

# 噪声标签下心脏 MRI 心肌瘢痕分割的鲁棒深度学习

Aida Moafi<sup>1,2</sup>, Danial Moafi<sup>3</sup>, Evgeny M. Mirkes<sup>4</sup>, Gerry P. McCann<sup>1,2</sup>,  
Abbas S. Alatrany<sup>1,2</sup>, Jayanth Ranjit Arnold<sup>1,2\*</sup>, and Mostafa Mehdipour  
Ghazi<sup>5\*</sup>

<sup>1</sup> Department of Cardiovascular Sciences, University of Leicester, Leicester, UK

<sup>2</sup> NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Leicester, UK

<sup>3</sup> Department of Information Engineering and Mathematics, University of Siena,  
Italy

<sup>4</sup> Department of Mathematics, University of Leicester, Leicester, UK

<sup>5</sup> Department of Computer Science, University of Copenhagen, Copenhagen, DK

**摘要** 心肌瘢痕从心脏 MRI 的精确分割对于临床评估和治疗计划至关重要。在这项研究中，我们提出了一种稳健的深度学习管道，通过对最先进的模型进行微调以实现完全自动的心肌瘢痕检测和分割。该方法通过使用 Kullback-Leibler 损失和大量数据增强明确解决了来自半自动注释的标签噪声、数据异质性和类别不平衡的挑战。我们在急性病例和慢性病例上评估了模型的表现，并证明其能够在有噪声标签的情况下产生准确且平滑的分割。特别是，我们的方法在诸如 nnU-Net 等最先进的模型中表现出色，并在一个分布外测试集上显示出强大的泛化能力，突显了它在各种成像条件和临床任务中的鲁棒性。这些结果为自动心肌瘢痕量化奠定了可靠的基础，并支持深度学习在心脏成像领域的更广泛应用。代码可在：<https://github.com/Danialmoa/YoloSAM> 获取。

**Keywords:** 深度学习 · 分割 · 心肌瘢痕 · 噪声标签。

## 1 介绍

心血管疾病仍然是一个全球性的健康问题，导致全世界近 32% 的死亡病例 [1]。在其表现形式中，心肌梗死 (MI) 是发病率和死亡率的主要原因，通常会使得患者留下永久性的心肌疤痕。疤痕范围是一个关键的预后因素，即使是 5% 的疤痕大小增加也会使死亡风险上升 19% [2]。疤痕负担还影响血管重建的结果，并预测室性心律失常和心力衰竭的可能性 [3,4]。此外，透壁性疤

\* Co-senior corresponding authors: jra14@leicester.ac.uk and ghazi@di.ku.dk

痕对于评估恢复潜力至关重要，因为透壁疤痕小于 50% 的区域更有可能恢复功能 [5,6,7]。鉴于其临床重要性，准确量化疤痕对于个性化治疗和风险分层是至关重要的。

晚期钆增强心血管磁共振成像 (LGE-CMR) 是非侵入性疤痕评估的金标准，提供高分辨率、对比增强的梗死组织可视化。然而，目前的分割方法，包括手动轮廓绘制和半自动阈值法，耗时费力，并且容易受到观察者间变异的影响 [8,9,10]。手动标注每个扫描可能需要长达 25 分钟 [8]，这在临床实践中限制了其可扩展性。像信号阈值与参考均值 (STRM) 和半高全宽 (FWHM) 这样的半自动方法，基于强度阈值分割疤痕组织，但对信噪比变化、兴趣区域选择以及梗死信号特征敏感 [11,12,13]。这些限制往往导致疤痕描记的一致性较差。

深度学习为自动疤痕分割提供了有前景的解决方案 [14]。基于 2D 和 3D U-Net 架构的方法 [15] 已在 EMIDEC 挑战赛 [16] 中展示了可行性，但与地面真实数据变异性及成像协议差异相关的挑战阻碍了临床应用 [17]。一项系统综述报告这些方法的 Dice 相似性得分高达 0.63 [18]，而人工标注本身也显示出有限的一致性，其内部和外部观察者 Dice 指数分别为 0.76 和 0.69 [19]。然而，LGE 序列及信号强度的变化进一步复杂化了分割问题。幅度和相位敏感反转恢复 (PSIR) 序列在空间分辨率和对比度特性上有所不同，尽管 PSIR 提供了更好的对比度，但引入了绝对强度值的不一致性 [20,21]。不同的扫描仪供应商、采集参数及后处理技术增加了异质性，复杂化模型泛化。此外，相对于整体图像而言，疤痕的小尺寸和间歇出现可能导致模型过度拟合背景像素。地面真实标签的质量对性能有重大影响，因为模型通常基于具有内在偏差的手动或半自动标注进行训练。基于阈值的方法如 STRM 和 FWHM，虽然提高了可重复性，但在弥漫性纤维化、微血管阻塞 (MVO) 或信号异质性 [22] 的情况下可能无法完全捕捉到疤痕的范围。这些挑战凸显了需要稳健且具有普遍适用性的深度学习模型来进行自动心肌疤痕分割。

我们开发了一个深度学习管道，用于通过微调最先进的模型 [23,24] 实现心肌疤痕的全自动分割。

我们的贡献有六个方面：(1) 我们提出了一种端到端的管道，实现了高精度的分割，尽管是在嘈杂标签上训练的，但仍然能够产生更平滑的结果。(2) 我们引入了 Kullback-Leibler (KL) 损失来减轻不完美手动或半自动注释的影响。(3) 我们指出了基于重叠度量评估小病变时的局限性，并提出周长相似性作为更为可靠的大小度量。(4) 我们使用数据增强和检测模型来指导

分割过程，解决类别不平衡和标签噪声问题。(5) 我们在一个异构、多队列的数据集上验证了我们的方法，展示了其在不同成像条件下的鲁棒性。(6) 我们对一个分布外的测试集（包括 MI 瘢痕）进行了分割精度评估，证明了它对于未见过数据的任务具有泛化能力。

## 2 方法

### 2.1 数据采集与标注

本研究利用了一个包含四项关于慢性再灌注缺血瘢痕的独立研究的多队列数据集，详见表 1。这些数据集来自多个中心，涵盖了不同的 MRI 采集协议，包括成像序列（如幅度和相位）、扫描仪制造商（如西门子和飞利浦）、磁场强度（如 1.5T 和 3T）以及空间分辨率（从  $1 \times 1 \times 6 \text{ mm}^3$  到  $2 \times 2 \times 10 \text{ mm}^3$ ）的变化。这些变化引入了图像质量、强度和对比度的异质性，对一致性分割性能构成了挑战。当标准 LGE 图像显示不一致时使用 PSIR 图像，因为它们在采集过程中对反转时间选择和操作员调整的依赖较小。这种方法增强了心肌结构识别的可靠性。

共选择了 348 名患者，使用商用的 Circle CVI 软件手动进行了 LGE 疤痕分割。为了确保高质量的标注，所有分割结果都由两名心脏 MRI 专家进行复审。为了可重复性，使用了广泛使用的 FWHM 方法来进行疤痕勾画。作为一种基于阈值的技术，FWHM 方法存在固有限制，包括对强度变化和部分体积效应的敏感性，这影响了分割准确性。标记过程涉及三个关键步骤：

表 1: 研究中使用的多队列数据集 [25] 的总结。

	# Images			# Subjects			Age	Sex
	Train	Valid	Test	Train	Valid	Test	Mean $\pm$ SD	Female
Cohort 1	450	74	91	49	9	11	59.44 $\pm$ 9.85	25%
Cohort 2	493	82	88	59	12	13	45.41 $\pm$ 21.20	65%
Cohort 3	304	55	63	38	8	9		
Cohort 4	471	54	41	49	8	6		
Cohort 5	81	32	18	12	4	2		
Cohort 6	93	22	34	10	3	4		
Cohort 7	129	27	26	15	3	3		
Cohort 8	117	37	49	12	4	5		

(1) 心肌轮廓描记，在此过程中手动追踪心外膜和心内膜边界以定义感兴趣的心肌区域。(2) 疤痕核心识别，在其中在心肌信号强度最高的区域内放置一个 ROI，对应于疤痕核心。(3) 优化与排除，在此步骤中进行最终修订以排除潜在的成像伪影并确保案例之间的一致性。

## 2.2 预处理和数据增强

标注的切片经过预处理，考虑了 DICOM 像素间距，并调整为  $256 \times 256$  像素。我们实现了不同的增强技术以提高模型对解剖变异和成像伪影的鲁棒性。除了使用翻转、旋转、缩放和变形等空间变换的标准增强之外，我们还整合了特定于医学图像的技术来模拟 MRI 中的现实变化。这些增强包括伽玛校正、自适应直方图均衡化、亮度调整以及添加瑞利噪声以模仿 MRI 特有的噪声特征。还引入了边界框抖动以模拟手动注释的变化并提高检测鲁棒性。

为了确保评估的稳健性和无偏性，我们在患者级别划分了数据集，将每个队列的数据分为训练（70%）、验证（15%）和测试（15%）子集。这防止了数据泄露，并保证子集能够代表不同的队列。鉴于疤痕分割中固有的类别不平衡，我们根据每位患者的总疤痕像素数对子集进行了分层划分。这种方法保持了各子集中疤痕负担的完整分布，避免了关键案例的代表性不足。然后逐层面处理每个 LGE-MRI 扫描，单个切片作为网络输入。由于 LGE 序列中的许多切片不包含心肌，这种做法有助于缓解背景不平衡，并确保在模型开发过程中仅考虑相关区域。

## 2.3 噪声标签下的检测与分割

图 1 说明了本研究开发的用于检测和分割带有噪声地面实况标签的心肌疤痕的全自动管道。该管道采用 You Only Look Once (YOLO) 架构 [23] 和默认配置进行疤痕边界框检测，作为 Segment Anything Model (SAM) [24] 的提示。这种方法实现了自动化的疤痕定位，并提供了一个简化的分割框架，消除了手动干预的需要。检测步骤对于关注微小和碎片化的疤痕区域至关重要。该方法有效地将解剖学知识整合到深度学习中，在传统方法难以处理的具有挑战性的成像场景和临床应用中提高了分割性能。

尽管具有可重复性，FWHM 半自动标签非常锐利，并且由于阈值伪影和部分体积效应而经常与真实解剖边界错位。因此，我们将它们视为结构化边界不确定性中的噪声，而不是错误。为了解决带噪声标签的问题，我们建

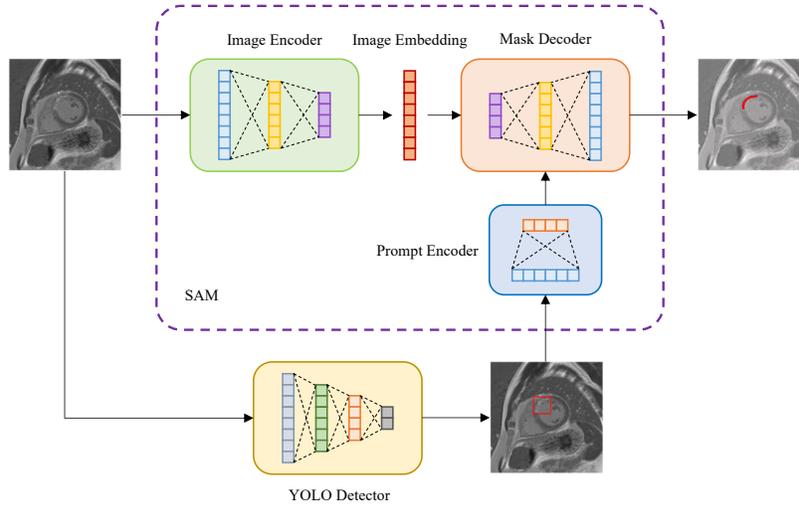


图 1: 提出的全自动心肌瘢痕检测和分割管道的概述。YOLO 模型被微调以生成自动边界框提示, 引导微调后的 SAM 进行局部化瘢痕的精确分割。使用 KL 损失和广泛的数据增强策略来缓解数据变化并处理噪声标签训练。

我们将 KL 散度作为额外的损失函数与 Dice 损失和交叉熵一起使用。我们将目标掩码应用高斯平滑, 其中  $\sigma = 2$ , 以获得软标签, 捕捉过渡边界区域中的不确定性。模型预测通过 Sigmoid 激活映射到  $[0,1]$  区间, 然后进行 log-softmax 操作以获取预测分布。将预测和软目标分数归一化为概率后, 计算预测和软目标之间的概率分布的 KL 散度, 惩罚在模糊区域附近的过于自信的预测。这三种损失函数以加权方式结合在一起, 权重系数分别为 0.2、0.2 和 0.6, 这些系数通过网格搜索确定。

### 2.4 评估指标

我们使用豪斯多夫距离 (HD)、Dice 相似系数 (DSC)、区域相似性 (AS) 和周长相似性 (PS) 来评估分割性能。由于 DSC 是一种对小误差敏感的重叠度量, 可能无法完全捕捉到微小、嘈杂 (半自动) 瘢痕标签的表现, 因此 AS 和 PS 提供了互补的基于形状的评价。AS 衡量了真实值 ( $T$ ) 与预测值 ( $Y$ ) 之间的像素计数相似性作为  $1 - \frac{||Y| - |T||}{(|Y| + |T|)}$ , 而 PS 则对边界像素计数应用相同的公式。

表 2: 测试分割准确性 (平均值  $\pm$  标准差) 的不同模型在所有队列中的表现。请注意, Box-SAM 和 Box-DeepLabV3+ 指的是使用地面实况边界框作为输入进行微调的模型。

	DSC	HD	AS	PS
nnU-Net	0.579 $\pm$ 0.338	11.786 $\pm$ 16.004	0.702 $\pm$ 0.345	0.748 $\pm$ 0.341
Box-SAM	0.653 $\pm$ 0.135	7.807 $\pm$ 7.922	0.793 $\pm$ 0.133	0.918 $\pm$ 0.080
Box-DeepLabV3+	0.621 $\pm$ 0.169	8.029 $\pm$ 8.818	0.838 $\pm$ 0.138	0.905 $\pm$ 0.095

### 3 结果

**基线比较。** 我们首先评估了用于分割的 nnU-Net[26], 结果汇总在表 2 中。接下来, 我们设计了一个实验, 使用真实边界框作为提示, 并对 SAM 和修改后的 DeepLabV3+[27] 进行微调, 该模型以 ResNet-50 为骨干网络并添加了处理边界框的额外输入层。这种方法显著提高了基线 nnU-Net 的表现, 如表所示。这些发现促使我们利用基于 YOLO 的边界框检测来实现 SAM 的自动化。

**YOLO-SAM.** 表 3 展示了基于微调的 YOLO-SAM 模型提出的方法的结果。尽管这个全自动管道引入了一点性能下降, 但它消除了对人工干预的需求, 与需要手动边界框输入的半自动方法不同。

图 2 展示了使用所提出的深度学习模型与 FWHM 基线真值标签对比的两个分段样本图像。可以看到, 自动分割结果与真值标签一致。然而, 预测标签中的轻微空间位移导致了由于第一样本中微小瘢痕区域重叠面积很小而产生的非常低的 Dice 分数 0.15。这表明虽然模型捕捉到了瘢痕的一般范围, 并且边界相似度高达 0.85, 但它在精确定位上存在问题。然而, 对预测标签进行更仔细检查后可以发现, 它们通常被放置在更具解剖学准确性的位置。这一观察不仅突显了噪声和半自动生成的真值标签的局限性, 还揭示了基于重叠度量如 Dice 分数评价分割准确性时存在的不足之处。相比之下, 基于大小的评估提供了更为稳健的评价, 通过独立于确切空间对齐的方式来衡量相似度。此外, 自动化预测产生的标签区域比半自动注释更加平滑和连贯, 后者往往表现为碎片化的小散件部分。尽管第二样本的 Dice 分数为

表 3: 测试不同队列中所提出方法 (YOLO-SAM) 的分割准确性 (均值  $\pm$  标准差)。

	DSC	HD	AS	PS
Cohort 1	0.579 $\pm$ 0.338	11.786 $\pm$ 16.004	0.702 $\pm$ 0.345	0.748 $\pm$ 0.341
Cohort 2	0.636 $\pm$ 0.258	9.362 $\pm$ 8.469	0.814 $\pm$ 0.243	0.865 $\pm$ 0.232
Cohort 3	0.525 $\pm$ 0.280	15.993 $\pm$ 18.012	0.753 $\pm$ 0.301	0.782 $\pm$ 0.299
Cohort 4	0.579 $\pm$ 0.406	8.931 $\pm$ 13.915	0.628 $\pm$ 0.412	0.670 $\pm$ 0.418
Cohort 5	0.596 $\pm$ 0.308	15.741 $\pm$ 15.535	0.779 $\pm$ 0.317	0.805 $\pm$ 0.271
Cohort 6	0.516 $\pm$ 0.373	15.388 $\pm$ 16.646	0.636 $\pm$ 0.373	0.680 $\pm$ 0.373
Cohort 7	0.666 $\pm$ 0.377	13.217 $\pm$ 24.006	0.767 $\pm$ 0.368	0.774 $\pm$ 0.357
Cohort 8	0.628 $\pm$ 0.361	7.599 $\pm$ 11.248	0.734 $\pm$ 0.347	0.778 $\pm$ 0.339
Total	0.601 $\pm$ 0.330	10.728 $\pm$ 14.217	0.753 $\pm$ 0.318	0.797 $\pm$ 0.311

0.52, 受到真值标签不连续性和大小影响, 但边界相似度仍然保持在较高的 0.98, 更好地反映了预测分割的质量。

我们还强调了预测分割在瘢痕检测和分类中的效率和临床影响。为此, 我们在测试集中计算了半自动 ground truth 和预测分割的瘢痕质量。半自动 (ground truth) 瘢痕的质量估计为 12.510 $\pm$ 8.784 克, 相比之下, 全自动 (预测) 瘢痕的质量为 11.337 $\pm$ 8.201 克, 表明估算非常接近。Wilcoxon 秩和检验没有揭示半自动和预测标签估计的瘢痕质量之间存在统计学上的显著差异 ( $p=0.42$ )。

**急性心肌梗死的推广。**为了评估在不同临床领域中的泛化能力, 我们在一个包含 50 名患者和 406 张图像的多中心急性心肌梗死队列数据集中对训练好的模型进行了评估。目标是比较急性与慢性病例之间的瘢痕模式差异, 并评估预测分割是否捕捉到了预期的临床区别。由于急性和慢性病例中瘢痕的空间特征, 包括大小、形状和连续性存在差异, 我们从预测的瘢痕区域中提取了多个形态学特征。这些特征包括瘢痕大小 (瘢痕区域内的像素数量)、独立病灶的数量 (连通分量的数量)、紧实度 (固体度) 定义为病灶面积与其凸包面积之比以及拉长程度 (圆度) 计算为  $4\pi \cdot \text{面积} / \text{周长}^2$ 。

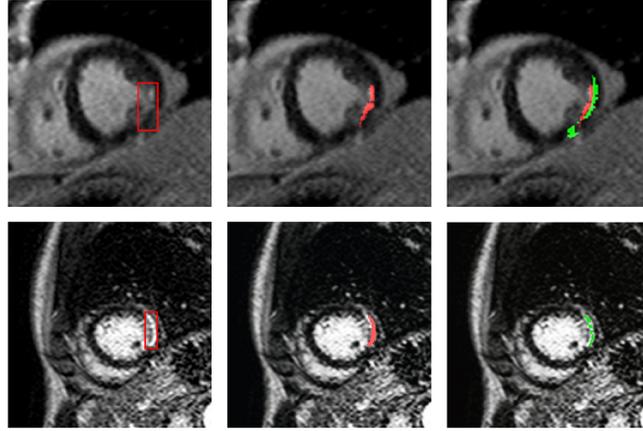


图 2: 提出的深度学习心肌瘢痕分割管线的结果。每一行显示一个样本, 包含三列: (1) 基于 YOLO 的边界框检测 (红色), (2) 基于 SAM 的自动分割 (红色), 和 (3) 半自动 FWHM 基准真实值 (绿色)。第一次预测的空间偏移导致尽管周长相似度为 0.85, Dice 分数却仅为 0.15, 这反映出模型对瘢痕范围的准确估计但定位不精确。第二次预测平滑, 实现了 0.98 的周长相似度, 但由于碎片化的基准真实值, Dice 分数较低, 为 0.52。

为了区分急性瘢痕和慢性瘢痕, 我们使用从预测分割中提取的这些特征训练了一个 XGBoost 分类器。然后, 我们应用 Shapley 方法来解释每个特征在分类过程中的重要性。图 3 显示了基于预测瘢痕检测急性病例的 Shapley 总结图, 强调了在急性病例中观察到的梗死负担和坚实度更大。提取特征的统计分析进一步证实了它们在区分急性瘢痕和慢性瘢痕中的重要性, 具体而言, 瘢痕大小、紧凑度、圆度以及连通组件数量的  $p$ -值分别为  $4.8 \times 10^{-7}$ ,  $1.4 \times 10^{-7}$ ,  $1.7 \times 10^{-4}$  和 0.03。

使用基于形状的分析对分段疤痕进行分析, 有可能开发出诊断分类器, 并预测患者风险和治疗结果。这些见解可以使心脏病专家快速区分不同类型的梗死, 这有助于诊断和治疗决策。特别是, 急性梗死经常表现出可能随着时间恢复的水肿区域, 而慢性梗死通常代表不可逆损伤。识别这些不同的模式可以指导更有效的治疗策略。该实验还强调了仅在具有更均匀且特征明确疤痕模式的慢性病例中训练模型的优势, 从而减少了来自急性疤痕常见碎片化和不明显外观的偏差。

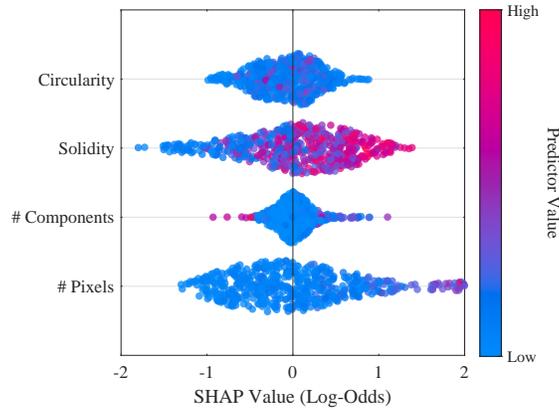


图 3: 基于预测的瘢痕分割中提取的形态特征来检测急性心肌梗死病例的 Shapley 总结图。

## 4 结论

我们开发了一个稳健的深度学习管道,通过微调最先进的 SAM 和 YOLO 模型来实现完全自动化的心肌瘢痕分割。我们的方法解决了医学图像分析中的关键挑战,包括来自半自动化注释的标签噪声、数据异质性和类别不平衡问题,采用了 KL 损失和广泛的数据增强策略。通过对急性病例和慢性病例的全面评估,该模型展示了即使存在嘈杂标签也能生成平滑且准确分割的能力。

检测模型的集成以指导分割显著提高了准确性,通过专注于瘢痕区域。此外,多样化的训练数据和系统的数据增强相结合使模型能够很好地推广到具有不同解剖学和病理特征以及不同采集参数(包括更高分辨率扫描(例如,  $0.6 \times 0.6 \times 8 \text{ mm}^3$ ))的未见测试数据。最后,使用分割的急性 MI 瘢痕进行的分类实验验证了模型区分急性与慢性病例的能力,即使仅在慢性数据上训练也是如此。这些发现强调了我们在自动心脏成像分析中的临床应用潜力。

## 致谢

A.M. 获得了莱斯特大学未来 100 名学生的奖学金。M.M.G. 获得了先驱人工智能中心 (DNRF 拨款 P1) 的支持。J.R.A. 获得了国家健康研究所

临床科学家奖 (CS-2018 – 18-ST2-007) 的支持。G.P.M. 获得了国家健康研究所研究教授奖 (RP-2017-08-ST2-007) 的支持。

## 参考文献

1. Lindstrom, M., DeCleene, N., Dorsey, H., Fuster, V., Johnson, C.O., LeGrand, K.E., Mensah, G.A., Razo, C., Stark, B., Varieur Turco, J., et al.: Global burden of cardiovascular diseases and risks collaboration, 1990-2021. *Journal of the American College of Cardiology* **80**(25) (2022) 2372–2425
2. Stone, G.W., Selker, H.P., Thiele, H., Patel, M.R., Udelson, J.E., Ohman, E.M., Maehara, A., Eitel, I., Granger, C.B., Jenkins, P.L., et al.: Relationship between infarct size and outcomes following primary PCI: patient-level analysis from 10 randomized trials. *Journal of the American College of Cardiology* **67**(14) (2016)
3. Kancharla, K., Weissman, G., Elagha, A.A., Kancherla, K., Samineni, S., Hill, P.C., Boyce, S., Fuisz, A.R.: Scar quantification by cardiovascular magnetic resonance as an independent predictor of long-term survival in patients with ischemic heart failure treated by coronary artery bypass graft surgery. *Journal of Cardiovascular Magnetic Resonance* **18**(1) (2016) 45
4. Disertori, M., Rigoni, M., Pace, N., Casolo, G., Masè, M., Gonzini, L., Lucci, D., Nollo, G., Ravelli, F.: Myocardial fibrosis assessment by LGE is a powerful predictor of ventricular tachyarrhythmias in ischemic and nonischemic LV dysfunction: a meta-analysis. *JACC: Cardiovascular Imaging* **9**(9) (2016) 1046–1055
5. Choi, K.M., Kim, R.J., Gubernikoff, G., Vargas, J.D., Parker, M., Judd, R.M.: Transmural extent of acute myocardial infarction predicts long-term improvement in contractile function. *Circulation* **104**(10) (2001) 1101–1107
6. Enein, F.A., Allaaboun, S., Khayyat, S., Andijani, M., Alkhuzai, M.M., Aljunied, A.A., Al Adhrai Sr, M., Khayyat, S.F., Andijani, M.A.: Association between myocardial scar burden and left ventricular ejection fraction in ischemic cardiomyopathy. *Cureus* **12**(12) (2020)
7. Holtackers, R.J., Emrich, T., Botnar, R.M., Kooi, M.E., Wildberger, J.E., Kretzner, K.F.: Late gadolinium enhancement cardiac magnetic resonance imaging: from basic concepts to emerging methods. In: *RöFo-Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*. Volume 194., Georg Thieme Verlag KG (2022) 491–504
8. Mewton, N., Liu, C.Y., Croisille, P., Bluemke, D., Lima, J.A.: Assessment of myocardial fibrosis with cardiovascular magnetic resonance. *Journal of the American College of Cardiology* **57**(8) (2011) 891–903

9. Khan, J.N., Nazir, S.A., Horsfield, M.A., Singh, A., Kanagala, P., Greenwood, J.P., Gershlick, A.H., McCann, G.P.: Comparison of semi-automated methods to quantify infarct size and area at risk by cardiovascular magnetic resonance imaging at 1.5 T and 3.0 T field strengths. *BMC Research Notes* **8** (2015) 1–15
10. Kramer, C.M., Barkhausen, J., Bucciarelli-Ducci, C., Flamm, S.D., Kim, R.J., Nagel, E.: Standardized cardiovascular magnetic resonance imaging (CMR) protocols: 2020 update. *Journal of Cardiovascular Magnetic Resonance* **22**(1) (2020)
11. Klem, I., Heiberg, E., Van Assche, L., Parker, M.A., Kim, H.W., Grizzard, J.D., Arheden, H., Kim, R.J.: Sources of variability in quantification of cardiovascular magnetic resonance infarct size-reproducibility among three core laboratories. *Journal of Cardiovascular Magnetic Resonance* **19**(1) (2016) 62
12. Heiberg, E., Engblom, H., Carlsson, M., Erlinge, D., Atar, D., Aletras, A.H., Arheden, H.: Infarct quantification with cardiovascular magnetic resonance using "standard deviation from remote" is unreliable: validation in multi-centre multi-vendor data. *Journal of Cardiovascular Magnetic Resonance* **24**(1) (2022) 53
13. Flett, A.S., Hasleton, J., Cook, C., Hausenloy, D., Quarta, G., Ariti, C., Muthurangu, V., Moon, J.C.: Evaluation of techniques for the quantification of myocardial scar of differing etiology using cardiac magnetic resonance. *JACC: Cardiovascular Imaging* **4**(2) (2011) 150–156
14. Chen, C., Qin, C., Qiu, H., Tarroni, G., Duan, J., Bai, W., Rueckert, D.: Deep learning for cardiac image segmentation: a review. *Frontiers in Cardiovascular Medicine* **7** (2020) 25
15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2015) 234–241
16. Lalande, A., Chen, Z., Pommier, T., Decourselle, T., Qayyum, A., Salomon, M., Ginhac, D., Skandarani, Y., Boucher, A., Brahim, K., et al.: Deep learning methods for automatic evaluation of delayed enhancement-MRI. the results of the EMIDEC challenge. *Medical Image Analysis* **79** (2022) 102428
17. Hoh, T., Margolis, I., Weine, J., Joyce, T., Manka, R., Weisskopf, M., Cesarovic, N., Fuetterer, M., Kozerke, S.: Impact of late gadolinium enhancement image acquisition resolution on neural network based automatic scar segmentation. *Journal of Cardiovascular Magnetic Resonance* **26**(1) (2024) 101031
18. Jathanna, N., Podlasek, A., Sokol, A., Auer, D., Chen, X., Jamil-Copley, S.: Diagnostic utility of artificial intelligence for left ventricular scar identification using cardiac magnetic resonance imaging—a systematic review. *Cardiovascular Digital Health Journal* **2**(6) (2021) S21–S29

19. Lalande, A., Chen, Z., Decourselle, T., Qayyum, A., Pommier, T., Lorgis, L., de La Rosa, E., Cochet, A., Cottin, Y., Ginhac, D., et al.: EMIDEC: a database usable for the automatic evaluation of myocardial infarction from delayed-enhancement cardiac MRI. *Data* **5**(4) (2020) 89
20. Stirrat, J., Joncas, S.X., Salerno, M., Drangova, M., White, J.: Influence of phase correction of late gadolinium enhancement images on scar signal quantification in patients with ischemic and non-ischemic cardiomyopathy. *Journal of Cardiovascular Magnetic Resonance* **17** (2015) 1–9
21. Schulz-Menger, J., Bluemke, D.A., Bremerich, J., Flamm, S.D., Fogel, M.A., Friedrich, M.G., Kim, R.J., von Knobelsdorff-Brenkenhoff, F., Kramer, C.M., Pennell, D.J., et al.: Standardized image interpretation and post-processing in cardiovascular magnetic resonance-2020 update: Society for cardiovascular magnetic resonance (SCMR): Board of trustees task force on standardized post-processing. *Journal of Cardiovascular Magnetic Resonance* **22**(1) (2020) 19
22. Hennemuth, A., Seeger, A., Friman, O., Miller, S., Klumpp, B., Oeltze, S., Peitgen, H.O.: A comprehensive approach to the analysis of contrast enhanced cardiacMR images. *IEEE Transactions on Medical Imaging* **27**(11) (2008) 1592–1610
23. Tian, Y., Ye, Q., Doermann, D.: YOLOv12: Attention-centric real-time object detectors. *arXiv preprint arXiv:2502.12524* (2025)
24. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: *Proceedings of the IEEE/CVF International Conference on Computer Cision*. (2023) 4015–4026
25. Author, A.: Anonymous title. *Anonymous Journal* (Anonymous Year)
26. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods* **18**(2) (2021) 203–211
27. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: *Proceedings of the European Conference on Computer Vision*. (2018) 801–818