

DOI:10.19651/j.cnki.emt.2519779

基于 YOLOv11n 改进的小目标航拍检测算法^{*}陈 磊^{1,2} 邢 智^{1,2} 粘山坡³ 薄敬东^{1,2} 常国涛^{1,2}(1. 华北理工大学电气工程学院 唐山 063210; 2. 河北省风光氢储安全监测与智能运行技术创新中心 唐山 063210;
3. 华北理工大学国有资产与实验室管理处 唐山 063210)

摘 要: 针对无人机航拍图像中目标尺寸小、遮挡严重、背景复杂等问题,传统目标检测算法常出现漏检与误检。为提升检测精度,在 YOLOv11n 基础上提出了一种融合注意力机制与特征调制机制的小目标检测算法—YOLOv11n-AFD。该方法通过集成空间条带注意力(SSA)模块、调制融合模块(MFM)与曼哈顿特征增强(MFE)模块,全面提升模型对小目标的感知能力与语义建模能力。在保持 YOLOv11n 框架不变的前提下,空间条带注意力建模水平与垂直空间依赖,增强结构感知;调制融合通过通道调制实现精细融合,突出关键信息;曼哈顿特征增强模块强化几何特征,抑制背景干扰,实现关键模块的深度增强。实验结果表明,改进后的 YOLOv11n-AFD 相较于原始 YOLOv11n 模型,在 VisDrone 数据集上准确率、召回率与检测精度 mAP@0.5 分别提升 1.8%、0.8% 和 1.4%,mAP@0.5:0.95 提升至 21.5%。对比其他算法表现出了良好的性能。

关键词: YOLOv11n; 小目标检测; 航拍图像; 空间条带注意力; 特征调制融合; 曼哈顿特征增强

中图分类号: TP391.4; TN919.8; TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.60

YOLOv11n-based improved algorithm for small object detection in
aerial imagesChen Lei^{1,2} Xing Zhi^{1,2} Nian Shanpo³ Bo Jingdong^{1,2} Chang Guotao^{1,2}(1. College of Electrical Engineering, North China University of Science and Technology, Tangshan 063210, China;
2. Hebei Province Innovation Center for Safety Monitoring and Intelligent Operation Technology of
Wind-Solar-Hydrogen-Storage Systems, Tangshan 063210, China; 3. State-owned Assets and
Laboratory Management Office of North China University of Science and Technology, Tangshan 063210, China)

Abstract: To address the challenges of small object size, severe occlusion, and complex backgrounds in UAV aerial images, traditional object detection algorithms often suffer from missed and false detections. To improve detection accuracy, this paper proposes a small object detection algorithm, YOLOv11n-AFD, based on YOLOv11n by integrating attention and feature modulation mechanisms. The method incorporates a Spatial Strip Attention (SSA) module, a Modulation Fusion Module (MFM), and a Manhattan Feature Enhancement (MFE) module to comprehensively enhance the model's perception and semantic representation of small objects. Within the unchanged YOLOv11n framework, the SSA models horizontal and vertical spatial dependencies to strengthen structural awareness; the MFM refines feature fusion through channel modulation to highlight key information; and the MFE reinforces geometric features while suppressing background interference for deeper enhancement. Experimental results show that YOLOv11n-AFD achieves improvements of 1.8% in precision, 0.8% in recall, and 1.4% in mAP@0.5 over the original YOLOv11n, with mAP@0.5:0.95 increasing to 21.5%, demonstrating superior performance compared with other algorithms.

Keywords: YOLOv11n; small object detection; aerial imagery; spatial strip attention; modulated feature fusion; Manhattan feature enhancement

0 引 言

当前,人工智能正逐渐成为社会形态质变的第一推动

力,以无人机为代表的低空经济产业得到了空前的重视和发展^[1]。无人机技术的快速发展,航拍图像已成为获取地面信息的重要手段,被广泛应用于农业监测、城市规划、环

收稿日期:2025-09-03

^{*} 基金项目:教育部产学研合作协同育人项目(230802495182120,231104697215835)、华北理工大学教育教学改革研究与实践项目(L2345)资助

环境保护以及灾害管理等领域^[2]。城市化进程的不断推进,汽车驾驶已经成为现代生活中不可或缺的交通方式^[3]。车辆检测是实现智能交通的关键部分^[4],视觉算法在辅助车辆路径规划、交通流量监控等方面具有广泛的应用价值,提升对车辆目标的检测与定位精度是智慧交通研究的重要内容^[5]。如何在复杂多变的道路环境中,实现对多类车辆的高效准确检测是道路监控、交通分析和违章管理智能化的关键,也影响智慧城市进程。作为环境感知的基础,车辆检测要求算法在复杂场景中具备高精度与实时性。目前,目标检测的主流算法主要可以分为两大类:二阶段检测器和一阶段检测器。二阶段算法,如 R-CNN^[6-8] 系列,通过候选区域生成器提取候选区域,然后进行分类,二阶段目标检测算法能保证不错的车辆检测效果,但其复杂的网络结构,导致这种类型的算法难以保证实时的车辆检测^[9];一阶段算法,例如 YOLO 系列^[10-13] 和 SSD^[14],通过卷积神经网络直接在图像上生成边界框并进行分类,速度较快但精度相对较低。

近年来,学者们针对 YOLO 算法在小目标检测中的局限性,提出了多种改进方案。冉宁等^[15] 提出 SBC-YOLOv8 算法,利用 SAHI 切片缓解目标密集与小尺寸问题,并结合 Space-to-Depth、BiFPN 与 CSP-OmniFusion 增强特征提取与多尺度融合,在多数数据集上取得显著提升,但在复杂地物干扰下的稳定性仍待加强;李天林等^[16] 提出 EM-YOLO,通过边界引导多尺度特征块与细节特征增强块提升特征表征,并针对遮挡改进检测头,构建 WPF-IoU 损失以优化训练,显著提升精度、召回率和 mAP,但其跨场景泛化能力仍需验证;李珺等^[17] 提出 WT-YOLO 航拍小目标检测算法,通过调整颈部特征图尺寸、引入 C3k2-WT 模块和 Focal-Modulation 机制,以及设计共享卷积检测头,实现了小目标高效检测与复杂场景下的鲁棒性提升,但在多层次特征融合方面仍有改进的空间。

本文提出的 YOLOv11n-AFD 结合空间条带注意力 (spatial strip attention, SSA)、调制融合模块 (modulation fusion module, MFM) 与曼哈顿特征增强模块 (Manhattan feature enhancement, MFM),实现了多层次特征建模,提升了小目标检测精度和复杂场景鲁棒性,体现了研究的现实意义和技术先进性。

1 YOLOv11n 算法原理

作为 YOLO 系列的演化版本,YOLOv11n 在模型结构设计、特征提取能力和检测精度方面均实现了显著优化,表现出较强的实际应用能力,尤其适用于需要高效检测的交通车辆识别场景。YOLOv11n 网络结构如图 1 所示。

YOLOv11n 延续了“Backbone-Neck-Head”的架构设计。在 Backbone 部分,YOLOv11n 引入了 C2f 模块,通过跨层连接与特征融合机制,在保持轻量化的同时提升了深层特征提取能力。这使其在复杂背景下仍能准确感知多样化目标,适

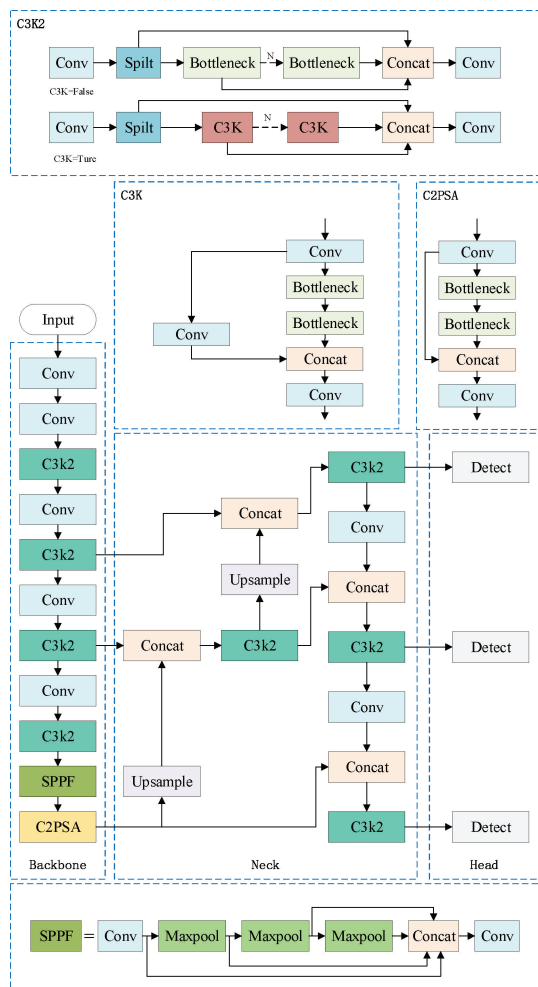


图 1 YOLOv11n 网络结构

Fig. 1 Structure of YOLOv11n network

用于不同类型、尺寸及密集重叠车辆的识别。Neck 部分结合了 FPN 与 PANet 结构,利用上采样和下采样操作实现多尺度特征的有效融合,兼顾浅层定位信息和深层语义理解。该结构显著提升了对小目标及尺度变化大的目标的检测效果,尤其适用于远距离或遮挡车辆的识别。Head 部分采用了解耦式检测头,将分类、置信度预测和边界框回归任务分离,分别独立训练和预测,从而减少任务间干扰,提高检测精度与稳定性。该检测头输出 3 个不同分辨率的特征图,分别对应不同尺度目标,实现对大中小车辆的全面检测。

此外,YOLOv11n 在训练和推理策略上也进行了多项优化。训练阶段引入了 Mosaic、MixUp、颜色扰动和随机裁剪等增强方法,并采用动态标签分配策略优化样本匹配过程,结合改进的 CIoU 损失提升边框回归精度。推理阶段支持导出 ONNX、TensorRT 等格式,适配边缘设备部署,满足智慧交通等应用对实时性强和精度高的双重要求。

2 改进的 YOLOv11n 算法

改进后的 YOLOv11n-AFD (attention and fusion

enhanced detection)算法在保持原有高效性和准确性的基础上,针对复杂交通场景中的应用进行了优化。在复杂动态环境下的车辆小目标、遮挡目标检测时仍存在一些局限性^[18]。因此,本文从主干网络、特征融合和检测头 3 个方面对其结构进行改进,全面增强其在多目标交通检测任务中的性能。改进后的整体结构如图 2 所示。

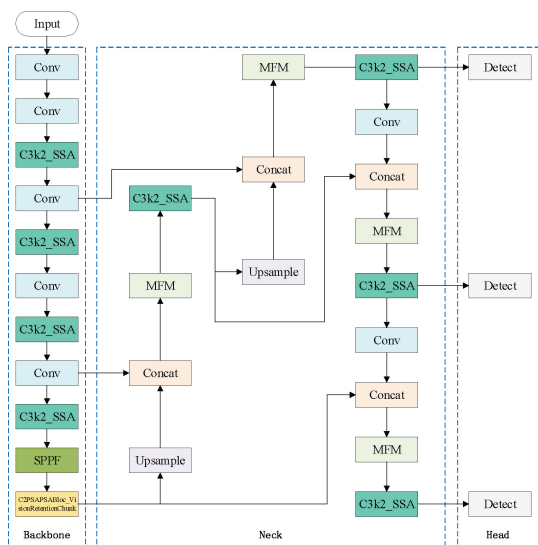


图 2 YOLOv11n-AFD 网络结构

Fig. 2 YOLOv11n-AFD network architecture

首先,在 Backbone 主干网络部分,原 YOLOv11n 中的 C3k2 模块被统一替换为 SSA 机制^[19]的 C3k2_SSA 模块。SSA 模块可分别沿水平与垂直方向捕捉近距离依赖信息,提升模型对车辆边缘和结构线条的空间感知能力。

其次,在 Backbone 末端的高层语义输出中,原来的 C2PSA 模块被升级为曼哈顿特征增强模块中的 C2PSAPSABloc_VisionRetentionChunk 模块,进一步融合了通道并行注意力与视觉保持机制^[20]。该模块增强了模型对小目标的聚焦,提升了远距离以及小尺度车辆的识别效果。

最后,在 Neck 部分保留了 FPN 与 PANet 的经典上下采样融合结构,并引入了多组多尺度特征融合模块^[21]中的 MFM 模块来替代传统的拼接与卷积操作。每次特征融合

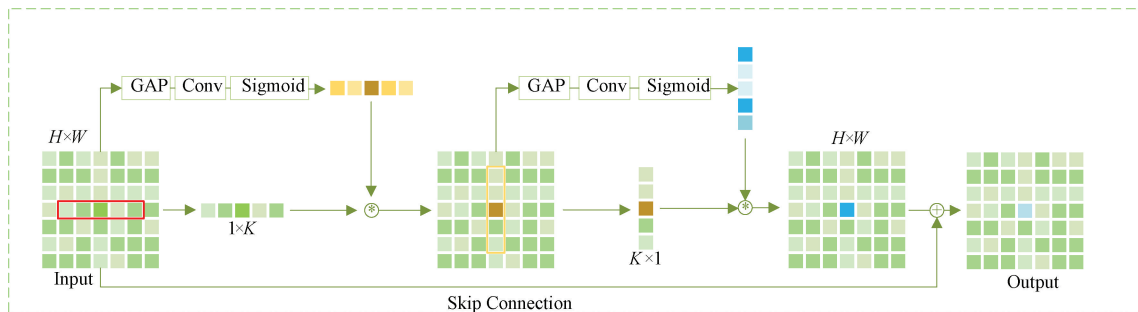
后通过 MFM 模块进行通道选择和信息强化,有效减少冗余信息、突出关键目标区域,从而增强融合特征的表达能力。同时,在多层融合路径中多次使用 C3k2_SSA 模块,使融合特征不断获得空间与语义的细化处理,增强了模型对不同尺度目标的鲁棒性,降低了复杂背景下的误检率。

此外,在 Head 检测头部分,模型延续了 YOLOv11n 原有的三尺度解耦检测头设计,分别面向小目标、中目标和大目标进行预测。不同于原始结构仅采用基础卷积融合,本文在每一尺度分支中均新增了一层 MFM 模块与一层 C3k2_SSA 模块。MFM 模块用于对特征图中的通道信息进行重新权重分配,突出目标区域特征,而 C3k2_SSA 模块则进一步细化空间建模能力,强化模型在复杂背景中的定位准确性。该结构的设计确保了最终输出特征图在进入 Detect 模块前,已具备充分的语义表达力与空间判别能力,从而有效提升了模型在密集目标、远距离目标和遮挡目标下的检测性能。

2.1 SSA 模块

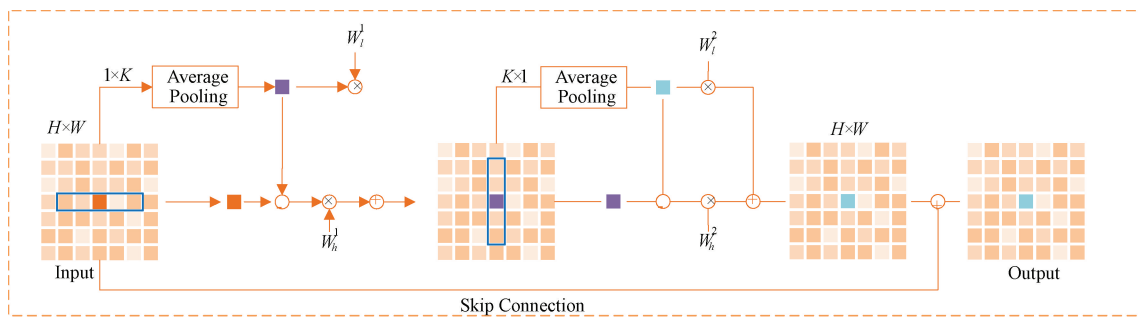
在目标检测等视觉任务中,如何有效提取图像的局部特征、提升特征表达能力,Cui等^[19]首次提出 SSA 和频域条带注意力(frequency strip attention, FSA)模块,用于空间域和频域特征的联合增强。SSA 通过空间条带注意力对特征在水平方向和垂直方向进行局部建模,而 FSA 则通过傅里叶变换或 DCT 等操作在频域提升细节恢复能力。基于任务特性,YOLOv11n-AFD 重点集成了 SSA 模块而舍弃 FSA,旨在兼顾特征增强与检测效率,如图 3 所示。

如图 3(a)所示, SSA 主要处理流程中, 首先输入特征 $F^H \in \mathbb{R}^{C \times H \times W}$ 被按行(水平条带)与按列(垂直条带)切分, 形成两个方向的条带块 $F^H \in \mathbb{R}^{C \times 1 \times H}$ 和 $F^V \in \mathbb{R}^{C \times H \times 1}$, 用于高效提取条带。然后水平条带经全局平均池化, 通道聚合后生成方向权重 $gH = \text{GAP}(F^H) \in \mathbb{R}^{C \times 1 \times W}$, 同理生成垂直方向权重 gV 。接着, 每个方向经过 1×1 卷积和 *sigmoid* 激活生成权重 $\alpha^H = \sigma(\text{Conv}1 \times 1(gH))$ 和 $\alpha^V = \sigma(\text{Conv}1 \times 1(gV))$ 。再方向权重与对应条带特征相乘 $\tilde{F}H = \alpha^H \otimes F^H$ 和 $\tilde{F}V = \alpha^V \otimes F^V$ 。最后增强后的条带特征重组为原尺寸, 与输入特征相加形成输出: $\hat{F} = F + \text{Fold}(\tilde{F}H) + \text{Fold}(\tilde{F}V)$ 。



(a) SSA结构

(a) Spatial strip attention structure



(b) FSA 结构
(b) Frequency strip attention structure

图 3 注意力结构

Fig. 3 Attention structure

图 3(b)展示了 FSA 的结构。它的主要不同在于操作域和条带加权策略,首先在频域条带处理上输入特征通过 DCT 等频域变换,在频域内按行、列分别平均池化,生成频域方向特征 $f^H = \text{AvgPool}_{1 \times K}(\text{DCT}(F))$, $f^V = \text{AvgPool}_{K \times 1}(\text{DCT}(F))$,然后在方向加权与重组上方向频域特征通过加减运算、权重乘积作用于频域特征 $\tilde{F}_{\text{freq}} = f^H \otimes W'_H + f^V \otimes W'_V$ 。最后逆变换恢复空间输出 $\hat{F} = F + \text{IDCT}(\tilde{F}_{\text{freq}})$ 。

从模型配置与框架结构上看,SSA 模块被深度集成于骨干网络与检测头多个关键层中,作为 C3k2_SSA 结构的核心构件,参与特征提取、融合和细粒度增强。根据框架图步骤,SSA 首次在骨干网络的低层特征提取阶段引入,用以对空间局部信息进行条带化编码,并增强方向敏感特征;随后在更深层继续叠加条带注意力,以在不同尺度特征上捕获丰富空间上下文。进入检测头阶段,SSA 模块在分支特征融合处发挥作用,与 MFM 模块协同提升多尺度特征的表征能力。

相比之下,FSA 模块虽在去雾等低级视觉任务中对频域细节补偿有良好效果,但在目标检测场景中引入频域建模可能引发计算开销显著增加,并带来冗余特征干扰检测主干的高效性。而 SSA 模块以空间域为主,通过条带状注意力分别在水平与垂直方向或两种方向上建模可由式(1)表示。

$$\text{SSA}(X) = \sigma(\text{Conv}_{\text{strip}}(X)) \otimes X \quad (1)$$

其中, $\text{Conv}_{\text{strip}}$ 表示条带卷积操作, σ 是激活函数, $\otimes$ 为逐元素乘积,在方向感知和局部依赖捕获上优势突出,并且因其设计更适合卷积型检测网络,在参数量与推理延时上更具优势。因此本文在 YOLOv11n 改进模型中选用 SSA 模块,通过高效空间特征增强,提升对小目标、遮挡目标和密集目标的检测性能。对此,SSA 模块的引入不仅在空间信息建模上优于单纯频域增强的 FSA,更在 YOLOv11n 检测场景中兼顾精度与效率,在骨干和头部网络的多层次、多尺度特征融合中体现其关键作用。

2.2 曼哈顿模块结构与差异增强机制

在深度神经网络中,传统的卷积运算倾向于融合邻域

信息以增强语义理解能力,但在处理空间细粒度变化或局部纹理差异时,往往会出现特征平滑、边界模糊等问题,导致模型在检测细小目标、边界目标或异质区域时表现不佳。为解决此问题,YOLOv11n-AFD 引入了一种基于空间差异感知机制的模块——曼哈顿模块^[20]。该模块核心思想是在特征层面度量局部差异,从而捕捉具有判别性的异质信息。模块结构如图 4 所示。

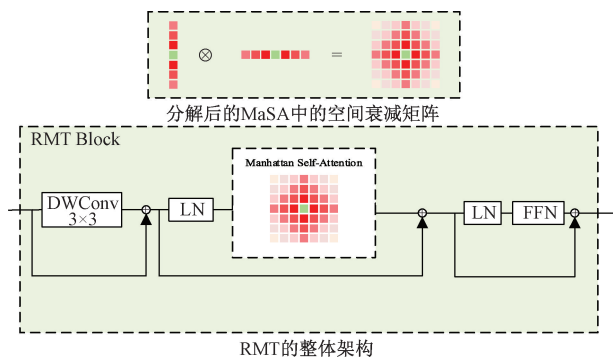


图 4 曼哈顿特征增强模块结构

Fig. 4 Manhattan feature enhancement module structure

具体而言,曼哈顿模块接受两个输入特征图 $X_1 \in \mathbb{R}^{C \times H \times W}$ 和 $X_2 \in \mathbb{R}^{C \times H \times W}$,这两个特征可以来自于不同尺度、不同路径或同一路径中不同阶段的输出。首先,模块对两者执行逐元素的绝对差运算,以提取局部差异信息并生成中间差异特征图 D : $D = |X_1 - X_2|$ 。

该操作是曼哈顿距离在局部空间内的一种具体实现。与欧氏距离相比,曼哈顿距离更强调方向性差异并避免了平方运算带来的梯度爆炸风险,因此更适合在浅层或低分辨率特征上进行局部响应增强。

为了进一步提升差异特征的语义表达能力,模块对差异图 D 进行轻量卷积处理。具体操作为一个 1×1 卷积层,随后连接归一化与 ReLU 激活函数,从而得到高维非线性表达特征: $F_D = \text{ReLU}(\text{BN}(\text{Conv}_{1 \times 1}(D)))$ 。

其中, 1×1 卷积的作用是进行通道混合和特征压缩,既保留了差异图中重要的空间信息,又提升了计算效率。

此外,为了在保持模型稳定性的同时融合差异信息,曼哈顿模块采用了残差连接策略,即将差异特征图 F_D 与输入特征之一 X_1 相加,得到最终输出特征图 $Y: Y = X_1 + F_D$ 。

这一结构既能保留原始输入的全局语义特征,又能引入局部增强信息,避免了直接替代可能带来的特征扰动,有效提升了训练收敛速度与模型表达能力。

2.3 MFM 模块

在现代图像目标检测等视觉任务中,如何高效融合不同来源的多尺度特征,一直是提升模型性能的关键问题。为此 Zhang 等^[21]提出了 MFM,该模块能够在融合多路特征时自适应生成权重,有效突出关键特征、抑制冗余信息。MFM 的整体结构清晰分为输入处理、全局感知、权重生成、特征调制及输出几个阶段,各步骤协同作用,共同完成高效融合,如图 5 所示。

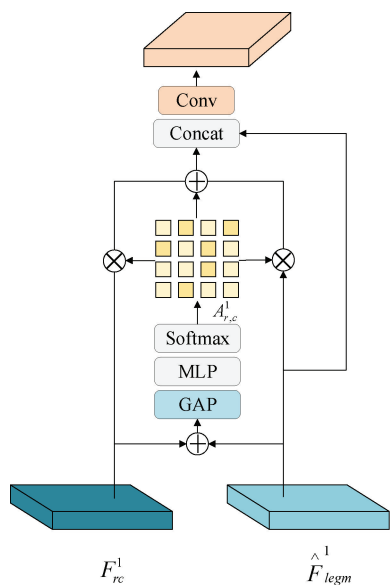


图 5 MFMM 结构

Fig. 5 MFM structure

在输入阶段, MFM 接收两路特征张量, 分别为重构特征 F_{rc}^1 与引导特征 $\hat{F}_{legm}^1$ (图 5 中左下方与右下方的输入块)。考虑到不同输入的通道维可能不一致, 模块通过 1×1 卷积(图 5 中 Conv 步骤)对输入特征进行升维或降维, 以保证后续拼接操作的维度一致性, 公式如式(2)、(3)所示。

$$F_{rc}^{1'} = Conv_{1 \times 1}(F_{rc}^1) \quad (2)$$

$$\hat{F}_{legm}^{1'} = Conv_{1 \times 1}(\hat{F}_{legm}^{1'}) \quad (3)$$

经过通道对齐后的特征在通道维度进行拼接(图 5 中

Concat 步骤), 得到融合输入 $F_{cat} = Concat(F'_{rc}, \hat{F}'_{legm})$ 。为了模拟 Transformer 的“多头”策略, MFM 将拼接后的特征划分成若干子空间(head), 以捕捉不同子空间的特征模式, 公式如式(4)所示。

$$F_{sb} = F_{cat} \cdot view(B, h, C', H, W) \quad (4)$$

其中, $h = 2$ 表示该模块划分成两个子空间, 分别对应不同特征来源或尺度。

在全局信息提取阶段(图 5 中 GAP 与 MLP 步骤),模块首先在子空间维数求和,公式如式(5)所示。

$$F_{sum} = \sum_{i=1}^h F_{sp,i} \quad (5)$$

其中, F_{sum} 指子空间在空间维度上的求和结果(融合后的特征), $F_{sp,i}$ 指第 i 个子空间特征。然后通过全局平均池化 (GAP) 对其压缩, 获得全局描述向量: $\mathbf{z} = \text{GAP}(F_{sum})$ 。

该向量作为全局感知的代表输入 MLP 网络,生成权重向量,最后通过 Softmax 在子空间维度归一化,以确保不同子空间的注意力权重之和为 1(图 5 中 Softmax 步骤)即 $W = \text{Softmax}_h(\text{MLP}(\mathbf{z})), \sum^h W_i = 1$ 。

进入特征调制阶段, MFM 将每个子空间特征与对应权重逐元素相乘(图 5 中两个乘号步骤), 并在子空间维度求和融合 $F_{fuse} = \sum_{i=1}^h W_i \cdot F_{sp,i}$; 该操作动态调整不同特征对最终融合结果的贡献, 突出关键特征、弱化冗余信息。

融合结果经过 1×1 卷积(图 5 中顶部 Conv 步骤), 输出最终特征 F_{out} 供后续模块使用 $F_{out} = Conv_{1 \times 1}(F_{fuse})$; 这种设计在 YOLOv11 中用以替代传统的 concat 操作, 相比简单拼接, MFMM 能够根据特征重要性进行有针对性的加权, 显著提升模型对目标的表达能力和检测精度。

3 实验结果与分析

3.1 实验环境

本文在 Windows11 操作系统下进行实验验证,CPU 型号英特尔 13 代 i5-13400F,显卡型号 NVIDIA RTX4060, CUDA 版本 11.8, Cudnn 版本 8.9.0, 编程语言 Python3.9,深度学习框架为 Pytorch2.3.1。

3.2 实验数据集

本文实验数据采用 VisDrone,是领先的无人机视觉基准数据集,旨在推动低空无人机平台上的目标检测、跟踪与识别等任务的发展。

该数据来源于各类无人机摄像头,覆盖了中国的 14 个不同城市,有不同环境和密度场景下的图像。包括了行人、人、自行车、汽车、面包车、卡车、三轮车、遮阳蓬三轮车、公交车及摩托车 10 个类别^[17]。按照训练集:验证集:测试集=7:1:2 的比例划分,数据集中有 6 040 张图片用于训练,863 张图片用于验证,1 726 张图片用于测试。

3.3 评价指标

为了全面评价改进后模型的性能,本文主要采用 3 种精度相关指标,即准确率(precision, P)、召回率(recall, R)、平均精度均值(mean average precision, mAP)。这些指标可从不同角度反映模型在目标检测任务中的表现。 P 和 R 计算公式如式(6)、(7)所示。

$$P = \frac{TP}{TP + FP} \tag{6}$$

$$R = \frac{TP}{TP + FN} \tag{7}$$

其中, TP 表示被正确检测为正例的样本数, FP 表示被错误检测为正例的样本数, FN 是未被检测出来的真实正样本数。较高的准确率说明模型在预测正样本时出错较少, 误检率低。在目标检测任务中, 准确率高有助于减少对无关目标或背景的误报。高召回率表示模型能够较好地发现数据集中存在的目标, 漏检较少, 对于实际应用中需要高覆盖率的场景尤为重要。

目标检测中 mAP 是最常见评价指标之一, 用于衡量模型在多个类别上的平均性能指标, 计算公式如式(8)、(9)所示。

$$mAP@0.5 = \frac{1}{m} \sum_{i=1}^m AP_i \tag{8}$$

$$mAP@0.5:0.95 = \frac{1}{10} \sum_{t=0.5}^m mAP_t \tag{9}$$

其中, m 是类别数, AP_i 是第 i 类的平均精度。平均精度($mAP@0.5$)交并比(intersection over union, IoU)阈值为 0.5 时, 通过计算每个类别精度-召回曲线下的面积(average precision, AP), 然后对所有类别取平均值得到。当 IoU 阈值为 0.5 时, 匹配标准相对宽松, 主要衡量检测框位置的基本准确性。严格平均精度($mAP@0.5:0.95$)进一步提高了评估难度, 它要求在 IoU 从 0.5~0.95 每隔 0.05 取一个阈值分别计算 AP , 然后求平均。该指标全面反映了检测框的精度及其对重叠度要求的鲁棒性, 是衡量目标检测算法质量的重要标准。

3.4 消融实验

本研究以 YOLOv11n 作为基线模型, 在 VisDrone 数据集上开展消融实验, 为验证各项改进的有效性, 本文设计了 8 组消融实验评估各改进模块对 YOLOv11n 的优化效果, 每组实验采用相同的训练策略, 实验结果如表 1 所示。

表 1 消融实验结果
Table 1 Ablation experiment results

方法	SSA	Manhattan	MFM	P	R	$mAP@0.5$	$mAP@0.5:0.95$
YOLOv11n				0.469	0.355	0.350	0.204
1	✓			0.468	0.361	0.357	0.210
2		✓		0.470	0.351	0.348	0.203
3			✓	0.475	0.364	0.362	0.211
4	✓		✓	0.492	0.354	0.363	0.213
5	✓	✓		0.470	0.360	0.358	0.209
6		✓	✓	0.469	0.360	0.355	0.208
7	✓	✓	✓	0.487	0.363	0.364	0.215

结果表明, 每个模块的改进均提升了模型的整体性能。在单模块引入中, SSA 模块有效增强了模型对空间条带特征的感知能力, P 基本持平, R 提升至 0.361, $mAP@0.5$ 提高至 0.357, $mAP@0.5:0.95$ 提升至 0.210, 表明其对召回率和边界框定位有正面效果。引入 Manhattan 模块后, P 提升至 0.470, R 略降至 0.351, $mAP@0.5$ 和 $mAP@0.5:0.95$ 基本持平于基线, 说明其在局部特征几何约束上起到一定作用但整体增益有限。引入 MFM 模块时, P 达到 0.475, R 提升至 0.364, $mAP@0.5$ 和 $mAP@0.5:0.95$ 分别提升至 0.362 和 0.211, 显示其在多模态特征融合中具备显著优势。

在双模块组合实验中, 不同模块间的协同进一步带来性能提升。首先, MFM+SSA 组合使得模型精确率提升至 0.492, R 维持在 0.354, $mAP@0.5$ 和 $mAP@0.5:0.95$ 分别达到 0.363 和 0.213, 显示出空间条带特征与多模态信息的协同融合有助于增强模型检测的综合能力。其次, SSA+Manhattan 组合下, P 和 R 分别为 0.470 和 0.360, $mAP@0.5$ 达到 0.358, $mAP@0.5:0.95$ 为

0.209, 表明空间信息与几何约束结合可在一定程度上提升小目标检测与遮挡条件下的表现。再次, Manhattan+MFM 组合时, P 为 0.469, R 为 0.360, $mAP@0.5$ 为 0.355, $mAP@0.5:0.95$ 为 0.208, 相比单独 MFM, 提升幅度较小, 显示两者间协同作用不及前两组明显。

最终, 在 MFM+SSA+Manhattan 三模块联合引入时, 模型综合能力达到最优, P 提升至 0.487, R 为 0.363, $mAP@0.5$ 和 $mAP@0.5:0.95$ 提高至 0.364 和 0.215。该结果充分表明三模块融合在空间条带信息、几何结构约束及多模态特征增强上形成互补, 有效改善目标检测的精度与稳健性, 特别是在复杂背景或小目标场景中表现更优。

3.5 不同算法对比试验

为全面评估所提出 YOLOv11n-AFD 模型在复杂交通场景中的检测性能, 本文在 VisDrone2019 数据集上选取了多种典型目标检测算法进行对比实验, 结果如表 2 所示。

从表 2 中可以看出, YOLO 系列模型整体在 $mAP@0.5$ 与 $mAP@0.5:0.95$ 指标上均优于传统检测器, 说明

表 2 在 VisDrone2019 数据集上不同模型对比试验
Table 2 Comparative experiments of different models
on the VisDrone2019 dataset

模型	$mAP@0.5$	$mAP@0.5:0.95$
SSD ^[22]	0.237	0.129
RetinaNet ^[23]	0.212	0.118
Faster-RCNN ^[24]	0.341	0.186
YOLOv3-tiny ^[25]	0.166	0.074
YOLOv5s	0.332	0.193
YOLOv6s	0.305	0.177
YOLOv8s	0.345	0.201
YOLOv10n ^[25]	0.339	0.199
YOLOv11s	0.351	0.205
YOLOv11n	0.350	0.204
MS-YOLO ^[26]	0.338	—
FMS-YOLOv10n ^[27]	0.362	—
YOLOv11n-AFD	0.364	0.215

YOLO 结构在复杂无人机航拍场景下具有更强特征表达与目标识别能力。YOLOv11 系列作为新一代检测模型在检测精度上表现出稳步提升。其中, YOLOv11n 模型在 $mAP@0.5$ 和 $mAP@0.5:0.95$ 上分别达到 0.350 与 0.204, 较前代 YOLOv10n 模型的 0.339 与 0.199 有一定程度提升, 表明 YOLOv11 的结构改进增强了模型的检测性能。



图 6 小目标车辆检测结果比较

Fig. 6 Comparison of small-target vehicle detection results

为进一步验证 YOLOv11n-AFD 模型在极端气象条件下的实用性与鲁棒性, 本文对 VisDrone 数据集样本进行降雨、下雪以及雾霾模拟数据增强, 并分别将 YOLOv11n 与 YOLOv11n-AFD 模型应用于同一组图像进行检测测试, 结果如图 7 所示。从图 7 中的降雨场景可知, YOLOv11n 模型虽能识别出部分车辆, 但在大雨干扰下出现了置信度显著下降的问题, 部分车辆的检测分值降至 0.60 以下, 甚至遗漏了边缘目标。而 YOLOv11n-AFD 模

在多种改进模型中, 本文提出的 YOLOv11n-AFD 在精度上取得了最优表现, 其 $mAP@0.5$ 和 $mAP@0.5:0.95$ 分别达到 0.364 与 0.215, 较原始 YOLOv11n 提升了约 4.0% 与 5.4%。相比之下, MS-YOLO 和 FMS-YOLOv10n 的 $mAP@0.5$ 分别为 0.338 与 0.362, 均低于 YOLOv11n-AFD 的 0.364; 而在更严格的 $mAP@0.5:0.95$ 指标上, YOLOv11n-AFD 同样保持领先。说明了本文改进在精度提升方面具有显著优势。

3.6 检测结果对比试验

为了验证本文提出的 YOLOv11n-AFD 算法有效性, 选取了典型场景图像, 对原始 YOLOv11n 模型与改进后模型进行比较, 结果如图 6 所示。如图 6(b) 所示, YOLOv11n 原始模型在面对遮挡严重、目标密集以及远处目标较小的城市交通环境时, 尽管能够基本框出车辆、人群与骑行者等目标, 但存在明显的置信度偏低、检测框定位偏差较大以及部分目标未检测出的情况, 尤其在远处密集行人区域和车辆出现较多漏检与误检。此外, 检测框的稳定性与判别性不够, 易被背景树木、建筑、天气等复杂环境干扰。相比之下, 图 6(c) 在目标定位精度与类别判别方面均有显著提升。首先, 模型对多数车辆的识别置信度明显提高, 如前景车辆的检测置信度普遍在 0.80 以上, 准确的框出了其边界, 显示出更强的特征提取能力; 其次, 对远处小尺度的目标也能实现稳定检测, 表现出良好的小目标感知能力; 同时, 检测框边界更加紧凑精准, 置信度分布更合理。

型在同样场景下展现出更强的鲁棒性, 对所有可见车辆均完成了正确检测, 且多数置信度高于 0.70, 可见其在条带遮挡与低对比度场景中仍具备良好的特征提取与结构感知能力。在图中的雾霾环境下, YOLOv11n 检测结果表现出边缘模糊、目标融合严重的问题, 多个车辆与大型车辆 (如公交、卡车) 被遗漏或误检, 部分目标置信度不足 0.30, 说明模型对低可见度环境下的特征辨识能力有限。相比之下, YOLOv11n-AFD 模型能稳定检测出大部分目标车

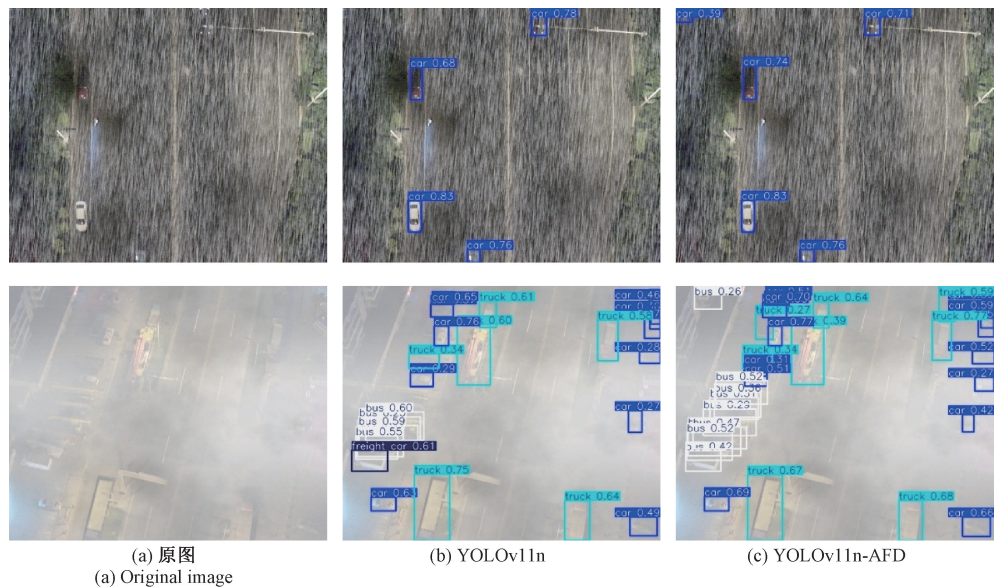


图 7 恶劣天气车辆检测结果比较

Fig. 7 Comparison of vehicle detection results in adverse weather conditions

辆,并对复杂结构具有更强的类别判别能力。检测框分布更合理,边界更紧凑,说明改进模块对特征提取和信息融合具有良好的增强效果。

此外,本文引入 Grad-CAM 可视化技术,对改进前后的模型在典型场景下的注意力响应区域进行对比分析,结果如图 8 所示。从图 8 中可以直观观察到,不同模型对输入图像中目标区域的关注程度存在显著差异。从城市交通的热力图对比可见,YOLOv11n 对部分近景目标具备一

定的追踪能力,但在远处行人、骑行者及车辆等小目标区域的激活响应较弱,热力图分布稀疏且响应强度不明显,说明其对小目标的捕捉能力有限。而改进后的 YOLOv11n-AFD 在同一场景下呈现出更密集的高响应区域,尤其在人群聚集与骑行者密集区域,显著提升了注意力集中度,说明其在复杂背景干扰下仍具备良好的小目标聚焦能力且在密集交通场景中具备更强的全局感知能力与小目标检测稳定性。

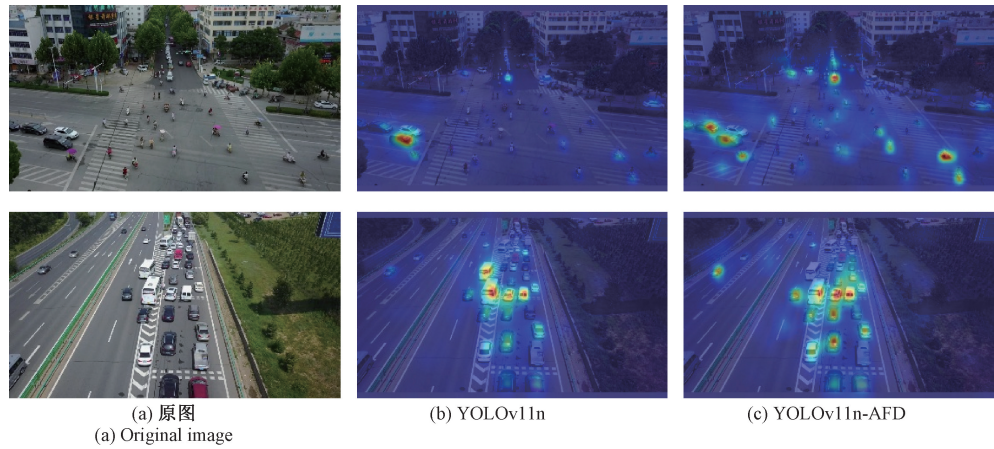


图 8 车辆可视化检测结果比较

Fig. 8 Comparison of visualized vehicle detection results

整体而言,改进后的模型在复杂城市道路场景中具备更高的目标检测准确性与鲁棒性,尤其在小目标与密集目标识别任务中表现优异,为实际应用提供了更具实效性的解决方案。

3.7 泛化实验对比

为验证本文改进方法的泛化性能与鲁棒性,选取

RSOD 遥感图像数据集开展模型对比实验。RSOD 数据集作为大规模光学遥感检测基准,涵盖飞机、油管、立架桥、操场。采用 7:1:2 的比例对数据集进行训练集、验证集、测试集划分,采用 P 、 R 、以及 $mAP@0.5$ 和 $mAP@0.5:0.95$ 进行对比分析,结果如表 3 所示。

实验结果显示,YOLOv11n-AFD 在指标上整体优于

表 3 泛化试验对比

Table 3 Generalization experiment comparison

模型	<i>P</i>	<i>R</i>	<i>mAP</i> @0.5	<i>mAP</i> @0.5:0.95
YOLOv11n	0.923	0.867	0.904	0.625
YOLOv11n-AFD	0.934	0.851	0.919	0.633

基础模型 YOLOv11n,其中精确率从 0.923 提升至 0.934,虽然召回率由 0.867 降至 0.851,但精度方面 *mAP*@0.5 提升了 1.5%;而在更严格的 *mAP*@0.5:0.95 指标上提升了 0.8%。从结果表明,YOLOv11n-AFD 显著增强了模型对小目标及复杂背景的识别能力,提升了检测的准确性和鲁棒性,从而实现了 RSOD 数据集目标的更优泛化性能。总体来看,YOLOv11n-AFD 的改进有效缓解了无人机航拍图像中目标尺寸小、遮挡严重等典型挑战,展现出更强的实用价值和推广潜力。

4 结 论

本研究提出了一种改进的 YOLOv11n 目标检测算法,首先,引入 SSA 模块,替代传统 C2f 中的 Bottleneck 结构,对小目标和遮挡目标具有显著提升的检测效果,以增强模型对空间关键特征的提取能力;其次,集成 MFM 模块,实现不同尺度和模态特征的高效融合,有效改善目标检测的精度与稳定性;此外,在保持模型轻量化的前提下,加入 Manhattan 模块引入几何先验和特征对齐机制,进一步提升边界框预测的准确度。实验结果表明,所提模型在 VisDrone 数据集上的 *mAP*@0.5 和 *mAP*@0.5:0.95 分别提升至 0.364 和 0.215,有效支持行人、车辆等多类别目标的精准识别,达到了无人机技术与智慧交通应用中对高精度、实时性目标检测的要求。与传统 YOLOv11n 模型相比,改进后的算法在参数量和计算量控制合理的基础上实现了检测精度的提升,能够更好地适用于复杂场景下的目标检测任务。该研究对智能交通、无人机监测等实际应用场景具有重要意义,并为高精度、低计算开销的目标检测模型设计提供了参考。

参考文献

[1] 郭润泽,孙备,孙晓永,等. 无人机弱光条件下多模态融合目标检测方法[J]. 仪器仪表学报, 2025, 46(1): 338-350.
GUO R Z, SUN B, SUN X Y, et al. Multimodal fusion object detection method for UAVs under low light conditions [J]. Chinese Journal of Scientific Instrument, 2025, 46(1): 338-350.

[2] MU AI M, WANG H J, MENG W J, et al. Small target detection in drone aerial images based on feature fusion[J]. Signal, Image and Video Processing, 2024, 18(Suppl 1): 585-598.

[3] ZHAO H J, WANG CH, LIN Y P, et al. On-road vehicle trajectory collection and scene-based lane change

analysis: Part I[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 18(1): 192-205.

[4] 李伟东,黄振柱,何精武,等. 改进行为克隆与 DDPG 的无人驾驶决策模型[J]. 计算机工程与应用, 2024, 60(14): 86-95.
LI W D, HUANG ZH ZH, HE J W, et al. Improved behavioral cloning and DDPG's driverless decision model[J]. Computer Engineering and Applications, 2024, 60(14): 86-95.

[5] 张浩晨,张竹林,史瑞岩,等. YOLO-CDC:优化改进 YOLOv8 的车辆目标检测算法[J]. 计算机工程与应用, 2025, 61(13): 124-137.
ZHANG H CH, ZHANG ZH L, SHI R Y, et al. YOLO-CDC: Improved YOLOv8 multi-scale vehicle object detection algorithm[J]. Computer Engineering and Applications, 2025, 61(13): 124-137.

[6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.

[7] REN SH 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, 2016, 39(6): 1137-1149.

[8] GIRSHICK R. Fast R-CNN[C]. IEEE International Conference on Computer Vision (ICCV), 2015: 1440-1448.

[9] 杨沙沙,徐新源. 基于 QARep-YOLOv8n 的城市物联网车辆目标检测方法[J]. 电子测量技术, 2025, 48(8): 46-54.
YANG SH SH, XU X Y. Urban internet of things vehicle object detection method based on QARep-YOLOv8n[J]. Electronic Measurement Technology, 2025, 48(8): 46-54.

[10] 郑玉珩,黄德启. 改进 MobileViT 与 YOLOv4 的轻量化车辆检测网络[J]. 电子测量技术, 2023, 46(2): 175-183.
ZHENG Y H, HUANG D Q. Lightweight vehicle detection network based on MobileViT and YOLOv4[J]. Electronic Measurement Technology, 2023, 46(2): 175-183.

[11] WANG CH Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.

[12] 刘玉蕾,褚丽莉,李波. 改进 YOLOv11 的模糊监控异常检测算法[J]. 电子测量技术, 2026, 49(3): 232-242.
LIU Y L, CHU L L, LI B. Optimized YOLOv11 for blurry surveillance anomaly detection[J]. Electronic Measurement Technology, 2026, 49(3): 232-242.

- [13] 何奕霏, 刘晓东, 赵兴, 等. 改进 YOLOv5s 的接触网小零件检测方法[J]. 电子测量与仪器学报, 2025, 39(9): 150-158.
HE Y F, LIU X D, ZHAO X, et al. Detection method for contact mesh small parts based on improved YOLOv5s [J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(9): 150-158.
- [14] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision. Cham: Springer International Publishing, 2016: 21-37.
- [15] 冉宁, 施高朗, 张少康, 等. 基于 YOLOv8 的遥感小目标检测算法[J]. 电子测量与仪器学报, 2025, 39(5): 197-207.
RAN N, SHI G L, ZHANG SH K, et al. Remote sensing small target detection algorithm based on YOLOv8[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(5): 197-207.
- [16] 李天林, 安毅, 陈岩. 基于边界特征与多尺度特征的车辆目标检测[J]. 电子测量技术, 2025, 48(16): 158-171.
LI T L, AN Y, CHEN Y. Vehicle detection based on boundary and multi-scale feature [J]. Electronic Measurement Technology, 2025, 48(16): 158-171.
- [17] 李珺, 丁彬彬, 史维娟, 等. 改进 YOLOv11 的无人机航拍图像检测算法[J]. 电子测量技术, 2025, 48(18): 111-121.
LI J, DING B B, SHI W J, et al. Improved UAV aerial image detection algorithm for YOLOv11 [J]. Electronic Measurement Technology, 2025, 48(18): 111-121.
- [18] 于承峰, 高松, 王鹏伟, 等. 改进的 YOLOv11 智能车辆动态环境目标检测算法[J]. 电子测量技术, 2025, 48(21): 55-66.
YU CH Y, GAO S, WANG P W, et al. Improved YOLOv11 intelligent vehicle dynamic environment object detection algorithm [J]. Electronic Measurement Technology, 2025, 48(21): 55-66.
- [19] CUI Y N, KNOLL A. Dual-domain strip attention for image restoration[J]. Neural Networks, 2024, 171: 429-439.
- [20] FAN Q H, HUANG H B, CHEN M R, et al. RMT: Retentive networks meet vision transformers [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 5641-5651.
- [21] ZHANG Y F, ZHOU S, LI H F. Depth information assisted collaborative mutual promotion network for single image dehazing[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 2846-2855.
- [22] SHI T, DING Y, ZHU W X. YOLOv5s_2E: Improved YOLOv5s for aerial small target detection [J]. IEEE Access, 2023, 11: 80479-80490.
- [23] YU W P, YANG T J N, CHEN CH. Towards resolving the challenge of long-tail distribution in UAV images for object detection [C]. IEEE/CVF Winter Conference on Applications of Computer Vision, 2021: 3258-3267.
- [24] 韦柳梅, 罗雪梅, 康健. 改进 YOLOv8 的航拍图像小目标检测算法[J]. 计算机工程与科学, 2025, 47(12): 2204-2215.
WEI L M, LUO X M, KANG J. Improved YOLOv8 small object detection algorithm in aerial image [J]. Computer Engineering & Science, 2025, 47(12): 2204-2215.
- [25] 刘臣杰, 刘巍, 杨雯迪, 等. DEPA-YOLO: 无人机视角下的小目标检测模型[J]. 计算机科学与探索, 2026, 20(1): 266-279.
LIU CH J, LIU W, YANG W D, et al. DEPA-YOLO: A drone-based small object detection model [J]. Journal of Frontiers of Computer Science and Technology, 2026, 20(1): 266-279.
- [26] 王春艳, 张成谦, 王祥, 等. 改进 YOLOv7-tiny 网络的多尺度无人机航拍小目标检测[J]. 测绘科学, 2023, 48(11): 189-199.
WANG CH Y, ZHANG CH Q, WANG X, et al. YOLOv7-tiny improved network multiscale UAV aerial photography small target detection incorporating swin-transformer [J]. Science of Surveying and Mapping, 2023, 48(11): 189-199.
- [27] 杨玉龙, 张晓玲. 改进 YOLOv10n 的无人机小目标检测算法[J/OL]. 计算机应用与软件, 1-9 [2025-09-01]. <https://link.cnki.net/urlid/31.1260.TP.20250714.1722.006>.
YANG Y L, ZHANG X L. Improved UAV small target detection algorithm for YOLOv10n [J/OL]. Computer Applications and Software, 1-9 [2025-09-01]. <https://link.cnki.net/urlid/31.1260.TP.20250714.1722.006>.

作者简介

陈磊, 硕士, 副教授, 主要研究方向为检测技术及智能装置、人工智能技术及应用、新能源发电及智能电网等。

E-mail: clei@ncst.edu.cn

邢智(通信作者), 硕士研究生, 主要研究方向为智能控制与模式识别。

E-mail: 15614887381@163.com

粘山坡, 硕士, 讲师, 主要研究方向为智能检测与控制技术等。

E-mail: shanpo1999@126.com

薄敬东, 硕士, 讲师, 主要研究方向为仪器与测试技术、自动控制技术等。

E-mail: 13930556261@163.com

常国涛, 硕士研究生, 主要研究方向为新能源发电与电能存储。

E-mail: 2480422631@qq.com