

文章编号 1004-924X(2025)22-3549-15

# RSF-DETR: 空频增强与上下文重构的 路面损伤检测

周冬梅<sup>1</sup>, 仵兵兵<sup>1</sup>, 刘小明<sup>2</sup>, 闫浩文<sup>1</sup>, 吴小所<sup>1\*</sup>  
(1. 兰州交通大学 电子与信息工程学院, 甘肃 兰州 730070;  
2. 青海理工学院 工学院, 青海 西宁 810016)

**摘要:** 针对路面损伤形式多样、检测精度不高及漏检率较高的问题, 本文提出一种基于 RT-DETR 模型的改进方法。首先, 基于空间域高频边缘增强与频域全局特征提取的联合思想, 设计了空频双域特征增强模块 FreSCal, 增强模型对目标信息和边缘信息的提取能力, 并提升目标区域与背景的区别能力。其次, 借鉴 CGRSeg 网络的上下文引导特征重构思想, 提出上下文引导空间特征重构金字塔网络 RSDFPN, 通过构建尺度感知的语义金字塔与动态特征融合机制, 显著增强模型对多尺度目标的特征融合能力。最后, 通过动态分组卷积混洗与 Transformer 的全局建模能力, 实现空间域高效特征增强与频域上下文融合, 提升模型对目标识别的检测精度。实验结果表明, 本文的改进方法在 RDD2022 和 UAV-PDD2023 两个主流数据集上均取得显著提升, mAP@0.5 指标较基线方法分别提升 1.9% 和 3.7%, 可为路面损伤检测提供一种有效的技术支持。

**关键词:** 路面损伤检测; 实时检测 Transformer; 空频双域; 上下文引导重构; 动态分组卷积混洗与 Transformer 协同优化模块

中图分类号: TP391.41 文献标识码: A

doi: 10.37188/OPE.20253322.3549 CSTR: 32169.14.OPE.20253322.3549

## RSF-DETR: Road damage detection with space frequency enhancement and context reconstruction

ZHOU Dongmei<sup>1</sup>, WU Bingbing<sup>1</sup>, LIU Xiaoming<sup>2</sup>, YAN Haowen<sup>1</sup>, WU Xiaosuo<sup>1\*</sup>

(1. College of Electronic and Information Engineering, Lanzhou Jiaotong University,  
Lanzhou 730070, China;

2. Engineering School, Qinghai Institute of Technology, Xining 810016, China)

\* Corresponding author, E-mail: wuxs\_laser@lzjtu.edu.cn

**Abstract:** Aiming at the problems of various pavement-damage forms, low detection accuracy and high miss-detection rate, this paper proposed an improved method based on the RT-DETR model. First, inspired by the joint idea of high-frequency edge enhancement in the spatial domain and global feature extraction in the frequency domain, the spatial - frequency dual-domain feature-enhancement module FreSCal

收稿日期: 2025-06-10; 修订日期: 2025-07-28.

基金项目: 国家重点研发计划 (No. 2022YFB3903604); 甘肃省科技计划项目 (No. 24JRZA104); “昆仑英才”人才引进行项目 (No. W2023-QLGKLYCZX-034); 国家自然科学基金项目 (No. 62161016); 科研培育计划项目 (No. 202301lwys021)

was designed to strengthen the model's ability to extract target and edge information and to improve its capacity to distinguish target regions from background. Secondly, drawing on the context-guided feature-reconstruction concept of the CGRSeg network, the context-guided spatial feature-reconstruction pyramid network RSDFPN was proposed. By building a scale-aware semantic pyramid and a dynamic feature-fusion mechanism, the model's capability to fuse features for multi-scale targets was significantly enhanced. Finally, through dynamic group convolution shuffling and the global modeling capacity of Transformer, efficient spatial-domain feature enhancement and frequency-domain context fusion were achieved, raising the model's detection accuracy for target recognition. The experimental results show that the improved method in this paper has achieved significant improvement on both RDD2022 and UAV-PDD2023 mainstream datasets, mAP@0.5 Compared with the baseline method, the indicators are increased by 1.9% and 3.7% respectively, which can provide an effective technical support for pavement damage detection.

**Key words:** pavement damage detection; Real-Time Detection Transformer (RT-DETR); space frequency dual domain; context guided reconstruction; Dynamic Group Convolution Shuffle Transformer(DGCST)

## 1 引言

在全球城市化加速进程中,道路设施面临高频荷载与环境侵蚀的多重威胁,传统人工检测因效率瓶颈难以满足需求,而基于深度学习的检测算法通过精准识别路面裂缝及坑洼损伤,为构建道路健康智能监测系统开辟了新路径。

近年来,路面损伤智能检测领域借助深度学习技术获得了显著的突破,而多种创新性的检测算法通过基于卷积神经网络(Convolutional Neural Network, CNN)的结构来提升路面损伤检测的性能。例如, Yang等<sup>[1]</sup>设计多尺度三重注意力网络,通过多尺度特征融合、通道注意力机制及空间注意力机制,来解决传统方法在像素级裂缝分割中感受野有限和局部特征处理不足等问题,提升了裂缝分割的准确性。Li等<sup>[2]</sup>构建了融合高低层语义信息的金字塔结构,并采用可变形卷积层替代标准卷积操作以适配道路损伤的不规则轮廓,同时在特征金字塔末端引入空洞卷积扩展感受野,显著改善了复杂路况下的检测效果。Sami等<sup>[3]</sup>基于YOLOv5框架改进,通过优化初始锚框参数、引入高效通道注意力机制以及焦点损失函数,解决了样本失衡和特征提取效率低下的问题,提升了模型对路面损伤的检测精度和速度。Wang等<sup>[4]</sup>提出将BiFPN和SimSPPF相结合来优化特征金字塔,并引入具有动态大卷积核注意力机制扩大模型感受野,提升了对道路缺陷的检测精度和速度。Hong等<sup>[5]</sup>提出了

一种基于深度学习的无人机图像高速公路裂缝分割方法,主要采用卷积神经网络构建模型,通过大量标注数据训练,学习裂缝与背景的特征差异,实现裂缝像素的精准分类。Li等<sup>[6]</sup>提出了一种基于区域引导网络的复杂背景图像道路损毁检测方法,通过区域引导网络将图像划分为多个区域,提取各区域特征并结合分类器识别道路损毁区域,解决了复杂背景图像中道路损毁检测精度低、易受背景干扰的问题。Han等<sup>[7]</sup>提出了MambaCrackNet,融合Visual Mamba模块与卷积神经网络,结合自监督预训练和多尺度特征融合,提升像素级裂缝分割精度,解决了传统方法精度不足和对标注数据依赖强的问题。

随着Transformer架构在计算机视觉领域的兴起,其也被逐渐应用于路面损伤检测中。Qing等<sup>[8]</sup>对DETR检测框架进行了适应性改进,通过引入动态调整的损失函数和多尺度特征融合策略,解决了裂缝形态复杂多样导致的检测精度不足问题,提升了模型对裂缝的表征能力和检测效率。Lin等<sup>[9]</sup>则采用纯Transformer架构的编解码网络,将输入图像解构为规则化网格序列后,采用压缩型多头自注意力机制进行全局上下文建模,利用跨层聚合架构的解码器整合全局和局部注意力以恢复详细特征,并进行像素级裂缝掩码预测,有效地捕捉了目标对象。Wang等<sup>[10]</sup>提出了一种基于Transformer优化构建的Road-TransTrack模型,用于道路损伤检测与计数,该

模型借助 Transformer 和自注意力机制处理连续帧序列特征,解决了现有模型在连续帧中难以准确计数和区分相同损伤的问题。Deng 等<sup>[11]</sup>提出了一种基于检测变换器的 DSR-Net,该网络采用独特的选择性回退查询机制,专注于提升道路裂缝检测的精度和效率,有效解决了传统方法在复杂背景下检测精度不足、易受干扰的问题。

早期的 DETR 基于 Transformer 编码器-解码器实现端到端检测:编码器通过全局注意力建模展平图像特征( $O(n^2)$ 复杂度),解码器利用可学习目标查询(Object Queries)执行集合预测,并以二分匹配损失取代 NMS。然而,随机初始化的查询令收敛过程迟缓,而高分辨率特征序列所带来的计算负担又严重制约了实时应用。则 RT-DETR<sup>[12]</sup>(Real-Time Detection Transformer)针对该瓶颈提出三项协同优化:首先,构建混合编码器(AIFI+CCFM),以尺度内交互与跨尺度融合替代单一的 Transformer 结构,显著压缩计算量并维持全局上下文建模;其次,引入 IoU 感知查询选择,依据质量评分动态筛选初始目标查询,直接加速解码器收敛;最后,通过特征解耦与集合预测机制简化后处理,彻底规避 NMS。这些改进使 RT-DETR 在保持端到端优势的同时,达成高精度实时检测。

尽管许多方法在路面损伤检测领域取得了一定进展,但对于路面损伤信息检测仍需要进一步优化,本文提出了一种基于 RT-DETR 模型的

改进方法 RSF-DETR,旨在提升路面损伤检测的精度。主要包括以下几点工作:

(1) 设计空频双域特征增强模块(FreSCal),结合空间域的 Scharr 算子梯度计算强化局部细节,利用频域的快速傅里叶变换与逆变换提取全局纹理信息,并通过可学习权重动态融合双域特征,增强模型对目标信息和边缘信息的提取能力,提升目标区域与背景的分度。

(2) 提出上下文引导空间特征重构金字塔网络(Residual Shuffle Dense Feature Pyramid Network, RSDFPN),构建尺度感知语义金字塔与动态特征融合机制,通过特征交互提取尺度感知的语义特征,减少信息丢失,提升模型对多尺度目标的特征融合能力。

(3) 引入动态分组卷积混洗与 Transformer 协同优化模块(Dynamic Group Convolution Shuffle Transformer, DGCST),通过动态分组卷积混洗减少冗余计算,打破通道隔离,增强特征多样性,同时引入基于 Transformer 的卷积前馈网络(ConvFFN),融合局部与全局特征建模能力,提升对图像的特征能力。

## 2 RT-DETR 模型

RT-DETR 是一个基于 Transformer 的实时端到端检测模型,由主干网络、混合编码器和解码器构成,其结构如图 1 所示。

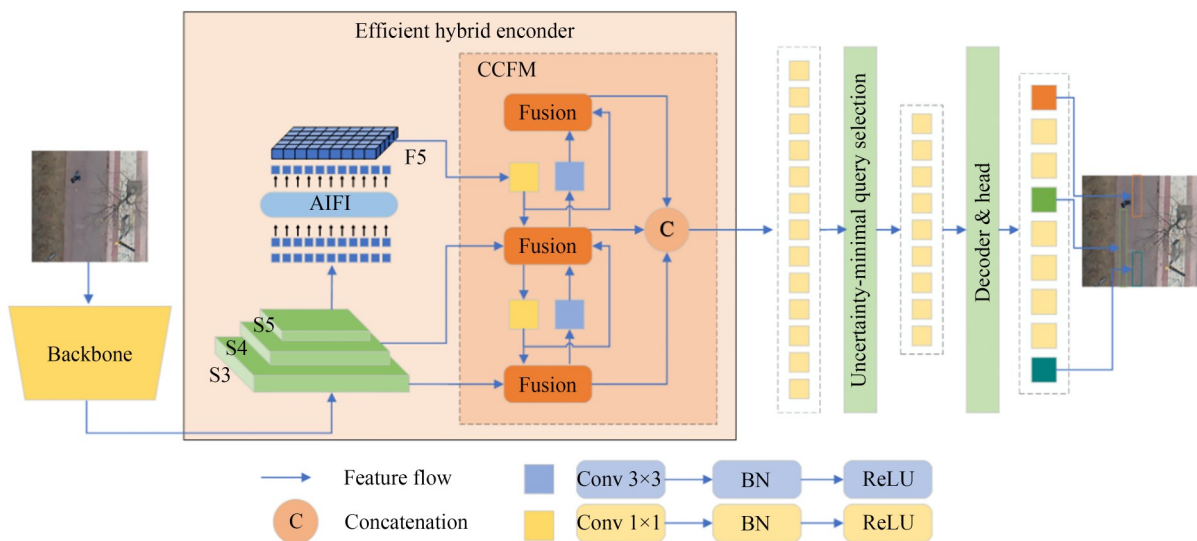


图 1 RT-DETR 模型结构图

Fig. 1 RT-DETR model structure diagram

该模型以 ResNet 为主干网络提取多尺度图像特征,通过混合编码器整合 AIFI 模块的尺度内交互与 CCFM 模块的跨尺度融合,有效平衡全局上下文建模与计算效率。其解码器采用 IoU 感知查询选择机制,动态筛选初始查询特征,并借助检测头进行迭代优化,直接输出目标边界框及置信度,在避免传统 NMS 后处理的同时实现实时检测。通过特征解耦、查询动态优化与集合预测机制,可以为路面损伤快速精准识别提供一种解决方案。

### 3 RSF-DETR 模型

#### 3.1 RSF-DETR 整体网络架构

为提升对路面损伤的检测能力,本文基于以 ResNet18 为 Basicblock 的 RT-DETR 方法进行改进,其结构如图 2 所示。

RSF-DETR 首先经空频双域特征增强模块

FreSCal 来改进主干网络提取图像特征,通过融合空间细节与频域纹理的互补特性,强化模型对局部细微结构的敏感度。

在混合编码器部分,AIFI 模块进行尺度内交互,并利用上下文引导空间特征重构金字塔网络 RSDFPN,构建尺度感知的语义金字塔与动态特征融合机制,来增强模型对多尺度目标的特征融合能力,通过动态组卷积混洗变换与 Transformer 协同优化 DGCST<sup>[13]</sup> 模块来提升特征信息交互。

在解码器部分,RSF-DETR 延续 RT-DETR 中的 IoU 感知查询选择策略,从图像特征中筛选出固定数量的特征作为解码器的初始查询对象。最后经过解码器借助辅助预测头对这些查询进行迭代优化,生成目标的边界框和置信度,从而输出最终的检测结果。

为清晰呈现 RSF-DETR 在 RT-DETR 基础上的具体改动,其结构对比如表 1 所示。

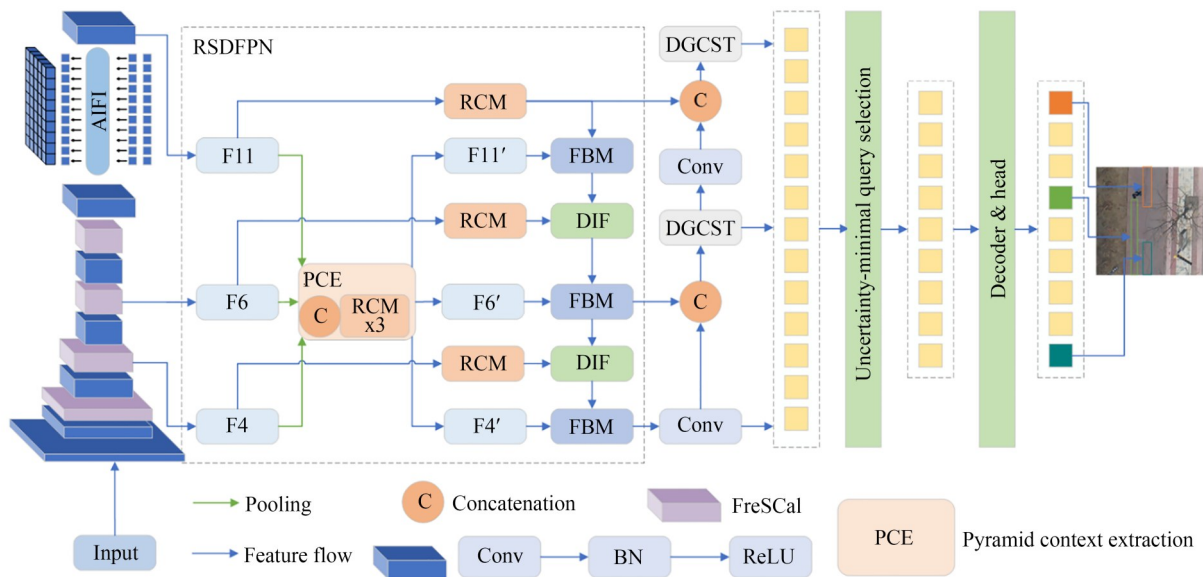


图 2 RSF-DETR 模型结构图

Fig. 2 RSF-DETR model structure diagram

表 1 RT-DETR 与 RSF-DETR 结构对比

Tab. 1 Comparison of RT-DETR and RSF-DETR Structures

组件	RT-DETR	RSF-DETR	位置	作用
主干网络	ResNet18	Conv+FreSCal 模块	主干网络	增强空间边缘细节与频域全局纹理信息提取能力
特征金字塔	CCFM (跨尺度特征融合模块)	RSDFPN 网络(替换 CCFM)	混合编码器	构建尺度感知语义金字塔,动态融合多尺度特征
特征交互模块	RepC3	DGCST 模块(替换 RepC3)	混合编码器	动态分组卷积混洗优化局部计算,融合 Transformer 全局建模

### 3.2 空频双域特征增强 FreSCal 模块

RT-DETR 以 ResNet 作为主干网络提取多尺度图像特征,主要依靠卷积操作来完成特征提取。在对路面目标信息提取时,该结构对边缘信息的提取能力不足,且对全局纹理的感知能力有限,难以精准捕捉裂缝等损伤的细节。因此,本文引入空频双域特征增强模块 FreSCal 模块改进主干网络来增强图像特征提取能力。

FreSCal 模块通过分别在空间域捕捉图像的边缘信息以及在频域中捕捉图像的低频和低频信息,从而在全局和局部尺度上提取信息,最后将空间域和频域的特征进行加权相加实现特征融合,使模型能同时考虑空间结构信息和频率信息,大幅增强了图像特征的表达能力,其结构如图 3 所示。

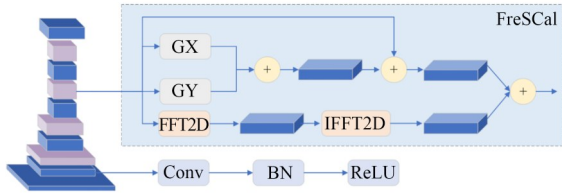


图 3 FreSCal 模块结构图

Fig. 3 FreSCal module structure diagram

首先使用 Scharr 算子来计算图像的  $x$  和  $y$  方向梯度,生成高频边缘特征图,这些特征图经过归一化处理后进行合并,以强化图像中裂缝、坑洞等损伤的局部空间结构,具体公式如式(1):

$$G_{\text{mag}} = \sqrt{G_x^2 + G_y^2}, G_{\text{norm}} = \frac{G_{\text{mag}}}{\max(G_{\text{mag}})}, \quad (1)$$

其中: $G_x, G_y$  表示水平和垂直方向梯度图, $G_{\text{mag}}$  表示梯度幅值, $G_{\text{norm}}$  表示归一化梯度图。随后通过卷积层逐级增强这些高频细节,同时引入残差连接来缓解梯度消失的问题,可以确保细节信息的完整传递,进而能够有效地捕捉图像中的细节和边缘信息。这时在频域分支中,模块通过快速傅里叶变换 FFT2D 将图像从空间域映射到频率域,并进行卷积操作以捕捉图像的全局特征和高频纹理细节,公式为:

$$F(I)(u, v) = \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} I(x, y) \cdot e^{-j2\pi\left(\frac{ux}{W} + \frac{vy}{H}\right)}, \quad (2)$$

其中: $F(I)$  表示图像( $I$ )的频域, $(u, v)$  表示频域坐标, $(x, y)$  表示空间域坐标, $W, H$  为图像尺寸,

$e^{-j2\pi\left(\frac{ux}{W} + \frac{vy}{H}\right)}$  为复正弦基函数,用于分解图像频率成分。

$$F\{I * k\}(u, v) = F(u, v) \odot K(u, v), \quad (3)$$

$$F(u, v) = F(I)(u, v) \odot K_{\text{freq}}(u, v), \quad (4)$$

其中, $K_{\text{freq}}(u, v)$  表示可学习频域卷积核,公式(3)利用卷积定理,频域乘积等价于空间域循环卷积,因此该操作实际上对原图施加了一个全局纹理滤波器。

之后,通过逆快速傅里叶变换 IFFT2D 将数据转换回空间域,并再次应用卷积层以进一步提取和整合特征,这样可以有效捕捉图像的低频和低频成分,帮助模型在不同尺度上提取信息,公式为:

$$I_{\text{recon}}(x, y) = \frac{1}{W \cdot H} \sum_{u=0}^{W-1} \sum_{v=0}^{H-1} F_{\text{out}}(u, v) \cdot e^{j2\pi\left(\frac{ux}{W} + \frac{vy}{H}\right)}, \quad (5)$$

$$\sum_{x, y} |I_{\text{recon}}[x, y]|^2 = \frac{1}{W \cdot H} \sum_{u, v} |\hat{F}(u, v)|^2, \quad (6)$$

其中, $I_{\text{recon}}$  表示从频域重建的空间域图像,式(6)根据帕塞瓦尔定理,以保证整体能量在空频变换过程中守恒。

最终将空间域和频域分支提取的特征进行加权融合,生成综合特征图,显著提升对道路质量信息的表达能力,公式为:

$$F_{\text{fused}} = \alpha \cdot F_{\text{spatial}} + \beta \cdot F_{\text{freq}}, \quad (7)$$

其中: $\alpha, \beta$  表示可学习权重,初始值各为 0.5,默认(0.5, 0.5)用于保证跨数据的平稳收敛, $F_{\text{spatial}}$  表示空间域分支提取的特征, $F_{\text{freq}}$  表示频域分支提取特征。

经过 FreSCal 模块处理后,模型通过跨步卷积逐步聚合多尺度特征,结合空间域和频域特征提取的优势,显著提升对道路损伤信息的捕捉能力以及与背景的分度,最终输出融合空频双域优势的特征图,为路面损伤检测提供鲁棒的特征表示。

### 3.3 上下文引导重构网络 RSDFPN

在特征融合阶段,RT-DETR 通过融合主干网络最后三个阶段的特征图来整合目标特征,但在应用于路面损伤检测的场景时,特征在多尺度融合中易被忽略或丢失,并且缺乏有效的上下文引导机制,无法充分利用全局上下文信息来优化特征表示。因此,采用上下文引导空

间特征重构 RSDFPN 网络来提升模型对目标的检测能力。

RSDFPN 网络受 CGRSeg<sup>[14]</sup> (Context-Guided Spatial Feature Reconstruction for Efficient Se-

matic Segmentation) 网络启发, 通过上下文引导的空间特征重构提升模型对多尺度特征的融合能力以及对小目标的检测性能, 其结构如图 4 所示。

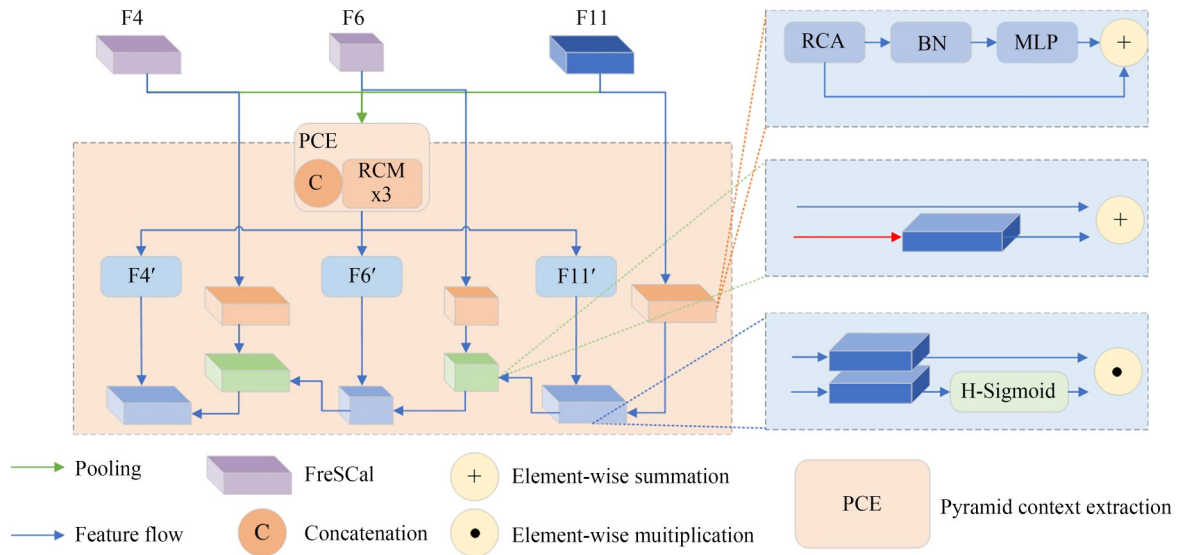


图 4 RSDFPN 网络结构图

Fig. 4 RSDFPN network structure diagram

模型首先通过不同尺度的特征提取生成不同分辨率的特征图, 随后对这些特征图 ( $F_4, F_6, F_{11}$ ) 进行池化操作形成金字塔特征, 并将这些金字塔特征送入多个堆叠的 RCM (Rectangular Self-Calibration Module) 模块中, 通过特征交互提取尺度感知的语义特征。

$$P = AP(F, r), \quad (8)$$

其中:  $AP(F, r)$  表示平均池, 它对特征  $F$  进行了一个给定因素的降采样,  $P$  表示包含金字塔形上下文的特征。

RCM 模块通过捕获金字塔上下文信息优化特征表示, 包含矩形自校准注意力 RCA (Rectangular Self-Calibration Attention)、批量归一化 BN (Batch Normalization) 和多层感知机 (Multi-Layer Perceptron, MLP)。RCA 通过水平及域的进行有效建模, 融入形状自校准机制来进行优化。并采用两个大核条带卷积分置水平、垂直方向分别对注意力图进行校准, 网络首先在水平方向上, 通过条带卷积对各元素所在行调整, 然后采用批量归一化对特征进行归一化处理, 其次使用

ReLU 实现非线性激活, 最后再通过垂直条带卷积对垂直形状进行校准, 从而实现对小目标的精准定位和特征优化。

在金字塔上下文提取完成后, RSDFPN 网络通过动态插值融合技术将低层次特征与高层次特征进行融合, 这一过程通过多特征融合模块 FBM (Feature Fusion Block) 实现, 并将来自主干网络的低层次空间特征与金字塔上下文模块提取的高层次特征进行结合, 经过动态插值技术使融合后的特征更好地保留小目标的语义信息, 同时增强了模型对复杂背景的适应能力。具体过程:

$$CGR = DIF(FBM(RCM(F), P)). \quad (9)$$

经过上下文引导的空间特征重构后的特征显著提升了模型对路面损伤目标的检测能力。

### 3.4 DGCST 模块

RT-DETR 的 RepC3 模块在特征融合上存在固定策略和动态调整不足的问题, 在对路面损伤检测中易导致目标特征丢失, 影响检测精度。因此, 采用 DGCST 模块替换 RepC3 模块,

减少冗余计算,并融合局部与全局特征建模,以提升特征提取效率和表达能力,其结构如图 5 所示。

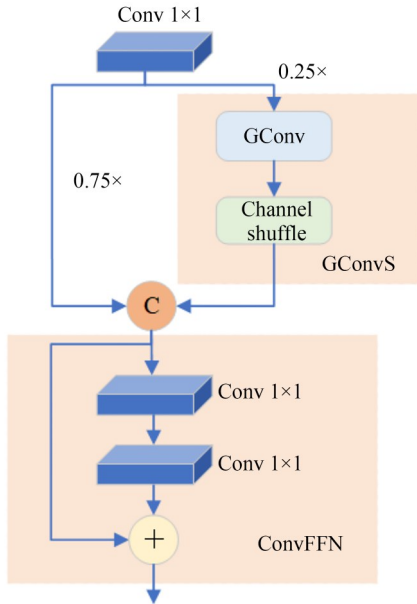


图 5 DGCST 模块结构图

Fig. 5 DGCST module structure diagram

该模块首先沿通道维度将输入特征图按 3:1 的比例分为两条路径,一部分用于分组卷积和通道混洗操作,通过先将输入特征图进行分组卷积,根据网络深度和通道数的调整分组策略,捕捉特征之间的相关性,减少冗余计算,公式为:

$$G = \min \left( G_{\max}, \frac{C_{\text{in}}}{k \cdot \sqrt{d}} \right), \quad (10)$$

其中:  $C_{\text{in}}$  为输入通道数,  $d$  为网络深度,  $k$  为可调比例系数,  $G_{\max}$  为最大分组数。

随后通过通道混洗在不引入额外参数的条件下重新排列通道顺序,打破组间隔离,促进跨组信息流动。与此同时,另外一部分直接进行卷积处理以保留原始特征信息,通过分组卷积和通道混洗的特征图与直接卷积处理后的特征图进行融合,整合两条路径的特征信息。将融合后的特征图引入基于 Transformer 的卷积前馈网络 ConvFFN,通过多尺度卷积层堆叠进一步捕捉特征图的局部关系和上下文信息,提升特征的代表能力,输出优化后的特征图。

## 4 实验

### 4.1 数据集介绍

本次实验使用的数据集为 RDD2022<sup>[15]</sup>,该开源数据集涵盖了来自中国、日本、美国、捷克、挪威以及印度等多个国家的道路破损图像。在本次研究中,仅选取中国的道路损伤图像,并剔除了一些未标记路面质量的样本,筛选出 4 373 张道路破损图像作为本次实验的数据集。这些图像涵盖了 5 种常见的路面病害类型,分别为: D00(纵向裂缝)、D10(横向裂缝)、D20(网状裂缝)、D40(坑洞)以及 Repair(修补)。在本次实验中,将 RDD2022 数据集中的中国道路损伤图像按照 7:1:2 的比例随机划分为训练集、验证集和测试集<sup>[16]</sup>,其中训练集包含 3 061 张图像,验证集包含 437 张图像,而测试集则包含 875 张图像。表 2 呈现了 RDD2022 中国道路损伤数据集中每种损伤类别所在的图像分布以及标注实例数量统计。

表 2 RDD2022 中国数据集详细信息

Tab. 2 Detailed information of RDD2022 Chinese dataset

Category	Number of images	Number of annotations
D00	2 453	4 104
D10	1 761	2 359
D20	750	934
D40	228	321
Repair	803	1 076

在分析实验中,选用的数据集为 UAV-PDD2023<sup>[17]</sup>,该数据集的图像数据来源于无人机对道路路面的巡查拍摄,标注了六种道路损坏类型,分别为: Longitudinal crack(纵向裂缝)、Transverse crack(横向裂缝)、Alligator crack(鳄鱼裂缝)、Oblique crack(斜裂缝)、Repair(修复)、Pothole(坑洼)。该数据集剔除部分未标记图像共 2 425 张路面损失图像,划分比例为 8:1:1,其中训练集、验证集和测试集分别为 1 940 张图像、242 张图像和 243 张图像。表 3 呈现了 UAV-PDD2023 数据集中每种损伤类别所在的图像分布以及标注实例数量统计。

选用从 RDD2022 数据集中选取的挪威路段图像,剔除未标注及损坏样本后保留 2 652 张清

表 3 UAV-PDD2023数据集详细信息

Tab. 3 Detailed information of UAV-PDD2023 dataset

Category	Number of images	Number of annotations
Longitudinal crack	2 152	5 398
Transverse crack	408	603
Alligator crack	843	1 686
Oblique crack	1 230	2 994
Repair	132	195
Pothole	180	282

晰图像,使用数据增强技术模拟雨天场景,生成对应的雨天图像。按 8:1:1 划分训练集、验证集和测试集,用于验证模型在雨况下的鲁棒性。原始图像与模拟雨天效果对比如图 6 所示。



图 6 原始图片和模拟雨天图片结果对比

Fig. 6 Comparison of the original image and simulated rainy day image results

## 4.2 实验环境配置

本次实验在 Ubuntu 20.04 操作系统下进行,硬件配置为 NVIDIA RTX 4090 GPU 和运行内存 24GB。使用 Python 3.8 和 PyTorch 2.0.0 深度学习框架进行模型训练,使用 cuda11.8 版本进行加速计算,输入图像尺寸设定为  $640 \times 640$  pixels,批量大小(batch size)为 8,配置 4 个工作进程(workers=4)以并行处理数据加载任务,初始学习率设为 0.001,训练轮次(epochs)设为 250,并采用随机梯度下降算法(Stochastic Gradient Descent, SGD)作为优化器来更新模型参数。

## 4.3 评价指标

本实验通过多种指标对模型性能进行评估,包括参数量(Params)、平均检测精度(mAP)、精确率(Precision)、召回率(Recall)、计算量(G-FLOPs)以及每秒检测帧数(FPS)。

精确率(Precision)是评估模型预测准确性的指标,反映了模型预测为正例的样本中实际为正

例的比例。 $TP$ 表示被模型正确预测为正例的样本数量, $FP$ 表示被错误预测为正例的样本数量,计算公式为:

$$precision = \frac{TP}{TP + FP}. \quad (11)$$

召回率(Recall)表示在所有实际为正例的样本中,被模型正确预测为正例的比例。 $FN$ 表示实际为正例但被模型错误预测为负例的样本数量,计算公式为:

$$recall = \frac{TP}{TP + FN}. \quad (12)$$

AP(Average Precision)是精确率-召回率( $P-R$ )曲线下面积,以  $R$  为横轴,以  $P$  为纵轴,绘制  $P-R$  曲线,计算公式为:

$$AP = \int_0^1 P(R) dR. \quad (13)$$

mAP 为各类别平均精度的均值,其中  $N$  为  $N$  个检测类别目标,计算公式如下:

$$mAP = \frac{\sum_{i=1}^N AP_i}{N}. \quad (14)$$

## 4.4 消融实验

在本次实验中,以 RT-DETR 模型作为基线模型,并通过参数量、精确率、召回率、GFLOPs 和检测精度(mAP@0.5, mAP@0.5:0.95)作为评价指标进行消融研究,根据实验结果来分析改进方法中各个模块对整体模型性能的影响,实验结果如表 4 所示。

根据表 4 的实验结果可知,经过 FreSCal 模块改进模型的主干网络后,模型参数量略有增加,GFLOPs 从 57.0 G 上升到 60.6 G,但 mAP@0.5 从 83.9% 提升至 84.6%,mAP@0.5:0.95 从 54.4% 提升至 54.7%,而且召回率和精确率也都提高。通过采用 RSDFPN 网络进行改进,可以看出参数量和计算量分别降低了 3% 和 15%,且 mAP@0.5 提升了 1.2%,mAP@0.5:0.95 提升了 0.6%,有效提升了模型对目标的区分和定位能力。使用 DGCST 模块替换混合编码器中的 RepC3 模块后,检测精度 mAP@0.5 提升了 1%,mAP@0.5:0.95 提升了 0.4%,且参数量和计算量均有所下降,这表明使用 DGCST 模块替换 RepC3 模块对整体模型产生了积极影响。当 FreSCal 模块与 RSDFPN 结合时,检测精度相比单独使用 FreSCal 模块 mAP@0.5, mAP@0.5

表 4 消融实验结果

Tab. 4 Results of ablation test

	FreSCal	RSDFPN	DGCST	Params/M	P	R	mAP@0.5	mAP@0.5:0.95	GFLOPs
0				19.9	0.847	0.787	0.839	0.544	57.0
1	✓			20.4	0.873	0.815	0.846	0.547	60.6
2		✓		19.2	0.852	0.808	0.851	0.550	48.2
3			✓	19.2	0.852	0.814	0.849	0.548	55.6
4	✓	✓		21.0	0.874	0.793	0.853	0.558	58.3
5	✓		✓	19.8	0.869	0.792	0.848	0.552	59.3
6		✓	✓	18.6	0.878	0.786	0.850	0.553	46.8
7	✓	✓	✓	20.3	0.875	0.794	0.858	0.561	56.9

:0.95分别提高了0.7%,1.1%,进一步增强了模型对目标的检测能力。当FreSCal模块与DGCST模块配合使用时,参数量和计算量比单独使用FreSCal模块有所降低。此外当RSDFPN与DGCST模块结合时,参数量和计算量也比之前有所减少,提升了模型的轻量化程度。

最后经过上述三个方法改进后相较于基线方法参数量虽略有增加,但计算量些微下降,同时精确率和召回率也都有所提高,且检测精度mAP@0.5, mAP@0.5:0.95最终分别达到了85.8%和56.1%。由此可见,改进后的模型整体性能显著提升。

为了更加直观地比较基线方法和改进方法的优劣,对两种方法使用热力图技术进行可视化分析,其可视化分析结果如图7所示。

在(a)部分的图中,可以看出基线方法大部分关注了目标区域,却漏失了中间一小部分没有关注,而改进方法对所检测的区域,则是大部

分的对目标都进行了关注,虽然也有部分细节遗漏,但比基线方法则更加全面地进行了关注所要检测的目标区域。而在(b)部分的图中,可知基线方法和改进方法都检测到了目标区域,但基线方法的检测效果却不如改进方法。

因此,改进方法引入空频双域特征增强FreSCal模块和上下文引导重构网络RSDFPN,通过捕捉高频细节和提升多尺度特征融合能力,能够覆盖更完整的目标区域,并且聚焦于目标区域,此外通过DGCST模块来增强信息交互,显著提升对目标区域的感知能力。

#### 4.5 对比实验

为验证RSF-DETR方法相对于其他主流算法的检测性能,选取参数量(Params)、检测精度(mAP@0.5, mAP@0.5:0.95)、计算量(GFLOPs)和每秒检测帧数(FPS)来作为评价指标,在RDD2022中国数据集上与Faster-RCNN<sup>[18]</sup>, SSD<sup>[19]</sup>, YOLOv5m, YOLOv6s<sup>[20]</sup>, YOLOv8m, YOLOv9m<sup>[21]</sup>, YOLOv10b<sup>[22]</sup>, YOLO11m, YOLO12m, Mamba-YOLO<sup>[23]</sup>, Deformable-DETR<sup>[24]</sup>, DAB-DETR<sup>[25]</sup>以及RT-DETR等主流算法进行对比实验。

根据表5实验结果,可以看出Faster-RCNN在模型参数量上明显高于其他模型,但在检测精度和计算量方面的表现不如部分其他模型,且其检测精度仅略优于YOLOv6s。改进方法相较于Faster-RCNN, SSD, YOLOv5m, YOLOv8m, Deformable-DETR, DAB-DETR和Mamba-Yolo等模型,参数量更低,计算量仅略高于Mamba-YOLO,但在检测精度方面均优于这些模型。相较于YOLOv6s,尽管改进方法的计算量和参数

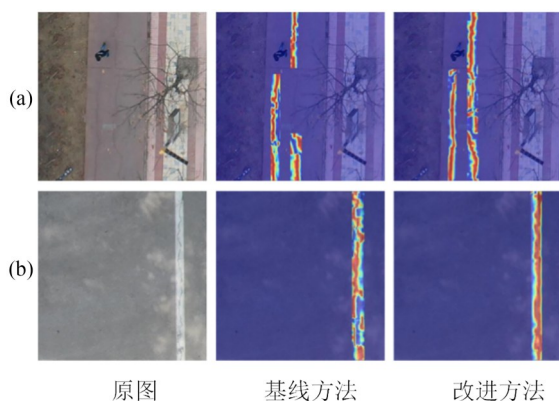


图7 基线方法和改进方法的热力图结果对比

Fig. 7 Comparison of thermodynamic chart results between baseline method and improved method

表 5 对比实验结果

Tab. 5 Comparison of experimental results

Model	Params/M	mAP@0.5	mAP@0.5:0.95	GFLOPs	FPSbs=1
Faster-RCNN	137.1	0.791	0.501	370	71.5
SSD	26.3	0.777	0.485	62	78.2
YOLOv5m	25.0	0.852	0.556	64.0	229.5
YOLOv6s	16.3	0.790	0.490	44.0	325.0
YOLOv8m	25.8	0.837	0.546	78.7	261.5
YOLOv9m	20.0	0.836	0.560	76.5	158.7
YOLOv10b	20.4	0.803	0.532	98.0	166.5
YOLO11m	20.0	0.828	0.552	67.7	198.8
YOLO12m	20.1	0.808	0.526	67.1	165.7
Mamba-YOLO	21.8	0.850	0.559	49.6	70.3
Deformable-DETR	40.1	0.853	0.552	123.0	45.1
DAB-DETR	43.7	0.852	0.547	65.3	47.6
RT-DETR	19.9	0.839	0.544	57.0	66.4
Ours	20.4	0.858	0.561	56.9	52.3

量略高,但检测精度优于YOLOv6s。相较于YOLOv8m, YOLO11m和YOLO12m,本文的改进方法参数量略高,但计算量更低,且检测精度更高。而与参数量相同的YOLOv10b相比,改进方法在检测精度和计算量上表现出更优的性能。

与基线模型RT-DETR相比时,改进方法的参数量略高,但计算量有所降低,且mAP@0.5和mAP@0.5:0.95分别提升至85.8%和56.1%,并根据表4可以看出改进方法在检测精度方面优于所有对比模型,表明其对于路面的各

种损伤的目标检测能力更强,尽管改进方法推理速度为52.3 FPS低于部分模型,但在实际应用场景中仍能满足检测速度的要求。

综上所述,RSF-DETR在计算复杂度、检测速度和检测精度之间实现了有效平衡,相较于其他模型具有显著优势。

#### 4.6 不同划分比例验证实验

将RDD2022数据集中的中国道路损伤图像按照8:1:1的比例随机划分为训练集、验证集和测试集<sup>[26]</sup>,实验结果如表6所示。

表 6 不同划分比例的实验结果

Tab. 6 Experimental results with different division ratios

Strategy	P	R	mAP@0.5	mAP@0.5:0.95
Baseline	0.850	0.816	0.844	0.556
Baseline+FreSCal	0.864	0.820	0.853	0.557
Baseline+FreSCal+RSDFPN	0.871	0.838	0.861	0.565
Baseline+FreSCal+RSDFPN+DGCST	0.872	0.831	0.866	0.577

由表6可知,改进方法在不同划分比例中检测精度也有所提升,检测精度mAP@0.5提高了2.2%,mAP@0.5:0.95提高了2.1%,证明了改进方法的有效性。

#### 4.7 可视化分析

为进一步体现改进方法的检测效果,本文选取了三张图片进行评估,并将其与部分主流算法进行可视化对比分析,其可视化结果如图8所示。

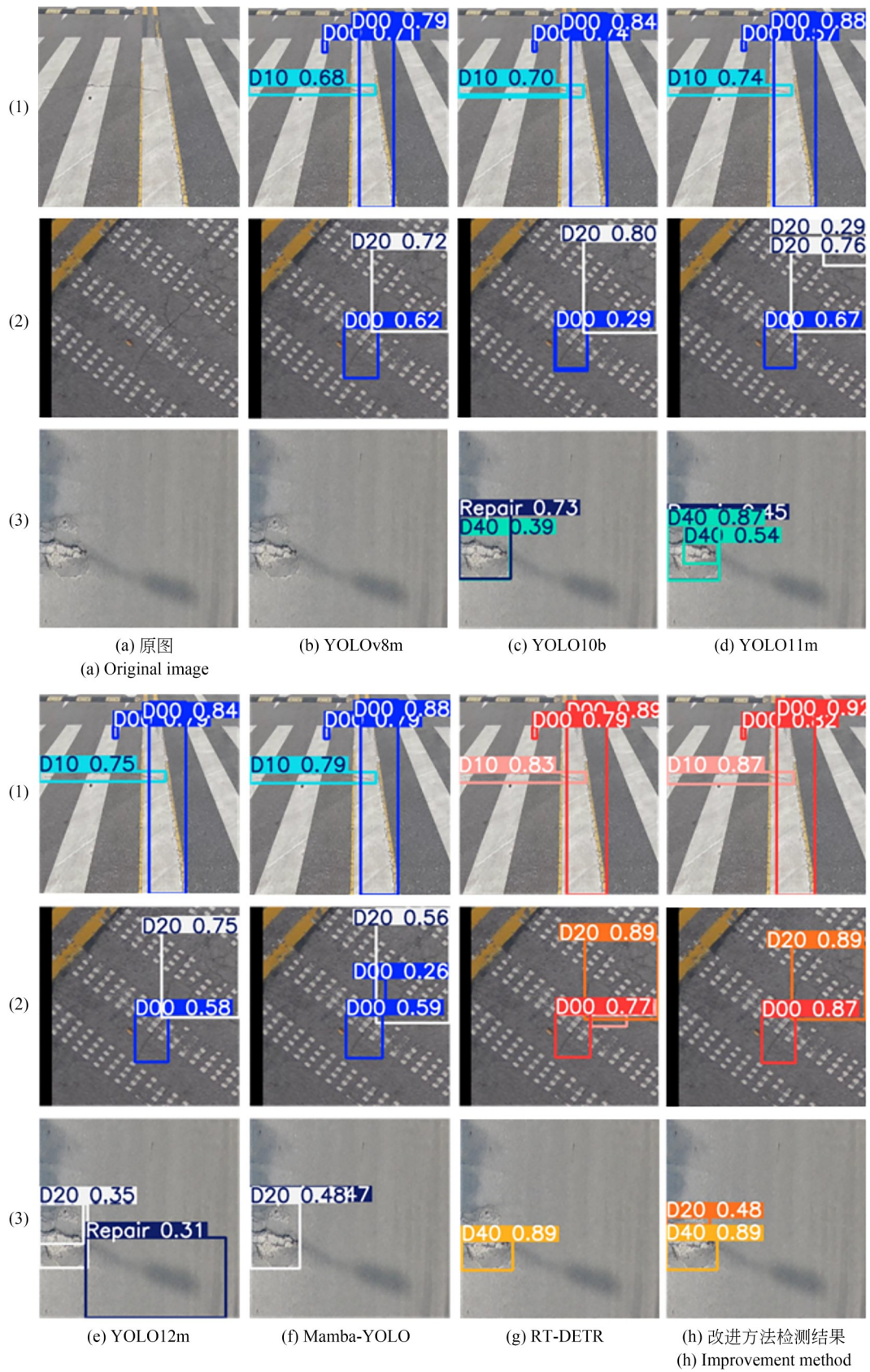


图 8 不同模型检测结果可视化分析

Fig. 8 Visual analysis of test results of different models

如图 8, (a) 为图片的三张原图, 在 (b) 部分中, 可以看出 YOLOv8m 在图片 (1), (2) 中均检测到了目标结果, 却在图片 (3) 中出现漏检现象, 并没有检测到任何信息, 这可能是因为在特征提取过程中, 未能充分有效地提取关键特征或在高层特征图阶段发生了特征丢失, 从而导致了漏检现象的发生。在 (c) 和 (e) 部分中可知, YOLOv10b 和 YOLO12m 在图片 (2), (3) 中均成功检测出了目标对象, 不过 YOLO12m 在图片 (3) 的检测结果中出现了误检现象, 而 YOLOv10b 则只检测出一种类别, 并且也出现误检现象。在 (d), (f) 部分中, YOLOv11m 和 Mamba-YOLO 在图片 (2), (3) 的检测结果中, 均出现了误检的现象, 无法精准地识别并检测出目标信息。

图 8 中的 (g) 和 (h) 分别为基线方法与改进方法的可视化检测结果, 在图片 (1) 中, 基线方法虽未出现漏检, 但其检测效果的确不如改进方法, 且改进方法针对小目标的检测效果要优于基线方法的检测效果, 并在与其他方法对比时, 也能更加精确的识别出目标。对比图片 (2) 的检测结果时, 可以看出基线方法可能由于受复杂背景的

影响, 错误地检测出了 D10 类别的目标, 而原图中并不存在该类别目标, 出现了较为明显的误检现象。相比之下, 改进方法精准地检测出了目标对象, 且未出现错误检测的现象。在对图片 (3) 进行检测时, 基线方法未能检测出图像左侧的 D20 类别的目标, 出现了漏检现象, 而改进方法则不受任何影响, 准确地识别并检测出了该目标。

在图 8 的可视化检测结果中, 改进方法相较于其他方法, 能够更为精确识别目标对象, 有效地减少了漏检和误检的情况, 且各目标的检测置信度也高于其他方法。

#### 4.8 分析实验

为验证改进方法的泛化性和鲁棒性, 本文在 UAV-PDD2023 数据集和选用 RDD2022 中的挪威路段模拟雨天图像数据集进行检测结果分析。

表 7 展示了改进方法在 RDD2022 中的挪威路段模拟雨天图像数据集上的消融实验结果, 可知改进方法相较于基线方法精确率和召回率有所提高, 检测精度  $mAP@0.5$  提高了 1.8%, 并且  $mAP@0.5:0.95$  提高了 1.4%。

表 7 改进方法在模拟雨天图像消融实验结果

Tab. 7 Improved method for simulating the experimental results of image ablation on rainy days

Strategy	P	R	$mAP@0.5$	$mAP@0.5:0.95$
Baseline	0.450	0.286	0.275	0.107
Baseline+FreSCal	0.458	0.286	0.283	0.111
Baseline+FreSCal+RSDFPN	0.479	0.293	0.286	0.116
Baseline+FreSCal+RSDFPN+DGCST	0.504	0.303	0.293	0.121

表 8 展示了改进方法在 UAV-PDD2023 数据集上的消融实验结果, 可知在基线方法上使用 FreSCal 模块改进主干网络时, 检测精度  $mAP@0.5$  相较于基线方法提高了 1.7%, 在此基础上使用 RSDFPN 网络进行改进后, 检测精度  $mAP@$

0.5 提高了 1.1%, 最后加入 DGCST 模块后, 检测精度  $mAP@0.5$  最终达到了 85.8%, 且检测精度  $mAP@0.5:0.95$  最终达到了 55.8%。实验结果显示改进方法提高了在 UAV-PDD2023 数据集上的检测精度。

表 8 改进方法在 UAV-PDD2023 消融实验结果

Tab. 8 Improvement method in UAV-PDD2023 ablation experiment results

Strategy	P	R	$mAP@0.5$	$mAP@0.5:0.95$
Baseline	0.845	0.768	0.821	0.519
Baseline+FreSCal	0.863	0.793	0.838	0.545
Baseline+FreSCal+RSDFPN	0.871	0.805	0.849	0.550
Baseline+FreSCal+RSDFPN+DGCST	0.888	0.816	0.858	0.558

表 9 和表 10 分别展示了基线方法和改进方法在 UAV-PDD2023 数据集中各具体类别的精确率、检测精度和召回率的实验结果,可知每个类别的检测精度 mAP@0.5 都有所提升,而且在总的检测精度中 mAP@0.5 提高了 3.7%, mAP@0.5:0.95 提高了 3.9%,并且召回率和精确率都有所提高。

表 9 基线方法在 UAV-PDD2023 数据集的检测结果

Tab.9 Detection results of the baseline method on the UAV-PDD2023 dataset

Type	P	R	mAP@0.5	mAP@0.5:0.95
Longitudinal crack	0.923	0.821	0.870	0.563
Transverse crack	0.826	0.630	0.721	0.512
Alligator crack	0.888	0.760	0.806	0.483
Oblique crack	0.877	0.804	0.804	0.544
Repair	0.869	0.808	0.808	0.442
Pothole	0.689	0.778	0.778	0.567
All	0.845	0.768	0.821	0.519

由上述分析结果可知改进方法在 UAV-PDD2023 数据集和选用 RDD2022 中的挪威路段模拟雨天图像数据集上的有效性,实验结果进一步验证了改进方法在路面损伤检测中的泛化性和鲁棒性。

## 5 结 论

该研究针对路面损伤检测改进了 RT-DETR 模型,致力于提升对于路面损伤的检测能力。模型的核心是首先引入空频双域特征增强 FreSCal

### 参考文献:

- [1] YANG L, BAI S L, LIU Y H, *et al.* Multi-scale triple-attention network for pixelwise crack segmentation[J]. *Automation in Construction*, 2023, 150: 104853.
- [2] LI S B, HUANG Y J. Damage detection algorithm based on faster-RCNN[C]. *2023 5th International Conference on Electronics and Communication, Network and Computer Technology (ECNCT)*. August 18-20, 2023, Guangzhou, China. IEEE, 2023:

表 10 改进方法在 UAV-PDD2023 数据集的检测结果

Tab.10 Detection results of the improved method on the UAV-PDD2023 dataset

Type	P	R	mAP@0.5	mAP@0.5:0.95
Longitudinal crack	0.934	0.869	0.911	0.590
Transverse crack	0.917	0.722	0.786	0.509
Alligator crack	0.887	0.782	0.828	0.510
Oblique crack	0.894	0.812	0.870	0.585
Repair	0.940	0.846	0.891	0.509
Pothole	0.757	0.867	0.861	0.646
All	0.888	0.816	0.858	0.558

模块,增强网络对边缘信息的提取能力,其次通过构建尺度感知的语义金字塔与动态特征融合机制的 RSDFPN 网络,增强模型多尺度特征融合的能力,最后使用 DGCST 模块来替换 RepC3 模块,增强了局部特征的表征能力,进一步改善了图像识别效果,RSF-DETR 以小幅降速换取显著精度提升,其 52.3 的 FPS 适配适用于对实时性要求适中的道路巡检中,但在高帧率车载场景可能受限。通过实验结果分析,改进后的方法在一定程度上提高了对路面损伤检测的精度。

### 作者贡献声明:

周冬梅:论文构思,获取资助,审核与编辑写作;

件兵兵:论文构思,方法论,验证,可视化呈现,初稿编写;

刘小明:数据管理,形式分析;

闫浩文:调查研究,项目管理;

吴小所:提供资源,研究监管和指导。

177-180.

- [3] SAMI A A, SAKIB S, DEB K, *et al.* Improved YOLOv5-based real-time road pavement damage detection in road infrastructure management[J]. *Algorithms*, 2023, 16(9): 452.
- [4] WANG X Q, GAO H B, JIA Z M, *et al.* BL-YOLOv8: an improved road defect detection model based on YOLOv8[J]. *Sensors*, 2023, 23(20): 8361.
- [5] HONG Z H, YANG F, PAN H Y, *et al.* Highway crack segmentation from unmanned aerial vehicle images using deep learning[J]. *IEEE Geoscience and*

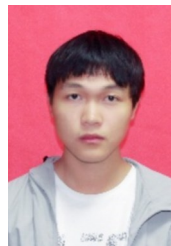
- Remote Sensing Letters*, 2021, 19: 6503405.
- [6] LI J, QU Z, WANG S Y, *et al.* A method of road damage detection for complex background images based on region guidance network[J]. *Pattern Recognition*, 2025, 168: 111780.
- [7] HAN C J, YANG H D, YANG Y W. Enhancing pixel-level crack segmentation with visual mamba and convolutional networks[J]. *Automation in Construction*, 2024, 168: 105770.
- [8] WANG Q Y, CHEN B. A novel transfer learning model for the real-time concrete crack detection[J]. *Knowledge-Based Systems*, 2024, 301: 112313.
- [9] LIN C M, TIAN D X, DUAN X T, *et al.* TransCrack: revisiting fine-grained road crack detection with a transformer design[J]. *Philosophical Transactions Series A, Mathematical, Physical, and Engineering Sciences*, 2023, 381(2254): 20220172.
- [10] WANG N N, SHANG L H, SONG X T. A transformer-optimized deep learning network for road damage detection and tracking[J]. *Sensors*, 2023, 23(17): 7395.
- [11] DENG Y, MA J, WU Z Y, *et al.* DSR-Net: Distinct selective rollback queries for road cracks detection with detection transformer[J]. *Digital Signal Processing*, 2025, 164: 105266.
- [12] ZHAO Y A, LV W Y, XU S L, *et al.* DETRs beat YOLOs on real-time object detection [C]. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 16-22, 2024, Seattle, WA, USA. IEEE, 2024: 16965-16974.
- [13] GONG W. Lightweight object detection: A study based on yolov7 integrated with shufflenetv2 and vision transformer [J]. *arXiv preprint arXiv: 2403.01736*, 2024.
- [14] NI Z L, CHEN X H, ZHAI Y J, *et al.* Context-guided Spatial Feature Reconstruction for Efficient Semantic Segmentation[M]. *Computer Vision-ECCV 2024*. Cham: Springer Nature Switzerland, 2024: 239-255.
- [15] 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, 11(4): 846-862.
- [16] SONG J, SUN Y, CAI Z Y, *et al.* DUSA-UNet: dual sparse attentive U-Net for multiscale road network extraction[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 5624514.
- [17] YAN H H, ZHANG J F. UAV-PDD2023: a benchmark dataset for pavement distress detection based on UAV images[J]. *Data in Brief*, 2023, 51: 109692.
- [18] REN S 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.
- [19] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: Single Shot Multibox Detector [M]. *Computer Vision-ECCV 2016*. Cham: Springer International Publishing, 2016: 21-37.
- [20] LI C, LI L, JIANG H, *et al.* YOLOv6: A single-stage object detection framework for industrial applications[J]. *arXiv preprint arXiv: 2209.02976*, 2022.
- [21] WANG C Y, YE H I H, MARK LIAO H Y. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information [M]. *Computer Vision-ECCV 2024*. Cham: Springer Nature Switzerland, 2024: 1-21.
- [22] CHEN H, CHEN K, DING G G, *et al.* YOLOv10: Real-Time end-to-end Object Detection [C]. *Advances in Neural Information Processing Systems 37*. December 10-15, 2024. Vancouver, BC, Canada. *Neural Information Processing Systems Foundation, Inc. (NeurIPS)*, 2024.
- [23] WANG Z Y, LI C, XU H Y, *et al.* Mamba YOLO: a simple baseline for object detection with state space model [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025, 39(8): 8205-8213.
- [24] ZHU X, SU W, LU L, *et al.* Deformable detr: Deformable transformers for end-to-end object detection [J]. *arXiv preprint arXiv: 2010.04159*, 2020.
- [25] LIU S, LI F, ZHANG H, *et al.* Dab-detr: Dynamic anchor boxes are better queries for detr[J]. *arXiv preprint arXiv: 2201.12329*, 2022.
- [26] LIU Y C, LIU F, HUANG Y C, *et al.* The real-time pavement distress detection system based on edge-cloud collaborative computing [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2025, 26(7): 10512-10522.
- [27] KHAN Z Y, NIU Z D. CNN with depthwise separable convolutions and combined kernels for rating prediction[J]. *Expert Systems with Applications*, 2021, 170: 114528.
- [28] MA N C, FAN J H, WANG W S, *et al.* Comput-

- er vision for road imaging and pothole detection: a state-of-the-art review of systems and algorithms [J]. *Transportation Safety and Environment*, 2022, 4(4): tdac026.
- [29] HU K L, CHEN W P, SUN Y Z, *et al.* PpNet: Pyramid pooling based network for polyp segmentation [J]. *Computers in Biology and Medicine*, 2023, 160: 107028.
- [30] 李智杰, 惠爱婷, 李昌华, 等. 面向遥感图像道路提取的多尺度上下文感知网络[J]. *光学精密工程*, 2025, 33(4): 610-623.
- LI Z J, HUI A T, LI C H, *et al.* Multi-scale context-aware network for road extraction in remote sensing images [J]. *Opt. Precision Eng.*, 2025, 33(4): 610-623. (in Chinese)
- [31] 黄丹丹, 高晗, 刘智, 等. 面向无人机平台的轻量化目标检测网络[J]. *光学精密工程*, 2023, 31(20): 3021-3033.
- HUANG D D, GAO H, LIU Z, *et al.* Lightweight target detection network for UAV platforms [J]. *Opt. Precision Eng.*, 2023, 31(20): 3021-3033. (in Chinese)
- [32] GAO N, JIANG X Y, ZHANG X H, *et al.* *Efficient Frequency-Domain Image Deraining with Contrastive Regularization*[M]. *Computer Vision-ECCV 2024*. Cham: Springer Nature Switzerland, 2024: 240-257.
- [33] HE K M, ZHANG X Y, REN S Q, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition [C]. *Computer Vision-ECCV 2014*. Cham: Springer International Publishing, 2014: 346-361.
- [34] DU L X, ZHANG R Y, WANG X T. Overview of two-stage object detection algorithms[J]. *Journal of Physics: Conference Series*, 2020, 1544(1): 012033.
- [35] DHILLON A, VERMA G K. Convolutional neural network: a review of models, methodologies and applications to object detection[J]. *Progress in Artificial Intelligence*, 2020, 9(2): 85-112.
- [36] MISRA D, NALAMADA T, ARASANIPALAI A U, *et al.* Rotate to attend: convolutional triplet attention module[C]. *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. January 3-8, 2021, Waikoloa, HI, USA. IEEE, 2021: 3138-3147.
- [37] AHMADI A, KHALESIS S, GOLROO A. An integrated machine learning model for automatic road crack detection and classification in urban areas[J]. *International Journal of Pavement Engineering*, 2022, 23(10): 3536-3552.
- [38] FENG C J, ZHONG Y J, GAO Y, *et al.* TOOD: task-aligned one-stage object detection [C]. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. October 10-17, 2021, Montreal, QC, Canada. IEEE, 2021: 3490-3499.
- [39] ABTAHI T, SHEA C, KULKARNI A, *et al.* Accelerating convolutional neural network with FFT on embedded hardware [J]. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2018, 26(9): 1737-1749.
- [40] CHEN J, HE Y. A novel U-shaped encoder - decoder network with attention mechanism for detection and evaluation of road cracks at pixel level[J]. *Computer-Aided Civil and Infrastructure Engineering*, 2022, 37(13): 1721-1736.
- [41] YAO S L, GUAN R W, HUANG X Y, *et al.* Radar-camera fusion for object detection and semantic segmentation in autonomous driving: a comprehensive review[J]. *IEEE Transactions on Intelligent Vehicles*, 2024, 9(1): 2094-2128.

#### 作者简介:



周冬梅(1975—),女,河南平顶山人,博士,副教授,1999年、2010年于兰州交通大学分别获得学士和硕士学位,主要从事无线电工程、人工智能方面的研究。E-mail: zhoudongmeit@mail.lzjtu.cn



仵兵兵(2001—),男,河南周口人,硕士研究生,2023年于河南牧业经济学院获得学士学位,主要从事目标检测算法方面的研究。E-mail: 1815987813@qq.com