

任洪娥,张佳源,王金聪,等.基于改进YOLOv12的无人机图像小目标检测方法[J].辽宁工程技术大学学报(自然科学版),2026,45(3):349-358.DOI:10.11956/j.issn.1008-0562.20250493

REN Hong'e,ZHANG Jiayuan,WANG Jincong,et al.Small target detection method of UAV image based on improved YOLOv12[J]. Journal of Liaoning Technical University(Natural Science),2026,45(3):349-358.DOI:10.11956/j.issn.1008-0562.20250493

基于改进YOLOv12的无人机图像小目标检测方法

任洪娥^{1,2}, 张佳源¹, 王金聪^{1,2}, 郭继峰³

(1. 东北林业大学 计算机与控制工程学院, 黑龙江 哈尔滨 150040; 2. 黑龙江省林业智能装备工程研究中心, 黑龙江 哈尔滨 150040; 3. 桂林航天工业学院 计算机科学与工程学院, 广西 桂林 541000)

摘 要: 针对无人机遥感图像中小目标检测精度低、背景干扰强、特征表达能力弱等问题, 提出一种融合多核特征提取与三重注意力机制的目标检测算法。在YOLOv12框架上, 引入MogaNet双域聚合机制增强多尺度特征提取能力, 嵌入CA注意力模块提升空间定位精度, 并设计三重注意力机制, 在空间、通道与尺度维度联合强化小目标关注度, 结合CMUNeXt特征融合策略优化跨层级信息传递。研究结果表明: 该算法在VisDrone2019和NWPU VHR-10数据集上的 $I_{mAP@0.5}$ 分别达到0.331和0.808, 较YOLOv12模型分别提升5.7%和15.0%, 显著改善了复杂背景下小目标的检测精度与鲁棒性。

关键词: 无人机遥感图像; 小目标检测; 三重注意力机制; YOLOv12模型; 多尺度特征提取

中图分类号: TP399

文献标志码: A

文章编号: 1008-0562(2026)03-0349-10

Small target detection method of UAV image based on improved YOLOv12

REN Hong'e^{1,2}, ZHANG Jiayuan¹, WANG Jincong^{1,2}, GUO Jifeng³

(1. College of Computer and Control Engineering, Northeast Forestry University, Harbin 150040, China;

2. Heilongjiang Forestry Intelligent Equipment Engineering Research Center, Harbin 150040, China;

3. School of Computer Science and Engineering, Guilin University of Aerospace Technology, Guilin 541000, China)

Abstract: Aiming at the problems of low detection accuracy, strong background interference and weak feature expression ability of small targets in unmanned aerial vehicle (UAV) remote sensing images, a target detection algorithm combining multi-core feature extraction and triplet attention mechanism is proposed. Based on the YOLOv12 framework, the MogaNet dual-domain aggregation mechanism is introduced to enhance the multi-scale feature extraction ability, the coordinate attention (CA) module is embedded to improve the spatial positioning accuracy, and the triple attention mechanism is designed to jointly strengthen the attention to small targets in the spatial, channel and scale dimensions, and the CMUNeXt feature fusion strategy is combined to optimize cross-level information transmission. The research results show that the index $I_{mAP@0.5}$ of the algorithm on the VisDrone2019 and NWPU VHR-10 datasets reaches 0.331 and 0.808, respectively, which is 5.7% and 15.0% higher than the YOLOv12 model, and significantly improves the detection accuracy and robustness of small targets in complex backgrounds.

Keywords: UAV remote sensing image; small target detection; triplet attention; YOLOv12 model; multi-scale feature extraction

收稿日期: 2025-11-19 修回日期: 2026-01-06 接受日期: 2026-01-13 责任编辑: 焦丽

基金项目: 国家自然科学基金项目(62466012); 黑龙江省自然科学基金项目(LH2024F047)

作者简介: 任洪娥(1962-), 女, 黑龙江哈尔滨人, 博士, 教授, 主要从事计算机视觉与模式识别方面的研究。E-mail: nefu_rhe@163.com

0 引言

随着无人机(unmanned aerial vehicle, UAV)平台技术和高分辨率可见光成像设备的不断普及^[1], 无人机可见光小目标检测逐渐成为智能遥感、安防监控、交通管理、灾害救援等应用场景^[2]的重要基础支撑。与常规目标检测任务不同, 无人机平台下的可见光小目标检测面临着更为严峻的技术挑战, 迫切需要更加智能、高效、鲁棒的解决方案。

无人机可见光图像中的小目标通常具有如下特征: 一是目标所占像素极小^[3], 通常为十几个甚至更少的像素点, 缺乏完整的结构信息, 导致目标在视觉上难以与背景区分; 二是局部像素级外观表征不足、全局形状信息模糊。以卷积神经网络为代表的深度网络框架为特征提取与目标检测及分类提供了新路径, 但针对无人机平台目标检测任务而言, 有限的像素级外观表征仍难以提供充足的细节判别信息。

无人机高空拍摄面临大视场、远距离、运动模糊、光照变化、大范围遮挡等问题^[3]。远距离大视场的复杂背景包含地表反光、树影等干扰因素, 增加了目标筛选难度; 无人机运动或风力因素导致的成像模糊, 显著提升了微小目标的分辨难度; 树木、建筑等造成的大范围遮挡, 易引发漏检; 多场景跨区域作业要求检测系统具备强适应性。因此, 提升复杂背景下小目标的感知、区分与定位能力, 是无人机视觉智能感知领域研究的核心难点。

小目标检测因其像素占比低、特征信息弱, 始终是计算机视觉领域的挑战。在无人机高空视角下, 视角变化、目标形变及大气扰动等问题进一步加剧了目标边缘模糊与特征辨识难度, 导致传统检测框架性能严重下降。针对上述问题, 现有研究主要从三个方向展开探索。其一, 优化多尺度特征融合机制^[4-6], 通过设计更精细的特征金字塔网络(FPN)来保留并增强浅层特征, 例如引入全局语义与局部注意力来强化特定尺度信息, 或利用神经架构搜索(NAS)与无损下采样模块来优化融合结构。这类方法普遍存在模型计算复杂度与内存消耗显著增加的问题, 难以在资源受限的无人机平台高效部署。其二, 利用上下文信息增强检测性能^[7-9], 通过对目标周围环境线索进行建模来弥补其自身特征的不足。例如, 采用注意力机制或图卷积网络来捕获长程依赖与特征关

联, 或设计专用模块来集成局部与全局上下文。但这类方法在复杂背景下, 易将相似纹理的噪声(如植被、建筑等)误判为有效上下文, 反而引入干扰, 导致误检。其三, 结合超分辨率技术^[10-15], 重建或生成高分辨率特征, 提升小目标的清晰度。尽管该策略直观有效, 但计算开销巨大, 且易引入干扰检测的伪影。

综上, 现有检测方法虽在一定程度上提升了检测性能, 但在处理无人机遥感图像时仍面临以下挑战: 复杂背景(如树木、建筑阴影等)产生的强干扰; 小目标像素极少导致的特征表达微弱; 密集分布下的小目标定位精度低。为系统解决上述问题, 提出一种基于YOLOv12的改进检测方法, 在复杂背景下, 实现对极微小目标的特征重构与精准定位。针对背景干扰强的问题, 引入MogaNet双域聚合机制, 利用其特有的门控聚合策略, 在提取特征时动态抑制环境噪声权重, 强化目标纹理信息, 减少误检。针对小目标特征弱的问题, 采用CMUNeXt特征融合策略, 通过大核卷积与反向瓶颈设计, 在深层网络中保留并增强微小目标的全局上下文信息, 防止特征在下采样中丢失。针对定位精度低与密集遮挡的问题, 设计三重注意力机制(Triplet Attention)与通道注意力(coordinate attention, CA)机制, 在空间、通道与尺度三个维度上联合重构特征, 精准聚焦密集分布的小目标中心, 提升边界框回归的准确性。

1 改进YOLOv12模型

YOLOv12在YOLOv11的基础上, 针对网络结构优化、特征提取能力提升, 以及小目标检测性能增强进行改进, 引入多种结构创新与注意力机制, 进一步提升了模型在复杂场景下的表现。

YOLOv12的网络架构由4个通用模块构成, 分别为Input、Backbone、Neck和Prediction, 并包含CBS、SPPCSPC、MogaNet模块、Triplet Attention单元, 以及CA注意力模块等核心组件。输入图像经过数据增强等预处理后进入Backbone进行多层次特征提取, 融合CA与Triplet Attention提升全局与局部信息表达能力; Neck部分在CMUNeXt的指导下进行高效特征融合, 最终由检测头输出目标类别与位置预测结果。改进的YOLOv12框架见图1, 旨在保持推理速度的同时, 显著提升复杂背景、多尺度与小目标场景下的检测精度。

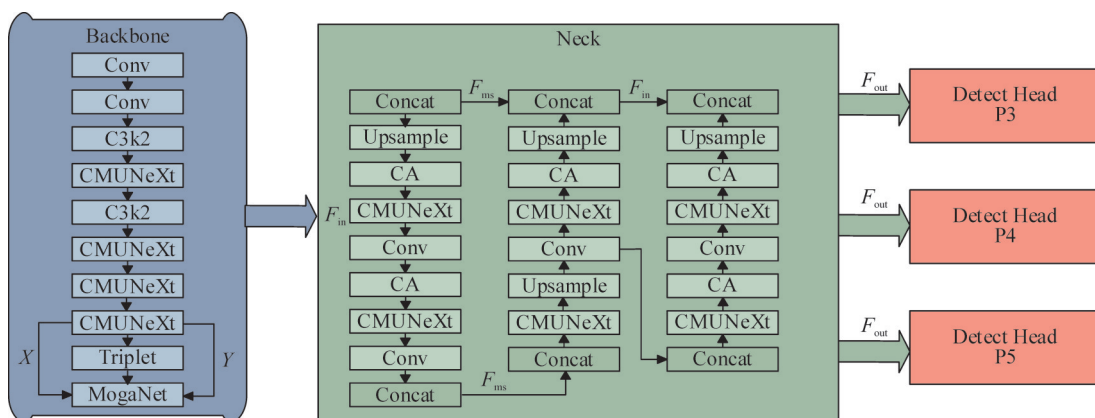


图1 改进的YOLOv12框架

Fig.1 improved YOLOv12 framework

1.1 MogaNet双域聚合机制

针对无人机航拍图像中背景纹理复杂易导致误检的问题,在主干网络中引入MogaNet模块(见图2)。与传统卷积不同,MogaNet采用双域聚合策略^[16],通过空间聚合和通道聚合两个分支,分别提取几何特征与语义特征。其核心优势在于引入了“门控”机制,根据输入特征的统计信息动态调整响应权重,有效过滤背景噪声并突出目标

特征。具体而言,给定输入特征映射 $X \in \mathbf{R}^{C \times H \times W}$,分别在空间域与通道域构建交互表示,其中空间聚合模块旨在建模像素之间的全局依赖关系。通过全局上下文编码器 $F_s(\cdot)$ 计算空间注意力权重,以聚焦或筛选关键区域信息,可表示为

$$A_s = \sigma(\text{Conv}_{1 \times 1}(F_s(X))), \quad (1)$$

式中, $\sigma(\cdot)$ 为sigmoid函数。

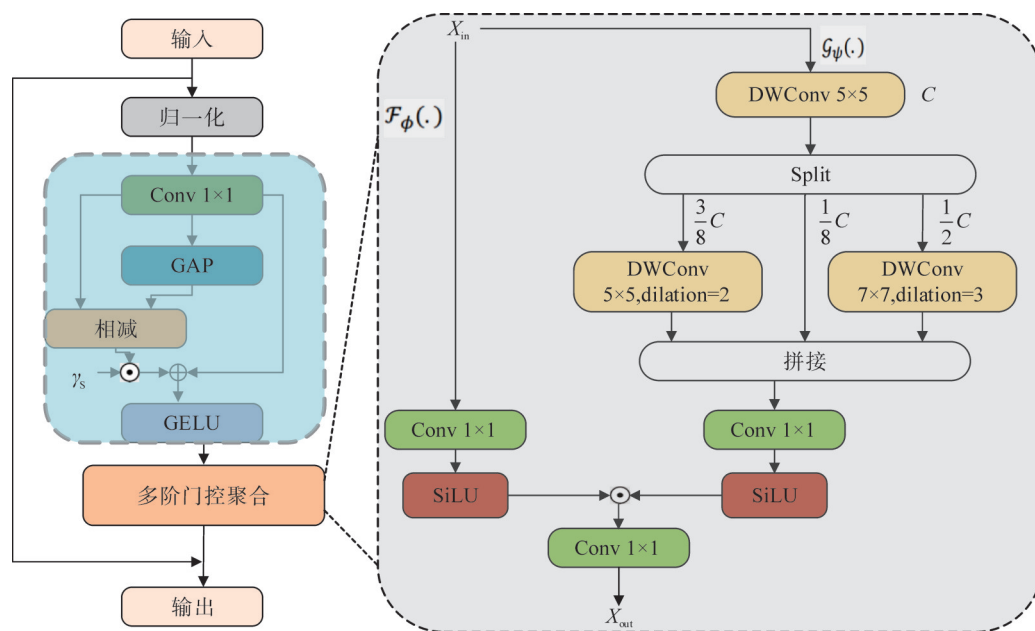


图2 MogaNet模型架构

Fig.2 MogaNet model architecture

利用 A_s 对输入进行加权和重标定,得到 $X_s = A_s \odot X$ 。通道聚合分支则通过全局特征描述子 $F_c(\cdot)$ 捕捉通道间的语义关联。

通过全局平均池化(global average pooling, GAP)与全局最大池化(global max pooling, GMP)提取统计信息 Z ,进而利用多层感知机

(multilayer perceptron, MLP)生成通道注意力权重,即

$$A_c = \sigma(\text{MLP}(Z)), \quad (2)$$

加权后的通道特征表示为

$$X_c = A_c \odot X. \quad (3)$$

MogaNet将两种交互特征融合,实现空间-通

道双域的特征增强与上下文建模。在多尺度与复杂背景下获得更具判别性的表示为

$$Y = \text{Fuse}(X_s, X_c) = X_s + X_c. \quad (4)$$

1.2 CA注意力机制

近年来, 诸多研究致力于进一步提升注意力机制的效率与表达能力。例如, 彭继慎等^[17]在 C2f-EF 模块中嵌入跨空间学习多尺度注意力机制 EMA, 以及 FasterNet Block, 优化了模型的运行效率; LIU 等^[18]提出 Dense Skip-Attention 模型, 在卷积神经网络 (convolutional neural network, CNN) 中引入密集跳跃连接的注意力模块, 使各注意力单元

之间建立交互, 强化了网络内部的特征建模能力。

CA 机制将二维特征图在水平与垂直方向上分别进行全局平均池化, 编码出目标在不同方向上的精确位置信息。这种设计对于无人机视角下密集分布的小目标 (如拥挤的人群、停车场车辆) 尤为重要。因为在密集场景下, 传统注意力机制容易将相邻目标的特征混淆, 而 CA 机制引入的坐标信息能够辅助模型更好地区分相邻目标的边界, 抑制冗余背景干扰, 从而显著提升模型在密集小目标检测中的定位精度与鲁棒性。CA 注意力机制模型架构见图 3。

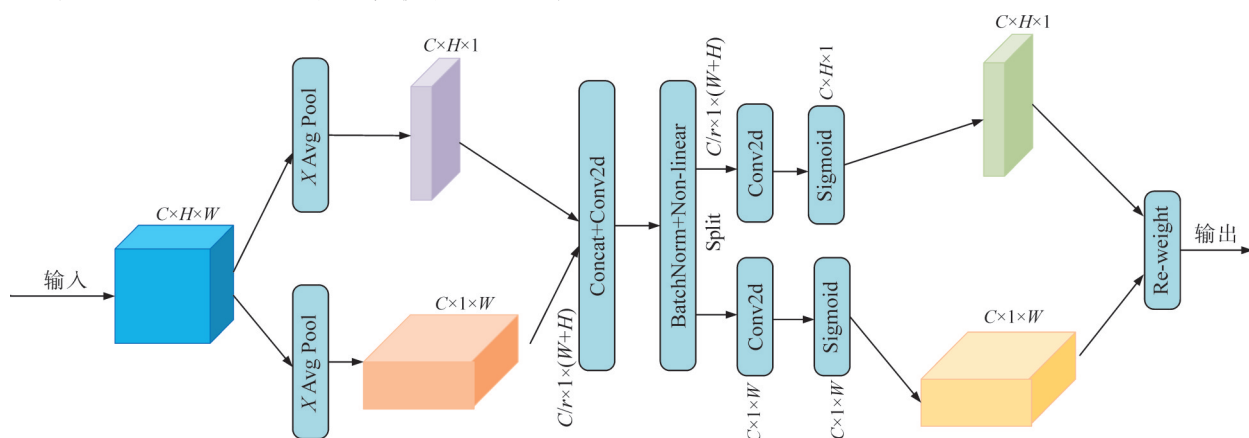


图 3 CA 注意力机制模型架构

Fig.3 CA attention mechanism model architecture

CA 注意力机制在垂直方向和水平方向分别执行全局平均池化, 提取方向感知的全局信息。其中, 垂直方向上的全局上下文可表示为

$$Z_h(c, h) = \frac{1}{W} \sum_{i=1}^W X(c, h, i), \quad (5)$$

水平方向上的全局上下文可表示为

$$Z_w(c, h) = \frac{1}{H} \sum_{j=1}^H X(c, j, w), \quad (6)$$

式 (5)、式 (6) 中: $Z_h \in \mathbf{R}^{C \times H \times 1}$; $Z_w \in \mathbf{R}^{C \times 1 \times H}$ 。

将两个方向的特征在通道维拼接并通过共享的 1×1 卷积 Conv_1 进行压缩映射, 表示为

$$f = \delta(\text{Conv}_1(Z_h \| Z_w)), \quad (7)$$

式中: $\delta(\cdot)$ 为非线性激活函数; “ $\|$ ” 表示拼接操作。

将融合特征 f 沿空间维度拆分为垂直分量 f_h 与水平分量 f_w , 通过 2 个独立的 1×1 卷积映射, 生成方向感知的注意力权重 $g_h \in \mathbf{R}^{C \times H \times 1}$ 和 $g_w \in \mathbf{R}^{C \times 1 \times H}$, 表示为

$$g_h = \sigma(\text{Conv}_h(f_h)), \quad (8)$$

$$g_w = \sigma(\text{Conv}_w(f_w)). \quad (9)$$

将两个方向的权重分别广播至输入特征的空间维度并进行逐元素相乘, 实现坐标感知的通道增强, 即

$$Y(c, h, w) = X(c, h, w) \odot g_h(c, h, 1) \odot g_w(c, 1, w). \quad (10)$$

该设计通过在通道注意力中显式编码空间位置信息, 使模型在捕捉长程依赖的同时具备精确定位能力, 尤其适用于密集小目标与复杂背景场景的检测任务。

1.3 三重注意力机制

在传统注意力机制的基础上, 进一步设计并引入一种更为精细的三重注意力机制, 以增强网络对遥感图像中复杂背景与小目标的感知能力。该机制融合了通道注意力、空间注意力和尺度注意力的信息增强路径, 从而实现更全面的特征重建与信息强化。三重注意力机制模型架构见图 4。

具体而言, Triplet Attention 模块通过并行路径对输入特征图进行三维度建模, 分别从通道相关性、空间分布以及尺度变化角度进行信息提取与加权, 有效捕捉目标的显著特征与上下文关联信

息。该模块被嵌入至主干网络的多个关键阶段,对不同层级的特征图进行引导与增强,使网络在

面对多尺度、密集、小尺寸目标时具备更强的判别力与鲁棒性。

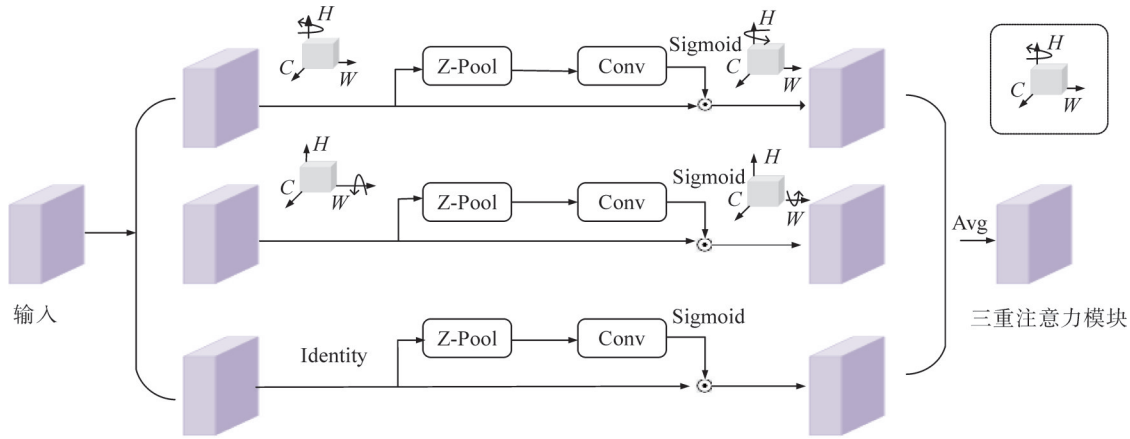


图4 三重注意力机制模型架构

Fig.4 triple attention mechanism model architecture

此外, Triplet Attention 在引入多维建模的同时,保持了较低的计算成本,适合在高效检测框架中部署。其联合注意机制不仅提升了小目标的检测精度,也有效抑制了背景干扰与伪目标响应,在遥感图像目标检测任务中具有更优的泛化能力和稳定性。

Triplet Attention 分别在通道-高度平面、通道-宽度平面、高度-宽度平面上构建注意力权重,计算过程可表示为

$$X_{CH} = \text{Pool}_W(X), A_{CH} = \sigma(\text{Conv}_{k \times k}(X_{CH})), \quad (11)$$

$$X_{CW} = \text{Pool}_H(X), A_{CW} = \sigma(\text{Conv}_{k \times k}(X_{CW})), \quad (12)$$

$$X_{HW} = \text{Pool}_C(X), A_{HW} = \sigma(\text{Conv}_{k \times k}(X_{HW})), \quad (13)$$

式(11)~式(13)中: Pool表示全局池化操作; $\text{Conv}_{k \times k}$ 表示 $k \times k$ 二维卷积。

将三种注意力权重融合,表示为

$$Y = \frac{1}{3}(A_{CH} \odot X + A_{CW} \odot X + A_{HW} \odot X). \quad (14)$$

式(14)实现了对输入特征的多维度加权,该机制可有效捕捉跨通道与空间的复杂依赖关系,提升特征表达的完整性与鲁棒性。

1.4 CMUNext特征融合结构

为缓解归纳偏差并高效提取全局上下文信息,在检测网络的Neck部分引入CMUNext特征融合模块。该模块融合了多尺度卷积与轻量化残差设计,在保证信息流充分传递的同时降低计算量,其特征融合结构见图5。

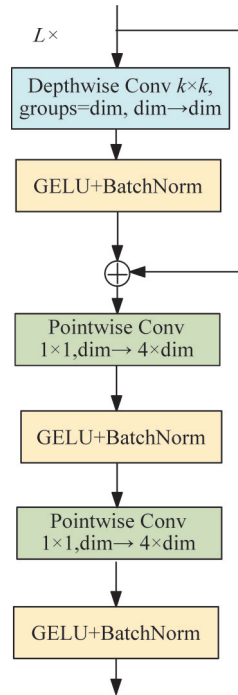


图5 CMUNext特征融合结构

Fig.5 CMUNext feature fusion architecture

CMUNext首先通过多尺度卷积组 $M(\cdot)$ 提取不同感受野的特征,计算公式为

$$F_{ms} = M(F_{in}) = \sum_{s \in S} \text{DWConv}_{k_s \times k_s}(F_{in}), \quad (15)$$

式中: S 为卷积核尺寸集合,以捕捉不同尺寸上下文信息; $\text{DWConv}(\cdot)$ 表示深度可分离卷积; $\delta(\cdot)$ 为激活函数; F_{in} 为输入特征。

通过通道融合卷积实现多尺度特征的统一映射,输出特征可表示为

$$F_{out} = F_{ms} + \text{Conv}_{1 \times 1}(\delta(\text{BN}(\text{DWConv}_{3 \times 3}(F_{ms})))), \quad (16)$$

式中, $\text{BN}(\cdot)$ 为批量归一化操作。

F_{out} 可直接送入检测头或后续特征融合模块, 该结构在减少计算开销的同时, 保证了跨尺度特征的有效整合, 并对小目标及密集目标的检测提供了更丰富的上下文信息支持。

2 实验设计与结果分析

2.1 实验环境及评价指标

实验在 Linux 操作系统平台上进行, 硬件配置与核心软件框架参数见表 1。模型输入的初始图像尺寸为 640×640 像素, 训练 100 次。

表 1 实验运行环境配置

Tab.1 experimental operating environment configuration	
实验环境	具体配置参数
CPU	Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50GHz
GPU	NVIDIA GeForce RTX 4090
编译器	Visual Studio Code
算法框架	Pytorch2.7.0+Python3.11+Cuda12.1

实验数据来自 AISKYEYE 团队所在的机器学习与数据挖掘实验室, 选用 VisDrone2019 数据集和 NWPU VHR-10 数据集 (2014 年发布)。其中, VisDrone2019 数据集由多种型号的无人机于不同城市与自然场景采集, 涵盖多种天气状况 (如晴天、阴天、雨天等)、光照条件 (强光、逆光、弱光等), 以及复杂的背景场景 (如城区、乡村、交通道路、运动场所等), 可为无人机平台下的智能感知与视觉分析提供数据支撑。

为全面评估目标检测模型的性能, 采用 I_{Pre} 、 I_{Rec} 、 $I_{\text{mAP}@0.5}$ 和 $I_{\text{mAP}@0.5:0.95}$ 作为核心评价指标。

I_{Pre} (准确率) 用于衡量预测为正样本的结果中, 真正正样本的占比。 I_{Rec} (召回率) 用于衡量所有真实目标被正确检测出的比例。 $I_{\text{mAP}@0.5}$ (mean average precision at IoU is 0.5) 是指 IoU 阈值为 0.5 时所有类别平均精度 (AP) 的均值。 $I_{\text{mAP}@0.5:0.95}$ 为 IoU 阈值在 0.50~0.95 (步长为 0.05) 范围内, 所有类别平均精度的均值, 用于评估模型的综合表现。

2.2 消融实验

消融实验通过逐步引入 CA 注意力机制、Triplet Attention、CMUNeXt 特征融合结构和 MogaNet 双域聚合机制, 系统验证各模块对模型检测性能的贡献。VisDrone2019 数据集实验结果见表 2。由表 2 可知, 在 VisDrone2019 数据集上, 实验 1 (基础配置) 的 $I_{\text{mAP}@0.5}$ 和 $I_{\text{mAP}@0.5:0.95}$ 分别为 0.274 和 0.155。随着组件的逐步叠加, 模型性能呈现阶

梯式提升, 实验 2 引入 CA 与 Triplet 双注意力机制后, 两项指标分别提升至 0.291 和 0.167。表明通道注意力机制能够有效优化通道维度的特征权重分配, 使模型对关键语义信息的捕捉更加精准。该模块通过建模空间-通道交互关系, 强化了特征的多维度判别能力, 显著改善了复杂场景下的目标定位和分类性能。实验 3 进一步集成 CMUNeXt 特征融合结构, $I_{\text{mAP}@0.5}$ 提升至 0.322, $I_{\text{mAP}@0.5:0.95}$ 提升至 0.186。这证实了 CMUNeXt 结构对跨层级特征的整合作用, 增强了模型对多尺度目标的适应能力和鲁棒性。实验 4 引入 MogaNet 后, $I_{\text{mAP}@0.5}$ 为 0.331, $I_{\text{mAP}@0.5:0.95}$ 为 0.191, 两项指标同步达到最优, 表明该模块通过轻量化的特征增强策略, 在可控计算开销的前提下, 进一步提升了特征的细粒度表征能力。

表 2 VisDrone2019 消融实验结果

Tab.2 ablation experiments on VisDrone2019						
实验组别	CA	Triplet	CMUNeXt	MogaNet	$I_{\text{mAP}@0.5}$	$I_{\text{mAP}@0.5:0.95}$
1					0.274	0.155
2	√	√			0.291	0.167
3	√	√	√		0.322	0.186
4	√	√	√	√	0.331	0.191

注: “√”表示对应模块在模型中被启用。

在 NWPU VHR-10 数据集上的表现同样印证了各模块的有效性, 如表 3 所示。

表 3 NWPU VHR-10 消融实验结果

Tab.3 ablation experiments on NWPU VHR-10						
实验组别	CA	Triplet	CMUNeXt	MogaNet	$I_{\text{mAP}@0.5}$	$I_{\text{mAP}@0.5:0.95}$
1					0.658	0.375
2	√	√			0.692	0.391
3	√	√	√		0.756	0.416
4	√	√	√	√	0.808	0.466

注: “√”表示对应模块在模型中被启用。

基础配置的 $I_{\text{mAP}@0.5}$ 和 $I_{\text{mAP}@0.5:0.95}$ 分别为 0.658 和 0.375。引入两个注意力机制后, $I_{\text{mAP}@0.5}$ 提升至 0.692, $I_{\text{mAP}@0.5:0.95}$ 提升至 0.391, 验证了通道注意力在遥感高分辨率场景下的特征增强效应, 显著提升了模型对复杂背景和多尺度目标的监测精度。后续引入 CMUNeXt 特征融合结构与 MogaNet 双重融合机制后, 模型性能持续优化并最终达到同步优化, 充分印证了各模块的协同增益效应与跨场景适配能力。

2.3 对比实验

为验证所提改进算法的有效性和性能优势, 保持配置环境与初始训练参数一致, 在

VisDrone2019数据集上,将本文算法与具有代表性的主流目标检测算法开展对比实验,实验结果见表4。本文算法在精确率、召回率等评价指标上均优于YOLO系列主流模型,性能稳定,鲁棒性优势明显。具体来说,改进算法的检测精度 I_{Pre} 较YOLOv8n提升0.027,较YOLOv5n提升0.040,体现出对误检样本的强过滤能力,有效降低低置信度误判问题; I_{Rec} 优于所有对比模型,反映了在精度优先策略下对真实目标的均衡捕获能力。在中等IoU检测能力($I_{mAP@0.5}$ 为0.321)上,本文算法较YOLOv5n提升0.034,较YOLOv8n提升0.033,优势源于对边界框定位精度的优化,使预测框与真实框的重合度匹配更精准。本文算法在 $I_{mAP@0.5:0.95}$ 指标上同样领先,表明其对多尺度、复杂场景目标的综合检测能力兼具精度与泛化性,验证了算法在目标检测任务中通过结构创新实现的系统性提升,为实际应用场景提供了更可靠的解决方案。

尽管本文算法的计算复杂度较YOLOv12模型有所增加,导致推理速度(494.8 FPS)低于YOLOv6

等轻量化模型,但该速度仍远超常规无人机视频流实时回传(通常为30 FPS或60 FPS)的处理需求,完全满足机载嵌入式平台的实时推理阈值。在安防监控、灾害救援等对漏检率“零容忍”的关键应用场景中,牺牲部分冗余的推理速度换取检测精度($I_{mAP@0.5}$)大幅提升,具有充分的合理性和使用价值,这表明该算法在保证工程实用性的前提下,实现了精度与速度的有效权衡。

在NWPU VHR-10数据集上的对比试验结果见表5。本文算法的浮点运算次数为 14.6×10^9 ,推理速度为488.9 FPS,相对较低。具体来看,其 I_{Pre} 达到0.860,较YOLOv8模型提高0.014,且在 I_{Rec} 上显著达到0.739,比YOLOv6提高0.009。在 $I_{mAP@0.5}$ 指标上,本文算法为0.808,较YOLOv8和YOLOv5分别提升0.039和0.052; $I_{mAP@0.5:0.95}$ 为0.466,较YOLOv6提高0.005,充分证明其在高分辨率遥感图像小目标检测中的鲁棒性与泛化能力。综上,本文算法在不同分辨率、不同场景和不同目标密度下均表现出优越性能,验证了其广泛的适应性和实用价值。

表4 VisDrone2019数据集性能比较
Tab.4 performance comparison of VisDrone2019 dataset

方法	I_{Pre}	I_{Rec}	$I_{mAP@0.5}$	$I_{mAP@0.5:0.95}$	浮点运算次数/ 10^9	推理速度/FPS
YOLOv5 ^[19]	0.393	0.297	0.287	0.163	7.1	1 471.0
YOLOv6 ^[20]	0.348	0.274	0.267	0.154	11.8	1 714.4
YOLOv8 ^[21]	0.406	0.307	0.298	0.170	8.1	1 549.8
YOLOv12 ^[22]	0.378	0.283	0.274	0.155	5.8	674.9
YOLOv13 ^[23]	0.391	0.292	0.280	0.159	5.0	1 205.1
FFCA ^[24]	0.395	0.295	0.283	0.161	5.8	668.8
APW-YOLOv8 ^[25]	0.419	0.320	0.319	0.188	12.4	1 268.6
YOLOv8-Ghost ^[26]	0.419	0.284	0.266	0.146	5.0	1 255.3
本文算法	0.433	0.335	0.331	0.191	14.6	494.8

表5 NWPU-VHR-10数据集性能比较
Tab.5 performance comparison of NWPU-VHR-10 dataset

方法	I_{Pre}	I_{Rec}	$I_{mAP@0.5}$	$I_{mAP@0.5:0.95}$	浮点运算次数/ 10^9	推理速度/FPS
YOLOv5 ^[19]	0.771	0.727	0.756	0.430	7.1	1 475.0
YOLOv6 ^[20]	0.825	0.730	0.795	0.461	11.8	1 691.2
YOLOv8 ^[21]	0.846	0.704	0.769	0.441	8.1	1 589.6
YOLOv12 ^[22]	0.821	0.600	0.658	0.375	5.8	667.5
YOLOv13 ^[23]	0.625	0.662	0.357	0.366	5.0	470.8
FFCA ^[24]	0.816	0.607	0.658	0.376	5.8	659.0
APW-YOLOv8 ^[25]	0.850	0.728	0.785	0.456	12.4	1 268.6
YOLOv8-Ghost ^[26]	0.704	0.689	0.699	0.381	5.0	1 231.4
本文算法	0.860	0.739	0.808	0.466	14.6	488.9

2.4 实验结果可视化

YOLOv12模型与本文改进模型在VisDrone2019

数据集和NWPU VHR-10数据集上的可视化对比结果见图6和图7。

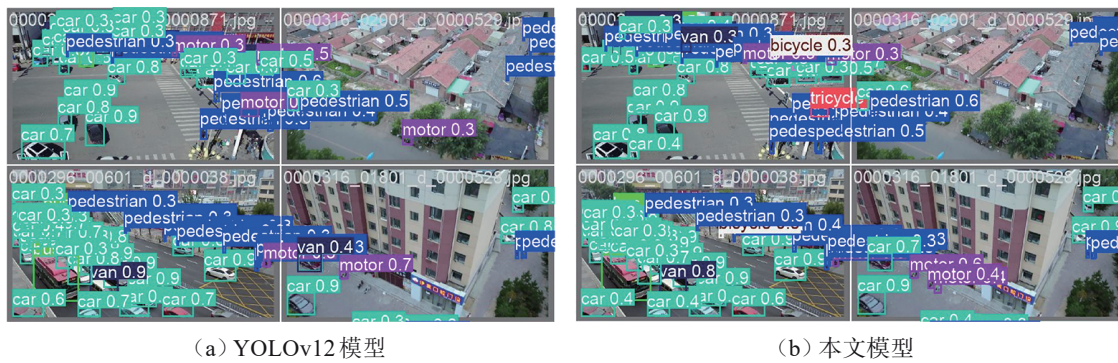


图 6 VisDrone2019 数据集实验结果

Fig.6 experimental results of VisDrone2019 data set

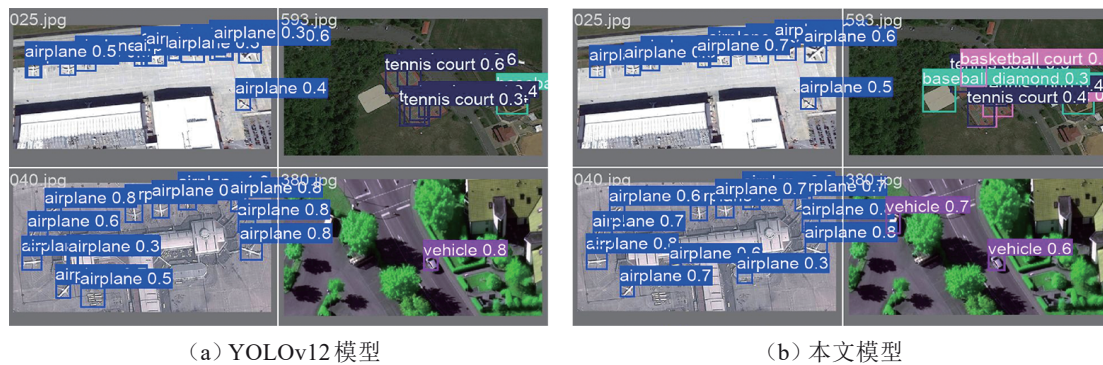


图 7 NWPU VHR-10 数据集实验结果

Fig.7 experimental results of NWPU VHR-10 data set

由图 6 可见,在密集遮挡场景下, YOLOv12 模型在处理停车场密集车辆和拥挤人群时,容易出现漏检和边界框重叠现象。而改进模型得益于 Triplet Attention 机制对空间与通道特征的联合增强,能够有效区分重叠目标的边界,显著降低了漏检率。

由图 7 可见,在微小目标场景下,由于 MogaNet 双域聚合机制有效抑制了树木、阴影等

复杂背景的干扰,配合 CMUNeXt 对多尺度特征的深度融合,改进模型能够精准捕捉到远距离、像素占比极低的行人和车辆。本文方法显著提升了模型在密集遮挡与微小尺度目标下的鲁棒性。

2.5 混淆矩阵分析

YOLOv12 模型与改进后 YOLOv12 模型在 NWPU VHR-10 数据集类上的识别性能见图 8。

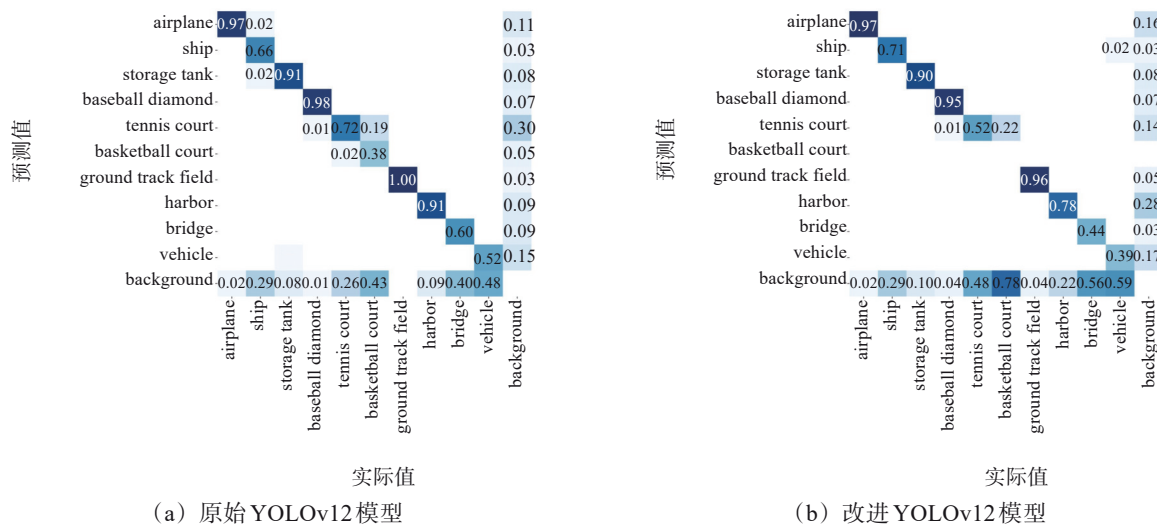


图 8 NWPU VHR-10 数据集类别识别性能

Fig.8 classification performance of the NWPU VHR-10 dataset

由图8(a)可见, YOLOv12模型对部分类别(如 storage tank、tennis court、airplane)的识别存在混淆与误检, 尤其在 storage tank 与 roundabout、tennis court 与 basketball court 等易混淆类别间分类错误率较高。图8(b)中, 改进后 YOLOv12模型的混淆矩阵主对角线区域颜色更深, 非对角线区域颜色更浅, 表明各类别的识别精度整体提升, 类别间混淆显著减少。尤其在上述易混类别上, 改进模型有效降低了错误率。这说明改进 YOLOv12模型在特征表达能力、类间差异性捕捉, 以及背景干扰抑制方面均有优化, 对 NWPU VHR-10数据集的分类准确率更高, 鲁棒性更强。

3 结论

针对无人机遥感图像中小目标检测精度低、背景干扰强等问题, 提出融合多核特征提取与三重注意力机制的改进型目标检测算法。在 VisDrone2019 和 NWPU VHR-10 数据集上的实验结果表明, 该算法的 I_{Pre} 、 I_{Rec} 、 $I_{mAP@0.5}$ 和 $I_{mAP@0.5:0.95}$ 等核心指标, 优于主流 YOLO 系列模型。消融实验验证了各模块对整体性能提升的独立贡献, 可视化分析进一步说明该算法在小目标、密集目标及复杂背景场景下的检测优势。

综合来看, 本文所提出的改进策略在保持较高推理效率(满足实时性要求)的同时, 实现了检测精度与鲁棒性的双重提升。针对 MogaNet 和 Triplet Attention 模块引入的额外计算开销, 未来的研究工作将集中在以下两个方向: 模型轻量化部署, 研究模型剪枝和量化技术, 在尽可能保持高精度的前提下, 进一步降低模型的参数量与运算功耗, 使其更适配算力受限的微型无人机边缘计算终端; 跨模态数据融合, 考虑单一可见光模态在夜间或恶劣天气下的局限性, 计划引入红外热成像数据, 探索跨模态特征融合检测算法, 以进一步拓展无人机全天候智能感知的应用边界。

参考文献(References):

- [1] GUEBSI R, MAMI S, CHOKMANI K. Drones in precision agriculture: a comprehensive review of applications, technologies, and challenges[J]. Drones, 2024, 8(11): 686.
- [2] 吴一全, 童康. 基于深度学习的无人机航拍图像小目标检测研究进展[J]. 航空学报, 2025, 46(3): 174-200.
WU Yiquan, TONG Kang. Research advances on deep learning-based small object detection in UAV aerial images[J]. Acta Aeronautica et Astronautica Sinica, 2025, 46(3): 174-200.
- [3] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision. October 22-29, 2017, Venice, Italy. IEEE, 2017: 2999-3007.
- [4] LEE J H, GWON G H, KIM I H, et al. A motion deblurring network for enhancing UAV image quality in bridge inspection[J]. Drones, 2023, 7(11): 657.
- [5] GAO T, NIU Q Q, ZHANG J, et al. Global to local: a scale-aware network for remote sensing object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5615614.
- [6] 薛光辉, 闫朝阳, 吴冕. PCSED-YOLO: 复杂环境下跨尺度多目标穿戴检测算法研究[J]. 计算机工程与应用, 2026, 62(5): 88-105.
XUE Guanghui, YAN Zhaoyang, WU Mian, et al. PCSED-YOLO: study on cross-scale multi-object wearable detection algorithm in complex environments[J]. Computer Engineering and Applications, 2026, 62(5): 88-105.
- [7] WANG X L, CHEN H. HPS-DETR: enhancing small object detection with lightweight feature extraction and transformer integration[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 5937420.
- [8] WU Q G, LI Y, YIN J R, et al. LGC-YOLO: local-global feature extraction and coordination network with contextual interaction for remote sensing object detection[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2025, 18: 15376-15393.
- [9] 周晓琴, 王长海, 王维. 结合上下文感知的无人机航拍目标检测模型[J]. 测绘科学, 2026, 51(1): 88-97.
ZHOU Xiaolin, WANG Changhai, WANG Wei. A UAV aerial object detection model integrating contextual perception[J]. Science of Surveying and Mapping, 2026, 51(1): 88-97.
- [10] LI Z, WANG Y C, ZHANG Y X, et al. Context feature integration and balanced sampling strategy for small weak object detection in remote sensing imagery[J]. IEEE Geoscience and Remote Sensing Letters, 2024, 21: 6009105.
- [11] 李璐, 陈清江. 基于超分辨率和YOLO的红外小目标检测[J]. 无线电工程, 2025, 55(8): 1571-1579.
LI Lu, CHEN Qingjiang. Infrared small target detection based on super-resolution and YOLO[J]. Radio Engineering, 2025, 55(8): 1571-1579.
- [12] ZHANG Y, LEI X, HU Q, et al. Learning cross-modality high-resolution representation for thermal small-object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5404815.
- [13] GAO Y X, WANG Y C, ZHANG Y X, et al. Feature super-resolution fusion with cross-scale distillation for small-object detection in optical remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2024, 21: 6008105.
- [14] 刘浩南, 周刚, 刘江涛, 等. 融合超分辨率重建和动态卷积的棉田虫害检测[J/OL]. 计算机工程, 2025: 1-17[2025-09-26]. <https://doi.org/10.19678/j.issn.1000-3428.0252566>.

- LIU Haonan, ZHOU Gang, LIU Jiangtao, et al. Cotton field pest detection based on super-resolution reconstruction and dynamic convolution[J/OL]. Computer Engineering, 2025: 1-17[2025-11-10]. <https://doi.org/10.19678/j.issn.1000-3428.0252566>.
- [15] 宋润泽,崔洪博,赵子龙,等.基于扩散超分重建的小目标检测方法[J/OL].系统工程与电子技术,2025: 1-9[2025-11-13]. <https://link.cnki.net/urlid/11.2422.TN.20251110.1048.006>.
- SONG Runze, CUI Hongbo, ZHAO Zilong, et al. Small target detection method based on diffusion super-resolution reconstruction[J/OL]. Systems Engineering and Electronics, 2025: 1-9[2025-11-13]. <https://link.cnki.net/urlid/11.2422.TN.20251110.1048.006>.
- [16] GUO M H, LU C Z, LIU Z N, et al. Visual attention network[J]. Computational Visual Media, 2023, 9(4): 733-752.
- [17] 彭继慎,马龙泽,孙梦宇,等.多尺度特征融合增强检测模型 MFFE-YOLO[J]. 辽宁工程技术大学学报(自然科学版),2024,43(5):625-632.
- PENG Jishen, MA Longze, SUN Mengyu, et al. A multi-scale feature fusion enhanced detection model MFFE-YOLO[J]. Journal of Liaoning Technical University(Natural Science), 2024, 43(5): 625-632.
- [18] LIU W J, WU G Q, WANG H, et al. Dense skip-attention for convolutional networks[J]. Scientific Reports, 2025, 15: 22710.
- [19] 张在岩,宋伟东,邬嘉晨.改进 YOLOv5 的复杂场景下水泥路面病害检测[J]. 辽宁工程技术大学学报(自然科学版),2025,44(1): 102-112.
- ZHANG Zaiyan, SONG Weidong, WU Jiachen. Disease detection of cement pavement based on improved YOLOv5 in complex scenarios[J]. Journal of Liaoning Technical University (Natural Science),2025,44(1):102-112.
- [20] 陈孟元,许瑞珩,杨苏朋,等.动态遮挡场景下基于改进 YOLOv6s 网络的 SLAM 算法[J]. 中国惯性技术学报,2025,33(8):802-811.
- CHEN Mengyuan, XU Ruiheng, YANG Supeng, et al. SLAM algorithm based on improved YOLOv6s network in dynamic occlusion scenarios[J]. Journal of Chinese Inertial Technology, 2025,33(8):802-811.
- [21] VARGHESE R, M S. YOLOv8: a novel object detection algorithm with enhanced performance and robustness[C]//2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems. April 18-19, 2024, Chennai, India. IEEE, 2024: 1-6.
- [22] CHEN Q Z. Traffic object detection using YOLOv12[J]. Open Access Library Journal, 2025, 12: e13991.
- [23] 段虹羽,孙志阳.基于 YOLOv13 的无人机类别检测方法[J]. 计算机时代,2025(12):14-18.
- DUAN Hongyu, SUN Zhiyang. A YOLOv13-based method for drone type detection[J]. Computer Era,2025(12):14-18.
- [24] RABBI J, RAY N, SCHUBERT M, et al. Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network[J]. Remote Sensing, 2020, 12(9): 1432.
- [25] 王宗阳,黄莉,江都.基于 APW-YOLOv8 的无人机高空图像小目标检测[J]. 计算机系统应用, 2025, 34(10): 195-205.
- WANG Zongyang, HUANG Li, JIANG Du. APW-YOLOv8-based detection of small targets in high-altitude UAV image[J]. Computer Systems & Applications, 2025, 34(10): 195-205.
- [26] HUANGFU Z M, LI S Q, YAN L H. Ghost-YOLO v8:an attention-guided enhanced small target detection algorithm for floating litter on water surfaces[J]. Computers, Materials & Continua, 2024, 80(3): 3713-3731.