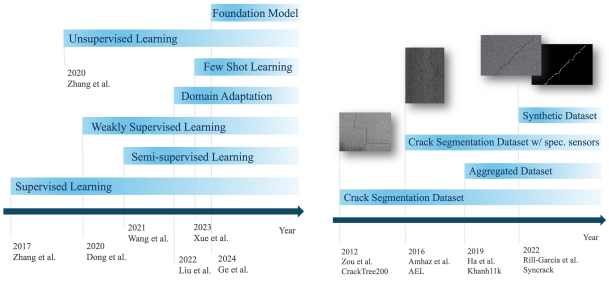
裂纹检测的深度学习：学习范式、普遍性和数据集的综述

Xinan Zhang[1] Haolin Wang[1] Yung-An Hsieh[1] Zhongyu Yang[1] Anthony Yezzi[1] Yi-Chang Tsai[1]
Georgia Institute of Technology[1]

Abstract

裂缝检测在土木基础设施中扮演着至关重要的角色，包括道路、建筑物等的检查，近年来深度学习在这一领域取得了显著进展。虽然在这一领域存在大量技术和综述论文，但新兴趋势正在重塑这一领域的格局。这些转变包括学习范式的转变（从完全监督学习到半监督、弱监督、无监督、小样本、领域自适应和微调基础模型）、泛化能力的提升（从单一数据集表现到跨数据集评估），以及数据集重新获取的多样化（从 RGB 图像到专业传感器数据）。在本次综述中，我们系统地分析了这些趋势并重点介绍了具有代表性的工作。此外，我们引入了一个通过 3D 激光扫描收集的新数据集，3DCrack，以支持未来的研究，并进行广泛的基准测试实验，为常用的深度学习方法，包括最新的基础模型，建立基准。我们的发现为基于深度学习的裂缝检测的演变方法和未来方向提供了见解。项目页面： https://github.com/nantonzhang/Awesome-Crack-Detection 。

1. 引言



(a) (b)

Figure 1. （a）基于深度学习的裂缝分割中新兴学习范式的时间线，以及（b）基于不同类别的裂缝分割的公开可用数据集的发布时间线。两者（a）和（b）都突出显示了首个代表性工作。

裂缝检测是土木基础设施中的一个关键任务 [? ? ?]，包括道路、建筑等。及时准确地识别裂缝可以进行主动维护，降低维修成本并防止严重的结构性故障。因此，裂缝检测一直是一个重要的研究领域，并在过去的几十年中经过了不同阶段的发展 [? ?]。

最初，裂缝检测依赖于人工调查，训练有素的检查员通过目测评估路面状况。虽然这种方法有效，但它劳动密集、耗时且容易产生主观错误。为了提高效率，研究人员开发了基于传统图像处理的方法 [? ? ?]，利用边缘检测、强度阈值和形态学操作等技术从图像中提取裂缝特征。但这些基于规则的方法往往难以应对光照、路面纹理和裂缝模式的变化。

机器学习（ML）的引入标志着一个重大的进步，因为 ML 模型可以从标记的数据集中学习特征表示，提高裂缝检测的准确性。然而，传统的 ML 方法仍然需要预先提取的初始特征，并且缺乏实际应用所需的可扩展性 [? ? ?]。随着深度学习（DL）的兴起，以卷积神经网络（CNN）和 transformers 的广泛应用为代表，裂缝检测进入了一个新纪元，实现了自动特征提取，并显著增强了检测性能 [? ?]。

深度学习已经通过多种方法应用于裂缝检测，包括分类、目标检测、语义分割等。分类模型判断一幅图像中是否包含裂缝，而目标检测方法通过在裂缝周围画出边界框来定位裂缝。然而，这两种方法都缺乏准确裂缝特征化所需的细粒度精度。在这些方法中，语义分割尤为突出，因为它能够精确地进行逐像素识别裂缝，使其成为详细裂缝识别和量化的理想选择。鉴于其重要性及在该领域的日益采用，本综述专注于基于深度学习的裂缝检测语义分割方法。如图所示，近年来该领域取得了快速进展，许多研究正在探索不同的学习范式和数据集，从监督学习到无监督学习；从追求单个数据集上的性能到可泛化性；从 RGB 相机数据到专业传感器数据，等等。

虽然已有几篇综述论文，如表 1 所示，但对最新趋势的系统分析仍急需。值得注意的是，本领域应抓住三个关键趋势以进行全面综述：

1. 学习范式的转变。正如图 3 所示，该领域的焦点已开始转向数据高效学习范式，如半监督学习，这引入了新的挑战和机遇。这些转变由日益增长的需求驱动，以克服有限标注数据、领域差距等挑战。分析包括监督学习（SL）、半监督学习（SSL）、弱监督学习（WSL）、无监督学习（USL）、少样本学习（FSL）、领域适应（DA）和参数高效微调（PEFT）基础模型在内的不同学习范式中的这些发展，可以为特征研究和实施提供明确而独特的指导。这些转变反映了更广泛的计算机视觉领域的趋势，同时融

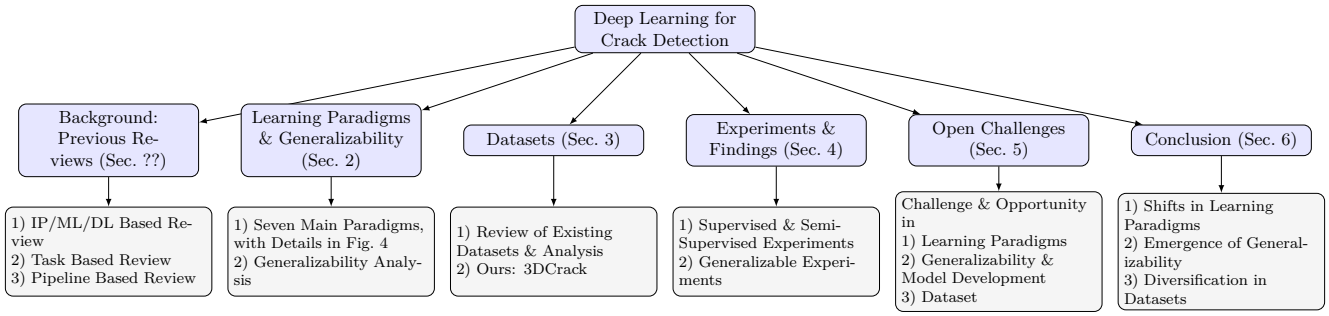


Figure 2. 本综述论文的结构。

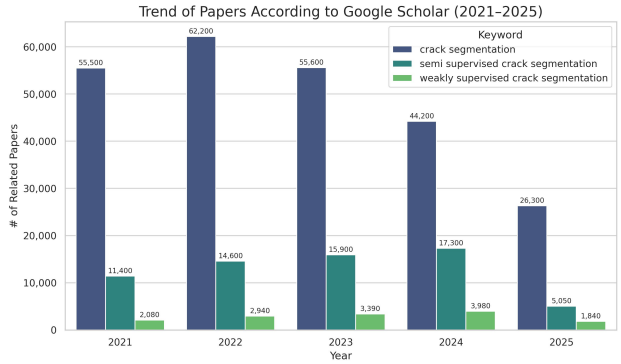


Figure 3. 截至 2025 年 6 月 3 日，通过 Google Scholar 使用不同关键词的最近五年相关论文趋势。虽然与裂缝分割相关的论文总数在 2022 年后有所下降，但半监督和弱监督学习等非完全监督学习范式在该领域受到越来越多的关注。

入了针对具体特征的设计。

2. 普适性的新兴。裂纹检测模型传统上是在单一数据集上进行训练和评估的，这限制了它们在现实世界中的零样本适用性。伴随学习范式和数据集开发的趋势，最近已经做出了努力，增强跨多个数据集的普适性，这方面的内容应被纳入系统评审，以洞察该领域的前沿。
3. 数据集的多样化。早期的工作使用标准 RGB 图像进行裂缝分割，但该领域已经开始采用基于专业传感器的数据集（如 3D 激光扫描），这些数据集在各种条件下为裂缝检测提供了更可靠和稳健的成像。聚合数据集和合成数据集的可用性也为基于深度学习的裂缝分割创造了新的机会。分析不同数据集的特征有助于弥合研究实践与现实世界应用之间的差距。

因此，在本次综述中，我们通过分析最近的代表性工作，提供针对各类别的学习范式及其泛化能力的结构化分析来解析这些趋势。此外，对当前可用的数据集也进行了系统性审查。我们还介绍了一个新的高分辨率、大规模、工业级裂缝检测数据集，旨在弥合研究与实际应用之间的差距。我们对不同学习环境中的广泛使用的深度学习模型（包括近期的基础模型）进行了基准测试，并获得了一些独特的发现。我们的目标是提供有关基于深度学习的裂缝检测的全面、最新的观点，指导这一快速发展的领域的未来研究。本文的结构如图 Fig. 2 所示。最后，我们的综述论文的贡献如下：

1. 对上述三种趋势进行了全面研究，包括对最近的代表性工作进行回顾以及对学习范式进行类别特定的结构化分析，讨论了每一类别在裂缝检测领域的优势和泛化能力。
2. 为了促进这一系统研究和未来的研究，我们引入了一个新的高分辨率、大规模、工业级数据集 3DCrack 用于裂缝语义分割，与类似的数据集相比，它具有更高的分辨率和更为多样化的路面状况。
3. 我们总结了当前的挑战和未来研究与实施的独特机会，并对代表性的基线模型进行了基准测试。

正如表 1 所示，早期关于裂纹检测领域的综述文章主要集中在图像处理（IP）技术上。这些方法因其简单性和可解释性而受到青睐，通常依赖于手工设计的特征，如边缘检测、阈值化和纹理分析。随着计算资源和标注数据的增加，该领域见证了向机器学习（ML）和深度学习（DL）方法的转变。这种演变在综述文献中得到了体现，从以 IP 为中心的调查转变为以数据驱动和学习为基础的最新研究。这一进展反映了技术的进步以及对可扩展、精确和自动化裂纹检测系统日益增长的需求。

最近的评论采取了多种视角。一些评论基于其基本方法对方法进行分类，例如图像分类（确定裂缝的存在性）、对象检测（使用边界框定位裂缝）、语义分割（为裂缝区域分配像素级标签） [? ? ?] 。其他评论则采用以流程为导向的视角，涵盖从数据采集（例如无人机、智能手机或激光传感器）到裂缝识别、量化和决策支持的整个工作流程 [? ? ?] 。第三种评论强调计算方面，包括网络架构设计、特征表示和模型优化策略 [? ?] 。

然而，目前尚未有现有的综述能够在一个层次结构中全面分析三个关键趋势——学习范式的转变、泛化能力的出现、数据集的多样化。本论文旨在填补这一空白，通过提供如图 2 & 4 所示的深入综述，并对这三个焦点领域进行讨论，这些领域是以深度学习为基础的裂缝检测发展的基础。

Review Paper	Description	Learning Paradigms
Mohan et al. (2018) [?]	Literature presents different techniques to automatically identify the crack and its depth using image processing techniques.	None (IP Only)
Wang et al. (2019) [?]	A number of literature in this research topic are collected for summarizing the research artwork status, and giving a review of the pavement crack image acquisition methods and 2D crack extraction algorithms.	SL
Cao et al. (2020) [?]	This paper reviews the three major types of methods used in road cracks detection: image processing, machine learning, and 3D imaging-based methods.	SL
Hsieh et al. (2020) [?]	Authors organize and provide up-to-date information on ML-based crack detection algorithms for researchers to more efficiently seek potential focus and direction.	SL
Munawar et al. (2021) [?]	This paper provides a review of image-based crack detection techniques that implement image processing and/or machine learning.	SL
Ali et al. (2022) [?]	A review of CNN implementation on civil structure crack detection, highlighting significant research for crack detection through classification and segmentation.	SL
Hamishebahar et al. (2022) [?]	A comprehensive literature review of deep learning-based crack detection studies and the contributions they have made to the field is presented.	SL
Li et al. (2022) [?]	It presents a comprehensive thematic survey of DL-based CIS techniques. Our review offers several contributions to the CIS area.	SL, WSL, USL
Kheradmandi et al. (2022) [?]	This literature review focuses heavily on three major types of approaches in the field of image segmentation, namely thresholding-based, edge-based, and data-driven-based methods.	SL
Zhou et al. (2023) [?]	It reviews recent developments in deep learning-based methods for crack segmentation and investigates the impact from different image types.	SL
Nguyen et al. (2023) [?]	This study has identified the bigger picture of DL methods for crack identification in asphalt pavement. The authors evaluated several DL-based crack identification algorithms from the literature.	SL, USL
Ai et al. (2023) [?]	This paper provides a comprehensive review of the research progress and prospects in computer vision frameworks for crack detection of civil infrastructures from multiple materials, including asphalt, concrete, and metal-like materials.	SL
Yuan et al. (2024) [?]	Based on the main research methods of the 120 documents, we classify them into three crack detection methods: fusion of traditional methods and deep learning, multimodal data fusion, and semantic image understanding.	SL
Xu et al. (2024) [?]	This article systematically summarizes recent advances in FSL algorithms and the corresponding applications in SHD for civil infrastructure.	FSL
Amirkhani et al. (2024) [?]	This survey aims to study the recent progress of vision-based concrete bridge defect classification and detection in the deep learning era.	SL, SSL, WSL, USL
Ours (2025)	DL for crack segmentation is the focus of this survey, with trends in learning paradigm, generalizability, and dataset comprehensively reviewed and categorized. A professional dataset, 3DCrack, is released.	SL, SSL, WSL, DA, FSL, USL, FM

Table 1. 与裂缝分割相关的综述论文摘要。SL、SSL、WSL、DA、FSL、USL 和 FM 分别代表监督学习、半监督学习、弱监督学习、领域适应、少样本学习、无监督学习和基础模型。

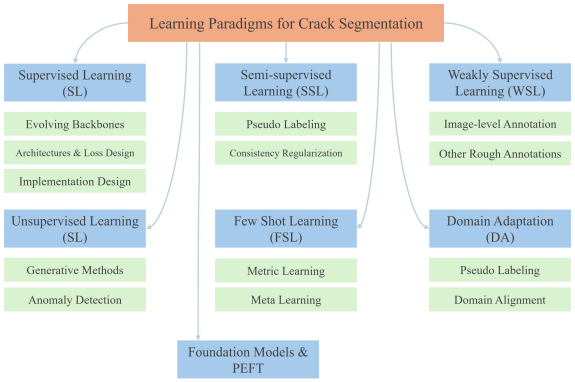


Figure 4. 基于 2 节中的学习范式分类进行审查。绿色框表示每个蓝色框中每个学习范式下的子组。

2. 学习范式 & 的泛化性

通过分割进行裂缝检测本质上是计算机视觉领域的一项语义分割任务。其主要目标是将输入图像划分为具有不同语义意义的区域——在本例中，是在像素层面上区分裂缝区域和非裂缝区域。与传统的分类或目标检测不同，语义分割提供细粒度的、逐像素的预测，这使得其特别适合于基础设施检测任务，如裂缝检测，其中几何精确检测至关重要。

随着深度学习的出现，这一领域见证了快速进展。卷积神经网络（CNN）以及最近基于 transformer 的模型被广泛采用，以捕捉多种环境下裂缝的复杂空间模式和结构变化。这些基于深度学习的方法在准确性和鲁棒性方面显著超越了传统图像处理技术 [? ? ?] 。

伴随着这一进展，各种学习范式相继出现，每一种都解决了不同的实际挑战，包括提高性能、提升泛化能力、处理有限标记数据等。这些范式包括监督学习，其中模型在像素级完全标注的数据集上进行训练；弱监督方法，通过使用更粗糙的标签如涂鸦或图像级标签来减少注释负担；半监督方法，利用标记和未标记数据；领域自适应，其目的是在训练和测试数据集属于不同数据分布的不同领域中自适应模型；小样本学习，使得模型能够在只有少量标记样本的情况下推广到新任务；以及无监督学习，发掘完全未标记数据的潜在结构。此外，受到更广泛的计算机视觉社区中泛化能力最新进展的启发，如对比语言-图像预训练（CLIP） [?] 和 Segment Anything Model（SAM） [?] ，基于基础模型和参数高效微调（PEFT）技术 [?] 的可泛化方法已开始在裂缝分割中出现，这些技术是为大模型的微调而设计的。

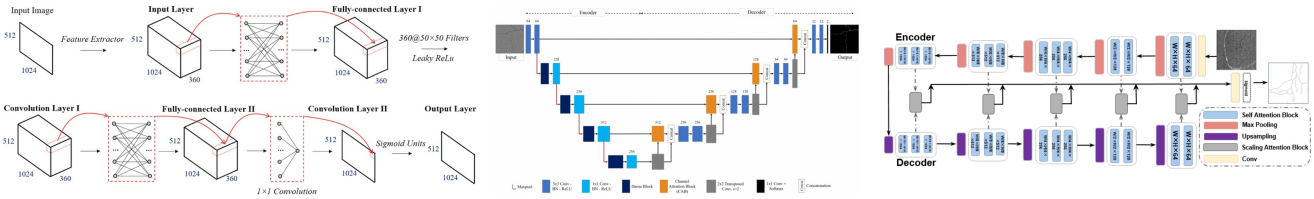
鉴于这些方法的日益普及和多样性，本节提供了一个关于基于深度学习的裂缝分割主要学习范式的结构化综述。同时，还讨论了每种范式的优势，特别是关于其推广潜力的优势。

在全监督裂缝分割中，问题可以如下表述。给定一个完全像素级标注的数据集 $\mathcal{D}_L = \{(x_1, y_1), \ldots, (x_l, y_l)\}$ ，目标是 $\min_\theta \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(f_\theta(x_i), y_i)$ ，其中 f_θ 是学习到的分割网络，旨在分割裂缝区域。

该领域近年来取得了显著进展。这些进展可以分为三个关键阶段，包括 1) 更强的骨干网络，2) 更复杂的结构设置和损失设计以提高性能，以及 3) 特定的现场实施设计。

演变中的主干网络。在基于深度学习的裂缝分割中，模型主干选择经历了从 CNN 到 Transformer 的转变。最初，浅层的 CNN，通常具有较少的层数和较简单的结构，被用作主干 [? ? ?] 。这些早期的基于 CNN 的模型已经显示出优于传统机器学习和图像处理技术的性能，这主要是因为它们能够直接从原始图像或预处理的初步特征中学习特征表示，如图 5 (a) 所示。随着该领域的进步和更多样化数据集的出现，如表 2 & 3 所示，研究人员越来越多地采用更深层次和更强大的 CNN 架构。ResNet、Xception 等主干网络 [? ? ? ?] 成为裂缝分割模型中的标准主干网络 [? ?] 。

注意力机制后来被整合到基于 CNN 的架构中，以进一步优化特征表示并增强定位能力 [? ? ? ? ?] 。例如，DAU-Net [?] 引入了一个通道注意力块，该块减少了噪声激活，并从编码器中强调显著特征，从而实现更精确的裂缝分割。类似的设计也已被提出 [? ? ? ?] ，证实了注意力在引导模型关注裂缝区域方面的有效性。



(a) (b) (c)

Figure 5. 关于骨干网络和架构，监督学习在裂缝分割中的演变。从 (a) CrackNet [?] 使用最基本的 CNN，到 (b) DAUNet [?] 利用编码器-解码器架构，再到 (c) CrackFormer [?] 应用变压器，已经在更深的骨干网络和更复杂的架构中见证了演变。

同时，基于 transformer 的骨干网也被引入了领域 [? ? ? ? ? ? ?] 。与 CNNs 不同的是，CNNs 由于其基于核的操作倾向于局部特征，而 transformers 通过注意机制具有捕捉长距离依赖和全局上下文的优势 [? ?] 。这种能力对于裂纹分割来说特别有价值，因为图像中远距离部分的上下文信息可以帮助区分复杂的情况，通过区分裂纹和非裂纹特征或其他噪声模式 [? ? ?] 。

复杂的架构和损失设计。与主干网络的演进同时进行的，还有架构设计的发展。早期基于 CNN 的方法通常使用简单的、顺序堆叠的卷积层 [? ?] (图 5 (a))。

随着时间的推移，编码器-解码器结构已成为主流 [? ? ? ?] ，如图 5 (b) & (c) 所示。这些结构受 U-Net [?] 的启发，被发现对裂缝分割任务非常有效。编码器通过连续的下采样来提取高级特征表示，而解码器则逐步上采样并重建空间分辨率以生成详细的分割掩码。这种架构通常会进一步适应 [? ? ? ? ?] 或与特定损失设计相结合 [? ? ? ?] 用于裂缝分割，使模型能够同时保留低级空间细节和高级语义上下文，这对于准确检测容易被忽视的狭窄和断开的裂缝至关重要。

实现导向设计。目前越来越强调适用于现实场景部署的模型。这包括设计注重效率和准确性的轻量级网络 [? ? ? ?] 。例如，SDDNet [?] 采用密集连接的可分离卷积，在显著减少参数数量的情况下实现了具有竞争力的性能，比同类模型的参数数量最多减少 88 倍，使其对于实时现场应用具吸引力。最近的一项研究，SCSegamba [?] ，引入了一个基于状态空间模型 Mamba [? ?] 的结构感知视觉网络，在轻量化设计下实现了最先进的结果。

另一个实际考虑因素是适应多种数据采集平台，例如无人机（UAV） [? ?] 。在 [?] 中，例如，提出了一种改进的相机校准方法，用于计算安装在无人机云台相机的全场尺度。该方法允许在不同飞行姿态间方便地索引比例因子，消除了重复校准的需要并有助于在不同场景中实现一致的测量。

作为基于深度学习的裂缝检测中的基础范式，监督学习推动了该领域的初步进展。其发展，包括在骨干网络和架构设计等领域的进步，奠定了坚实的基础，并为替代学习范式提供了宝贵的参考。然而，其主要限制之一在于对大量标记数据的高度依赖，这些数据的获取既昂贵又耗时 [? ?] 。因此，业界越来越倾向于探索更加数据高效的方法，如半监督学习、弱监督学习、领域适应和少样本学习。这些方法旨在利用监督学习的优势，同时减少其对广泛手动标注的依赖。

2.1. 半监督学习

裂缝分割中的半监督学习（SSL）开始受到关注，因为研究人员希望在保持高分割性能的同时，减少对昂贵的像素级标注数量的依赖。

在 SSL 中，有一个包含全部标注样本的子集 $\mathcal{D}_L = \{(x_1, y_1), \ldots, (x_l, y_l)\}$ ，与另一个包含未标记样本的子集 $\mathcal{D}_U = \{x_{l+1}, \ldots, x_n\}$ 成对出现。通常 $l \ll n$ ，这意味着未注释的数据构成了整个数据集的主要部分。目标是利用这两组数据来训练一个模型，其性能优于仅使用标记数据训练的模型。这在现实世界的应用中特别重要，因为人工标注裂缝既耗时又费力。如图 6 所示，该领域常用的方法可以分为两大趋势：1）伪标记；2）一致性正则化。值得注意的是，这些单一方法之上的另一个显著趋势是它们的结合，以实现性能改进。

伪标签。伪标签，又称为自训练，是半监督学习中广泛采用的一种技术 [? ? ? ?] ，在裂缝检测任务中表现出了显著的潜力 [?] 。其核心思想是利用一小部分标记的裂缝图像来训练一个初始的监督模型，然后用该模型对未标记数据进行预测。基于如置信水平 [?] 、判别网络的决策 [?] 等精心选择的标准，对满足要求的预测将被选为伪标签，并被纳入训练集，作为额外的标记数据进行进一步的微调 [? ?] 。

例如，作者们在 [?] 中采用这种伪标签策略，通过利用一个改进的 U-Net [?] 架构作为核心分割网络进行裂缝分割。训练过程以监督方式开始，模型从一小部分标注的裂缝图像中学习。一旦初始模型训练完毕，它就会被用来在更大规模的未标注数据集上生成预测。预测的概率被视为置信评分，只有那些评分高于某一阈值的预测结果才会被添加到原始训练集中。在经过迭代地对扩充数据集进行重新训练后，该半监督模型在仅使用 40 % 标注数据的情况下与监督模型相比达到了竞争力的准确率。

在 [?] 中，作者首先以完全监督的方式训练一个分割网络，并训练一个判别器来区分真实值与预测值。然后他们使用训练好的分割网络在未标记的数据上进行预测。训练好的判别器可以提供一个置信度评分，该评分可以决定预测是否良好。如果是的话，就会在交叉熵损失中使用。类似地，在 [?] 中，作者应用对抗学习的理念来训练一个判别器，为未标记数据的预测提供一

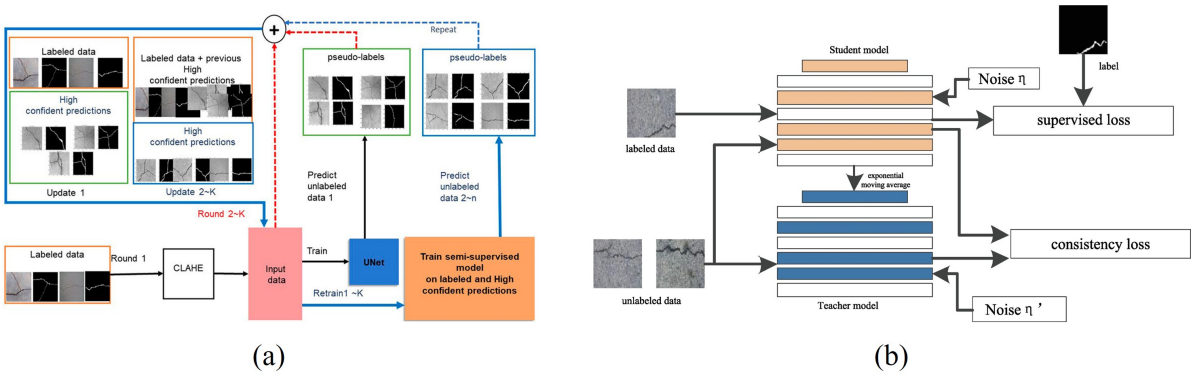


Figure 6. 半监督裂缝分割的代表性工作，(a) 用于伪标签 [?]，(b) 用于一致性正则化 [?]。

个置信度图。他们进一步使用多尺度特征来增强性能。

另一组工作探索结合多种方法的混合方法，包括伪标签，这将在下面的组合方法中讨论。

一致性正则化。一致性正则化是半监督学习中的一个基本原则，由于其简单和有效性，在裂缝检测任务中得到了广泛应用 [? ?]。其核心思想是，当输入图像受到小扰动或者采用不同模型以实现稳健性能追求相同的预测目标时，模型的预测应该保持稳定或一致。

在 [?] 中，作者使用 EfficentNet [?] 来提取多尺度裂纹特征信息以进行裂纹分割。两个模型被初始化为具有相同结构的学生网络和教师网络。学生网络中的权重直接使用梯度下降进行更新，而教师网络则通过学生模型的指数移动平均权重进行更新。此外，噪声被添加到输入数据中，学生和教师网络通过输出的一致性损失进行正则化。

在 [?] 中，它集成了分形维度分析 [?]，这是一种用于表征结构复杂性如何在不同尺度变化的方法，以及半监督学习。他们首先使用分形维度分析初步确定候选裂缝区域，然后使用裂缝相似性学习网络（CrackSL-Net）来像 SimCLR [?] 那样学习裂缝图像区域的语义相似性，从而在半监督环境中增强鲁棒性和准确性。

结合方法。另一种方法是结合伪标签和一致性正则化，以进一步提高性能 [? ?]。

在 [?] 中，作者提出了一种交叉教师伪监督框架和交叉增强策略。该方法使用两对教师-学生模型通过各自教师模型生成的伪标签相互监督。为了提高所提算法的性能，在训练过程中对输入、特征和网络应用扰动。在 [?] 中，训练过程分为三个连续阶段：1）在有标签数据上的监督训练，2）在输入扰动下对齐预测的一致性正则化，以及 3）基于预测确定性生成伪标签。

半监督学习提供了一种有效的解决方案，能够减少对大规模完全标注数据集的依赖。通过在训练过程中利用标注数据和未标注数据，这种方法在减少标注负担的同时提升了模型性能。在裂缝分割的背景下，半监督学习相比于监督学习表现出了具有竞争力的结果 [? ? ?]，有效地弥合了有限标注数据与高性能模型需求之间的差距。

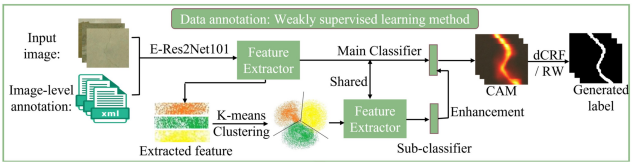


Figure 7. 一项具有代表性的工作是使用图像级弱监督 [?]。作者利用表示裂缝存在或不存在的二值图像级标签来指导弱监督过程。

2.2. 弱监督学习

近年来，弱监督学习作为全监督方法的另一种数据高效替代方案，受到了越来越多的关注。弱监督裂缝分割旨在通过使用较弱形式的标注（包括图像级标签、边界框、涂鸦（粗线）标注等）训练模型来缓解这一困难。图 7 展示了一个具有图像级弱监督的代表性工作。与半监督学习不同，弱监督学习通常依赖于粗略的标注来减轻标注负担，而不是使用数量较少的精确标注。

图像级标注。图像级标注是裂缝检测和其他领域最常用的弱监督信号之一，主要是因为相比于密集的像素级标签，它们更易于获取。这些标注通常与基于类激活映射（CAM）的方法结合使用，后者旨在通过使用梯度统计的已训练模型来定位与感兴趣类别相关的判别区域。换句话说，CAM 有助于可视化当模型做出决策时图像的哪些区域受到关注。诸如 GradCAM 及其他技术构成了该领域中许多弱监督方法的基础。

在 [?] 中，作者利用指示裂缝存在或不存在的二进制图像级标签来引导弱监督过程。他们介绍了一种名为 Crack-CAM 的方法，该方法首先使用 CAM 生成粗略的像素级图，然后利用密集条件随机场 (DenseCRF) [?] 和随机游走 (RW) 算法 [?] 对这些图进行细化。细化后的标签随后被用来以完全监督的方式训练分割模型，有效地弥合了弱监督和完全监督之间的差距。

基于一个相似的理念，Huangfu 等人提出了一种全面的改进策略，称为 AT-CAM（仿射变换和伪标签细化）。他们的方法将 XGradCAM 与通过仿射变换的几何数据增强相结合。它进一步结合了双钩去噪和动态

范围压缩机制，以抑制激活图中的噪声并强调裂缝的结构特征。与基线相比，该框架在分割准确性上实现了 7.2 % 的相对改进。

在 [?] 中引入了一种替代策略，作者通过对图像块而非整个图像或像素进行标注来降低标注成本。一个分类模型在这些块级标签上进行训练，而生成的 CAM 用于推断像素级监督。为了提高生成标签的空间一致性，应用了 DenseCRF [?] 。这种基于块的标注策略显著减少了标注负担，大约降低了 80 % ，同时仍能够训练出有效的分割模型。

像 [? ? ? ? ? ?] 这样的作品也将图像级或区域级标注的想法应用到他们的像素级分割任务中。这些方法共同展示了将基于 CAM 的定位、细化算法和最小监督相结合以实现高质量裂缝分割的潜力，即使在缺乏详细像素级标注的情况下。

其他粗略标注。除了图像和补丁级别的标注外，框选监督和诸如涂鸦和关键点这类粗略的像素级标注也提供了一种平衡标注工作量和真实值指导的中间形式监督，正因如此，裂缝检测和其他任务 [? ? ? ?] 开始采用这种弱标注形式。这些弱监督形式使得可以利用额外的先验知识或图像处理策略推导出像素级标签。

在 [? ?] 中，作者研究了基于边界框的裂缝分割监督。在 [?] 中，为了将粗略的边界框转化为像素级的伪标签，他们采用了一个多步骤的细化流程。首先，抑制背景噪声，并识别框内的高置信种子区域。接下来，应用区域增长算法扩展种子区域并生成粗略的分割图。最后，使用 GrabCut 算法 [?] 迭代细化分割，从而逐步改进伪标签。该方法有效地弥合了边界框标注和全监督之间的差距，而无需昂贵的逐像素标注。

像 [? ?] 这样的工作则使用粗略标注。Tang 等人提出了一种基于粗略标注裂缝和非裂缝区域的方法。该方法不依赖于精确的轮廓，而是利用近似的标注来提取三种不同类型的监督信号：1）裂缝像素标签，表示必须被分类为裂缝的像素；2）非裂缝像素标签，指示必须被标记为背景的像素；3）孤立像素掩码，它编码了不确定但信息丰富的区域，这些区域可能提供额外的上下文线索。这些标签是使用经典的图像处理技术提取的，从而使网络可以同时使用强监督和弱监督信号进行训练。这种混合方法有效地平衡了标注工作量与分割性能。

通过减少对繁琐人工标注的依赖，弱监督裂缝分割促进了分割模型在标注资源有限的真实应用中的部署，并为利用最初未标注的图像集合开辟了新的机遇。然而，实现可与全监督方法相媲美的性能仍然是一个重大的研究挑战，这推动了设计更有效的学习框架、损失函数和专为弱监督量身定制的正则化策略的持续努力。[? ? ?] 。

2.3. 领域适应

裂缝分割中的领域自适应趋势变得越来越显著，这驱动着模型需求，使其能够跨越不同的成像条件、路面类型和传感器设置进行泛化，而无需针对每个新领域进行广泛的重新训练或标注。在这种情况下，源领域数据集是完全标注的，而目标领域通常很少或没有标注。从不同区域收集的裂缝图像，由于光照条件、材料和摄像头设置的变化，往往会表现出显著的外观差异，使领域转换成为一个关键的挑战。然而，这种情况在现实世界中是非常实际的，因为由于不同的条件，实际数据与标注数据有所不同，因此它是一个值得探索的有意义方向。类似于 SSL，已经出现了两大类方法：1）伪标签和 2）领域对齐，如图 8 所示。

伪标签。它已被广泛采用为裂缝检测和其他语义分割任务中的领域适应策略 [? ? ?] 。在文献中也被称为自训练 [?] 。其核心思想是通过利用从源领域学习到的信息，为目标领域生成合理的伪标签，从而使分割网络能够适应并在未见过的目标领域数据上表现更好，而无需真实标注。这个机制促进了知识的转移，并减轻了通常在不同数据集之间困扰泛化能力的领域转换问题。

在 [?] 中，作者展示了基于伪标签的方法在裂缝检测的无监督领域适应中是有效的。此外，通过在 DAFormer 框架中结合掩码图像一致性 [?] ，获得了额外的性能提升，这强调了在适应新领域时一致性正则化的重要性。同样，STDASeg [?] 利用基于图像混合的领域混合模块实现伪标签，作为 CNN 和 Transformer 之间互学习方案的自训练流程。在 [?] 中探索了一个相关概念，作者提出了多源集合标签来辅助领域适应。其他研究则通过使用领域对齐技术增强这一机制 [? ?] ，如下所述。

领域对齐。领域对齐是裂缝检测和分割领域适应中另一种广泛采用的技术。其核心思想是通过在不同层面对齐源域和目标域来减少领域间隙，通常是在输入空间、特征空间或输出空间进行对齐，从而使在源域上训练的模型能够更好地泛化到目标域，而不受领域间分布变化的严重影响。

在 [?] 中，作者提出了 DACrack，一个综合框架，在输入、特征和输出三个层面进行领域对齐。为了处理输入层面的差异，他们设计了前景增强模块（FEM），该模块在源域和目标域之间交换低级视觉特征。这种交换过程鼓励模型关注与裂缝相关的结构相似性，同时尽量减少来自无关纹理的干扰，有效地作为一种对比性注意。在特征层面，特征适应模块（FAM）通过对抗学习来对齐跨域的中间特征分布，帮助模型学习领域不变的表示。最后，在输出空间，他们引入输出适应模块（OAM），利用变分自编码器（VAE）[?] 来建模和对齐裂缝的聚类流形结构，确保输出分布在各个领域之间保持一致。

除了像素外观，诸如深度等辅助信号也被探索用于弥合领域差距。在 [?] 中，作者提出使用深度预测中的视差作为领域差异的新指标。通过估计源域和目标域之间预测深度图的不一致性，他们在优化过程中引入了一个深度引导的正则化项。这种方法强调了利用超越 RGB 图像的附加模式的潜力，表明通过整合更丰富、互补的信息可以加强领域适应性。

生成对抗网络（GANs）[?] 在领域适应任务中也找到了广泛的应用。GANs 旨在通过对抗训练学习一个

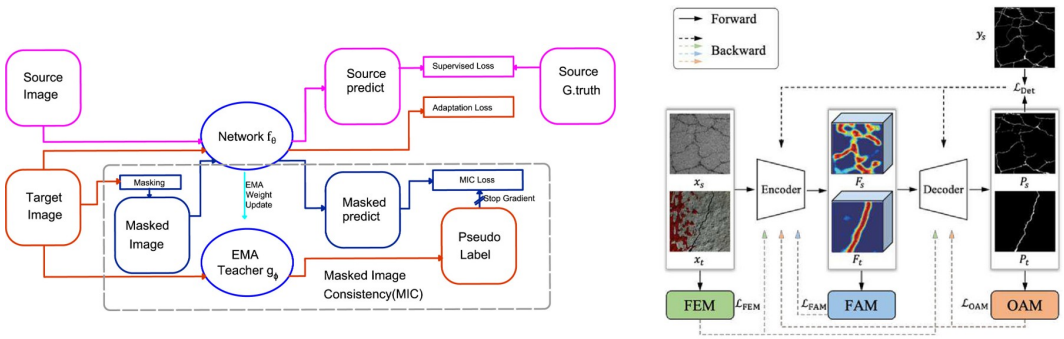


Figure 8. 基于领域自适应的裂纹分割中的代表性工作，(a) 用于伪标签 [?]，(b) 用于领域对齐 [?]。

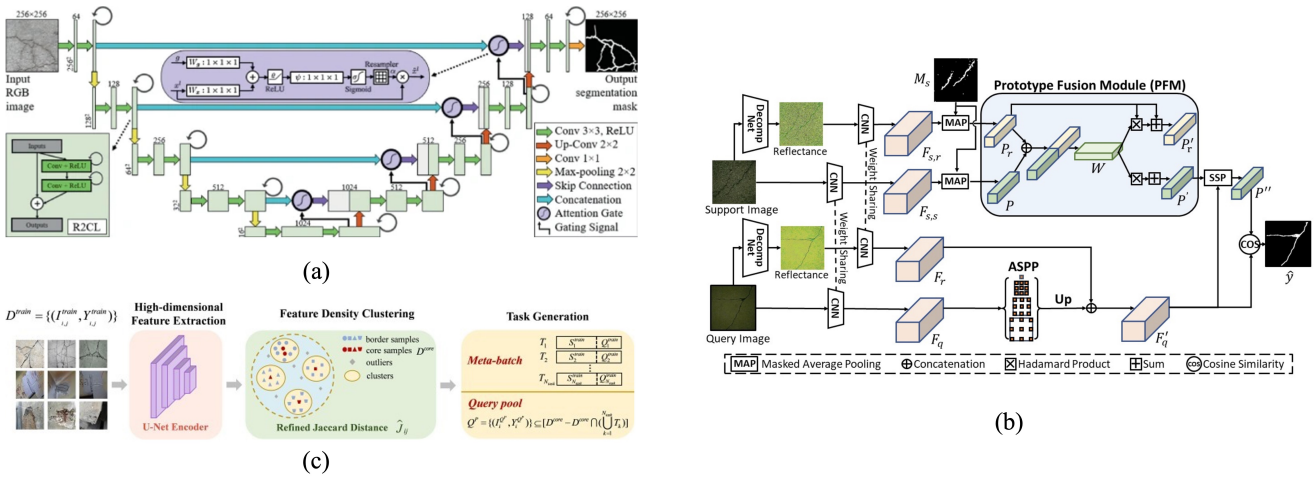


Figure 9. 基于小样本学习的裂缝分割中具有代表性的工作，(a) 用于迁移学习 [?]，(b) 用于度量学习 [?]，(c) 用于元学习 [?]。

生成模型，可以在不同的数据分布之间进行映射。一个值得注意的子类，CycleGAN [?]，使图像到图像的翻译无需配对样本，在无配对领域适应中特别有价值。例如，在 [?] 中，CycleGAN 被用于合并无阴影和有阴影数据集的特征，有效地创建了一个具有领域不变特征的适应数据集，以利于在不同光照条件下的裂缝检测。同样地，[?] 使用 CycleGAN 进行特征级对齐，而不需要目标领域中的注释。他们的研究调查了转移方向，发现双向转移，即将源和目标特征都映射到一个共享的辅助空间，能够更好地实现独立同分布 (i.i.d.) 结构，减轻过度适应的风险，并在整个层级网络层中保持关键特征。

总的来说，这些工作突出了领域对齐技术日益复杂化的发展，从简单的对抗性特征匹配演变为多层次、多模态和结构感知的适应策略。

2.4. 少样本学习

小样本学习是另一种学习范式，其中模型被训练以仅使用少量标记样本来推广到新任务。这在裂缝检测中尤其有用，因为收集大规模标注数据是困难或昂贵的 [? ?]。一个典型的问题形式化是 N 类 K 样本，其中模型任务是给定每个类别仅 K 个样本时执行 N 类分类或分割。在该领域中，代表性工作可以分为三类：1) 迁移学习，2) 度量学习，和 3) 元学习，如图所示 9。

迁移学习。在裂缝分割的背景下，由于像素级标注成本高昂且费力，少样本学习正逐渐受到关注。迁移学习一直是基于深度学习的裂缝检测中广泛使用的技术 [? ?]。在少样本裂缝检测的背景下，可以在一个不同但相关领域的大型数据集上预训练模型，并在一个小型裂缝数据集上进行微调 [? ?]。例如，R2AU-Net [?] 利用迁移学习和用户反馈进行少样本适配。它通过使用手动修正的示例进行增量微调动态优化模型权重。

度量学习。为了更好地利用少样本设置，基于度量的方法受到关注。这些方法使用距离度量 [? ? ?] 比

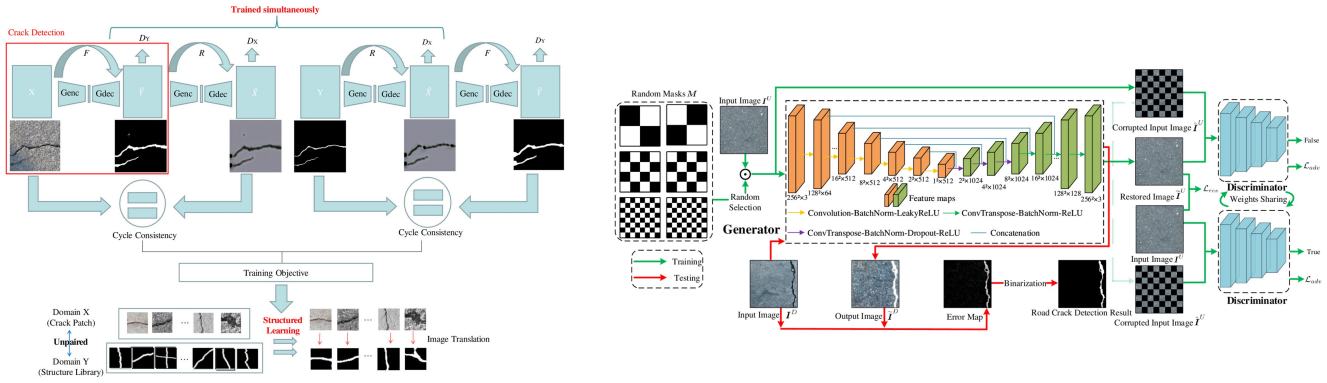


Figure 10. 基于无监督学习的裂缝分割的代表性工作，(a) 通过生成进行检测 [?]，(b) 用于异常检测 [?]

较支持集和查询集之间的特征表示。在裂缝分割中，这通过从两个集合中提取特征体积并通过相似性比较进行分割来实现。例如，CrackNex [?] 引入了一个基于原型的框架，其中提取并融合了支持和反射率原型。最终的分割是通过查询特征与这些原型之间的余弦相似性来获得的 [? ?] 。

元学习。另一方面，元学习，通常被描述为“学习去学习”，其目的是训练可以以最少的数据快速适应新任务的模型。在裂缝分割中，元学习框架通常涉及在多个分割任务上进行阶段性训练，以模拟小样本条件。通过这种方式，模型学习到一套通用的策略，可以快速适应新的裂缝分割问题，即使只提供非常有限的标注数据 [? ?] 。这使得模型可以获得适用于裂缝分割的一般化适应策略 [? ? ?] 。例如，在 [?] 中，作者提出了一种基于特征密度聚类的任务生成方法，在应用无关模型元学习（MAML）之前，选择核心样本形成查询池 [?] ，这提高了元训练的可解释性和有效性。

在裂缝检测中应用小样本学习有助于减少对大型训练数据集的依赖，同时仍能保持合理的准确性和鲁棒性，使其成为实际和可扩展的解决方案，适用于现实场景。

2.5. 无监督学习

无监督学习旨在探索大量未标记的数据，以发现潜在模式、底层结构或有意义的表示，而不依赖于明确的注释或人为提供的标签。在裂缝分割的背景下，无监督方法尤其具有吸引力，因为裂缝的人工标记是劳动密集型的、主观的，并且可能容易出现不一致。目前已经出现了两大类方法，即 1) 通过生成检测和 2) 异常检测，如图 10 所示。

早期的研究表明，在不使用深度学习的情况下进行无监督的裂缝分割是可行的。它们通常依赖聚类方法 [? ?] 或传统的图像处理技术 [?] 来区分裂缝区域和非裂缝区域。

通过生成进行检测。在基于深度学习的无监督裂缝分割中，生成模型被探索用于学习输入裂缝图像与分割图之间的逐像素翻译，仅用非配对数据集进行训练，其中裂缝图像并未与专用裂缝图对齐 [? ?] 。在 [?] 中，网络由两个 GAN 组成：一个学习将裂缝图像片段翻译为表示裂缝的结构化曲线，另一个学习进行反向翻译。两个 GAN 同时训练，并通过循环一致性进行正则化，类似于 CycleGAN [?] 。已经表明，该双网络可以被训练以将破裂的图像翻译为具有相似结构模式的类似于真实值（GT）的图像，这可以直接用于裂缝检测。

异常检测。另一方面，一些方法如 [? ?] 采用异常检测的方法，其中裂缝区域是要检测的异常。例如，UP-CrackNet [?] 首先基于未损坏的路面图像准备随机遮蔽的训练图像，然后训练一个生成网络，在遮蔽图像的条件下，用完好的路面模式恢复缺失的区域。在推理过程中，输入图像同样被一组多尺度的方形遮罩遮蔽，通过计算输入图像与其重建对应物之间的逐像素差异创建错误图。这张错误图突显了裂缝，因为生成网络已经学会仅恢复未损坏的区域。

无监督学习是裂缝分割的另一种实际方法，特别是因为它完全消除了对人工标注的需求。然而，请注意，即使去除了标注的负担，无监督学习仍然需要经过精心准备的数据。随着研究的进展，无监督学习继续作为一个有前景的途径向可扩展和无标注的裂缝检测发展。

2.6. 基础模型 & PEFT

在计算机视觉的广泛领域中，泛化性正成为一个越来越重要的话题，并且最近在裂缝分割领域引起了显著关注。泛化的目标是在开发模型时，不仅要在与训练集相似的数据上具有高准确性，还要在面对多样的真实世界条件时具备鲁棒性，如不同的路面纹理、光照变化、成像设备和地理位置等。这种日益增长的关注是由于意识到在狭窄或同质数据集上训练的深度学习模型在应用于未见过的环境时往往会出现显著的性能下降。因此，研究人员开始有意识地设计以泛化性为关键目标的裂缝分割方法，这表明一种从纯粹以性能为导向的模型转向优先考虑鲁棒性和适应性的模型的范式转

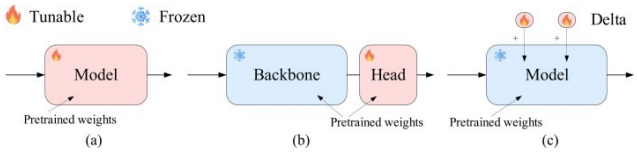


Figure 11. 最近一项关于裂缝分割泛化的代表性工作 [?] 是基于参数高效微调 SAM [?] 。

变。

这一转变中的一个显著影响因素是由 Segment Anything Model (SAM) [?] 代表的视觉基础模型的出现。SAM 是在大规模数据集 SA-1B（由 1100 万张图像和 11 亿个掩码组成）上训练的，能够以零样本的方式进行通用的图像分割。SAM 已被应用并进一步微调于一些专业领域，并展示了成功，例如医学图像分割 [? ? ?] ，以及裂缝检测。在 [?] 中，作者提出了 CrackSAM，在其中应用了两种参数高效的微调方法来微调 SAM：适配器和低秩适配（LoRA） [?] ，如图 11 所示。同样地， [?] 采用 LoRA 对 SAM 进行微调，证明了与传统裂缝分割模型相比，这一策略可以显著提高八个不同裂缝数据集的分割准确性。

在此趋势的基础上，[?] 引入了 MCrack1300 数据集，旨在针对砌体缺陷检测。作者提出了两种基于 SAM 的新颖的全自动方法。两种方法都通过 LoRA 微调 SAM 的编码器：一种连接编码器到替代解码器，另一种则利用可学习的自生成提示器来指导分割。这两种策略在基准测试上均取得了显著的改进，展示了基于提示的微调的多功能性。

在 [?] 中，引入了一种选择性微调策略，其中仅对 SAM 架构内的归一化参数进行微调。这种选择性微调不仅在泛化性能方面优于完全微调和其他 PEFT 技术，还具有更高的计算效率，使其成为现实世界部署的一个吸引人的选择。

进一步扩展 SAM 在边缘设备上裂缝检测的应用，[?] 提出了 CrackESS，这是一种用于检测和分割混凝土裂缝的综合系统。CrackESS 首先使用 YOLOv8 [?] 模型生成自我提示以提供初始定位，然后应用基于 LoRA 的微调 SAM 进行分割。生成的掩膜通过专门的裂缝掩膜细化模块（CMRM）进行进一步优化，从而提高了分割精度。

总的来说，这些开创性的工作代表了最早和最有前途的努力之一，即为裂缝分割调整和微调 SAM。它们的成功凸显了利用大规模视觉模型和参数高效调整方法来解决泛化挑战的潜力，为在多样化和现实世界条件下的裂缝检测开辟了新的研究方向。

2.7. 泛化性分析

监督学习一直是基于深度学习的裂缝分割的主要范式，是实现该领域重大突破的基础方法。虽然传统的监督学习并不明确关注跨领域的泛化能力，但它提供了大多数模型构建所依赖的关键训练框架。此外，近期出现的在更大数据集上通过监督目标进行训练的基础模型表明，即使在设计为广泛泛化的系统中，监督学习仍然是基石。

基于这一基础，替代学习范式逐渐受到欢迎，尤其是因为它们可以提高模型的一般化能力并减少对劳动密集型标注的依赖以实现数据效率。半监督学习利用一小部分有标记的数据和大量无标记的数据，即使在监督有限的情况下也能实现有效训练。弱监督学习通过使用粗略或间接的标签来指导学习过程，从而进一步减少标注要求。小样本学习旨在从每个类别仅少量标记样本中进行泛化，这在现实世界场景中尤为有价值，因为在这些场景中，收集大量标注数据集是不切实际的。

领域自适应解决了领域迁移的挑战，其中在一个数据集上训练的模型可能由于特性不同在另一个数据集上表现不佳。它使得模型能够从一个有标签的源领域适应到一个不同的、通常没有标签的目标领域，通过在图像、特征或标签层面上对齐分布来实现。无监督学习则代表了一种更为高效的注释策略，模型直接从无标签数据学习潜在的结构和模式，但除非结合辅助监督，否则通常在特定任务上的准确性有限。

最后，基础模型代表了这些趋势的综合。通过使用监督目标在大量且多样的数据集上进行预训练，它们具备在不同领域中的强大泛化能力。

在裂缝分割中，这些模型和学习范式的适应性为在各种复杂环境中的数据效率和泛化能力提供了一个有前景的方向。

3. 数据集

随着裂缝检测领域的快速发展，越来越多的数据集被引入以支持研究和开发。图 1 中总结了一个时间轴，显示了基于不同类别的公开数据集的发布：1）通用裂缝分割数据集，2）使用 RGB 相机以外的专用传感器的数据集，3）结合现有数据集的综合数据集，4）合成数据集。表 2 & 3 中提供了一个详尽的公开数据集列表。

如表格所示，大多数公开可用的数据集是使用 RGB 相机收集的。这些数据集通常包括标注了裂缝位置的道路或结构表面，并已成为深度学习模型基准测试的基础。

后来，使用专业级传感器如 3D 激光扫描仪收集的数据在实际部署中获得了广泛关注，因为其在不同的照明和表面条件下具有增强的鲁棒性。早期的研究 [? ? ? ?] 开始利用 2012 年左右收集的数据，从那时起，它被研究人员和实践者广泛采用 [? ? ?] 。一项 2017 年报告的调查显示，美国有 18 个州使用 3D 自动数据收集，另有 17 个州计划使用这项技术 [?] 。它也被联邦公路管理局（FHWA）以及美国各州的交通运输部门认可为成熟技术 [? ?] 。然而，尽管这些数据集非常有价值，它们却很少被公开共享。虽然诸如 AEL [?] 和 FIND [?] 这样的数据集提供了用专业传感器收集的标注数据，但在数量、分辨率或多样性方面仍有不足，限制了其对需要在路面接缝、标记等干扰下保持鲁棒性的深度学习模型的实用性。因此，学术研究和实际

Dataset	Year	Description	Image Characteristics
CrackTree200 [?]	2012	206 pavement images containing various types of cracks, including distractions such as shadows and occlusions.	600 × 800
CrackIT [?]	2014	84 pavement surface images collected using an optical device during a traditional road survey.	1536 × 2048, 1 pixel ≈ 1 mm ^{2} , grayscale
CFD [?]	2016	118 urban road crack images captured using a smartphone with fixed camera settings in Beijing, containing common noise such as shadows.	320 × 480
Crack500 [?]	2016	pavement images collected on the Temple University campus using a smartphone.	2448 × 3264, 3-channel (RGB)
AEL [?]	2016	269 pavement images (68 annotated) collected using five acquisition systems (AigleRN, ESAR, LCMS, LRIS, TEMPEST2). Also referred to as the Sylvie Chambon dataset.	No uniform size
CSSC [?]	2017	954 concrete crack images collected via web search; crack shapes are stochastically distributed.	100 × 100 and 130 × 130
FCN [?]	2018	800+ images of pavement and concrete wall cracks with varying widths (1–100 pixels) and complex shapes. Includes internet and real-world images from Harbin, China.	224 × 224; resolution: 72–300 dpi
CRKWH100 [?]	2018	100 pavement images captured using a line-array camera under visible-light illumination.	512 × 512, ground sampling distance: 1 mm
CrackTree260 [?]	2018	260 pavement images (expanded from CrackTree200 [?]) captured using an area-array camera under visible-light illumination.	600 × 800
CrackLS315 [?]	2018	315 pavement images captured using a line-array camera under laser illumination.	512 × 512, ground sampling distance: 1 mm
Stone331 [?]	2018	331 stone surface images captured using an area-array camera under visible-light illumination.	512 × 512
DeepCrack [?]	2019	537 crack images from internet and real-world sources, covering varied textures, materials (asphalt, concrete), and crack widths (1–180 pixels).	544 × 384, 3-channel (RGB)
GAPs384 [?]	2019	384 asphalt pavement crack images selected from the German Asphalt Pavement Distress (GAPs) dataset [?].	1920 × 1080, 1.2 mm/pixel, 8-bit grayscale
KolektorSDD [?]	2019	52 crack images contained in 399 electrical commutator surface images.	1408 × 512
Khanh11k [?]	2019	112,000 images with varied crack conditions and environments, merged from multiple datasets including CrackTree200 [?], Crack500 [?], GAPs384 [?], CFD [?], and AEL [?].	448 × 448
UAV75 [?]	2019	75 images captured by unmanned aircraft systems (UAS), containing various cracks and planking patterns (visually similar to cracks).	512 × 512
DIC [?]	2020	530 image patches from 8 laboratory stone masonry wall specimens, captured for digital image correlation (DIC) under various loading levels and support conditions.	256 × 256

Table 2. 关于裂缝分割的公开可用数据集总结（表 1/2）。

Dataset	Year	Description	Image Characteristics
Masonry [?]	2021	11,491 image patches (4,057 with cracks) from 469 masonry surface photos with varied resolutions, crack types, and background noise.	224 × 224
Ceramic [?]	2021	167 ceramic crack images with varied crack shapes, ceramic types, and imaging conditions (e.g., scale, angle, illumination).	256 × 256, 3-channel (RGB)
CrSpEE [?]	2021	2,229 images with cracks and spalling on various material types and at different scales.	147 × 288 to 4600 × 3070
BCL [?]	2021	11,000 image patches from 50+ in-service bridges: 5,769 non-steel crack, 2,036 steel crack, 3,195 noise; under varied weather, lighting, and imaging conditions.	256 × 256
Conglomerate [? ?]	2021	10,995 crack images synthesized from nine public crack datasets [? ? ? ? ? ?] via normalization, noise removal, and resizing.	512 × 512
LCW [? ?]	2021	3,817 bridge inspection images with varied lighting, background materials, crack widths, and scales. Named as Labeled Cracks in the Wild (LCW) dataset.	512 × 512
Syncrack [?]	2022	600 synthetic crack images (expandable) generated under varying noise levels using the proposed crack generator.	480 × 320, 3-channel (RGB)
TopoDS [?]	2022	692 stone masonry images expanded from DIC [?], including cracks on lab specimens and real-world damaged buildings.	256 × 256
CrackSeg9k [?]	2022	9,255 crack images refined from ten public crack datasets [? ? ? ? ? ? ? ?] via resizing, noise removal, distortion, and refinement.	400 × 400
S2DS [?]	2022	232 crack images from the 743-image Structural Defect Dataset (S2DS), collected at real inspection sites using various camera platforms.	1024 × 1024
FIND [?]	2023	2,500 crack image patches from bridge deck and roadways obtained by a laser scanning device, containing surface elevation information.	256 × 256, image types: raw intensity, raw range, filtered range, and fused
CrackMap [?]	2023	120 road crack images collected using a vehicle-mounted RGB camera, representing typical road scenes.	256 × 256, 3-channel (RGB)
Crack900 [? ?]	2023	914 sets of visible and infrared (IR) images captured from masonry walls using solar heating as the sole heat source.	384 × 288, image types: RGB, IR, RGB_IR, RGB_T, fused
TUT [?]	2024	1,408 crack images from mobile phones (1,270) and online sources (138), featuring diverse, real-world backgrounds and complex crack patterns.	640 × 640
SegCODEBRIM [?]	2024	420 crack segmentation labels added to CODEBRIM, a concrete defect dataset for bridges.	1500 × 844
MCrack1300 [?]	2024	1,300 masonry crack images sourced from Crack900, online images, and mobile phone photos, covering diverse brick types and crack patterns.	640 × 640
OMNICRACK30K [?]	2024	30,017 crack images synthesized from 20 public datasets [?], with duplicates and noise removed, and class imbalance addressed.	No uniform size
Ours	2025	1,634 pavement 3D images from asphalt and concrete surfaces, collected using a mainstream survey system under varied environments and crack severities; each pixel encodes pavement elevation.	512 × 624, 4 mm/pixel, 8-bit grayscale

Table 3. 裂缝分割的公开可用数据集摘要（表 2/2）。

实施之间仍然存在显著差距。

而且，数据集不仅在数量上增加，还在规模上扩大。较新的数据集往往通过聚合多来源的数据来包含数千甚至数万张标注图像 [? ? ?]，这使得可以训练更大和更复杂的模型，以提高性能并在多种场景下进行泛化 [?]。

同时，人们对合成数据集的兴趣日益增加。这些数据集是通过计算机图形、仿真或生成模型生成的，并提供对裂纹特征、环境条件和真实标签 [? ?] 进行细粒度控制。合成数据可以通过提供多样化、均衡且完全标记的训练样本来补充真实世界的数据集，尤其对于数据难以获取或注释的任务特别有用。

3.1. 一个新数据集：3DCrack

如 3 中介绍的，3D 激光扫描已成为获取高分辨率、全车道路面数据的主流技术之一，在不同光照条件下提供强大的性能。与基于 2D 强度的成像不同，3D 激光系统能够抵御阴影、污渍或低对比度光照产生的噪声，使裂缝和表面缺陷更加明显 [? ? ? ? ? ? ?]。由于这些优势，3D 路面数据已被交通运输部门（DOTs）广泛采用，用于实际裂缝检测 [? ?]。然而，缺乏具有常见干扰的大规模、高分辨率数据集仍是进一步研究和实施的瓶颈。

为了弥合这一差距，我们发布了 3DCrack，一个新的 3D 路面图像数据集，以支持基于深度学习的裂缝检测（图 12）。数据是使用佐治亚理工交通感知车（GTSV）收集的，该车辆配备了双 3D 线激光传感器。这些传感器以最高 60 英里每小时的速度捕捉宽达 4 米的路面轮廓。每帧生成一个密集的 1,000 × 2,080 点云，插值后分辨率为 1 毫米 ×1 毫米，高度精度为 0.5 毫米 [? ?]。通过将高度重分为 8 位灰度进行压缩来将点云转换为 3D/距离图像，较暗的像素表示凹陷，并通过高斯高通滤波进行校正以抑制大规模高度趋势 [? ?]。

3DCrack 数据集包括从乔治亚州的 SR 275、US 80 和 I-16 收集的 1,139 张训练图像、245 张验证图像和 250 张测试图像，涵盖柔性和混凝土路面类型。该数据集包含多样的裂缝状况，包括不同几何形状如横向、纵

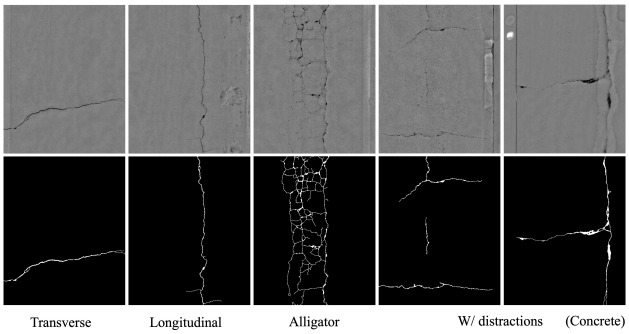


Figure 12. 来自我们 3DCrack 数据集的示例。每个距离图像都配有细粒度的像素级注释。裂缝可以分为横向、纵向和复合（包括鳄鱼纹）裂缝。请注意，最后两个示例展示了一些常见的干扰因素，如路面标线和接缝，它们不被视为裂缝。最后一列包括了一个混凝土路面表面的例子，而其他的是柔性表面。

Method	Metric	DeepCrack				3DCrack			
		25 %	50 %	75 %	100 %	25 %	50 %	75 %	100 %
ResUNet	IoU	0.694	0.717	0.725	0.725	0.456	0.648	0.694	0.703
DeepCrack		0.667	0.674	0.701	0.706	0.435	0.504	0.639	0.662
Deeplabv3+		0.667	0.708	0.707	0.718	0.385	0.56	0.632	0.635
XcepUNet		0.673	0.684	0.702	0.708	0.41	0.482	0.622	0.661
SegFormer		0.664	0.701	0.701	0.709	0.413	0.56	0.632	0.635
UniMatch		0.665	0.672	0.702	0.675	0.267	0.346	0.551	0.567
UniMatchv2		0.696	0.700	0.707	0.732	0.566	0.47	0.611	0.665
CrackSAM		0.723	0.748	0.748	0.751	0.564	0.588	0.592	0.6
ResUNet	Precision	0.831	0.866	0.87	0.879	0.607	0.81	0.82	0.827
DeepCrack		0.838	0.834	0.873	0.859	0.57	0.686	0.762	0.805
Deeplabv3+		0.838	0.839	0.855	0.834	0.52	0.705	0.736	0.733
XcepUNet		0.843	0.838	0.874	0.895	0.6	0.659	0.827	0.803
SegFormer		0.795	0.818	0.844	0.829	0.53	0.705	0.736	0.733
UniMatch		0.789	0.819	0.815	0.87	0.306	0.353	0.608	0.658
UniMatchv2		0.821	0.784	0.816	0.85	0.736	0.576	0.694	0.753
CrackSAM		0.847	0.823	0.849	0.86	0.661	0.685	0.691	0.703
ResUNet	Recall	0.825	0.82	0.827	0.817	0.552	0.727	0.778	0.782
DeepCrack		0.793	0.799	0.788	0.809	0.558	0.586	0.738	0.741
Deeplabv3+		0.779	0.831	0.817	0.853	0.521	0.683	0.763	0.767
XcepUNet		0.788	0.804	0.787	0.783	0.486	0.566	0.668	0.739
SegFormer		0.822	0.839	0.82	0.844	0.563	0.683	0.763	0.767
UniMatch		0.837	0.811	0.853	0.767	0.639	0.654	0.747	0.725
UniMatchv2		0.84	0.886	0.863	0.852	0.66	0.626	0.765	0.811
CrackSAM		0.844	0.901	0.873	0.862	0.716	0.736	0.745	0.748
ResUNet	Dice	0.808	0.825	0.833	0.833	0.557	0.747	0.788	0.795
DeepCrack		0.784	0.787	0.81	0.812	0.538	0.602	0.735	0.758
Deeplabv3+		0.786	0.819	0.818	0.827	0.496	0.673	0.74	0.741
XcepUNet		0.79	0.794	0.811	0.817	0.503	0.582	0.72	0.754
SegFormer		0.785	0.813	0.816	0.821	0.521	0.673	0.74	0.741
UniMatch		0.783	0.786	0.815	0.793	0.361	0.434	0.66	0.675
UniMatchv2		0.812	0.816	0.822	0.84	0.673	0.585	0.72	0.771
CrackSAM		0.828	0.848	0.848	0.85	0.676	0.699	0.704	0.711

Table 4. 在不同数据集上训练了不同数量的数据后的模型性能。最佳性能以粗体显示，second best 用下划线标出。

向、鳄鱼纹和复合裂缝，以及路面接缝和其他视觉干扰等挑战，确保在真实场景中训练模型的鲁棒性。每张图像的尺寸为 512 × 624 像素，每像素代表 4 毫米的分辨率。裂缝的真实标签由经过培训的标注者手动标注，以确保标注质量如 [?] 。

4. 实验 & 发现

为了促进未来的研究，我们在 3DCrack 数据集上对几种代表性模型进行基准测试，以建立基线性能。考虑了两种实验设置。第一种评估在不同训练数据量下的监督学习方法，为监督学习和半监督学习方法提供参考点。第二种专注于泛化性能，其中模型在一个大规模集合上训练，并在具有不同表面条件的独立集合上测试，为通用裂缝检测方法提供基线。

为了评估模型性能，我们采用四个常用指标：交并比（IoU）、精度、召回率和 Dice 系数。这些指标定义如下：TP 代表真正例，FP 代表假正例，FN 代表假负例。

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{Dice} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{4}$$

在接下来的实验中，我们对 3DCrack 和 DeepCrack [?] 数据集进行相同的训练和测试设置，以量化每个模型在不同数据集上的表现，并突出我们 3DCrack 特有的特征和挑战。原始的 DeepCrack 没有验证集，因此我们将原始的测试集以 1:1 的比例划分为验证集和测试集。

4.1. 监督 & 半监督实验

4.1.1. 实验设置

在监督实验中，我们选择了一组具有代表性的模型，依据是这些模型在图像分割任务中已被证明的有效性以及相对容易的实现和微调。具体而言，我们包括了具有两种不同主干编码器的 U-Net [?] 架构：ResNet [?] 和 Xception [?] 。ResNet 主干以其残差学习能力而广为人知，能够帮助缓解梯度消失问题并改善特征传播。另一方面，Xception 主干则利用深度可分离卷积来提高模型的效率和准确性。此外，我们引入了专门用于裂缝检测的模型 DeepCrack [?] ，该模型在识别图像中精细线状结构方面表现出色。Deeplabv3+ [?] 因其强大的语义分割能力而被纳入，尤其是它使用的空洞空间金字塔池化和编码器-解码器设计，可以实现精确的边界定位。由于 SegFormer [?] 在近期基准测试中的强劲表现，也被评估在内，它结合了基于变压器的编码和轻量级解码器，实现了高效和准确的分割。最近一个用于裂缝分割的基础模型 CrackSAM [?] 也包括在此，且将在下文详述。

对于半监督实验，前述模型通过限制有标签训练样本的数量以直接的方式进行调整，使我们能够评估它们在标签稀缺条件下的鲁棒性。为了补充这些基线，我们还包括了专为半监督语义分割设计的最新模型家族 UniMatch v1 & v2 [? ?] 。UniMatch 通过一致性正则化和伪标签策略，有效利用未标记的数据，并在有限监督条件下展示了优越的性能。此外，UniMatch v2 采用了更新且更强大的编码器 DINOv2 [?] 。

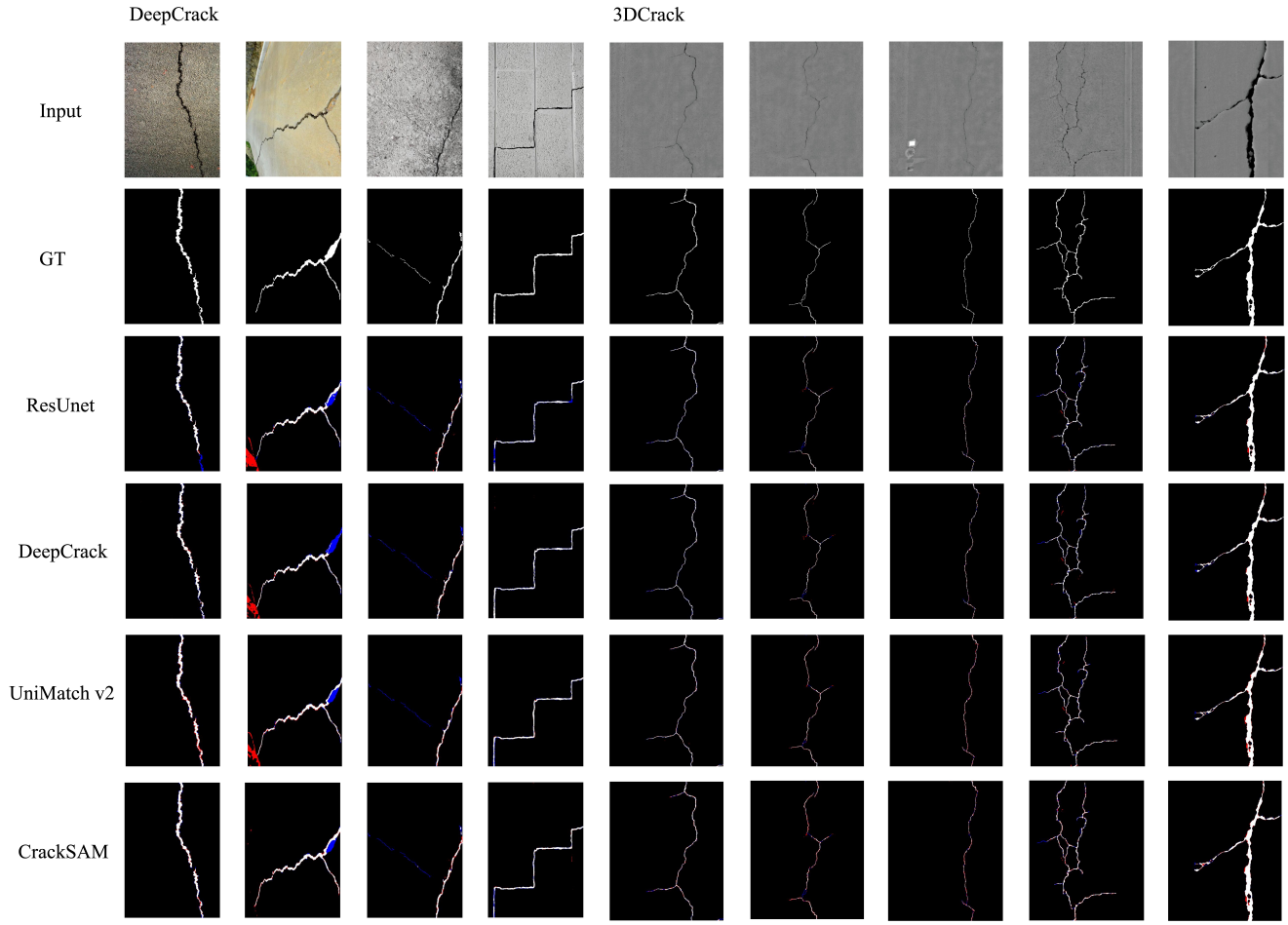


Figure 13. 来自不同模型在 DeepCrack 和 3DCrack 数据集上的裂缝分割结果。假阳性以红色突出显示，假阴性以蓝色突出显示。

模型根据其论文中指定的原始训练策略和 Dice 损失函数进行训练。在训练过程中应用提前停止，每个模型最多进行 100 个 epoch 的训练。

4.1.2. 研究结果

总体表现。正如在表 4 和图 13 中所示，所有选定的模型在训练后都能够有效地分割裂缝。其中，ResUNet 在使用完整的训练集进行训练时，在两个数据集上都表现出强劲的整体分割性能，在 DeepCrack 上的 Dice 得分为 0.833，在 3DCrack 上为 0.795。最近的 UniMatch v2 模型由于使用了强大的 DINOv2 视觉编码器，也表现良好。在训练过程中应用提前停止，每个模型的最大训练次数为 100 个 epoch。

两个专门的裂缝检测模型，DeepCrack 和 CrackSAM，表现出显著不同的行为。尽管它们是为此任务专门设计的，DeepCrack 稍微落后于像 ResUNet 这样的模型，这可能是由于其浅显和简单的架构所致。相比之下，CrackSAM 在各项指标、训练数据量和数据集方面持续获得高分。这强调了基于深度学习的裂缝检测的一项基本进展：更强大的骨干网络对性能是有益的。CrackSAM 从"Segment Anything Model (SAM)" [?] 继承了强大的语义理解能力，使其能够更好地将裂缝与常见的非裂缝干扰区分开来。这一优势通过图 13 第二列的一个示例得到了说明，其中 CrackSAM 比其他模型更有效地抑制了非裂缝区域的误报。

数据集差异。总体而言，模型在 3DCrack 数据集上的表现比在 DeepCrack 上的表现要差，这突出显示了这两个数据集之间的领域差异。尽管 DeepCrack 包含常见的干扰因素，如阴影和非裂缝区域，它通常提供了更清晰和更聚焦于裂缝的视图。相比之下，由于其自上而下的视角和对路面表面的更广泛覆盖，3DCrack 包含了更细微、更隐蔽的发丝裂缝，如图 12 中的示例所示。这些裂缝往往更难检测，导致性能显著下降。例如，ResUNet 的 IoU 分数从使用 100 % 训练数据的 DeepCrack 上的 0.725 下降到 3DCrack 上的 0.703。同样地，尽管 CrackSAM 具有优越的语义建模能力，但它在 3DCrack 中捕捉小尺度裂缝细节的能力有所降低，这些裂缝在放大背景中不太容易与背景区分，如表 4

Method	Metric	Zero Shot		Many Shot	
		DeepCrack	3DCrack	DeepCrack	3DCrack
ResUNet	IoU	0.441	0.279	0.633	0.67
DeepCrack		0.387	0.181	0.557	0.584
CrackSAM		0.502	0.3	0.711	0.6
Cycle GAN		0.621	–	–	–
ResUNet	Precision	0.589	0.298	0.848	0.824
DeepCrack		0.64	0.207	0.793	0.723
CrackSAM		0.551	0.319	0.834	0.699
Cycle GAN		0.776	–	–	–
ResUNet	Recall	0.731	0.729	0.736	0.732
DeepCrack		0.602	0.633	0.681	0.682
CrackSAM		0.903	0.758	0.846	0.742
Cycle GAN		0.768	–	–	–
ResUNet	Dice	0.568	0.375	0.742	0.767
DeepCrack		0.514	0.27	0.674	0.686
CrackSAM		0.644	0.392	0.816	0.711
Cycle GAN		0.745	–	–	–

Table 5. 在不同数据集上经过组合数据集训练的具有普遍化能力的模型表现。最佳表现用粗体显示，second best 用下划线标出。

和图 13 所示。

训练数量效应。训练数据量的影响在 Tab. 4 和 Fig. 14 中的定性结果中清晰可见。所有模型都显著受益于更多的标记数据，提高了它们正确分割裂缝、抑制噪声和处理多样化裂缝模式的能力。特别是，训练数据的增加有助于模型更好地区分真实裂缝和虚假的模式。在 Fig. 14 中，这种改进在视觉上得以体现：红色区域（误报）缩小，而蓝色区域（漏报）逐渐被白色（正报）所替代，表明随着数据量增加，模型的准确性和召回率得到增强。

实验结果还表明，当训练数据从 75 % 增加到 100 % 时，大多数模型的性能趋于平稳，这表明探索更有效的策略以利用更多数据仍然是未来研究的一个重要方向。

低数据表现。在标记数据有限的情况下，CrackSAM 和 UniMatchv2 脱颖而出。UniMatchv2 由于其半监督学习框架和使用强大的 DINOv2 主干网络，而在表现上更胜一筹，这相比于 UniMatchv1 中基于 ResNet 的主干网络提供了更丰富的特征表示。CrackSAM 表现也相当优秀，受益于在强大先验知识的 SAM 基础模型上的微调。其在对象级语义理解方面的预训练能力使其在最少监督下识别裂缝时特别高效。这些结果表明，结合预训练语义先验或为半监督学习设计的模型在数据匮乏的条件下具有明显的优势。

4.2. 通用实验

4.2.1. 实验设置

在这个实验中，ResUNet [? ?] 和 DeepCrack [?] 的训练方式与之前实验相同。CrackSAM [? ?] 和 CycleGAN [?] 的训练遵循原始论文。在这些具有通用性的实验中，有两种不同的设置，零样本和多样本。测试集保持不变，由 DeepCrack [?] 和 3DCrack 的测试集组成。零样本实验的训练集不包含这两个数据集的任何图像，而多样本设置的训练数据则包含了这

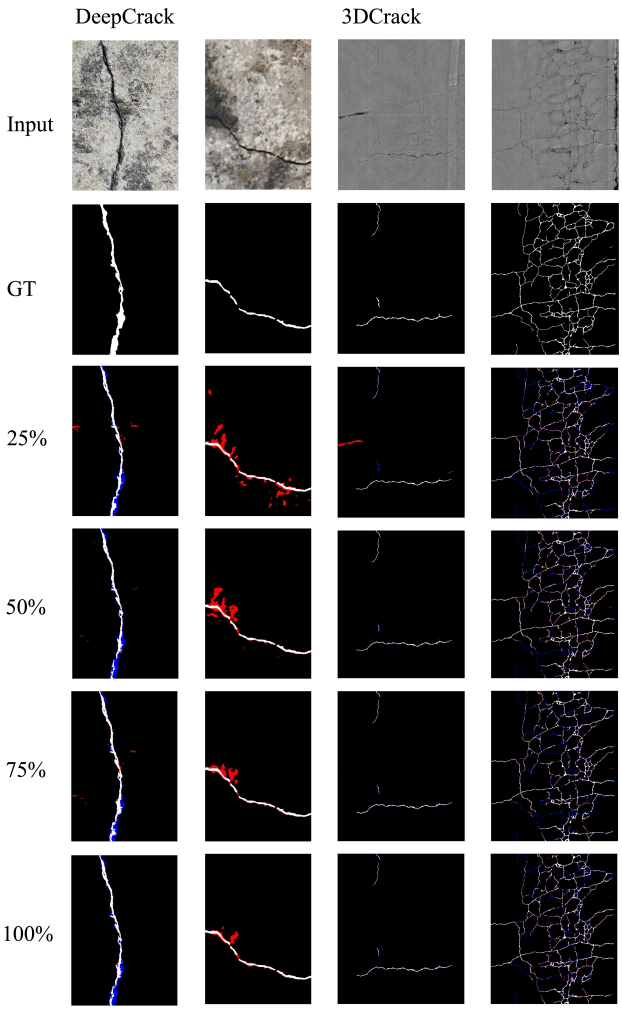


Figure 14. 使用不同训练数量的 ResUNet 在 DeepCrack 和 3DCrack 数据集上的裂缝分割结果。假阳性用红色高亮显示，假阴性用蓝色高亮显示。

两个数据集的训练集。

零样本。在零样本研究中，我们探讨了文献中常用的两组不同方法。第一组着重于利用大规模数据集，通过广泛的数据暴露使模型能够在多种不同的场景中泛化 [?] 。这些方法通常依赖于具有更大参数数量的高容量架构，使它们能够从各种输入条件中学习复杂的特征表示。在我们的实验中，我们评估了三个这样的模型，分别是 ResUnet [? ?] 、DeepCrack [?] 和 CrackSAM [? ?] ，它们代表了大型分割网络的不同家族，在从 Khanh11k [?] 删除属于 DeepCrack 的数据后，这些网络以其在充分多样化的数据上进行训练时的强泛化性能而著称。我们还包含了 CycleGAN [?] 以评估未配对的图像到图像翻译在无监督的情况下能多好地从现实输入中生成可靠的分割图；具体来说，我们打乱标注以创建基于上面提到的精心整理的 Khanh11k 的未配对训练数据。

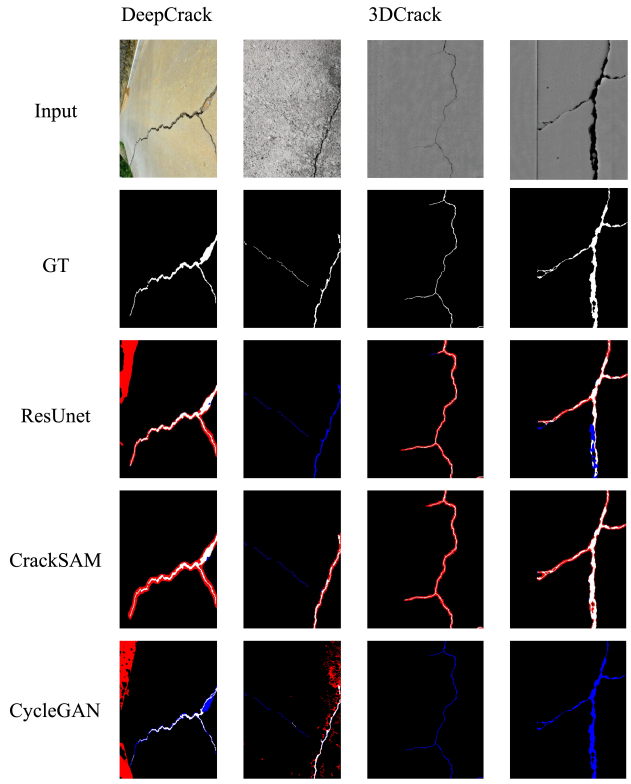


Figure 15. ResUNet、CrackSAM 和 CycleGAN 的零样本性能。假阳性用红色突出显示，假阴性用蓝色突出显示。

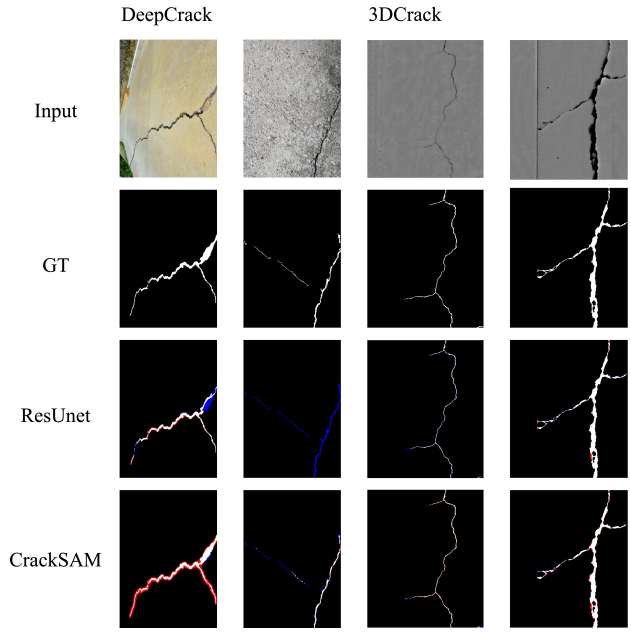


Figure 16. ResUNet 和 CrackSAM 的多次识别性能。误报以红色突出显示，漏检以蓝色突出显示。

多次训练实验。为了进一步探索数据多样性和数据量对分割性能的影响，我们通过从 DeepCrack 和 3DCrack 数据集的整个训练子集中增加额外的标注样本，增强 Khanh11k 数据集 [?] 的训练集，进行多次训练实验。这个设置使我们能够模拟一个现实的、完全监督的训练场景，其中模型可以访问丰富的、异质的数据源。其目的是评估接触更多样化的裂缝外观和成像条件是否能提高裂缝分割模型的泛化能力和鲁棒性。

4.2.2. 研究结果

零样本。据表 5 和图 15 所示，所有选定的模型均展示了进行零样本裂缝分割的能力。然而，尽管有一些积极的结果，零样本性能仍然受到限制，这是由于泛化差距，特别是在存在域偏移的情况下。观察到的最普遍的问题是过分和不足现象：模型要么错过了精细裂缝结构（不足），要么由于视觉相似性错误地分割了非裂缝区域（过分）。

ResUNet 在监督场景中表现出色，但在这种设置下表现不佳。它经常无法识别正确的裂缝区域，容易被诸如表面纹理、道路标记或阴影等背景模式分散注意力。这表明，如果没有针对数据集的特定微调，其推广语义边界的能力是有限的。

在这种零样本环境中，CrackSAM 相比于 ResUNet 表现出更为稳健的分割能力。由于其建立在 SAM 的基础上，它在裂缝区域的定位上有了改善，并且表现出较少的干扰。然而，它仍然表现出一个一致的倾向，即过度分割裂缝周围的区域，导致边界膨胀并偶尔包括无关的路面细节。

在这种设置中，CycleGAN 在 DeepCrack 数据集上表现最佳。在无监督的图像到图像翻译过程中，CycleGAN 通过领域自适应将输入图像风格化为与训练分布相似的样子，从而能够在领域之间架起视觉鸿沟。这使得它能够更接近训练期间看到的裂缝模式。然而，CycleGAN 的方式纯粹依靠风格驱动，没有任务特定的目标，因此在噪声和误报方面仍然容易受到影响，特别是在纹理与裂缝相似的区域。此外，其在 3DCrack 数据集上的性能崩溃，因为该数据集中的裂缝更细、更淡，且嵌入在大面积的非裂缝背景中。这表明基于风格迁移的模型在处理缺乏强视觉显著性的细粒度结构特征时存在关键弱点。

总之，尽管零样本分割为减少标注负担提供了一个有吸引力的方向，但目前的方法在准确处理裂缝边界和适应具有不同视觉特征的新领域方面仍然面临重大挑战。未来的工作可能集中在整合先验知识 [? ?] 、不确定性建模 [? ?] 或对比学习策略 [? ?] 以增强零样本的鲁棒性，并减轻过冲/欠冲误差。

多次拍摄。如表格 5 和图 16 所示，与零次拍摄设置相比，所有模型的分割精度均有显著提高。性能提升在 Dice 分数和 IoU 方面尤为显著，非裂缝区域的假阳性减少，裂缝边界的界定更好。尤其是 ResUNet，表现出在背景噪声和视觉干扰下的鲁棒性显著增强，这表明增加的训练多样性帮助模型更好地学习可以跨数据集泛化的判别特征。

此外，所有模型在识别更细微裂缝方面都有所提高，这些裂缝在零样本条件下以前常被低估。这表明，增加的标注裂缝的数量和变异性提高了模型捕捉定义复杂或不太显眼裂缝模式的微妙视觉线索的能力。

然而，尽管有这些改进，将表格 5（多次训练）中的结果与表格 4 （个别数据集的全监督训练）中的结果进行比较时，并没有获得性能提升。这表明简单地增加来自多个领域的训练数据的数量和多样性，并不会自动带来成比例的性能提升。一个可能的原因是数据集之间的领域差异：DeepCrack、3DCrack 和 Khanh11k 中的其他数据集在成像视角、裂缝样式、背景纹理和干扰类型上存在差异。基线模型可能仍然难以协调这些差异，这凸显了需要针对任务的设计，包括定制的架构、损失函数等。

这些发现指出了未来研究的一个重要方向：开发能够更有效地利用多样和领域多变的训练数据的模型。这可能涉及诸如特征对齐 [? ? ?] 和课程学习 [? ?] 等技术，这些技术关注数据集之间不变的裂缝特征。解决这一限制对于构建真正稳健和可泛化的裂缝分割系统至关重要。

5. 开放性挑战

尽管近年来取得了快速进展和有前景的结果，但正如我们的综述和实验所示，自动裂缝检测任务仍远未完全解决。在下文中，我们概述了一些继续影响裂缝检测研究格局的关键开放挑战。

5.1. 学习范式

数据的高效利用。我们的研究综述显示，发展能够更高效利用可用数据的学习策略是一个日益增长的趋势。尽管在数据集扩展和大型预训练模型的采用上持续取得进展，总会有边缘情况、未见场景和变体无法被现有数据充分捕捉。获取一个包括大量新变体的全面、高质量的数据集仍然资源密集，尤其当需要像素级注释时。为了解决这些挑战，未来的研究可以进一步探索数据高效学习范式，如 SSL、WSL、USL、FSL、DA 以及带有 PEFT 的基础模型。这些方法已展示出有前景的结果，具有进一步最大化模型性能的潜力，同时减少对大型标注数据集的依赖，从而增强在实际应用中的可扩展性和快速适应性。

基准测试和标准化。随着该领域继续探索从监督学习到基础模型适应的广泛学习范式，标准化基准的需求变得越来越关键。没有一致的数据集、评估指标和任务定义，难以在方法之间进行公平比较或得出关于其优缺点的可靠结论。目前，许多研究使用自定义的评估协议，这妨碍了可重复性并减缓了集体进步。建立具有明确定义的训练和测试集、标准化注释格式和统一性能指标的共享基准数据集，将能够进行更严格和透明的评估 [? ?] 。本研究通过新发布的 3DCrack 数据集和定制的基准设计为这一目标做出贡献。此外，维护排行榜和组织挑战赛可以进一步激励创新，并为裂缝检测和分割的进展提供一个集中的参考点。

5.2. 通用性 & 模型开发

扩展到更大模型和基础架构。最近在计算机视觉领域的进展是由大规模模型和基础架构（如 transformers [?] 、CLIP [?] 、DINO [?] 和 SAM [?] ）推动的，这些架构在任务和领域间表现出了显著的泛化能力。受此趋势的启发，一些初步努力已经开始尝试将这些基础模型应用于裂缝检测 [?] 。尽管结果令人鼓舞，但这个领域仍然未被充分探索，总体性能并不完美，如我们在第 4.2 节中所展示的。具体来说，应该消除由干扰性图案引起的误报和粗略检测，以支持后续任务如裂缝量化 [?] 。

未来的工作可以研究如何更好地适应基础模型以应对不同场景下的裂缝检测，可能通过结合特定领域的微调 [? ? ? ? ?] 、提示调节 [?] 或用于薄结构分割的专用模块 [? ?] 。

裂缝特定的架构和目标设计。与一般物体不同，裂缝通常是细长的，并且常常是不连续结构，使得用现成的模型分割非常具有挑战性。许多研究已经引入了裂缝感知架构或损失设计，这些设计考虑了裂缝的几何和拓扑特征 [? ? ? ?] 。这些设计在整体性能上表现出明显的优势。同样，在这个方向上持续的研究是必要的，以开发明确地考虑裂缝独特空间模式的模型，从而减少误报和漏报，得到更精细的结果。

实时和高效的架构。随着模型规模的增加以获得更高的准确性，平衡性能与计算效率的需求也在增长，尤其是在移动检测机器人或设备等实际场景中进行部署时。对于实际应用而言，实时推理至关重要。轻量级骨干架构（例如，MobileNet [?] 、EfficientFormer [?] 、CSegamba [?])、剪枝 [?] 、量化 [?] 和知识蒸馏 [?] 都是可以探索的方向。设计能够保持竞争性准确性，同时实现低延迟和小内存占用的模型将是未来的一大挑战。

多模态融合。裂缝检测有机会通过整合超越标准 RGB 图像的数据受益。多模态融合，例如，可以涉及结合各种传感器数据，如深度、红外信息 [? ?] 。融合多种模态有可能提高在不同环境条件下的稳健性，并提供更丰富的上下文理解。能够对齐和处理异构输入的深度学习模型，如视觉-语言模型 [? ?] 或传感器/特征融合网络 [? ? ?] ，代表了未来研究的一个有前途的领域。挑战包括数据源的同步、特定模态的噪声以及对大型注释多模态数据集的需求。

5.3. 数据集

数据稀缺。尽管已有几个公开可用的裂缝分割数据集存在，但它们的规模仍然有限。这些数据集通常包含几百到几千张图像，与其他计算机视觉任务中的大规模数据集（例如，COCO [?] ，Cityscapes [?] ）相比，这个数量相对较少。为了应对这一限制，研究人员尝试将多个数据集汇聚成一个统一的集合，通过增加训练的多样性和数量来提升性能。然而，与其他领域相比，总体上的数据量和多样性仍然不足，限制了模型的鲁棒性和泛化能力。

行业层面数据的缺乏。另一个关键缺口在于专业收

集和标注的数据集的有限可用性。大多数公共数据集依赖于使用消费级相机在不受控条件下收集的 RGB 图像。相比之下，使用专业传感器（例如，3D 激光扫描）获取的数据提供了更高的保真度，并且更能代表真实的检测场景。我们的工作尝试通过发布一个具有细粒度、专家标注裂缝掩膜的数据集来弥合这一差距。然而，仍然强烈需要来自不同环境和传感器的额外高质量数据。考虑到像素级标注所需的巨大努力，未来的研究也可以考虑探索使用半自动化标注工具 [？？？] 以扩大数据集的生成。

数据质量和标注方法。除了数据数量，质量也是另一个需要关注的问题。人工标注本质上容易受到主观性和标注者之间差异的影响。影响因素包括对裂缝定义的不同解释、感知严重程度的变化以及标注协议的不一致性，这些都可能导致基础标签的噪声。建立明确的标注标准、质量保证的评估指标，并引入基于共识或概率的标注策略，可以帮助提高标注一致性。

另一方面，替代的注释策略值得进一步探索。粗略注释为加速数据生成提供了一个有前景的方向，特别是在弱监督学习等学习范式的支持下。当注入专家参与时，也可能带来益处，因为这不仅确保了注释质量，还通过迭代反馈和模型优化促进了持续学习。

6. 结论

在本文中，我们提出了深度学习裂缝检测的一项全面回顾，重点关注七个主要学习范式、对模型广泛适应性的日益关注，以及数据集在该领域的多样化。我们对学习范式的变化进行了分类和讨论，并附详细的子类别和代表性工作，强调了该领域如何从传统的监督方法进展到更数据效率较高的范式，如半监督、少样本方法等。适应性能力的出现被认为是一个关键趋势，反映了对模型在多样化和未知条件下可靠工作的需求日益增长。

在数据方面，我们观察到一个明显的趋势，即数据集变得更大、更多样化，并具有丰富的传感器。作为该努力的一部分，本文发布了一个高分辨率、大规模、工业级裂缝分割数据集 3DCrack，具有多样化的路面状况，为社区提供了一个宝贵的资源，用于基准测试和未来开发。

虽然裂缝检测的趋势与更广泛的计算机视觉领域密切相关，但该任务需要专门的设计和考虑。通过这篇综述和我们的实证研究，我们还确定并讨论了尚未解决的开放挑战。这些挑战涵盖学习范式、泛化能力和模型开发，以及数据集，为未来研究勾画出关键方向。我们希望这项工作为该领域的研究人员和从业者提供有用的参考，并鼓励进一步努力，朝向稳健、可扩展和可泛化的裂缝检测系统发展。

作者衷心感谢佐治亚理工学院的 Zsolt Kira 教授、Mi Zhou 博士、Jacob Biros，以及丰田研究院的 Muhammad Zubair Irshad 博士的有益讨论和宝贵反馈。