面向远程学习环境的基于摄像头的心率变异性预测

Kegang Wang[a, c], Yantao Wei[a*], Jiankai Tang[b, c], Yuntao Wang[b, c*] Mingwen Tong[a], Jie Gao[a], Yujian Ma[d], Zhongjin Zhao[a]

* 共同通讯作者[a] 中国武汉华中师范大学，430079[b] 人因工程国家重点实验室，北京，100094。[c] 清华大学计算机科学与技术系，北京，100084。[d] 中国上海华东师范大学，200241

arXiv:2305.04161v3 [cs.CV] 4 Nov 2024

摘要——**近年来，随着互联网视频的广泛应用，远程光电容积描记（rPPG）在情感计算领域日益受到关注**

从面部视频中恢复血容量脉搏（BVP）信号是一项具有挑战性的任务，需要通过一系列预处理、图像算法和后处理来恢复波形。心率指标不仅用于情感计算，心率变异性（HRV）指标更为重要

通过远程光电容积描记获取心率变异性指标的难点，在于算法必须精确预测血容量脉搏峰值位置。本文收集了远程学习情感与生理（RLAP）数据集，其中包含来自58名受试者、总时长超过32小时的高度同步视频和标签。这是一个公开数据集，其BVP标签经过精心设计，更适合用于训练心率变异性模型。我们利用RLAP数据集训练了一个名为Seq-rPPG的新模型。该模型基于一维卷积，实验结果表明，这种结构更适合处理心率变异性任务；其心率变异性表现优于所有其他基线方法，同时在计算效率方面也展现出显著优势。

索引词—**远程光电容积描记、数据集、情感计算、远程学习。**

一、引言

近年来，远程生理感知领域涌现出许多公开数据集[7], [42], [19], [28], [17], [1], [10], [37], [15], [31], [23], [22], [9],其中许多数据集致力于为运动、年龄、族裔、性别和光照条件更多样化的场景提供基准测试。

然而，大多数数据集使用两台独立设备分别采集视频和生理信号标签，缺乏确保这两种信号严格同步的机制。经过仔细检查发现，部分数据并未精确同步；另有一些数据存在帧率波动，导致从某个时间点开始出现延迟；还有一些数据集采用人工粗略同步，但仍存在 100 至 200 毫秒的误差。通常，心率变异性分析的时间尺度为数十毫秒，而这些误差会降低数据集用于心率变异性训练的适用性。此外，心率变异性需要更纯净的 BVP信号，因此，数据集具备高视频质量同样至关重要，McDuff等人(2017)[16] 和 Yu等人(2019)[40] 已有研究表明，压缩视频格式会降低BVP信号的质量，因此采集无损格式的数据至关重要。

先前研究[39], [5], [29] 表明，与同步相关的问题会使模型训练更具挑战性，尤其是在采用均方误差（MSE）作为损失函数时。由于 MSE 对延迟过于敏感，因此已有研究提出了多种对延迟不敏感或具有延迟不变性的损失函数[39], [5], [29], [2], [41]。然而，专门用于解决这一问题的数据集仍然稀缺，尤其是高度同步的数据集，至今仍难以汇编。

在前人工作的基础上，我们收集了远程学习情感与生理（RLAP）数据集，用于远程或在线学习场景。该数据集包含高质量视频和高度同步的BVP标签。我们的目标是通过高质量数据提高心率变异性准确性，并拓展远程光电容积描记在情感计算中的应用[8]，，尤其是在学生情绪分析中的应用，并实现更高精度的心搏间期（IBI）预测，扩展至基于大语言模型的健康模型[33]。有关RLAP的基本信息及其与其他数据集的比较，见表I。

基于移动设备的本地远程光电容积描记算法尤具前景，因为这类算法计算成本较低，且能充分保护隐私，因此许多研究[39],[3], [12], [2], [5]都致力于开发轻量级端到端网络。本文设计了一种基于一维卷积神经网络（1D CNN）的算法，通过简单的线性映射将视频帧编码为一维特征，并提取BVP信号。实验结果表明，该方法不仅能显著提高心率变异性估计的准确性，还能大幅降低计算开销。本文的贡献如下：

- 我们构建了高质量、高度同步的公开数据集RLAP，专为远程学习场景中的情绪场景而设计。此外，我们还将用于采集该数据集的工具PhysRecorder开源。

- 研究人员提出了一种基于 1D CNN 的轻量级算法 Seq-rPPG。该算法在心率变异性任务中展现出显著优势，且计算开销较低。

表 I：主要数据集及我们的新数据集 RLAP 的基本信息

数据集	参与者	帧	小时	PPG	信号偏移	无损格式
AFRL[7][1]	25	97.2M[2]	25	✓	0	✓
PURE[28]	10	106K	1	✓	0	✓
UBFC[1]	42	75K	0.7	✓	>0.5秒[3]	✓
MMSE-HR[42]	58	435K	4.8		0	
MAHNOB-HCI[37]	30	25.2M[4]	19.4		0	
VIPL-HR[19]	107	2.14M[5]	19.8	✓	>0.5s	
MMPD[31]	33	1.15M	10.6	✓	<0.2s	
RLAP	58	353万	32.7	✓	0	✓

[1]作者未说明这是一个公开数据集。[2]使用九台RGB摄像机以120 fps同步记录。[3]只有部分视频存在明显偏移。[4]使用一台RGB摄像机和五台黑白摄像机以60 fps同步记录。[5]使用估计的30 fps，尽管实际 fps 在15到30之间波动。

II. 数据集

RLAP数据集的采集分为两个步骤：1）检查摄像机支持的编解码器和传输标准，以确保尽可能采用无损的存储格式；检查脉搏血氧仪的应用程序接口，以确保实时采集 BVP 信号；并开发可同时采集这两类信号的软件。2 ）按照预先制定的数据采集流程搭建采集环境，与每位参与者沟通，并在取得签署的知情同意后采集数据。

A. 程序编码

我们使用 Logitech C930c网络摄像头采集支持 MJPG 和 YUY2（YUV422）格式的视频。默认情况下，它采用 MJPG 格式，以 1920x1080@30fps 的帧率传输视频（适用于一般场景）；同时，通过应用程序接口指定 YUY2 格式，以 640x480@30fps 的帧率传输原始图像（适用于远程光电容积描记专用场景）。

我们使用 HID 驱动程序，通过 USB接口从脉搏血氧仪读取原始信号，具体采集 BVP 片段。我们编写的应用程序同时采集摄像机和脉搏血氧仪的信号，并为每个数值或视频帧分配 UNIX时间戳，以确保数据严格对齐，避免 BVP峰值定位错误。

B. 数据集采集

在数据记录过程中，受试者完成了一系列任务或观看视频。完成指定任务后，受试者休息一段时间，然后实验人员为其安排下一个任务。全部58名受试者（16名男性和42名女性）均为中国学生，主要是硕士研究生。分配给每名受试者的任务分为三个部分。第一部分是远程光电容积描记任务（rPPG 任务），包括四种场景：一般放松场景、拉上窗帘的黑暗放松场景、玩计时游戏的紧张场景，以及朗读鲁迅文章的演讲场景。

第二部分是情绪诱导任务，要求受试者观看六段视频，各段视频之间穿插短暂休息：观看自然景观、玩益智游戏、观看喜剧、观看幻觉图像、阅读学术论文，以及观看打哈欠视频。第三部分是学习投入任务，包括完成三项学习活动，每项活动之后都安排了短暂休息：观看视频并根据视频内容学习和答题（简单）、阅读文本并根据文本内容学习和答题（困难），以及观看公开课讲座（无练习）。RLAP提供了超过32小时的视频（353万帧）。有关RLAP的更多详情见表II。数据采集工作流程示意图和部分样本见图1

数据采集环境面朝窗户，并设有室内人工光源。受试者坐在计算机前，距离摄像机约一米。采集视频期间，受试者用右手握住鼠标或笔完成任务，并在左手佩戴 CMS50E 脉搏血氧仪。我们要求他们尽量减少左手移动，以确保信号采集稳定。受试者的头部不受固定，可以自然移动。同时，检查员使用附近的另一台计算机连接参与者的网络摄像头和脉搏血氧仪，并监督数据采集过程。

III. 算法

心率变异性任务需要精确地估计峰值，这是一项对时间敏感的工作。考虑到需要捕捉尽可能多的频率特征，模型架构针对时间序列任务进行了专门设计，并应采用较大的时间窗口。与大多数通用时空建模方法（如差分 2D 卷积（DIFF 2D CNN）[3], [13], [12], 3D 卷积（3D CNN）[2], [39], [16], 和时空图（STMap & MSTMap）[18], [20], [21],）不同，我们的方法主要侧重于时间特征。这包括沿时间维度使用一维卷积（1D CNN），并确保

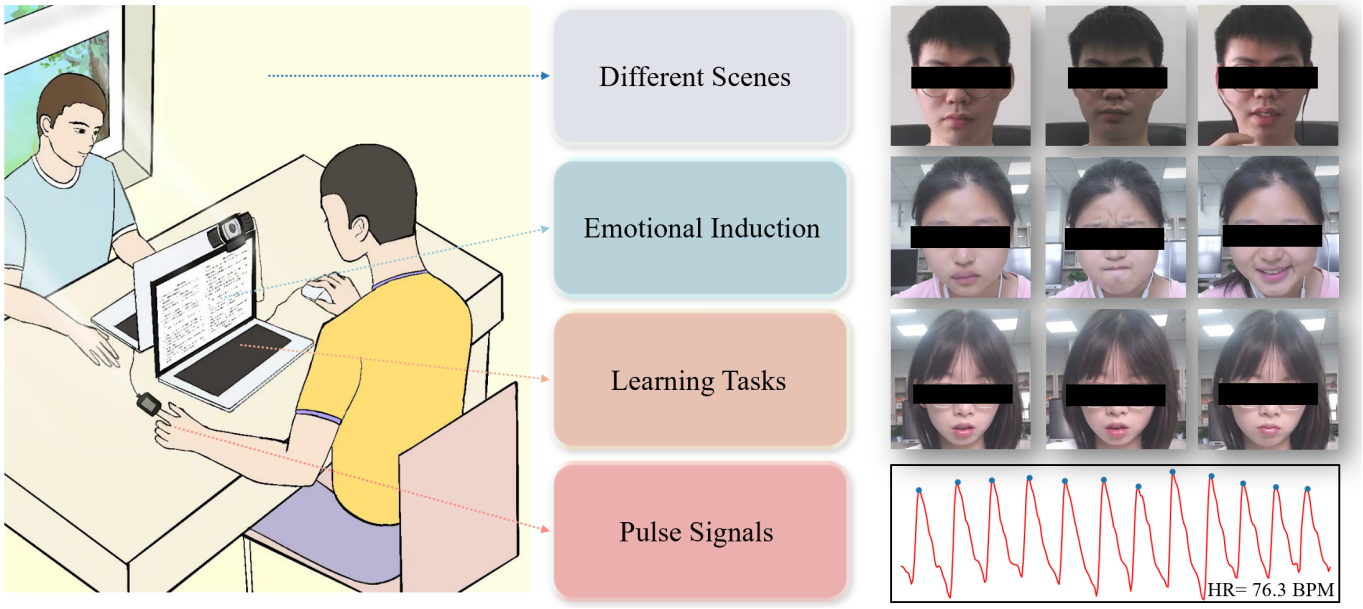


图1：RLAP数据集概览。RLAP数据集包含58名学生的样本，涵盖各种场景、情绪和学习参与度。在完成这些任务时，参与者的脉搏信号由脉搏血氧仪同步记录。

表II：数据采集流程

子数据集	任务或场景	目标	时长（秒）	摄像机编解码器	视频编解码器	分辨率
场景任务	放松	-	120	YUY2[1]	RGB,MJPG,H264	640×480
	放松（深色）	-	120			
	玩游戏	移动手部	120			
	阅读文章	面部活动	120			
情绪任务	自然风光	宁静	120	MJPG	MJPG,H264	1920×1080
	益智游戏	专注	180			
	喜剧	快乐	120			
	错觉图片	困惑	20			
	学术论文	困倦	60			
	打哈欠视频	困倦	60			
学习任务	基于视频的学习	视频参与度	240	MJPG	MJPG，H.264	1920×1080
	基于教材的学习	文本参与度	480			
	观看公开课程	参与度低	420			

[1]YUY2 是网络摄像头使用的一种 RAW 传输格式，受带宽限制，只能以 480p@30fps 运行。

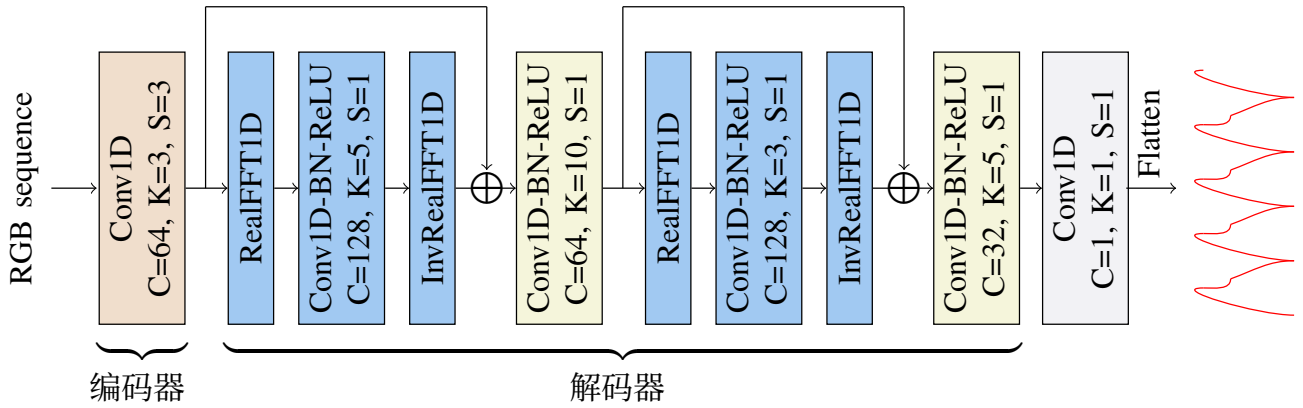


图 2：Seq-rPPG 网络，由编码器和解码器组成。编码器是一个单层 1x1 卷积，而解码器由四层时域和频域卷积交替构成。

训练阶段BVP标签与视频之间的严格对齐。

手工设计算法 [25], [24], [6], [38], [34]基于不同的反射分量。Shafer [26]假设 RGB 图像中的 BVP信号由不同频率光线的线性组合产生，而皮肤反射光包含镜面反射分量和漫反射分量。在后处理中，平稳信号和噪声会被滤除，而由血红蛋白浓度波动产生的周期信号则会被保留。因此，我们的算法分为两个部分。第一部分将原始视频（$450 \times 8 \times 8 \times 3$）的 RGB 通道和帧编号相结合，并合并宽度和高度，以获得一个1350×64 RGB 序列。使用大小为 3、步幅为 3 的卷积核对 RGB 序列进行混合，从而得到一个 450×64粗略信号。第二部分使用带有激活函数的多个卷积滤波器对粗略信号进行非线性滤波，从而获得 BVP信号。这两个部分可以视为一个编码器-解码器，这种结构此前已用于 Seq2Seq 网络的机器翻译任务[30]。在我们看来，rPPG 任务与机器翻译具有相似之处，其中 rPPG 算法将视频“翻译”为 BVP信号，因此该算法被称为 Seq-rPPG。

我们的设计灵感源自快速傅里叶卷积（FFC）[4], [27], ，它在处理周期信号（例如音频信息）方面非常有效。我们设计了一个频谱变换模块，并将其加入 1D CNN，以提升其性能。最终模型在四个时域卷积神经网络层与频谱变换之间交替切换（见图2）。

我们使用快速傅里叶变换（FFT）和 1D CNN 实现了频谱变换模块。对于信号$\mathbf{Y} \in \mathbb{R}^{N \times C}$，频谱滤波层首先对每个通道执行实数快速傅里叶变换（RFFT），得到频域信号$\mathbf{Y}_{Freq} \in \mathbb{Z}^{\lfloor \frac{N+1}{2} \rfloor \times C}$，将其分解为实部$\mathbf{Y}_{Real}$ 和虚部$\mathbf{Y}_{Img}$，然后沿通道维度将它们组合为$\mathbf{Y}_{Comb} \in \mathbb{R}^{\lfloor \frac{N+1}{2} \rfloor \times 2C}$。接下来，卷积层将输出重新分解为实部和虚部。新的输出被转换为复数，并通过逆实数快速傅里叶变换（IRFFT）还原为时域信号。输出信号通过残差连接与原始信号混合。请注意，整个过程中通道数保持不变。

IV. 实验与结果

所使用的实验平台配备 AMD Ryzen 9 5950X CPU、Nvidia RTX 3090 GPU，并运行 Windows 11 操作系统。我们从中选取了五种基线模型，其中，所提出的模型、PhysNet[39], DeepPhys[3],EfficientPhys[13], 和 TS-CAN[12] 均基于 TensorFlow 2.6 进行训练。PhysFormer[41] 则基于 PyTorch 2.0 进行训练。尽管代码中详细介绍了多种配置，但 TS-CAN 和 DeepPhys 均采用 36x36 的分辨率，与原文保持一致。对于预训练模型，我们测试了其移动端 CPU 在 Raspberry Pi 4B 上的推理性能（CPU：Cortex-A72，4 核；操作系统：Debian 11）。

Seq-rPPG 使用以下训练参数：Adam 优化器、批量大小为 32，以及平均绝对误差（MAE）损失。其他算法的参数则按照原始文献中的设置。

A. 数据集与指标

除 RLAP 数据集外，测试还使用了另外两个数据集：UBFC[1], 包含来自 42 名受试者的 42 段视频片段，每段时长为 1 分钟；PURE[28], 包含来自 10 名受试者的 59 段视频，每段时长为 1 分钟。对于 PURE 数据集，每位受试者完成了六种头部运动：保持静止、说话、头部相对于摄像机平面的缓慢和快速平移，以及小幅和中幅头部旋转。

我们评估了四项任务的准确性：心率（HR）、NN 间期标准差（SDNN）、NN50 间期百分比（pNN50）以及相邻 NN 间期差值均方根（RMSSD）。心率采用 Welch 方法测量，并结合 30–210 BPM 的带通滤波器。心率变异性分析使用 HeartPy 工具包[36], [35]进行。

每项任务均采用两种指标：平均绝对误差（MAE）和均方根误差（RMSE）。

B. 数据集内测试

我们根据受试者将 RLAP 数据集随机划分为训练集、验证集和测试集。数据集部分提供了具体的划分详情。所有算法均在完整测试集上进行测试。在 RLAP 数据集中，由于视频时长不一且较长，所有心率任务均采用 30 秒移动窗口，而其他任务则使用完整视频。对于 UBFC[1] 和 PURE[28] 数据集，心率和心率变异性任务均使用完整视频。测试结果见表III。

在数据集内测试中，Seq-rPPG 在心率任务上的表现与当前最先进的基线方法相近，其平均绝对误差指标优于当前最先进的基线方法，而均方根误差略逊于 TS-CAN 和 PhysNet。在心率变异性相关任务（正常心搏间期标准差、NN50间期百分比、相邻NN间期差值均方根）中，Seq-rPPG 展现出显著优势，优于所有基线方法，充分体现了该架构在心率变异性任务中的巨大潜力。

C. 跨数据集测试

我们使用 UBFC[1] 和 PURE[28]作为独立的测试集。使用 UBFC 作为测试集时，训练集为 RLAP 和 PURE。使用 PURE 作为测试集时，训练集为 RLAP 和 UBFC。详情请参阅表IV 和V 。

在跨数据集测试中，无论是在 UBFC-rPPG[1] 数据集还是 PURE[28] 数据集上，Seq-rPPG 在所有任务中的表现都优于其他基线方法。这表明 Seq-rPPG 不仅具备出色的心率变异性预测能力，还展现出稳健的泛化能力和迁移能力。此外，我们还测试了使用不同训练数据集时的结果。例如，在 UBFC-rPPG 上测试时，使用 RLAP 训练的模型通常优于使用 PURE 训练的模型。相反，当在

表 III：RLAP 上的数据集内测试。**粗体**：最佳结果。

方法	HR		SDNN		NN50间期百分比		相邻NN间期差值均方根	
	MAE↓	均方根误差↓	MAE↓	均方根误差↓	MAE↓	均方根误差↓	MAE↓	均方根误差↓
Seq-rPPG	**1.07**	4.15	**12.7**	**18.7**	**0.137**	**0.168**	**22.9**	**30.6**
DeepPhys[3]	1.52	4.40	61.1	71.1	0.367	0.396	92.5	103
TS-CAN[12]	1.23	**3.59**	43.0	56.1	0.267	0.304	65.1	80.3
PhysNet[39]	1.12	4.13	30.2	37.3	0.293	0.319	61.0	67.7
PhysFormer[41]	1.56	6.28	22.8	28.1	0.267	0.296	48.7	54.7
CHROM[6]	6.86	15.57	56.1	65.1	0.398	0.420	98.9	109
POS[38]	4.25	12.06	78.0	83.1	0.502	0.518	142	149
独立成分分析[24]	6.05	13.3	77.3	82.8	0.505	0.524	136	145

心率：心率，**正常心搏间期标准差**：NN间期标准差，**NN50间期百分比**：NN50间期百分比，**相邻NN间期差值均方根**：相邻差值均方根，**平均绝对误差**：平均绝对误差，**均方根误差**：均方根误差。

表IV：在 UBFC-rPPG 上进行跨数据集测试，并比较不同训练集。**加粗**：最佳结果。

方法	训练集	HR		SDNN		NN50间期百分比		相邻NN间期差值均方根	
		MAE↓	均方根误差↓	MAE↓	均方根误差↓	MAE↓	均方根误差↓	MAE↓	均方根误差↓
Seq-rPPG		**0.918**	**1.42**	**5.15**	**9.04**	**0.078**	**0.131**	**11.1**	**18.2**
DeepPhys[3]	RLAP	1.30	2.39	31.2	42.2	0.357	0.388	60.8	72.8
TS-CAN[12]	RLAP	1.11	1.51	25.5	30.9	0.349	0.373	54.3	62.1
PhysNet[39]	RLAP	0.962	1.47	13.9	17.4	0.214	0.234	28.7	33.1
PhysFormer[41]	RLAP	1.05	1.55	10.4	13.2	0.159	0.179	20.3	24.5
Seq-rPPG	PURE	1.22	1.84	18.8	29.0	0.237	0.274	35.1	45.9
DeepPhys[3]	PURE	1.97	5.09	51.0	60.9	0.487	0.514	90.4	101.4
TS-CAN[12]	PURE	1.17	1.71	38.1	48.5	0.423	0.446	70.2	79.5
PhysNet[39]	PURE	1.29	1.83	27.3	31.5	0.388	0.413	57.5	63.9
PhysFormer[41]	PURE	1.60	3.07	22.4	26.1	0.357	0.376	48.1	52.5
CHROM[6]	PURE	6.10	19.6	22.7	29.0	0.336	0.359	49.0	59.2
POS[38]	PURE	2.54	8.97	43.6	49.9	0.524	0.539	93.4	100
独立成分分析[24]	PURE	1.59	2.55	44.8	52.3	0.488	0.508	91.5	101

心率：心率，**正常心搏间期标准差**：NN间期标准差，**NN50间期百分比**：NN50间期百分比，**相邻NN间期差值均方根**：相邻差值均方根，**平均绝对误差**：平均绝对误差，**均方根误差**：均方根误差。

表 V：交叉　数据集测试：在 PURE 上比较不同训练集。**粗体**：最佳结果　.

方法	训练集	HR		SDNN		NN50间期百分比		相邻NN间期差值均方根	
		MAE↓	均方根误差↓	MAE↓	均方根误差↓	MAE↓	均方根误差↓	MAE↓	均方根误差↓
Seq-rPPG		**0.318**	**0.597**	**11.2**	**19.6**	**0.111**	**0.138**	**17.7**	**26.1**
DeepPhys[3]	RLAP	3.10	8.63	93.5	99.4	0.485	0.514	150	158
TS-CAN[12]	RLAP	2.19	6.69	74.8	86.2	0.404	0.442	120	136
PhysNet[39]	RLAP	0.420	0.666	21.3	33.3	0.224	0.257	38.3	50.1
PhysFormer[41]	RLAP	1.56	9.45	20.6	31.1	0.203	0.230	35.6	45.3
Seq-rPPG	UBFC	21.5	35.0	83.2	92.0	0.394	0.443	116	130
DeepPhys[3]	UBFC	8.31	15.3	88.0	93.2	0.631	0.644	145	150
TS-CAN[12]	UBFC	16.1	23.8	96.9	99.6	0.681	0.687	156	158
PhysNet[39]	UBFC	8.82	19.2	59.6	66.7	0.385	0.422	96.5	107
PhysFormer[41]	UBFC	16.3	28.2	70.5	75.5	0.461	0.495	109	114
CHROM[6]	UBFC	2.76	13.3	47.4	64.6	0.276	0.327	73.4	95.5
POS[38]	UBFC	0.402	0.658	64.3	74.2	0.483	0.508	127	145
独立成分分析[24]	UBFC	0.805	3.36	76.5	85.6	0.461	0.492	144	159

心率：心率，**正常心搏间期标准差**：NN间期标准差，**NN50间期百分比**：NN50间期百分比，**相邻NN间期差值均方根**：相邻差值均方根，**平均绝对误差**：平均绝对误差，**均方根误差**：均方根误差。

在 PURE 上，使用 RLAP 训练的模型比使用 UBFC-rPPG 训练的模型表现更好。这表明 RLAP数据集质量更高，泛化能力更强，因此更适合作为训练集。

D. 计算开销

我们在一台典型的移动设备上测试了多种算法的平均帧处理时间：Raspberry Pi 4B（CPU：Cortex-A72，4 核）。与表现最佳的 PhysNet 和 TS-CAN 相比，Seq-rPPG 的计算开销更低（约为它们的 1/30），同时性能相近或更优。详情请见表VI 。

表VI：移动端 CPU 上的计算开销

模型	分辨率	每帧 FLOPs（M）	每帧时间（毫秒）
Seq-rPPG	8x8	0.26	0.36
DeepPhys[3]	36x36	52.16	9.09
TS-CAN[12]	36x36	52.16	10.72
PhysNet[39]	32x32	54.26	9.69
PhysFormer[41]	128x128	323.80	150

五、讨论

*A. RLAP*数据集

表格 IV 和 V 表明，Seq-rPPG模型和所有基线模型在使用RLAP数据集进行训练时表现更好，尤其是在心率变异性任务中，这证实了我们在数据采集过程中所倡导的观点。与其他数据集相比，例如UBFC-rPPG[1] 数据集，该数据集相对简单，缺乏复杂场景，且BVP标签未能实现完全同步。这种简单性导致泛化能力下降，而BVP标签的偏移则会影响心率和心率变异性任务的训练。因此，该数据集不适合作为训练集，因为在UBFC-rPPG上训练的模型表现较差。PURE[28] 数据集的结果明显更好，该数据集包含六种不同的场景，并且BVP信号与视觉内容尽可能实现同步匹配，这验证了我们关于数据集采集过程中同步重要性的假设。然而，我们的RLAP数据集取得了最佳表现；与PURE相比，RLAP不仅包含复杂场景和同步标签，还涵盖了各种能够引起心率变异性变化的情绪和参与度任务，同时拥有大得多的数据量——总计32.7小时的视频，而PURE仅包含1小时的视频。

B. Seq-rPPG 方法

在数据集内测试和跨数据集测试中，Seq-rPPG 在心率和心率变异性相关任务中表现出色。这种表现归功于 Seq-rPPG 的两个关键特性：较大的上下文窗口，以及独特的时频多层一维解码器架构。通常，rPPG算法会考虑联合时空建模，即输入高分辨率多帧图像，在时间和空间维度上进行建模。然而，受计算资源限制，这些模型通常无法捕捉长程时间关联，这一点可从它们的表现中看出。相对较短的时间窗口，例如 PhysFormer[41] 中的 160 帧和 PhysNet[39] 中的 128 帧。扩展此类空间聚焦模型中的时间关联会显著增加计算负载，时间复杂度为 $O(N^3)$。然而，Seq-rPPG 在这方面表现出色。与一些近期通过简单的编码器将视频输入编码为一维信号、专注于空间[32], [11], [14], 的方法不同，Seq-rPPG 即使在计算资源有限的情况下也能增大时间窗口，将复杂度降低至 $O(N)$。此外，其快速傅里叶变换卷积层（也称为频谱变换层）能够构建全局感受野，这意味着有效时间窗口等于模型输入的长度。总之，Seq-rPPG 最大限度地扩展了时间窗口，以建立长程关联，更好地捕捉周期性特征，从而提升性能。

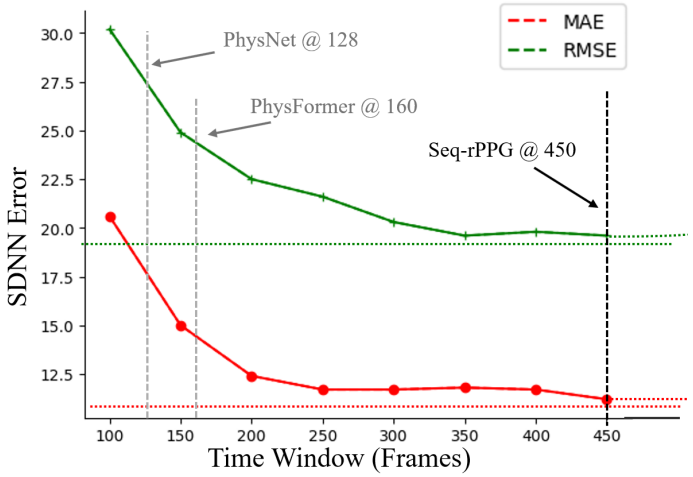


图3：Seq-rPPG 的时间窗口大小与 SDNN 误差之间的关系。

为进一步验证远程光电容积脉搏波算法对时间窗口的敏感性，我们为 Seq-rPPG 配置了多种参数，时间窗口范围从 100 帧到最多 450 帧。我们在 RLAP 上训练，并在 PURE 上测试，观察时间窗口变化时的 SDNN 误差曲线，如图3所示。起初，误差迅速下降，但当窗口大小继续增至 400 帧以上时，性能提升便不再显著。因此，我们为 Seq-rPPG 选择了 450 帧的窗口大小，对应 15 秒的视频输入。作为参考，图中还标出了 PhysNet[39] 和 PhysFormer[41] 的时间窗口，这也解释了 Seq-rPPG 为何在心率变异性任务上表现更佳。

如图 4所示，在 PURE 数据集中，局部头部运动往往会导致模型输出出现失真。通过绘制模型输出的波形可以观察到，较短的时间窗口应对失真的能力较弱，因此输出中包含更多噪声。Seq-rPPG模型采用

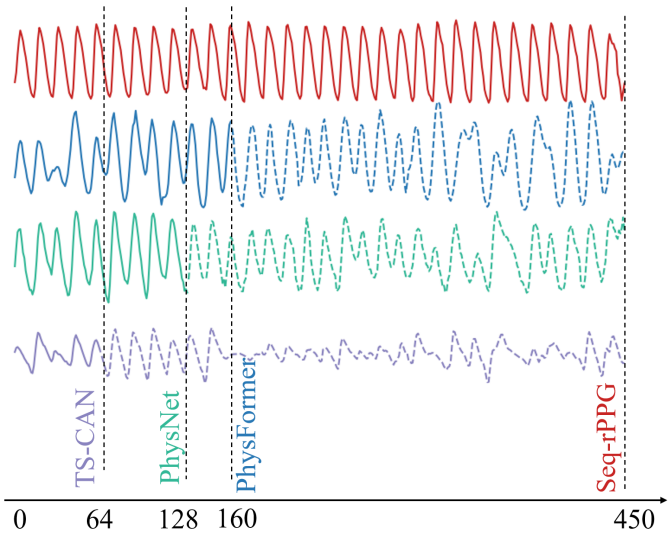


图4：四种模型在头部运动期间输出的BVP波形，实线表示时间窗口的大小。

450 帧的较大窗口，表明其具备强大的能力，可修复头部运动产生的噪声；其解码器架构能够生成真实的 BVP波形。

由于Seq-rPPG采用一维结构，因此能够以较低的计算成本获得较大的时间窗口。它在准确性方面具有显著优势，同时计算开销极低，仅为其他基线方法的3%，如表VI所示。因此，我们认为这是一种极具潜力的结构，能够在保持轻量化的同时实现高准确率。

VI. 结论

在本研究中，我们收集了RLAP公共数据集，该数据集适用于远程学习和情感计算，并包含专为心率变异性任务设计的BVP信号。在以往的远程光电容积描记数据集中，许多研究未关注标签与视频之间的严格同步，导致帧率波动和信号延迟引起脉搏峰值偏移，给心率变异性任务带来了重大挑战。RLAP解决了这一问题，并公开了其数据采集工具PhysRecorder，有助于未来收集更多高质量数据集。

我们提出了 Seq-rPPG 算法，该算法在 HRV任务中表现出色。分析表明，该算法的时间窗口和时间感受野对HRV任务至关重要，这为未来 rPPG算法的设计提供了指导。Seq-rPPG 还是一种轻量级算法，可以在移动设备上轻松实时运行，从而促进 rPPG算法的广泛应用。

致谢

本研究得到人因工程国家重点实验室基金（项目编号：HENKL.2024W06）和国家自然科学基金的支持（NSFC，项目编号：62366043、62472244）、清华大学科研计划项目、北京市自然科学基金（项目编号：QY23124、QY24248）的支持。国家自然科学基金项目（项目编号：62277029）和教育部人文社会科学研究项目（项目编号：20YJC880100）也为本研究提供了支持。本文所述涉及人类受试者的实验程序已获机构审查委员会批准。

参考文献

[1]Serge Bobbia、Richard Macwan、Yannick Benezeth、Alamin Mansouri 和 Julien Dubois。用于远程光电容积描记的无监督皮肤组织分割。《模式识别快报》，124:82–90，2019。[2] Deivid Botina-Monsalve、Yannick Benezeth 和 Johel Miteran。Rtrppg：一种用于实时远程光电容积描记的超轻量3D卷积神经网络。载于*IEEE/CVF*计算机视觉与模式识别会议研讨会论文集（*CVPRW*），第2146–2154页，2022年6月。[3] Weixuan Chen 和 Daniel McDuff。DeepPhys：利用卷积注意力网络进行基于视频的生理测量。载于欧洲计算机视觉会议论文集（*ECCV*），2018年9月。[4]Lu Chi、Borui Jiang 和 Yadong Mu。快速傅里叶卷积。载于 H. Larochelle、M. Ranzato、R. Hadsell、M.F. Balcan 和 H. Lin（编），神经信息处理系统进展，第33卷，第4479–4488页。Curran Associates, Inc.，2020。[5] J. Comas、A. Ruiz 和 F. Sukno。结合时间导数模块和时间平移不变损失的高效远程光电容积描记。载于*2022*年*IEEE/CVF*计算机视觉与模式识别会议研讨会（*CVPRW*），第2181–2190页，美国加利福尼亚州洛斯阿拉米托斯，2022年6月。IEEE计算机学会。[6] Gerard de Haan 和 Vincent Jeanne。基于色度的rPPG鲁棒脉率。*IEEE*生物医学工程汇刊，60(10):2878–2886，2013。[7] Justin R. Estepp、Ethan B. Blackford 和 Christopher M. Meier。利用多成像器阵列，在运动伪影干扰下恢复非接触式成像光电容积描记测得的脉率。载于*2014*年*IEEE*国际系统、人类与控制论会议（*SMC*），第1462–1469页，2014。[8] Shalom Greene、Himanshu Thapliyal 和 Allison Caban-Holt。压力检测情感计算综述：评估压力检测技术以改善健康状况。《*IEEE*消费电子杂志》，5(4):44–56，2016。[9] Amogh Gudi、Marian Bittner 和 Jan van Gemert。利用干净真实值进行评估的实时网络摄像头心率及其变异性估计。《应用科学》，10(23)，2020。[10] Xiaobai Li、Iman Alikhani、Jingang Shi、Tapio Seppanen、Juhani Junttila、Kirsi Majamaa-Voltti、Mikko Tulppo 和 Guoying Zhao。OBF数据库：用于远程生理信号测量的大型人脸视频数据库

及心房颤动检测。载于*2018*年第*13*届*IEEE*国际人脸与手势识别会议（*FG 2018*），第242–249页，2018。[11] Mingxaun Liu、Jiankai Tang、Haoxiang Li、Jiahao Qi、Siwei Li、Kegang Wang、Yuntao Wang 和 Hong Chen。Spiking-physformer：采用并行脉冲驱动变换器的基于摄像机的远程光电容积描记。*arXiv*预印本 *arXiv:2402.04798*，2024。[12] Xin Liu、Josh Fromm、Shwetak Patel 和 Daniel McDuff。用于设备端非接触式生命体征测量的多任务时序移位注意力网络。载于 H. Larochelle、M. Ranzato、R. Hadsell、M.F. Balcan 和 H. Lin（编），神经信息处理系统进展，第33卷，第19400–19411页。Curran Associates, Inc.，2020。[13] Xin Liu、Brian Hill、Ziheng Jiang、Shwetak Patel 和 Daniel McDuff。EfficientPhys：实现简单、快速且准确的基于摄像机的心脏测量。载于*IEEE/CVF*冬季计算机视觉应用会议论文集，第5008–5017页，2023。[14] Xin Liu、Yuntao Wang、Sinan Xie、Xiaoyu Zhang、Zixian Ma、Daniel McDuff 和 Shwetak Patel。Mobilephys：个性化的基于移动摄像机的非接触式生理感测。*ACM*交互、移动、可穿戴与泛在技术汇刊，6(1):1–23，2022。[15] Daniel McDuff、Miah Wander、Xin Liu、Brian L Hill、Javier Hernandez、Jonathan Lester 和 Tadas Baltrusaitis。Scamps：用于摄像机生理信号测量的合成数据。*arXiv*预印本 *arXiv:2206.04197*，2022。[16] Daniel J. McDuff、Ethan B. Blackford 和 Justin R. Estepp。视频压缩对使用成像光电容积描记进行远程心脏脉搏测量的影响。载于*2017*年第*12*届*IEEE*国际人脸与手势识别会议（*FG 2017*），第63–70页，2017。[17]Rita Meziati Sabour、Yannick Benezeth、Pierre De Oliveira、Julien Chappe 和 Fan Yang。UBFC-Phys：用于社会压力心理生理学研究的多模态数据库。*IEEE*情感计算汇刊，第1–1页，2021。[18] Xuesong Niu、Hu Han、Shiguang Shan 和 Xilin Chen。Synrhythm：从通用到特定地学习深度心率估计器。载于*2018*年第*24*届国际模式识别会议（*ICPR*），第3580–3585页，2018。[19] Xuesong Niu、Hu Han、Shiguang Shan 和 Xilin Chen。VIPL-HR：用于在较少约束条件下从人脸视频估计脉搏的多模态数据库。载于亚洲计算机视觉会议，第562–576页。施普林格，2018。[20] Xuesong Niu、Shiguang Shan、Hu Han 和 Xilin Chen。RhythmNet：通过时空表征端到端地从人脸估计心率。*IEEE*图像处理汇刊，29:2409–2423，2020。[21] Xuesong Niu、Zitong Yu、Hu Han、Xiaobai Li、Shiguang Shan 和 Guoying Zhao。通过交叉验证特征解耦实现基于视频的远程生理测量。载于 Andrea Vedaldi、Horst Bischof、Thomas Brox 和 Jan-Michael Frahm，编辑，《计算机视觉——ECCV 2020》，第295–310页，尚，2020年。Springer International Publishing。[22] Ewa M. Nowara、Tim K. Marks、Hassan Mansour和Ashok Veeraraghavan。驾驶期间的近红外成像光电容积描记。《IEEE智能交通系统汇刊》，23(4):3589–3600，2022年。[23] Ewa Magdalena Nowara、Tim K. Marks、Hassan Mansour和Ashok Veeraraghavan。Sparseppg：利用近红外基于摄像头的生命体征估计实现驾驶员监测。载于*2018*年*IEEE/CVF*计算机视觉与模式识别会议研讨会（*CVPRW*），第1353–135309页，2018年。[24] Ming-Zher Poh、Daniel J McDuff和Rosalind W Picard。使用网络摄像头进行非接触式、多参数生理测量的进展。《IEEE生物医学工程汇刊》，58(1):7–11，2010年。[25] Ming-Zher Poh、Daniel J McDuff和Rosalind W Picard。利用视频成像和盲源分离进行非接触式自动心脏脉搏测量。《光学快报》，18(10):10762–10774，2010年。[26] Steven A. Shafer。使用颜色分离反射分量。《色彩研究与应用》，10:210–218，1985年。[27] Ivan Shchekotov、Pavel Andreev、Oleg Ivanov、Aibek Alanov和Dmitry Vetrov。FFC-SE：用于语音增强的快速傅里叶卷积。《arXiv预印本 arXiv:2204.03042》，2022年。[28] Ronny Stricker、Steffen Müller和Horst-Michael Gross。移动服务机器人上的非接触式视频脉率测量。载于第*23*届*IEEE*机器人与人类交互通信国际研讨会，第1056–1062页，2014年。[29] Weiyu Sun、Xinyu Zhang、Ying Chen、Yun Ge、Chunyu Ji和Xiaolin Huang。BYHE：一种用于提升端到端视频心率测量网络的简单框架。《arXiv预印本 arXiv:2207.01697》，2022年。[30] Ilya Sutskever、Oriol Vinyals和Quoc V Le。使用神经网络进行序列到序列学习。《神经信息处理系统进展》，27，2014年。[31] Jiankai Tang、Kequan Chen、Yuntao Wang、Yuanchun Shi、Shwetak Patel、Daniel McDuff和Xin Liu。MMPD：多领域移动视频生理数据集。《arXiv预印本 arXiv:2302.03840 》，2023 年。[32] Jiankai Tang 、Xinyi Li、Jiacheng Liu、Xiyuxing Zhang、Zeyu Wang和Yuntao Wang。面向不同肤色数百名受试者的基于摄像头的远程生理感测。《arXiv预印本 arXiv:2404.05003》，2024年。[33] Jiankai Tang、Kegang Wang、Hong ming Hu、Xiyuxing Zhang、Peiyu Wang、Xin Liu和Yuntao Wang。ALPHA：使用大语言模型进行异常生理健康评估。《arXiv预印本 arXiv:2311.12524》，2023年。[34] Sergey Tulyakov、Xavier Alameda-Pineda、Elisa Ricci、Lijun Yin、Jeffrey F. Cohn和Nicu Sebe。用于从

现实条件下面部视频中进行心率估计的自适应矩阵补全。载于*2016*年*IEEE*计算机视觉与模式识别会议（*CVPR*），第2396–2404页，2016年。[35] Paul Van Gent、Haneen Farah、Nicole Nes和Bart van Arem。面向人为因素的心率分析：用于噪声自然场景心率数据的开源工具包的开发与验证。载于第*6*届*HUMANIST*会议论文集，第173–178页，2018年。[36] Paul Van Gent、Haneen Farah、Nicole Van Nes和Bart Van Arem。Heartpy：一种用于分析噪声信号的新型心率算法。《交通研究F辑：交通心理学与行为》，66:368–378，2019年。[37] Wenjin Wang、Albertus C. den Brinker、Sander Stuijk和Gerard de Haan。远程光电容积描记的算法原理。《IEEE生物医学工程汇刊》，64(7):1479–1491，2017年。[38] Wenjin Wang、Albertus C. den Brinker、Sander Stuijk和Gerard de Haan。远程光电容积描记的算法原理。《IEEE生物医学工程汇刊》，64(7):1479–1491，2017年。[39]Zitong Yu、Xiaobai Li和Guoying Zhao。使用时空网络从面部视频中测量远程光电容积描记信号。《arXiv预印本 arXiv:1905.02419》，2019年。[40] Zitong Yu*、Wei Peng*、Xiaobai Li、Xiaopeng Hong和Guoying Zhao。从高度压缩的面部视频中进行远程心率测量：一种结合视频增强的端到端深度学习解决方案。载于国际计算机视觉大会（*ICCV*），2019年。[41] Zitong Yu、Yuming Shen、Jingang Shi、Hengshuang Zhao、Philip Torr和Guoying Zhao。PhysFormer：结合时间差分变换器的基于面部视频的生理测量。载于*CVPR*，2022年。[42] Zheng Zhang、Jeff M. Girard、Yue Wu、Xing Zhang、Peng Liu、Umur Ciftci、Shaun Canavan、Michael Reale、Andy Horowitz、Huiyuan Yang、Jeffrey F. Cohn、Qiang Ji和Lijun Yin。用于人类行为分析的多模态自发情绪语料库。载于*IEEE*计算机视觉与模式识别会议论文集（*CVPR*），2016年6月。