

DOI:10.19651/j.cnki.emt.2518716

UltraLight CrackNet: 基于 VMamba 的 轻量化裂缝分割网络*

成荣 朱文忠 王文

(四川轻化工大学计算机科学与工程学院 宜宾 644000)

摘要: 裂缝检测在土木基础设施维护中具有关键作用。传统人工视觉检测方法存在诸多缺陷,推动了裂缝检测技术的持续发展。然而,现有裂缝检测技术仍面临复杂背景干扰、特征多样性干扰及高计算资源需求的挑战。本研究挖掘 Mamba 模型在视觉任务中的潜力,提出一种超轻量裂缝检测网络(UltraLight CrackNet),其包含 3 个核心模块:并行轻量化视觉 Mamba 模块(通过高效建模长程依赖关系提取深层语义特征)、多尺度残差视觉状态空间模块(增强多尺度特征表征能力),以及改进的语义-细节融合模块(优化编码器-解码器架构的跳跃连接机制)。实验表明,该方法仅需 0.13 M 参数量与 1.96 G 浮点运算量,在超轻量模型设计下,于 DeepCrack 和 Crack500 数据集分别取得 87.85% 和 77.92% 的平均交并比 mIoU,达到最优性能;在 SteelCrack 数据集获得可比结果,且参数量较现有对比模型中参数量最小的模型降低 87.85%。

关键词: 裂缝检测;视觉 Mamba;轻量化模型;语义分割;特征融合

中图分类号: TP391.41;TN919.8 **文献标识码:** A **国家标准学科分类代码:** 520.6040

UltraLight CrackNet: A VMamba-based lightweight network for crack segmentation

Cheng Rong Zhu Wenzhong Wang Wen

(School of Computer Science and Engineering, Sichuan University of Science and Engineering, Yibin 644000, China)

Abstract: Crack detection is crucial in the maintenance of civil infrastructure. The many drawbacks of traditional manual visual inspection methods have led to the continuous development of crack detection methods. However, existing crack detection techniques face the challenges of complex backgrounds and feature diversity interference, and the high computational resource requirements. This study exploits the potential of Mamba for visual tasks and proposes an UltraLight CrackNet, which consists of a parallel lightweight visual Mamba block for efficiently modelling long-distance dependencies and extracting deep semantic features, a multi-scale residual visual state space block for enhanced multi-scale feature representation, and an enhanced semantics and detail infusion module for optimising skip connections within the encoder-decoder architecture. The experimental results show that our method requires only 0.13 M parameters and 1.96 G FLOPs, and achieves the optimal performance on DeepCrack and Crack500 datasets with ultra-lightweight model design, with the mean intersection over union (mIoU) of 87.85% and 77.92%, respectively, and obtains comparable results on SteelCrack dataset, and the number of parameters is 87.85% lower than that of the model with the smallest number of parameters among the available comparison models.

Keywords: crack detection; vision Mamba; lightweight model; semantic segmentation; feature fusion

0 引言

在复杂运行环境下,土木工程基础设施在其服役期内

不可避免地经历各类缺陷的产生、积累与扩展。这些缺陷不仅对结构安全构成重大风险,还会缩短其使用寿命。其中,裂缝作为结构性能退化的关键表征指标,需予以重点监

收稿日期:2025-04-28

* 基金项目:四川省科技计划重点研发项目(2023YFS0371)、企业信息化与物联网测控技术四川省高校重点实验室开放基金(2024WYJ03)、四川省智慧旅游研究基地(ZHYJ24-01)、四川轻化工大学研究生创新基金(Y2024115)项目资助

测^[1]。定期裂缝检测是获取基础设施裂缝信息的成熟且广泛应用的方法。当前裂缝检测主要依赖人工目视检查,该过程耗时耗力,且因人类判断的主观性易导致检测结果的不确定性与不一致性^[2]。鉴于传统方法的局限性,研究者与工程师日益关注融合先进人工智能与计算机视觉技术实现裂缝的智能化、高效化与精准化检测^[3]。基于研究任务性质,人工智能驱动裂缝检测方法通常分为三类:图像级裂缝分类、框级裂缝检测及像素级裂缝分割。早期裂缝分割模型主要基于卷积神经网络(convolutional neural networks,CNN)构建。尽管 CNN 擅长图像特征提取,但其建模长程依赖关系的能力受限于卷积操作的局部性特性。由于裂缝像素常沿长度方向全局分布于图像中,基于 CNN 的裂缝分割网络在有效建模裂缝特征的全局特性方面面临挑战。相比之下,基于自注意力机制构建的 Transformer 架构因其全局上下文建模能力而具有显著优势^[4]。在裂缝分割领域,Transformer 既可作为补充模块与 CNN 协同工作,也可完全替代 CNN 构建网络,从而增强模型捕获全局上下文信息的能力。这种融合 Transformer 的框架不仅展现出卓越的裂缝分割全局建模能力,还在捕获长程上下文特征方面具有显著优势^[5]。然而,尽管 Transformer 在裂缝分割中表现优异,但其计算复杂度与空间复杂度随输入序列长度的增加呈平方级数增长,这为边缘设备的实时处理带来了挑战。

近期,基于状态空间模型(state space model,SSM)的 Mamba 架构^[6]展现出成为 Transformer 替代方案的潜力。该架构在一维序列建模任务中表现出比 Transformer 更低的计算复杂度和更高的效率。Mamba 强大的全局建模能力与线性增长的计算复杂度相结合,使其成为图像处理领域潜在的高效模型。在计算机视觉领域,Mamba 已成功用于医学图像分割^[7]和遥感图像分割^[8]等任务。同时文献^[9]首次将 Mamba 架构引入全景分割任务,显著提升了全景分割的效率和抗畸变能力,但其对复杂纹理场景的挑战仍需进一步突破。

综上所述,当前裂缝检测方法仍面临以下核心挑战:

1)复杂背景与特征多样性干扰:裂缝常存在于具有复杂背景的场景中,其形态、宽度和分布呈现高度多样性。传统模型易受背景干扰,难以全面提取裂缝的细粒度特征。

2)高计算资源需求:多数深度学习模型(尤其是基于 Transformer 的架构)存在较高计算复杂度,难以满足边缘设备实时检测与部署的需求。

3)新型架构框架应用不足:尽管基于 SSM 的 Mamba 框架凭借高效的全局建模能力与线性复杂度特性,在视觉任务中展现出显著潜力,但其在裂缝检测领域的应用尚未得到充分探索。

本文主要贡献:

1)提出基于视觉 Mamba 的轻量化裂缝检测网络(UltraLight CrackNet)。该模型仅需 0.13 M 参数量与

1.96 G 浮点运算量(floating point operations,FLOPs),据本研究所知这是当前参数量最小的裂缝检测模型。在 DeepCrack 与 Crack500 数据集上达到最优性能,并在 SteelCrack 数据集上展现出优异的泛化能力与检测精度。

2)通过构建并行轻量化视觉 Mamba 模块(parallel lightweight visual mamba,PLVM)以高效建模长程依赖关系、多尺度残差视觉状态空间模块(multi-scale residual visual state space,MSRVSS)以增强多尺度特征表征,以及改进的语义-细节融合模块(enhanced semantics and detail infusion,E-SDD)以优化编码器-解码器特征融合机制,在平衡轻量化与效率的同时显著提升裂缝检测性能。

3)在 3 个公开数据集上开展全面消融实验与多维度对比实验,结果表明:所提方法以超轻量模型设计取得可比甚至最优性能,为视觉 Mamba 在裂缝检测领域的应用提供了新基准与理论参考。

1 相关工作

1.1 轻量化网络

随着基于深度学习的裂缝检测模型日益复杂化,研究者常通过引入附加模块提升检测鲁棒性。然而,这些模块的叠加不仅增加模型复杂度,还会导致检测过程中计算开销与响应延迟的显著上升。因此,设计轻量化网络对实现高效计算、降低资源消耗至关重要,特别是在需要部署于低成本设备的实际应用场景中。由于裂缝检测任务与通用计算机视觉任务存在差异性,其网络结构优化需特别关注裂缝特征的精细化提取,例如需使网络架构能够处理复杂裂缝形态、边缘细节及多尺度裂缝信息。为此,许多研究探索如何通过神经架构搜索与自动机器学习等技术,基于特定任务场景自动搜索最优模型架构。这些技术可在不过度增加计算复杂度的前提下自动优化模型结构,从而在裂缝检测等任务中取得更优效果。针对裂缝检测中的复杂背景干扰问题,Yu 等^[10]提出基于 U-Net 的轻量化网络(DSU-Net),通过深度可分离卷积构建轻量化解码器以降低模型复杂度,并设计 3 个功能模块提升检测精度。Zhou 等^[11]提出基于拆分交换卷积(split exchange convolution,SEConv)与多尺度特征交换(multi-scale feature exchange,MSFE)模块的轻量化模型,仅需 1.3 M 参数量与 8 G 浮点运算量即达到与现有方法相当的检测性能。Hou 等^[12]提出结合 MobileNet 与 MobileNet-SSD 的轻量化网络,通过迁移学习实现路面状态检测。

1.2 裂缝检测方法

裂缝检测与识别算法通常可分为两类:手动特征工程与自动特征工程。手动特征工程方法主要包括特征检测算法(如 Sobel 边缘检测、Canny 边缘检测等传统图像处理技术),这些方法在早期被广泛应用于裂缝检测。此外,许多传统机器学习检测算法采用滑动窗口技术定位潜在裂缝区域,并通过人工设计的特征提取器完成裂缝特征抽取。例

如,Zhang 等^[13]对采集的图像片段实施监督式深度卷积神经网络训练,实现图像片段的裂缝分类。

近年来,随着自注意力机制与 Transformer 架构在计算机视觉领域的广泛应用,Kang 等^[14]通过在 SegNet 中集成多头注意力机制,设计出适用于复杂场景的裂缝分割网络。König 等^[15]基于 U-Net 架构引入残差连接与注意力门控机制,提升裂缝分割精度。需指出,上述研究多采用全监督学习范式,需依赖大量人工标注的像素级裂缝标签。而弱监督学习通过降低标注需求,可处理标签信息不完整场景。例如,Dong 等^[16]在人工分类图像块后,利用类激活图(class activation mapping, CAM)与密集条件随机场(dense conditional random field,dCRF)生成伪标签。

2 本文方法

2.1 总体架构

如图 1 所示,本文模型采用经典 U 型架构,包含编码器、解码器及跳跃连接模块,共由 6 个层级构成。编码器前 3 层通过卷积模块提取浅层特征;每个卷积模块首先使用 3×3 标准卷积进行特征提取,随后应用组归一化(Group Norm)与 GELU 激活函数,最终通过最大化池化操作将特征图空间维度缩减至一半。第 4~6 层聚焦于深层语义特征提取,在各层级引入 PLVM 模块以高效捕获复杂特征。解码器采用与编码器对称的结构设计,并在跳跃连接中引入 E-SDI 模块,通过跨尺度融合多层次特征以增强目标区域的细粒度表征能力。

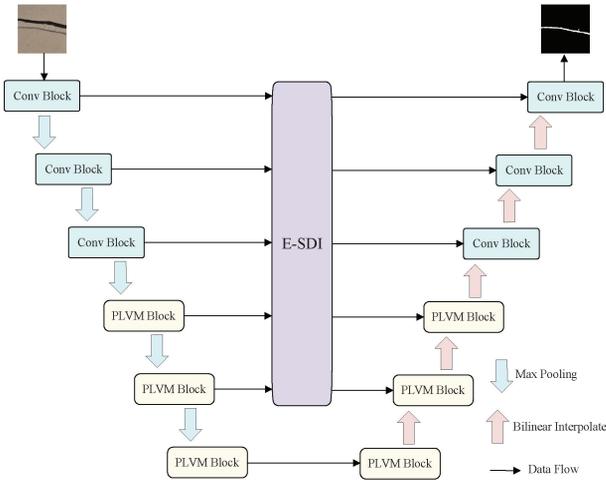


图 1 所提出方法的总体架构

Fig.1 Overall architecture of the proposed method

2.2 并行轻量化视觉 Mamba 模块

本文提出 PLVM 模块(如图 2 所示),旨在更高效建模长程依赖关系并提取深层语义特征。PLVM 通过 MSRVS 模块增强特征表征,同时保持低参数量与计算复杂度。其核心由若干关键模块构成:首先,输入特征 \mathbf{X}_in^C 经 RMSNorm 层归一化以稳定训练过程;随后将输入通道划

分为四部分,分别通过轻量化 MSRVS 模块进行特征提取。该模块结合 SSM 的线性复杂度与多尺度卷积运算的高效特征提取能力,显著提升长程依赖捕获性能。经并行处理的特征经重拼接、二次归一化后,通过线性变换映射至输出维度,最终生成输出特征 \mathbf{X}_{out}^C 。PLVM 模块的设计在复杂度与性能间实现了有效平衡。实验结果表明,该模块在降低计算开销的同时能够提升特征表示细化能力,为裂缝分割的精度提供了重要支撑。其计算流程如式(1)~(3)所示。

$$\mathbf{Y}_1^{C/4}, \mathbf{Y}_2^{C/4}, \mathbf{Y}_3^{C/4}, \mathbf{Y}_4^{C/4} = Split [RMSNorm(\mathbf{X}_{in}^C)] \quad (1)$$

$$\mathbf{VM}_i \mathbf{Y}_i^{C/4} = MSRVS(\mathbf{Y}_i^{C/4}) \quad i = 1, 2, 3, 4 \quad (2)$$

$$\mathbf{X}_{out}^C = Linear\{RMSNorm[Concat(\mathbf{VM}_i \mathbf{Y}_i^{C/4} | i = 1, 2, 3, 4)]\} \quad (3)$$

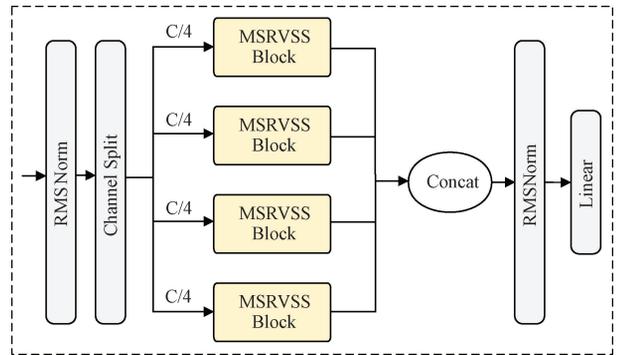


图 2 PLVM 模块结构

Fig.2 The structure of the PLVM block

2.3 多尺度残差视觉状态空间模块

MSRVSS 作为模型提取深层语义信息的核心模块,融合多尺度特征提取、残差学习与状态空间建模策略,以增强裂缝检测中的细节捕捉与全局特征建模能力。如图 3(a)所示,该模块由多尺度前馈网络(multi-scale feed-forward network, MSFFN)和二维选择性扫描(2D selective scan, SS2D)模块^[17]等子模块构成,同时引入残差学习机制以增强信息传递与梯度流动。特别地,MSRVSS 在残差连接部分引入可学习调节因子 Scale,其通过参数化实现,可对残差信号进行缩放,从而更精准地控制残差信息的融合程度。该机制与传统残差连接不同,既避免了直接叠加可能导致的特征冲突问题,又提供了额外的灵活性。MSRVSS 模块的计算流程表述如下:

$$\mathbf{X}_{out_1} = SiLU(Linear(\mathbf{X}_{in})) \quad (4)$$

$$\mathbf{X}_{out_2} = DWConv(Linear(\mathbf{X}_{in})) \quad (5)$$

$$\mathbf{X}'_{out_2} = LN(SS2D(SiLU(\mathbf{X}_{out_2}))) \quad (6)$$

$$\mathbf{X}_{out} = \mathbf{X}_{out_1} \otimes \mathbf{X}'_{out_2} \quad (7)$$

$$\mathbf{Out} = Scale \cdot \mathbf{X}_{in} \oplus MSFFN(\mathbf{X}_{out}) \quad (8)$$

其中,Linear 表示线性层运算,SiLU 为 SiLU 激活函数,(depthwise separable convolution,DWConv)指代深度可分离卷积,(layer normalization,LN)表示层归一化操作, \otimes 代表逐元素乘法, \oplus 表示逐元素加法,Scale 为残差连

接调节因子。

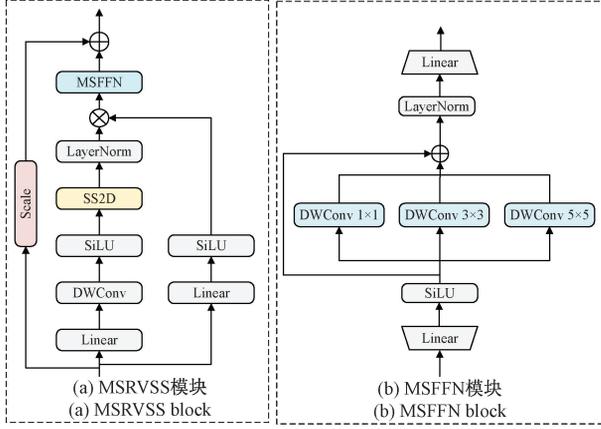


图 3 MSRVSS 与 MSFFN 结构

Fig. 3 The structure of MSRVSS and MSFFN

为提升模型多尺度特征提取能力, MSFFN 模块(如图 3(b)所示)实现了局部与全局上下文信息的有效捕获。具体而言, 给定输入特征张量 $\mathbf{X}_{in} \in R^{H \times W \times C}$, MSFFN 首先通过全连接层将输入通道扩展至原四倍, 经 SiLU 函数激活后获得 $\mathbf{X}'_{in} \in R^{H \times W \times 4C}$, 该操作可更充分地捕捉输入特征的潜在表征。随后, \mathbf{X}'_{in} 经多尺度卷积模块处理, 通过不同核尺寸的深度可分离卷积捕获多尺度空间特征。最终输出 \mathbf{X}_{out} 经层归一化处理, 并通过输出线性层恢复特征维度, 得到最终输出 $\mathbf{Out} \in R^{H \times W \times C}$ 。该过程定义如下:

$$\mathbf{X}'_{in} = \text{SiLU}(\text{Linear}(\mathbf{X}_{in})) \quad (9)$$

$$\mathbf{X}_{out} = \sum_{ks \in KS} \text{DWConv}_{ks}(\mathbf{X}'_{in}) + \mathbf{X}'_{in} \quad (10)$$

$$\mathbf{Out} = \text{Linear}(\text{LN}(\mathbf{X}_{out})) \quad (11)$$

其中, DWConv_{ks} 表示核尺寸为 KS 的深度可分离卷积, KS 定义为包含 $\{1 \times 1, 3 \times 3, 5 \times 5\}$ 的并行卷积核集合。

MSFFN 的引入使 MSRVSS 模块增强了自适应能力并优化多尺度特征提取, 从而显著提升不同尺寸与形态裂缝区域的检测性能。

2.4 增强语义与细节融合模块

跳跃连接的核心功能在于将编码器提取的多阶段特征直接传递至解码器, 从而补偿因下采样导致的空间信息损失。然而, 不同阶段特征在语义表征与分辨率上存在显著差异, 直接融合可能导致信息冗余或不足。为增强模型对裂缝区域特征的捕获能力, E-SDI 模块在原始 (semantics and detail infusion, SDI) 模块^[18]基础上进行改进, 具体引入文献^[19]提出的空间注意力桥 (spatial attention bridge, SAB) 与通道注意力桥 (channel attention bridge, CAB) 机制。其中, SAB 负责捕获裂缝区域多尺度空间信息, CAB 专注于建模通道间全局依赖关系, 二者协同作用可显著提升跳跃连接中的特征传递效能, 进而增强解码器对裂缝区域细节的还原能力。

如图 4 所示, 针对编码器生成的五个阶段特征图 $f_i^0 \in$

$R^{\frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}} \times C_i}$, $0 \leq i \leq 4$, E-SDI 模块分别通过 SAB 与 CAB 计算空间注意力得分与通道注意力得分。

其中, f_i^0 表示第 i 层特征图, C_i 代表从 Stage0 ~ Stage4 的通道数, 本文设置为 $[8, 16, 24, 32, 48]$ 。该过程的数学表达如式(12)所示。

$$f_i^1 = \phi_{\text{CAB}}(\phi_{\text{SAB}}(f_i^0)) \quad (12)$$

其中, ϕ_{SAB} 与 ϕ_{CAB} 分别表示 SAB 与 CAB 的计算函数, f_i^1 为第 i 层处理后的特征图。随后, 通过 1×1 卷积将 f_i^1 的通道数统一调整至维度 C (超参数 C 设置为 8), 生成特征图 $f_i^2 \in R^{\frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}} \times C}$, $0 \leq i \leq 4$ 。将 f_i^2 作为第 i 层的目标参考特征, 通过调整各 j 层特征图分辨率至与 f_i^2 匹配, 生成融合特征图。

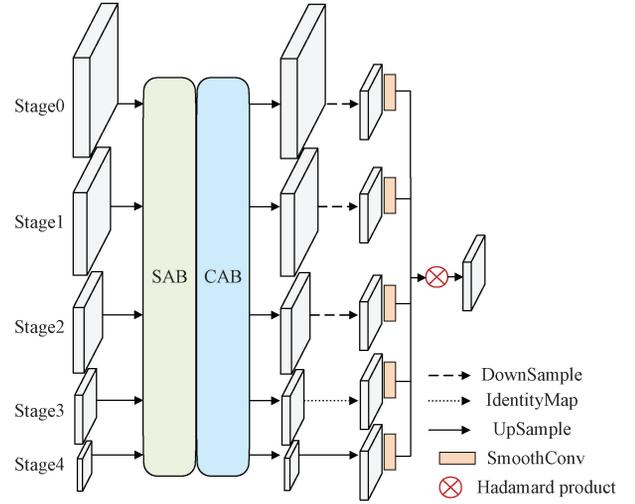


图 4 E-SDI 模块结构

Fig. 4 The structure of the E-SDI module

2.5 损失函数

裂缝检测任务面临严重的类别不平衡问题, 其中非裂缝像素数量远多于裂缝像素。当采用标准二元交叉熵损失函数训练模型时, 其性能在高度类别失衡场景下可能显著下降导致模型将更倾向于分割背景像素。Dice 损失在正负样本极度失衡的图像分割任务中具有优势, 但单独使用会导致网络训练过程中收敛困难。为使模型更聚焦于难检测的裂缝区域, 同时规避单一使用 Dice 损失引发的训练不稳定性问题, 本文采用二元交叉熵 (binary cross-entropy, BCE) 与 Dice 损失的复合损失函数, 具体计算方法如下:

$$L_{\text{BCE}}(\mathbf{Y}, \hat{\mathbf{Y}}) = -\mathbf{Y} \log(\hat{\mathbf{Y}}) - (1 - \mathbf{Y}) \log(1 - \hat{\mathbf{Y}}) \quad (13)$$

$$L_{\text{Dice}}(\mathbf{Y}, \hat{\mathbf{Y}}) = 1 - \frac{2\mathbf{Y}\hat{\mathbf{Y}} + \epsilon}{\mathbf{Y} + \hat{\mathbf{Y}} + \epsilon} \quad (14)$$

$$Loss = \alpha L_{\text{BCE}} + \beta L_{\text{Dice}} \quad (15)$$

其中, \mathbf{Y} 与 $\hat{\mathbf{Y}}$ 分别为真实标注掩码与预测掩码, ϵ 用于避免 $\mathbf{Y} + \hat{\mathbf{Y}} = 0$ 时的未定义值, α 和 β 为超参数, 在本文中均

设为 1。

3 实验结果与分析

3.1 数据集

由于裂缝的形成机制与特征因建筑材料差异而显著不同(如混凝土裂缝与钢结构裂缝在形态、成因及检测难度上存在本质区别),为全面评估本文方法在不同类型裂缝检测中的有效性,本研究选取多个代表性数据集进行实验验证。包括 DeepCrack^[20]、Crack500^[21] 和 SteelCrack^[22] 数据集,分别涵盖混凝土裂缝与钢结构裂缝的典型样本。

1) DeepCrack: 共包含 537 幅图像,分辨率均为 544 pixel×384 pixel。该数据集多样性丰富,涵盖混凝土与沥青路面裂缝,同时包含不同宽度裂缝及复杂环境条件。

2) Crack500: 包含 1 896 幅训练图像、348 幅验证图像与 1 124 幅测试图像,分辨率均为 640 pixel×360 pixel。数据集内裂缝形态与宽度差异显著,背景中多含复杂线条与斑点特征,对检测算法提出更高挑战。

3) SteelCrack: 数据集图像直接采集自多个钢结构工程现场,提供精细人工标注,所有图像及标注分辨率均为 512 pixel×512 pixel。

实验中,首先对所有训练集图像进行预处理,包括随机水平与垂直翻转、随机旋转,并将 DeepCrack 与 Crack500 数据集裁剪至 384×384 固定尺寸以适配网络训练。具体参数配置如表 1 所示。

表 1 所采用的三个数据集详细信息

Table 1 Details of the three adopted datasets

数据集	尺寸	训练集	验证集	测试集
DeepCrack	384×384	300	50	187
Crack500	384×384	1 896	348	1 124
SteelCrack	512×512	3 300	525	530

3.2 实验细节

实验硬件环境配置为: Intel Xeon Platinum 8352V CPU 与单块 NVIDIA RTX 4090 GPU(24 GB 显存);软件环境包括 CUDA 12.2、Python 3.8.20 及 PyTorch 1.13.0。训练过程中采用 AdamW 优化器,初始学习率设为 0.006,权重衰减率为 0.01,一阶矩衰减率为 0.9、二阶矩衰减率为 0.999,并结合余弦退火学习率调度器进行动态调整。总训练轮次为 100 epochs,批次大小设置为 8。

3.3 评价指标

本文采用精确率、召回率、F1 分数以及 mIoU 作为模型性能评估指标,各指标计算公式如式(16)~(19)所示。

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

$$F1 = \frac{2PrRe}{Pr + Re} \quad (18)$$

$$mIoU = \frac{TP}{FN + TP + FP} \quad (19)$$

其中, TP 表示模型正确预测为正类的正样本数量, FP 表示模型错误预测为正类的负样本数量, FN 表示模型错误预测为负类的正样本数量, Pr 表示精确率, Re 表示召回率。

3.4 模型复杂性

为评估本研究方法在模型轻量化方面的有效性,选取具有代表性的前沿模型开展复杂度对比实验。对比模型涵盖当前裂缝分割任务的领先方法,包括基于卷积神经网络的 U-Net^[23]、SegNet^[24] 与 DeepLabv3+^[25], 同时纳入 EfficientNet^[26]、MobileNetv3^[27] 等轻量化模型,以及 DeepCrack^[28]、LinkCrack^[29]、HrSegNet-B48^[30] 和 CrackSegFormer^[31] 等裂缝检测专用模型。此外,考虑融合 Transformer 架构的 TransUNet^[32], 共计十类对比模型。该选型策略可从多维度综合评估本文模型在性能与效率上的优势。选用模型复杂度双核心指标: 表征计算负荷的浮点运算次数 FLOPs、反映存储需求的参数量(Params)及衡量推理速度的帧率(frames per second, FPS)。

所有模型均以形状为 512×512×3 的张量进行测试,结果如表 2 所示,本文方法在模型复杂度方面展现显著优势。具体而言,本方法 FLOPs(1.96 G)与 Params(0.13 M)均远低于其他对比模型。相较于 U-Net、DeepLabv3+ 等常规模型,其计算开销与存储需求显著更高。此外, MobileNetv3 等轻量化模型虽在 FLOPs(0.70 G)与 Params(1.07 M)上表现较优,但本方法参数量仍较其降低约 87.9%, 实现了超轻量化裂缝检测模型。后续对比实验发现, MobileNetv3 等轻量化模型的裂缝检测性能与本文方法存在较大差距。尽管本方法 FPS 相对较低,但其 FLOPs 与 Params 的显著缩减,在计算与存储开销方面均超越传统模型及轻量化模型,突显其轻量化优势。

表 2 模型复杂度对比

Table 2 Comparison of models complexity

方法	FLOPs/ G	Params/ M	FPS	Memory/ MB
U-Net	160.74	17.26	76	1 578
SegNet	170.26	29.48	69	1 699
DeepCrack	80.49	14.72	143	876
DeepLabv3+	136.05	37.05	81	2 155
TransUNet	129.25	93.23	33	3 668
EfficientNet	4.74	6.98	86	1 256
MobileNetv3	0.70	1.07	144	318
LinkCrack	61.02	3.42	64	1 144
HrSegNet-B48	9.08	5.43	125	464
CrackSegFormer	6.76	3.71	101	643
Proposed	1.96	0.13	29	496

3.5 对比实验

为更全面验证本文方法优越性,体现其在保持超轻量化特性的同时实现高精度与高效率的最优平衡,在 DeepCrack、Crack500 及 SteelCrack 数据集上分别评估对比模型与本文方法的性能。

DeepCrack 实验结果如表 3 所示,本文方法以 87.73% 的精确率、86.79% 的 F1 分数及 87.85% 的平均交并比取得最优结果。值得注意的是,基于 Transformer 架构的 TransUNet 参数量高达 93.23 M,计算量达 129.25 G,其复杂度显著高于本文方法。

表 3 DeepCrack 数据集对比结果

Table 3 Comparative results on DeepCrack dataset %

方法	Pr	Re	F1	mIoU
U-Net	85.38	84.75	85.06	86.45
SegNet	81.73	82.16	81.94	84.04
DeepCrack	86.81	86.16	86.49	87.60
DeepLabv3+	86.23	79.81	82.90	84.79
TransUNet	84.17	87.44	85.77	87.01
EfficientNet	71.20	83.25	76.76	80.21
MobileNetv3	70.88	83.21	76.55	80.06
LinkCrack	86.89	86.52	86.71	87.77
HrSegNet-B48	84.59	88.83	86.66	87.72
CrackSegFormer	82.43	84.76	83.58	85.28
Proposed	87.73	85.87	86.79	87.85

Crack500 实验结果如表 4 所示,本文方法以 74.25% 的 F1 与 77.92% 的 mIoU 领先其他方法。HrSegNet-B48 以 73.93% 的 F1 与 77.68% 的 mIoU 紧随其后,而 CrackSegFormer 性能最低 (F1 = 69.98%, mIoU = 74.96%)。特别指出,本文方法仅使用 0.13 M 参数即达成上述优异结果,既验证了方法有效性,也凸显了其高效性与

表 4 Crack500 数据集对比结果

Table 4 Comparative results on Crack500 dataset %

方法	Pr	Re	F1	mIoU
U-Net	72.90	70.22	71.54	76.09
SegNet	73.12	69.64	71.34	75.97
DeepCrack	74.72	71.77	73.22	77.23
DeepLabv3+	73.04	73.87	73.45	77.34
TransUNet	70.18	74.42	72.24	76.47
EfficientNet	71.73	72.09	71.91	76.30
MobileNetv3	69.20	74.28	71.65	76.06
LinkCrack	74.40	72.80	73.59	77.47
HrSegNet-B48	73.74	74.13	73.93	77.68
CrackSegFormer	67.67	72.46	69.98	74.96
Proposed	74.89	73.63	74.25	77.92

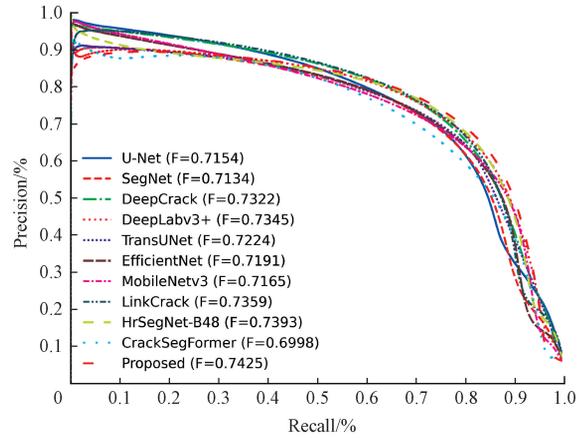


图 5 Crack500 数据集的 PR 曲线

Fig. 5 PR curves on Crack500 dataset

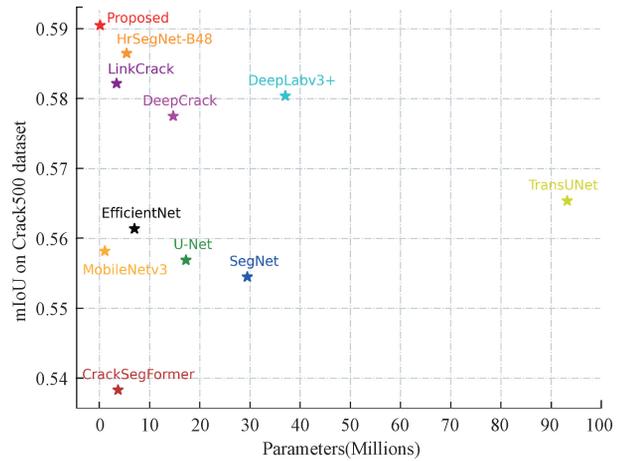


图 6 Crack500 数据集上模型参数量与性能关系的可视化

Fig. 6 Visualization of the relationship between model parameters and performance on the Crack500 dataset

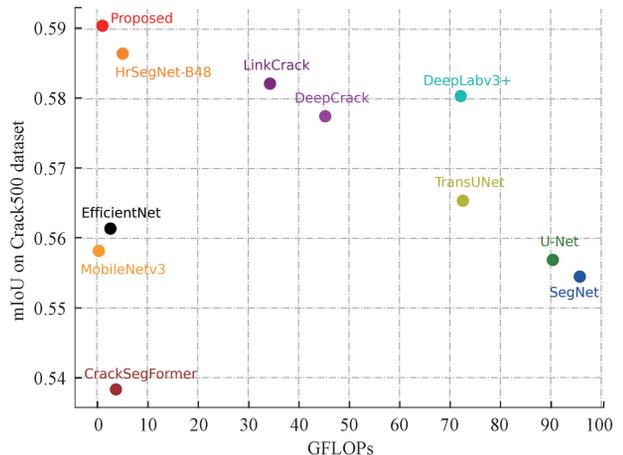


图 7 Crack500 数据集上模型计算量与性能关系的可视化

Fig. 7 Visualization of the relationship between model computational cost and performance on the Crack500 dataset

强泛化能力。如图 5 所示,精确率-召回率 (precision-recall, PR) 曲线进一步表明本文方法在保持高召回率的同时,精确率显著优于其他方法。如图 6 和 7 所示,本文方法在参数量、计算复杂度与分割性能间实现了最佳平衡。

SteelCrack 实验结果如表 5 所示,相较于混凝土裂缝数据集,SteelCrack 数据集因裂缝形态更复杂、背景噪声更高,对模型鲁棒性提出更大挑战。本文方法取得 Pr = 90.91%、Re = 86.14%、F1 = 88.46% 及 mIoU = 89.48%。虽召回率略低,但其 90.91% 的精确率优势显著。相较于 DeepLabv3+ (F1 = 89.95%, mIoU = 90.72%),本文方法凭借 0.13 M 参数量(降低 99.6%)与 1.96 G 计算量(降低 98.6%),在 SteelCrack 数据集上达到可比拟的性能表现,显著体现了超轻量化优势。尽管所提方法在精确率上具备显著优势(90.91%),但其超轻量化设计使得特征提取能力、感受野范围及噪声鲁棒性存在一定局限,受限于召回率偏低(86.14%),致使其整体性能较略低于 DeepLabv3+。

表 5 SteelCrack 数据集对比结果

Table 5 Comparative results on SteelCrack dataset %

方法	Pr	Re	F1	mIoU
U-Net	90.30	89.12	89.71	90.51
SegNet	88.00	84.82	86.38	87.81
DeepCrack	90.42	88.95	89.68	90.49
DeepLabv3+	90.88	89.05	89.95	90.72
TransUNet	88.24	88.20	88.22	89.28
EfficientNet	71.53	82.71	76.71	80.73
MobileNetv3	73.01	84.17	78.20	81.74
LinkCrack	90.38	89.20	89.78	90.58
HrSegNet-B48	89.65	87.37	88.50	89.51
CrackSegFormer	85.90	85.51	85.70	87.27
Proposed	90.91	86.14	88.46	89.48

通过引入 95% Hausdorff 距离(数值越小精度越高)进一步评价模型的分割效果,表 6 显示了所有对比方法在 3 个数据集的表现。该指标通过 95% 分位数降低异常值影响,量化边界对齐精度。本文方法性能均衡且居前列,展现跨数据集泛化能力。相较 CrackSegFormer 在 Crack500 数据集的波动和 MobileNetv3 的场景依赖性,其稳定性与复杂度平衡更优,得益于多尺度特征融合设计,尤其在裂纹复杂的 Crack500 数据集表现突出。

通过十类模型在 3 个数据集上的对比实验,本文方法综合性能显著领先。图 8 所示为 3 个数据集上的裂缝检测结果可视化对比,从左至右依次为:DeepCrack、Crack500 和 SteelCrack 数据集,每个数据集选取两个样本进行展示,进一步验证了本文方法的有效性。

3.6 消融实验

本节通过一系列消融实验评估所提 PLVM 与 E-SDI

表 6 3 个数据集上的 95% Hausdorff 距离对比

Table 6 Comparison of 95% Hausdorff distances on three datasets

方法	DeepCrack	Crack500	SteelCrack
U-Net	24.24	56.98	34.21
SegNet	43.99	54.48	35.65
DeepCrack	16.39	48.76	30.19
DeepLabv3+	20.53	38.87	19.53
TransUNet	29.47	72.49	32.68
EfficientNet	19.44	47.67	32.89
MobileNetv3	17.36	46.44	26.06
LinkCrack	14.44	43.26	27.19
HrSegNet-B48	13.95	47.85	28.06
CrackSegFormer	21.14	76.86	37.50
Proposed	14.08	38.32	31.18

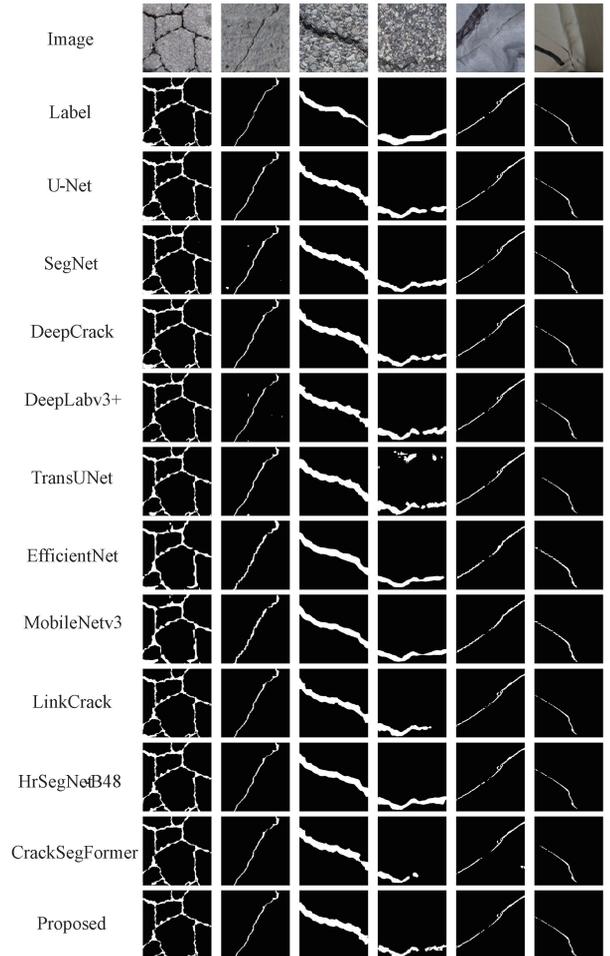


图 8 不同数据集分割结果示意图

Fig. 8 Illustration of segmentation results on different datasets

模块在裂缝检测中的有效性。如表 7 所示,基线模型(Base)仅使用标准卷积模块(Conv Block),在 DeepCrack

数据集上获得 $F1=84.92\%$ 、 $mIoU=86.33\%$ 的性能。引入 PLVM 模块后(Base+PLVM),模型性能显著提升, $F1$ 增至 85.73% , $mIoU$ 达到 86.98% ,证明 PLVM 模块增强特征提取能力的有效性。随后引入 E-SDI 模块(Base+E-SDI),通过空间与通道注意力机制优化跳跃连接特征传

递。最终组合 PLVM 与 E-SDI 模块(Proposed),在 3 个数据集上均取得最优性能,例如 DeepCrack 数据集 $F1$ 达到 86.79% , $mIoU$ 达 87.85% ,验证两模块的互补性与协同效应。实验表明,PLVM 与 E-SDI 的协同作用可增强模型特征表征能力,显著提升裂缝检测精度与鲁棒性。

表 7 基于 3 个数据集的消融实验

Table 7 Ablation experiments on three datasets

实验设置	DeepCrack				Crack500				SteelCrack			
	Pr	Re	F1	mIoU	Pr	Re	F1	mIoU	Pr	Re	F1	mIoU
Base	84.10	85.76	84.92	86.33	74.25	71.09	72.64	76.84	87.89	87.45	87.67	88.84
Base+PLVM	85.92	85.54	85.73	86.98	75.08	73.14	74.10	77.82	86.98	86.83	86.90	88.22
Base+E-SDI	85.12	85.15	85.13	86.51	74.21	71.33	72.74	76.90	89.08	87.48	88.27	89.32
Proposed	87.73	85.87	86.79	87.85	74.89	73.63	74.25	77.92	90.91	86.14	88.46	89.48

如图 9 所示,分别对 PLVM 与 E-SDI 模块进行裂缝检测结果可视化。绿色实线框表示模型添加 PLVM 模块后,其对裂缝边缘和背景噪声的区分能力,且裂缝边缘更加平

滑;红色虚线框表示融合 E-SDI 模块后的模型有效抑制背景噪声误检的积极作用。

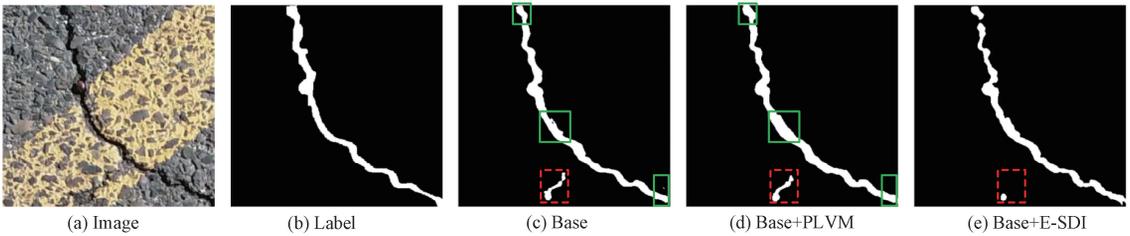


图 9 PLVM 和 E-SDI 模块的效果可视化

Fig. 9 Visualisation of the effects of PLVM and E-SDI modules

为评估 MSRVS 设计的有效性,在 DeepCrack 数据集上对比不同配置的性能(如表 8 所示)。以 VMamba 提出的 Vanilla VSS 模块为基准(仅含基础视觉状态空间模块),其 $F1$ 为 85.82% , $mIoU$ 为 87.05% 。引入残差连接调节因子 Scale(Vanilla VSS+Scale)后,模型 $F1$ 提升至 86.30% , $mIoU$ 增至 87.45% ,表明动态残差调节可优化特征融合效果。添加多尺度卷积优化的 MSFFN 模块(Vanilla VSS+MSFFN)后, $F1$ 与 $mIoU$ 分别提升至 86.50% 与 87.61% 。最终,集成 Scale 与 MSFFN 的完整 MSRVS 模块(Proposed)在 DeepCrack 数据集取得最优性能($F1=86.79\%$, $mIoU=87.85\%$)。

表 8 MSRVS 模块的有效性

Table 8 Effectiveness of the MSRVS block

实验设置	DeepCrack			
	Pr	Re	F1	mIoU
Vanilla VSS	86.27	85.36	85.82	87.05
Vanilla VSS+Scale	87.71	84.93	86.30	87.45
Vanilla VSS+MSFFN	87.50	85.52	86.50	87.61
Proposed	87.73	85.87	86.79	87.85

如表 9 所示,融合 SAB 与 CAB 的 E-SDI 模块(Proposed)相较于单独使用 SAB+CAB(SCAB)或 SDI 模块, $mIoU$ 分别提升 1.83% 与 0.7% ,有力佐证 E-SDI 模块在跳跃连接特征传递中的关键作用。

表 9 E-SDI 模块的有效性

Table 9 Effectiveness of the E-SDI block

实验设置	DeepCrack			
	Pr	Re	F1	mIoU
Base+PLVM+SCAB	82.58	86.58	84.53	86.02
Base+PLVM+SDI	86.92	84.96	85.93	87.15
Proposed	87.73	85.87	86.79	87.85

为评估 PLVM 模块中不同并行连接策略的影响,针对 3 种设置开展消融实验。假设输入 PLVM 模块的特征图通道数为 C ;设置 A 采用非并行标准连接(单 MSRVS 模块直接处理全通道 C);设置 B 采用双路并行结构(每个 MSRVS 模块输入通道数为 $C/2$);设置 C 采用四路并行结构(每个 MSRVS 模块输入通道数为 $C/4$)。如表 10 所示,相较于设置 A,设置 B 与设置 C 的参数量分别显著降低 66.23% 与 83.12% ,且总体浮点运算次数(FLOPs)仅小

幅下降。性能方面,设置 B 以更少参数量实现与设置 A 相当的结果,而设置 C 在参数量与计算量最小的情况下,以 86.79% 的 F1 与 87.85% 的 mIoU 取得最优分割性能。实验结果表明,四路并行结构(设置 C)既保持了计算高效性,又实现了精度最优平衡,故本文采用设置 C 作为 PLVM 模块的最终结构设计。

表 10 PLVM 中不同并行连接结构的消融实验
Table 10 Ablation experiments in PLVM with different parallel connections

设置	DeepCrack					
	Params	FLOPs	Pr	Re	F1	mIoU
A	0.77	1.47	86.73	86.81	86.77	87.83
B	0.26	1.23	87.84	84.85	86.32	87.47
C	0.13	1.10	87.73	85.87	86.79	87.85

本文模型各层通道数的不同设置可能对性能产生影响,为此针对四类通道数设置开展消融实验。具体包括 ABCD 四类设置,分别为:[8,16,24,32,48,64]、[8,16,32,64,128,256]、[16,32,64,128,256,512]以及[32,64,128,256,512,1024],实验结果如表 11 所示。

表 11 不同通道数组合的消融实验

Table 11 Ablation experiments with different channel number combinations

设置	DeepCrack					
	Params	FLOPs	Pr	Re	F1	mIoU
A	0.13	1.10	87.73	85.87	86.79	87.85
B	0.56	1.80	87.86	86.69	87.27	88.24
C	2.10	6.63	87.89	86.01	86.94	87.97
D	8.12	25.49	85.40	84.87	85.14	86.51

参数量与 FLOPs 随通道数增加呈阶梯式增长,分别从最低 0.13 M/1.10 G 至最高 8.12 M/25.49 G。中等通道数设置[8,16,32,64,128,256]以 F1=87.27%、mIoU=88.24%取得最优性能,表明适度通道扩展可优化特征表征。而最大通道数设置[32,64,128,256,512,1024]虽然计算复杂度最高,但性能表现最差(F1=85.14%、mIoU=86.51%),揭示过度扩展通道易引发过拟合与效率下降问题。值得注意的是,最小通道数设置[8,16,24,32,48,64]仍保持竞争力(F1=86.79%、mIoU=87.85%),在模型复杂度与性能间实现最优平衡,故本文选用该设置作为最终通道数方案。

4 结 论

本文提出一种基于视觉 Mamba 架构的超轻量化裂缝检测模型,通过引入 PLVM 模块与 MSRVSS 模块实现裂缝特征的深度表征与多尺度信息高效捕获,同时结合

E-SDI 模块优化跳跃连接,显著提升解码器对裂缝区域的细粒度还原能力。在损失函数设计上,BCE 与 Dice 损失的组合有效缓解了类别不平衡问题并优化模型性能。实验结果表明,该模型在参数量仅为 0.13 M 的前提下,于多个公开数据集上均表现优异,兼具高检测精度与轻量化优势,为资源受限场景下的裂缝检测提供了高效解决方案。

尽管本文方法在裂缝检测任务中展现出显著的性能与效率优势,但仍存在若干不足。首先,针对存在大量复杂背景干扰或极细小裂缝区域的检测能力仍需提升;其次,对于不同材质或环境条件下的裂缝,本方法未能实现最高检测精度(尤其在具有高反光金属表面的 SteelCrack 数据集上),表明其在材质自适应特征提取方面存在局限。未来研究应着重于提升视觉 Mamba 架构在复杂环境下的泛化能力,探索与其他深度学习技术的融合路径,并进一步探索更高效的轻量化模型设计与边缘端部署优化方案,以推动实时工程化应用。

参考文献

- [1] 邓露,褚鹤鸣,龙砺芝,等. 基于深度学习的土木基础设施裂缝检测综述[J]. 中国公路学报,2023,36(2): 1-21.
DENG L, CHU H H, LONG L ZH, et al. Review of deep learning-based crack detection for civil infrastructures[J]. China Journal of Highway and Transport, 2023, 36(2): 1-21.
- [2] 贾晓芬,江再亮,赵佰亭. 裂缝小目标缺陷的轻量化检测方法[J]. 湖南大学学报(自然科学版),2024,51(6): 52-62.
JIA X F, JIANG Z L, ZHAO B T. Lightweight detection method for small crack target defects[J]. Journal of Hunan University (Natural Sciences), 2024, 51(6): 52-62.
- [3] 周建新,张媛,贾梓涵,等. 改进 YOLOv8 的路面缺陷检测算法[J]. 电子测量技术,2024,47(19):146-154.
ZHOU J X, ZHANG Y, JIA Z H, et al. The pavement defect detection algorithm of YOLOv8 was improved[J]. Electronic Measurement Technology, 2024, 47(19): 146-154.
- [4] 金宇锋,陶重霖. 基于 Transformer 的融合信息增强 3D 目标检测算法[J]. 仪器仪表学报,2023,44(12): 297-306.
JIN Y F, TAO CH B. Fusion information enhanced method based on transformer for 3D object detection[J]. Chinese Journal of Scientific Instrument, 2023, 44(12): 297-306.
- [5] 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. Pavement crack segmentation model based on pixel intensity order transform and UNetFormer [J]. *Electronic Measurement Technology*, 2024, 47(11): 151-159.
- [6] GU A, DAO T. Mamba: Linear-time sequence modeling with selective state spaces [J]. *ArXiv preprint arXiv: 2312.00752*, 2023.
- [7] 彭圆圆, 黎浩洋, 李文, 等. LTDA-Mamba: 基于 CNN-Mamba 混合网络的视网膜血管分割算法[J]. *光学学报*, 2025, 45(7): 163-174.
- PENG Y Y, LI H Y, LI W, et al. LTDA-Mamba: Retinal vessel segmentation based on a hybrid CNN-Mamba network [J]. *Acta Optica Sinica*, 2025, 45(7): 163-174.
- [8] 厉芳婷, 张过, 周方栋, 等. 基于双流并行全向扫描 Mamba 的遥感影像建筑物变化检测[J]. *武汉大学学报(信息科学版)*, 2024, 49(9): 1574-1585.
- LI F T, ZHANG G, ZHOU F D, et al. Building change detection based on dual-stream parallel omnidirectional scan mamba network [J]. *Geomatics and Information Science of Wuhan University*, 2024, 49(9): 1574-1585.
- [9] XU J Y, XU CH, ZHAO J P, et al. Mamba4PASS: Vision mamba for panoramic semantic segmentation [J]. *Displays*, 2025, 89: 103058.
- [10] YU Y B, ZHANG Y G, YU J Y, et al. Lightweight decoder u-net crack segmentation network based on depthwise separable convolution [J]. *Multimedia Systems*, 2024, 30(5): 295.
- [11] ZHOU Q, QU ZH, JU F R. A lightweight network for crack detection with split exchange convolution and multi-scale features fusion [J]. *IEEE Transactions on Intelligent Vehicles*, 2022, 8(3): 2296-2306.
- [12] HOU Y, SHI H Y, CHEN N, et al. Vision image monitoring on transportation infrastructures: A lightweight transfer learning approach [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 24(11): 12888-12899.
- [13] ZHANG L, YANG F, ZHANG Y D, et al. Road crack detection using deep convolutional neural network [C]. 2016 IEEE International Conference on Image Processing (ICIP). IEEE, 2016: 3708-3712.
- [14] KANG D H, CHA Y J. Efficient attention-based deep encoder and decoder for automatic crack segmentation [J]. *Structural Health Monitoring*, 2022, 21 (5): 2190-2205.
- [15] KÖNIG J, JENKINS M D, BARRIE P, et al. A convolutional neural network for pavement surface crack segmentation using residual connections and attention gating [C]. 2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019: 1460-1464.
- [16] DONG ZH M, WANG J J, CUI B, et al. Patch-based weakly supervised semantic segmentation network for crack detection [J]. *Construction and Building Materials*, 2020, 258: 120291.
- [17] LIU Y, TIAN Y J, ZHAO Y ZH, et al. Vmamba: Visual state space model [J]. *Advances in Neural Information Processing Systems*, 2024, 37: 103031-103063.
- [18] PENG Y P, CHEN D Z, SONKA M. U-net v2: Rethinking the skip connections of u-net for medical image segmentation [C]. 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI). IEEE, 2025: 1-5.
- [19] RUAN J CH, XIANG S CH, XIE M Y, et al. Malunet: A multi-attention and light-weight unet for skin lesion segmentation [C]. 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2022: 1150-1156.
- [20] LIU Y H, YAO J, LU X H, et al. DeepCrack: A deep hierarchical feature learning architecture for crack segmentation [J]. *Neurocomputing*, 2019, 338: 139-153.
- [21] YANG F, ZHANG L, YU S J, et al. Feature pyramid and hierarchical boosting network for pavement crack detection [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2019, 21 (4): 1525-1535.
- [22] HE Z L, CHEN W, ZHANG J, et al. Crack segmentation on steel structures using boundary guidance model [J]. *Automation in Construction*, 2024, 162: 105354.
- [23] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]. *Medical Image Computing and Computer-assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer International Publishing, 2015: 234-241.
- [24] BADRINARAYANAN V, KENDALL A, CIPOLLA R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481-2495.
- [25] 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.
- [26] TAN M X, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C]. International Conference on Machine Learning. PMLR, 2019: 6105-6114.
- [27] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3 [C]. IEEE/CVF International Conference on Computer Vision, 2019: 1314-1324.
- [28] ZOU Q, ZHANG ZH, LI Q Q, et al. Deepcrack: Learning hierarchical convolutional features for crack detection [J]. IEEE Transactions on Image Processing, 2018, 28(3): 1498-1512.
- [29] LIAO J H, YUE Y H, ZHANG D J, et al. Automatic tunnel crack inspection using an efficient mobile imaging module and a lightweight CNN[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(9): 15190-15203.
- [30] LI Y SH, MA R G, LIU H, et al. Real-time high-resolution neural network with semantic guidance for crack segmentation[J]. Automation in Construction, 2023, 156: 105112.
- [31] 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.
- [32] CHEN J N, LU Y Y, YU Q H, et al. Transunet: Transformers make strong encoders for medical image segmentation [J]. ArXiv preprint arXiv: 2102.04306, 2021.

作者简介

成荣, 硕士研究生, 主要研究方向为计算机视觉和深度学习。

E-mail: 323085406201@stu.suse.edu.cn

朱文忠(通信作者), 硕士, 教授, 主要研究方向为深度学习、物联网技术及应用、最优控制等。

E-mail: zwz@suse.edu.cn

王文, 硕士研究生, 主要研究方向为时序数据预测和深度学习。