

基于YOLOv10的实时无人机路面病害检测方法

陈 甫*

(重庆电子科技职业大学 智慧健康学院, 重庆 401331)

摘 要: 针对无人机高空视角下路面裂缝细长断连与坑洼边界形变复杂导致的漏检与定位偏差问题, 文章提出一种基于YOLOv10的端到端路面病害检测网络C2B-DeformYOLO。该方法在基准网络上引入C2GD Neck, 通过全局信息收集与连续条件门控实现跨尺度特征的自适应融合, 增强裂缝目标的连贯表征。同时设计BOS Conv, 在有限计算预算下对高响应区域进行稀疏偏移采样与重建, 提高对非凸边界与几何失配的建模能力并兼顾推理效率。在无人机路面数据与公开RDD2022数据集上进行了实验验证, 结果表明所提方法不仅在保持实时性的同时, 还提升了检测精度与跨域泛化表现。

关键词: 无人机; 路面病害检测; YOLOv10; 多尺度特征融合; 稀疏形变卷积

中图分类号: TP391.41 **文章编号:** 1001-5868(2026)04-0749-07

Real-time UAV Road Damage Detection Built on YOLOv10

CHEN Fu

(School of Smart Health, Chongqing College of Electronic Technology, Chongqing 401331, CHN)

Abstract: This study proposes C2B-DeformYOLO, an end-to-end real-time unmanned aerial vehicle (UAV) road damage detector built on YOLOv10, with targeted architectural improvements to address two core issues: cross-scale discontinuous representations of elongated cracks and accurate localization of nonconvex pothole boundaries. First, a Continuous Conditioned Gather Distribute neck (C2GD Neck) is introduced, which performs global information gathering over multi-level features and generates continuous conditioning weights to adaptively allocate and fuse features in a bidirectional manner. This design strengthens coherent responses for thin crack patterns and enhances long-range dependency modeling across scales. Second, Budgeted Offset Sparse Convolution (BOS Conv) is presented, which conducts sparse offset sampling and reconstruction on high-response regions under a constrained computation budget, improving geometric alignment and boundary deformation modeling while preserving efficient inference. Experiments on a UAV road damage dataset and the public RDD2022 benchmark demonstrated consistent gains in detection accuracy and cross-domain generalization over the baseline, without sacrificing real-time performance.

Key words: UAV; road damage detection; YOLOv10; multi-scale feature fusion; sparse deformable convolution

0 引言

随着交通基础设施建设的快速推进, 路面养护已成为保障道路安全运行的关键环节^[1-2]。无人机

凭借机动性强与覆盖范围广的优势, 在路面病害巡检任务中展现出巨大潜力^[3-4]。传统的人工巡检方式效率低且存在安全隐患^[2-3], 难以满足大规模路网的维护需求。因此, 利用计算机视觉技术实现路面

收稿日期: 2026-03-08.

* 通信作者: 陈 甫

病害的自动化检测与评估,已成为智能交通维护领域的重要研究方向^[2,5]。

近年来,基于深度学习的目标检测算法在工业巡检中得到广泛应用^[6-7]。以Faster R-CNN为代表的双阶段检测器虽然检测精度较高,但复杂的候选区域生成过程限制了推理速度,难以满足实时巡检需求^[8]。以YOLO系列为代表的单阶段检测器通过回归方式实现了端到端检测,在速度与精度之间取得了较好平衡^[9-10]。然而,标准卷积操作采用固定几何形状的感受野,在路面病害检测任务中存在固有局限性^[11]。对于细长且延展范围较大的裂缝目标,矩形感受野往往会引入较多无关背景信息,从而降低特征表征的信噪比。对于形态不规则的坑洼目标,固定卷积核也难以有效贴合其边界形状,因而限制了对其几何特征的精确建模。此外,尽管现有形变卷积算子能够在一定程度上提升模型的几何适应能力^[11-13],但其非结构化访存模式在嵌入式平台上通常存在较低的计算效率,从而制约了相关算法在实际场景中的部署应用^[13-16]。

需要指出的是,无人机高空巡检场景下的路面病害目标在成像过程中呈现出明显区别于近距离视角的视觉特征。裂缝通常表现为细长且跨尺度分布的线状结构,在光照变化、阴影干扰或局部遮挡条件下易出现表现断裂现象,从而导致检测器将同一裂缝误判为多个离散碎片目标^[2,4-5,17]。此外,坑洼病害往往具有边界不规则、形态非凸等特点,常规检测器在进行矩形边界框回归时,难以同时兼顾目标中心区域的定位准确性与边界轮廓的贴合效果,进而影响检测结果的精度与稳定性。综上,在满足端侧实时运行要求的前提下,如何增强模型对长距离裂缝连续性特征的表征能力,并提升其对不规则坑洼几何形态的自适应建模能力,已成为当前无人机路面巡检任务中亟待解决的关键问题。

针对上述问题,本文提出了一种面向无人机路面巡检场景的C2B-DeformYOLO端到端目标检测方法。该方法以去除非极大值抑制(NMS)后处理的YOLOv10为基础框架^[17],旨在构建兼顾全局结构感知能力与局部几何表征精度的实时检测模型。首先,在特征融合阶段设计连续性条件化全局分布颈部网络(C2GD Neck),通过提取场景级统计分布信息并生成相应的门控系数,将携带长程先验的全局语义信息注入高分辨率特征层,从而增强模型对裂缝连续性结构的表征能力,并在特征空间中缓解

裂缝断连带来的检测误差。其次,在特征提取阶段引入预算受限偏移稀疏卷积(BOS Conv),该算子通过掩码预测器筛选高响应区域,并结合组共享偏移与插值重建机制,在降低计算开销的同时提升对不规则病害形状的自适应建模能力,从而保证模型在嵌入式平台上的推理效率与部署可行性。

在UAV-PDD2023^[18]、HighRPD^[19]及RDD2022^[5]等多个数据集上的实验结果表明,所提方法在保持低计算复杂度的前提下显著提升了检测性能。在主基准数据集上,本方法的召回率达到89.0%,平均精度均值mAP@0.5达到80.5%,优于RT-DETR及YOLOv12n等主流模型。同时,消融实验验证了C2GD Neck对长裂缝召回率的提升作用,以及BOS Conv在精度与延迟权衡上的优势。所提方法具有较高的工程实用价值,能够满足资源受限条件下的大范围路面实时巡检需求。

1 基准网络模型

如图1所示,YOLOv10网络结构整体由主干网络Backbone、颈部网络Neck和检测头Head三部分组成。该模型最大的创新在于引入了双重标签分配策略,在训练阶段同时采用一对多和一对一匹配,在推理阶段仅使用一对一支,从而去除了非极大值抑制NMS后处理步骤,实现了端到端的高效检测。此前,DETR通过集合预测与二分图匹配构建了端到端目标检测框架^[20]。

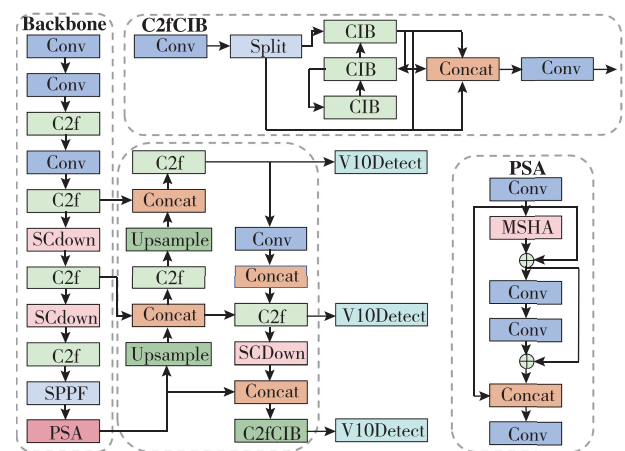


图1 YOLOv10结构示意图

主干网络Backbone主要负责从输入图像中提取多尺度特征。它延续了CSPNet^[21]的设计理念,利用C2f模块通过丰富的梯度流增强特征提取能力。同时结合SCDown模块进行空间下采样,有效

保留了图像的细粒度信息并减少了计算冗余。SPPF模块^[10]位于主干末端,通过不同核大小的最大池化操作扩展感受野,以捕获全局上下文信息。

在多尺度特征融合方面,FPN通过自顶向下路径与横向连接构建特征金字塔^[22],颈部网络Neck进一步采用路径聚合网络PANet^[23]结构。该结构包含自顶向下的上采样路径和自底向上的下采样路径,实现了深层语义特征与浅层定位信息的双向融合。通过跨层连接机制,不同尺度的特征图在通道维度上进行拼接,增强了模型对多尺度目标的检测能力。

检测头Head部分采用解耦式设计,分别处理类别分类与边界框回归任务。尽管YOLOv10通过端到端机制显著降低了推理延迟,但其主要依赖矩形框的交并比作为匹配依据。在面对无人机视角下路面裂缝的细长弯曲形态及坑洼的不规则边缘时,标准的矩形感受野难以精确对齐目标几何特征,且缺乏针对长程断连目标的全局关联机制,这限制了其在复杂路面巡检场景下的进一步应用。

2 改进的C2B-DeformYOLO网络

针对无人机路面巡检场景中裂缝细长易断连及坑洼形态不规则导致的检测难题,同时兼顾嵌入式平台上的实时运行需求,本文在YOLOv10n基础上引入两项核心改进,构建了C2B-DeformYOLO端到端目标检测网络。改进后的网络结构主要包含连续性条件化全局分布颈部网络即C2GD Neck和预算受限偏移稀疏卷积即BOS Conv两个关键组件,如图2所示。将所提出的改进模型命名为DeformYOLO,其中“Deform”表示“deformable”,意在突出网络对目标形变、尺度变化及几何结构变化的自适应建模能力。DeformYOLO并非单纯指某一现有标准模型,而是本文在YOLOv10n基础上,通过引入连续性条件化全局分布颈部网络和预算受限偏移稀疏卷积后形成的改进检测框架,其整体

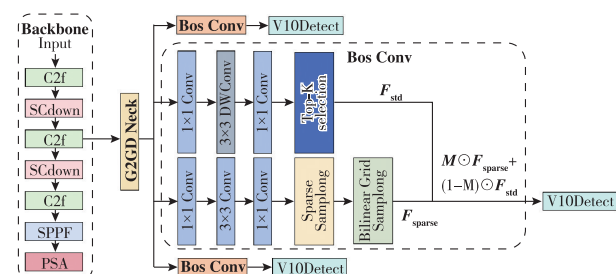


图2 C2B-DeformYOLO改进结构图

架构旨在通过特征流的全局连续性建模解决断连问题,并利用硬件友好的稀疏计算算子提升对非凸几何形状的适应能力。

2.1 网络整体架构

C2B-DeformYOLO网络保留了原主干网络的高效特征提取能力,重点重构了特征融合与对齐阶段。主干网络输出的多尺度特征首先输入至C2GD Neck模块。该模块取代了传统的路径聚合结构,负责在特征融合过程中注入全局统计分布信息,以强化跨尺度长程目标的特征连通性。随后,经过增强的特征图进入检测头之前的空间感知阶段。在此阶段的高分辨率分支引入BOS Conv算子,利用其预算门控机制对病害边缘进行稀疏采样与形变校正。最终,对齐后的特征图被送入解耦检测头,直接输出目标的类别置信度与边界框坐标,完成端到端检测流程。

2.2 连续性条件化全局分布颈部网络

为解决裂缝在局部感受野下呈现的特征断裂问题,本文设计了C2GD Neck模块。如图3所示,该模块摒弃了仅依赖局部像素邻域的传统融合方式,转而采用全局收集、连续性条件化融合及尺度敏感分发注入的三步流程。

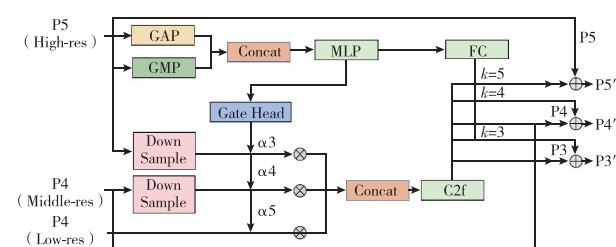


图3 C2GD Neck架构示意图

首先是全局收集阶段。该模块在最高层特征图上应用全局平均池化与最大池化操作,提取场景级的纹理统计分布向量。这一全局向量不仅包含类别先验,还编码了路面背景与线性病害的整体分布特征。

其次是连续性条件化融合阶段。已有SE与CBAM分别通过通道重标定和通道-空间联合注意力增强有效特征^[24-25]。不同于上述注意力方式,本文模块利用提取的全局向量生成一组门控系数,对不同尺度的特征通道进行自适应加权。若全局分布指示存在长程线结构,门控机制将显著增强沿裂缝走向的特征响应,同时抑制非相关的背景噪声通道。这一过程在特征空间建立了一种基于统计规

律的连续性条件,使离散的局部特征在全局上下文的约束下产生关联。

最后是尺度敏感分发注入阶段。经过处理的全局上下文信息被回灌至高分辨率特征层,这种自顶向下的注入机制使得原本在底层特征图中因信噪比低而被淹没的微小裂缝片段获得了上下文支持,从而被重新激活为有效的前景特征。C2GD Neck通过上述机制在特征层面缝合了视觉上断续的裂缝,有效提升了长裂缝目标的召回率。

2.3 预算受限偏移稀疏卷积

针对坑洼不规则边界与标准卷积矩形感受野的几何失配问题,以及传统形变卷积在PCIe硬件上推理效率低下的矛盾,本文提出了BOS Conv算子。该算子通过预算门控与组共享偏移机制,实现了计算效率与几何适应性的平衡。

预算门控机制主要解决全图形变计算带来的资源浪费问题。该机制设定一个计算预算超参数,利用轻量级掩码预测器预先评估特征图各位置的响应强度。对于平坦路面等背景区域,算子退化为标准卷积或直接跳过计算。仅在病害边缘及纹理复杂的高响应区域,算子才触发形变采样过程。这种稀疏计算策略显著减少了非必要的浮点运算量。

组共享偏移与插值重建技术则致力于提升硬件亲和性。BOS Conv摒弃了逐点独立预测偏移量的做法,改为在特征通道组之间共享同一组偏移坐标。这不仅大幅压缩了偏移量预测分支的参数量,还减少了非结构化内存访问的频率。同时,算子利用嵌入式平台优化的双线性插值指令进行特征重采样。通过这种设计,BOS Conv能够根据坑洼的实际形状自适应调整感受野,使回归出的矩形框中心点更加精准,同时确保模型在端侧部署时具备低延迟与高吞吐量的实时性能。

3 实验结果及分析

3.1 数据集构建与实验环境

为全面验证所提方法的有效性,本文构建了包含三个不同特性的实验数据集。主基准采用UAV-PDD2023数据集,该数据集由无人机低空采集,包含六类典型路面病害,用于评估模型在标准场景下的综合检测性能。为验证模型在大范围巡检场景下对微小目标的捕捉能力,引入了HighRPD数据集,该数据集包含大量高分辨率切片图像,主要考

察算法对细长裂缝与微小坑洼的召回能力。此外,选用RDD2022数据集进行实验,主要目的是验证所提方法在视角畸变场景下的泛化能力。该数据集的来源复杂、采集条件多样,包含车辆移动采集过程中形成的不同拍摄角度、路面尺度变化以及明显的透视形变,这使得目标在图像中的外观表现具有较强的不确定性。相比于视角较为固定、成像条件相对单一的数据集,RDD2022更能够反映真实道路检测任务中由于相机姿态变化和场景透视效应带来的挑战。因此,在该数据集上开展实验,有助于评估模型在复杂视角畸变条件下的特征建模能力及跨场景泛化性能。

为保证实验结果的可比性,所有实验均在统一的硬件与软件环境下进行。实验平台为搭载NVIDIA RTX 4090 GPU的高性能计算服务器,深度学习框架采用PyTorch。模型训练阶段使用随机梯度下降(SGD)优化器,初始学习率设为0.01,动量参数设为0.937,并结合Mosaic^[10]数据增强策略,以增强模型对复杂场景的适应能力。

3.2 评价指标

为全面评估模型性能,选取精确率(P)、召回率(R)、平均精度均值(mAP)、参数量(Parameters)和浮点运算量(GFLOPs)作为评价指标。其中,mAP@0.5用于表征当交并比阈值为0.5时模型的检测精度,能够反映目标检测的总体效果;mAP@0.5:0.95则按照COCO评价协议在多个交并比阈值下综合评价模型性能^[26],可用于衡量模型的精确定位能力;参数量和GFLOPs分别从模型规模与计算成本两个方面表征其复杂度,以验证模型在端侧部署场景下的轻量化可行性。

3.3 对比实验

除了本文所提机制,还选取了Faster R-CNN, RT-DETR^[27],以及YOLOv8至YOLOv12^[28-29]等多款主流检测器进行对比。在UAV-PDD2023, HighRPD, RDD2022三个数据集上验证了所提模型,如图4所示,本文方法在三个数据集及其平均设置下均位于帕累托前沿附近,在相近计算量条件下取得更高的检测精度,或在相近精度条件下降低计算开销,体现出更优的精度-效率权衡。而mAP和PR曲线如图5所示,不同数据集的曲线形态与最佳置信度区间存在差异,但整体趋势一致;其中HighRPD的曲线整体更靠近右上角、RDD2022相对更低,表明模型在不同数据分布下的可分性与阈

值敏感性不同,同时与表格中的总体性能排序保持一致。其可视化结果如图6所示。

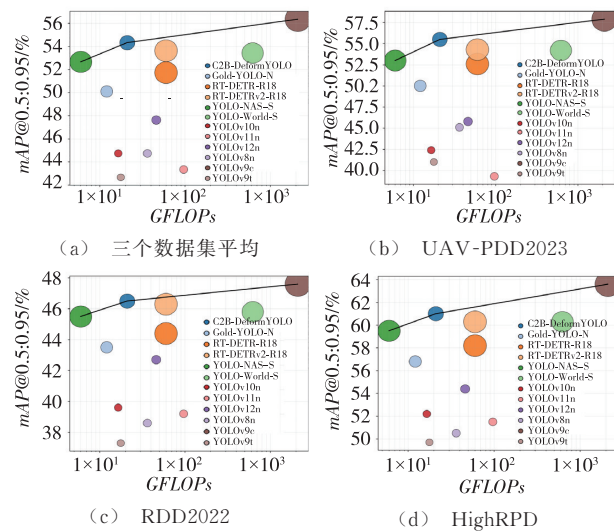


图4 不同数据集上的性能-计算量帕累托前沿对比散点图

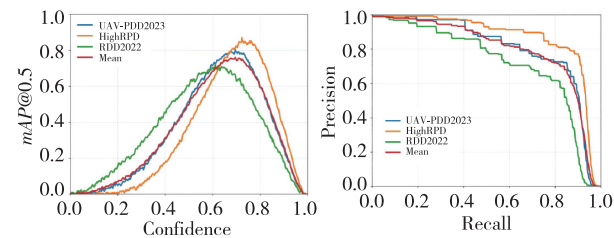


图5 $mAP@0.5$ - Confidence 与 Precision - Recall(P-R)曲线

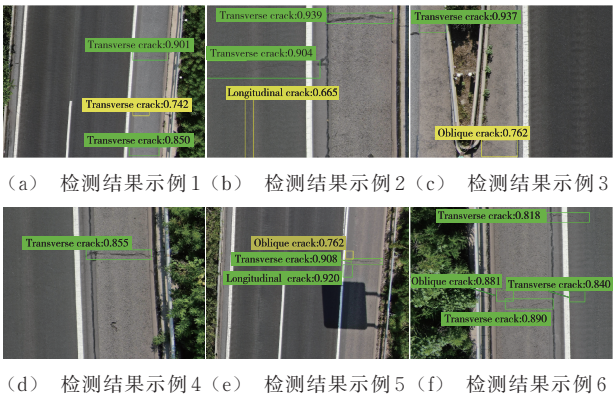


图6 UAV-PDD2023数据集检测结果

如表1所示,在主基准 UAV-PDD2023 数据集上,实验结果表明本文方法在各项指标上均处于领先地位。C2B-DeformYOLO 的召回率达 89.0%, $mAP@0.5$ 达 80.5%, $mAP@0.5:0.95$ 达 55.5%。相较于原生 YOLOv10n,本文方法在保持参数量仅增加至 3.82 M 的前提下, $mAP@0.5$ 提升了 12.7%,充分证明了特征对齐机制对低信噪比病害的检测增益。与拥有 25.50 M 参数的 YOLOv9c^[29] 大模型

相比,本文方法在精度接近的同时,计算量 GFLOPs 仅为前者的 1/10,展现了极佳的效能比。

表1 UAV-PDD2023数据集对比实验结果

方法	R / %	$mAP@0.5 / %$	$mAP@0.5:0.95 / %$	参数量 / M	GFLOPs
YOLOv8n	78.0	64.7	45.1	2.58	36.3
YOLOv9t	75.0	61.7	41.0	2.02	17.7
YOLOv10n	80.0	67.8	42.4	2.25	16.5
YOLOv11n	79.0	65.9	39.3	2.58	96.4
YOLOv12n	84.0	72.8	45.8	3.14	46.1
YOLO-World-S	88.0	78.8	54.2	17.93	623.1
YOLOv9c	90.0	82.2	57.9	25.50	2 102.1
Gold-YOLO-N	85.5	74.5	50.0	5.62	12.1
RT-DETR-R18	86.0	76.0	52.6	20.06	60.0
RT-DETRv2-R18	87.0	78.0	54.3	20.06	60.0
YOLO-NAS-S	87.0	77.0	53.0	19.06	6.0
C2B-DeformYOLO	89.0	80.5	55.5	3.82	21.0

如表2所示,在 HighRPD 数据集上,针对大尺度图像中的细小病害,本文方法的优势更为显著。得益于全局上下文的注入,模型召回率高达 91.0%, $mAP@0.5$ 达到 87.0%。相比 YOLOv11n 和 YOLOv12n,本文方法在召回率上分别提升了 7.0% 和 5.0%,表明 C2GD Neck 模块有效缓解了长裂缝的断连漏检问题。

表2 HighRPD大尺度场景对比实验结果

方法	R / %	$mAP@0.5 / %$	$mAP@0.5:0.95 / %$	参数量 / M	GFLOPs
YOLOv8n	83.0	75.1	50.5	2.58	36.3
YOLOv9t	82.0	74.2	49.7	2.02	17.7
YOLOv10n	85.0	77.1	52.2	2.25	16.5
YOLOv11n	84.0	76.7	51.5	2.58	96.4
YOLOv12n	86.0	80.8	54.4	3.14	46.1
YOLO-World-S	90.0	85.7	60.3	17.93	623.1
YOLOv9c	92.0	88.5	63.6	25.50	2 102.1
Gold-YOLO-N	88.0	82.7	56.8	5.62	12.1
RT-DETR-R18	89.0	84.6	58.2	20.06	60.0
RT-DETRv2-R18	90.0	86.5	60.3	20.06	60.0
YOLO-NAS-S	90.0	85.4	59.5	19.06	6.0
C2B-DeformYOLO	91.0	87.0	61.0	3.82	21.0

如表3所示,在跨域 RDD2022 数据集上,尽管存在视角差异,C2B-DeformYOLO 依然保持了稳健的性能, $mAP@0.5$ 达到 72.0%。这一结果优于 RT-DETRv2-R18 的 71.0% 和 YOLOv10n 的 64.0%,验证了 BOS Conv 算子在应对不同几何形变时具有良好的鲁棒性,能够适应车载视角下的透视变换。

表3 RDD2022跨域场景对比实验结果

方法	$R/\%$	$mAP@0.5/\%$	$mAP@0.5:0.95/\%$	参数量/ M	$GFLOPs$
YOLOv8n	74.0	62.1	38.6	2.58	36.3
YOLOv9t	73.0	61.7	37.3	2.02	17.7
YOLOv10n	75.0	64.5	39.6	2.25	16.5
YOLOv11n	75.0	63.4	39.2	2.58	96.4
YOLOv12n	78.0	67.6	42.7	3.14	46.1
YOLO-World-S	80.0	70.3	45.8	17.93	623.1
YOLOv9c	82.0	73.2	47.6	25.50	2 102.1
Gold-YOLO-N	79.0	68.1	43.5	5.62	12.1
RT-DETR-R18	80.0	69.3	44.4	20.06	60.0
RT-DETRv2-R18	81.0	71.5	46.3	20.06	60.0
YOLO-NAS-S	80.0	70.6	45.5	19.06	6.0
C2B-DeformYOLO	81.5	72.0	46.5	3.82	21.0

3.4 消融实验

为系统评估 C2B-DeformYOLO 中各核心模块的贡献,共设计了7组消融实验,结果见表4。其中,实验1作为基线模型,仅保留YOLOv10的主干网络及常规卷积模块,其在多个数据集上的平均 $mAP@0.5$ 达到74.2%,召回率为76.1%。在实验3中引入C2GD Neck后,召回率提升至77.8%,优于引入原版GD-Neck的实验2。这表明连续性条件化融合机制能够有效利用全局分布信息,将断裂的裂缝特征重新连接,从而显著降低漏检率。在实验5中引入BOSConv替代高分辨率层的标准卷积, $mAP@0.5:0.95$ 提升至50.0%,且参数量仅比基线增加1.57 M。对比实验4中引入的DCNv3^[13],BOSConv在实现同等精度提升的同时, $GFLOPs$ 降低了2.1,验证了预算门控与组共享机制在计算效率上的优势。最终,在实验7中同时集成C2GD Neck与BOSConv,模型达到了最优性能,召回率攀升至79.6%, $mAP@0.5:0.95$ 达到52.0%。相较于基线模型,全量模型在参数量仅增加1.04 M的微小代价下,实现了精度的全面突破。

表4 消融实验结果

序号	模块组合	$P/\%$	$R/\%$	$mAP@0.5/\%$	$mAP@0.5:0.95/\%$	参数量/ M	$GFLOPs$
1	Baseline	82.6	76.1	74.2	47.8	2.25	16.5
2	+原版GD-Neck	83.1	76.9	75.6	48.9	3.37	21.0
3	+C2GD-Neck	83.7	77.8	76.8	50.1	3.55	21.3
4	+DCNv3	83.0	77.3	76.2	49.7	3.28	23.8
5	+BOSConv	83.4	77.5	76.6	50.0	3.86	21.7
6	C2GD-Neck+DCNv3	84.1	78.7	78.0	51.2	3.45	24.2
7	C2GD-Neck+BOSConv	84.6	79.6	78.8	52.0	3.82	21.0

4 结论

本文针对无人机路面巡检任务中裂缝目标细长且易发生断连、坑洼目标形态不规则等关键难点,提出了一种基于改进YOLOv10的端到端目标检测方法C2B-DeformYOLO。针对路面病害目标的视觉表征特点及端侧部署需求,分别从特征融合机制和卷积算子设计两个层面进行了改进。在特征融合阶段,引入C2GD Neck模块,通过全局信息收集与连续性条件化融合机制,将场景级纹理分布先验注入特征表示过程中,从而在特征空间中建立长程依赖关系,增强模型对裂缝连续性结构的表征能力,有效缓解传统局部卷积易造成的漏检和断连问题。在特征提取阶段,提出BOS Conv算子,利用预算门控与组共享偏移机制,在提升网络对不规则坑洼边界几何适应能力的同时,降低了传统形变卷积在嵌入式平台上的计算开销,从而兼顾了检测精度与推理效率。

实验结果表明,C2B-DeformYOLO在保持低计算复杂度的前提下显著提升了检测性能。在主基准UAV-PDD2023数据集上,该方法的召回率达到89.0%, $mAP@0.5$ 达到80.5%,在综合性能上优于对比的主流检测器。在HighRPD大尺度数据集上,模型对细小病害的召回率高达91.0%,证明了其在复杂大场景下缓解漏检的能力。在RDD2022跨域数据集上, $mAP@0.5$ 保持在72.0%,显示出算法对不同视角及透视形变的良好鲁棒性。此外,该模型参数量控制在3.82 M,浮点运算量为21.0 $GFLOPs$,实现了高精度检测与端侧实时推理的有效平衡。未来工作将进一步探索多传感器数据融合及轻量化模型在嵌入式边缘设备上的部署与优化,并结合Grad-CAM等可视化方法分析模型对裂缝连续区域与坑洼边界的关注机制。

参考文献:

- [1] 欧建西,张健钦,段彬,等. 沥青路面养护决策技术研究进展[J]. 北京工业大学学报, 2025, 51(10): 1229-1240.
Ou Jianxi, Zhang Jianqin, Duan Bin, et al. Progress in research on asphalt pavement maintenance decision-making technology[J]. Journal of Beijing University of Technology, 2025, 51(10): 1229-1240.
- [2] 易钰程,王靖智,朱路,等. 路面缺陷智能检测系统与方法综述[J]. 华东交通大学学报, 2023, 40(5): 19-31.
Yi Yucheng, Wang Jingzhi, Zhu Lu, et al. Review of the intelligent pavement defect detection system and methods[J].

- Journal of East China Jiaotong University, 2023, 40(5): 19-31.
- [3] 罗旭东, 吴一全, 陈金林. 无人机航拍影像目标检测与语义分割的深度学习研究方法研究进展[J]. 航空学报, 2024, 45(6): 235-264.
- Luo Xudong, Wu Yiquan, Chen Jinlin. Research progress on deep learning methods for object detection and semantic segmentation in UAV aerial images[J]. Acta Aeronautica et Astronautica Sinica, 2024, 45(6): 235-264.
- [4] 王晓燕, 王禧钰, 李 杰, 等. 无人机视角下的道路损伤检测算法 MAS-YOLOv8n[J]. 光电工程, 2024, 51(10): 240170.
- Wang Xiaoyan, Wang Xiyu, Li Jie, et al. MAS-YOLOv8n road damage detection algorithm from the perspective of drones[J]. Opto-Electronic Engineering, 2024, 51(10): 240170.
- [5] Arya D, Maeda H, Ghosh S K, et al. RDD2022: a multi-national image dataset for automatic road damage detection[J]. Geoscience Data Journal, 2024: gdj3. 260.
- [6] Hütten N, Alves Gomes M, Hölken F, et al. Deep learning for automated visual inspection in manufacturing and maintenance: a survey of open-access papers[J]. Applied System Innovation, 2024, 7(1): 11.
- [7] 罗 元, 王薄宇, 陈 旭. 基于深度学习的目标检测技术的研究综述[J]. 半导体光电, 2020, 41(1): 1-10.
- Luo Yuan, Wang Boyu, Chen Xu. Research progresses of target detection technology based on deep learning[J]. Semiconductor Optoelectronics, 2020, 41(1): 1-10.
- [8] Ren S, He K, 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, 2017, 39(6): 1137-1149.
- [9] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [10] Bochkovskiy A, Wang C Y, Liao H M. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. 2020: arXiv: 2004. 10934.
- [11] Zhu X, Hu H, Lin S, et al. Deformable ConvNets V2: more deformable, better results[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 9300-9308.
- [12] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks [C]//2017 IEEE International Conference on Computer Vision (ICCV), 2017: 764-773.
- [13] Wang W, Dai J, Chen Z, et al. InternImage: exploring large-scale vision foundation models with deformable convolutions[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 14408-14419.
- [14] Ahn S, Chang J W, Kang S J. An efficient accelerator design methodology for deformable convolutional networks [EB/OL]. 2020: arXiv: 2006. 05238.
- [15] 王利翔, 林珊玲, 林志贤, 等. 基于Zynq平台的图像目标检测系统[J]. 半导体光电, 2023, 44(1): 147-152.
- Wang Lixiang, Lin Shanling, Lin Zhixian, et al. Image target detection system based on zynq platform[J]. Semiconductor Optoelectronics, 2023, 44(1): 147-152.
- [16] 贾 亮, 林铭文, 戚丽瑾, 等. 面向无人机航拍图像的多尺度目标检测研究[J]. 半导体光电, 2024, 45(3): 501-508.
- Jia Liang, Lin Mingwen, Qi Lijin, et al. Multiscale target detection for UAV aerial images [J]. Semiconductor Optoelectronics, 2024, 45(3): 501-508.
- [17] Wang A, Chen H, Liu L, et al. YOLOv10: real-time end-to-end object detection [C]//Neural Information Processing Systems Foundation, 2024: 107984-108011.
- [18] Yan H, Zhang J. UAV-PDD2023: a benchmark dataset for pavement distress detection based on UAV images[J]. Data in Brief, 2023, 51: 109692.
- [19] He J, Gong L, Xu C, et al. HighRPD: a high-altitude drone dataset of road pavement distress [J]. Data in Brief, 2025, 59: 111377.
- [20] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//Computer Vision - ECCV 2020, 2020: 213-229.
- [21] Wang C Y, Mark Liao H Y, Wu Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 1571-1580.
- [22] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 936-944.
- [23] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [24] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [25] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//Computer Vision-ECCV 2018, 2018: 3-19.
- [26] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[C]//Computer Vision - ECCV 2014, 2014: 740-755.
- [27] Zhao Y, Lv W, Xu S, et al. DETRs beat YOLOs on realtime object detection [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.
- [28] Tian Y, Ye Q, Doermann D. YOLOv12: attention-centric real-time object detectors[EB/OL]. 2025: arXiv: 2502. 12524.
- [29] Wang C Y, Yeh I H, Mark Liao H Y. YOLOv9: learning what you want to learn using programmable gradient information[C]//Computer Vision-ECCV 2024, 2025: 1-21.

作者简介:

陈 甫(1983—),男,重庆市人,副教授,研究领域主要为计算机科学、软件工程、人工智能、大数据。