

基于改进 RT-DETR 的道路缺陷检测算法^{*}

李 涛¹ 孙祥娥²

(1. 长江大学电子信息与电气工程学院 荆州 434023; 2. 长江大学地球物理与石油资源学院 武汉 430058)

摘要: 道路破损增加了交通事故的发生概率,严重威胁交通安全。因此,实时监测路面状况对于保障道路安全和有效管理基础设施至关重要。针对现有道路缺陷检测中精度不足和小目标检测困难的问题,本文提出了一种基于改进 RT-DETR 的道路缺陷检测算法。首先,通过引入部分卷积(PConv)对 RT-DETR 主干网络进行重构,从而有效降低计算开销;其次,在主干网络中融合三重注意力机制,提升模型对多维特征的感知能力,进而更精准地捕捉图像细节。接着,采用双向特征金字塔网络(BiFPN)优化 CCFM 特征融合模块,并引入 S2 特征,增强小目标的检测性能。最后,利用 DySample 上采样算子获取更多局部细节与语义信息,进一步提升模型对小目标的检测能力。实验结果表明,改进后的算法在 RDD2022 数据集上的 mAP@50 较原始 RT-DETR 模型提升了 3.6%,且参数量减少了 12.5%,检测速度达到 66 fps。与其他目标检测算法相比,改进算法在检测精度和速度方面均表现出显著优势,具有更好的实际应用前景。

关键词: 道路缺陷检测;RT-DETR;BiFPN;深度学习;目标检测

中图分类号: TN911.73; TP183 文献标识码: A 国家标准学科分类代码: 510.6040

Road defect detection based on an optimized RT-DETR model

Li Tao¹ Sun Xiangge²

(1. College of Electronic Information and Electrical Engineering, Yangtze University, Jingzhou 434023, China;

2. College of Geophysics and Petroleum Resources, Yangtze University, Wuhan 430058, China)

Abstract: Road damage increases the likelihood of traffic accidents, posing a serious threat to traffic safety. Therefore, real-time monitoring of road conditions is crucial for ensuring road safety and effectively managing infrastructure. To address the issues of insufficient detection accuracy and small target detection challenges in existing road defect detection methods, this paper proposes an improved RT-DETR-based road defect detection algorithm. First, partial convolution (PConv) is introduced to reconstruct the RT-DETR backbone network, effectively reducing computational overhead. Second, a triplet attention mechanism is integrated into the backbone network to enhance the model's sensitivity to multi-dimensional features, enabling more precise capture of image details. Next, a BiFPN-based feature pyramid network is employed to optimize the CCFM feature fusion module, and S2 features are introduced to improve the detection performance of small targets. Finally, the DySample upsampling operator is utilized to capture more local details and semantic information, further enhancing the model's ability to detect small targets. Experimental results show that the improved algorithm achieves a 3.6% increase in mAP@50 on the RDD2022 dataset compared to the original RT-DETR model, with a 12.5% reduction in the number of parameters and a detection speed of 66 fps. Compared with other object detection algorithms, the improved algorithm demonstrates significant advantages in both detection accuracy and speed, making it more suitable for practical applications in road defect detection.

Keywords: road defect detection;RT-DETR;BiFPN;deep learning;object detection

0 引言

在保障交通安全和提升出行体验方面发挥着至关重要的作用。道路缺陷,如裂缝、坑洞和龟裂,不仅影响驾驶的舒适性,还可能增加交通事故的风险,并加重车辆维修成本。因

收稿日期:2025-03-11

* 基金项目:国家自然科学基金(62273060)项目资助

此,及时、准确地检测和修复这些缺陷,对于确保道路安全具有重要的意义^[1]。

传统的检测方法多依赖人工巡检和基于规则的图像处理技术,但这些方法通常效率较低、准确度有限,并且容易受环境因素如光照和天气的影响。近年来,深度学习技术的兴起为路面缺陷检测带来了显著的进展。通过自动化特征提取和模式识别,深度学习能够高效地从海量图像中提取缺陷特征,大幅提高了检测的准确性和自动化水平。

目标检测技术在深度学习中的应用可分为两大类:基于区域的两阶段检测模型和基于回归的单阶段检测方法。两阶段模型通常包括候选区域生成与特征提取和分类两大过程,典型算法如 Fast R-CNN^[2]、R-FCN^[3] 和 Mask R-CNN^[4]。这类方法在复杂背景下特别适合小目标检测,并且对遮挡和复杂环境具有较好的适应性,但其缺点是计算量大、速度较慢,尤其在实时检测任务中可能无法满足高效性要求。相比之下,基于回归的一阶段检测方法通过一次前向传播直接输出目标的类别与位置,具有较高的处理速度和简化的架构,适用于实时应用。然而,这类方法通常依赖非极大值抑制(non-maximum suppression, NMS)技术来消除冗余边界框,这会导致处理效率低,且难以应对小目标的漏检问题^[5-6]。

Transformer 模型最初在自然语言处理领域表现优异,近年来被逐渐应用于计算机视觉领域。DETR (detection transformer, DETR) 首次将 Transformer 应用于计算机视觉领域,直接对图像进行集合预测,取消了 anchor 机制和 NMS 处理,简化了目标检测流程,在目标检测领域表现出了卓越的学习能力^[7]。然而,DETR 在训练时需要大量时间,限制了其在实时应用中的使用。许多研究人员对 DETR 进行了改进,但注意力机制本身引入了更多的参数和计算复杂性,阻碍了 DETR 模型的实时应用。Zhao 等^[6] 提出了实时检测模型 RT-DETR,这是一款高实用性的端到端实时目标检测模型,保持高精度的同时实现了实时性能,在速度和精度方面优于同等规模的 YOLO (you only look once, YOLO) 系列^[8-10] 检测模型。

尽管深度卷积神经网络(CNN)在道路缺陷检测领域中取得了显著进展,但仍然面临许多挑战。李松等^[11] 提出了基于 YOLOv8 的 BOT 模块,能够同时提取道路损伤图像的全局与局部特征,并结合 CA 注意力机制与 C2fGhost 模块显著提升了检测精度。然而,传统的 YOLO 模型在小目标检测上表现较差,特别是在低分辨率图像上,容易产生误检和漏检。Yang 等^[12] 使用全卷积网络 (fully convolutional networks, FCN) 进行裂缝检测,能够识别多种裂缝类型,但 FCN 对不同尺度目标的适应性差,且在处理复杂背景时表现欠佳。Kang 等^[13] 采用 Faster R-CNN^[14] 算法进行道路裂缝检测,但计算量大,速度较慢,难以满足实时检测需求。与此同时,Du 等^[15] 提出的加权双

向特征金字塔网络(bidirectional feature pyramid network, BiFPN)在多尺度特征融合上取得了较好的进展,但其仍然面临着小目标检测精度不足的问题。牛慧余等^[16] 通过改进 Mask R-CNN 主干网络提高了裂缝特征提取能力,然而,该模型对计算资源的需求较高,不适用于实时应用。在道路缺陷检测领域,实时性尤为关键,特别是在车载设备与无人机巡检中。车载设备如自动驾驶汽车需快速响应路面变化,处理延迟需低于 50 ms,但算力受限(几 TOPS),且需低功耗(<10 W)。无人机巡检则需实时处理高清图像以识别微小缺陷,要求单帧处理不超过 100 ms(30 fps),算力同样有限(约 10 TOPS),且需控制功耗(<20 W)以维持长时间飞行。

针对当前道路缺陷检测领域面临的小目标检测效果差,对不同尺度目标适应性差,以及难以满足实时检测需求的问题,本文在借鉴上述研究成果的基础上对 RT-DETR 模型进行了改进,提出了以下几点创新:

1) 利用部分卷积(partial convolution, Pconv)^[17] 和三重注意力机制(Triplet Attention)^[18] 重构了骨干网络,有效地增强了模型的特征提取能力和小目标检测能力,减少了模型的参数量,提高了模型的检测精度。

2) 设计改进版的跨尺度特征融合模块(cross-scale feature fusion module, CCFM)增强了模型对不同尺度目标的,采用双向特征金字塔网络(BiFPN)^[19] 替代传统的特征金字塔 FPN(feature pyramid network, FPN),并引入额外的高分辨率特征图 S2,与 S3 融合,显著提高了小目标的检测精度。

3) 引入动态上采样算子 (dynamic upsampling, DySample)^[20],进一步提升特征金字塔网络的性能,增强了对小目标的检测能力。

1 RT-DETR 模型

RT-DETR^[6] (real-time detection transformer, RT-DETR) 是一种实时端到端目标检测框架,该框架基于 Transformer 架构,创新性地摒弃了传统目标检测方法中的 NMS,在保持高检测精度的同时显著提升推理速度,特别适用于对实时性要求较高的应用场景。

RT-DETR 的整体架构由骨干网络、高效混合编码器和解码器三部分组成。骨干网络采用残差网络(residual network, ResNet)^[21] 或 HGNet^[22],负责提取多尺度特征图。高效混合编码器通过基于注意力的同尺度特征交互模块(attention-based intra-scale feature interaction, AIFI)和基于 CNN 的跨尺度特征融合模块(cross-scale feature fusion module, CCFM)协同工作,实现高效语义信息捕获。具体而言,AIFI 在单尺度上捕获特征以降低计算复杂度,而 CCFM 则融合多尺度特征,生成高质量特征表达,为解码器提供精准输入。本文选择 RT-DETR-R18 作为基线模型,其结构设计如图 1 所示。

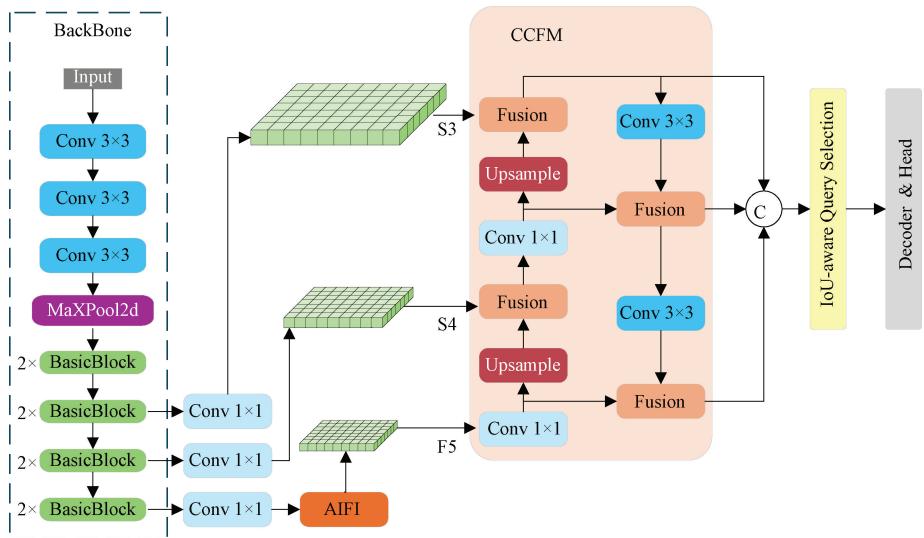


图 1 RT-DETR-R18 模型原理图

Fig. 1 Diagram of the RT-DETR-R18 model architecture

在解码阶段, RT-DETR 引入不确定性最小化查询选择机制, 通过联合优化分类置信度和定位置信度, 有效提升检测精度。同时, 解码器支持动态调整层数, 以灵活适应不同的实时检测需求, 无需重新训练模型, 从而进一步增强了模型的适配性和实用性^[6]。

2 改进的 RT-DETR 模型

本文对 RT-DETR-R18 网络进行了若干改进, 旨在提升算法对道路缺陷检测的精度和速度, 减少模型的参数量。在追求检测精度最大化提升的同时, 通过模型轻量化设计

与计算效率优化策略, 确保推理速度维持在原始基线水平, 本文在主干网络采用三重注意力快速特征提取模块(triple-attention faster feature extraction module, TAFaster)替换原本的 BasicBlock 结构。在特征融合阶段, 采用双向金字塔跨尺度融合模块(bidirectional feature pyramid network with cross-scale feature fusion module, BiFPN-CCFM)优化算法的多尺度特征融合能力, 并引入 S2 特征, 提高模型对小目标的敏感度。最后, 应用 DySample 上采样算子来进一步优化特征金字塔网络的性能, 增强了对小目标的检测效果。改进后的 RT-DETR 结构如图 2 所示。

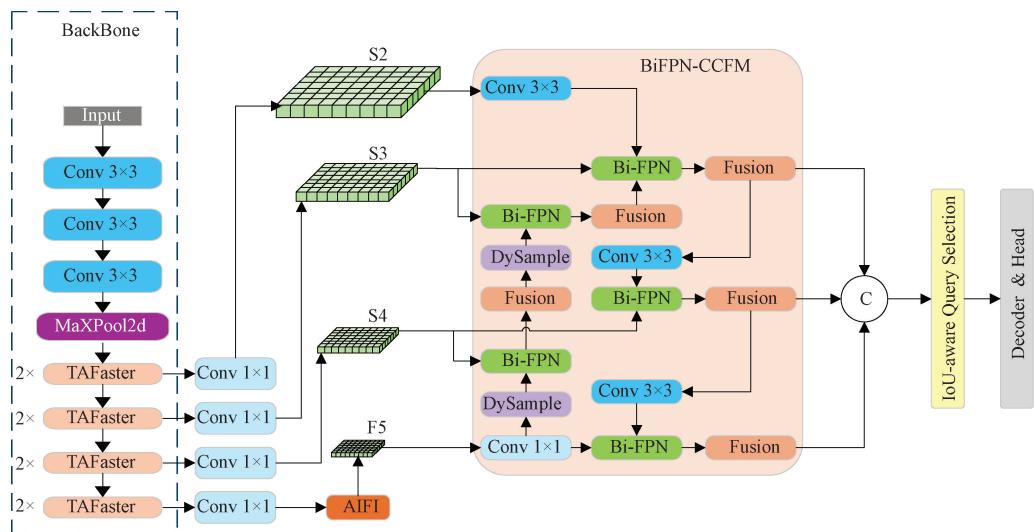


图 2 改进后的 RT-DETR 模型

Fig. 2 The improved RT-DETR model

2.1 三重注意力快速特征提取模块(TAFaster)

为了在减少 RT-DETR 原模型参数量的同时提升检测精度, 本文提出了一种三重注意力快速特征提取模块

TAFaster, 该模块结合了三重注意力机制和轻量级主干网络 FasterNet 中的 FasterBlock 模块。

FasterBlock 应用了一种创新的 PConv 技术, 通过减

少计算冗余和内存访问, 实现高效的特征提取。PConv 层在部分通道上应用传统的卷积进行空间特征的提取, 其他

通道则保持不变, 以此实现高效的利用计算资源。图 3 为 TAFaster 模块以及 PConv 模块与传统卷积的对比。

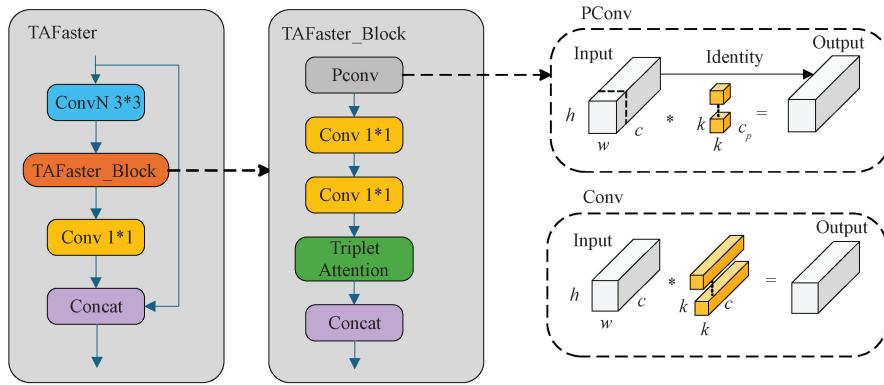


图 3 TAFaster 模块以及 PConv 与传统卷积的对比

Fig. 3 TAFaster module and comparison of PConv with traditional convolution

PConv 的 FLOPs 和内存访问量为:

$$h \times w \times k^2 \times c_p^2 \quad (1)$$

$$h \times w \times 2c_p + k^2 \times c_p^2 \approx h \times w \times 2c_p \quad (2)$$

普通卷积的浮点运算量和内存访问量为:

$$h \times w \times k^2 \times c^2 \quad (3)$$

$$h \times w \times 2c + k^2 \times c^2 \approx h \times w \times 2c \quad (4)$$

在深度卷积运算中, 输入特征图的维度由通道容量 c 、垂直高度 h 和水平宽度 w 共同表征。其中, 常规卷积操作采用 $k \times k$ 尺寸的卷积核, 其通道维度保持为 c 。而 PConv 通过通道压缩机制, 将通道数据缩减为 c_p , 且满足 c_p 与 c 的比例关系 $c_p/c = 1/4$ 。基于该结构设计, PConv 的计算效率显著提升: 其浮点运算量仅为传统卷积的 $1/16$, 同时内存访问开销也降低至常规卷积的 25% 。

为增强模型对多模态特征的表达能力和微小目标检测精度, 本研究设计了一种三维注意力模块 Triplet Attention。该模块突破传统注意力机制的设计框架, 采用三分支异构结构实现多粒度特征融合: 顶端分支通过空间旋转与卷积操作捕捉通道与高度($C-H$)之间的交互, 增强垂直特征融合能力; 中间分支专注于通道与宽度($C-W$)的关联分析, 提升水平感知精度; 底端分支则保留经典空间注意力机制以强化局部特征对比。这种三支路协同设计不仅实现了

跨维度特征交互的全面覆盖, 还通过异构结构避免了单一分支的过拟合风险。在注意力权重生成过程中, 该模块创新性地采用复合池化策略 Z-Pool, 该策略通过融合平均池化与最大池化的输出特征, 将四维特征张量压缩至二维表示空间。与传统的全局平均池化相比, Z-Pool 在有效保留关键信息的同时显著降低了计算复杂度, 且避免了降维操作可能导致的特征信息丢失。进一步地, 模块通过旋转操作建立高度(H)与宽度(W)维度间的隐式依赖关系, 这种跨维度交互机制使其能够捕捉到传统注意力机制难以建模的特征关联模式。最终, 3 个支路的输出特征经过均值融合与卷积变换, 生成具有统一维度的注意力特征表示。实验结果表明, 在 ImageNet-1k 图像分类、MSCOCO 和 PASCAL VOC 目标检测等任务中, Triplet Attention 模块相比于 SE Attention (squeeze-and-Excitation Attention) 模块和卷积块注意力模块 (convolutional block attention module, CBAM) 等主流方法, 有更优的性能。尤其在资源受限场景下, 其仅引入极少的额外参数和计算开销, 却能显著提升模型对微小目标的检测精度, 并通过梯度加权类激活映射 (gradient-weighted class activation mapping, Grad-CAM) 可视化验证了其更强的特征聚焦能力^[23]。Triplet Attention 模块的原理如图 4 所示。

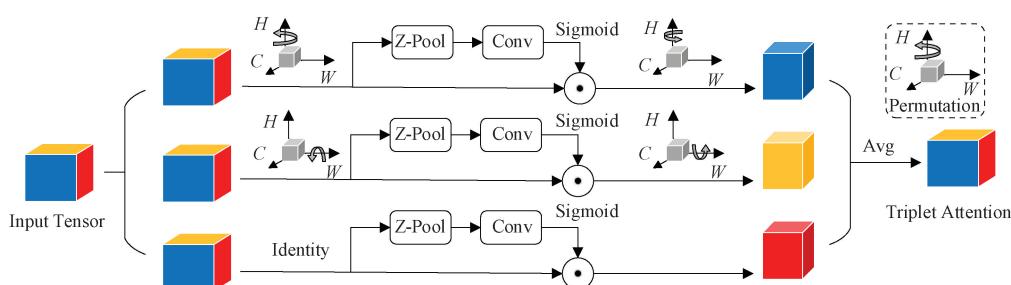


图 4 三重注意力模块

Fig. 4 Triplet attention module

2.2 双向金字塔跨尺度融合模块 BiFPN-CCFM

随着神经网络层次的加深,简单地增加网络层数未必能提升精度,反而可能导致梯度消失或梯度爆炸等问题。谷歌 2020 年发布的加权双向特征金字塔网络 BiFPN 相比于 PAN^[24] 网络,其特点在于引入了双向信息流,信息不仅可以从顶部向底部传递,还能从底部向顶部回流。这种双向流动的特性使得不同层级之间的特征可以相互调整和交换,进而优化特征的融合效果,提升模型的性能。

如图 5(a)所示,PAN 结构通过横向连接改进了信息传递,但其单向特性可能导致浅层特征的丢失。为解决这一问题,BiFPN 采用双向信息流(图 5(b))使得特征在双向流动中得以更全面地整合,保留更多细节。

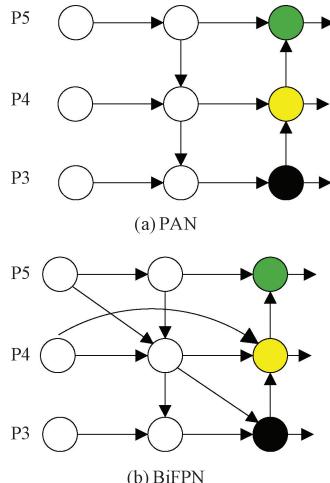


图 5 PAN 网络与 BiFPN 网络结构图

Fig. 5 PAN and BiFPN architecture diagram

RT-DETR 检测框架通过融合主干网络深层 3 个阶段的特征表示,建立了一种通用的目标检测架构。在道路缺陷检测领域,由于远处细微裂缝目标仅占图像极小比例的像素区域,浅层特征信息在提取的过程中很容易丢失,为解决这一问题,本研究改进并引入了双向特征金字塔网络 BiFPN,改进后的网络架构如图 6 所示。

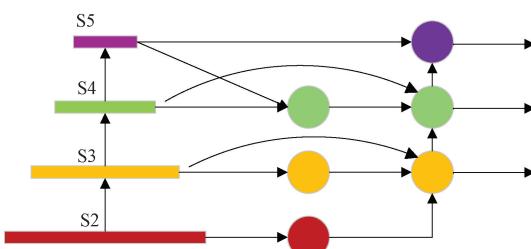


图 6 改进的 BiFPN 结构

Fig. 6 Improved BiFPN architecture

在改进的 BiFPN 架构中,首先从骨干网络中引入能捕获更多局部细节的 160 pixel×160 pixel 高分辨率特征图,其次,通过将 S2 层提取的高频细节特征与 S3 层进行跨尺

度融合。这种多层级特征整合机制增强了模型对细微目标的识别能力。

2.3 基于点采样的上采样 Dysample 模块

在路面缺陷检测场景中,缺陷区域往往像素占比小,且常常存在污渍遮挡或者图像光照不均等情况,导致难以提取目标缺陷特征,从而影响模型的检测精度和鲁棒性。为了应对这一挑战,本文引入基于点采样的动态上采样器 Dysample,旨在增强模型对小缺陷的检测能力以及抗干扰能力,Dysample 的设计原理如图 7 所示,公式如式(5)~(8)所示。

$$\chi' = \text{grid_sample}(\chi, S) \quad (5)$$

$$S = G + O \quad (6)$$

$$O = 0.25(\text{linear}(\chi)) \quad (7)$$

$$O = 0.25\sigma(\text{linear}_1(\chi)) \times \text{linear}_2(\chi) \quad (8)$$

式中: χ 表示输入特征, χ' 表示特征, O 为生成偏移量, σ 表示 Sigmoid 函数, G 为原始采样, S 为上采样特征。

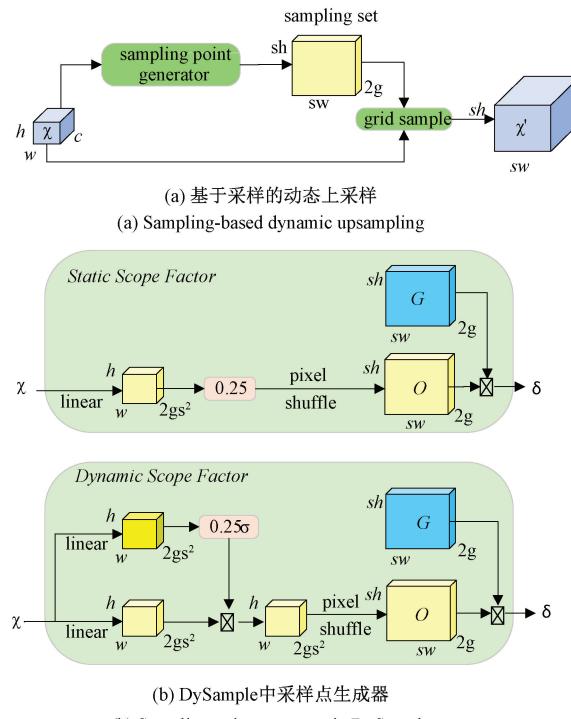


图 7 Dysample 网络结构图

Fig. 7 Dysample network architecture

如图 7(a)所示,输入特征 χ 通过采样点发生器生成的采样集 S 进行重采样。图 7(b)是采样点发生器结构图,其采集器由两部分构成,分别是生成的偏移量 G 和原始采样网格 G ,偏移量 O 的生成机制结合了线性变换与像素随机播放。以图 7(b)上半部分所示静态因子采样方法为例,给定一个上采样因子 S 和一个大小为 $c \times h \times w$ 的特征映射,特征映射通过一个输入通道和输出通道分别为 c 和 $2s^2$ 的线性层,生成一个大小为 $2s^2 \times h \times w$ 的偏移量 O 。

然后, 使用像素级随机方法将其转换为 $2 \times sh \times sw$, 其中 2 表示 x 和 y 坐标, 最后生成一个大小为 $c \times sh \times sw$ 的上采样特征图^[25]。Dysample 方法通过引入邻近像素采样机制, 有效增加了样本多样性, 从而缓解类别不平衡问题, 抑制虚假预测。与传统上采样方法相比, Dysample 能够更好地保持图像高频细节, 避免简单像素复制导致的模糊和锯齿伪影, 显著提升重建质量。

3 实验与结果

3.1 数据集与实验环境

本文所用的 RDD2022^[26] 为全球道路损伤检测挑战赛使用的数据集, 由中国、印度、日本、美国和捷克等多个国家的道路图像组成。本文选取 4 398 张分辨率为 512 pixel \times 512 pixel 的中国道路图像作为实验对象。其中包含 2 401 张无人机拍摄的道路图片和 1 977 张车载相机拍摄的道路图像。本文研究的道路缺陷类型有 5 种, 包括纵向裂缝 D00、横向裂缝 D10、网状裂缝 D20、坑洼 D40 和道路修补 Repair, 不同道路缺陷类型在数据集中分布如表 1 所示。

表 1 数据集中缺陷类型及数量

Table 1 Defect types and quantities in the dataset

缺陷类型	标签	数量
网状裂缝	D20	756
纵向裂缝	D00	3 270
横向裂缝	D10	1 895
坑洼	D40	255
路面修补	Repair	821

不同类型的道路缺陷实例如图 8 所示。数据集按照 8:1:1 的比例划分为训练集、验证集和测试集。



图 8 数据集中不同道路缺陷示例图

Fig. 8 Examples of different road defects in the dataset

本文实验使用 CentOS 7.9 操作系统, 处理器型号为 Intel Xeon Silver 4210R, 采用 PyTorch 2.4.0 + cu118 深度学习框架, 搭载 5 张 NVIDIA GeForce RTX 3080 GPU, 每张显卡的显存为 10 GB, CUDA 版本为 11.8。实验超参数设置如表 2 所示。

表 2 实验超参数

Table 2 Experimental hyperparameters

参数	参数值
Learning rate	0.000 1
Image size	640 \times 640
Optimizer	AdamW
Batch size	8
Epoch	200

3.2 评价指标

本文采用了 COCO 数据集类型的评价方法, 构建了轻量化-精度双轨评估体系, 采用参数量(Params)、每秒 10 亿的浮点运算次数(giga floating-point operations per second, GFLOPs) 和每秒检测帧数(frames per second, FPS) 三维指标衡量模型效率, 其中 Params 反映空间复杂度(值越小内存占用越低), GFLOPs 表征计算复杂度, FPS 体现实时性(值越高推理速度越快); 平均精度均值(mean average precision, mAP) 来检验模型的分类的准确性和物体的定位性能, 其计算如式(9)所示。

$$mAP = \frac{1}{classes} \sum_{i=1}^{classes} \int_0^1 P(R)d(R) \quad (9)$$

其中, P 表示精确度, R 表示召回率。

另外在 COCO 数据集中, 针对 small、medium、large 3 种不同尺寸目标制定了测量标准: 像素小于 32 pixel \times 32 pixel 的为“小目标”, 像素在 32 pixel \times 32 pixel \sim 96 pixel \times 96 pixel 之间的为“中等目标”, 像素大于 96 pixel \times 96 pixel 的为“大目标”。

3.3 消融实验

为评估改进模型在道路缺陷检测任务中的实际效果, 本研究在同一的实验条件下, 以 RT-DETR 原模型作为基准模型, 进行了一些列消融实验, 分别验证了 Pconv 模块、Triplet Attention 模块、BiFPN 模块和 Dysample 模块这 4 种改进方法带来的性能增益。实验结果如表 3 所示。

分析表中消融实验结果可知, 在引入 PConv 后, 模型的 mAP@0.5 从 86.1% 提升至 86.3%, 在保证检测精度的前提下, 参数量减少了 14.67%, 浮点运算量减少了 15%, 检测速度提升至 70.4 fps。这一结果的原因是 PConv 通过在保持特征图尺寸不变的情况下减少了通道数量, 从而提高了模型的计算效率, 减少了计算开销并提升了检测速度。

在 PConv 的基础上, 进一步引入了 Triplet Attention

表 3 消融实验结果

Table 3 Ablation experiment results

Pconv	Triplet Attention	BiFPN	Dysample	Map@0.5/%	Params/M	GFLOPs	FPSbs=1
×	×	×	×	86.1	19.92	60	56.7
√	×	×	×	86.3	17.00	50.9	70.4
√	√	×	×	88.9	17.00	51.0	68.1
√	√	√	×	89.5	17.42	57.8	66.0
√	√	√	√	89.7	17.43	57.8	66.0

模块,使得 mAP@0.5 进一步提升至 88.9%。这一提升主要得益于该模块增强了对小目标和细粒度特征的关注,尤其是在车载相机拍摄的道路图像中,裂纹往往占据较少像素且背景复杂。Triplet Attention 帮助模型更好地识别这些难以察觉的道路裂纹,提升了检测精度。

随后,在引入 Triplet Attention 的基础上,在模型的颈部网络中加入了 S2 特征参与特征融合,并且在跨尺度特征融合层中添加了 BiFPN。尽管这一改动导致模型参数略有增加,但 mAP@0.5 从 88.9% 提升至 89.5%,表明 BiFPN 优化了跨尺度特征融合,提升了多尺度目标的检测能力。

为了进一步增强模型对小目标的检测能力,本文使用 Dysample 替换掉了原有跨尺度融合层中的上采样模块,进而使 mAP@0.5 再度提升了 0.2%。实验结果表明随着 PConv、Triplet Attention、BiFPN 和 DySample 上采样算子的逐步引入,最终模型的 mAP@0.5 提升至 89.7%,较基础模型提高了 3.6%,并且检测速度提升至 66.0 fps,较基础模型提高了 16.4%。与此同时,参数量为 17.43 M,浮点运算量为 57.9 GFLOPs。这一结果证明,逐步引入这些模块能够在保证较低计算开销的同时,显著提升模型的检测精度和速度,平衡了检测精度、计算开销和检测速度之间的矛盾。

3.4 检测效果可视化对比

为全面评估改进模型的检测性能,本文从 RDD2022 数据集中随机选取一张车载相机拍摄的路况图像和一张无人机拍摄的路况图像进行检测与识别实验。图 9 为车载相机拍摄的路面场景。在该场景中,RT-DETR 原始算法对细小裂纹的检测置信度较低,尤其是在图像上半部分中间区域出现漏检现象,暴露出其在细微结构检测方面的局限性。相较而言,改进后的 RT-DETR 算法在该场景中表现出色,几乎未出现漏检或误检。图 10 展示了无人机拍摄的路况图像。在这一场景下,尽管原始模型与改进模型均实现了较高的检测水平,但改进模型凭借更高的置信度和更稳定的检测性能。

改进算法通过引入 Pconv 和 Triplet Attention 优化骨干网络,显著增强了特征提取能力,并有效改善了小目标检测性能。同时,结合重新设计的双向金字塔跨尺度融合模块 BiFPN-CCFM,进一步提升了模型在细小裂纹及复杂

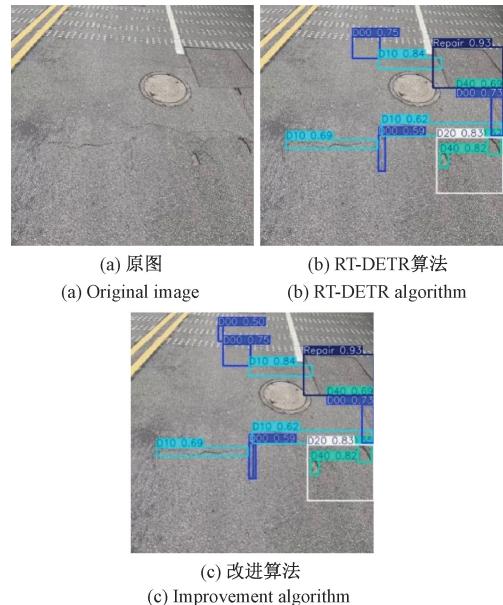


图 9 车载相机拍摄的路面场景

Fig. 9 Road surface scene captured by in-vehicle camera

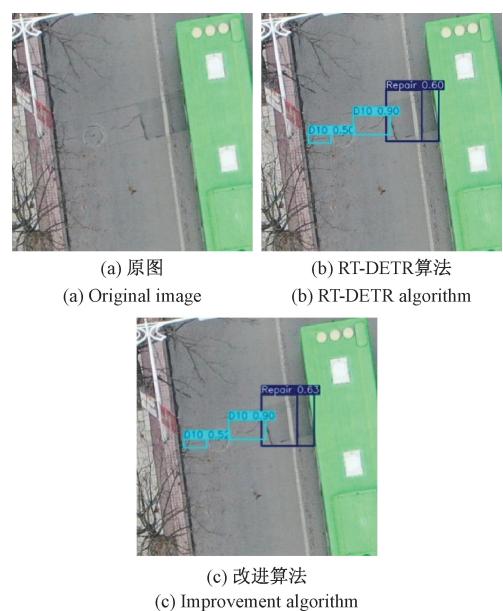


图 10 无人机拍摄的路面场景

Fig. 10 Road surface scene captured by drone

目标检测任务中的表现。可视化实验结果表明,改进模型在多场景下展现出更高的稳定性和优越性,验证了其在检测精度与鲁棒性上的显著提升。

3.5 主流算法对比实验

为评估改进模型的性能,本文将提出的算法与主流目标检测模型进行了对比实验,包括二阶段有锚框模型 Faster R-CNN,一阶段有锚框模型 SSD,以及一阶段无锚框模型 YOLOv8m 和 YOLOv11m。实验在相同数据集和实验条件下进行,所有模型均已训练至收敛状态,其对比实验结果如表 4 所示。

表 4 主流算法对比实验

Table 4 Comparison of mainstream algorithms

模型	mAP@0.5/%	Para/ 10^6	FPS	GFLOPs
Faster R-CNN	72.2	41.14	12.9	78.13
SSD	73.7	24.83	41.2	88.14
YOLOv8m	86.0	25.84	55.6	78.7
YOLOv11m	86.4	20.06	60.0	67.7
RE-DETR	86.1	19.92	56.7	60.0
本文	89.7	17.43	66.0	57.8

实验结果表明,改进算法(OURS)在所有模型显著优于 YOLOv11m(86.4%)和其他模型。尽管精度提高,改进模型的每秒帧数(FPS)仍达到了 66.0 fps,优于 YOLOv8m(55.6 fps)和 YOLOv11m(60.0 fps),显示出较好的实时性。此外,本文模型的参数量为 17.43 M,相较于其他模型具有较小的规模,且 GFLOPs 为 57.8,计算复杂度较低,体现了其在计算效率上的优势。这表明,改进算法在保证较高精度的同时,仍能兼顾实时性和资源消耗,具有较好的应用前景。

3.6 泛化实验

VOC 2007 数据集包含 9 963 张图像,涵盖 dog、person、sofa、tvmonitor 等 20 类常见物体,图像背景丰富多样,目标形状、纹理、颜色差异大。而道路缺陷检测数据集聚焦道路表面裂缝、坑洼等,背景相对单一,但道路材质、天气等因素致图像特征存在差异。为验证模型在通用目标检测上的能力并考察其向道路缺陷检测任务的迁移潜力,在该数据集上开展实验,其中训练集和验证集分别包含 5 011 张和 4 952 张图像。PASCAL VOC 2007 数据集常用于目标检测任务的评估,并且其广泛的类别和复杂的图像内容使其成为评估模型泛化能力的理想选择。

实验结果如表 5 所示,改进后的算法在大多数类别上均实现了轻微的性能提升。例如,在 aeroplane 类别中,mAP@50 从 84.9% 提升至 85.1%;在 car 类别中,从 83.5% 增加到 84.0%。在部分类别(如 cat 和 horse)中,改进后的算法则取得了更为显著的提升,分别提高了 0.6% 和 1.3%。这些结果表明,改进后的 RT-DETR 算法在不

同类别的检测上更加精准,且具备较强的适应能力。

表 5 VOC 2007 数据集不同算法 mAP@50 对比

Table 5 mAP@50 comparison of algorithms on VOC 2007

类别	RT-DETR/%	本文/%
aeroplane	84.9	85.1
bicycle	80.1	80.5
bird	75.8	76.0
boat	64.3	64.6
bottle	58.8	59.5
bus	80.5	80.5
car	83.5	84.0
cat	85.5	86.1
chair	53.5	54.7
cow	82.3	83.2
diningtable	59.2	61.5
dog	81.4	81.9
horse	82.2	83.5
motorbike	80.3	80.9
person	81.5	81.9
pottedplant	48.9	49.5
sheep	71.4	72.4
sofa	65.7	66.8
train	85.9	86.6
tvmonitor	74.1	74.6
Average	74.0	74.7

此外,改进后的算法在 person、dog 和 train 等常见类别上也展现了稳定的提升,进一步验证了该算法在数据集中的广泛泛化能力。尽管在某些类别(如 bottle 和 boat)中,改进幅度较小,但总体来看,改进后的算法在多类别目标检测中仍展现出了良好的稳健性。

4 结 论

本文提出了一种基于改进 RT-DETR 的道路缺陷检测算法,针对现有方法在检测精度和小目标检测方面的不足,提出了创新性改进。通过引入 PConv 重构 RT-DETR 主干网络,提升了特征提取能力,同时通过 Triplet Attention 增强了模型对多尺度、多维度特征的感知能力,从而进一步提高了检测精度。采用 BiFPN 改进的 BiFPN-CCFM,显著提升了对小目标的检测性能,并通过 DySample 上采样算子的引入,增强了对图像局部细节和语义信息的捕捉能力。

实验结果表明,改进算法在 RDD2022 数据集上的 mAP@50 较原始模型提升了 3.6%,同时在检测速度上达到了 66 fps,参数量减少了 12.5%。这些改进表明,提出的算法在保持高检测精度的同时,具有更低的计算开销和更

优的实时性,具备较强的实际应用潜力。

虽然现有道路缺陷检测模型在某些领域已取得良好效果,但是依然面临诸如缺陷形态和尺寸差异大、复杂环境干扰等挑战。未来的研究应着重于进一步优化算法的实时性能,同时通过引入更多的路面数据集,提升算法的泛化能力。此外,随着硬件加速技术的发展,模型的推理速度和检测精度有望得到更大幅度的提升,为道路缺陷检测的实际应用提供更强有力的技术支持。

参考文献

- [1] FAN L, WANG D, WANG J, et al. Pavement defect detection with deep learning: A comprehensive survey[J]. IEEE Transactions on Intelligent Vehicles, 2024, 9(3): 4292-4311.
- [2] GIRSHICK R. Fast R-CNN [C]. 2015 IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [3] DAI J, LI Y, HE K, et al. R-FCN: Object detection via region-based fully convolutional networks [J]. Advances in Neural Information Processing Systems, 2016, 29: 379-387.
- [4] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]. IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [5] ZOU ZH X, CHEN K Y, SHI ZH W, et al. Object detection in 20 years: A survey[J]. Proceedings of the IEEE, 2023, 111(3): 257-276.
- [6] ZHAO Y, LYU W Y, XU SH, et al. DETRs beat YOLOs on real-time object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.
- [7] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 213-229.
- [8] 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.
- [9] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [10] 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.
- [11] 李松,史涛,井方科.改进YOLOv8的道路损伤检测算法[J].计算机工程与应用, 2023, 59(23): 165-174.
- [12] LI S, SHI T, JING F K. Improved road damage detection algorithm of YOLOv8 [J]. Computer Engineering and Applications, 2023, 59(23): 165-174.
- [13] YANG X C, LI H, YU Y T, et al. Automatic pixel-level crack detection and measurement using fully convolutional network[J]. Computer-Aided Civil and Infrastructure Engineering, 2018, 33(12): 1090-1109.
- [14] KANG D, BENIPAL S S, GOPAL D L, et al. Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning [J]. Automation in Construction, 2020, 118: 103291.
- [15] 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, 2017, 39 (6): 1137-1149.
- [16] DU F J, JIAO SH J. Improvement of lightweight convolutional neural network model based on YOLO algorithm and its research in pavement defect detection[J]. Sensors, 2022, 22(9): 3537.
- [17] 牛慧余,包腾飞,李扬涛,等.基于改进Mask R-CNN的混凝土坝裂缝像素级检测方法[J].水利水电科技进展,2023,43(1):87-92,98.
- [18] NIU H Y, BAO T F, LI Y T, et al. Pixel-level detection method for cracks in concrete dam based on improved mask R-CNN [J]. Progress in Water Resources and Hydropower Science and Technology, 2023, 43(1): 87-92, 98.
- [19] CHEN J R, KAO SH H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021-12031.
- [20] 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, 2021: 3138-3147.
- [21] TAN M X, PANG R M, LE Q V. EfficientDet: Scalable and efficient object detection[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10778-10787.
- [22] LIU W Z, LU H, FU H T, et al. Learning to upsample by learning to sample [C]. IEEE/CVF International Conference on Computer Vision, 2023: 6027-6037.
- [23] HE K M, ZHANG X, REN SH Q, et al. Deep residual learning for image recognition [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition,

- 2016: 770-778.
- [22] CHEN J T, LEI B W, SONG Q Y, et al. A hierarchical graph network for 3D object detection on point clouds[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 392-401.
- [23] 于平平, 闫玉婷, 唐心亮, 等. 输电线路场景下的施工机械多目标跟踪算法[J]. 计算机应用, 2025, 45(7): 2351-2360.
- YU P P, YAN Y T, TANG X L, et al. Multi object tracking algorithm for construction machinery in transmission line scenarios [J]. Computer Applications, 2025, 45(7): 2351-2360.
- [24] LIU SH, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [25] 王慧云, 赵俊生, 王禹, 等. 面向无人边防的复杂环境遮挡小目标检测算法[J]. 电子测量技术, 2024, 47(21): 168-177.
- WANG H Y, ZHAO J SH, WANG Y, et al. Complex environment occlusion small target detection algorithm for unmanned border defense[J]. Electronic Measurement Technology, 2024, 47(21): 168-177.
- [26] 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.

作者简介

李涛, 硕士研究生, 主要研究方向为计算机视觉、目标检测。

E-mail: 207217113@qq.com

孙祥娥(通信作者), 博士, 教授, 主要研究方向为计算机视觉、自然语言处理、语音信号处理方法及应用研究、现代信号处理方法及应用研究、地震波速度求取方法研究、VSP 数据处理方法研究及应用等。

E-mail: xinges2000@yangtzeu.edu.cn