



DOI:10.12404/j.issn.1671-1815.2405207

引用格式:张学辉,李笑航,田学昭,等.基于轻量化多尺度动态融合卷积模块的裂缝分割方法[J].科学技术与工程,2025,25(18):7719-7728.

Zhang Xuehui, Li Xiaohang, Tian Xuezhao, et al. Segmentation method based on light measurement multi-scale dynamic fusion module[J]. Science Technology and Engineering, 2025, 25(18): 7719-7728.

建筑科学

基于轻量化多尺度动态融合卷积模块的 裂缝分割方法

张学辉^{1,2}, 李笑航^{1*}, 田学昭³, 安军海^{1,2}, 赵双双¹

(1. 河北科技大学建筑工程学院, 石家庄 050018; 2. 河北省岩土与结构体系防灾减灾技术创新中心(筹), 石家庄 050018; 3. 中国二十二冶集团有限公司, 石家庄 050000)

摘要 裂缝检测对维护建筑物结构安全至关重要,近年来,基于深度学习的卷积神经网络为裂缝检测提供了新的解决方案。但这是以巨大的计算资源为代价的,因此在实际应用中存在实时性差、检测效率低的问题。针对此问题,提出了一种基于U-Net架构的轻量化多尺度动态融合卷积模块(multi-scale dynamic fusion convolution, MSFC),以提高裂缝分割的效率。为验证所提方法的有效性,构建了一个包含2 045张裂缝图像的数据集Crack2045,并在该数据集上进行了实验。实验结果表明:与原始U-Net模型相比,采用MSFC模块的模型在保持准确率不下降的情况下减少了78.51%的参数数量和63.75%的计算量。同时,MSFC模块具有一定的泛化性,能够无缝集成到不同的语义分割模型中。研究结果不仅为裂缝检测提供了一种高效的深度学习方法,也为资源受限环境下的模型部署提供了新的可能性。

关键词 深度学习; 裂缝分割; U-Net; 轻量化模型

中图分类号 TU17; 文献标志码 A

Segmentation Method Based on Light Measurement Multi-scale Dynamic Fusion Module

ZHANG Xue-hui^{1,2}, LI Xiao-hang^{1*}, TIAN Xue-zhao³, AN Jun-hai^{1,2}, ZHAO Shuang-shuang¹

(1. School of Civil Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China;

2. Innovation Center of Disaster Prevention and Mitigation Technology for Geotechnical and Structural Systems of Hebei Province (Preparation), Shijiazhuang 050018, China; 3. China MCC22 Group Co., Ltd., Shijiazhuang 050000, China)

[Abstract] Crack detection is crucial to maintaining the structural safety of buildings. In recent years, convolutional neural networks based on deep learning have provided new solutions for crack detection. However, this comes at the cost of huge computing resources, so there are problems of poor real-time performance and low detection efficiency in practical applications. To address this problem, a lightweight MSFC (multi-scale dynamic fusion convolution module) based on the U-Net architecture was proposed to improve the efficiency of crack segmentation. To verify the effectiveness of the proposed method, a dataset Crack2045 containing 2 045 crack images was constructed and experiments were conducted on this dataset. The experimental results show that compared with the original U-Net model, the model using the MSFC module reduces 78.51% of the parameters and 63.75% of the computational complexity while maintaining the same accuracy. At the same time, the MSFC module has a certain degree of generalization and can be seamlessly integrated into different semantic segmentation models. This study not only provides an efficient deep learning method for crack detection, but also provides new possibilities for model deployment in resource-constrained environments.

[Keywords] deep learning; crack segmentation; U-Net; lightweight model

收稿日期: 2024-07-11 修订日期: 2025-03-19

基金项目: 河北省自然科学基金(E2024208050); 河北科技大学教育教学改革研究与实践项目(2022-YB02)

第一作者: 张学辉(1980—),男,汉族,河北河间人,博士,副教授。研究方向:智能建造。E-mail:zhxuehui@126.com。

*通信作者: 李笑航(2000—),男,汉族,河北邯郸人,硕士研究生。研究方向:智能建造。E-mail:15630462389@163.com。

裂缝是建筑物在使用过程中经常出现的问题,可能会引起严重的安全隐患。因此,及早发现裂缝,对于维护建筑物的结构安全至关重要^[1]。传统方法通常依赖于专业人员进行目视检查和手工测量,但存在主观差异的问题,并且手工检查需要大量的时间和人力资源,特别是对于大型建筑或大面积的检查任务。为了满足当前对裂缝检测的需求,研究者们基于图像处理技术提出了阈值分割、边缘检测、结合特定理论等图像分割方法^[2],以实现更快更准的裂缝检测,但这些方法仍需克服复杂场景和泛化等问题。随着深度学习技术在计算机视觉领域的广泛应用^[3],为解决图像分割问题提供了新的方法,并且在裂缝检测^[4]方面取得了显著进展。

基于深度学习技术的像素级裂缝检测最广泛使用的架构主要为 SegNet^[5]、Deeplab^[6]、FCN (fully convolutional networks)^[7] 和 UNet^[8] 等,几乎所有的裂缝分割架构都是基于这些架构开发的。大量的土木工程专业的研究者对这些模型进行了优化和改进,例如,邓砚学等^[9] 针对 SegNet 网络难以精确分割沥青路面微小病害的问题,使用残差网络 (residual network, ResNet) 作为主干网络,并加入空洞卷积层,提高网络对细小病害的识别性能。朱俊彬等^[10] 提出了一种基于 VGG-16 (visual geometry group network with 16 layers) 和深度特征融合的全卷积神经网络用于裂缝检测,通过侧边输出和 Focal Loss 损失函数提高了模型的抗干扰性和收敛速度。张修杰等^[11] 通过在 DeepLabv3+ 模型中嵌入注意力机制并改进主干网络,提出了优化模型 DeepLabv3+ (N-S) 用于混凝土梁裂缝分割及特征量化。然而,这样的成功在很大程度上依赖于密集的计算,这给在资源受限环境中有效部署带来了挑战。因此,研究者们已经探索了各种模型压缩策略和网络架构设计来提高网络效率。

模型压缩策略包括网络剪枝、权重量化和知识蒸馏。具体来说,网络剪枝^[12] 是通过剪除已有模型中不重要的神经元连接,使其得更小。权重量化^[13] 主要关注将网络权重从浮点类型转换为整数类型以节省计算资源。知识蒸馏^[14] 在训练有素的大型教师网络的指导下生成小型学生网络。这些压缩技术的共同点是,它们被视为后处理步骤,因此它们的性能通常受给定初始模型的限制。与此同时,这些方法在实现高压缩率时会导致准确性急剧下降。

在网络设计中,U-Net 作为一种经典的全卷积神经网络结构,在图像分割任务中取得了重要的突

破。它采用了编码器-解码器结构,并利用跳跃连接有效地保留了多尺度的上下文信息,使得模型能够更好地捕捉细节和整体信息。然而,为了进一步提升裂缝检测效率和适用性,学者们尝试对 U-Net 进行改进。孟庆成等^[15] 采用 MobileNet_V2 结构对 U-Net 进行优化,以实现模型的轻量化,适用于资源受限的环境中。钟志峰等^[16] 通过减少 U-Net 的深度来降低模型参数量。这种方法可以简化网络结构,减少模型复杂度,提高模型的运行效率。朱立学等^[17] 提出了一种利用多尺度串联空洞卷积组合的方法,以增大感受野并加强对细节的敏感度,提高图像分割的准确性和效率。尽管这些轻量化的改进方法在特定网络设计和数据集上取得了成功,但它们缺乏泛化能力,难以适用于其他网络架构和任务。

卷积操作的过度使用是导致模型参数量庞大的主要原因。因此,开发一种高效的卷积模块,对于构建轻量且性能优越的网络至关重要。MobileNet^[18] 将标准卷积操作分为深度卷积和逐点卷积两个步骤,提出了深度可分离卷积模块 (depthwise separable convolutional module, DSC) 在减少参数量的同时加快了训练速度。TiedBlockConv^[19] 在相等的通道块上共享相同的卷积滤波器,以单个滤波器产生多个响应并提出了共享卷积模块 (tied convolutional module, Tied)。GhostNet^[20] 考虑了特征图之间的冗余,使用廉价操作如深度卷积来学习冗余特征并提出了幽灵卷积模块 (ghost convolutional module, Ghost)。李润龙等^[21] 通过引入分组卷积和 Channel-shuffle 技术,减少了网络结构的参数量并提高了网络运行效率。

以上的卷积模块和改进方法在一定程度上可以降低模型的参数量和计算量,但在裂缝图像分割等下游任务中会显著影响准确率。因此,现提出轻量化多尺度动态融合卷积模块 (multi-scale dynamic fusion convolution, MSFC)。MSFC 模块采用了多尺度深度可分离卷积,并结合注意力机制用于通道动态融合。以在不降低模型预测准确率的前提下实现模型轻量化,既满足工程应用中对实时性和计算效率的需求,又保证高精度分割。

1 基础网络架构

本文中采用 U-Net^[8] 作为基础网络架构,U-Net。模型的输入为原始灰度图像,输出为与原始输入等大的预测图像,是一种端到端的网络模型。如图 1 所示,输入数据先通过下采样路径,在下采样路径中先后经历卷积模块、最大池化,逐渐减少输入

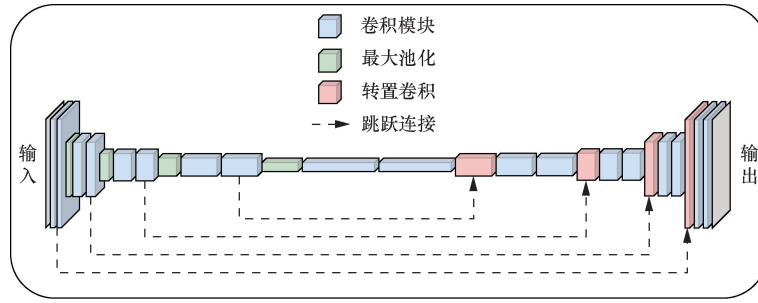


图1 U-Net 架构图

Fig. 1 U-Net architecture diagram

图像的尺寸,上采样路径先后通过转置卷积、卷积模块来自下采样的特征图,恢复图像尺寸,在上采样过程中,通过跳跃连接的方式进行特征融合,将上下采样中相同通道的网络层进行拼接,使模型在上采样过程中获得更多的信息,以达到更好的分割效果。

2 多尺度动态融合卷积模块

原始基础网络采用的常规卷积模块参数量较大,因此训练出的模型也较大。为了在不降低模型预测准确率的前提下实现轻量化,本文研究提出轻量化多尺度动态融合卷积模块(multi-scale dynamic fusion convolution, MSFC)。该模块采用深度可分离卷积(depthwise separable convolutional module, DSC)代替常规卷积。深度可分离卷积是一种成本较低的卷积操作,通过减少参数数量和计算量以提高模型的计算效率,但会导致准确度下降。为了弥补下降的准确度,使用了多种尺度的卷积核,从 3×3 到 11×11 的卷积核,如图2所示,其中较小的卷积核主要用于捕获图像中的细微特征,如裂缝的细小分支和边缘。较大的卷积核则有助于捕捉更广泛的上下文信息,为模型提供对整张图像的深入理解,这在处理大尺度裂缝和复杂纹理背景时尤为重要。设计原理基于金字塔模块^[22],该理论指出多尺度特征的整合对于提高图像分割的准确性至关重要。这种多尺度卷积核的组合显著提高了模型对裂缝特征的识别能力。

为了更好地融合不同尺度卷积核捕获到的特征,采用了如图3所示的SE(squeeze-and-excitation)机制^[23]。首先,对于每批次的多个通道特征图,通过全局平均池化计算其平均值,得到一个全局特征。然后,将全局特征输入到全连接层中,计算每个通道的权重,进而对特征图通道进行加权。加权后的特征图通道可以保留原特征图重要特征,并且可以抑制无用的特征。SE机制旨在解决卷积过程中不同通道权重分配不均匀的问题。在传统的卷积操作中,每个通道被默认视为同等重要,然而在实

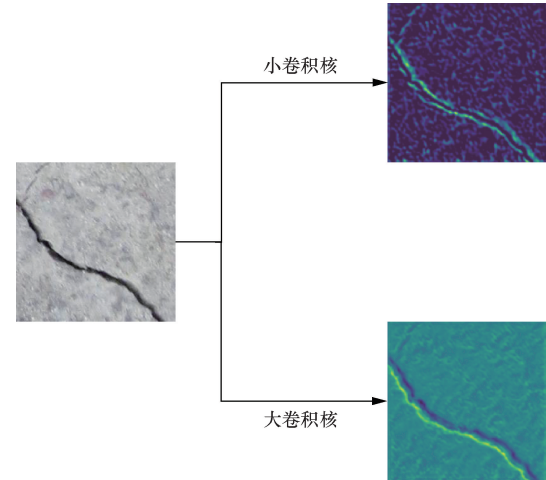


图2 不同大小卷积核的特征映射

Fig. 2 Feature maps of different sizes of convolutional kernels

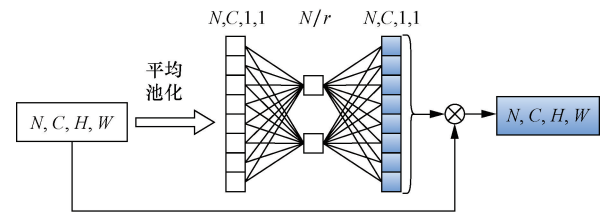


图3 SE 机制结构图

Fig. 3 Structure diagram of SE mechanism

际问题中,各通道的重要性存在差异。本文研究中在每种尺度下的卷积核后引入SE机制,通过学习不同通道的重要性来调整权重,然后将不同尺度下得到的特征图在通道维度上拼接,再次经过SE机制,进一步学习不同尺度中通道的重要性。

具体来说,MSFC主要分为3个步骤(图4):减少通道、空间多尺度编码器以及通道动态融合。以卷积神经网络第 K 层为例,将输入 X 记为 $X^k \in \mathbf{R}^{N \times C \times H \times W}$,其中 N, C, H, W 分别代表照片数量、照片通道数、照片高度及照片宽度。以下将分别介绍每一步的算式。

步骤1 减少通道。

$$X_d^k = W_d^k \otimes X^k \quad (1)$$

式(1)中: W_d^k 为 1×1 卷积核; \otimes 表示卷积计算。

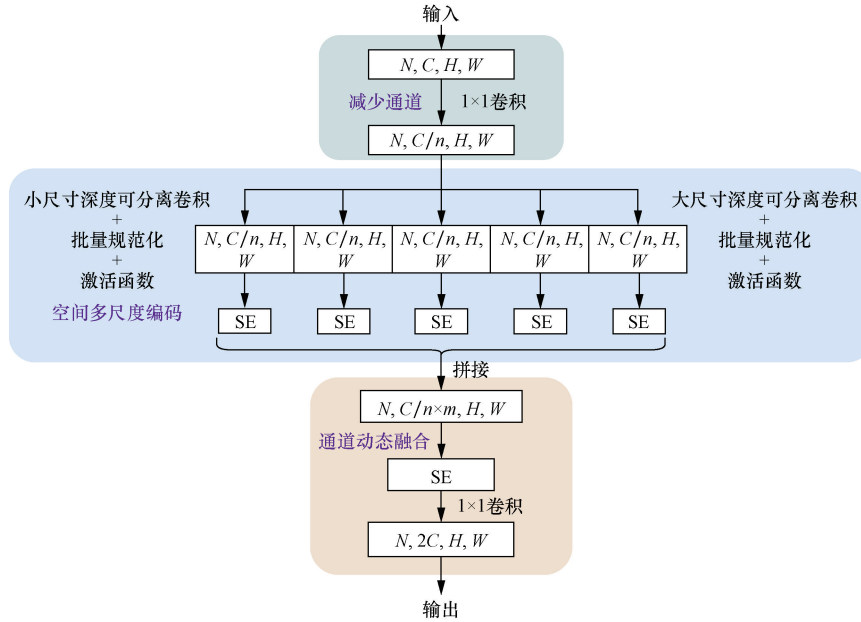


图4 MSFC 模块结构的设计细节

Fig. 4 Design details of MSFC module structure

得到特征映射 $X_d^k \in \mathbf{R}^{N \times C/n \times H \times W}$, 其中 n 为可调节的参数, 决定计算量大小。本文中 n 取 4, 即将通道减少为原来的 1/4。

步骤 2 空间多尺度编码器。

$$W_s^k = \{W_3^k, W_5^k, W_7^k, W_9^k, W_{11}^k\} \quad (2)$$

$$X_s^k = W_s^k \otimes X_d^k \quad (3)$$

$$X_s^k = \{X_1^k, X_2^k, X_3^k, X_4^k, X_5^k\} \quad (4)$$

$$M_s^k = \text{SE}(X_s^k) \quad (5)$$

式中: W_s^k 为 m 个不同大小的卷积核, m 取 5; SE 为 SE 机制。

本文研究中使用 $3 \times 3, 5 \times 5, 7 \times 7, 9 \times 9, 11 \times 11$ 5 种大小的卷积核, 对 X_d^k 卷积计算得到 5 种尺度的特征映射 $X_s^k \in \mathbf{R}^{N \times C/n \times m \times H \times W}$ 。然后在同一尺度中对不同通道的重要性进行学习, 每种尺度的特征映射 X_s^k 后加上 SE 机制, 得到 5 个 $M_s^k \in \mathbf{R}^{N \times C/n \times H \times W}$, 最终在通道上进行拼接得到 $M^k \in \mathbf{R}^{N \times C/n \times m \times H \times W}$ 。

步骤 3 通道动态融合。

$$M_e^k = \text{SE}(M^k) \quad (6)$$

$$X_o^k = W_d^k \otimes M_e^k \quad (7)$$

再次对不同尺度通道的重要性进行学习, M^k 后加上 SE 机制得到 $M_e^k \in \mathbf{R}^{N \times C/n \times m \times H \times W}$, 最后通过 1×1 的卷积核融合所有特征并扩展通道数为 $2C$ 得到最终输出 $X_o^k \in \mathbf{R}^{N \times 2C \times H \times W}$ 。

3 实验内容

3.1 数据集

本文研究中使用的数据集为 Crack2045, 由 2045

张多种裂缝场景和尺度的图像组成。如表 1 所示, 包括公开数据集 Deepcrack^[24]、Concrete-crack^[25] 及自制数据集 Crack723。

Deepcrack 数据集^[24] 包含了 537 张分辨率为 544 像素 \times 384 像素的裂缝图像, 采集自墙体、路面以及互联网, 覆盖了多种裂缝场景和尺度。Concrete-crack 数据集^[25] 为中东技术大学建筑混凝土裂缝图像数据集, 共 458 张, 每张图像大小为 4 032 像素 \times 3 024 像素。这些照片被切片成 4×10^4 张大小为 227 像素 \times 227 像素的图像, 并从裂缝图像中选择 410 张进行像素级标注。Crack723 数据集为自制数据集, 由建筑结构检测加固公司采集自一座需进行加固设计的砌体结构办公楼。这个自制数据集包含了 45 张大小为 5 120 像素 \times 3 840 像素的裂缝图像, 并使用 Photoshop 软件对裂缝进行像素级标注, 再经过切片处理和筛选最终得到 723 张大小为 256 像素 \times 256 像素的图像。

将以上 3 个数据集的大小统一处理为 256 像素 \times 256 像素, 并按照 7:3 的比例划分将其划分为训练集和测试集, 最终得到训练集 1 436 张, 测试集 609 张。表 1 中的三种数据集示例如图 5 所示。

表 1 数据集

Table 1 The dataset

数据集	裁剪前数量	裁剪前大小	裁剪后数量	裁剪后大小
Deepcrack	537	544 \times 384	912	256 \times 256
Concrete-crack	410	227 \times 227	410	256 \times 256
Crack723	44	5 120 \times 3 840	723	256 \times 256
Crack2045	991	—	2 045	256 \times 256

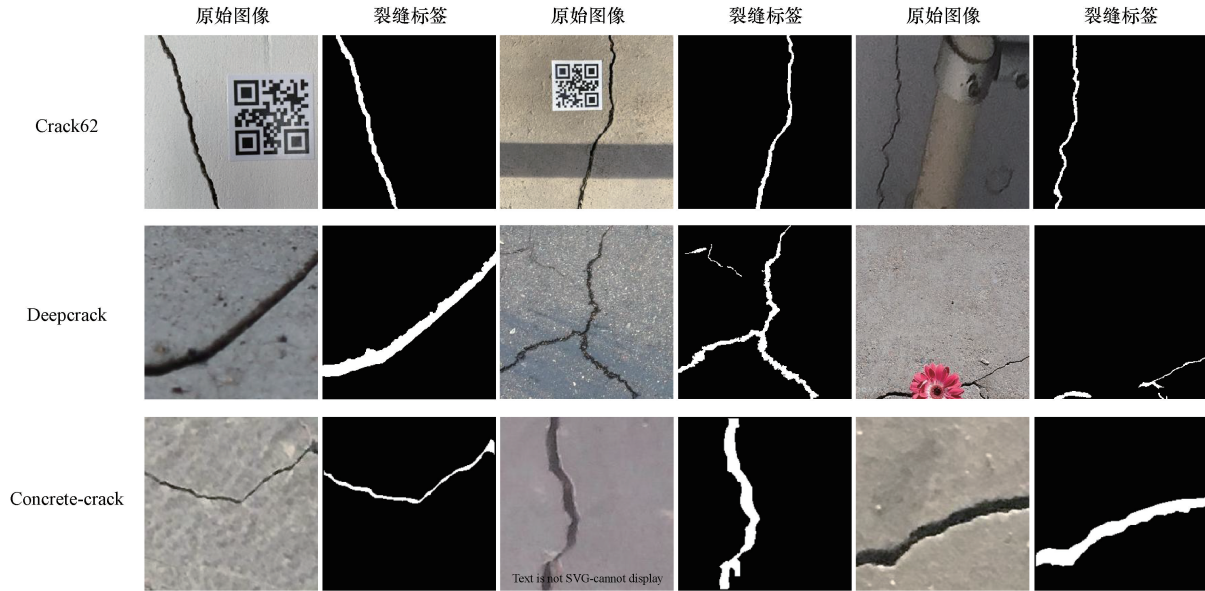


图5 数据集示例图片

Fig. 5 Dataset example images

3.2 实验环境与配置

为了确保实验的可靠性和结果的可比性,提供的详细本地实验配置如表2所示。

本文研究中采用 PyTorch 深度学习框架,CPU 为 Intel(R) Xeon(R) CPU @ 2.00 GHz, GPU 采用了 NVIDIA Tesla P100,采用 200 个 Epochs 来训练深度学习模型,使用了 Adam 优化器来优化模型参数,并采用相同的随机种子进行训练和测试。这些训练参数的选择经过了仔细的考虑,旨在保证模型性能的同时,尽可能地减少训练时间和资源消耗。

表2 实验配置

Table 2 Experimental setup

硬件和训练参数	配置
CPU	Intel(R) Xeon(R) CPU @ 2.00 GHz
GPU	NVIDIA Tesla P100
Cuda	12.1
PyTorch	2.1.2
Python	3.10
Epoch	200
Optimizer	Adam

3.3 评估指标

为了对比评价本文所提出卷积模块的分割性能,主要使用了以下性能评价指标,包括精确度 (P_{recision})、召回率 (R_{ecall})、 F_1 分数 ($F_1\text{-score}$)、交并比 (intersection over union, IoU)。其数学公式为

$$P_{\text{recision}} = \frac{TP}{TP + FP} \quad (8)$$

$$R_{\text{ecall}} = \frac{TP}{TP + FN} \quad (9)$$

$$F_1\text{-score} = \frac{2TP}{2TP + FP + FN} \quad (10)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (11)$$

式中:TP 为模型将样本中的正样本正确预测为正样本的数量;FP 为模型将样本中的负样本错误预测为正样本的数量;FN 为模型将样本中的正样本错误预测为负样本的数量。

除了分割性能,模型的效率和适用性同样至关重要。在实际应用中,还需考虑模型参数量 (Params),即指神经网络或模型内部可学习参数的数量,以及计算量 (FLOPs),指模型在单个输入上进行推理时执行的数学操作数量。

以上这些评估指标将帮助全面了解和比较所提出模块的性能。

3.4 超参数优化

在模型训练中,超参数的选择对于模型性能至关重要。它们直接影响模型的性能表现。本文研究中使用 Crack2045 数据集对 U-Net 进行多个实验来确定最优超参数。首先使用训练集训练模型并记录表现最佳轮次,然后在测试集上确认模型效果。重复训练和测试过程,直至找到最佳超参数组合。

3.4.1 Batch sizes

在模型训练中,批量大小也是一个关键超参数,它决定了每次迭代中用于更新模型参数的样本数量。由于硬件配置的限制,本文研究中将批量大小设定为 32。在实际实验中需要根据硬件性能和任务需求进行进一步调整,以找到最适合研究目标的批量大小。

3.4.2 学习率

学习率(learning rate, LR)是模型训练中的一个关键超参数,它决定了网络优化过程中权重更新的幅度,并可能对网络收敛和最终性能产生重大影响。在本研究中,通过使用4种学习率(1×10^{-1} 、 1×10^{-2} 、 1×10^{-3} 和 1×10^{-4})来训练网络以获得最佳训练性能和最佳学习率,并使用相同损失函数和训练批次大小。

如表3所示,当学习率设置为 1×10^{-3} 时, F_1 分数(83.00%)和交并比(73.03%)达到其最大值。相反,当学习率设置为 1×10^{-1} 时,4种评估指标值均为最低。因此, 1×10^{-3} 被确定为具有最佳学习率。

表3 不同学习率训练的模型测试结果

Table 3 Test results of models trained with different learning rates

学习率	召回率/%	F_1 分数/%	交并比/%
1×10^{-1}	80.54	79.29	68.46
1×10^{-2}	83.82	82.23	71.97
1×10^{-3}	86.25	83.00	73.03
1×10^{-4}	83.14	81.99	71.98

注:加粗数值为最优数据。

3.4.3 损失函数

在模型训练中,损失函数的选择也至关重要,需要根据数据集的特点来选择适合的损失函数。在裂缝图像中,正样本(目标区域)通常比负样本(背景)少得多。为了处理样本不平衡的问题,本文中采用了Dice损失函数和二元交叉熵(binary cross entropy, BCE)损失函数的组合形式。

BCE损失函数是用来评估当前训练得到的概率分布(预测值)与真实分布(输入值或真实值)的差异情况,其计算公式为

$$L_{BCE} = -[y \ln \hat{y} + (1 - y) \ln(1 - \hat{y})] \quad (12)$$

式(12)中: L_{BCE} 为BCE损失值; y 为真实标签; \hat{y} 为模型的预测输出。

Dice损失函数由Milletari等^[26]首次提出,旨在应对语义分割中正负样本强烈不平衡的场景。Dice损失函数来自Dice系数(Dice coefficient),是一种用于评估两个样本相似性的度量函数,取值范围在0~1,取值越大表示两个样本相似度越高。其计算公式为

$$L_{Dice} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (13)$$

式(13)中: L_{Dice} 为Dice损失值; $|X \cap Y|$ 为X和Y之间交集的元素个数; $|X \cup Y|$ 为X和Y之间并集的元素个数;X为真实分割图像的像素标签;Y为模型预测分割图像的像素类别。

为了确定损失函数的最佳组合,本文研究对3种不同Dice + BCE组合的损失函数进行了评估。根据3.4.1节和3.4.2节的结果,设置学习率为 1×10^{-3} ,批量大小为32。

如表4所示,综合比较分析3种不同加权损失函数组合的结果后,确定0.5Dice + 0.5BCE为损失函数的最佳组合。最终U-Net的训练和测试的最优超参数如表5所示。

训练的Batch sizes为32,测试的Batch sizes为1,学习率为 1×10^{-3} 。损失函数为0.5Dice + 0.5BCE。在后续的实验固定下这些参数。

表4 不同损失函数组合训练模型的测试结果

Table 4 Test results of models trained with different combinations of loss functions

损失函数配置	召回率/%	F_1 分数/%	交并比/%
0.3Dice + 0.7BCE	85.28	82.12	71.82
0.5Dice + 0.5BCE	86.25	83.00	73.03
0.7Dice + 0.3BCE	84.90	82.97	72.98

注:加粗数值为最优数据。

表5 最优超参数

Table 5 Key hyperparameters

训练超参数	配置
训练Batch sizes	32
测试Batch sizes	1
损失函数配置	0.5Dice + 0.5BCE
学习率	1×10^{-3}

4 实验结果和分析

4.1 MSFC采用不同卷积核组合对比实验

对MSFC模块采用不同卷积核组合的情况进行了对比实验,卷积核组合包括[3, 5, 7]、[3, 5, 7, 9]、[3, 5, 7, 9, 11]及[3, 5, 7, 9, 11, 13]。

实验结果如表6所示,MSFC模块采用不同卷积核组合的模型在测试集上评估指标,发现采用[3, 5, 7, 9, 11]卷积核组合模型的交并比最大。且所用参数量和计算量均略高于前两种卷积核组合。然而对于[3, 5, 7, 9, 11, 13]虽然卷积核种类增加了,但是交并比降低0.02%。

表6 MSFC采用不同卷积核组合训练的模型测试结果

Table 6 Test results of model trained with different combinations of convolutional kernels in MSFC

卷积模块	召回率/%	F_1 分数/%	交并比/%	参数量/ 10^6	计算量/ 10^6
[3, 5, 7]	84.94	82.33	71.92	1.26	4 017.23
[3, 5, 7, 9]	83.66	82.56	72.39	1.45	4 448.79
[3, 5, 7, 9, 11]	83.94	83.12	73.15	1.67	4 968.84
[3, 5, 7, 9, 11, 13]	84.41	83.17	73.13	1.93	5 595.06

注:加粗数值为最优数据。

综合比较分析,[3,5,7,9,11]卷积核组合为最优选择,之后实验中的 MSFC 模块均采用该组合。

4.2 不同卷积模块对比实验

为了验证所提出的 MSFC 模块的有效性,以 U-Net 架构为基础,第一层均为普通卷积模块 (standard convolutional module, BASE),其他层分别采用普通卷积模块、深度可分离卷积模块 (DSC)、幽灵卷积模块 (Ghost)、共享卷积模块 (Tied) 和本文所提出 MSFC 模块进行对比实验,其中 BASE 卷积模块使用标准的卷积操作,通过多个卷积核在输入图像或特征图上进行卷积,具有较高的参数量和计算量。DSC 模块将标准卷积操作分为深度卷积和逐点卷积两个步骤,显著减少了参数量和计算量。Ghost 卷积模块在应用标准卷积后,通过线性变换生成冗余特征图,从而减少参数量和计算量。Tied 卷积模块在多个层之间共享卷积核参数,减少了参数量,同时保持了一定的特征提取能力。

实验结果如表 7 所示, BASE 卷积模块由于其高参数量和计算量,特征提取能力强,导致较高的召回率、 F_1 分数和交并比,但训练和推理的效率较低。DSC 模块通过分离卷积操作显著减少了参数量和计算量,提升了计算效率,但特征提取能力有所下降,导致较低的召回率、 F_1 分数和交并比。Ghost 卷积模块通过生成冗余特征图进一步减少了参数量和计算量,但可能导致特征表达能力不足,表现略低于标准卷积模块。Tied 卷积模块通过共享卷积核参数减少了参数量,并保持了较好的特征提取能力,但在某些情况下可能会限制模型的灵活性。MSFC 模块通过结合多尺度卷积核进行特征提取,提升了特征表达能力,在性能和效率之间取得了较好的平衡。

综合比较分析, MSFC 模块在不降低模型预测准确率的前提下实现了模型轻量化。

表 7 采用不同卷积模块训练的模型测试结果

Table 7 Test results of models trained with different convolutional module

卷积模块	召回率/ %	F_1 分数/ %	交并比/ %	参数量/ 10^6	计算量/ 10^6
BASE	86.25	83.00	73.03	7.77	13 710.13
DSC	81.62	81.29	70.93	1.52	3 644.98
Ghost	83.49	81.13	70.82	1.12	3 495.56
Tied	83.87	83.11	73.12	2.48	8 274.31
MSFC	83.94	83.12	73.15	1.67	4 968.84

注:加粗数值为最优数据。

4.3 不同卷积模块可视化对比分析

如图 6 所示,选取了测试集中具有代表性的裂缝图片,并对比了不同卷积模块在裂缝检测任务中

的表现。在红色框标记的区域内, MSFC 模块展现出了更高的细节捕捉能力和整体结构保持性,能够准确识别并连接裂缝特征,其他模块可能不足以捕获复杂的裂缝特征,导致模型测试效果不佳。

以此可以得出 DSC、Ghost、Tied 以及 BASE 模块在应对复杂裂缝时存在一定不足,容易导致裂缝出现断裂或部分细节缺失的情况。然而, MSFC 模块提取裂纹的整体结构具有更好的连续性,取得了比其他卷积模块更好的预测结果。

同时,为了更直观地证明 MSFC 模块的效果,如图 7 所示, U-Net 模型分别采用 BASE 模块和 MSFC 模块进行充分训练后,网络输出层对同一张测试图像特征映射的对比。

图 7(a) 为输入图像,图 7(c) 显示了 MSFC 模块在输出层中产生的特征映射,呈现出与图 7(b) 的 BASE 卷积模块不同的特征映射。采用 BASE 模块对于裂缝和非裂缝的特征捕捉对比明显不清晰。而采用 MSFC 模块的输出层中的所有特征映射,对于裂缝和非裂缝的特征捕捉对比明显清晰。这一对比清晰地展示了两类卷积模块在网络训练过程中对捕获有益特征能力的差异, MSFC 模块能够更有效地平衡局部细节和全局上下文信息,从而进一步提升分割性能。

4.4 在其他模型上的对比实验

为了全面验证本研究算法的效果,比较了 5 种常用的语义分割模型: FCN-8s^[7]、Segnet^[5]、Res-U-net^[27]、Attention U-net^[28] 和 U-net ++^[29], 以及采用了 MSFC 模块的这些模型,其中 MSFC 模块采用 4.1 节中实验所得出的最优 [3, 5, 7, 9, 11] 卷积核组合这些模型各具特点:全卷积网络 (FCN-8s) 作为经典的端到端像素级分类模型; Segnet 采用了编码-解码结构,并通过最大池化索引实现高分辨率的语义分割输出; Res-U-net 结合了残差网络 (ResNet34) 和 U-Net 以解决梯度消失问题; Attention U-net 在 U-Net 中引入了注意力机制,提升了对重要特征的关注能力; 而 U-net ++ 通过引入多尺度特征和密集连接进一步增强了特征提取和分割性能。

实验结果如表 8 所示,使用 MSFC 模块后模型的参数量和计算量均显著减少,同时性能保持稳定或略有提升。如 Attention U-net 的参数量从 7.85×10^6 减少到 1.76×10^6 , 计算量从 $13 996.61 \times 10^6$ 减少到 $5 255.31 \times 10^6$, 召回率从 83.71% 提升至 85.10%, F_1 分数从 82.71% 提升至 82.98%, 交并比从 72.74% 提升至 72.91%。这表明 MSFC 模块在提升模型性能的同时,能够大幅降低模型复杂度,使其在资源受限环境中表现出色。

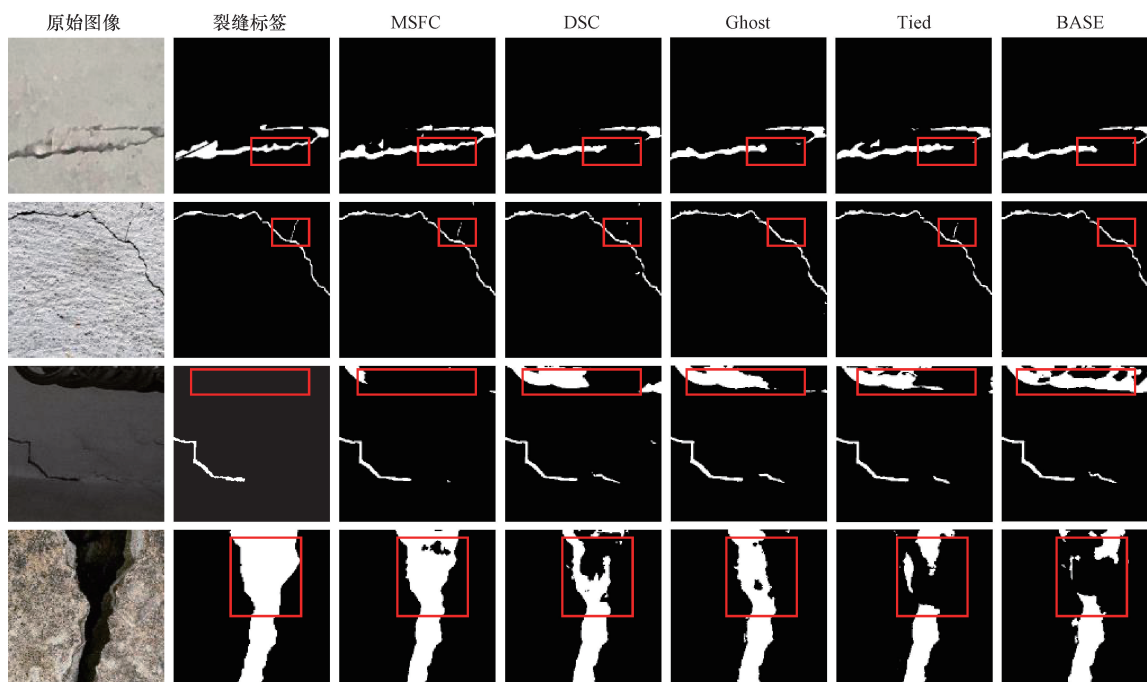


图6 不同卷积模块训练的模型的预测结果对比

Fig. 6 The prediction results of models trained with different convolutional modules compared

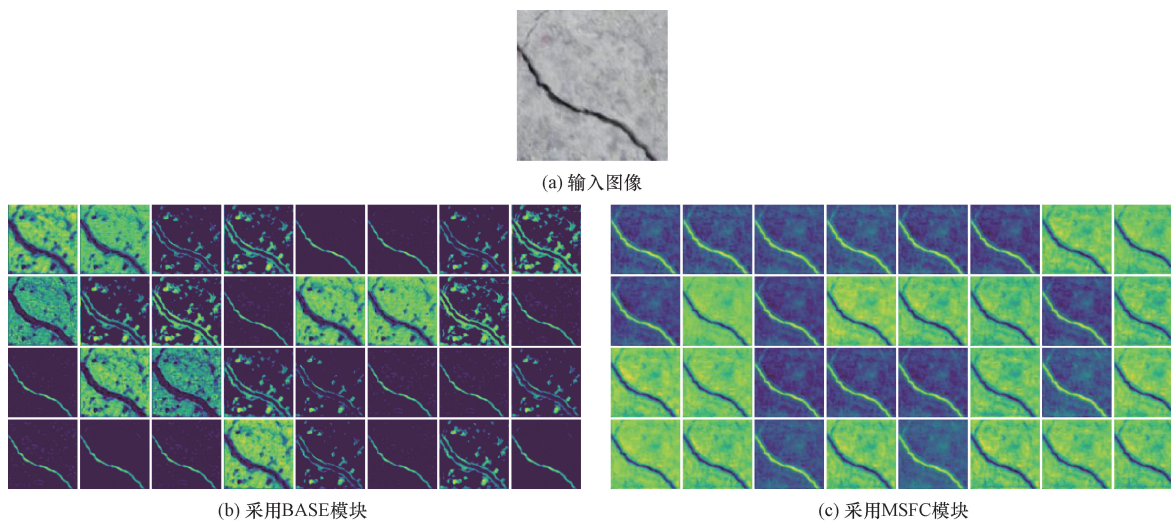


图7 网络输出层对同一张测试图像特征映射的对比

Fig. 7 The comparison of feature mappings for the same test image in the network output layer

表8 不同模型测试结果

Table 8 Test results of models trained with different convolutional module

算法	召回率/%	F_1 分数/%	交并比/%	参数量/ 10^6	计算量/ 10^6	
BASE	FCN-8s	74.75	72.05	59.77	4.72	4 281.18
	Segnet	74.07	77.12	64.77	7.07	6 103.24
	Res-U-net	85.95	83.17	73.24	24.52	14 502.99
	Attention-U-net	83.71	82.71	72.74	7.85	13 996.61
	U-net ++	85.08	83.53	73.43	8.83	33 926.68
MSFC	FCN-8s	77.04	72.53	60.22	0.35	924.55
	Segnet	75.84	77.11	64.68	0.98	1 400.56
	Res-U-net	84.05	83.12	73.12	16.27	12 708.92
	Attention-U-net	85.10	82.98	72.91	1.76	5 255.31
	U-net ++	85.78	83.52	73.52	2.17	15 902.88

综上所述,MSFC 模块具有一定的泛化性,能够无缝集成到不同的语义分割模型中,适应多种应用场景,并有效处理多尺度特征,增强模型的鲁棒性和适应性。

5 结论

对于本文设计的 MSFC 模块在裂缝分割任务上进行的一系列试验,得到以下主要结论。

(1) 基于 U-Net 架构提出的轻量化多尺度动态融合卷积模块 (MSFC), 通过对深度可分离卷积和注意力机制的创新性整合, 在保持准确率不下降的情况下减少了 78.51% 的参数量和 63.75% 的计算量。该方法为资源受限环境下裂缝检测的工程应用提供了新的可能性, MSFC 模块源代码可在 <https://github.com/hanglixiao/MSFC> 获取。

(2) 在裂缝分割任务中, 模型性能的提升并非仅仅依赖于特征数量的增加。关键在于如何充分捕捉并利用不同尺度的特征。MSFC 模块在整合全局图像信息方面展现了卓越的性能。相较于仅使用单一尺度卷积核的方法, MSFC 模块能够更好地协调局部细节特征与全局上下文信息之间的关系。

(3) 对于多尺度卷积的思想, 盲目增加卷积核种类并不会使模型性能显著提升, 反而可能会伴随着更多的参数量和计算量, 未来的工作将探索更加灵活的卷积核配置策略, 以增强模型的泛化能力。

(4) MSFC 模块具有很强的泛化性, 能够无缝集成到不同的语义分割模型中, 参数量和计算量均显著减少, 同时其性能指标保持稳定或有所提升, 适应多种应用场景, 具有广泛的应用前景和实际价值, 对推动土木工程的智能化发展具有重要意义。

参 考 文 献

- [1] König J, Jenkins M, Mannion M, et al. What's cracking? a review and analysis of deep learning methods for structural crack segmentation [J]. *Detection and Quantification*, 2022. DOI: 10.48550/arXiv.2202.03714.
- [2] 李丽. 复杂背景下的路面裂缝检测算法研究 [D]. 西安: 陕西师范大学, 2019.
Li Li. Research on pavement crack detection algorithm under complex background [D]. Xi'an: Shaanxi Normal University, 2019.
- [3] Olga R, Jia D, Hao S, et al. Imagenet large scale visual recognition challenge [J]. *International Journal of Computer Vision*, 2015, 115(3): 211-252.
- [4] 宋泽冈, 刘艳莉, 张长兴. 基于机器视觉的桥梁裂缝检测应用及发展综述 [J]. *科学技术与工程*, 2023, 23(30): 12796-12805.
Song Zegang, Liu Yanli, Zhang Changxing. Review on application and development of bridge crack detection based on machine vision [J]. *Science Technology and Engineering*, 2023, 23(30): 12796-12805.
- [5] 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, 2017: 2481-2495.
- [6] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 2018: 834-848.
- [7] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C]//*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston: CVPR, 2015: 1-7.
- [8] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [C]//*Lecture Notes in Computer Science, Medical Image Computing and Computer-Assisted Intervention*. Dalian: MICCAI, 2015: 234-241.
- [9] 邓砚学, 张志华, 张新秀. 一种新的沥青路面灌封裂缝自动提取方法 [J]. *科学技术与工程*, 2022, 22(16): 6687-6694.
Deng Yanxue, Zhang Zhihua, Zhang Xinxiu. A novel approach for automatic extraction asphalt pavement sealed crack [J]. *Science Technology and Engineering*, 2022, 22(16): 6687-6694.
- [10] 朱俊彬, 杜斌, 许世敏. 基于融合多尺度多通道特征的深度监督网络实现裂缝检测 [J]. *科学技术与工程*, 2024, 24(13): 5595-5603.
Zhu Junbin, Du Bin, Xu Shimin. Crack detection using deep supervised networks with multi-scale and multi-channel feature fusion [J]. *Science Technology and Engineering*, 2024, 24(13): 5595-5603.
- [11] 张修杰, 袁嘉豪, 岳学军, 等. 基于优化 DeepLabv3+ 的混凝土梁裂缝分割及特征量化 [J]. *科学技术与工程*, 2023, 23(9): 3794-3803.
Zhang Xiujie, Yuan Jiahao, Yue Xuejun, et al. Crack segmentation and feature quantification of concrete beams based on optimized DeepLabv3+ [J]. *Science Technology and Engineering*, 2023, 23(9): 3794-3803.
- [12] Liu Z, Li J G, Shen Z Q, et al. Learning efficient convolutional networks through network slimming [C]//*Proceedings of the IEEE international Conference on Computer Vision*. New York: IEEE, 2017: 2736-2744.
- [13] Han S, Mao H Z, Dally W J. Deep compression: compressing deep neural networks with pruning, trained quantization and Huffman coding [J]. *ArXiv preprint ArXiv*, 2015: 1510.00149.
- [14] Zhao B R, Cui Q, Song R J, et al. Decoupled knowledge distillation [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE, 2022: 11953-11962.
- [15] 孟庆成, 李明健, 万达, 等. 基于 M-Unet 的混凝土裂缝实时分割算法 [J]. *土木与环境工程学报(中英文)*, 2024, 46(1): 215-222.
Meng Qingcheng, Li Mingjian, Wanda, et al. Real-time segmentation algorithm for concrete cracks based on M-Unet [J]. *Chinese Journal of Civil and Environmental Engineering*, 2024, 46(1): 215-222.
- [16] 钟志峰, 何佳伟, 侯瑞洁, 等. 改进 UNet 的轻量化道路图像语义分割算法 [J]. *现代电子技术*, 2022, 45(19): 71-76.
Zhong Zhifeng, He Jiawei, Hou Ruijie, et al. Improved UNet lightweight road image semantic segmentation algorithm [J].

- Modern Electronic Technology, 2022, 45(19): 71-76.
- [17] 朱立学, 伍荣达, 付根平, 等. 基于多尺度串联空洞卷积的轻量化 UNet 香蕉图像分割[J]. 农业工程学报, 2022, 38(13): 194-201.
Zhu Lixue, Wu Rongda, Fu Genping, et al. Lightweight UNet for banana image segmentation based on multi-scale cascaded dilated convolution [J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(13): 194-201.
- [18] Howard A, Zhu M, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[J]. arXiv: Computer Vision and Pattern Recognition, 2017. DOI: 10.48550/arXiv.1704.04861.
- [19] Wang X D, Stella X Y. Tied block convolution: leaner and better cnns with shared thinner filters [C]//Proceedings of the AAAI Conference on Artificial Intelligence. San Francisco: AAAI, 2021: 10227-10235.
- [20] Han K, Wang Y, Tian Q, et al. GhostNet: more features from cheap operations [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 1-9.
- [21] 李润龙, 王运圣, 徐识溥, 等. 基于轻量化卷积神经网络的改进模型与验证 [J]. 科学技术与工程, 2020, 20(28): 11653-11658.
Li Runlong, Wang Yunsheng, Xu Shipu, et al. A new module based on light-weight convolutional neural network [J]. Science Technology and Engineering, 2020, 20(28): 11653-11658.
- [22] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York, IEEE: 2017: 6230-6239.
- [23] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York, IEEE: 2018: 7132-7141.
- [24] Liu Y, Yao J, Lu X, et al. DeepCrack: a deep hierarchical feature learning architecture for crack segmentation [J]. Neurocomputing, 2019, 2019: 139-153.
- [25] Dung C V, Anh L D. Autonomous concrete crack detection using deep fully convolutional neural network [J]. Automation in Construction, 2019, 2019: 52-58.
- [26] Milletari F, Navab N, Ahmadi S A. V-Net: fully convolutional neural networks for volumetric medical image segmentation [C]//Fourth International Conference on 3D Vision (3DV). Stanford: IEEE, 2016: 1-10.
- [27] Diakogiannis F I, Waldner F, Caccetta P, et al. ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 2020: 94-114.
- [28] Wang S, Li L, Zhuang X. AtU-NET: attention U-Net for brain tumor segmentation [C]//International MICCAI Brainlesion Workshop. Cham: Springer, 2022: 302-311.
- [29] Zhou Z, Siddiquee M M R, Tajbakhsh N, et al. UNet++: a nested U-Net architecture for medical image segmentation [C]//4th Deep Learning in Medical Image Analysis (DLMIA) Workshop. Cham: Springer, 2018: 3-11.