

# YOLO-MAT: 融合旋转感知注意力与自适应特征过滤的无人机目标检测

邓承志<sup>1,2†</sup>, 武瑛博<sup>1,2</sup>, 吴朝明<sup>1,2</sup>, 孙小惟<sup>1,2</sup>, 汪胜前<sup>1,2</sup>

(1. 江西水利电力大学 信息工程学院, 南昌 330099;

2. 江西水利电力大学 智慧水利江西省重点实验室, 南昌 330099)

**摘要:** 针对无人机航拍图像中因目标方向任意旋转、背景环境复杂以及目标尺寸微小等因素导致的检测精度下降问题, 提出一种基于 YOLOv12 架构的轻量化无人机目标检测网络 YOLO-MAT, 该网络融合旋转感知注意力与自适应特征过滤机制。首先, 提出一种多路径旋转感知注意力模块 (MRAC2f), 通过引入旋转不变注意力机制 (RAM), 增强模型对旋转目标的鲁棒特征表征能力; 其次, 设计一种自适应加权多尺度特征过滤融合模块 (AMFF), 集成双域协同注意力 (DDCA) 与拉普拉斯边缘增强器 (LEE), 在抑制浅层背景噪声的同时增强高频细节特征, 并利用可学习权重实现多尺度特征的自适应融合; 最后, 构建一种高分辨率小目标检测头, 进一步提升模型对微小目标的检测性能。在 VisDrone2019 和 NWPU VHR-10 数据集上的实验结果表明, 相较于基准模型 YOLOv12, YOLO-MAT 在模型参数量减少 3.2% 的同时, 平均精度均值 (mAP) 分别提升 6.7% 和 9.3%, 可实现轻量化设计与检测精度的有效平衡。与其他主流检测算法相比, YOLO-MAT 在检测精度方面具有明显优势, 可为无人机实时目标检测提供一种高效的解决方案。

**关键词:** YOLOv12; 自适应特征过滤; 小目标检测; 无人机; 旋转不变注意力; 特征融合

中图分类号: TP391.41

文献标志码: A

DOI: 10.13195/j.kzyjc.2025.1017

引用格式: 邓承志, 武瑛博, 吴朝明, 等. YOLO-MAT: 融合旋转感知注意力与自适应特征过滤的无人机目标检测 [J]. 控制与决策.

## YOLO-MAT: integrating rotation-aware attention and adaptive feature filtering for UAV object detection

DENG Cheng-zhi<sup>1,2†</sup>, WU Ying-bo<sup>1,2</sup>, WU Zhao-ming<sup>1,2</sup>, SUN Xiao-wei<sup>1,2</sup>, WANG Sheng-qian<sup>1,2</sup>

(1. Jiangxi University of Water Resources and Electric Power, School of Information Engineering, Nanchang 330099, China; 2. Jiangxi University of Water Resources and Electric Power, Key Laboratory of Smart Water Conservancy in Jiangxi Province, Nanchang 330099, China)

**Abstract:** To address the issue of declining detection accuracy caused by arbitrary object rotation, complex backgrounds, and small target sizes in UAV aerial images, this paper proposes YOLO-MAT, a lightweight object detection network based on the YOLOv12 architecture that integrates rotation-aware attention and adaptive feature filtering. The main contributions are as follows: First, a Multi-path Rotation-aware Attention C2f module (MRAC2f) is proposed, which enhances the model's robustness in representing rotated targets through a Rotation-Invariant Attention Mechanism (RAM). Second, an Adaptive Multi-scale-feature Filtering and Fusion module (AMFF) is designed, which integrates a Dual Domain collaborative attention (DDCA) mechanism and a Laplace Edge Enhancer (LEE) to suppress background noise in shallow features while enhancing high-frequency details, and employs learnable weights to achieve adaptive fusion of multi-scale features. Third, a high-resolution Tiny Head is constructed to further improve the detection performance for small targets. Experimental results on the VisDrone2019 and NWPU VHR-10 datasets demonstrate that compared to the baseline model YOLOv12, YOLO-MAT reduces the number of parameters by 3.2% while increasing the mean Average Precision (mAP) by 6.7% and 9.3%, respectively, achieving an effective balance between lightweight design and detection accuracy. Compared with other mainstream detection algorithms, YOLO-

收稿日期: 2025-09-27; 录用日期: 2026-01-09.

基金项目: 国家自然科学基金项目 (61865012); 江西省自然科学基金项目 (20252BAC250011).

责任编委: 程龙.

†通信作者. E-mail: dengcz@nit.edu.cn.

MAT exhibits clear advantages in detection accuracy, offering an efficient solution for real-time UAV object detection tasks.

**Keywords:** YOLOv12; adaptive feature filtering; small target detection; UAV; rotation-invariant attention; feature fusion

## 0 引言

随着无人机 (UAV) 平台与深度学习目标检测技术的快速发展和深度融合, 面向无人机航拍图像的目标检测受到学术界和工业界的高度关注, 并已成功应用于精准农业、动物监测、城市管理、应急救援等领域<sup>[1]</sup>. 无人机拍摄角度多变、复杂背景干扰、光照和天气变化等给无人机小目标检测带来许多困难, 极易造成错检和漏检. 如何快速、准确地检测出无人机航拍图像中的小目标, 是当前无人机目标检测领域的热点和难点.

基于深度学习的无人机目标检测算法可分为两类<sup>[2]</sup>: 双阶段检测器 (如 Fast R-CNN 系列<sup>[3]</sup>) 和单阶段检测器 (如 YOLO 系列<sup>[4]</sup>、SSD 系列<sup>[5]</sup>). 双阶段检测器虽检测精度较高, 但其计算复杂度和推理延迟较大, 难以满足无人机平台的实时处理要求<sup>[6]</sup>. 相比之下, 单阶段检测器通过预定义锚框<sup>[7]</sup>或锚框无关机制<sup>[8]</sup>直接预测目标, 具备更快的推理速度, 因而更适用于无人机目标检测任务. 直接将通用单阶段检测器应用于无人机目标检测效果非常有限, 研究者提出了诸多针对性改进方案. Wei 等<sup>[9]</sup>通过在 RT-DETR 网络中使用可变形注意力机制增强对高密度场景中的小目标检测能力. Gao 等<sup>[10]</sup>设计并行空洞卷积模块与注意力上下采样分支模块, 提出一种高效的无人机航拍小目标检测算法, 增强无人机小目标检测能力. Chen 等<sup>[11]</sup>设计了一种适配无人机的目标检测算法, 通过基于特征金字塔网络的多维特征自适应融合模块增强浅层特征利用. Xie 等<sup>[12]</sup>提出了一种基于密度引导的两阶段目标检测框架 (DG-TSOD), 用于识别小物体. Zhang 等<sup>[13]</sup>提出了 MISATrack 模型, 结合了频域编码器和尺度频率线性注意力机制, 极大地简化了表示学习过程, 同时又大幅降低了计算成本, 从而实现了更稳健的结果. Zhu 等<sup>[14]</sup>通过测量连接旋转边界框的中心点与特定边界中心点的线与水平线之间的角度来实现对旋转物体的精确定位. Du 等<sup>[15]</sup>对 YOLO11 的骨干网络结构中的基本构建单元进行优化, 并引入 SimAM 全局注意力机制, 实现对旋转目标的检测.

上述方法在一定程度上提升了无人机目标检测性能, 在实际部署中仍存在诸多挑战. 一方面, 无人机飞行中姿态变化导致成像视角持续变化, 目标的方位也随之变化, 现有模型多针对水平目标优化, 对

旋转目标适用性不好. 另一方面, 高空广视角拍摄使得检测目标占比极小, 有效特征稀少, 导致检测精度显著下降. 另外, 无人机航拍图像背景广阔复杂, 背景干扰易降低模型对真实目标的辨识置信度. 因此, 设计一种对旋转目标敏感、小目标检测能力强, 且能抵抗复杂背景干扰的无人机航拍图像目标检测模型, 具有重要理论价值与现实意义.

为更好地解决上述问题, 提出一种融合旋转感知注意力与自适应特征过滤的无人机目标检测网络 YOLO-MAT, 主要创新如下:

1. 设计一种多路径旋转感知注意力模块 (Multi-Path Rotation-Invariant Attention Module C2f, MRAC2f), 增强模型对旋转目标的表征能力.
2. 设计自适应加权多尺度特征过滤融合模块 (Adaptive Multi-scale-feature Filtering and Fusion, AMFF), 提升多尺度特征融合的质量, 减少复杂背景对检测目标的影响.
3. 设计高分辨率小目标检测头 (Tiny Head), 提升对小目标的检测能力.

## 1 YOLO-MAT 模型

YOLOv12<sup>[16]</sup>是当前先进的单阶段目标检测网络, 尤其在小目标检测任务中表现卓越. 其主干网络由 Conv 模块, C3K2 模块<sup>[17]</sup>以及 A2C2f 模块<sup>[18]</sup>组成. 其中, C3K2 模块由 YOLOv11<sup>[19]</sup>提出, 通过引入多核卷积结构以提取多尺度特征信息, 增强了模型对不同尺寸目标的感知能力. YOLOv12 的核心创新之一是引入了区域注意力机制<sup>[20]</sup>, 该机制采用十字形窗口自注意力结构, 沿水平和垂直方向分别计算注意力权重, 能够在避免复杂计算操作的同时获得更大的感受野. 另一重要创新是设计了残差高效层聚合网络, 该结构通过引入过渡层调整通道维度, 生成统一的中间特征图, 再经后续模块处理并连接, 构建具有瓶颈结构的高效特征融合路径, 从而显著提升了原始特征的集成与保留能力. 颈部网络延续了先前版本的设计理念, 采用特征金字塔网络<sup>[21]</sup>与路径聚合网络<sup>[22]</sup>相结合的方式, 实现对主干网络所提取特征的多尺度融合与增强. 然而, 在处理具有方向旋转多样、背景干扰强烈的无人机航拍小目标检测任务时, YOLOv12 仍存在一定局限性.

本文提出的 YOLO-MAT 网络整体架构如图 1

所示. 在主干网络设计多路径旋转感知注意力模块 (MRAC2f), 结合多级并行处理架构, 以提升对目标方向变化与复杂背景的适用性; 在颈部网络设计自适应加权多尺度特征过滤融合模块 (AMFF), 通过选

择性增强不同层级与支路的特征, 以提升模型复杂背景下小目标的特征表征能力; 在检测头设计高分辨率小目标检测头 (Tiny Head), 强化浅层特征利用, 以提升小目标的检测精度.

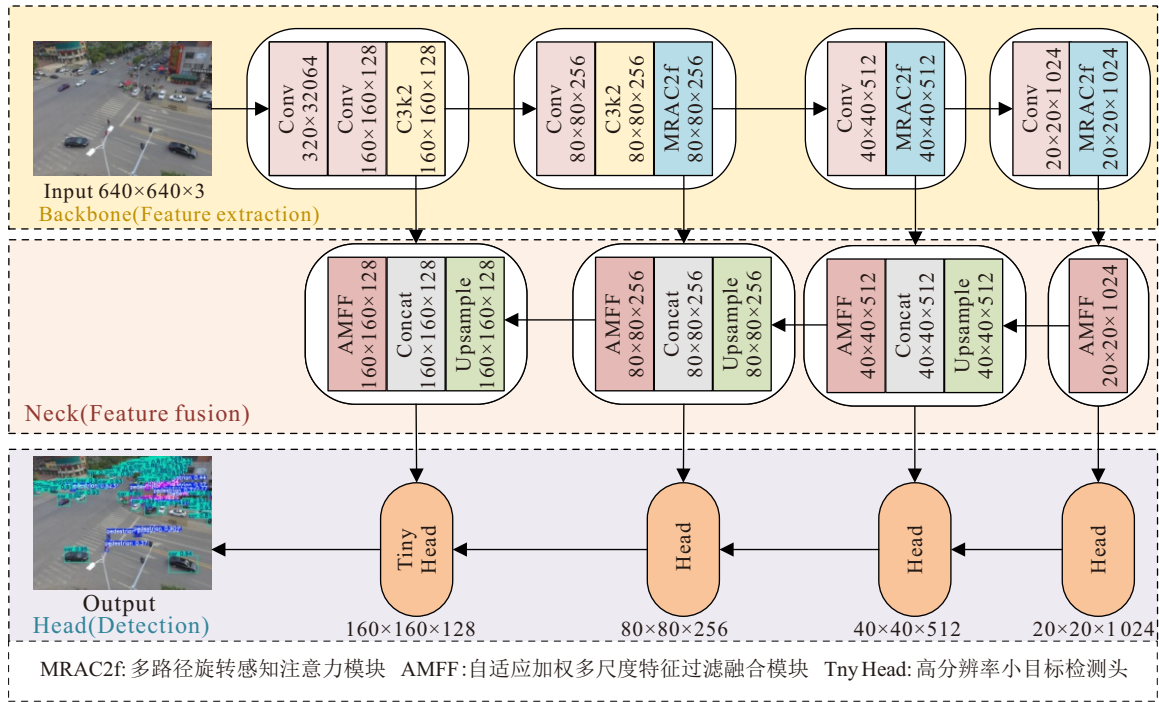


图1 YOLO-MAT 网络结构模型图

### 1.1 多路径旋转感知注意力模块 (MRAC2f)

为了更精准地捕获无人机图像中目标方向旋转的特征, 本文设计一种全新的多路径旋转感知注意力模块 (MRAC2f), 其结构如图 2 所示. 首先, 设计了旋转不变注意力机制 (Rotational-Invariant Attention Mechanism, RAM), 增强模型对旋转目标的特征鲁棒性. 其次, 设计了深度旋转不变特征提取模块 (Deep Rotation-Invariant Feature Extraction, DRFE), 融入 RAM 并通过多路径并行处理方式, 保留丰富的空间细节与高级语义信息, 增强模型的抗干扰能力.

设输入特征  $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$ , 其中  $C$  为输入通道

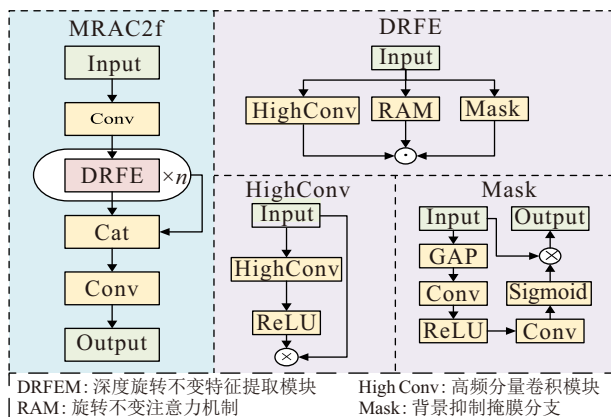


图2 MRAC2f 模块结构图

数,  $H$  和  $W$  为特征图的高度和宽度. MRAC2f 具体操作如下:

首先, 使用一个  $1 \times 1$  的卷积层对输入特征图进行变换, 压缩通道数以降低参数量:

$$\mathbf{Y}_0 = \text{Conv}(\mathbf{X}) \in \mathbb{R}^{C_* \times H \times W}. \quad (1)$$

其中,  $C_* = \max(8, C/2)$  输出通道数, 最小值约束设置 8, 旨在确保网络即使输入通道数较小时仍具备足够的特征表征能力.

随后, 通过  $n$  个级联的 DRFE 模块, 每级 DRFE 模块提取不同抽象层级的特征, 从而构建从细节到语义的连续表征空间:

$$\mathbf{Y}_k = \mathbf{G}_{\text{DRFE}}^{(k)}(\mathbf{Y}_{k-1}), \quad k = 1, 2, \dots, n. \quad (2)$$

其中,  $\mathbf{G}_{\text{DRFE}}(\cdot)$  表示 DRFE 的映射函数;  $\mathbf{Y}_k$  表示融合  $k$  次非线性变换后的高级语义信息.

最后, 将各级输出沿通道维度进行拼接, 并通过一个线性投影层将通道数恢复至  $C$ .

$$\mathbf{Y}' = \bigoplus_{k=0}^n \mathbf{Y}_k, \quad \mathbf{Y} = \text{Conv}(\mathbf{Y}'). \quad (3)$$

其中,  $\bigoplus$  表示按通道维度拼接.

#### 1.1.1 深度旋转不变特征提取模块 (DRFE)

DRFE 模块借鉴特征金字塔思想, 通过构建并行的多级特征提取路径, 显式地保留并融合从浅层

细节到深层语义的跨层级特征,为旋转感知计算提供信息更丰富、细节更完备的特征表征,克服原单一支路结构在复杂检测场景中的表征局限性. DRFE模块以上层卷积输出特征  $\mathbf{X} \in \mathbb{R}^{C^* \times H \times W}$  为输入,通过三条并行分支提取高频细节、旋转不变注意力和背景抑制掩码.

高频细节提取分支用于提取特征中的高频分量,以强化边缘和角点等细节特征,提升模型对小尺寸目标的特征提取能力.

$$\mathbf{X}_{hf} = \text{ReLU}(\text{Conv}_f(\mathbf{X})). \quad (4)$$

其中, ReLU为激活函数;  $\text{Conv}_f$ 为高频卷积.

旋转不变注意力分支主要是增强模型对旋转目标特征的提取能力,具体操作流程在下节介绍.

$$\mathbf{X}_{ram} = \text{F}_{\text{RAM}}(\mathbf{X}). \quad (5)$$

其中,  $\text{F}_{\text{RAM}}$ 表示旋转不变注意力机制.

背景抑制掩码分支利用全局上下文信息生成空间-通道联合抑制系数,增强模型的抗背景干扰能力.

$$\mathbf{X}_{mask} = \sigma(\text{Conv}(\text{ReLU}(\text{Conv}(\text{GAP}(\mathbf{X}))))). \quad (6)$$

其中, GAP为全局平均池化,  $\sigma$ 为 Sigmoid 激活函数.

最后,将上述三条分支的输出进行融合,实现深度旋转不变特征的提取.

$$\mathbf{Y} = \text{Conv}(\text{ReLU}(\text{Conv}((\mathbf{X} \odot \mathbf{X}_{mask}) \oplus \mathbf{X}_{hf}))) \odot \mathbf{X}_{ram}. \quad (7)$$

其中,  $\odot$ 表示点乘.

### 1.1.2 旋转不变注意力机制 (RAM)

为解决旋转特征表征问题,本文提出一种旋转不变注意力机制 (RAM). RAM通过频域-空间域联合分析,将目标的旋转变化解耦为方向可分性的频域分量,并借助可学习的方向滤波器组与逆变换对齐,实现了对旋转目标的内在固有表征增强. RAM结构如图3所示,包含多尺度方向特征提取、频域方向分析、小波特征增强、双重注意力机制、逆旋转与特征融合等五个环节.

在多尺度方向特征提取阶段,首先采用分组卷积与空洞卷积构建多尺度方向特征提取器.具体操作如下:设输入特征  $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$ ,分组卷积操作将输入通道平均分成8组,每组独立进行  $3 \times 3$ 卷积运算,即

$$\text{GroupConv}(\mathbf{X}) = \text{Concat}[\text{Conv}_{3 \times 3}(\mathbf{X}_{gi})]_{i=1}^8. \quad (8)$$

其中,  $\mathbf{X}_{gi}$ 表示第*i*组输入特征,每组包含  $C/8$ 个通道.

分组卷积后接入 ReLU 激活函数与分组归一化 (GroupNorm),确保不同方向的特征分布一致性.同时为了扩大感受野并捕获到多尺度上下文信息,采用空洞率为2的  $3 \times 3$ 空洞卷积 (DilatedConv):

$$\text{F}_{\text{dir}} = \text{DilatedConv}_{\text{dil}=2}(\text{GroupNorm}(\text{ReLU}(\text{GroupConv}(\mathbf{X}))). \quad (9)$$

在频域方向分析阶段,对目标的旋转特性进行频域解耦.首先,将提取到的八组特征通过 GAP 压缩空间维度,保留目标的整体结构信息并消除位置

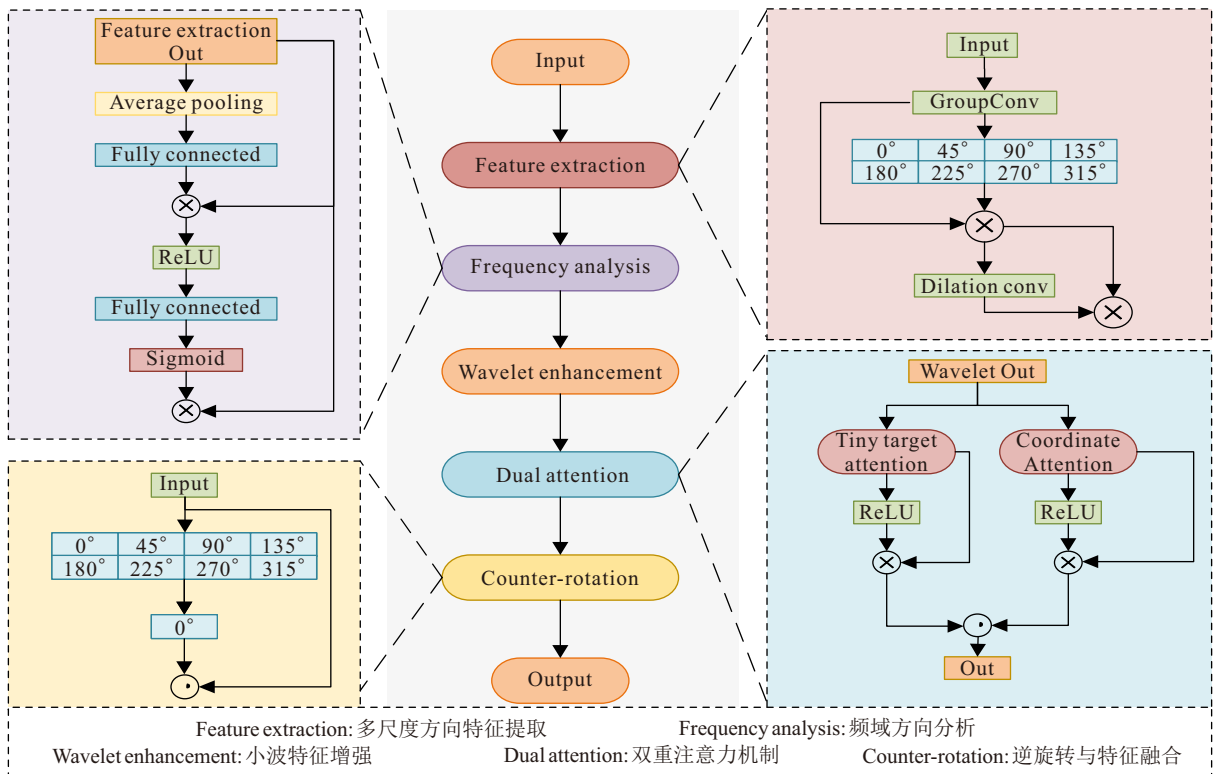


图3 RAM 模块结构图

干扰. 随后, 通过两层多层感知机 (MLP) 将特征投影至 8 组方向 ( $0^\circ, 45^\circ, 90^\circ, \dots, 315^\circ$ ) 并生成权重空间. 最后, 将八组方向的特征图进行加权融合.

$$W = \sigma(\text{MLP} \otimes \text{ReLU}(\text{MLP} \otimes \text{GAP}(F_{dir}))), \quad (10)$$

$$F_{align} = \sum_{k=1}^8 W_k \cdot R_{\theta_k}(F_{dir}). \quad (11)$$

其中,  $W_k$  为第  $k$  个方向上的权重,  $R_{\theta_k}(F_{dir})$  表示特征图旋转  $\theta_k$  角度后的结果.

在小波特征增强阶段, 由于 Haar 小波具有计算效率高、适于捕捉边缘特征, 且易于集成于深度学习框架选择, 因此本文选择使用 Haar 小波来处理特征. 通过 Haar 小波分解将特征分解为四个子带, LL、LH、HL、HH. 对 LH 和 HL 进行锐化以增强边缘, 抑制 HH 中的噪声, 对 LL 子带应用自适应门控以保留结构信息:

$$F_{wave} = F_{WEC}(F_{align}). \quad (12)$$

其中,  $F_{WEC}$  表示小波变换特征增强器.

对 LL 子带应用可学习的自适应门控, 实现结构信息的智能保留:

$$LL_{enhanced} = \sigma(\text{Conv}_{1 \times 1}(\text{ReLU}(\text{Conv}_{1 \times 1}(LL)))). \quad (13)$$

在双重注意力机制阶段分别聚焦小目标与空间上下文信息, 其中小目标注意力通过通道压缩-扩展策略增强微小目标响应:

$$A_{small} = \sigma(W_{s2} \otimes \text{ReLU}(W_{s1} \otimes F_{wave})). \quad (14)$$

其中,  $A_{small}$  表示小目标注意力机制;  $W_{s1}$  与  $W_{s2}$  表示权重系数.

坐标注意力则建模全局空间关系:

$$A_{coord} = [\sigma(W_{c2} \otimes \text{ReLU}(W_{c1} \otimes \frac{1}{HW} \sum_{i,j} F_{wave}))]. \quad (15)$$

其中,  $A_{coord}$  表示坐标注意力机制;  $W_{c1}$  与  $W_{c2}$  表示权重系数.

通过点乘得到最终的旋转注意力图:

$$A_{rot} = A_{small} \odot A_{coord}. \quad (16)$$

其中,  $A_{rot}$  表示双重注意力机制的输出.

在逆旋转与特征融合阶段, 通过逆变换将注意力图还原至原始视角, 并通过逐元素相乘调制原始特征, 完成整个旋转不变特征增强过程:

$$\mathbf{X}_{RAM} = \mathbf{X} \odot \sum_{k=1}^8 w_k \cdot R_{-\theta_k}(A_{rot}). \quad (17)$$

其中,  $R_{-\theta_k}$  表示逆旋转操作;  $\mathbf{X}_{RAM}$  表示旋转不变注意力机制的输出.

## 1.2 自适应加权多尺度特征过滤融合 (AMFF)

在颈部特征融合阶段, YOLOv12 在处理背景复杂、小目标密集的无人机图像时存在明显局限. 其一, 简单的逐元素相加或通道拼接方式融合特征, 对浅层特征中背景噪声的抑制有限. 其二, 融合权重往往固定不变, 无法根据输入图像自适应地调整.

为克服上述缺陷, 本文提出一种自适应加权多尺度特征过滤融合模块 (AMFF), 整体结构如图 4 所示. AMFF 引入“过滤”与“自适应”两大核心机制. 在“过滤”阶段, 通过双域协同注意力 (Dual Domain collaborative attention, DDCA) 与拉普拉斯边缘增强器 (Laplace Edge Enhancer, LEE) 构建特征过滤器 (Multi-scale Feature Filtering, MFF), 从空间和频域两个维度滤除浅层特征中的背景噪声并增强有益的高

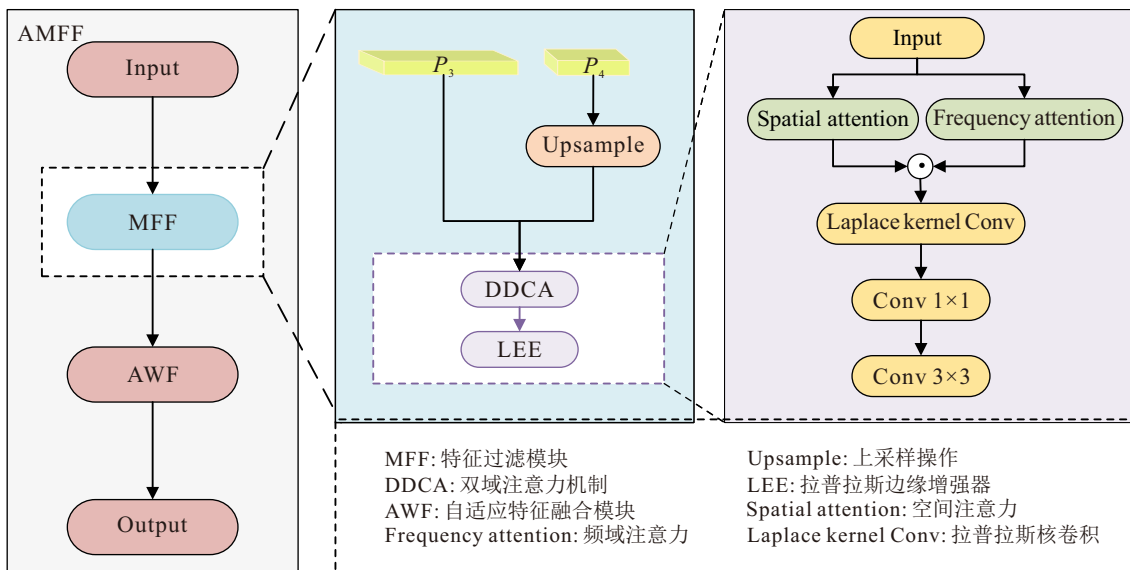


图4 AMFF 模块结构图

频细节. 在“自适应”阶段, 提出可学习的自适应加权融合机制 (Adaptive weighted fusion, AWF), 通过可学习的权重参数, 动态调整不同层级特征的融合贡献度, 使网络能够根据具体场景自主决定依赖深层语义信息还是浅层细节信息.

### 1.2.1 特征过滤器 (MFF)

现有检测器通常仅聚焦于空间域或通道域的单维度建模来增强特征响应, 难以有效解耦和增强领域中蕴藏的关键信息. 尤其对于极小目标, 其特征信号微弱且在多次卷积操作中易被平滑滤波抑制.

为此, 本文设计了双域协同注意力机制 (DDCA), 通过一个共享的瓶颈结构联合生成通道级频域注意力图  $F_{att}$  (用于增强与高频边缘/纹理相关的重要特征通道) 和空间注意力图  $S_{att}$  (用于抑制背景主导的低频区域), 从而实现频域-空间特征的协同解耦与增强, 并通过逐元素相乘得到最终的双域注意力.

$$F_{att} = \sigma(W_2^f \text{ReLU}(W_1^f(\text{GAP}(\mathbf{X})))), \quad (18)$$

$$S_{att} = \sigma(W_3^s \text{ReLU}(W_1^s(\text{GAP}(\mathbf{X})))), \quad (19)$$

$$A_{DDCA} = F_{att} \odot S_{att}. \quad (20)$$

其中,  $W$  表示可学习权重;  $A_{DDCA}$  表示双域协同注意力机制的输出.

此外, 针对小目标的边缘、角点等高频特征在深层网络中易被平滑滤波削弱的问题, 本文使用 LEE 增强梯度对比. LEE 采用优化的拉普拉斯核  $K_{lap}$  提取显著梯度响应, 并借助轻量瓶颈结构与空间调制卷积实现可学习的特征增强.

$$H = K_{lap} * A_{DDCA}, \quad K_{lap} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}, \quad (21)$$

$$\psi = \sigma(\text{Conv}_{3 \times 3}(\text{ReLU}(\text{Conv}_{1 \times 1}(\text{ReLU}(\text{Conv}_{1 \times 1}(H))))) \quad (22)$$

其中,  $\psi$  表示特征过滤器的输出.

### 1.2.2 自适应加权融合 (AWF)

经过 MFF 处理后, 浅层特征中的背景噪声被有效滤除, 小目标特征占比得以增强, 信噪比显著提高. 然而, 静态融合机制不适用于目标尺度、背景复杂度的无人机目标检测, 难以在多样场景下始终保持最优特征组合, 限制了模型的泛化性能. 因此, 本文使用 AWF, 可根据输入图像的上下文内容, 动态地学习并分配不同层级特征在融合过程中的权重系数.

具体而言, 对于待融合的浅层细节特征与深层语义特征, 通过引入可训练的参数权重, 使网络能够自主判断当前场景下应更依赖细节信息还是语义信

息. 以第三、四层输出为例, 首先对  $P_4$  层进行上采样操作使其恢复至与  $P_3$  形状一致, 然后对  $P_3$  和  $P_4$  分别进行背景造成过滤以及目标细节特征增强, 最后通过自适应权重将两层特征进行加权融合.

$$P'_3 = \text{Conv} \left( \frac{\omega_1 \times \text{MFF}(P_3) + \omega_2 \times \text{Upsample}(\text{MFF}(P_4))}{\omega_1 + \omega_2 + \varepsilon} \right). \quad (23)$$

其中, Upsample 表示上采样操作,  $\omega_1$ 、 $\omega_2$  表示权重系数,  $\varepsilon$  极小常数,  $P'_3$  表示  $P_3$  层输出.

权重参数  $\omega_1$ 、 $\omega_2$  的训练通过梯度下降与分类、回归损失联合优化, 确保其能够根据具体场景自主决定依赖深层语义信息还是浅层细节信息. 不依赖额外的监督信号, 而是与整个检测网络的优化目标一致. 通过端到端的训练过程, 网络学习到在不同场景下如何平衡细节信息与语义信息的融合比例. 这种设计使 AMFF 模块在测试阶段能够根据输入图像的内容特性, 自动调整融合策略, 实现真正的自适应融合.

### 1.3 高分辨率小目标检测头 (Tiny Head)

随着网络深度的增加, 特征图的感受野不断扩大, 但空间分辨率却显著降低. 这种分辨率的损失对于常规尺寸目标的检测影响尚可接受, 但对于无人机航拍图像中广泛存在的极小目标而言, 则是致命的. 如图 5 所示, 以 YOLOv12 的检测头设计为例, 其多尺度检测机制依赖于来自网络不同深度的特征图:  $P_3$ 、 $P_4$  和  $P_5$  层对应的下采样步长分别为 8、16 和 32. 这些层分别擅长检测中、大型目标, 但对于像素尺寸可能小于  $8 \times 8$  的极小目标, 它们在  $P_3$  层上仅能保留极其有限的信息, 而在更深的  $P_4$  和  $P_5$  层上, 这些微小目标几乎完全湮没于背景噪声中, 导致严重的漏检和定位偏差.

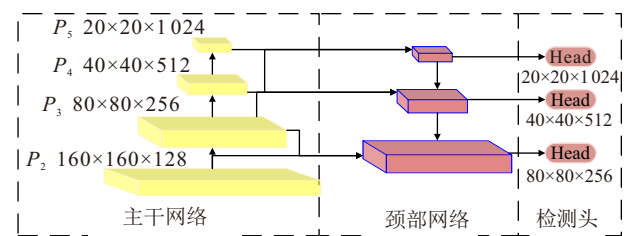


图5 YOLOv12 检测头结构

为解决上述问题, 本文创新性地设计了一个专为极小目标检测设计的高分辨率检测头. 如图 6 所示, 该检测头直接利用来自主干网络浅层的特征图 (其下采样步长为 4, 对应特征图尺寸为  $160 \times 160$ ), 保持高空间分辨率. 首先, 它极大缓解了微小目标在特征提取过程中的信息衰减, 使  $4 \times 4$  像素级的超小

目标仍然能够在特征图上保留可辨识的结构信息和边缘梯度,为分类提供充分依据.其次,高分辨率特征图为预测框回归提供了更为精细的空间坐标参考,从而显著提升了边界框的定位精度,减少了回归误差.

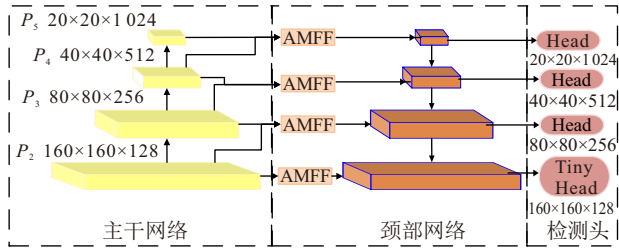


图6 YOLO-MAT 检测头结构

值得注意的是,高分辨率特征虽然富含细节,但缺乏深层的语义信息.为此,本文通过AMFF模块将来自深层的、富含语义信息的特征经过上采样和过滤后,与浅层高分辨率特征进行自适应融合.这种设计使得高分辨率检测头不仅“看得清”,而且“认得准”,即在保留细节优势的同时,融入对目标类别的高层语义理解,以更适用无人机目标检测.

## 2 实验结果与分析

### 2.1 实验环境

本文实验在64位Windows11操作系统下进行,硬件配置包括Intel i5-13400 CPU和NVIDIA RTX4060 GPU(显存为8GB).软件环境基于Pytorch 2.2.2 框架,并搭配CUDA 11.8 加速库,Python版本为3.8.3.实验参数设置如下表1.

表1 训练参数设置

参数名称	参数值
Epoch	300
Batch-size	8
初始学习率	0.01
优化器	SGD
图片分辨率	640×640

### 2.2 数据集

VisDrone2019公共数据集由天津大学AISKYEYE团队创建,采集自多场景下无人机拍摄的图像,涵盖不同城市、天气条件与光照环境.该数据集包含10,209张静态图像,包括行人、自行车、汽车等10种目标类别.NWPU VHR-10公共数据集是由西北工业大学团队创建,涵盖10个类别的地理空间物体.数据集包含800张图像,包含650张包含目标的图像和150张背景图像,包括飞机、舰船、油罐等10种目标类别.

### 2.3 评价指标

实验选取精确率(Precision)、召回率(Recall)、平均精度均值(mAP)为评价指标.Precision表示模型预测为正样本中实际为正样本的比例.Recall衡量模型识别出的正样本占真正样本总数的比例.mAP是综合性能评价指标,表示所有类别的平均精度(AP)的平均值.计算公式如下:

$$\text{Precision} = TP / (TP + FP), \quad (24)$$

$$\text{Recall} = TP / (TP + FN), \quad (25)$$

$$\text{mAP} = \int_0^1 \text{Precision}(\text{Recall})d(\text{Recall})/N. \quad (26)$$

其中,TP指实际为正样本且被正确预测为正样本的数量,FP指实际为负样本但被错误预测为正样本的数量,FN指实际为正样本但被错误预测为负样本的数量,N表示目标总类别数.

### 2.4 对比实验分析

为验证算法的有效性,将YOLO-MAT与YOLOv5、YOLOv8、YOLOv10、YOLOv12以及改进的YOLO-HV<sup>[23]</sup>进行对比,结果如表2所示(最优指标已加粗并带下划线).与基准模型相比,本文提出的YOLO-MAT在多项关键指标上均有显著提升,参数量由2.6M降至2.54M,模型体积由5.8MB压缩至5.2MB,GFLOPs由6.5降低至4.9,同时推理速度提升至178 FPS.在检测精度方面,mAP相较于YOLOv12提升6.7%.以上结果表明,YOLO-MAT在保持模型轻量化的同时,有效兼顾了检测精度与推理效率,能够更好地满足无人机平台对实时目标检测任务中高精度与高效性的双重需求.与改进的YOLO-HV模型相比,本文模型检测精度虽然没有过多的提升,但是本文使用的参数量和计算量与之相比大大减少,更方便部署于无人机.

为更直观展示模型的优越性,对检测结果进行了可视化分析.图7(a)所示的低光照与树木遮挡场景中,基准模型未能检测出右下角被遮挡目标,而本文模型仍可准确检出.图7(b)中,基准模型在行人、摩托车等小目标密集区域出现漏检.图7(c)显示,在强光照条件下,基准模型将汽车误检为三轮车、将厢式货车误检为汽车,而本文模型可以准确有效检测出各个目标.进一步表明,本文模型在不同天气和遮挡条件下均表现出更优的检测鲁棒性和准确性;而基准模型存在明显的漏检误检.

### 2.5 消融实验

为了验证YOLO-MAT模型中各模块的有效性,

表2 不同目标检测算法在 VisDrone2019 数据集的对比结果

模型	参数(M)	GFLOPs	FPS	模型大小	mAP@50(small object)						
					All	Pedestrian	People	Bicycle	Tricycle	Awing-Tricycle	Motor
YOLOv5	2.6	7.7	149	5.3MB	0.22	0.295	0.232	0.0344	0.0964	0.0491	0.292
YOLOv8	2.7	8.7	132	6.23MB	0.30	0.320	0.221	0.084	0.183	0.109	0.333
YOLOv10	2.69	8.2	149	5.9MB	0.30	0.300	0.25	0.072	0.178	0.109	0.331
YOLOv12	2.6	6.5	161	5.8MB	0.32	0.343	0.274	0.0988	0.195	0.11	0.363
YOLO-HV	38.5	111.9	—	—	0.381	—	—	—	—	—	—
Faster R-CNN	41.4	—	—	—	0.218	0.209	0.148	0.073	0.14	0.088	0.212
Cascade R-CNN	26.29	—	—	—	0.232	0.222	0.148	0.076	0.148	0.086	0.214
RetinaNet	69.2	—	—	—	0.139	0.13	0.079	0.014	0.063	0.042	0.118
DMNet	23.8	—	—	—	0.303	0.285	0.204	<b>0.159</b>	0.206	0.12	0.292
DETR	42.42	112.1	65	72.1MB	0.269	—	—	—	—	—	—
RT-DETR	32	103.5	67	66.2MB	0.28	0.31	0.215	<b>0.02</b>	0.171	0.087	0.319
YOLO-MAT(Ours)	<b>2.54</b>	<b>4.9</b>	<b>178</b>	<b>5.2MB</b>	<b>0.387</b>	<b>0.346</b>	<b>0.295</b>	0.142	<b>0.252</b>	<b>0.156</b>	<b>0.388</b>

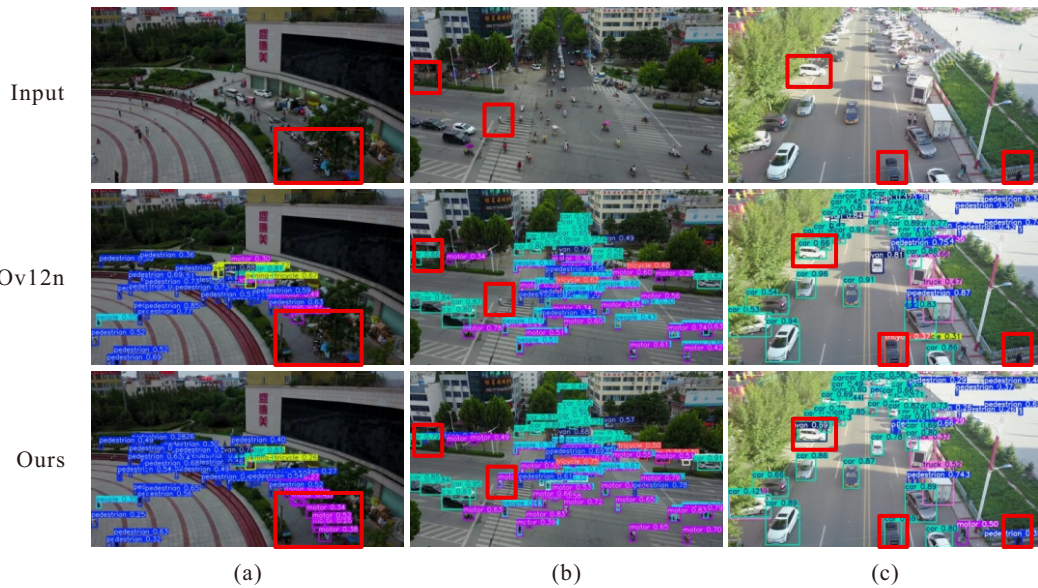


图7 在 VisDrone2019 数据集检测结果可视化效果图

基于基准模型进行消融实验. 实验以堆叠方式进行, 逐步将各模块添加到基准模型中, 以评估其贡献. 从表3可以看出, 将 A2C2f 模块替换为 MRAC2f 模块, 参数减少了 0.2M, mAP 分别提升 1.9% 和 7.8%, 表明 MRAC2f 模块在精度和参数量之间取得了平衡. 使用 AMFF 模块后, 模型参数减少 0.22M, mAP 分别提升 1.3% 和 3.8%, 说明该结构在精度和参数量

之间也达到良好平衡. 添加高分辨率小目标检测头使 mAP 进一步提高 1.6% 和 3.4%, 凸显了该组件在提升小目标检测能力方面的重要作用. 各模块组合后模型性能持续提升, 表明彼此兼容, 整体方案可行.

## 2.6 模块有效性验证

### 2.6.1 MRAC2f 模块的有效性

图8展示了 MRAC2f 模块与基准模型的检

表3 不同数据集上消融实验的对比结果

MRAC2f	AMFF	Tiny Head	VisDrone2019			NWPU VHR-10			Params(M)
			Precision	Recall	mAP50	Precision	Recall	mAP50	
			0.425	0.318	0.32	0.828	0.782	0.829	2.6
√			0.442	0.322	0.339	0.868	0.857	0.907	2.4
	√		0.431	0.326	0.333	0.848	0.839	0.867	<b>2.38</b>
		√	0.442	0.322	0.336	0.838	0.827	0.863	2.62
√	√		0.44	0.337	0.343	0.9	0.845	0.912	2.46
√		√	0.445	0.324	0.34	0.91	0.859	0.903	2.48
	√	√	0.436	0.328	0.348	0.922	0.856	0.917	2.45
√	√	√	<b>0.463</b>	<b>0.339</b>	<b>0.388</b>	<b>0.934</b>	<b>0.865</b>	<b>0.922</b>	2.54

测效果对比. 为分析该模块的旋转不变性, 在 VisDrone2019 数据集中选取同一输入图像, 对其进行了八个不同角度的旋转变换, 并分别进行检测. 从结果中可以看出, 在未旋转条件下, 基准模型在左侧

出现漏检, 而 MRAC2f 模块能够实现完整检测. 在不同旋转角度下, 基准模型均出现明显的漏检和误检现象; 相比之下, MRAC2f 模块检测性能显著优于基准模型, 表明该方法具备更强的旋转鲁棒性.

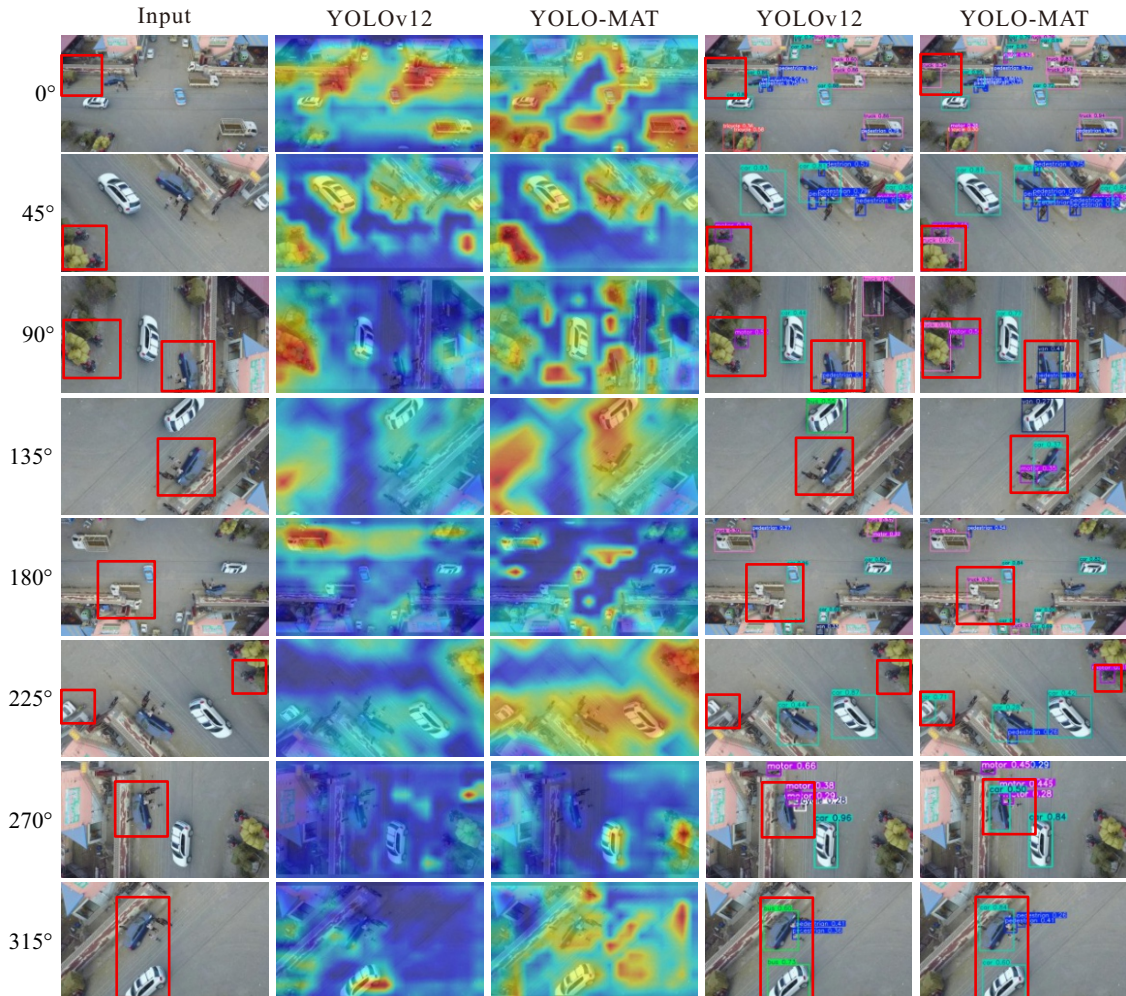


图8 MRAC2f 模块在 VisDrone2019 数据集检测结果可视化效果图

为说明 MRAC2f 模块中所采用方向数量与检测性能的关系, 对不同方向组设置进行了实验对比分析. 如表 4 所示, 在不同方向数配置下, 模型检测精度相较于基准模型均有明显提升, 表明引入方向感知机制的有效性. 具体而言, 当方向数由四组增至八组时, 尽管其参数规模仅出现轻微增长, 但是模型在检测精度上取得了更为显著的提升. 该结果表明, 采用八组方向能够在模型复杂度与特征分辨能力之间实现更优的平衡, 从而有效提升对旋转目标的表征性能.

表4 不同方向组数在 NWPU VHR-10 数据集对比实验

维度个数	mAP	参数(M)
0(YOLOv12)	0.829	2.6
2	0.853	<b>2.46</b>
4	0.877	2.50
8(Ours)	<b>0.922</b>	2.54

## 2.6.2 AMFF 模块的有效性

图 9 展示了 AMFF 模块与基准模型的可视化对比结果. 在 VisDrone2019 数据集中选取三张具有不同挑战性的图像进行测试, 可见原模型在不同场景下均出现不同程度的漏检. 具体而言: 在 9(a) 光照良好的条件下, 原模型对右下角的面包车出现漏检; 在 9(b) 小目标密集且角度异常的情况下, 原模型对中间密集小目标出现严重漏检; 在 9(c) 光照不足的场景中, 原模型未能检测出右侧被遮挡的卡车. 而添加 AMFF 模块后, 检测结果有明显的改进, 进一步证明了 AMFF 模块的有效性.

为验证 AWF 机制的有效性, 对不同层级的融合权重设置进行了消融实验分析. 如表 5 所示, 采用固定权重系数的融合策略在检测精度 (mAP) 上均显著低于本文提出的 AWF 机制. 该结果充分表明, 通过

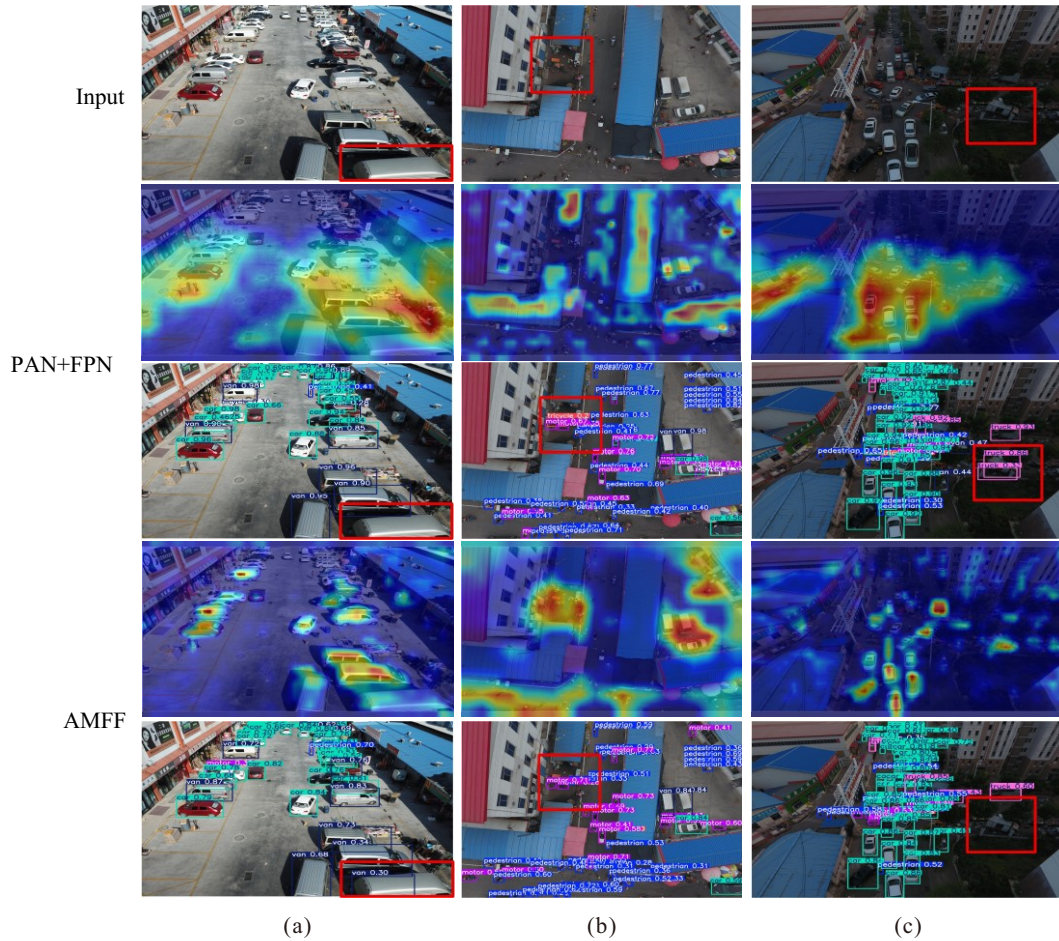


图9 AMFF 模块在 VisDrone2019 数据集检测结果可视化效果图

可学习参数动态调整各层级特征贡献度的自适应策略,能够更有效地融合多尺度信息,从而在不同场景下实现更优的检测性能.

表5 不同融合权重在 NWPU VHR-10 数据集对比实验

$\omega_1$	$\omega_2$	mAP
1/2	1/2	0.864
1/3	2/3	0.89
2/3	1/3	0.885
Adaptive	Adaptive	<b>0.922</b>

### 2.6.3 Tiny Head 模块的有效性

图 10 展示了添加高分辨率小目标检测头后的检测结果. 10(a) 光照良好的条件下,原模型对左侧小目标存在漏检并将有顶三轮车误检成三轮车. 10(b) 曝光的情况下,原模型未检测到远处的行人并将电线杆误检成行人. 10(c) 光照条件不足的情况下,原模型未检测到右侧被遮挡的目标并将左侧门楼顶端误检成行人. 针对小目标密集的图像测试显示,原检测头在不同条件下均出现漏检和误检;而增加高分辨率小目标检测头后,检测效果显著改善.

## 3 结论

针对无人机航拍图像目标检测中存在的旋转目

标敏感、小目标特征弱化以及复杂背景干扰等问题,提出了一套系统性的解决方案. 首先,通过设计 MRAC2f 模块,融合方向敏感卷积与频域权重预测,实现了对旋转目标的鲁棒表征. 其次,通过构建 AMFF 模块,有效抑制了背景噪声干扰,并增强小目标的高频特征. 最后,设计的  $160 \times 160$  高分辨率小目标检测头 (Tiny Head),显著缓解了深层网络中的特征退化现象. 在公开数据集上的实验验证了本文方法的有效性和优越性. 尽管 YOLO-MAT 在实验中表现出色,未来将进一步深入研究多模态融合机制以应对极端天气条件,探索基于极坐标或流形的连续旋转表征方法,突破当前离散方向建模的局限性.

### 参考文献 (References)

- [1] 闫建红, 冉云霄. 基于 YOLOv8 的轻量化无人机图像目标检测算法[J]. 图学学报, 2024, 45(6): 1328-1337. (Yan J H, Ran T X. Lightweight UAV image target detection algorithm based on YOLOv8[J]. *Journal of Graphics*, 2024, 45(6): 1328-1337.)
- [2] 李利霞, 王鑫, 王军, 等. 基于特征融合与注意力机制的无人机图像小目标检测算法[J]. 图学学报, 2023, 44(4): 658-666. (Li L X, Wang X, Wang J, et al. Small object detection algorithm in UAV image based on feature fusion and

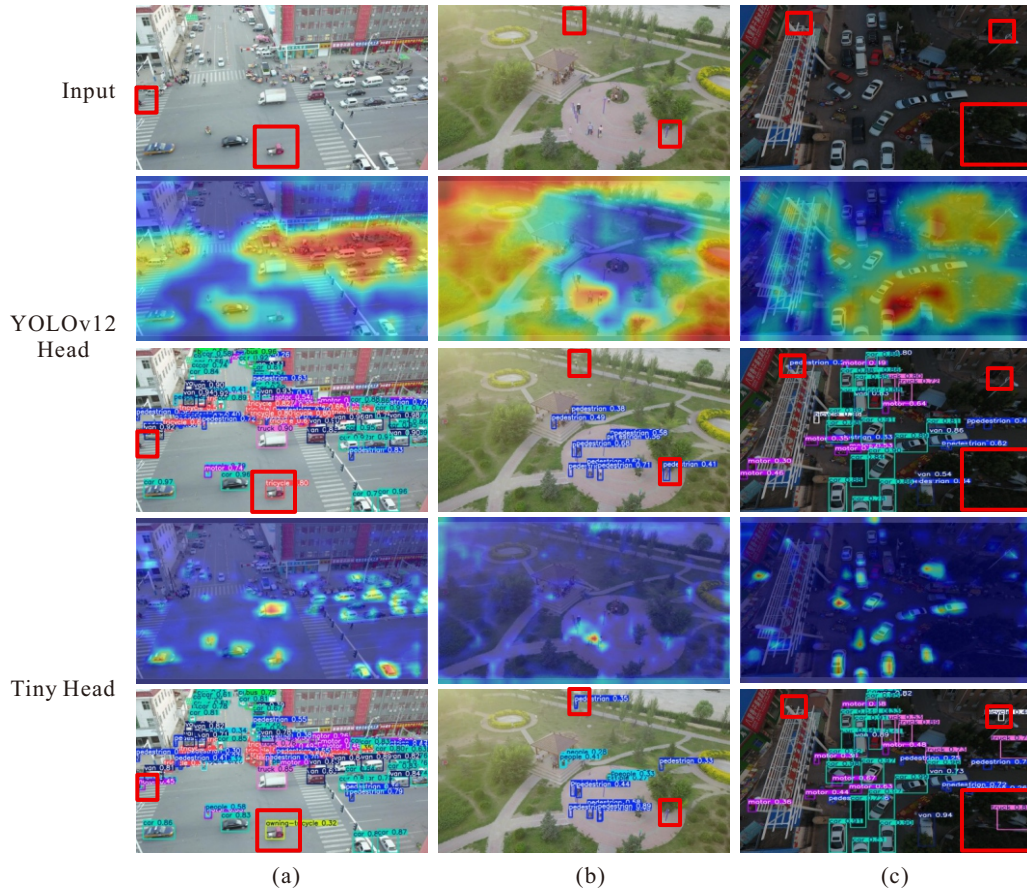


图10 Tiny Head 在 VisDrone2019 数据集检测结果可视化效果图

- attention mechanism[J]. *Journal of Graphics*, 2023, 44(4): 658-666.)
- [3] Xu X Y, Zhao M, Shi P X, et al. Crack detection and comparison study based on faster R-CNN and mask R-CNN[J]. *Sensors*, 2022, 22(3): 1215.
- [4] Murat A A, Kiran M S. A comprehensive review on YOLO versions for object detection[J]. *Engineering Science and Technology, an International Journal*, 2025, 70: 102161.
- [5] 胡冬波, 赵吉文, 张晓虎. 基于图像增强局部上采样 SSD 的直线电机定子非接触位置检测方法[J]. *控制与决策*, 2023, 38(6): 1629-1636.  
(Hu D B, Zhao J W, Zhang X H. Research on non-contact position detection method of linear motor mover based on image enhanced local upsampling SSD[J]. *Control and Decision*, 2023, 38(6): 1629-1636.)
- [6] 李琼, 考月英, 张莹, 等. 面向无人机航拍图像的目标检测研究综述[J]. *图学学报*, 2024, 45(6): 1145-1164.  
(Li Q, Kao Y Y, Zhang Y, et al. Review on object detection in UAV aerial images[J]. *Journal of Graphics*, 2024, 45(6): 1145-1164.)
- [7] Zhang H B, Zhou X W, Lan X G, et al. A real-time robotic grasping approach with oriented anchor box[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021, 51(5): 3014-3025.
- [8] Fu X, Huang K J, Sidiropoulos N D, et al. Anchor-free correlated topic modeling[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 41(5): 1056-1071.
- [9] Wei X L, Yin L, Zhang L L, et al. DV-DETR: Improved UAV aerial small target detection algorithm based on RT-DETR[J]. *Sensors*, 2024, 24(22): 7376.
- [10] 高卫峰, 易宇轩, 黄玲玲, 等. 一种高效的无人机航拍小目标检测算法[J]. *控制与决策*, 2025, 40(8): 2525-2533.  
(Gao W F, Yi Y X, Huang L L, et al. An efficient algorithm for small object detection in unmanned aerial vehicle images[J]. *Control and Decision*, 2025, 40(8): 2525-2533.)
- [11] 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.)
- [12] Xie B J, Wang Y J, Han M H, et al. Density-guided two-stage small object detection in UAV images[J]. *Expert Systems with Applications*, 2026, 297: 129346.
- [13] Zhang D W, Wang Y B, Wu Y X, et al. Multifrequency integration and scale-frequency linear attention for aerial tracking[J]. *IEEE Transactions on Instrumentation and Measurement*, 2025, 74: 5041113.
- [14] Zhu H F, Huang Y H, Xu Y, et al. Unmanned aerial vehicle (UAV) object detection algorithm based on keypoints representation and rotated distance-IoU

- loss[J]. *Journal of Real-Time Image Processing*, 2024, 21(2): 58.
- [15] Du S, Chen Y Q, Ling J, et al. Visual rotation detection algorithm for power grid UAV inspection targets based on YOLO-OBB[J]. *IEEE Access*, 2025, 13: 202354-202361.
- [16] Wang C H, Siang Y S, Yang X S, et al. Dynamic object removal and background reconstruction in aerial images via global alignment and YOLOv12[J]. *The Visual Computer*, 2025, 41(14): 12091-12107.
- [17] Liang E Q, Wei D P, Li F, et al. Object detection model of vehicle-road cooperative autonomous driving based on improved YOLO11 algorithm[J]. *Scientific Reports*, 2025, 15: 32348.
- [18] Liang Z H, Zhu T T, Teng G, et al. YOLO-RGDD: A novel method for the online detection of tomato surface defects[J]. *Foods*, 2025, 14(14): 2513.
- [19] Gong X J, Yu J, Zhang H Y, et al. AED-YOLO11: A small object detection model based on YOLO11[J]. *Digital Signal Processing*, 2025, 166: 105411.
- [20] Zhao D P, Wang C X, Gao Y, et al. Semantic segmentation of remote sensing image based on regional self-attention mechanism[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 8010305.
- [21] 牛为华, 郭迅. 基于改进 YOLOv8 的船舰遥感图像旋转目标检测算法[J]. *图学学报*, 2024, 45(4): 726-735. (Niu W H, Guo X. Rotating target detection algorithm in ship remote sensing images based on YOLOv8[J]. *Journal of Graphics*, 2024, 45(4): 726-735.)
- [22] 崔克彬, 焦静颐. 基于 MCB-FAH-YOLOv8 的钢材表面缺陷检测算法[J]. *图学学报*, 2024, 45(1): 112-125. (Cui K B, Jiao J Y. Steel surface defect detection algorithm based on MCB-FAH-YOLOv8[J]. *Journal of Graphics*, 2024, 45(1): 112-125.)
- [23] Xu S Z, Zhang M J, Chen J Y, et al. YOLO-HyperVision: A vision transformer backbone-based enhancement of YOLOv5 for detection of dynamic traffic information[J]. *Egyptian Informatics Journal*, 2024, 27: 100523.

### 作者简介

邓承志 (1980-), 男, 教授, 博士, 主要研究方向为人工智能、图像处理、目标检测, E-mail: [dengcz@nit.edu.cn](mailto:dengcz@nit.edu.cn);

武瑛博 (2001-), 男, 硕士, 主要研究方向为图像处理和目标检测, E-mail: [wuyb\\_wio@foxmail.com](mailto:wuyb_wio@foxmail.com);

吴朝明 (1979-), 男, 副教授, 博士, 主要研究方向为人工智能、机器视觉和高光谱影像处理, E-mail: [wzm@nit.edu.cn](mailto:wzm@nit.edu.cn);

孙小惟 (1982-), 女, 讲师, 硕士, 主要研究方向为模式识别和高光谱影像处理, E-mail: [2013994451@nit.edu.cn](mailto:2013994451@nit.edu.cn);

汪胜前 (1965-), 男, 教授, 博士, 主要研究方向为图像处理 and 模式识别, E-mail: [2008994050@nit.edu.cn](mailto:2008994050@nit.edu.cn).