



利用掩码自编码器预训练的热红外视频呼吸监测方法*

毛如石¹, 吕学斌^{2, Δ}, 张伟^{1, 3}

1. 四川大学华西医院 生物医学大数据中心(成都 610041); 2. 四川大学计算机学院(成都 610065);
3. 四川大学华西医院 心理卫生中心(成都 610041)

【摘要】目的 通过热红外视频构建时空图, 结合自监督学习方法自动捕获呼吸引发的口鼻区域温度变化趋势, 实现长期非接触式呼吸监测, 为情绪、应激反应、呼吸系统疾病等早期预警提供技术支撑。**方法** 针对现有方法对热红外温度变化趋势表征能力不足的问题, 提出两阶段自监督重建时空图-下游微调方法: 首先通过仿射变换矩阵完成普通摄像头与热红外摄像头的空间配准, 实现关键点跟踪并构建时空图; 考虑到呼吸信号敏感区域的个体差异, 先通过自监督学习挖掘深度表征, 再对隐向量微调以适配下游监测任务, 同时提出进出气量比与呼吸时长平均绝对误差两项新评估指标。**结果** 数据集实验结果表明, 该两阶段训练方法可精准捕获呼吸信号趋势, 端到端特征学习的波形拟合精度显著优于传统方法及现有流行模型, 核心性能指标分别为: 平均绝对误差 0.07 ± 0.02 、均方根误差 0.69 ± 0.11 、皮尔逊相关系数 0.15 ± 0.04 、进出气量比 $0.40 \pm 0.12/0.26 \pm 0.05$ 、呼吸时长平均绝对误差 $0.79 \pm 0.19/0.79 \pm 0.10$ 。**结论** 基于掩码自编码器(masked autoencoder)的自监督预训练波形分析方法在呼吸监测中展现频域与时域双重优势; 同时发现呼吸阶段脉搏信号的形态学差异, 呼气阶段光电容积脉搏波(photoplethysmography, PPG)信号峰度显著降低、偏度减小。该发现为从PPG形态特征维度评估心肺耦合功能与自主神经调节提供了新视角。

【关键词】 自编码器 呼吸频率 生理监测

A Method for Respiratory Monitoring in Thermal Infrared Video Using Masked Autoencoder Pretraining

MAO Rushi¹, LÜ Xuebin^{2, Δ}, ZHANG Wei^{1, 3}. 1. West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu 610041, China; 2. School of Computer Science, Sichuan University, Chengdu 610065, China; 3. Mental Health Center, West China Hospital, Sichuan University, Chengdu 610041, China

Δ Corresponding author, E-mail: lvxb@scu.edu.cn

This study was supported by STI 2030-Major Projects (No. 2021ZD0200500).

[Abstract] Objective By constructing spatiotemporal maps from thermal infrared videos and applying self-supervised learning methods, temperature change trends in the oral and nasal regions caused by breathing can be automatically detected, enabling long-term, non-contact breathing monitoring. This offers technical support for early warning of emotions, stress responses, and respiratory system diseases. **Methods** To address the limitations of existing methods in characterizing temperature variation trends in thermal infrared, a two-stage self-supervised reconstruction spatio-temporal graph and downstream fine-tuning method is proposed. First, spatial registration between the ordinary camera and the thermal infrared camera is achieved using an affine transformation matrix, enabling key point tracking and the construction of a spatio-temporal graph. Considering individual differences in the sensitive areas of the respiratory signal, deep representations are initially extracted through self-supervised learning, and the latent vectors are then fine-tuned to adapt to downstream monitoring tasks. Additionally, two new evaluation metrics are introduced: the average absolute error of the ratio of inhaled to exhaled gas volumes and the average absolute error of respiratory duration. **Results** The experimental results on the dataset show that this two-stage training method can accurately capture the trend of respiratory signals. The waveform fitting accuracy of end-to-end feature learning is significantly better than that of traditional methods and current popular models. The core performance indicators are as follows: mean absolute error 0.07 ± 0.02 , root mean square error 0.69 ± 0.11 , Pearson correlation coefficient 0.15 ± 0.04 , ratio of inhalation to exhalation volume $0.40 \pm 0.12/0.26 \pm 0.05$, and mean absolute error of respiratory duration $0.79 \pm 0.19/0.79 \pm 0.10$. **Conclusion** The self-supervised pre-training waveform analysis method based on a masked autoencoder demonstrates advantages in both the frequency and time domains for respiratory monitoring. Additionally, morphological differences in the pulse signal during the respiratory phase were identified. The kurtosis of the photoplethysmography (PPG) signal significantly decreased, and the skewness was reduced during the exhalation phase. This finding offers a new perspective for evaluating cardiopulmonary coupling function and autonomic nerve regulation from the morphological feature dimension of PPG.

[Key words] Autoencoder Respiratory rate Monitoring, physiologic

* 科技创新2030-“脑科学与类脑研究”重大项目(No. 2021ZD0200500)资助

Δ 通信作者, E-mail: lvxb@scu.edu.cn

出版日期: 2026-05-20

呼吸监测是情绪、应激反应和健康状态评估的关键指标。自主呼吸不仅受代谢需求支配,也受情绪影响。在睡眠医学领域,阻塞性睡眠呼吸暂停容易诱发儿童行为问题^[1],也会导致高血压及其他心血管疾病风险^[2-3]。大脑中与呼吸同步的活动也能够调节认知功能,而这种调节取决于呼吸周期的相位特性^[4]。因此,持续性的呼吸监测能为情绪、应激反应及整体健康状态评估提供精准且关键的量化依据。

传统的呼吸监测方法,如胸带式或腹带式传感器,需要与被测者身体紧密接触,这会在一定程度上干扰被测者的正常活动^[5]。非接触式的方法可以实现长期的无感监测。雷达传感器通过发射回波信号来检测呼吸过程引起的胸部位移,并分析接收信号的变化形式,从而有望提供连续和长期非接触式呼吸监测^[6-7]。

近年来,随着热成像(thermal infrared, IR)技术的发展,研究人员提出使用热成像视频来进行非接触式呼吸监测^[8-9]。人体核心温度大于面温,环境温度通常小于面温,在呼吸时会导致口鼻处温度变化,热成像技术可以通过非接触探测红外能量,并将其转换为电信号。尽管在大部分自然场景下,热成像视频难以获取绝对准确的温度信息,但却能够获取温度变化的趋势。研究者通常使用红外热像仪采集人脸图像序列,然后使用关键点跟踪模型在运动图像序列中标记感兴趣区域(region of interest, ROI),得到口鼻下区域温度趋势序列,滤波得到呼吸波形^[10-12]。这种方式对跟踪模型的性能要求高,由于缺少专用于热红外人脸关键点的数据集,还需要手工进行大量精确样本标注。此外,温度与呼吸幅度并不呈现线性关系:在呼吸行为的变化临界处会呈现线性温度变化;但是在持续吸气或吐气的过程中,口鼻处会趋于某个温度且保持近似恒温趋势,此时温度变化幅度与真实呼吸幅度无关。

面对上述挑战,本研究提出一种基于少样本粗标注的方法,在RGB图像上完成对口鼻下区的ROI检测跟踪,

进而点对点映射到红外对应的区域。此外,本研究提出了一种基于掩码自编码器预训练的深度学习方法,首次通过重建时空图(spatial temporal map, STMap)特征,从而获取热成像中的深度表征,主动学习ROI趋势特征,进而能够准确持续进行呼吸监测。本研究的代码可通过<https://github.com/mraz-mrs/Contactless-Rsp>获取。

1 资料与方法

1.1 数据集

本研究收集了27例20~40岁研究对象的头部运动视频,男女比例为5:4。采集的视频图像数据包括普通摄像头、热红外、三维,生理数据包括呼吸带、脉搏、心电、眼动。其中生理数据采样率为2000 Hz,热红外视频和普通摄像头帧率分别为25 FPS和30 FPS。实验包含六个阶段:静坐,演讲,慢速水平摇头,快速水平摇头,快速俯仰摇头,随机方位摇头。每阶段持续时间为30 s。为了确保实验的可重复性,在演讲阶段,被试者需要根据固定问题回答;在头动阶段,被试者需要面对屏幕,根据屏幕前的小球运动调整头部角度。排除明显受到伪影干扰、数据缺失的数据之后,把每条数据切分为6 s片段,总共纳入了206条片段。按照6:2:2的比例将数据划分为训练集、测试集和验证集。本研究遵循知情同意原则,并经四川大学华西医院生物医学伦理委员会审查批准,批准号为2025年审(1136)号,批件有效期一年。

1.2 整体框架

模型架构如图1所示。本研究首先利用普通摄像头图像,构造仿射变换矩阵,从而能够将从自然图像中提取的关键点映射至热红外图像中。之后,进行掩码自编码器(masked autoencoder)^[13-14]预训练,针对热成像视频的成像特点,将其构造为STMap,利用掩码自编码器进行预训练,获取深度表征后,使用全连接层进行监督微调。在本研究方法中,使用的encoder为8层的Attention Block,

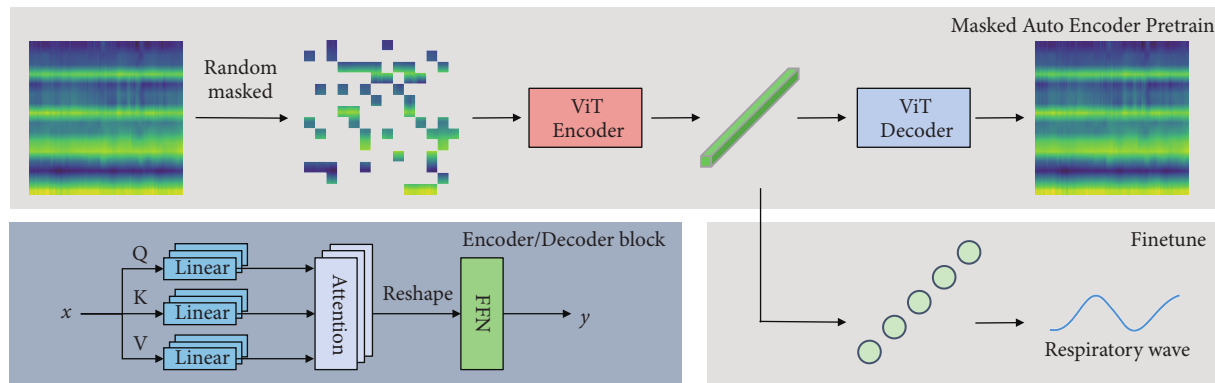


图1 两阶段模型架构图,包括掩码自编码器预训练和微调

Fig 1 Two stage Framework, including masked autoencoder pretrain and finetune

decoder为6层Attention Block。

1.3 少样本粗标注ROI跟踪方法

RGB图像中人脸关键点跟踪的方法已成熟。考虑到普通摄像头和热成像摄像头内参不同,且人脸视频中摄像头位置固定,人脸位置移动小,本研究通过少量标注RGB-IR人脸关键点对,进而通过最小二乘法求解仿射变换矩阵。

$$\begin{bmatrix} P_{x,IR} & P_{y,IR} \end{bmatrix} = \begin{bmatrix} P_{x,RGB} & P_{y,RGB} & 1 \end{bmatrix} \times A \quad (1)$$

$$A = \begin{bmatrix} sc & -ss \\ ss & sc \\ tx & ty \end{bmatrix}$$

其中, A 为变换矩阵, sc 、 ss 、 tx 、 ty 分别表示缩放、旋转、平移参数, $P_{i,j}$ 为来自 $j \in \{RGB, IR\}$ 的坐标值 $i \in \{x, y\}$ 。对于每段视频, 只需对第一帧锚定 $n > 4$ 个RGB-IR点对。由于关键点对数量 n 大于 A 中的未知参数, 该方程组为超定系统, 使用最小二乘法拟合最优解, 即寻找使所有点对映射误差平方和最小的变换矩阵。本研究选取眉心、左面颊、右面颊、鼻尖、下巴共5个面部特征点作为RGB-IR锚定点对, 这部分特征点在头部多方位运动时空间位置相对稳定, 且避开面部边缘像素, 能有效降低摄像头畸变影响; 同时5个点的数量满足超定系统要求, 可通过最小二乘法拟合得到全局最优变换矩阵。

1.4 STMap构造

构造了仿射变换矩阵后, Mediapipe^[15]跟踪面部关键点, 从而得到热成像面部关键点序列。取鼻尖、下巴、左嘴角、右嘴角的外接矩形框, 得到序列 $X_{raw} \in R^{C \times T \times W \times H}$ 。将序列 X_{raw} 分块得 $X_{patch} \in R^{C \times T \times N_s \times W/N_s \times H/N_s}$, 进一步对每个块取均值得 $X_{STMap} \in R^{C \times T \times N_s^2}$, 三次样条插值后得 $X \in R^{C \times T \times T}$ 。其中, $C = 1$ 为热成像视频通道数, $T = 224$ 为视频帧数, $N_s = 4$ 为STMap分块的数量。STMap将原始视频中口鼻区域的时空信息压缩到了一张二维图上, 其中横纵坐标分别表示时间和空间位置, 像素值表示该时空位置的温度值。

1.5 掩码自编码器预训练

重建新的STMap是确保编码器学习到呼吸信号的周期性特征。进行掩码自编码器预训练时, 将输入STMap X 划分为 $N_p \times N_p$ 个块(patch), 经过embedding层后, 对这些块随机比例掩码, 比例为 r , 得 $X_{unmasked} \in R^{C \times N \times d}$ 。其中, $N = [N_p \times N_p \times (1-r)]$, d 为模型的维度。Vit Encoder和Vit Decoder分别由8个和6个Multi-Head Self-Attention Block(MHSA)堆叠而成。

$$Z = \text{Encoder}(X_{unmasked}) \quad (2)$$

$$\hat{X} = \text{Concat}(\text{Decoder}(Z), X_{masked}) \quad (3)$$

其中, $X = [X_{unmasked}, X_{masked}]$, $Z \in R^{C \times N \times d}$ 。像素损失函数仅在被掩码的像素空间内, 计算重建图像与原始图像之间的均

方误差(mean square error, MSE)。预训练的损失函数如下:

$$L_{pixel} = \text{MSE}(X_{unmasked}, \hat{X}_{unmasked}) \quad (4)$$

其中, $\text{MSE}(\cdot)$ 表示均方根误差, $\hat{X}_{unmasked} \in R^{C \times N \times d}$ 表示重建的未掩码的STMap。仅对掩码块计算损失能够迫使模型从稀疏上下文推理全局语义。若对所有块计算损失, 模型会优先拟合可见块, 掩码重建任务被弱化, 无法学到全局结构与语义关联。

1.6 下游微调

预测呼吸波形阶段由两部分构成: 一是在重建任务中预训练完成的编码器, 二是呼吸预测头, 采用一个线性层作为预测头, 用于预测呼吸信号。

$$Y_{pred} = \text{Linear}(Z) \quad (5)$$

其中, $Y_{pred} \in R^T$ 是模型输出的呼吸波形预测结果。微调阶段的目标是将预训练好的表征用于具体任务, 输入全部token, 让模型基于完整输入提取特征并完成下游任务。损失函数如下:

$$L_{MSE} = \text{MSE}(Y_{gt}, Y_{pred}) \quad (6)$$

其中, $Y_{gt} \in R^T$ 是呼吸波形标签。

1.7 实验参数

训练分为两阶段: 预训练与微调。数据集按照6:2:2划分为训练集、测试集与验证集。共纳入27名被试, 其中固定5名被试构成独立验证集; 剩余22名被试通过随机抽样, 分别选取18名和4名被试, 构建训练集与测试集, 随机重复采样五次。

预训练阶段: 模型训练总轮次为2000, 初始学习率设置为0.001, 批次大小为32, 学习率调度策略采用余弦退火调度器(cosine annealing scheduler)。

微调阶段: 为适配下游任务, 模型训练总轮次延长至500, 初始学习率调整为0.009, 批次大小为32, 采用余弦退火调度器控制学习率衰减。架构上, 移除预训练阶段用于掩码重建的解码器, 替换为单层线性层作为任务头, 以完成下游任务的特征映射与输出。

1.8 评价指标

针对模型输出波形与真实参考波形的频率特征一致性, 采用快速傅里叶变换将时域波形转换至频域, 提取频域信号中的峰值成分。此外, 为了更好应用于临床研究, 本研究首次提出了气体进出量比例和呼吸时长平均绝对误差, 评估了进气量、出气量、吸气时间和呼气时间与真实值的差异。指标计算规则如下:

$$MAE = \frac{1}{B} \sum_{i=1}^B |A_{pred,i} - A_{gt,i}| \quad (7)$$

$$RMSE = \sqrt{\frac{1}{B} \sum_{i=1}^B (A_{pred,i} - A_{gt,i})^2} \quad (8)$$

$$PCC = \frac{1}{B} \sum_{i=1}^B \frac{Cov(Y_{gt,i}, Y_{pred,i})}{\sigma_{Y_{gt,i}} \cdot \sigma_{Y_{pred,i}}} \quad (9)$$

$$R_j = \frac{|V_{j,pred} - V_{j,gt}|}{V_{j,gt}} \quad (10)$$

$$D_{MAE,j} = |D_{j,pred} - D_{j,gt}| \quad (11)$$

其中, B 为批次大小, $A_{pred,i}$ 和 $A_{gt,i}$ 分别表示批次中第 i 个样本的频谱峰值的估计值和真实值, $Cov(\cdot)$ 表示协方差, PCC 为皮尔逊相关系数 (Pearson correlation coefficient), $j \in \{inspiratory, expiration\}$, V_j 表示气体体积, $R_j/-$ 和 $D_{MAE,j}/s$ 分别表示进出气量比和呼吸时长平均绝对误差。进出气量比与呼吸时长平均绝对误差可直接量化呼吸通气效率与呼吸节律特征, 能评估肺泡残留气量及

呼吸的异常趋势。

2 结果

2.1 对比实验

本研究纳入了 4 种流行的方法。其中 Filter 方法是在验证集上使用分块数量为 3×3 的 STMap 的中央块序列进行二阶 Butterworth 带通滤波, 通带范围为 $0.15 \sim 0.50$ Hz, 其余 3 种是常见的深度学习模型^[16-18]。结果如表 1 所示。本研究提出方法的平均绝对误差 (mean absolute error, MAE) 为 0.07 ± 0.02 , 均方根误差 (root mean square error, RMSE) 为 0.69 ± 0.11 , PCC 为 0.15 ± 0.04 , 进出气量比分别为 0.04 ± 0.12 和 0.26 ± 0.05 , 呼吸时长平均绝对误差分别为 (0.79 ± 0.19) s 和 (0.79 ± 0.10) s。图 2 展示了所提出方法在

表 1 对比实验

Table 1 Comparative experiment

Model	MAE	RMSE	PCC	R_{in}	R_{ex}	D_{in}/s	D_{ex}/s
Filter	0.10	1.26	0.10	0.76	1.52	0.58	0.79
ResNet 18 ^[15]	0.07 ± 0.01	0.68 ± 0.05	0.04 ± 0.04	0.33 ± 0.07	0.30 ± 0.16	0.66 ± 0.10	0.63 ± 0.06
ConvNeXt ^[16]	0.07 ± 0.01	0.70 ± 0.01	0.01 ± 0.04	0.34 ± 0.13	0.28 ± 0.09	0.62 ± 0.22	0.79 ± 0.36
ViT ^[17]	0.07 ± 0.02	0.75 ± 0.10	0.07 ± 0.07	0.40 ± 0.08	0.29 ± 0.09	0.66 ± 0.14	0.93 ± 0.27
Ours	0.07 ± 0.02	0.69 ± 0.11	0.15 ± 0.04	0.40 ± 0.12	0.26 ± 0.05	0.79 ± 0.19	0.79 ± 0.10

MAE: mean absolute error; RMSE: root mean square error; PCC: Pearson correlation coefficient; R_{in} : ratios of inspiratory volume; R_{ex} : ratios of expiratory volume; D_{in} : mean absolute errors of inspiratory duration; D_{ex} : mean absolute errors of expiratory duration.

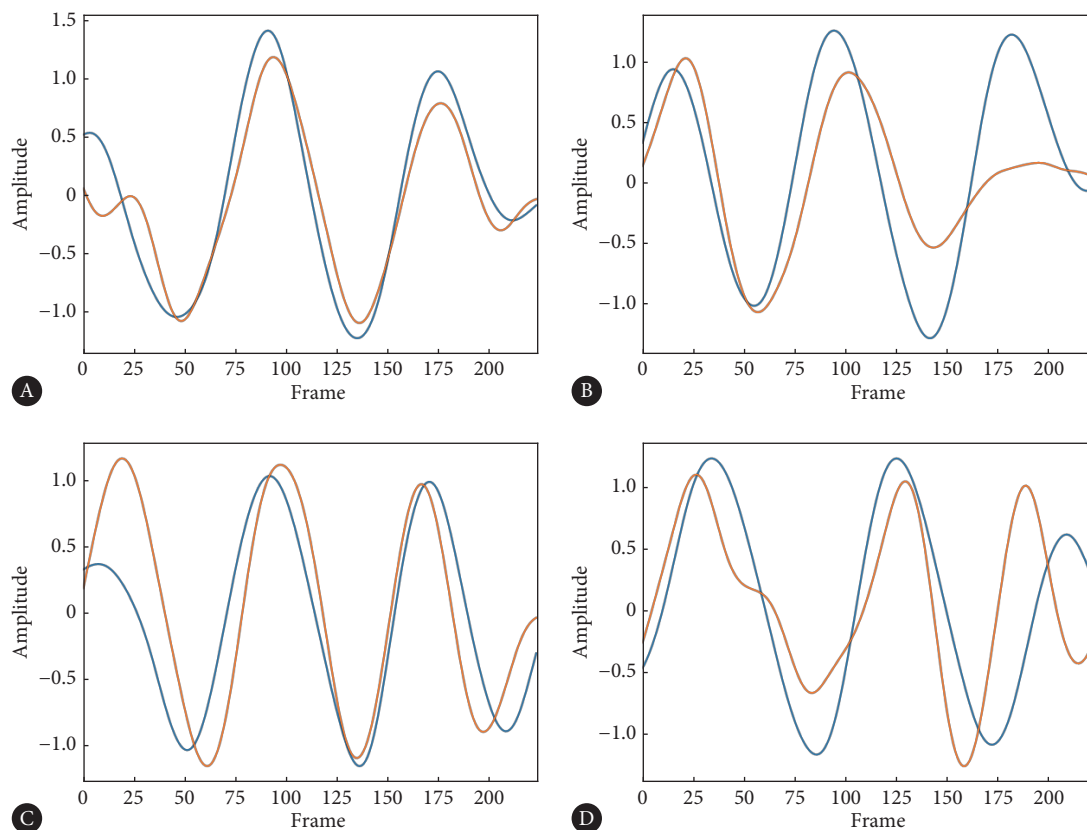


图 2 输出可视化

Fig 2 Visualization

A, Segment 1; B, segment 2; C, segment 3; D, segment 4. The orange and blue waveform are ground truth and prediction respectively

一段热红外视频中的推理结果。

实验结果表明传统带通滤波方法在波形任务中缺乏对复杂特征的学习能力,无法精准拟合波形的细节变化,尤其是RMSE和进出气量比指标方面显著差于深度学习方法。本方法通过掩码自编码预训练学习STMap的深度表征,实现了对波形深层特征的自动挖掘;而其他深度学习模型缺乏预训练阶段的特征迁移,难以学习到鲁棒特征表示,导致性能损失。

2.2 消融实验

本研究采用消融实验来验证所提出的基于掩码自编码器预训练方法的有效性。为明确掩码比例与STMap分块数量对模型性能的影响,本研究测试了不同掩码比例 r 与不同分块 N_s 下的模型性能,结果如图3所示。

结果表明,MAE受到的影响较小,而PCC对于参数变化更加敏感。当 $r = 90\%$, $N_s = 4$ 时,模型性能达到最优。一方面,STMap是根据三次样条插值而得,更大的掩码比例能够使得模型更加聚焦于捕获温度趋势特征而非拟合插值信息;另一方面,合适的分块数量既能够保留特征局部的完整性,也能够通过块均值计算降低噪声影响。

此外,本研究分别使用3种模型,包括完全端到端训练的模型、预训练-微调-冻结Encoder和预训练-微调。如表2所示,掩码自编码器预训练后微调的模型在RMSE和PCC上显著优于完全从头训练的模型,而其余指标均差异不大。冻结Encoder后,PCC显著恶化,说明预训练的Encoder虽然学到了通用特征,但下游波形任务仍需对Encoder进行微调以适配任务特异性需求。若冻结Encoder,模型无法根据下游任务的信号特点调整特征表示,导致误差剧增、波形相关性骤降。时域指标在远程呼吸监测中很难优化,主要原因可能是在吸气或者呼气过程中,温度趋于不变,而在吸气呼气的过渡阶段,温度变化幅度剧烈。通过两阶段学习,模型能够更好地捕获细微的温度变化进而更好在时域中拟合呼吸波形。

2.3 呼吸阶段对脉搏信号的调控作用研究

对呼吸两阶段下的脉搏信号形态特征差异展开研

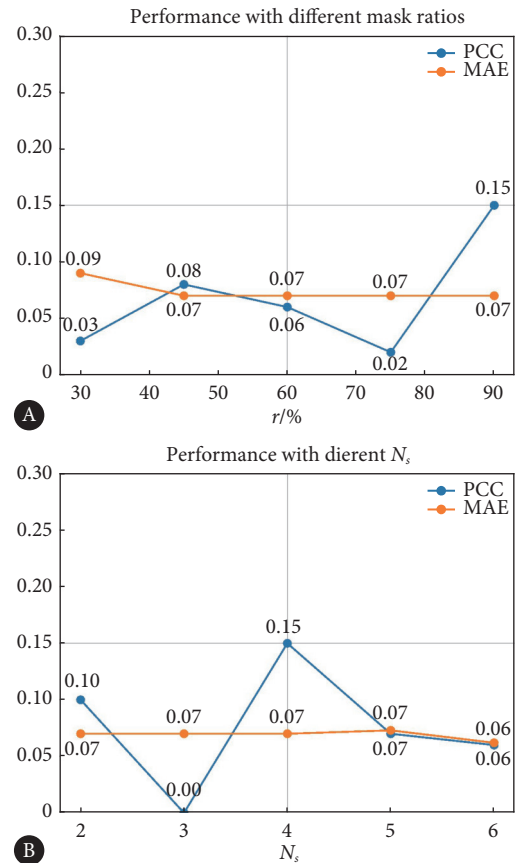


图3 不同设置下的模型性能

Fig 3 Model performance under different settings

PCC: Pearson correlation coefficient; MAE: mean absolute error. A, Varying mask ratios r with fixed $N_s = 4$; B, varying block number N_s with fixed $r = 0.9$.

究。脉搏信号使用光电容积脉搏波法(photoplethysmography, PPG)通过监测血氧饱和度变化测量。考虑到进行相关性分析时需要高质量的生理信号样本,本研究排除16个信号失真、运动伪影导致信号异常的样本,纳入了11例被试的高质量样本数据。对于每段呼吸-脉搏同步样本,使用NeuroKit2^[19]拆分呼气 and 吸气片段,计算脉搏特征,再进行样本级平均。Wilcoxon配对秩和检验结果如表3所示。图4展示了呼吸阶段的PPG片段。

结果表明,呼气 and 吸气阶段,导致脉搏信号在形态学

表2 模块消融实验

Table 2 Ablation experiment of modules

Model			MAE	RMSE	PCC	R_{in}	R_{ex}	D_{in}	D_{ex}
Pretrain	Finetune	Freeze Encoder							
×	×	×	0.07 ± 0.02	0.75 ± 0.10	0.07 ± 0.07	0.40 ± 0.08	0.29 ± 0.09	0.66 ± 0.14	0.93 ± 0.27
√	√	√	0.07 ± 0.01	0.70 ± 0.02	0.07 ± 0.03	0.38 ± 0.21	0.28 ± 0.08	0.64 ± 0.26	0.92 ± 0.14
√	√	×	0.07 ± 0.02	0.69 ± 0.11	0.15 ± 0.04	0.40 ± 0.12	0.26 ± 0.05	0.79 ± 0.19	0.79 ± 0.10

MAE: mean absolute error; RMSE: root mean square error; PCC: Pearson correlation coefficient; R_{in} : ratio errors of inspiratory volume; R_{ex} : ratio errors of expiratory volume; D_{in} : mean absolute errors of inspiratory duration; D_{ex} : mean absolute errors of expiratory duration.

表 3 呼气阶段和吸气阶段的脉搏信号Wilcoxon配对秩和检验结果

Table 3 3 Results of the Wilcoxon matched-pairs signed-rank test for PPG signals during the expiratory phase and inspiratory phase

Variable	Mean (in)	Mean (ex)	Diff-mean (ex-in) (95% CI)	Statistics	P
Entropy	-1.11	-1.13	0.09 (-0.02, 0.19)	0	< 0.001
Heart rate power	0.47	0.81	0.34 (0.16, 0.54)	0	< 0.001
Kurtosis	-0.57	-0.85	-0.28 (-0.68, 0.10)	0	< 0.001
Max	1.89	1.97	0.08 (-0.08, 0.25)	3	< 0.01
Mean	-0.03	0.24	0.06 (0.04, 0.08)	1	< 0.01
Min	-1.20	-1.30	-0.11 (-0.26, 0.04)	1	< 0.01
Peak_duration_mean	0.16	0.18	0.02 (-0.02, 0.05)	14	0.10
Peak_height_mean	2.38	2.59	0.21 (-0.16, 0.59)	10	< 0.05
PPG_amp_max	3.25	3.30	0.05 (-0.07, 0.18)	23	0.41
PPG_amp_mean	2.46	2.65	0.19 (-0.19, 0.59)	14	< 0.05
PPG_amp_min	1.60	1.73	0.13 (-0.68, 0.90)	29	0.76
PPG_amp_std	0.74	0.66	-0.08 (-0.45, 0.32)	30	0.83
Range	3.08	3.27	0.19 (0.08, 0.29)	0	< 0.001
Rise_slope_mean	6.22	6.57	0.35 (-0.60, 1.28)	19	0.24
Skewness	0.67	0.50	-0.16 (-0.45, 0.13)	3	< 0.01
Std	0.93	0.98	0.05 (0.02, 0.08)	1	< 0.01
Sys_peak_mean	1.27	1.46	0.18 (-0.09, 0.45)	7	< 0.05
Sys_peak_std	0.71	0.56	-0.13 (-0.52, 0.24)	28	0.70
Var	0.91	1.01	0.10 (0.06, 0.14)	1	< 0.01

Mean (in): mean value of the feature in the inspiratory phase; Mean (ex): mean value of the feature in the expiratory phase; Diff-mean (ex-in): mean of per-sample differences between expiratory and inspiratory feature values; Statistics: Wilcoxon signed-rank test statistic, the minimum of the sum of positive and negative signed ranks of expiratory-inspiratory feature differences.

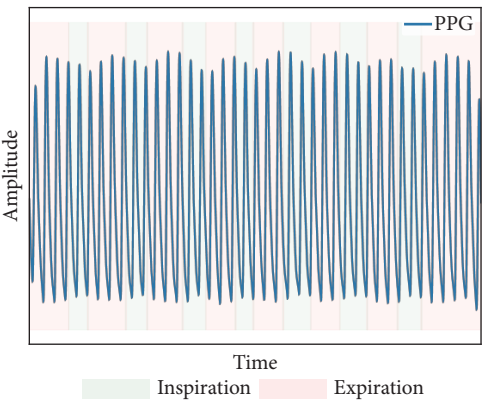


图 4 根据呼吸阶段裁剪PPG片段
Fig 4 Clip PPG segments according to respiratory phases
PPG: photoplethysmography.

上有较大差异。主要体现在呼气阶段导致范围(Range)扩大,峰值(Max)上升,谷值(Min)下降,功率(Heart rate power)上升。呼气阶段PPG信号峰度(Kurtosis)显著低于吸气阶段,提示呼气时脉搏波形态更平坦、分布更分

散。偏度(Skewness)在呼气时亦显著降低,表明脉搏波向左偏移,即上升支变陡、下降支相对延长。

3 讨论

本研究的核心创新在于将计算机视觉领域的掩码自编码器自监督预训练范式迁移至波形任务,并通过预训练学习STMap深度表征微调适配任务特异性的流程,解决了传统方法与端到端深度学习模型在波形特征挖掘上的不足。这一尝试不仅为呼吸监测类任务提供了新的技术路径,更验证了自监督学习在跨模态特征迁移中的潜力。自监督预训练学习到的全局局部特征关联能力能有效适配呼吸监测任务,为后续跨模态自监督研究提供了参考。

在评估指标方面,本研究及各类基线模型在频域特征建模方面均表现良好,然而在时域特征PCC评估中包括基线模型在内普遍表现欠佳。可能的原因在于两方面:①口鼻区域的热辐射动态具有显著的阶段性特征:仅

在呼吸相转换节点呈现剧烈变化,而在持续的吸气或呼气平台期则趋于稳态或仅表现为微幅波动,即STMap像素强度与呼吸信号之间在呼吸交界点处存在线性映射关系,而在呼吸持续过程中则呈现非线性。该内在属性使得模型易于捕获具有周期性的频域节律特征,却难以在时域特征上优化。②PCC在评估周期性生理信号的时域一致性时,对相位偏移表现出高度敏感性。即使模型生成的信号在频率与振幅上与真实信号高度吻合,微小的时序相位差亦会导致PCC值显著衰减,存在评估偏差。此外,本研究在进出气量比和呼吸时长平均绝对误差两项新指标表现优异,在临床监测中提前预警COPD、肺大泡等呼吸系统疾病^[20-21]和情绪、应激状态评估^[22]有潜在重要意义。此外,研究表明PPG波形形态的变化也被用于心血管疾病的早期筛查^[23-24]。因此,本研究发现的呼吸调制下的PPG形态差异,可能为从形态学维度评估心肺耦合功能提供了新的量化视角,这对于开发集成式可穿戴健康监测系统具有潜在意义。

本研究仍存在一些局限性。首先,STMap构造步骤较为复杂且长度固定,导致难以进行动态地监测。传统卷积神经网络复杂度较低,但通常输入的尺寸固定,这一限制源于原始架构末尾的全连接层。如果加上全局平均池化可以缓解,但是会损失细粒度的特征信息。尽管Transformer系列模型具有 $O(N^2)$ 复杂度,但在前向传播时,注意力层和FFN层本身并不会对序列长度进行变换,因此具有很好的外推性潜力。在未来的研究中,可以考虑将STMap的时间维度建模为序列长度,以适配不同的视频长度增强其外推性,实现动态长度输入。此外,本研究纳入的数据量有限,场景较为单一,需要考虑到复杂自然场景下的呼吸监测任务,例如睡眠监测、活体检测等。在探索呼吸阶段对脉搏形态特征的调控作用时,对多数指标的评价仅依赖点估计结果的直观对比,以此判定组间差异,导致结论易受随机波动干扰,严谨性不足。从统计学角度,需经更规范的假设检验与显著性分析验证。

* * *

作者贡献声明 毛如石负责论文构思、正式分析、调查研究、研究方法和初稿写作,吕学斌负责数据审编、研究方法、验证和审读与编辑写作,张伟负责经费获取、研究项目管理、监督指导和审读与编辑写作。所有作者已经同意将文章提交给本刊,且对将要发表版本进行最终定稿,并同意对工作的所有方面负责。

Author Contribution MAO Rushi is responsible for conceptualization, formal analysis, investigation, methodology, and writing--original draft. LÜ Xuebin is responsible for data curation, methodology, validation, and writing--review and editing. ZHANG Wei is responsible for funding acquisition, project administration, supervision, and writing--review and

editing. All authors consented to the submission of the article to the Journal. All authors approved the final version to be published and agreed to take responsibility for all aspects of the work.

利益冲突 本文作者张伟是本刊编委会编委。该文在编辑评审过程中所有流程严格按照期刊政策进行,且未经其本人经手处理。除此之外,所有作者均声明不存在利益冲突。

Declaration of Conflicting Interests ZHANG Wei is a member of the Editorial Board of the journal. All processes involved in the editing and reviewing of this article were carried out in strict compliance with the journal's policies and there was no inappropriate personal involvement by the author. Other than this, all authors declare no competing interests.

参 考 文 献

- [1] 赵昕,李婉妹,周锦,等.阻塞性睡眠呼吸暂停与儿童行为问题的关联研究. *四川大学学报(医学版)*, 2025, 56(5): 1373-1379. doi: [10.12182/20250960101](#).
ZHAO X, LI W S, ZHOU J, *et al.* Association between obstructive sleep apnea and behavioral problems in children. *J Sichuan Univ (Med Sci)*, 2025, 56(5): 1373-1379. doi: [10.12182/20250960101](#).
- [2] 李小英,张晨旭,曾萍,等.阻塞性睡眠呼吸暂停综合征对无心脑血管疾病老年人夜间动态血压的影响. *四川大学学报(医学版)*, 2025, 56(2): 464-469. doi: [10.12182/20250360110](#).
LI X Y, ZHANG C X, ZENG P, *et al.* Effect of obstructive sleep apnea syndrome on nocturnal ambulatory blood pressure monitoring results in older adults without cardiovascular or cerebrovascular diseases. *J Sichuan Univ (Med Sci)*, 2025, 56(2): 464-469. doi: [10.12182/20250360110](#).
- [3] 黄通达,植良娥,税荣春,等.成人阻塞性睡眠呼吸暂停相关性心血管事件的研究进展. *实用医院临床杂志*, 2025, 22(1): 190-192. doi: [10.3969/j.issn.1672-6170.2025.01.042](#).
HUANG T D, ZHI L E, SHUI R C, *et al.* Research progress of cardiovascular events associated with obstructive sleep apnea in adults. *Practical Journal of Clinical Medicine*, 2025, 22(1): 190-192. doi: [10.3969/j.issn.1672-6170.2025.01.042](#).
- [4] NAKAMURA N H, FUKUNAGA M, OKU Y. Respiratory modulation of cognitive performance during the retrieval process. *PLoS One*, 2018, 13(9): e0204021. doi: [10.1371/journal.pone.0204021](#).
- [5] LIU S, GAO R X, FREEDSON P S. Non-invasive respiration and ventilation prediction using a single abdominal sensor belt//*IEEE Signal Processing in Medicine and Biology Symposium*, 2011: 1-5. doi: [10.1109/SPMB.2011.6120113](#).
- [6] WANG T, SUN W, SHIMA K. Continuous body movement-robust respiration rate detection using frequency-modulated continuous wave millimeter wave radar based on non-negative matrix factorization algorithm. *Biomedical Signal Processing and Control*, 2025, 102: 107243. doi: [10.1016/j.bspc.2024.107243](#).
- [7] TESTA O, CICCHETTI R, PISA S, *et al.* A novel FMCW radar scheme with millimeter motion detection capabilities suitable for cardio-respiratory monitoring. *Sensors*, 2025, 25(9): 2765. doi: [10.3390/s25092765](#).
- [8] GIOIA F, PURA F, GRECO A, *et al.* Contactless estimation of respiratory frequency using 3D-CNN on thermal images. *IEEE Journal of Biomedical and Health Informatics*, 2025, 29(10): 7387-7395. doi: [10.1109/JBHI.2025.3567141](#).
- [9] KHEDDAR H, HIMEUR Y, AMIRA A. BreathAI: transfer learning-based thermal imaging for automated breathing pattern recognition. *IEEE International Conference on Image Processing (ICIP)*, 2025: 2612-2617. doi: [10.1109/ICIP55913.2025.11084681](#).
- [10] ALVES R, Van MEULEN F, OVEREEM S, *et al.* Thermal cameras for continuous and contactless respiration monitoring. *Sensors*, 2024, 24(24): 8118. doi: [10.3390/s24248118](#).
- [11] JAGADEV P, NAIK S, INDU GIRI L. Contactless monitoring of human respiration using infrared thermography and deep learning. *Physiol Meas*,

- 2022, 43(2): 025006. doi: [10.1088/1361-6579/ac57a8](https://doi.org/10.1088/1361-6579/ac57a8).
- [12] VITAZKOVA D, FOLTAN E, KOSNACOVA H, *et al.* Advances in respiratory monitoring: a comprehensive review of wearable and remote technologies. *Biosensors*, 2024, 14(2): 90. doi: [10.3390/bios14020090](https://doi.org/10.3390/bios14020090).
- [13] HE K, CHEN X, XIE S, *et al.* Masked autoencoders are scalable vision learners. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022: 16000-16009. doi: [10.1109/CVPR52688.2022.01553](https://doi.org/10.1109/CVPR52688.2022.01553).
- [14] LIU X, ZHANG Y, YU Z, *et al.* RPPG-mae: self-supervised pretraining with masked autoencoders for remote physiological measurements. *IEEE Transactions on Multimedia*, 2024, 26: 7278-7293. doi: [10.1109/TMM.2024.3363660](https://doi.org/10.1109/TMM.2024.3363660).
- [15] LUGARESI C, TANG J, NASH H, *et al.* Mediapipe: a framework for building perception pipelines. ArXiv, 2019: arXiv:1906.08172. doi: [10.48550/arXiv.1906.08172](https://doi.org/10.48550/arXiv.1906.08172).
- [16] HE K, ZHANG X, REN S, *et al.* Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 770-778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [17] LIU Z, MAO H, WU C Y, *et al.* A ConvNet for the 2020s. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 11976-11986. doi: [10.1109/CVPR52688.2022.01167](https://doi.org/10.1109/CVPR52688.2022.01167).
- [18] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al.* An image is worth 16x16 words: transformers for image recognition at scale. ArXiv, 2021: arXiv:2010.11929. doi: [10.48550/arXiv.2010.11929](https://doi.org/10.48550/arXiv.2010.11929).
- [19] MAKOWSKI D, PHAM T, LAU Z J, *et al.* NeuroKit2: a Python toolbox for neurophysiological signal processing. *Behav Res Methods*, 2021, 53(4): 1689-1696. doi: [10.3758/s13428-020-01516-y](https://doi.org/10.3758/s13428-020-01516-y).
- [20] GALDI F, PEDONE C, MCGEE C A, *et al.* Inhaled high molecular weight hyaluronan ameliorates respiratory failure in acute COPD exacerbation: a pilot study. *Respir Res*, 2021, 22(1): 30. doi: [10.1186/s12931-020-01610-x](https://doi.org/10.1186/s12931-020-01610-x).
- [21] TAN D, WANG B, CAO P, *et al.* High flow nasal cannula oxygen therapy versus non-invasive ventilation for acute exacerbations of chronic obstructive pulmonary disease with acute-moderate hypercapnic respiratory failure: a randomized controlled non-inferiority trial. *Crit Care*, 2024, 28(1): 250. doi: [10.1186/s13054-024-05040-9](https://doi.org/10.1186/s13054-024-05040-9).
- [22] JERATH R, BEVERIDGE C. Respiratory rhythm, autonomic modulation, and the spectrum of emotions: the future of emotion recognition and modulation. *Front Psychol*, 2020, 11: 1980. doi: [10.3389/fpsyg.2020.01980](https://doi.org/10.3389/fpsyg.2020.01980).
- [23] PATEL A, ANUSSHREE R, SWARNA M, *et al.* Detection of atrial fibrillation using PPG signals. 2024 IEEE 21st India Council International Conference (INDICON), 2024: 1-6. doi: [10.1109/INDICON63790.2024.10958406](https://doi.org/10.1109/INDICON63790.2024.10958406).
- [24] ABRISHAM K P, ALIPOUR K, TARVIRDIZADEH B, *et al.* Noninvasive assessment of arterial stiffness using photoplethysmography: feature analysis and machine learning-based estimation of carotid-femoral pulse wave velocity. *IEEE Access*, 2025(13): 185657-185673. doi: [10.1109/ACCESS.2025.3626252](https://doi.org/10.1109/ACCESS.2025.3626252).

(2025-11-20 收稿, 2026-03-17 修回)

编辑 余琳



开放获取 本文使用遵循知识共享署名—非商业性使用 4.0 国际许可协议 (CC BY-NC 4.0), 详细信息请访问

<https://creativecommons.org/licenses/by-nc/4.0/>。

OPEN ACCESS This article is licensed for use under Creative Commons Attribution-NonCommercial 4.0 International license (CC BY-NC 4.0). For more information, visit <https://creativecommons.org/licenses/by-nc/4.0/>.

© 2026 《四川大学学报(医学版)》编辑部

Editorial Office of *Journal of Sichuan University (Medical Sciences)*