

基于 ST-UNet 和目标特征的混凝土裂缝检测^①



范昊坤, 刘向阳

(河海大学 数学学院, 南京 211106)

通信作者: 刘向阳, E-mail: liuxy@hhu.edu.cn

摘要: 混凝土裂缝对结构承载能力、耐久性和防水性有负面影响, 因此早期裂缝检测具有重要意义. 大数据和深度学习的快速发展, 为裂缝智能检测提供了有效的方法. 针对裂缝检测过程中图像正负样本不平衡, 裂缝区域色彩深沉和低亮度的特点, 提出一种基于 ST-UNet (Swin Transformer U-Net) 和目标特征的裂缝检测方法. 该算法在网络中引入 CBAM 注意力机制, 使网络更加关注图像中对裂缝检测起决定作用的像素区域, 增强裂缝图像的特征表达能力; 使用 Focal+Dice 混合损失函数代替单一交叉熵损失函数, 处理样本图像正负样本分布不均的问题; 设计 APSD 正则化项优化损失函数, 针对裂缝区域色彩深沉、低亮度的问题, 降低检测的漏检率与误检率. 裂缝检测结果表明: *IoU* 指标提升 22%, *Dice* 指数提升 17%, 该算法是有效可行的.

关键词: U-Net; 语义分割; 损失函数; 正则化项; 裂缝检测

引用格式: 范昊坤, 刘向阳. 基于 ST-UNet 和目标特征的混凝土裂缝检测. 计算机系统应用, 2024, 33(9): 77-84. <http://www.c-s-a.org.cn/1003-3254/9632.html>

Concrete Crack Detection Based on ST-UNet and Target Features

FAN Hao-Kun, LIU Xiang-Yang

(School of Mathematics, Hohai University, Nanjing 211106, China)

Abstract: Concrete cracks have negative impacts on the structural load-bearing capacity, durability, and waterproofing. Therefore, early crack detection is of paramount importance. The rapid development of big data and deep learning provides effective methods for intelligent crack detection. To address the issues of imbalanced positive and negative samples, as well as the challenges posed by deep colors and low luminance in crack areas during the crack detection process, this study proposes a crack detection method based on Swin Transformer U-Net (ST-UNet) and target features. This algorithm introduces the CBAM attention mechanism into the network, enabling the network to focus more on the pixel regions in the image that are crucial for crack detection, thereby enhancing the feature representation capability of crack images. The Focal+Dice mixed loss function replaces the single cross-entropy loss function to address the problem of uneven distribution of positive and negative sample images. Additionally, the design of the APSD regularization term optimizes the loss function, addressing the issues of deep colors and low luminance in crack areas and reducing both missed rates and false rates in detection. The results of crack detection show a 22% improvement in *IoU* and a 17% increase in the *Dice* index, indicating the effectiveness and feasibility of the algorithm.

Key words: U-Net; semantic segmentation; loss function; regularization term; crack detection

① 基金项目: 国家自然科学基金重点项目 (41830110)

收稿时间: 2024-03-20; 修改时间: 2024-04-16; 采用时间: 2024-05-06; csa 在线出版时间: 2024-07-26

CNKI 网络首发时间: 2024-07-29

混凝土结构常因外部环境、过度使用、施工、材料或养护问题而产生裂缝,这些裂缝不仅会导致建筑耐久度下降,还会破坏建筑体结构,产生安全隐患.传统的人工检测方法存在专业人员依赖性强、效率低和主观性强等问题^[1-3],影响了裂缝检测的准确性.近年来,随着计算机技术的发展,图像处理技术已被更多地用于检测缺陷和裂缝^[4].

在较早的研究中,裂缝检测算法大多基于传统的数字图像处理技术算法,如阈值分割法、形态学方法、渗流模型和边缘检测^[5-7].这些方法依赖于人工获取特征,虽然取得了较为丰厚的成果,但是获取特征依赖于人工,最终检测结果主观性强.

近年来,基于深度学习的检测方法在裂缝检测领域取得了显著的成果.汪超宇等^[8]通过引入通道交互空间注意力与金字塔池化,提高了检测精度.两阶段检测模型 Faster R-CNN^[9]在识别混凝土路面裂缝中,研究了不同条件对模型检测效果的影响.Ronneberger 等^[10]提出了具有 U 型对称结构的 U-Net 网络,实现了高低层网络信息融合.基于全卷积网络的像素级裂缝自动检测和测量方法^[11],实现了在像素层面上同时识别和测量不同的裂缝.Chen 等^[12]提出一种更高效的编解码结构用以裂缝检测.MSAR^[13]提出了 Swin Transformer 作为新的视觉网络架构,更适合用于图像分割.利用阈值法和改进的迪杰斯特拉连接算法进行精确提取特征^[14],在 U-Net 网络进行裂缝检测,可以有效提升检测精度.Hu 等^[15]提出了 SENet,他们在通道维度上对特征层进行特征提取,从而提高了通道特征信息的表达能力.Fu 等^[16]提出双重注意网络,他们通过自注意机制自适应地融合局部语义特征,更有效地捕获空间和通道维度的全局依赖关系.Transformer^[17-19]是借鉴语音识别中注意力机制取得成功的网络模型,它基于空间和通道轻量注意力机制在经典语义分割网络中用以强化特征图的自适应性,增强网络模型表达能力.

以上工作提高了对混凝土裂缝图像中细长形状裂缝的检测效果,但对不明显、低亮度、色彩深沉的细小裂缝存在漏检与误检现象.因此,针对混凝土结构裂缝区域的像素呈现出色彩深沉且明显低亮度的特点,本文提出了一种基于 ST-UNet 和目标特征的混凝土裂缝检测方法.该方法基于 Swin Transformer U-Net 语义分割模型^[20],在编码层与解码层的连接位置添加 CBAM 注意力模块,同时采用 Focal+Dice 作为损失函

数^[21],并引入基于目标特征的 APSD (average pixel and standard deviation) 正则化项来改进损失函数,该正则化项使模型更加关注裂缝图像中低亮度、不明显的裂缝区域,从而缓解模型对混凝土裂缝的误检、漏检问题,提高检测效果.

1 研究方法

1.1 网络架构

针对混凝土结构裂缝区域的像素呈现出色彩深沉且明显低亮度的特点,本文提出了一种基于 ST-UNet 和目标特征的混凝土裂缝检测方法,在 ST-UNet 模型中引入 CBAM 模块,网络结构如图 1 所示.作为 Swin Transformer 和 U-Net 的混合体,ST-UNet 继承了 U-Net 的优秀结构,构建了以 CNN 残差网络为主编码器和以 Swin Transformer 为辅助编码器的双编码器结构.

在编码器阶段,先在辅助编码器中采用 ViT (vision Transformer) 将图像划分为不同的 patch,类比序列数据的 token,然后传递到辅助编码器中,利用 Swin Transformer 模块获得图像的全局特征,同时在主编码器中将原始图像送入 ResNet50 网络,获得图像的深度特征;然后,从辅助编码器的全局特征中提取通道相关性,将其嵌入到主编码器的局部特征中,作为全局线索引导主编码器进行全局编码,从而突出整个特征图中重要且更具代表性的通道,对更多的全局特征进行编码,在编码器与解码器的过渡阶段,添加 CBAM 注意力模块,使来自编码器的特征信息更加关注图像中对裂缝检测起决定作用的像素区域,增强裂缝图像的特征表达能力.在解码器阶段,对特征图进行 4 次上采样,每次都将在编码器中低分辨率特征图通过跳跃连接层,拼接到对应解码器上采样生成的特征图中,最后得到检测结果.

1.2 CBAM 模块

Swin Transformer 最初被设计用于自然语言处理等序列数据任务,所以在应用于图像处理等非序列数据时,其注意力机制无法充分捕捉到图像中的局部细节特征,进而影响整体模型的预测准确性.而 CBAM 注意力机制^[22]可以缓解这一问题.CBAM 结合了卷积和注意力机制,能够有效地捕获图像中的空间和通道信息,通道注意力模块 (channel attention module, CAM) 使网络能够学习对整体网络性能具有影响的通道信息,空间注意力模块 (spatial attention module, SAM) 使网络能够更好地捕捉位置关系信息,两个模块

的结合使得网络能够更好地选择和强化图像中的重要特征通道,并在不同位置上自适应地调整特征图的权重.因此,在主干网络之后加入CBAM注意力模块,这有助于增强模型对图像非序列数据的处理能力,提高模型对图像局部细节特征的提取能力.CBAM位置如

图1所示.

CBAM模块包含2个独立的子模块,通道注意力模块(CAM)和空间注意力模块(SAM),可以通过它们分别推断出通道与空间上的注意力权重,并与输入特征相乘对特征进行自适应调整.CBAM结构如图2所示.

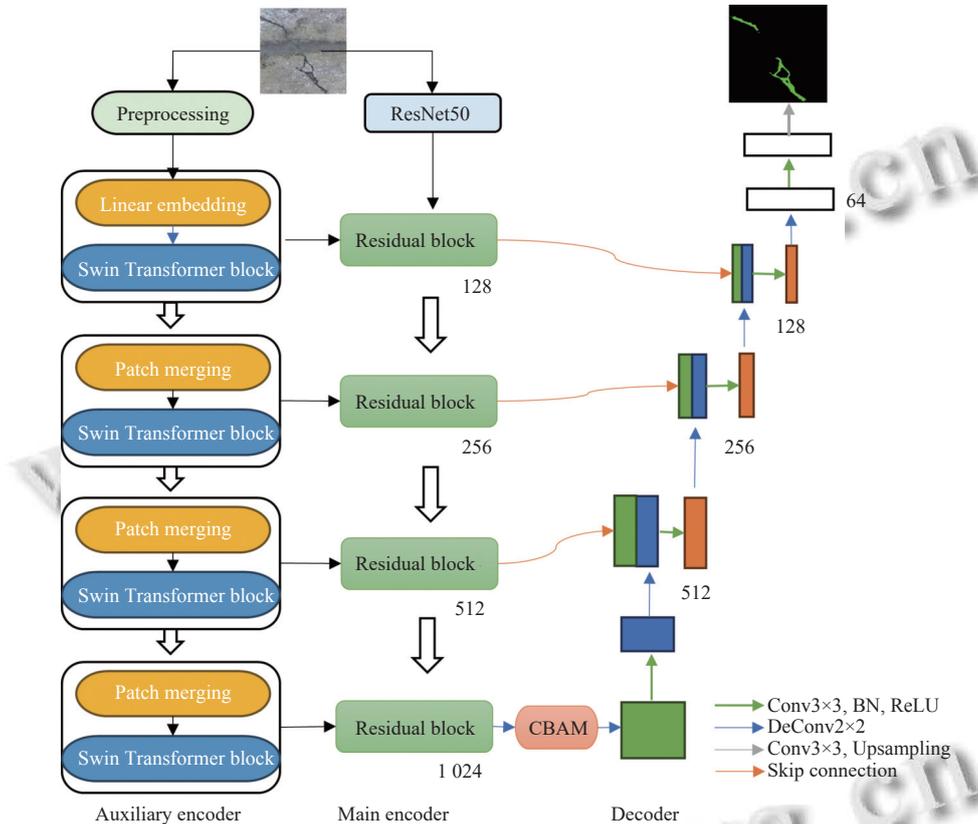


图1 基于多重互信息约束的高表现力语音转换模型框架图

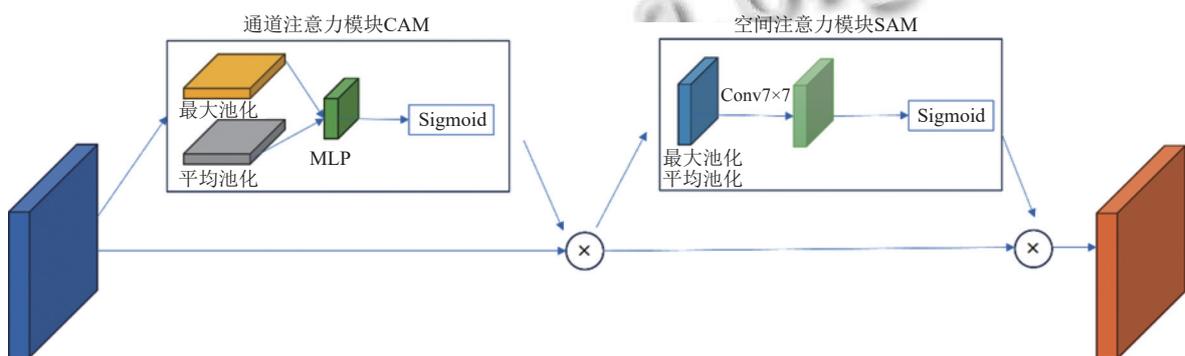


图2 CBAM注意力模块

CAM 是对输入的特征 F 分别进行平均池化和最大池化,得到两个特征,即 F_{avg}^c 和 F_{max}^c .然后再利用多层感知器 (MLP) 进行处理,将 MLP 输出的特征进行相加,再经过 Sigmoid 激活操作,得到 M_c .最后,将 M_c 和输入特征图 F 相乘,生成 SAM 需要的输入特征 F' .

CAM 计算公式如下:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma(W_1 W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c)) \quad (1)$$

其中, $\sigma()$ 表示 Sigmoid 函数, $AvgPool$ 表示平均池化,

$MaxPool$ 表示最大池化, W_0 和 W_1 表示MLP中的参数.

SAM是将CAM输出的特征 F' 作为本模块的输入特征. 首先做一个基于通道维度的平均池化和最大池化, 得到两个特征 F_{avg}^c 和 F_{max}^c , 之后将其进行堆叠. 然后经过 7×7 卷积调整通道数, 经过Sigmoid生成权重系数 M_s , 再相乘得到最终的特征. SAM计算公式如下:

$$\begin{aligned} M_s(F) &= \sigma(f^{7 \times 7}([AvgPool(F'); MaxPool(F')])) \\ &= \sigma(f^{7 \times 7}([F_{avg}^s \cdot F_{max}^s])) \end{aligned} \quad (2)$$

其中, $\sigma()$ 表示Sigmoid函数, $AvgPool$ 表示平均池化, $MaxPool$ 表示最大池化.

2 损失函数

2.1 Dice 损失函数

在裂缝检测过程中, 存在图像背景与裂缝分布不平衡的情况. 为了减轻这种不平衡对模型的影响, 引入Dice损失作为模型的损失函数, Dice损失是一种用来评估两个样本相似性的度量函数, 取值在0-1之间, 值越大表明越相似. 其计算公式为:

$$L_{Dice} = -\frac{1}{n} \sum_{i=1}^n \frac{2 \sum_j y_{ij} \hat{y}_{ij} + c}{\sum_j y_{ij} + \sum_j \hat{y}_{ij} + c} \quad (3)$$

其中, y_{ij} 为第 i 个样本的第 j 个类别的真实标签, \hat{y}_{ij} 为第 i 个样本的第 j 个类别的预测值, n 为样本数, m 为类别数, c 为平滑系数.

2.2 Focal 损失

交叉熵损失用于度量模型的预测概率分布与实际标签之间的差异, 它反映了模型的预测结果与实际标签的匹配程度. 通过最小化交叉熵损失, 深度学习模型能够更好地进行分类任务. 其计算公式为:

$$L_{CE} = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m y_{ij} \log(\hat{y}_{ij}) \quad (4)$$

其中, y_{ij} 为第 i 个样本的第 j 个类别的真实标签, \hat{y}_{ij} 为第 i 个样本的第 j 个类别的预测值, n 为样本数, m 为类别数.

然而, 在裂缝分割任务中, 存在正负样本比例不平衡问题, 传统的交叉熵损失并不适用. 因此, 使用Focal损失来代替传统的交叉熵损失. Focal损失由交叉熵损失改进而来, 根据不同类别的样本分布使用不同的权重, 为占比较小的类别分配更高的权重, 解决正负样本分布不均的问题. 对于本文的二分类任务, Focal损失函数的计算公式为:

$$L_{Focal} = \begin{cases} -\alpha(1-p)\log(p)y = 1 \\ -(1-\alpha)p^\gamma \log(1-p)y = 0 \end{cases} \quad (5)$$

其中, y 表示预测样本的真实标签, p 表示模型预测样本属于正类别(即 $y=1$)的概率, α 是一个平衡因子, 用于调整正类别和负类别之间的权重, γ 是一个调节参数, 用于控制难易样本的权重.

2.3 APSD 正则化项

使用上述混合损失函数进行实验后, 发现模型对于一般的细长狭小裂缝有出色的检测效果. 然而, 在处理裂缝区域色彩深沉、亮度较低、不明显的图像时, 出现了边缘模糊、预测结果不连续、误检和漏检等问题. 为了缓解这些问题带来的影响, 我们设计了APSD正则化项. APSD正则化项通过最小化预测过程中图像裂缝区域的像素均值与像素标准差, 对损失函数进行修正, 减少对低对比度背景和深色裂缝的误检率与漏检率, 从而提高检测效果.

对于混凝土裂缝图像, 裂缝部分通常只占整个图像的一小部分, 且裂缝区域的颜色较深, 亮度较低. 因此, 在模型训练过程中, 计算预测图像中标签为目标的像素点在原始图像中的平均像素值, 将其作为像素值分数. 分数越低表示预测结果包含更多的深色裂缝区域.

此外, 考虑到混凝土裂缝的另一个特点, 即裂缝通常在图像中集中分布在某个区域, 区域的像素值接近, 标准差较小. 因此, 在训练过程中, 计算预测图像中标签为目标的像素点在原始图像中的标准差, 与实际图像中裂缝区域的标准差进行比较, 从而获得标准差分数. 因此, APSD正则化项的计算公式为:

$$P_i = \sum_{x,y} \frac{I_i(x,y) \times M_i(x,y)}{T_i} \quad (6)$$

$$S_i = \sqrt{\sum_{x,y} \frac{(I_i(x,y) - P)^2 \times M_i(x,y)}{T_i}} \quad (7)$$

$$L_{APSD} = \frac{1}{n} \sum_{i=1}^n P_i + \lambda \times \frac{1}{n} \sum_{i=1}^n S_i \quad (8)$$

其中, $I_i(x,y)$ 表示第 i 张图像在像素点 (x,y) 处的像素值, $M_i(x,y)$ 表示第 i 张图像的掩码函数, 它将为预测为正样本的像素点返回1, 否则返回0, T_i 表示所有预测为正样本的像素点的数量, λ 为平衡系数, P_i 表示第 i 张图像中预测为正样本像素点的均值, S_i 表示第 i 张图像中预测为正样本像素点的标准差.

2.4 一种结合了 Focal 损失、Dice 损失和 APSD 正则化项的损失函数

本文为了应对裂缝检测过程中正负样本不平衡的问题,缓解实验过程中模型对具有低亮度、不明显、色彩深沉特点的裂缝的漏检、误检问题,设计了一种新的损失函数.首先,将 Focal 损失函数与 Dice 损失函数相结合,以适应正负样本分布不均的裂缝图像.同时针对裂缝区域色彩深沉和低亮度的特点,为损失函数增加 APSD 正则化项来提高模型对裂缝的特征提取能力,使模型更加注意低亮度、色彩深沉的区域,忽略背景部分,减少漏检、误检的情况,提高模型检测精度.其计算公式为:

$$L = L_{\text{Focal}} + \phi \times L_{\text{Dice}} + \varphi \times L_{\text{APSD}} \quad (9)$$

其中, ϕ 为平衡系数, φ 为正则化项参数.在联合损失函数中,平衡系数 ϕ 通过交叉验证确定,通过尝试不同的平衡系数组合进行实验,最后选择表现最好的组合作为平衡系数;固定平衡系数 ϕ ,对正则化项参数 φ 通过反复使用二分法找到最佳正则化项参数.本文初始筛选范围为[0, 100].

3 实验分析

3.1 实验数据集

在实际应用场景下,采集并构建了混凝土裂缝数据集.数据集共有原始图像 840 张,通过随机旋转、缩放、几何平移等方法对数据集进行数据增强,最终得到 3200 张图像,其中训练集图像 2900 张,测试集 300 张图像.数据集包含多场景多尺度的混凝土裂缝,所有裂缝区域均由人工标注完成.

3.2 实验环境

实验基于深度学习框架 PyTorch 来实现,硬件环境为 GPU NVIDIA A10,采用 Python 3.7 语言环境.采用 SGD 优化器,其中动量项设置为 0.9,权重衰减为 $1\text{E}-4$,初始学习率设置为 0.01,并采用了“Poly”衰减策略.批次大小设置为 4,迭代轮次设置为 100.主编码器采用了半压缩的 ResNet50,辅助编码器采用 Swin Transformer 的 Tiny 配置.

3.3 评价指标

本文实验使用平均交并比 (IoU)、召回率 (Recall) 和 Dice 系数来对模型效果进行定量评估,这些指标为语义分割实验综合度量指标.其计算公式为:

$$IoU = \frac{TP}{TP + FN + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Dice = \frac{2 \times TP}{2 \times TP + FN + FP}$$

其中, TP 为正确预测为裂缝的像素个数, FP 为错误预测背景为裂缝的像素个数, FN 为错误预测裂缝为背景的像素个数.

3.4 图像分辨率实验

本文数据集中图像分辨率不统一,为了保证输入的裂缝图像分辨率一致,需要对数据集图像统一分辨率.本文使用 3 种不同分辨率的图像进行实验对比,分别是 192×192 像素、 224×224 像素和 256×256 像素.实验结果如表 1 所示.

表 1 不同图像分辨率的实验结果

分辨率(像素)	IoU	Recall	Dice
192×192	0.572	0.637	0.681
224×224	0.576	0.649	0.693
256×256	0.583	0.658	0.713

从表 1 可以看出,图像分辨率越大,实验结果越好.相比于其他两组图像, 256×256 像素的图片在所有指标上都获得了最优的效果.因此,图像分辨率对实验结果影响较小,本文实验将所有图像的分辨率统一到 256×256 像素.

3.5 训练过程分析实验

图 3 为本文模型在训练过程中训练损失的变化情况.图像中横坐标表示训练的 epoch 次数,纵坐标表示损失值.可以看出,本文模型在训练过程中的损失值下降,并在训练 60 个 epoch 后趋于收敛.

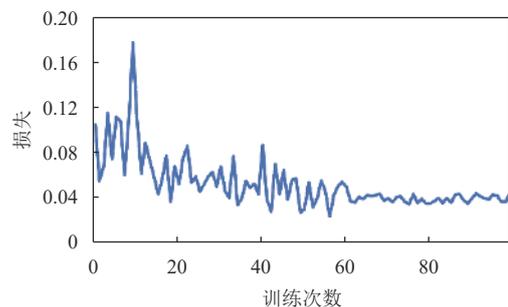


图 3 损失曲线

为了验证正则化项参数 φ 对实验的影响,在数据集上分别采用不同的参数 φ 进行训练,并用测试集测试,实验结果如图 4 所示.

从图 4 的结果可以看出,当正则化项参数 φ 取 0.2 时,模型在 IoU、Recall 和 Dice 评价指标上达到了

最佳性能. 随着系数的增大, 模型性能开始迅速下降, 可能因为正则化项对损失函数的惩罚过于强烈, 导致了模型过度拟合或者丧失了一些细节信息. 因此, 本文选择正则化项参数 $\varphi=0.2$ 进行实验.

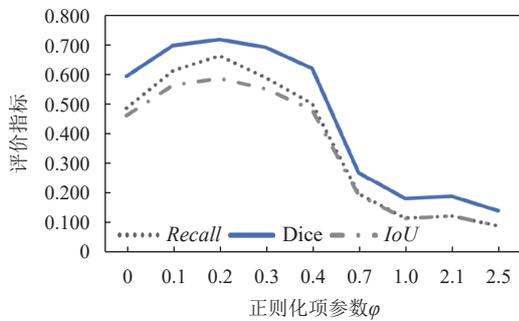


图 4 不同参数正则化项实验结果折线图

3.6 与其他网络对比实验

为了验证改进模型的有效性, 我们在混凝土裂缝数据集上进行了 4 种深度学习模型的实验, 分别是 Mask R-CNN 模型、U-Net 模型、TransUnet 模型, 以及提出的改进模型. 不同模型在裂缝数据集上实验结果如表 2 所示.

由表 2 结果可知, 在运算速度方面, 带有 Transformer 模块或 Swin Transformer 的语义分割模型在运算速度

方面低于纯 CNN 结构的语义分割模型. 其中, 本文方法在数据集上处理一张图片需要 250 ms, 略高于其他模型, 但是考虑到实际裂缝检测任务主要是针对无人机拍摄的图片进行检测, 因此从运算速度和运算复杂度的角度来看, ST-UNet 符合任务场景需求. 由表 2 的结果可知, 在检测效果方面, 目标检测算法 Mask R-CNN 在本实验数据的前提下, 不如经典语义分割模型的效果. 在语义分割模型的比较中, 我们提出的方法在各项评估指标上都表现出良好的效果. 具体来说, 本文模型在 IoU、Recall 和 Dice 系数相对于对于最佳效果的 TransUnet 模型提升了 28.9%、33.7% 和 20.6%, 具有明显的提升. 这表明我们的改进方法在裂缝检测任务中是有效的.

表 2 不同模型裂缝检测实验结果

方法	IoU	Recall	Dice	Time (ms)
Mask R-CNN	0.070	0.040	0.070	110
U-Net	0.364	0.478	0.504	39
TransUnet	0.414	0.436	0.566	125
本文方法	0.583	0.658	0.713	250

为了更直观地对比其他模型与本文模型对裂缝图像的分割效果, 选取部分裂缝图像进行对比分析, 具实验结果如图 5 所示.

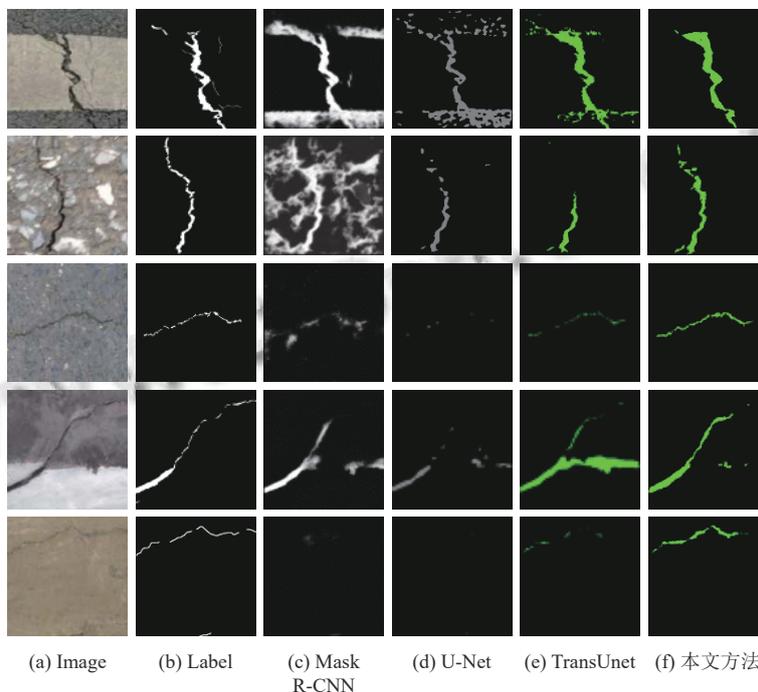


图 5 不同模型裂缝检测结果

从图 5 中可以看出, 在本文的裂缝检测实验中, 目标检测算法 Mask R-CNN 出现了许多误检和漏检现

象, 并且裂缝轮廓模糊. 经典的语义分割模型 U-Net 和 TransUnet, 在表现上优于 Mask R-CNN, 但在对特殊裂

缝检测时也出现了漏检与误检问题,在第1幅图像和第4幅图中将背景部分错误检测为裂缝,在其余图像中表现出较多的漏检.通过对比发现,本文提出的模型分割出的裂缝更加完整,有效减少了误检和漏检的情况.在第1幅图和第4幅图中,U-Net误将背景上的白色条纹和灰色背景相邻的部分标记为裂缝,而本文方法正确的检测到裂缝.在第3幅和第5幅图中,在低亮度裂缝图像检测中,U-Net出现了漏检,TransUnet也只分割出部分轮廓,而本文提出的模型较为完整的检测出了裂缝.综上所述,本文所提出的模型分割效果更加精准,验证了算法的优越性.

3.7 消融实验

为了验证本文所提出的改进的ST-UNet网络的有效性,在基础网络上增加CBAM模块进行消融实验,得到的实验结果如表3所示.

表3 模型消融实验

方法	<i>IoU</i>	<i>Recall</i>	Dice
未加入CBAM的ST-UNet	0.549	0.623	0.684
本文方法	0.583	0.658	0.713

从表3的结果可以看出,改进的ST-UNet模型在加入CBAM注意力模块后,模型的性能得到了提升,在*IoU*、*Recall*和Dice分别提高了6.2%、7.2%和4.3%,验证了CBAM模块可以关注到裂缝的细节特征,使模型能够充分提取图像的特征,也证实了模型所作出的改进是有效的.

为了验证本文所提出的损失函数的有效性,分别使用不同的损失函数在数据集上进行消融实验,并在测试集上测试,得到的实验结果如表4所示.

表4 不同损失函数的消融实验结果

损失函数			评价指标		
CE	Focal	Dice	<i>IoU</i>	<i>Recall</i>	Dice
√	—	—	0.363	0.373	0.499
—	√	—	0.426	0.446	0.552
—	—	√	0.402	0.413	0.527
√	—	√	0.459	0.483	0.590
—	√	√	0.476	0.497	0.609

从表4的数据可以看到,单独使用CE、Focal、Dice作为损失函数实验效果较差,评价指标在整体上都低于结合后的损失函数.选择CE+Focal作为损失函数,在性能上比单独使用Focal在*IoU*、*Recall*和Dice分别提高了7.7%、8.3%和6.9%.而本文提出使用Focal+Dice作为损失函数,与CE+Local相比,在*IoU*、*Recall*和Dice分别提高了3.7%、2.9%和3.2%.这说明使用Focal+Dice作为损失函数,可以发挥这两

类损失函数的协同作用,优化了模型性能,验证了本文所提出的改进的损失函数的有效性.

为了验证本文所提出的正则化项的有效性,在数据集上进行了消融实验,并在测试集上测试,得到的实验结果如表5所示.

表5 正则化项消融实验结果

损失函数	<i>IoU</i>	<i>Recall</i>	Dice
CE+Dice	0.459	0.483	0.590
Focal+Dice	0.476	0.497	0.609
CE+Dice+APSD	0.560	0.609	0.693
Focal+Dice+APSD	0.583	0.658	0.713

从表5的数据可以明显看出,将正则化项添加到CE+Dice损失函数中,*IoU*、*Recall*和Dice分别提高了22.1%、26.2%和17.3%;将正则化项添加到改进的损失函数中,*IoU*、*Recall*和Dice分别提高了21%、32%和17%.这说明了本文提出的正则化项能够有效提高模型的检测性能,验证了本文所提出的正则化项的有效性.

为更直观地对比其他模型与本文模型对裂缝图像的分割效果,选取部分裂缝图像进行对比分析,如图6.

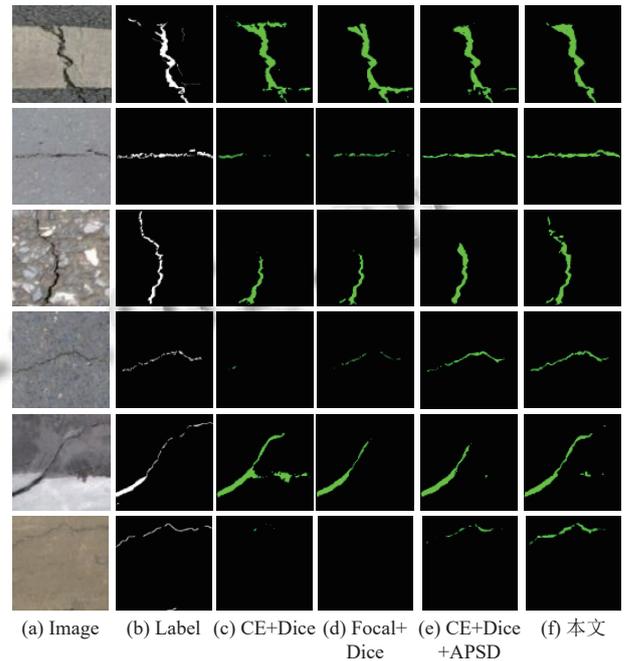


图6 不同模型裂缝检测结果

从图6中可以看出,改进后的损失函数在检测效果上要优于其他损失函数.在第1幅图像和第5幅图像中,改进后的模型消除了误检的背景部分;在2幅和第3幅图像中,本文模型检测裂缝更加完整;在第4幅和第6幅图像中,改进后的模型检测到了其他模型的漏检部分,提高了检测图像的准确性.因此本文提出的增加

正则化项的损失函数在一定程度上解决了裂缝检测中的漏检、误检问题,验证了本文改进方法的有效性。

4 结论与展望

本文提出了一种基于 ST-UNet 和目标特征的混凝土裂缝检测方法,该方法首先在 ST-UNet 网络中加入 CBAM 模块,然后使用 Focal+Dice 作为损失函数,优化模型在正负样本不平衡图像上的分割性能,最后针对裂缝图像中裂缝区域像素呈现出色彩深沉且低亮度的特点,提出了 APSD 正则化项来修正损失函数,减少裂缝的漏检和误检,提高了分割的精度。通过消融实验与对比实验,验证了提出方法的有效性。下一步研究工作要继续优化模型算法,改进图像预处理算法以适应小数据的训练,同时也要增大真实混凝土裂缝数据集容量,使模型能够适合大数据集训练,增加模型的适用性。

参考文献

- 1 田浩江,路娜,崔二洋. 基于改进对比学习的道路裂缝图像分类. 计算机系统应用, 2023, 32(2): 310–315. [doi: [10.15888/j.cnki.csa.008944](https://doi.org/10.15888/j.cnki.csa.008944)]
- 2 宋榕榕,王财勇,田启川,等. 多层次特征融合和注意力机制的道路裂缝模型. 计算机工程与应用, 2023, 59(13): 281–288. [doi: [10.3778/j.issn.1002-8331.2203-0531](https://doi.org/10.3778/j.issn.1002-8331.2203-0531)]
- 3 刘凡,王君锋,陈峙宇,等. 基于并行注意力 UNet 的裂缝检测方法. 计算机研究与发展, 2021, 58(8): 1718–1726. [doi: [10.7544/issn1000-1239.2021.20210335](https://doi.org/10.7544/issn1000-1239.2021.20210335)]
- 4 Zhang KG, Zhang YT, Cheng HD. CrackGAN: Pavement crack detection using partially accurate ground truths based on generative adversarial learning. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(2): 1306–1319. [doi: [10.1109/TITS.2020.2990703](https://doi.org/10.1109/TITS.2020.2990703)]
- 5 Alphonse AS, Starvin MS. A novel and efficient approach for the classification of skin melanoma. Journal of Ambient Intelligence and Humanized Computing, 2021, 12(12): 10435–10459. [doi: [10.1007/s12652-020-02648-x](https://doi.org/10.1007/s12652-020-02648-x)]
- 6 Riyanto I. Morphology based melanoma skin segmentation method. The Scientific Explorer, 2021, 11(3): 7–13.
- 7 Lakshmi BV, Sridevi K. A qualitative approach on denoising and segmentation algorithms for melanoma images. Annals of the Romanian Society for Cell Biology, 2021, 25(2): 1384–1393.
- 8 汪超宇,杜震洪,汪愿愿. 结合通道交互空间组注意力与金字塔池化的高分影像语义分割网络. 浙江大学学报(理学版), 2024, 51(2): 131–142, 152.
- 9 Hacıfendioglu K, Başağa HB. Concrete road crack detection using deep learning-based faster R-CNN method. Iranian Journal of Science and Technology, Transactions of Civil Engineering, 2022, 46(2): 1621–1633. [doi: [10.1007/s40996-021-00671-2](https://doi.org/10.1007/s40996-021-00671-2)]
- 10 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention. Munich: Springer, 2015. 234–241.
- 11 Feng CC, Zhang H, Wang HR, et al. Automatic pixel-level crack detection on dam surface using deep convolutional network. Sensors, 2020, 20(7): 2069. [doi: [10.3390/s20072069](https://doi.org/10.3390/s20072069)]
- 12 Chen HS, Lin HP, Yao MH. Improving the efficiency of encoder-decoder architecture for pixel-level crack detection. IEEE Access, 2019, 7: 186657–186670. [doi: [10.1109/ACCESS.2019.2961375](https://doi.org/10.1109/ACCESS.2019.2961375)]
- 13 Liu Z, Lin YT, Cao Y, et al. Swin Transformer: Hierarchical vision Transformer using shifted windows. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021. 9992–10002.
- 14 Dais D, Bal IE, Smyro E, et al. Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning. Automation in Construction, 2021, 125: 103606. [doi: [10.1016/j.autcon.2021.103606](https://doi.org/10.1016/j.autcon.2021.103606)]
- 15 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141.
- 16 Fu J, Liu J, Tian HJ, et al. Dual attention network for scene segmentation. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 3141–3149.
- 17 Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale. Proceedings of the 9th International Conference on Learning Representations. OpenReview.net, 2021.
- 18 Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with Transformers. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 213–229.
- 19 邵凯,王明政,王光宇. 基于 Transformer 的多尺度遥感语义分割网络. 智能系统学报, 2024. [doi: [10.11992/tis.202304026](https://doi.org/10.11992/tis.202304026)]
- 20 He X, Zhou Y, Zhao JQ, et al. Swin Transformer embedding UNet for remote sensing image semantic segmentation. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 4408715.
- 21 Milletari F, Navab N, Ahmadi SA. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. Proceedings of the 4th International Conference on 3D Vision (3DV). Stanford: IEEE, 2016. 565–571.
- 22 Woo S, Park J, Lee JY, et al. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 3–19.

(校对责编:张重毅)