

# GSM-CrackFormer: 基于高斯尺度混合模型的多应用场景裂缝检测方法

黄廷辉, 李升典

(桂林电子科技大学计算机与信息安全学院 桂林 541000)  
(glhth@guet.edu.cn)

**摘要:** 传统图像处理方法或机器学习方法解决裂缝检测问题通常仅适应特定场景。随着场景切换, 此类方法的检测精度会受到显著的影响, 在多应用场景下缺乏鲁棒性。为了适应多应用场景, 在原裂缝检测方法 CrackFormer 基础上进行改进, 提出一种基于高斯尺度混合模型的检测方法——GSM-CrackFormer。首先通过高斯尺度混合模型构建描述裂缝特征的高斯分布的模块; 然后结合门控机制设计信号转换模块, 将由分布生成的裂缝特征信息转化为锐化裂缝语义特征的指导信号, 通过新颖的上下采样策略进一步平衡模型感受野与其捕获细节特征能力之间的关系; 最后调整损失函数, 缓解裂缝像素与非裂缝像素之间不平衡的问题。在多样化场景数据集 CrackSeg9k 上进行训练和评估的实验结果表明, 所提方法优于文中所对比的方法, 其全局最佳(ODS)指标达到 0.784, 单图最佳(OIS)指标达到 0.785, 平均交并比(MIoU)达到 0.828。

**关键词:** 裂缝检测; 图像处理方法; 高斯尺度混合模型; 裂缝检测场景; 上下采样策略  
**中图法分类号:** TP391.41 **DOI:** 10.3724/SP.J.1089.2024.20196

## GSM-CrackFormer: Crack Detection Method in Multi-Application Scenarios Based on Gaussian Scale Mixed Model

Huang Tinghui and Li Shengdian

(College of Computer and Information Security, Guilin University of Electronic Technology, Guilin 541000)

**Abstract:** Traditional image processing methods or machine learning approaches for addressing crack detection problems typically cater to specific scenarios. As the scene changes, the detection accuracy of such methods is significantly affected, lacking robustness across multiple application scenarios. To adapt to diverse application contexts, improvements have been made upon the original crack detection method, CrackFormer, leading to the proposal of a detection approach based on the Gaussian Scale Mixture (GSM) model, known as GSM-CrackFormer. Firstly, a Gaussian Scale Mixture model is constructed to describe Gaussian distributions related to crack features. Subsequently, a signal transformer is designed, incorporating a gating mechanism to convert crack feature information generated by the distributions into guiding signals for enhancing semantic features. Additionally, a novel up-down sampling strategy is introduced to further balance the relationship between model receptive fields and its capability to capture detailed features. By adjusting the loss function, the imbalance issue between crack and non-crack pixels is effectively mitigated. Finally, we conduct experiments on the CrackSeg9k dataset to evaluate the performance of our proposed method. Our

收稿日期: 2023-04-02; 修回日期: 2023-11-28. 基金项目: 广西自然科学基金面上项目(2019GXNSFAA245053); 广西科技重大专项(AA19254016); 桂林电子科技大学研究生教育创新计划(2023YCX068). 黄廷辉(1970—), 男, 硕士, 副教授, 硕士生导师, 主要研究方向为物联网系统信息安全和协同处理、云计算、边缘计算、图像处理; 李升典(1998—), 男, 硕士研究生, 主要研究方向为图像处理、计算机视觉。

experimental results demonstrate that GSM-CrackFormer outperforms the methods compared in the paper, achieving a global best (ODS) of 0.784, a monograph best (OIS) of 0.785, and a mean intersection ratio (MIoU) of 0.828.

**Key words:** crack detection; image processing method; Gaussian scale mixture model; crack detection scenario; up-down sampling strategy

裂缝是基础设施或工艺品中常见的缺陷,它不仅会给钢筋、混凝土等基础设施的结构安全带来威胁,还会影响玻璃、瓷砖等工艺品的外观和功能,因此,及时了解裂缝的恶化程度,准确、高效地定位裂缝位置对于基础设施或工艺品都至关重要.已有的人工检测裂缝的方法既费时又存在主观性,为了提高检测效率和降低成本,人们迫切需要采用自动化方法来检测裂缝.传统机器学习方法通过捕获裂缝区域与背景区域的天然差异性、裂缝细长形状和不规则走向等特征来检测裂缝,这些方法主要依赖于工程特征的提取,需要手动设计和选择特征,并且效果受限于规则制定的合理性,因此在实际应用中,这些方法的检测效果并不理想.由于深度学习的方法能够从训练数据中自主学习特征,不需要手动设计和选择特征,因此能够显著地提高目标分割的性能;但这些方法受限于训练样本的规模与多样性,往往只能适用于单一场景(如墙、道路或玻璃等),一旦切换到陌生“场景”,检测精确度会显著变化.尽管传统检测方法受限于规则,但在某些场景切换下,其检测效果反而比深度学习方法好.总体而言,深度学习方法的精度更高,但在多样化场景中的鲁棒性还需要进一步提升.针对上述问题,本文借鉴文献[1]的思路,将广泛应用于图像领域的高斯尺度混合(Gaussian scale mixture, GSM)模型与现代卷积神经网络(convolutional neural network, CNN)中流行的卷积算子相结合,兼具传统检测方法与深度学习方法两者的优点.本文提出描述裂缝特征分布的模块 GSM-Mask 和处理模块 GSM-Del,其中, GSM-Mask 模块通过 GSM 构建关于裂缝特征的高斯分布, GSM-Del 模块利用门控机制将高斯分布生成的裂缝特征信息转换为锐化语义特征的指导信号;提出一种上下采样策略,其由混合池化层<sup>[2]</sup>和参照交替卷积思想<sup>[3]</sup>设计的交替卷积层组成,用于平衡模型感受野和模型捕获细节特征的能力,在此基础上引入 BCELoss<sup>[4]</sup>,以有效地提高模型训练的稳定性和泛化性.

## 1 相关工作

### 1.1 传统裂缝检测方法

传统裂缝检测方法依赖于手动设计的特征,如 Sobel 算法<sup>[5]</sup>, Canny 算法<sup>[6]</sup>, 集成差分进化<sup>[7]</sup>和花授粉算法<sup>[8]</sup>的差分进化花授粉算法(differential flower pollination, DFP)<sup>[9]</sup>方法等.这些方法仅对局部进行编码,缺乏裂缝的全局视图,无法应对复杂的裂缝检测场景,但它们也奠定了传统裂缝检测方法的基础.文献[10]基于 Canny 算法提出一种关于金属表面缺陷的半自动检测方法,仅需处理较少的热图就能输出缺陷的边缘图;文献[11]通过激光扫描仪获得点云,从相邻点的强度梯度和 2 个法向量的倾斜角中提取属于裂缝和破损区域的点,并将其转换为二值图,最后基于 Sobel 算法提取裂缝边缘;文献[12]采用鲁棒主成分分析和支持向量机技术对信号进行去噪、分类,用于检测尺寸原件小于 30 mm 的微裂纹;文献[13]根据裂缝图像中每个像素与其附近像素之间的强度顺序,将灰度图像转换为对比度不变的四通道图像,该方法保留了裂缝固有的曲线结构特征,并提高了分割的健壮性.

### 1.2 基于深度学习的裂缝检测方法

近年来,深度学习在计算机视觉领域取得了巨大突破. Crack-SegNet<sup>[14]</sup>和 CrackU-Net<sup>[15]</sup>都是采用编码器-解码器体系结构.其中,编码器通过使用卷积、池化等操作提取高级语义;解码器利用池化索引或跳跃连接融合编码器所提取的高级语义特征,以从高级表示中恢复丢失的空间信息.文献[2]用混合池化替换传统的空间池化,在低级特征图使用空间注意力(spatial attention, SA)捕获裂缝的空间结构信息,使用通道注意力捕获高级上下文的特征;文献[16]提出一种基于多尺度注意力和犹豫模糊集的路面裂缝检测方法,利用多尺度特征融合模块获取多尺度上下文信息,结合双重注意力模块增强捕捉裂缝细节的能力和抑制背景噪声的效果,然后利用犹豫模糊集在多属性决策中的优势计算裂缝的隶属度,最终通过犹豫模糊

度量对分割后的二值图像进行判断; 文献[17]利用生成对抗网络检测裂缝, 旨在解决检测方法将所有像素视为背景的问题; 文献[18]提出一种用于细粒度裂纹检测的方法, 由类似 SegNet 的编码器-解码器架构的新颖注意力模块组成, 通过缩放注意力模块融合编码器-解码器块的输出, 达到抑制非语义特征同时锐化语义特征的目的。

### 1.3 GSM 模型

GSM 模型作为一种经典的统计模型, 在图像恢复、去噪、压缩、特征提取和信号处理等领域得到了广泛应用。在信号处理领域, 文献[19]利用 GSM 模型作为降水回波高通小波系数的先验模型, 表征降水回波的空间分布统计特征; 在图像处理领域, GSM 模型的显著特征是对图像的每个像素进行建模, 得到具有不同尺度的高斯分布的加权和; 文献[20]使用同步稀疏编码扩展 GSM 模型, 获得了非局部扩展, 实验结果表明, 该方法恢复的图像不仅保持边缘的清晰度, 还抑制了不良的伪影; 文献[21]充分利用 GSM 模型对系数邻域相关性的概括能力, 实现图像去噪效果; 文献[22]基于 GSM 先验提出一种新颖的、基于最大后验估计框架的高光谱图像重构方法, 与之前使用手工定制尺度先验的 GSM 模型不同, 该方法是通过 CNN 学习尺度先验和局部均值。

## 2 GSM-CrackFormer 方法

### 2.1 GSM-CrackFormer 方法框架

为了适应多应用场景, 本文改进了原裂缝检测方法 CrackFormer<sup>[18]</sup>, 提出一种基于 GSM 模型 CrackFormer 方法——GSM-CrackFormer。

裂缝检测作为二分类问题, 其检测效果可以表现为检测框或语义分割形式, GSM-CrackFormer 方法属于图像分割领域。在其输入部分, 添加以 GSM 模型为原型设计的 GSM-Mask 模块和转化模块 GSM-Del; 其主干部分是在 CrackFormer<sup>[18]</sup>的基础上改进而来, 在下采样阶段的前 3 轮中, 使用混合池化<sup>[2]</sup>代替最大池化操作, 同时捕获不同位置之间长短距离的依赖关系; 在上采样阶段的后 3 轮中, 借鉴交替卷积思想<sup>[3]</sup>设计交替卷积模块替换双线性插值, 提高模型对细节特征的恢复能力。最后, 将指导信号与方法主干部分在各阶段所获得的裂缝特征进行融合。图 1 所示为 GSM-CrackFormer 方法的结构, 其中待检测图像(图 1 中右上角)会分别传入到以 GSM 模型所设计的模块(图 1 中标记为 GSM-Mask 和 GSM-Del)以及方法主干部分(改进后的 CrackFormer<sup>[18]</sup>, 图 1 左边部分), 最后将 2 部分所捕获到的特征信息汇聚得到预测结果(图 1 中右下角)。

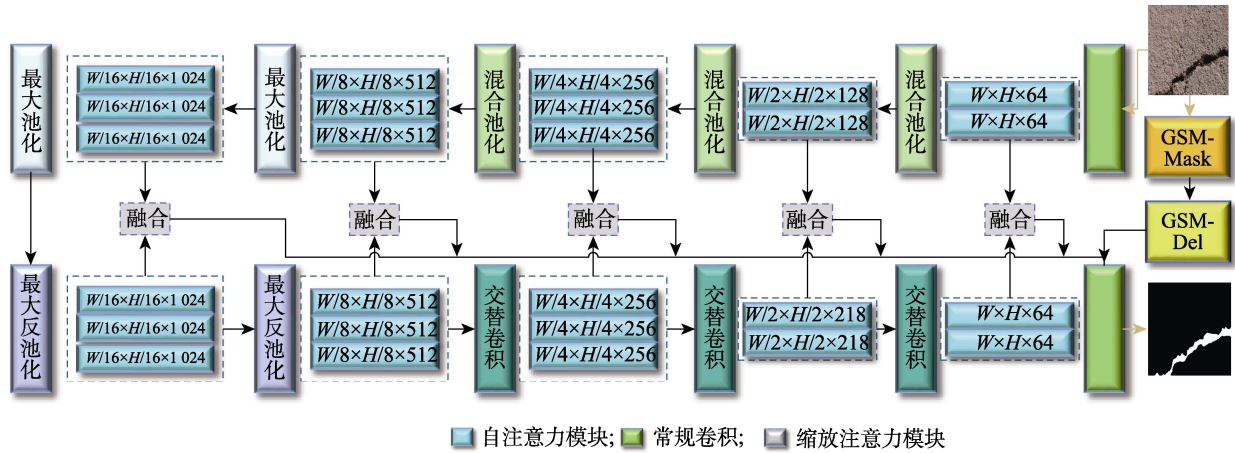


图 1 GSM-CrackFormer 方法结构

### 2.2 GSM-Mask 模块

GSM 模型是一种在图像处理中广泛应用的高维数据概率模型, 具备强大的自适应性, 能够自动调整模型参数并同时学习多个特征, 从而提取更加丰富的图像特征, 非常适用于裂缝检测等相关工作。本文将 GSM 模型应用于裂缝检测, 通过学习裂缝特征分布并从中提取特征信息, 能够更好地考虑

数据的统计特性, 提取更加准确的特征信息。

本文借鉴文献[22]的设计思路, 将捕获裂缝特征问题描述为最大后验概率估计问题。假设裂缝图像为  $Y \in \mathbb{R}^{C \times H \times W}$ , 裂缝特征信息为  $X \in \mathbb{R}^{N \times H \times W}$ , 其中  $C$  与  $N$  均表示裂缝图像与裂缝特征的维度。  $y \in \mathbb{R}^P$  与  $x \in \mathbb{R}^Q$  分别为裂缝图像与裂缝特征信息的矢量化形式 ( $P = CHW$ ,  $Q = NHW$ ), 则有

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}) \times p(\mathbf{x}) \quad (1)$$

假设似然项服从高斯分布, 引入测量矩阵  $\mathbf{A} \in \mathbb{R}^{P \times Q}$ , 则有

$$p(\mathbf{y}|\mathbf{x}) = (\sqrt{2\pi}\sigma)^{-1} \exp\left(-\frac{\|\mathbf{y} - \mathbf{Ax}\|_2^2}{2\sigma^2}\right) \quad (2)$$

假设裂缝特征信息中每个像素  $x_i$  是独立同分布, 利用 GSM 模型表征可得先验项

$$p(\mathbf{x}) = \prod_i p(x_i) = \prod_i \int_0^\infty p(x_i|\theta_i) p(\theta_i) d\theta_i.$$

其中,  $p(x_i|\theta_i)$  服从均值为  $\mu_i$ 、方差为  $\theta_i^2$  的高斯分布, 公式为

$$p(x_i|\theta_i) = (\sqrt{2\pi}\theta_i)^{-1} \exp\left(-\frac{\|x_i - \mu_i\|_2^2}{2\theta_i^2}\right) \quad (3)$$

在 GSM 模型中, 尺度先验  $p(\theta_i)$  通常采用手动定制的算子进行表示, 如无信息先验分布中的杰弗里先验  $p(\theta_i) \propto \theta_i^{0.5} (1 - \theta_i)^{0.5}$ . 为了方便起见, 统一将其记为  $p(\theta_i) \propto \exp(-F(\theta_i))$ , 其中,  $F(\theta_i)$  表示能量函数. 由于 GSM 模型本身的复杂性以及能量函数的不确定性, 单独求解  $p(x_i)$  的解析解是比较困难的, 因此, 本文采用联合估计  $\mathbf{x}$  与  $\boldsymbol{\theta}$  来逼近最优解的方法, 将式(1)更新为

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}) \times p(\mathbf{x}, \boldsymbol{\theta}) = p(\mathbf{y}|\mathbf{x}) \times p(\mathbf{x}|\boldsymbol{\theta}) \times p(\boldsymbol{\theta}) = p(\mathbf{y}|\mathbf{x}) \times \prod_i (p(x_i|\theta_i)) \times p(\boldsymbol{\theta}) \quad (4)$$

对式(4)两边同时取对数, 并将式(2)(3)代入, 可得目标函数

$$\begin{aligned} (\mathbf{x}, \boldsymbol{\theta}) &= \operatorname{argmax}_{\mathbf{x}, \boldsymbol{\theta}} \log_n(p(\mathbf{y}|\mathbf{x})) + \\ &\sum_{i=1}^Q \log_n(p(x_i|\theta_i)) + \log_n(p(\boldsymbol{\theta})) = \\ &\operatorname{argmax}_{\mathbf{x}, \boldsymbol{\theta}} \log_n(\sqrt{2\pi}\sigma)^{-1} + \sum_{i=1}^Q \log_n(\sqrt{2\pi}\theta_i)^{-1} - \\ &\frac{\|\mathbf{y} - \mathbf{Ax}\|_2^2}{2\sigma^2} - \sum_{i=1}^Q \frac{(x_i - \mu_i)^2}{2\theta_i^2} + F(\boldsymbol{\theta}) \rightarrow \\ &\operatorname{argmin}_{\mathbf{x}, \boldsymbol{\theta}} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \sigma^2 \sum_{i=1}^Q \frac{(x_i - \mu_i)^2}{\theta_i^2} + R(\boldsymbol{\theta}) = \\ &\operatorname{argmin}_{\mathbf{x}, \boldsymbol{\theta}} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \sum_{i=1}^Q \ell_i (x_i - \mu_i)^2 + R(\boldsymbol{\theta}) \end{aligned} \quad (5)$$

其中,  $R(\boldsymbol{\theta}) = \operatorname{argmin}_{\mathbf{x}, \boldsymbol{\theta}} 2\sigma^2 F(\boldsymbol{\theta}) + 2\sigma^2 \log_n(\sqrt{2\pi}\theta_i)^{-1} +$

$$2\sigma^2 \sum_{i=1}^Q \log_n(\sqrt{2\pi}\theta_i)^{-1}, \quad \ell_i = \sigma^2/\theta_i^2.$$

推导式(5)时, 为了简化运算, 引入  $\mathbf{l}$  表示  $\boldsymbol{\theta}$ , 并将目标函数转化为迭代更新  $\mathbf{x}$  与  $\mathbf{l}$ . 值得注意的是,  $\mathbf{l}$  本质上仍然与  $\boldsymbol{\theta}$  相关, 而  $\boldsymbol{\theta}$  的求解受到能量函数的约束. 此外, 随着  $\mathbf{x}$  的更新, 模型中的其他参数(如测量矩阵  $\mathbf{A}$ 、均值  $\boldsymbol{\mu}$  和方差  $\sigma$  等)也应相应更新. 本文借鉴了文献[22]的思路, 引入 CNN, 除了  $\mathbf{x}$  之外, 其余参数均利用 CNN 自主地从数据中学习更新, 最终, 式(5)更新为

$$\mathbf{x} = \operatorname{argmin}_{\mathbf{x}, \boldsymbol{\theta}} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \sum_{i=1}^Q \ell_i (x_i - \mu_i)^2 + R(\boldsymbol{\theta}) \quad (6)$$

经随机梯度下降法, 可得目标函数为

$$\mathbf{x}_{t+1} = \mathbf{x}_t - 2\delta \{ \mathbf{A}^T (\mathbf{Ax}_t - \mathbf{y}) + \xi(\mathbf{x}_t)(\mathbf{x}_t - \boldsymbol{\mu}_t) \} \quad (7)$$

其中,  $t$  表示迭代轮次;  $\delta$  表示为梯度下降步长;  $\xi(\cdot)$  表示 CNN 模块, 用于更新  $\mathbf{l}$ . 虽然本文的推导流程与文献[22]相似, 且目标函数相同, 但本文在具体模块实现上与文献[22]存在差异. 下面通过迭代流程, 逐一介绍各个模块的设计.

本文采用多轮迭代更新方式优化  $\mathbf{x}_t$ , 从而得到裂缝特征信息  $\mathbf{x}_t$ . GSM-Mask 模块流程如图 2 所示.

(1)  $\mathbf{A}$  与  $\mathbf{A}^T$  矩阵. 作为迭代更新的第 1 个环节,  $\mathbf{A}$  与  $\mathbf{A}^T$  矩阵用于提取式(1)似然项的高斯分布均值. 这 2 个矩阵具有相同的结构, 包含 2 个卷积层和 1 个残差连接层. GSM-Mask 模块中, 裂缝特征信息  $\mathbf{x}_t$  率先传入  $\mathbf{A}$  矩阵进行处理, 得到高斯分布均值; 然后将该均值与  $\mathbf{y}$  进行差运算得到差异部分, 并将其传入  $\mathbf{A}^T$  矩阵, 最终得到预估结果.  $\mathbf{A}$  与  $\mathbf{A}^T$  的结构图如图 3a 所示.

(2) 特征交融. 特征交融的目的是采集裂缝特征信息, 以便后续获取分布均值与方差信息, 其与文献[22]方法在实现上存在显著差异. 此部分由 3 个卷积层组成密集卷积层, 首先通过采用 `unfold` 函数对裂缝特征信息  $\mathbf{x}_t$  进行局部块分解, 通过 `view` 函数重新组装局部块, 可以在避免使用卷积情况下将特征张量转化为所需的形状, 与直接使用卷积相比, 这样做可以避免特征丢失并减少内存开销; 然后通过密集卷积层进行特征重用. 该部分的最终输出结果标记为  $\mathbf{M}$ , 结构如图 3b 所示.

(3)  $\xi(\cdot)$  与  $\boldsymbol{\mu}$ . 本文提出一种使用非对称轻量 SegNet 来获取高斯分布的均值和方差的方法, 此网络采用编码器-解码器结构, 其中, 每个编码块与解码块均仅包含 1 个卷积层. 编码器部分由 3 个编码块组成, 其中, 后 2 个编码块之间添加了平均池化层以进行下采样. 解码器部分由 2 个解码块和

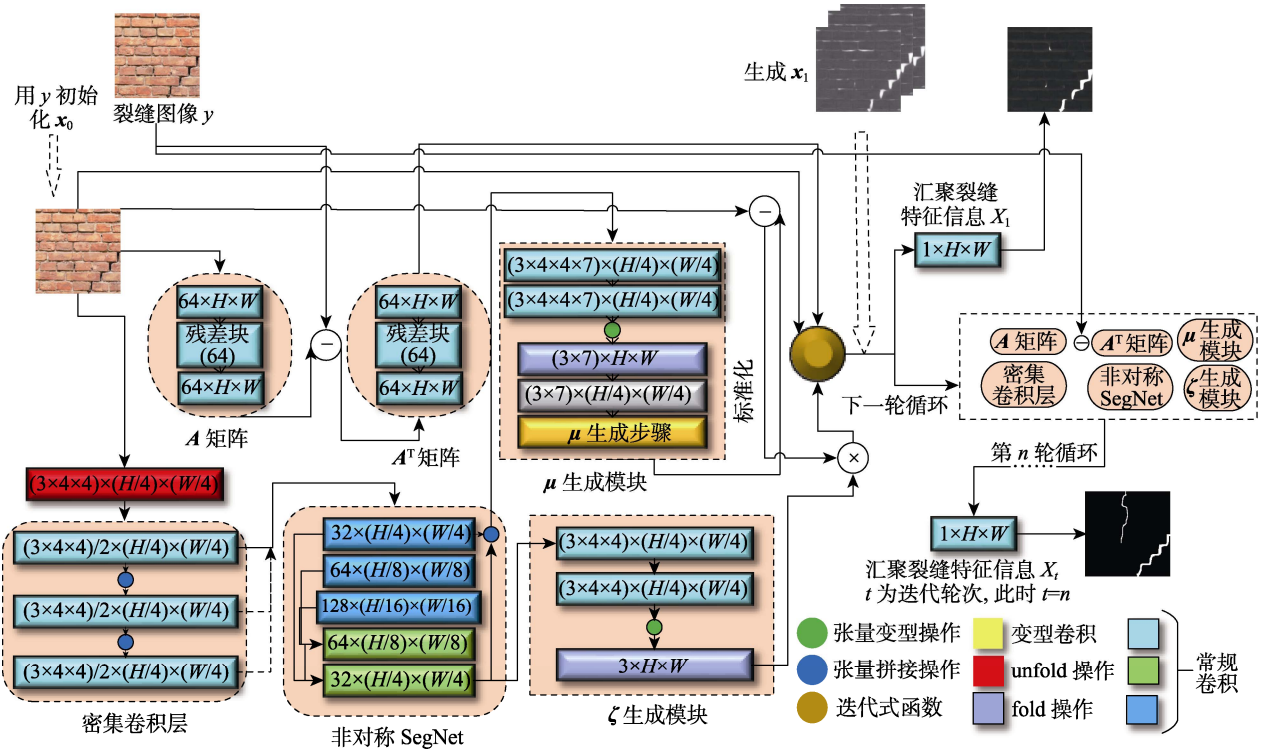


图 2 GSM-Mask 模块流程

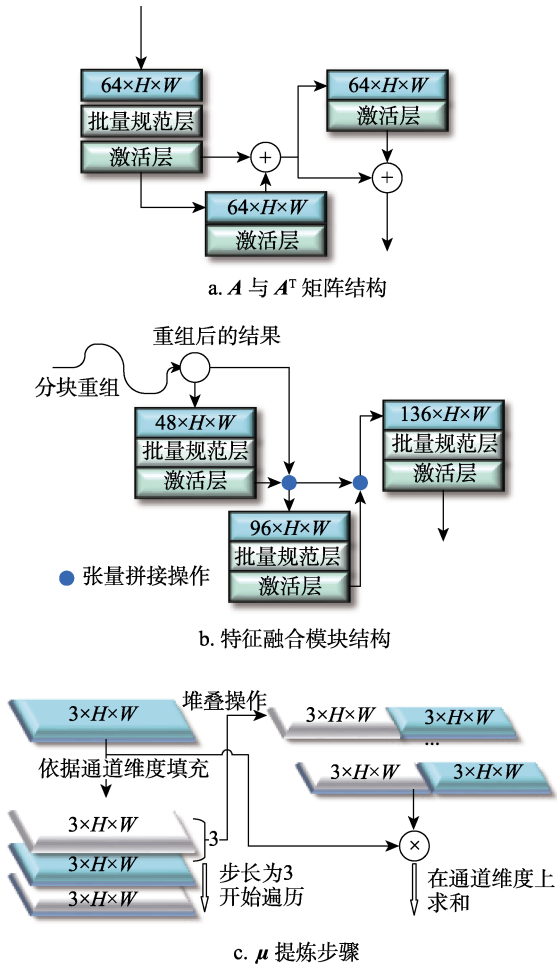
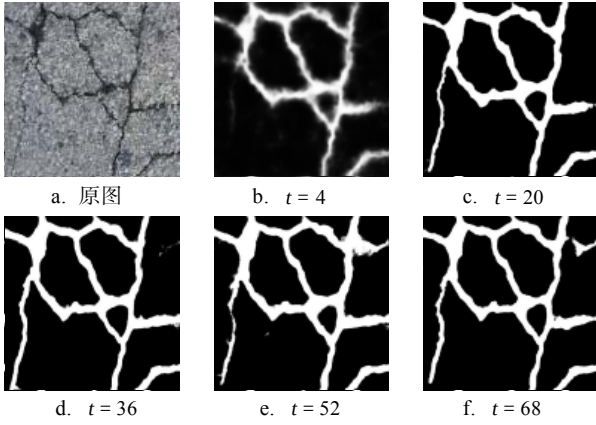


图 3 GSM-Mask 模块结构

2 个分别表示  $\xi(\cdot)$  和  $\mu$  的模块组成。在对  $\xi(\cdot)$  的提炼步骤中, 首先编码器通过 3 个编码块进一步提炼从密集卷积层获得的特征  $M$ , 得到 3 个不同尺度下的特征信息, 分别记为  $f_1^e$ ,  $f_2^e$  和  $f_3^e$ ; 将特征信息  $f_2^e$  与  $f_3^e$  进行融合, 随后再融合  $f_1^e$ , 达到多尺度特征融合的目的; 将融合结果通过 reshape 函数展开特征信息, 最后通过 fold 函数将展开后的特征信息重新组装, 组装后的特征信息作为高斯分布的方差信息  $\xi(x_t)$ 。在  $\mu$  的提炼步骤中, 前期步骤跟  $\xi(\cdot)$  提炼方法一致, 获得重新组装后的特征信息; 然后借鉴 im2col 的设计思想, 将卷积操作转化为矩阵间的乘法。  $\mu$  的提炼步骤如图 3c 所示。首先对原裂缝特征  $x_t$  在高度和宽度维度上进行填充, 以保证边缘像素计算的正确性; 然后通过循环从填充后的张量中提取子张量, 并沿着通道维度进行堆叠; 最后将重新组装的特征信息与堆叠的原始裂缝特征信息进行逐元素相乘, 并将结果沿着通道维度进行求和, 此结果作为高斯分布的均值信息  $\mu_t$ 。

综合上述迭代过程获得的中间结果, 根据式 (7) 迭代更新  $x_t$ , 最后通过卷积操作实现特征汇聚构成裂缝特征图  $X_t$ 。图 4 所示 GSM-Mask 模块每轮迭代提取的裂缝特征图, 其中,  $t$  为迭代轮次。

图4  $t$  不同时 GSM-Mask 模块的结果

### 2.3 GSM-Del 模块

尽管学习到的高斯分布能够映射出裂缝特征,但并不能完全满足裂缝检测模型的预测需求,不过,这种分布可以为后续裂缝检测模型提供指导信号.因此,本文提出一种结合门控机制和注意力机制的方法,能够将从 GSM-Mask 模块获取到的特征信息转化为锐化语义特征的指导信号.门控机制能够锐化现有特征的显著性,同时尽可能地补充尚未感知的残余特征. GSM-Del 模块的输入包括前一模块捕获的裂缝特征信息  $X_t$  和裂缝图像  $Y$ . 在  $X_t$  的指导下,强化  $Y$  与  $X_t$  共有的特征  $O$ ,并通过 SA<sup>[23]</sup>和高效通道注意力(efficient channel attention, ECA)<sup>[24]</sup>机制提取尚未被提取的潜在特征  $R$ ,最终汇聚得到锐化裂缝语义特征的指导信号,进一步提高后续模型的分割性能.该方法表示为

$$\begin{aligned} O &= \text{Conv}(X_t[X_t \leq 0] \cdot Y), \\ R &= \text{SA}(X_t[X_t > 0] \cdot Y) \cdot \text{ECA}(X_t[X_t > 0] \cdot Y), \\ S &= \text{Conv}(R \cdot Y + (1 - R) \cdot O). \end{aligned}$$

其中,  $S$  表示所求的指导信号.

GSM-Mask 步骤如下:

输入. 裂缝图像  $Y$ , 从分布中所提取的特征信息  $X$ .

输出. 指导信号  $S$ .

Step1. 对  $X$  强转换  $X[x \leq 0.5] = 0$ ,  $X[x > 0.5] = 1$ .

Step2. 获取  $Y$  与  $X$  共有特征信息  $m = Y \cdot X$ .

Step3. 获取残余特征信息  $r = Y \cdot (1 - X)$ .

Step4. 通过空间以及通道注意力机制进一步精练残余特征  $R = \text{SA}(r) \cdot \text{ECA}(r)$ .

Step5. 提炼共有特征  $O = \text{Conv}(m)$ .

Step6. 汇聚残余特征以及共有特征,同时通过卷积网络改变输出维度,得到锐化裂缝语义特征的指导信号  $S = \text{Conv}(R \cdot Y + (1 - R) \cdot O)$ .

### 2.4 CrackFormer 的改进

本文提出上下采样策略替换原 CrackFormer<sup>[18]</sup>的采样策略.在下采样阶段,利用混合池化层<sup>[2]</sup>代替原先单一的池化操作.通过消融实验证明,混合池化层能有效地提高方法的表征能力;在上采样阶段,采用交替卷积设计思想<sup>[3]</sup>,提出交替卷积模块,能够有效地提高图像的保真度,尽可能将采样结果图还原成真实特征图;为了平衡计算开销和效率问题,仅在前3轮下采样和后3轮上采样阶段进行替换操作.最终,整个模型的输出表示为

$$f_i = \zeta(f_{d_i}, f_{u_i}, a_i), i \in \{1, 2, 3, 4, 5\},$$

$$F = \text{Conv}(\text{cat}(S, f_1, f_2, f_3, f_4, f_5)).$$

其中,  $f_{u_i}$  与  $f_{d_i}$  分别表示上下采样的中间结果;  $a_i$  表示缩放注意力层<sup>[18]</sup>的输出结果;  $\zeta(\cdot)$  为融合操作;  $f_i$  表示融合的中间结果;  $F$  为最终检测结果.本文策略的结构图 5 所示.

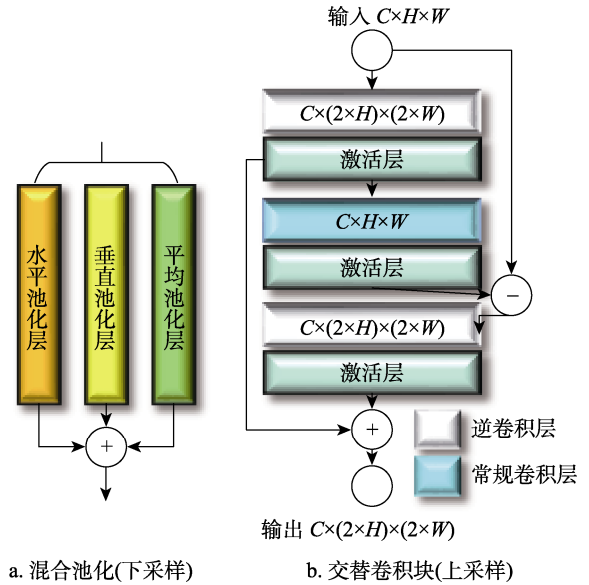


图5 上下采样策略模块结构

### 2.5 损失函数

裂缝检测场景中本身存在前景背景不平衡问题,给模型带来不稳定性.为了解决这个问题,本文采用文献[4]提出的将加权二元交叉(binary cross entropy, BCE)损失函数与 Dice 相似系数相结合的损失函数.

BCE 损失函数可以有效地惩罚模型预测错误的像素点,使模型更加关注预测正确的像素点,定义为

$$l_{\text{BCE}} = -\frac{1}{n} \sum_{i=1}^n \omega_p y_i \log_n(p_i) + \omega_f (1 - y_i) \log_n(1 - p_i).$$

其中,  $\omega_p$  与  $\omega_f$  分别表示调整模型对于正样本(裂缝区域)与负样本(非裂缝区域)的关注程度;  $p$  表示模型预测结果;  $y$  是真实样本;  $n$  为任意常数.

Dice 相似系数可以更好地反映模型对于少数类别的分割效果, 定义为

$$l_{\text{DICE}} = 1 - \frac{2 \sum py + \varepsilon}{\sum p + \sum y + \varepsilon}.$$

其中,  $\varepsilon$  表示光滑系数, 用于避免出现分母为 0 的情况, 保证了损失函数的稳定性. 综上, 最终的损失函数为  $l = \lambda_{\text{BCE}} l_{\text{BCE}} + \lambda_{\text{DICE}} l_{\text{DICE}}$ , 其中,  $\lambda_{\text{BCE}}$  与  $\lambda_{\text{DICE}}$  分别表示  $l_{\text{BCE}}$  和  $l_{\text{DICE}}$  的权重.

### 3 实验及结果分析

#### 3.1 数据集

本文采用哈佛大学提供的数据集 CrackSeg9k<sup>[25]</sup> 进行实验, 它是迄今为止所构建最大、最多样的开源裂缝分割数据集, 总共包含 9 255 幅图像, 由多个开源数据集组成. 该数据集在表面、背景、光照、曝光、裂缝宽度和类型(线性、分支、网状、无裂缝)上具有很大的多样性, 涉及的场景包括沥青道路、水泥道路、瓷砖表面、墙体表面、玻璃等各类常见的生活场景, 最大限度地符合多场景检验要求. 为了进一步验证 GSM-CrackFormer 方法的优越性, 还引入了广泛应用于裂缝检测中的 3 组经典数据集 CrackTree260<sup>[26]</sup>, CrackLS315<sup>[27]</sup> 和 Stone331<sup>[28]</sup>. 这 3 组数据集均可在开源社区中获得 (<https://github.com/qinnzou/DeepCrack>). 其中, CrackTree260 数据集包含 260 幅道路图像, 并且都是在可见光照下拍摄的; CrackLS315 数据集包含 315 幅由线阵相机在激光照射下拍摄的沥青路面图像, 由于对比度极低, 在这个数据集上训练比其他数据集上训练困难很多; Stone331 数据集采自石材切割表面, 由于所采集的石材分割表面十分光滑, 因此人的肉眼很难察觉到. 本文中, 所有数据集均按照 7 : 3 划分为训练集和测试集.

#### 3.2 实验设置

GSM-CrackFormer 方法基于公共深度学习框架 PyTorch 实现. 在模型训练中, batchsize 设置为 4, Shuffle 策略为 True, 并采用随机梯度下降法 (stochastic gradient descent) 对网络参数进行更新, 动量设置为 0.9. 总共进行了 500 个 epoch 的训练, 初始学习率为  $10^{-3}$ , 并使用了 StepLR 策略调整第 20, 50 和 100 个 epoch 的学习率. 所有的实验均在

Ubuntu 18.04 操作系统下进行, 使用单卡 NVIDIA TESLA V100S 32 GB 进行训练和测试.

#### 3.3 评价指标

在评估模型性能时, 本文选用裂缝检测任务中常用的指标有全局最佳(optimal dataset scale, ODS), 单图最佳(optimal image scale, OIS), 平均交并比(mean intersection over union, MIoU)以及计算量(floating-point operations per second, FLOPs).

#### 3.4 参数设定

本文模型的目标函数由 BCE 损失函数与 Dice 相似系数 2 部分组成. 针对 BCE 损失函数中的权重设定, 为了方便求得最优参数, 在实验过程中固定 Dice 系数的权重  $\lambda_{\text{DICE}}$  与 BCE 损失函数的整体权重  $\lambda_{\text{BCE}}$ , 再假定  $\omega_f = 1$ , 随着权重  $\omega_p$  的变化, 分别记录模型最终收敛后的测试结果, 结果如表 1 所示. 可以看出, 当  $\omega_p$  与  $\omega_f$  的比值趋向 20 时, 模型的表现效果最好. 同理, 在求解 BCE 损失函数与 Dice 相似系数之间的权重关系时, 先固定  $\omega_p$  与  $\omega_f$  的比值, 再固定  $\lambda_{\text{DICE}} = 1$ , 此时模型最终收敛后的结果如表 2 所示. 可以看出,  $\lambda_{\text{BCE}}$  与  $\lambda_{\text{DICE}}$  的比值趋向 6 时, 模型的表现效果最好.

表 1  $\omega_f = 1$  时  $\omega_p$  不同取值的测试结果对比

| $\omega_p$ | ODS            | OIS            | MIoU           |
|------------|----------------|----------------|----------------|
| 24         | 0.764 3        | 0.767 6        | 0.786 2        |
| 22         | 0.777 5        | 0.775 6        | 0.808 3        |
| 20         | 0.784 5        | <b>0.785 6</b> | <b>0.818 2</b> |
| 18         | <b>0.785 3</b> | 0.781 6        | 0.811 3        |
| 16         | 0.781 1        | 0.783 6        | 0.801 6        |
| 14         | 0.779 6        | 0.780 6        | 0.808 7        |
| 12         | 0.774 1        | 0.775 6        | 0.788 8        |
| 10         | 0.770 3        | 0.771 0        | 0.791 2        |
| 8          | 0.761 3        | 0.765 3        | 0.783 2        |

注: 粗体表示最优值.

表 2  $\lambda_{\text{DICE}} = 1$  时  $\lambda_{\text{BCE}}$  不同取值的测试结果对比

| $\lambda_{\text{BCE}}$ | ODS            | OIS            | MIoU           |
|------------------------|----------------|----------------|----------------|
| 14                     | 0.767 1        | 0.770 8        | 0.793 9        |
| 12                     | 0.774 4        | 0.775 9        | 0.806 2        |
| 10                     | 0.780 9        | 0.781 7        | 0.812 3        |
| 8                      | 0.781 1        | 0.782 6        | 0.812 9        |
| 6                      | <b>0.784 5</b> | <b>0.785 6</b> | <b>0.818 2</b> |
| 4                      | 0.783 3        | 0.784 6        | 0.814 3        |
| 2                      | 0.780 1        | 0.781 7        | 0.810 6        |

注: 粗体表示最优值.

### 3.5 对比实验

选取 5 种经典方法进行对比实验, 分别是 Canny<sup>[6]</sup>, RCF<sup>[29]</sup>, DeepCrack<sup>[27]</sup>, CrackFormer<sup>[18]</sup> 和 SegCrack9k<sup>[25]</sup> 上表现最好的基准模型 DeepLabV3<sup>[30]</sup> (添加 DINO<sup>[31]</sup> 模块辅助生成特征图). 其中, Canny 是一种经典的边缘检测算法; RCF 基于 VGG16 进行拓展, 通过在每个阶段生成特征图并进行多阶段融合实现边缘检测; DeepCrack 是 SegNet 模型的拓展; CrackFormer 是一种用于细粒度裂缝检测的方法, 在单一裂缝检测场景中表现优异; 而 DeepLabV3 则使用 ResNet-101 作为预训练模型.

6 种方法在 CrackSeg9k 数据集上的实验结果如表 3 所示. 可以看出, GSM-CrackFormer 方法优于对比方法, ODS 和 OIS 指标较基准模型提升了约 0.06, MIoU 提升了约 0.047. 6 种方法对 CrackSeg9k

数据集中 10 个场景随机采样的可视化结果如图 6 所示. 从图 6 中可以看出, Canny 和 RCF 这 2 种边缘检测方法在最后第 2 列单一且较为光滑的玻璃或瓷砖的应用场景中, 表现较好, 但是它们在第 1 列~第 8 列的沥青等道路场景中的表现很差; 与

表 3 6 种方法在 CrackSeg9k 数据集上的结果对比

| 方法                          | ODS            | OIS            | MIoU           | 参数量          | FLOPs         |
|-----------------------------|----------------|----------------|----------------|--------------|---------------|
| Canny <sup>[6]</sup>        | 0.081 5        | 0.102 4        | 0.480 9        |              |               |
| RCF <sup>[29]</sup>         | 0.326 4        | 0.355 3        | 0.554 5        | 15.6M        | 187.9G        |
| DeepCrack <sup>[27]</sup>   | 0.340 5        | 0.442 6        | 0.575 8        | 35.8M        | 468.6G        |
| CrackFormer <sup>[18]</sup> | 0.741 2        | 0.743 5        | 0.795 5        | 8.2M         | 128.3G        |
| DeepLabV3 <sup>[30]</sup>   | 0.723 4        | 0.725 6        | 0.771 6        | 138.6M       | 560.9G        |
| GSM-CrackFormer             | <b>0.784 5</b> | <b>0.785 6</b> | <b>0.818 2</b> | <b>10.5M</b> | <b>207.6G</b> |

注. 粗体表示最优值.

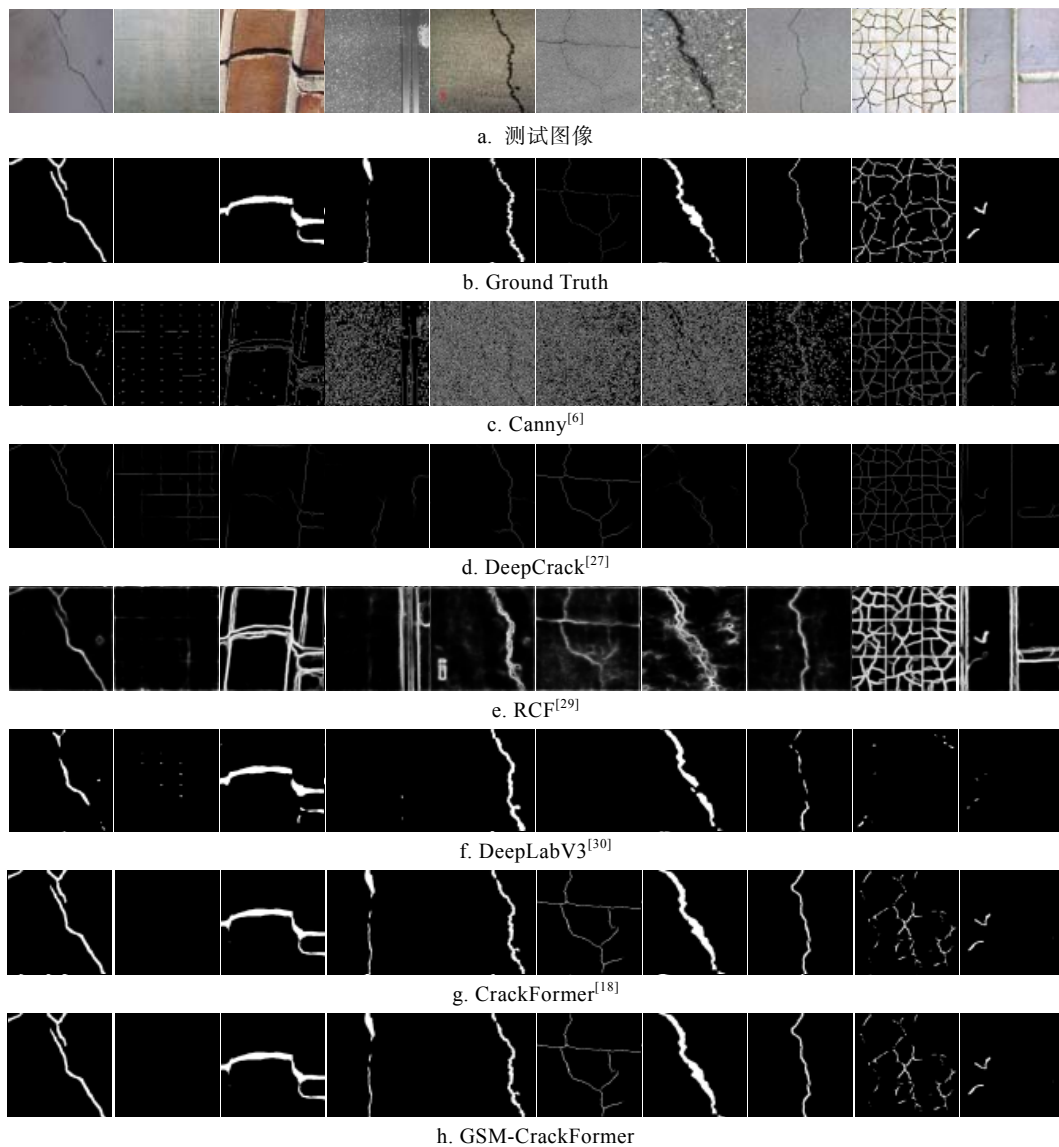


图 6 6 种方法在 CrackSeg9k 数据集上的预测结果

Canny 和 RCF 方法相比, Deepcrack, DeepLabV3 和 CrackFormer 这 3 种应用于分割中的方法恰好相反; GSM-CrackFormer 方法面对混合的应用场景获得结果都很清晰且贴近裂缝的 Ground Truth, 同时对于第 2 列所示混有非裂缝的干扰图的测试数据, 除了 CrackFormer 和 GSM-CrackFormer 方法外, 其余方法都出现误判结果。

实验结果表明, GSM-CrackFormer 方法优于对比方法, 包括目前在 CrackSeg9k 数据集上表现最好的方法。

### 3.6 消融实验

进一步, 在 CrackSeg9k<sup>[25]</sup>, CrackTree260<sup>[26]</sup> CrackLS315<sup>[27]</sup>和 Stone331<sup>[28]</sup>数据集上进行消融实验, 验证各模块的增益和方法的优异性。

(1) 在 CrackSeg9k 数据集上的实验结果表 4 所示。首先将 GSM-Mask 和 GSM-Del 这 2 个模块联合训练, 单独作为预测方法(简称为 GSM), 可以看出, ODS 为 0.721 2, OIS 为 0.721 5, 表明仅联合这 2 个模块作为独立方法, 就几乎优于目前大部分方法。然后改进原 CrackFormer 方法, 分别替换了该方法的上下采样策略、损失函数以及添加 GSM 模块, 可以看出, GSM-CrackFormer 方法对裂缝检测方法提高表征裂缝特征能力均有贡献。其中, GSM 模块完美契合 CrackFormer 模型, 在 ODS 和 OIS 上相较于 CrackFormer 分别提升约 0.022 4, 0.026。其他改进也有相应的提升, 引入结合交替卷积与混合池化操作的上下策略后, 分别提升 0.016 9 和 0.019 7; 替换损失函数后, 提升约 0.012 9 和 0.014 8。

表 4 CrackSeg9k 数据集上的消融实验对比

| 方法          | 采样策略 | 损失函数 | GSM | ODS     | OIS     |
|-------------|------|------|-----|---------|---------|
| GSM         | ×    | ×    | √   | 0.721 2 | 0.721 5 |
|             | √    | ×    | ×   | 0.758 1 | 0.763 2 |
|             | ×    | √    | ×   | 0.754 1 | 0.758 3 |
| CrackFormer | √    | √    | ×   | 0.767 5 | 0.768 2 |
|             | ×    | ×    | √   | 0.763 6 | 0.769 5 |
|             | √    | √    | √   | 0.784 5 | 0.785 6 |

(2) 在数据集 CrackTree260, CrackLS315 和 Stone331 上的实验结果如表 5 所示。可以看出, GSM-CrackFormer 方法在数据集 CrackTree260 和 Stone331 上的表现略逊于 CrackFormer 方法, 但总体而言优于其他对比方法。此外, 由于 CrackSeg9k 数据集在一定程度上可视为 CrackTree260, CrackLS315

和 Stone331 这 3 组数据集的超集, 为了公平验证 GSM-CrackFormer 方法的优异性和鲁棒性, 能更准确地捕获裂缝的本质特征, 对所有参与评估的方法进行了调整。原先的训练集随机采样自 3 组经典数据集的方式改为按相同数量随机采样自 CrackSeg9k 数据集, 这样有助于扩大应用场景范围, 使各参与比较的方法所训练的样本更加多样化, 同时也更加复杂化(存在一定比例的干扰图), 而测试集保持不变。接着, 针对各对比方法对输入样本尺寸的要求, 分别对训练集的图像尺寸进行调整。最后, 进行了 5 轮实验, 最终结果取这 5 轮实验的平均值。可以看出, GSM-CrackFormer 方法受数据集影响不大, 检测指标仅下降 1%~4%; 而其他对比方法得到的检测指标普遍下降较多, 尤其是 RCF 方法, 与本身在 Stone331 数据集上训练且测试的情况下获得指标相比, 在 Stone331 数据集上的指标下降最多, 其中, ODS 下降约 22%, OIS

表 5 4 种方法在 3 个数据集上的消融实验对比

| 数据集          | 方法                           | ODS            | OIS            |
|--------------|------------------------------|----------------|----------------|
| CrackTree260 | RCF <sup>[29]</sup>          | 0.857 1        | 0.863 3        |
|              | DeepCrack <sup>[27]</sup>    | 0.852 1        | 0.864 2        |
|              | CrackFormer <sup>[18]</sup>  | <b>0.881 4</b> | <b>0.883 5</b> |
|              | GSM-CrackFormer              | 0.875 8        | 0.873 4        |
|              | RCF <sup>*[29]</sup>         | 0.742 6        | 0.752 1        |
|              | DeepCrack <sup>*[27]</sup>   | 0.715 6        | 0.734 1        |
|              | CrackFormer <sup>*[18]</sup> | 0.805 1        | 0.812 3        |
| CrackLS315   | GSM-CrackFormer <sup>*</sup> | <b>0.856 1</b> | <b>0.861 2</b> |
|              | RCF <sup>[29]</sup>          | 0.786 1        | 0.810 3        |
|              | DeepCrack <sup>[27]</sup>    | 0.780 6        | 0.807 9        |
|              | CrackFormer <sup>[18]</sup>  | 0.851 4        | 0.861 5        |
|              | GSM-CrackFormer              | <b>0.868 1</b> | <b>0.874 9</b> |
|              | RCF <sup>*[29]</sup>         | 0.701 2        | 0.742 6        |
|              | DeepCrack <sup>*[27]</sup>   | 0.695 2        | 0.731 7        |
| Stone331     | CrackFormer <sup>*[18]</sup> | 0.756 8        | 0.765 9        |
|              | GSM-CrackFormer <sup>*</sup> | <b>0.815 2</b> | <b>0.835 2</b> |
|              | RCF <sup>[29]</sup>          | 0.788 6        | 0.815 6        |
|              | DeepCrack <sup>[27]</sup>    | 0.856 3        | 0.876 6        |
|              | CrackFormer <sup>[18]</sup>  | <b>0.879 9</b> | <b>0.883 3</b> |
|              | GSM-CrackFormer              | 0.861 2        | 0.876 4        |
|              | RCF <sup>*[29]</sup>         | 0.612 0        | 0.681 2        |
|              | DeepCrack <sup>*[27]</sup>   | 0.768 2        | 0.779 8        |
|              | CrackFormer <sup>*[18]</sup> | 0.790 9        | 0.800 2        |
|              | GSM-CrackFormer <sup>*</sup> | <b>0.823 6</b> | <b>0.834 8</b> |

注: 粗体表示最优值, \*号表示训练集为 CrackSeg9k。

下降约 16%，这可能与该方法未能适应检测石材切割表面等场景有关；DeepCrack 和 CrackFormer 方法也都有接近 10% 的下降。实验结果表明，不论从检测精度还是鲁棒性角度看，GSM-CrackFormer 方法均优于目前表现优异的裂缝检测模型。

## 4 结 语

GSM-CrackFormer 方法的设计旨在适应多样化的裂缝检测场景。本文基于 GSM 模型构建了 GSM-Mask 模块，该模块能够生成关于裂缝特征的高斯分布，为裂缝检测方法的研究与改进提供了新的思路。GSM 模型的强大之处在于它能够在每个像素点上建模，具备出色的自适应性，能够自动地调整模型参数并同时学习多个特征。因此，GSM-CrackFormer 方法能够提取更加丰富的图像特征，使其在众多对比方法中脱颖而出。此外，本文采用了门控机制设计信号转化模块 GSM-Del，将由高斯分布生成的裂缝特征转化为锐化的裂缝语义特征的指导信号。这一巧妙的设计不仅在增强已有的显著性的同时，尽可能地补充了尚未发掘的剩余特征。最后，GSM-CrackFormer 方法是在原裂缝检测方法 CrackFormer 的基础上进行改进的。CrackFormer 方法本身已经是一个表现优异的方法。本文通过引入上下采样策略以弥补原方法可能存在的细节丢失的缺陷，同时更换了损失函数，进一步引导方法集中关注裂缝区域，从而提高了方法检测裂缝的能力。实验结果表明，在 CrackSeg9k 数据集上，GSM-CrackFormer 方法的 ODS 为 0.784 5，OIS 为 0.785 6，MIoU 为 0.818 2，均超过对比的检测方法。

与目前主流方法相比，尽管 GSM-CrackFormer 方法能够在 CrackSeg9k 数据集中涉及多个场景下检测到裂缝，不会像其他方法一样出现全黑现象，但对于某些比较模糊的场景，精度还有待提高；此外，正如表 3 所示，该方法的 FLOPs 还是较高，尚不适用于一些资源有限的移动设备或嵌入式系统等。未来，将提升现有方法的检测精度，并且尝试增加模型的应用场景(如夜间检测)，进一步提高鲁棒性。此外，对推理模块的优化也是今后工作的重点。

## 参考文献(References):

- [1] Su Z, Liu W Z, Yu Z T, *et al.* Pixel difference networks for efficient edge detection[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 5117-5127
- [2] Zhou Q, Qu Z, Cao C. Mixed pooling and richer attention feature fusion for crack detection[J]. Pattern Recognition Letters, 2021, 145: 96-102
- [3] Haris M, Shakhnarovich G, Ukita N. Deep back-projection networks for super-resolution[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 1664-1673
- [4] Taghanaki S A, Zheng Y F, Zhou S K, *et al.* Combo loss: handling input and output imbalance in multi-organ segmentation[J]. Computerized Medical Imaging and Graphics, 2019, 75: 24-33
- [5] Sobel I. Neighborhood coding of binary images for fast contour following and general binary array processing[J]. Computer Graphics and Image Processing, 1978, 8(1): 127-135
- [6] Canny J. A computational approach to edge detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, PAMI-8(6): 679-698
- [7] Storn R, Price K. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces[J]. Journal of Global Optimization, 1997, 11(4): 341-359
- [8] Yang X S. Flower pollination algorithm for global optimization[C] //Proceedings of the 11th International Conference on Unconventional Computing and Natural Computation. Heidelberg: Springer, 2012: 240-249
- [9] Hoang N D, Bui D T, Liao K W. Groutability estimation of grouting processes with cement grouts using differential flower pollination optimized support vector machine[J]. Applied Soft Computing, 2016, 45: 173-186
- [10] Pech-May N W, Ziegler M. Detection of surface breaking cracks using flying line laser thermography: a canny-based algorithm[J]. Engineering Proceedings, 2021, 8(1): Article No.22
- [11] Phan A T T, Huynh T N. Pavement crack extraction method from mobile laser scanning point cloud[J]. Advances in Civil Engineering, 2022, 2022: Article No.6317008
- [12] Song Q S, Yan G P, Tang G W, *et al.* Robust principal component analysis and support vector machine for detection of microcracks with distributed optical fiber sensors[J]. Mechanical Systems and Signal Processing, 2021, 146: Article No.107019
- [13] Shi T Y, Boutry N, Xu Y C, *et al.* Local intensity order transformation for robust curvilinear object segmentation[J]. IEEE Transactions on Image Processing, 2022, 31: 2557-2569
- [14] Ran R, Xu X H, Qiu S H, *et al.* Crack-SegNet: surface crack detection in complex background using encoder-decoder architecture[C] //Proceedings of the 4th International Conference on Sensors, Signal and Image Processing. New York: ACM Press, 2021: 15-22
- [15] Huyen J, Li W, Tighe S, *et al.* CrackU-net: a novel deep convolutional neural network for pixelwise pavement crack detection[J]. Structural Control and Health Monitoring, 2020, 27(8): Article No.e2551
- [16] Li C, Wen Y, Shi Q X, *et al.* A pavement crack detection method based on multiscale attention and HFS[J]. Computational Intelligence and Neuroscience, 2022, 2022: Article No.1822585
- [17] Zhang K G, Zhang Y T, Cheng H D. CrackGAN: pavement

- crack detection using partially accurate ground truths based on generative adversarial learning[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 22(2): 1306-1319
- [18] Liu H J, Miao X Y, Mertz C, *et al.* CrackFormer: transformer network for fine-grained crack detection[C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2021: 3783-3792
- [19] Kou Leilei, Wang Zhenhui, Shen Feifei, *et al.* High resolution interpolation for weather radar data based on Gaussian-scale mixtures model in wavelet domain[J]. *Acta Meteorologica Sinica*, 2019, 77(1): 142-153(in Chinese)  
(寇蕾蕾, 王振会, 沈菲菲, 等. 基于小波域高斯尺度混合模型的天气雷达图像高分辨率插值[J]. *气象学报*, 2019, 77(1): 142-153)
- [20] Dong W S, Shi G M, Ma Y, *et al.* Image restoration via simultaneous sparse coding: where structured sparsity meets Gaussian scale mixture[J]. *International Journal of Computer Vision*, 2015, 114(2-3): 217-232
- [21] Liu Qiaohong, Li Bin, Lin Min. Image denoising with dual-directional filter bank GSM model and non-local mean filter[J]. *Optics and Precision Engineering*, 2014, 22(10): 2806-2814(in Chinese)  
(刘巧红, 李斌, 林敏. 结合复方向滤波器组高斯尺度混合模型及非局部均值滤波的图像去噪[J]. *光学精密工程*, 2014, 22(10): 2806-2814)
- [22] Huang T, Dong W S, Yuan X, *et al.* Deep Gaussian scale mixture prior for spectral compressive imaging[C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2021: 16216-16225
- [23] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2021: 13713-13722
- [24] Wang Q L, Wu B G, Zhu P F, *et al.* ECA-Net: efficient channel attention for deep convolutional neural networks[C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2020: 11534-11542
- [25] Kulkarni S, Singh S, Balakrishnan D, *et al.* CrackSeg9k: a collection and benchmark for crack segmentation datasets and frameworks[C] // *Proceedings of the European Conference on Computer Vision*. Heidelberg: Springer, 2023: 179-195
- [26] Zou Q, Cao Y, Li Q Q, *et al.* CrackTree: automatic crack detection from pavement images[J]. *Pattern Recognition Letters*, 2012, 33(3): 227-238
- [27] Zou Q, Zhang Z, Li Q Q, *et al.* DeepCrack: learning hierarchical convolutional features for crack detection[J]. *IEEE Transactions on Image Processing*, 2019, 28(3): 1498-1512
- [28] König J, Jenkins M D, Mannion M, *et al.* Optimized deep encoder-decoder methods for crack segmentation[J]. *Digital Signal Processing*, 2021, 108: Article No.102907
- [29] Liu Y, Cheng M M, Hu X W, *et al.* Richer convolutional features for edge detection[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2017: 3000-3009
- [30] Chen L C, Zhu Y K, Papandreou G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation[C] // *Proceedings of the 15th European Conference on Computer Vision*. Heidelberg: Springer, 2018: 801-818
- [31] Caron M, Touvron H, Misra I, *et al.* Emerging properties in self-supervised vision transformers[C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2021: 9650-9660