

采用自注意力抗干扰网络的视频房颤检测

陈静^{1,2}, 杨学志^{2,3}, 陈鲸¹, 刘雪南^{1,2}

- (1. 合肥工业大学计算机与信息学院, 安徽 合肥 230009;
2. 合肥工业大学工业安全与应急技术安徽省重点实验室, 安徽 合肥 230009;
3. 合肥工业大学软件学院, 安徽 合肥 230009)

摘 要: 房颤的早期发现与诊断是降低房颤以及并发症风险的关键。视频光电体积描记术(VPPG)技术为房颤筛查提供了新途径, 但易受到现实场景中运动干扰。现有 VPPG 房颤检测方法存在运动干扰时会造成脉冲信号失真, 从而发生误判。为解决以上问题, 提出一种抗干扰视频房颤检测模型。该模型使用注意力编码器网络, 从包含运动干扰的脉冲信号中提取鲁棒的脉冲潜在特征, 径向基分类网络根据潜在特征实现房颤检测。注意力编码器将复杂脉冲信号映射到高维子空间, 重点关注有效信息, 提取稳健潜在特征。径向基分类网络在房颤标签监督下提高房颤识别能力, 输出可靠结果。在 200 名测试者参与的自建数据集上进行实验, 结果表明该模型在各类场景中均表现优异。在静态场景中, 检测精度较最优对比算法提高了 8.1%, 敏感性提高了 7.5%。在动态场景中, 对比算法准确度均大幅下降, 所提模型精度相比提升了 16.5%, 特异性提升了 18.3%。模型具有良好的抗运动干扰能力, 可有效地消除运动干扰影响, 提高现实场景中视频房颤检测精度。

关 键 词: 视频光电体积描记术; 房颤检测; 抗运动干扰; 注意力编码器; 潜在特征

中图分类号: TP 391

DOI: 10.11996/JGJ.2095-302X.2023020313

文献标识码: A

文章编号: 2095-302X(2023)02-0313-11

Video atrial fibrillation detection using self-attentional anti-interference network

CHEN Jing^{1,2}, YANG Xue-zhi^{2,3}, CHEN Jing¹, LIU Xue-nan^{1,2}

- (1. School of Computer Science and Information Engineering, Hefei University of Technology, Hefei Anhui 230009, China;
2. Anhui Key Laboratory of Industrial Safety and Emergency Technology, Hefei Anhui 230009, China;
3. School of Software, Hefei University of Technology, Hefei Anhui 230009, China)

Abstract: Early detection and diagnosis of atrial fibrillation is the key to reducing the risk of atrial fibrillation and complications. Although the video photoplethysmography (VPPG) technology provides a new approach to atrial fibrillation screening, it is susceptible to motion interference in real-world scenarios. When the existing VPPG atrial fibrillation detection method encounters motion interference, the pulse signal will be distorted, resulting in misjudgment. To solve the above problems, an anti-interference video atrial fibrillation detection model was proposed. The model employed an attention encoder network to extract robust spiking latent features from spiking signals

收稿日期: 2022-06-23; 定稿日期: 2022-08-22

Received: 23 June, 2022; Finalized: 22 August, 2022

基金项目: 安徽省科技重大专项项目(JZ2019AKKZ0469); 安徽高校协同创新项目(PA2019AGXC0114)

Foundation items: Anhui Province Science and Technology Major Special Project (JZ2019AKKZ0469); Anhui University Collaborative Innovation Project (PA2019AGXC0114)

第一作者: 陈静(1997-), 女, 硕士研究生。主要研究方向为视频信号处理、视频医学。E-mail: chen_j_hfut@163.com

First author: CHEN Jing (1997-), master student. Her main research interests cover video signal processing, video medicine.
E-mail: chen_j_hfut@163.com

通信作者: 杨学志(1970-), 男, 教授, 博士。主要研究方向为数字信号处理、计算机视觉。E-mail: xzyang@hfut.edu.cn

Corresponding author: YANG Xue-zhi (1970-), professor, Ph.D. His main research interests cover digital signal processing, computer vision.
E-mail: xzyang@hfut.edu.cn

containing motion disturbances. A radial basis classification network then performed atrial fibrillation detection based on these latent features. The attention encoder mapped complex impulse signals into high-dimensional subspaces, focusing on effective information and extracting robust latent features. Furthermore, Radial Basis Classification Network enhanced atrial fibrillation recognition ability under the supervision of atrial fibrillation labels and output reliable results. Experiments were carried out on a self-built dataset with 200 testers, and the results show that the proposed model performed well in various scenarios. In static scenes, the detection accuracy was 8.1% higher than the optimal comparison algorithm, and the sensitivity was 7.5% higher. In dynamic scenes, where the accuracy of the comparison algorithms was greatly reduced, the accuracy of the proposed model was improved by 16.5%, and the specificity was improved by 18.3%. The model demonstrated good anti-motion interference ability, effectively eliminating the influence of motion interference and improving the detection accuracy of video atrial fibrillation in real scenes.

Keywords: video photoplethysmography; atrial fibrillation detection; anti-motion interference; attention encoder; potential characteristics

房颤是最常见的心率失常, 规则有序的心房电活动消失, 出现快速无序的颤动波, 是严重的心房电活动紊乱。目前我国有超过 1 000 万的房颤患者, 房颤患病率随年纪增长呈上升趋势^[1]。房颤可致中风风险提高 5 倍, 房颤诱发的中风一年内的致死率高达 50%。通常情况, 房颤患者在发病前无任何症状。因此, 房颤的早期发现与诊断是提高治疗质量, 降低房颤及并发症风险的关键。

在临床上, 心电图(electrocardiography, ECG)是目前房颤诊断的金标准。在心电图记录中无明显的重复 P 波, RR 间期(RR interval)不规则, 可诊断为临床房颤。但由于操作复杂、成本高, 很难实现大规模人群的房颤初筛。光电容积描记术(photoplethysmography, PPG)是一种光学心血管监测技术, 在最新的指南^[2]中被推荐为房颤筛查方法之一。近年来, PPG 技术已在智能腕带和手表等智能可穿戴设备上得到应用。华为公司一项基于 PPG 技术的房颤筛查研究^[3]涉及 187 912 人(平均年龄 35 岁, 86.7%为男性), 房颤检出率为 91.6%, 验证了 PPG 手表在长期房颤监测中的可行性, 在主动房颤筛查方面取得了显著的效果。

近几年得益于人工智能和计算机视觉技术的高速发展, 视频光电容积描记术(video photoplethysmography, VPPG)已成为 PPG 技术的前沿^[4]。VPPG 技术使用消费级相机记录视频中的皮肤颜色变化, 从视频中提取脉搏脉冲信号, 已成功用于检测心率^[5]、心率变异性、血氧饱和度等。与 PPG 技术相比, VPPG 技术具有以下优点: ①更舒适: 避免了传感器和皮肤的密切接触, 实现非接触式检测; ②更便捷: 可以在日常移动终端(如智能手机和笔记本电脑)上开发和校准, 可用于长期健康

监测; ③更高效: 每次仅使用一个设备便可对多人进行筛查。因此, 不希望直接接触皮肤(如新生儿筛查、皮肤过敏者)或缺乏检测工具(如户外急救)时, VPPG 技术是首选。

针对 VPPG 脉冲信号的房颤筛查, 研究学者提出多种检测方法。基于 VPPG 脉冲信号的房颤检测可分为 3 部分: ①VPPG 信号预处理; ②特征提取; ③房颤分类。自 VPPG 技术^[6]首次提出并成功应用于面部视频检测心率, 此后出现众多方法提高实际应用中鲁棒性。包括基于时域模型的独立分量分析^[7]、主成分分析^[8]等, 将观察到的颜色轨迹分解为不相关的基向量, 再将具有脉冲特征的向量组合成脉冲信号; 基于频域模型的带通滤波^[9]、小波分解^[10]和经验模态分解^[11]等, 将观察到的颜色轨迹分解成不同频段的分量, 选择位于人体脉搏频率带(0.7~4.0 Hz)内的分量作为脉冲信号^[12]。从视频中提取脉搏信号方法已经得到了很好地研究, 所以提高房颤检测性能的关键步骤是高质量房颤特征提取。

研究表明, VPPG 脉冲信号在房颤期间存在可变的脉冲间隔(inter-pulse intervals, IPIs), 其部分特征与房颤高度相关。常用的特性如下: ①可变性特征: 描述 IPIs 序列的离散度, 包括平均值、标准差(standard deviation, SD)、相邻间期差值的均方根(root mean squared successive difference, RMSSD)和 IPIs 大于 x ms 的差异百分比(percentage of successive normal cardiac interbeat intervals greater than x msec, PNN x); ②不规则特征: 揭示了 IPIs 波动的不可预测性程度, 包括香农熵(shannon entropy, ShanEn)和样本熵(sample entropy, SampEn); ③光谱特征: 描述 IPIs 变化的频率, 包括 0.04~0.15 Hz(LF(Low frequency))之间的总光谱

功率、0.15~0.4 Hz(HF(High frequency))之间的总光谱功率、归一化比值(LF/HF)。

以往房颤检测方法,多数基于 ECG 信号或 PPG 脉冲信号。COUDERC 等^[13]利用脉冲谐波强度(pulse harmonic intensity, PHS)指标度量 PPG 信号变异性,通过分析心率变异性时域指标和 PHS 指标,综合判断房颤结果。CORINO 等^[14]根据样本熵和形状相似度指标进行房颤筛查。CHENG 等^[15]对 PPG 信号进行数据增强后获得时频图,并送入卷积神经网络(convolutional neural networks, CNN)和长短期记忆网络(long short term memory, LSTM)混合网络中进行房颤识别。郭一楠等^[16]对 ECG 信号进行希尔伯特黄变换得到二维时频图,使用 DenseNet 神经网络从时频图中学习房颤相关的信息,以实现分类。KWON 等^[17]提出由一维卷积神经网络(one dimensional convolutional neural network, 1D-CNN)和递归神经网络(recurrent neural networks, RNN)构建深度网络,定义房颤诊断结果置信水平的度量标准,比较真实与错误诊断的置信水平。随着 VPPG 技术方法的不断发展, YAN 等^[18]首次提出基于智能手机的房颤检测方法,从手机摄像头采集的面部视频中提取 VPPG 脉冲信号,利用自相关分析算法计算波形相关性,当出现三次以上连续不规则信号时判定存在房颤。SHI 等^[19]提出多区域特征融合算法,首先提取面部多区域 VPPG 脉冲信号,使用生物医学统计方法针对不同区域提取 IPIs 特征,对多区域特征筛选得到融合特征,再送入分类器中进行房颤分类。以上方法都是通过分析房颤脉冲信号与正常脉冲信号的差异特征进行房颤检测。然而, VPPG 脉冲信号在实际场景中包含干扰噪声,如头部运动将会导致脉冲信号产生类似房颤发作时的不规则波形,对房颤检测具有误导性,影响检测结果。房颤脉冲信号与运动干扰信号之间的相似性使这些方法的应用变得困难。

为了解决上述问题,本文设计了一种自注意力抗干扰房颤分类模型,重点解决现实场景中因运动干扰信号与脉冲信号相似,导致房颤检测困难问题。该模型以自编码器为基本架构,借助自注意力机制强大的信息获取能力,设计了结合时频分析的自注意力编码器房颤分类网络。该网络以 VPPG 脉冲信号的时频图为输入,时频图比一维时域信号蕴含更多有效信息。PPG 脉冲信号抗干扰能力强,在网络训练阶段,同步 PPG 脉冲信号作为自注意力编码器网络标签,提高网络模型抗干扰能力。编码器

得到的稳健潜在特征,再经过房颤结果标签提高特征的房颤识别能力。实验结果表明,在实际应用场景中,该模型抗干扰能力和房颤检测精度有明显上升。

1 自注意力抗干扰房颤分类网络

1.1 视频预处理

视频数据大,直接运算难度大,为加快网络的训练速度和提升训练的精度,需要从面部视频中提取 VPPG 脉冲信号。具体步骤如下:

(1) 人脸检测与跟踪。采用 LI 等^[20]提出的人脸跟踪算法抑制人脸刚性运动。首先检测视频第一帧中的矩形面部区域;然后,使用判别响应图拟合检测视频每帧图像中的面部地标^[21];再通过 KLT (Kanade-lucas-tomasi)算法追踪视频中面部地标。

(2) 目标区域选择。从跟踪的面部区域中选择感兴趣区域(region of interest, ROI),避免不相关区域的干扰。研究表明^[22],面部中脸颊区域检测到的 VPPG 脉冲信号质量最佳,因为该区域包含丰富的毛细血管。ROI 使用脸颊边缘 4 个面部来标记,视频每帧图像 ROI 是大小为 40×120 像素的矩形。

(3) 色度空间转化。将 ROI 色度空间由 RGB 转换到 CHROM,可以增强抗面部运动干扰能力^[23]。每个像素的 RGB 信号被线性组合成 2 个信号, $D=3R-2G$ 和 $E=1.5R+G-1.5B$,其中 R , G 和 B 分别是 RGB 空间的归一化值,组合系数是基于肤色标准化方法得到。信号送入通带频率为 0.8~4.0 Hz 的带通滤波器,滤波可去除与人体脉冲无关的频率分量。最后,滤波后的信号组成原始脉冲信号 $F=D_f-\alpha E_f$,其中 D_f 和 E_f 为经过滤波后的组合信号,其中系数 $\alpha=\sigma(D_f)/\sigma(E_f)$, $\sigma(\cdot)$ 为求标准差。

原始脉冲信号是一维时域信号,只体现信号幅值随时间的变化,而信号频率变化同样包含丰富信息。时频分析将一维时域信号映射到二维时频平面,全面反映信号的时频联合特征^[24]。小波变换^[25]具有多分辨率分析的特点,并且在时频域均具有表征信号局部特征的能力,可聚焦到信号的细节信息。Morlet 小波^[26]在时间与频率的局部化之间有着很好的平衡。

原始脉冲信号经过时频分析,所得到的时频图包含时频域信息。对比图 1(a)和(b)可知,房颤患者的脉搏信号中各频率成分幅值在不同时刻波动明显,该波动和房颤发作时心脏跳动不齐,不规律高度相关。因此,该特征差异有助于房颤分类。

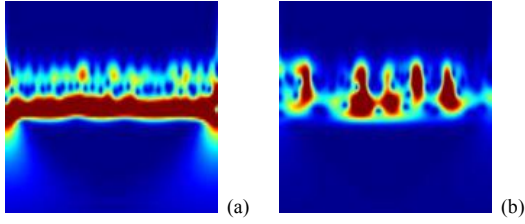


图1 时频图特征(224×224 像素)((a)正常测试者;
(b)房颤患者)

Fig. 1 Time-frequency diagram features (224×224 px)
((a) Normal tester; (b) AF patients)

1.2 抗干扰房颤分类网络

VPPG 脉冲信号在现实场景中易受到面部运动干扰,从面部视频中提取的原始脉冲信号包含脉冲信号和运动信号 2 部分。运动信号产生的不规则波形与房颤波形相似会导致误检,所以需要将原始脉冲信号中的 2 种信号分开。先前的研究表明^[27],自编码器可以从包含噪声的信号中重构高质量信号。自编码器网络包括编码器和解码器,编码器从输入信号中提取重点有效信息,得到潜在特征;解码器根据潜在特征信息重构原信号。可知,潜在特征是输入信号的有效表达。因此,自编码器可作为信号特征提取器,基于重构信号优化过程,得到信号有效潜在特征。

本文模型以自编码器为基本架构,设计了一种抗运动干扰的房颤分类网络,模型结构如图 2 所示。模型由 2 部分组成:基于自编码器的特征提取网络和房颤分类网络。编码器网络,在同步 PPG 脉冲信号监督下得到最优重构信号和鲁棒的潜在特征。编码器优化过程中,编码器不断学习 VPPG 脉冲信号中与 PPG 脉冲信号共有的脉冲信息,实现 VPPG 脉冲信号与运动信号分离。编码器网络优化后,得到潜在特征包含全部脉冲信息,抗运动干扰能力强,可用于房颤检测。

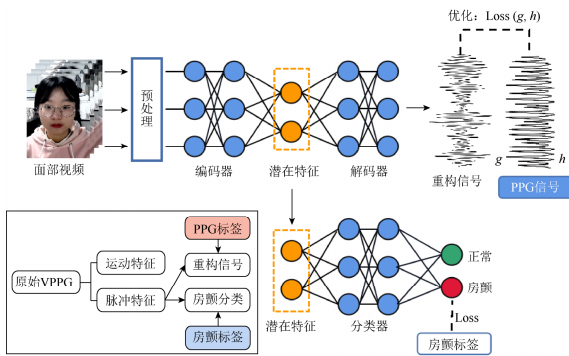


图2 抗干扰房颤分类网络结构

Fig. 2 Anti-interference atrial fibrillation classification network structure

网络优化过程中,误差函数选用均方误差,即

$$Loss_E = \frac{1}{N} \sum_{n=1}^N (g_n - h_n)^2 \quad (1)$$

其中, g_n 为解码器第 n 个重构的信号; h_n 为第 n 个输入信号的同步 PPG 脉冲信号; N 为输入信号总数。同步 PPG 脉冲信号作为标签有以下优点: ①PPG 脉冲信号是由指夹式仪器直接接触测试者得到,信号抗干扰能力强; ②PPG 脉冲信号中脉搏信息更加丰富,包含 VPPG 脉冲信号蕴含的全部脉搏信息。

房颤分类网络中,输入为编码器网络提取的鲁棒脉冲特征,输出为房颤分类结果。分类网络在房颤结果标签监督下,增强潜在特征的房颤识别能力,提高房颤分类精度,实现现实场景中面部视频抗干扰房颤检测。网络优化过程中,损失函数被设计为预测和分类标签之间的交叉熵,即

$$Loss_C = -\frac{1}{N} \sum_{n=1}^N y_n \cdot \log_2[c_w(p_n)] \quad (2)$$

其中, $c_w(\cdot)$ 为分类器的映射函数; y_n 是从临床诊断结果中获得的热编码标签(即非房颤样本为[0 1],房颤样本为[1 0]); p_n 为输入的第 n 个潜在特征。考虑到信号重构任务和房颤分类任务收敛速度不同,采用几何损失策略联合学习,减小损失函数差距过大造成的影响,且同步提高潜在特征的抗运动能力和房颤识别能力

$$Loss = \sqrt{Loss_E} \times \sqrt{Loss_C} \quad (3)$$

本文提出的抗干扰房颤分类网络,编码器网络获取稳健的房颤潜在特征,分类网络准确识别房颤,解决了头部运动造成的房颤特征提取困难、房颤检测精度低问题,提高了现实场景中视频房颤检测可靠性。

1.3 注意力去噪编码器

基于自编码器设计的注意力去噪编码器,具体结构如图 3 所示。编码器由多头注意力层、激活层和线性层组成;解码器组成元素与解码器相同,但连接存在差异,网络各模块功能如下。

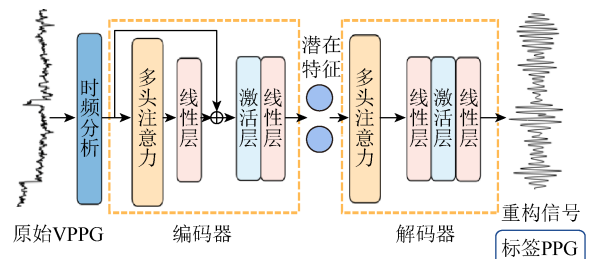


图3 注意力去噪编码器结构

Fig. 3 Attention denoising encoder structure

编码器输入为脉冲信号时频图, 先通过多头注意力层关注多个特征空间信息。线性层确保多头注意力层输出维度与输入维度一致, 保证残差连接及归一化操作。增加残差连接, 可以解决梯度消失和权重矩阵的退化问题。激活函数进行非线性映射, 加强注意力层学习到的重点部分, 抑制无关信息。选用 GELU 激活函数公式为

$$GELU(x) = x \cdot \int_{-\infty}^x \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \quad (4)$$

其中, x 为激活层的输入数据。输出线性层降低维度, 提取重要信息, 降低计算复杂度。通过 2 个线性层先将数据映射到高维空间再映射到低维空间的过程, 可学习到更加抽象的特征。编码器最终提取到脉冲信号中的有效特征信息。

解码器由潜在特征重构脉冲信号, 优化重构信号过程同时实现潜在特征优化。优化后潜在特征剔除了原始 VPPG 脉冲信号中部分运动干扰信息。多头注意力层关注潜在特征中多空间信息, 增强模型表达能力。激活函数与编码器中一致, 使用非线性映射使注意力关注重要信息。输出线性层依据原始输入信号维度得到重构信号。

多头注意力是编码器网络中重要组成部分, 具体结构如图 4 所示。多头注意力结构基于自注意力机制^[28], 将输入信号映射到多个子空间, 重点关注意子空间中脉冲信号的共有信息, 以提取鲁棒的脉冲信号特征。

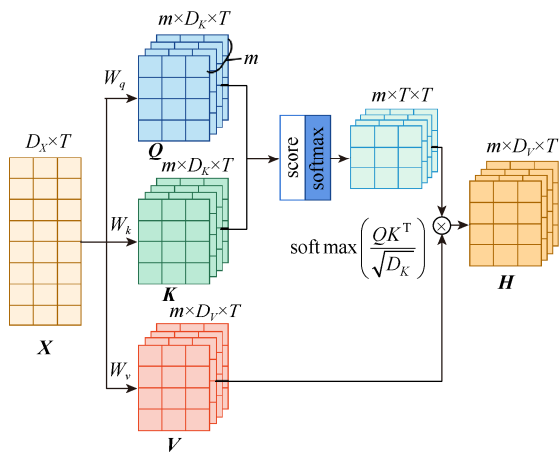


图 4 多头注意力模型

Fig. 4 Multihead attention model

将输入信号分别经过 m 次线性变化映射到 3 个空间, 得到查询矩阵 Q 、键矩阵 K 、值矩阵 V 。经过键值计算得到的输出为

$$H = \text{Attention}((K, V), Q) = \text{softmax}(s(K, Q))V \quad (5)$$

其中, $s(K, Q)$ 为注意力打分函数, 计算输出注意力分布 $\text{softmax}(\cdot)$ 为按列进行归一化的函数。注意力打分函数选择缩放点积模型, D_K 为矩阵 Q 的列数 (即向量维度), 保证梯度稳定性, 即

$$s(K, Q) = QK^T / \sqrt{D_K} \quad (6)$$

多头注意力将输入脉冲信号经过多组线性变化, 并行计算每组注意力, 关注输入脉冲信号多个子空间的信息。输出包含不同子空间的编码信息, 可以提高模型特征表达能力。

1.4 径向基分类网络

径向基房颤分类网络(radial basis function atrial fibrillation classification network, RBF-AF), 使用局部激活函数, 增强了模型对输入信号中房颤特征的辨别能力。强大的拟合能力可真实刻画面部脉搏信号与房颤检测的深层次关系, 提高实际场景中应用可行性。RBF-AF 网络隐藏层依赖径向基函数的神经元实现非线性映射, 将低维的输入数据转化到高维空间, 实现低维线性不可分到高维空间的线性可分。输出层采用线性优化策略, 输出各神经元的线性加权和, 具体结构如图 5 所示。

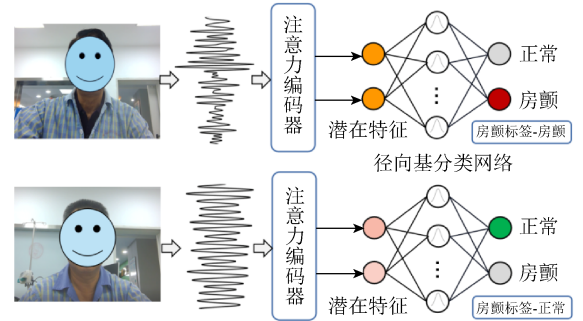


图 5 径向基房颤分类网络结构

Fig. 5 Radial basis function atrial fibrillation classification network structure

网络输入为注意力编码器提取的潜在特征, 潜在特征经过径向基函数映射到高维空间, 在房颤结果标签监督下扩大特征差异, 实现高精度房颤分类。分类网络神经元数分别为: 输入层节点个数为输入潜在特征的维数, 即 128 个; 隐藏层节点个数设为 512 个; 输出层节点个数为房颤分类结果数, 即 2 个。

隐藏层的高斯径向基函数是一种非线性函数, 具有中心对称衰减、局部分布特点, 是一种现成的相似性度量表示法。房颤信号与正常信号波形存在差异, 高斯径向基函数可体现波形差

异, 即

$$\varphi(\mathbf{X}) = \exp\left(-\frac{(\mathbf{X} - \mathbf{c})^2}{\sigma^2}\right) \quad (7)$$

其中, $\mathbf{X}$ 为输入向量; $\mathbf{c}$ 为中心点; σ^2 为方差。输出值与距离中心点的距离有关, 距离中心点越远, 输出越小。方差影响神经元感受域的大小。经过高斯径向基神经网络输出信号为

$$F(\mathbf{X}) = \sum_{h=1}^H w_h \exp\left(-\frac{\|\mathbf{X} - \mathbf{X}^h\|^2}{2\sigma^2}\right) \quad (8)$$

其中, $\mathbf{X}^h$ 为第 h 个隐藏层神经元径向基函数的中心; w_h 为第 h 个隐藏层神经元输出的权重, 隐藏层神经元的总数为 H 。径向基函数的中心、方差以及隐藏层到输出层的权值这 3 个参数决定网络性能的优良。由于径向基神经网络的输出神经元为线性单元, 梯度下降算法可收敛于全局最优解。

2 实验结果与分析

2.1 实验数据

目前还没有基于面部视频的房颤检测公开数据集。因此, 房颤实验是在本团队采集的数据上验证。本团队与中国科学技术大学附属第一医院心内科建立合作关系, 创建了从 2019 年 7 月至 2021 年 9 月非接触式房颤检测技术研究数据集。包括 200 名受试者(房颤患者与非房颤患者各 100 名)的 1 200 个面部视频。所有房颤患者在心电图检测中均检出房颤发作, 以此为房颤诊断依据, 数据集具体特征信息见表 1。

表 1 数据集特征表

Table 1 Dataset feature Table

特征	房颤患者(100 例)	正常测试者(100 例)
年龄(岁)	62.76±10.29	63.32±10.99
性别(男性: 女性)	59: 41	44: 56
心率(次/分钟)	72.14±11.06	78.35±13.43
刘海(%)	34.0	27.0
胡子(%)	5	9
眼镜(%)	9	7

数据采集会记录每个受试者 6 个面部视频(静态和动态各 3 个)和 PPG 信号。每段视频为 60 s, 采集摄像头为网络摄像头(C930 Logitech, 30 fps, 1920×1080 P)。数据采集时, 受试者在光照充足的环境中, 面部正对摄像头进行面部视频采集, 同时佩戴医疗级指夹式脉搏采集仪(型号为 ZJEPWS-

20D)同步记录 PPG 信号。受试者在静态视频中保持面部静止, 并在动态视频中进行典型的面部动作, 如说话、微笑和头部扭曲。

公开数据集 PURE(Pulse rate detection dataset)采集了 10 名测试者(8 名男性, 2 名女性)的 60 段面部视频, 包括 10 段静态视频和 50 段动态视频, 每段视频时长 60 s。每名测试者有 6 种不同测试状态, 包括静止、说话、转头(小幅度和大幅度)、头部平移(小幅度和大幅度)。视频图像大小为 640×480 P, 采样率为 30 Hz。面部视频采集同时使用接触式脉搏血氧仪(CMS50E, Contec Medical)采集 PPG 脉冲信号。

2.2 评价指标

2.2.1 重构信号评价指标

IPI 序列是由脉冲信号的峰-峰间隔得到, 同步采集的 VPPG 和 PPG 脉冲信号该特征相同, 2 个信号具有相同的 IPI 序列。因此, 2 个信号 IPI 序列误差可体现重构信号质量。其中, 第 k 个重构的 VPPG 信号 IPI 序列为 $R_{VPPG}^k(t)$ 和同步 PPG 信号 IPI 序列为 $R_{PPG}^k(t)$, 用于描述 IPI 误差的 3 个统计量如下:

(1) 平均误差(mean of the errors, M_E)。描述重构 VPPG 脉冲信号与标签信号误差的算术平均, 即

$$M_E = \frac{\sum_k \sum_t [R_{VPPG}^k(t) - R_{PPG}^k(t)]}{KT} \quad (9)$$

其中, T 为脉冲信号样本中 IPI 的个数; K 为样本的个数。

(2) 平均绝对误差(mean of the absolute errors, M_{AE})。描述所有恢复 VPPG 信号与标签信号绝对误差的平均值, 即

$$M_{AE} = \frac{\sum_k \sum_t |R_{VPPG}^k(t) - R_{PPG}^k(t)|}{KT} \quad (10)$$

(3) 误差标准差(standard deviation of the errors, SD_E)。描述信号的离散程度, 即

$$SD_E = \sqrt{\frac{\sum_k \sum_t [R_{VPPG}^k(t) - R_{PPG}^k(t) - M_E]^2}{KT - 1}} \quad (11)$$

2.2.2 房颤分类指标

房颤分类有 4 种结果: 真阳性(true positive, TP), 表示房颤病例正确分类; 真阴性(true negative, TN), 表示非房颤病例正确分类; 假阳性(false positive, FP), 表示非房颤病例错误分类; 假阴性(false negative, FN), 表示房颤病例错误分类。基

于以上结果,分类模型性能通过以下5个指标评价:

(1) 敏感性(sensitivity, S_E)。描述在房颤样本中,检测为房颤的概率,即

$$S_E = \frac{TP}{TP + FN} \quad (12)$$

(2) 特异性(specificity, S_P)。描述在非房颤样本中,检测出非房颤的概率,即

$$S_P = \frac{TN}{TN + FP} \quad (13)$$

(3) 阳性预测值(positive predictive value, P_{PV})。描述所有检测为房颤的样本中,实际房颤样本所占的比例,即

$$P_{PV} = \frac{TP}{TP + FP} \quad (14)$$

(4) 阴性预测值(negative predictive value, N_{PV})。其描述所有检测为非房颤的样本中,实际非房颤样本所占的比例,即

$$N_{PV} = \frac{TN}{TN + FN} \quad (15)$$

(5) 准确性(accuracy, A_{cc})。其所有被正确分类的样本数的占比,即

$$A_{cc} = \frac{TP + TN}{TP + TN + FP + FN} \quad (16)$$

2.3 实验参数与环境

实验数据按测试者进行划分,同一测试者的所有样本只出现在训练集或测试集中,即测试集和训练集中不包含同一测试者样本。模型通过五倍交叉验证进行了训练和测试。本文房颤分类模型,编码器和解码器中多头注意力模块参数一致,每次同步进行10个自注意力块。第一层线性层节点数为1024,第二层线性层节点数为128。

实验软件环境是基于开源的Pytorch深度学习框架,计算机硬件配置为Window10操作系统,因特尔酷睿i7-9700K CPU,储存大小为16 GB,显卡是GTX1650 SUPRE.模型训练时批处理数为3个,使用SGD优化器,动量moment为0.9,学习率是0.00088。

2.4 实验内容

为验证所提抗干扰房颤分类网络的实际性能,设计了分别在静态视频、动态视频2种场景下的对比实验以及消融实验。选取了3种典型的房颤检测方法进行了比较,方法介绍如下:

(1) 多区域脉冲信号融合算法(Region)。提取脸颊、额头等多区域的脉冲信号,针对不同区域提取

特征,再对多区域特征进行融合,最后送入分类器中进行房颤分类。

(2) 基于样本熵结合波形相似度算法(Similarity)。提取脉冲信号样本熵和形状相似度2个指标,将2个特征送入二分类网络中进行房颤分类。

(3) 基于RR间期LSTM。提取脉冲信号RR间期作为特征,将该特征送入LSTM网络中进行房颤分类。

2.4.1 静态场景实验结果与分析

静态场景下房颤分类对比实验使用自建数据集中静态视频,即测试者面部无明显运动。实验结果分析见表2。

表2 静态场景对比实验结果

Table 2 Experimental results of static scene comparison

方法	S_E	S_P	P_{PV}	N_{PV}	A_{cc}
Similarity	0.807	0.808	0.819	0.822	0.864
LSTM	0.812	0.819	0.834	0.831	0.875
Region	0.827	0.832	0.853	0.849	0.891
本文	0.889	0.894	0.901	0.918	0.934

注:加粗数据为最优值

实验结果表明,静态视频无明显运动干扰时,该4种方法均可以实现房颤初筛功能。对比方法中表现最差的房颤检测准确度也达到86.4%,已基本满足房颤初筛要求。然而,本文算法的房颤准确度为93.4%,检测结果可靠性更高。

2.4.2 动态场景实验结果与分析

动态场景下房颤分类对比实验和消融实验使用自建数据集中动态视频,即测试者存在说话、微笑等面部动作。由表3可知,在动态场景下所有算法精度均有所下降,但本文模型仍具备一定优势,精度下降幅度最小。对比方法Similarity相比静态场景精度下降了18.75%,分析原因是面部运动产生运动干扰信号,直接影响脉冲信号波形,使基于波形相似度的特征有效性严重下降。对比方法LSTM根据脉冲信号RR间期特征分类,但运动干扰产生的不规则信号直接影响脉冲信号RR间期。在头部运动过程中,面部各区域光照分布发生改变,对比方法Region并未针对各区域进行光照补偿等操作,所提取的多区域特征受光照变化影响。3种对比房颤检测特异性最高仅72.6%,说明易将运动干扰信号误判为房颤信号,房颤检测不够鲁棒。本文方法中注意力去噪编码器可将脉冲信号中运动信号与脉冲信号分离,得到高质量的脉冲信号,特异性相

比最优对比方法提升了 18.3%。为进一步展示实验结果,选取一段包含运动干扰的视频,使用各对比方法提取脉冲信号。图 6 为各方法提取的脉冲信号。原始脉冲信号中包含 3 处强运动干扰,如图 6(a)中虚框所示位置,图 6(b)为同步 PPG 信号。比较 3

种对比方法脉冲信号与 PPG 信号可知,Similarity 方法和 LSTM 方法在运动干扰处波形均存在严重失真,产生运动伪峰。Region 方法有所提升,但波形在连续运动干扰处,仍有一定影响。本文方法得到的脉冲信号形态规整,无明显运动干扰痕迹。

表 3 动态场景实验结果
Table 3 Dynamic scene experimental results

方法	IPI 误差指标			房颤分类指标				
	SD _E	M _E	M _{AE}	S _E	S _P	P _{PV}	N _{PV}	A _{cc}
Similarity	199.1	96.2	188.3	0.712	0.657	0.735	0.724	0.702
LSTM	175.6	70.5	162.1	0.731	0.697	0.754	0.763	0.721
Region	169.1	68.4	150.3	0.782	0.726	0.794	0.773	0.752
本文	142.1	51.6	115.8	0.846	0.859	0.869	0.851	0.876

注:加粗数据为最优值

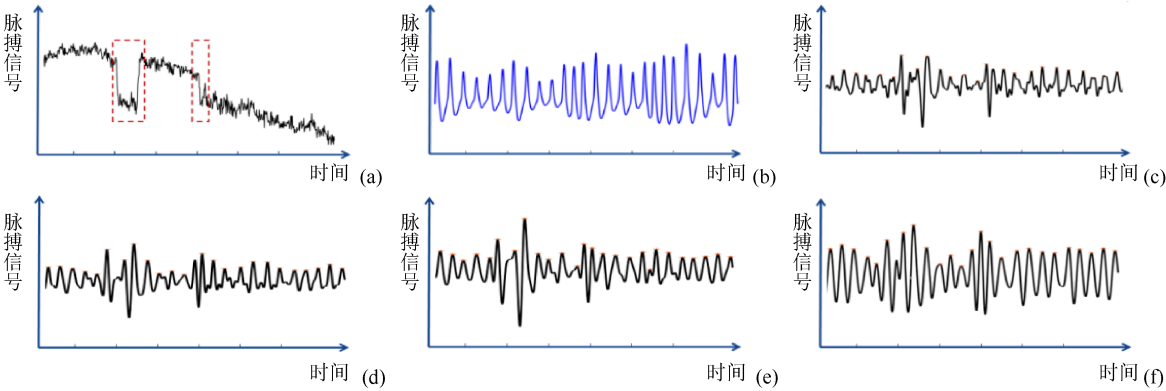


图 6 各对比方法脉冲信号图((a)原始信号;(b)同步 PPG 信号;(c) Similarity 方法;(d) LSTM 方法;(e) Region 方法;(f)本文方法)

Fig. 6 Pulse signal diagram for each comparison method ((a) Original signal; (b) PPG signal; (c) Similarity method; (d) LSTM method; (e) Region method; (f) Ours)

2.4.3 消融实验结果与分析

网络中注意力编码器基于多头注意力机制提取输入信号的多维空间信息,多头注意力子空间数影响提取效果。在静态、动态视频上分别探究子空间数对分类精度影响。实验结果表明,子空间数为 10 时效果达到最佳,故实验设置子空间数为 10 个,如图 7 所示。

为进一步验证本文模型中各模块的必要性,进行了动态场景中的消融实验。设计了 5 组实验,选取一段动态场景视频,使用全连接网络、去噪自编码器、卷积自编码器 3 种对比网络进行特征提取,与本文注意力编码器所提取的潜在特征分别输入同一径向基分类网络中,对比各方法房颤特征提取能力。为验证高斯径向基分类网络有效性,将注意力编码器潜在特征分别输入 RBF-AF 网络和三层卷积网络(表 4 第 4 行实验)中比较房颤分类精度。

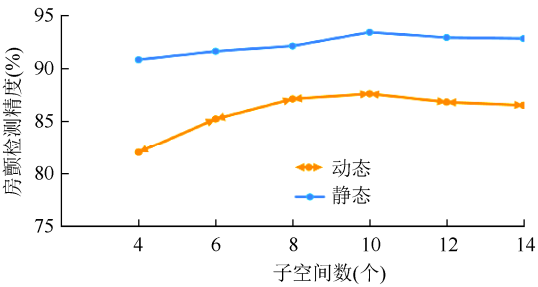


图 7 子空间数实验结果图

Fig. 7 Plot of experimental results for subspace numbers

为验证本文双重构损失函数组合效果,进行 2 个实验:一是仅采用交叉熵损失函数,编码器中不添加重构损失函数(表 4 第 5 行实验),二是交叉熵和重构损失函数组合。

消融实验结果由表 4 可知,3 种对比网络的潜在特征明显受到运动干扰,与标准 PPG 脉冲信号 IPI 序列误差较大。分析原因可能为:全连接网络仅依靠隐藏层节点学习输入信号的全局特征,当信

号构成复杂, 网络学习到的特征可信度低。去噪自编码器为避免输入信号中包含存在噪声, 只学习部分输入信号, 减少了输入信号包含的信息, 去除部分运动干扰同时也丢弃了部分脉冲信息。卷积自编码器使用卷积层和池化层提取特征, 脉冲信号中运动干扰幅度一般大于脉冲信号幅度, 卷积操作对输入信号信息量变化大的信息敏感。本文利用注意力

编码器提取输入信号多个子空间信息, 在同步 PPG 脉冲信号引导下, 重点关注鲁棒的房颤特征。本文构建的高斯径向基分类网络房颤分类精度高于 CNN, 对相同输入的潜在特征 RBF-AF 网络房颤识别能力更强, 分类精度更高。双损失函数组合方式对重构信号质量有明显改善, 增强房颤特征表达, 提高房颤识别精度。

表 4 消融实验结果

Table 4 Ablation experimental results

方法	IPI 误差指标			房颤分类指标				
	SDE	ME	MAE	SE	Sp	Ppv	Npv	Acc
全连接网络+RBF-AF	196.2	95.8	183.8	0.710	0.717	0.732	0.701	0.701
去噪编码器+RBF-AF	170.7	72.4	165.1	0.821	0.726	0.729	0.794	0.759
卷积编码器+RBF-AF	171.5	72.9	163.3	0.811	0.708	0.711	0.782	0.763
注意力编码器+CNN	160.8	65.4	155.1	0.837	0.752	0.720	0.802	0.794
本文 注意力编码器+ Loss 1	166.8	69.5	160.1	0.823	0.748	0.713	0.794	0.812
注意力编码器+RBF-AF	142.1	51.6	115.8	0.846	0.859	0.869	0.851	0.876

注: 加粗数据为最优值

为更加生动展示消融实验结果, 选取一段动态视频使用各方法提取脉冲信号。视频同步采集的 PPG 信号及 IPI 序列作为基准, 图 8 展示了消融实验各方法得到的脉冲信号和脉冲信号对应的 IPI 序列。

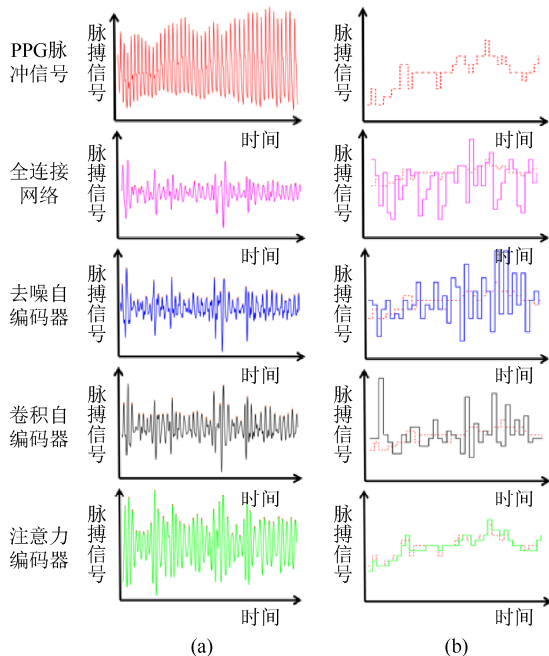


图 8 各对比方法重构信号和 IPI 序列((a)重构信号; (b) IPI 序列)

Fig. 8 Each comparison method reconstructed signal and IPI sequences ((a) Reconstructed signal; (b) IPI sequences)

观察各组脉冲信号形态和 IPI 序列可知, 全连接网络所得脉冲信号与标准 PPG 脉冲信号差异较大, IPI 序列形态差异大、波动大, 潜在特征包含

脉冲特征以外的运动干扰特征。因此, 使用该潜在特征进行房颤分类时, 分类结果可信度降低, 分类精度下降。去噪自编码器和卷积自编码器所得 IPI 序列总体在基准线附近, 存在一定波动。说明 2 个网络具有去噪抗干扰能力, 但在强烈运动干扰处仍不足。本文注意力编码器结果, 脉冲信号形态规整, IPI 序列波动小, 与基准序列相似。综合以上分析可知, 在运动干扰场景中, 所提抗干扰房颤检测模型抗干扰能力强、可提高检测精度、增强结果可信度。

2.4.4 交叉数据集实验

为评估本文模型的泛化能力, 将本文模型与 3 种房颤分类算法在其他数据集进行测试。由于没有基于 VPPG 的房颤检测公开数据集, 选择包含面部视频和同步 PPG 信号的 PURE 数据集。假设所有测试者无房颤, 检测各方法的特异性, 图 9 为实验结果。

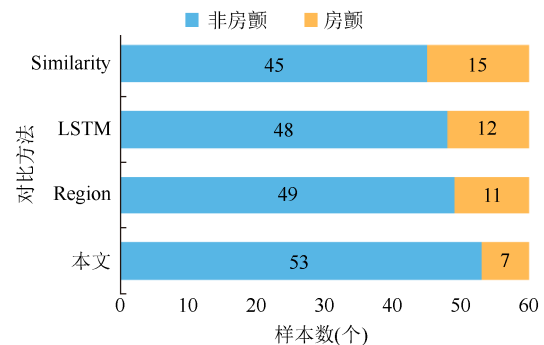


图 9 交叉数据集实验结果

Fig. 9 Cross dataset experimental results

实验使用 PURE 数据集中共 60 个样本, Similarity 方法将 15 个正常样本误判为房颤, 特异性为 0.750。Region 方法和 LSTM 方法分别错误分类 11 个和 12 个, 特异性分别为 0.817 和 0.800。本文方法错误分类 7 个, 特异性为 0.883。PURE 数据集中有 10 个静态视频, 因此特异性相比自建数据集动态视频结果均略有上升。

2.4.5 特征可视化

脉冲信号经过各模块后得到的特征特性不同, 将各层特征可视化。t-SNE 聚类^[29]可以将高维空间映射到可视化的二维空间。图 10 展示了基于 120 个测试样本在一次交叉验证的结果, 其中绿点表示正常样本, 红点表示房颤样本, 黑点表示被错误分类的样本。

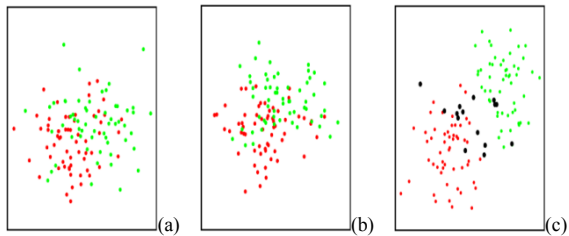


图 10 特征可视化((a)原始信号特征; (b)注意力编码器网络特征; (c)径向基网络特征)

Fig. 10 Feature visualization ((a) Original signal feature; (b) Attention encoder network feature; (c) Radial basis network feature)

对比图中的特征分布, 随着网络模型的学习, 正常样本与房颤样本特征差异愈加明显。原始脉冲信号由于运动噪声, 存在信号伪峰, 与房颤信号特征类似, 因此原始脉冲信号中 2 种特征很难区分。注意力编码器网络去噪后的特征, 减小运动噪声影响, 特征差异逐渐展现, 但仍然存在部分相似特征难以区分。通过高斯径向基函数映射到高维特征空间后的特征分布, 2 类样本特征分布明显, 利于房颤分类。最终错误分类样本为 15 例, 分类精确度为 87.5%。

实验结果证明, 注意力编码器抗干扰房颤分类网络提高了现实场景下的面部视频房颤检测精确度, 拓展了 VPPG 脉冲信号实际应用范围。

2.4.6 信号时长影响实验

为探究面部视频长度对房颤检测结果的影响, 设计了一组使用不同长时的脉冲信号检测房颤实验。实验使用自建数据集中静态视频, 模拟使用该检测系统作为房颤初筛工具, 受试者配合检测下的场景, 具体实验结果如图 11 所示。

从图中可知, 随着信号时间的增加, 房颤检测精度逐步提升。当时长大于 50 s 时, 分类精度趋于稳定。因此, 在实际场景中运用时需要获取足够长的面部视频, 以保证检测结果的有效性。

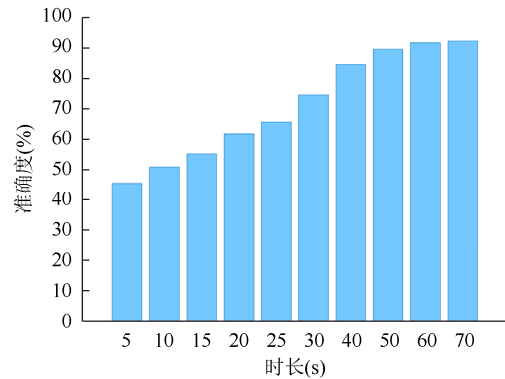


图 11 时长实验结果

Fig. 11 Duration results

3 结束语

房颤的早期发现与诊断是降低房颤及并发症风险的关键。VPPG 技术为房颤筛查提供了新途径, 但易受到现实场景中运动干扰。本文提出了一种抗干扰视频房颤检测网络。该模型使用注意力编码器提取鲁棒的潜在特征, 再由径向基分类网络实现房颤分类。将包含运动信号的原始脉冲信号映射到多个子空间, 重点关注脉冲信息, 提取抗运动干扰的潜在特征。径向基分类网络在房颤标签监督下增强潜在特征的房颤识别能力, 输出可靠分类结果。实验结果表明, 本文模型在各类场景中均表现优异。在静态场景中, 检测精度相比对比算法提高了 8.1%, 敏感性提高了 7.5%。在动态场景下, 对比算法准确度大幅下降, 本文检测精度相比最优算法提升了 16.5%, 特异性相比提升了 18.3%。本文优于现有 VPPG 房颤分类方法, 在实际场景中抗运动干扰能力强, 检测结果可靠。

虽然本文模型能提高动态场景中房颤检测精度, 增强特征抗运动干扰能力。但仍存在不足, 该模型仅在肤色较浅的人群中进行了测试, 不清楚是否适用于其他肤色的人群。未来考虑通过迁移学习扩展到其他肤色人群。

参考文献 (References)

- [1] CHUNG M K, ECKHARDT L L, CHEN L Y, et al. Lifestyle and risk factor modification for reduction of atrial fibrillation: a scientific statement from the American heart association[J]. Circulation, 2020, 141(16): e750-e772.
- [2] HINDRICKS G, POTPARA T, DAGRES N, et al. 2020 ESC

- guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): the task force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC[J]. *European Heart Journal*, 2020, 42(5): 373-498.
- [3] GUO Y T, WANG H, ZHANG H, et al. Mobile photoplethysmographic technology to detect atrial fibrillation[J]. *Journal of the American College of Cardiology*, 2019, 74(19): 2365-2375.
 - [4] SUN Y, THAKOR N. Photoplethysmography revisited: from contact to noncontact, from point to imaging[J]. *IEEE Transactions on Bio-Medical Engineering*, 2016, 63(3): 463-477.
 - [5] 周双, 杨学志, 金兢, 等. 采用自适应信号恢复算法的非接触式心率检测[J]. *中国图象图形学报*, 2019, 24(10): 1670-1682.
ZHOU S, YANG X Z, JIN J, et al. Non-contact heart rate detection using self-adaptive signal recovery algorithm[J]. *Journal of Image and Graphics*, 2019, 24(10): 1670-1682 (in Chinese).
 - [6] VERKRUYSE W, SVAASAND L O, NELSON J S. Remote plethysmographic imaging using ambient light[J]. *Optics Express*, 2008, 16(26): 21434-21445.
 - [7] FAVILLA R, ZUCCALÀ V C, COPPINI G. Heart rate and heart rate variability from single-channel video and ICA integration of multiple signals[J]. *IEEE Journal of Biomedical and Health Informatics*, 2019, 23(6): 2398-2408.
 - [8] WANG W J, STUIJK S, DE HAAN G. Exploiting spatial redundancy of image sensor for motion robust rPPG[J]. *IEEE Transactions on Biomedical Engineering*, 2015, 62(2): 415-425.
 - [9] POH M Z, MCDUFF D J, PICARD R W. Advancements in noncontact, multiparameter physiological measurements using a webcam[J]. *IEEE Transactions on Biomedical Engineering*, 2011, 58(1): 7-11.
 - [10] LIU X N, YANG X Z, JIN J, et al. Self-adaptive signal separation for non-contact heart rate estimation from facial video in realistic environments[J]. *Physiological Measurement*, 2018, 39(6): 06NT01.
 - [11] GHODRATIGOHAR M, GHANADIAN H, AL OSMAN H. A remote respiration rate measurement method for non-stationary subjects using CEEMDAN and machine learning[J]. *IEEE Sensors Journal*, 2020, 20(3): 1400-1410.
 - [12] LIU X N, YANG X Z, WANG D L, et al. Detecting pulse rates from facial videos recorded in unstable lighting conditions: an adaptive spatiotemporal homomorphic filtering algorithm[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 1-15.
 - [13] COUDERC J P, KYAL S, MESTHA L K, et al. Detection of atrial fibrillation using contactless facial video monitoring[J]. *Heart Rhythm*, 2015, 12(1): 195-201.
 - [14] CORINO V, IOZZIA L, MARIANI A, et al. Identification of atrial fibrillation episodes using a camera as contactless sensor[C]//2017 Computing in Cardiology Conference. New York: IEEE Press, 2017: 1-4.
 - [15] CHENG P, CHEN Z C, LI Q Z, et al. Atrial fibrillation identification with PPG signals using a combination of time-frequency analysis and deep learning[J]. *IEEE Access*, 8: 172692-172706.
 - [16] 郭一楠, 邵慧杰, 巩敦卫, 等. 基于希尔伯特黄变换和深度卷积神经网络的房颤检测[J]. *电子与信息学报*, 2022, 44(1): 99-106.
GUO Y N, SHAO H J, GONG D W, et al. Atrial fibrillation detection based on Hilbert-Huang transform and deep convolutional neural network[J]. *Journal of Electronics & Information Technology*, 2022, 44(1): 99-106 (in Chinese).
 - [17] KWON S, HONG J, CHOI E K, et al. Deep learning approaches to detect atrial fibrillation using photoplethysmographic signals: algorithms development study[J]. *JMIR MHealth and UHealth*, 2019, 7(6): e12770.
 - [18] YAN B P, LAI W H S, CHAN C K Y, et al. Contact-free screening of atrial fibrillation by a smartphone using facial pulsatile photoplethysmographic signals[J]. *Journal of the American Heart Association*, 2018, 7(8): e008585.
 - [19] SHI J G, ALIKHANI I, LI X B, et al. Atrial fibrillation detection from face videos by fusing subtle variations[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(8): 2781-2795.
 - [20] LI X B, CHEN J, ZHAO G Y, et al. Remote heart rate measurement from face videos under realistic situations[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press: 4264-4271.
 - [21] ASTHANA A, ZAFEIRIOU S, CHENG S Y, et al. Robust discriminative response map fitting with constrained local models[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press: 3444-3451.
 - [22] KUMAR M, VEERARAGHAVAN A, SABHARWAL A. DistancePPG: robust non-contact vital signs monitoring using a camera[J]. *Biomedical Optics Express*, 2015, 6(5): 1565-1588.
 - [23] DE HAAN G, JEANNE V. Robust pulse rate from chrominance-based rPPG[J]. *IEEE Transactions on Biomedical Engineering*, 2013, 60(10): 2878-2886.
 - [24] WANG C, ZHU M X, WANG X, et al. Time-frequency analysis of electroencephalogram signals in a cognitive decision-making task[C]//The 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society. New York: IEEE Press: 3469-3472.
 - [25] KUMAR A, TOMAR H, MEHLA V K, et al. Stationary wavelet transform based ECG signal denoising method[J]. *ISA Transactions*, 2021, 114: 251-262.
 - [26] COHEN M X. A better way to define and describe Morlet wavelets for time-frequency analysis[J]. *NeuroImage*, 1999: 81-86.
 - [27] HUMPHREYS G W, SUI J. Attentional control and the self: the self-attention network (SAN)[J]. *Cognitive Neuroscience*, 2016, 7(1-4): 5-17.
 - [28] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all You need[C]//The 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.
 - [29] VAN DER MAATEN L, HINTON G. Visualizing data using t-SNE[J]. *Journal of Machine Learning Research*, 2008, 9: 2579-2625.