

基于改进 YOLOv10-MCS 多模态协同感知的轻量化负障碍物检测算法

范忆梅¹, 蒋存皓¹, 范贤杰², 伍彩霞³, 周力强³

(1. 福州理工学院 计算与信息科学学院, 福州 350506;

2. 鲁东大学 集成电路学院, 山东 烟台 264025; 3. 福州理工学院 电子工程学院, 福州 350506)

摘要: 针对复杂环境负障碍物检测存在的特征模糊性显著、多尺度漏检率高及环境干扰鲁棒性不足等问题, 提出基于 YOLOv10 框架的多模态协同感知改进模型 YOLOv10-MCS; 该算法通过感受野注意力卷积模块替换主干特征提取网络的传统卷积操作, 利用动态多分支感受野与空间注意力机制强化低对比度边缘特征提取; 构建上下文引导块完成全局语义与局部细节的自适应融合, 有效解决边界模糊导致负障碍物漏检的问题; 设计跨尺度特征融合模块重构颈部特征融合网络, 采用通道统一化策略与跨层级拼接机制, 在优化多尺度特征一致性的同时实现轻量化架构; 联合全局注意力机制的通道-空间重校准显著抑制背景干扰; 实验结果表明, YOLOv10-MCS 模型精确率达 88.13%, mAP 提升至 85.80%, 浮点运算次数降至 5.7 GFLOPs, 相较于原模型, 精确率提升 5.96%, 平均精度均值提升 3.3%, 计算量降低 32.1%; YOLOv10-MCS 通过跨模态特征交互机制攻克负障碍物检测难题, 其高精度轻量化特性为复杂场景目标检测提供了新的技术路径, 在自动驾驶感知系统与机器人动态避障场景中具备工程实践价值。

关键词: 负障碍物检测; YOLOv10; 跨尺度特征融合模块; 上下文引导块; 全局注意力机制; 感受野注意力卷积模块

YOLOv10-MCS: A Lightweight Negative Obstacle Detection Algorithm Based on Multimodal Collaborative Sensing

FAN Yimei¹, JIANG Cunhao¹, FAN Xianjie², WU Caixia³, ZHOU Liqiang³

(1. School of Computing and Information Science, Fuzhou Institute of Technology, Fuzhou 350506, China;

2. College of Integrated Circuits, Ludong University, Yantai 264025, China;

3. School of Electronic Engineering, Fuzhou Institute of Technology, Fuzhou 350506, China)

Abstract: To address the challenges of significant feature ambiguity, high multi-scale missed detection rates, and insufficient environmental interference robustness in negative obstacle detection under complex environments, an improved multimodal collaborative sensing model named YOLOv10-MCS is proposed based on the YOLOv10 framework. In the backbone network, the receptive-field attention convolution (RFACnv) module replaces conventional convolution operations, leverages dynamic multi-branch receptive fields and spatial attention mechanisms to enhance low-contrast edge feature extraction. The context guided block (CGB) enables adaptive fusion of global semantics and local details, effectively resolving boundary ambiguity-induced missed detections. The cross-scale feature fusion Module (CCFM) reconstructs the neck network using channel normalization and cross-layer concatenation, optimizing multi-scale feature consistency while achieving lightweight design. Integrated channel-spatial recalibration via Global attention mechanism (GAM) significantly suppresses background interference. Experimental result shows that the YOLOv10-MCS model achieves 88.13% precision, 85.80% mean average precision (mAP), and 5.7 GFLOPs computational cost. Compared to the benchmark model, these metrics represent a 5.96% precision improvement, 3.3% mAP gain, and 32.1% computation reduction. YOLOv10-MCS establishes a new technical framework for complex scene object detection through cross-modal feature interaction. The proposed framework with high-precision lightweight architecture demonstrates deployment potential in autonomous driving perception systems and

收稿日期: 2025-05-15; 修回日期: 2025-05-30。

基金项目: 福建省中青年教育科研一般项目(JAT241223); 福州理工学院校级科研基金重点项目(FTKY2024012)。

作者简介: 范忆梅(1998-), 女, 硕士, 助教。

引用格式: 范忆梅, 蒋存皓, 范贤杰, 等. 基于改进 YOLOv10-MCS 多模态协同感知的轻量化负障碍物检测算法[J]. 计算机测量与控制, 2025, 33(6): 288-297, 320.

robotic systems for dynamic obstacle avoidance.

Keywords: negative obstacle detection; YOLOv10; CCFM; CGB; GAM; RFACnv

0 引言

随着智能驾驶辅助系统(ADAS, advanced driver assistance systems)的广泛应用,道路环境感知已成为自动驾驶技术的关键性挑战。其中,路面异常检测作为环境感知的核心环节,直接关系到车辆的路径规划与决策控制能力。当前研究主要聚焦于正向障碍物(如车辆、行人)与道路标识的检测,相关技术已趋于成熟。然而,针对路面凹陷缺陷(即负向障碍物,NOBs, negative obstacles)的检测仍存在技术瓶颈。负障碍物的产生与道路长期损耗密切相关,高频次的交通荷载导致材料疲劳累积,加之雨水渗透加速沥青层结构劣化,促使局部坑洼区域扩展。在雨雪等极端天气下,路面湿滑特性与负向障碍物形成耦合风险,而驾驶者需集中精力观察周围车辆,难以准确评估坑洼危害程度,从而增加交通事故风险并造成车辆损坏。传统人工检测方法依赖主观经验,效率低、成本高且难以满足实际需求。因此,开发鲁棒的负向障碍物检测方法对完善自动驾驶安全冗余机制具有重要现实意义。

传统路面负障碍物检测长期依赖人工目视检查,存在检测效率低、结果可重复性差等固有缺陷。随着智能检测技术的发展,当前主流研究方法可分为:基于振动的检测方法、基于三维重建的检测方法和基于计算机视觉的检测方法^[1]。其中基于振动的检测方法通常包括三阶段处理流程:传感器部署与信号捕获、噪声抑制与频域转换预处理、时频域特征驱动的统计学习算法。研究实践表明,不同学者在传感装置选型上呈现显著差异性,文献[2]提出基于物联网加速度计与无监督阈值分析的坑洞检测系统,通过无人地面车辆实时采集振动信号,实验验证其可精准辨识多尺度负障碍物并量化几何参数;文献[3]采用智能手机单轴加速度传感器实现基础振动感知,文献[4]融合加速度计-陀螺仪进行多维运动状态解析,文献[5]则集成加速度计与GPS构建多源异构传感系统,通过时频域联合特征工程优化机器学习分类性能。然而,该方法受传感器和车辆的影响导致负障碍物几何形态检测精确方面有局限性,且易受复杂环境振动噪声干扰。

基于三维重建的检测方法主要依托多模态传感范式实现,涵盖专用3D传感器与2D传感器构建三维信息的技术路径。文献[6]基于3D点云数据开发双尺度聚类算法,通过中值滤波-阈值分割生成二值化裂缝特征,根据距离和角度两个指标实现裂缝形态参数化建模;文献[7]提出基于移动车辆立体视觉的坑洞测量方法,通过关键点匹配与块匹配结合的快速立体匹配算

法处理运动模糊图像,并开发面积-深度-体积多参数测量模型;文献[8]设计多视图立体成像架构,集成深度可分离卷积改进U-Net网络,在毫米级重建精度的三维路面数据集上完成像素级破损分割与体积测算。此类方法在负障碍物几何形态解析精度方面具有显著优势,但受限于多传感器时空同步误差与设备标定复杂度,且高精度3D传感硬件成本限制了大规模部署应用。

基于视觉的检测方法使用图像或视频作为输入,并应用图像处理和深度学习技术来确定路面上是否存在负障碍物。文献[9]构建边缘检测-YOLO融合架构,采用灰度转换与背景掩膜预处理抑制多目标干扰,结合轮廓特征增强策略提升复杂场景检测鲁棒性;文献[10]系统评估YOLO系列^[11-14]模型在坑洞检测中的性能,基于665幅高分辨率道路图像验证YOLOv4-tiny以78.7%的*mAP*实现精度与实时性最优平衡;文献[15]改进YOLOv3模型架构,通过ResNet101-CIoU联合优化与对抗样本训练,在含水坑洞细分场景下保持81.2%抗干扰检测精度;文献[16]创新设计双阶段位置感知网络,通过高召回区域定位与部件分类协同机制,在公共数据集实现95.2%精确率突破;文献[17]研发预池化卷积神经网络,在首卷积层前增设图像优化模块,基于9.6万张多环境数据验证98.95%识别精度与跨材质泛化能力。相较于三维重建方法,视觉检测技术具有显著成本优势与形态解析普适性,但存在几何参数量化精度不足的局限。基于目前的研究与挑战,本文提出YOLOv10-MCS改进模型,通过多尺度特征融合机制与自适应注意力模块协同优化,突破复杂场景下负障碍物精准定位与形态解析的技术瓶颈,为道路智能养护决策与自动驾驶前瞻感知提供可靠技术路径。

1 YOLOv10 算法原理

YOLOv10(You Only Look Once version 10)是目标检测领域中的一种高效算法,该版本基于YOLO系列的实时检测框架,进一步优化了检测速度和精度^[18]。YOLOv10算法的主干特征提取网络中保留了YOLOv8与YOLOv9的轻量化和高效的特征提取理念,继续使用网络中的C2f和SPPF结构,并在此基础添加了空间-通道解耦下采样模块(SCDown, spatial-channel decoupled downsampling module)、基于秩的块设计(C2fCIB)以及部分自注意力机制(PSA, position-wise spatial attention)3个模块。其中SCDown模块将点卷积与深度卷积串联,实现空间与通道的分离下采样,解决了传统下采样操作中空间压缩与通道扩展耦合而引发

的计算冗余问题；C2fCIB 模块采用倒置的拓扑结构将深度可分离卷积前置，同时引入动态通道压缩机制，通过学习门控系数控制 1×1 卷积的通道缩减率；PSA 模块通过特征通道分割与轻量化注意力设计实现了全局建模能力与计算效率的平衡，有效解决了传统自注意力机制计算复杂度过高的问题。YOLOv10 算法在颈部特征融合网络中采用了路径聚合网络（PAN，path aggregation network）聚合不同尺度的特征，并用于改进传统特征金字塔（FPN，feature pyramid network），使模型能够更加高效的处理不同的目标，实现多尺度融合。YOLOv10 的检测头部预测网络是目标检测模型的最后一个模块，其根据主干特征提取网络和颈部特征融合网络提供的特征信息进行预测。在 YOLOv10 检测头包含了一对一预测头（One-to-One Head）和一对多预测头（One-to-Many Head），其中一对一预测头使用在模型的

推理过程中，为每个对象生成单一最佳预测结果，并且不需要使用非极大值抑制（NMS），有效的减少延迟，从而提升推理效率。一对多预测头应用在模型的训练过程中，为每个对象生成多个预测结果，极大程度提高丰富的监督信息，有效提升学习精度。YOLOv10 算法的整体结构如图 1 所示。

2 基于 YOLOv10 的多模态特征融合算法

2.1 感受场注意力卷积模块

在当前的研究中，空间注意力机制已经被广泛应用于提高卷积神经网络的性能，然而仍然存在一定的局限性，即空间注意力机制本质上解决了卷积核参数共享的问题，但是对于大尺寸卷积核而言，空间注意力机制生成的特征图中仍有部分信息缺失，因此文献 [19] 提出了感受场注意力卷积模块（RFACConv，receptive-filed

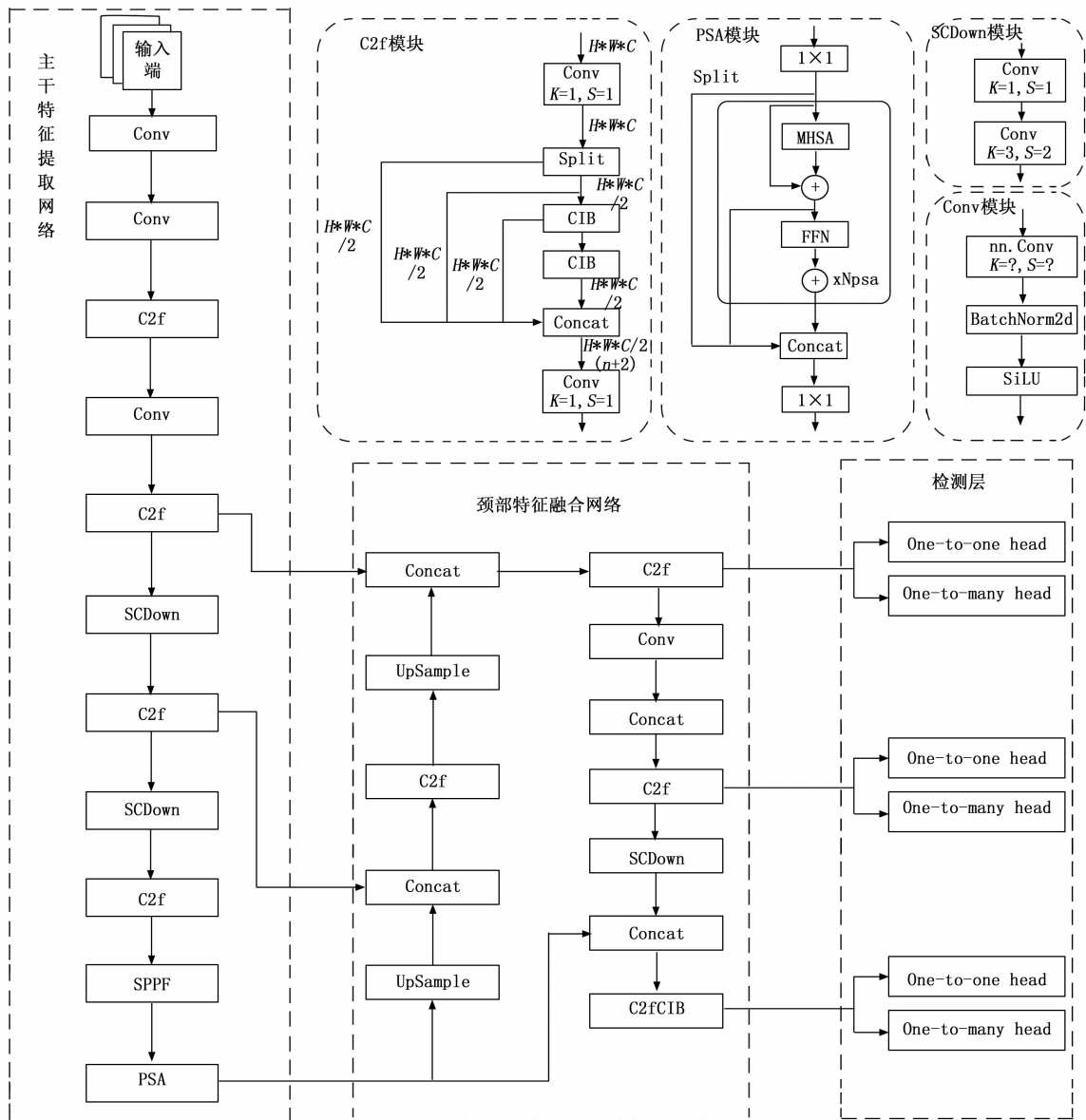


图 1 YOLOv10 算法原理图

attention convolution), 该模块将空间注意力机制与卷积操作相结合, 优化卷积核的工作方式, 提升卷积操作的性能。RFACnv 的卷积操作如图 2 所示, 模块重点关注感受野空间特征, 使用与卷积核 (即图中的 K_j) 同样大小的感受野进行处理, 并引入注意力机制 (即图中的 A_{ij}), 使网络为每个感受野生成特定的权重, 将感受野与经过注意力机制指导后的卷积核相乘求和便能得到特征图, 该方式能使网络更高效的学习图像中的局部特征并有效解决参数共享问题, 有效的提升网络性能。感受场注意力卷积模块即使使用大尺寸卷积核, 也能根据感受野注意力机制动态调整卷积核的权重, 尽可能降低系统资源的消耗。

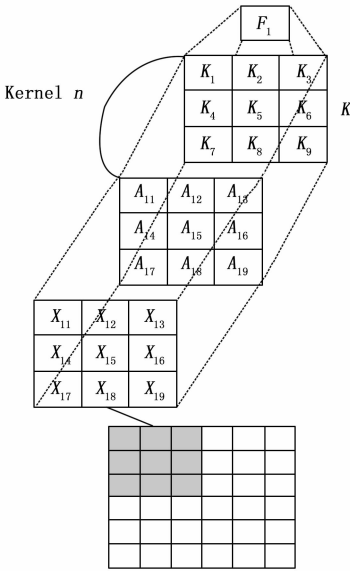


图 2 RFACnv 卷积操作过程图

2.2 上下文引导块

上下文引导块 (CGB, context guided block) 是由 Wu 等人^[20]在 CGNet 中首次提出, 其设计灵感源自人类视觉系统利用上下文信息解析复杂场景的认知机制。其结构如图 3 所示, CGB 用于由局部特征提取器 (FLOC, local feature extractor)、周围上下文提取器 (FSUR, surrounding context extractor)、联合特征提取器 (FJOI, joint feature extractor)、全局上下文提取器 (fglo, global context extractor) 组成。其中局部特征提取器 floc 设计了一个 3×3 的卷积层用于学习输入数据的局部特征, 为后续跨区域关联提供基础表征。周围上下位提取器 fsur 通过扩展卷积 (DConv, dilated convolution) 扩大感受野, 从更广阔的邻域范围聚合多尺度上下文特征。联合特征提取器 fjoie 对局部特征与周围上下文特征进行通道拼接, 经归一化

(BN, batch normalization) 和参数化修正线性单元 (PReLU, Parametric ReLU) 实现特征融合与非线性增强, 从而平衡局部精细表征与全局语义关联。全局上下文提取器 fglo 使用全局平均池化 (GAP, global average pooling) 压缩空间维度生成全局上下文描述符, 并借助两个全连接层 (FC, fully connected layer) 建模通道间长期依赖关系, 以提升模型对图像整体语义的感知能力。

在信息处理流程中, 局部特征提取器输出的高分辨率细节特征首先与周围上下文提取器提供的宽域语义特征进行互补性融合, 随后联合特征提取器通过归一化与自适应激活进一步优化特征分布。最终, 全局上下文提取器将整合后的特征映射至高层语义空间, 形成从局部到全局的层次化场景理解。该方法通过多粒度上下文引导机制, 显著增强了网络在复杂场景下的鲁棒性与判别性。

2.3 跨尺度特征融合模块

跨尺度特征融合模块 (CCFM, cross-scale feature fusion module) 是 RT-DETR 框架的核心创新组件, 可用于改进 YOLOv10 的颈部特征融合网络。如图 4 所示, 该模块以主干特征提取网络最后输出的三级多尺度特征图 (高、中、低分辨率) 作为输入, 采用双向跨尺度融合机制实现局部细节与全局语义的协同增强。针对高分辨率特征图, 通过轻量化卷积保留边缘、纹理等细粒度空间信息, 便于小目标检测。针对低分辨率特征图进行上下文编码, 提取目标的类别判别性特征。最后通过通道拼接 (C, concatenation) 与 1×1 卷积压缩实现多尺度特征融合, 在降低计算复杂度的同时构建层次化特征金字塔, 融合后的特征图经通道对齐后输入检测头。

2.4 全局注意力机制

针对传统注意力机制 (如 CBAM^[21]、SENet^[22] 等) 在跨维度全局交互建模上的局限性, 全局注意力机制 (GAM, global attention mechanism) 采用了顺序通道-空间注意力机制, 有效捕获全局上下文信息并保留多尺度特征完整性。GAM 模块包含通道注意力子模块 (Channel Attention) 和空间注意力子模块 (Spatial Attention), 其结构如图 5 所示。

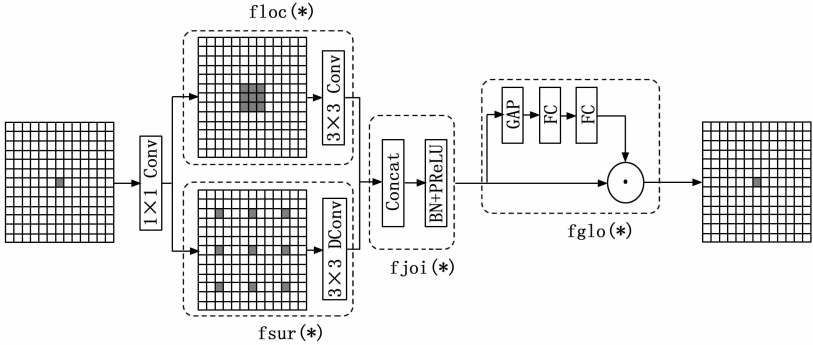


图 3 CG Block 结构图

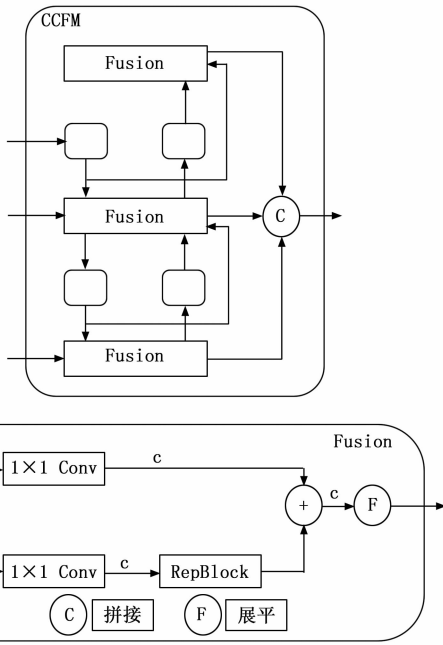


图 4 CCFM 结构图

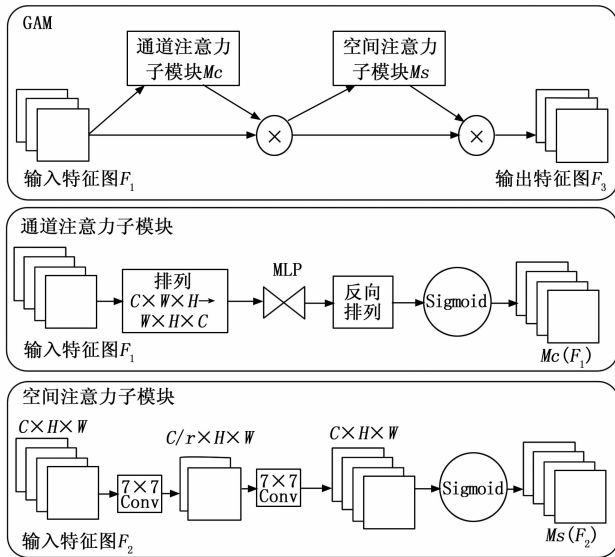


图 5 GAM 结构图

1) 通道注意力子模块:

输入的特征图 $F_1 \in R^{C \times W \times H}$ 首先通过维度重排策略进行跨维度交互建模, 将特征维度从通道 $\times$ 宽度 $\times$ 高度 ($C \times W \times H$) 重塑为宽度 $\times$ 高度 $\times$ 通道 ($W \times H \times C$), 建立空间维度与通道维度的显式关联。接着采用一个双层的多层感知机 (MLP, multilayer perceptron) 对跨维度依赖关系进行非线性映射, 通过式 (1) 实现通道—空间联合注意力学习最后反向排列恢复原始维度, 输出通道增强特征图 F_2 。该设计突破传统通道注意力仅依赖全局池化的统计表征限制, 通过维度置换与多层感知机实现通道与空间维度的双向信息交互。

$$Mc(F_1) = \sigma(MLP(Reshape(F_1))) \quad (1)$$

2) 空间注意力子模块:

为增强细粒度空间定位能力, 空间注意力子模块中使用两个卷积层融合空间信息, 同时为了保留更多的特征图信息, 在该模块中去除了池化操作, 将卷积后的结果同样使用激活函数 Sigmoid 处理最终得到特征图 F_3 。该方式摒弃池化操作, 采用全分辨率卷积能有效保留边缘、纹理等细粒度信息。

2.5 多模态协同感知算法 YOLOv10-MCS

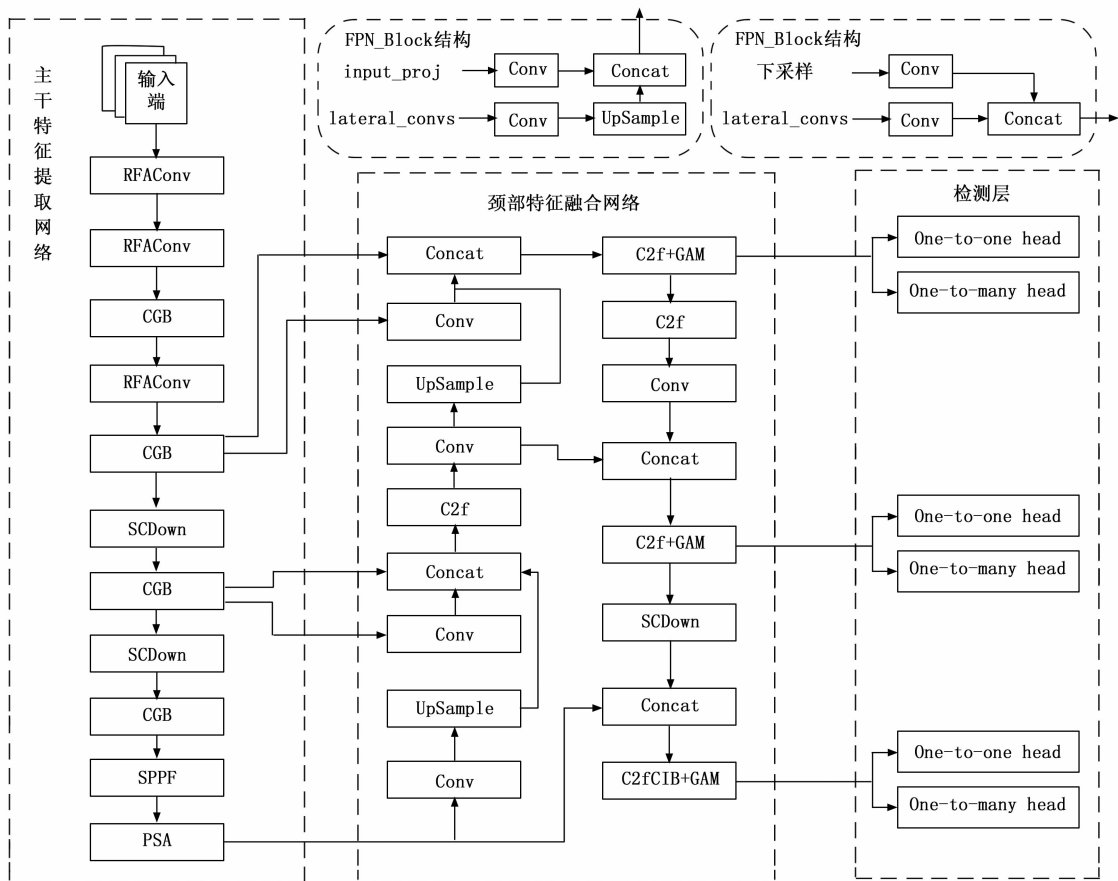
针对负障碍物检测任务中存在的特征模糊性、尺度敏感性及环境干扰等问题, 本文提出基于多模态协同感知的 YOLOv10-MCS 网络架构。该模型结构具体设计如图 6 所示, 通过引入感受野注意力卷积模块、上下文引导块、跨尺度特征融合模块及全局注意力机制, 实现了对负障碍物特征的深度挖掘与高效融合, 详细改进如下:

1) RFACnv 重构主干特征提取网络:

为提升负障碍物边缘特征的提取能力, 将传统卷积操作替换为感受野注意力卷积模块 (RFACnv)。该模块通过融合多分支感受野与空间注意力机制, 实现自适应特征增强。具体操作为: 在主干特征提取网络构建并行的多分支卷积核 ($3 \times 3/5 \times 5$ 等不同尺度), 在每个感受野分支内部嵌入空间注意力单元, 使网络能够依据局部上下文信息动态调整各尺度特征的贡献权重; 然后通过跨感受野的注意力竞争机制, 优先激活包含显著边缘梯度变化的特征区域。相较于传统空间注意力模块 (如 CBAM 仅进行全局特征加权), RFACnv 的创新优势在于将注意力计算与多尺度特征提取过程深度融合, 在卷积操作的底层实现感受野选择与空间聚焦的联合优化, 从而精准强化负障碍物与地面交界处的低对比度边缘响应, 同时抑制背景噪声干扰。

2) CGB 替换 C2f 模块:

为强化负障碍物与场景的上下文关联性, 采用上下文引导块 CGB 替代原始 C2f 结构。CGB 通过双路径特征互补机制与动态权重融合策略实现多尺度特征协同, 具体操作为: 首先主路径采用级联空洞卷积 (扩张率分别为 $2/4/6$) 构建多级感受野, 在保持特征分辨率的同时捕获跨区域的全局语义关联; 其次辅助路径设计为深度可分离卷积与局部注意力单元的级联结构, 针对边缘模糊区域实施高密度特征采样; 最后引入基于通道敏感度的自适应融合门控, 通过可学习参数动态调节两路径的贡献权重。相较于传统多尺度模块 (ASPP 依赖固定扩张率、Inception 堆叠静态分支等), CGB 的核心优势在于将全局语义推理与局部特征增强解耦为独立路径, 通过门控机制实现场景自适应的跨尺度特征合成, 从而在负障碍物边界模糊区域形成梯度增强效应, 有效解决



传统结构因尺度单一导致的边缘弥散问题。

3) CCFM 重构颈部特征融合网络:

为解决多尺度特征融合中的信息衰减问题,采用跨尺度特征融合模块 (CCFM) 重构颈部特征融合网络。该模块采用双向特征传播机制,通过自上而下的语义引导和自下而上的细节补偿构建特征融合通路。具体操作为:将传统金字塔结构解耦为双分支架构,分别为 FPN_Block 结构和 PAN_Block 结构,其中 FPN_Block 结构通过通道规范化策略将深层特征图的通道数统一映射至 256 维,并采用上采样操作提升空间分辨率;利用卷积层对主干网络提取的浅层特征进行非线性投影,消除层级间特征分布差异,最终通过跨层级特征拼接实现深层语义信息与浅层细节信息的高效融合。PAN_Block 结构基于横向连接机制构建逆向特征传播路径,通过卷积操作调整特征通道数为 256 维,并采用下采样操作降低空间分辨率;通过特征拼接将低分辨率特征与通道调整后的特征进行多尺度关联,进一步结合 C2f 结构与 GAM 优化特征表达,增强模型对多尺度目标的捕捉能力。CCFM 模块通过统一特征通道维度与动态分辨率调控,在提升多尺度特征一致性的同时实现模型轻量化设计,有效平衡了检测精度与计算效率的协同优化需求。

4) GAM 引导多维度特征增强:

在颈部特征网络输出端引入 GAM，通过级联通道—空间注意力实现特征多维重校准，具体操作为：构建通道—空间注意力级联架构，首阶段采用门控式通道筛选，通过时间步记忆机制保留历史通道权重信息，抑制持续性背景干扰通道；在空间注意力阶段设计异构卷积组（ 3×3 与 7×7 并行卷积核），通过宽窄感受野互补捕获多粒度空间关联，优先强化离散目标的边缘连续性特征；最后引入残差注意力掩码约束，确保特征重校准过程中主体语义的完整性。相较于传统双注意力模块（如 CBAM 的并行独立计算模式），GAM 的核心优势在于通过维度递进式校准机制，在通道维度实施粗粒度特征过滤后，在空间维度进行细粒度特征聚焦，形成“通道净化—空间锐化”的级联优化链路，从而在复杂道路场景中有效区分高相似性背景，有效降低误检率。

3 实验结果与分析

3.1 实验环境及参数配置

本文实验所使用服务器系统为 64 位的 Win11 专业工作站版, CPU 使用的是 12th Gen Intel (R) Core (TM) i9-12900K, 内存为 128 G, GPU 选用的是 NVIDIA GeForce RTX 3090, 实验环境配置详情如表 1 所示。

表 1 实验环境配置相关参数

名称	具体信息
操作系统	Win11 专业工作站版 64 位
CPU	12th Gen Intel Core i9-12900K
GPU	NVIDIA GeForce RTX 3090 * 1
内存	128 GB
Python	3.9.7
Pytorch	2.1.2
CUDA	11.8

为验证 YOLOv10-MCS 负障碍物检测模型的有效性，在模型训练的超参数方面，本文设置训练轮次为 400 轮，批处理大小 8，并行进程数 8，使用随机梯度下降（SGD，stochastic gradient descent）作为优化器，初始学习率在 $1 \times 10^{-5} \sim 1 \times 10^{-1}$ 之间，学习变化率在 0.01~1 之间，部分参数设置如表 2 所示。

表 2 实验超参数设置

超参数	值
训练轮次	400
优化器	SGD
批次大小	8
并行进程数	8
初始学习率	0.01
学习变化率	0.01
框损失增益	7.5
类损失增益	0.5
DFL 损失增益	1.5
动量	0.937

3.2 数据集构建

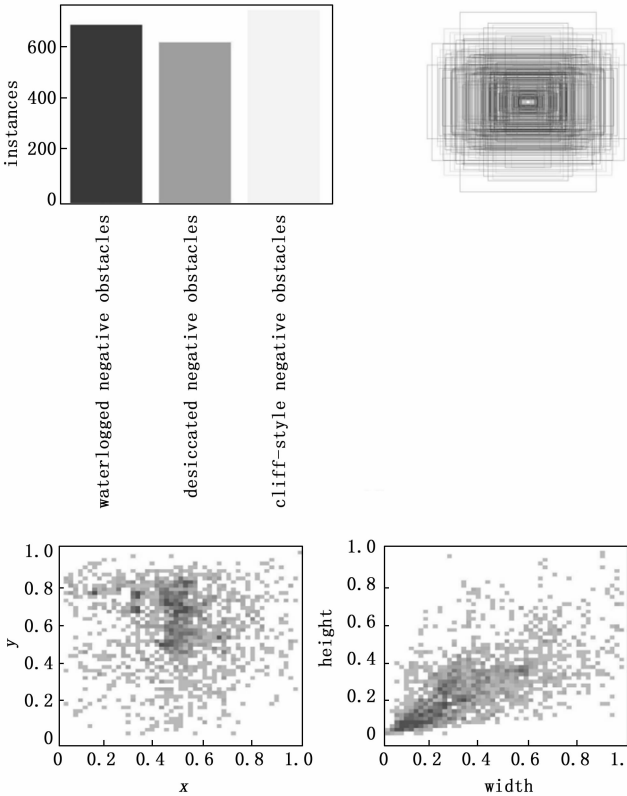
本文采用多源异构数据融合策略构建负障碍物检测专用数据集。数据来源涵盖自动驾驶开源数据集、网络开源平台样本及自主采集设备获取的道路场景图像，重点筛选包含积水型负障碍物（Waterlogged Negative Obstacles）、干涸型负障碍物（Desiccated Negative Obstacles）及断崖型负障碍物（Cliff-Style Negative Obstacles）三类典型负障碍物的有效样本。通过 Make Sense 平台对所有图像进行矩形框标注，明确标定负障碍物区域边界，生成标准化 YOLO 格式文本标签，最终构建包含 2 000 张高分辨率图像的数据集，标注过程及最终标注得到的各类型障碍物数量如图 7 所示。在 YOLOv10 模型训练过程中，生成了对训练数据集目标边界框（Bounding Box）的统计分析图，如图 7（b）所示。其中关键的一幅图展示了归一化处理后的目标空间分布特性：图 7（b）左下图表示目标中心点在图像宽度（X 轴）和高度（Y 轴）方向上的归一化坐标分布直方图；图 7（b）右下图表示目标边界框的归一化宽度与高度比例的分布直方图。该图直观反映了训练集中目标在图像平面上的位置偏好以及目标尺寸的宽高比特

征，为评估数据集标注质量和模型潜在学习难点提供了重要参考。

本文将其命名为 Triple Hazardous Negative Obstacles Dataset（TriHazard-NOD），按 7：1：2 比例划分为训练集、验证集与测试集，典型样本示例如图 8 所示。



(a) 数据集标注过程



(b) 各类型障碍物分布图

图 7 数据集标注过程与类型分布图

3.3 评价指标

本文使用精确率（P，precision）、平均精度均值（mAP，mean average precision）以及浮点运算次数（FLOPs）对模型进行评价，其中：

精确率指标用来表示在负障碍物检测模型预测的多种目标中，将障碍物类别预测准确的比例，计算公式如（2）所示：

$$Precision = \frac{TP}{TP + FP}$$
 (2)

式中，TP 为 True Positive，表示在当前模型中正确识别负障碍物类型的预测框数量，FP 为 False Positive，

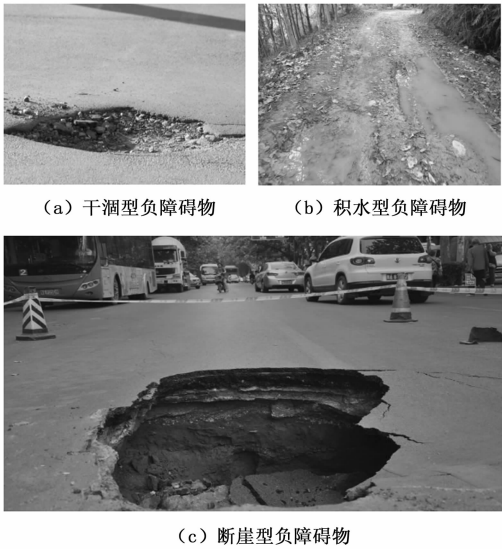


图 8 典型样本示例图

表示将其他类型的障碍物识别为负障碍物的标签框数量,即预测错误的情况。该评价指标值越高, FP 值越小,说明模型能将路面的负障碍物识别出来并且较少出现误判。

平均精度均值 (mAP) 指标用来衡量模型在多个类别检测中的综合能力。该指标是在计算之前需要先计算单个类别不同召回率 ($Recall$) 下的平均精确率,即精确率-召回率 ($P-R$, precision-recall) 曲线下的面积。计算公式分别为:

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$AP = \int_0^1 p(r) dr \tag{4}$$

$$mAP = \frac{\sum_{i=1}^k AP_i}{k} \tag{5}$$

式中, FN 为 False Negative, 表示当前模型未能正确识别出存在的负障碍物的标签框数量, AP 表示数据集中单一类别的精度值, 该值的取值范围为 $[0, 1]$, mAP 则对所有类别的 AP 取平均值, 其中 k 表示类别的数量, 该指标可以取值高说明模型在多个类别和不同的阈值下均有较好的表现, 综合能力较强, 在目标检测中, 该指标作为评价模型的一个重要指标。

浮点运算次数 ($FLOPs$, floating-point operations per second) 指标表示模型完成一次推理所需要的浮点计算次数, 该指标可以用来评估模型的计算复杂程度, 计算公式如 (6) 所示:

$$FLOPs = 2 \times K_H \times K_W \times C_{in} \times O_H \times O_W \tag{6}$$

式中, K_H 和 K_W 分别为卷积核的高度和宽度, C_{in} 为输入的通道数, C_{out} 为输出的通道数, O_H 和 O_W 分别为输出的特征图的高度和宽度。该指标的值低说明模型适用

于移动端且适合实时检测。

3.4 消融实验

为验证各改进模块的有效性及其协同作用, 本文设计渐进式消融实验, 通过控制变量法评估 RFACnv、CGB、GAM 及 CCFM 模块对检测性能的独立贡献与组合效应。实验在自制数据集 TriHazard-NOD 上进行, 基准模型 (实验 1) 采用原始 YOLOv10 架构, 实验 2-4 依次单独添加 RFACnv、CGB、CAM 模块, 评估其对基础模型特征提取、上下文建模及注意力机制的独立改进效果; 实验 5-7 分别组合模块, 探究多模块协同作用下的性能增益与计算成本平衡关系; 实验 8 集成全部改进模块, 验证多模态协同感知架构的全局优化能力。逐步引入改进模块后性能演化如表 3 所示。

表 3 消融实验

	RFACnv	CGB	GAM	CCFM	Precision / %	mAP / %	FLOPs ($\times 10^9$)
1	—	—	—	—	82.17	82.50	8.4
2	✓				83.22	84.39	8.6
3		✓			82.49	82.94	6.5
4			✓		86.20	84.57	9.7
5	✓	✓			86.19	83.66	6.7
6	✓		✓		84.75	82.77	9.9
7	✓	✓	✓		87.87	84.44	8.0
8	✓	✓	✓	✓	88.13	85.80	5.7

实验 1 采用基准 YOLOv10 架构实现基线性能, 精确率为 82.17%, $mAP@0.5$ 达 82.5%, 浮点运算次数为 8.4 GLOPs (1 GFLOPs= 10^9 FLOPs)。实验 2 引入 RFACnv 重构主干特征提取网络后, 精确率提升至 83.22%, mAP 提升 1.89%, 表明感受场注意力卷积模块有效强化了负障碍物边缘特征的增强效果; 实验 3 采用 CGB 模块替代主干特征提取网络中的 C2f 模块, 浮点运算次数降至 6.5 GFLOPs, 降幅达 22.6%, 在保持 82.49% 精确率的同时验证该轻量化设计, 可显著提升计算效率, 为计算资源极其受限的场景提供了高性价比的解决方案; 实验 4 在颈部特征融合网络特征图输出前添加 GAM 模块, 精确率提升至 86.20%, 证明全局注意力机制通过有效增强目标相关特征通道并抑制背景干扰通道, 显著优化了特征表示, 从而对复杂背景干扰的抑制效果, 并直接贡献了最高的单模块精度增益, 但该设计伴随计算量显著增加, 增幅达 15.5%。

实验 5 在主干特征提取网络中采用 RFACnv 替代传统的卷积操作, 并联合 CGB 替换网络原有 C2f 结构, 该组合精确率达到 86.19%, mAP 提升 1.16%, 浮点运算次数下降至 6.7 GFLOPs, 降幅 20.2%, 证明该组合在精度与计算量间达到优异平衡, 在计算资源中度受限且对精度有较高要求的场景下展现出显著优势; 实验

6 采用 RFACnv 重构主干特征提取网络,并在颈部特征融合网络中引入 GAM,该组合精确率为 84.75%,*mAP* 提升 0.6%,但该组合浮点运算次数增幅高达 17.86%,相较于其他模型,该组合在综合性能提升幅度方面表现相对有限,却额外引入了更多的计算资源消耗,凸显了模块组合策略对效率与精度平衡的关键影响。实验 7 在实验 5 的基础上引入 GAM,精确率达到 87.87%,*mAP* 提升 1.94%,该组合体现 GAM 提供的精细化特征选择能力与跨尺度信息相结合,对多目标检测场景适应性的优势,有效减少了复杂场景中的误检和漏检。实验 8 集成全部模块的 YOLOv10-MCS 精确率达到 88.13%,同时 *mAP* 提升 3.3%,浮点运算次数降低 32.1%,实验结果从 3 个维度验证了多模态协同感知机制的技术优势,即在检测精度层面实现量化突破,在计算效率层面达成显著优化,更关键的是构建了精度与效率的协同提升范式,并提供了覆盖不同计算需求谱系(从实验 3 的极致轻量到实验 8 的高性能轻量)的可配置选项。消融实验表明,各功能模块在模型优化过程中均产生独立可量化的性能贡献,特别是 GAM 模块在关键特征增强和背景噪声抑制方面展现核心作用,完整验证了多模态架构在检测性能与计算效率双重维度的优越性,且这些模块的组合方式为实际应用提供了灵活性,开发者可根据目标硬件平台的算力约束(如边缘设备需更低 FLOPs)和任务对精度的要求(如安全关键场景需更高 *mAP*),通过启用或禁用特定模块(如 CGB 用于显著压缩计算量,GAM 用于显著提升精度但需权衡算力成本,RFACnv 用于平衡性提升),动态地调整模型结构,实现特定场景下的最优性能权衡。

3.5 对比实验

为了进一步 YOLOv10-MCS 算法在负障碍物检测中的可行性与有效性,将其与常见的几种检测算法进行对比,包含 YOLOv5, YOLOv7^[23], YOLOv8, YOLOv9^[24], YOLOv10。对比实验在同一数据集,相同条件的实验环境及参数下进行,实验结果如表 4 所示。

表 4 对比实验

Model	<i>Precision</i> /%	<i>mAP</i> /%	FLOPs($\times 10^9$)
YOLOv5	82.46	82.38	4.2
YOLOv7	83.09	82.61	105.1
YOLOv8	82.25	83.82	8.2
YOLOv9	86.69	85.81	266.1
YOLOv10	82.17	82.50	8.4
YOLOv10-MCS	88.13	85.80	5.7

由表 4 对比实验结果可知,本文改进的模型 YOLOv10-MCS 在负障碍物检测任务中展现出显著优势。作为工业级轻量化标杆, YOLOv5 以 4.2 GFLOPs 保持

较高部署效率 (*Precision* 82.46%, *mAP* 82.38%),然而在负障碍物检测场景中,其浅层网络对边缘模糊的负障碍物(如坑洼、阴影遮挡)特征提取能力不足,导致漏检率偏高,难以满足自动驾驶场景的安全阈值需求。YOLOv7 算法在 YOLOv5 的基础上通过复杂特征融合将 *mAP* 提升至 82.61%,而 FLOPs 剧增至 105.1 G,是 YOLOv5 的 16 倍;该设计在车载嵌入式平台面临严重延迟问题,无法满足负障碍物检测的毫秒级响应要求。YOLOv8 将 FLOPs 压缩至 8.2 G(较 YOLOv7 降低 92%),但过度轻量化导致精度下降,精确率仅 82.25%(低于 YOLOv5);在夜间负障碍物检测中,其对低对比度目标的识别能力薄弱,体现出轻量化与精度难以兼顾的缺陷。作为对比模型中性能最高者 (*mAP* 85.81%, *Precision* 86.69%), YOLOv9 依赖 266.1 GFLOPs 的复杂计算实现多尺度特征融合;尽管对负障碍物检测效果显著,但其计算量超出车载芯片算力极限,仅适用于服务器端离线处理,无法满足驾驶辅助系统实时检测的核心需求。作为本研究的基线模型, YOLOv10 (8.4 GFLOPs) 试图平衡实时性与精度,但 82.17% *Precision* 和 82.50% *mAP* 在负障碍物检测中表现平庸;实验表明其对动态遮挡障碍物(如部分陷入地面的轮胎)的识别存在明显盲区,侧面体现传统单模态架构的特征提取瓶颈。本文提出的 YOLOv10-MCS 算法通过多模态协同感知机制 *mAP* 较基线 YOLOv10 提升 3.3%, *Precision* 达 YOLO 系列最高的 88.13%, FLOPs 降低至 5.7 G, 低于所有对比轻量模型。综合各项评价指标可得,本文提出的 YOLOv10-MCS 算法在负障碍物检测任务中同步实现检测精度突破,计算成本优化,场景适应性增强(GAM 模块有效抑制复杂地形背景干扰, RFACnv 强化负障碍物边缘特征),满足驾驶辅助系统对负障碍物的高精度、低延时和轻量化的检测需求。

3.6 训练过程与结果可视化

YOLOv10-MCS 算法在训练过程中各项性能指标变化如图 9 所示,训练损失函数(包括边界框回归损失 *train/box_loss*、分类损失 *train/cls_loss* 及分布式焦点损失 *train/df_loss*) 在 400 个训练轮次内分别从初始值 3.3、4.3、4.0 稳定下降至 1.0、0.6、1.2,降幅达 67%~75%。验证集损失函数 (*val/box_loss*、*val/cls_loss*、*val/df_loss*) 亦呈现同步下降趋势,最终分别稳定在 1.3、0.8、1.5,验证损失与训练损失的差异始终保持在 0.3 以内,表明模型在训练集和验证集上均实现了有效优化,且未出现显著过拟合现象。精确率 *metrics/precision* (B) 从 0.0 稳步提升至 0.88 并趋于平稳,目标检测平均精度 *metrics/mAP50* (B) 最优达到 0.85,表明模型在分类与定位任务中均取得显著性

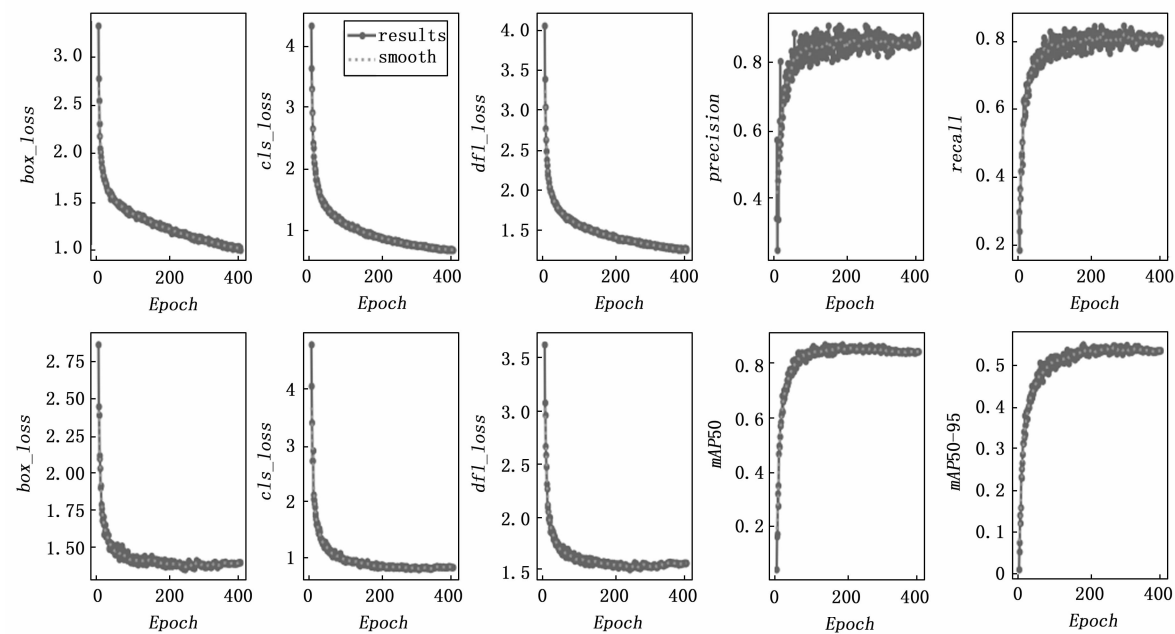


图 9 YOLOv0-MCS 各项性能指标图

能提升。

本文提出的 YOLOv10-MCS 算法在保持原有检测框架优势的基础上, 实现了对负障碍物更精准的定位与分类, 部分典型场景的检测结果如图 10 所示。



图 10 部分负障碍物检测结果示例图

4 结束语

本文构建多模态协同感知的 YOLOv10-MCS 网络架构, 系统解决了负障碍物检测任务中特征模糊性、尺度敏感性及环境干扰等关键问题。通过引入感受野注意力卷积模块 (RFAConv) 强化边缘特征提取, 采用上下文引导块 (CGB) 增强场景关联性, 构建跨尺度特征融合模块 (CCFM) 优化多层级信息交互, 并结合全局注意力机制 (GAM) 实现多维度特征重校准, 建立了一套融合注意力引导、上下文关联与跨尺度优化的特征挖掘与协同增强框架。实验结果表明, 集成全部改进模块的 YOLOv10-MCS 模型相较于基准模型, 在检测精度与计算效率上取得显著突破, 精确率提升至 88.13%, mAP 提高 3.3%, 同时浮点运算量降低 32.1%, 实现了

检测任务中精度与效率的高效平衡。消融实验进一步验证了各模块对模型性能提升的独立贡献及组合增益效应, 其中 RFAConv 与 CGB 的组合在保证精度的同时降低 20.2% 计算量, 而全局注意力机制则展现出对复杂背景干扰的显著抑制能力。本文提出的多模态协同感知机制不仅为负障碍物检测提供了新的解决方案, 也为复杂场景下的目标检测任务提供了可迁移的技术框架, 可拓展应用于自动驾驶、机器人导航等领域。未来将进一步探索模型轻量化与跨模态特征学习的深度融合, 推动算法在边缘计算场景下的实际部署能力。

参考文献:

[1] YOUNGMOK K, YOUNGGIL K, SEUNGYONG S, et al. Review of recent automated pothole-detection methods [J]. Applied Sciences, 2022, 12 (11): 5320 - 5320.

[2] SINGH P, WIJETHUNGA R, SADHU A, et al. Expert evaluation system for pothole defect detection [J]. Expert Systems With Applications, 2025: 127280.

[3] DU R, QIU G, GAO K, et al. Abnormal road surface recognition based on smartphone acceleration sensor [J]. Sensors, 2020, 20 (2): 451.

[4] AZZA A, ANIS K, TAREK A, et al. RoadSense: smartphone application to estimate road conditions using accelerometer and gyroscope [J]. IEEE Sensors Journal, 2017, 17 (13): 4231 - 4238.

[5] CHAO W, ZHEN W, SIMON H, et al. An automated machine-learning approach for road pothole detection using smartphone sensor data [J]. Sensors, 2020, 20 (19): 5564.

(下转第 320 页)