目标检测算法, 改进算法的具体网络结构如图 1 所示。

1) 改进 Wise-IoU 损失函数。采用 Wise-IoU v3 损失函数, 根据目标尺寸动态调整梯度增益, 降低低质量边界框的负面影响, 进一步提升模型整体小

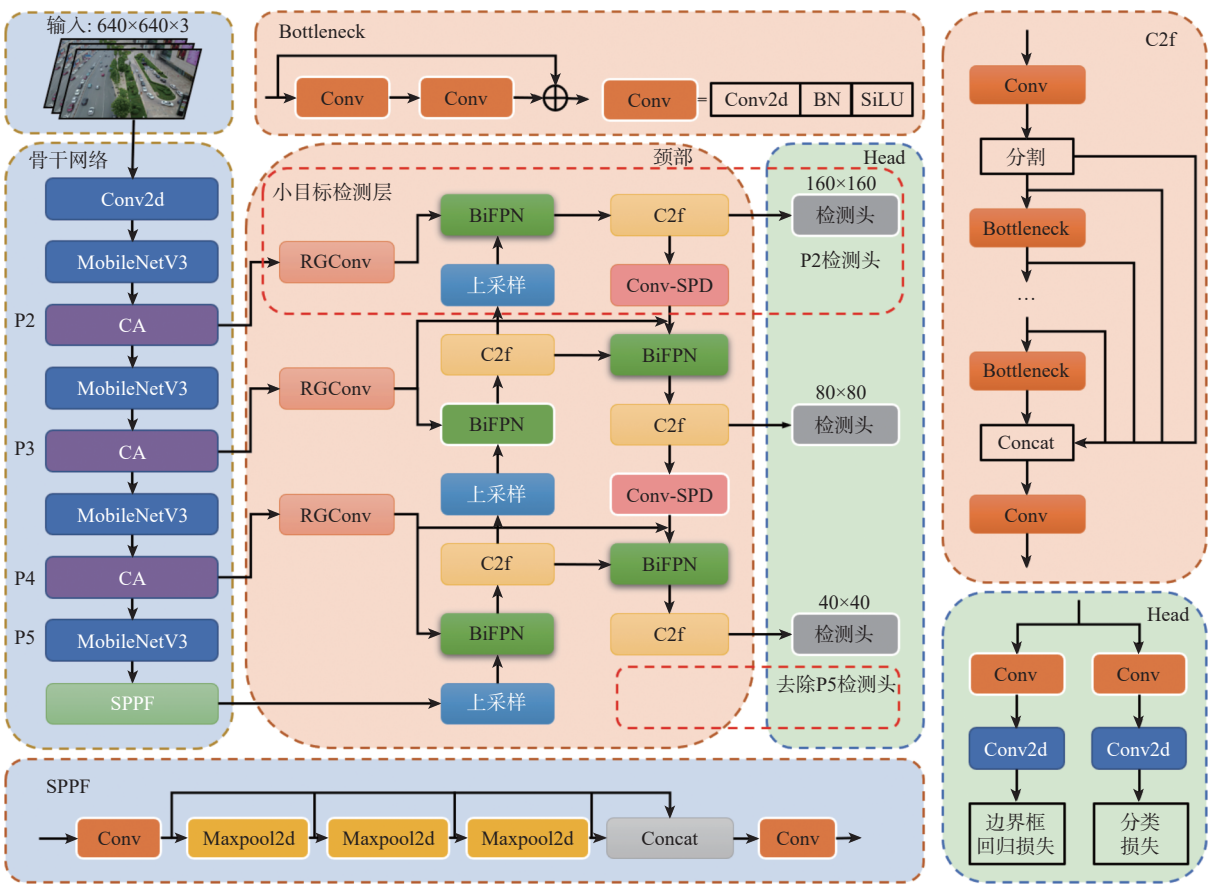


图 1　YOLOv8n-WMCBS 算法网络结构

Fig. 1　Network architecture of YOLOv8n-WMCBS algorithm

目标检测性能。

2）改进轻量化的骨干网络。采用MobileNetV3-Small作为轻量化的基础特征提取骨干网络，融合CA机制构建MobileNetV3-CA模块，增强浅层特征的空间定位能力，确保网络移动端部署友好性。

3）引入小目标检测层。在原始P3～P5检测头基础上新增高分辨率的P2小目标检测层与检测头，移除感知能力不高的P5大目标低分辨率检测头，减轻计算负担的同时，提升网络对高分辨率小目标的感知能力。

4）设计多尺度特征增强模块。设计RGConv模块作为特征传递路径，结合BiFPN形成新的颈部网络模块，增强多尺度特征丰富性与融合能力，实现跨层级语义-细节信息互补。

5）改进下采样。为解决原网络频繁下采样易导致的小目标细节信息丢失问题，本文采用改进非跨步空深卷积(convolution-space-to-depth, Conv-SPD)取代传统跨步卷积，充分保留高分辨率小目标空间信息。

2 改进算法设计

2.1 损失函数改进

原算法采用完整交并比(complete intersection over union, CIoU)[14]和分布焦点损失(distribute focal loss, DFL)[15]形成边界框损失回归，DFL方法以交叉熵形式，使网络更快聚焦到目标位置和临近区域，随后由CIoU损失函数对预测框和真实框进行损失计算。然而，无人机视角目标检测任务中，CIoU损失函数存在明显不足：①未考虑低质量与高质量样本的平衡，对低质量样本的惩罚可能会对模型整体训练产生较大影响；②当预测框与真实框无重叠(交并比为0)时，梯度计算失效，导致困难样本无法被有效优化。针对上述问题，提出Wise-IoU v3[16]作为边界框损失函数，可在降低低质量样本产生有害梯度的同时，显著提升模型对小目标和困难样本的回归精度。

Wise-IoU v3通过基于低质量预测框离群度指标β生成一个非单调的动态聚焦系数r，并将r应用于Wise-IoU v1中，由此得到Wise-IoU v3。当β值较小，基于该值的动态聚焦权重r也越小时，表示预测框的质量较高，此时，降低高质量预测框在损失函数中的权重，相反，若β值较大，则表示一个低质量的预测框，此时，为其分配一个小的梯度增益，从而降低该低质量样本产生的有害梯度。因此，Wise-IoU v3通过这种梯度增益分配策略，动态优化了损失中低质量和高质量预测框的权重，使该模型聚焦于平均质量的样本，从而提升模型整体检测性能。该函数公式为

$$\begin{cases} R_{\mathrm{WIoU}}=\exp\left[\dfrac{(b_{cx}^{\mathrm{gt}}-b_{cx})^2+(b_{cy}^{\mathrm{gt}}-b_{cy})^2}{c_w^2+c_h^2}\right] \\ L_{\mathrm{IoU}}=1-b_{\mathrm{IoU}} \\ L_{\mathrm{WIoUv1}}=R_{\mathrm{WIoU}}L_{\mathrm{IoU}} \\ \beta=\dfrac{L_{\mathrm{IoU}}^{*}}{L_{\mathrm{IoU}}}\in[0,+\infty) \\ r=\dfrac{\beta}{\delta\alpha^{\beta-\delta}} \\ L_{\mathrm{WIoUv3}}=rL_{\mathrm{WIoUv1}} \end{cases} \tag{1}$$

式中：R_{WIoU}为惩罚因子；b_{cx}^{gt}、b_{cy}^{gt}和b_{cx}、b_{cy}分别为真实框和预测框的中心坐标；c_w和c_h分别为真实框与预测框最小外接矩形的宽和高；b_{IoU}为真实框与预测框交集面积除以并集面积，即交并比；L_{IoU}为基础损失，反映边界框的重叠质量；L_{WIoUv1}和L_{WIoUv3}分别为加权交并比损失v1版本和加权交并比损失v3版本；L_{IoU}^{*}为移动平均交并比损失；α和δ为调节超参数，分别控制损失缩放强度和曲线陡峭度。

2.2 骨干网络改进

2.2.1 MobileNetV3

MobileNetV3[17]为Google团队提出的轻量级卷积神经网络，以低参数量和计算复杂度为特点，旨在提高模型在移动端设备上的速度和准确性，这也使其更适合计算资源受限的嵌入式设备等边缘计算设备。该网络继承了此前版本的深度可分离卷积和倒残差瓶颈结构，在上一版本瓶颈结构基础上引入轻量级挤压和激励(squeeze and exciation, SE)[18]注意力模块和改进非线性Hard-swish激活函数。

MobileNetV3网络将倒残差瓶颈与SE注意力机制结合，结构如图2所示。输入图像经1×1卷积扩张，随后通过深度卷积进行升维，并应用SE注意力机制，筛选对网络性能更好的特征通道权重，抑制对网络性能不好的特征通道权重，得到各通道的权重，输入特征图与得到的权重相乘，优化生成的特征图，使用1×1逐点卷积减少通道数并输出，增强每个通道的特征信息捕捉能力。当步长为1且输入和输出通道数相等时，进行残差连接，而当步长为2时，直接输出下采样的特征图。

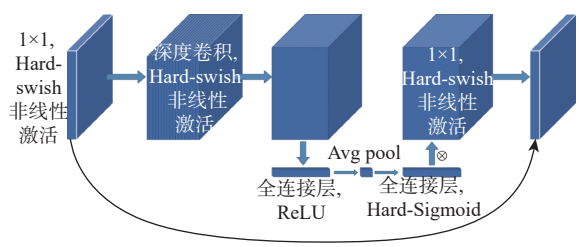


图2 瓶颈结构示意图

Fig. 2 Schematic diagram of the bottleneck structure

考虑到无人机等边缘设备计算资源的限制, MobileNetV3 网络有着参数量小、特征提取能力强、较高的移动端设备适配性等特点, 本文采用该网络 Small 版本作为改进的骨干网络进行特征提取, 该网络由 11 个 Bottleneck 瓶颈层、3 个卷积层组成, 为进一步减轻计算负担, 去除最后 4 层, 保留该网络核心部分。

2.2.2　坐标注意力

大量移动网络设计研究表明, 通道注意力(如 SE 等注意力模块)对模型性能有显著的提升, 然而, 这些注意力通常忽略位置信息。MobileNetV3 作为骨干网络, 轻量化的同时, 瓶颈结构可能忽略细粒度的空间位置信息, 故在此之后引入 CA 机制[19], 其专为移动网络设计而提出, 仅全局平均池化和少量 1×1 卷积, 与 MobileNetV3 的轻量化设计高度兼容。

受 Cai 等[20]在改进 YOLOv8 的无人机水稻图像生长阶段识别研究中启发, 根据 CA 模块在 MobileNetV3 骨干网络不同层之间的影响, 经验证, 在输出特征图尺度分别为 160 像素×160 像素、80 像素×80 像素、40 像素×40 像素处应用该模块时得到最优效果。CA 模块结构如图 3 所示。

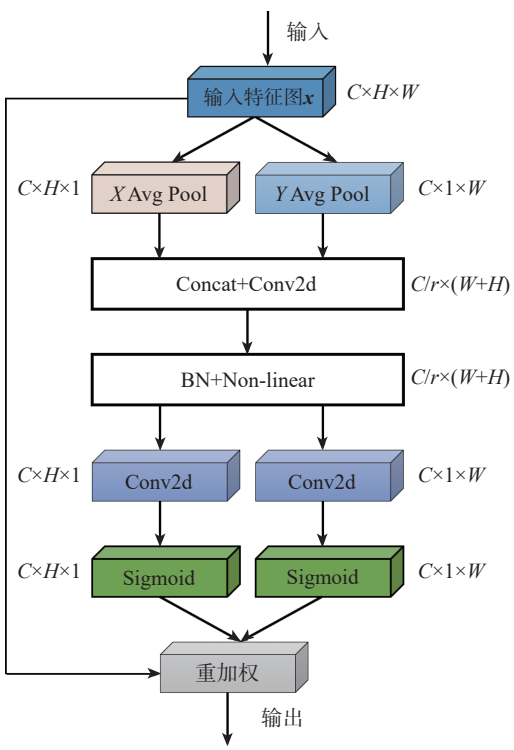


图 3　CA 结构

Fig. 3　CA structure

CA 模块将位置信息嵌入通道注意力中, 尤其是在需要空间选择性注意力的视觉任务中, 与传统的通道注意力机制不同, CA 将特征张量分别沿 2 个垂直和水平方向聚合形成 2 个特征图, 随后, 特征图被分别编码为注意力图, 于是, 每个注意力图都能捕获输入特征图沿一个方向的远程依赖关系, 又能保留沿另一个空间方向的精确位置信息。具体而言, 给定输入特征图 $\boldsymbol{x}$, 经过 2 个大小为 $(H,1)$ 和 $(1,W)$ 的池化核全局平均池化后, 分别沿水平坐标和垂直坐标对每个通道进行编码, 以生成 $C\times H\times 1$ 和 $C\times 1\times W$ 大小的特征图, 输出表示为

$$z_{\mathrm{c}}^{h}(h)=\frac{1}{W}\sum_{0\leqslant i\leqslant W}\boldsymbol{x}_{\mathrm{c}}(h,i) \tag{2}$$

$$z_{\mathrm{c}}^{w}(w)=\frac{1}{H}\sum_{0\leqslant j\leqslant H}\boldsymbol{x}_{\mathrm{c}}(j,w) \tag{3}$$

式中: z_{c}^{h}和z_{c}^{w}分别为通道沿高度 h 和宽度 w 方向的输出; $\boldsymbol{x}_{\mathrm{c}}$ 为输入特征图。

将生成的 2 个聚合特征图拼接后, 通过共享 1×1 卷积进行降维处理, 并生成新特征图, 输出为

$$\boldsymbol{f}=\sigma(F_1([z_{\mathrm{c}}^{h},z_{\mathrm{c}}^{w}])) \tag{4}$$

式中: $\boldsymbol{f}$为对水平和垂直方向空间信息编码的中间特征图; σ为 Sigmoid 非线性激活函数; F_1 为 1×1 卷积转换。

将中间特征图沿空间维度分为单独张量$\boldsymbol{f}^{h}$、$\boldsymbol{f}^{w}$, 通过 1×1 卷积对其进行维度增强, 并分别应用 Sigmoid 激活函数, 以得到 2 个方向的权重:

$$\begin{cases}\boldsymbol{g}_{\mathrm{c}}^{h}=\sigma(\boldsymbol{F}_h(\boldsymbol{f}^{h}))\\ \boldsymbol{g}_{\mathrm{c}}^{w}=\sigma(\boldsymbol{F}_w(\boldsymbol{f}^{w}))\end{cases} \tag{5}$$

式中: $\boldsymbol{g}_{\mathrm{c}}^{h}$和$\boldsymbol{g}_{\mathrm{c}}^{w}$分别为高度和宽度方向的权重; $\boldsymbol{F}_h$ 和 $\boldsymbol{F}_w$ 分别为沿高度和宽度方向的卷积变换。

将注意力权重 $\boldsymbol{g}_{\mathrm{c}}^{h}$、$\boldsymbol{g}_{\mathrm{c}}^{w}$与输入特征图 $\boldsymbol{x}_{\mathrm{c}}$ 相乘进行加权运算, 得到经过 CA 模块的 $\boldsymbol{Y}\in\mathbf{R}^{C\times H\times W}$ 特征图输出:

$$\boldsymbol{y}_{\mathrm{c}}(i,j)=\boldsymbol{x}_{\mathrm{c}}(i,j)\boldsymbol{g}_{\mathrm{c}}^{h}(i)\boldsymbol{g}_{\mathrm{c}}^{w}(j) \tag{6}$$

2.3　颈部网络改进

2.3.1　整体网络结构改进

为解决原检测层应对小目标检测任务的不足, 加入 P2 (160×160) 小目标检测层与检测头, 该检测层未经过下采样, 具有高分辨率、特征图尺寸大等特点, 充分保留了小目标的边缘与纹理特征。此外, 将 P5 (20×20) 检测头删除, 减少小尺度特征图的后续计算, 仅保留 P2、P4, 降低计算开销与模型参数量。P2 检测层的加入, 结合使小目标上下文增强的 BiFPN 特征融合模块, 不仅可加强网络模型对小目标检测位置信息和特征信息的融合, 提高微小目标的检测精度, 又可有效降低计算资源需求。最终, 将 3 组检测头 (P2、P3、P4) 中多尺度信息输入后续 Head 部分进行预测。

2.3.2 动态 RGConv

骨干网络追求轻量化设计，牺牲了部分检测精度，虽然 CA 模块的加入增强了网络对关键特征的聚焦，提高了特征表达能力，但该骨干网络不同尺度输出通道数少，与原算法通道数差异较大，故与后续 BiFPN 模块跨尺度融合时通道对齐困难，难以充分发挥 BiFPN 模块性能。

于是，本文设计了基于 Ghost[21] 卷积的动态 RGConv 模块作为骨干网络与颈部网络连接的特征传递途径，其中，RGConv 模块结构如图 4 所示，其中，$\boldsymbol{\Phi}_k$ 为第 k 个廉价线性变换或廉价操作分支，用于由主卷积分支输出特征生成补充特征；残差部分为深度可分离卷积（depthwise separable convolution，DSConv）。

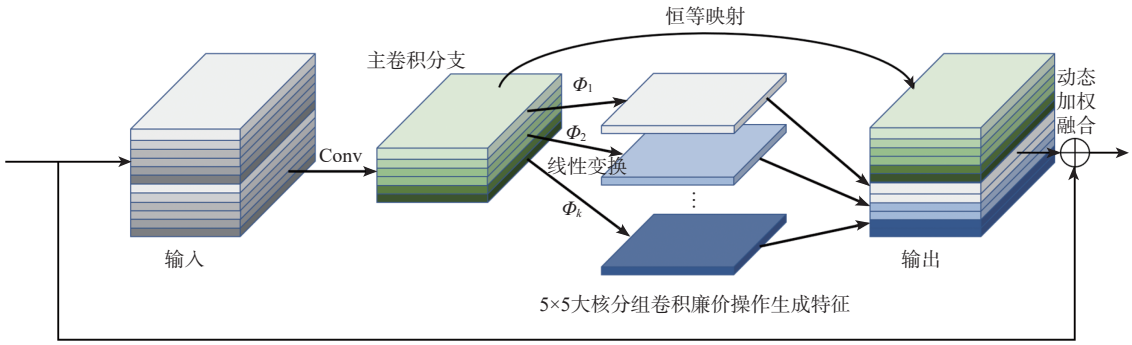


图 4 RGConv 模块

Fig. 4 RGConv module

主路径幻影卷积廉价操作生成更为丰富的特征信息，当输入输出不同时，残差路径采用 DSConv（3×3 深度卷积进行空间特征提取，1×1 逐点卷积进行通道调整）以低参数量和低计算成本增强特征表达能力，并实现特征对齐，引入可学习的通道级权重参数，通过 Sigmoid 函数动态控制主路径与残差路径的贡献。改进的 RGConv 模块特征提取与输出表示为

$$\begin{cases} \boldsymbol{F}_{\text{main}} = \varphi(\text{BN}(\boldsymbol{W}_{\text{main}}\boldsymbol{X})) \\ \boldsymbol{F}_{\text{cheap}} = \varphi(\text{BN}(\boldsymbol{W}_{\text{dw}}\boldsymbol{F}_{\text{main}})) \\ \boldsymbol{F}_{\text{ghost}} = \text{Concat}(\boldsymbol{F}_{\text{main}}, \boldsymbol{F}_{\text{cheap}}) \end{cases} \tag{7}$$

式中：$\boldsymbol{X}$ 为 RGConv 模块输入特征图；$\boldsymbol{F}_{\text{main}}$、$\boldsymbol{F}_{\text{cheap}}$、$\boldsymbol{F}_{\text{ghost}}$ 分别为主卷积分支输出、廉价操作分支输出、拼接后的幻影卷积输出特征图；$\boldsymbol{W}_{\text{main}} \in \mathbf{R}^{(C_{\text{out}}/2)\times C_{\text{in}}\times 1\times 1}$、$\boldsymbol{W}_{\text{dw}} \in \mathbf{R}^{(C_{\text{out}}/2)\times 1\times 5\times 5}$ 分别为主卷积核、深度卷积核，分别生成 $C_{\text{out}}/2$ 的特征，C_{in}、C_{out} 分别为输入、输出通道数；φ 为 SiLU 激活函数；Concat 为拼接函数；BN 为批量归一化。

当输入输出不匹配或步长大于 1 时，引入残差路径，最终双路径经可学习权重参数 $\boldsymbol{\lambda} \in \mathbf{R}^{C_{\text{out}}\times 1\times 1}$ 动态融合得到最终输出：

$$\begin{cases} \boldsymbol{F}_{\text{res}} = \boldsymbol{W}_{\text{pw}}(\boldsymbol{W}_{\text{res_dw}}\boldsymbol{X}) \\ \boldsymbol{\lambda} = \sigma(\boldsymbol{W}_{\boldsymbol{\lambda}}) \\ \boldsymbol{Y} = \boldsymbol{\lambda} \odot \boldsymbol{F}_{\text{ghost}} + (\boldsymbol{I} - \boldsymbol{\lambda}) \odot \boldsymbol{F}_{\text{res}} \end{cases} \tag{8}$$

式中：$\boldsymbol{F}_{\text{res}}$ 为残差路径输出特征图；$\boldsymbol{W}_{\text{pw}}$ 为点卷积核；$\boldsymbol{W}_{\text{res_dw}}$ 为残差深度卷积核；⊙表示广播相乘；$\boldsymbol{W}_{\boldsymbol{\lambda}}$ 为权重向量；$\boldsymbol{Y}$ 为最终输出特征图。

2.3.3 改进的 BiFPN

在原 YOLOv8 算法模型中，颈部网络由 FPN 与 PAN 组合而成，该网络仅将得到的特征进行简单聚合处理，现阶段已不足以满足小目标检测任务需求，因此，本文采用改进的 BiFPN[22] 作为颈部网络。

BiFPN 模块由自下而上、自上而下的特征信息传递及多尺度特征融合 3 部分组成。该模块通过引入可学习权重来获取不同输入特征的重要性，同时，迭代应用于自下而上和自上而下的多尺度特征融合，其引入特征信息的双向流，即同时引入水平和垂直连接，以更好地融合和利用不同尺度的特征，因此，在处理复杂场景时具有更强的鲁棒性，可解决在不同尺度上特征提取时信息丢失和冗余的问题[23]。

结合加入的小目标检测层，并移除大目标检测头，将 BiFPN 通过向现有的 FPN 和 PAN 框架引入额外 2 个横向连接创新，保留了从骨干网络提取的经 RGConv 模块增强后的原始特征，并将其整合到检测特征图中。其中，FPN、PAN 及本文改进的 BiFPN 模块结构如图 5 所示。

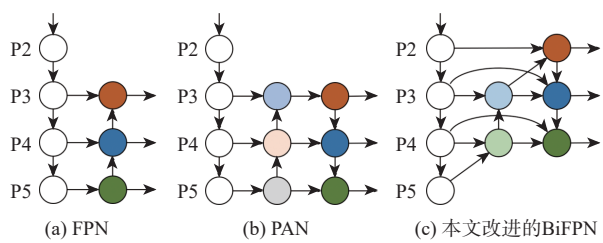


图 5 FPN、PAN 及 BiFPN 结构

Fig. 5 Structure of FPN, PAN and BiFPN

2.3.4　非跨步 Conv-SPD

传统神经网络频繁下采样, 大量的冗余像素信息被卷积或池化操作跳过, 易导致模型损失关键信息, 从而降低小目标检测性能。于是, 本文采用 Conv-SPD 替代原传统跨步卷积。相较由 Sunkara 等提出的 Conv-SPD[24] 模块, 本文选择将普通非跨步卷积作用在空深 (space-to-depth, SPD) 模块前, 普通卷积在高分辨率下提取局部语义过滤噪声, 保留高分辨率小目标信息, SPD 模块对已提取的局部语义特征进行重组, 实现真正意义上的信息保留下采样。此外, 原 SPD 模块最后阶段仅进行简单的通道维度拼接, 无法区分不同子特征图的贡献, 易导致关键特征被噪声淹没或冗余特征未被抑制。为解决该问题, 采用可学习加权融合, 初始化时, 4 个特征图均具有为 1 的权重值, 即初始状态不做任何偏好设置, 通过训练逐渐学习每个子特征图的重要性, 并得到不同权重值, 将 4 个子特征图进行加权融合, 可有效解决特征冗余、表达力不足等问题。改进 Conv-SPD 模块结构如图 6 所示。

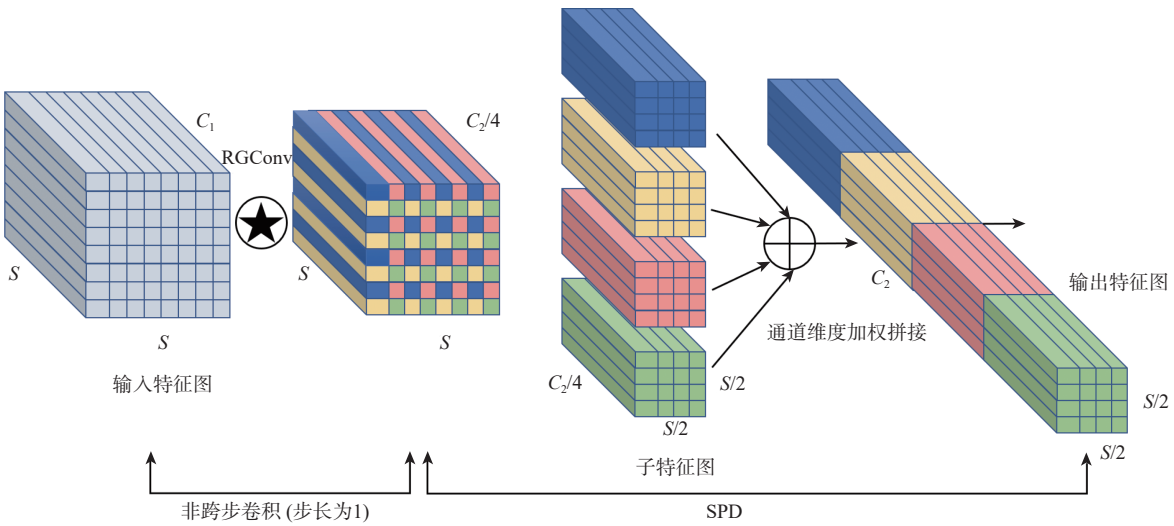


图 6　Conv-SPD 模块结构

Fig. 6　Conv-SPD module structure

Conv-SPD 主要由 2 个模块组成, 即普通非跨步卷积 (步长为 1) 和 SPD 模块。输入特征图经非跨步卷积过滤噪声, SPD 模块将特征图的空间信息转换为深度信息, 主要过程为: 首先, 将输入特征图划分为多个子特征图, 减少空间维度, 同时, 扩展通道数得到深度信息; 然后, 沿通道维度拼接, 通道数变为原来的 4 倍; 最后, 完成特征图从空间信息到深度信息的转化。其中, 若输入特征图维度为 (S, S, C_1), SPD 模块采用 2 倍尺度缩减因子对输入特征图进行下采样, 分为 4 个维度为 ($S/2$, $S/2$, $C_2/4$) 的子特征图, 进行通道维度的权重拼接, 完成输出维度为 ($S/2$, $S/2$, C_2) 的采样后特征图。

3　实验结果与分析

3.1　实验设置

3.1.1　数据集

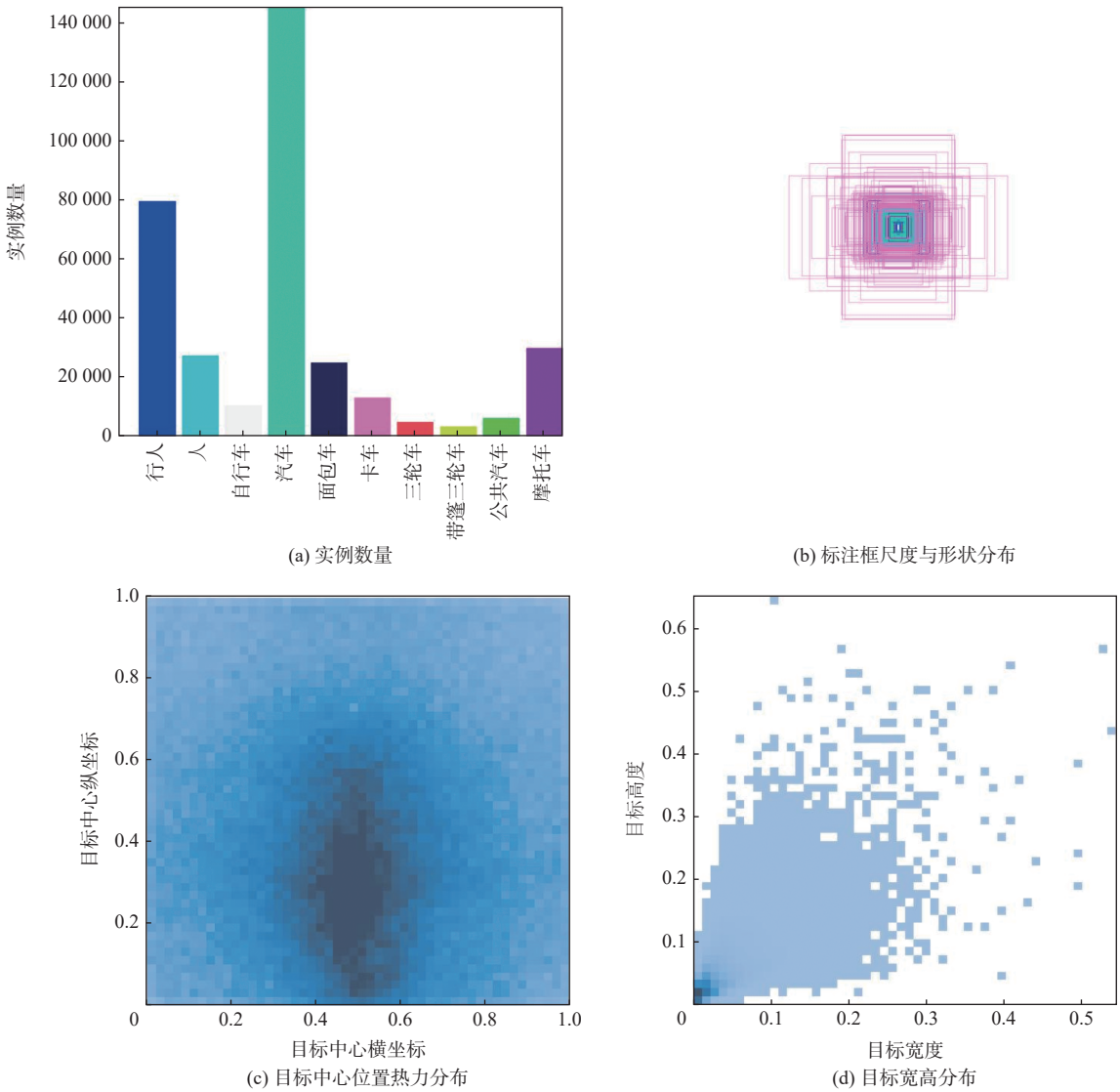
采用 VisDrone-DET 数据集[25] 进行模型训练, 由天津大学的机器学习和数据挖掘实验室的 Aiskyeye 团队进行收集开发。该数据集取自十几个不同城市真实场景, 包括城市、农村和公路等, 覆盖范围广, 由无人机视角下多角度、多场景、不同天气和光照条件下取得, 共有 8 629 张无人机航拍图像, 划分 6 471 张为训练集、548 张为验证集、1 610 张为测试集, 多达 260 万个实例中包含 10 种不同对象类别, 如行人、自行车、汽车和公共汽车等。

图 7 为该数据集的实例数量、实例类别分布及数据集各类目标标注的宽高分布。可以看到, 小目标居多, 且各类别的比例不均衡。

3.1.2　实验环境与评估指标

实验环境为租用服务器, GPU 为 NVIDIA GeForce RTX 4090D, 24 GB 显存, CPU 为 18 vCPU AMD EPYC 9754 128-Core Processor, 60 GB 内存, 操作系统为 Ubuntu22.04, 采用 python 3.12 版本, 模型训练依赖 CUDA 12.4、PyTorch 2.5.1 和 8.3.90 版本 Ultralytics 库深度学习框架。实验参数如表 1 所示。

采用精确度 P、召回率 R、平均精度均值(mean average precision, mAP)、参数量、浮点运算量、推理帧率及模型文件大小为评估指标。精确度 P 表示评估模型预测为正样本的能力; 召回率 R 表示评估模型能找到正样本的能力; mAP 为评估模型计算



(a) 实例数量 (b) 标注框尺度与形状分布

(c) 目标中心位置热力分布 (d) 目标宽高分布

图 7 VisDrone 数据集标注情况

Fig. 7 VisDrone dataset annotation situation

表 1 实验参数配置

Table 1 Experimental parameter configuration

初始输入图像尺寸/(像素×像素)	批次大小	训练轮次	工作线程数	优化器	动量系数	权重衰减系数	patience	初始学习率	最终学习率
640×640	16	100	8	AdamW	0.937	0.000 5	50	0.01	0.01

注: patience表示模型若经历50轮次后性能仍未提升则立即停止训练。

所有类别的平均精确度均值; 参数量为模型可训练的参数数量, 表示该算法模型占用内存的大小; 推理帧率表示每秒能处理图像数, 反映模型的检测速度; 浮点运算量表示每秒浮点运算次数, 反映模型的计算复杂度; 模型文件大小指模型训练后产生的pt文件, 表示训练好的模型权重和偏置等参数的文件磁盘占用大小, 直接关系到模型部署到算力有限的无人机等边缘设备上的难易程度及模型加载速度和推理速度。其中, P、R、单一类别的平均精度(average precision, AP)、mAP、推理帧率的计算公式分别为

$$\begin{cases} P = \dfrac{n_{\mathrm{TP}}}{n_{\mathrm{TP}}+n_{\mathrm{FP}}} \times 100\% \\ R = \dfrac{n_{\mathrm{TP}}}{n_{\mathrm{TP}}+n_{\mathrm{FN}}} \times 100\% \\ \mu_{\mathrm{AP}} = \displaystyle\int_0^1 P(R)\mathrm{d}R \\ \mu_{\mathrm{mAP}} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\mu_{\mathrm{AP},i} \\ \mu_{\mathrm{FPS}} = \dfrac{1\,000}{t_{\mathrm{pre}}+t_{\mathrm{inf}}+t_{\mathrm{post}}} \end{cases} \tag{9}$$

式中: n_{TP} 表示真实样本为正例、模型预测为正的样本数; n_{FP} 表示真实样本为负例、模型预测为正的样本数; n_{FN} 表示真实样本为正例、模型检测为负的样本数; N 为目标类别数; $P(R)$ 为由 P-R 曲线与坐标

轴围成的面积; $\mu_{AP,i}$ 为第 i 类样本的平均检测精度; t_{pre} 为前处理时间; t_{inf} 为推理时间; t_{post} 为后处理时间。

3.2　改进网络结构与损失函数消融实验分析

为验证本文模块的有效性, 在 VisDrone 2019 数据集上设计了多组消融实验, 与原算法进行对比分析。

3.2.1　整体颈部网络重设计实验分析

为验证重新设计颈部网络的合理性, 进行改进算法与原 YOLOv8n 算法的实验对比分析, 其中, YOLOv8n-P2 算法表示在原 YOLOv8n 网络中加入小目标检测层和检测头 P2, 并删去 P5 检测头。实验结果如表 2 所示。

表 2　颈部网络重设计实验结果

Table 2　Neck network redesign experiment results

模型	P/%	R/%	mAP0.5/%
YOLOv8n	41.7	32.2	31.9
YOLOv8n-P2	45.6	34.9	35.3
YOLOv8n-WMCBS	48.0	36.4	36.5

从表 2 可以看出, 经过重新设计的颈部网络整体结构在 P、R、mAP0.5(IoU 阈值为 0.5 时的平均精度均值)上, 相较于 YOLOv8n 分别提升 6.3%、4.2%、4.6%, 对比 YOLOv8n-P2, 在 P、R、mAP0.5 上分别提升 2.4%、1.5%、1.2%。说明最终 YOLOv8n-WMCBS 算法改进了原算法 FPN-PAN 整体颈部特征融合网络结构, 保留了 P3、P4 深层特征图, 在考虑丰富语义的同时, 充分利用了新增的 160 像素×160 像素尺度(P2)小目标检测层的细节信息, 显著提高了模型的目标检测性能, 增强了模型的小目标敏感性。

3.2.2　损失函数改进实验分析

为验证改进损失函数后的模型更适合小目标尺度检测任务, 与原算法 CIoU 损失函数进行对比实验分析, +CIoU 表示 YOLOv8n 算法采用 CIoU 损失函数, +Wise-IoU v3 表示 YOLOv8n 算法采用 Wise-IoU v3 损失函数。实验结果如表 3 所示。

表 3　损失函数改进实验结果

Table 3　Experimental results of loss function improvement

模型	边界框回归损失	P/%	R/%	mAP0.5/%
+CIoU	1.331	41.4	32.1	31.9
+Wise-IoU v3	1.256	43.5	32.4	32.1

从表 3 可以看出, 改进损失函数为 Wise-IoU v3 后, P、R、mAP0.5 相较于原应用 CIoU 损失函数时分别提升 2.1%、0.3%、0.2%, 虽然检测精度提升并不显著, 但边界框损失 box_loss 显著下降, 这说明改进后模型预测边界框与目标实际边界框位置偏差越小, 因此, 改进 Wise-IoU v3 函数计算损失, 对提升模型整体小目标检测性能有重要意义。

3.3　综合消融实验分析

以 YOLOv8n 为基线模型, 在 VisDrone 2019 数据集上对改进模块进行单独添加和总添加的综合消融实验分析, 进一步验证各个模块的作用, 实验结果如表 4 所示。其中, A、B、C、D、E 分别表示在 YOLOv8n 的基础上各自加上 Wise-IoU v3、Mobile-CA、小目标检测层、RG-BiFPN、Conv-SPD 改进后的实验结果。可以看出, 虽然算法 A 的 mAP0.5 值提升并不显著, 仅提升 0.2%, 但该损失函数边界框惩罚策略对小目标检测有着重要意义; 算法 B 在牺牲一定精度的条件下, 参数量与模型文件大小都降低, 有助于在实际无人机硬件上的部署; 而算法 C、D 在牺牲部分检测速度的情况下, 模型文件大小和参数量都进一步降低, 检测精度显著提高, mAP0.5 值分别提升了 3.4%、3.7%; 算法 E 的 mAP0.5 值提升 0.4%, 且其改进非跨步卷积后, 检测速度显著提高。与原 YOLOv8n 基准模型相比, 算法 F 的 mAP0.5 和 mAP0.5:0.95(IoU 阈值从 0.5 到 0.95, 步长为 0.05 时各阈值下 mAP 的平均值)值分别提高 4.6% 和 2.9%, 达到 36.5% 和 21.3%, 参数量和模型文件大小相较原算法分别降低 40.5% 和 35.5%, 以更小的模型文件大小和参数量达成更高的检测精度, 且在该

表 4　综合消融实验结果

Table 4　Comprehensive ablation experiment results

算法	Wise-IoU v3	Mobile-CA	+P2, −P5	RG-BiFPN	Conv-SPD	mAP0.5/%	mAP0.5:0.95/%	推理帧率/(帧·s^{-1})	参数量	模型文件大小/MB
YOLOv8n						31.9	18.4	326.9	3.01×10^6	6.2
A	√					32.1	18.5	313.7	3.01×10^6	6.2
B		√				28.5	15.4	297.5	2.58×10^6	5.4
C			√			35.3	21.1	229.2	2.01×10^6	4.3
D				√		35.6	21.2	223.7	2.44×10^6	7.9
E					√	32.3	18.9	387.3	3.15×10^6	6.5
F	√	√	√	√	√	36.5	21.3	199.3	1.79×10^6	4.0

硬件设备中推理帧率达到 199.3 帧/s，满足无人机移动设备的实时性要求。

3.4 对比实验

为更直观地体现改进后模型整体性能，将基线模型 YOLOv8n 与改进算法在同实验平台同训练参数条件下进行训练，并输出数据对比，采用 Origin 绘图软件对输出结果绘制曲线，如图 8 所示。可以看出，前 50 个迭代轮次模型收敛速度较快，检测精度也随之迅速上升，而后 50 个迭代轮次收敛速度和检测精度皆逐渐趋于平稳。改进算法的 mAP0.5 达到 36.5%，相比原 YOLOv8n 算法提升 4.6%，精确度 P 和召回率 R 较原算法分别提升 6.3% 和 4.2%，达到 48.0% 和 36.4%。边界框损失相较于原算法显著降低，充分表明损失函数改进的有效性。因此，上述改进对无人机等资源受限设备的目标检测任务具有明显贡献。

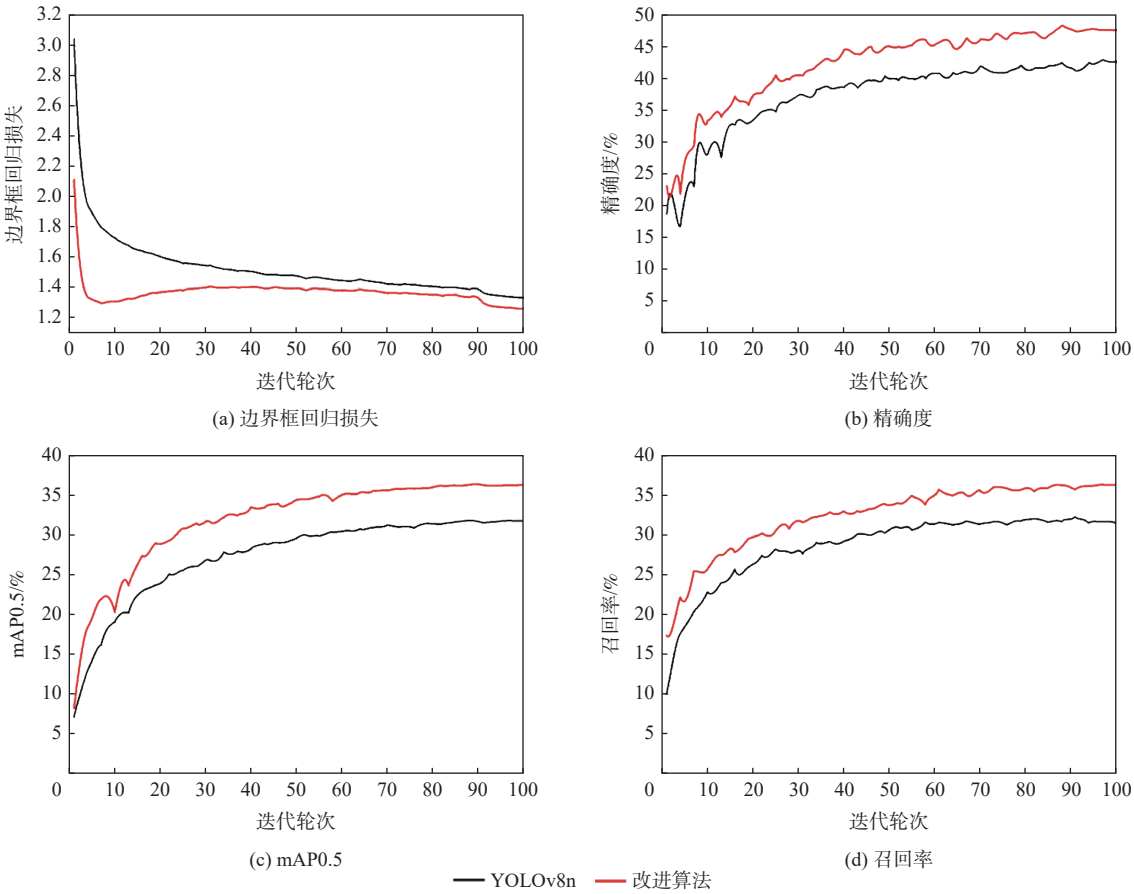


图 8 模型评价指标对比曲线

Fig. 8 Comparison curves of model evaluation metrics

此外，进行改进算法与其余主流目标检测算法 YOLOv4[26]、YOLOv5s[27]、YOLOv6s、FPN[28]、YOLOv11n、YOLOv12n 及基线算法 YOLOv8n 在 VisDrone 2019 数据集上的各类别检测效果对比实验，结果如表 5 所示。可知，对于该数据集的 10 种目标类别，改进算法整体明显优于原算法。且在行人、人、汽车、面包车及摩托车类别中检测效果均优于其余算法，尤其在行人、人和摩托车小目标检测对象中，AP 值分别达到 41.1%、33.6% 和 41.3%，为该类别最优。因此可以得出，改进算法 YOLOv8n-WMCBS 在该数据集上的综合表现优异，有着比原算法更强的检测性能。

为验证模型整体性能，进行上述算法参数量、浮点运算量和 mAP 的对比实验，结果如表 6 所示。可知，在 VisDrone 2019 数据集中，改进算法相比其余算法，参数量最小，模型最小，在无人机等移动设备中更具优势；浮点运算量相较原算法略有增加，但 mAP0.5 与 mAP0.5:0.95 值分别得到相当可观的提升。因此，改进算法较其余算法有着更为轻量化的优势，以更小的参数量和模型文件大小取得了比原算法更优异的检测效果，可有效解决无人机等边缘设备资源受限难以部署目标检测模型的问题。

为进一步验证改进算法的泛化能力，将原算法 YOLOv8n 与改进算法 YOLOv8n-WMCBS 在 VOC-2007 数据集上进行对比实验分析，实验结果如表 7

表 5　VisDrone 2019 数据集上不同算法对比实验结果

Table 5　Comparison experiments results of different algorithms on the VisDrone 2019 dataset

算法	AP/%										mAP0.5/%
	行人	人	自行车	汽车	面包车	卡车	三轮车	带篷三轮车	公共汽车	摩托车	
YOLOv4[26]	24.8	12.6	8.6	64.3	22.4	22.7	11.4	7.6	44.3	21.7	30.7
YOLOv5s[27]	26.8	13.9	9.0	71.0	36.8	37.4	16.3	**18.4**	**56.1**	28.3	31.4
YOLOv6s	24.4	12.9	6.9	71.0	37.5	**39.0**	15.1	18.0	56.0	26.7	30.7
FPN[28]	33.0	25.8	**13.9**	69.4	40.0	34.3	**27.4**	13.4	49.1	37.6	35.6
YOLOv11n	34.3	26.4	7.8	75.2	38.6	27.3	21.6	11.9	48.1	36.9	32.8
YOLOv12n	40.2	32.2	**10.3**	73.3	35.9	27.5	18.1	10.7	41.5	39.7	33.0
YOLOv8n(原算法)	33.4	27.4	7.4	74.7	38.0	26.3	20.8	10.9	45.3	34.3	31.9
改进算法	**41.1**	**33.6**	8.7	**79.1**	**40.8**	31.3	24.2	14.0	50.7	**41.3**	**36.5**

注: 加粗数值表示该类别最优值。

表 6　算法性能对比

Table 6　Comparison of algorithm performance

算法	mAP0.5/%	mAP0.5:0.95/%	参数量	浮点运算量/次
YOLOv4[26]	30.7	17.9	2.450×10^{7}	8.79×10^{10}
YOLOv5s[27]	31.4	18.3	9.100×10^{6}	2.38×10^{10}
YOLOv6s	30.7	18.0	1.630×10^{7}	4.40×10^{10}
YOLOv11n	32.8	19.1	2.590×10^{6}	6.50×10^{9}
YOLOv12n	33.0	19.4	2.570×10^{6}	6.50×10^{9}
YOLOv8n	31.9	18.5	3.010×10^{6}	$\mathbf{8.20\times10^{9}}$
改进算法	**36.5**	**21.3**	$\mathbf{1.790\times10^{6}}$	1.31×10^{10}

注: 加粗数值表示性能最优。

表 7　VOC2007 数据集上泛化性实验结果

Table 7　Generalization experiment results on the VOC2007 dataset

模型	*P*/%	*R*/%	mAP0.5/%	推理帧率/(帧·s^{-1})
YOLOv8n	75.3	67.5	75.1	561
改进算法	76.5	68.9	75.9	323

所示。可以看出, 相较于原 YOLOv8n 算法, 改进算法在 VOC2007 数据集中, *P*、*R*、mAP0.5 分别提升了 1.2%、1.4%、0.8%, 推理帧率为 323 帧/s, 有着良好的实时性, 说明改进算法在其余数据集上仍有较为出色的效果, 模型泛化能力强。

3.5　可视化分析

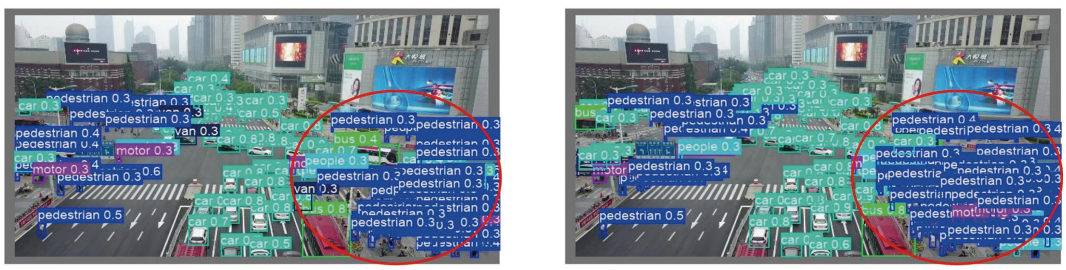
为更全面评估改进算法在不同场景下的检测性能, 在所有数据集图像中选取 4 类具有代表性的场景, 包括夜晚低光照场景、小目标场景、密集复杂场景及高空场景, 并对以上 4 个场景进行 YOLOv8n 算法与改进算法的可视化对比分析, 可视化效果对比如图 9 所示, 其中, 左侧为 YOLOv8n 算法检测效果, 右侧为改进算法检测效果。可以看出, YOLOv8n 算法在不同场景中检测性能都有待提升, 尤其是小目标检测场景, 存在大量的漏检和误检问题。相较于 YOLOv8n 算法, 改进算法即使在夜晚昏暗场景中仍能检测出大量如行人、自行车、摩托车等密集小目标; 在小目标场景中, 检测对象(如行人等)常具有尺度小且多变的特点, 改进算法表现出比原算法更为优异的效果, 显著降低了模型检测漏检率; 而在人群车辆密集且易形成遮挡的复杂密集场景下, 改进算法仍能保持高检测精度与低漏检率; 在高空场景下, 检测效果对比进一步验证了本文算法的改进有效性。



(a) 夜晚低光照场景检测效果对比



(b) 小目标场景检测效果对比



(c) 密集复杂场景检测效果对比



(d) 高空场景检测效果对比

京审字 (2026) G第2942号

图 9 不同场景下可视化效果对比

Fig. 9 Comparison of visualization effects in different scenarios

4 结 论

本文提出一种基于轻量化 YOLOv8 的无人机视角目标检测算法，改进了原网络检测层，加入 P2 检测头与检测层、移去 P5 检测头与检测层；设计了动态 RGConv 模块，形成改进的 RG-BiFPN 颈部网络；改进了 Conv-SPD 代替传统下采样，使算法模型聚焦高分辨小目标信息。得出以下结论：

1）改进算法在可视化分析中表现出优异的检测性能，有效改善了 YOLOv8n 算法存在的误检与漏检问题，且在数据集中各类别目标 AP 均有提升，如行人、人、摩托等小目标检测精度分别提升 7.7%、6.2%、7.0%。

2）改进算法与原算法相比，在 VisDrone 2019 数据集中满足实时性要求下，参数量和模型文件大小分别降低 40.5% 和 35.5%，mAP 提升 4.6%。

3）改进算法以更有优势的模型文件大小达成了更高的精度提升，提升了模型在资源受限无人机平台的部署适配性。

本文聚焦于无人机航拍检测算法模型创新。未来的重要方向是将本文模型部署到实际的无人机边缘计算平台上，并综合运用模型压缩与加速技术，进行详尽的功耗和实时性分析，以推动其在实际应用中的落地。为解决更为复杂场景下目标检测精度不高的问题，后续将对改进算法进行多源数据融合及并行双分支网络结构设计，使无人机对实际生活中复杂应用场景具有更为出色的检测性能。

参考文献（References）

[1] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 779-788.

[2] Tan M X, Pang R M, Le Q V. EfficientDet: scalable and efficient object detection[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 10778-10787.

[3] 张瑞芳, 杜伊婷, 程小辉. 无人机视角下小目标检测算法[J]. 激光与光电子学进展, 2025, 62(4): 0437002.
Zhang R F, Du Y T, Cheng X H. Algorithm for small target detection from an unmanned aerial vehicle view[J]. Laser & Optoelectronics Progress, 2025, 62(4): 0437002(in Chinese).

[4] 李金凤, 崔少华, 潘立慧, 等. 无人机视角下改进 YOLOv8 的红外小目标检测算法[J]. 河北师范大学学报 (自然科学版), 2024, 48(6): 558-564.
Li J F, Cui S H, Pan L H, et al. Infrared small target detection algorithm based on improved YOLOv8 from the perspective of UAV[J]. Journal of Hebei Normal University (Natural Science Edition), 2024, 48(6): 558-564(in Chinese).

[5] 潘玮, 韦超, 钱春雨, 等. 面向无人机视角下小目标检测的 YOLOv8s 改进模型[J]. 计算机工程与应用, 2024, 60(9): 142-150.
Pan W, Wei C, Qian C Y, et al. Improved YOLOv8s model for small object detection from perspective of drones[J]. Computer Engineering and Applications, 2024, 60(9): 142-150(in Chinese).

[6] Zhang X, Liu C, Yang D G, et al. RFAConv: receptive-field attention convolution for improving convolutional neural networks[EB/OL]. (2026-02-28)[2026-06-20]. https://arxiv.org/abs/2304.03198.

[7] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the Computer Vision-ECCV 2018. Berlin: Springer, 2018: 3-19.

[8] Lau K W, Po L M, Rehman Y A U. Large separable kernel attention: rethinking the large kernel attention design in CNN[EB/OL]. (2023-10-20)[2025-06-20]. https://arxiv.org/abs/2309.01439.

[9] Yuan Z J, Shao P W, Li J R, et al. YOLOv8-ACU: improved YOLOv8-pose for facial acupoint detection[J]. Frontiers in Neurorobotics, 2024, 18: 1355857.

[10] Rezatofighi H, Tsoi N, Gwak J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 658-666.

[11] Ma S L, Xu Y. MPDIoU: a loss for efficient and accurate bounding box regression[EB/OL]. (2023-07-14)[2025-06-20]. https://arxiv.org/abs/2307.07662.

[12] Jocher G. YOLOv8n[EB/OL]. (2023-09-27)[2025-06-20]. https://github.com/ultralytics/ultralytics.

[13] Wang C Y, Mark Liao H Y, Wu Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2020: 1571-1580.

[14] Li X, Wang W H, Wu L J, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection [EB/OL]. (2020-06-08)[2025-06-20]. https://arxiv.org/abs/2006.04388.

[15] Zheng Z H, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[16] Tong Z J, Chen Y H, Xu Z W, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism[EB/OL]. (2023-01-24)[2025-06-20]. https://arxiv.org/abs/2301.10051.

[17] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 1314-1324.

[18] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7132-7141.

[19] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 13708-13717.

[20] Cai W X, Lu K B, Fan M T, et al. Rice growth-stage recognition based on improved YOLOv8 with UAV imagery[J]. Agronomy, 2024, 14(12): 2751.

[21] Han K, Wang Y H, Tian Q, et al. GhostNet: more features from cheap operations[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 1577-1586.

[22] Chiley V, Thangarasa V, Gupta A, et al. RevBiFPN: the fully reversible bidirectional feature pyramid network[EB/OL]. (2023-04-28)[2025-06-20]. https://arxiv.org/abs/2206.14098.

[23] Xu W Y, Cui C, Ji Y C, et al. YOLOv8-MPEB small target detection algorithm based on UAV images[J]. Heliyon, 2024, 10(8): e29501.

[24] Sunkara R, Luo T. No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects[C]//Proceedings of the Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2023: 443-459.

[25] Du D W, Wen L Y, Zhu P F, et al. VisDrone-DET2020: the vision meets drone object detection in image challenge results[C]//Proceedings of the Computer Vision-ECCV 2020 Workshops. Berlin: Springer, 2020: 692-712.

[26] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2025-06-05]. https://arxiv.org/abs/2004.10934.

[27] Jocher G. Ultralytics YOLOv5[EB/OL]. (2020-05-29)[2025-06-05]. https://github.com/ultralytics/yolov5.

[28] Lin T Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 2117-2125.

UAV perspective target detection algorithm based on lightweight YOLOv8n

SONG Junyu[1, 2], PANG Yan[1, 2, *]

(1. State Key Laboratory of Structural Analysis, Optimization and CAE Software for Industrial Equipment, Dalian University of Technology, Dalian 116024, China;
2. Liaoning Provincial Key Laboratory of Frontier Technology of Aerospace Vehicles, Dalian University of Technology, Dalian 116024, China)

Abstract: In order to ensure the accuracy of target detection within limited resources, a lightweight YOLOv8n target detection algorithm YOLOv8n-WMCBS from the perspective of an unmanned aerial vehicle (UAV) was proposed. Firstly, Wise-IoU v3 is used as the overall loss function of the network, and the low-quality prediction frame is dynamically focused. Secondly, the MobileNetV3 network and the coordinate attention (CA) backbone network were introduced to reduce the number of parameters and the computational complexity. In order to preserve the specific characteristics and position data of the small target and lessen the model's computational load, the big target detection layer is eliminated and the small target detection layer and detection head are added concurrently. Then, the residual ghost convolution (RGConv) was designed to further enhance the feature richness of the backbone network, and was combined with the bidirectional feature pyramid network (BiFPN) module as a new neck network module to enhance the dynamic feature fusion of the neck network. Finally, in order to fully preserve the high-resolution small target information, an enhanced convolution-space-to-depth (Conv-SPD) is suggested to replace the initial step-down down sampling. Experimental results show that in the VisDrone 2019 dataset, while the number of parameters and the model size are reduced by 40.5% and 35.5%, the mAP0.5 and mAP0.5:0.95 are increased by 4.6% and 2.9%, respectively, compared with the original algorithm, which has stronger generalization ability and is more suitable for resource-constrained equipment such as UAVs.

Keywords: YOLOv8; unmanned aerial vehicle; loss function; small-object detection layer; lightweight networks

Received: 2025-06-30; **Accepted**: 2025-11-14; **Published Online**: 2025-12-30 16:10
URL: link.cnki.net/urlid/11.2625.V.20251230.1323.004

* **Corresponding author**. E-mail: ypang@dlut.edu.cn