液晶与显示
Chinese Journal of Liquid Crystals and Displays

文章编号:1007-2780(XXXX)XX-0001-18

LCDA-Net:面向航拍小目标检测的低参数协同方向感知网络

王艳玲,牛　娜,刘亚秋*,郑涵文
(东北林业大学 计算机与人工智能学院,黑龙江省 哈尔滨市 150040)

摘要:航拍图像中小目标尺度小、分布密集且背景复杂等问题,给目标检测带来了显著挑战。针对单阶段检测框架在航拍小目标场景下特征易衰减、上下文与方向信息建模不足以及检测头适应性有限的问题,提出了一种低参数协同方向感知网络模型(LCDA-Net,Low-Parameter Collaborative Direction-Aware Network)。该方法以YOLOv11为基线,通过低参数协同双向路径网络增强多尺度特征的语义与细节协同,缓解小目标信息在下采样过程中的丢失;引入上下文感知感受野模块,从方向上下文、动态感受野、通道重标定与细节保留等方面提升航拍小目标的结构与边缘表征;并构建小目标感知动态检测头,在层级、空间与任务维度上实现自适应特征增强与任务解耦。实验在VisDrone2019和AI-TODv2等公开小目标数据集上进行。实验结果表明,LCDA-Net在VisDrone2019数据集上相较于YOLOv11-S将mAP50由39.3%提升至48.1%,提高8.8个百分点,同时参数量减少47.8%。在AI-TODv2数据集上,AP50和AP50:95分别达到64.0%和29.4%,其中针对极小(APvt)、小(APt)和中小(APs)目标的检测精度分别提高8.1、9.4和8.0个百分点。上述结果表明,LCDA-Net能够有效增强不同尺度目标的检测能力,并在检测精度与模型复杂度之间取得良好平衡。

关　键　词:小目标检测;航拍图像;低参数检测网络;多尺度特征融合;动态检测头
中图分类号:TP391.41　　**文献标识码**:A　　**doi**:10.3724/CJLCD.2026-0224　　**CSTR**:32172.14.CJLCD.2026-0224

LCDA-Net:low-parameter collaborative direction-aware network for aerial small-object detection

WANG Yanling, NIU Na, LIU Yaqiu*, ZHENG Hanwen

(*School of Computer Science and Artificial Intelligence Northeast Forestry University, Harbin, China*)

Abstract: Small objects in remote sensing images are characterized by their limited scale, dense distribution, and complex backgrounds, posing significant challenges to object detection. To address the problems of feature attenuation, insufficient contextual and directional information modeling, and limited detection-head adaptability in single-stage detection frameworks for remote sensing small-object scenarios, a Low-Parameter Collaborative Directional Attention Network(LCDA-Net) is proposed. Built upon YOLOv11, the proposed method employs a low-parameter collaborative bidirectional path network to enhance the semantic and detail collaboration of multi-scale features, thereby mitigating the loss of small-object information during downsampling. A context-aware receptive field module is introduced to improve

收稿日期:2026-10-12;**修订日期**:2026-11-04.
基金项目:黑龙江省自然科学基金项目(No.PL2025F002)
Supported by Heilongjiang Provincial Natural Science Foundation of China (No. PL2025F002)
*通信联系人,E-mail:yaqiuLiu@nefu.edu.cn

the structural and edge representations of remote sensing small objects by incorporating directional contextual modeling, dynamic receptive fields, channel recalibration, and detail preservation. In addition, a small-object-aware dynamic detection head is developed to achieve adaptive feature enhancement and task decoupling across hierarchical, spatial, and task dimensions. Experiments are conducted on publicly available small-object datasets, including VisDrone2019 and AI-TODv2. Experimental results show that, compared with YOLOv11-S, LCDA-Net improves the mAP50 from 39.3% to 48.1% on the VisDrone2019 dataset, representing an increase of 8.8 percentage points, while reducing the number of parameters by 47.8%. On the AI-TODv2 dataset, LCDA-Net achieves AP50 and AP50:95 of 64.0% and 29.4%, respectively, with the detection accuracy for very tiny (APvt), tiny (APt), and small (APs) objects improved by 8.1, 9.4, and 8.0 percentage points, respectively. These results demonstrate that LCDA-Net effectively enhances the detection of objects at different scales while achieving a favorable balance between detection accuracy and model complexity.

Key words: small-object detection; remote sensing images; low-parameter detection network; multi-scale feature fusion; dynamic detection head

1 引　　言

小目标检测是目标检测领域中最具挑战性的研究方向之一[1],广泛应用于无人机感知[2][3]、远距离监控[4]及复杂场景理解[5]等任务。由于小目标在图像中尺度极小、语义信息弱且易受背景干扰,其判别特征在网络前向传播过程中容易被下采样和特征融合操作所淹没,导致检测精度受限[6][7]。针对复杂背景下小目标特征不明显的问题,已有研究通过引入难点注意力感知机制,增强难以区分的目标特征响应,以改善红外小目标检测性能[8]。如何在控制模型参数规模与部署开销的前提下,增强网络对小目标的感知与表达能力,仍是当前研究的关键问题。

本文选择 YOLOv11[9]作为基线模型,主要考虑其在检测精度、模型复杂度与推理效率之间具有较好的平衡。相较于部分结构较为复杂的检测模型,YOLOv11具有较成熟的网络结构和较完善的工具链,其多尺度特征提取与融合机制也为航拍小目标检测的轻量化改进提供了良好的基础。近年来,YOLO系列模型也逐渐应用于无人机航拍小目标检测。张志豪等针对无人机航拍场景对YOLOv11进行改进,进一步探索了基于该系列网络的小目标检测方法[10]。吕学涵等进一步针对无人机航拍图像中的小目标检测任务,对YOLOv11s网络进行改进,为基于YOLOv11系列开展无人机小目标检测研究提供了相关参考[11]。

围绕航拍小目标检测中的多尺度特征融合、细节增强及检测头优化,已有研究提出了多种改进方法。在基于Transformer的航拍目标检测研究中,李大湘等提出位置敏感Transformer模型,通过增强目标空间位置信息的表达,提高航拍图像中的目标检测能力[12]。Zhang等人在FFCA-YOLO[13]中提出SCAM模块通过建模跨空间与通道的全局上下文关系,抑制复杂背景干扰并增强航拍图像中小目标的判别特征。Yao等提出FCM[14]融合浅层空间与深层语义信息,并结合MKP扩展多尺度感受野。Hui[15]等人针对遥感无人机场景小目标易模糊、易漏检的问题,提出了DRSR用于增强低分辨率小目标细节表征,并设计了MGDA以自适应调整锚框回归尺度。Liu[16]等人提出了轻量化骨干网络(RepViT-TD backbone)和轻量化检测头(detection head with deformable convolution, DCN);其中骨干网络通过调整计算资源分配到前两阶段提高对小目标的细粒度特征提取能力;检测头结合普通卷积和DCN,增强特征适应性和动态感受野,同时保持低复杂度。此外,白晨帅等通过融合残差特征与几何感知信息,增强遥感图像中目标的结构特征表达,为复杂遥感场景下的目标检测提供了新的思路[17]。Wu[18]等人通过引入MBFA-CSP增强多尺度上下文表征,并设计MLFF以有效融合低层位置信息与高层语义信息,同时结合Shape-

IoU提升航拍图像中细长与复杂目标的定位精度。此外，在高效视觉计算架构方面，光子神经网络[19]、光学神经形态图像处理[20]及可重构集成光子计算芯片[21]等研究也为高吞吐量视觉信息处理提供了新的技术思路。不过，这些工作主要聚焦于硬件层面的高效计算架构，与本文面向航拍小目标检测的网络结构设计属于不同研究维度，因此本文仍以轻量化网络结构改进为主，相关高效计算架构可作为后续部署优化的参考方向。

综上，现有研究从多尺度特征融合、上下文建模、细节增强及检测头优化等方面改善了小目标检测性能，但在轻量化约束下，如何进一步实现多尺度特征的有效协同、方向上下文信息的充分利用，以及检测头对不同尺度目标的自适应建模，仍有进一步研究空间。具体而言，多尺度特征交互仍需兼顾高分辨率细节保留与高层语义表达；方向结构与长程上下文信息的联合建模有待加强；检测头也需要根据不同层级的特征及目标分布灵活调整特征增强策略。上述问题共同构成了本文的研究动机。

针对上述问题，本文提出一种低参数协同方向感知网络LCDA-Net，以YOLOv11为基线框架，从多尺度特征传递、方向上下文建模和检测头自适应三个方面进行协同改进。首先，设计低参数协同双向路径网络LCPAN，通过自顶向下与自底向上的双向特征交互，增强高分辨率细节与高层语义之间的协同表达。其次，引入上下文感知感受野模块CARM，通过跨轴上下文建模、动态感受野调节、通道重标定及细节保留等机制，增强复杂背景下目标结构与边缘特征的表征能力。最后，构建小目标感知动态检测头SO-DyHead，通过多尺度特征融合以及层级、空间与任务维度的自适应建模，提高检测头对不同尺度目标的特征感知与分类定位能力。三个模块分别作用于特征传递、上下文增强和检测决策环节，形成互补的协同设计，在提升小目标检测能力的同时兼顾模型复杂度。基于上述设计，本文的主要贡献如下：

1）提出了一种面向航拍图像小目标检测的低参数协同方向感知网络LCDA-Net，通过结构协同设计在保持较低参数规模的同时提升了检测性能。

2）设计了低参数协同双向路径网络（Low-Parameter Collaborative PAN，LCPAN），通过以P4/16作为主干终止尺度并在检测端重构高层特征，增强了高分辨率细节保留与多尺度语义交互能力。

3）提出了上下文感知感受野模块（Context-Aware Receptive-field Module，CARM），通过跨轴上下文建模、动态感受野调节、通道重标定和细节增强，提高了复杂背景下的小目标表征能力。

4）构建了小目标感知动态检测头（Small-Object-Aware Dynamic Head，SO-DyHead），从层级、空间和任务三个维度对特征进行自适应增强，从而提升了分类与回归过程对小目标的适应性。

2 LCDA-Net算法

2.1 概述

LCDA-Net是一种面向复杂场景小目标检测的低参数协同方向感知网络，整体由低参数协同双向路径特征网络、上下文感知感受野模块以及小目标感知动态检测头三部分构成，如图1所示。网络在保持高效推理的前提下，通过自顶向下与自底向上的特征协同融合，充分利用不同尺度语义与细节信息，缓解小目标在深层特征中易丢失的问题。在特征增强阶段，引入方向感知的上下文聚合机制，从水平与垂直两个维度建模长程依赖，并结合动态感受野与细节保留策略，强化目标结构与边缘表达。在检测阶段，采用小目标感知的动态检测头，根据尺度与空间特性自适应调整特征增强与注意力分配，实现分类与回归任务的有效解耦。整体设计在精度、计算复杂度与模型规模之间取得良好平衡。

2.2 低参数协同双向路径网络

本文对YOLOv11网络整体设计思想的特征提取与多尺度建模方式进行了有针对性地重构，提出了一种低参数协同双向路径网络的结构。其核心目标并非单纯追求更深的语义层级，而是围绕“小目标特征如何在网络中被有效保留与利用”这一问题，对Backbone与Head的职责边界进行了重新划分。在Backbone阶段，LCPAN通过削减最深层的下采样路径，使特征提取过程不再

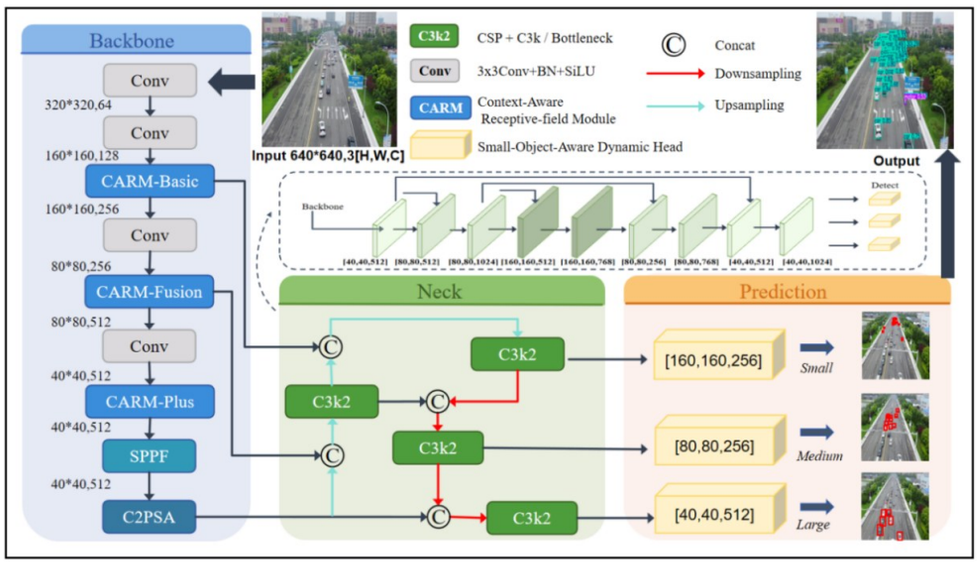


图1 LCDA-Net总体网络结构以及Neck中的多尺度特征融合结构

Fig. 1 Overall architecture of LCDA-Net and the multi-scale feature fusion structure in the neck

延伸至传统的P5/32级别，而是以P4/16作为主干网络的终止尺度。这一调整显著减少了小目标在连续stride=2下采样过程中发生的空间信息塌缩，使得目标在中高分辨率特征图上仍然保有足够的像素表达与结构完整性，从结构层面提升了对密集分布与尺度较小目标的建模友好度。同时，Backbone深度与通道规模的整体收缩也带来了参数量与模型大小的同步下降，使模型在训练稳定性与部署效率方面具备更现实的工程优势。

在语义建模策略上，LCPAN并非简单削弱高层语义表达，而是通过将SPPF与C2PSA等上下文聚合与注意力模块前移至512通道的中尺度特征层，使全局感受野与长程依赖关系在更高空间分辨率条件下建立。相较于将此类模块置于极低分辨率的深层特征上，这种设计能够在引入全局语义的同时保留更丰富的局部结构信息，使注意力机制既服务于目标级语义建模，又不过度牺牲定位精度。由此，中尺度特征不仅承担了空间细节表达的职责，也在一定程度上兼具了高层语义抽象能力，形成了更加均衡的特征表征。

在Head阶段，仍然保持YOLOv11原生的多尺度检测范式，通过上采样、跨层特征拼接以及二次下采样，构建出三层用于检测的特征图。不同之处在于，最高层检测特征并非直接来自Backbone的原生输出，而是通过Head内部的特征重构过程形成的“虚拟高层语义”。这种设计在结构兼容性上保持了与原始检测头的一致性，使训练流程与损失计算无需额外修改，同时也为高层语义的灵活建模提供了空间。总体而言，LCPAN以高分辨率细节保留为基础，以中尺度语义增强为核心，并通过检测端的多尺度重构补偿缺失的高层表征，从而在小目标检测精度、参数规模与尺度表达能力之间实现更合理的平衡。

2.3 小目标感知动态检测头

为提升复杂场景中小目标的检测性能，本文在动态检测头框架下提出SO-DyHead（如图2所示）。首先利用EfficientMultiScaleFusion对不同尺度特征进行对齐与加权融合，以增强跨尺度信息交互；随后引入LightweightSpatialAttention强化关键空间区域响应；同时结合AdaptiveSmall-ObjectEnhance对高频细节信息进行自适应增强，以提升小目标表征能力，最后分别生成分类与回归任务的通道权重，实现任务相关的特征重标定。整体上，该方法围绕层级、空间与任务三个维度建模动态注意力机制，在保证计算效率的前提下增强小目标特征表达能力，其内部模块如图3所示。

在层级维度（πL）上，对来自不同尺度的特征进行空间对齐，并引入可学习的尺度权重以建模各层级特征的重要性。设第i个尺度特征为F_i，

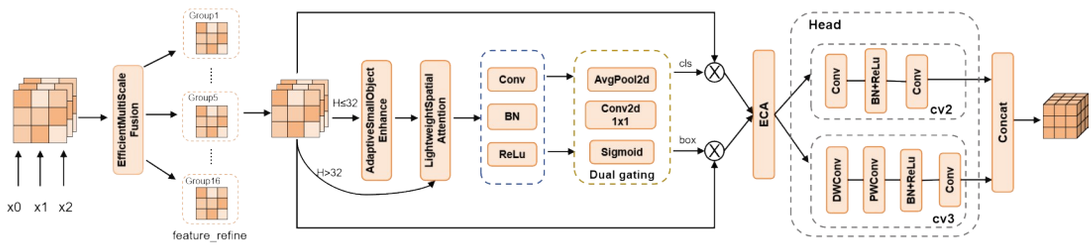


图2 SO-DyHead的整体结构图

Fig. 2 Overall architecture of SO-DyHead

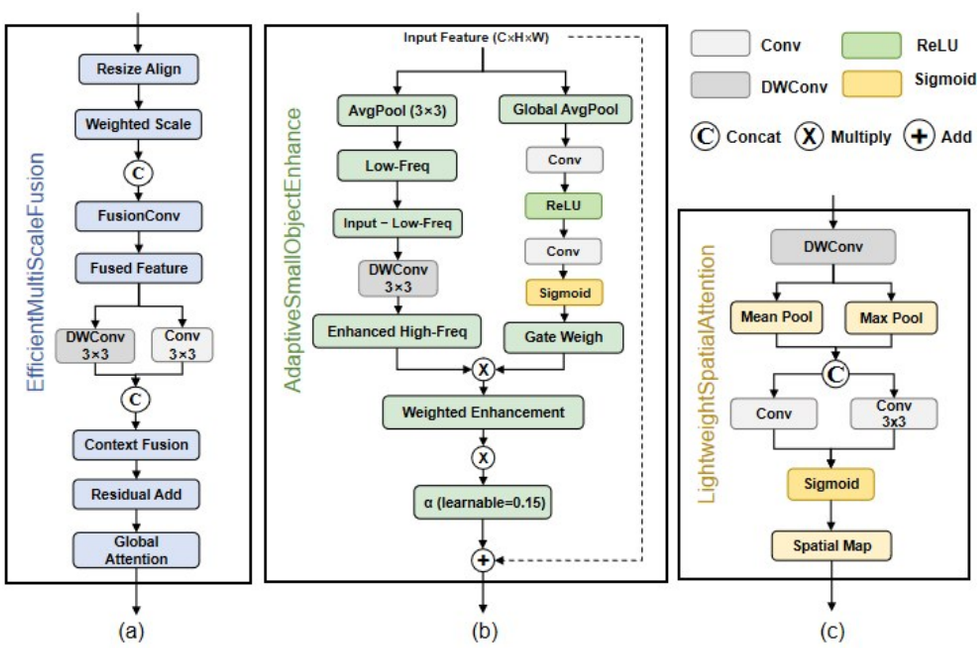


图3 SO-DyHead内置的模块。(a)为EfficientMultiScaleFusion；(b)为AdaptiveSmallObjectEnhance；(c)为LightweightSpatialAttention。

Fig. 3 Built-in modules of SO-DyHead. (a) EfficientMultiScaleFusion; (b) AdaptiveSmallObjectEnhance; (c) LightweightSpatialAttention.

对齐后的融合特征表示为：

$$F_L = \phi\left(\mathrm{Concat}\left(w_i \cdot R(F_i)\right)\right), \tag{1}$$

其中R(·)表示尺度对齐操作，w_i为可学习权重，Φ(·)为轻量级融合变换。该设计使模型能够自适应选择更有利于当前目标尺度分布的特征层级。

在空间维度(πS)上，首先通过空间注意力对特征显著区域进行重标定：

$$F_S = F_L \odot A_S, \tag{2}$$

其中A_S为空间注意力权重图。进一步考虑到小目标在深层特征中易丢失局部细节，本文引入自适应高频增强机制，通过残差形式强化局部结构信息：

$$F_{SO} = F_S + \alpha \cdot G(F_S) \odot \sigma\left(g(F_S)\right), \tag{3}$$

其中$G(\cdot)$表示高频特征提取操作，$\sigma(\cdot)$为门控函数，α为可学习增强系数。该模块仅作用于高分辨率特征层，以避免噪声放大和额外计算开销。

在任务维度(πC)上，为缓解分类与回归任务之间的特征竞争，SO-DyHead采用任务解耦的通道注意力机制。基于共享特征F_{SO}，分别生成分类与回归的通道权重：

$$F^{\mathrm{cls}} = F_{SO} \odot A_C^{\mathrm{cls}}, F^{\mathrm{reg}} = F_{SO} \odot A_C^{\mathrm{reg}}. \tag{4}$$

该设计在不引入重型分支的情况下，实现了任务条件化的特征重标定。

综上，SO-DyHead在不改变动态检测头基本范式的前提下，通过πL－πS－πC的协同建模，

实现了对小目标更加友好的特征表示，并在效率与精度之间取得良好平衡。

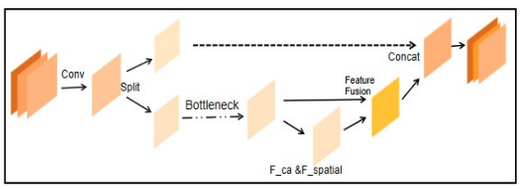


图4 CARMBasic特征图变化流程图

Fig. 4 Feature transformation flowchart of CARM-Basic

2.4 上下文感知感受野模块

针对航拍图像中目标尺度差异大、分布密集且背景复杂的问题，现有特征增强方法对方向上下文与局部细节的协同利用仍存在不足。为此，本文提出上下文感知感受野模块（Context-Aware Receptive-field Module，CARM），通过多种互补的上下文建模机制增强目标特征的判别能力。具体而言，跨轴注意力单元分别聚合水平方向与垂直方向的上下文信息，建立长程空间依赖，以增强目标结构及其邻域关系的表达；动态感受野组件通过并行多尺度卷积分支与可学习权重，自适应融合不同感受野下的局部特征，以适应不同尺度目标的结构差异；通道注意力机制对特征通道进行自适应重标定，强化与目标识别相关的语义信息；细节保留单元则通过高频信息提取增强边缘与纹理响应，减少局部细节信息的弱化。为控制额外计算开销，CARM结合轻量卷积映射、深度卷积及可学习权重等操作，实现不同上下文信息的自适应融合，从而兼顾特征表达能力与模型复杂度。

需要说明的是，跨轴注意力主要对水平与垂直方向的上下文关系进行显式建模，能够增强复杂空间布局下的目标结构表征，但并不等同于显式的任意角度旋转等变机制。因此，对于大角度旋转目标的适应能力，仍需通过专门的旋转目标实验进一步验证。

在跨轴注意力中，融合后的跨轴上下文特征先经过1×1卷积进行通道映射，再通过Sigmoid激活生成最终注意力图，如图5所示。也就是把“垂直上下文”和“水平上下文”进行联合建模，生成一个同时感知两个轴向信息的空间-通道响应图，公式如下：

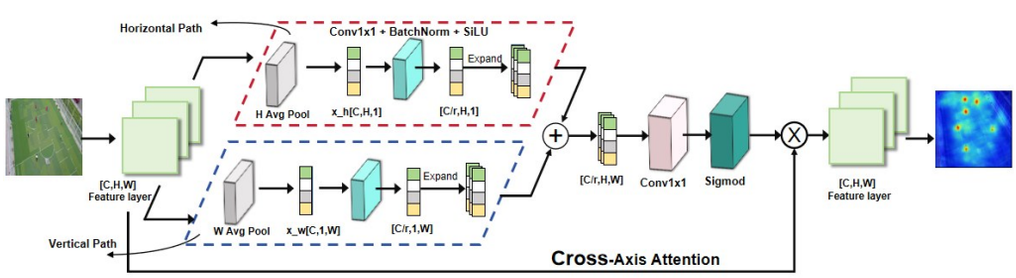


图5 跨轴注意力

Fig. 5 Cross-Axis attention

$$A=\sigma\left(\mathrm{Conv}_{1\times1}\left(\hat{\mathbf{X}}_h+\hat{\mathbf{X}}_w\right)\right). \tag{5}$$

对于动态分支权重是先对输入特征做全局平均池化，再用1×1卷积生成各分支的权重，最后通过Softmax归一化。公式如下：

$$\alpha=\mathrm{Softmax}\left(\mathrm{Conv}_{1\times1}\left(\mathrm{GAP}(X)\right)\right). \tag{6}$$

若有K个分支，则：

$$\alpha=[\alpha_1,\alpha_2,\cdots,\alpha_K],\ \sum_{k=1}^{K}\alpha_k=1. \tag{7}$$

目的是让网络根据当前输入内容自适应决定“当前更应该相信哪个感受野分支”。即它不是固定把多尺度特征平均融合，而是让权重由数据驱动地学习出来，因此能够在目标尺寸变化较大的场景中更灵活地调整特征表达。

为适应不同的计算预算与性能需求，CARM被设计为三个渐进式版本：基础版（CARM-Basic）集成了通道与局部空间上下文增强，如图4所示；融合版（CARM-Fusion）进一步引入跨轴全局上下文与多尺度融合；增强版（CARM-Plus）则完整整合所有上下文组件，并引入原始特征残差连接以稳定训练，如图6所示。在YOLO骨干网络中的嵌入实验表明，该模块能够有效提升模型

对不同尺度目标的检测精度，特别是在背景复杂、目标尺度多变的场景中，其多层次上下文感知能力显著增强了特征的判别性。CARM以模块化设计提供了即插即用的上下文增强方案，为提升轻量级视觉模型的场景适应能力提供了有效途径。

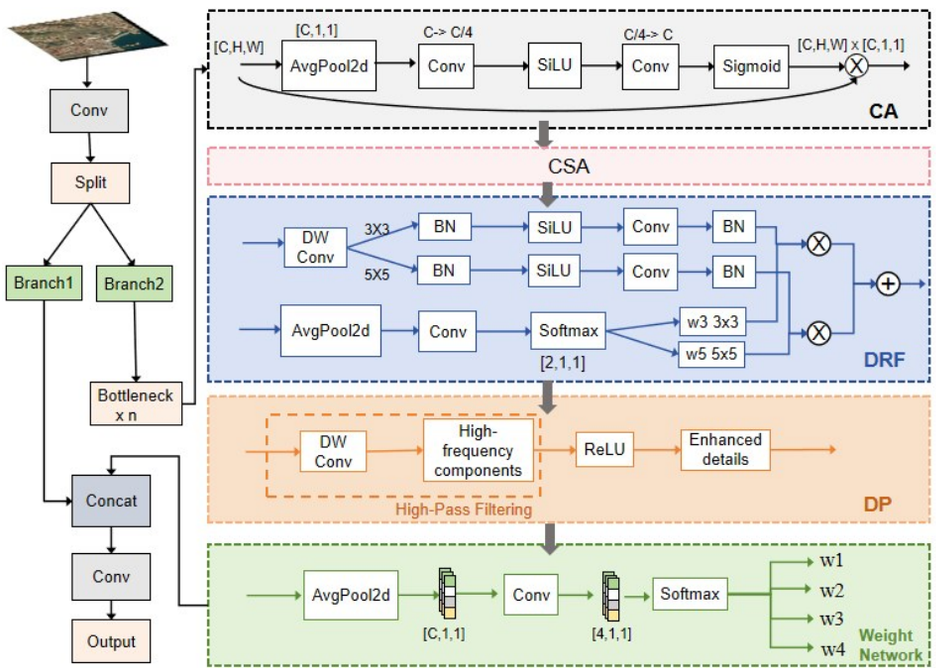


图6 CARM-Plus的整体结构图

Fig. 6 Complete architecture of CARM-Plus

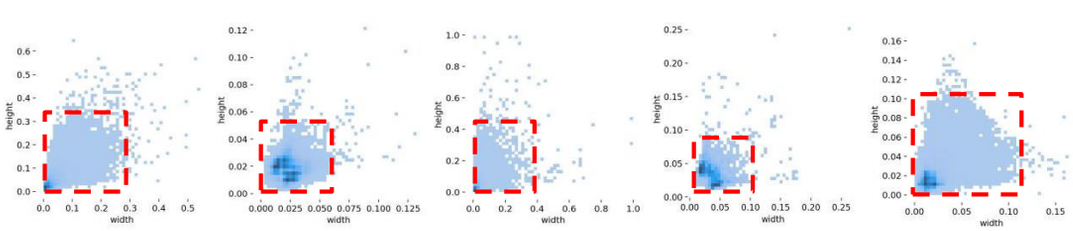
3 实验与结果分析

3.1 数据集概述

本研究主要采用五个公开的小目标数据集，涵盖无人机、遥感及航空场景，这些数据集均具有目标尺寸微小、背景复杂及类别多样化的特征。目标实例的分布情况如图7所示。

1) *VisDrone2019－DET*：VisDrone2019[22]是一个大规模公开无人机视角数据集，包含10,209张静态图片，涵盖城市、乡村、道路等多种复杂场景，标注了超过260万个目标边界框。数据集包含行人、车辆、自行车等10个类别，93.4%的图像包含小型实例，97.8%包含中型实例，60.6%包含大型实例[23]。

2) *AI－TODv2*：AI-TODv2[24]是面向微小目标检测的专用数据集。其图像中目标的平均大小仅为12.8像素，远小于通用数据集，精准聚焦“微小目标”这一核心挑战。其包含28,036张图像多时相航拍图像（涵盖昼夜光照变化特征），共8个类别，这一数据不仅有效验证了模型在弱



(a) (b) (c) (d) (e)

图7 数据集目标分布情况。(a)VisDrone2019;(b)USOD;(c)CoDrone;(d)VEDAI;(e)AI-TODv2。

Fig. 7 Distribution of target instances in the dataset. (a) VisDrone2019; (b) USOD; (c) CoDrone; (d) VEDAI; (e) AI-TODv2.

光条件下的性能连续性，更通过微小物体检测挑战，证实了算法在复杂光照尺度耦合场景中具备全面的检测能力。

3) VEDAI：Vehicle Detection in Aerial Imagery[25]是一个用于车辆检测的航空影像数据集，包含高分辨率彩色图像与红外图像，又分为1024x1024分辨率类型和512x512分辨率类型。共标注了9个类别：汽车、卡车、皮卡、拖拉机、露营车、船只、摩托车和公交车以及其它，其特点是目标尺寸极小且背景复杂。本实验使用的是1，249张1024x1024分辨率的彩色图像，并将数据集以7:2:1的比例划分。

4) *CoDrone*：CoDrone[26]包含了10004张高分辨率图像，按照5:2:3进行划分。拥有水平边界框和旋转框两种标注类型，共13种标注类别。遵循常见的水平检测设置，本文实验在CoDrone中使用水平边界框注释。

5) *USOD*：USOD[16]数据集取自于UNICORN 2008[27]。包含总计3000张图像，涵盖43378个车辆实例。训练集与测试集的比例为7：3，其中尺寸小于16×16的物体占比为96.3%，尺寸小于32×32的物体占比为99.9%。

3.2 实验环境

本节所有实验均在单张NVIDIA GeForce RTX 3090 GPU上完成，实验框架为PyTorch。LCDA-Net采用SGD优化器进行训练，初始学习率设为0.01，动量设为0.937，权重衰减设为0.0005，batchsize为8，训练轮数为300，输入分辨率统一设为640×640。除了在VisDrone数据集上的实验，本文的LCDA-Net都采用了归一化Wasserstein距离[28](NWD)作为损失函数，NWD将边界框间的距离建模为Wasserstein距离，从而降低IoU对小目标的敏感性。

3.3 评价指标

为验证本文提出方法的有效性，我们采用精确率、召回率、平均精确率、平均精确率均值及方法模型参数数量等指标对算法进行评估。

精确率表示算法模型在判定为阳性的样本中正确预测的比例，其计算公式为：

$$P=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}\ . \tag{8}$$

召回率表示被标记为正例的样本中被算法模型正确预测的比例。它被定义为：

$$R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}\ . \tag{9}$$

准确率-召回率曲线和坐标x轴下方的区域称为AP。平均准确率均值(mAP)是数据集中所有类别AP值的平均值。计算公式为：

$$\mathrm{AP}=\int_0^1 P(y)dy\ , \tag{10}$$

$$\mathrm{mAP}=\frac{\sum_n^N \mathrm{AP}_n}{N}\ . \tag{11}$$

在上式中，TP表示真阳性样本的数量，FP表示假阳性样本的数量，FN表示假阴性样本的数量，y为准确率-召回率曲线。

此外，Params是模型训练中需要训练的参数总数，FLOPs是可以执行的计算量，用于度量模型的复杂度。

3.4 对比实验

如表1所示，为验证LCDA-Net在航拍目标检测任务中的性能优势，本文选取了具有代表性的基础目标检测模型进行对比实验，以评估不同检测框架的整体性能，并为后续分析所提方法的改进效果提供参考。实验结果表明，两阶段检测方法通常具有较高的参数量和计算开销，而单阶段方法在模型效率方面具有一定优势。相比基线YOLOv11-S，本文提出的Ours-S在保持较低模型规模的同时，实现了更优的检测性能，其mAP50和mAP50:95分别达到0.481和0.296，较基线模型分别提高8.8和6个百分点，表明所提方法能够在控制模型规模的基础上有效提升目标检测精度。

为进一步评估LCDA-Net与现有改进型检测方法的性能差异，本文选取了近年来具有代表性的优化模型进行对比，重点分析不同改进策略在航拍目标检测任务中的实际表现。如表2所示，本文方法在pedestrian、people等具有代表性的小目标类别上取得了更优或具有竞争力的检测结果，表明所提出的网络结构能够有效适应航拍图像中小目标的检测需求。

我们进一步在不同数据集上验证了LCDA-Net在多种典型场景下的小目标检测效果。如图8所示，在远距离港口、高速公路等场景中，LCDA-Net均表现出较好的检测性能。而通过在VisDrone2019数据集上的检测结果对比，如图9所示，可以看出相较于YOLOv11s和YOLOv8s，

表1 不同目标检测方法在 VisDrone 数据集上的性能对比

Tab. 1 Performance comparison of different object detection methods on the VisDrone dataset

Model	ImgSize	mAP_{50}	$mAP_{50:95}$	Params /M	GFLOPS	Model Size/MB
Two-stage methods						
Faster R-CNN	1 024×1 024	0.325	0.193	60.06	288.55	450.56
Cascade R-CNN	800×800	0.361	0.189	71.68	200.42	527.32
One-stage methods						
YOLOv5-S	640×640	0.392	0.234	9.12	23.8	18.5
YOLOv8-N	640×640	0.339	0.196	3.01	8.1	6.3
YOLOv8-S	640 × 640	0.396	0.238	11.13	28.5	22.5
YOLOv8-M	640×640	0.431	0.262	25.85	78.7	52
YOLOv9-S	640×640	0.399	0.241	7.17	26.7	15.2
YOLOv9-M	640 × 640	0.441	0.269	20.02	76.5	40.8
YOLOv10-N	640×640	0.334	0.196	2.27	6.6	5.8
YOLOv10-S	640×640	0.396	0.239	7.22	21.5	16.6
YOLOv10-M	640×640	0.431	0.263	15.32	59	33.5
YOLOv11-N	640×640	0.33	0.193	2.58	6.3	5.5
YOLOv11-S	640×640	0.393	0.236	9.42	21.3	18.3
YOLOv11-M	640×640	0.445	0.272	20.04	67.8	40.5
YOLOv12-N	640×640	0.328	0.191	2.56	6.4	5.5
YOLOv12-S	640×640	0.395	0.237	9.23	21.2	18.9
YOLOv12-M	640×640	0.437	0.266	20.11	67.3	40.8
YOLOv13-N	640×640	0.324	0.187	2.45	6.2	5.4
YOLOv13-S	640×640	0.376	0.223	9.00	20.7	18.6
YOLOv26-N	640×640	0.329	0.181	2.37	**5.5**	5.4
YOLOv26-S	640×640	0.391	0.223	9.47	20.9	20.3
End-to-end methods						
DETR	multiple	0.254	0.123	41.28	57.25	474.05
RT-DETR-R18	640×640	0.436	0.256	20.09	60.90	307
RT-DETR-R34	640×640	0.453	0.266	31.33	92.95	479.2
Ours						
Ours-N	640×640	0.426	0.256	**1.26**	11.7	**2.9**
Ours-S	640×640	**0.481**	**0.296**	4.92	44.7	10.3

表2 小模型与大模型检测器在 VisDrone 上的性能对比(仅展示典型小目标主导类别)

Tab. 2 Performance comparison between small-scale and large-scale detectors on VisDrone (only representative small-object-dominated categories are reported)

Methods	PED	PEO	BIC	TRI	AT	mAP_{50}	$mAP_{50:95}$	Params	GFLOPS
small size									
FBRT-YOLOs (AAAI 2025)[17]	0.462	0.358	0.15	**0.311**	0.17	0.423	0.257	2.89M	22.9
AntiUAV (Drones 2025)[29]	0.41	0.345	0.097	0.223	0.118	0.354	0.209	2.86M	12.2
MCA-YOLO(GRSL 2024)[30]	0.441	0.399	0.118	0.229	0.107	0.376	0.2	6.37M	12.3
SuperYOLO(RGB) (TGRS 2023)[31]	0.45	0.296	0.104	0.199	0.105	0.392	0.214	4.83M	16.61
LDTNet (TGRS 2025)[19]	0.089	0.047	0.013	0.045	0.03	0.214	0.103	4.85M	18.85
Ours-N	**0.483**	**0.389**	**0.154**	0.308	**0.177**	**0.426**	**0.257**	**1.26M**	**11.7**
big size									
FFCA-YOLO (TGRS 2024)[16]	0.539	0.426	0.196	0.328	0.163	0.452	0.261	7.14M	51.4
FBRT-YOLOm (AAAI 2025)[17]	0.494	0.384	0.189	0.354	0.206	0.455	0.279	7.36M	58.7
QueryDet (CVPR 2022)[32]	0.289	0.188	0.12	0.159	0.076	**0.500**	0.27	39.59M	—
YOLC (TITS 2024)[33]	0.285	0.169	0.121	0.183	0.119	0.491	0.277	25.66M	102.42
Detectors (CVPR 2021)[34]	0.056	0.022	0.025	0.079	0.048	0.231	0.142	123.12M	116.86
CEASC(CVPR 2023)[35]	0.239	0.142	0.112	0.199	0.112	0.46	0.267	43.06M	**40.32**
Ours-S	**0.554**	**0.445**	**0.21**	**0.365**	**0.203**	0.481	**0.296**	**4.92M**	44.7

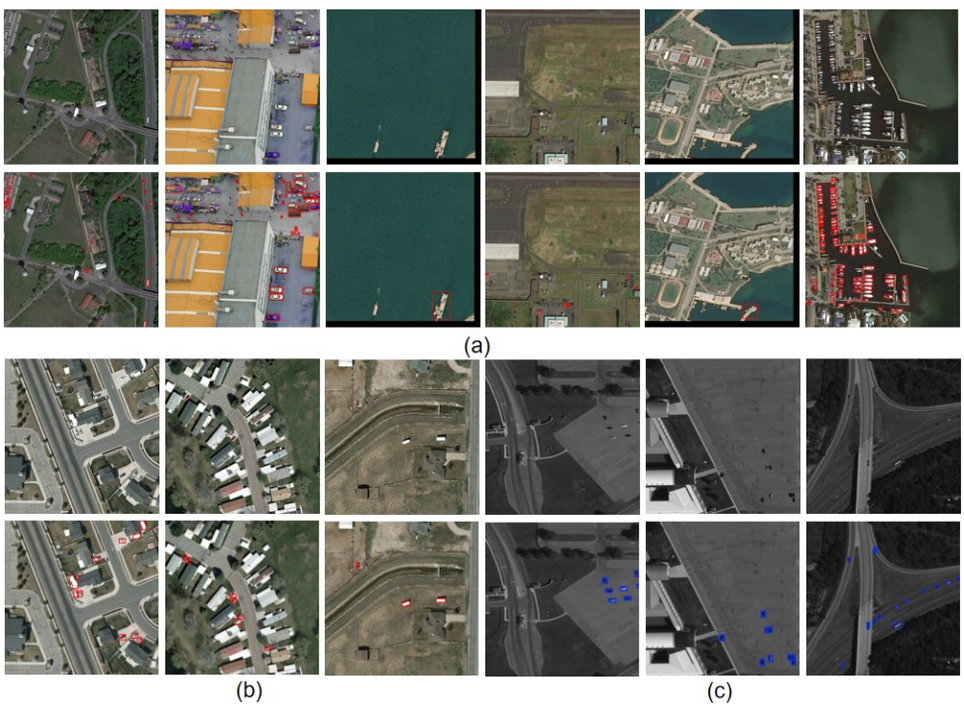


图8 LCDA-Net在AI-TOD、USOD及VEDAI中针对典型小目标场景(如港口、高速公路)的检测结果。(a) AI-TOD数据集结果;(b) VEDAI数据集结果;(c)USOD数据集结果。

Fig. 8 Detection results of LCDA-Net on typical small-object scenarios, such as harbors and highways, in the AI-TOD, USOD, and VEDAI datasets. (a) results on the AI-TOD dataset; (b) results on the VEDAI dataset; and (c) results on the USOD dataset.

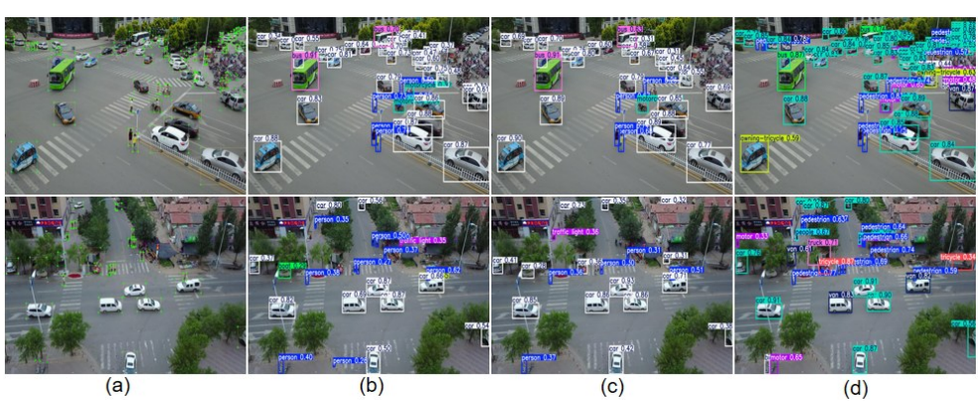


图9 在数据集VisDrone2019上各模型检测结果。(a)原始标签;(b)YOLOv11s;(c)YOLOv8s;(d)Ours。

Fig. 9 Detection results of different models on the VisDrone2019 dataset. (a) ground-truth annotations; (b) YOLOv11s; (c) YOLOv8s; and (d)Ours.

所提出的方法在复杂交通场景中对小目标和密集目标的检测更加准确,漏检和误检现象明显减少。尤其是在低空远距离行人和车辆等小目标区域,模型能够给出更加完整且紧密的检测框,同时在遮挡和背景干扰较强的情况下仍保持良好的鲁棒性,整体检测结果更加接近真实标注,进一步验证了模型在小目标检测任务中的有效性。

在AITODv2数据集上,我们进一步对比了不同尺度目标的检测性能。从表3可以看出,Ours-S在各尺寸等级上均表现出稳定且显著的提升。对于更具挑战性的小目标(APs),Ours-S检测精度为0.366,相较于L-FFCA-YOLO的0.327和HIC-YOLOv5的0.237,分别提升了约12%和54%。在中等尺寸目标(APm)上,Ours-

表3 不同检测器在AITODv2上多尺度目标检测性能对比

Tab. 3 Comparison of small-object detection performance of different detectors on AITODv2

Method	ImageSize	AP50	AP75	AP50:95	APvt	APt	APs	APm	Param/M
ATSS[36]	640×640	0.281	0.054	0.105	0.012	0.076	0.173	0.312	51.06
RetinaNet[37]	800×800	0.294	0.066	0.113	0.022	0.076	0.18	0.4	36.25
CenterNet[38]	512×512	0.236	0.026	0.073	0.015	0.065	0.115	0.186	14.21
CenterNet	800×800	0.308	0.038	0.098	0.019	0.096	0.14	0.227	14.21
CEASC	640×640	0.346	0.085	0.143	0.015	0.117	0.238	0.356	43.05
Detectors	640×640	0.191	0.072	0.09	0	0.03	0.209	0.379	123.22
LEAF-YOLO[39]	640×640	0.265	0.077	0.116	0.030	0.109	0.177	0.239	**4.27**
HIC-YOLOv5[40]	640×640	0.399	0.124	0.179	0.047	0.165	0.237	0.355	9.32
PDWT-YOLO[41]	640×640	0.403	—	0.165	—	—	—	—	6.20
L-FFCA-YOLO	640×640	0.614	0.181	0.268	0.124	0.258	0.327	0.403	5.06
YOLOv11-S	640×640	0.477	0.137	0.203	0.052	0.190	0.286	0.364	9.42
Ours-S	640×640	**0.640**	**0.231**	**0.294**	**0.133**	**0.284**	**0.366**	**0.464**	4.92

S的0.464也高于L-FFCA-YOLO的0.403与HIC-YOLOv5的0.355，相对提升15%与31%。即便是尺度更小的极端目标（APvt），Ours-S也能获得0.133的精度，大幅优于HIC-YOLOv5的0.047。整体而言，Ours-S在仅4.92M参数的轻量设计下，实现了从极小到大尺寸目标的全面覆盖与精度领先，充分证明了其多尺度特征提取能力的鲁棒性与通用性。

如图10所示，在CoDrone数据集上的检测效果对比，可以看出相比Baseline和FFCA-YOLO，所提出的方法在复杂场景中对小目标和密集目标的检测更加准确。Baseline存在明显漏检与定位不准问题，FFCA-YOLO虽有所改善，但在远距离和遮挡区域仍存在检测不完整现象；而LCDA-Net能够更好地捕捉密集分布的小目标，检测框更加完整且贴合目标边界，在复杂背景和多尺度变化下表现出不错的鲁棒性与细节感知能力。

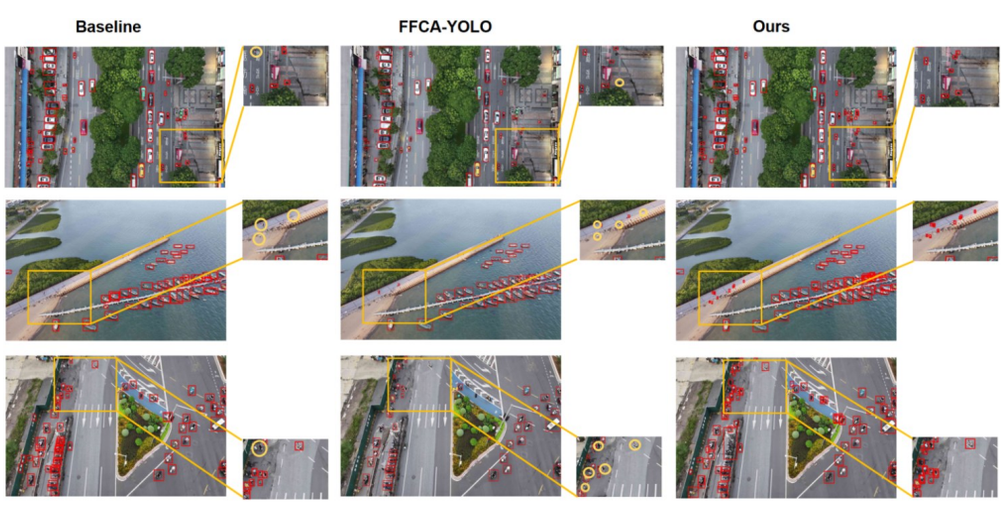


图10 在数据集CoDrone上不同模型的检测结果区别（黄色圆圈内为模型漏检处）

Fig. 10 Comparison of detection results of different models on the CoDrone dataset (Yellow circles indicate missed detections)

为进一步验证模型在不同小目标场景中的适应性，本文在USOD、CoDrone和VEDAI三个数据集上进行了对比实验。由表4可见，本文方法在USOD上取得0.923的mAP50，较次优方法0.906提升1.9%；在CoDrone上，mAP50和mAP50:90分别达到0.311和0.159，较对比方法分别提升

4.0%和8.2%;在VEDAI上,本文方法的mAP50达到0.625,与最优结果持平,同时mAP50:90提升至0.401。上述结果表明,所提方法在多种小目标场景下均表现出较好的检测能力。

表4 相同检测模型在不同数据集上的对比实验

Tab. 4 Performance comparison of different detectors across multiple datasets

Method	USOD				CoDrone		VEDAI(RGB)	
	Params/M	GFLOPs	mAP_{50}	$mAP_{50:90}$	mAP_{50}	$mAP_{50:90}$	mAP_{50}	$mAP_{50:90}$
FCOS[42]	50.8	109.23	—	—	—	—	0.610	0.380
GFL[43]	32.05	81.94	—	—	—	—	—	—
EfficientDet[44]	18.53	48.24	—	—	—	—	0.587	0.342
TAM-TR[45]	41.93	400.1	0.91	0.35	—	—	—	—
HR-FPN[46]	32.06	282.2	0.906	**0.565**	—	—	—	—
YOLOv3[47]	103.67	246	0.901	0.356	0.283	0.147	0.615	0.377
FFCA-YOLO	7.12	51.2	0.898	0.344	0.299	0.145	**0.625**	0.360
FBRT-YOLO	7.36	58.7	0.871	0.328	—	—	—	—
MCA-YOLO	6.27	—	0.762	0.247	0.266	0.121	0.269	0.121
YOLOv8m	25.84	78.7	0.897	0.353	0.284	0.145	0.623	0.386
YOLOv11s	9.41	**21.3**	0.88	0.337	0.261	0.132	0.616	0.368
Ours-S	**4.92**	44.6	**0.923**	0.355	**0.311**	**0.159**	**0.625**	**0.401**

从表5中可以看出,LCDA-Net在实时性方面表现较为优秀,其FPS达到117.65,推理时间为7.3ms,优于大多数对比方法,如HR-FPN和YOLOv8m,说明模型在保证精度的同时具备较强的实时检测能力。同时,相比FBRT-YOLO等轻量模型,其在精度提升的前提下仍保持较高的推理速度,体现出较好的效率与性能平衡。但LCDA-Net相比于MCA-YOLO,吞吐率更高,但单张推理时延不是最低,说明在极端追求速度的场景下仍存在一定局限。此外,引入多种注意力与动态模块虽然提升了检测性能,但也带来了一定计算开销,使得模型在低算力设备上的极致优化空间仍有待进一步提升。

表5 在数据集NWPU-VHR-10测试集上的FPS实验

Tab. 5 FPS experiments on the test set of the NWPU-VHR-10 dataset

Methods	FPS	inference/ms	mAP_{50}	Params/M
HR-FPN	31.15	25.5	0.83	32.06
YOLOv8m	25.45	34.7	0.883	25.84
FBRT-YOLO	90.93	6.797	0.858	7.36
MCA-YOLO	69.20	**4.79**	0.886	6.27
FFCA-YOLO	48.48	18.4	0.871	7.12
LCDA-Net	**117.65**	7.3	**0.887**	**4.92**

3.5 消融实验

3.5.1 结构消融

如图11所示,低层特征如P1、P2具有更高的空间分辨率,主要负责表征小目标细节信息;而高层特征如P4、P5语义信息更丰富,但空间分辨率较低,更适合检测大目标。对于以小目标为主的数据集,保留或强化低层特征具有关键意义,否则容易导致小目标信息丢失,从而影响整体检

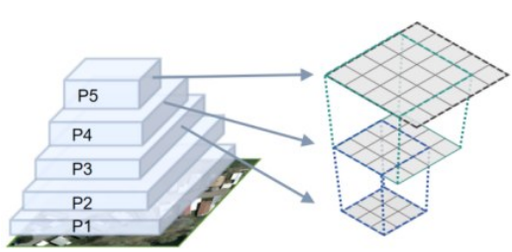


图11 典型特征金字塔结构中不同层级特征图与目标尺度之间的对应关系

Fig. 11 Correspondence between feature maps at different levels and object scales in a typical feature pyramid structure

测性能。通过表1可见，HS-FPN[48]实现了FPN的变体，有效但是仍然有限；直接删除P5后，模型性能仅发生轻微变化，说明在以小目标为主的数据集上，最深层低分辨率特征的边际收益相对有限，而高分辨率特征对小目标检测更为关键；在此基础上进一步引入P2后，模型性能有所提升，说明高分辨率特征对小目标检测更为重要。然而，仅增加低层特征仍不足以充分补偿高层语义表达的不足。相比之下，LCPAN不仅保留了更多高分辨率细节信息，还通过双向特征融合与检测端高层特征重构增强了语义表达，因此取得了更优的检测结果。如图12所示，LCPAN对于目标存在的位置更为敏感和准确。

表6 数据集AITODv2、VisDrone2019上对模型进行结构调整对比

Tab. 6 Comparison of structural adjustments to the model on the AITODv2 and VisDrone2019 datasets

Dataset		AITODv2							VisDrone		
Base	Methods	AP50	AP75	AP50:95	APvt	APt	APs	APm	mAP_{50}	$mAP_{50:95}$	Param/M
Cascade_RCNN	HS-FPN	—	—	—	—	—	—	—	0.361	0.189	71.68
YOLOv11n	default	0.386	0.103	0.161	0.035	0.144	0.235	0.317	0.33	0.193	2.58
	delete P5	0.379	0.101	0.159	0.045	0.154	0.224	0.300	0.335	0.196	1.85
	add P2 &delete P5	0.440	0.125	0.186	0.058	0.179	0.245	0.341	0.361	0.215	1.93
	LCPAN	**0.502**	**0.159**	**0.224**	**0.062**	**0.209**	**0.287**	**0.403**	**0.404**	**0.244**	**1.58**



图12 不同结构下网络响应的位置

Fig. 12 Target response locations under different network structures

3.5.2 损失消融

从损失函数的信息约束角度来看，不同损失函数对目标定位特征的关注重点存在差异。传统基于IoU的边界框回归损失主要利用预测框与真实框的重叠关系进行优化，对于极小目标，微小的位置偏移可能导致IoU明显变化。NWD则将边界框表示为概率分布，并通过Wasserstein距离衡量预测框与真实框之间的几何分布差异，主要反映目标中心位置及尺寸方面的偏差。相较于直接依据重叠比例进行度量，该方法能够降低极小目标定位对微小位置偏移的敏感性。

然而，NWD的几何距离度量并不直接包含边界框的交并比信息，也不直接建模目标类别语义、复杂背景纹理及相邻实例之间的关系。因此，NWD不能单独解决类别混淆、背景干扰等问题，其优化结果也不一定在所有数据集上带来更高的定位精度。基于此，本文通过损失消融实验，分析NWD在不同目标尺度分布和数据集上的适用性，并结合检测指标评价其实际影响。

从表7结果可以看出，归一化Wasserstein距

表7 NWD损失函数对于模型在小目标数据集的影响

Tab. 7 Impact of the NWD loss function on the model over small-object datasets

Scale	NWD	Precision	Recall	mAP_{50}	$mAP_{50:95}$
USOD					
S	×	0.908	0.853	0.901	0.34
	√	0.918	0.885	0.923	0.355
N	×	0.891	0.84	0.882	0.316
	√	0.899	0.854	0.893	0.327
VisDrone 2019					
S	×	0.577	0.459	0.481	0.296
	√	0.575	0.464	0.48	0.293
N	×	0.527	0.413	0.426	0.256
	√	0.52	0.4	0.417	0.249

离损失(NWD)函数在VisDrone数据集上的表现并不佳,因此在此数据集上我们仍使用默认损失函数;但对于尺寸小于16×16的物体占比达到96.3%的数据集USOD而言,其对于模型检测精度的提升明显,说明其对数据分布具有一定敏感性。因此本文开展该实验旨在验证基于分布距离的回归损失在小目标定位中的有效性,并进一步说明相比损失层优化,LCDA-Net通过特征建模(CARM、LCPAN)和检测头设计(SO-DyHead)带来的性能提升更加稳定可靠。未来可考虑将NWD与LCDA-Net进行自适应融合,以在不同尺度场景下进一步提升定位精度。

3.5.3 模块消融

为分析各模块对整体性能的贡献,本文进一步在VisDrone2019上开展了模块消融实验,如表8所示。该模型利用方向感知机制增强长程依赖建模能力,同时通过小目标感知模块提升高分辨率特征的表达能力,从而在复杂场景下有效改善密集小目标的检测性能,并在精度与计算开销之间取得良好平衡。

从表8消融实验结果可以看出,各模块在完整框架中的作用具有明显互补性。其中,LCPAN带来了最显著的单独增益,说明高分辨率特征保留对小目标检测至关重要;CARM进一步提升了复杂背景下的表征能力;SO-DyHead在单独使用时提升有限,但在LCPAN与CARM提供更优输入特征的基础上能够进一步释放检测端的动态适配能力,最终使完整模型达到最优性能。

表8 在数据集VisDrone2019上的消融实验

Tab. 8 Ablation experiments on the VisDrone2019 dataset

Modules			Precision	Recall	mAP_{50}	$mAP_{50:95}$	Params/M
CARM	LCPAN	SO-DyHead					
			0.496	0.385	0.393	0.236	9.42
✔			0.512	0.386	0.4	0.24	9.56
	✔		0.559	0.448	0.466	0.286	5.41
		✔	0.497	0.39	0.392	0.238	8.85
✔	✔		0.561	0.454	0.474	0.291	5.49
	✔	✔	0.568	0.458	0.474	0.291	4.85
✔		✔	0.518	0.389	0.396	0.237	9.0
✔	✔	✔	**0.578**	**0.459**	**0.481**	**0.296**	**4.92**

图13对比了引入模块后对特征响应的影响,可以看出基线模型的注意力分布较为分散,对小目标区域关注不足;引入LCPAN和CARM后,模型能够更集中地聚焦于潜在目标区域,背景干扰明显减弱;进一步结合完整模型后,目标区域的响应更加突出且边界更清晰,尤其在密集小目标场景中表现更为明显,对关键目标的定位更加准确,最终检测结果也更加完整,验证了各模块在增强特征表达与提升小目标感知能力方面的有效性。

3.5.4 检测误差分析

为进一步分析LCDA-Net的检测误差及其相较于基线模型的改进效果,本文对YOLO11-S与LCDA-Net在VisDrone2019验证集上的归一化混淆矩阵进行了对比,如图13所示。可以看出,LCDA-Net在全部10个类别上的对角线数值均高于基线模型,表明各类别的正确预测比例均有所提高。其中,tricycle、motor、pedestrian和people的对角线数值分别提高了0.18、0.18、0.17和0.17,说明所提方法对这些类别的识别表现改善较为明显。

从漏检情况来看,LCDA-Net中各类别对应的背景列数值均低于基线模型,表明被预测为背景的真实目标比例有所下降。这说明改进模型对部分原本难以识别的目标产生了更有效的检测响应。此外,混淆矩阵仍显示部分外观相似的类别之间存在混淆,例如van被预测为car的比例由0.36变为0.39,说明车辆类别之间的区分仍有进一步提升空间。背景行也表明模型仍存在一定的背景误检现象。

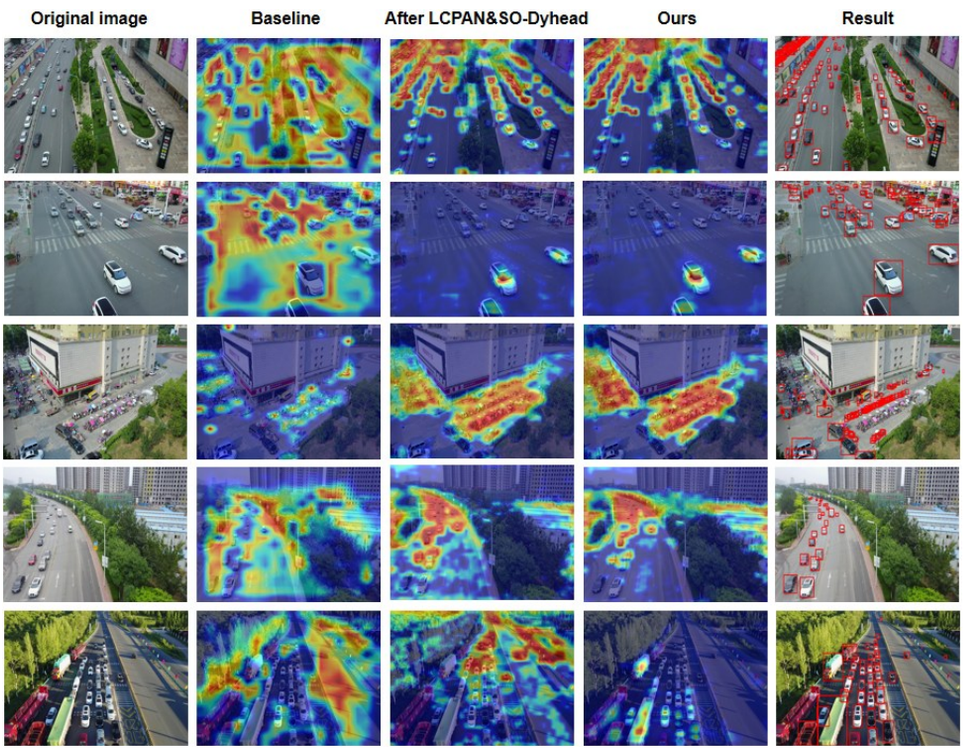


图 13　改进后的的模型对目标区域的响应效果展示，颜色越亮表示模型对该区域的关注度越高。

Fig. 13　Visualization of the response of the improved model to target regions, where brighter colors indicate higher attention of the model to those regions.

综合上述分析，LCDA-Net 在类别级正确预测比例和漏检控制方面表现出一定优势，但对相似类别的区分以及复杂背景下的误检抑制仍有改进空间。结合整体检测评价指标，可以进一步从类别级误差分布与总体检测性能两个方面评价所提方法的有效性。

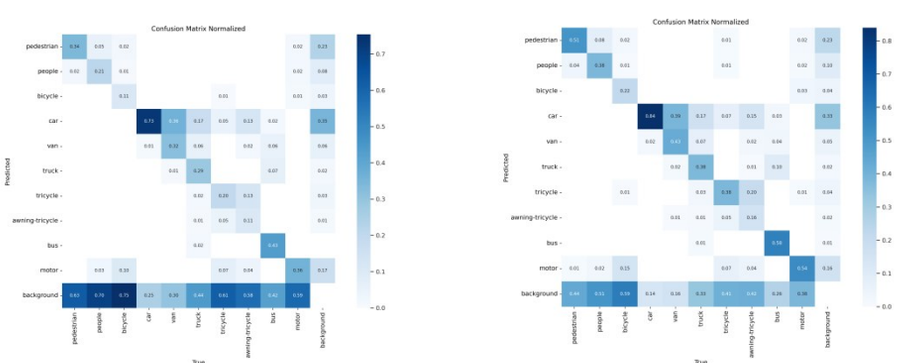


图 14　YOLO11-S 与 LCDA-Net 在 VisDrone2019 数据集上的归一化混淆矩阵对比

Fig. 14　Comparison of normalized confusion matrices between YOLO11-S and LCDA-Net on the VisDrone2019 dataset

4　结　　论

本文提出了一种面向航拍图像小目标检测的低参数协同方向感知网络 LCDA-Net。该方法通过 LCPAN、CARM 和 SO-DyHead 三个关键模块的协同设计，增强了高分辨率细节保留能力、复杂背景下的上下文表征能力以及检测端对小目标特征的自适应建模能力。实验结果

表明,本文方法在多个小目标主导数据集上取得了较优性能,验证了所提结构设计的有效性。

尽管如此,LCDA-Net仍存在一定局限。一方面,模型在获得更高检测精度的同时带来了额外计算开销;另一方面,当前方向建模主要面向水平与垂直轴,对旋转目标或任意方向目标的适应能力仍有待增强。未来工作将围绕更低计算开销的结构设计、旋转场景下的方向建模以及复杂密集场景中的鲁棒检测展开。

参 考 文 献:

[1] DAI Y, LI X, ZHOU F, *et al*. One-stage cascade refinement networks for infrared small target detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-17.

[2] WANG Q, YU C. AMFE-YOLO: A Small Object Detection Model for Drone Images[J]. *IET Image Processing*, 2025, 19(1): e70110.

[3] ZHOU S, ZHOU H. Detection based on semantics and a detail infusion feature pyramid network and a coordinate adaptive spatial feature fusion mechanism remote sensing small object detector[J]. *Remote Sensing*, 2024, 16(13): 2416.

[4] SU Z, Li X, Zhao F, *et al*. Long-Distance Small Target Recognition Based on Dynamic Frame Difference Method [C]//Proceedings of the 2024 International Conference on Image Processing, Multimedia Technology and Maching Learning. 2024: 173-179.

[5] JIANG W, Han D, Han B, *et al*. YOLOV8-FDF: A small target detection algorithm in complex scenes[J]. *IEEE Access*, 2024, 12: 119223-119237.

[6] LI C, Chang Y, Yang S, *et al*. A directional enhanced adaptive detection framework for small targets [J]. *Electronics*, 2024, 13(22): 4535.

[7] HUI Y, WANG J, LI B. STF-YOLO: a small target detection algorithm for UAV remote sensing images based on improved SwinTransformer and class weighted classification decoupling head[J]. *Measurement*, 2024, 224.

[8] 王伯霄, 宋延嵩, 董小娜. 难点注意力感知红外小目标检测网络[J]. 中国光学(中英文), 2024, 17(3): 538-547.
WANG Bo-xiao, SONG Yan-song, DONG Xiao-na. Indistinguishable points attention-aware network for infrared small
object detection[J]. *Chinese Optics*, 2024, 17(3): 538-547.

[9] KHANAM R, HUSSAIN M. Yolov11: an overview of the key architectural enhancements[J]. 2024.

[10] 张志豪, 厉小润, 陈淑涵. 基于改进YOLO11的无人机航拍图像小目标检测算法[J]. 液晶与显示, 2025,40(6): 915-930.
ZHANG Zhihao, LI Xiaorun, CHEN Shuhan. Small object detection algorithm in UAV aerial images based on improved YOLO11[J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(6): 915-930.

[11] 吕学涵, 李富, 祁铭瑞, 等. 基于改进YOLOv11s的无人机小目标检测算法[J]. 液晶与显示, 2025,40(11): 1744-1756.
XuehanLÜ, LI Fu, QI Mingrui, *et al*. Target detection algorithm based on improved YOLOv11s UAV aerial image [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(11): 1744-1756.

[12] 李大湘,辛嘉妮,刘颖. 位置敏感Transformer航拍图像目标检测模型[J]. 光学精密工程,2024,32(05):727-739.
LI Daxiang,XIN Jiani,LIU Ying. Position-sensitive Transformer aerial image object detection model[J]. *Optics and Precision Engineering*,2024,32(05):727-739.

[13] ZHANG Y, YE M, ZHU G, *et al*. FFCA-YOLO for small object detection in remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-15.

[14] XIAO Y, XU T, XIN Y, *et al*. FBRT-YOLO: faster and better for real-time aerial image detection [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025, 39(8): 8673-8681.

[15] HUI Y, WANG J, LI B. DSAA-YOLO: UAV remote sensing small target recognition algorithm for YOLOV7 based on dense residual super-resolution and anchor frame adaptive regression strategy[J]. *Journal of King Saud*

University － Computer and Information Sciences, 2024, 36(1).

[16] LIU D, ZHANG J, QI Y, *et al*. Exploring lightweight structures for tiny object detection in remote sensing images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 1-15.

[17] 白晨帅,白晓凤,邬开俊等 . 超融合残差行进几何感知的遥感目标检测[J]. 光学精密工程,2025,33(08):1289-1302.
BAI Chenshuai, BAI Xiaofeng, WU Kaijun, *et al*. Remote sensing object detection algorithm based on ultra fusion residual marching geometric perception[J]. *Optics and Precision Engineering*, 2025, 33(08): 1289-1302.

[18] WU Q, YOU X, HUANG W, *et al*. M2FE-YOLO: multibranch and multilevel feature enhancement network for remote sensing object detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 1-19.

[19] Li G. H. Y., Leefmans C. R., Williams J. *et al*. *Deep learning with photonic neural cellular automata. Light Sci Appl* 13, 283 (2024). https://doi.org/10.1038/s41377-024-01651-7

[20] Dong Y., Bai Y., Zhang Q. *et al*. *High－throughput optical neuromorphic graphic processing at millions of images per second. eLight* 5, 29 (2025). https://doi.org/10.1186/s43593-025-00106-9

[21] Wang Y., Liao K., Zhang K. *et al*. *Reconfigurable versatile integrated photonic computing chip. eLight* 5, 20 (2025). https://doi.org/10.1186/s43593-025-00098-6

[22] DU D, ZHU P, WEN L, *et al*. VisDrone-DET2019: The vision meets drone object detection in image challenge results[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). IEEE, 2019: 213-226.

[23] CHEN L, LIU C, LI W, *et al*. DTSSNet: dynamic training sample selection network for UAV object detection [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-16.

[24] WANG J, YANG W, GUO H, *et al*. Tiny object detection in aerial images[C]//2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021: 3791-3798.

[25] RAZAKARIVONY S, JURIE F. Vehicle detection in aerial imagery: A small target detection benchmark[J]. *Journal of Visual Communication and Image Representation*, 2016, 34: 187-203.

[26] CHEN P, OUYANG T, LUO K, *et al*. CoDrone: autonomous drone navigation assisted by edge and cloud foundation models[J]. *IEEE Internet of Things Journal*, 2025.

[27] LEONG C, ROVITO T, MENDOZA-SCHROCK O, *et al*. Unified coincident optical and radar for recognition (UNICORN) 2008 dataset[J]. 2020-04-15]. https://github. com/AFRL-RY/data-unicorn-2008, 2019.

[28] WANG J, XU C, YANG W, *et al*. A normalized Gaussian Wasserstein distance for tiny object detection[J/OL]. *arXiv preprint*, 2021. [2026-08-18]. https://arxiv. org/abs/2105. 09914.

[29] YASMEEN A, DAESCU O. Recent research progress on ground-to-air vision-based anti-UAV detection and tracking methodologies: A review[J]. *Drones*, 2025, 9(1): 58.

[30] CHEN Z, LUO A, DING L, *et al*. A robust and efficient multiscenario object detection network for edge devices [J]. *IEEE Geoscience and Remote Sensing Letters*, 2025, 22: 1-5.

[31] ZHANG J, LEI J, XIE W, *et al*. SuperYOLO: super resolution assisted object detection in multimodal remote sensing imagery[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-15.

[32] YANG C, HUANG Z, WANG N. QueryDet: cascaded sparse query for accelerating high-resolution small object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 13668-13677.

[33] LIU C, GAO G, HUANG Z, *et al*. Yolc: you only look clusters for tiny object detection in aerial images[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2024, 25(10): 13863-13875.

[34] QIAO S, CHEN L-C, YUILLE A. Detectors: detecting objects with recursive feature pyramid and switchable atrous convolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 10213-10224.

[35] DU B, HUANG Y, CHEN J, *et al*. Adaptive sparse convolutional networks with global context enhancement for faster object detection on drone images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 13435-13444.

[36] ZHANG S, CHI C, YAO Y, *et al*. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 9759-9768.
[37] 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: 2980-2988.
[38] DUAN K, BAI S, XIE L, *et al*. Centernet: keypoint triplets for object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6569-6578.
[39] Nguyen H H, Hoang M S. Leaf-yolo: Lightweight edge-real-time small object detection on aerial imagery[J]. *Intelligent Systems with Applications*, 2025, 25: 200484.
[40] TANG S, ZHANG S, FANG Y. HIC-YOLOv5: improved YOLOv5 for small object detection[C]//2024 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2024: 6614-6619.
[41] ZHANG L, XIONG N, PAN X, *et al*. Improved object detection method utilizing yolov7-tiny for unmanned aerial vehicle photographic imagery[J]. *Algorithms*, 2023, 16(11): 520.
[42] TIAN Z, SHEN C, CHEN H, *et al*. Fcos: fully convolutional one-stage object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9627-9636.
[43] LI X, WANG W, WU L, *et al*. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection[J]. *Advances in neural information processing systems*, 2020, 33: 21002-21012.
[44] TAN M, PANG R, LE Q V. Efficientdet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
[45] XU J, *et al*. TAM-TR: text-guided attention multi-modal transformer for object detection in UAV images[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2025, 227: 170-184.
[46] CHEN Z, JI H, ZHANG Y, *et al*. High-resolution feature pyramid network for small object detection on drone view [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024, 34(1): 475-489.
[47] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. *arXiv preprint* arXiv:1804.02767, 2018.
[48] SHI Z, *et al*. HS-FPN: high frequency and spatial perception FPN for tiny object detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(7): 6896-6904.

作者简介：



王艳玲，女，硕士研究生，2023年于重庆理工大学获得学士学位，主要从事遥感图像处理方面的研究。E-mail:iris_wyl@163.com



刘亚秋，男，博士，教授，2004年于哈尔滨工业大学获得博士学位，主要从事信息控制与高性能计算领域及目标检测研究。E-mail:yaqiuLiu@nefu.edu.cn