

DOI:10.3969/j.issn.1003-5060.2024.04.008

# 基于权重优化卷积神经网络的非接触心率检测

王盼孺<sup>1,2</sup>, 杨学志<sup>2,3</sup>, 刘雪南<sup>2,3</sup>, 李龙伟<sup>4</sup>, 王定良<sup>4</sup>

(1. 合肥工业大学 计算机与信息学院, 安徽 合肥 230601; 2. 工业安全与应急技术安徽省重点实验室, 安徽 合肥 230601; 3. 合肥工业大学 软件学院, 安徽 合肥 230601; 4. 中国科学技术大学第一附属医院(安徽省立医院) 心血管内科, 安徽 合肥 230002)

**摘要:** 心率是反映人体健康状况和运动状态的重要指标。针对传统心率检测的局限性, 近年来提出了许多非接触式的检测方法, 其在配合条件下的效果较好, 但是在有运动干扰时准确率明显下降。针对该问题, 文章结合计算机视觉与深度学习前沿理论, 提出一种基于权重优化卷积神经网络的非接触心率检测方法, 通过优化卷积神经网络的结构, 提高网络的抗噪性能, 获得更准确的心率值。首先输入相对稳定的面部视频, 然后将输入的面部视频逐帧按行取像素平均值并时域扩展得到各行子脉搏波, 再使用主成分分析(principal component analysis, PCA)法与带通滤波器对各行子脉搏波组成的脉搏矩阵进行处理, 最后将所得特征矩阵输入权重优化卷积神经网络学习, 预测心率值。为了验证该方法的性能优势, 使用自采数据集中的2200份人脸视频样本进行实验分析, 实验结果表明, 文中所提方法与现有的非接触心率检测方法相比, 具有更高的准确率、更强的鲁棒性。

**关键词:** 非接触; 主成分分析(PCA); 权重优化; 卷积神经网络; 心率

中图分类号: TP391

文献标志码: A

文章编号: 1003-5060(2024)04-0479-09

## Non-contact heart rate detection based on weight optimized convolutional neural network

WANG Panru<sup>1,2</sup>, YANG Xuezhizhi<sup>2,3</sup>, LIU Xuenan<sup>2,3</sup>, LI Longwei<sup>4</sup>, WANG Dingliang<sup>4</sup>

(1. School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China; 2. Anhui Province Key Laboratory of Industry Safety and Emergency Technology, Hefei 230601, China; 3. School of Software, Hefei University of Technology, Hefei 230601, China; 4. Department of Cardiovasology, The First Affiliated Hospital of University of Science and Technology of China(Anhui Provincial Hospital), Hefei 230002, China)

**Abstract:** Heart rate is an important indicator of the health and exercise status of the human body. Due to the limitations of traditional heart rate detection, many non-contact detection methods have been proposed in recent years, which have good results under cooperative conditions. But the accuracy is significantly reduced when there is motion disturbance. For this problem, this paper combines computer vision and deep learning frontier theory to propose a non-contact heart rate detection method based on weight optimized convolutional neural network. By optimizing the network structure of the convolutional neural network, the noise immunity performance of the network is improved, and a more accurate heart rate value is obtained. Firstly, the stable facial video is input. Secondly, the pixel average of the input facial video frame is taken by frame and row by row, and the time domain expansion is made to obtain each row of sub-pulse waves. The pulse matrix composed of the sub-pulse waves is processed by using principal component analysis(PCA) and band-pass filter. Finally, the resulting feature matrix is input into the weight optimized convolutional neural network to learn to predict the heart rate. To verify the performance advantages of this algorithm, 2200 face video samples from the self-collected dataset were used for experimental analysis. The experimental results showed that the

收稿日期: 2022-02-27; 修回日期: 2022-04-26

基金项目: 安徽省科技重大专项资助项目(201903C080200100); 安徽高校协同创新资助项目(GXXT-2019-003)和智能互联系统安徽省实验室(合肥工业大学)资助项目(PA2021AKSK0111)

作者简介: 王盼孺(1997—), 男, 安徽合肥人, 合肥工业大学硕士生;

杨学志(1970—), 男, 安徽合肥人, 博士, 合肥工业大学教授, 博士生导师, 通信作者, E-mail: xzyang@hfut.edu.cn.

proposed algorithm has higher accuracy and better robustness than the existing non-contact heart rate detection methods.

**Key words:** non-contact; principal component analysis(PCA); weight optimization; convolutional neural network; heart rate

心率是心脏对外界负荷的反应,对心血管疾病的各个方面有重要的指导意义,同时它也是运动强度最直观的体现,心率高低直接影响到运动的效果与目的<sup>[1]</sup>,心率控制在运动中有着显著作用。早期对于心率的检测主要通过专业的医学仪器,一般有心电测量仪<sup>[2]</sup>和脉搏仪<sup>[3]</sup>等,但是该类仪器价格昂贵且不方便携带,接触人体甚至有创的检测方式令患者较为不适<sup>[4]</sup>。

随着非接触生理参数检测<sup>[5]</sup>的兴起,人们使用图像光电容积描记术(image photoplethysmography, IPPG)<sup>[6]</sup>来检测心率,该技术利用成像设备观测人体皮肤表面颜色变化对血液容积的影响来提取脉搏信号,得到心率值,不仅容易实现,而且成本低廉、安全高效。

文献[7]首次验证了非接触心率检测的可行性。文献[8]提出了一种基于面部区域分块选择的方法,将整个面部区域划分为若干个子区域,通过排除已被污染的子区域,选择出合适的感兴趣区域(region of interest, ROI),然后提取纯净的脉搏信号,然而若面部区域存在整体较大幅度的运动干扰,无法选取到合适感兴趣区域,该方法则无法准确检测心率。文献[9]提出一种基于独立成分分析和小波变换的方法,该方法首先使用独立成分分析法得到脉搏信号,然后通过小波变换去噪,然而,小波分解的局限性使其只能滤除心率频率范围之外的噪声,无法处理由运动干扰导致的处于心率频率范围内的尖锐噪声。文献[10]使用深度学习对心率进行预测,将面部视频信号的时域和频域特征输入卷积神经网络分类心率。文献[11]将面部视频的不同时空表示输入卷积神经网络获取脉搏信号,同时使用模拟心率信号预训

练来更好地增加样本规模,初始化模型。此方法虽然在一定程度上解决了数据缺乏的问题,但是预训练模型参数较为繁琐,另外若使用在复杂多样的现实场景中,预训练很难获取到初始样本的共同特征,则微调的效果达不到预期。文献[12]尝试直接使用人脸视频作为输入来检测心率,通过级联全卷积神经网络实现对心率的估计。但是此方法对于硬件要求过高,且处理数据与计算时间太长,只适用于理论研究与方法分析。文献[13]提出利用人脸视频时空上下文信息,构建底层时空图的方法滤除背景信号噪声,另外在神经网络训练时,引入邻近不同心率值之间的数值关系。但该方法只考虑了环境噪声的影响,忽视了面部表情变化等运动造成的信号干扰,同时引入的门控循环单元对于异常值的兼容性较差,容易受到极端值的影响而大幅降低心率预测的准确率。综上所述,目前的非接触心率检测方法效果均不理想,大多无法满足复杂多变场景下的心率检测需求。

针对以上问题,本文提出了一种基于权重优化卷积神经网络的非接触心率检测方法。通过对面部脉搏矩阵的构建与处理,有效保留了图像信息并去除噪声;提出权重优化卷积神经网络,在卷积层后增加了权重优化层,优化网络结构,提高了神经网络学习原始心率特征和辨别运动干扰的能力;在数据规模较大的自采数据集上,对于心率的预测比其他方法更准确,鲁棒性更强。

### 1 本文方法

本文提出了一种基于权重优化卷积神经网络的非接触心率检测方法,方法流程如图 1 所示。

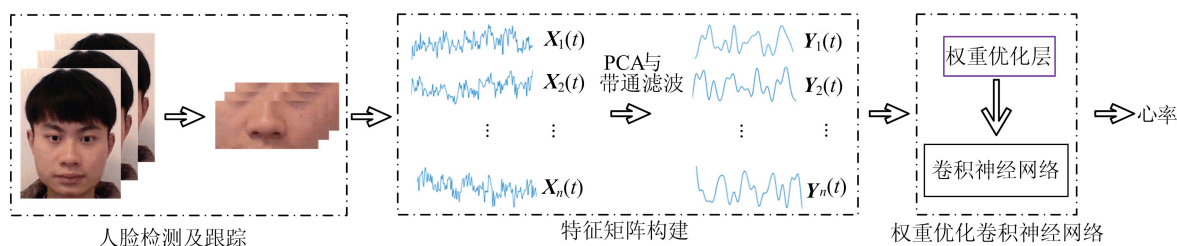


图 1 检测方法流程

该方法主要分为以下2个部分。

1) 数据预处理。这部分主要包括人脸检测及跟踪和特征矩阵的构建。首先使用人脸检测与跟踪算法获得相对稳定的面部视频,随后将面部视频逐帧按行取像素平均值并时域扩展得到各行子脉搏波,使用主成分分析法与带通滤波器对各行子脉搏波组成的脉搏矩阵进行处理,获得输入卷积神经网络的特征矩阵。

2) 权重优化卷积神经网络。对原有卷积神经网络结构做了改进,在卷积层后加入权重优化层,引入3项重要指标和Critic赋权法对每次卷积后得到的特征图进行权重优化,调整不同行脉搏特性向量的权重大小,随后再将优化后的特征图经过池化和全连接等操作得到最终的心率值。

### 1.1 数据预处理

#### 1.1.1 脉搏矩阵的构建

在实际的数据采集过程中,往往会有不同程度的面部晃动和头部运动,为了获得较为稳定的面部视频图像,受文献[14]启发,本文首先使用OpenCV的ViolaJones人脸检测器获取人脸区域,并利用判别响应图拟合(discriminative response map fitting, DRMF)算法和Kanade-Lucas-Tomasi(KLT)算法分别进行人脸面部特征点的定位与跟踪,最后得到校正后的稳定图像。

考虑到在RGB三色光中,绿色光的反射光信息最多,对皮肤血液流动的变化最敏感,对应RGB颜色模型中,G通道的像素值相对于R、B通道更能反映出人体血液容积的变化信息,因此本文所用像素值均为三基色分离后G通道的像素值。

同时由于头发、眼部等区域不包含心率信息且容易引入更多运动干扰,本文采用眼部以下、唇部以上的脸颊区域作为感兴趣区域。

对面部视频的感兴趣区域进行处理,得到各点像素值,然后按行分块,计算出每一帧视频图像各行的像素平均值,并按照帧数时域扩展,得到各行子脉搏波 $\mathbf{X}_i$ ,如图2所示,从而一个面部视频就得到了一个脉搏矩阵,即

$$\boldsymbol{\delta} = \begin{pmatrix} \mathbf{X}_1(t) \\ \mathbf{X}_2(t) \\ \vdots \\ \mathbf{X}_n(t) \end{pmatrix} \quad (1)$$

其中: $t$ 为视频帧数; $n$ 为每一帧图像被分割的块数。

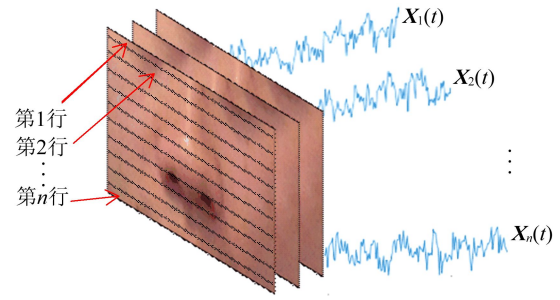


图2 脉搏矩阵构建

#### 1.1.2 时域滤波

上述得到的子脉搏波均含有不同程度的噪声干扰,需要通过带通滤波器滤除心率频带范围外的高频低频成分。本次所用数据集样本中,受试者的心率范围集中在50~110次/min之间,考虑到巴特沃斯带通滤波器的频率响应曲线具有最大限度平坦幅度,没有起伏的特点,本实验使用通频带为0.83~1.67 Hz的三阶巴特沃斯带通滤波器对各行子脉搏波滤波,可降低由呼吸和姿势变化等引起的低频噪声和由人体不自主微小运动等引起的高频噪声的影响,得到滤波后的脉搏矩阵 $\boldsymbol{\delta}'$ 。

#### 1.1.3 主成分分析法

主成分分析(principal component analysis, PCA)法是一种应用广泛的数据处理方法,它可以将原有的数据通过线性组合的方法表示在另一个新的基底坐标上,被广泛应用于数据压缩、消除冗余和降低噪声等方面。滤波后的脉搏矩阵 $\boldsymbol{\delta}'$ 集合了面部视频大部分心率信息,然而其数据规模过于庞大和复杂,为了消除冗余数据,提高网络模型学习效率,同时降低由心脉冲击引起的头部运动和其他面部表情变化等运动信号的干扰,本文使用PCA法对滤波后的脉搏矩阵 $\boldsymbol{\delta}'$ 压缩处理。PCA法计算步骤如下。

- 1) 设有一个 $n$ 维矩阵 $\mathbf{M}$ 。
- 2) 取出 $\mathbf{M}$ 的每一行减掉该行的均值,即中心化处理。
- 3) 求出协方差矩阵。
- 4) 计算出协方差矩阵的特征值及其对应的特征向量。
- 5) 按照特征值的大小将其所对应的特征向量从上到下排列成矩阵,取前 $m$ 行组成矩阵 $\mathbf{P}$ 。
- 6)  $\mathbf{N}=\mathbf{PM}$ 即为降到 $m$ 维后的矩阵。

对于 $n$ 维脉搏矩阵 $\boldsymbol{\delta}'$ ,通过PCA法计算主成分贡献率得到前 $m$ 个主要成分,形成最终的特征矩阵 $\boldsymbol{\theta}$ ,即

$$\theta = \begin{pmatrix} Y_1(t) \\ Y_2(t) \\ \vdots \\ Y_m(t) \end{pmatrix} \quad (2)$$

其中:  $t$  为视频帧数;  $m$  为特征矩阵的维数。

1.2 权重优化卷积神经网络

卷积神经网络<sup>[15-17]</sup>是一种典型的前馈神

经网络,它可以从原始图像有效的信息中获取隐藏特征,避免了人工选取特征的繁琐过程,能够更好地反映图像之间的内在规律,被广泛应用于图像识别、数字图像处理等领域。本文所用权重优化卷积神经网络以卷积神经网络为基础,在每一个卷积层后增加权重优化层,总体结构如图 3 所示。

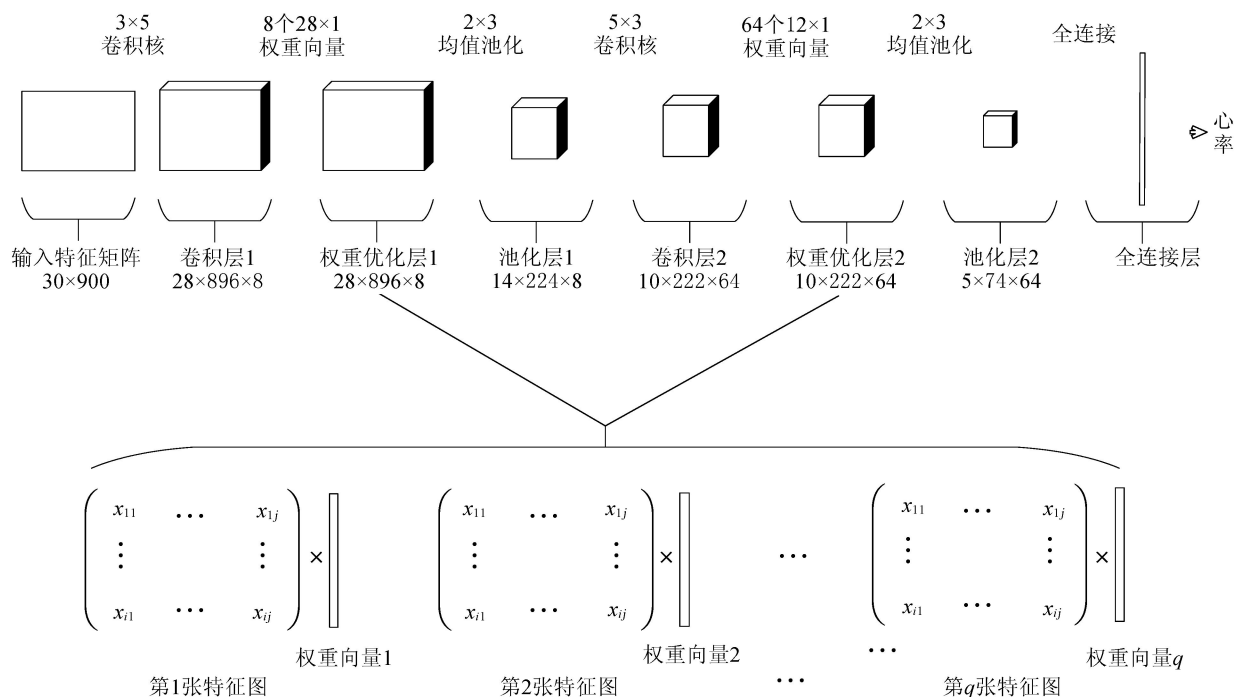


图 3 权重优化卷积神经网络结构

卷积神经网络中最重要的一个步骤是卷积层对输入图像特征的提取,然而当输入的图像含有运动噪声时,噪声所含特征会被神经网络当作心率特征学习记忆,从而影响最终的检测结果。为了帮助卷积神经网络更好地辨别运动噪声,提高对心率特征的提取能力,本文提出了权重优化层的概念,它的主要作用是优化卷积层输出的特征图,降低特征图中噪声特征所占的比例,从而提高网络的抗噪性能。

权重优化层主要工作原理如下:对于输入的一张  $i$  行  $j$  列的特征图,它是由  $i$  个含有心率和噪声特征的脉搏特性向量  $Z(t)$  组成,首先计算每一个脉搏特性向量的信噪比、R-R 间期方差和 PNN70 这 3 项指标的大小。

信噪比  $S_{NR}$  反映的是脉搏特性向量中信号与噪声的比例,计算公式如下:

$$S_{NR}(Z(t)) = \frac{1}{\lg\{\max(\text{abs}(\text{FFT}(Z(t))))\}}$$

$$\left[ \sum_{f=0}^{\infty} \text{abs}(\text{FFT}(Z(t))) - \max(\text{abs}(\text{FFT}(Z(t)))) \right] \quad (3)$$

其中:  $\max(\text{abs}(\text{FFT}(Z(t))))$  为每个脉搏特性向量经过快速傅里叶变换后的频谱峰值;  $\sum_{f=0}^{\infty} \text{abs}(\text{FFT}(Z(t)))$  为整个频谱的幅值和;  $\text{abs}$  为幅值;  $\max$  为求取最大值函数。

R-R 间期表示脉搏特性向量中 2 个相邻主波波峰之间的时限,它的方差  $S_{R-R}^2$  能够反映出向量的规整程度,同时其也是临床中判断心率是否正常的指标之一,计算公式如下:

$$S_{R-R}^2 = \frac{\sum_{k=1}^n (p_k - \bar{p})^2}{n-1} \quad (4)$$

其中:  $p_k$  为第  $k$  个主波波峰和第  $k+1$  个主波波峰之间的时限;  $\bar{p}$  为所有相邻主波波峰之间时限的平均值;  $n$  为整个时限序列的总个数。

PNN70 表示脉搏特性向量中 R-R 间期大于

70 ms 的个数占总体的比例,反映异常心跳的权重大小,辨别出真实的脉搏信号,计算公式如下:

$$P_{\text{NN70}} = \frac{A}{B} \quad (5)$$

其中: $A$  为相邻主波波峰之间的时限大于 70 ms 的个数; $B$  为整个时限序列的总个数。

在得到了特征图中每一个脉搏特性向量的信噪比、R-R 间期方差和 PNN70 后,引入 Critic 赋权法对这 3 项指标进行权重划分。Critic 赋权法是一种客观的权重分配方法,该方法通过计算各指标的对比强度与指标之间冲突性来判断指标应当被分配的权重大小。其中对比强度是指同一个指标在各行向量之间取值差异的大小,用标准差表示。标准差越大表明该指标的数值差异越大,越能反映出更多的信息,则它的权重占比应当越大;指标之间的冲突性以相关系数的形式表现,与其他指标的相关性越强,该指标与其他指标的冲突性就越小,反映出不同的信息就越少,则它的权重占比应当越小。

对于  $i$  个脉搏特性向量,3 项评价指标形成原始数据矩阵  $\mathbf{T}$  如下:

$$\mathbf{T} = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ \vdots & \vdots & \vdots \\ t_{i1} & t_{i2} & t_{i3} \end{pmatrix} \quad (6)$$

其中: $t_{ik}$  为第  $b$  行向量第  $c$  项评价指标的数值。

为了消除不同量纲对最终结果的影响,首先对上述指标进行无量纲化处理。信噪比的数值越大表明噪声比例越小,信噪比使用正向化处理方式;R-R 间期的方差和 PNN70 的数值越小表明噪声比例越小,因此使用逆向化处理方式。正向化和逆向化处理方式如下:

$$t_{ik}' = \begin{cases} \frac{t_{ik} - t_{\min}}{t_{\max} - t_{\min}}, & \text{正向化处理;} \\ \frac{t_{\max} - t_{ik}}{t_{\max} - t_{\min}}, & \text{逆向化处理} \end{cases} \quad (7)$$

其中: $t_{ik}$  和  $t_{ik}'$  分别为第  $b$  行向量第  $c$  项指标无量纲化处理前、后的数值; $t_{\max}$  为第  $c$  项指标的最大值; $t_{\min}$  为第  $c$  项指标的最小值。

计算各指标的对比强度和指标之间冲突性,计算公式如下:

$$\bar{t}_c = \frac{1}{i} \sum_{b=1}^i t_{bc}' \quad (8)$$

$$S_c = \sqrt{\frac{\sum_{b=1}^i (t_{bc}' - \bar{t}_c)^2}{i - 1}} \quad (9)$$

$$R_c = \sum_{d=1}^3 (1 - r_{cd}) \quad (10)$$

其中: $\bar{t}_c$  为第  $c$  项指标的平均值; $S_c$  为第  $c$  项指标的标准差; $R_c$  为第  $c$  项指标的冲突性; $r_{cd}$  为指标  $c$  与指标  $d$  的相关系数。则第  $c$  项指标的客观权重  $W_c$  如下:

$$W_c = \frac{S_c R_c}{\sum_{c=1}^3 (S_c R_c)} \quad (11)$$

计算信噪比、R-R 间期方差和 PNN70 这 3 项指标对应的客观权重为  $W_1$ 、 $W_2$ 、 $W_3$ ,则最终对特征图中脉搏特性向量所做的权重分配如下:

$$W_b' = W_1 S_{\text{NR}_b} + W_2 S_{\text{R-R}_b}^2 + W_3 P_{\text{NN70}_b} \quad (12)$$

其中: $W_b'$ 、 $W_1 S_{\text{NR}_b}$ 、 $W_2 S_{\text{R-R}_b}^2$  和  $P_{\text{NN70}_b}$  分别为第  $b$  行脉搏特性向量的最终分配权重、无量纲化处理后的信噪比、R-R 间期方差和 PNN70 的值。

对每一张特征图每一行的脉搏特性向量做相同的处理,就得到了该特征图对应的权重向量  $\mathbf{W}'$ 。

在经过权重优化层对特征图中各行脉搏特性向量的权重重新优化后,能让真实心率特征所占的比例大大增加,增强了网络的学习能力和抗运动噪声性能,提高了心率检测的准确性。

本文所用卷积神经网络的层数较少,主要由于深层的神经网络往往会有过拟合的现象发生,而心率是一个相对低层的特征,使用较少的卷积神经网络层数不仅能够提高网络学习效率,也避免了学习过程中对于深层特征的过度关注。每一层网络使用 ReLU 激活函数对特征的非线性性能进行补充。此外,还使用 0.2 的 Dropout 层防止过拟合现象的发生,并将学习率设定为一个变量,每经过一定的训练周期,学习率会乘以一个固定的系数,以避免在预测数据的收敛过程中因梯度来回震荡导致无法取到最小值的问题。选用 L2 范数损失函数<sup>[18]</sup>,即最小平方误差,保证能得到解,且在数据有变化时的解更加稳定、效率更高。它的表达式如下:

$$L_2(\hat{y}, y) = \sum_{q=0}^N (y^{(q)} - \hat{y}^{(q)})^2 \quad (13)$$

其中: $N$  为样本数; $y^{(q)}$  为第  $q$  个样本的真实心率值; $\hat{y}^{(q)}$  为第  $q$  个样本的预测心率值。

## 2 实验结果与分析

### 2.1 数据集

本文所选用的数据集均来自中国科学技术大

学第一附属医院(安徽省立医院)自采数据,采集设备包括 ThinkPad 笔记本电脑 1 部(操作系统为 Windows10,CPU 为锐龙 R7-4800H,内存为 16 GiB,分辨率为  $1\ 920 \times 1\ 080$ ),医院专用脉搏血氧仪(科瑞康 O2Ring 型号)1 个。采集过程中,受试者坐在笔记本电脑前适当位置,使用电脑自带的前置摄像头记录下受试者的面部视频,视频帧率 30 帧/s,1 080 p 显示,并同步使用脉搏血氧仪采集受试者当前状态下的心率作为真值。为方便后续实验,在同一场景下采集了受试者面部静止、面部自由活动 2 种情况下的样本数据,部分样本视频展示如图 4 所示。

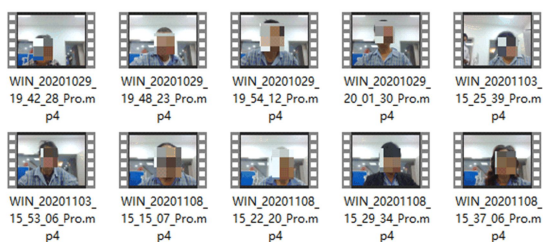


图 4 部分样本视频展示

为了进一步扩大实验数据规模,提高深度学习模型的泛化能力,本文使用数据增强技术,将所有采集到的面部视频数据按帧倒序排列,得到一组新的数据,使得总数据量变为原来的 2 倍,总计 2 200 份。其中包括 550 位受试者,每位受试者分别含有面部静止、面部运动和各自按帧倒序排列的 4 份面部视频样本。本文所用医院自采数据集样本分布见表 1 所列。

表 1 数据集样本分布

年龄/岁	10~<30	30~<50	50~<70	70~<90	合计
男性人数	19	50	123	95	287
女性人数	15	46	104	98	263

## 2.2 评价标准

本文复现了文献[9,11,13]所用方法,并与本文方法进行对比分析,且采用 5 项指标对最终结果定量评价。

第 1 项指标为平均绝对误差  $M_e$ ,是绝对误差的平均值,是样本预测值与真实值最直接误差的表现,计算公式如下:

$$H_{Re}(i) = H_{Rp}(i) - H_{Rg}(i) \quad (14)$$

$$M_e = \frac{1}{N} \sum_{i=1}^N H_{Re}(i) \quad (15)$$

其中: $N$  为实验样本总数; $H_{Rp}(i)$  为对于第  $i$  个样本的预测值; $H_{Rg}(i)$  为第  $i$  个样本的真实值(本文即为脉搏血氧仪测量的心率值)。

第 2 项指标为标准误差  $S_{De}$ ,用来衡量数据的变化量,评估其分散性, $S_{De} \in [0, \infty)$ ,其值越低,表示数据越趋于平均值,计算公式如下:

$$S_{De} = \sqrt{\frac{\sum_{i=1}^N (H_{Re}(i) - M_e)^2}{N}} \quad (16)$$

第 3 项指标为均方根误差  $R_{MSE}$ ,又被称为标准误差,它对一组测量中较大或较小的误差反映非常敏感,能够很好地反映出测量的精密密度,计算公式如下:

$$R_{MSE} = \sqrt{\frac{\sum_{i=1}^N (H_{Rp}(i) - H_{Rg}(i))^2}{N}} \quad (17)$$

第 4 个指标是平均绝对百分比误差  $M_{eRate}$ ,它是对预测准确度的一种量化指标,其值越低表示预测值越接近真实值, $M_{eRate} \in [0, \infty)$ ,计算公式如下:

$$M_{eRate} = \frac{1}{N} \sum_{i=1}^N \frac{|H_{Re}(i)|}{H_{Rg}(i)} \quad (18)$$

第 5 项指标为皮尔逊相关系数  $\rho$ ,它被用于判定预测值与真实值之间的线性相关关系<sup>[19]</sup>,取值范围为  $\rho \in [-1, 1]$ , $-1$  表示负相关, $1$  表示正相关,计算公式如下:

$$\rho = \frac{\sum_{i=1}^N (H_{Rg}(i) - \bar{H}_{Rg})(H_{Rp}(i) - \bar{H}_{Rp})}{\sqrt{\sum_{i=1}^N (H_{Rg}(i) - \bar{H}_{Rg})^2} \sqrt{\sum_{i=1}^N (H_{Rp}(i) - \bar{H}_{Rp})^2}} \quad (19)$$

## 2.3 结果与分析

### 2.3.1 特征矩阵的维度选择

前文利用 PCA 法将滤波后得到的脉搏矩阵  $\delta'$  降维,获得神经网络输入端的特征矩阵,这是为了消除冗余数据,去除部分噪声干扰的同时提高神经网络的学习效率。为了验证特征矩阵的最佳维度,本文首先针对特征矩阵不同维度对最终结果的影响进行了相关对比实验和讨论。

本文分别选取相同场景面部静止和运动 2 种状态下各 1 100 份样本数据进行实验,采用五折交叉验证的方法,计算平均值(后续实验均采用此种验证方法),实验结果见表 2、表 3 所列。

从表 2、表 3 可以看出,不管在运动还是静止状态下,当特征矩阵的维数为 30 时,各项指标均表现为最优值,且表现效果随着维度远离 30 变得

越来越差。因此,本文后续实验验证均基于特征矩阵的维数为 30 进行。

表 2 面部静止状态实验结果

维数	$M_e$	$S_{De}$	$R_{MSE}$	$M_{eRate}/\%$	$\rho$
20	1.32	1.98	1.63	1.64	0.987
25	0.88	1.66	1.26	0.98	0.990
30	0.57	1.10	0.99	0.69	0.994
35	0.96	1.63	1.06	1.22	0.982
40	1.03	1.85	1.62	1.54	0.975

表 3 面部运动状态实验结果

维数	$M_e$	$S_{De}$	$R_{MSE}$	$M_{eRate}/\%$	$\rho$
20	4.00	4.59	3.86	4.88	0.938
25	3.42	3.90	3.50	4.77	0.955
30	2.98	3.55	3.13	3.99	0.968
35	3.15	3.93	3.99	4.24	0.942
40	3.86	4.76	4.19	5.92	0.933

### 2.3.2 静态场景下的心率检测

在面部静止状态下,运动干扰较小,本次实验选用相同场景总共 1 100 份静止状态下的面部视频数据进行验证,结果见表 4 所列。

表 4 静态场景下的方法性能分析

方法	$M_e$	$S_{De}$	$R_{MSE}$	$M_{eRate}/\%$	$\rho$
文献[9]	0.96	1.52	1.43	1.86	0.957
文献[11]	0.84	1.41	1.25	1.44	0.979
文献[13]	0.73	1.23	1.14	1.27	0.984
本文方法	0.57	1.10	0.99	0.69	0.994

从表 4 可以看出,所述 4 种方法的各项指标差异不大,在面部静止状态下心率的预测误差较小,但本文在 5 项指标上均有小幅度地提升,其中  $M_e$  小于 0.60,  $S_{De}$  小于 1.20,  $R_{MSE}$  约等于 1.00,  $M_{eRate}$  低于 0.70%,  $\rho$  大于 0.990。

本文方法在面部静止状态下神经网络心率预测散点结果如图 5 所示。从图 5a 的线性回归图可以看出,散点基本分布在线性回归线周围,且回归线斜率基本为 1。图 5b 的 Bland-Altman 图展示了预测值与心率真实值的误差,黑线表示预测值与真实值的平均绝对误差,红线为 95% 置信区间的分界线,所述样本点大部分分布在距离平均值误差为 2 的范围内,总体平均误差也控制在 1 以内。分析原因可能在于实验中受试者的面部无法保持绝对的静止,会存在细小的运动变化,而本文在特征矩阵的构建和神经网络结构 2 个方面做了优化,可以更好地去掉面部运动干扰,与其他

方法相比心率检测的准确性更高。

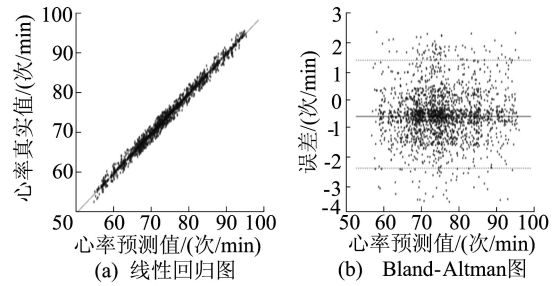


图 5 静态场景下的结果散点图

### 2.3.3 动态场景下的心率检测

面部运动有很多种,如表情变化、头部晃动等,均会对心率检测的准确性造成较大影响,主要原因在于面部运动会引入许多噪声信号,导致原始心率信号失真,也影响了神经网络对真实心率特征的提取与辨别。本次实验选用相同场景总共 1 100 份运动状态下的面部视频数据进行验证,结果见表 5 所列,散点结果如图 6 所示。

表 5 动态场景下的方法性能分析

方法	$M_e$	$S_{De}$	$R_{MSE}$	$M_{eRate}/\%$	$\rho$
文献[9]	5.08	5.65	6.17	6.93	0.798
文献[11]	4.65	5.11	5.81	6.15	0.851
文献[13]	4.26	4.98	5.63	5.88	0.863
本文方法	2.98	3.55	3.13	3.99	0.968

由表 5 可知:较大的运动干扰对心率预测影响较大,文献[9]的  $M_e$ 、 $S_{De}$ 、 $R_{MSE}$  均大于 5,对于其中效果较好的文献[13],这 3 项指标也均大于 4;而在  $M_{eRate}$  这一指标上,上述文献[9, 11, 13]均高于 5%,最低为 5.88%, $\rho$  也是大幅度降低,只能保持在 0.8 附近;而本文方法在各项指标上均有较大幅度提升, $M_e$ 、 $S_{De}$ 、 $R_{MSE}$  均小于 4,  $M_{eRate}$  低于 4%, $\rho$  也超过了 0.96。

图 6a 为运动状态下心率预测的线性回归散点图。从图 6a 可以看出:文献[9, 11, 13]中散点虽然也大多分布在回归线周围,但是偏差范围较大,整体分布不太平均;而本文方法的散点不仅对线性回归线的拟合程度更高,样本分布更加紧凑,而且很少有散点距离回归线非常远的异常预测点。在图 6b 中,文献[9, 11, 13]的 95% 置信区间较宽,样本比较离散,并且绝对误差的平均值也较大,而本文方法样本多数分布在误差平均值周围,95% 置信区间的范围更小,心率检测的准确率更高。

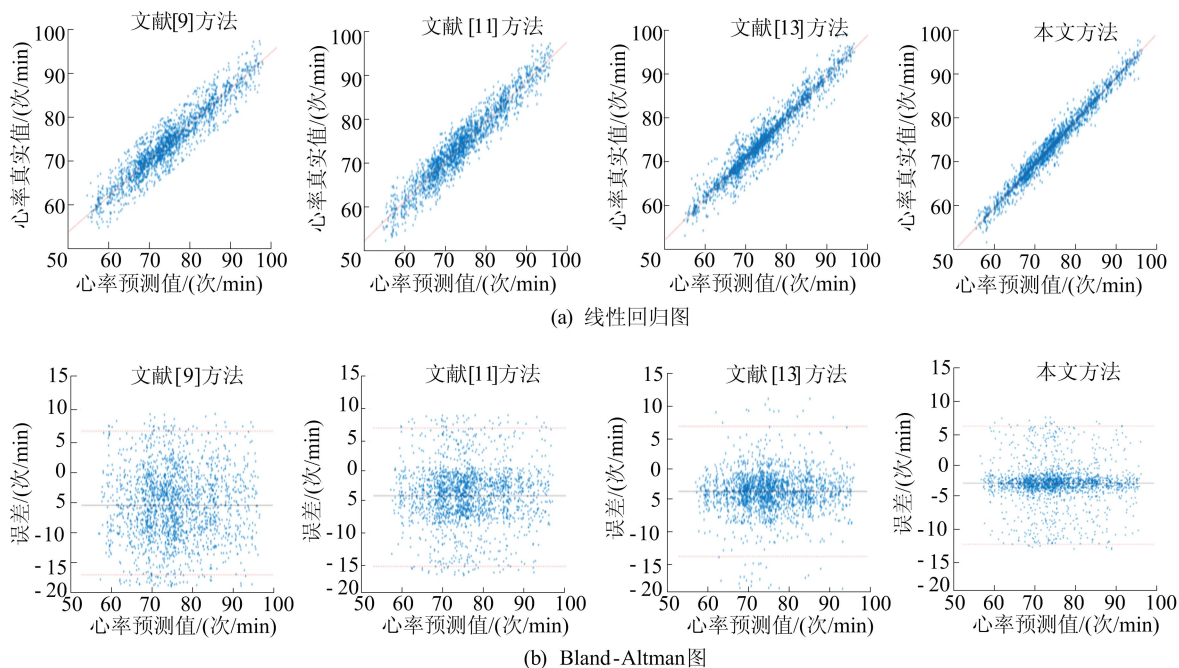


图 6 动态场景下的结果散点图

对实验结果分析如下。

1) 文献[9]利用独立成分分析和小波变换获取心率信号去除噪声,然而小波变换的局限性在于其只能滤除心率频率范围之外的噪声,而由运动干扰导致的尖锐噪声可能会覆盖整个心率频段,部分心率频率范围内的噪声会对最终的结果产生较大影响。

2) 文献[11]将面部视频的不同时空表示输入卷积神经网络回归心率信号,同时使用了预训练的方式增加样本数量。但是预训练模型参数较为繁琐,而且计算量太大,心率检测的时间过长,极大地限制了应用场景。同时,在复杂多变的现实场景中,预训练很难获取到初始样本的共同特征,所得的效果无法达到预期。

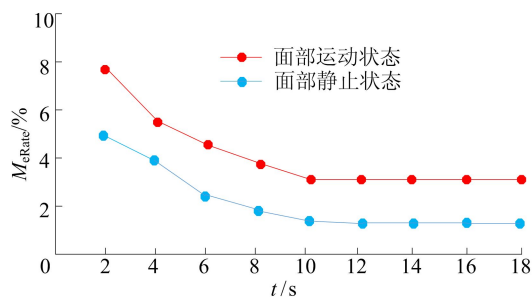
3) 文献[13]通过引入不同心率值之间的数值关系,同时改进面部视频信号的时空表示解决数据缺乏和网络不均衡的问题,所取得的效果较为显著。但该方法只考虑了环境噪声的影响,忽视了面部表情变化等运动造成的信号干扰,同时引入的门控循环单元虽然实现了高效的心率检测,但是对于异常值的兼容性较差,容易受到极端值的影响而大幅度降低准确率。

4) 本文基于权重优化卷积神经网络的非接触心率检测方法,在预处理阶段充分保留原始图像信息的同时滤除运动噪声,对卷积神经网络结构进行优化,在卷积层后加入权重优化层,引入3项重要指标和Critic赋权法对每次卷积后得到

的特征图进行权重优化,调整不同行脉搏特性向量的权重大小,进一步提高了神经网络学习原始心率特征和辨别运动干扰的能力,准确性和鲁棒性均大大提高。

#### 2.3.4 方法稳定性验证

在实际的检测过程中,稳定性受到各种复杂场景的影响,需要考虑检测时间的长短,较短的检测时间可能会导致心率检测的准确性有所降低。为了验证方法的稳定性,本实验将之前采集的面部视频进行分割,分别保留2、4、6、8、10、12、14、16、18 s面部静止与运动状态下的视频,以平均绝对百分比误差 $M_{eRate}$ 作为检验稳定性的唯一指标,从而验证方法在不同心率检测时间下的准确性,结果如图7所示。

图 7 不同检测时间下 $M_{eRate}$ 变化趋势

由图7可知,随着检测时间的增加,面部静止和运动状态下心率预测的平均绝对百分比误差 $M_{eRate}$ 会越来越小,在10 s后基本趋于稳定,在低

于10 s时,误差会显著增加,特别是在面部运动状态下更为明显。原因在于心率是一个周期特征,在面部视频时长太少的情况下,由于运动噪声的影响,在仅有的信号中无法提取出准确的周期信息。因此,为了保证本文方法的准确性,建议视频时长应保持在10 s以上。

### 3 结 论

心率是人体重要的生理指标,近年来 IPPG 技术的迅速发展使得非接触心率检测成为可能,并以其高效准确的特点被广泛运用,但这其中还存在许多难点与挑战<sup>[20]</sup>。本文提出了一种基于权重优化卷积神经网络的非接触心率检测方法,通过对面部脉搏矩阵的构建与处理,有效保留了图像信息并去除噪声,提出权重优化卷积神经网络,在卷积层后增加了权重优化层,优化网络结构,进一步提高了神经网络学习原始心率特征和辨别运动噪声的能力。通过实验验证,相比现有的检测方法准确率大大提高,稳定性也更好,为在现实运动状态中的非接触心率检测提供了一个选择。然而,本文方法在一些特定的场景,如强光、黑暗等光照影响比较大的条件下还无法准确检测心率,下一步将侧重于展开这方面的研究,提高本文方法在更多复杂场景中的实用性。

### [参 考 文 献]

- [1] WU X S. Significance of heart rate in cardiovascular disease [J]. Chinese Journal of Internal Medicine, 2006, 45 (7): 601-602.
- [2] MAZOMENOS E B, BISWAS D, ACHARYYA A, et al. A low-complexity ECG feature extraction algorithm for mobile healthcare applications[J]. IEEE Journal of Biomedical and Health Informatics, 2013, 17(2): 459-469.
- [3] HANNING C D, ALEXANDER-WILLIAMS J M. Pulse oximetry: a practical review[J]. British Medical Journal, 1995, 311(7001): 367-370.
- [4] LIM Y G, KIM K K, PARK S. ECG measurement on a chair without conductive contact[J]. IEEE Transactions on Biomedical Engineering, 2006, 53(5): 956-959.
- [5] 马良. 基于普通摄像头的非接触式生理参数检测技术研究[D]. 济南: 山东大学, 2017.
- [6] 郑鲲, 孔江萍, 周晶, 等. iPPG 技术及生理参数检测的教育应用综述[J]. 计算机工程与应用, 2021, 57(5): 25-35.
- [7] VERKRUYSSSE W, SVAASAND L O, NELSON J S. Remote plethysmo-graphic imaging using ambient light[J]. Optics Express, 2008, 16(26): 21434-21445.
- [8] KUMAR M, VEERARAGHAVAN A, SABHARWAL A. Distance PPG: robust non-contact vital signs monitoring using a camera[J]. Biomedical Optics Express, 2015, 6(5): 1565-1588.
- [9] 刘祎, 欧阳健飞. 基于人脸视频的非接触式心率测量方法[J]. 纳米技术与精密工程, 2016, 14(1): 76-79.
- [10] HSU G S, AMBIKAPATHI A, CHEN M S. Deep learning with time-frequency representation for pulse estimation from facial videos[C]//Proceedings of 2017 IEEE International Joint Conference on Biometrics. [S. l.]: IEEE, 2017: 383-389.
- [11] NIU X S, HAN H, SHAN S G, et al. Synrhythm: learning a deep heart rate estimator from general to specific[C]//Proceedings of the 24th International Conference on Pattern Recognition. [S. l. ; s. n. ], 2018: 3580-3585.
- [12] SPETLIK R, FRANCO V, CECH J, et al. Visual heart rate estimation with convolutional neural network [C]//Proceedings of British Machine Vision Conference. [S. l. ; s. n. ], 2018: 1-12.
- [13] NIU X S, SHAN S G, HAN H, et al. RhythmNet: end-to-end heart rate estimation from face via spatial-temporal representation[J]. IEEE Transactions on Image Processing, 2019, 29(10): 2409-2423.
- [14] 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.
- [15] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.
- [16] 常亮, 邓小明, 周明全, 等. 图像理解中的卷积神经网络[J]. 自动化学报, 2016, 42(9): 1300-1312.
- [17] 林景栋, 吴欣怡, 柴毅, 等. 卷积神经网络结构优化综述[J]. 自动化学报, 2020, 46(1): 24-37.
- [18] 张飞翔, 余学儒, 何卫锋, 等. 结合改进的损失函数与多重范数的人脸识别[J]. 计算机工程与应用, 2020, 56(24): 144-150.
- [19] ADLER J, PARMRYD I. Quantifying colocalization by correlation: the pearson correlation coefficient is superior to the Mander's overlap coefficient[J]. Cytometry Part A, 2010, 77(8): 733-742.
- [20] 姚丽峰. 基于 PPG 和彩色视频的非接触式心率测量[D]. 天津: 天津大学, 2012.

(责任编辑 李 凯)