

# 融合分层大核卷积与高效注意力的轻量化 RT-DETR道路缺陷检测

杨宗才, 李星星\*

五邑大学电子与信息工程学院, 广东 江门

收稿日期: 2026年1月18日; 录用日期: 2026年3月6日; 发布日期: 2026年3月19日

## 摘要

当前道路缺陷检测算法难以兼顾高精度与实时性要求, 且传统卷积神经网络因局部感受野限制, 对细长裂缝等特征的捕捉能力有限。为此, 本文提出一种基于改进RT-DETR的轻量化道路缺陷检测算法。该方法以端到端的RT-DETR为基线, 在消除后处理延迟的同时增强全局特征建模。具体改进包括: 1) 采用轻量级StarNet替换原ResNet骨干, 结合大核深度可分离卷积与线性门控机制, 在显著降低参数量和计算成本的前提下保持多尺度特征提取能力; 2) 为缓解下采样导致的细微裂缝特征模糊与丢失, 使用分层大核卷积(HLKConv)替换标准下采样层; 3) 在骨干网络高层引入高效多尺度注意力模块(EMA), 通过跨通道与跨空间交互建模, 提升模型对多尺度、不规则缺陷的感知与聚焦能力。在RDD2022数据集上的实验结果表明, 改进后模型的mAP@0.5达到85.79%, 较原始RT-DETR-R18提升0.67%, 而参数量和计算量分别大幅降低55.9%和56.5%。相较于其他主流算法, 本方法在维持高检测精度的同时, 显著降低了硬件资源需求, 更适用于计算资源受限的边缘设备进行实时道路巡检。

## 关键词

道路缺陷检测, RT-DETR, 轻量化网络, StarNet, HLKConv, 注意力机制

# Lightweight Road Defect Detection Based on RT-DETR Fusing Hierarchical Large Kernel Convolution and Efficient Attention

Zongcai Yang, Xingxing Li\*

School of Electronic and Information Engineering, Wuyi University, Jiangmen Guangdong

Received: January 18, 2026; accepted: March 6, 2026; published: March 19, 2026

\*通讯作者。

文章引用: 杨宗才, 李星星. 融合分层大核卷积与高效注意力的轻量化 RT-DETR 道路缺陷检测[J]. 人工智能与机器人研究, 2026, 15(2): 516-526. DOI: 10.12677/airr.2026.152050

## Abstract

Current road defect detection algorithms struggle to balance the requirements of high accuracy and real-time performance. Furthermore, traditional Convolutional Neural Networks (CNNs) are constrained by local receptive fields, limiting their ability to capture features of slender cracks. To address these issues, this paper proposes a lightweight road defect detection algorithm based on an improved RT-DETR. Leveraging the end-to-end RT-DETR as a baseline, the proposed method enhances global feature modeling while eliminating post-processing latency. The specific improvements include: 1) Replacing the original ResNet backbone with the lightweight StarNet, which combines large-kernel depth-wise separable convolutions with a linear gating mechanism to maintain multi-scale feature extraction capabilities while significantly reducing parameter count and computational costs; 2) Utilizing Hierarchical Large-Kernel Convolution (HLKConv) to replace standard downsampling layers, thereby mitigating the blurring and loss of fine crack features caused by downsampling; and 3) Incorporating the Efficient Multi-Scale Attention (EMA) module into the high-level layers of the backbone to enhance the model's perception and focus on multi-scale and irregular defects through cross-channel and cross-spatial interaction modeling. Experimental results on the RDD2022 dataset demonstrate that the improved model achieves an mAP@0.5 of 85.79%, an increase of 0.67% over the original RT-DETR-R18, while the parameter count and computational volume are substantially reduced by 55.9% and 56.5%, respectively. Compared with other mainstream algorithms, this method significantly lowers hardware resource requirements while maintaining high detection accuracy, making it highly suitable for real-time road inspection on resource-constrained edge devices.

## Keywords

Road Defect Detection, RT-DETR, Lightweight Network, StarNet, HLKConv, Attention Mechanism

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

道路基础设施的安全与健康状况对社会经济发展与公众出行至关重要[1]。传统的道路缺陷(如裂缝、坑槽、网裂等)检测主要依赖人工巡检,存在效率低、成本高、安全性差且主观性强等问题。随着计算机视觉技术的快速发展,基于深度学习的自动缺陷检测方法已成为研究主流,其具备自动化、高效率和高精度的潜力,对于实现道路养护的智能化、预防性维护具有重大应用价值[2]。

当前,基于卷积神经网络的目标检测模型在道路缺陷检测中取得了广泛应用。早期的研究主要集中在以 Faster R-CNN 为代表的两阶段检测算法上。例如, Wang 等人[3]提出了基于 Faster R-CNN 的道路损伤检测和分类方法,在竞赛中获得 0.6255 的平均 F1 得分。Fang 等人[4]提出了一种新颖的混合方法用于原始图像中的裂纹检测,结合 Faster R-CNN 模型和贝叶斯概率分析以实现稳健的裂纹检测。这类两阶段算法虽然在检测精度上表现优异,但其网络结构复杂、计算量大,导致推理速度较慢,难以满足车载设备实时巡检的需求。为了解决实时性问题,以 SSD 和 YOLO 系列为代表的单阶段检测算法逐渐成为研究热点。Yang 等人[5]构建了一种基于单发多框检测器(SSD)的路面裂纹自动检测网络。该网络通过引入感受野增强模块,强化了特征提取能力,从而在保证实时性的前提下,显著提升了裂纹识别精度。Lu 等人

[6]集成了深度分离变形卷积(DSDCM)、起始模块(IM)与特征再校准模块(FRM), 提出一种改进的单发多框检测器(ISSD), 用于解决桥梁裂纹检测的挑战。随后, 随着 YOLO 系列的迭代, Hu 等人[7]提出了一种基于改进版 YOLOv5 和车载图像的新型探测方法用于车辆安装图像中的路面裂纹检测。Wu 等人[8]提出一种增强型道路裂缝检测算法 YOLOv8-VOS, 该算法在 RDD2022 数据集上的平均准确率比原始 YOLOv8 提升了 2%。Gao 等人[9]为提升建筑裂纹检测的精度和实时效率提出了一种改进的 YOLOv11 算法, 实验结果表明, 增强后的模型检测精度(mAP@0.5)达到 88.6%, 相较于原始 YOLOv11 模型提升了 4.6%。然而, 上述基于 CNN 的检测器在道路场景中仍面临两大核心瓶颈: 第一, 受限与卷积核的局部感受野, CNN 难以有效捕捉长距离的上下文信息, 而道路裂缝往往具有细长、跨度大的几何特征, 导致模型在处理长裂缝或密集网裂时容易出现漏检或断裂现象; 第二, CNN 类算法均依赖非极大值抑制(NMS)进行后处理以去除冗余框, 这不仅增加了推理延迟, 且 NMS 的超参数在面对不同尺度的路面病害时难以自适应调节, 影响了最终的检测稳定性。针对上述局限性, RT-DETR [10] (Real-Time DEtection TRansformer)模型应运而生。作为首个实时的端到端目标检测器, RT-DETR 摒弃了传统的 NMS 后处理操作, 利用 Transformer 架构独有的自注意力机制(Self-Attention), 使其具备了强大的全局特征建模能力。这使得模型能够从全局视角理解图像信息, 从而更准确地识别形态各异的道路缺陷。虽然 RT-DETR 在精度和速度的平衡上超越了许多 YOLO 模型, 但其原始网络参数量依然较大, 在计算资源受限的边缘设备上部署仍存在挑战。

因此, 本文面向道路缺陷检测的实际部署需求, 提出一种基于改进 RT-DETR 的轻量化检测方法。本文旨在不牺牲精度的前提下, 大幅降低模型复杂度与计算成本, 并提升对复杂缺陷特征的感知能力。主要贡献包括:

(1) 骨干网络轻量化: 采用专为边缘设备设计的极轻量级主干网络 StarNet [11], 替代 RT-DETR 原有的 ResNet 骨干。StarNet 通过大核深度可分离卷积与线性门控机制构建高效基础模块, 在极大压缩参数量的同时保持了良好的多尺度特征提取能力, 为模型的高效推理奠定基础。

(2) 下采样策略优化: 针对裂缝目标细长且纹理复杂的特性, 引入了分层大核卷积[12] (HLKConv)以替换标准下采样卷积。HLKConv 通过深度可分离卷积、深度可分离空洞卷积与逐点卷积的三阶段分层计算, 有效缓解了传统下采样造成的特征细节丢失与块状伪影问题。

(3) 特征增强机制: 针对道路缺陷形态多样、尺度变化大的特点, 在 StarNet 的高层特征阶段引入了高效多尺度注意力模块 EMA [13]。该模块通过分组交叉协方差计算, 以极低开销建模通道与空间远程依赖, 有效增强网络对关键缺陷特征的聚焦与表征能力, 提升对小目标及不规则缺陷的检测鲁棒性。

(4) 端到端轻量化模型: 将上述改进集成于 RT-DETR 框架中, 构建一个完整的、无需 NMS 的轻量化端到端检测器。该模型在保持 RT-DETR 实时性与高精度优势的基础上, 进一步优化了模型效率, 更适用于计算资源紧张的边缘部署场景。

## 2. RT-DETR 模型

RT-DETR 的整体结构包含骨干网络、高效混合编码器以及解码器三部分。其中, 骨干网络通常采用 CNN 架构(如 ResNet 或 HGNet), 用于提取图像的多尺度特征。高效混合编码器由基于注意力的尺度内特征交互模块(AIFI)与基于 CNN 的跨尺度特征融合模块(CCFM)共同构成: AIFI 专注于在单尺度高层语义特征上通过自注意力机制捕获实体间的关联, 而 CCFM 则用于融合不同尺度的特征, 从而形成高质量的特征表达, 为解码阶段提供丰富而精准的输入。在解码器部分, RT-DETR 引入不确定性最小化查询选择机制, 通过联合优化分类与定位置信度以提升检测准确性; 同时, 其支持在推理过程中动态调整解码器层数, 无需重新训练即可灵活平衡检测速度与精度, 增强了模型在实际应用中的适应性与实用性。

在本文工作中, 我们选择以 ResNet18 为骨干网络的 RT-DETR-R18 作为基准模型, 其结构如图 1 所示。这一选择主要基于在道路缺陷检测这一实际应用场景中, 模型需要在计算资源受限的边缘设备上实现实时或准实时检测的需求。ResNet18 作为经典的轻量级卷积网络, 在模型参数量、计算复杂度与特征提取能力之间取得了良好的平衡。相比于更深的 ResNet-50 或 HGNetv2 等网络, 其显著降低了计算负担, 确保了较高的推理帧率; 同时, 其基本的残差结构又能有效提取多层次的特征, 为后续的高效混合编码器提供必要的多尺度特征输入。

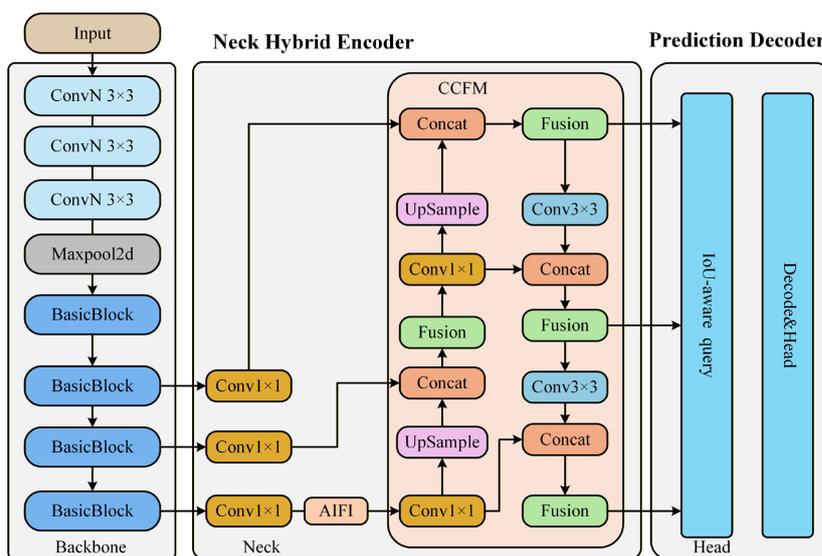


Figure 1. Architecture of the RT-DETR-R18 network  
图 1. RT-DETR-R18 网络结构

### 3. 改进的 RT-DETR 模型

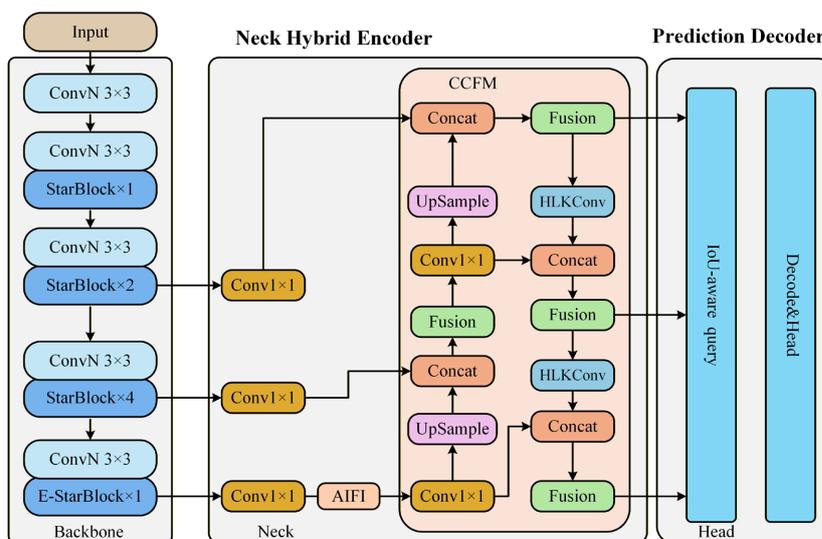


Figure 2. Architecture of the improved RT-DETR Network  
图 2. 改进的 RT-DETR 网络结构

改进后的 RT-DETR 整体网络结构如图 2 所示。该模型保留了原 RT-DETR 的高效混合编码器与解码

器架构, 主要针对特征提取阶段进行了轻量化与增强重构。具体改进主要体现在三个层面: 首先, 利用轻量级 StarNet 替换原有的 ResNet 骨干网络; 其次, 在特征下采样阶段, 使用 HLKConv 替代标准卷积以保留细微特征; 最后, 在骨干网络的高层特征输出端构建了融合 EMA 注意力机制的 E-StarBlock。整个网络依然保持端到端的推理模式, 无需 NMS 后处理。

### 3.1. StarNet 主干网络

为在 RT-DETR-R18 的基础上进一步实现轻量化目标, 满足道路缺陷检测任务对低功耗与高实时性的部署需求, 本文采用更为高效的轻量级主干网络 StarNet 替代原有的 ResNet-18。该网络的核心是由大核深度可分离卷积构成的轻量化构建块(StarBlock), 能够在显著降低模型参数量与计算复杂度的同时, 保持出色的多尺度特征提取能力。如图 3 所示, 每个 StarBlock 首先利用  $7 \times 7$  深度卷积进行局部特征建模, 随后通过两路  $1 \times 1$  卷积生成特征并进行线性门控交互, 再经由逐点卷积与二次深度卷积实现高效融合, 最终通过残差连接增强优化稳定性。整个网络由一层  $3 \times 3$  卷积构成的 stem 层及四个下采样阶段组成, 每个阶段包含一个卷积下采样层与多个前述 Block, 最终输出适用于 RT-DETR 高效混合编码器的多尺度特征。StarNet 凭借其精巧的架构设计, 在维持较强表征能力的前提下, 进一步压缩了模型体积、提升了推理速度, 使其更适于部署于计算资源受限的嵌入式或移动设备, 从而为轻量化道路缺陷检测系统的实际落地提供了更优的骨干网络选择。

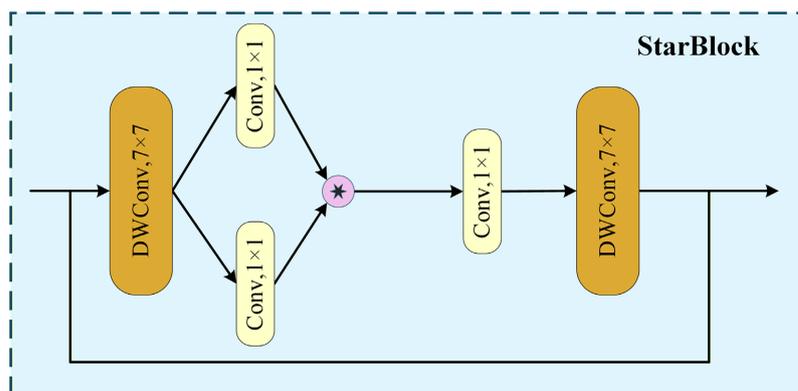


Figure 3. StarBlock module structure  
图 3. StarBlock 模型结构

### 3.2. EMA 模块

图 4 展示了本文构建的 E-StarBlock, 其核心为高效多尺度注意力模块(EMA)。为增强网络对道路缺陷多样形态与尺度的特征捕获能力, 本文在语义信息更丰富的高层特征阶段引入该模块。其核心设计思想是通过分组策略与空间维度分解, 以较低的计算开销建立跨通道与跨空间的远程依赖关系。具体流程如下: 首先, 输入特征在通道维度被均匀分组; 随后, 分别沿高度和宽度方向进行自适应池化以捕获长程空间上下文, 并通过  $1 \times 1$  卷积融合后经 Sigmoid 函数生成方向感知的空间权重。与此同时, 另一分支通过组归一化与  $3 \times 3$  卷积并行提取经空间权重调制的特征与局部细节特征。最后, 模块通过计算两组特征的交叉协方差矩阵, 生成能够反映特征通道间全局相关性的注意力图, 并与原始特征进行加权融合。通过上述设计, EMA 模块能够协同建模通道间的全局依赖与空间上的长程交互, 显著增强了网络对多尺度、不规则形态道路缺陷的语义信息捕获与聚焦能力, 从而为后续的检测解码器提供了判别性更强、更具鲁棒性的特征表达。

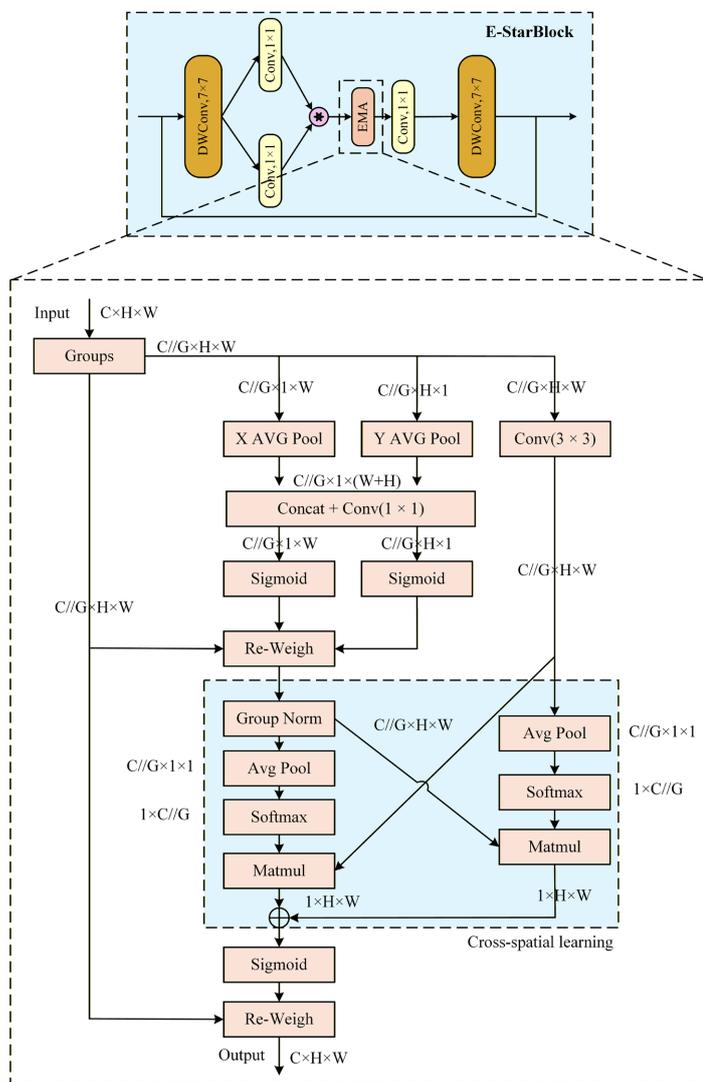


Figure 4. E-StarBlock module structure  
图 4. E-StarBlock 模型结构

### 3.3. HLKConv

针对路面裂缝细微且拓扑结构复杂的特点, 常规卷积在下采样时难免造成关键细节丢失, 而直接使用大空洞率卷积虽能捕捉长距离信息, 却往往因稀疏采样导致网格伪影。因此, 本文在 RT-DETR 的下采样阶段引入了分层大核卷积(HLKConv)以替代标准卷积, 该结构如图 5 所示。不同于传统的单步大核计算, HLKConv 设计了一种“先密后疏”的级联特征提取路径, 在降低计算冗余的同时, 兼顾了局部纹理的连续性与全局语义的完整性。具体而言, 该模块将大感受野的特征建模解耦为三个递进的步骤: 局部密集感知、长程稀疏探索以及特征聚合映射。假设输入特征图为  $X \in \mathbb{R}^{W \times H \times C}$ , 计算流程如下: 首先, 利用小尺寸的深度可分离卷积(DWConv)提取空间上连续的局部特征。这一步至关重要, 它能捕获裂缝的边缘细节, 并为后续的空洞卷积提供密集的像素信息, 填补采样盲区:

$$X_{dw} = \text{DWConv}_{k \times k}(X) \tag{1}$$

其中  $k$  为分解后的卷积核尺寸。随后, 基于生成的  $X_{dw}$ , 利用较大空洞率为  $r$  的深度可分离空洞卷积

(DWDCnv)进一步捕获长距离上下文信息。由于输入并未直接使用原始特征, 而是经过了第一步的平滑与特征填充, 从而有效抑制了空洞卷积带来的伪影问题:

$$X_{dwd} = DWDCnv_{k \times k, r}(X_{dw}) \tag{2}$$

最后, 采用通道拼接操作(Concat)融合局部细节与长程语义, 并通过 $1 \times 1$ 的逐点卷积(PWConv)进行通道混合与维度映射, 生成最终的输出特征 $Y$ :

$$Y = PWConv_{1 \times 1}(\text{Concat}[X_{dw}, X_{dwd}]) \tag{3}$$

通过这种分层设计, HLKConv 有效模拟了先看清局部纹理, 再关联周围环境的感知过程, 在降低计算量的同时, 增强了模型对不规则、细长裂缝的完整性检测能力。

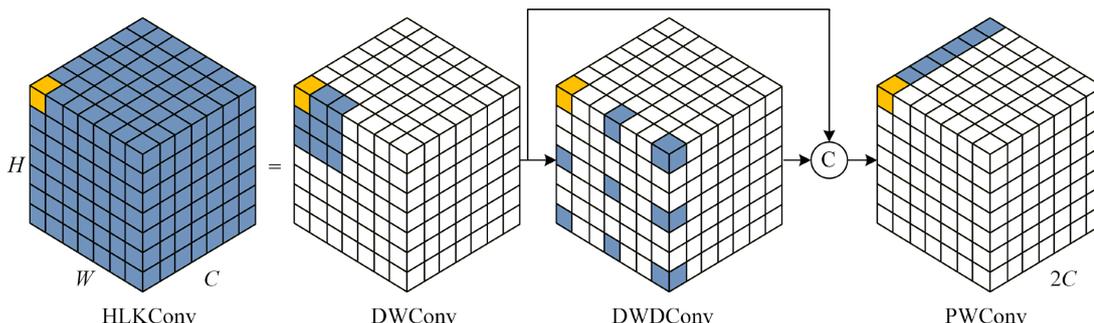


Figure 5. HLKConv module structure [12]  
图 5. HLKConv 模型结构[12]

## 4. 实验与结果

### 4.1. 环境配置

本实验的硬件配置为一张 NVIDIA GeForce RTX 4090 D 显卡, 其显存容量为 24 GB。软件环境方面, 采用 Python 3.10.16 编程语言, 并基于 CUDA 12.1 使用 PyTorch 2.2.2 深度学习框架。实验参数具体设置如下: 输入图像尺寸统一调整为  $640 \times 640$ , 使用 AdamW 优化器进行训练, 总训练轮数(epochs)为 200, 批量大小(batch size)为 8, 初始学习率设置为 0.0001。

### 4.2. 数据集

本文选用的数据来源于道面损坏检测基准数据集 RDD2022, 该数据集被广泛用于道路病害自动检测研究。实验中, 我们使用了其中的中国子集, 共计 4370 张有效图像。数据涵盖了五种典型的道路缺陷类型, 包括纵向裂缝(D00, 3680 处)、横向裂缝(D10, 2145 处)、网状裂缝(D20, 852 处)、坑洼(D40, 286 处)以及道路修补(Repair, 948 处)。为进行模型训练与评估, 我们将全部数据按 8:1:1 的比例随机划分为训练集、验证集和测试集。

### 4.3. 评价指标

为全面评估所提模型的性能, 本文采用目标检测领域广泛认可的评估指标, 主要包括: 平均精度均值(mAP)、模型参数量(Parameters)与计算复杂度(GFLOPS)。其中, mAP 以交并比(IoU)阈值为 0.5 (即 mAP@0.5)进行计算, 能够综合反映模型在多种道路缺陷类别上的检测精度与鲁棒性。参数量用于衡量模型自身的规模与存储开销, 而 GFLOPS 则用于量化模型前向推理过程的理论计算负担。这三项指标的结合, 旨在从检测精度、模型轻量化程度及计算效率三个核心维度, 对模型性能进行综合考量, 以满足实

际道路巡检应用对准确性、部署便捷性与实时性的综合要求。

#### 4.4. 消融实验

如表 1 所示, 我们在 RDD2022 数据集上对所提的逐项改进策略进行了系统的消融实验验证。为适应边缘部署的严格资源限制, 我们采用轻量级 StarNet 骨干网络替代了原始的 ResNet-18。这一调整显著降低了模型的复杂程度, 使其参数量和计算量分别下降了 52.7% 和 54.7%。尽管 mAP@0.5 指标出现 0.41% 的轻微下降, 但该步骤为后续实现实时推理奠定了关键基础。在 StarNet 的高效架构之上, 我们进一步以分层大核卷积(HLKConv)取代了标准下采样层。该设计使得模型体积进一步缩减至 8.51 M 参数, 计算量降至 24.0 GFLOPs, 同时 mAP@0.5 回升至 84.93%。这一结果表明, HLKConv 通过分层计算降低参数冗余与通过扩大感受野捕捉细节特征的双重优势, 有效部分弥补了轻量化骨干网络带来的精度损失。最后, 通过将 EMA 模块集成到 StarNet 的高层特征中, 最终模型实现了多维度特征的协同增强。在 HLKConv 所构建的高效特征流基础上, EMA 机制强化了对不规则缺陷的语义聚焦能力, 最终将 mAP@0.5 提升至 85.79%。与原始 RT-DETR-R18 基准相比, 优化后的模型在检测精度上提升了 0.67%, 同时参数量和计算成本分别降低了 55.9% 与 56.5%, 在保持高精度的同时大幅提升了推理效率, 实现了双重突破。

**Table 1.** Ablation study results

**表 1.** 消融实验结果

Method	mAP@0.5	Parameters (M)	GFLOPS
RT-DETR-R18	85.12	19.88	57.0
+StarNet	84.71	9.41	25.8
+HLKConv	85.55	18.98	55.2
+StarNet+HLKConv	84.93	8.51	24.0
+StarNet+EMA	85.60	9.66	26.6
+StarNet+HLKConv+EMA	85.79	8.76	24.8

#### 4.5. 不同算法对比实验

在不同算法的对比实验中, 本文算法在综合性能评估中表现出色。如表 2 所示, 在检测精度方面, 本文算法达到了最优的 mAP@0.5 指标, 不仅显著超越传统 Faster R-CNN 模型, 也优于 YOLOv3-tiny、YOLOv5n 和 YOLOv8n 等 YOLO 系列算法, 同时较 RT-DETR-R18 也有进一步提升。在模型复杂度控制上, 本文算法在参数量和计算量方面均大幅低于 Faster R-CNN 与 RT-DETR-R18, 虽然高于 YOLOv5n 和 YOLOv8n, 但取得了更为优异的检测精度, 实现了精度与效率的更好平衡。总体而言, 本文算法在保持较高检测性能的同时有效控制了模型复杂度, 展现了良好的实用性与部署潜力。

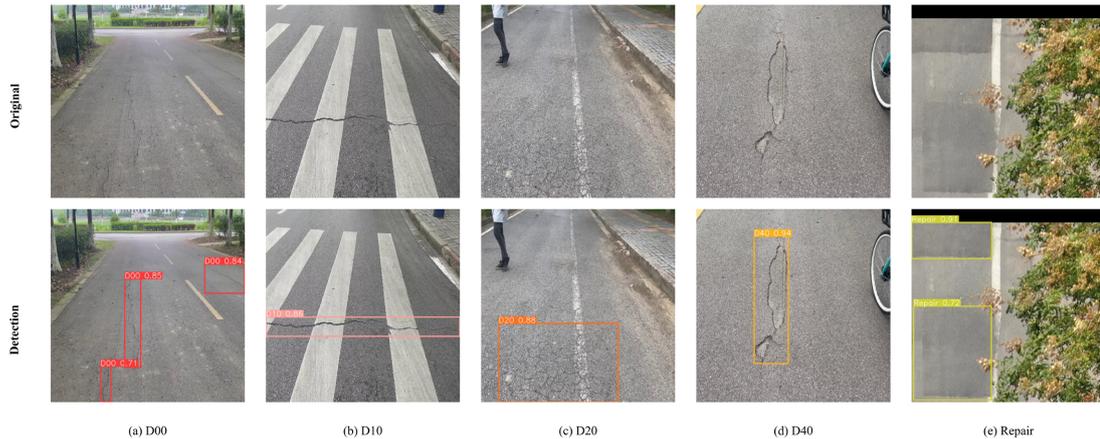
**Table 2.** Comparison of different algorithm results

**表 2.** 不同算法结果对比

Model	mAP@0.5	Parameters (M)	GFLOPs
Faster R-CNN	67.71	136.77	401.2
YOLOv3-tiny	77.6	12.13	18.9
YOLOv5n	81.4	2.50	7.1
YOLOv8n	83.2	3.01	8.1
RT-DETR-R18	85.12	19.88	57.0
OURS	85.79	8.76	24.8

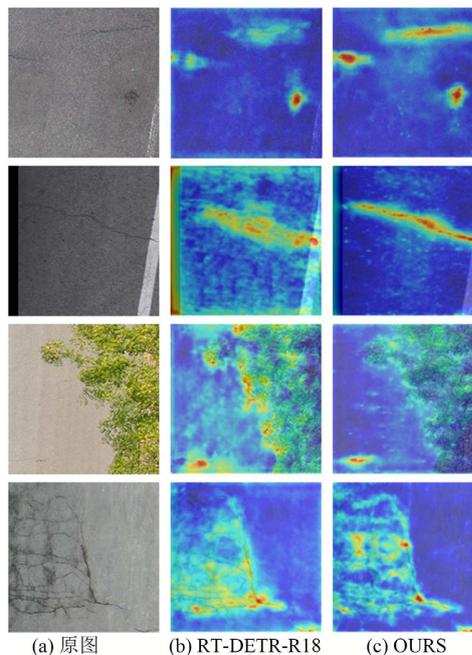
### 4.6. 检测结果可视化

为了直观验证算法的实际检测效能, 图 6 展示了模型在五种道路缺陷类型的可视化推理结果。可以看出, 无论是针对细长的纵向与横向裂缝(图 6(a)、图 6(b)), 还是形态复杂的网状裂缝与坑槽(图 6(c)、图 6(d)), 模型均能给出紧密且高置信度的检测框。特别值得注意的是, 在图 6(b)中, 尽管存在斑马线强纹理干扰, 模型仍精准识别了横向裂缝; 在图 6(e)中, 面对树木阴影遮挡, 模型依然准确检测出修补区域。这些结果表明, 改进后的轻量化模型不仅具备强大的特征提取与抗干扰能力, 且对不同尺度及光照条件下的道路病害具有良好的鲁棒性。



**Figure 6.** Visualization results of the proposed algorithm on different road defects, (a) D00, (b) D10, (c) D20, (d) D40, (e) Repair  
**图 6.** 本文算法在不同道路缺陷上的可视化结果, (a) 纵向裂缝, (b) 横向裂缝, (c) 网状裂缝, (d) 坑洼, (e) 道路修补

### 4.7. 特征热力图可视化分析



**Figure 7.** Different scene visualization comparison of multi-scale aggregated features  
**图 7.** 不同场景下多尺度聚合特征的可视化对比

为了验证改进模块在特征提取层面的内在机理,本文基于 GradCAMPlusPlus [14]技术对基准模型(RT-DETR-R18)与本文改进模型(OURS)的多尺度聚合特征进行了热力图可视化对比。如图 7 所示,我们选取了四种具有代表性的道路裂缝场景进行分析。观察发现,基准模型受限于局部感受野,其热力图在裂缝处常呈碎片化分布,且极易受环境噪声干扰。相比之下,改进模型展现出优异的特征重构能力:得益于 HLKConv 的分层大核机制,模型在处理断续及含噪裂缝(第一、二行)时,成功建立了像素间的长距离依赖,将原本断裂、模糊的局部特征修复为连续且锐利的线性结构。而在极具挑战的植被阴影干扰场景中(第三行),EMA 模块通过通道重校准机制发挥了关键作用,有效屏蔽了背景中的高频纹理噪声,使得热力图在阴影区域不产生误响应,仅对真实病害产生高置信度激活。此外,面对形态复杂的网状裂缝(第四行),改进模型同样能够清晰勾勒出病害的完整骨架脉络。上述可视化结果直观地证实了本文提出的轻量化策略并非简单的模块堆叠,而是从根本上增强了模型对裂缝几何拓扑的感知连续性以及在非结构化环境下的检测鲁棒性。

## 5. 结论

本文针对道路缺陷检测任务中存在的模型参数量大、推理延迟高以及细长裂缝特征难以捕捉等问题,提出了一种基于改进 RT-DETR 的轻量化端到端检测算法。通过引入专为移动端设计的 StarNet 替代 ResNet 骨干网络,并引用分层大核卷积(HLKConv)重构下采样模块,结合高效多尺度注意力(EMA)模块,我们成功构建了一个适应边缘侧部署的高性能模型。StarNet 的线性门控机制与大核卷积在大幅降低计算成本的同时保证了基础特征提取能力,HLKConv 则通过多阶段的特征分解与聚合策略,有效克服了传统下采样造成的特征细节丢失与网格伪影问题,强化了模型对细长裂缝拓扑结构与长程依赖的捕捉能力;而 EMA 模块则通过跨维度交互有效弥补了轻量级网络在捕捉长距离依赖和复杂几何特征上的不足,显著增强了对不规则裂缝的识别鲁棒性。在 RDD2022 数据集上的实验结果表明,改进后的算法在将参数量和计算量分别压缩至 8.76 M 和 24.8 G 的同时,实现了 85.79% 的 mAP@0.5,在综合性能上不仅优于原始 RT-DETR-R18,也超越了 YOLOv8n 等主流检测器。该研究有效地解决了模型复杂度与检测性能之间的矛盾,证明了在资源受限环境下实现高精度、实时道路智能化巡检的可行性,为未来的嵌入式设备部署提供了具有实际应用价值的解决方案。

## 参考文献

- [1] Yu, J., Jiang, J., Fichera, S., Paoletti, P., Layzell, L., Mehta, D., *et al.* (2024) Road Surface Defect Detection—From Image-Based to Non-Image-Based: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, **25**, 10581-10603. <https://doi.org/10.1109/tits.2024.3382837>
- [2] Mei, Q. and Gül, M. (2020) A Cost-Effective Solution for Pavement Crack Inspection Using Cameras and Deep Neural Networks. *Construction and Building Materials*, **256**, Article 119397. <https://doi.org/10.1016/j.conbuildmat.2020.119397>
- [3] Wang, W., Wu, B., Yang, S. and Wang, Z. (2018) Road Damage Detection and Classification with Faster R-CNN. 2018 *IEEE International Conference on Big Data (Big Data)*, Seattle, 10-13 December 2018, 5220-5223. <https://doi.org/10.1109/bigdata.2018.8622354>
- [4] Fang, F., Li, L., Gu, Y., Zhu, H. and Lim, J. (2020) A Novel Hybrid Approach for Crack Detection. *Pattern Recognition*, **107**, Article 107474. <https://doi.org/10.1016/j.patcog.2020.107474>
- [5] Yang, J., Fu, Q. and Nie, M. (2020) Road Crack Detection Using Deep Neural Network with Receptive Field Block. *IOP Conference Series: Materials Science and Engineering*, **782**, Article 042033. <https://doi.org/10.1088/1757-899x/782/4/042033>
- [6] Lu, G., He, X., Wang, Q., Shao, F., Wang, J. and Jiang, Q. (2022) Bridge Crack Detection Based on Improved Single Shot Multi-Box Detector. *PLOS ONE*, **17**, e0275538. <https://doi.org/10.1371/journal.pone.0275538>
- [7] Hu, H., Li, Z., He, Z., Wang, L., Cao, S. and Du, W. (2024) Road Surface Crack Detection Method Based on Improved YOLOv5 and Vehicle-Mounted Images. *Measurement*, **229**, Article 114443.

- <https://doi.org/10.1016/j.measurement.2024.114443>
- [8] Wu, H.Y., Kong, L.Y. and Liu, D.H. (2024) Crack Detection on Road Surfaces Based on Improved YOLOv8. *IEEE Access*, **12**, 190850-190864. <https://doi.org/10.1109/access.2024.3517632>
- [9] Gao, X., Cao, C. and Yi, X. (2025) Using the Improved YOLOv11 Model to Enhance Computer Vision Applications for Building Crack Detection Algorithms. *Scientific Reports*, **15**, Article No. 38843. <https://doi.org/10.1038/s41598-025-22160-6>
- [10] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., *et al.* (2024) DETRs Beat YOLOs on Real-Time Object Detection. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 16965-16974. <https://doi.org/10.1109/cvpr52733.2024.01605>
- [11] Ma, X., Dai, X., Bai, Y., Wang, Y. and Fu, Y. (2024) Rewrite the Stars. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 5694-5703. <https://doi.org/10.1109/cvpr52733.2024.00544>
- [12] Zhang, G., Xu, G., Chen, S., Wang, H. and Zhang, X. (2025) Learning Dynamic Local Context Representations for Infrared Small Target Detection. *IEEE Transactions on Geoscience and Remote Sensing*, **63**, 1-13. <https://doi.org/10.1109/tgrs.2025.3605402>
- [13] Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., *et al.* (2023) Efficient Multi-Scale Attention Module with Cross-Spatial Learning. 2023 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, 4-10 June 2023, 1-5. <https://doi.org/10.1109/icassp49357.2023.10096516>
- [14] Chattopadhyay, A., Sarkar, A., Howlader, P. and Balasubramanian, V.N. (2018) Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks. 2018 *IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Tahoe, 12-15 March 2018, 839-847. <https://doi.org/10.1109/wacv.2018.00097>