

DOI: 10.13382/j.jemi.B2609303

远程脉搏波信号多通道线性融合模型*

李阳捷 何星延 伏长虹 洪弘

(南京理工大学电子工程与光电技术学院 南京 210094)

摘要:针对传统远程脉搏波信号提取方法在复杂场景下鲁棒性不足,以及深度学习方法计算成本高、可解释性差的问题,提出一种结合机器学习与线性模型优势的轻量化解决方案。该方法主要包含4个步骤。首先,对面部区域进行多感兴趣区域(region of interest, ROI)划分,提取多通道红、绿、蓝(red, green, blue, RGB)时序信号。其次,通过相关滤波实现视频信号与参考脉搏波信号的时间对齐。随后,利用轻量化网络学习各通道的最优线性融合权重,并在时域与频域联合约束下实现脉搏波信号重建。在此过程中,模型保留了传统线性组合的可解释性,同时引入数据驱动的权重学习机制,实现多通道信号的自适应融合。最后,结合高分辨率频谱分析与时频脊线提取方法完成心率估计与动态追踪。在VitalVideo、MMPD、UBFC-Phys、iBVP及PhyRec等数据集上的实验结果表明,该模型在不同场景下均取得稳定性能。在VitalVideo数据集上,信噪比(signal-to-noise ratio, SNR)为6.15 dB,平均绝对误差(mean absolute error, MAE)为 1.45 min^{-1} ;在PhyRec数据集上,平均绝对误差为 0.29 min^{-1} 。在综合评估中,平均绝对误差为 8.45 min^{-1} ,均方根误差(root mean square error, RMSE)为 10.43 min^{-1} ,信噪比为-0.55 dB,准确率(accuracy, ACC)为0.84。结果表明,多通道线性融合策略在保持线性模型可解释性的同时,引入数据驱动建模能力,实现了对信号重建质量与心率估计稳定性的提升,能够在复杂环境下兼顾模型可解释性与性能表现,适用于非接触式生理信号测量场景。

关键词: 远程脉搏波提取;多感兴趣区域;线性融合网络;时频分析;心率追踪;信号对齐

中图分类号: TN911.7;TH789

文献标识码: A

国家标准学科分类代码: 510.4030

Multi-channel linear fusion model for remote photoplethysmography signal

Li Yangjie He Xingyan Fu Changhong Hong Hong

(School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: Conventional remote photoplethysmography (rPPG) methods suffer from limited robustness in complex scenarios, while deep learning approaches often involve high computational cost and poor interpretability. To address these issues, a lightweight solution that integrates the advantages of machine learning and linear modeling is developed. The proposed method first partitions the facial region into multiple regions of interest (ROI) to extract multi-channel (RGB) temporal signals, followed by correlation-based filtering to achieve temporal alignment between video signals and reference pulse signals. A lightweight network is then employed to learn the optimal linear fusion weights across channels, and pulse signal reconstruction is performed under joint constraints in both the time and frequency domains. In this process, the model preserves the interpretability of traditional linear combination methods while introducing a data-driven weight learning mechanism, enabling adaptive fusion of multi-channel signals. Finally, high-resolution spectral analysis and ridge-based time-frequency tracking are applied for heart rate estimation and dynamic tracking. Experimental results on multiple public datasets, including VitalVideo, MMPD, UBFC-Phys, iBVP, and PhyRec, demonstrate consistent performance across different scenarios. On the VitalVideo dataset, the proposed method achieves a signal-to-noise ratio (SNR) of 6.15 dB and a mean absolute error (MAE) of 1.45 min^{-1} . On the PhyRec dataset, the MAE reaches 0.29 min^{-1} . In the overall evaluation, the method achieves an MAE of 8.45 min^{-1} , a root mean square error (RMSE) of 10.43 min^{-1} , an SNR of -0.55 dB, and an accuracy (ACC) of 0.84. These

收稿日期: 2026-04-28 Received Date: 2026-04-28

* 基金项目: 国家自然科学基金(62471232)、江苏省重点研发项目(BE2023819)资助

results indicate that the proposed multi-channel linear fusion strategy improves signal reconstruction quality and heart rate estimation stability while maintaining model interpretability. The method effectively balances interpretability and performance under complex conditions, making it suitable for non-contact physiological signal measurement.

Keywords: remote photoplethysmography extraction; multiple regions of interest; linear fusion network; time-frequency analysis; heart rate tracking; signal alignment

0 引言

远程脉搏波 (remote photoplethysmography signal, rPPG) 提取技术凭借其非接触式测量优势,在健康监测、人机交互等领域受到广泛关注。与此同时,传统基于脉搏波的接触式测量方法在可穿戴传感器设计^[1]、无创血压估计^[2-3]以及电子血压计检测^[4]等方面已取得了丰富成果。这些研究为生理信号获取与建模提供了重要基础,但在实际应用中仍受到接触限制与使用场景约束。相比之下,rPPG 技术能够通过普通视频设备实现远程生理参数感知,具有更好的使用便捷性与应用潜力。

现有 rPPG 方法中,传统方法多基于生理建模与线性信号处理框架。典型方法包括独立成分分析 (independent component analysis, ICA)^[5]、基于肤色模型的色度 (chrominance-based method, CHROM) 方法^[6]以及投影空间优化的皮肤正交平面法 (plane-orthogonal-to-skin, POS)^[7]、判别式特征提取法 (discriminative signature based extraction, DIS)^[8]、此外,局部群不变性法 (local group invariance, LGI)^[9]通过局部群组不变性进一步提升了光照鲁棒性。这类方法可以总结为对面部 ROI 的 RGB 信号进行线性组合,以增强脉搏波成分并抑制噪声干扰。其优点在于物理意义明确、计算复杂度低,但由于融合系数依赖经验性假设或人工设计,在复杂场景(如运动干扰与光照变化)下的泛化能力有限。

近年来,基于深度学习的 rPPG 方法取得了显著进展。以 DeepPhys^[10]为代表的端到端模型,通过卷积神经网络学习视频中的时空特征,在受控环境下表现出优于传统方法的性能。后续工作如 PhysNet^[11]和 EfficientPhys^[12]等进一步优化了网络结构,提高了计算效率。然而,这类方法通常存在参数规模大、计算成本高的问题,不利于在资源受限设备上的部署;同时,其“黑箱”特性导致模型可解释性较弱。此外,在分布外场景(如剧烈运动或复杂光照)中,深度模型容易受到非线性干扰,导致性能下降。

综上所述,传统方法具有良好的稳定性与可解释性,但表达能力有限;深度学习方法具备较强的特征建模能力,但存在计算开销大和鲁棒性不足的问题。针对上述不足,本文提出一种融合两类方法优势的多通道线性融合框架:在保持线性融合结构的前提下,引入数据驱动机

制学习最优融合权重,从而在保证模型可解释性的同时提升其泛化性能。

本文首先将面部划分为 17 个 ROI,并构建 51 维多通道信号输入;在此基础上,设计轻量化网络学习从多通道到单通道的线性投影,实现对 rPPG 信号的自适应融合。同时,为保证监督信号的准确性,提出基于相关滤波的时间对齐方法,以消除视频与参考脉搏波信号之间的系统延时。

1 方法

针对复杂场景下远程脉搏波信号易受运动干扰和光照变化影响的问题,本文提出一种基于多通道线性融合网络的 rPPG 信号提取框架。该方法以视频序列为输入,通过多区域信号建模与自适应融合,实现对高质量脉搏波信号的稳健恢复,并进一步完成心率估计与追踪。

具体而言,所提出方法主要包括 3 个关键环节:首先,通过视频信号与参考脉搏波信号的时间对齐,构建严格同步的监督数据;其次,对面部区域进行多 ROI 划分,提取具有空间差异性的多通道 RGB 信号;最后,设计多通道线性融合网络,对多源信号进行加权融合,生成单通道 rPPG 信号,并在此基础上实现心率的频域分析与动态追踪。方法整体流程如图 1 所示。

1.1 视频与标签对齐

由于视频采集系统与接触式脉搏波测量设备在采集与处理过程中存在固有延时,直接使用未对齐的参考信号进行监督将导致模型学习到错误的时序关系。因此,在网络训练之前,有必要对视频信号与参考脉搏波信号进行精确时间对齐。

以接触式传感器获得的脉搏波信号 $s_c(t)$ 作为真值,并计算出频率 f_{HR} 。另外,从视频中构建用于精确对齐的 rPPG 参考信号。具体而言,基于人脸关键点检测方法 Mediapipe^[13]提取面部皮肤区域,并对该区域内像素的 RGB 值进行空间平均,得到随时间变化的三通道信号:

$$\bar{\mathbf{C}}(t) = (\mathbf{R}(t), \mathbf{G}(t), \mathbf{B}(t))^T = \text{mean}(\mathbf{c}_{ij}) = \frac{1}{|\text{ROI}(t)|} \times \sum_{i,j \in \text{ROI}(t)} (\mathbf{R}_{ij}(t), \mathbf{G}_{ij}(t), \mathbf{B}_{ij}(t))^T \quad (1)$$

在此基础上,为提高对齐的鲁棒性,选用 3 种传统经典 rPPG 方法: POS、CHROM 和 DIS,分别提取出备选 rPPG 信号 $s_i(t)$, $i = 1, 2, 3$,并从中选择 SNR 最高的信号

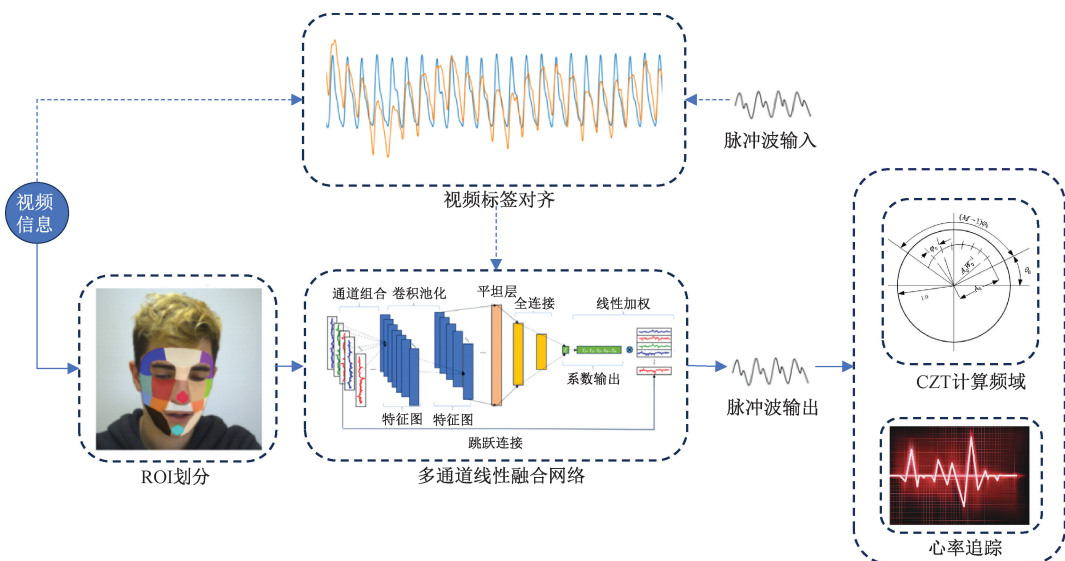


图 1 基于多通道信号融合网络的远程脉搏波信号提取方法流程

Fig. 1 Flowchart of the remote photoplethysmography signal extraction method based on multi-channel signal fusion network

$s_r(t)$, 作为视频侧的参考信号。信噪比定义如式(2)所示。

$$SNR(i) = 10 \lg \frac{\sum_{f_{\min}=0.8}^{f_{\max}=3.5} (U_i(f)P(f))^2}{\sum_{f_{\min}=0.8}^{f_{\max}=3.5} ((1 - U_i(f))P(f))^2} \quad (2)$$

式中: $P(f)$ 表示脉搏波信号的功率谱; $U_i(f)$ 是以 f_{HR} 为中心; 宽度 Δf (汉宁窗 3 dB 主瓣宽度) 的频谱权重窗。

通过互相关函数计算两信号在不同时间延迟下的相似性:

$$C(\tau) = \int_{t=0}^T s_c(\tau) \times s_r(t - \tau) dt \quad (3)$$

$$\tau_{\text{delay}} = \arg \max_{\tau} C(\tau) \quad (4)$$

互相关函数 $C(\tau)$ 的最大值对应两信号的最佳时间对齐位置, 据此得到时间延迟 τ_{delay} 。通过对 $s_c(t)$ 进行相应的时间平移, 得到与视频信号严格对齐的标签信号:

$$s'_c = s_c(t - \tau_{\text{delay}}) \quad (5)$$

最后, 截取对齐后重叠的有效时间段, 作为后续网络训练的输入与监督信号。经该步骤处理后, 视频信号与参考脉搏波信号在时间轴上实现精确同步, 如图 2 所示。

1.2 面部多 ROI 划分

由于面部不同区域在血流分布、皮肤特性及运动敏感性等方面存在显著差异, 各区域所包含的 rPPG 信号质量不一致。为充分利用空间信息并提升信号鲁棒性, 本文对面部区域进行多 ROI 划分, 从而构建具有空间多样性的多通道信号表示。所提出的 ROI 划分方法主要包括 3 个步骤: 细粒度区域划分、可见性与质量筛选以及区域合并。

首先, 基于人脸关键点检测结果, 采用 Delaunay 三角

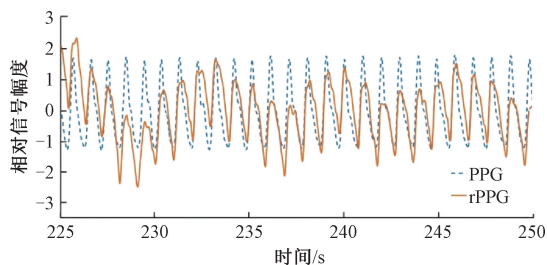


图 2 严格对齐的远程脉搏波信号与参考信号

Fig. 2 Strictly aligned remote photoplethysmographic (rPPG) signal and reference signal

剖分方法将面部划分为若干小三角区域, 从而获得更精细的空间信号表示, 如图 3 所示。每个小区域用于提取局部 RGB 平均信号, 保留空间差异信息。



图 3 面部小区域示意图

Fig. 3 Schematic diagram of small facial regions

其次, 为消除三维面部结构在二维投影中带来的遮挡与失真影响, 对小区域进行可见性筛选。通过比较三

角区域到摄像头的相对深度,优先保留可见区域,并利用蒙版避免重复采样;同时,根据区域法向量与视角方向的夹角,剔除倾斜角度过大的区域,以减少由反射衰减带来的低信噪比信号(阈值设为 $\pi/4$)。筛选结果如图4所示。

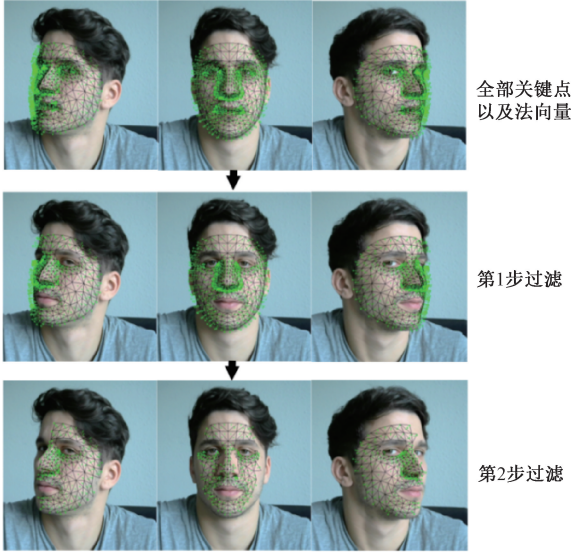


图4 两步过滤法

Fig. 4 Two-step filtering method

在此基础上,为获得具有一致信号特性的区域划分方案,对各小区域信号 $s_i(t)$ 与参考信号 $s_c(t)$ 的相关性

进行统计分析:

$$\rho_i = \sum_{t=0}^N \frac{s_i(t) - \mu(s_i(t))}{\sigma(s_i(t))} \times \frac{s_c(t) - \mu(s_c(t))}{\sigma(s_c(t))} \quad (6)$$

式中: μ 为均值; σ 为标准差。归一化处理消除了不同区域间幅值差异的影响,使得相关性系数能够真实反映各区域的信号质量。

统计结果表明,不同区域的信号质量存在显著差异,其中额头与脸颊区域相关性较高,而眼周、鼻翼等区域易受运动干扰。基于该分布特性,将空间相邻且信号特性相似的小区域进行合并,形成若干具有代表性的 ROI。

最后,为确保每个 ROI 具有清晰且稳定的空间边界,采用凸包方法对同一区域内的特征点进行包络,生成不重叠的闭合区域。最终,将面部划分为 17 个 ROI,并基于式(1)计算各区域的 RGB 均值信号 $s_i(t)$,作为后续多通道融合网络的输入。划分结果如图5所示。该策略在保证空间分辨率的同时,有效提升了输入信号的信噪比与一致性。

1.3 多通道信号融合网络

1) 网络结构及特点

针对多 ROI 多通道信号融合过程中传统方法依赖固定投影系数、难以适应复杂场景变化的问题,本文提出一种多通道线性融合网络(multi-channel fusion, MCF)。该方法的核心目标并非直接通过深度网络生成 rPPG 信号,而是利用数据驱动方式学习多通道信号的最优线性融合权重,从而在保持模型可解释性的前提下提升信号质量。

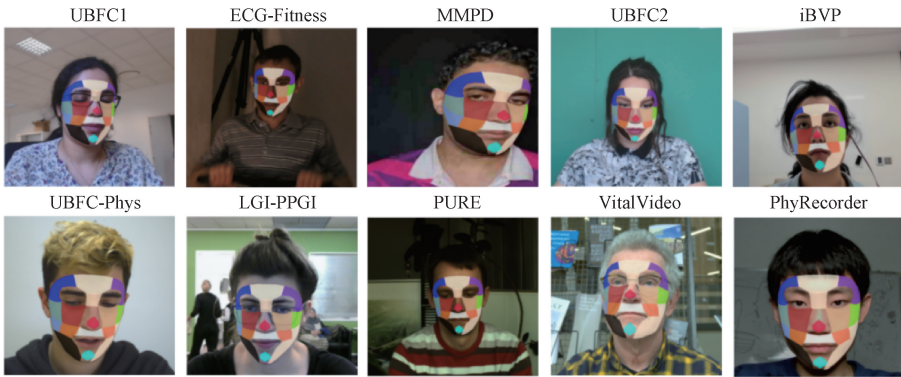


图5 蒙版凸包法在不同数据集的最终效果

Fig. 5 Final effect of the mask convex hull algorithm in different datasets

相比于 DeepPhys 等黑箱网络, MCF 学到的 51 维权重向量,实际上反映了不同面部区域在不同场景下的贡献度。其核心优势在于通过数据驱动方式为 17 个面部感兴趣区域分别学习融合权重,而这一权重分布并非随机或无意义的数值,而是具有可解释性。具体而言,模型最终输出的 51 维权重向量反映了不同面部区域在不同场景下对脉搏波信号重建的贡献程度。

设输入为来自 17 个 ROI 的 RGB 信号,共构成 51 维多通道时间序列,本文将 rPPG 信号建模为该高维信号在一组最优投影向量上的线性映射。不同于传统方法中基于经验设计的固定投影系数,本文通过轻量化神经网络自适应学习该投影向量,使其能够根据不同场景自动调整各通道的贡献权重。

在网络结构设计上, MCF 由通道组合层、卷积池化

层、全连接层以及线性融合层构成,如图 6 所示。通道组合层用于对输入多通道信号进行重组与压缩,以增强通道间相关性并减少冗余信息;卷积池化层用于提取时序特征并进行降维;全连接层将提取的特征映射回与输入一致的通道维度,从而生成融合权重向量。

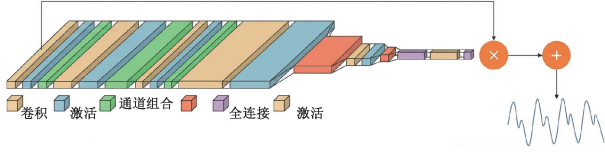


图 6 线性信号融合网络层次

Fig. 6 Hierarchy of the linear signal fusion network

为保证最终输出满足线性融合约束,本文在网络末端引入跳跃连接结构,将原始输入信号与网络输出的权重向量进行逐通道线性加权求和,生成单通道 rPPG 信号。该设计的关键在于网络仅负责学习融合权重,而不直接参与非线性信号生成,从而有效避免深度模型引入的非线性失真问题。

通过上述结构设计,MCF 在保持线性模型物理可解释性的同时,引入了数据驱动的自适应能力,使其能够在复杂场景(如运动干扰和光照变化)下实现更加稳健的信号融合。该结构本质上可视为一种受约束的线性子空间学习过程。

2) 输入信号预处理

为提高模型训练的稳定性与输入信号的一致性,对多通道 ROI 信号进行统一预处理。首先对各区域 RGB 信号进行线性插值至 30 Hz,以统一采样率;随后采用 0.5~3.5 Hz 带通滤波提取脉搏波有效频段,抑制低频漂移与高频噪声。

在此基础上,将信号按固定长度(256 帧)进行分窗处理,构建训练样本。将每一时间窗内的第 i 个 ROI、第 c 个颜色通道的信号,记为 S_c^i ,并进行归一化处理:

$$\widetilde{S}_c^i = \frac{S_c^i - \min(S_c^i)}{\max(S_c^i) - \min(S_c^i)} \quad (7)$$

归一化操作用于消除不同区域间由于光照强度与反射特性差异带来的幅值不一致问题,从而保证各通道在统一尺度下参与融合。同时,该处理能够有效抑制由于局部运动或区域漂移引起的缓慢幅值突变,避免个别异常通道对网络学习产生干扰。

3) 数据增强策略

为提升模型在不同心率分布及信号缺失场景下的鲁棒性,本文引入两种数据增强策略:重采样增强与随机失效模拟。

首先,针对数据集中高心率样本分布不足的问题,采用基于重采样的增强方法,通过对原始信号进行时间尺

度变换,将部分 $60 \sim 90 \text{ min}^{-1}$ 区间样本扩展至 $120 \sim 180 \text{ min}^{-1}$ 范围,并叠加高斯噪声后重新滤波,以增加样本多样性并降低过拟合风险。

其次,为模拟实际应用中局部区域信号丢失或质量退化的情况,引入随机失效增强策略。具体而言,对约 20% 的训练样本,随机选择单一通道并在约 1 s 时间范围内进行数值替换(基于邻域插值),从而构造局部信息缺失情形。该策略能够促使模型学习在部分通道失效时仍保持稳定输出。

通过上述数据增强方法,有效提升了模型对复杂场景的适应能力。

4) 损失函数

为提高模型输出信号与参考信号在波形结构及频谱分布上的一致性,本文构建了一个联合时域与频域约束的组合损失函数。该函数从信号相关性与频谱结构两个角度对模型输出进行约束,提升 rPPG 信号重建质量。

在时域上,采用负皮尔逊相关系数(pearson correlation coefficient, PCC)作为损失函数,用于衡量预测信号与参考信号之间的线性相关性。其定义为:

$$L_{\text{Pearson}} = 1 - \frac{\text{Cov}(\hat{S}_{\text{rPPG}}, S_{\text{PPG}})}{\sigma_{\hat{S}_{\text{rPPG}}} \sigma_{S_{\text{PPG}}}} \quad (8)$$

式中: $\hat{S}_{\text{rPPG}}$ 为模型预测信号; S_{PPG} 为参考脉搏波信号; Cov 与 σ 分别表示协方差与标准差。该损失函数能够有效约束信号整体形态的一致性,使模型关注波形变化趋势而非绝对幅值。

在频域上,为保证预测信号在主频分布上的准确性,引入频域相关性损失。通过对预测信号与参考信号进行离散傅里叶变换(discrete fourier transform, DFT),约束其频谱能量分布的一致性,定义为:

$$L_{\text{Frequency}} = \frac{1}{N} \sum_{k=1}^N \left| \frac{\mathcal{F}(\hat{S}_{\text{rPPG}}(f_k))}{\mathcal{F}(S_{\text{PPG}}(f_k))} \right|^2 \quad (9)$$

其中, $\mathcal{F}(\cdot)$ 为离散傅里叶变换。该损失项能够强化模型对主频成分(对应心率信息)的建模能力,从而提高心率估计的准确性。

综合上述两部分,最终损失函数定义为:

$$L_{\text{total}} = L_{\text{Pearson}} + L_{\text{Frequency}} \quad (10)$$

该组合损失函数在时域与频域两个层面同时对模型进行约束:前者保证信号波形的整体一致性,后者强化频谱结构尤其是主频信息的准确性。两者相互补充,使模型在复杂噪声环境下仍能够学习到稳定且具有生理意义的 rPPG 信号表示。

1.4 基于时频图的心率估计与追踪

在获得融合后的 rPPG 信号后,需要进一步从中提取

心率信息。由于心率在实际场景中随时间动态变化,该问题可建模为短时信号的时变频率估计问题。为此,本文采用高分辨率频谱分析方法获取信号的时频表示,并结合频率轨迹提取算法实现稳定的心率追踪。

1) 基于线性调频 Z 变换 (chirp z-transform, CZT) 的高分辨率频谱分析

在短时窗条件下,传统离散傅里叶变换的频率分辨率受限于信号长度,难以满足精细心率估计的需求。为提高特定频段内的频率分辨率,本文采用 CZT^[14] 对 rPPG 信号进行频谱分析。

与 DFT 在单位圆上等间隔采样不同, CZT 允许在复平面指定路径上进行频谱计算,从而能够在感兴趣频段 (如 0.8~3.5 Hz) 内以更高分辨率进行频谱分析。该特性使其在不增加时间窗长度的情况下,提高心率估计的精度。

通过对信号进行短时 CZT 变换,可得到对应的时频表示,为后续心率追踪提供基础。如图 7 所示的对比,在相同的时间复杂度下, CZT 相对 FFT 频率分辨率可以提高 5 倍左右。

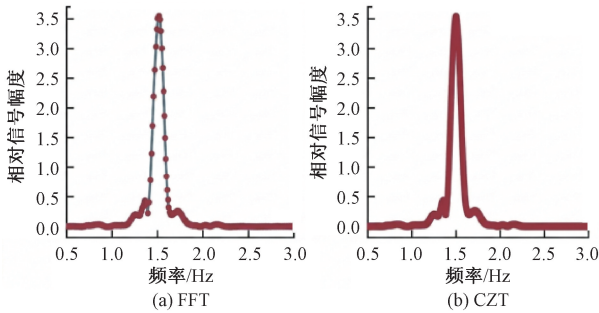


图 7 频域分辨率

Fig. 7 Frequency domain resolution

2) 心率追踪算法

在获得信号的时频表示后,心率估计可转化为在时频图中提取主频轨迹的问题。理想情况下,该轨迹对应能量最大的频率分量,并在时间上保持连续和平滑变化。为此,本文采用基于时频脊线提取^[15]的心率追踪方法。

首先,在每一时刻的频谱中选取局部能量极大值作为候选频率点,以抑制噪声干扰:

$$p_n[d]: \begin{cases} [\Delta_k | X_n[k]|]_{k=p_n[d]} = 0, & Q_n[p_n[d]] \equiv |X_n[p_n[d]|, \\ [\Delta_k^2 | X_n[k]|]_{k=p_n[d]} < 0, \end{cases} \quad (11)$$

式中: $p_n[d]$ 表示 n 时刻的所有脊线的频域采样点,显然 $p_n[d] \subset \{K_0, \dots, K_N\}$, 该公式定义了频域 $X_n[k]$ 的位置条件; $Q_n[p_n[d]]$ 表示脊线频域对应的幅值。

在此基础上,构建结合信号幅值与频率连续性的优

化目标函数如式 (12) 所示,并通过动态规划在全局范围内搜索最优频率路径。这一优化函数会在时频平面上选择能量最大、最稳定的路径,以避免噪声的影响。在实际操作中, $\hat{\omega}_n$ 是基于过去时刻的频率状态, $\hat{\omega}_{n-1}, \hat{\omega}_{n-2}, \dots, \hat{\omega}_{n_0}$ 逐步更新的。

$$\hat{\omega}_n = \nu \left(\underset{k \in [k_0, \dots, k_N]}{\operatorname{argmax}} \left(F \left[\begin{matrix} t_n, Q_n[p_n], \\ p_n, \hat{\omega}_{n-1}, \hat{\omega}_{n-2}, \dots, \hat{\omega}_{n_0} \end{matrix} \right] \right) \right) \quad (12)$$

$$F[\dots] = \log Q_m(p_n[d]) +$$

$$w_2(\nu(p_n[k]), m[\hat{\omega}], \operatorname{IQR}[\hat{\omega}], \beta) +$$

$$w_1(\nu(p_n[d]) - \hat{\omega}_{n-1}, m[\Delta\hat{\omega}], \operatorname{IQR}[\Delta\hat{\omega}], \alpha) \quad (13)$$

式中: $\nu(p_n[d])$ 表示当前的频率位置; $\log Q_m(p_n[d])$ 是对幅值的对数处理; 为提高鲁棒性, 引入基于中位数 $m[f(t)]$ 与四分位距 $\operatorname{IQR}[f(t)]$ 的约束, 刻画心率的典型分布范围:

$$m[f(t)] \equiv \operatorname{perc}_{0.5}[f(t)],$$

$$\operatorname{IQR}[f(t)] \equiv \operatorname{perc}_{0.75}[f(t)] - \operatorname{perc}_{0.25}[f(t)] \quad (14)$$

并通过如式 (15) 所示的惩罚项抑制异常频率跳变, 确保结果的平滑性与稳定性。

$$w_1(\Delta\xi, m[\Delta\hat{\omega}], \operatorname{IQR}[\Delta\hat{\omega}], \alpha) = -\alpha \left| \frac{\Delta\xi - m[\Delta\hat{\omega}]}{\operatorname{IQR}[\Delta\hat{\omega}]} \right|$$

$$w_2(\xi, m[\hat{\omega}], \operatorname{IQR}[\hat{\omega}], \beta) = -\beta \left| \frac{\xi - m[\hat{\omega}]}{\operatorname{IQR}[\hat{\omega}]} \right| \quad (15)$$

其中, α 控制频率变化的抑制程度, 而 β 控制频率本身的相对偏离。最终, 通过迭代优化得到的时频脊线对应信号的瞬时频率, 即为心率随时间的变化轨迹。实验结果表明, 该方法能够在存在噪声干扰的情况下实现较稳定的心率追踪 (图 8)。

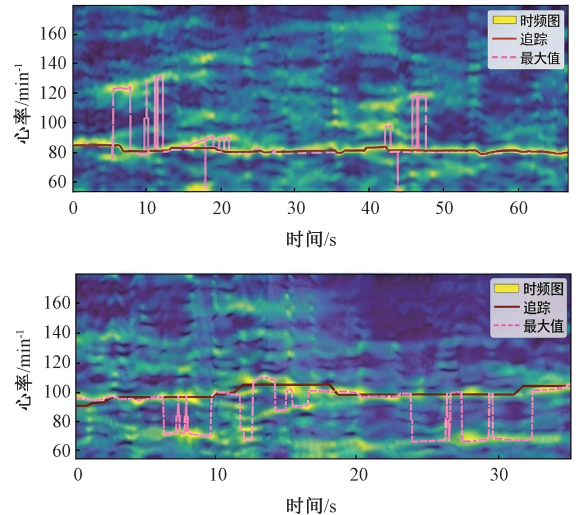


图 8 频域追踪算法效果示意图

Fig. 8 Schematic diagram of the frequency domain tracking algorithm effect

2 实 验

2.1 实验设置

为全面评估所提出方法的有效性,本文在多个公开 rPPG 数据集上开展实验,并从心率估计精度与信号重建质量两个方面进行定量分析。所使用的数据集包括公开数据集 ECG-Fitness^[16]、iBVP^[17]、LGI-PPGI, MMPD^[18]、PURE^[19]、UBFC-Phys^[20]、UBFC-rPPG^[21]、VitalVideo^[22], 和一个自制数据集 PhyRec(已在 github 上公开),覆盖了不同光照条件、运动强度及设备环境下的多种复杂场景。

在训练策略上,本文采用跨数据集训练与独立测试相结合的评估方式。具体而言,选取 PURE、ECG-Fitness、UBFC-rPPG (UBFC1 与 UBFC2) 以及 LGI-PPGI 作为训练集,并从中采用基于被试者划分的方式随机划分 10% 的样本作为验证集,其余用于模型训练。在 VitalVideo、MMPD、UBFC-Phys、iBVP 以及 PhyRec 数据集上进行独立测试,以验证模型在分布外场景下的泛化能力。

模型训练在 Linux 5.15 平台上完成,硬件配置为 Intel Xeon 8336C CPU(64 核,2.30 GHz)、128 GB 内存及 NVIDIA A5000 GPU。模型基于 PyTorch 2.3.0 框架实现,并采用 Adam 优化器进行参数更新,初始学习率设为 1×10^{-3} 。训练过程中,批大小设置为 512,每个 epoch 后在验证集上评估模型性能,并根据验证损失选择最优模型参数。

训练过程中损失函数的收敛曲线如图 9 所示。可以观察到,模型在约 25 个 epoch 后趋于稳定,因此选取验证集损失最低的第 25 轮参数作为最终模型。

2.2 误差计算及结果分析

为从多角度评估所提出方法的性能,本文在时域与频域分别引入多种评价指标,对信号重建质量与心率估计精度进行综合分析。

在时域上,采用皮尔逊相关系数衡量预测信号与参考信号之间的线性相关性,用于评估波形结构的一致性。

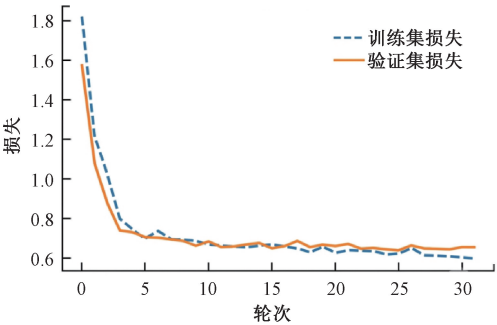


图 9 训练损失下降曲线

Fig. 9 Training loss reduction curve algorithm effect

其定义如式(16)所示。

$$PCC = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \tag{16}$$

在频域上,采用平均绝对误差 (MAE) 与均方根误差 (RMSE) 对心率估计结果进行量化。其中,MAE 反映整体估计精度,而 RMSE 对较大误差更加敏感,能够体现极端情况下的性能稳定性。其定义分别为:

$$MAE = \frac{1}{N} \sum_{i=1}^N |HR_i^{pred} - HR_i^{gt}| \tag{17}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (HR_i^{pred} - HR_i^{gt})^2} \tag{18}$$

此外,引入准确率 (ACC) $ACC \Delta x$,用于衡量估计误差不超过给定阈值 Δx (本文设为 5 min^{-1}) 时的样本占比,从而反映心率追踪的稳定性:

$$ACC(\Delta x) = \frac{|\{i \mid |HR_i^{pred} - HR_i^{gt}| \leq \Delta x, 1 \leq i \leq N\}|}{N} \tag{19}$$

同时,本文采用信噪比 (SNR) 作为补充指标,用于评估算法在复杂噪声环境下对脉搏波信号的恢复能力。

为验证所提出方法的有效性,本文将其与典型传统方法 DIS、POS 以及深度学习方法 DeepPhys 进行对比实验。以下结果均在窗长 20 s,步长 10 s,采样频率为 30 Hz 的参数状态下计算。表 1 给出了各方法在不同测试数据集上的性能结果。

表 1 本文方法 (MCF) 与现有方法在各个数据集上效果对比

	VitalVideo		MMPD		UBFC-Phys		iBVP		PhyRec	
	SNR	MAE	SNR	MAE	SNR	MAE	SNR	MAE	SNR	MAE
DIS	-3.15	2.69	<u>-7.24</u>	<u>15.96</u>	-8.35	21.49	-5.15	4.75	-4.52	5.27
POS	<u>4.12</u>	1.43	-7.43	16.30	<u>-7.01</u>	21.00	<u>-3.54</u>	4.50	7.30	<u>0.52</u>
DeepPhys	1.95	2.42	-9.73	19.54	-9.93	<u>20.04</u>	-3.97	<u>4.35</u>	1.67	2.62
MCF	6.15	<u>1.45</u>	-6.80	15.93	-5.57	18.61	-2.37	4.04	<u>4.19</u>	0.29

从表 1 可以看出,MCF 在多数数据集上均取得了最优或次优的结果。在 VitalVideo 数据集上,MCF 在 SNR 和 MAE 两项指标上均优于对比方法,表明其在复杂场景下具有更强的信号恢复能力。在 MMPD 与 UBFC-Phys 等包含显著运动干扰的数据集中,MCF 同样表现出更稳定的误差水平,相较于深度学习方法在部分场景中的性能波动,具有更好的鲁棒性。在 MMPD 数据集上,本文提出的 MCF 方法取得了 15.93 min^{-1} 的 MAE,略优于 DIS 方法的 15.96 min^{-1} ,但两者差距极小。对此,分析认为可能存在两方面原因。首先,MMPD 数据集本身具有高度的场景多样性,涵盖了不同的光照条件、运动类型及采集设备,这对任何算法都构成了严峻挑战。在剧烈的头部运动和复杂光照耦合的场景下,即使是基于线性融合的优化策略,其性能也可能趋近于传统方法的性能上限,导致差距难以进一步拉开。其次,一个更深层次的原因在于对齐策略。如表 2 所示,MCF 在整体评估中的皮尔逊相关系数达到 0.50,显著高于 DIS 方法的 0.35,这表明 MCF 预测的信号波形与真实脉搏波具有更高的一致性。然而,MMPD 数据集中可能存在更为复杂的、非线性的系统时延,而本文目前采用的基于全局互相关的线性时间对齐方法,可能无法在这些极端样本上完美校正标签与信号的同步关系,从而部分掩盖了 MCF 在波形质量上的真实优势,最终反映在 MAE 指标上的提升不够显著。尽管如此,MCF 在该数据集上依然取得了所有对比方法中最优的 MAE 和 SNR 结果,验证了其在复杂场景下的稳定性和泛化能力。

在 iBVP 与 PhyRec 数据集上,MCF 在保持较低 MAE 的同时,实现了相对较高的 SNR,说明其在不同信号质量条件下均能够有效提取脉搏波主成分。整体而言,MCF 在多个数据集上的表现更加均衡,体现出良好的跨场景泛化能力。

表 2 进一步汇总了不同方法在综合指标上的表现。从整体结果来看,MCF 在 MAE、RMSE、PCC 以及 ACC 等关键指标上均取得了最优结果。MCF 的 MAE 为 8.45, RMSE 为 10.43,相较于传统方法 DIS 均有明显改善。在信号质量方面,MCF 的 SNR 达到 -0.55 dB ,与 POS 方法相当,同时优于 DeepPhys。

表 2 本文方法与现有方法在不同指标上的对比

Table 2 Comparison of various indicators between the proposed method and existing methods

	MAE	RMSE	SNR	PCC	ACC($\Delta 5\text{ min}^{-1}$)
DIS	9.18	11.71	-5.46	0.35	0.76
POS	8.61	10.86	-2.27	0.44	0.79
DeepPhys	9.24	12.83	-6.57	0.23	0.65
MCF	8.45	10.43	-0.55	0.50	0.84

在时域相关性方面,MCF 的 PCC 达到 0.50,显著高

于对比方法。这一结果得益于本文提出的信号对齐策略(1.1 节)。该策略能够有效减小预测信号与参考信号之间的时间偏移,从而提升波形一致性。相比之下,未显式进行时间对齐处理的方法在该指标上表现较弱。

此外,在准确率指标 $\text{ACC}(5\text{ min}^{-1})$ 上,MCF 达到 0.84,表明其在大多数时间段内能够实现较为准确且稳定的心率估计。综合来看,MCF 在信号保真度与心率估计精度之间取得了良好的平衡。

综上所述,实验结果表明,所提出的多通道线性融合方法在不同数据集与复杂场景下均具有良好的稳定性与泛化能力。通过结合线性模型的可解释性与数据驱动方法的自适应能力,该方法在保证计算效率的同时,实现了对远程脉搏波信号的高质量重建与准确估计。

3 结 论

本文针对复杂场景下远程脉搏波信号易受运动干扰与光照变化影响的问题,提出了一种基于多感兴趣区域的多通道线性融合方法。该方法在保持线性模型可解释性的前提下,引入数据驱动机制学习最优融合权重,实现了多通道信号的自适应加权融合。

在方法设计上,首先通过基于互相关的时间对齐策略,有效消除了视频信号与参考脉搏波信号之间的系统延时;其次,基于 Delaunay 三角剖分与凸包约束构建多 ROI 表示,提高了空间信息利用效率与信号一致性;在此基础上,设计轻量化线性融合网络,实现对多通道信号的稳健融合;最后,结合高分辨率频谱分析与时频脊线提取方法,实现了稳定的心率估计与动态追踪。

在多个公开数据集上的实验结果表明,所提出方法在心率估计精度(MAE、RMSE)、信号质量(SNR)及波形一致性(PCC)等指标上均取得了具有竞争力的性能。相较于传统方法与深度学习方法,MCF 在不同场景下表现出更稳定的性能,尤其在存在运动干扰与光照变化的复杂环境中,展现出更好的鲁棒性与泛化能力。

综上,本文方法通过融合线性建模与数据驱动学习的优势,在保证计算效率与可解释性的同时,实现了对远程脉搏波信号的高质量重建,为无接触式生理信号测量提供了一种有效且可部署的解决方案。

参考文献

[1] 曹从军, 赖凯旋, 王旭升, 等. 基于柔性压电薄膜的可穿戴脉搏传感器设计[J]. 电子测量与仪器学报, 2024, 38(10): 35-47.

Cao Congjun, Lai Kaixuan, Wang Xusheng, et al. Design of wearable pulse sensor based on flexible piezoelectric film[J]. Journal of Electronic Measurement

- and Instrumentation, 2024, 38(10): 35-47.
- [2] 闫硕, 吴阳, 王世锋, 等. 基于多生理参数的连续无创血压预测[J]. 电子测量技术, 2025, 48(8): 126-132.
- Yan Shuo, Wu Yang, Wang Shifeng, et al. Continuous non-invasive blood pressure prediction based on multiple physiological parameters [J]. Electronic Measurement Technology, 2025, 48(8): 126-132.
- [3] 吴海燕, 季忠, 李孟泽. 基于脉搏波的无创连续血压监测模型簇研究[J]. 仪器仪表学报, 2020, 41(7): 152-160.
- Wu Haiyan, Ji Zhong, Li Mengze. Research on non-invasive continuous blood pressure monitoring model cluster based on pulse wave [J]. Chinese Journal of Scientific Instrument, 2020, 41(7): 152-160.
- [4] 林俊辉, 高改丽, 陈湘强. 基于脉搏波原理的电子血压计的检测与探讨[J]. 国外电子测量技术, 2024, 43(6): 48-51.
- Lin Junhui, Gao Gaili, Chen Xiangqiang. Detection and discussion of electronic sphygmomanometer based on pulse wave principle [J]. Foreign Electronic Measurement Technology, 2024, 43(6): 48-51.
- [5] Comon P. Independent component analysis, a new concept [J]. Signal Processing, 1994, 36(3): 287-314.
- [6] De Haan G, Jeanne V. Robust pulse rate from chrominance-based rPPG [J]. IEEE Transactions on Biomedical Engineering, 2013, 60(10): 2878-2886.
- [7] Wang W J, Den Brinker A C, Stuijk S, et al. Algorithmic principles of remote PPG [J]. IEEE Transactions on Biomedical Engineering, 2017, 64(7): 1479-1491.
- [8] Wang W, Den Brinker A C, Stuijk S, et al. Discriminative signatures for remote-PPG [J]. IEEE Transactions on Biomedical Engineering, 2020, 67(5): 1462-1473.
- [9] Pilz C S, Zaunseder S, Krajewski J, et al. Local group invariance for heart rate estimation from face videos in the wild [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. Salt Lake City: IEEE, 2018: 1254-1262.
- [10] Chen W X, McDuff D. DeepPhys: Video-based physiological measurement using convolutional attention networks [C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018: 349-365.
- [11] Yu Z T, Li X B, Zhao G Y. Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks [PP/OL]. [2026-04-28]. <https://doi.org/10.48550/arXiv.1905.02419>.
- [12] Liu X, Hill B, Jiang Z H, et al. EfficientPhys: Enabling simple, fast and accurate camera-based vitals measurement [C]//Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa: IEEE, 2023: 4997-5006.
- [13] Lugaresi C, Tang J Q, Nash H, et al. MediaPipe: A framework for building perception pipelines [PP/OL]. [2026-04-28]. <https://doi.org/10.48550/arXiv.1906.08172>.
- [14] Bluestein L. A linear filtering approach to the computation of discrete Fourier transform [J]. IEEE Transactions on Audio and Electroacoustics, 1970, 18(4): 451-455.
- [15] Iatsenko D, McClintock P V E, Stefanovska A. On the extraction of instantaneous frequencies from ridges in time-frequency representations of signals [PP/OL]. [2026-04-28]. <https://doi.org/10.48550/arXiv.1310.7276>.
- [16] Špetlík R, Franc V, Matas J. Visual heart rate estimation with convolutional neural network [C]//Proceedings of the British Machine Vision Conference. Newcastle: BMVA Press, 2018: 3-6.
- [17] Joshi J, Cho Y. iBVP Dataset: RGB-thermal rPPG dataset with high resolution signal quality labels [J]. Electronics, 2024, 13(7): 1334.
- [18] Tang J K, Chen K Q, Wang Y T, et al. MMPD: Multi-domain mobile video physiology dataset [C]//2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2023: 1-5.
- [19] Ferrari A, Spagnolo G O, Gnesi S. PURE: A dataset of public requirements documents [C]//2017 IEEE 25th International Requirements Engineering Conference (RE). IEEE, 2017: 502-505.
- [20] Sabour R M, Benezeth Y, De Oliveira P, et al. UBFC-Phys: A multimodal database for psychophysiological studies of social stress [J]. IEEE Transactions on Affective Computing, 2023, 14(1): 622-636.

[21] Bobbia S, Macwan R, Benezeth Y, et al. Unsupervised skin tissue segmentation for remote photoplethysmography [J]. Pattern Recognition Letters, 2019, 124: 82-90.

[22] Toye P J. Vital Videos: A dataset of videos with PPG and blood pressure ground truths [PP/OL]. [2026-04-28]. <https://doi.org/10.48550/arXiv.2306.11891>.

作者简介



李阳捷,现为南京理工大学电子工程与光电技术学院通信工程专业本科生,主要研究方向为视频信号处理,新型生物医学传感。

E-mail: 365315162@qq.com

Li Yangjie is currently a B. Sc. candidate in the School of Electronic and Optical Engineering, Nanjing University of Science and Technology. His main research interests include video signal processing and novel biomedical sensing.



何星延,2025 年获南京理工大学硕士学位。研究方向为图像/视频处理及远程光体积描记技术。

E-mail:740325411@qq.com

He Xingyan received the M. Sc. degree from Nanjing University of Science and Technology in 2025. His research interests include video and image processing, and physiological signal detection, with a focus on remote photoplethysmography.



伏长虹(通信作者),2008 年获香港理工大学获博士学位。现为南京理工大学电子工程与光电技术学院副教授。主要研究方向为信号处理、图像视频编码、视频生理监测和多模态融合的远程健康感知。

E-mail: enchfu@njust.edu.cn

Fu Changhong (Corresponding author) received the Ph. D. degree from the Hong Kong Polytechnic University in 2008. He is currently an associate professor in School of Electronic and Optical Engineering at Nanjing University of Science and Technology. His main research interests include the area of signal processing, image and video compression, video-derived physiological monitoring and multimodal fusion for robust remote health sensing.



洪弘,2010 年获南京大学博士学位。现为南京理工大学电子工程与光电技术学院教授、副院长。研究方向为微波技术生物医学应用及音频/视频/雷达多模态信号处理。

E-mail:hongnju@njust.edu.cn

Hong Hong received the Ph. D. degree from Nanjing University in 2010. He is currently a professor and the Vice Dean with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology. His research interests include biomedical applications of microwave technology, audio/visual/radar signal processing.