

DOI: 10.13382/j.jemi.B2609321

基于生成式动画与生物物理嵌入的 rPPG 面部视频合成算法*

李沂杭^{1,2} 程世超^{1,2} 赵昶辰^{1,2} 张建海^{1,2}

(1. 杭州电子科技大学计算机学院 杭州 310018; 2. 全省(浙江)脑机协同智能技术及应用重点实验室 杭州 310018)

摘要: 远程光电容积脉搏波描记法作为一种非接触式生理监测技术,在临床监护与人机交互领域具有重要价值。然而,深度学习模型在 rPPG 任务中的性能高度依赖于大规模且多样化的标注数据。现有的基准数据集(如 PURE 和 UBFC-rPPG)受限于受试者规模且运动场景较为单一,导致模型在面对复杂非线性运动干扰时泛化能力不足。为此,提出一种基于生成式动画与生物物理嵌入的面部视频合成框架。该框架采用运动生成-信号嵌入的二阶段策略。首先,利用生成式框架 LivePortrait 将驱动视频中的大幅度动作与表情迁移至静态图像,构建高保真的动态面部视频,有效模拟了真实场景中的宏观位移噪声;随后,提出一种基于层分离的生物物理嵌入模型,将生理信号植入视频中。该模型利用高斯大核算子将图像分解为基准层与细节层,在低频分量中引入乘性调制,以更符合皮肤光学特性和血流变化的真实表现,在保留皮肤细节纹理的同时,确保了微弱生理信号在光谱与时域维度的物理准确性。在 UBFC-rPPG 和 PURE 数据集上的实验结果表明,在样本受限的条件下,加入本研究生成的合成数据后,PhysNet 在 UBFC-rPPG 数据集上的平均绝对误差(MAE)由 1.11 bpm 降至 0.78 bpm,信噪比(SNR)由 5.50 dB 提升至 5.99 dB;在 PURE 数据集上的 MAE 由 1.16 bpm 降至 0.99 bpm,SNR 由 8.17 dB 提升至 8.36 dB。结果表明,本研究生成的合成数据能够有效提升 rPPG 模型的心率估计精度与信号质量,为解决 rPPG 领域的数据稀缺问题提供了一种兼具物理确定性与运动真实感的可行路径。

关键词: 远程光电容积描记术;合成视频生成;数据增强;生物物理建模

中图分类号: TN911.73;TP391 **文献标识码:** A **国家标准学科分类代码:** 310.61

rPPG facial video synthesis via generative animation and biophysical embedding

Li Yihang^{1,2} Cheng Shichao^{1,2} Zhao Changchen^{1,2} Zhang Jianhai^{1,2}

(1. School of Computer Science, Hangzhou Dianzi University, Hangzhou 310018, China; 2. Zhejiang Provincial Key Laboratory of Brain Computer Collaborative Intelligence Technology and Applications, Hangzhou 310018, China)

Abstract: Remote photoplethysmography (rPPG), as a non-contact physiological monitoring technique, is of great value in clinical monitoring and human-computer interaction. However, the performance of deep learning models in rPPG tasks heavily relies on large-scale and diverse annotated data. Existing benchmark datasets, such as PURE and UBFC-rPPG, are limited by subject scale and relatively simple motion scenarios, which leads to insufficient generalization when models face complex nonlinear motion interference. Therefore, this paper proposes a facial video synthesis framework based on generative animation and biophysical embedding. The framework adopts a two-stage strategy of motion generation and signal embedding. First, the generative framework LivePortrait is used to transfer large-scale motions and facial expressions from driving videos to static images, constructing high-fidelity dynamic facial videos and effectively simulating macroscopic displacement noise in real scenes. Then, a layer-separated biophysical embedding model is proposed to implant physiological signals into videos. This model uses a large-kernel Gaussian operator to decompose the image into a

收稿日期: 2026-04-30 Received Date: 2026-04-30

* 基金项目: 国家自然科学基金面上项目(62573171)、浙江省自然科学基金(LZ25F030005,LY24F020015)、浙江省属高校基本科研业务费专项资金(GK269910299001-009)、全省(浙江)脑机协同智能技术及应用(含中央引导地方科技发展资金项目编号:2025ZY01045、2026ZY01008;计划编号:2025E10015)项目资助

base layer and a detail layer, and introduces multiplicative modulation into the low-frequency component, which better conforms to skin optical characteristics and blood-flow variations. While preserving detailed skin texture, it ensures the physical accuracy of weak physiological signals in both spectral and temporal dimensions. Experimental results on the UBFC-rPPG and PURE datasets show that, under sample-limited conditions, after introducing the synthetic data generated by this paper, the MAE of PhysNet on the UBFC-rPPG dataset decreases from 1.11 bpm to 0.78 bpm, and the SNR increases from 5.50 dB to 5.99 dB; on the PURE dataset, the MAE decreases from 1.16 bpm to 0.99 bpm, and the SNR increases from 8.17 dB to 8.36 dB. The results demonstrate that the synthetic data generated by this paper can effectively improve the heart-rate estimation accuracy and signal quality of rPPG models. This study provides a feasible approach with both physical determinacy and motion realism for addressing data scarcity in the rPPG field.

Keywords: rPPG; synthetic video generation; data augmentation; biophysical modeling

0 引言

远程光电容积脉搏波描记法 (remote photoplethysmography, rPPG)^[1]作为一种非接触式生理信号测量技术,近年来在计算机视觉与生物医学工程领域引起了广泛关注。其核心物理机制基于光学/生理学原理:当心脏搏动引导血液流经面部毛细血管网时,皮肤组织对特定波长光线的吸收率会随血容量的周期性变化而发生微弱波动^[2]。通过普通消费级摄像头捕捉这些肉眼难以察觉的颜色变化,即可提取出包含心率、心率变异性以及呼吸频率等关键生理指标^[3]。相比于传统的接触式传感器(如心电图或指夹式脉搏血氧仪),rPPG具有无创、非接触、易于部署等显著优势^[4],在新生儿监护^[5]、老年人远程医疗^[6]、驾驶员疲劳监测^[7],以及心理压力评估等场景中展现出巨大的应用潜力^[8]。

早期的 rPPG 信号估计方法主要依赖于手工特征和线性分解技术。文献[9-10]先后提出了基于盲源分离的算法,如独立成分分析和主成分分析,旨在从红、绿、蓝3个通道中解耦出脉搏信号。随后,基于物理光学模型的算法,如 CHROM^[11]、POS^[2]及 PBV^[12]等,通过预定义的投影方向来抑制由于光照变化和运动引起的噪声^[13]。

近年来,随着深度神经网络^[14]在特征提取任务中的崛起,端到端的深度学习模型逐渐成为 rPPG 领域的主流^[15]。通过卷积神经网络提取空间特征^[16],结合长短期记忆网络^[17]或时间偏移卷积网络 (TS-CAN)^[18]处理时域波动,深度模型在处理复杂光照和大幅度动作干扰方面表现出了优于传统算法的非线性建模能力。

尽管深度学习模型性能强劲,但其高度依赖于大规模且标注的多样的场景如(非均匀动态光照、大幅度非线性运动以及复杂的背景环境)的标注数据。目前,学术界广泛使用的基准数据集主要包括 PURE^[19]和 UBFC-rPPG^[20]。PURE 数据集记录了受试者在静坐、办公及头部旋转等特定任务下的面部视频,为早期运动鲁棒性研究提供了基础^[21]。UBFC-rPPG 数据集则侧重于受控环境下的高信噪比信号采集。

然而,这些数据集在实际应用中暴露出明显的局限

性。首先是样本规模受限,这些数据通常仅涵盖几十名受试者,且多为实验室环境下的受控采集,导致模型难以学习到真实场景中极其多样的皮肤质感、年龄特征及复杂的环境光照分布。其次是运动场景单一:虽然 PURE 包含了一定的运动,但其动作类型较为机械且重复,无法模拟现实生活中非线性、随机的面部表情变化及大范围的姿态切换。这种数据分布的狭窄性导致模型在训练集上表现优异,但在面对跨场景、跨受试者的实际测试时,往往会出现严重的性能退化,即泛化性问题。

为了弥补小规模数据集的缺陷^[22],数据增强成为提升模型泛化能力的必要手段。传统的数据增强方法(如镜像翻转、随机裁剪、色彩抖动等)虽然在通用的目标检测或图像分类任务中效果显著,但在 rPPG 领域却面临挑战。rPPG 信号属于极其微弱的时空波动,简单的空间变换无法模拟皮肤下血管灌注的动态过程,而过度的颜色抖动则会直接破坏生理信号的物理完整性。

另一类方法尝试利用计算机图形学^[23]或生成对抗网络^[24]合成虚拟人脸数据。尽管这类方法可以生成大规模的样本,但合成图像往往存在纹理过拟合、面部解剖结构扭曲或生理波动分布不自然等问题。尤其是在运动建模方面,如何生成既具有真实复杂的宏观位移,又保留微弱物理确定性的生理波动视频,仍然是当前合成数据增强领域亟待解决的核心难题。

本文提出一种基于生成式动画与生物物理嵌入的 rPPG 数据增强框架,该框架的核心逻辑在于构建了一套“运动生成-信号嵌入”的解耦二阶段流水线。在动态面部建模阶段,本框架利用基于隐式关键点表示的 LivePortrait^[25]引擎,通过深度重定向与缝合技术将驱动视频中的复杂非线性动作(如大幅度头部旋转、面部微表情)精确迁移至静态源图像,生成高保真的动态面部视频;相比于先前的一阶运动模型,该引擎能有效克服大幅动作下的面部塌陷与背景扭曲,且这种先动后嵌的顺序规避了深度渲染网络在像素重构过程中对微弱搏动信号产生的平滑效应,从底层数值层面保障了信号的完整性。在生理信号嵌入阶段,本文提出一种结合空间异质性与物理规律的精密植入机制。首先,利用 MediaPipe FaceMesh^[26]实时追踪动态视频中的 468 个面部地标点,

精确定位额头与双颊等高血管灌注区域,并引入基于距离变换的注意力加权算法生成具有生理梯度的空间掩码,模拟皮肤血流灌注从中心向边缘平滑递减的物理特征,消除植入边界的视觉伪影。随后,本框架采用层分离技术将视频帧分解为承载次表面散射信息的低频基准层与保留解剖纹理的高频细节层,并遵循朗伯反射定律^[27]对基准层实施乘性调制,使信号振幅随局部照度自然衰减。在 PURE 和 UBFC-rPPG 数据集上的对比实验证明,该方法通过提供兼具运动真实感与物理确定性的增强样

本,有效弥补了现有数据集动作多样性与样本量的不足,显著提升了心率估计的稳健性与泛化能力。

1 方法

本文提出的基于生成式动画与生物物理嵌入的 rPPG 面部视频合成算法整体框架如图 1 所示。该框架主要包括生理信号处理、视频生成与特征提取、生理信号嵌入以及后处理输出 4 个模块。

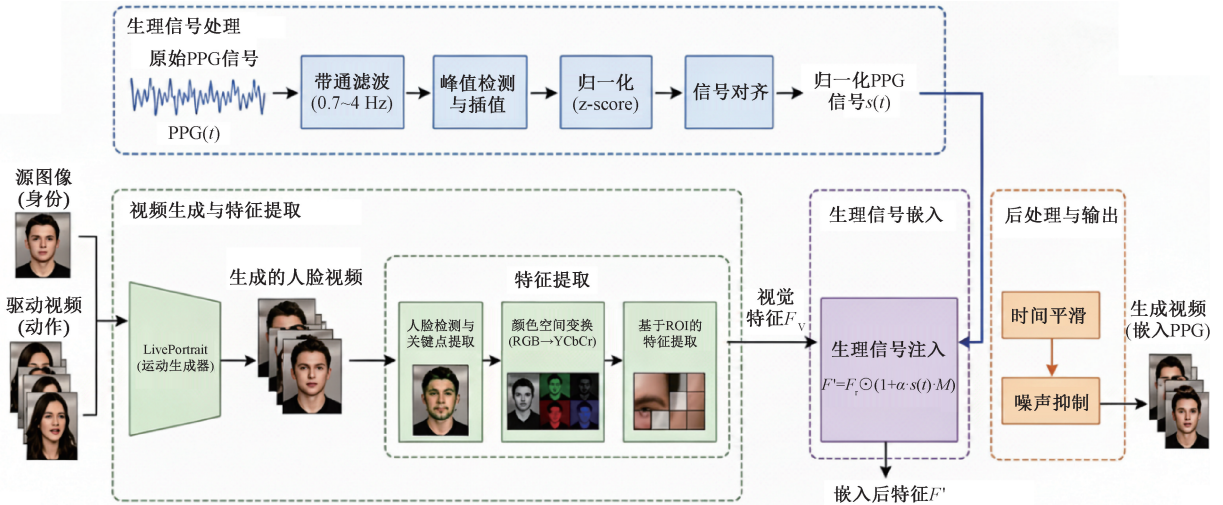


图1 本文提出的基于生成式动画与生物物理嵌入的 rPPG 数据增强框架图

Fig. 1 Framework of the proposed rPPG data augmentation based on generative animation and biophysical embedding

1.1 基于生成式框架的动态面部建模

本文的输入端仅由少量的静态源图像 I_s 与包含目标动作的驱动视频 V_d 组成。为了模拟真实场景中面部复杂的动态变化,本文引入 LivePortrait 作为核心运动生成引擎。该引擎通过隐式关键点重定向技术,将驱动视频 V_d 中的头部姿态偏移、面部表情变化及细微肌肉动作迁移至源图像 I_s 上,从而生成具有高度真实感的动态面部基准序列。

本文在架构设计上严格遵循先运动建模,后生理增强的二阶段流水线。这一顺序的选择具有关键的生物物理学意义:LivePortrait 等神经渲染网络在重建像素时,本质上是一个基于高维特征映射与非线性解码的渲染过程。这种神经渲染机制倾向于优化大尺度的视觉连贯性,往往会将极微弱的皮肤颜色波动^[27] (即信号强度通常低于 1 个灰度级的 rPPG 信号) 误判为高频噪声并进行抑制或平滑处理。若在生成动作之前嵌入信号, rPPG 分量极易在神经渲染器的特征瓶颈处丢失。通过后置嵌入策略,确保了生理信号能够以物理确定性的方式叠加在已生成的动态像素上,从而最大程度地保障了信号的幅度完整性与时域频率纯净度。此外,该流程实现了宏

观运动特征与微观生理特征的有效解耦控制。通过同一张源图像配合多组驱动视频,我们可以批量合成具有一致生理参数(如同质的 PPG 信号分布)但运动强度与类型各异的训练样本。这种高度可控的数据生成模式显著增强了合成训练集的分布多样性,为提升 rPPG 估计算法在剧烈运动场景下的泛化性能提供了坚实的数据基础。

1.2 基于生理先验的空间定位与掩码生成

为了在合成视频中精确模拟真实人体的生理信号分布,本文提出了一种结合解剖学先验与几何变换的空间掩码生成方案。该方案旨在解决合成视频中常见的信号分布均匀化与边缘伪影问题,增强生理信号的物理真实感。首先,实现动态区域追踪与皮肤分割利用 MediaPipe FaceMesh^[26] 模型从生成的动态序列中逐帧提取 468 个面部地标点。考虑到面部微血管分布的异构性,本文通过关键点坐标精确界定了额头 ROI_h 及左右双颊 ROI_{cheek} 区域。解剖学研究表明,这些区域的毛细血管网最为丰富,是获取高信噪比 rPPG 信号的核心探测点^[28]。为了进一步排除眼睛、嘴巴及毛发等非皮肤区域的干扰,本文结合 YCrCb^[29] 颜色空间过滤与形态学闭运算生成了二值化皮肤掩码 M_s ^[30]。由于地标点随 LivePortrait 生成的

动作实时更新,该掩码在时域上能够与面部运动保持高度的一致性。其次,构建基于空间平滑先验的注意力加权机制。本文中距离变换权重并不假设 ROI 的几何中心必然对应真实血管灌注最强区域。由于普通 RGB 视频中难以直接获得个体化的皮下血管分布与灌注图,本文将距离变换作为一种 ROI 内部平滑先验与边界衰减约束使用,其主要作用是避免生理信号在 ROI 边界处发生突变,并降低非皮肤区域混入带来的伪影。额头与双颊区域被选为候选 ROI,是基于既有 rPPG 研究中这些区域通常具有较高脉搏可观测性的经验依据;但 ROI 内部的实际信号强度仍会受到个体血管分布、皮肤厚度、姿态以及局部光照条件的共同影响。因此,本文引入基于距离变换(distance transform, DT)的权重分配策略,对识别出的 ROI 区域应用欧氏距离变换 $DT(\cdot)$,并利用灌注指数 α 进行非线性映射,生成生理注意力掩码:

$$M_a = \left(\frac{DT(P_{ROI})}{\max(DT(P_{ROI}))} \right)^\alpha \quad (1)$$

式中: P_{ROI} 代表特定 ROI 内的像素集合。通过调节 α , 可以精细化控制生理信号从 ROI 中心向外周的衰减速率。最后,生成最终的空间分布掩码 M 。为了确保信号仅在有效的皮肤区域内波动,将皮肤掩码 M_s 与生理注意力掩码 M_a 进行融合,并辅以高斯平滑算子 G_σ 以进一步消除潜在的阶跃跳变。最终的空间分布掩码 M 定义为:

$$M = G_\sigma(M_s \odot M_a) \quad (2)$$

这种梯度化处理并非对个体真实灌注分布的精确重建,而是在候选皮肤 ROI 内提供连续、平滑且边界衰减的信号植入权重。该设计可以有效减弱嵌入信号在 ROI 边缘处的突变伪影,并与后续局部照度调制共同提高合成视频的自然度与物理合理性。

1.3 光谱特性建模与动态生理调制

为了使合成视频在底层物理逻辑上符合相机捕捉皮肤反射光的规律,本文基于 Shafer 双色反射模型^[31]对 rPPG 信号的光谱分布进行了精确建模。

不同波长的光在皮肤组织中的穿透深度与血红蛋白吸收率存在显著差异。研究表明,含氧血红蛋白在 550 nm(绿光波段)附近具有明显的吸收峰值,因此绿光通道携带了最强的搏动能量。为了模拟这一光学特性,引入光谱权重向量 w 对原始信号 $s(t)$ 进行多通道映射^[32]:

$$[w_R, w_G, w_B] = [0.22, 0.46, 0.32] \quad (3)$$

这一权重比例确保了合成视频在执行经典的基于色度或空间正交投影的提取算法(如 POS 或 CHROM)时,能够呈现出与真实皮肤一致的色度波动特征。

此外,在真实生理场景中,心血管系统受自主神经系统调节,表现为心跳信号随呼吸周期产生的节律性波动。为了提升合成数据集的复杂度与稳健性,本文引入了呼

吸诱导强度波动(respiratory induced intensity variation, RIIV)调制项^[33]。由于原始传感器采集频率(125 Hz)与视频帧率(30 fps)不一致,首先通过线性插值将生理信号对齐至视频时间轴。随后,将对齐后的信号 $s(t)$ 与模拟呼吸频率(约 0.2 Hz)的低频正弦波进行乘性耦合,得到增强后的生理信号 $s'(t)$ 为:

$$s'(t) = s(t) \times (0.9 + 0.1 \sin(2\pi \times 0.2t)) \quad (4)$$

本文采用固定的 0.2 Hz 呼吸调制频率,主要出于控制变量和实验可比性的考虑。该频率对应约 12 次/min,处于成人静息呼吸的常见范围。本文关注的是生理信号嵌入框架的有效性,而非个体化呼吸动力学建模。该调制方式模拟了呼吸导致的胸腔压力变化对外周血容量的影响。通过引入这种非平稳生理特征,生成的合成视频能够有效训练模型从复杂的频率背景中解耦并提取纯净的心率信号。

1.4 基于层分离的增强型生物物理嵌入模型

为了真实还原光线在皮肤组织中的传播特性,本文提出了一种基于层分离的生理信号嵌入模型,旨在模拟次表面散射效应^[34],同时保持面部纹理的自然度。

rPPG 信号本质上来源于皮肤真皮层内血液容积变化引起的漫反射波动。这一物理过程主要体现在光影的低频特征中,而毛孔、皱纹及毛发等高频纹理并不随心跳产生像素级的波动。基于此观察,利用大核高斯算子 G_k 将 LivePortrait 生成的原始动态帧 F 分解为代表弥散光信息的基准层 F_{base} 和代表纹理细节的细节层 F_{detail} 如下:

$$F_{base} = G_{k(F)} F, F_{detail} = F - F_{base} \quad (5)$$

其中, F_{base} 承载了主要的光学特征变化,用于植入生理信号,而 F_{detail} 则保留了面部的解剖学特征,确保信号嵌入过程不会导致表面纹理产生不自然的闪烁伪影。

在完成空间掩码生成与信号预处理后,本研究通过一个增强型生物物理模型将生理信号植入动态序列中。该模型的设计核心在于模拟皮肤组织的次表面散射效应并维持视频的辐射定标一致性。最终的合成帧 F' 通过式(6)重构。

$$F'_{c,i,j,t} = F_{base,c}(i,j,t) \times (1 + \gamma \times w_c \times M \times s' \times L) + F_{detail,c}(i,j,t) + \epsilon \quad (6)$$

式中: $c \in \{R, G, B\}$ 为通道索引; γ 在 $[0.01, 0.05]$ 范围内随机采样; w_c 为光谱权重; $M(i,j)$ 为空间分布掩码; $s'(t)$ 为经呼吸调制的生理信号; $L(i,j,t)$ 为提取自基准层的局部照度因子; 加性高斯白噪声项 ϵ 服从均值为 0、标准差为 1.0×10^{-4} 的正态分布。

该嵌入模型从光谱、空间、时域和图像分层 4 个方面约束生理信号的植入过程,以保证合成变化符合面部 rPPG 信号的基本规律。在光谱维度上,本文通过通道权重 w_c 使不同颜色通道具有不同的搏动响应,其中绿光通道承担较强的生理调制,以模拟血红蛋白吸收特性导致

的通道差异。在空间维度上,生理信号仅通过空间分布掩码 $M(i,j)$ 作用于有效皮肤 ROI,避免在眼睛、嘴唇、毛发等非皮肤区域产生不合理波动。在时域维度上,嵌入信号 $s'(t)$ 来源于真实 PPG 信号并结合呼吸诱导强度变化,因此能够保留心搏周期性及低频呼吸调制特征。此外,本文仅对低频基底层 F_{base} 进行乘性调制,而保持高频细节层 F_{detail} 不变,从而避免毛孔、皱纹和面部轮廓等纹理细节随心跳同步闪烁。由于 $s'(t)$ 、 $M(i,j)$ 和 $L(i,j,t)$ 均经过归一化处理,且全局嵌入强度 γ 被限制在较小范围内,最终植入的生理变化主要表现为弱幅度、低频率的局部亮度和色度波动,不会对面脸结构、身份特征和整体视觉质量造成明显破坏。

根据光学原理,相机捕捉到的皮肤表面反射光强度是入射光照度 L 与皮肤漫反射率 ρ 的乘积。rPPG 信号本质上是血液容积变化引起的反射率微小波动 $\Delta\rho$ 。因此,由生理特征引起的像素变化量 ΔI 必须满足 $\Delta I = L \cdot \Delta\rho$ 。式(6)中的乘性项准确模拟了生理信号作为反射率属性受局部光照驱动的物理过程。

在真实场景中,面部光照分布通常是不均匀的。若采用加性模型,在光照极弱的阴影区域(L 趋于 0)仍会植入恒定振幅的信号,导致暗部区域产生不自然的“自发光”现象。本文的乘性机制确保了生理信号随照度 L 的减弱而自然衰减,从而维持了合成视频在复杂光环境下的物理保真度。此外,通过仅对低频基准层 F_{base} 进行调制而保持细节层 F_{detail} 静止,有效避免了皮肤纹理随心跳产生的同步闪烁伪影。

在乘性模型下,生理信号的有效振幅随光照降低,而加性传感器噪声 ϵ 在全图分布相对均匀。这使得阴影区域的信噪比(signal-tonoise ratio, SNR)自然下降。这种设计为深度学习模型提供了一个极具挑战性且符合实际物理规律的训练环境,能够显著提升神经网络在处理低照度、高噪声视频时的稳健性。

2 实验

2.1 数据集

本文采用广泛应用于生理计算研究的 BIDMC 数据集^[35]作为合成信号的底层来源。该数据集包含 53 段通过接触式传感器采集的 8 min PPG 记录,原始采样频率为 125 Hz。在实验过程中,本文按照合成视频的目标帧率 30 fps 对信号进行了重采样处理。对于每一段生成的视频,截取对应信号的前 L 秒序列作为生理真值,其中 L 为生成序列的总时长。

基于上述生理信号,本文利用所提框架构建了名为 Physio-Ultra Synthetic 的增强型数据集。具体而言,本文首先利用大模型生成 10 张具有不同性别、年龄和面部外

观特征的静态人脸图像,并将其作为 LivePortrait 的源图像输入。这些图像不对应真实个体,用于构建 10 个虚拟受试者身份。驱动视频来源于 PURE 数据集。为保证合成视频的运动分布与公开基准数据集保持一致,本文为每个虚拟受试者设置 6 种动作场景,分别为静止、说话、慢速平移、快速平移、小幅旋转和中等幅度旋转,每段视频时长为 1 min。随后,本文将 BIDMC 数据集中的 PPG 生理信号与上述虚拟身份及动作序列进行组合,并通过 LivePortrait 完成动作迁移,生成包含头部旋转、眨眼、说话及面部姿态变化等复杂运动特征的合成面部视频序列。合成视频帧序列如图 2 所示。



图 2 合成视频帧序列示意图

Fig. 2 Example frame sequences from generated synthetic videos

PURE 数据集包含 10 名受试者(包括男性和女性)在 6 种不同动作任务下的面部视频,总计 60 段序列。视频采用工业相机采集,分辨率为 640×480 ,帧率为 30 fps。该数据集的特点是包含了多种受控的头部运动(如平移、旋转及交谈),为评估模型在运动干扰下的性能提供了理想环境。生理真值通过指夹式脉搏血氧仪同步采集,包含心率和原始 PPG 波形。

UBFC-rPPG 数据集专门设计用于评估 rPPG 算法。它由 42 段面部视频组成,受试者在参与一项旨在诱发心率变化的数学游戏任务时被录制。视频分辨率为 640×480 ,采用无压缩格式以保留微弱的皮肤颜色变化。数据集提供了由 CMS50E 血氧仪测得的同步 PPG 信号及心率值。该数据集因其真实的心率波动范围而成为评估 rPPG 模型泛化能力的核心基准。

为保证实验评估的公平性并避免身份信息泄露,本文对 PURE 和 UBFC-rPPG 数据集均采用受试者独立的数据划分方式,按照 7 : 2 : 1 的比例划分为训练集、验证集和测试集。其中,训练集用于模型参数学习,验证集用于超参数选择与模型保存,测试集仅用于最终性能评估。所有对比实验均采用相同的数据划分方式,以保证不同方法之间结果的可比性。

2.2 指标

为了客观评估本文提出的数据增强策略在提升心率估计精度及信号质量方面的有效性,本文采用了 4 项国

际通用的核心评价指标。其中,平均绝对误差(mean absolute error, MAE)与均方根误差(root mean square error, RMSE)用于衡量心率估计值的绝对精度,而相关系数 r 与SNR则用于评估估计信号与原始生理信号的相关性及信噪比。所有分析均基于滑动窗口机制执行,窗口长度设定为5 s,步长为1 s。

MAE用于衡量估计心率 HR_{est} 与生理真值 HR_{gt} 之间的平均绝对偏差。该指标对异常误差的敏感度较为线性,能够直观反映系统的整体测量精度:

$$MAE = \frac{1}{n} \sum_{k=1}^n |HR_{est}(k) - HR_{gt}(k)| \quad (7)$$

式中: k 表示第 k 个滑动窗口; n 为总窗口数。MAE的数值越低,代表估计精度越高。

RMSE与MAE相比,RMSE通过对误差项进行平方处理,更加强调大幅度离群误差对结果的影响。在rPPG研究中,该指标能有效评估模型在面对剧烈运动干扰时的稳健性:

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (HR_{est}(k) - HR_{gt}(k))^2} \quad (8)$$

其中,较低的RMSE值意味着模型的心率预测结果具有更高的稳定性与可靠性。

相关系数 r 用于衡量估计心率序列与生理真值序列之间的线性相关程度,其取值范围为 $[-1, 1]$:

$$r = \frac{\sum (HR_{est} - \overline{HR_{est}})(HR_{gt} - \overline{HR_{gt}})}{\sqrt{\sum (HR_{est} - \overline{HR_{est}})^2 \sum (HR_{gt} - \overline{HR_{gt}})^2}} \quad (9)$$

其中, r 值越接近1,表明合成信号增强后的模型在心率波动趋势的追踪上与真值保持了高度的一致性。

SNR^[11]是衡量rPPG信号质量的核心指标。它通过计算频率域内脉搏信号功率与声功率的比值来评估信号的清晰度:

$$SNR = 10 \lg \left(\frac{P_{\text{signal}}}{P_{\text{noise}}} \right) \quad (10)$$

式中: P_{signal} 为心率基频及其二阶谐波区域内的频谱功率之和; P_{noise} 为生理心率范围0.7~4.0 Hz内除信号功率外的其余噪声功率之和。较高的SNR值证明了本文提出的生物物理嵌入模型能够有效模拟清晰的生理搏动特征,同时抑制非搏动性噪声。

2.3 模型架构与实现细节

为了评估合成数据增强对不同模型性能的提升作用,本文选取了两种具有代表性的rPPG估计模型作为实验基准:经典的端到端3D卷积神经网络PhysNet^[16]和采用双分支解耦架构的MSFNet^[36]。PhysNet侧重于利用3D卷积捕捉基础的时空波动特征,而MSFNet则通过独立的脉搏与噪声分支、全局-局部交互设计以及时频域联合约束机制,展现出更强的信号解耦与抗运动干扰能力。

通过在这两种不同设计理念的模型上进行测试,可以更全面地验证合成数据集的泛化有效性。所有算法均基于PyTorch框架实现,输入视频尺寸统一调整为 $192 \times 128 \times 150$ 帧。实验在配备NVIDIA GeForce RTX 4090 GPU的高性能工作站上运行,采用Adam优化器,初始学习率设定为 1×10^{-4} ,并利用滑动窗口机制对预测结果进行分段评估。

2.4 实验分析

为了验证本文提出的合成数据增强策略对神经网络性能的提升作用,在PURE和UBFC-rPPG两个公开基准数据集上分别进行了对比实验。实验对比了仅使用传统数据增强与引入本文合成数据和只使用合成数据后,PhysNet与MSFNet两种模型在心率估计精度及信号质量上的表现。

表1对比了在PURE数据集上不同训练策略的性能。对于PhysNet模型,在加入合成数据训练后,平均绝对误差MAE从1.16 bpm显著降低至0.99 bpm,同时信噪比也有提升。对于采用双分支架构的MSFNet,性能同样得到了稳步提升,MAE从1.24 bpm优化至1.18 bpm。

表1 PURE数据集上的结果

Table 1 Result on PURE

方法	MAE/ bpm	RMSE/ bpm	r	SNR/dB
PhysNet/Real&Synth	0.99	1.26	0.95	8.36
PhysNet/Real	1.16	1.63	0.93	8.17
PhysNet/Synth	1.27	2.01	0.86	6.43
MSFNet/Real&Synth	1.18	1.74	0.94	5.27
MSFNet/Real	1.24	1.92	0.90	4.92
MSFNet/Synth	1.66	3.94	0.88	4.71

表2对比了各模型在UBFC-rPPG数据集上的性能。从表2可以观察到,在引入合成数据进行联合训练后,两种模型的各项指标均得到了显著提升。对于PhysNet模型,MAE从1.11 bpm降低至0.78 bpm,降幅约30%,同时Pearson相关系数也从0.76提升至0.84,表明模型在追踪心率真实波动趋势上的能力得到了提升。对于更复杂的MSFNet模型,各项指标都得到了显著的提升。

表2 UBFC数据集上的结果

Table 2 Result on UBFC

方法	MAE/bpm	RMSE/bpm	r	SNR/dB
PhysNet/Real&Synth	0.78	1.10	0.84	5.99
PhysNet/Real	1.11	1.50	0.76	5.50
PhysNet/Synth	0.85	1.07	0.85	5.70
MSFNet/Real&Synth	0.20	0.48	0.91	3.09
MSFNet/Real	2.57	6.98	0.80	2.82
MSFNet/Synth	0.33	0.59	0.89	2.24

本文提出的合成数据增强策略在 PURE 和 UBFC-rPPG 两个主流基准数据集上均表现出显著的性能增益。无论是基础的 PhysNet 还是双分支架构的 MSFNet, 在引入合成数据进行联合训练后, MAE、RMSE、 r 以及 SNR 均得到了全方位的优化。这充分验证了利用 LivePortrait 模拟宏观非线性运动与基于生物物理模型植入物理确定性信号的有效性, 证明该方法能够显著增强神经网络在复杂运动干扰下的特征解耦能力。

3 结 论

本文针对 rPPG 任务中训练数据稀缺及运动干扰敏感的问题, 提出了一种结合生成式人脸动画与生物物理嵌入的增强型合成视频生成框架。该框架打破了传统数据增强方法的局限, 采用“先运动建模, 后生理增强”的二阶段策略, 利用 LivePortrait 框架构建了具有高度真实感的动态面部基准。通过引入层分离技术与基于双色反射模型的乘性调制机制, 本文成功将 BIDMC 数据库中的高质量生理信号以符合物理光学规律的方式植入视频序列中。在 PURE 和 UBFC-rPPG 数据集上的实验结果一致表明, 引入合成增强数据能够显著提升不同架构神经网络在复杂运动场景下的心率估计精度及信号追踪能力, 有效证明了合成数据在弥合虚实性能鸿沟方面的巨大潜力。尽管本研究取得了积极成果, 但仍存在一定的局限性。目前生成的合成视频规模相较于深度学习所需的理想数据量仍有进一步扩充的空间, 且受试者的面部特征分布仍待细化。在未来的工作中, 将致力于构建更大规模、涵盖更多极端环境变量的合成数据集。同时, 将考虑开展跨数据集的迁移训练与验证研究, 旨在利用大规模高保真合成数据进一步探索 rPPG 模型在异构场景下的普适性, 为实现更稳健、通用的非接触式生理监测系统提供支撑。

参考文献

[1] 牛雪松, 韩璇, 山世光. 基于 rPPG 的生理指标测量方法综述 [J]. 中国图象图形学报, 2020, 25(11): 2321-2336.
Niu Xuesong, Han Hu, Shan Shiguang. Review of physiological measurement methods based on rPPG [J]. Journal of Image and Graphics, 2020, 25(11): 2321-2336.

[2] Wang Wenjin, Den Brinker A C, Stuijk S, et al. Algorithmic principles of remote PPG [J]. IEEE Transactions on Biomedical Engineering, 2016, 64(7): 1479-1491.

[3] 陈嘉涛, 张泓凯, 黄燕平, 等. 基于视频的生理参数测量技术及研究进展 [J]. 计算机工程与应用, 2022,

58(6): 1-13.
Chen Jiatao, Zhang Hongkai, Huang Yanping, et al. Video-based physiological parameter measurement technology and research progress [J]. Computer Engineering and Applications, 2022, 58(6): 1-13.

[4] 凌振宝, 张铭, 熊文激, 等. 基于脉搏波传导时间的无袖带血压测量仪设计 [J]. 电子测量与仪器学报, 2012, 26(12): 1080-1085.
Ling Zhenbao, Zhang Ming, Xiong Wenji, et al. Design of cuffless blood pressure measurement instrument based on pulse wave transit time [J]. Journal of Electronic Measurement and Instrumentation, 2012, 26(12): 1080-1085.

[5] 淳新益, 郑秀娟, 张畅, 等. 遗传算法优化 BP 神经网络的非接触式血压估计方法 [J]. 电子测量与仪器学报, 2023, 35(7): 53-59.
Chun Xinyi, Zheng Xiujuan, Zhang Chang, et al. Non-contact blood pressure estimation method based on BP neural network optimized by genetic algorithm [J]. Journal of Electronic Measurement and Instrumentation, 2023, 35(7): 53-59.

[6] 张瑞, 朱天军, 邹志亮, 等. 驾驶员疲劳驾驶检测方法研究综述 [J]. 计算机工程与应用, 2022, 58(21): 53-66.
Zhang Rui, Zhu Tianjun, Zou Zhiliang, et al. Review of driver fatigue driving detection methods [J]. Computer Engineering and Applications, 2022, 58(21): 53-66.

[7] Huang Powei, Wu B J, Wu Bingfei. A heart rate monitoring framework for real-world drivers using remote photoplethysmography [J]. IEEE Journal of Biomedical and Health Informatics, 2020, 25(5): 1397-1408.

[8] Liu Boxiang, Zheng Xiujuan, Wu Y I. Remote heart rate estimation in intense interference scenarios: A white-box framework [J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-14.

[9] Poh M Z, McDuff D J, Picard R W. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation [J]. Optics Express, 2010, 18(10): 10762-10774.

[10] Lewandowska M, Ruminski J, Kocejko T, et al. Measuring pulse rate with a webcam: A non-contact method for evaluating cardiac activity [C]//Federated Conference on Computer Science and Information Systems (FedCSIS). IEEE, 2011: 405-410.

[11] De Haan G, Jeanne V. Robust pulse rate from chrominance-based rPPG [J]. IEEE Transactions on Biomedical Engineering, 2013, 60(10): 2878-2886.

[12] De Haan G, Van Leest A. Improved motion robustness of

- remote-PPG by using the blood volume pulse signature [J]. *Physiological Measurement*, 2014, 35(9): 1913-1926.
- [13] 赵昶辰, 梅培义, 冯远静. 面向远程光体积描记的人脸检测与跟踪 [J]. *中国图象图形学报*, 2020, 25(11): 2439-2450.
- Zhao Changchen, Mei Peiyi, Feng Yuanjing. Face detection and tracking algorithm for remote photoplethysmography [J]. *Journal of Image and Graphics*, 2020, 25(11): 2439-2450.
- [14] Zhao Changchen, Cao Pengcheng, Hu Meng, et al. WTC3D: An efficient neural network for noncontact pulse acquisition in internet of medical things [J]. *IEEE Transactions on Industrial Informatics*, 2024, 21(2): 1547-1556.
- [15] 李江鹏. 基于 CNN-rPPG 的生命体征监测研究 [J]. *中国医疗设备*, 2026, 41(3): 49-54.
- Li Jiangpeng. Research on vital signs monitoring based on CNN-rPPG [J]. *China Medical Devices*, 2026, 41(3): 49-54.
- [16] Yu Zitong, Li Xiaobai, Zhao Guoying. Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks [PP/OL]. V2. (2019-07-31) [2024-04-30]. <https://doi.org/10.48550/arXiv.1905.02419>.
- [17] Niu Xuesong, Han Hu, Shan Shiguang, et al. VIPL-HR: A multi-modal database for pulse estimation from less-constrained face video [C]//*Asian Conference on Computer Vision*. Springer, 2018: 562-576.
- [18] Liu Xin, Fromm J, Patel S, et al. Multi-task temporal shift attention networks for on-device contactless vitals measurement [J]. *Advances in Neural Information Processing Systems*, 2020, 33: 19400-19411.
- [19] Stricker R, Müller S, Gross H M. Non-contact video-based pulse rate measurement on a mobile service robot [C]//*The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2014: 1056-1062.
- [20] Bobbia S, Macwan R, Benezeth Y, et al. Unsupervised skin tissue segmentation for remote photoplethysmography [J]. *Pattern Recognition Letters*, 2019, 124: 82-90.
- [21] 赵昶辰, 居峰, 冯远静. 多视角神经网络非接触式脉搏信号提取 [J]. *中国图象图形学报*, 2020, 25(11): 2428-2438.
- Zhao Changchen, Ju Feng, Feng Yuanjing. Non-contact pulse signal extraction using multi-view neural network [J]. *Journal of Image and Graphics*, 2020, 25(11): 2428-2438.
- [22] 张森乐. 基于卷积神经网络的非接触式视频心率测量方法研究 [D]. 合肥: 合肥工业大学, 2020: 9-10.
- Zhang Senle. Research on non-contact video heart rate measurement method based on convolutional neural network [D]. Hefei: Hefei University of Technology, 2020: 9-10.
- [23] McDuff D, Hernandez J, Wood E, et al. Advancing non-contact vital sign measurement using synthetic avatars [PP/OL]. V1. (2020-10-24) [2024-04-30]. <https://doi.org/10.48550/arXiv.2010.12949>.
- [24] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets [C]//*Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2014, 2: 2672-2680.
- [25] Guo Jianzhu, Zhang Dingyun, Liu Xiaoqiang, et al. Liveportrait: Efficient portrait animation with stitching and retargeting control [PP/OL]. V2. (2025-02-28) [2024-04-30]. <https://doi.org/10.48550/arXiv.2407.03168>.
- [26] Kartynnik Y, Ablavatski A, Grishchenko I, et al. Real-time facial surface geometry from monocular video on mobile GPUs [PP/OL]. V1. (2019-07-15) [2024-04-30]. <https://doi.org/10.48550/arXiv.1907.06724>.
- [27] McDuff D. Camera measurement of physiological vital signs [J]. *ACM Computing Surveys*, 2023, 55(9): 1-40.
- [28] Verkrusse W, Svaasand L O, Nelson J S. Remote plethysmographic imaging using ambient light [J]. *Optics Express*, 2008, 16(26): 21434-21445.
- [29] Chai D, Ngan K N. Face segmentation using skin-color map in videophone applications [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2002, 9(4): 551-564.
- [30] Chen Weixuan, McDuff D. DeepPhys: Video-based physiological measurement using convolutional attention networks [C]//*Computer Vision - ECCV 2018*. Springer, 2018, 11206: 356-373.
- [31] Shafer S A. Using color to separate reflection components [J]. *Color Research & Application*, 1985, 10(4): 210-218.
- [32] Lin Xueming. Using blood volume pulse vector to extract rPPG signal in infrared spectrum [D]. Eindhoven: Eindhoven University of Technology, 2014.
- [33] Meredith D J, Clifton D, Charlton P, et al. Photoplethysmographic derivation of respiratory rate: A review of relevant physiology [J]. *Journal of Medical Engineering & Technology*, 2012, 36(1): 1-7.
- [34] Jensen H W, Marschner S R, Levoy M, et al. A practical model for subsurface light transport [C]//*Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, 2001: 511-518.
- [35] Pimentel M A, Johnson A E, Charlton P H, et al.

Toward a robust estimation of respiratory rate from pulse oximeters [J]. *IEEE Transactions on Biomedical Engineering*, 2016, 64(8): 1914-1923.

- [36] Zhao Changchen, Wang Hongsheng, Chen Huiling, et al. JAMSNet: A remote pulse extraction network based on joint attention and multi-scale fusion [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 33(6): 2783-2797.

作者简介



李沂杭, 2023 年于河北科技大学获得学士学位, 现为杭州电子科技大学硕士研究生, 主要研究方向为视觉非接触式生理信号监测、智慧医疗。

E-mail: 242050134@hdu.edu.cn

Li Yihang received his B. Sc. degree from Hebei University of Science and Technology in 2023. He is currently a M. Sc. candidate at Hangzhou Dianzi University. His main research interests include visual non-contact physiological signal monitoring and smart healthcare.



程世超, 2013 年于河南师范大学获得学士学位, 2019 年于大连理工大学获得博士学位, 现为杭州电子科技大学计算机学院副教授, 主要研究方向为图像处理、可学习优化方法、计算机视觉、脑机接口。

E-mail: sccheng@hdu.edu.cn

Cheng Shichao received her B. Sc. degree from Henan Normal University in 2013, and Ph. D. degree from Dalian University of Technology in 2019. She is currently an associate professor with the School of Computer Science, Hangzhou Dianzi University. Her main research interests include image processing, learnable optimization methods, computer vision,

and brain-computer interfaces.



慧医疗。

E-mail: cczhao@hdu.edu.cn

Zhao Changchen (Corresponding author) received his B. Sc. degree from Shandong University of Science and Technology in 2011, and Ph. D. degrees from National Chung Hsing University and Beihang University in 2018. He is currently an associate professor with the School of Computer Science, Hangzhou Dianzi University. His main research interests include visual contactless physiological signal monitoring, cardiovascular disease risk prediction, lower-limb rehabilitation robots, and smart healthcare.



张建设, 1997 年于济南大学获得学士学位, 2004 年于浙江大学获得硕士学位, 2008 年于浙江大学获得博士学位, 现为杭州电子科技大学计算机学院教授, 主要研究方向为人工智能、类脑计算及脑-机接口技术。

E-mail: jhzhzhang@hdu.edu.cn

Zhang Jianhai received his B. Sc. degree from University of Jinan, in 1997, M. Sc. degree from Zhejiang University in 2004, and Ph. D. degree from Zhejiang University in 2008. He is currently a professor with the School of Computer Science, Hangzhou Dianzi University. His main research interests include artificial intelligence, brain-inspired computing, and brain-computer interface technology.