FEMS-YOLO: 基于特征增强多尺度融合的无人航拍图像目标检测

张开玉[1,2]，王浩杰[1,2]，卢 迪[1,2†]

(1. 哈尔滨理工大学 测控技术与通信工程学院，哈尔滨 150080；
2. 哈尔滨理工大学 模式识别与信息感知黑龙江省重点实验室，哈尔滨 150080)

摘 要: 随着无人机技术在城市安防、交通监管和应急救援等领域的快速发展, 无人机图像的目标检测与识别技术为多行业应用提供了可靠的技术支持. 低空视角的目标检测任务面临小目标密集、尺度变化大、背景复杂等挑战. 针对上述问题, 本文提出一种改进的 YOLO11 无人机图像目标检测算法. 首先, 设计 CSP-MS 模块通过分层融合和异构卷积结构实现多尺度特征的表达; 其次, 设计特征增强多尺度特征聚合金字塔模块, 通过空洞卷积与跨层融合机制提高模型对复杂场景的感知能力; 最后引入轻量级动态任务对齐检测头, 降低模型参数量的同时提升对小尺寸目标的检测精度. 模型在 VisDrone 数据集上 mAP0.5 和 mAP0.5:0.95 指标分别提升 10.2% 和 6.7%, 在 CODrone 数据集上分别提升 5.4% 和 3.7%, 实验结果表明, 改进模型在小目标、复杂背景和多尺度目标场景中均具有显著性能优势, 体现出较强的泛化能力和实用价值.

关键词: 无人机航拍图像；目标检测；YOLO11；多尺度特征提取；特征增强多尺度特征聚合金字塔

中图分类号: TP391.41 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2025.0943

引用格式: 张开玉, 王浩杰, 卢迪. FEMS-YOLO: 基于特征增强多尺度融合的无人航拍图像目标检测 [J]. 控制与决策, xxxx, x(x): xxxx-xxxx.

FEMS-YOLO: An UAV image target detection based on feature-enhanced multi-scale fusion

ZHANG Kai-yu[1,2]，*WANG Hao-jie*[1,2]，*LU Di*[1,2†]

(1. 150080；2. 150080)

Abstract: With the rapid development of unmanned aerial vehicle (UAV) technology in urban security, traffic monitoring, and emergency response, object detection and recognition based on UAV imagery have become essential for supporting a wide range of intelligent applications. However, UAV-based object detection from low-altitude perspectives remains highly challenging due to the dense distribution of small objects, significant scale variations, and complex background interference. To address these issues, this paper proposes an enhanced UAV-oriented YOLO11 object detection algorithm. First, a CSP-MS module is designed to improve multi-scale feature representation through hierarchical fusion and heterogeneous convolution structures. Second, an feature-enhanced multi-scale aggregation pyramid is introduced, which combines dilated convolutions with cross-layer fusion to strengthen the model's perception capability in complex scenes. Finally, a lightweight dynamic task-aligned detection head is integrated to reduce model parameters while improving detection accuracy for small objects. Experiments on the VisDrone dataset demonstrate improvements of 10.2% in mAP0.5and 6.7% in mAP0.5:0.95.On the CODrone dataset, the proposed method achieves gains of 5.4% and 3.7%, respectively. Overall, the results show that the improved model delivers notable advantages in detecting small objects, handling complex backgrounds, and managing multi-scale targets, highlighting its strong generalization capability and practical applicability.

Keywords: UAV aerial images；object detection；YOLO11；multi-scale feature extraction；feature-enhanced multi-scale feature aggregation pyramid

0 引 言

随着近年来无人机技术的发展, 无人机技术广泛应用于精准农业, 交通运输, 城市交通等领域[1-3]. 无人机系统突破了传统固定摄像设备的视角限制,

收稿日期: 2025-09-09；**录用日期:** 2026-01-12.
†通信作者. E-mail: zhangkaiyu@hrbust.edu.cn.

凭借其低成本、高灵活性和独特的多视角采集能力等优势, 能够从低空获取各类复杂场景的高分辨率图像数据. 对无人机图像中各类目标高效精准地识别与定位, 为多个领域应用提供智能化解决方案.

现有目标检测算法可依据是否依赖锚框与候选区域生成过程划分为二阶段检测算法与一阶段检测算法. 典型的二阶段检测算法包括 R-CNN[4]、Faster R-CNN[5] 等, 其核心思想是首先利用区域提议网络 (Region Proposal Network, RPN) 生成候选区域 (Region Proposals), 随后对候选区域进行分类和边界框回归, 从而获得较高的检测精度. 然而, 该类方法因候选区域生成过程带来的高计算复杂度和参数开销, 推理速度相对较慢. 一阶段检测算法则将目标检测任务建模为端到端的回归问题, 直接预测目标类别及其位置, 无需生成候选框. 代表性方法包括 YOLO[6] 和 SSD[7], 其中基于 YOLO 的算法因其推理速度快、检测精度较高而成为当前研究的热点.

为了使上述目标检测技术更好地适用于无人机图像目标检测任务, Zhu 等人[8] 在 YOLOv5 中设计 Transformer 检测头和引入了卷积注意力模块, 提高了在高密度场景中小目标的检测准确性, 但也增加了额外的计算开销; 侯等人[9] 通过在 YOLOv8 骨干网络中引入 SPDConv 与 Biformer 注意力机制增强了模型对小目标细节信息的提取能力, 用重参数化网络优化颈部网络从而提高模型推理速度, 但模型的参数量也大幅度提高; Zhou 等人[10] 引入非稀疏的注意力机制并将高层信息与底层信息融合, 有效提取与任务相关的非语义特征, 拓展了网络的感受野, 提高模型在复杂场景下目标的检测性能; Jiang 等人[11] 同时将 AKConv 与 EMA 模块加入到 YOLOv8 的特征提取模块 C2f 中, 结合二者的优势, 以增强特征提取能力并降低参数量; 引入 BiFPN 模块, 通过加权特征融合保留更多浅层特征信息, 有效缓解遮挡带来的漏检问题; Zhang 等人[12] 提出了 Drone-YOLO 设计了 Sandwich 特征融合模块, 增强浅层与深层特征融合能力, 提高了多个场景下小目标的检测能力, 并通过 RepVGG 重参数化卷积模块提高模型的学习特征的效率; Zhang[13] 等人通过引入 SPD 卷积模块和平均空间金字塔池化模块增强了小目标的特征提取能力, 并通过嵌入归一化注意力机制 (NAM) 抑制无效特征, 从而提高了目标检测的效率. Shen 等人[14] 引入感受野混合注意力卷积, 不增加额外计算开销的同时提高模型的特征提取能力, 减少因目标遮挡引起的漏检.

尽管无人机图像目标检测近年来取得了显著进展, 但在实际场景应用中仍存在局限性与挑战: (1) 无人机在低空进行拍摄, 图像中小目标实例数量众多, 且目标尺寸相对较小, 导致特征表达不充分, 严重影响检测精度; (2) 无人机拍摄过程中受飞行高度、视角变化等因素影响, 同一类别目标在图像中的尺度差异较大, 容易引发漏检与误检; (3) 无人机具备较强的灵活性, 可在多种复杂环境中获取图像, 导致图像中背景元素复杂、干扰信息多, 从而增加了模型对目标与背景区分的难度, 影响检测鲁棒性.

综合已有研究可以发现, 当前无人机目标检测方法多通过引入复杂注意力机制如 Biformer 或卷积结构来增强特征提取能力, 从而提升小目标的检测性能. 然而, 这类方法往往伴随着模型参数量和计算复杂度的显著增加, 限制了其在资源受限场景下的应用. 同时, 这些方法在多尺度特征建模方面仍存在不足, 主干网络在不同特征提取阶段普遍采用固定尺寸卷积核, 未能充分适配特征金字塔中不同层级特征的表征需求. 此外, 现有研究通过改进特征融合结构如 BiFPN、Sandwich 特征融合模块等加强浅层与深层特征的交互, 以缓解多尺度特征利用不足的问题, 但此类融合方式主要依赖特征加权或简单的跨层拼接, 对融合特征的复用与选择性增强能力有限. 但在背景复杂的场景下, 难以有效抑制冗余特征, 关键判别特征难以充分突出, 导致检测性能提升仍受限.

针对以上问题, 本文设计了 FEMS-YOLO 网络: 一种基于 YOLO11 的高效多特征融合的无人机图像目标检测算法: (1) 设计 CSP-MS 模块 (Cross Stage Partial Multi-Scale Module), 将多尺度模块与 CSPNet 相结合, 使用分层融合机制提升模型的特征提取能力, 并在主干网络不同阶段采用异构卷积核提高模型多尺度特征提取能力; (2) 设计特征增强多尺度特征聚合金字塔模块 (FEMS-DFPN, (Feature-enhanced Multi-Scale Feature with Dual Path Aggregation), 融合不同尺度的特征, 随后经过并行空洞卷积进行特征增强, 获取无人机目标的全局和局部信息, 将融合特征融入多路特征聚合金字塔网络, 并扩散到各个检测层, 提升在复杂背景下的检测能力; (3) 引入动态任务对齐检测头, 通过特征参数共享来降低模型参数量, 提高模型的对小尺寸目标检测性能.

1 本文算法

1.1 FEMS-YOLO 无人机图像目标检测网络

本文以 YOLO11 为基线模型进行改进, 提出 FEMS-YOLO 网络, 结构如图 1 所示. 分别对主干网

络、颈部网络和检测头进行改进. 主干网络设计CSP-MS 模块进行特征提取, 增强多尺度目标检测能力, 颈部网络增加小目标检测层, 设计特征增强多尺度特征聚合金字塔模块, 融合多个尺度特征, 提升无人机图像在复杂背景下检测能力. 最后引入动态任务对齐检测头减少参数量同时提升模型对小目标的定位的准确性.

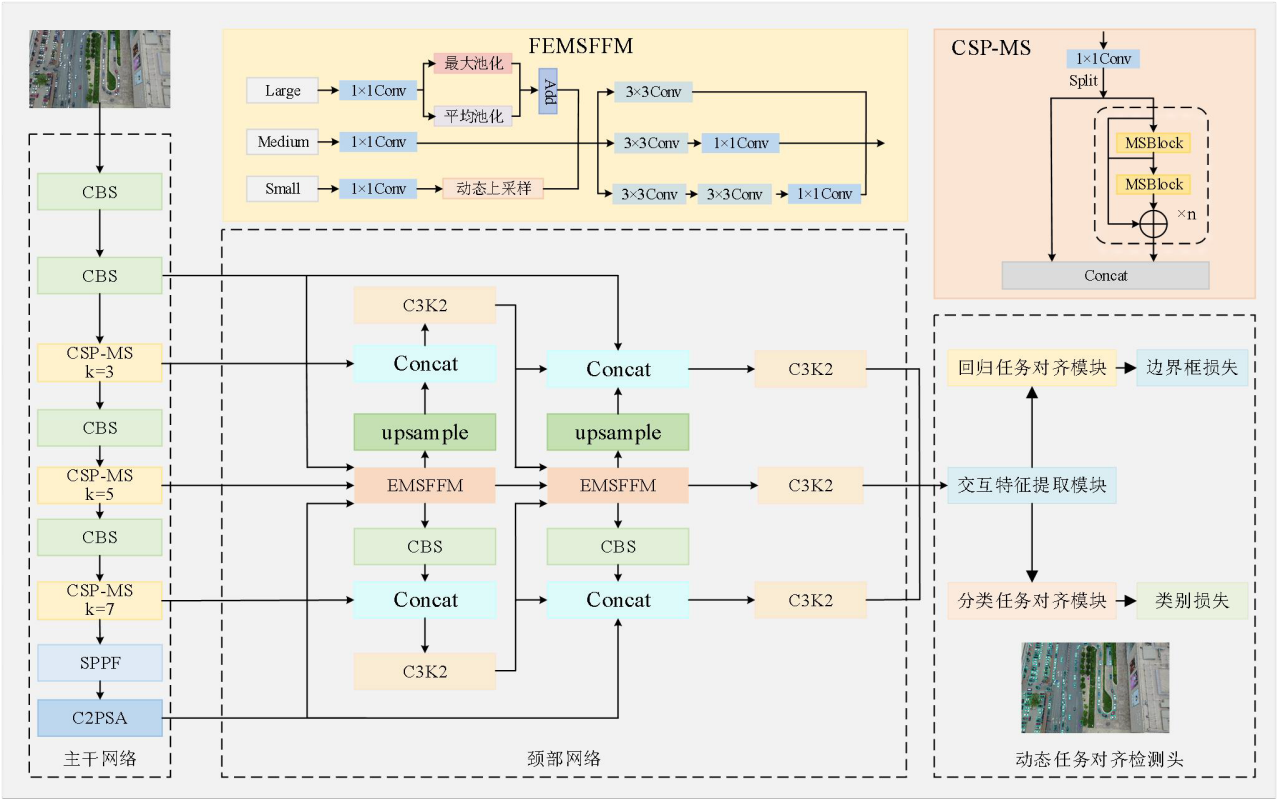

图1　FEMS-YOLO 网络结构

1.2　CSP-MS 模块

现有 YOLO 系列主干网络的不同特征提取阶段均使用固定尺寸的卷积核. 在特征金字塔网络中, 各层级特征具有差异化的表征特性: 浅层高分辨率特征主要包含细粒度空间信息, 而深层低分辨率特征包括高级语义信息, 负责大尺度目标的识别. 而在深层网络阶段持续采用统一的小尺寸卷积核会显著制约有效感受野, 致使不能充分捕获大尺度目标所需的全局上下文信息. 难以实现多尺度语义特征的最优提取. 针对该问题, 设计了 CSP-MS 模块, CSP-MS 将多尺度模块[15](Muti-scale Block, MSBlock) 融入 CSPNet. 该结构在通道维度上将输入特征划分为两个子分支; 第一分支直接用于后续特征融合, 第二分支经由多个堆叠的多尺度模块进行深度特征提取. 该设计充分保留了 CSPNet 的梯度分流与特征重用机制, 同时结合 MSBlock 的多尺度特征提取优势, 进一步提升了特征层级之间的协同表达能力.CSP-MS 模块如图 2 所示.

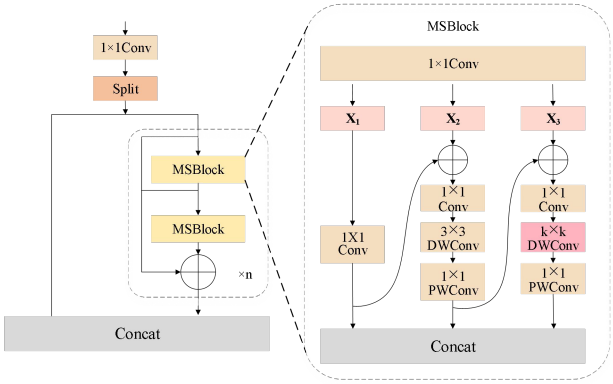

图2　CSP-MS 模块

多尺度模块中设计了分层特征融合机制, 输入特征X首先通过 1×1 卷积进行通道扩展后划分为 3 个子通道X_1, X_2, X_3, 如公式 1 所示.

$$[X_1, X_2, X_3] = \text{Split}(X). \tag{1}$$

每个子通道送入独立的分支路径, 分别对应 1×1, 3×3 和 k×k 的深度可分离卷积. 将前一尺度的输出引入当前分支, 以增强尺度间的信息交互. 如公式 2 所示.

$$Y_1 = \mathrm{PWConv}(\mathrm{Conv3} \times 3^{\mathrm{DW}}(X_1)),$$
$$Y_2 = \mathrm{PWConv}((\mathrm{Conv3} \times 3^{\mathrm{DW}}(X_2 + Y_1)),$$
$$Y_3 = \mathrm{PWConv}(\mathrm{Conv}K \times K^{\mathrm{DW}}(X_3 + Y_2)). \quad (2)$$

其中$\mathrm{Conv}^{\mathrm{DW}}$代表逐深度卷积, PWConv代表逐点卷积.

将三个路径的的输出沿通道维度进行拼接得到最终输出Y_{out}. 如公式 3 所示.

$$Y_{\mathrm{out}} = \mathrm{Concat}(Y_1, Y_2, Y_3). \quad (3)$$

通过设计多尺度并行卷积结构, 可同时提取局部与上下文信息, 有效增强模型对无人机图像中小目标和复杂场景目标的表征能力. 在骨干网络特征提取阶段中 k×k 的卷积分别采用大小为 3×3, 5×5 和 7×7 卷积核, 提高模型对无人图像中多尺度目标特征的表达能力.

1.3 特征增强多尺度特征聚合金字塔模块

为了解决在复杂场景下检测效果不佳的问题, 本文对颈部网络进行改进, 颈部网络改进如图 3 所示. 相比 YOLO11 网络结构如图 3(a) 所示, 本文设计特征增强多尺度特征聚合金字塔模块 (FEMS-DFPN) 如图 3(b) 所示, 引入一个小目标检测层并将高效多尺度融合模块 (FEMSFFM) 加入到双路径特征聚合金字塔网络 (DPFN) 中. 通过将不同尺度特征融合, 并将多尺度特征扩散至各个检测层, 提升复杂场景下的检测精度与鲁棒性.

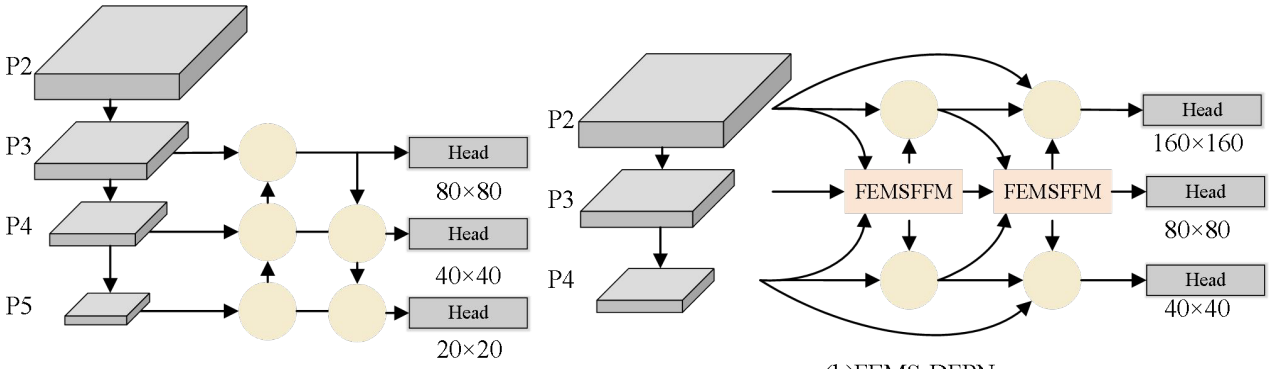

图3 颈部网络改进

1.3.1 小目标检测层

由于拍摄高度导致无人机图像中存在大量小尺寸目标, 本文引入 160×160 分辨率小目标检测层, 该设计保留更多的浅层信息, 增强模型对细节信息的敏感性, 提高对小目标定位的准确性. 删除了 20×20 分辨率的大目标检测层, 有效减少在特征提取阶段对小目标语义信息丢失和浅层特征信息的冗余, 该设计减少下采样对小目标特征的影响, 强化网络对细节特征表达能力, 在降低参数复杂度的同时改善小目标检测性能.

1.3.2 特征增强多尺度特征融合模块

在 YOLO11 的颈部网络结构中单向特征传递路径使模型在浅层局部细节与深层语义信息的融合方面存在不足, 影响了跨层级特征的有效表达. 导致无人机图像小目标特征在传递过程存在丢失, 为解决该问题, 设计特征增强多尺度特征融合模块 (Feature-enhanced Multi-Scale Feature Fusion Module, FEMSFFM), 通过融合三个不同尺度的特征, 并使用不同空洞率的空洞卷积进行特征增强, 提高复杂背景下的检测性能. 网络结构如图 4 所示在该模块中首先使用 1×1 卷积将不同尺度输入特征F_i的通道调整至中等特征图维度, 如公式 (4) 所示.

$$\overline{F}_i = Conv_{1\times1}(F_i),\ i \in \{s, m, l\}. \quad (4)$$

其中$Conv_{1\times1}$表示 1×1 卷积的卷积操作, $F_i,\ i \in \{s, m, l\}$表示三个来自主干网络中大中小三个不同尺度的特征. 随后, 对不同尺度的特征分别处理, 处理大尺度特征图$\overline{F}_l'$时使用最大池化和平均池化级联结构来进行下采样, 如公式 (5) 所示, 这种结构可以保留无人机图像目标的主要特征, 在降低空间维度的同时有效的提高模型在复杂背景下对细节特征理解能力.

$$\overline{F}_l' = \mathrm{Add}[\mathrm{Avg}(\overline{F}_l), Max(\overline{F}_l)]. \quad (5)$$

$Avg()$, $Max()$分别表示平均池化和最大池化.

对于小尺度特征图$\overline{F}_s$, 采用轻量级的动态上采样[16], 通过点采样的上采样方式, 该采样方式首先通过采样点生成器生成采样点集, 随后经过点采样, 根据生成的采样点集对插值后的特征图进行重采样, 得到最终的上采样图. 通过动态偏移调整采样位置, 更精准地保留无人机小目标的边缘信息, 动态上采样过程如公式 (6) 所示.

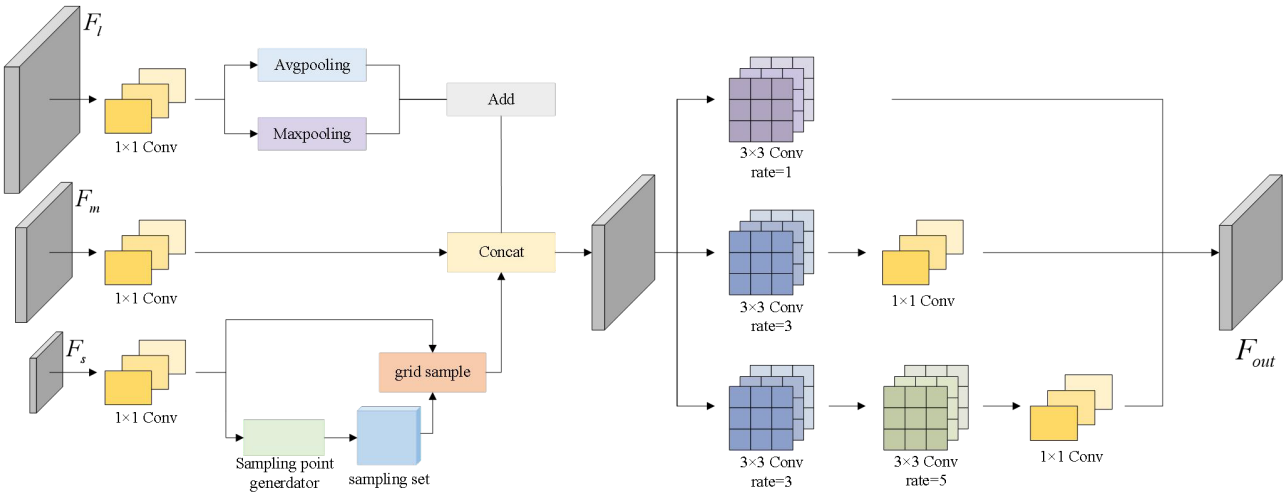

图4　FEMSFFM 结构图

$$\overline{F}_s' = \text{grid\_sample}(\overline{F}_s, S). \tag{6}$$

S表示特征$\overline{F}_s'$经过点采样生成器产生的采样集

将三个尺度的特征图在通道维度拼接得到输出$\overline{F}'$, 如公式 (7) 所示.

$$\overline{F}' = Concat(\overline{F}_s', \overline{F}_m', \overline{F}_l'). \tag{7}$$

随后, 将融合后的特征划分为三个分支, 每个分支采用空洞率不同的 3×3 空洞卷积进行特征提取, 以增强网络对复杂场景的适应能力. 空洞率较大的卷积能够捕获丰富的上下文信息, 从而提升网络在复杂场景下无人机图像检测的鲁棒性; 较小的卷积提取无人机图像小目标的细节特征. 最终, 将三个分支的特征融合以生成网络输出, 实现对局部细节与全局信息的兼顾. 特征增强过程可表示为:

$$\overline{F}_1' = \text{Conv}_{3\times3}^{(d=1)}(\overline{F}'), \tag{8}$$

$$\overline{F}_2' = \text{Conv}_{1\times1}[\text{Conv}_{3\times3}^{(d=3)}(\overline{F}'), \tag{9}$$

$$\overline{F}_3' = \text{Conv}_{1\times1}\{\text{Conv}_{3\times3}^{(d=5)}[\text{Conv}_{3\times3}^{(d=3)}(\overline{F}')], \tag{10}$$

$$F_{out} = Concat(\overline{F}_1', \overline{F}_2', \overline{F}_3'). \tag{11}$$

$\text{Conv}_{3\times3}^{(d=3)}$, $\text{Conv}_{3\times3}^{(d=5)}$分别表示空洞率为 3 和 5 的 3×3 卷积

1.3.3　双路径特征聚合金字塔网络

YOLO11 的特征金字塔网络 PAN-FPN 主要实现自顶向下的 FPN 路径, 以及自底向上的增强路径, 在背景复杂场景检测任务中由于特征融合不充分、信息传导路径短、跨层连接不足等问题, 会导致信息损失和识别能力下降. 本文引入双路特征聚合金字塔网络 (Dual Path Aggregation Net,DPAN), 网络结构图如图 5 所示.

DPAN 的构建两条并行的特征传播路径以实现多尺度特征的双向扩散与深度融合. 一条为自顶向下路径对小尺度特征图最近邻上采样后与中等尺度特征图拼接, 将高层语义信息向低层传递, 增强小目标检测中的上下文理解能力; 另一条为自底向上路径通过 3×3 的卷积核对大尺度特征图进行下采样后与中等尺度特征图拼接, 反向注入精细空间细节, 缓解因过度抽象而导致的边界模糊问题. 这种并行的双路径设计打破了传统 FPN 仅依赖单向信息流的限制, 实现了特征在不同层级间的互补增强, 从而更有效地保留了图像的空间结构与语义信息.

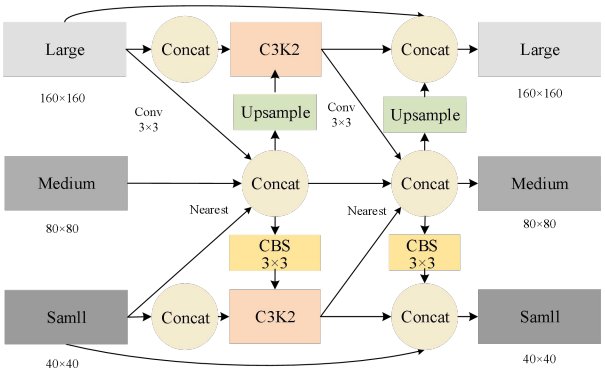

图5　双路特征聚合金字塔网络

DPAN 金字塔网络通过将不同特征尺度特征图进行特征融合, 然后对融合后的特征图分别进行上采样和下采样, 随后重复上一步操作, 将融合后的多尺度特征扩散到各个检测, 此外, 该聚合网络还增加了来自主干网络的跳跃连接, 将浅层细节特征融入到检测层中,, 避免了梯度消失风险, 同时增强了模型对细粒度纹理、边缘等局部特征的敏感性. 本文将特征增强多尺度特征融合模块与双路径聚合特征金字塔网络相结合, 有效缓解了特征提取与融合过程中信息损失的问题.

1.4　动态任务对齐检测头

在 YOLO11 检测头结构中, 处理边界框回归和分类的任务的解耦头是相互独立的. 这种独立的结

构使模型增加了额外的参数量, 为了解决这些问题, 本文引入了一种参数共享的动态任务对齐检测头 (Dynamic Task-aligned Head, DTH)[17], 使模型在不额外增加参数量的同时提升对小目标的检测性能. 动态任务对齐检测头如图 6 所示共分为三个部分:

1) 交互特征提取模块. 对特征金字塔中不同尺度的特征分别进行特征提取, 为分类和回归任务提供多尺度的特征表示. 特征提取过程如公式 (12) 所示.

$$X_{\text{inter}} = SiLU(GN(C(X^{\text{fpn}}))). \tag{12}$$

其中GN代表组归一化, $SiLU$表示 3×3 的卷积. X^{fpn}代表特征金字塔 P3, P4, P5 中的特征. $SiLU$表示激活函数

2) 回归任务对齐模块. 将提取到的交互特征一个分支通过卷积层, 获得可变形卷积所需的偏移量和调制标量, 另一个分支通过层注意力机制, 如图 7 所示.

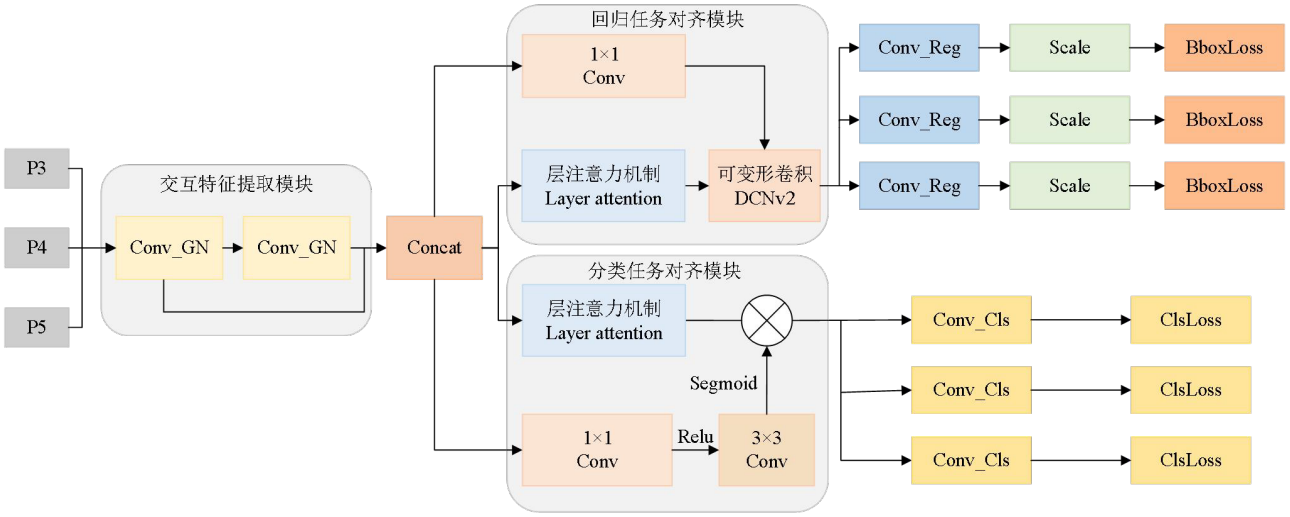

图6 动态任务对齐检测头

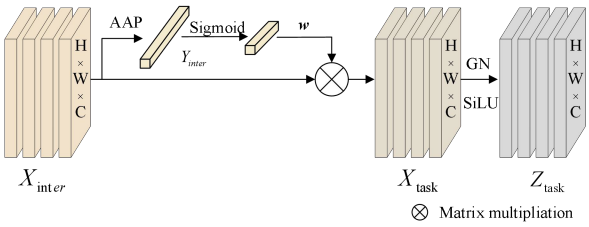

图7 层注意力机制

层注意力机制首先通过自适应平均池化 (Adapt average pooling, AAP) 和 Sigmoid 函数对特征图进行压缩, 生成权重向量, 与原始特征相乘, 经过组归一化和激活函数输出特征Z_{task}. 层注意力机制从交互特征中提取与分类和定位任务最相关的特征以用于它们各自的目的. 层注意力机制公式表示如下:

$$w = Sigmoid(C_2(SiLU(C_1(Y_{\text{inter}})))). \tag{13}$$

其中Y_{inter}是由X_{inter}经过平均池化得到, C_1, C_2表示两次 3×3 卷积,

$$X_{task} = w \cdot X_{inter}, \tag{14}$$

$$Z_{task} = SiLU(GN(X_{\text{task}})). \tag{15}$$

其中Z_{task}表示用于特定任务的特征.

随后, 将获取到的与回归任务相关的特征经过可变形卷积动态调整, 使模型可以更加准确的定位小目标边界框. 增强模型在检测小目标时的准确性. 最后通过 Conv_Reg 层对边界进行回归, 引入 Scale 层, 通过可学习的缩放因子调整特征图, 使不同检测头的输出一致.

3) 分类任务对齐模块. 通过卷积和激活函数生成交互特征的分类概率, 从层注意力机制处理后的交互特征中获取分类任务特征, 随后将分类概率与分类任务特征相乘, 使模型更精确感知无人机图像目标类别的置信度. 最后通过 Conv_Cls 层得到最终的分类输出.

2 实验结果与分析

本实验基于 Pytorch 框架实现, 设备为 Intel(R) Core(TM) i5-9400F 和 NVIDIA GeForce RTX 2080Ti. 学习率设置为 0.001. 共训练 200 个 epoch, Batchsize 设置为 2. 采用 SGD 随机梯度下降法优化模型训练, 优化器的权重衰减系数为 0.0005, 并使用余弦退火算法在训练过程中对学习率进行逐步调整. 采用的评价指标为精确率 (Precision,P), 召回率 (Rall,R), 平均精度均值 (mean Average Precision,mAP), 参数量 (Pramaeters), 计算复杂度 (Floating Point Operations, FLOPs) 和推理速度 (Frames Per Second, FPS.

本文使用 VisDrone2019[18] 数据集来对改进后

的模型进行训练评估. 该数据集是由天津大学AISKYEYE团队采集并制作, 数据集中包含了无人机在多个场景下不同角度拍摄捕获的真实图像, 涵盖行人、汽车、卡车、三轮车等10个类别. 划分为训练集包含6710张图像, 验证集包含548张图像, 测试集包含1610张图像. CODrone[19]是为无人机视觉研究设计的高分辨率多场景数据集, 采集自五个不同城市的真实无人机飞行数据, 涵盖港口、码头、工业区及城市密集区等多种复杂环境, 同时包括在正常光线、低光和夜间等不同光照条件下的图像. 包含5002张训练集图像, 2000张验证集图像. 经过对数据集进行分析, 这些数据集表现出类不平衡且小尺寸目标的比例较高的问题, 因此这些数据集非常适合无人机图像目标检测算法的研究.

2.1　消融实验

为了验证各改进模块对检测性能的影响, 本文在VisDrone2019数据集上以YOLO11s为基线模型对改进模块进行逐步消融, 进行一系列消融实验. 实验结果如表1所示. 用“√”表示该模块替换YOLO11s中的原始结构, FEMS-DFPN表示颈部网络改进, CSP-MSD对主干网络下采样进行优化, TDH代表改进后的动态任务检测头.

表1　消融实验结果

Baseline	P2	FEMS-DFPN	CSP-MS	DTH	P(%)	R(%)	mAP0.5(%)	mAP0.5:0.95(%)	Prama(M)	GFLOPs
√					49.8	37.8	38.7	23.0	9.4	21.7
√	√				51.7	42.2	43.1	26.1	2.8	24.6
√		√			55.6	43.2	45.6	27.6	4.8	31.4
√			√		51.6	38.2	39.6	24.0	9.5	25.3
√				√	51.7	40.0	41.2	25.1	8.7	25.1
√		√	√		56.5	46.0	48.2	29.4	4.9	33.8
√		√		√	57.9	45.4	48.0	29.1	3.4	36.6
√			√	√	54.0	42.3	44.0	26.3	8.2	27.7
√		√	√	√	58.6	46.3	48.9	29.7	4.4	38.6

实验结果表明, 加入FEMS-DFPN, 精确率与召回率等评价指标显著提升, 表明该模块通过对不同尺度特征的通道调整与融合, 使网络能够充分整合浅层细节特征与深层语义特征之间的互补信息. 提高了特征金字塔的表达丰富度, 还有效减少了特征冗余与信息损失, 从而显著提升了在复杂无人机场景下的检测精度与鲁棒性. 加入CSP-MS模块后, 参数量仅提升了0.1M, mAP50和mAP0.5:0.95分别提升0.9%和1%, 说明该模块在保持较低计算代价的同时, 有效弥补了固定卷积核结构下的感受野不足与语义表达局限, 各分支输出之间的逐层特征交互使得浅层特征与深层特征实现了更充分的协同表达. 现了对不同尺度目标特征的自适应表达. 单独引入DTH时, 模型参数量减少0.7M的同时mAP50提升2.5%, 实验结果中, 该结构能够在保持模型轻量化的同时, 显著提高小目标检测精度, 验证了DTH模块设计的有效性与通用性. 该模块通过参数共享和任务任务对齐实现了分类与回归任务的深度协同, 有效提升了模型对无人机航拍图像中小目标的表征能力与检测鲁棒性. 同时融入这三种改进策略时, 尽管计算量有所提升但是参数量降低了53%, 相对于基线模型mAP50与mAP0.5:0.95分别提升了10.2%和6.7%, 充分体现了该模型在无人机图像检测任务中的优势.

2.2　消融实验可视化对比

为了进一步验证各改进模块对模型性能的贡献, 以热力图的形式进行了消融实验可视化, 各改进模热力图可视化如图所示.

通过对比可发现, 不同模块对检测性能的提升具有针对性与互补性: CSP-MS模块的引入使模型在远处目标的检测效果显著提升, 热力区域在图像深处目标位置更加集中, 说明该模块增强了网络的多尺度特征提取能力. FEMS-DFPN模块在遮挡目标区域表现突出, 能够有效识别被部分遮挡的行人和物体, 说明其对特征融合与上下文信息提取的改进提升了模型复杂场景下的鲁棒性. DTH模块的加入提升了模型对小目标的检测能力, 从热力图可见其在近景小目标区域响应增强, 证明该模块有助于细

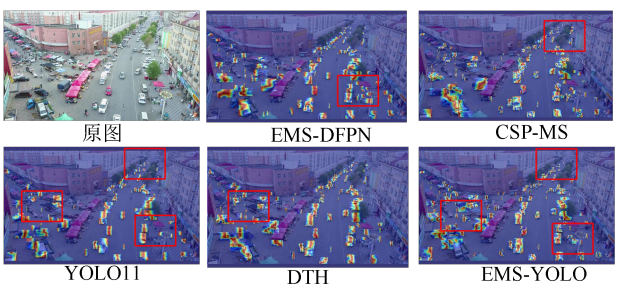

图8　消融实验热力图

粒度特征的保持与增强. FEMS-YOLO 模型结合了上述三个模块的优点, 在远距离目标、遮挡场景及小目标区域均表现出更为均衡且突出的检测效果, 热力分布更加全面, 体现了整体检测性能的协同提升.

2.3 FEMS-DFPN 模块在不同尺度目标实验

为进一步验证删除大目标检测层对不同尺度目标检测性能的影响, 本文在 VisDrone 数据集上统计了大、中、小三个不同尺度目标的 AP_{large}、AP_{medium}、AP_{small}, 结果如表 2 所示.

表2 FEMS-DFPN 模块在不同尺度目标实验结果

Method	AP_{large}	AP_{medium}	AP_{small}	F1
YOLO11	40.6	33.7	12.6	42.8
P2	42.1	34.8	16.0	46.1
FEMS-DFPN	44.2	36.5	17.4	48.1

实验结果如表 2 所示, YOLOv11 在大目标检测上 AP_{large} 达到 40.6% 具有较高精度, 但对小目标的检测性能 AP_{small} 为 12.6%. 相对较弱在引入 P2 小目标检测层模型中, AP_{large} 提升至 42.1%, AP_{small} 提升至 16.0%, 说明该结构改进并未因删除大目标检测层导致大目标效果变弱, 反而在一定程度上增强了多尺度特征表达能力. 在进一步加入 FEMS-DFPN 模型使小目标检测精度达到 17.4, 较基线模型提升了 4.8%; 同时 AP_{large} 提升至 44.2%. 该结果表明, 表明 FEMS-DFPN 模块能够充分地融合跨层级特征信息, 增强了跨层级尺度特征的复用性, 提升特征表达能力. 显著提升中小目标的检测性能, 从而在检测任务中表现出更强的鲁棒性.

2.4 不同网络模型对比实验

为验证本文所提出的 FEMS-YOLO 在无人机图像场景下的检测性能, 本研究对比主流目标检测算法在 VisDrone 无人机数据集上进行了系统评估. 对比实验结果如表 3 所示.

表3 不同网络模型对比实验结果

Methods	P(%)	R(%)	mAP0.5(%)	mAP0.5:0.95(%)	Prama(M)	GFLOPs	FPS
RetinaNet[20]	37.6	28.5	31.4	—	9.1	82.3	68.6
YOLOv5s	47.4	34.8	35.5	21.0	9.1	24.1	178.2
YOLOv8s	51.0	37.2	39.0	23.2	11.1	28.6	161.4
YOLOv11s	49.8	37.8	38.7	23.0	9.4	21.7	190.8
YOLOv12s	49.5	36.8	37.9	22.6	9.0	19.3	147.3
TPH-YOLO[7]	58.0	42.7	45.5	27.0	60.43	145.7	33.4
Drone-YOLO[12]	—	—	44.3	27.0	10.9	—	—
UAV-YOLOv8s[21]	54.4	45.6	47.0	29.2	10.3	—	—
文献[21]	56.3	44.6	46.8	28.8	12.73	42.8	101.5
文献[23]	55.4	43.2	44.8	27.0	9.34	33.6	155.7
FEMS-YOLO	58.6	46.3	48.9	29.7	4.4	38.6	120.2

由表 3 可以看出, 本方法在无人机图像检测任务中取得了显著性能提升. 与其他对比模型相比, 所提出的改进模型在精确率、召回率、mAP0.5 和 mAP0.5:0.95 指标上分别达到 58.6%、46.3%、48.9% 和 29.7%, 优于现有方法. 相较于 YOLOv11s 模型, mAP0.5 和 mAP0.5:0.95 分别提升了 10.2% 和 6.7%. 但由于 FEMS-DFPN 模块中多次执行跨层级特征融合, 模型的计算复杂度有所增加, 达到 38.6G, 略高于文献 [23] 中的 33.6G. 尽管计算复杂度并非最优, 本文模型在精确率等主要指标上均取得领先, FEMS-YOLO 模型在该对比中同样展现出优越的实时性能, 其推理速度达到 120.2FP, 明显优于大多数高精度模型 TPH-YOLO 与文献 [21], 在保证较高精度的前提下实现了较快的推理速度, 能够满足无人机实时处理. 明在 FEMS-YOLO 保持检测精度同时具有更强的实时处理能力. 本文模型在性能与效率之间取得了更好的平衡.

2.5 VisDrone 数据集上不同类别对比实验

为验证所提方法在复杂无人机视角图像中的目标检测性能, 在 VisDrone 数据集与多个主流检测模型进行横向对比并进行系统分析, 对比实验结果如表 4 所示.

从表 4 可以看出各方法在 VisDrone 数据集上 10 个目标类别上的 AP 结果及其整体 mAP. 表格中的数据展示了本文模型与多个模型在不同目标类别上的对比实验, 本文方法在 VisDrone 无人机图像目标检测任务中表现出优越的检测精度与多个场景下目标检测适应能力, 本文提出的 FEMS-YOLO 模型在自行车、三轮车等多个小目标类别中达到最优效果的同时在大尺寸目标 Bus、Car 两个类的精度保持以 64.6%, 85.3% 的高精度, 超越所有对比模型, 验证了其对多尺度的广泛适应性, 能兼顾大目标和小目

表4　Visdrone 数据集上不同类别对比实验结果

Method	TARGET CLASS (AP/%)										mAP 50(%)
	Pedestrian	Person	Bicycle	Car	Van	truck	Tricycle	Awning tricycle	Bus	Motor	
RetinaNet	26.0	16.7	7.9	67.3	33.4	29.8	14.0	15.5	47.4	27.0	31.4
YOLOv5s	41.2	30.5	11.5	77.4	35.9	27.6	25.5	15.3	54.1	42.8	35.5
YOLOv8s	42.2	32.0	12.9	79.7	44.3	36.4	26.7	16.0	56.6	44.2	39.0
YOLO11s	42.1	32.9	11.8	79.4	45.7	36.2	24.9	14.7	56.1	42.7	38.6
YOLO12s	39.4	31.2	11.8	78.5	43.9	35.5	27.4	14.6	54.9	41.6	22.6
文献[7]	29.0	16.7	15.6	68.9	49.7	45.1	27.0	**24.7**	61.8	30.9	36.9
文献[23]	53.3	42.0	17.7	84.3	49.0	36.7	33.1	17.7	61.2	53.2	44.8
文献[24]	**55.7**	45.3	21.4	84.8	49.4	42.3	32.0	19.1	63.5	53.1	46.7
文献[25]	52.3	42.7	18.5	84.7	51.1	41.8	33.0	18.6	59.6	54.1	45.6
FEMS-YOLO	54.8	**45.4**	**23.3**	**85.3**	**53.3**	**45.4**	**39.2**	23.1	**64.6**	**56.7**	**48.9**

标的定位准确性. 表明本文模型多尺度特征提取和跨层特征融合机制有效捕获了不同尺度目标特征, 同时融合了目标的全局上下文信息以及局部细节信息, 提升了模型对不同尺度目标定位的准确性以及复杂场景下无人机图像检测的鲁棒性.

2.6　检测结果可视化对比

本文模型与 YOLOv8s、YOLO11s 和 YOLO12s 在 VisDrone 数据集上, 分别针对小目标场景、多尺度场景、复杂场景及昏暗场景进行了检测效果对比, 可视化对比如图 9 所示. 小目标场景下展示了对无人机在高空视角下拍摄图像目标检测, 该图中存在大量的小目标, 对检测结果对比可知, 其他模型存在目标漏检, 而本文模型精准检测出了图片中的多数车辆目标. 在多尺度场景下其他模型对远处小尺寸目标漏检严重, 在高密度区域下的复杂场景表现出更强的目标分离能力, 能够较为准确地区分重叠行人、车辆和三轮车. 检测结果边界清晰、置信度高, 且误检与漏检率明显较低. 弱光区域的目标检测中表现出较高的准确性, 能够有效抑制背景干扰与误检, 充分体现了其对复杂光照条件的适应能力.

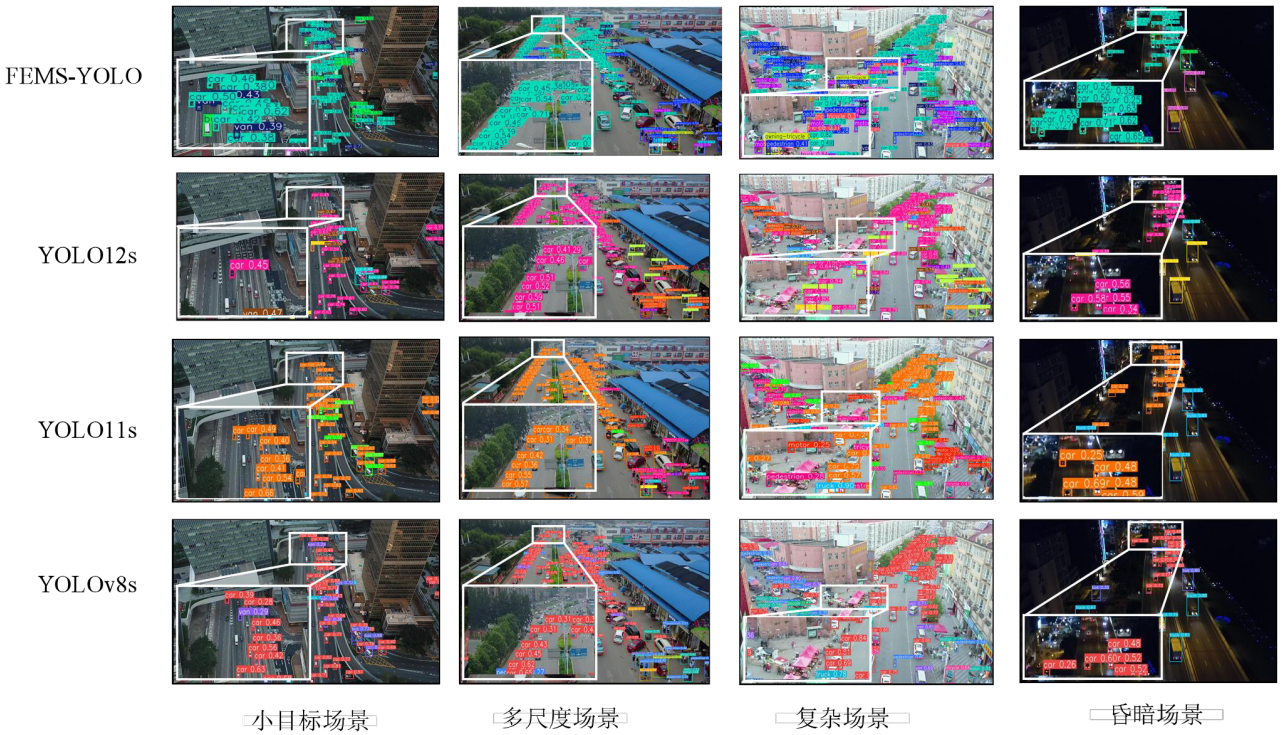

图9　检测结果可视化对比

2.7　CODrone 数据集上对比实验

为进一步评估所提模型的泛化性能, 本文选取 CODrone 数据集, 并与 YOLO 系列模型进行对比实验, 实验结果如表 5 所示. 其中, 本文模型在 mAP0.5 上达到 33.1%, 相比 YOLO11s 提高了 6.6%; 在 mAP0.5:0.95 上达到了 17.2%. 实验结果表明, 本文方法在各项指标上均优于对比模型综上, 本文方法在 CODrone 数据集上展现出优越的检测精度, 展

示出较强的泛化能力. 进一步验证了所提改进策略在真实复杂场景中的有效性与实用价值.

表5 CODrone 数据集上对比实验结果

Methods	P(%)	R(%)	mAP0.5(%)	mAP0.5:0.95(%)	Prama(M)	GFLOPs
YOLOv8s	42.2	29.4	27.6	14.4	11.1	28.6
YOLOv10s	41.0	27.7	27.2	13.7	8.04	24.4
YOLO11s	41.5	30.8	27.7	14.5	9.4	21.7
YOLO12s	43.9	29.2	27.3	14.0	9.1	19.6
FEMS-YOLO	45.7	34.9	33.1	17.2	4.4	38.6

3 结 论

本文针对无人机图像中小目标数量多、尺度变化大及背景复杂等问题, 提出了一种基于 YOLO11 的高效多特征融合目标检测算法. 首先, 设计 CSP-MS 模块, 采用分层融合机制与异构卷积核设计, 强化了主干网络的特征提取能力, 提升了模型对多尺度目标的检测效果. 随后, 设计特征增强多尺度特征聚合金字塔模块, 将不同尺度特征进行融合, 并通过并行空洞卷积增强融合特征的表征能力, 增强模型在复杂背景下的鲁棒性, 最后, 引入动态任务对齐检测头, 在有效降低参数量的同时提升对小目标检测性能.

通过实验证明, 相比于 YOLO11s 本文模型在 VisDrone 数据集上 mAP0.5 和 mAP0.5:0.95 指标上分别提升 10.2% 和 6.7%, 在 CODrone 数据集在 mAP0.5 和 mAP0.5:0.95 指标分别提升 5.4% 和 3.7%, 取得良好性能并验证了其良好的泛化能力. 特别是在小目标密集、背景复杂与尺度变化大的无人机拍摄场景中, 本文方法展现出良好的目标识别能力和应用潜力. 未来的研究会继续探索更加轻量化研究, 以在保持高精度的前提下降低模型计算开销, 提升模型在嵌入式设备和无人机平台上的部署效率, 为无人机图像目标检测算法在实际场景中的应用提供更强的支撑.

参考文献 (References)

[1] Jasim A N, Fourati L C. Unmanned aerial vehicles (UAV) and its evolution in smart agriculture: A review[C]. 2024 IEEE International Conference on Agrosystem Engineering, Technology & Applications. Kuala Lumpur, 2025: 42-47.

[2] Pamucar D, Gokasar I, Ebadi Torkayesh A, et al. Prioritization of unmanned aerial vehicles in transportation systems using the integrated stratified fuzzy rough decision-making approach with the hamacher operator[J]. Information Sciences: An International Journal, 2023, 622(C): 374-404.

[3] Wlodarczyk D, Saber T. Assessing potential of UAV-based increase in urban disasters communication on traffic congestion mitigation[C]. 2024 International Conference on Information and Communication Technologies for Disaster Management. Setif, 2024: 1-7.

[4] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, 2014: 580-587.

[5] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[6] 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. Las Vegas, 2016: 779-788.

[7] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot MultiBox detector[C]. Computer Vision – ECCV 2016. Cham: Springer, 2016: 21-37.

[8] Zhu X K, Lyu S C, Wang X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]. 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal, 2021: 2778-2788.

[9] 侯颖, 吴琰, 寇旭瑞, 等. 改进 YOLOv8 的无人机航拍图像小目标检测算法[J]. 计算机工程与应用, 2025, 61(11): 83-92.
(Hou Y, Wu Y, Kou X R, et al. Small object detection algorithm for UAV images based on improved YOLOv8[J]. Computer Engineering and Applications, 2025, 61(11): 83-92.)

[10] Zhou S L, Zhou H J, Qian L. A multi-scale small object detection algorithm SMA-YOLO for UAV remote sensing images[J]. Scientific Reports, 2025, 15: 9255.

[11] 蒋伟, 王万虎, 杨俊杰. AEM-YOLOv8s: 无人机航拍图像的小目标检测[J]. 计算机工程与应用, 2024, 60(17): 191-202.
(Jiang W, Wang W H, Yang J J. AEM-YOLOv8s: Small target detection algorithm for UAV aerial images[J]. Computer Engineering and Applications, 2024, 60(17): 191-202.)

[12] Zhang Z X, Zhang Z X. Drone-YOLO: An efficient neural network method for target detection in drone images[J]. Drones, 2023, 7(8): 526.

[13] 张信佳, 王芳. 基于多层次特征融合和注意力机制的无人机图像小目标检测算法[J]. 计算机工程, https://doi.org/10.19678/j.issn.1000-3428.0069729.

[14] 谌海云, 肖章勇, 郭勇, 等. 基于改进 YOLOv8s 的无人机航拍目标检测算法[J]. 电光与控制, 2024, 31(12): 55-63.
(Chen H Y, Xiao Z Y, Guo Y, et al. A UAV aerial target detection algorithm based on improved YOLOv8s[J]. Electronics Optics & Control, 2024, 31(12): 55-63.)

[15] Chen Y M, Yuan X B, Wang J B, et al. YOLO-MS:

Rethinking multi-scale representation learning for real-time object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(6): 4240-4252.

[16] Liu W Z, Lu H, Fu H T, et al. Learning to upsample by learning to sample[C]. 2023 IEEE/CVF International Conference on Computer Vision. Paris, Piscataway: IEEE, 2024: 6004-6014.

[17] Xu Z, Luo X L, Gao X J, et al. DTH-YOLO: Enhanced YOLOv8n with dynamic task-aligned head for mousehole detection[J]. IEEE Access, 2025, 13: 44912-44927.

[18] Du DW, Zhu P F, Wen L Y, et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Seoul, 2019: 213-226.

[19] Ye K, Tang H D, Liu B W, et al. More clear, more flexible, more precise: A comprehensive oriented object detection benchmark for UAV[J/OL]. 2025, arXiv: 2504.20032.

[20] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]. 2017 IEEE International Conference on Computer Vision. Venice, 2017: 2999-3007.

[21] Wang G, Chen Y F, An P, et al. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios[J]. Sensors, 2023, 23(16): 7190.

[22] 陈志旺, 肖迪创, 吕昌昊, 等. 基于多尺度融合和高分辨特征增强的无人机航拍目标检测[J]. 控制与决策, 2025, 40(7): 2290-2299.
(Chen Z W, Xiao D C, Lv C H, et al. UAV aerial target detection based on multi-scale fusion and high-resolution feature enhancement[J]. Control and Decision, 2025, 40(7): 2290-2299.)

[23] 张轩宇, 周思航, 黄健, 等. 基于高阶空间特征提取的无人机航拍小目标检测算法[J]. 计算机工程与应用, 2025, 61(12): 210-221.
(Zhang X Y, Zhou S H, Huang J, et al. High-order spatial feature extraction based small target detection for UAV aerial photographs[J]. Computer Engineering and Applications, 2025, 61(12): 210-221.)

[24] Ma C J, Fu Y Y, Wang D Y, et al. YOLO-UAV: Object detection method of unmanned aerial vehicle imagery based on efficient multi-scale feature fusion[J]. IEEE Access, 2023, 11: 126857-126878.

[25] Li Y H, Zhang X Y, Zhou Z G. DBS-YOLO: A vehicle detection model based on improved YOLOv8 for UAV aerial scenes[C]. 2024 5th International Conference on Computer Vision, Image and Deep Learning. Zhuhai, 2024: 1432-1438.

作者简介

张开玉 (1978−), 男, 副教授, 博士, 主要研究方向为图像处理, E-mail: zhangkaiyu@hrbust.edu.cn;

王浩杰 (2001−), 男, 硕士生, 主要研究方向为无人机目标检测, E-mail: 13939288498@163.com;

卢迪 (1971−), 女, 教授, 博士 主要研究方向为数据融合与图像处理等, E-mail: ludizeng@hrbust.edu.cn.