全局与局部对比学习用于心脏 MRI 和心电图的联合表示

Alexander Selivanov[1], Philip Müller[1,2], Özgün Turgut[1,2], Nil Stolt-Ansó[1,4], and Daniel Rückert[1,2,3,4]

[1] Chair for AI in Healthcare and Medicine, Technical University of Munich (TUM) and TUM University Hospital, Munich, Germany
[2] School of Medicine, Klinikum rechts der Isar, TUM, Germany
[3] Department of Computing, Imperial College London, UK
[4] Munich Center for Machine Learning (MCML), Munich, Germany
alexander.selivanov@tum.de

摘要 心电图（ECG）是一种广泛使用且成本效益高的检测心脏电气异常的工具。然而，它不能直接测量如心室容积和射血分数等关键的功能参数，这些参数对于评估心脏功能至关重要。心脏磁共振成像（CMR）是这些测量的金标准，可以提供详细的结构和功能洞察，但其成本高昂且不太容易获取。为了解决这一差距，我们提出了 PTACL（**P**atient 和 **T**emporal**A**lignment**C**ontrastive**L**earning），这是一种多模态对比学习框架，通过整合来自 CMR 的时空信息来增强 ECG 表示。PTACL 使用全局患者级对比损失和局部时间级对比损失。全局损失通过对同一患者的 ECG 和 CMR 嵌入进行拉近，同时将不同患者的嵌入推开以对齐患者级别的表示。局部损失通过对比编码后的 ECG 片段与相应的编码后 CMR 帧来强制执行每个患者的细粒度时间对齐。这种方法不仅丰富了 ECG 表示，使其包含超越电活动的诊断信息，并且比仅靠全局对齐传递更多模态间的洞察，而且无需引入新的可学习权重。我们在来自英国生物样本库 27,951 名受试者的配对 ECG-CMR 数据上评估 PTACL。与基线方法相比，PTACL 在两项临床相关任务中表现更好：(1) 检索具有类似心脏表型的患者；(2) 预测由 CMR 导出的心脏功能参数，如心室容积和射血分数。我们的结果突显了 PTACL 通过 ECG 增强非侵入性心脏诊断的潜力。代码可在以下位置获取：https://github.com/alsalivan/ecgcmr

Keywords: 对比学习 · 时间对齐 · 心电图 · 磁共振成像

1 介绍

心血管疾病（CVD）仍然是全球主要的死亡原因之一 [28]。心电图（ECG）是一种非侵入性和成本效益高的工具，用于检测心脏的电气异常（例如，心

律失常 [1])。然而，ECG 无法直接评估如心室容积、心肌质量或射血分数等结构性和功能性心脏特性。相比之下，心脏磁共振成像（CMR）是这些评估的金标准 [11]，但它昂贵、耗时且需要专业技能，限制了其可访问性。

为了整合两种模式的信息，研究人员已将对比学习应用于全局对齐 ECG 和 CMR 嵌入 [5,18,25]，使同一患者的表示更接近而不同患者则远离。然而，这些方法将 ECG 和 CMR 视为整体表示，忽略了更精细的时间关系。由于 CMR 捕获心脏周期中的动态序列，而 ECG 记录连续的电活动，仅进行全局对齐缺乏有效整合两种模式所需的时间精度。

为了解决这一限制，我们引入了 PTACL（**P**atient 和 **T**emporal**A**ignment**C**ontrastive**L**earning），一个结合全局和局部对齐的对比学习框架。全局损失在患者级别对 ECG 和 CMR 嵌入进行对齐，而局部损失通过对比来自单个心跳表示的编码令牌——即 ECG 衍生片段——与同一心脏阶段对应的 CMR 帧表示来强制细粒度的时间对齐（图 1）。值得注意的是，局部对齐完全不需要参数，不引入任何额外的学习参数。我们的方法是完全自监督的，并在来自 UK Biobank [21] 的 27,951 名受试者的配对 ECG-CMR 数据上进行训练。我们在患者检索和表型回归上评估了 PTACL，以评估 ECG 嵌入捕捉 CMR 衍生的心脏功能的程度。我们的结果显示局部对比对齐改进了 ECG 表示，增强了其用于评估心脏结构和功能的诊断效用。

2 相关工作

自监督学习在单模态和多模态表示学习方面都取得了进展。对于单模态任务，掩码自动编码器（MAE）[7] 已被证明能够从未标记的数据中学习有意义的表示。它们已被广泛应用于诸如心电图（ECG）[4,16,19,26,30,32]、语音和音频 [8] 等信号，以及医学成像领域，在该领域三维 MAE [6,24] 能够从体积扫描 [15,22,33] 中学习时空特征。

在多模态学习中，大多数对比方法都受到 CLIP [17] 的启发，它将来自不同模态的每个患者的表示对齐到一个共享嵌入空间中，在拉近相似样本的同时推开不相关的样本。这种框架已在医疗领域得到应用，例如心电图-心脏磁共振成像融合 [5,18,25] 和 X 光 – 文本关联 [23,27,34]，使零样本分类和改进的诊断洞察成为可能。然而，这些方法主要侧重于全局对齐，将每种模态视为单一表示，并且可能会忽略细粒度的局部交互。

为了解决这一问题，最近的研究开始探索医学应用中的局部对比学习。例如，GLORIA [9] 将全局-局部学习方案应用于图像-文本配对，而 Seibold

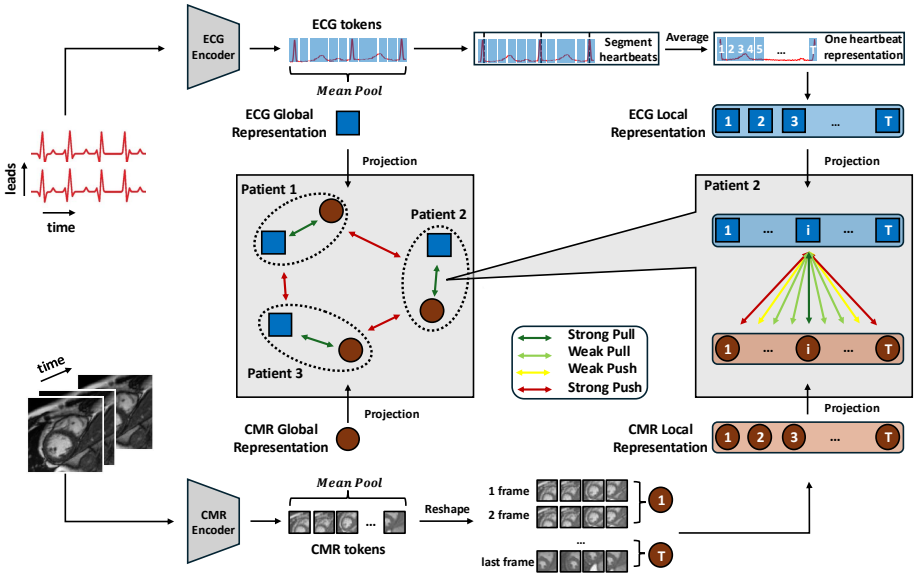

图 1. PTACL 概述。我们使用全局对比损失（左侧正方形）和局部对比损失（右侧正方形）的组合，在心电图和心脏磁共振成像配对数据上训练一个多模态模型。全局损失通过将来自同一患者的表示拉得更近，同时将来自不同患者的表示推开，从而对齐患者级别的 ECG 和 CMR 嵌入。局部损失通过对比从单一心跳表示中编码的心电图片段及其对应的心脏磁共振成像帧，强制执行细粒度的时间对齐。在图表中，1 到 T 的表示始于舒张期末（ED）阶段。令牌中的叠加层（CMR 和 ECG）仅用于可视化目的。

等人。[20] 则通过对比局部图像特征与放射学报告来改进 CLIP，LoVT [14] 将图像子区域与文本嵌入对齐。这些方法反映了视觉-语言研究领域的更广泛趋势 [12, 29, 31]。

3 方法

我们的框架由三个阶段组成：单模态预训练（SMP）（章节 3.1）、多模态对比学习（章节 3.2）和推理。首先，我们使用 MAE [7] 在未标记数据上预训练单独的 ECG 和 CMR 编码器。然后，我们使用大量的增强方法联合训练这两个编码器，应用全局对比损失进行患者级别对齐，以及局部对比损失进行细粒度时间对齐。局部损失将 ECG 导出的时间段——来自多导 ECG 单心跳表示的固定长度段——与 CMR 在相同心脏相位的相应编码帧进行对比。最后，我们在多个下游任务上评估预训练的 ECG 编码器。

3.1 心电图和心脏磁共振成像的预训练

为了整合 ECG 和 CMR，每种模式首先通过单模态预训练（SMP）独立学习有意义的表示。这一步使得编码器能够在多模态对比学习之前捕捉到特定于某种模式的特征。遵循先前的工作 [8,24,32]，我们在每种模式上分别对 ECG 和 CMR 编码器进行预训练，两种情况下均使用 MAE 框架 [7]。

我们将多导联心电图信号通过时间维度上的 1D 卷积嵌入，卷积核大小为 p_t，将每个导联转换成一系列不重叠的时间基标记。该卷积独立应用于每个导联，并在所有导联之间共享权重。为了学习时序和导联结构，在随机屏蔽部分标记之前添加可学习的位置嵌入。

对于 CMR，我们将 2D+T 短轴序列视为多帧输入，并应用一个核大小为 $[t, p, p]$ 的 3D 卷积，其中 t 定义了时间窗口，p 控制空间补丁的大小。这生成了一串时空标记序列，每个标记代表心脏在一段时间内的局部区域。为了学习空间和时间结构，在掩码之前我们添加可学习的位置嵌入。

在这两种情况下，变压器编码器仅处理可见标记，而轻量级变压器解码器重构被屏蔽的标记。均方误差（MSE）损失仅在被屏蔽的标记上计算，比较重构输出与原始未损坏输入。

3.2 多模态对比预训练

为了学习鲁棒的跨模态表示，我们采用了两种对比损失：一种是用于患者级别对齐的全局损失，另一种是用于细粒度时间对齐的局部损失。全局损失确保来自同一患者的嵌入相似，同时将不同患者的嵌入推开。局部损失通过匹配由心电图导出的时间嵌入与对应的 CMR 帧进一步优化对齐。心电图局部嵌入由 T 段组成，每段跨越单个心跳的多导联心电图中的 p_t 时间步长，而 CMR 嵌入则代表心脏序列中的编码帧。

全局对比损失： 对于一批 B，令 $z_{\mathcal{E}}, z_{\mathcal{C}}$ 为 ECG 和 CMR 嵌入，通过对标记表示进行平均池化，然后通过单独的投影层获得。我们使用 InfoNCE 损失来强制跨模态对齐，遵循 [17]。该损失对称地定义用于 ECG $\to$ CMR 和 CMR $\to$ ECG 方向。ECG $\to$ CMR 对比损失为：

$$\mathcal{L}_{\mathrm{ECG}\to\mathrm{CMR}} = \frac{1}{B}\sum_{i=1}^{B} -\log \frac{\exp\left(\cos\left(z_{\mathcal{E}}^{(i)}, z_{\mathcal{C}}^{(i)}\right)/\tau\right)}{\sum_{j=1}^{B}\exp\left(\cos\left(z_{\mathcal{E}}^{(i)}, z_{\mathcal{C}}^{(j)}\right)/\tau\right)}, \tag{1}$$

其中 τ 是温度参数，$z_{\mathcal{E}}$ 和 $z_{\mathcal{C}}$ 是投影嵌入，而 $\cos(\cdot,\cdot)$ 表示余弦相似度。最终的全局对比损失 $\mathcal{L}_{\text{global}}$ 是两个方向的加权组合。

局部对比损失： CMR 采集由心电门控引导，其中在多次心跳期间连续记录成像数据，同时捕捉心电信号。检测到的 R 波峰值稍后用于对齐和重建单个代表性的心跳周期。这种生理对齐激发了我们局部对比损失函数的制定。

局部嵌入形成： 对于 CMR，每一帧被划分为表示 $p \times p$ 图像块的空间标记和捕捉 t 帧序列的时间标记。局部嵌入是通过对每个时间标记内的空间标记进行平均获得的，生成 $z_{\mathcal{C}}^{(i,t)}$，其中 t 索引 CMR 帧。对于 ECG，每个标记表示来自单一导联长度为 p_t 的片段。为了在患者之间标准化心搏表现形式，我们提取每颗检测到的心脏跳动之间的连续 R 波峰间的标记。这些变长序列被插值到固定长度的 T 个标记。最后，我们计算多个心跳的平均值以获得患者 i 的心电图局部嵌入，记为 $z_{\mathcal{E}}^{(i,k)}$，其中 k 索引心电图时间片段。

时间对齐和局部损失： 遵循 [10]，我们使用对称的对齐矩阵 $\mathcal{P}_{k,t}$ 强制执行 $z_{\mathcal{E}}^{(i,k)}$(ECG) 和 $z_{\mathcal{C}}^{(i,t)}$(CMR) 局部嵌入之间的时序对齐：

$$\mathcal{P}_{k,t} \propto \begin{cases} \delta_{k,t} & \sigma = 0 \quad \text{(hard alignment)} \\ \exp\left(-\frac{1}{2}\left(\frac{\text{dist}_{k,t}}{\sigma}\right)^2\right) & \sigma > 0 \quad \text{(soft alignment)} \end{cases} \tag{2}$$

这里 $\sum_{t=1}^{T} \mathcal{P}_{k,t} = 1$ 和 $\delta_{k,t}$ 确保了 $\sigma = 0$ 的精确匹配，而对于 $\sigma > 0$，则使用基于归一化时间距离 $\text{dist}_{k,t}$ 的高斯加权来软化对齐（图 2）。

局部对比损失对 ECG $\to$ CMR 和 CMR $\to$ ECG 方向定义为对称。每个患者、每个片段的局部损失对于 ECG $\to$ CMR：

$$\ell_{\text{ECG}\to\text{CMR}}^{(i,k)} = -\sum_{t=1}^{T} \mathcal{P}_{k,t} \log \left[\frac{\exp\left(\cos(z_{\mathcal{E}}^{(i,k)}, z_{\mathcal{C}}^{(i,t)})/\tau\right)}{\sum_{t'=1}^{T} \exp\left(\cos(z_{\mathcal{E}}^{(i,k)}, z_{\mathcal{C}}^{(i,t')})/\tau\right)} \right]. \tag{3}$$

总的局部对比损失对于 ECG $\to$ CMR 是通过对批次中所有心电图片段 k 和患者的 $\ell_{\text{ECG}\to\text{CMR}}^{(i,k)}$ 进行平均得到的。最终的局部对比损失 $\mathcal{L}_{\text{local}}$ 是两个方向的加权组合。

总损失： 最终目标结合了全局和局部对比损失，以确保在患者和时间层面上的对齐。总损失为：

$$L = \mathcal{L}_{\text{global}} + \beta \cdot \mathcal{L}_{\text{local}} \tag{4}$$

其中 β 控制局部对比损失的相对贡献。

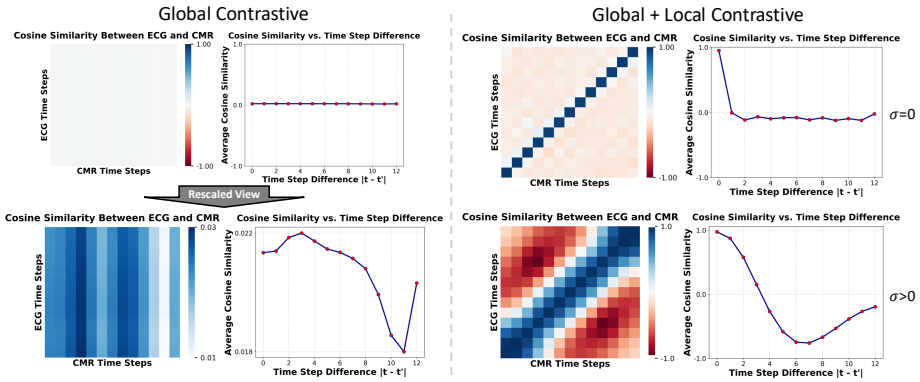

图 2. 局部对比损失对两种模态在推理过程中时间步长之间的时间对齐的影响。左侧仅使用全局对比学习，显示了模态之间没有明显的时间对应关系。右侧展示了添加局部对比损失（PTACL）的效果，导致对齐度提高。对于 $\sigma = 0$，模型强制执行严格的一对一时序匹配，而对于 $\sigma > 0$，高斯加权使对齐更加平滑。我们的方法实现了 ECG 和 CMR 时序步长之间的时间对齐，这在仅使用全局方法的情况下是不存在的。

4 实验

我们使用来自英国生物库的配对 12 导联静息心电图和短轴心脏磁共振数据进行模型评估 [21]，其中 27,951 名患者用于训练，5,991 名患者用于测试。CMR 序列包含 50 帧，围绕心脏中心裁剪中间短轴切片（84×84）并进行最小-最大归一化。心电图是 10 秒的 12 导联记录，采样率为 500 Hz。心电图通过 8 层小波变换进行预处理，遵循 [13, 32]，用于去除基线漂移和噪声，然后进行 z 分数归一化。CMR 增强包括随机调整大小的裁剪、时间采样、旋转、翻转、颜色抖动、高斯模糊和噪声，而心电图增强包括傅里叶替代变换、随机裁剪（2500 点）、抖动和重缩放。我们评估 10 个源自 CMR 的心脏表型 [2]：左心室/右心室舒张末期和收缩末期容积（EDV、ESV）、每搏输出量（SV）、射血分数（EF）、左心室质量（M）和左心室心输出量（CO）。

4.1 相似患者检索

为了评估我们学习到的 ECG 嵌入的临床实用性，我们测试它们检索具有相似 CMR 衍生表型患者的能力。对于每个查询 ECG，我们使用余弦相似度对 CMR 嵌入数据库中的所有患者进行排序。如果候选人的表型在查询值的 $\pm 0.5\sigma$ 范围内，则算作匹配，其中 σ 是该表型在整个测试数据中的标准

差。我们比较了两种嵌入策略：一个基线模型（全局）和 PTACL（全局+局部）。虽然我们的全局模型已经表现出强大的检索性能，但加入局部损失进一步提高了精确度和排序效果。这种改进具有临床意义，因为它能够实现更准确的患者分层。表 1 总结了关键心脏表型的 Precision@k(P@k)，平均排名 (MnR) 和中位数排名 (MdR)。供参考，括号中的数值表示检索 k 随机患者时的性能，仅报道了 $k = 1$ 和 $k = 10$ 以节省空间。

表 1. 从 CMR 衍生的心脏表型检索结果使用 ECG 嵌入。我们将基线（全局 △）和 PTACL（全局+局部 □）模型进行比较，报告性能为 △/□（随机），括号内是随机检索的结果。匹配被定义为候选者表型与查询差异不超过 $\pm 0.5\,\sigma$ 的情况。在 PTACL 中包含局部损失始终能够提高患者检索的精度，并降低所有表型的排名分数，展示了其增强临床患者分层的潜力。

	P@1↑	**P@3↑**	**P@5↑**	**P@10↑**	**P@15↑**	**锰 R↓**	**MdR↓**
EDV_{LV}	0.492 / **0.510** (0.291)	0.462 / **0.463**	**0.445** / 0.444	0.426 / **0.427** (0.287)	**0.416** / 0.415	3.1 / **2.9**	2.0 / **1.0**
ESV_{LV}	0.501 / **0.524** (0.303)	0.469 / **0.480**	0.456 / **0.461**	0.438 / **0.442** (0.284)	0.429 / **0.432**	3.0 / 3.0	1.0 / 1.0
SV_{LV}	0.458 / **0.472** (0.294)	0.417 / **0.423**	0.404 / **0.405**	0.387 / **0.388** (0.296)	**0.380** / 0.377	3.4 / **3.2**	2.0 / 2.0
EF_{LV}	0.431 / **0.436** (0.278)	0.392 / **0.396**	**0.378** / 0.377	0.355 / **0.359** (0.282)	0.349 / **0.351**	3.9 / **3.8**	2.0 / 2.0
CO_{LV}	0.439 / **0.453** (0.295)	0.407 / **0.417**	0.391 / **0.394**	0.377 / **0.378** (0.284)	0.368 / 0.368	3.9 / **3.6**	2.0 / 2.0
M_{LV}	0.547 / **0.561** (0.293)	0.516 / **0.522**	0.498 / **0.504**	0.479 / **0.486** (0.294)	0.471 / **0.474**	2.6 / 2.6	1.0 / 1.0
EDV_{RV}	0.500 / **0.516** (0.280)	0.467 / **0.477**	0.450 / **0.459**	0.433 / **0.441** (0.287)	0.426 / **0.430**	3.0 / **2.8**	2.0 / **1.0**
ESV_{RV}	0.518 / **0.531** (0.290)	0.485 / **0.494**	0.471 / **0.476**	0.455 / **0.460** (0.281)	0.446 / **0.451**	2.8 / **2.6**	1.0 / 1.0
SV_{RV}	0.458 / **0.473** (0.285)	0.420 / **0.428**	0.403 / **0.408**	0.386 / **0.390** (0.289)	0.380 / **0.382**	3.8 / **3.7**	2.0 / 2.0
EF_{RV}	0.433 / **0.443** (0.288)	0.397 / **0.405**	0.383 / **0.390**	0.370 / **0.373** (0.290)	0.364 / **0.365**	4.3 / **4.2**	2.0 / 2.0

4.2 心脏表型的回归

我们评估了不同的预训练策略在各种心脏表型上的回归性能，使用 R^2 分数。表 2 展示了基于 ECG、CMR 或配对的 ECG-CMR 数据训练的模型的结果。CMR 单模态预训练（SMP）揭示了成像信息的上限，而 ECG SMP 性能较低则突显了单独使用 ECG 的挑战。多模态学习旨在缩小这一差距。在自监督方法中，MAE [7] 在两种模式下始终优于 SimCLR [3]，尽管全监督单模态训练仍然是一个稳健的基础模型。我们的全局模型作为一个强大的基础模型，已经超越了之前的工作。添加局部对比损失（PTACL）进一步提高了所有心脏表型的性能。重要的是，PTACL 在不引入额外可学习参数的情况下实现了这些改进，使其成为一个计算效率高的增强。

表 2. 回归性能（R^2）的不同预训练方法。模型通过线性探测 (LP) 或全微调 (FN) 使用注意力池化 (AP) 进行评估。我们的 Global 模型优于之前的多模态方法。添加局部对比损失 (PTACL) 在所有表型上进一步提高结果，而无需引入额外的可学习参数。

输入	方法	评估	左心室舒张末期容积	左心室舒张末期容积	LVSV	LVEF	LVCO	LVM	RVEDV	RVESV	RVSV	RVEF
				CMR R^2 ↑								
CMR	Random Init	LP	0.015	0.017	0.009	0.008	0.003	0.021	0.018	0.020	0.009	0.007
CMR	SimCLR [3]	LP	0.572	0.612	0.466	0.513	0.418	0.628	0.618	0.681	0.463	0.520
CMR	MAE [7]	LP	0.809	0.772	0.678	0.489	0.571	0.824	0.799	0.784	0.657	0.492
CMR	Supervised	FN	0.814	0.789	0.665	0.470	0.536	0.823	0.801	0.808	0.623	0.473
				心电图 R^2 ↑								
ECG	Random Init ($\times 10^{-5}$)	LP	-4.86	-3.89	-7.03	-38.8	-7.61	18.3	0.936	3.61	-9.66	0.983
ECG	SimCLR [3]	LP	0.278	0.261	0.196	0.085	0.156	0.343	0.308	0.317	0.205	0.130
ECG	MAE [7]	LP	0.439	0.425	0.332	0.218	0.252	0.524	0.466	0.481	0.333	0.254
ECG	Supervised	FN	0.458	0.443	0.333	0.198	0.252	0.516	0.481	0.491	0.340	0.237
				多模态（心电图+心脏磁共振成像）R^2 ↑								
ECG	ECCL [5]*	FN	0.372	0.348	0.270	0.176	0.212	0.449	0.397	0.397	0.281	0.203
ECG	CMAE [18]†	LP	0.451	0.380	0.316	0.103	0.281	0.536	0.490	0.445	0.320	0.129
ECG	MMCL [25]‡	FN+AP	0.498	0.497	0.360	0.245	-	0.597	0.527	0.534	0.375	0.248
ECG	PTACL (w/o SMP)[1]	LP	0.359	0.334	0.256	0.114	0.187	0.406	0.381	0.386	0.253	0.149
ECG	PTACL (w/ CosSim++)[2]	LP	0.509	0.496	0.383	0.252	0.286	0.614	0.537	0.553	0.387	0.295
ECG	Global[3]	LP	0.507	0.499	0.377	0.258	0.281	0.612	0.534	0.553	0.380	0.297
心电图	**PTACL**[3]	**线性规划**	**0.514**	**0.500**	**0.386**	**0.255**	**0.289**	**0.616**	**0.540**	**0.557**	**0.389**	**0.299**
心电图	**PTACL**[3]	**功能+近似值**	**0.544**	**0.537**	**0.408**	**0.283**	**0.310**	**0.645**	**0.568**	**0.582**	**0.408**	**0.315**

† 预训练于长轴切片，并在多模态预训练期间使用重构损失。

‡ 预训练于三个短轴切片。

* 预训练在一个中间短轴切片上。R^2 值基于报告的皮尔逊相关系数计算为 r^2。

[1] 在单个中间短轴切片上进行预训练。不包括单一模态预训练 (SMP) 的 I 相。

[2] 预训练于单个中间短轴切片的时间序列。使用局部损失最大化余弦相似性以匹配时间步长（无负样本）。

[3] 预训练于单个中间短轴切片的时间上。标准差未显示，且在 ∼0.001–0.002。带有 SMP 的 I 相。

4.3 实现细节

我们的方法采用基于变压器的架构来处理两种模式。CMR 模型有 4 层，8 个注意力头，一个 patch$[2, 12, 12]$，以及一个 90%的掩码比例（13M 参数）。ECG 模型有 6 层，8 个头，一个 50 大小的 patch，以及一个 75%的掩码比例（19M 参数）。两者都使用一层、四个头的解码器。在多模态预训练期间，前两层被冻结，分别为 CMR 和 ECG 留下 6.3M 和 12.6M 个可训练参数。对比损失遵循 $\mathcal{L}_{out}^{sup}$ 变体来自 [10]。我们使用 $T = 13, \beta = 1.0$。回归使用了 $\sigma = 0.1$ 和全局/局部温度为 0.1/1.0，检索使用了 $\sigma = 0.0$ 和温度为 0.07/0.07。线性探测使用了普通最小二乘回归。

5 结论与讨论

我们介绍了 PTACL，这是一个对比学习框架，通过患者层面和时间步长的对齐来增强 ECG 表示，与 CMR 相匹配。与之前仅关注全局对齐的方法不同，PTACL 捕捉更精细的时间对应关系，从而在检索和表型回归方面表现更好，而且无需增加额外的可学习参数。我们的方法只依赖于 CMR 中的

单个中间短轴切片，表明即使是最少量的成像数据也能显著丰富 ECG。然而，PTACL 依赖配对的 ECG-CMR 数据，这可能限制了其适用性。未来的工作可以将 PTACL 扩展到其他模态，如超声波，探索三模态学习，或开发策略以减少对配对多模态数据集的依赖。

Acknowledgments. 此项研究是在英国生物银行资源下编号为 87802 的申请中进行的。本项目由欧洲研究理事会 (ERC) 资助的 Deep4MI 项目 (884622) 提供资金支持。A. S. 通过巴伐利亚自由州的 EVUK 计划 ("下一代 Al 集成诊断") 获得资助。

Disclosure of Interests. 作者声明与本文内容相关的不存在任何利益冲突。

参考文献

1. Ansari, Y., Mourad, O., Qaraqe, K., Serpedin, E.: Deep learning for ECG arrhythmia detection and classification: An overview of progress for period 2017–2023. Front. Physiol. **14**, 1246746 (2023)
2. Bai, W., Suzuki, H., Huang, J., Francis, C., Wang, S., Tarroni, G., et al.: A population-based phenome-wide association study of cardiac and aortic structure and function. Nat. Med. **26**(10), 1654–1662 (2020)
3. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: ICML. pp. 1597–1607. PmLR (2020)
4. Chien, H.Y.S., Goh, H., Sandino, C.M., Cheng, J.Y.: MAEEG: Masked auto-encoder for EEG representation learning. In: NeurIPS Workshop (2022), https://arxiv.org/abs/2211.02625
5. Ding, Z., Hu, Y., Li, Z., Zhang, H., Wu, F., Xiang, Y., et al.: Cross-modality cardiac insight transfer: A contrastive learning approach to enrich ECG with CMR features. In: MICCAI 2024. vol. LNCS 15003, pp. 109–119. Springer (2024)
6. Feichtenhofer, C., Fan, H., Li, Y., He, K.: Masked autoencoders as spatiotemporal learners. In: Koyejo, S., et al. (eds.) Advances in NeurIPS. vol. 35, pp. 35946–35958 (2022)
7. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: CVPR. pp. 16000–16009 (2022)
8. Huang, P.Y., Xu, H., Li, J., Baevski, A., Auli, M., Galuba, W., et al.: Masked autoencoders that listen. In: NeurIPS. vol. 35, pp. 28708–28720 (2022)
9. Huang, S.C., Shen, L., Lungren, M.P., Yeung, S.: GLoRIA: A multimodal global-local representation learning framework for label-efficient medical image recognition. In: ICCV. pp. 3942–3951 (2021)

10. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., et al.: Supervised contrastive learning. In: Larochelle, H., et al. (eds.) Advances in NeurIPS. vol. 33, pp. 18661–18673 (2020)

11. Leiner, T., Bogaert, J., Friedrich, M.G., Mohiaddin, R., et al.: SCMR position paper (2020) on clinical indications for cardiovascular magnetic resonance. J. Cardiovasc. Magn. Reson. **22**(1), 76 (2020)

12. Ma, Y., Xu, G., Sun, X., Yan, M., Zhang, J., Ji, R.: X-CLIP: End-to-end multi-grained contrastive learning for video-text retrieval. In: ACM MM. pp. 638–647. ACM (2022)

13. Martis, R.J., Acharya, U.R., Min, L.C.: ECG beat classification using PCA, LDA, ICA and discrete wavelet transform. Biomed. Signal Process. Control **8**(5), 437–448 (2013)

14. Müller, P., Kaissis, G., Zou, C., Rueckert, D.: Joint learning of localized representations from medical images and reports. In: Avidan, S., et al. (eds.) ECCV. pp. 685–701. Springer (2022)

15. Munk, A., Ambsdorf, J., Llambias, S.N., Nielsen, M.: AMAES: Augmented masked autoencoder pretraining on public brain MRI data for 3D-native segmentation. arXiv preprint arXiv:2408.00640 (2024)

16. Na, Y., Park, M., Tae, Y., Joo, S.: Guiding masked representation learning to capture spatio-temporal relationship of electrocardiogram. In: ICLR (2024)

17. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., et al.: Learning transferable visual models from natural language supervision. In: ICML. pp. 8748–8763. PmLR (2021)

18. Radhakrishnan, A., Friedman, S.F., Khurshid, S., Ng, K., Batra, P., Lubitz, S.A., et al.: Cross-modal autoencoder framework learns holistic representations of cardiovascular state. Nat. Commun. **14**(1), 2436 (2023)

19. Sawano, S., Kodera, S., Takeuchi, H., Sukeda, I., Katsushika, S., et al.: Masked autoencoder-based self-supervised learning for electrocardiograms to detect left ventricular systolic dysfunction. In: NeurIPS Workshop (2022)

20. Seibold, C., Reiß, S., Sarfraz, M.S., Stiefelhagen, R., Kleesiek, J.: Breaking with fixed set pathology recognition through report-guided contrastive training. In: MICCAI. pp. 690–700. Springer (2022)

21. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., et al.: UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. **12**(3), e1001779 (2015)

22. Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., et al.: Self-supervised pre-training of Swin transformers for 3D medical image analysis. In: CVPR. pp. 20730–20740 (2022)

23. Tiu, E., Talius, E., Patel, P., Langlotz, C.P., Ng, A.Y., Rajpurkar, P.: Expert-level detection of pathologies from unannotated chest X-ray images via self-supervised learning. Nat. Biomed. Eng. **6**(12), 1399–1406 (2022)

24. Tong, Z., Song, Y., Wang, J., Wang, L.: VideoMAE: Masked autoencoders are data-efficient learners for self-supervised video pre-training. Advances in NeurIPS **35**, 10078–10093 (2022)

25. Turgut, Ö., Müller, P., Hager, P., Shit, S., Starck, S., Menten, M.J., et al.: Unlocking the diagnostic potential of electrocardiograms through information transfer from cardiac magnetic resonance imaging. Med. Image Anal. **101**, 103451 (2025)

26. Wang, G., Wang, Q., Iyer, G., Nag, A., John, D.: Unsupervised pre-training using masked autoencoders for ECG analysis. In: BioCAS. pp. 1–5 (2023)

27. Wang, Z., Wu, Z., Agarwal, D., Sun, J.: MedCLIP: Contrastive learning from unpaired medical images and text. In: Goldberg, Y., et al. (eds.) EMNLP. p. 3876. ACL (2022)

28. World Health Organization: WHO Mortality Database (2025), https://www.who.int/data/data-collection-tools/who-mortality-database, accessed: Jan. 31, 2025

29. Yang, J., Bisk, Y., Gao, J.: TACo: Token-Aware cascade contrastive learning for video-text alignment. In: ICCV. pp. 11562–11572 (2021)

30. Yang, S., Lian, C., Zeng, Z., Xu, B., Su, Y., et al.: Masked self-supervised ECG representation learning via multiview information bottleneck. Neural Comput. Appl. **36**(14), 7625–7637 (2024)

31. Yao, L., Huang, R., Hou, L., Lu, G., Niu, M., Xu, H., et al.: FILIP: Fine-grained interactive language-image pre-training. In: ICLR (2022)

32. Zhang, H., Liu, W., Shi, J., Chang, S., Wang, H., He, J., et al.: MaeFE: Masked autoencoders family of electrocardiogram for self-supervised pretraining and transfer learning. IEEE Trans. Instrum. Meas. **72**, 1–15 (2023)

33. Zhang, Y., Chen, C., Shit, S., Starck, S., Rueckert, D., Pan, J.: Whole heart 3D+T representation learning through sparse 2D cardiac MR images. In: MICCAI. vol. 15001, pp. 359–369. Springer, Cham (2024)

34. Zhang, Y., Jiang, H., Miura, Y., Manning, C.D., Langlotz, C.P., et al.: Contrastive learning of medical visual representations from paired images and text. In: Lipton, Z., et al. (eds.) MLHC. vol. 182, pp. 2–25. PMLR (2022)