

DOI:10.3969/j.issn.1005-202X.2025.04.009

医学信号处理与医学仪器

基于改进三维卷积网络的非接触式生理参数检测方法

徐展宇, 陈兆学

上海理工大学健康科学与工程学院, 上海 200093

【摘要】远程光电容积描记法(rPPG)是从面部视频中测量心率等生理参数的方法,针对现有的心率测量方法难以同时兼顾高准确率和轻量化的问题,提出一种改进的三维卷积网络模型实现基于面部视频的非接触式生理参数检测。在预处理时,使用YuNet模型替代传统人脸检测器,从而快速且精确地识别人脸区域。此外,将注意力机制和残差模块嵌入到三维卷积网络中提取通道和空间的关键特征,并使用长短期记忆网络作为时期记忆模块捕捉数据中的长期依赖关系。实验结果表明,所提出Res-CHATM模型在公开数据集UBFC-rPPG和PURE进行心率评估交叉实验时分别取得MAE=2.19 BPM, RMSE=7.02 BPM, C=0.95以及MAE=1.65 BPM, RMSE=3.44 BPM, C=0.98的优异效果,进一步验证了模型预测值与真实值之间的一致性以及融合模块的有效性,展示了高效轻量化模型在rPPG技术中的潜力。

【关键词】非接触式;心率检测;混合注意力机制;信号处理

【中图分类号】R318; TP391

【文献标志码】A

【文章编号】1005-202X(2025)04-0479-10

Non-contact physiological parameter detection method based on improved three-dimensional convolution network

XU Zhanyu, CHEN Zhaoxue

School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

Abstract: Remote photoplethysmography is a method of measuring physiological parameters such as heart rate from facial video. For overcoming the difficulties in achieving both high accuracy and lightweight by the existing heart rate measurement methods, an improved three-dimensional convolution network model is proposed to realize non-contact physiological parameter detection in facial video. In the pre-processing, YuNet model takes place of the traditional face detector, so that the face region can be recognized quickly and accurately. In addition, attention mechanisms and residual modules are embed into three-dimensional convolution network to extract key channel and spatial features, with long short-term memory networks used as period memory modules to capture long-term dependencies in the data. The experimental results show that the proposed Res-CHATM model achieves excellent results of MAE=2.19 BPM, RMSE=7.02 BPM, C=0.95, and MAE=1.65 BPM, RMSE=3.44 BPM, C=0.98 in the cross experiments on public datasets UBFC-rPPG and PURE for heart rate estimation. The consistency between the predicted value and the real value and the effectiveness of the fusion module are further verified, demonstrating the potential of efficient lightweight model in remote photoplethysmography.

Keywords: non-contact type; heart rate detection; hybrid attention mechanism; signal processing

前言

在医学领域,人体的健康状况与多种生理指标参数密切相关,如心率(Heart Rate, HR)、呼吸频率(Respiration Frequency, RF)、心率变异性(Heart Rate

Variability, HRV)等,针对这些人体生理参数的检测技术已成为现代医学的重要诊断手段。HR与心脏功能密切相关,是评估心血管健康的重要指标,也是反映整体健康状况的重要依据。据2022年《中国心血管健康与疾病报告》显示,心血管病的发病率呈逐年上升趋势,预计目前患者数量已达3.3亿^[1]。而且心血管病是我国居民的主要死亡原因,致死率约为40%。RF是评估呼吸状况的重要医学指标,研究表明,许多呼吸功能障碍性疾病在早期都会表现出RF异常现象,如呼吸暂停综合征,这是一种由心肺综合功能异常导致的病症,患者在睡眠中可能出现呼吸困难,进而导致心率异常,严重时甚至会导致死亡^[2]。

【收稿日期】2024-11-05

【基金项目】国家中医药管理局中医药创新团队及人才支持计划(ZYYCXTD-D-202208)

【作者简介】徐展宇,硕士研究生,研究方向:图像处理、深度学习, E-mail: xuzhanyu@foxmail.com

【通信作者】陈兆学,博士,副教授,研究方向:医学图像处理,医学信号处理, E-mail: chenzhaoxue@163.com

HRV 常用于评估心血管自主神经功能,是预测心肌梗死、慢性心力衰竭、心率失常、高血压、糖尿病等与自主神经功能相关疾病的重要指标^[3],也可用于精神分裂症、抑郁症等精神疾病的辅助诊断和评估^[4]。

目前,有多种方法可以测量这些生理参数,其中最主要的两种方法分别是光电容积描记法(Photoplethysmography, PPG)和远程光电容积描记法(remote Photoplethysmography, rPPG)^[5-6]。PPG 技术由于其检测成本低且操作简单,得到广泛应用。但其缺点是需要与皮肤接触,且对运动干扰的抵抗性较差,使得适用场景严重受限。在此背景下,rPPG 作为一种新兴的非接触式检测技术,通过相机捕捉面部细微颜色变化,以提取 HR、RF、HRV 等生理参数。由于其低成本、高安全性、连续测量且不受物理距离限制等优势,rPPG 技术在远程医疗、情绪分析和面部安全等领域展现出巨大的应用潜力^[7-9]。因此利用 rPPG 技术实现一种轻量化、低成本且精准的生理参数测量方法,对推动现代医学诊疗水平的提升具有十分重要的意义。

目前,国内外众多学者从人脸视频中重建 rPPG 信号以进行心率测量的技术方面已取得了显著进展。2008年,Verkruysse 等^[10]根据血红蛋白更易吸收绿光,发现 G 通道比 R 和 B 通道包含更强的血液容积脉搏波(Blood Volume Pulse, BVP)信号,首次证明了 rPPG 技术的可行性。2010年,Poh 等^[11]提出独立成分分析法(ICA),将 RGB 信号分解成 3 个独立分量,然后通过分析频率谱,从中选择出最合适的分量作为 BVP 信号。2013年,de Haan 等^[12]提出基于肤色假设的方法(CHROM),验证了该方法在处理不同肤色类型和运动伪影时的鲁棒性。2015年,Wang 等^[13]提出一种基于图像皮肤像素分布的新方法(Spatial Subspace Rotation, 2SR),通过测量皮肤像素子空间的时间旋转提取脉搏波信号。2016年,Wang 等^[14]提出更符合皮肤光学特性的模型,同时提出一种可以替代 CHROM 的方法:Plane-Orthogonal-to-Skin (POS)算法。这些方法主要是在早期开发的,依赖于人工程序,不需要训练数据集,因此被称为“传统方法”。近年来,随着深度学习技术的不断发展,在 rPPG 测量方面得到广泛应用。2018年,Chen 等^[15]提出了首个端对端的深度卷积模型 DeepPhys,用于预测视频心率。Špetlík 等^[16]提出两步式卷积模型 HR-CNN,先从面部图像序列提取 PPG 信号,然后通过 PPG 信号预测心率。相较于二维卷积网络(2DCNN),三维卷积网络(3DCNN)捕获 rPPG 信号中的时间信息能力更强,Yu 等^[17]在 2019 年提出了首个端到端的深度时空网络模型 PhysNet,将时间上下

文特征考虑在内,从而在面部视频中预测得到精确的 rPPG 信号。2020年, Lee 等^[18]提出转导元学习模型:Meta-rPPG,该模型在公开数据集上的验证显示其鲁棒性更强。Liu 等^[19]提出多任务卷积注意力网络 TS-CAN,该模型具有延时低、精度高等优点,对其它领域的研究具有借鉴意义。2021年,Hu 等^[20]提出时域注意力网络模型 ETA-rPPGNet,解决视频冗余问题,同时设计时域注意模块增强时间信息的相关性。Gideon 等^[21]提出自监督网络模型 Gideon2021,该模型通过稀疏注意力模块结合多尺度时间学习来挖掘弱信号中的准周期模式,在跨数据集评估中表现出出色的泛化能力。虽然这种方法具有开创性,但与最先进的监督方法相比,其准确性较低。2022年, Liu 等^[22]提出两种分别以 CNN 和 Transformer 为基础的高效神经模型 EfficientPhys,该模型不仅消除了对预处理步骤的需要,而且在公开数据集上表现优异。Yu 等^[23]提出了视频 Transformer 模型 PhysFormer,能自适应聚合时空特征,增强 rPPG 信息。虽然 Transformer 在捕获远程空间和时间相关性方面表现出色,但弊端同样明显,如可能忽略预定义感兴趣区域(Region Of Interest, ROI)之外的全局上下文线索,或倾向于以粗粒度提取特征,亦或忽视长期时间关系。2023年, Speth 等^[24]提出非对比无监督模型 SiNC,该模型即使在缺乏真值信号的情况下也能进行多种生理参数测量。但这种无监督方法没有考虑利用部分或低质量标签来进一步改善 rPPG 信号质量。

为解决现有心率测量模型难以同时实现高准确性与轻量化的问题,本文提出一种改进三维卷积网络模型,实现高效轻量化的生理参数测量,本文主要进行了如下工作:(1)使用 YuNet 模型作为人脸检测器,以提升检测精度和运行速度^[25]。(2)将混合注意力模块(Hybrid Attention Module, HAM)改进为 3D-HAM,并嵌入三维卷积网络中,增强了模型在处理通道和空间特征时的性能,使其关注包含更强心率信号的区域^[26];同时使用长短期记忆网络(Long Short-Term Memory network, LSTM)作为时期记忆层(Term Memory, TM)以捕获数据中的长期依赖关系^[27],从而显著提升了模型的整体性能。(3)在公开数据集上进行评估实验,验证了本文模型在准确性和泛化能力方面的优越性,并进行了可视化与一致性分析,证明了本文模型预测值与真实测量值之间的较强一致性。此外,通过消融实验和模型计算成本实验,进一步验证了融合模块的有效性和模型的轻量化特点。

1 理论与方法

本文方法的工作流程如图1所示,完整流程可分为预处理模块、主干网络模块以及后处理模块3个模块。此流程图详细说明了从面部视频实现生理参数测量的完整过程,首先,原始视频数据经过预处理模块,生成包含 n 段、每段 T 帧的 ROI 人脸图像序列。随后,这些图像序列与标签数据被输入到主干网络,输出 n 段 T 帧长度的 rPPG 信号。最后,这些信号被送入后处理模块,通过分析频谱,预测出 HR、RF 或 HRV。

1.1 预处理模块

如图1所示,预处理模块首先对视频数据进行分帧,然后选择合适的颜色空间获取图像序列,使用人脸检测器检测人脸并选择合适的 ROI。此外,该模块还包括预处理方法、ROI 图像尺寸和视频帧数 T 的设置。这一系列操作完成后,将得到形状为 $[3, T, W, H]$ 的4维人脸图像数据。预处理方法包括3种:(1)原始图像:仅对输入视频进行人脸检测,不对图像数据进行其他修改。(2)标准化:对检测到的人脸图像进行标准化数据分布,减少异常值的影响。(3)帧差法:通过分析视频连续帧之间的差异来预处理数据。本文还对不同人脸检测器在单帧人脸图像3种情况下的检测效果及其所需时间进行了对比,详见表1。

结果显示, MTCNN、Dlib-hog 均能有效检测到正脸和歪头,但在侧脸识别方面存在局限。Opencv-haar 表现最差,只能检测到正脸。Opencv-dnn 无法检测歪头情况,而 Dlib-cnn 和 YuNet 则在所有测试情况下均显示良好的检测性能。YuNet 在检测速度上显著优于其他模型,仅需 0.05 s。因此,鉴于其优越的识别精度和高效的检测速度,本文选择 YuNet 作为人脸检测器。

1.2 后处理模块

如图1所示,后处理模块的工作流程如下:模型

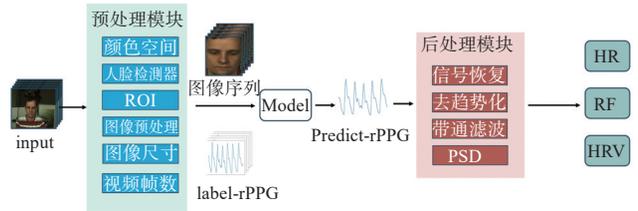


图1 面部视频测量生理参数流程

Figure 1 Facial video measurement of physiological parameters

表1 检测效果对比

Table 1 Comparison of detection performance

检测器	正脸	歪头	侧脸	t/s
MTCNN	√	√	×	0.65
Dlib-hog	√	√	×	0.02
Dlib-cnn	√	√	√	2.68
Opencv-haar	√	×	×	0.24
Opencv-dnn	√	×	√	0.04
YuNet	√	√	√	0.05

√表示检测到人脸,×表示未能

输出的 Predict-rPPG 信号首先经过差分信号恢复、去趋势化以及带通滤波处理。随后,通过对处理后的信号进行功率谱密度(PSD)分析,计算并获得对应的生理参数。其中 HR 频段设置为 0.75~2.5 Hz, RF 频段设置为 0.1~0.5 Hz, HRV 的低频功率(Low Frequency, LF)设置为 0.04~0.15 Hz, 高频功率(High Frequency, HF)设置为 0.15~0.4 Hz。

1.3 主干网络设计

本文所设计的主干网络完整结构如图2所示,主要由以下几部分组成:卷积层 Conv3D、最大池化层 MaxPool3D、注意力层 HAM_Block、残差卷积块 Res-Conv、上采样层 Unsample3D、全局平均池化层 GAP、时期记忆层 TM 和聚合层 Poly。

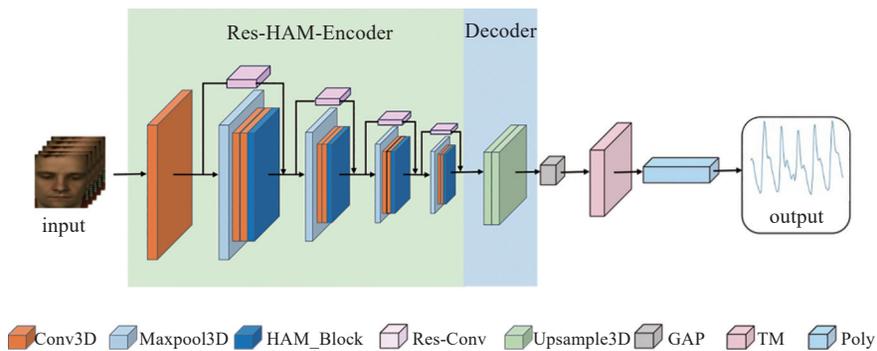


图2 改进的三维卷积网络

Figure 2 Improved three-dimensional convolution network

由图2可见,本文的模型是基于编码器-解码器模型进行构建和优化的,其简化后的模型结构如图3所示。首先,将预处理后的四维数据 $[3, T, W, H]$ 输入模型,并通过改进的注意力编码器(Res-HAM-Encoder)处理数据。在此阶段,模型提取数据的时空特征,并生成维度为 $[64, T/4, W/8, H/8]$ 的特征图。接着,通过

解码器(Decoder)执行上采样操作,将特征图的时间维度恢复至原始输入帧数 $[64, T, W/8, H/8]$ 。随后,利用GAP层将特征图的空间维度压缩至 $[64, T, 1, 1]$ 。而后通过TM层进一步捕获数据中的长期依赖关系。此特征图接着经过Poly层映射为 $[1, T, 1, 1]$ 。最后,通过降低维度,得到长度为 $[1, T]$ 的rPPG信号。

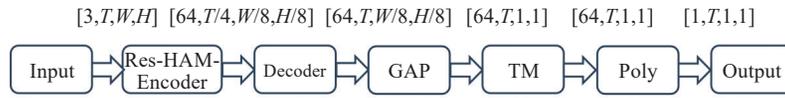


图3 主干网络简化结构

Figure 3 Simplified structure of backbone network

1.3.1 注意力编码器 改进前后的注意力编码器结构如图4所示,图4a为以三维卷积层为核心的原始编码器(3D-Encoder),图4b为融合注意力与残差结构后的残差注意力编码器(Res-HAM-Encoder)。由图4b可知Res-HAM-Encoder主要由注意力层HAM-Block、卷积层Conv3D、最大池化层MaxPool3D和残差卷积

块Res-Conv组成,通过这种融合,模型在执行图像卷积处理过程中,能够更有效地评估并利用通道与空间信息的重要程度,从而增强重要特征,抑制背景或无关信息,并集中注意力于那些包含更强生理信号的区域。

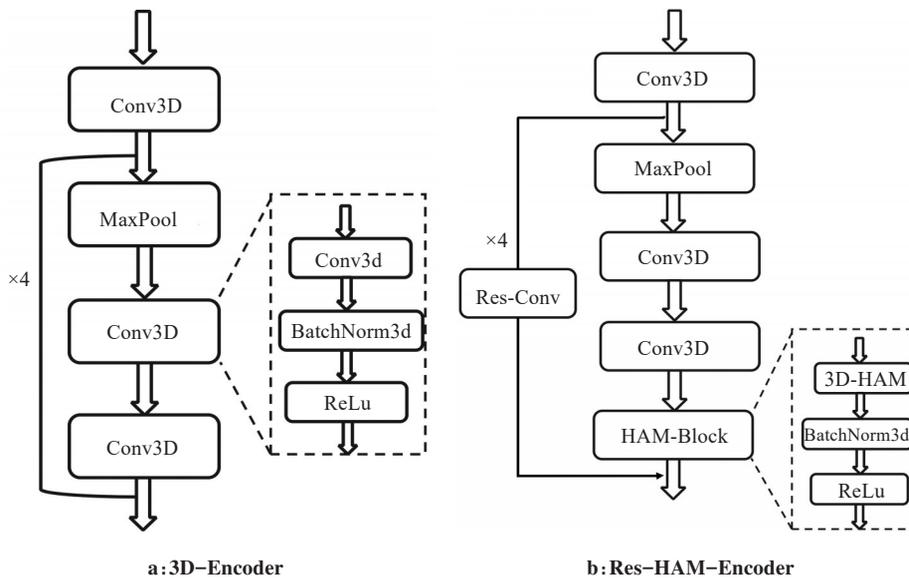


图4 编码器改进对比

Figure 4 Encoder improvement

注意力层中最重要的模块是3D-HAM模块。HAM是Li等^[26]在2022年提出的一个轻量、高效的混合注意力模块,旨在平衡模型的性能和复杂度。但HAM是基于2DCNN的,因此对该模块改进并融合到3DCNN中,通过在每次HAM计算时增加一个维度:深度 D (即视频帧数 T),构建3D-HAM模块,如图5a所示。首先,将3D-HAM融入到残差模块,输入特征 $F_{in} \in R^{C \times D \times W \times H}$ 经过通道注意力模块(Channel Attention Module, CAM)得到通道注意张量 A_c , A_c 与

F_{in} 相乘得到通道细化特征 F' 。然后用通道分离技术将 F' 切分为两组不同的特征:重要特征 F'_1 和次要特征 F'_2 ,接着两组特征分别通过空间注意力模块(Spatial Attention Module, SAM),从输入特征 F'_1 和 F'_2 获得空间注意张量 A_{s1} 和 A_{s2} ,将 A_{s1} 和 A_{s2} 分别与 F'_1 和 F'_2 相乘后获得对应空间细化特征 F''_1 和 F''_2 ,融合获得最终的细化特征 F'' ,最后与原始输入特征相加获得输出特征 $F_{out} \in R^{C \times D \times W \times H}$,也就是原输入特征 F_{in} 经过3D-HAM模块处理后,所获得的最终结果。其中

⊗表示矩阵相乘, ⊕表示矩阵相加。

CAM在图像处理中主要用于聚焦重要的信息,即确定图像中哪些部分是重要的,如图5b所示。通过平均池化和最大池化操作聚合空间特征,生成不同特征张量 F_c^{avg} 和 F_c^{max} ,然后将它们送入自适应机制(如图5b中 Adaptive mechanism Block),获得特征张量 $F_c^{add} \in R^{C \times D \times 1 \times 1}$,该张量不仅保留了原始特征信息并自适应的调整特征倾向,而且丰富了特征表示。此外使用一维卷积捕捉通道间的联系,有效避免通道数据的降维。通道注意张量 A_c 与输入特征 F_{in} 相乘得到通道细化特征 F' 。该模块的总过程如式(1)、(2)所示:

$$F_c^{add} = \left(\frac{1}{2} (F_c^{avg} \oplus F_c^{max}) \oplus \alpha \otimes F_c^{avg} \oplus \beta \otimes F_c^{max} \right) \quad (1)$$

$$A_c(F_{in}) = \sigma(C1D_{1 \times k}(F_c^{add})) \quad (2)$$

其中, σ 表示 Sigmoid 函数, $C1D_{1 \times k}$ 表示核尺寸为 k 的一维卷积。

SAM旨在增强模型对图像中不同空间区域的关

注度。主要集中于识别在图像中哪些区域是重要的。Li等提出带有通道分离技术的SAM,如图5c所示。该技术引入超参数 λ 与通道细化特征 F' 的通道维度相乘,从而得到重要特征组的通道维度。定义两个形状与通道注意力特征一致的掩码数组,一种是重要的通道置1,而次要的通道置0。另一种则相反,即重要的通道置0,次要的通道置1,称为重要掩码和次要掩码。将它们分别与 F' 相乘,得到重要特征 F'_1 和次要重要特征 F'_2 。对 F'_1 和 F'_2 进行两种池化操作分别得到两组特征张量 $F_{S1,2}^{max}$ 和 $F_{S1,2}^{avg}$,将结果拼接得到两组输出特征,接着使用二维卷积获得空间注意张量 A_{S1} 和 A_{S2} ,将 A_{S1} 和 A_{S2} 分别与 F'_1 和 F'_2 相乘后获得对应空间细化特征 F''_1 和 F''_2 ,逐元素相加获得最终的细化特征 $F'' \in R^{C \times D \times W \times H}$,该模块的总过程如式(3)所示:

$$A_{S1,2}(F') = \Phi \left(C2D_{7 \times 7} \left(\left[F_{S1,2}^{max}; F_{S1,2}^{avg} \right] \right) \right) \quad (3)$$

其中, Φ 表示先后使用 BatchNorm、ReLu 和 Sigmoid 函数, $C2D_{7 \times 7}$ 表示 7×7 的二维卷积核。

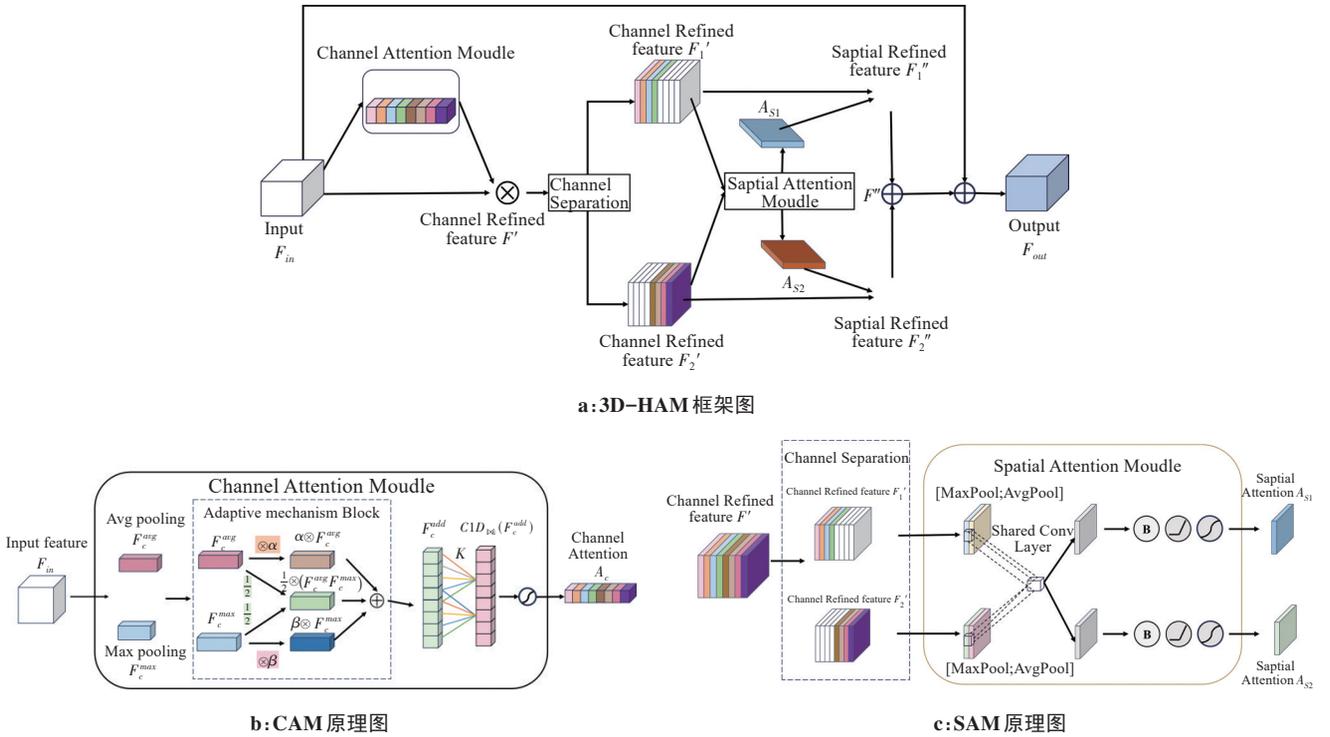


图5 3D-HAM框架与组件原理图

Figure 5 Framework and component schematics of 3D-HAM

1.3.2 解码器 解码器中包括两个上采样层,上采样层与卷积层(图4a)的结构类似,由三维转置卷积函数 ConvTranspose3d、批归一化函数 BatchNorm3d 和激活函数 ELU 组成。在上采样层中使用 ELU 作为激活函数可以更好地保留并恢复原始图像的细节和结构,使数据被压缩的时间维度恢复至原始时间长度,

同时卷积的空间和时间填充需要保持一致的大小。

1.3.3 全局平均池化层 该层由一个自适应平均池化函数 AdaptiveAvgPool3d 组成,对输入特征进行自适应调整以适应不同尺寸的输入,提高模型的泛化能力。

1.3.4 时期记忆层 由于 3DCNN 主要针对视频数据

提取短期时空特征,其在捕捉数据中的长时依赖关系方面存在限制。为了有效补充3DCNN在长期序列数据处理上的不足,本文引入了LSTM模型^[27]作为时期记忆层接收3DCNN处理后的PPG信号。通过其高效的信息管理机制提取并保留视频数据中的长期依赖关系。通过这种3DCNN与LSTM的结合,模型在实现精确心率预测的同时,还能有效处理视频数据在时间维度上的长距离依赖,显著提升了整体模型的性能。

1.3.5 聚合层 该层使用三维卷积函数Conv3d聚合多个特征图的信息,提高模型的性能和效率。模型在这一步输出特征经过降维后就得到与真实信号长度一致的预测rPPG信号。

2 实验设计

2.1 数据集

在UBFC-rPPG^[28]和PURE^[29]两个公开的数据集上进行实验与评估,它们都记录了各种现实世界条件的变化,如运动节奏、光照条件、肤色和背景等因素的变化,为信号处理和机器学习算法的研究带来了复杂的挑战。

UBFC-rPPG数据集记录了42个处于静止状态下的受试者视频和PPG信号。视频以30 FPS的帧率,640×480 pixels进行保存,使用CMS50E血氧仪以30 Hz的采样率记录受试者的PPG金标准真值。

PURE数据集共记录了10名受试者的头部图像序列和PPG信号,该数据集包含6种不同且受控的头部运动(01-稳定、02-说话、03-慢速平移、04-快速平移、05-小幅旋转、06-中幅旋转)。视频以30 FPS的帧率拍摄,保存为640×480 pixels的图像序列,使用CMS50E血氧仪以60 Hz的采样率记录受试者的PPG金标准真值。实验中对PPG真值进行降采样,使真实值与视频帧数相对应。

2.2 参数设置

本文实验在Windows操作系统上进行,GPU为RTX 4060,使用Pytorch框架版本为2.1.2以及python语言版本为3.10的开发模型,模型的初始学习率为9e-3,批次大小为4,使用AdamW优化器和OneCycleLR学习率调度器训练20个epoch。使用调度器可以加速模型的训练,并避免模型陷入局部最小值。在预处理模块以YuNet模型作为人脸检测器,以检测到的人脸作为ROI区域,将图像尺寸(W, H)统一调整为128×128的大小,并设置视频帧数 T 为160帧。因此模型的输入为 $[3, 160, 128, 128](C \times T \times H \times W)$ 。在重建rPPG信号的任务中,选择合适的损失函数是网络学习的关键,为了重建与真值PPG信号具备最

大趋势相似性和最小峰值定位误差的rPPG信号,本文选择NegPeaLoss函数作为损失函数,如式(4)所示:

$$\text{loss} = 1 - \frac{T \sum_{i=1}^T p_i g_{t_i} - \sum_{i=1}^T p_i \sum_{i=1}^T g_{t_i}}{\sqrt{\left(T \sum_{i=1}^T p_i^2 - \left(\sum_{i=1}^T p_i \right)^2 \right) \left(T \sum_{i=1}^T g_{t_i}^2 - \left(\sum_{i=1}^T g_{t_i} \right)^2 \right)}} \quad (4)$$

其中, p_i 和 g_{t_i} 分别表示第 i 帧的预测值与真实值。

2.3 评价指标

对于HR评估任务,本文遵循Li等^[30]和Niu等^[31-32]的研究,使用平均绝对误差(Mean Absolute Error, MAE)、均方根误差(Root Mean Square Error, RMSE)和Pearson相关因子(C)作为指标评估心率测量的准确性。对于HRV特征评估任务,包括RF、归一化单位的LF(n.u.)、HF(n.u.)以及LF/HF功率比,遵循文献[33]中概述的方法,采用标准偏差(Standard Deviation, STD)、RMSE和C作为评估指标。

MAE代表心率真实值与预测值的平均绝对误差,如式(5)所示;RMSE代表心率真实值与预测值的均方根误差,如式(6)所示;C代表心率真实值与预测值的Pearson相关因子,如式(7)所示:

$$\text{MAE} = \frac{1}{N} \sum_{n=1}^N |HR_{gt}^n - HR_p^n| \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^N (HR_{gt}^n - HR_p^n)^2} \quad (6)$$

$$C = \frac{\sum_{n=1}^N (HR_p^n - \overline{HR_p})(HR_{gt}^n - \overline{HR_{gt}})}{\sqrt{\left(\sum_{n=1}^N HR_p^n - \overline{HR_p} \right)^2 \left(\sum_{i=1}^N HR_{gt}^n - \overline{HR_{gt}} \right)^2}} \quad (7)$$

其中, N 表示数据集中的总样本数, HR_{gt}^n 和 HR_p^n 表示第 n 个样本的平均心率预测值与真实值。MAE、RMSE和STD数值越小表示误差越低,模型性能越优越;C数值越大,表示预测值与真实值之间的相关性越强,模型效果越佳。

3 结果分析

为了验证本文模型的性能,进行了以下实验:(1)将CHATM和Res-CHATM模型与已有研究方法的HR评估结果进行对比。(2)在交叉数据集上进行消融实验,以验证模型的有效性及其改进点。(3)对比不同模型在处理相同数据规模时的复杂度。(4)进行HRV特征评估实验。(5)验证模型的预测效果,并采用Bland-Altman散点图^[34]对整体预测效果进行一致性分析。

3.1 心率评估实验

表2列出了本文模型与传统方法及其他深度学

习模型的对比实验结果,所有模型处理的数据大小均为[3,160,128,128]。从表2的数据分析可知,深度学习模型,尤其是融合了Transformer的模型,如PhysFormer和EfficientPhys,相较于传统方法展现了更稳定的性能和更强的鲁棒性。在数据集内实验时,PhysFormer模型在UBFC数据集上性能最好,其MAE比Res-CHATM低0.59 BPM, SiNC模型在PURE数据集上性能最好,其MAE比Res-CHATM低0.63 BPM,而其余方法的表现均不如本文模型。

但在交叉实验中,Res-CHATM表现出更优越的性能和泛化能力,在UBFC-rPPG训练且在PURE

测试时(UBFC→PURE), MAE=2.19 BPM, RMSE=7.02 BPM, C=0.95;而在PURE训练UBFC-rPPG测试时(PURE→UBFC), MAE=1.65 BPM, RMSE=3.44 BPM, C=0.98。而PhysFormer模型在PURE→UBFC取得的最佳效果为MAE=2.35 BPM, RMSE=6.26 BPM, C=0.94; SiNC模型在的UBFC→PURE取得的最佳效果为MAE=4.02 BPM, RMSE=11.24 BPM, C=0.86,均低于本文模型的预测效果。分析表2发现,CHATM模型较Res-CHATM模型的性能表现仅相差0.02~0.19 BPM,因此下述实验对这两个模型进一步分析。

表2 HR评估结果对比

Table 2 Comparison of HR estimation results

模型	训练	测试					
		UBFC			PURE		
		MAE/BPM	RMSE/BPM	C	MAE/BPM	RMSE/BPM	C
ICA ^[11]	-	5.17	11.76	0.65	4.77	16.07	0.72
CHROM ^[12]	-	2.36	9.23	0.87	5.77	12.93	0.81
2SR ^[13]	-	6.90	18.50	0.65	2.44	3.06	0.98
POS ^[14]	-	2.11	9.11	0.87	3.67	11.82	0.88
DeepPhys ^[15]	UBFC	6.27	10.82	0.64	5.54	18.51	0.66
	PURE	2.61	9.02	0.88	0.83	1.54	0.99
PhysNet ^[17]	UBFC	1.78	4.39	0.97	4.49	11.70	0.80
	PURE	4.25	12.49	0.84	2.20	3.88	0.93
TS-CAN ^[19]	UBFC	1.70	2.72	0.99	3.69	13.80	0.82
	PURE	2.39	8.85	0.89	2.48	9.01	0.98
EfficientPhys ^[22]	UBFC	1.14	1.81	0.99	5.47	17.04	0.71
	PURE	2.07	6.32	0.94	1.11	5.94	0.97
PhysFormer ^[23]	UBFC	0.50	0.71	0.99	2.93	10.11	0.90
	PURE	2.35	6.26	0.94	1.10	1.75	0.99
SiNC ^[24]	UBFC	0.59	1.83	0.99	4.02	11.24	0.86
	PURE	4.88	12.61	0.69	0.61	1.84	0.99
CHATM	UBFC	1.11	2.68	0.99	2.38	7.90	0.94
	PURE	1.80	4.01	0.98	1.28	1.80	0.98
Res-CHATM	UBFC	1.09	2.67	0.99	2.19	7.02	0.95
	PURE	1.65	3.44	0.98	1.24	1.81	0.98

3.2 消融实验

为了验证本文模型的有效性及其改进点,本文在交叉数据集上进行详尽的消融实验,其结果汇总于表3。在PhysNet模型基础上分别融合TM、3D-HAM、Res-Conv,观察各组合对模型性能的影响。实验结果表明,当PhysNet分别融合TM和3D-HAM时,在交叉数据集的心率评估实验中,其表现均优于原始模型。特别地,PhysNet同时融合TM、3D-HAM(CHATM),其在MAE和RMSE上的误差较低,相关因子C值较高,分别达到0.94和0.98。融合Res-Conv(RES-CHATM)进一步提升了CHATM模型的性能,使其在心率评估任务中的表现更加出色。结果证实

了TM、3D-HAM和Res-Conv模块的有效性,能够显著提升模型在交叉数据集上的泛化能力和准确性。

3.3 模型成本计算

为验证模型的复杂度,本文对比分析了不同网络模型在处理大小为[3,160,128,128]的视频输入时的参数量(Param)和乘法累加运算(MACs),详细结果见表4。通过分析表3、表4数据可知,引入残差结构虽然提升了CHATM模型在心率预测任务中的效果,但也显著增加了Param和MACs。CHATM模型在性能和计算成本之间实现了较好的平衡,适用于对计算资源有要求但需要较高准确性的场景。而Res-CHATM模型在准确性上表现最佳,但同样较基

表3 消融实验
Table 3 Ablation study

模型	UBFC→PURE			PURE→UBFC		
	MAE/BPM	RMSE/BPM	C	MAE/BPM	RMSE/BPM	C
PhysNet	4.49	11.70	0.80	4.25	12.49	0.84
+TM	3.45	13.80	0.82	2.23	6.58	0.94
+3D-HAM	2.60	7.64	0.94	2.59	6.14	0.94
+TM+3D-HAM(CHATM)	2.38	7.90	0.94	1.80	4.01	0.98
+Res+CHATM(Res-CHATM)	2.19	7.02	0.95	1.65	3.44	0.98

表4 模型计算成本
Table 4 Model computational costs

模型	Param/M	MACs/G
DeepPhys ^[15]	7.504	120.000
PhysNet ^[17]	0.700	70.120
TS-CAN ^[19]	7.504	120.000
EfficientPhys ^[22]	7.439	60.687
PhysFormer ^[23]	7.318	50.607
CHATM	0.869	71.070
Res-CHATM	1.640	128.260

准模型有较大的计算资源优化。较低的计算成本不仅减少了对硬件资源的需求,还提高了模型的运行效率和响应速度,显示了其在高效、轻量化rPPG技术应用中的巨大潜力。

3.4 HRV 特征评估实验

使用CHATM模型在UBFC-rPPG数据集内进行

HRV 特征评估,结果如表5、表6所示。CHATM模型在RF评估任务中与ICA模型相比,STD降低了37.2%,且在LF、HF指标上表现出最高的准确性,与Gideon2021模型相比,STD均降低了36.6%,在LF/HF指标上,STD降低了57.8%,相关因子C提升了17.8%,反映出该模型在HRV特征提取中的精度和一致性明显优于传统方法和基准模型。

表5 RF评估结果对比
Table 5 Comparison of RF estimation results

模型	RF		
	STD/Hz	RMSE/Hz	C
ICA ^[11]	0.086	0.089	0.102
POS ^[14]	0.109	0.107	0.087
PhysNet ^[17]	0.148	0.152	0.081
Gideon2021 ^[21]	0.098	0.098	0.103
CHATM	0.054	0.058	0.722

表6 HRV特征评估结果对比
Table 6 Comparison of HRV characterization and its assessment results

模型	LF(n.u.)			HF(n.u.)			LF/HF		
	STD	RMSE	C	STD	RMSE	C	STD	RMSE	C
ICA	0.243	0.240	0.159	0.243	0.240	0.159	0.655	0.645	0.226
POS	0.171	0.169	0.479	0.171	0.169	0.479	0.405	0.399	0.518
PhysNet	0.149	0.169	0.519	0.149	0.169	0.519	0.511	0.616	0.581
Gideon2021	0.142	0.139	0.694	0.142	0.139	0.694	0.687	0.691	0.684
CHATM	0.090	0.090	0.812	0.090	0.090	0.812	0.290	0.346	0.806

3.5 可视化与一致性分析

如图6所示,展示了模型的部分预测信号与真实信号的可视化结果,以及整体预测效果的一致性分析。其中图6a和图6c为使用CHATM模型在UBFC→PURE时得到的rPPG信号对比和Bland-Altman散点图,图6b和图6d为PURE→UBFC时得到的rPPG信号对比和Bland-Altman散点图。由图6a、b可见,本文模型的重建rPPG信号在周期性和峰值点上与真实值之间存在较强的相关性。图6c中PURE数据集的HR差值为(0.83±7.85)BPM,95%一致性区间为[-14.56,16.22]BPM,仅2个样本点超出了一致

性界限。图6d中UBFC-rPPG数据集的HR差值为(-0.63±3.32)BPM,95%一致性区间为[-7.13,5.88]BPM,仅2个样本点超出了界限。由此可以看出,本文模型的预测值与真实测量值之间具有较强的一致性。

4 结论

本文提出了两种改进的三维卷积网络模型,旨在解决现有心率测量模型在轻量化和高精度之间的矛盾问题,以实现从面部视频中精确重建rPPG信号。其中Res-CHATM模型在公开数据集UBFC-rPPG和

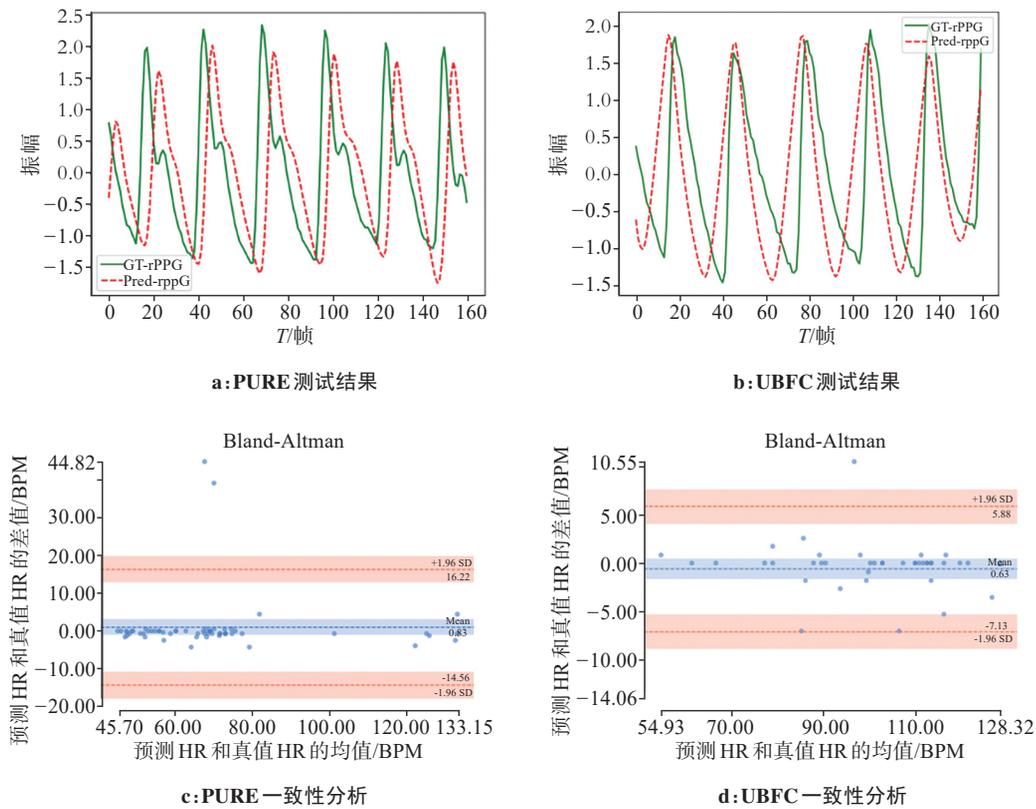


图6 心率预测效果验证

Figure 6 Verification of HR prediction performance

PURE 上进行了心率评估交叉实验得到最佳性能: MAE=2.19 BPM, RMSE=7.02 BPM, C=0.95 以及 MAE=1.65 BPM, RMSE=3.44 BPM, C=0.98。消耗计算资源较少的 CHATM 模型性能表现为: MAE=2.38 BPM, RMSE=7.90 BPM, C=0.94 以及 MAE=1.80 BPM, RMSE=4.01 BPM, C=0.98, Bland-Altman 一致性误差分别为 0.83 BPM 和 -0.63 BPM。本文模型能够重建较精准的 rPPG 信号,且具备良好的泛化能力。在 UBFC 数据集上进行 RF 和 HRV 特征评估实验同样取得良好的效果。目前,本文模型主要用于从面部视频中重建 rPPG 信号,以进行平均 HR、RF 和 HRV 特征分析,未来计划将这该模型扩展到实时应用上,以便在各种医疗和健康监测场景中提供更全面的生理参数分析,从而提升其实际应用价值。

【参考文献】

[1] 马丽媛,王增武,樊静,等.《中国心血管健康与疾病报告2022》要点解读[J].中国全科医学,2023,26(32):3975-3994.
Ma LY, Wang ZW, Fan J, et al. Interpretation of report on cardiovascular health and diseases in China 2022[J]. Chinese General Practice, 2023, 26(32): 3975-3994.
[2] 李金玉. 探讨对阻塞性睡眠呼吸暂停综合征(OSAS)患者生活方式的护理干预方法及效果[J]. 世界睡眠医学杂志, 2019, 6(8): 1148-1150.
Li JY. To explore the nursing intervention method and effect of life

style in patients with obstructive sleep apnea syndrome(OSAS)[J]. World Journal of Sleep Medicine, 2019, 6(8): 1148-1150.
[3] 洪必莹,何森,陈晓平. 心率变异性研究进展[J]. 华西医学, 2013, 28(4): 614-618.
Hong BY, He S, Chen XP. Research progress on heart rate variability [J]. West China Medical Journal, 2013, 28(4): 614-618.
[4] da Silva TD, Massetti T, Crocetta TB, et al. Heart rate variability and cardiopulmonary dysfunction in patients with Duchenne muscular dystrophy: a systematic review[J]. Pediatr Cardiol, 2018, 39(5): 869-883.
[5] Hertzman AB. Photoelectric plethysmography of the fingers and toes in man[J]. Exp Biol Med, 1937, 37(3): 529-534.
[6] Poh MZ, McDuff DJ, Picard RW. Advancements in noncontact, multiparameter physiological measurements using a webcam[J]. IEEE Trans Biomed Eng, 2011, 58(1): 7-11.
[7] Sun ZD, Junttila J, Tulppo M, et al. Non-contact atrial fibrillation detection from face videos by learning systolic peaks[J]. IEEE J Biomed Health Inform, 2022, 26(9): 4587-4598.
[8] Sabour RM, Benezeth Y, De Oliveira P, et al. UBFC-Phys: a multimodal database for psychophysiological studies of social stress[J]. IEEE Trans Affect Comput, 2023, 14(1): 622-636.
[9] Juefei-Xu F, Wang R, Huang YH, et al. Countering malicious DeepFakes: survey, battleground, and horizon[J]. Int J Comput Vis, 2022, 130(7): 1678-1734.
[10] Verkruyse W, Svaasand LO, Nelson JS. Remote plethysmographic imaging using ambient light[J]. Opt Express, 2008, 16(26): 21434-21445.
[11] Poh MZ, McDuff DJ, Picard RW. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation[J]. Opt Express, 2010, 18(10): 10762-10774.
[12] de Haan G, Jeanne V. Robust pulse rate from chrominance-based rPPG[J]. IEEE Trans Biomed Eng, 2013, 60(10): 2878-2886.
[13] Wang WJ, Stuijk S, de Haan G. A novel algorithm for remote photoplethysmography: spatial subspace rotation[J]. IEEE Trans Biomed Eng, 2016, 63(9): 1974-1984.
[14] Wang WJ, den Brinker AC, Stuijk S, et al. Algorithmic principles

- of remote PPG[J]. IEEE Trans Biomed Eng, 2017, 64(7): 1479-1491.
- [15] Chen WX, McDuff D. DeepPhys: video-based physiological measurement using convolutional attention networks[C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 356-373.
- [16] Špetlík R, Franc V, Čech J, et al. Visual heart rate estimation with convolutional neural network[C]//British Machine Vision Conference. Newcastle, UK: BMVA Press, 2018: 1-12.
- [17] Yu Z, Li X, Zhao G. Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks [C]//Proceedings of the British Machine Vision Conference (BMVC). Cardiff, UK: BMVA Press, 2019: 1-12.
- [18] Lee E, Chen E, Lee CY. Meta-rPPG: remote heart rate estimation using a transductive meta-learner[C]//Computer Vision-ECCV 2020. Cham: Springer International Publishing, 2020: 392-409.
- [19] Liu X, Fromm J, Patel S, et al. Multi-task temporal shift attention networks for on-device contactless vitals measurement[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates, Inc., 2020: 19400-19411.
- [20] Hu M, Qian F, Guo D, et al. ETA-rPPGNet: effective time-domain attention network for remote heart rate measurement[J]. IEEE Trans Instrum Meas, 2021, 70: 1-12.
- [21] Gideon J, Stent S. The way to my heart is through contrastive learning: remote photoplethysmography from unlabelled video[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 3975-3984.
- [22] Liu X, Hill B, Jiang ZH, et al. EfficientPhys: enabling simple, fast and accurate camera-based cardiac measurement[C]//2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Piscataway, NJ, USA: IEEE, 2023: 4997-5006.
- [23] Yu ZT, Shen YM, Shi JG, et al. PhysFormer: facial video-based physiological measurement with temporal difference transformer [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 4176-4186.
- [24] Speth J, Vance N, Flynn P, et al. Non-contrastive unsupervised learning of physiological signals from video[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2023: 14464-14474.
- [25] Wu W, Peng HY, Yu SQ. YuNet: a tiny millisecond-level face detector[J]. Mach Intell Res, 2023, 20(5): 656-665.
- [26] Li GQ, Fang Q, Zha LL, et al. HAM: hybrid attention module in deep convolutional neural networks for image classification[J]. Pattern Recognit, 2022, 129: 108785.
- [27] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Comput, 1997, 9(8): 1735-1780.
- [28] Bobbia S, Macwan R, Benezeth Y, et al. Unsupervised skin tissue segmentation for remote photoplethysmography[J]. Pattern Recognit Lett, 2019, 124: 82-90.
- [29] Stricker R, Müller S, Gross HM. Non-contact video-based pulse rate measurement on a mobile service robot[C]//The 23rd IEEE International Symposium on Robot and Human Interactive Communication. Piscataway, NJ, USA: IEEE, 2014: 1056-1062.
- [30] Li XB, Chen J, Zhao GY, et al. Remote heart rate measurement from face videos under realistic situations[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2014: 4264-4271.
- [31] Niu XS, Shan SG, Han H, et al. RhythmNet: end-to-end heart rate estimation from face *via* spatial-temporal representation[J]. IEEE Trans Image Process, 2020, 29: 2409-2423.
- [32] Niu XS, Yu ZT, Han H, et al. Video-Based remote physiological measurement *via* cross-verified feature disentangling[C]//Computer Vision-ECCV 2020. Cham: Springer International Publishing, 2020: 295-310.
- [33] Li XB, Alikhani I, Shi JG, et al. The OBF database: a large face video database for remote physiological signal measurement and atrial fibrillation detection[C]//2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). Piscataway, NJ, USA: IEEE, 2018: 242-249.
- [34] Kaur P, Stoltzfus JC. Bland - Altman plot a brief overview[J]. Int J Acad Med, 2017, 3(1): 110-111.

(编辑:薛泽玲)