

通过多窗口对比学习来学习 ECG 表示

Yi Yuan^{*†}, Joseph Van Duyn^{*†}, Runze Yan[‡], Zhuoyi Huang[¶], Sulaiman Vesal[¶], Sergey Plis[§],
Xiao Hu[‡], Gloria Hyunjung Kwak[‡], Ran Xiao[‡], Alex Fedorov[‡]

^{*}Department of Quantitative Theory & Methods

[†]Department of Mathematics

[‡]Center for Data Science, Nell Hodgson Woodruff School of Nursing

Emory University, Atlanta, GA, USA

[§] Department of Computer Science, Georgia State University, Atlanta, GA, USA

[¶] Microsoft

摘要—心电图 (ECG) 分析是心血管疾病诊断的基础,但深度学习模型的表现往往受限于标注数据的有限获取。自监督对比学习作为一种强大的方法,已经出现用于从无标签信号中学习鲁棒的心电图表示。然而,大多数现有方法仅生成成对增强视图,并未能充分利用心电图记录中的丰富时间结构。在这项工作中,我们提出了一种多窗口对比学习框架。我们从每个 ECG 实例中提取多个时间窗口以构建正样本对并通过统计最大化它们的一致性。受到慢特征分析原理的启发,我们的方法明确鼓励模型学习跨越时间保持不变且具有生理意义的特征。我们在 PTB-XL 数据集上通过广泛的实验和消融研究验证了我们的方法。我们的结果表明,多窗口对比学习在多标签超类分类中始终优于传统的两视图方法,实现了更高的 AUROC (0.891 vs. 0.888) 和 F1 分数 (0.680 vs. 0.679),同时所需预训练周期减少了高达四倍 (32 vs. 128),总计算时间减少 14.8%。尽管每个样本处理多个窗口,我们实现了训练周期数和总计算时间的显著减少,使得我们的方法适用于训练基础模型。通过广泛的消融研究,我们确定了最佳设计选择并展示了在各种超参数上的鲁棒性。这些发现确立了多窗口对比学习作为一种高效且可扩展的心电图自动化分析范式,并为生物医学时序数据中的自监督表示学习提供了一个有前景的通用框架。

Index Terms—心电图, 自监督学习, 对比学习, 深度学习

I. 介绍

心血管疾病 (CVD) 仍然是全球首要的死亡原因。心电图 (ECG) 分析在现代临床护理中扮演着关键角色,因为 ECG 是一种广泛可获取的、非侵入性的且成本效益高的工具,用于检测心脏异常和进行心血管疾病风险分层 [1]。然而,准确解读心电图耗时且需要专业

技能,并容易受到人为错误的影响 [2]。深度学习 (DL) 领域的最新进展显著提高了对细微病理的 ECG 分析能力,从而有助于更早诊断 [3]–[5]。然而,深度学习模型依赖于大型标记数据集限制了它们在现实世界医疗环境中的实际部署,因为高质量的专业注释稀缺且获取成本高昂 [6]。

自监督学习 (SSL) 在缺乏手动标签的情况下已成为表示学习的强大范式 [7]–[11]。特别是,对比学习方法如 SimCLR [10]、MoCo [12] 及其对时间序列数据的适应 [6], [13] 通过最大化同一实例的不同增强视图之间的协议,展示了强大的性能,从而从无标签数据中学习稳健且通用的特征。

大多数生物医学时间序列分析框架 [13]–[16] 仅为每个输入生成两个增强视图。这种常见方法未能充分利用心电图信号中发现的丰富时序结构。在临床实践中,典型的 10 秒心电图记录捕捉到多个心脏周期,这些周期的节奏和形状在整个记录过程中通常相似。如果存在病理特征,它们也倾向于出现在整个信号中。因此,在一次心电图中可以观察到正常和异常的心脏动力学在多个独立的时间窗口内。仅关注成对比较的传统对比方法未能利用这种持续存在的生理结构。心电图的每个段落都是由心脏的基本电活动产生的。因此,我们期望一个使用多段信息的模型能够学习更好地反映真实生理来源的特征。

最近的理论工作 [17] 表明,通过使用每个样本的两个以上正视图可以改进对比学习。通过使用多个视图,

这些方法可以捕捉到每个样本内部更多的变异性，从而实现更快和更有效的学习。多视角对比学习可以改善自然图像的学习表示，但迄今为止，它尚未广泛应用于生理时间序列数据，如心电信号。

心脏信号中的时间模式也提供了学习稳定特征的机会。这一想法与慢特征分析 (SFA) [18] 的概念相关，其目标是找到随时间变化缓慢的特征，从而捕捉到时间序列数据的主要结构。对于心电图分析而言，这一点很重要，因为心跳节奏和心律不齐通常会持续几个周期，而噪声和伪影则是短暂且不一致的。

受 SFA [18], [19] 和多视角对比学习 [17] 的启发，我们提出了一种用于心电图表示学习的**多窗口对比学习**框架。我们从每个心电图中提取多个时间上多样化的窗口，并通过算术或几何平均值来最大化这些成对之间的相似性。我们认为这允许模型捕捉到在整个心电图记录过程中持续存在的缓慢且具有生理相关性的特征。与传统的仅限于两两视角的对比方法不同，我们的多窗口策略直接利用了临床心电图的时间冗余和结构。

我们在大规模的 PTB-XL 数据集 [20] 上验证了我们的方法，展示了在多标签超类分类中相对于强大基线的一致性改进。通过广泛的消融和统计分析，我们表明多窗口对比学习能够实现快速收敛，提高线性评估性能，并且对超参数选择具有鲁棒性。我们的研究结果支持多视角对比目标中的慢特征分析原则。我们在 PTB-XL 上实现了对比方法的最新性能，并将训练加速了 14.8%，提高了样本效率。

II. 方法

A. 多窗口对比学习

我们提出了一种多窗口对比学习目标 [17]，该目标通过从每个心电图记录中提取 M 个时间窗口来扩展标准的两视图对比框架。与依赖单一正对不同，我们在所有窗口对之间计算相似性并通过统计量（如算术或几何平均）进行聚合。这鼓励编码器学习在局部扭曲和较长的时间偏移下都稳定的特征。

在心电图中，心律和形态模式通常在整个心脏周期内持续存在，而噪声和伪影则是暂时的。在多个窗口间强制相似性利用了这种生理冗余。我们的方法进一步受到慢特征分析 (SFA) 原理 [18] 的启发，旨在提取随时间变化缓慢的特征。通过强制多个窗口之间的一致性，模型捕获稳定的 cardiac 动态而不是过度拟合局部

波动。这种多窗口策略直接利用了心电图的时间冗余，并生成了在时间上不变且具有生理意义的表现形式。

我们的框架如图 1 所示。对于大小为 N 的批次中的每个心电图，我们采样 M （可能重叠）的时间窗口，结果得到 NM 个窗口。每个窗口被独立编码，生成一个 L2 归一化的嵌入矩阵 $\mathbf{Z} \in \mathbb{R}^{(NM) \times d}$ ，其中 d 是嵌入维度。我们计算一个成对相似性矩阵：

$$\mathbf{S} = \frac{1}{\tau} \mathbf{Z} \mathbf{Z}^\top, \quad (1)$$

其中 τ 是温度参数， S_{ab} 是表示 z_a 和 z_b 之间的缩放余弦相似性。

令 $M_{\text{pos}} \in \{0, 1\}^{(NM) \times (NM)}$ 为一个二进制掩码，其中当 z_a 和 z_b 是同一信号的不同窗口时 $M_{\text{pos}}(a, b) = 1$ ，否则为 0。对于每个锚点 z_a ，定义 $\mathcal{P}(a)$ 为索引集 b ，其中 $M_{\text{pos}}(a, b) = 1$ （即与 z_a 同一样本的所有其他窗口）。自我相似性掩码由 $a = b$ 定义。

为了超越标准的两视图对比损失，我们引入了一种多窗口公式，该公式使用选定的统计量（例如算术或几何平均值）汇总每个锚点的所有正样本对。聚合选择反映了不同的假设：算术平均值假设加性贡献但可能过度重视异常值配对，而几何平均值对外部值的敏感度较低。还可以在潜在空间中考虑其他充分统计量 [17] 但它尚未被证明能提高性能。

得到的批量平均多窗口 InfoNCE [7] 对比损失是：

$$\mathcal{L}_{\text{geo}} = -\frac{1}{NM} \sum_{a=1}^{NM} \left[\frac{1}{M-1} \sum_{b \in \mathcal{P}(a)} S_{ab} - \log \left(\sum_{c \neq a} e^{S_{ac}} \right) \right], \quad (2)$$

$$\mathcal{L}_{\text{arith}} = \frac{1}{NM} \sum_{a=1}^{NM} \left[\log(M-1) + \log \left(\sum_{c \neq a} e^{S_{ac}} \right) - \log \left(\sum_{b \in \mathcal{P}(a)} e^{S_{ab}} \right) \right]. \quad (3)$$

对于特殊情况 $M = 2$ ，每个锚点 z_a 只有一个正样本 z_{a+} ，和 $M-1 = 1$ ，因此两种形式都简化为标准的 InfoNCE 损失 [7]。

$$\mathcal{L}_{\text{geo}} = -\frac{1}{2N} \sum_{a=1}^{2N} \left[S_{a,a+} - \log \left(\sum_{c \neq a} e^{S_{ac}} \right) \right]. \quad (4)$$

B. 损失函数的推导

我们现在提供详细的推导。我们展示如何将聚合统计量引入 InfoNCE。带几何平均聚合的 InfoNCE 定义如下：

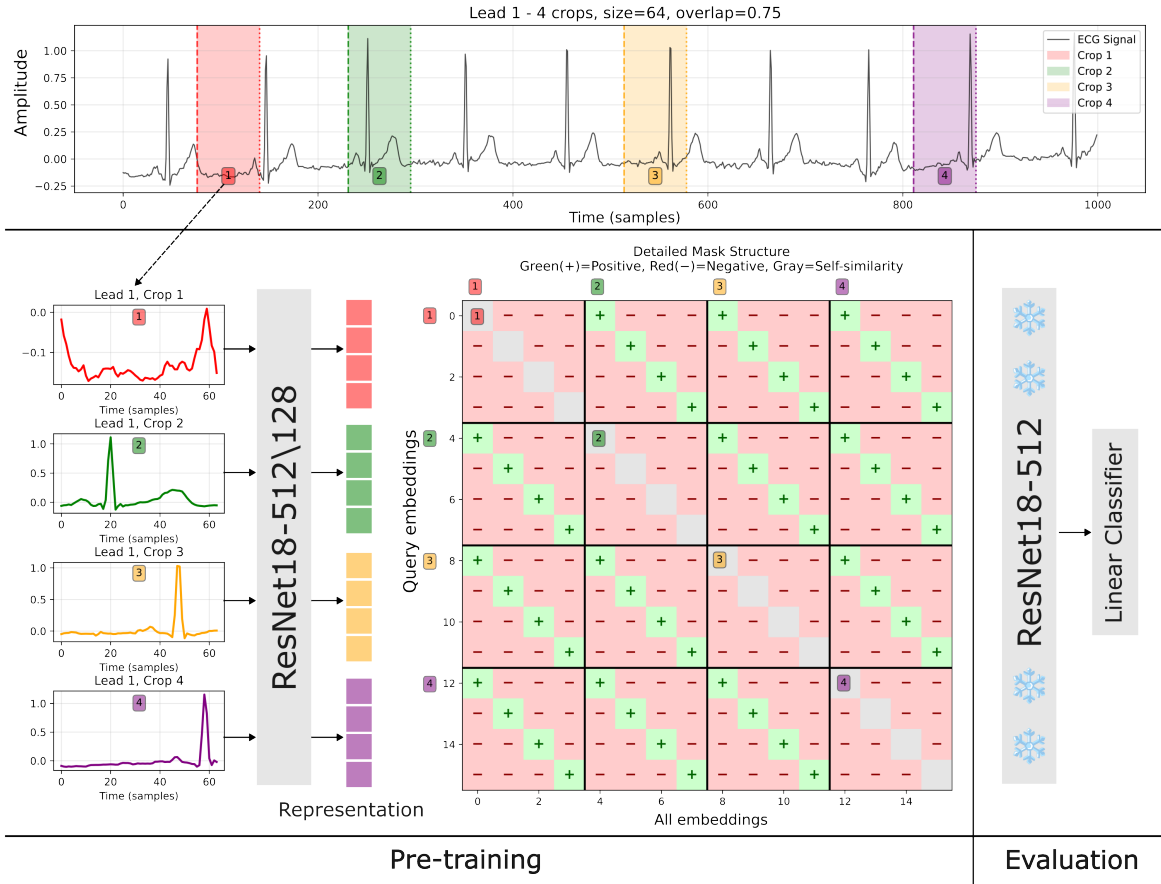


图 1: 多窗口对比学习用于心电图信号。(顶部) 简化可视化, 来自 12 导联心电图记录的单个导联 (1000 个时间点). 在我们的框架中, 时间剪辑应用于所有导联, 并且所有导联作为单个输入使用. (左侧) 每种彩色作物的波形, 通过 ResNet-18 主干网络编码成 512 维嵌入 (投影到 128 维度用于对比学习)。(正确) 多窗口对比相似性矩阵: 绿色 “+” 条目表示正对 (来自同一记录的不同窗口), 红色 “-” 条目表示负对 (来自不同记录的窗口), 而灰色对角线条目 (自相似性) 被排除在损失之外。请注意, 此处标记的正对也用于计算对比目标中的分母。在评估过程中, 预训练编码器被冻结, 并且在线性分类器上训练学习到的表示。

算术平均聚合的 InfoNCE 定义如下:

$$\begin{aligned}
 \mathcal{L}_{\text{arith}} &= -\frac{1}{NM} \sum_{a=1}^{NM} \log \left(\frac{1}{M-1} \sum_{b \in \mathcal{P}(a)} p_{ab} \right) \\
 &= -\frac{1}{NM} \sum_{a=1}^{NM} \log \left(\frac{1}{M-1} \sum_{b \in \mathcal{P}(a)} \frac{e^{S_{ab}}}{\sum_{c \neq a} e^{S_{ac}}} \right) \\
 &= -\frac{1}{NM} \sum_{a=1}^{NM} \log \left(\frac{1}{M-1} \cdot \frac{1}{\sum_{c \neq a} e^{S_{ac}}} \sum_{b \in \mathcal{P}(a)} e^{S_{ab}} \right) \\
 &= -\frac{1}{NM} \sum_{a=1}^{NM} \left[-\log(M-1) - \log \sum_{c \neq a} e^{S_{ac}} + \log \left(\sum_{b \in \mathcal{P}(a)} e^{S_{ab}} \right) \right] \\
 &= \frac{1}{NM} \sum_{a=1}^{NM} \left[\log(M-1) + \log \left(\sum_{c \neq a} e^{S_{ac}} \right) - \log \left(\sum_{b \in \mathcal{P}(a)} e^{S_{ab}} \right) \right].
 \end{aligned}$$

其中 $p_{ab} = \frac{e^{S_{ab}}}{\sum_{c \neq a} e^{S_{ac}}}$ 。

我们的多窗口损失通过统计聚合每个锚点的所有 $M-1$ 正样本, 而不是独立评估每一对正样本。这将正样本的计算从 $O(M^2 N^2)$ 减少到 $O(MN)$, 通过计算一

表 I: 保留测试线性评估性能在多标签超类分类中的平均值 [95%置信区间]。

窗口	批量大小	损失类型	作物	重叠	周期 ↓	损失函数	F1 分数 ↑	AUROC ↑	回忆 ↑	精度 ↑
2	256	Geometric	64	0.00	128	Geometric	0.679 [0.667, 0.690]	0.888 [0.884, 0.893]	0.660 [0.640, 0.680]	0.710 [0.685, 0.735]
4	768	Geometric	64	0.50	64	Geometric	0.675 [0.666, 0.685]	0.887 [0.886, 0.888]	0.665 [0.645, 0.684]	0.700 [0.689, 0.710]
6	256	Geometric	64	0.75	32	Geometric	0.675 [0.669, 0.680]	0.890 [0.887, 0.892]	0.656 [0.637, 0.676]	0.704 [0.685, 0.723]
8	256	Geometric	64	0.50	32	Geometric	0.680 [0.672, 0.688]	0.891 [0.889, 0.893]	0.660 [0.642, 0.678]	0.710 [0.699, 0.722]

个单一的 $MN \times MN$ 相似性矩阵，并将每个锚点的正样本总结为 MN 摘要。我们假设它提高了梯度更新的信噪比，与具有更高视图多重性的多视角目标观察到的估计器方差减少和互信息边界更紧一致。[17].

C. 数据集

对于所有实验，我们使用了 PTB-XL 数据集 [20]，这是一个大规模的公开可用的数据集，包含 21,837 临床 12 导联心电图记录（每段 10 秒，采样率为 100Hz），涵盖了广泛的年龄段并且性别平衡。所有模型均使用每份心电图记录的完整 12 导联作为单一输入进行训练和评估。每个心电图由最多两位心脏病专家根据 SCP-ECG 标准进行标注，诊断标签被组织成超类、子类和细粒度多标签陈述，涵盖所有 71 个类别。我们在超类级别进行了分类实验。PTB-XL 中用于多标签分类的五个超类是：正常心电图（NORM）、心肌梗死（MI）、ST/T 改变（STTC）、传导障碍（CD）和肥大（HYP）。遵循数据集推荐的评估协议，我们使用了标准分层划分：1-8 折用于训练，第 9 折用于验证，第 10 折作为保留测试集。为了确保鲁棒性，所有结果都是通过五次不同随机种子运行取平均值得出。

D. 实验设置

我们的实验工作流程包括两个主要阶段：第一阶段是自监督预训练，第二阶段是线性评估。

我们首先使用对比学习方法在 PTB-XL 数据集上对 ResNet 18 的一种 1D 变体进行预训练。ResNet18 模型编码 12 导联 ECG 时间窗口，并将时间序列投影到 512 维表示中。然后通过一个一层线性头嵌入此表示进入 128 维空间以计算对比目标的相似度。我们将这种架构记为 ResNet18-512/128。在预训练期间，输入大小由指定窗口裁剪大小（32, 64, 128, 256 时间点）定义，而在线性评估中，我们使用所有 1000 个时间点。我们不应用其他任何数据增强，仅随机分层时间裁剪并允许重叠。对于预训练，模型使用 PyTorch [21] 框架进行了 500 个时期的混合精度训练。AdamW [22] 优化器被

采用，初始学习率为 0.01，权重衰减为 1×10^{-4} ， ϵ 为 1×10^{-8} 。学习率在前 10 步使用线性预热调度，然后余弦退火至整个训练过程中的最终学习率为 1×10^{-6} 。

预训练后，我们从最后一个检查点（最终轮次）中提取模型权重。我们移除 128 维的投影头，并冻结编码器直到最后一个自适应平均池化层，该层返回 512 维的表示（我们将其记为 ResNet18-512）。然后我们训练一个线性分类器来评估学习到的 512 维特征的质量。对于线性评估，我们使用相同的优化器和学习率调度程序训练分类器 90 轮次，并通过 PyTorch 的多标签软边缘损失优化多标签一对全分类目标。模型选择基于 F1 指标的最佳验证性能。

III. 结果

在本节中，我们展示了所提出的多窗口对比学习框架在多标签超类分类任务上的经验评估和消融实验。

A. 超类分类的线性评估性能

表 I 总结了多标签超类分类任务上的线性评估的保留测试集结果。所有值均以平均值和 95% 置信区间报告，基于多次独立运行计算得出。所评估的指标包括 F1 分数、AUROC、召回率和精确度。

在所有配置中，我们观察到模型性能稳健且一致。最高值 AUROC 是通过配置 8 窗口、批量大小 256、裁剪尺寸 64、重叠 0.50 和 32 训练周期 (0.891[0.889, 0.893]) 实现的，最佳 F1 分数 (0.680[0.672, 0.688])、召回率 (0.660[0.642, 0.678]) 和精度 (0.710[0.699, 0.722])。然而，使用 Wilcoxon 符号秩检验 ($p > 0.05$) 并进行 Holm 校正的成对比较显示，最佳表现方法之间在性能上没有统计学上的显著差异。

一个显著的发现是，增加窗口数量使模型能够在更少的训练周期数下达到最优或接近最优性能。具体来说，对于 6 和 8 窗口，在仅 32 个周期时就观察到了饱和现象。同时，配置较少窗口的情况需要更多的周期才能收敛。这一观察结果表明了多窗口策略的优势，因为更高的视图多重性加速了收敛速度而没有牺牲性能。

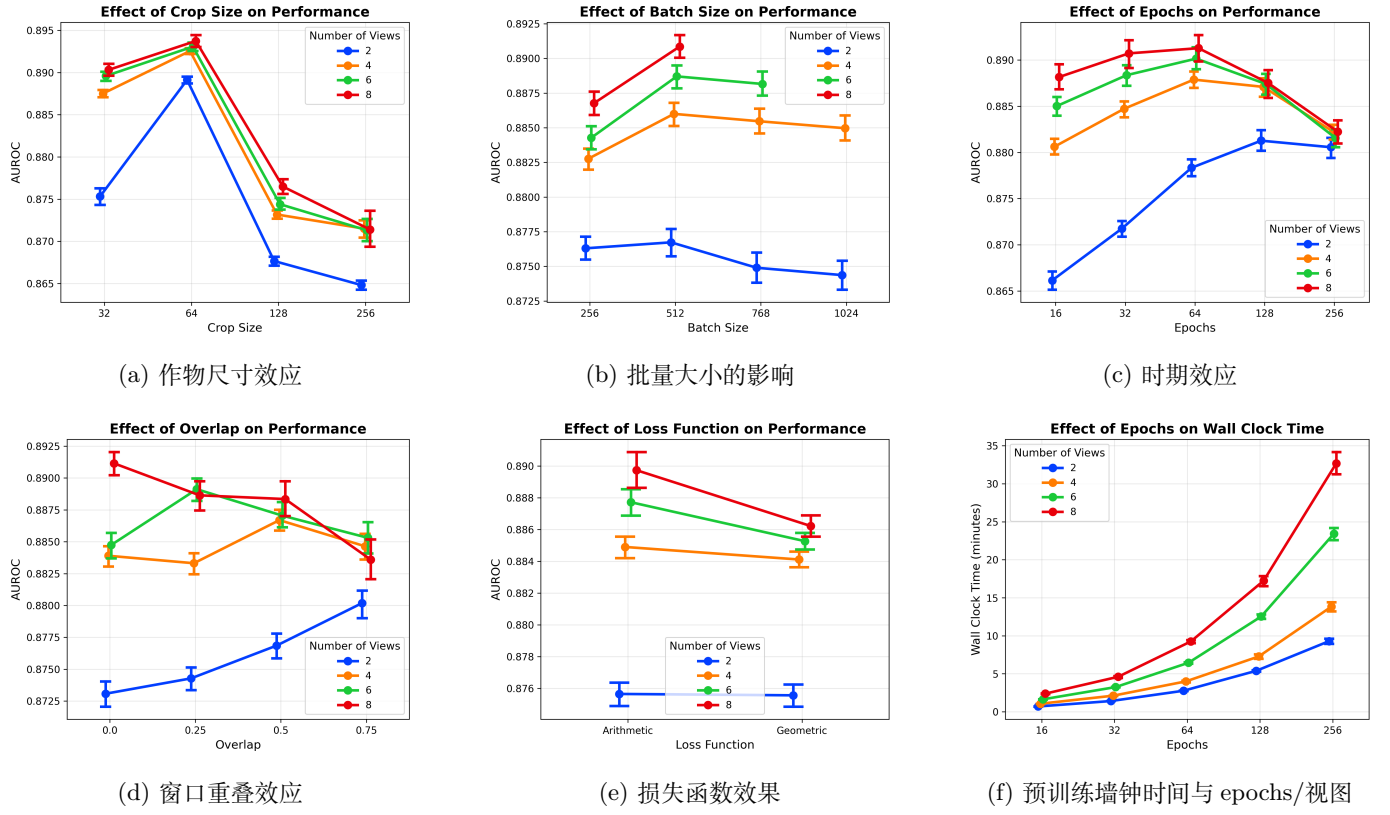


图 2: 消融实验和实时时钟时间结果对于多窗口对比学习框架验证 AUROC 和实际耗时显示了广泛的网格搜索覆盖范围: (视图数量 = 2, 4, 6, 8), (作物尺寸 = 32, 64, 128, 256), (重叠 = 0.0, 0.25, 0.5, 0.75), (EPOCHS_数量 = 16, 32, 64, 128, 256), (种子 = 0, 42, 123, 555, 789), 损失函数 (损失函数 = 几何, 算术), 和批量大小 (批量大小 = 256, 512, 768, 1024). 每个子图报告了五个随机种子的均值和 95% 置信区间。(a) 图像大小对模型性能 (AUROC) 的影响。(b) 批量大小对 AUROC 的影响。(c) 训练轮数对 AUROC 的影响。(d) 窗口重叠对 AUROC 的影响。(e) 损失函数对 AUROC 的影响。(f) 训练轮数和视图数量对墙钟时间的影响。

B. 消融研究

我们进行了一系列消融研究, 以系统地评估每个超参数的贡献。每次消融通过一次改变一个参数来孤立某个因素的影响。所有结果都在验证集上报告。

1) **作物大小的影响**: 作物大小对多窗口对比学习框架的影响如图 2a 所示。模型在作物大小为 64 时达到了最高的 AUROC。令人惊讶的是, 这种性能出现在时间窗口捕捉到大约一个心动周期时。这可能表明我们的表示方法能够在一次心跳中捕获这些生理过程的动态变化。更大的作物 (128, 256) 和非常短的作物 (32) 都表现出较低的性能, 可能是分别由于捕捉到了多次心跳或缺乏足够的时间上下文。作物大小 64 表现优于所有其他测试的作物大小 ($p < 0.00001$), 效果显著 ($\eta^2 = 0.5491$), 突显了作物大小作为最关键超参数的地位。

2) **批大小的影响**: 如图 2b 所示, 随着批量大小从 256 增加到 512, AUROC 有所提高。然而, 进一步将批量大小增至 768 和 1024 会导致性能下降, 这可能是由于收益递减以及在较大批量大小下梯度估计不那么有效的原因。批量大小的影响程度较小 ($\eta^2 = 0.0144$, $p < 0.00001$)。所有实验均使用了 AdamW 优化器。替代优化算法的潜在影响仍然是未来研究中的一个开放问题。

3) **epoch 数量的影响**: 图 2c 展示了模型性能作为训练周期的函数。虽然 AUROC 通常会随着额外的周期而增加, 但最优的周期数强烈受到窗口数量的影响。对于两个窗口的数量, 最佳的 AUROC 性能是在 128 个周期时达到的。当窗口数量超过 4 时, 最优性能会在更早的时间内达到, 通常在 32 到 64 个周期之间。这表明利用更多的窗口可以使模型更快地收敛。观察到的周期效

表 II: 与之前关于**超类多标签分类** (F1 和 AUROC) 的工作比较。所有方法均在 PTB-XL 上进行训练和评估, 除非另有说明。“Architecture” 表示模型主干及嵌入维度 (例如, ResNet50/2048)。“Epochs” 报告为预训练/线性评估。“repr.” 被特定引用标注为重现的方法。N/R 意味着未报告该指标。

方法	架构	周期	F1 分数	AUROC
SimCLR repr. by [13]	ResNet18/256	400/100	N/R	0.648
时钟系统 repr. by [13]	ResNet18/256	400/100	非/反应	0.784
3 千克 repr. by [13]	ResNet18/256	400/100	非/反	0.751
TSSL [13]	ResNet18/256	400/100	N/R	0.872
SimCLR repr. by [14]	ResNet50/2048	300/10	0.624	0.866
Our SimCLR (2 views, ours)	ResNet18-512/128	128/90	0.679 [0.667, 0.690]	0.888 [0.884, 0.893]
我们的多窗口 (8 个视图)	ResNet18-512/128	32/90	0.680 [0.672, 0.688]	0.891 [0.889, 0.893]

果很小 ($\eta^2 = 0.0574, p < 0.00001$)。

4) 重叠的影响: 图 2d 总结了窗口重叠对模型性能的影响。观察到的效果大小可以忽略不计 ($\eta^2 = 0.0063, p < 0.00001$), 因此重叠的实际影响有限。然而, 对于两个窗口, 它可以稍微提高性能, 而对于更多的窗口, 性能可能会下降。

5) 损失函数的影响: 如图 2e 所示, 算术损失函数在所有窗口数量下始终达到比几何损失更高的平均 AUROC。这一模式表明, 算术聚合在保持学习表示的判别结构方面略微更有效。然而, 损失函数之间的差异量很小, 效应大小可以忽略不计 ($\eta^2 = 0.0075, p < 0.00001$)。

6) 细节和统计分析 & 交互作用。: 我们进行了超过 7,200 次实验来验证作物大小、重叠度、周期数、损失函数、窗口数量和批量大小的影响。其中, 作物大小 ($\eta^2 = 0.5491$) 和窗口数量 ($\eta^2 = 0.1526$) 表现出较大的效应量, 并且是最具影响力的超参数。交互分析确定了作物大小 $\times$ 窗口数量的交互作用具有中等效应 ($\eta_p^2 = 0.0648$)。

C. 与之前工作的比较

表 II 提供了我们多窗口对比学习框架与代表性自监督方法在 PTB-XL 上的多标签超类分类的直接比较。为了可比性, 所有列出的方法均采用 ResNet 基础架构, 并遵循相同的预训练和评估协议。

TSSL [13] 使用具有 256 维嵌入的 ResNet18 主干, 在预训练中训练了 400 个周期, 并在线性评估中训练了 100 个周期。在其他最近的对比方法中, CLOCS [15] 实现了 AUROC 为 0.784, 而 3KG [16] 应用基于图的对比学习, 并报告了 AUROC 为 0.751。

我们的 SimCLR 基线模型 (2 个视图, ResNet18-512/128) 使用稍小的嵌入, 并训练了 128 个预训练周期, 最多进行 90 个时期的线性评估 (需要注意的是, 通常基于验证性能提前停止线性评估, 但我们为了保持一致性报告了 90 个时期的评估结果)。该基线模型达到了 AUROC 为 0.888[0.884, 0.893] 和 F1 分数为 0.679[0.667, 0.690], 已经超过了之前的成果。

我们的多窗口 (8 个视角) 方法进一步提高了性能, 达到 AUROC 为 0.891[0.889, 0.893] 和 F1 值为 0.680[0.672, 0.688], 仅需 32 预训练周期和最多 90 个线性评估周期 (同样可能提前停止)。这一结果突显了当前最先进的性能, 并且与之前的工作相比显著减少了所需的预训练周期。

供参考, 实现的 SimCLR 基线 [14] 使用更大的 ResNet50 主干网络 (2048 维嵌入) 和 300 预训练周期实现了 AUROC 0.866 和 F1 0.624。这两个值都低于我们的结果。此外, 该工作 [14] 还报告了在更大规模的多机构心电图数据集 (例如, 宁波 [23], CODE-15 [24], 查普曼 [25]) 上进行预训练后的指标。为了专注于方法本身, 我们的主要比较将预训练和评估都限制在 PTB-XL 数据集上。

总结来说, 在具有可比性的 ResNet 架构和一致的单数据集评估下, 我们的多窗口方法在 AUROC (0.891 对比 0.872 用于 TSSL [13] 和 0.866 用于 SimCLR) 和 F1 (0.680 对比 0.679 用于我们的 SimCLR 基线和 0.624 [14]) 方面取得了最高成绩, 并且随着视图数量的增加, 预训练周期显著减少。

D. 计算效率分析

表 III 和图 2f 说明了多窗口对比学习中的权衡。每条心电图增加视窗数量会增加每个周期所需的时钟时

表 III: 时钟时间 (分钟) 对于不同数量的视图和训练周期。数值以五次实验种子的平均值和 95%置信区间 (CI) 显示。

Windows	16 epochs	32 epochs	64 epochs	128 epochs	256 epochs
2	0.7 $\pm$ 0.0	1.4 $\pm$ 0.0	2.8 $\pm$ 0.0	5.4 $\pm$ 0.1	9.3 $\pm$ 0.4
4	1.1 $\pm$ 0.0	2.1 $\pm$ 0.0	4.0 $\pm$ 0.1	7.3 $\pm$ 0.3	13.8 $\pm$ 0.6
6	1.6 $\pm$ 0.0	3.2 $\pm$ 0.0	6.4 $\pm$ 0.1	12.5 $\pm$ 0.3	23.4 $\pm$ 0.8
8	2.4 $\pm$ 0.0	4.6 $\pm$ 0.1	9.3 $\pm$ 0.2	17.2 $\pm$ 0.7	32.7 $\pm$ 1.5

间。然而，尽管每个周期的成本增加，但实现最佳性能所需的实际总时钟时间在使用更多视窗的情况下更低。具体来说，在表 I 中，我们的实验显示 8 窗口模型仅需 32 个周期就能达到其最优验证 AUROC (0.891)，而 2 窗口基线则需要 128 个周期才能达到最佳 AUROC (0.888)。使用两个视窗训练 128 个周期大约需要 5.4 分钟 (见表 III)，而 8 窗口配置仅需 32 个周期，总共只需要 4.6 分钟，总时钟训练时间提高了约 14.8%。此外，8 视窗配置 (见图 2) 不仅收敛速度更快，而且始终产生最高的整体模型性能，在 AUROC 方面超过了所有其他设置。

至关重要的是，当考虑到验证频率降低、检查点减少以及由快速收敛启用的超参数优化周期大幅加快时，这种时间节省变得更加显著。在我们的实验中，我们使用了一块 NVIDIA A100 GPU (80 GB)。对于双视图 SimCLR 基线 and 我们的 $M = 8$ 多窗口方法，在批量大小为 1024 且裁剪大小为 64 的训练过程中，观察到峰值内存使用量为 12.35GB。因此，在相同设置下，两种方法的有效内存占用是可比较的。对于在较小批量大小 256 下的双视图 SimCLR 配置，内存使用量降至 1.2GB。

IV. 结论

在本研究中，我们引入了一个多窗口对比学习框架用于心电图的自监督表示学习，利用了每个记录中的多个时间上不同的视图。在 PTB-XL 数据集上，我们的方法在多标签超类分类任务中始终优于传统的两视图对比方法，实现了更高的 AUROC (0.891 vs. 0.888) 和 F1 分数 (0.680 vs. 0.679)。值得注意的是，多窗口学习使得性能稳健，并且仅需四分之一的预训练周期 (32 vs. 128)，同时将总训练时间减少了 14.8% (4.6 vs. 5.4 分钟)。我们的消融研究确定了裁剪大小和窗口数量是最重要的超参数。此外，我们的框架在使用 ResNet18 架构时超越了最近的对比基线，并允许减少训练周期数。

该方法本身是模型不可知的，并且可以轻松应用于其他深度学习架构和生物医学时间序列。

然而，仍存在若干限制。本研究中的所有训练和评估均局限于 PTB-XL 数据集，对于其他心电图队列或具有更广泛人群和设备变异性的真实世界数据的适用性尚未测试。我们的框架仅在 12 导联心电信号上进行评估，并且仅用于多标签超类分类。其对单导联心电图、其他生理信号或其他下游任务的有效性尚未探索。

致谢

A.F., Y.Y. 和 J.V.D. 得到了埃默里大学奈尔霍奇森伍德鲁夫护理学院的支持。S.P. 获得了 NIH R01DA040487 和 NSF 2112455 的部分支持，以及 NIH 2R01EB006841 的支持。此外，R.X. 和 A.F. 还得到了美国国立卫生研究院转化医学推进中心的 Georgia 临床与转化科学联盟 (CTSA) BERD 基金 (拨款号 UL1TR002378) 的部分支持。内容仅由作者负责，并不一定代表美国国立卫生研究院的官方观点。

参考文献

- [1] C. W. Tsao, A. W. Aday, Z. I. Almarzooq, C. A. Anderson, P. Arora, C. L. Avery, C. M. Baker-Smith, A. Z. Beaton, A. K. Boehme, A. E. Buxton *et al.*, “Heart disease and stroke statistics—2023 update: a report from the american heart association,” *Circulation*, vol. 147, no. 8, p. e93, 2023.
- [2] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, “Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network,” *Nature medicine*, vol. 25, no. 1, pp. 65–69, 2019.
- [3] Z. I. Attia, P. A. Noseworthy, F. Lopez-Jimenez, S. J. Asirvatham, A. J. Deshmukh, B. J. Gersh, R. E. Carter, X. Yao, A. A. Rabinstein, B. J. Erickson *et al.*, “An artificial intelligence-enabled ecg algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction,” *The Lancet*, vol. 394, no. 10201, pp. 861–867, 2019.
- [4] Z. Wu and C. Guo, “Deep learning and electrocardiography: systematic review of current techniques in cardiovascular disease diagnosis and management,” *BioMedical Engineering OnLine*, vol. 24, no. 1, p. 23, 2025.
- [5] X. Liu, H. Wang, Z. Li, and L. Qin, “Deep learning in ecg diagnosis: A review,” *Knowledge-Based Systems*, vol. 227, p. 107187, 2021.
- [6] T. Mehari and N. Strodthoff, “Self-supervised representation learning from 12-lead ecg data,” *Computers in biology and medicine*, vol. 141, p. 105114, 2022.
- [7] A. v. d. Oord, Y. Li, and O. Vinyals, “Representation learning with contrastive predictive coding,” *arXiv preprint arXiv:1807.03748*, 2018.

- [8] R. D. Hjelm, A. Fedorov, S. Lavoie-Marchildon, K. Grewal, P. Bachman, A. Trischler, and Y. Bengio, "Learning deep representations by mutual information estimation and maximization," in *International Conference on Learning Representations*, 2019. [Online]. Available: <https://openreview.net/forum?id=Bklr3j0cKX>
- [9] L. Jing and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 11, pp. 4037–4058, 2020.
- [10] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PmLR, 2020, pp. 1597–1607.
- [11] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar *et al.*, "Bootstrap your own latent-a new approach to self-supervised learning," *Advances in neural information processing systems*, vol. 33, pp. 21 271–21 284, 2020.
- [12] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9729–9738.
- [13] W. Chen, H. Wang, L. Zhang, and M. Zhang, "Temporal and spatial self supervised learning methods for electrocardiograms," *Scientific Reports*, vol. 15, no. 1, p. 6029, 2025.
- [14] S. Kim, "Learning general representation of 12-lead electrocardiogram with a joint-embedding predictive architecture," *arXiv preprint arXiv:2410.08559*, 2024.
- [15] D. Kiyasseh, T. Zhu, and D. A. Clifton, "Clocs: Contrastive learning of cardiac signals across space, time, and patients," in *International Conference on Machine Learning*. PMLR, 2021, pp. 5606–5615.
- [16] B. Gopal, R. Han, G. Raghupathi, A. Ng, G. Tison, and P. Rajpurkar, "3kg: Contrastive learning of 12-lead electrocardiograms using physiologically-inspired augmentations," in *Machine learning for health*. PMLR, 2021, pp. 156–167.
- [17] A. Shidani, R. D. Hjelm, J. Ramapuram, R. Webb, E. G. Dhekane, and D. Busbridge, "Poly-view contrastive learning," in *The Twelfth International Conference on Learning Representations*, 2024. [Online]. Available: <https://openreview.net/forum?id=iHcTLIor0m>
- [18] L. Wiskott and T. J. Sejnowski, "Slow feature analysis: Unsupervised learning of invariances," *Neural computation*, vol. 14, no. 4, pp. 715–770, 2002.
- [19] C. Feichtenhofer, H. Fan, B. Xiong, R. Girshick, and K. He, "A large-scale study on unsupervised spatiotemporal representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 3299–3309.
- [20] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze, W. Samek, and T. Schaeffter, "Pt看-xl, a large publicly available electrocardiography dataset," *Scientific data*, vol. 7, no. 1, pp. 1–15, 2020.
- [21] A. Paszke, "Pytorch: An imperative style, high-performance deep learning library," *arXiv preprint arXiv:1912.01703*, 2019.
- [22] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2019. [Online]. Available: <https://openreview.net/forum?id=Bkg6RiCqY7>
- [23] J. Zheng, H. Chu, D. Struppa, J. Zhang, S. M. Yacoub, H. El-Askary, A. Chang, L. Ehwerhemuepha, I. Abudayyeh, A. Barrett *et al.*, "Optimal multi-stage arrhythmia classification approach," *Scientific reports*, vol. 10, no. 1, p. 2898, 2020.
- [24] Y.-J. Chen, C.-L. Liu, V. S. Tseng, Y.-F. Hu, and S.-A. Chen, "Large-scale classification of 12-lead ecg with deep learning," in *2019 IEEE EMBS international conference on biomedical & health informatics (BHI)*. IEEE, 2019, pp. 1–4.
- [25] J. Zheng, J. Zhang, S. Danioko, H. Yao, H. Guo, and C. Rakovski, "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," *Scientific data*, vol. 7, no. 1, p. 48, 2020.