

DOI:10.19651/j.cnki.emt.2519303

基于MDPR-DeepLabV3+的无人机巡检航拍 路面破损分割方法*

韩建峰^{1,2} 南汝君^{1,2} 宋丽丽^{1,2} 房建东^{1,2} 徐子涵^{1,2}

(1.内蒙古工业大学信息工程学院 呼和浩特 010080; 2.内蒙古自治区智能感知与系统工程重点实验室 呼和浩特 010080)

摘要: 针对无人机巡检中路面破损形态差异显著、在复杂环境干扰下分割效果不佳的问题,本文提出一种改进模型MDPR-DeepLabV3+,用于无人机路面破损分割。首先,采用MobileNetV2网络替换原主干网络,提升运行效率;其次,在编码器构建DFSP模块,通过渐进式特征累积和跨尺度注意力交互,实现局部细节和全局上下文的高效整合;此外,添加PSA_M注意力模块强化破损边缘信息;最后,在解码器提出双路径特征解耦耦合模块RCD,促进深层次信息互补,增强特征多样性。实验结果表明,所提模型在自制无人机巡检航拍路面破损数据集上mIoU、mPA和mPrecision分别达到78.47%、92.03%和83.93%。在公开数据集Crack500上进一步验证了本文模型的有效性,表明其在复杂环境中的路面破损识别精度和鲁棒性方面有较好的表现。

关键词: MDPR-DeeplabV3+;无人机;破损检测;语义分割;特征融合

中图分类号: TN911.73; TN919.8 **文献标识码:** A **国家标准学科分类代码:** 520.20

Road surface damage segmentation method for unmanned aerial vehicle inspection aerial photography based on MDPR-DeepLabV3+

Han Jianfeng^{1,2} Nan Rujun^{1,2} Song Lili^{1,2} Fang Jiandong^{1,2} Xu Zihan^{1,2}

(1. School of Information Engineering, Inner Mongolia University of Technology, Hohhot 010080, China;

2. Inner Mongolia Autonomous Region Key Laboratory of Intelligent Perception and System Engineering, Hohhot 010080, China)

Abstract: To address the significant morphological differences in road surface damage and the unsatisfactory segmentation effects under complex environmental interference in UAV inspections, we propose an improved model, MDPR-DeepLabV3+, for road surface damage segmentation. Firstly, the original backbone network is replaced with the MobileNetV2 network to enhance operational efficiency. Secondly, a DFSP module is constructed in the encoder to efficiently integrate local details and global context through progressive feature accumulation and cross-scale attention interaction. In addition, a PSA_M attention module is added to strengthen the information of damaged edges. Finally, a residual channel decoupling (RCD) module is proposed in the decoder to promote the complementarity of deep and shallow layer information and enhance feature diversity. The experimental results demonstrate that the proposed model achieves mIoU, mPA, and mPrecision of 78.47%, 92.03% and 83.93%, respectively, on a self-made UAV inspection dataset for road damage. The effectiveness of this model is further validated on the publicly available dataset Crack500, indicating strong performance in terms of pavement damage recognition accuracy and robustness in complex environments.

Keywords: MDPR-DeeplabV3+; UAV; damage detection; semantic segmentation; feature fusion

0 引言

公路建设是我国交通强国战略和基础设施建设的重要组成部分^[1]。路面破损问题随环境、交通负荷及自然条件

变化日益突出,定期、准确地检测、定位和测量破损至关重要^[2]。传统人工巡检方式存在人力成本高、效率低、安全隐患大以及易受天气条件限制等问题,而无人机(unmanned aerial vehicle, UAV)在大范围区域覆盖和复杂环境的任务

收稿日期:2025-07-07

* 基金项目:内蒙古自治区自然科学基金项目(2025LHMS06009)、内蒙古自治区科技计划项目(2025YFHH0156, 2023YFJM0002)资助

中展现出较高的机动性和灵活性,能够有效弥补传统巡检方式的不足。配合高分辨率图像采集设备,无人机可高效获取道路表面长距离的高精度图像,便于实现路面破损的像素级分割,为后续破损参数计算提供可靠数据保障,显著提升巡检效率。

传统路面破损检测主要依赖颜色分布、梯度信息和纹理等底层特征,检测精度和效率不足,难以适应复杂多变的道路场景。随着深度学习技术的快速发展,卷积神经网络(convolutional neural networks, CNN)在路面破损分割领域取得显著进展。例如,HRNet^[3]通过持续保持高分辨率特征并融合多尺度信息,显著提升了细节捕捉能力,而PSPNet^[4]则利用金字塔池化模块获取丰富的上下文信息,增强了对复杂场景的理解;胡波等^[5]提出一种基于DeepLabV3+的后处理算法,通过连通域判别细化和端点聚类算法,提高隧道裂缝的准确识别,但该方法在处理复杂环境下仍存在漏检风险;杨萍等^[6]通过用优化后的MobileNetV2降低了参数量并提高了运行速度,同时引入条形池化增强模型对上下文信息的捕获能力,但由于公开数据集的稀缺,其对不同裂缝类型的泛化能力不足;于天河等^[7]提出SW-Net,一个低成本高精度的道路裂缝自动分割与量化系统,解决道路裂缝检测中的高成本与高精度矛盾,降低裂缝的量化误差;Nguyen等^[8]将检测和分割集成到一个框架,在提升性能的同时降低了计算需求;赵志宏等^[9]提出融合注意力与GhostUNet的裂缝检测方法,可精确识别裂缝并兼顾实时性,但因未充分考虑损失函数影响,存在漏检问题;尹学辉等^[10]引入渐进式上下文交互,改善了细节还原能力,但Unet++的全局特征表征仍有局限;段中兴等^[11]提出的EMFCHBNet通过多尺度特征融合提升了特征提取能力,但极细裂缝检测仍有改进空间。尽管CNN具有强大的特征提取能力,但其感受野有限、全局关联建模不足等局限,仍制约着其在路面破损精确分割上的进一步提升。

近年来,以自注意力机制为核心的视觉Transformer被引入计算机视觉领域,在全局建模上展现出独特优势。Li等^[12]通过分层变换器和联合损失设计,提升了裂缝检测性能,但复杂场景中的裂缝检测效果不佳;Yu等^[13]提出结合Swin Transformer与双分支解码器的CSTF网络,改善了细裂缝的分割能力,但对其他破损类型的检测能力尚未验证;周云等^[14]提出基于改进的Segformer模型的无人机单目视频裂缝定位方法,显著提高了道路裂缝的检测效率;Shan等^[15]提出高效裂缝分割系统Mix Transformer来处理无人机捕获的路面图像,在边界恢复方面取得良好效果。虽然上述方法能够有效利用全局依赖关系,但如何平衡全局与局部的关系仍需进一步探讨。因此,研究者尝试结合Transformer的全局建模能力与CNN的局部细节捕获优势,提出一系列混合架构模型。姚成武等^[16]设计的UNetFormer分割模型,有效增强了对细小及边缘裂缝特

征的关注度;Zhang等^[17]设计的ISTD-CrackNet融合CNN与Transformer的优势,使得裂缝断裂情况减少。混合模型虽能提升分割性能,但CNN与Transformer的结合会大幅增加模型复杂度,且在复杂环境干扰下全局语义与局部细节特征的平衡问题亟待解决。

针对目前大多数研究工作主要集中在裂缝检测,而对其他类型破损的研究相对较少。此外,由于各类破损形态差异较大,常常存在边缘模糊、背景干扰等问题,分割效果不佳。因此,本文基于无人机巡检获取的航拍图像展开研究,提升巡检效率,提出MDPR-DeepLabv3+模型。主要贡献包括:

1)将Xception主干网络替换为MobileNetV2,结合Dice损失与Focal损失以提升模型处理样本不平衡数据的能力。

2)在浅层特征输入解码器前,引入PSA_M注意力机制,恢复破损边缘信息。

3)在编码器设计DFSP模块,即将深层特征分支与Dense ASPP处理的分支在特征融合模块FFM中进行整合,有效融合局部细节与全局信息,提升对细小破损的分割能力。

4)在解码器提出RCD模块,增强浅层与深层特征之间的交互作用,实现更优的特征融合效果。

1 路面破损分割网络MDPR-DeepLabV3+

尽管DeepLabV3+模型^[18]在路面破损语义分割中展现了较高的准确性,但在识别细小或复杂裂缝以及处理不同尺度破损的能力方面仍存在局限性。

此外,DeepLabV3+模型的参数量较大,导致其计算复杂性增高,从而限制了其在资源受限环境中的应用。因此,本文以DeepLabV3+为基础模型进行改进,提出MDPR-DeepLabv3+模型。该模型主要由四部分组成:轻量化主干网络MobileNetV2、编码器中的DFSP模块、处理浅层特征的PSA_M模块,以及增强深浅层信息融合的RCD模块。其整体结构如图1所示。

1.1 引入MobilenetV2骨干网络

DeepLabV3+的原始主干网络Xception虽然在复杂语义分割中表现优异,但因网络层次深,导致其模型较大、计算资源消耗高。为提高运行效率,选用Google 2018年提出的MobileNetV2作为主干网络,其在保持低计算复杂度和参数量的同时,仍能保持较高特征表达能力与识别精度^[19]。

MobileNetV2主要由倒残差结构、线性瓶颈模块和深度可分离卷积3部分构成,其结构组成如表1所示。倒残差结构先通过 1×1 卷积升维,在高维空间学习丰富的特征表达;再利用深度可分离卷积提取空间特征,降低计算复杂度;最后用点卷积降维。倒残差结构中的残差连接仅在输入输出通道一致且步幅为1时加入,以保证信息流传递和

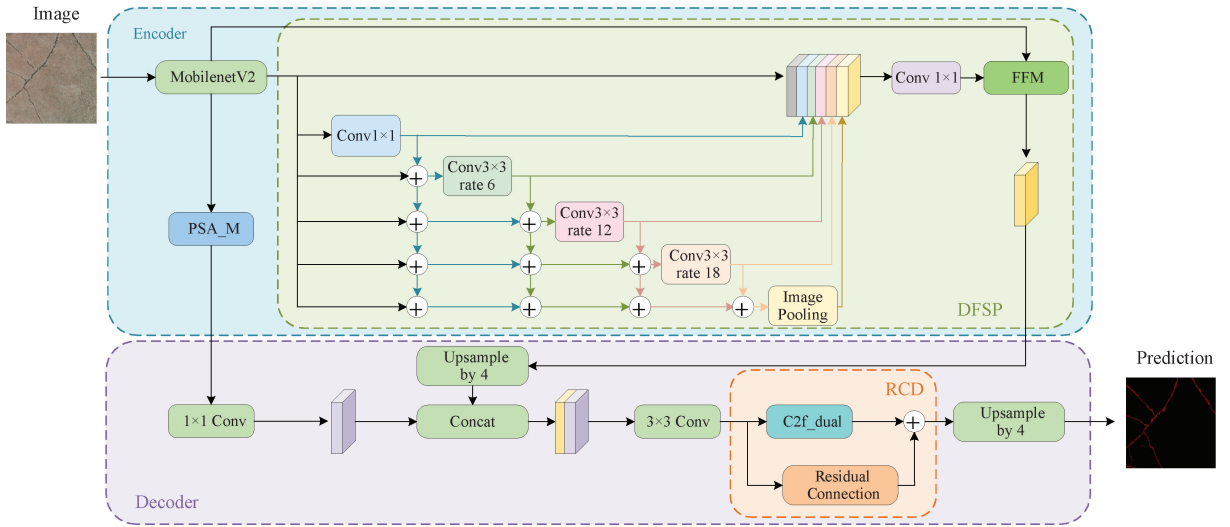


图 1 MDPR-DeepLabV3+ 结构

Fig. 1 MDPR-DeepLabV3+ structure

表 1 MobileNetV2 网络结构组成

Table 1 MobileNetV2 network structure

算子	扩张系数	输出通道数	重复次数	步长
Conv2d	—	32	1	2
Bottleneck	1	16	1	1
Bottleneck	6	24	1	2
Bottleneck	6	32	1	2
Bottleneck	6	32	2	1
Bottleneck	6	64	4	2
Bottleneck	6	96	3	1
Bottleneck	6	160	3	2
Bottleneck	6	320	1	1
Conv2d 1×1	—	1 280	1	1
Avgpool	—	1 280	1	—
Fully connected layer	—	4	1	—

梯度传播。线性瓶颈模块用线性激活函数替代 ReLU，避免低维空间特征损失，提升表达能力。这些设计使 MobileNetV2 兼顾高效性与精度，适合无人机巡检航拍图像的路面破损检测，故选择其作为模型特征提取网络。

1.2 DFSP 模块

在 DeepLabV3+ 中，标准空洞空间金字塔池化 (atrous spatial pyramid pooling, ASPP) 虽通过不同空洞率的并行卷积提取特征，但各分支信息相互独立，仅通过简单拼接后的一维卷积融合，难以有效整合多尺度信息。此外，尽管 ASPP 引入全局平均池化分支获取全局语义，但其对复杂场景中远距离依赖的建模能力依然有限。针对上述问题，本文基于 ASPP 提出 DFSP 模块，该模块由密集连接的 ASPP (Dense ASPP) 和特征融合模块 (feature fusion module, FFM) 构成。

Dense ASPP 引入了密集连接机制，使得不同空洞率特征互补，减少信息损失，捕获更丰富的多尺度细节，有效实现跨尺度特征增强；并将 MobileNetV2 深层特征输入 ASPP，深层特征分支为难分辨目标提供额外结构与纹理信息，增强模型细节感知能力。Dense ASPP 结构如图 2 所示。

同时，DFSP 创新性地采用 FFM 模块以增强全局语义理解。FFM 模块主要由交叉注意力机制、特征交互层和通道嵌入层 3 部分构成，连接主干网络深层特征图分支 FI (feature image) 与 Dense ASPP 的输出特征图 PM (prediction map), $FI \in \mathbb{R}^{H \times W \times C}$, $PM \in \mathbb{R}^{H \times W \times C}$, FFM 如图 3 所示。

首先将特征图平推为 $\mathbb{R}^{N \times C}$ 的特征序列，其中 $N = H \times W$ 。通过 Transformer 架构中的交叉注意力原理^[20]，实现两组特征序列间的全局信息交换，其计算复杂度 $O(N^2)$

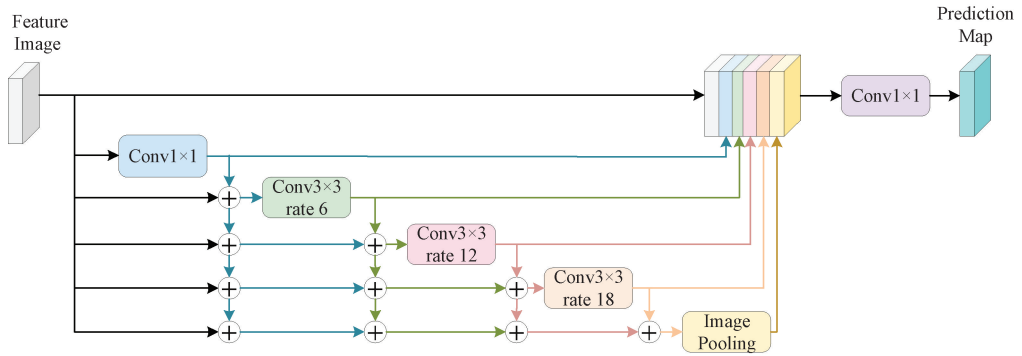


图2 Dense ASPP 结构

Fig. 2 Dense ASPP structure

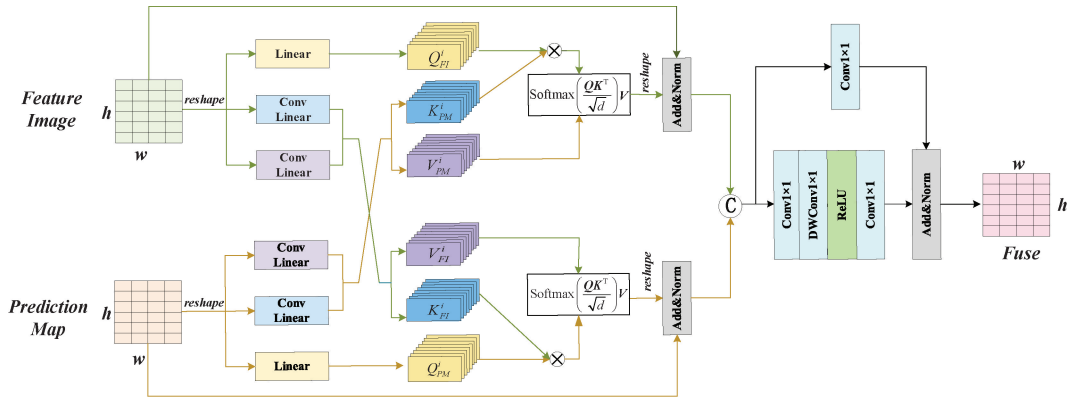


图3 FFM 结构

Fig. 3 FFM structure

如式(1)所示。

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V} \quad (1)$$

其中, $\mathbf{Q} \in \mathbb{R}^{N \times C}$, $\mathbf{K} \in \mathbb{R}^{N \times C}$, $\mathbf{V} \in \mathbb{R}^{N \times M}$ 由输入特征序列通过线性层生成, $N = H \times W$ 表示输入序列的长度。 C 是序列 \mathbf{Q} 和 \mathbf{K} 的维数, M 是序列 \mathbf{V} 的维数, 在 FFM 模块中 $C = M$, d 为每个注意力头中的特征向量的维数。采用序列缩减策略, 通过滑动卷积对特征序列进行下采样(下采样因子为 R), 得到更小的关键向量 $\mathbf{K} \in \mathbb{R}^{(N/R) \times C}$ 和值向量 $\mathbf{V} \in \mathbb{R}^{(N/R) \times C}$, 使得矩阵乘法复杂度从 $O(N^2)$ 降低到 $O((N^2/R))$, 有效降低整体计算复杂度和内存使用。

通过多头交叉注意机制扩展模型表达能力, 使其更适合处理复杂的输入关系。首先获得 FI 和 PM 的向量 $\mathbf{Q}_{FI}^i \in \mathbb{R}^{N \times d_{\text{head}}}$, $\mathbf{K}_{PM}^i \in \mathbb{R}^{(N/R) \times d_{\text{head}}}$, $\mathbf{V}_{PM}^i \in \mathbb{R}^{(N/R) \times d_{\text{head}}}$, $\mathbf{Q}_{PM}^i \in \mathbb{R}^{N \times d_{\text{head}}}$, $\mathbf{K}_{FI}^i \in \mathbb{R}^{(N/R) \times d_{\text{head}}}$ 和 $\mathbf{V}_{FI}^i \in \mathbb{R}^{(N/R) \times d_{\text{head}}}$, 其中, i 为注意力头部, d_{head} 为每个特征向量的维数, 通过式(2)计算 FI 和 PM 的注意矩阵:

$$\begin{cases} \mathbf{S}_{FI}^i = \text{Softmax}\left(\frac{\mathbf{Q}_{FI}^i \mathbf{K}_{PM}^i}{\sqrt{d_{\text{head}}}}\right), i = 1, 2, \dots, \text{heads} \\ \mathbf{S}_{PM}^i = \text{Softmax}\left(\frac{\mathbf{Q}_{PM}^i \mathbf{K}_{FI}^i}{\sqrt{d_{\text{head}}}}\right) \end{cases} \quad (2)$$

将 FI 注意矩阵与 PM 注意矩阵相乘, 得到修正后的 FI 特征图。随后, 将修正后的 FI 特征图沿通道维度进行拼接, 得到 $\mathbf{Z}_{FI} \in \mathbb{R}^{N \times C}$ 。同理, $\mathbf{Z}_{PM} \in \mathbb{R}^{N \times C}$ 也通过相同方式得到, 上述过程如式(3)所示。

$$\begin{cases} \mathbf{Z}_{FI} = \text{Concat}(\mathbf{S}_{FI}^i \times \mathbf{V}_{PM}^i), i = 1, 2, \dots, \text{heads} \\ \mathbf{Z}_{PM} = \text{Concat}(\mathbf{S}_{PM}^i \times \mathbf{V}_{FI}^i) \end{cases} \quad (3)$$

将最终得到的 \mathbf{Z}_{FI} , \mathbf{Z}_{PM} 特征图转换为 $\mathbb{R}^{H \times W \times C}$, 并添加至原始特征图中, 获得交互式增强的 FI、PM 特征图。通过残差卷积块对融合后的特征进行通道维度重构, 将两个交互增强的特征图进行有效融合, 融合后的特征图 $\mathbf{Fuse} \in \mathbb{R}^{H \times W \times C}$ 。

DFSP 模块通过 Dense ASPP 整合多尺度上下文信息, 增强路面破损感知; FFM 模块进行特征交互实现层级间的全局融合, 提升复杂背景下的语义理解, 进而提高检测准确率。

1.3 PSA_M 注意力机制结构

无人机巡检航拍的路面破损图像存在背景复杂、边界模糊、光照与阴影干扰显著等特点, 采用改进的 PSA 注意力机制^[21] PSA_M, 处理传入解码器前的浅层特征信息, 其采用混合带状池化(mixed pooling module, MPM)替换原 PSA 中空间维度分支的全局平均池化, 既能增强边界信息

关注,又能提升细长条状路面破损的分割效果。

MPM 由自适应平均池化层和条形池化(strip pooling module, SPM)^[22]组成,MPM 结构如图 4 所示。MPM 并行使用两个自适应平均池化:一个池化层用于提取全局上下文信息,另一个池化层则用于提取局部细节;条形池化层通过条带状特征聚合增强空间感知能力,有效地捕获路面破损的长距离依赖关系。在 SPM 中,输入特征图尺寸为 $H \times W$,通道数为 C 。对特征图的任一通道, $x \in \mathbb{R}^{H \times W}$ 经水平池化尺寸变为 $H \times 1$,经垂直池化后尺寸变为 $1 \times W$,然后分别对两种池化结果的特征值进行平均。SPM 结构如图 5 所示。水平、垂直条形池化计算公式如式(4)、(5)所示。

$$y_i^h = \frac{1}{W} \sum_{0 \leq j \leq W} x_{i,j} \quad (4)$$

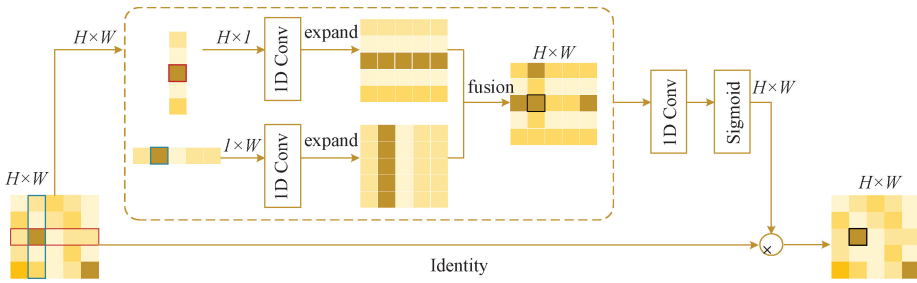


图 5 条形池化结构

Fig. 5 Strip pool structure

将两个输出结果分别在垂直和水平方向扩展至相同尺寸后,对应位置逐元素相加融合,获得的特征图尺寸与原始输入图像的大小一致,如式(6)所示。

$$y_{i,j} = y_i^h + y_j^v \quad (6)$$

PSA_M 利用极化滤波,包含通道与空间两个自注意力分支,在通道保持 $C/2$ 维度,空间保持 $[H, W]$ 维度,减小降维的信息损失。同时,在通道与空间分支结合 Softmax 和 Sigmoid 函数以适应两分支不同的特征建模需求,提升细粒度回归输出分布的拟合精度。最后两个分支结果融合后输出,具体结构如图 6 所示。

在通道维度计算的自注意力机制 $A^{ch}(X) \in \mathbb{R}^{C \times 1 \times 1}$ 由式(7)得到:

$$A^{ch}(X) = F_{SG}[C_z((\sigma_1(C_v(X)) \times F_{SM}(\sigma_2(C_q(X)))))] \quad (7)$$

其中, C_v, C_q, C_z 为 1×1 卷积, σ_1, σ_2 为重塑算子, F_{SG} 为 Sigmoid 函数用于注意力映射, F_{SM} 为 Softmax 函数对信息进行增强。

最后,输入特征 X 与通道维度自注意力 A^{ch} 融合,输出特征 $Z^{ch} \in \mathbb{R}^{C \times H \times W}$,如式(8):

$$Z^{ch} = A^{ch}(X) \odot^{ch} X \quad (8)$$

其中, \odot^{ch} 为在通道维度进行矩阵点乘的操作符。

在空间维度的自注意力 $A^{sp}(X) \in \mathbb{R}^{1 \times H \times W}$ 由式(9)

$$y_j^v = \frac{1}{H} \sum_{0 \leq i \leq H} x_{i,j} \quad (5)$$

其中, $x_{i,j}$ 表示特征图中第 i 行 j 列元素, y_i^h 为水平条形池化在第 i 行的输出结果, y_j^v 为垂直条形池化在第 j 列的输出结果, W 为宽度, H 为高度。

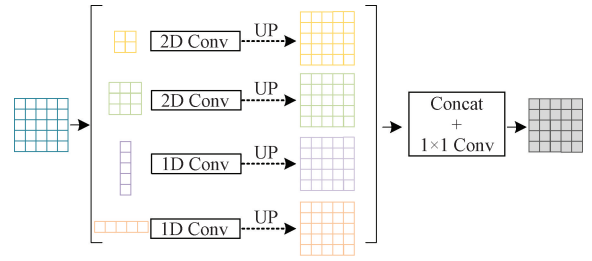


图 4 混合带状池化结构

Fig. 4 Mixed banded pooling structure

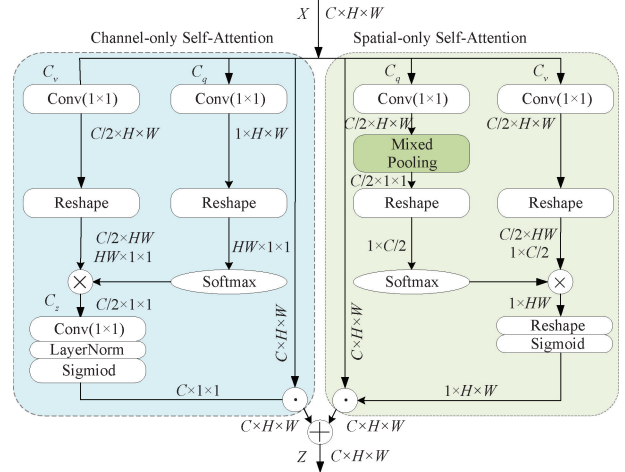


图 6 PSA_M 结构

Fig. 6 PSA_M structure

得到:

$$A^{sp}(X) = F_{SG}[\sigma_3(F_{SM}(\sigma_1(F_{MP}(C_q(X)))) \times \sigma_2(C_v(X)))] \quad (9)$$

其中, F_{MP} 函数为混合带状池化,转换空间尺寸大小。输入特征 X 与空间维度自注意力 A^{sp} 融合,输出特征 $Z^{sp} \in \mathbb{R}^{C \times H \times W}$ 如式(10)所示。

$$Z^{sp} = A^{sp}(X) \odot^{sp} X \quad (10)$$

其中, \odot^{sp} 为在空间维度进行矩阵点乘的操作符。

两个分支通过并行连接, Z^{ch} , Z^{sp} 融合输入特征 X 得到输出特征 $Z \in \mathcal{R}^{C \times H \times W}$, 如式(11)所示。

$$PSA(X) = Z^{ch} + Z^{sp} = A^{ch}(X) \odot^{ch} X + A^{sp}(X) \odot^{sp} X \quad (11)$$

通过 PSA_M 增强空间边界和细节信息的提取, 提高模型对各类破损的分割效果。

1.4 RCD 模块

C2f(concatenate to fuse)模块由 Ultralytics 团队随 YOLOv8 提出, 是一种高效的特征融合与轻量化卷积结构, 能以低计算成本实现信息融合并改善梯度流动, 有效解决传统卷积融合中信息交互不足的问题。相较于 DeepLabV3+ 解码器仅依赖普通卷积融合深浅层语义信息, 存在融合深度不足、多级语义互动不充分等局限, C2f 通过分片处理保留原始特征细节, 递进卷积强化非线性表达, 拼接融合聚合多阶段特征信息, 显著提升了融合效果。然而, 尽管 C2f 模块具有通用性, 其递进卷积仍采用标准的 3×3 卷积, 难以兼顾坑槽与裂缝的多尺度需求, 且仅通过分支拼接融合, 易丢失浅层边缘细节。

为此, 本文在解码器 3×3 卷积后提出双路径特征解耦融合 RCD(residual channel decoupling)模块, 基于 C2f 框架, 将其递进卷积单元替换为分组卷积 Dual Conv, 同时新增残差分支 $residual_{conv}$ 保留基础融合特征基线, 通过可学习权重的残差连接, RCD 模块能够动态融合 C2f_Dual 增强特征与残差特征, 从而避免特征丢失、提升多样性。相较于 C2f, RCD 通过 Dual Conv 解耦增强与残差连接的双重优化, 促进深浅层信息互补, 二者详细对比如图 7 所示。

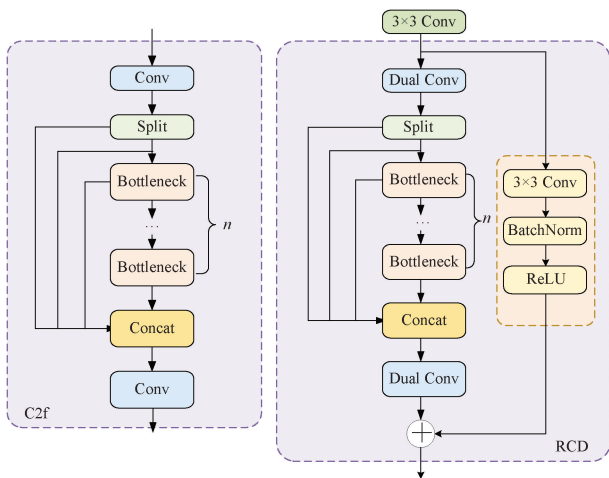


图 7 C2f 与 RCD 结构

Fig. 7 C2f and RCD structure

DualConv 通过结合 3×3 和 1×1 卷积处理相同输入特征通道^[23], 如图 8 所示。其采用组卷积对输入和输出特

征图分组, 同一组输入通道可被组内不同的卷积核并行处理, 不仅减少了计算开销和参数量, 还有效提升了信息流处理效率与特征提取效果。

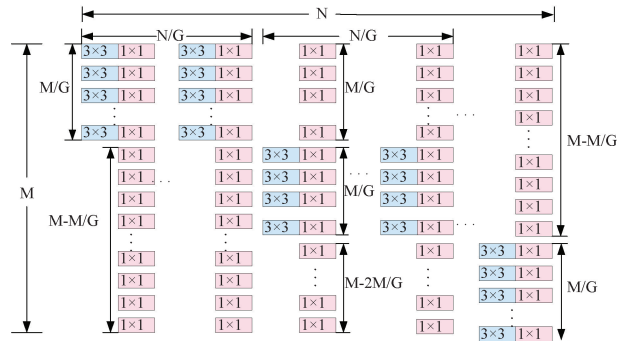


图 8 Dual Conv 结构

Fig. 8 Dual Conv structure

残差连接通过跳跃连接机制缓解了深层网络的梯度消失问题, 将浅层空间结构特征与深层语义特征进行加权融合。尽管引入额外卷积层, 但通过组卷积技术和可学习权重机制, 能够增强对小面积破损、细小裂缝等细节的捕捉能力, 融合过程如式(12)所示。

$$O = \sigma(W_{res}) \otimes residual_{conv}(X) + (1 - \sigma(W_{res})) \otimes C2f_{dual}(X) \quad (12)$$

其中, X 为输入特征, $residual_{conv}$ 表示残差路径变换, $C2f_{dual}$ 为改进的特征融合模块输出, W_{res} 为可学习权重, σ 为 Sigmoid 激活函数, \otimes 为张量积。

2 网络模型训练与评价指标

2.1 网络模型训练

实验使用 Windows10 系统, 处理器为 Intel i7-10700F 16 GB, GPU 为 NVIDIA GeForce RTX3090 24 GB, 编程环境 Python3.10.16, 基于 PyTorch2.0 框架训练模型。训练参数如表 2 所示。

表 2 模型训练参数

Table 2 Model training parameters

参数	数值/类别
训练轮次	300
输入尺寸	1 024 × 1 024
批次大小	2
动量	0.9
工作线程数	2
最大学习率	5×10^{-4}
学习率衰减类型	cos
学习率优化器	Adam

2.2 无人机巡检航拍路面破损数据集

针对现有路面破损公开数据集类别单一、数据量小且

缺乏无人机航拍视角下的问题,构建了无人机巡检航拍路面破损数据集。选用大疆 M300 RTK 无人机,搭配禅思 P1 相机拍摄路面破损图像,原始图像分辨率为 $8\ 192 \times 5\ 460$,将原始图像裁剪为 $1\ 024\ \text{pixel} \times 1\ 092\ \text{pixel}$ 的切片,以防止训练中图像压缩造成的关键信息丢失。此外,数据集按 8:2 的比例划分为训练验证集和测试集。在训练过程中,动态抽取 20% 的训练样本作为验证集,以便调整模型

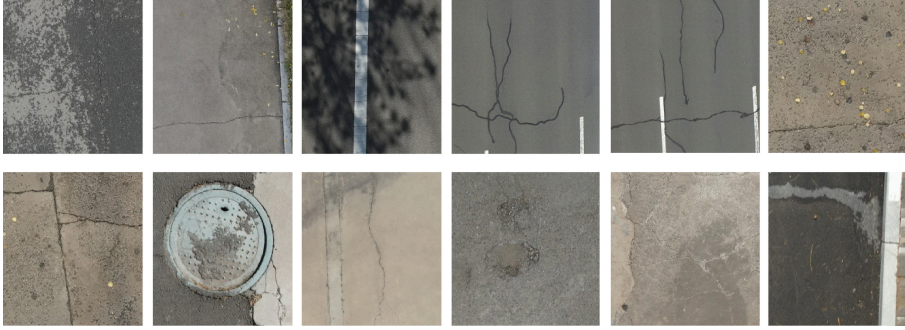


图 9 无人机巡检航拍下的路面破损实例

Fig. 9 An example of road damage under UAV inspection aerial photography

2.3 混合损失函数

在无人机巡检路面破损分割任务中,由于破损区域占比小、类别不平衡,传统交叉熵损失易因背景像素主导导致模型预测偏向背景。为此,本文结合 Focal 损失与 Dice 损失,Focal 损失能够增强对占比小类别的关注,减少背景干扰;Dice 损失从区域层面优化分割质量,改善边缘和细节识别。结合后的损失函数既能解决样本不平衡问题,又能稳定训练过程。

Focal 损失与 Dice 损失计算公式为式(13)、(14):

$$L_{Dice} = 1 - \frac{\sum_{i=1}^N p_i y_i + \epsilon}{\sum_{i=1}^N p_i + y_i + \epsilon} - \frac{\sum_{i=1}^N (1 - p_i)(1 - y_i) + \epsilon}{\sum_{i=1}^N (2 - p_i - y_i) + \epsilon} \quad (13)$$

$$L_{Focal} = -\alpha_i (1 - p_i)^\gamma \log_e(p_i) \quad (14)$$

其中, N 为类别数, y_i 为第 i 个像素点的真实标签, p_i 为第 i 个像素点的预测概率, ϵ 为用于确保数值稳定性的极小常数, γ 为调节权重因子的超参数, γ 总是大于 0, α_i 为权重因子。

混合损失函数计算式为式(15):

$$L = L_{Dice} + L_{Focal} \quad (15)$$

其中, L 为总损失, L_{Dice} 为 Dice 损失函数, L_{Focal} 为 Focal 损失函数。

2.4 评价指标

为对比各模型分割性能,选用 3 个指标衡量分割效果,分别为平均交并比(mean intersection over union, mIoU)预测与真实区域重叠比例的类别平均值,衡量分割精度;平均像素准确率(mean pixel accuracy, mPA)反映整体分类准确度;平均精度(mean precision, mPrecision)关注

超参数并监测过拟合情况。数据集使用 Labelme 进行标注并转换为 PNG 格式的标签。

数据集覆盖高速公路、国道、省道及校园道路等多种路段,涵盖不同飞行高度/速度、时间、天气等复杂环境因素,包含路面杂物、阴影遮挡、道路标线重叠等噪声干扰,包含 1 885 张图像,覆盖水泥路、沥青路,破损类型包括裂缝、坑槽、修补这 3 类,部分代表性图像如图 9 所示。

正例预测准确性,计算方法如式(16)~(18)所示。

$$mIoU = \frac{1}{N} \sum_{i=1}^N \frac{y_{ii}}{\sum_{j=1}^N y_{ij} + \sum_{j=1}^N y_{ji} - y_{ii}} \quad (16)$$

$$mPA = \frac{1}{N} \sum_{i=1}^N \frac{y_{ii}}{\sum_{j=1}^N y_{ij}} \quad (17)$$

$$mPrecision = \frac{1}{N} \sum_{i=1}^N \frac{y_{ii}}{\sum_{j=1}^N y_{ji}} \quad (18)$$

其中, N 为类别数; y_{ij} 是第 i 类像素点被预测为 j 类的像素数, y_{ji} 为第 j 类像素点被预测 i 类的像素数; y_{ii} 表示正确分类的像素数。

3 实验结果与分析

3.1 实验结果对比

为验证模型分割性能,选取 DeepLabV3+ 结合 4 种不同主干网络,同时与 HRNet、PSPNet、UNet、UNet++、Segformer^[24] 以及本文提出的方法,共计 10 种模型在自制数据集进行对比实验。根据表 3 的结果,本文方法在 mIoU 与 mPA 分别达到 78.47% 和 92.03%,显著优于其他对比模型。虽然 mPrecision 略低于 Segformer 的 86.14%,但依然优于其他对比模型,表明 Segformer 在预测正例的准确性(即预测为破损的区域中真实破损的比例)上仍具有一定优势。虽然 Segformer 在准确性上表现优秀,但其参数大小为 53 505 KB;而本文方法在实现高性能的同时,参数量为 30 235 KB,显示出更优的计算效率。综上,本文方法的整体性能优于其他对比模型,适合无人机巡检场景下的路面破损检测需求。

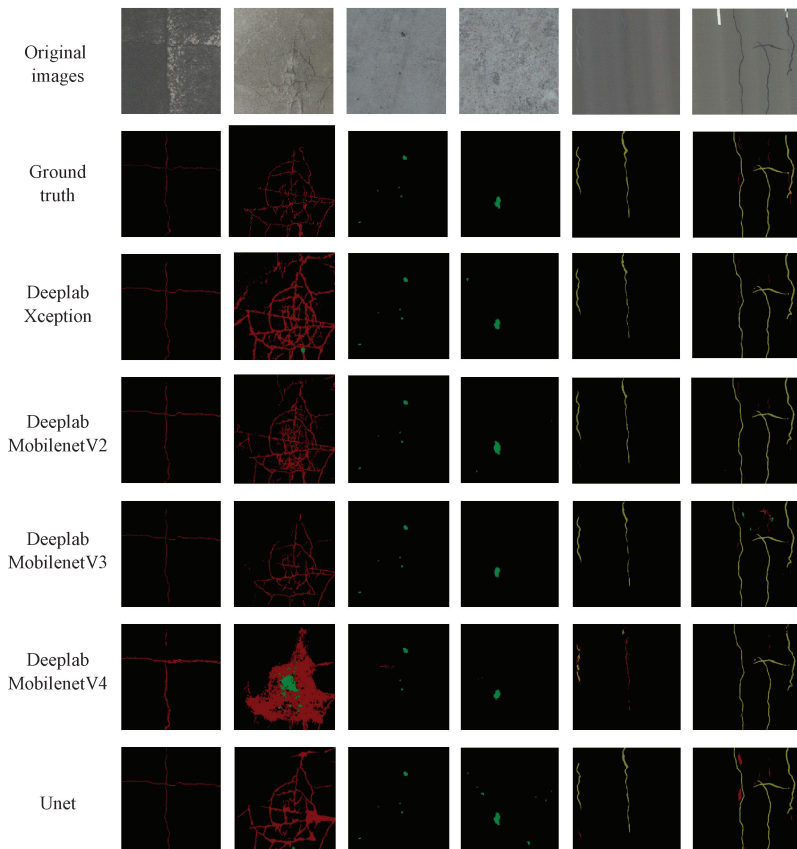
表3 自制数据集上不同模型的性能对比

Table 3 Performance comparison of different models on self-made datasets

模型	mIoU/%	mPA/%	mPrecision/%	Params/KB
Xception-DeeplabV3+	70.93	82.24	81.03	214 745
MobilenetV2-DeeplabV3+	69.13	84.90	75.20	22 980
MobilenetV3-DeeplabV3+	64.56	76.28	78.59	19 593
MobilenetV4-DeeplabV3+	49.83	69.38	59.23	30 478
HRnet	60.87	76.73	71.26	258 248
PSPnet	61.94	83.62	68.11	182 807
Unet	71.96	86.71	78.00	97 249
Unet++	63.24	81.01	76.45	142 336
Segformer	74.21	81.57	86.14	53 505
MDPR-DeepLabV3+	78.47	92.03	83.93	30 235

图10展示了在自制数据集上不同分割方法下的性能对比情况,其中第1~2列、3~4列、5~6列分别对应裂缝、坑槽、修补这3类破损的分割结果。第1列中,Deeplab Xception、Deeplab MobilenetV2、Deeplab MobilenetV3、Deeplab MobilenetV4、Unet++、Segformer模型存在裂缝断裂和边界粗糙现象;第2列中,Deeplab Xception、Deeplab MobilenetV4、Unet、Segformer模型将部分块状裂缝错误分类为坑槽,而Unet++模型将块状裂缝中部和底部部分区域误分类为修补;第3列中,Deeplab

MobilenetV4和Deeplab MobilenetV2模型对坑槽存在错检和漏检;第4列中,本文方法与PSPnet模型分割结果与标签近似,其他模型存在漏检和错检;第5列中,Deeplab MobilenetV2、Deeplab MobilenetV4、Unet、Unet++、PSPnet模型将修补部分误检为裂缝,Deeplab MobilenetV3、Segformer存在漏检,本文模型与真实标签最为接近;第6列中,PSPnet模型对修补区域的分割边界较粗,Unet++和本文模型可以将较细裂缝检测出来,其他模型均存在漏检和错检现象。



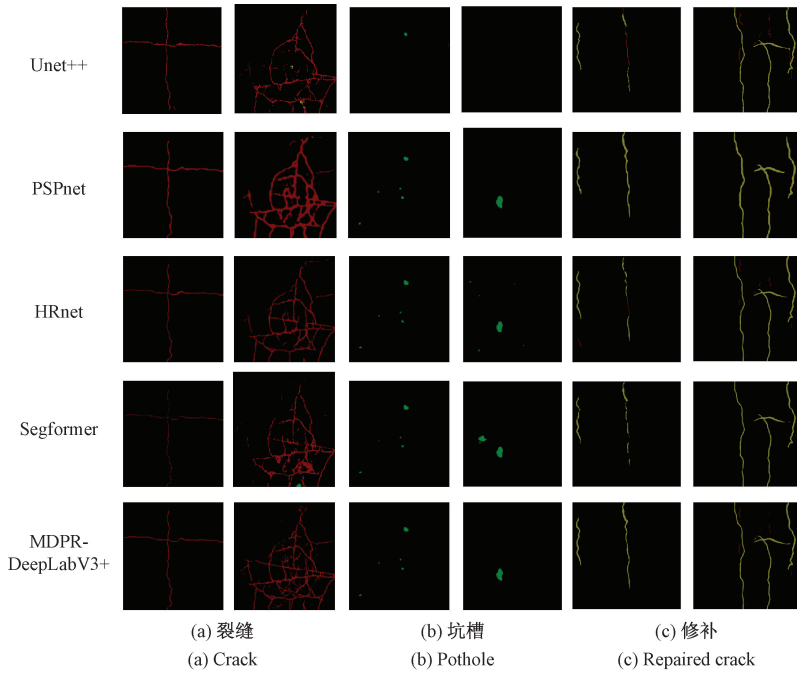


图 10 不同模型分割结果可视化

Fig. 10 Segmentation result visualization for various models

综上,本文模型在复杂背景下的路面破损识别及细小裂缝捕捉方面,分割效果均优于对比模型。图 10 的实际分割效果与表 1 的实验各项指标相互印证,验证了该模型在路面破损分割中的良好性能。

3.2 消融实验

为探究各模块对模型性能的具体贡献,在基线模型中分别添加 DFSP、PSA_M、RCD 模块,进行了多组消融实验,

结果如表 4 所示。实验 1 为基线模型,其参数量较大。实验 2 采用 MobilenetV2 主干网络,参数量降至约基线的 1/10,但 mIoU 降至 69.13%,表明单独替换主干虽提升效率,但需配合其他模块保证精度。实验 3 构建的 DFSP 模块使 mIoU 和 mPA 分别提升至 74.73% 和 91.91%,增强了多尺度特征提取能力。实验 4 中 PSA_M 模块在恢复裂缝边缘信息上效果显著,mPA 提升至 87.09%。

表 4 模型消融实验

Table 4 Model ablation experiment

实验序号	MobilenetV2	DFSP	PSA_M	RCD	mIoU/%	mPA/%	mPrecision/%	Params/KB
1					70.93	82.24	81.03	214 745
2	✓				69.13	84.90	75.20	22 980
3	✓	✓			74.73	91.91	78.69	26 454
4	✓		✓		71.04	87.09	77.52	22 766
5	✓			✓	74.82	85.24	84.09	26 509
6	✓	✓	✓		78.17	90.49	82.32	26 477
7	✓		✓	✓	76.55	86.80	84.81	26 532
8	✓	✓	✓	✓	78.47	92.03	83.93	30 235

实验 5 提出 RCD 模块促进深浅层信息互补,mPrecision 达到 84.09%。实验 6 中 DFSP 与 PSA_M 联合使用使 mIoU 提升至 78.17%,实验 7 中 PSA_M 与 RCD 组合在细节识别上达成最高的 mPrecision 84.81%。实验 8 中为 4 个模块协同作用的结果,虽然 mPrecision 略逊于实验 7,但 mPA 达到 92.03%、mIoU 达到 78.47%,

显著提升了模型的检测性能。消融实验结果表明,将 4 个改进部分引入到模型中能有效提升模型在路面破损分割任务中的性能表现。

3.3 泛化性实验

为充分验证所提方法的性能,本文将模型与 HRNet、PSPNet、UNet 等 9 种主流方法,以及最新改进的

DeepLabV3+模型文献[25-26]在公开数据集 Crack500 上进行了对比,结果如表 5 所示。这些模型代表了路面破损检测的先进方法,与最新改进的 DeepLabV3+模型进行对比,有助于验证本文方法的竞争力及其实际应用效果。结果表明,本文模型的 mPA 略低于 Segformer 的 95.79%,这说明针对 Crack500 数据集,模型存在少量背景像素被误分类为破损的情况,细粒度分类精度稍逊于 Segformer 的全局建模能力。不过,本文所提出的方法在 mIoU 和 mPrecision 上分别达到了 80.93%和 93.41%,显著优于其他对比模型,表现出在分割区域边界和减少

误检方面的良好能力。此外,本文模型的参数量为 30 233 KB,相较于 Xception-DeeplabV3+、HRNet 等模型,显示出良好的计算复杂度。这表明,尽管更大参数量的模型能够提供高性能,但在实际应用中往往会增加计算资源的需求。同时,文献[25]的 mIoU 为 57.21%、mPA 为 75.09%、mPrecision 为 70.60%,参数量为 14 878 KB;而文献[26]的对应指标为 72.50%、66.00%和 61.00%,参数量为 1 996 KB。尽管这两篇文献改进的 DeepLabV3+模型的参数量较小,但在多个指标上的性能仍低于本文方法。

表 5 Crack500 数据集上不同模型的性能对比

Table 5 Comparison of quantitative results of different models on Crack500 dataset

模型	mIoU/%	mPA/%	mPrecision/%	Params/KB
Unet	72.95	85.48	79.76	97 248
Xception-DeeplabV3+	76.14	89.88	81.28	214 733
MobilenetV2-DeeplabV3+	74.90	94.40	78.16	22 978
MobilenetV3-DeeplabV3+	70.07	93.76	73.84	19 591
MobilenetV4-DeeplabV3+	68.99	89.73	73.58	30 476
PSPnet	74.45	89.12	79.66	182 802
HRnet	75.62	89.66	80.78	258 242
Unet++	78.43	85.55	83.32	142 336
Segformer	79.07	95.79	81.75	53 503
文献[25]	57.21	75.09	70.60	14 878
文献[26]	72.50	66.00	61.00	1 996
MDPR-DeepLabV3+	80.93	93.41	84.83	30 233

综上所述,本文方法在保持较高分割精度的同时,有效平衡了模型轻量化与准确度之间的矛盾。这不仅使得模型在自制数据集上表现优异,还能够适应多种场景下的路面破损检测需求。

4 结 论

针对无人机巡检中路路面破损分割所面临的形态差异显著、复杂环境干扰等挑战,本文在 DeepLabV3+基础上提出改进的 MDPR-DeepLabV3+分割方法。该模型采用轻量级 MobileNetV2 作为主干网络,通过引入线性瓶颈结构和倒残差模块,在大幅减少计算量与参数量的同时,有效提升特征提取能力;在编码器中构建 DFSP 模块,将 ASPP 进行密集连接融合多尺度上下文信息,同时借助特征融合模块 FFM 中交叉注意力、特征交互等结构,实现不同层级特征的全局整合与高效交互;PSA_M 模块采用混合带状池化,强化对目标边界特征的关注;解码阶段提出残差卷积路径与 C2f_Dual 模块组成的双路径结构模块 RCD,以可学习的残差权重自适应调节信息融合比例,增强特征表达与信息流动,优化深层与浅层特征的融合效果。实验结果表明,本文模型能够有效应对无人机巡检航

拍图像中路路面破损形态差异大、难以分割的问题,显著提升了检测精度,并展现出良好的泛化能力。未来将在保持检测精度的前提下,进一步轻量化网络结构,实现无人机实时巡检路面破损分割。

参考文献

[1] 吴钢荣,侯丁戈,雷超. 高速公路路面病害检测技术研究综述[J]. 科技视界,2024,14(16):76-79.
WU G R, HOU D G, LEI CH. Summary of research on highway pavement disease detection technology[J]. Science and Technology Vision, 2024,14(16):76-79.

[2] 桂彦,叶文倩,王建新,等. 基于 CNN 和尺度自适应 Transformer 融合网络的路面裂缝分割方法[J]. 中国公路学报,2024,37(12):418-432.
GUI Y, YE W Q, WANG J X, et al. Pavement crack segmentation method based on CNN and scale adaptive Transformer fusion network [J]. China Journal of Highways, 2024,37(12):418-432.

[3] WANG J D, SUN K, CHENG T H, et al. Deep high-resolution representation learning for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine

- Intelligence,2020: 3349-3364.
- [4] ZHAO H SH, SHI J P, QI X J, et al. Pyramid scene parsing network[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2881-2890.
- [5] 胡波,陈翰新,任松,等. 一种基于分割掩码的隧道裂缝病害自动识别后处理算法[J]. 测绘学报, 2024, 53(9): 1715-1724.
- HU B, CHEN H X, REN S, et al. A post-processing algorithm for automatic identification of tunnel crack diseases based on segmentation mask[J]. Surveying and Mapping Journal, 2024, 53(9): 1715-1724.
- [6] 杨萍,张汐. 改进 DeepLabv3+的道路表面裂缝检测方法[J]. 计算机工程, 2025, 51(4): 261-270.
- YANG P, ZHANG X. Improved DeepLabv3 + method for road surface crack detection[J]. Computer Engineering, 2025, 51(4): 261-270.
- [7] 于天河,徐博超,侯善冲,等. 基于图像的道路裂缝分割及量化方法研究[J]. 仪器仪表学报, 2024, 45(9): 77-91.
- YU T H, XU B CH, HOU SH CH, et al. Research on image-based road crack segmentation and quantification methods [J]. Chinese Journal of Scientific Instrument, 2024, 45(9): 77-91.
- [8] NGUYEN N H T, PERRY S, BONE D, et al. Two-stage convolutional neural network for road crack detection and segmentation[J]. Expert Systems with Applications, 2021, 186: 115718.
- [9] 赵志宏,郝子晔,何朋. 融合注意力机制与 GhostUNet 的路面裂缝检测方法[J]. 电子测量技术, 2023, 46(24): 164-171.
- ZHAO ZH H, HAO Z Y, HE P. Pavement crack detection method combining attention mechanism and GhostUNet[J]. Electronic Measurement Technology, 2023, 46(24): 164-171.
- [10] 尹学辉,傅林琳,周尚波. 渐进式上下文交互和注意力机制的混凝土路面裂缝检测网络[J]. 计算机应用, 2025, 45(10): 3353-3362.
- YIN X H, FU L L, ZHOU SH B. Progressive context interaction and attention mechanism for concrete pavement crack detection network [J]. Computer Applications, 2025, 45(10): 3353-3362.
- [11] 段中兴,何宇超,张旭生,等. 有效特征提取和级联优化的路面裂缝检测方法[J]. 计算机辅助设计与图形学学报, 2024, 36(12): 2020-2028.
- DUAN ZH X, HE Y CH, ZHANG X SH, et al. An effective feature extraction and cascade optimization method for pavement crack detection[J]. Journal of Computer Aided Design and Graphics, 2024, 36(12): 2020-2028.
- [12] LI H B, ZHANG H W, ZHU H, et al. Automatic crack detection on concrete and asphalt surfaces using semantic segmentation network with hierarchical Transformer [J]. Engineering Structures, 2024, 307: 117903.
- [13] YU ZH W, CHEN Q Y, SHEN Y G, et al. Robust pavement crack segmentation network based on transformer and dual-branch decoder[J]. Construction and Building Materials, 2024, 453: 139026.
- [14] 周云,张文杰,陈祎林,等. 基于无人机单目视频的道路裂缝定位与量化方法[J/OL]. 工程力学, 1-14[2025-06-05]. <https://link.cnki.net/urlid/11.2595.o3.20241108.1437.028>.
- ZHOU Y, ZHANG W J, CHEN Y L, et al. Road crack location and quantification method based on UAV monocular video [J/OL]. Engineering Mechanics, 1-14[2025-06-05]. <https://link.cnki.net/urlid/11.2595.o3.20241108.1437.028>.
- [15] SHAN J H, HUANG Y, JIANG W, et al. GLoU-MiT: Lightweight global-local Mamba-guided U-mix transformer for UAV-based pavement crack segmentation[J]. Advanced Engineering Informatics, 2025, 65: 103384.
- [16] 姚成武,常琛. 基于像素强度顺序变换和 UNetFormer 的裂缝分割模型[J]. 电子测量技术, 2024, 47(11): 151-159.
- YAO CH W, CHANG CH. Crack segmentation model based on pixel intensity order transformation and UNetFormer [J]. Electronic Measurement Technology, 2024, 47(11): 151-159.
- [17] ZHANG Z Y, ZHUANG Y Y, SONG W D, et al. ISTD-CrackNet: Hybrid CNN-transformer models focusing on fine-grained segmentation of multi-scale pavement cracks[J]. Measurement, 2025, 251: 117215.
- [18] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]. European Conference on Computer Vision (ECCV), 2018: 801-818.
- [19] SANDLER M, HOWARD A, ZHU M L, et al. Mobilenetv2: Inverted esiduals and linear bottlenecks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. ArXiv preprint arXiv: 1706.03762, 2017.
- [21] LIU H J, LIU F Q, FAN X Y, et al. Polarized self-

- attention: Towards high-quality pixel-wise mapping[J]. *Neurocomputing*, 2022, 506: 158-167.
- [22] HOU Q, ZHANG L, CHENG M M, et al. Strip pooling: Rethinking spatial pooling for scene parsing[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 4003-4012.
- [23] ZHONG J CH, CHEN J Y, MIAN A. DualConv: Dual convolutional kernels for lightweight deep neural networks[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 34(11): 9528-9535.
- [24] XIE E Z, WANG W H, YU ZH D, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers [J]. *Advances in Neural Information Processing Systems*, 2021, 34: 12077-12090.
- [25] LI F F, MOU Y L, ZHANG Z Y, et al. A novel model for the pavement distress segmentation based on multi-level attention DeepLabV3 + [J]. *Engineering Applications of Artificial Intelligence*, 2024, 137: 109175.
- [26] QU ZH, MU G Q, YUAN B. A lightweight network with dual encoder and cross feature fusion for cement pavement crack detection [J]. *CMES: Computer Modeling in Engineering & Sciences*, 2024, 140(1): 255-273.

作者简介

韩建峰,教授,主要研究方向为图像处理与图像识别、信号处理、计算机视觉。

E-mail: hanjianfeng@imut.edu.cn

南汝君,硕士研究生,主要研究方向为信号与信息处理。

E-mail: nanrujun0920@163.com

宋丽丽(通信作者),副教授,主要研究方向为信号处理。

E-mail: songlili@imut.edu.cn