OEE | 光电工程 Opto-Electronic Engineering

Article

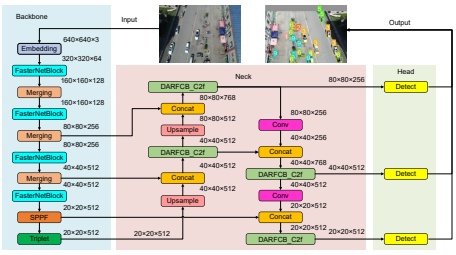
2026 年 1 月, 第 53 卷, 第 1 期

面向无人机航拍图像的注意力融合目标检测算法

徐昊天[1,2], 曾　臻[1,2*], 安嵩阳[1,2], 李　肖[1,2]

[1]湖北工业大学机械工程学院, 湖北 武汉 430068;

[2]现代制造质量工程湖北省重点实验室, 湖北 武汉 430068



摘要： 针对无人机航拍图像中小目标丰富、目标尺度存在差异以及背景复杂导致检测精度低的问题，本文提出了一种基于 YOLOv8n 的注意力融合目标检测算法。首先，采用 FasterNet 架构对骨干网络进行改进，通过部分卷积弥补传统卷积和池化操作中细节信息的丢失，提升小目标的检测能力；其次，引入三元组注意力机制 (Triplet)，增强特征图信息捕捉，有效地应对多样本和目标尺度多样性的问题；最后，基于感受野注意力卷积 (RFAConv) 和卷积块注意力模块 (CBAM)，设计了 DARFCB 结构，强化边缘特征提升模型在复杂环境下的检测精度。在 VisDrone2019 数据集上的实验结果表明，本文算法相较基准模型，*P*、*R* 和 mAP50 值分别提升了 4.9%、3.9% 和 6.1%，有效提高了检测精度。此外，本文算法在 WiderPerson 和 PASCAL VOC 数据集上的实验也展示了良好的泛化能力，进一步说明其能够为无人机航拍图像目标检测任务提供有效支持。

关键词： 目标检测；无人机航拍图像；FasterNet；三元组注意力；感受野注意力卷积

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250204 | CSTR: 32245.14.oee.2026.250204 | 中图分类号：TP391.4 | 文献标识码：A

徐昊天，曾臻，安嵩阳，等. 面向无人机航拍图像的注意力融合目标检测算法 [J]. 光电工程，2026，**53**(1): 250204

Xu H T, Zeng Z, An S Y et al. Attention-fused target detection algorithm for UAV aerial images[J]. *Opto-Electron Eng*, 2026, **53**(1): 250204

1　引　言

无人机航拍图像目标检测技术在农业、环境监测、电力设施检测等领域具有广泛应用。通过该技术，可高效地识别和监测目标，提高工作效率，并为精准决策提供数据支持。然而，在实际应用中仍面临一些挑战。例如，无人机拍摄时飞行高度和视角的不同，其图像通常包含更加丰富的小目标[1]。此外，背景复杂和目标尺度差异大也加大了检测难度。因此，研究高性能的检测模型具有重要的现实意义。

基于传统机器学习的检测方法依赖手工特征和固定规则，通常对单一目标效果较好，但在多类别目标检测和复杂场景下往往效果不佳，难以适应不同环境的检测需求。此外，使用滑动窗口进行图像分类的方式容易产生冗余，影响检测速度和性能。与传统方法不同，基于深度学习的目标检测方法通过卷积神经网络能够自动从图像中提取特征，从而简化传统流程，有效提高检测效率。目前，该算法可以分为两类，一类是基于候选框和分类器的双阶段目标检测算法，如 SPPNet[2]、R-CNN[3]、Faster R-CNN[4] 等，双阶段目标检测算法需要生成候选

收稿日期：2025-07-11；修回日期：2025-10-03；录用日期：2025-10-09；网络出版日期：2026-01-25

基金项目：湖北省创新群体项目 (2022CFA006)

*通信作者：曾臻，zengzhen@hbut.edu.cn

版权所有©2026 中国科学院光电技术研究所

区域，进行分类定位，最终输出位置和类别信息。在检测精度表现优异，但实时性较差。另一类是基于回归的单阶段目标检测算法，如 SSD[5]、RetinaNet[6]、YOLO 系列[7-9] 等，单阶段目标检测算法直接计算图像生成检测结果，简化了检测流程，有效缩短了训练和推理时间[10]，在实时性方面更具优势。

针对无人机航拍图像目标检测精度低的问题，许多学者提出了针对性的改进措施，主要包括多尺度特征融合、轻量化网络模型、数据增强和特征优化等。例如，Zhang 等[11] 通过引入全局-局部特征引导 (GLGG) 模块捕获全局信息，利用改进的并行采样特征融合 (PSFF) 模块和逆残差特征增强 (IFE) 模块，增强了模型对小目标的特征提取能力，但是所提出的模型体积庞大，无法实现部署；Zeng 等[12] 通过设计具有关联坐标注意力的混合注意力模块以及 SEB (simple and efficient bottleneck) 模块，并构建多层特征融合结构，有效提高了 YOLOv5 模型对小目标的检测精度；张润梅等[13] 采用 K-means++聚类算法优化锚框尺寸，引入无参注意力机制 (SimAM)，重构 SPPCSPC 模块，使用轻量级卷积模块 (GhostConv) 代替普通卷积，改进后的 YOLOv7 网络能够有效地应用于无人机视角下复杂环境的目标检测；Sun 等[14] 结合传统路径聚合网络 (PANet) 与收集和分配机制，提出新的特征融合架构-分布式路径聚合网络 (GD-PAN)，虽然提高了模型处理多尺度目标的能力，但也增加了模型的参数量和运算量；Hui 等[15] 提出的自适应残差特征块 (ARFB) 能有效地利用低光和正常光图像之间的特征相关性并共享它们之间的权重，提高了模型在弱光环境下的检测性能；Jiang 等[16] 通过用微小目标预测头代替大目标预测头，引入多尺度特征提取模块 (MSFEM) 和新的双向稠密特征金字塔网络 (BDFPN)，提高了模型的多尺度表示能力；朱立志等[17] 提出了一种基于 YOLOv8n的检测方法，通过设计特征融合模块提高模型对小目标的检测精度，同时减少了参数量。

综上所述，现有算法主要针对特定问题 (如多尺度、小目标) 或特定场景 (如暗光、雾天无人机航拍图像) 进行改进，未能综合处理这些问题。对于无人机航拍图像背景复杂、被检目标尺寸不一且小目标丰富的特点，关键在于提取小目标特征并补偿图像中的特征丢失。因此考虑到平台资源消耗以及检测精度的问题，本文对 YOLOv8n 算法进行改进，以期在无人机航拍图像的目标检测任务中提供一种有效途径。

2 无人机航拍目标检测算法改进

YOLOv8 算法主要由输入端、骨干网络、颈部网络和检测头四个部分组成。首先，输入端对图像进行尺寸调整，然后送入骨干网络进行特征提取。接着，颈部网络对提取的特征进行融合，YOLOv8 的颈部网络采用 FPN (feature pyramid network) 和 PAN (path aggregation network) 组合，构成 FPN-PAN 结构，增强了多尺度特征的融合能力。最后，检测头负责目标检测和分类任务。该模型根据网络宽度和深度，划分了 N、S、M、L、X 五种不同的尺度。

综合考虑计算资源等因素，本文选取 YOLOv8n 作为基准模型。改进后的网络结构如图 1 所示。主要改进点包括以下三个方面：

1) 使用一种高效的神经网络架构 (faster neural network, FasterNet)[18] 替换原骨干网络，利用部分卷积的优势，弥补下采样过程中信息的丢失，提升小目标的检测精度。

2) 在骨干网络与颈部网络之间引入三元组注意力机制 (Triplet)[19]，分别作用于通道和空间维度，有效捕捉重要特征，增强模型对不同尺寸目标的识别能力。

3) 基于感受野注意力卷积 (receptive-field attention convolution, RFAConv)[20] 和卷积块注意力模块 (convolutional block attention module, CBAM)[21]，设计全新的 DARFCB 结构。通过动态扩展感受野并引入多维度注意力机制，提升模型对边缘特征的提取能力，一定程度上解决了因重叠、遮挡导致的误检、漏检问题。

2.1 骨干网络改进

在 YOLOv8 中，骨干网络通过 CSPNet (cross stage partial network) 进行特征提取，这一过程虽然能有效减少分类和回归阶段的参数量，提取更深层次的语义特征，但提取过程中进行多次下采样操作会降低图像分辨率，并且由于卷积和池化操作过程中可能出现的细节损失，也限制了小目标的检测效果。为了解决这一问题，本文使用 FasterNet 架构对骨干网络进行改进，改进后的骨干结构如图 2 所示。

FasterNet 架构包括四个层级阶段，每个阶段前会设置一个嵌入层 Embedding (4×4 卷积，步长 4) 或合并层 Merging (2×2 卷积，步长 2)，用于空间下采样和通道数

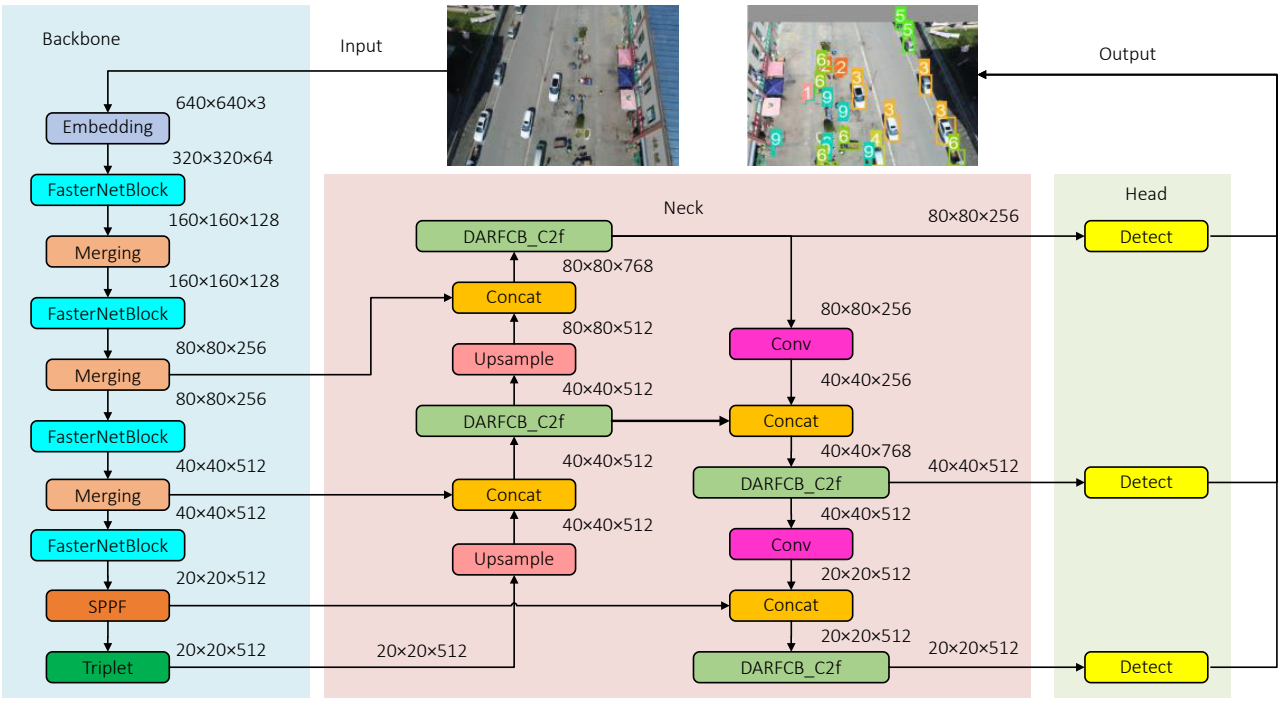


图 1 改进后的 YOLOv8n 结构图

Fig. 1 Structure diagram of improved YOLOv8n

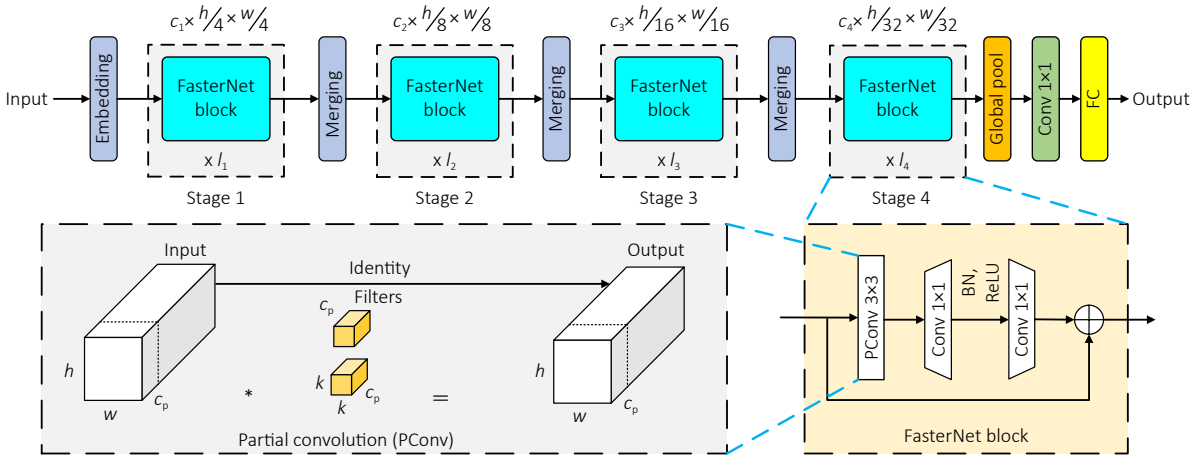


图 2 基于 FasterNet_T0 的骨干网络结构图

Fig. 2 Structure diagram of backbone network based on FasterNet_T0

扩展。每个阶段由多个 FasterNet block 堆叠而成，每个 Block 包含一个部分卷积 (partial convolution, PConv) 和两个逐点卷积 (pointwise convolution, PWConv)，其中 PConv 仅对输入特征图的部分通道进行处理，保留未处理通道以减少计算量并避免细节损失，PWConv 通过 1×1 卷积实现跨通道信息交互，使局部特征与全局信息融合。这种组合结构形成了 T-shaped 受限感受野，与传统卷积相比，更聚焦于小目标特征的中心区域，有效提升小目标的检测精度。这是因为小目标通常占据特征图的较小区域，而其中心位置的特征尤为关键。此外，每个 Block 后设有归一化层 (batch normalization) 和激活层 (GELU 或 ReLU)。本文使用的是 T0 模型，激活层使用 GELU 函数，并将其与原 YOLOv8n 的 SPPF 模块融合。

图 2 中左下角展示了 PConv 的原理，其计算负载和内存访问分别如下式所示，h 和 w 表示特征图的高度和

宽度，k 是卷积核的大小，c_p^2是用于计算的输入通道子集的大小。

$$FLOPs_{\mathrm{PConn}} = h\times w\times k^2\times c_p^2, \tag{1}$$

$$\begin{aligned} MemoryAccess_{\mathrm{PConv}} &= h\times w\times 2c_p + k^2\times c_p^2 \\ &\approx h\times w\times 2c_p \end{aligned}. \tag{2}$$

2.2 引入三元组注意力机制

YOLOv8 算法尽管使用了多尺度特征图，但由于分类和回归任务特征表示的互补性不足，未能充分利用不同尺度之间的信息。为了解决这一问题，本文引入三元组注意力机制 Triplet。该机制主要通过三个分支适配不同尺寸的目标，前两分支处理 C 与 H、C 与 W 的交互，关联通道特征与空间位置，避免大目标因空间广而导致特征割裂；第三分支处理 $H\times W$ 的空间依赖，增强小目标的局部关注，同时跨维度交互可减少小目标被背景掩盖的情况。该设计突破了传统注意力模块中通道注意力与空间注意力独立计算的局限，提升了特征表达能力，实现了对尺寸差异较大目标的更全面描述。此外，Z-pool 操作保留了张量中丰富的特征信息，增强了特征的鲁棒性。再加上卷积核尺寸的灵活调整，使得模型能够更好地适应不同尺寸目标的特征提取需求。操作细节可见图 3 所示。

在通过三元组注意力机制获得的精细注意力应用到输入张量$X\in\mathbb{R}^{C\times H\times W}$上时，该过程可表示为式 (3)[19]，其中$\sigma$代表 sigmod 激活函数，$\psi_1$、$\psi_2$和$\psi_3$代表示该机制中三个分支的标准二维卷积层。

$$y = \frac{1}{3}\left(\widehat{\chi_1}\sigma\left(\psi_1\left(\widehat{\chi_1^*}\right)\right)+\widehat{\chi_2}\sigma\left(\psi_2\left(\widehat{\chi_2^*}\right)\right)+\chi\sigma\left(\psi_3\left(\widehat{\chi_3}\right)\right)\right). \tag{3}$$

2.3 DARFCB-C2f

C2f 模块通过特征图分割和 Bottleneck 层优化信息流，减少特征冗余，保持了模型的准确性。然而，在无人机航拍图像目标检测中，背景与目标之间的区分度往往较低，以及重叠、遮挡较多，容易导致漏检和误检。因此，本文基于 RFA 和 CBAM 结构，设计全新的 DARFCB 结构，并引入到颈部网络中的 Bottleneck 层，形成 DARFCB_C2f 结构，DARFCB_C2f、DARFCB 以及 RFA 的详细结构如图 4 所示。

以输入特征图$\boldsymbol{X}\in\mathbb{R}^{C\times kH\times kW}$($C$ 为通道数，H，W 分别为高、宽)，卷积核大小$k\times k$为例来介绍改进后的 DARFCB 结构。首先采用 Group Conv 高效提取特征，将原始特征图的每个空间位置映射为对应$k\times k$感受野的特征，维度扩展为$C\times kH\times kW$。这一过程依靠 RFA 实现。

$$\begin{cases} \boldsymbol{F}_{\mathrm{rf}} = GroupConv_{k\times k}(\boldsymbol{X}) \\ \boldsymbol{F}_{\mathrm{rf}}\in\mathbb{R}^{C\times kH\times kW} \end{cases}. \tag{4}$$

接着，对感受野空间特征$\boldsymbol{F}_{\mathrm{rf}}$进行全局平均池化，聚合感受野空间特征的通道信息，得到每个通道的全局统计量。

$$\begin{cases} \boldsymbol{F}_{\mathrm{gap}} = GlobalAvgPool(\boldsymbol{F}_{\mathrm{rf}}) \\ \qquad = \dfrac{1}{kH\times kW}\displaystyle\sum_{i=1}^{kH}\sum_{j=1}^{kW}\boldsymbol{F}_{\mathrm{rf}}(:,i,j), \\ \boldsymbol{F}_{\mathrm{gap}}\in\mathbb{R}^{C\times 1\times 1} \end{cases} \tag{5}$$

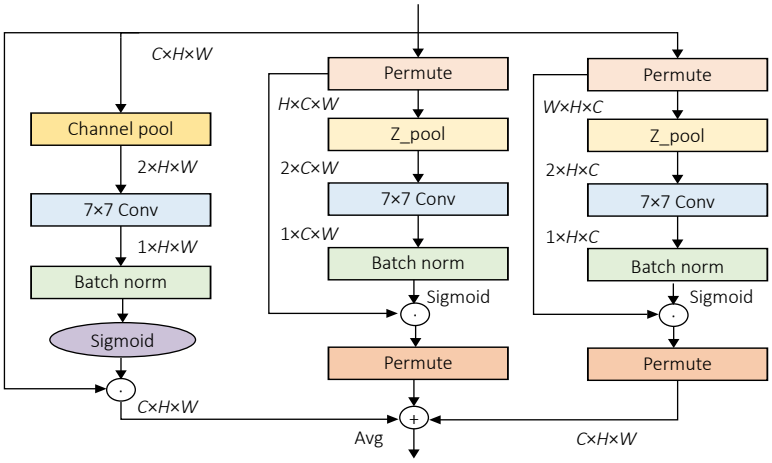


图 3 Triplet Attention 结构图

Fig. 3 Structure diagram of triplet attention

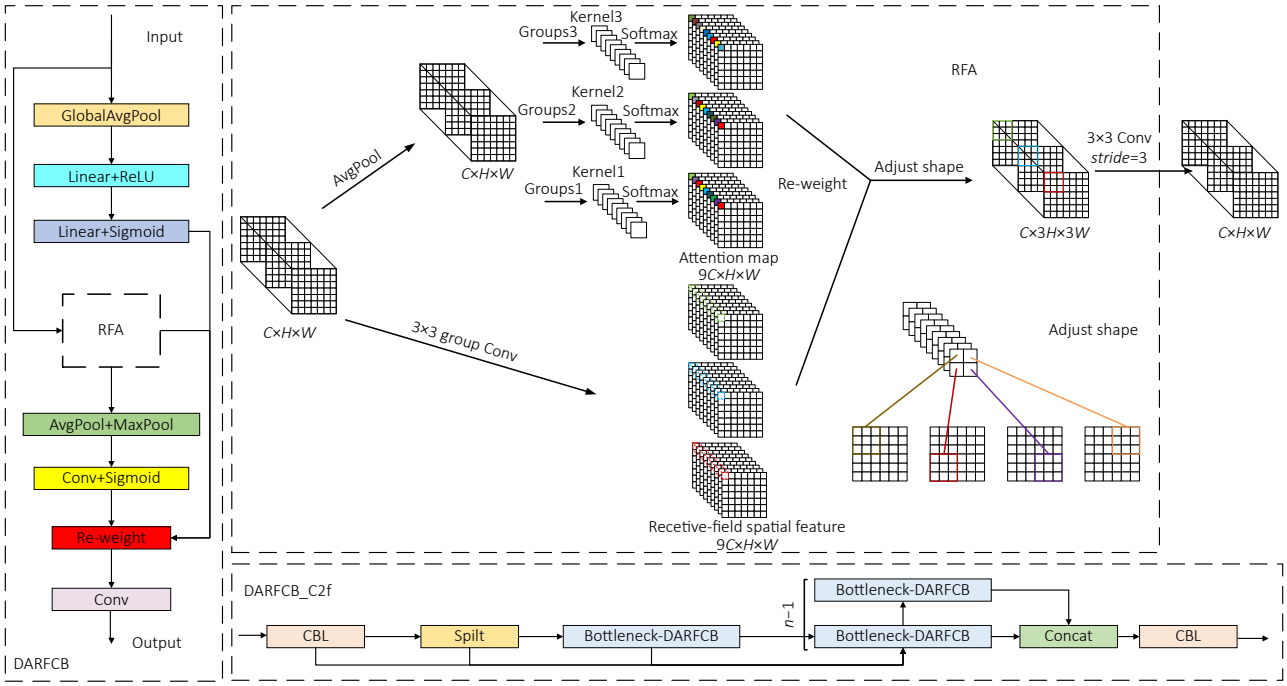


图 4 DARFCB_C2f 的详细结构图

Fig. 4 Detailed structural diagram of DARFCB_C2f

通过 MLP 对$\boldsymbol{F}_{\text{gap}}$降维 (缩减率$r$) 后升维，增强通道间依赖：

$$\begin{cases}\boldsymbol{F}_{\text{mlp1}}=ReLU\left(\boldsymbol{W}_1\cdot\boldsymbol{F}_{\text{gqp}}\right)\\ \boldsymbol{W}_1\in\mathbb{R}^{(C/r)\times C}\end{cases}, \tag{6}$$

$$\begin{cases}\boldsymbol{F}_{\text{mlp2}}=\boldsymbol{W}_2\cdot\boldsymbol{F}_{\text{mlp1}}\\ \boldsymbol{W}_2\in\mathbb{R}^{C\times(C/r)}\end{cases}. \tag{7}$$

用 Sigmoid 将权重归一化到 [0,1]，得到通道注意力图M_{c}：

$$\begin{cases}\boldsymbol{M}_{\text{c}}=\sigma\left(\boldsymbol{F}_{\text{mlp2}}\right)\\ \boldsymbol{M}_{\text{c}}\in\mathbb{R}^{C\times1\times1}\end{cases}. \tag{8}$$

对$\boldsymbol{F}_{\text{rf}}$的通道维度做平均和最大池化，得到两个单通道特征图：

$$\begin{cases}\boldsymbol{F}_{\text{s_avg}}=AvgPool_{\text{channel}}\left(\boldsymbol{F}_{\text{rf}}\right)=\dfrac{1}{C}\displaystyle\sum_{c=1}^{c}\boldsymbol{F}_{\text{rf}}(c,:,:)\\ \boldsymbol{F}_{\text{s_avg}}\in\mathbb{R}^{1\times kH\times kW}\end{cases}, \tag{9}$$

$$\begin{cases}\boldsymbol{F}_{\text{s_max}}=MaxPool_{\text{channel}}\left(\boldsymbol{F}_{\text{rf}}\right)=max\boldsymbol{F}_{\text{rf}}(c,:,:)\\ \boldsymbol{F}_{\text{s_max}}\in\mathbb{R}^{1\times kH\times kW}\end{cases}. \tag{10}$$

将两个池化结果拼接，通过 7×7 卷积 (扩大感受野) 提取空间依赖：

$$\begin{cases}\boldsymbol{F}_{\text{s_concat}}=Concat\left(\boldsymbol{F}_{\text{s_avg}},\boldsymbol{F}_{\text{s_max}}\right)\\ \boldsymbol{F}_{\text{s_concat}}\in\mathbb{R}^{2\times kH\times kW}\end{cases}, \tag{11}$$

$$\begin{cases}\boldsymbol{F}_{\text{s_conv}}=\text{Conv}_{7\times7}\left(\boldsymbol{F}_{\text{s_concat}}\right)\\ \boldsymbol{F}_{\text{s_conv}}\in\mathbb{R}^{1\times kH\times kW}\end{cases}. \tag{12}$$

用 Sigmoid 归一化，得到空间注意力图M_{s}：

$$\begin{cases}\boldsymbol{M}_{\text{s}}=\sigma\left(\boldsymbol{F}_{\text{s-conv}}\right)\\ \boldsymbol{M}_{\text{s}}\in\mathbb{R}^{1\times kH\times kW}\end{cases}. \tag{13}$$

DARFCB 将通道和空间注意力同时作用于感受野空间特征 (区别于 CBAM 的串行分离操作)，通过广播机制逐元素加权：

$$\begin{cases}\boldsymbol{F}_{\text{weighted}}=\boldsymbol{F}_{\text{rf}}\otimes\boldsymbol{M}_{\text{c}}\otimes\boldsymbol{M}_{\text{s}}\\ \boldsymbol{F}_{\text{weighted}}\in\mathbb{R}^{C\times kH\times kW}\end{cases}. \tag{14}$$

通过卷积 (步长) 将加权后的感受野特征映射回原始维度：

$$\begin{cases}\boldsymbol{F}_{\text{out}}=Conv_{k\times k}\left(\boldsymbol{F}_{\text{weighted}},\ stride=k\right)\\ \boldsymbol{F}_{\text{out}}\in\mathbb{R}^{C\times H\times W}\end{cases}. \tag{15}$$

表 1 是 DARFCB_C2f 模块首次应用于改进网络 (即输入特征图$\boldsymbol{X}\in\mathbb{R}^{512\times40\times40}$，$k=3$) 的完整计算流程，细化至每个子步骤的名称、尺寸和逻辑说明。

DARFCB 的核心是从传统的空间特征转向感受野空间特征，这是因为目标边缘通常表现为局部空间范围内的灰度/像素值突变 (如物体轮廓与背景的交界处)，而感受野空间特征恰好聚焦于这种局部区域。DARFCB 通过 Group Conv 快速提取感受野空间特征 (替代了低效的

Unfold 方法)，使注意力能直接作用于每个感受野内的边缘细节，避免了传统空间注意力中“权重共享”导致的边缘信息模糊。此外通过协同加权通道与空间注意力、高效聚合边缘信息，并与卷积操作紧密结合，实现了对复杂环境中目标边缘特征的精准提取。

表 1 DARFCB_C2f 各操作步骤的特征图维度变化

Table 1 DARFCB_C2f variation of feature map dimensions for each operation step

输出特征图名称	输出尺寸(C×H×W)
感受野扩展特征$\boldsymbol{F}_{\mathrm{rf}}$	512×120×120
通道全局统计量$\boldsymbol{F}_{\mathrm{gap}}$	512×1×1
降维通道特征$\boldsymbol{F}_{\mathrm{mlp1}}$	32×1×1
升维通道特征$\boldsymbol{F}_{\mathrm{mlp2}}$	512×1×1
通道注意力权重$\boldsymbol{M}_{\mathrm{c}}$	512×1×1
空间平均特征$\boldsymbol{F}_{\mathrm{s_avg}}$	1×120×120
空间最大特征$\boldsymbol{F}_{\mathrm{s_max}}$	1×120×120
拼接空间特征$\boldsymbol{F}_{\mathrm{s_concat}}$	2×120×120
卷积空间特征$\boldsymbol{F}_{\mathrm{s_conv}}$	1×120×120
感受野空间权重$\boldsymbol{M}_{\mathrm{s}}$	1×120×120
加权后感受野特征$\boldsymbol{F}_{\mathrm{weighted}}$	512×120×120
最终优化特征$\boldsymbol{F}_{\mathrm{out}}$	512×40×40

3 实验设计及结果分析

3.1 数据集与实验环境

本文实验使用 VisDrone2019 数据集[22]，共 8629 张图片，其中训练集 6471 张、验证集 548 张、测试集 1610 张，涵盖 pedestrian、people、bicycle、car、van、truck、tricycle、awning-tricycle、bus、motor 十个类别。该数据集小目标密集、尺度跨度大且背景复杂，为验证本文改进算法在复杂航拍场景下的检测性能提供了理想基准。

依据 COCO 评价准则，目标按边界框面积划分为小目标 (<32 pixel×32 pixel)、中目标 (32 pixel×32 pixel~96 pixel×96 pixel) 与大目标 (>96 pixel×96 pixel)。其中，行人 (pedestrian) 与人群 (people) 以小目标为主；自行车 (bicycle)、摩托车 (motor) 及三轮类 (tricycle/awning-tricycle) 亦多呈微小尺度；而汽车 (car)、面包车 (van)、卡车 (truck) 与公交车 (bus) 则跨小、中、大三种尺度分布。目标长宽比集中于 0.5~3.0，呈“等高宽”(如车辆) 或“细高型”(如行人) 两种典型形态。此外，数据集涵盖晴天、阴天、夜晚等多气象条件，并包含密集人群、车辆遮挡及复杂城市场景，进一步增强了实验的说服力。

硬件环境为 Windows 10、16 GB 内存、TUF-RTX 3080Ti 显卡、Intel i7-12700K 处理器，软件环境为 Pytorch 深度学习框架。优化器为 SGD，初始学习率 0.01，批次大小 16，输入图像尺寸 640 pixel×640 pixel，其他超参数默认为默认值，总训练迭代次数 300 次。

3.2 评价指标

实验采用精度 P、召回率 R、平均识别精度mAP来衡量模型的精度。P 表示模型预测为正类别的样本中实际为正类别的比例，R 表示模型成功预测出的正类别样本占所有正类别样本的比例，mAP50表示交并比IOU为 0.5 时所有类别的平均精度均值，mAP50-95是在IOU阈值从 0.5 到 0.95，步长为 0.05 时计算的一组mAP的平均值。计算公式如下：

$$P=\frac{TP}{TP+FP}, \tag{16}$$

$$R=\frac{TP}{TP+FN}, \tag{17}$$

$$mAP=\frac{\sum_{i=1}^{n}\int_{0}^{1}P(R)\mathrm{d}R}{n}, \tag{18}$$

式中：TP表示模型将正样本检测正确的数量；FP表示模型将正样本检测错误的数量；FN表示模型将负样本检测错误的数量；n 为样本类别数；$P(R)$指的是曲线，即不同阈值下的 P 和 R 点连成的曲线。除了模型精度以外，实验还采用浮点运算数FLOPs、参数量Params作为衡量模型复杂度的指标。Params表示网络模型训练时产生的总参数量，FLOPs和Params数值越小代表模型复杂度越低。

3.3 消融实验

为了验证 3 种改进方式对算法检测性能的提升效果，进行了不同改进方法之间的消融实验，A、B、C 分别表示使用 FasterNet 作为骨干网络、引入 Triplet 机制、使用 DARFCB_C2f 模块。实验结果如表 2 所示，单独引入 3 种改进方法时，均能提高模型的各项指标。特别是当使用 DARFCB_C2f 模块，模型的 P 值达到了 47.8%，mAP50 值提升至 36.4%，说明该创新的模块对算法性能的优化作用十分明显。同时使用 DARFCB_C2f 模块以及引入 Triplet 模块后模型的浮点运算数以及参数量都无明显增加，尽管进行骨干网络改进后，浮点运算

量和参数量有所增加，但仍在合理的性能范围内。当引入三种改进方法后，*P* 值提升至 51.7%，*R* 值为 37.9%，mAP50 值达到 39.3%。这些数值在各个指标上均优于任何单一改进方法或两两组合的效果，模型的精度和鲁棒性得到了最大化的提升，性能也达到最佳。

3.4 对比实验

本节通过在相同实验环境和数据集下，进行了一系列对比实验，全面评估了改进算法的有效性和先进性。主要进行了以下几类对比实验：引入不同注意力机制的对比实验、C2f 模块中引入不同结构的对比实验、不同目标检测算法的对比实验，以及改进算法的泛化能力对比实验。

3.4.1 不同注意力模块的对比实验

为了验证 Triplet 注意力机制对于基准算法的提升效果，本文在 YOLOv8n 模型上进行了与其他四种常见的能够有效捕获图像中不同尺度特征的注意力机制的对比实验。实验结果如表 3 所示，表中依次分别为基于注意力的内部尺度特征交互 (attention-based intrascale feature interaction, AIFI)，双层路由注意力 (bi-level routing attention, BRA)，通道先验卷积注意力 (channel prior convolutional attention, CPCA)，多尺度渐进通道注意力 (multi-scale progressive channel attention, MPCA)，三元组注意力。

根据表 3 的结果可以看出，引入 Triplet 注意力机制后，模型在多个关键指标上表现突出。具体来说，引入 Triplet 机制 *P* 值达到 47.6%，显著优于其他注意力机制 (AIFI: 45.1%、BRA: 44.4%、CPCA: 45.3%、MPCA: 45.6%)，增强了对正样本的识别能力。同时，Triplet 机制在 mAP50 值上也达到了 35.1%，超越了其他机制，提升了整体检测准确性。尽管 MPCA 和 MLCA 在召回率上略有优势 (分别为 34.2% 和 34.0%)，但 Triplet 在保持较高召回率的同时，也显著提高了精确率，实现了更优的精度-召回率平衡。综合考虑各项指标后，Triplet 注意力机制更有效地提升了网络的整体性能。

3.4.2 C2f 模块中引入不同结构的对比实验

为了验证设计的 DARFCB 结构对于 C2f 模块的提升效果，本文分别与倒置残差移动块 (inverted residual mobile block, iRMB)、多尺度构建块 (multi-scal block, MS-Block) 以及混合局部通道注意力 (mixed local channel attention, MLCA) 进行对比实验。由于 DARFCB 基于 RFA 和 CBAM 两种结构设计，因此也与 RFA 和 CBAM 模块进行了对比实验。实验结果如表 4 所示，DARFCB 模块在所有模块中表现最佳，其 *P*、*R* 和 mAP50 值分别达到了 47.6%、34.0% 和 35.1%，且 FLOPs 和 Params 分

表 2 消融实验结果

Table 2 Ablation experiment results

YOLOv8n	A	B	C	*P*/%	*R*/%	mAP50/%	mAP50-95/%	FLOPs/G	Params/M
√				46.8	35.0	33.2	20.7	8.1	3.0
√	√			47.6	37.5	37.8	22.3	18.5	6.7
√		√		47.8	35.0	36.4	21.5	8.0	3.1
√			√	47.6	34.0	35.1	19.7	8.1	3.0
√	√	√		48.2	34.1	36.1	21.6	10.8	4.1
√	√		√	49.2	35.0	37.6	21.2	10.7	4.2
√		√	√	47.8	35.2	36.8	21.9	8.3	3.0
√	√	√	√	51.7	37.9	39.3	24.3	10.8	4.2

表 3 五种注意力的性能对比

Table 3 Performance comparison of five types of attention

注意力机制	*P*/%	*R*/%	mAP50/%	FLOPs/G	Params/M
AIFI	45.1	33.7	34.0	8.0	2.9
BRA	44.4	33.9	34.0	8.1	3.3
CPCA	45.3	33.6	34.3	8.3	3.1
MPCA	45.6	34.2	34.3	8.1	3.3
Triplet	47.6	34.0	35.1	8.1	3.0

表 4 C2f 改进模块对比实验结果

Table 4 C2f improvement module comparison experiment results

Module name	*P*/%	*R*/%	mAP50/%	FLOPs/G	Params/M
iRMB	43.4	32.5	32.5	7.2	2.7
MS-Block	44.7	33.7	33.9	7.7	2.9
MLCA	45.8	33.6	34.5	8.1	3.0
RFAConv	44.7	33.4	33.1	8.0	3.1
CBAM	44.5	33.0	33.4	7.1	3.5
DARFCB	47.6	34.0	35.1	8.1	3.0

别为 8.0 G 和 3.1 M，整体性能优于其他模块。相比之下，引入 iRMB 模块虽然浮点运算量和参数量较小，但其 *P* 值和 mAP50 值明显低于其他模块，仅为 43.4% 和 32.5%，且 *R* 值也较低。虽然其他模块均对算法性能有所提升，但 DARFCB 模块仍在各项指标上占据优势。

3.4.3 不同算法的对比实验

为了验证改进后模型的性能，本文选取一些经典模型以及近几年一些优秀的算法，包括 DSM-YOLOv5 (depthwise separable multiplex YOLOv5)[23]、PECS-YOLO (P2-SPPELAN-CSPPC-SEAttention-YOLO)[24]等，分别进行了对比实验。实验结果如表 5 所示，本文提出的模型在 *P* 达 51.7%，较 YOLOv8n 提升 4.9%，超越 YOLOv7-tiny 和 DSM-YOLOv5；*R* 为 37.9%，仅略低于 YOLOv6s；mAP50 指标以 39.3% 位居榜首，较 YOLOv8n 提升 6.1%，超过 DSM-YOLOv5、PECS-YOLO 等改进算法。在计算成本上，其 FLOPs 和 Params 虽略高于 YOLOv8n，但仍低于传统模型如 FasterR-CNN 和 SSD，并优于 YOLOv5s等主流模型。整体来看，改进的模型提高了检测精度，且计算成本较低，具备较好的应用价值。表中 “–” 表示原论文中没有提供此数据。

3.4.4 改进算法与 YOLOv8n 各类别的对比实验

为了展示改进算法在各类目标检测精度上的提升，本文通过表 6 对比了改进前后算法在不同类别目标的检测表现。结果表明，改进后的算法在各类别的精度均有所提升，尤其是在自行车和遮阳三轮车等小尺寸、密集目标上，精度提升显著。尽管在小目标检测方面取得了进展，但整体检测精度并未因此下降。因此，改进算法在密集场景下对小目标检测效果良好，同时也能有效检测不同尺寸的目标。

3.4.5 改进算法的泛化能力对比实验

为了验证改进后模型在不同数据集上的泛化能力，本实验采用拥挤场景下的行人检测数据集 WiderPerson[25]

表 5 不同算法实验结果对比

Table 5 Comparison of experimental results of different algorithms

Model	*P*/%	*R*/%	mAP50/%	FLOPs/G	Params/M
SSD	36.8	26.8	22.9	86.6	22.8
Faster R-CNN	44.7	33.0	28.8	210.3	42.3
YOLOv5s	43.6	32.9	32.6	16.4	7.1
YOLOv6s	45.3	38.2	34.9	11.2	4.5
YOLOv7-tiny	48.2	34.1	36.7	13.1	5.8
YOLOv8n	46.8	35.0	33.2	8.1	3.0
DSM-YOLOv5	51.1	37.0	36.8	–	5.6
Ref.[17]	–	–	30.2	9.0	1.38
PECS-YOLO	–	–	36.8	11.9	2.6
Ours	51.7	37.9	39.3	10.8	4.2

表 6 改进算法与 YOLOv8n 各类别的 mAP50 对比

Table 6 Comparison of mAP50 between the improved algorithm and YOLOv8n for each category

Model/category	mAP50/%									
	Pedestrain	People	Bicycle	Car	Van	Truck	Tricycle	Awning-tricycle	Bus	Motor
YOLOv8n	36.6	30.0	9.2	76.7	40.6	34.1	26.1	13.2	50.8	39.1
Ours	41.8	32.7	13.5	80.6	45.8	36.6	28.5	16.5	58.4	44.8

和 PASCAL VOC[26] 数据集上对改后的模型进行泛化性验证。将 YOLOv5s、YOLOv7-tiny、YOLOv8n 以及本文模型在上述两种数据集上进行实验对比，结果如表 7 所示。可知本文模型在 WiderPerson 数据集上，*P* 和 *R* 值虽未达到最高，但 mAP50 值为最高，表明其平均精度较好；而在 PASCAL VOC 数据集上，本文模型所有指标均为最高，进一步证明了其具有较强的泛化能力。

3.5 实验结果可视化

为了评估三种改进方法的针对性，本文在 VisDrone2019 测试集中选取“小目标”、“多尺度-多样本”以及“暗光环境”三种典型无人机航拍场景，统一采用 Grad-CAM 对输出层进行热力图可视化，结果见图 5。色谱沿用“红热蓝冷”色谱：暖色 (红-橙-黄) 表示高激活与强关注；冷色 (蓝-青-紫) 表示低激活与弱关注。横向对比同一场景可得，以 FasterNet 为骨干，小目标区域暖色骤增，细粒度特征关注度明显提升。引入 Triplet 机制，多尺度-多样本图像暖色覆盖面积扩大，跨尺度特征关联增强。在颈部网络中嵌入 DARFCB-C2f 模块，暗光环境下整体红色区域面积增加。说明三种改进方法分别对于“小目标”、“多尺度-多样本”以及“复杂环境”产生了更具针对性的高激活响应，直观地验证了它们各自的有效性与针对性。

为了更直观地展示本文模型的优势，从 VisDrone2019 测试集中选取了一组能够展示无人机多高度、多视角以及背景复杂的典型图像，与基准模型进行对比，结果如图 6 和图 7 所示。结果表明，改进后的模型在不同场景、各类别上的检测精度均显著优于原网络，同时根据图 7 中低空垂直俯视与侧视示例显示，该模型也减少了漏检和误检现象。整体而言，改进后的算法不仅提升了目标的检测能力，也提高了模型在复杂环境下的适应性和鲁棒性。

表 7 泛化能力对比实验结果

Table 7 Generalisation ability comparison experiment results

Model	WiderPerson			PASCAL VOC		
	P/%	*R*/%	mAP50/%	*P*/%	*R*/%	mAP50/%
YOLOv5s	85.3	82.8	89.1	80.1	71.9	80.3
YOLOv7-tiny	87.2	82.2	89.1	80.6	72.1	80.5
YOLOv8n	87.1	82.2	90.1	81.8	76.2	81.9
Ours	87.0	82.8	90.2	83.4	76.8	84.3

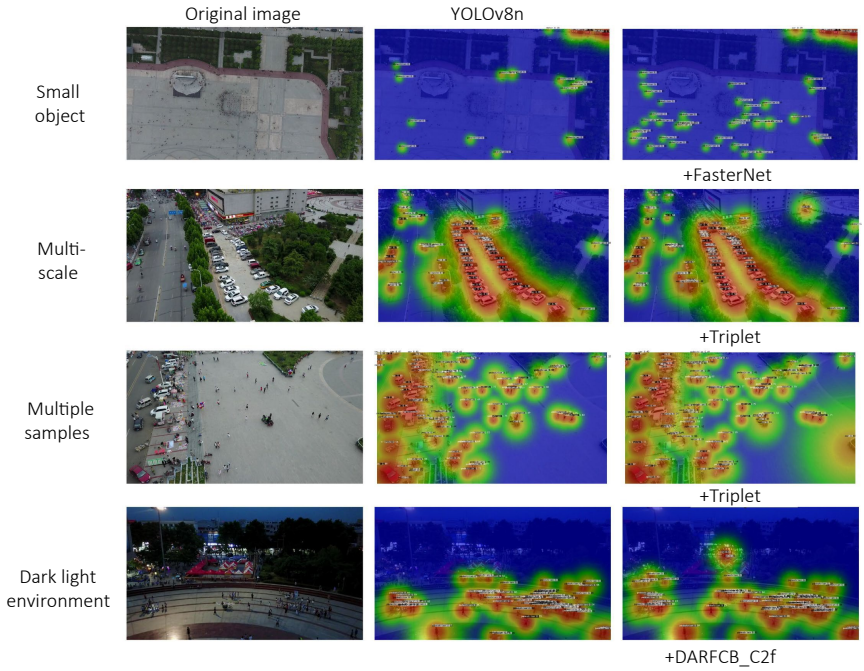


图 5 各改进方法在不同场景下的热力图对比

Fig. 5 Comparison of heatmaps of various improvement methods across different scenarios

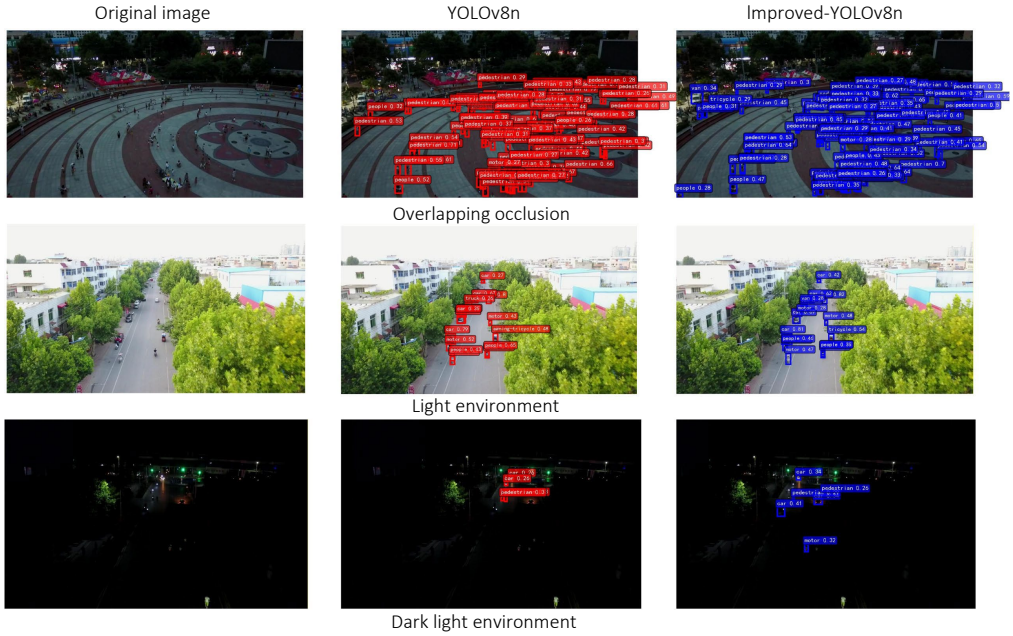


图 6 无人机航拍图像复杂背景下的检测结果对比

Fig. 6 Comparion of detection results for UAV aerial images in complex backgrounds

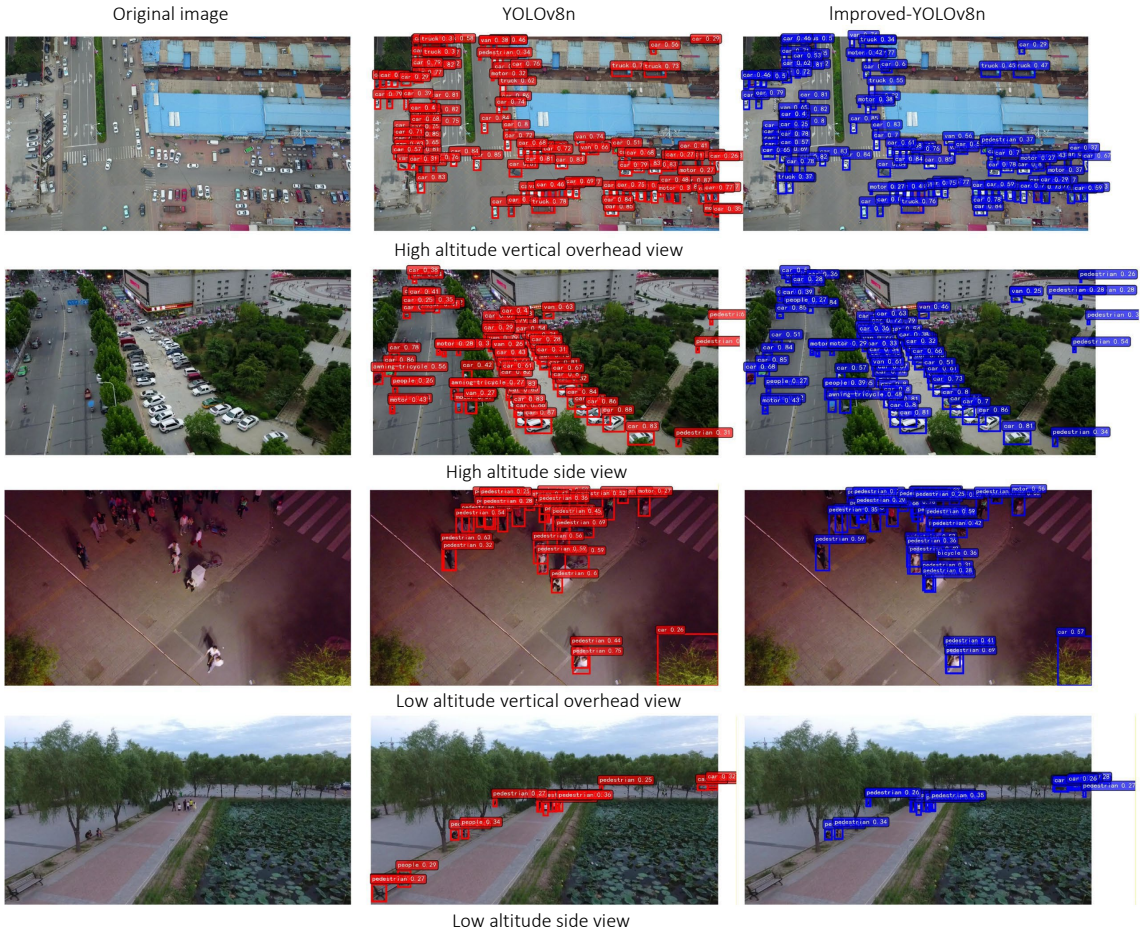


图 7 无人机航拍图像多高度多视角下的检测结果对比

Fig. 7 Comparison of detection results for UAV aerial images in multiple heights and multiple perspectives

4 结 论

针对无人机航拍图像中目标尺度存在差异、背景复杂以及小目标丰富导致检测精度低的问题，本文提出了一种优化解决方案。首先，采用基于部分卷积操作的 Fasternet 架构对原有骨干网络进行改进，以提升小目标的特征提取能力。其次，在骨干网络与颈部网络之间引入三元组注意力机制，增强模型对不同尺度目标的适应性，从而更好地应对无人机航拍图像中的多样本和尺度差异问题。此外，设计 DARFCB_C2f 模块，通过强化边缘特征的提取，进一步提高模型的表现。实验结果表明，改进后的算法在 VisDrone2019 数据集上，检测精度、召回率和平均精度方面均显著优于基准算法，同时有效减少了漏检和误检现象。此外，在其他两个数据集上的实验结果也优于一些经典算法，表明该模型具有较强的适用性和泛化性。未来的研究将致力于在提高检测精度的基础上，实现模型的轻量化设计。

作者贡献

徐昊天：研究资料的收集、分析，论文写作；曾臻：指导论文选题、论文修改；安嵩阳、李肖：负责实验部分、论文格式的修改。

利益冲突

所有作者声明无利益冲突

参考文献

1. Luo X D, Wu Y Q, Chen J L. Research progress on deep learning methods for object detection and semantic segmentation in UAV aerial images[J]. *Acta Aeronaut Astronaut Sin*, 2024, **45**(6): 028822 (in Chinese).
罗旭东, 吴一全, 陈金林. 无人机航拍影像目标检测与语义分割的深度学习方法研究进展[J]. 航空学报, 2024, **45**(6): 028822.
2. He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Trans Pattern Anal Mach Intell*, 2015, **37**(9): 1904−1916.
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*, 2014: 580–587. https://doi.org/10.1109/CVPR.2014.81.
4. Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, **39**(6): 1137−1149.
5. Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]//*14th European Conference on Computer Vision – ECCV 2016*, 2016: 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.
6. Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2017: 2999–3007. https://doi.org/10.1109/ICCV.2017.324.
7. Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//*2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016: 779–788. https://doi.org/10.1109/CVPR.2016.91.
8. Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]// *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017: 6517–6525. https://doi.org/10.1109/CVPR.2017.690.
9. 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]// *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023: 7464–7475. https://doi.org/10.1109/CVPR52729.2023.00721.
10. Li K X, Liu X F, Chen Q, et al. YOLOv8-GAIS: improved object detection algorithm for UAV aerial photography[J]. *Opto-Electron Eng*, 2025, **52**(4): 240295.
李凯璇, 刘晓锋, 陈强, 等. YOLOv8-GAIS: 一种改进的无人机航拍目标检测算法[J]. 光电工程, 2025, **52**(4): 240295.
11. Zhang Y Z, Wu C Y, Zhang T, et al. Self-attention guidance and multiscale feature fusion-based UAV image object detection[J]. *IEEE Geosci Remote Sens Lett*, 2023, **20**: 6004305.
12. Zeng S, Yang W Z, Jiao Y Y, et al. SCA-YOLO: a new small object detection model for UAV images[J]. *Vis Comput*, 2024, **40**(3): 1787−1803.
13. Zhang R M, Xiao Y F, Jia Z N, et al. Improved YOLOv7 algorithm for target detection in complex environments from UAV perspective[J]. *Opto-Electron Eng*, 2024, **51**(5): 240051.
张润梅, 肖钰霏, 贾振楠, 等. 改进 YOLOv7 的无人机视角下复杂环境目标检测算法[J]. 光电工程, 2024, **51**(5): 240051.
14. Sun F X, He N, Li R J, et al. GD-PAN: a multiscale fusion architecture applied to object detection in UAV aerial images[J]. *Multimedia Syst*, 2024, **30**(3): 143.
15. Hui Y M, Wang J, Li B. WSA-YOLO: weak-supervised and adaptive object detection in the low-light environment for YOLOV7[J]. *IEEE Trans Instrum Meas*, 2024, **73**: 2507012.
16. Jiang L J, Yuan B X, Du J W, et al. MFFSODNet: multiscale feature fusion small object detection network for UAV aerial images[J]. *IEEE Trans Instrum Meas*, 2024, **73**: 5015214.
17. Zhu L Z, Wei H. Unmanned aerial vehicle small target detection algorithm based on YOLOv8n[J]. *Laser Optoelectron Prog*, 2025, **62**(10): 1037009.
朱立志, 韦慧. 基于 YOLOv8n 的无人机小目标检测算法[J]. 激光与光电子学进展, 2025, **62**(10): 1037009.
18. Chen J R, Kao S H, He H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 12021–12031. https://doi.org/10.1109/CVPR52729.2023.01157.
19. Misra D, Nalamada T, Arasanipalai A U, et al. Rotate to attend: convolutional triplet attention module[C]//*Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2021: 3138–3147. https://doi.org/10.1109/WACV48630.2021.00318.
20. Zhang X, Liu C, Yang D G, et al. RFAConv: innovating spatial attention and standard convolutional operation[Z]. arXiv: 2304.03198, 2023. https://doi.org/10.48550/arXiv.2304.03198.
21. Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 3–19. https://doi.org/10.1007/978-3-030-

01234-2_1.
22. 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 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019: 213–226. https://doi.org/10.1109/ICCVW.2019.00030.
23. Chen W B, Jia X J, Zhu X B, et al. Target detection for UAV image based on DSM-YOLO v5[J]. *Comput Eng Appl*, 2023, **59**(18): 226−233.
陈卫彪, 贾小军, 朱响斌, 等. 基于 DSM-YOLO v5 的无人机航拍图像目标检测[J]. 计算机工程与应用, 2023, **59**(18): 226−233.
24. Wang S M, XU H Y, Zhu X Z, et al. Lightweight small object detection algorithm for aerial photography based on improved YOLOv8n: PECS-YOLO[J]. *Comput Eng*, 2025, **51**(9): 280−293.
王舒梦, 徐慧英, 朱信忠, 等. 基于改进 YOLOv8n 的航拍轻量化小目标检测算法: PECS-YOLO[J]. 计算机工程, 2025, **51**(9): 280−293.
25. Zhang S F, Xie Y L, Wan J, et al. WiderPerson: a diverse dataset for dense pedestrian detection in the wild[J]. *IEEE Trans Multimedia*, 2020, **22**(2): 380−393.
26. Everingham M, Van Gool L, Williams C K I, et al. The PASCAL visual object classes (VOC) challenge[J]. *Int J Comput Vis*, 2010, **88**(2): 303−338.

作者简介



徐昊天 (2000-)，男，湖北工业大学机械工程学院硕士研究生，主要研究目标检测、信号处理领域。
E-mail：102300016@hbut.edu.cn



【通信作者】曾臻 (1986-)，男，湖北工业大学机械工程学院讲师，主要研究图像处理、传感器领域。
E-mail：zengzhen@hbut.edu.cn

Attention-fused target detection algorithm for UAV aerial images

Xu Haotian[1,2], Zeng Zhen[1,2*], An Songyang[1,2], Li Xiao[1,2]

[1] School of Mechanical Engineering, Hubei University of Technology, Wuhan, Hubei 430068，China;
[2] Hubei Key Laboratory of Modern Manufacturing Quality Engineering, Wuhan, Hubei 430068，China

Abstract

Objective Unmanned aerial vehicle (UAV) imagery poses enduring challenges for object detection, largely because targets are frequently small, densely distributed, and captured under heterogeneous imaging conditions. In contrast to ground-level scenes, aerial views often exhibit pronounced scale variation, viewpoint changes, partial occlusions, and severe background clutter arising from man-made structures, vegetation, road textures, and illumination-induced shadows. In addition, motion blur and sensor noise are common due to platform dynamics and altitude-dependent resolution constraints. Collectively, these factors exacerbate the loss of fine-grained cues during feature downsampling and increase both missed detections and false alarms, particularly when small objects occupy only a few pixels and their boundaries are confounded with background patterns.

Methods This work presents an attention-fused object detection approach built upon an improved YOLOv8n framework to alleviate the above limitations. Although YOLOv8n is attractive for edge deployment owing to its lightweight design, its baseline configuration can underperform on UAV datasets when high-frequency details are attenuated by successive convolution and pooling operations and when multi-scale feature aggregation is insufficiently discriminative. To strengthen representation learning while retaining computational efficiency, we introduce three complementary modifications: (i) a backbone replacement based on FasterNet to better preserve small-object structure, (ii) an inter-stage Triplet attention mechanism to model coupled channel-spatial dependencies, and (iii) a neck-level enhancement module that integrates receptive-field attention convolution (RFAConv) with the convolutional block attention module (CBAM) via a newly designed DARFCB-C2f block.

First, the original YOLOv8n backbone is replaced with FasterNet to improve the retention of fine details that are critical for detecting small targets. FasterNet employs partial convolution (PConv) in conjunction with pointwise convolution (PWConv) to balance representational capacity and computational cost. By applying convolution to only a subset of channels, PConv reduces redundant computation while maintaining local structural information, whereas PWConv performs channel mixing to sustain expressive feature

Foundation item: Hubei Innovation Group Program (2022CFA006)
*Correspondence: Zeng Z, E-mail: zengzhen@hbut.edu.cn

interactions. Notably, the combination of PConv and PWConv yields a T-shaped restricted receptive-field pattern that biases attention toward central feature-map regions. Given that small objects typically occupy limited spatial extents, preserving and emphasizing central responses can mitigate the information loss induced by downsampling. As a result, the revised backbone enhances small-object feature fidelity without substantially increasing model complexity.

Second, to better accommodate the coexistence of multiple target categories and large scale disparities within a single aerial frame, Triplet attention is inserted between the backbone and the neck. Conventional attention modules often compute channel and spatial attention independently, which may limit their ability to capture joint interactions across dimensions. Triplet attention addresses this by modeling dependencies through three coordinated branches: channel-spatial height, channel-spatial width, and spatial height-width. This triadic formulation enables richer coupling between channel and spatial representations, thereby providing a more comprehensive description of salient object patterns across scales. In dense UAV scenes, where small and medium-sized objects may appear simultaneously and compete for representational resources, the proposed attention insertion improves feature selectivity and enhances robustness to scale variation.

Third, we redesign the neck feature-fusion blocks by proposing a DARFCB structure and constructing a DARFCB-C2f module that replaces the standard C2f blocks in YOLOv8n. The design leverages RFAConv to expand the effective receptive field and to perform attention-guided feature fusion, which is beneficial for incorporating contextual cues without sacrificing boundary-sensitive information. Such boundary cues are particularly important in UAV imagery, where object contours are frequently degraded by blur, low resolution, and background interference. In addition, CBAM is integrated to refine intermediate representations via channel and spatial attention, amplifying informative regions while suppressing irrelevant responses. Together, RFAConv and CBAM reinforce edge-related features and improve the reliability of multi-scale aggregation, leading to more stable detection under complex environmental conditions.

Results and Discussions Comprehensive experiments are conducted on the VisDrone2019 benchmark, which is widely used for UAV object detection and contains challenging scenarios including dense traffic, crowded pedestrian regions, substantial scale variation, and cluttered backgrounds. Ablation studies and module replacement analyses demonstrate that each proposed component contributes measurable gains, and their combination yields the strongest overall performance. Relative to the baseline YOLOv8n detector, the proposed method improves precision (P) by 4.9%, recall (R) by 3.9%, and mAP50 by 6.1%. These improvements reflect a meaningful reduction in both false positives and false negatives, indicating that the attention-fused design effectively enhances the discriminability and completeness of detection results in UAV scenes.

To assess cross-domain robustness, additional evaluations are performed on the WiderPerson and PASCAL VOC datasets, which differ from VisDrone2019 in terms of scene composition, object density, and imaging characteristics. The proposed model maintains strong performance across these datasets, suggesting that the introduced backbone redesign, attention fusion strategy, and neck enhancement provide transferable benefits rather than dataset-specific overfitting. This generalization evidence supports the applicability of the approach to practical UAV-driven tasks such as aerial surveillance, traffic monitoring, and search-and-rescue operations.

Conclusions In summary, this paper proposes an attention-fused YOLOv8n-based detector for UAV aerial imagery that integrates three synergistic improvements: a FasterNet backbone to preserve small-target details and alleviate downsampling-induced information loss, triplet attention to capture richer multi-dimensional dependencies for multi-scale recognition, and a DARFCB-C2f neck module that combines RFAConv and CBAM to strengthen edge features and improve robustness in cluttered environments. Experimental results on VisDrone2019 and cross-dataset validations on WiderPerson and PASCAL VOC collectively confirm that the proposed modifications substantially improve detection accuracy while retaining the efficiency advantages of a lightweight one-stage detector.

Keywords: object detection; UAV aerial images; FasterNet; triplet attention; receptive-field attention convolution

Xu H T, Zeng Z, An S Y et al. Attention-fused target detection algorithm for UAV aerial images[J]. *Opto-Electron Eng*, 2026, **53**(1): 250204; DOI: 10.12086/oee.2026.250204

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文