

基于面部视频心率估计的 MUSIC 算法分析及仿真复现

李珈锋¹, 龙小毅¹, 高嘉骏¹, 韩星月¹, 董金城¹

(1 四川大学, 四川 成都, 610000)

摘要: 由于传统测心率的方法舒适度与便携性较差, 适应场景较少, 近年来, 非接触式心率检测的需求逐渐增大。早期的大量学者基于传统测心率的光电容积体积分描记法, 利用 MUSIC 算法来实现非接触式心率估计。首先对面部划分感兴趣区域 ROI, 提取各区域的 rPPG 信号进行分析, 然后用 MUSIC 算法进行去噪和估计, 最终得出心率, 但由于光照、运动伪影等因素的影响使得估计结果准确性有限。针对上述问题, 对 MUSIC 算法的原理进行分析, 对比得出可行性, 并进行相关性分析, 以此剔除有噪声影响的数据, 提高准确性, 最后进行仿真复现。

关键词: MUSIC 算法; 面部视频; rPPG 信号; 心率估计

Analysis and Simulation Reproduction of MUSIC Algorithm Based on Facial Video Heart Rate Estimation

Li Jia-feng¹, Long Xiao-yi¹, Gao Jia-jun¹,

Han Xing-yue¹, Dong Jin-cheng¹

(1 Sichuan University, Sichuan Chengdu, 610000)

Abstract : Due to the poor comfort and portability of traditional heart rate measurement methods and limited adaptability to different scenarios, the demand for non-contact heart rate detection has gradually increased in recent years. Many early scholars used the MUSIC algorithm to achieve non-contact heart rate estimation based on the traditional photoelectric volumetric recording method for measuring heart rate. Firstly, the ROI of the area of interest is divided into different regions, and the rPPG signals of each region are extracted for analysis. Then, the MUSIC

algorithm is used for denoising and estimation to obtain the heart rate. However, due to factors such as lighting and motion artifacts, the accuracy of the estimation results is limited. In response to the above issues, analyze the principle of MUSIC algorithm, compare its feasibility, and conduct correlation analysis to eliminate data with noise effects, improve accuracy, and finally simulate and reproduce.

Key words: MUSIC algorithm; facial video; rPPG signal; heart rate estimation

1 引言

心率变化与心脏疾病密切相关，所以对心率的检测尤为重要。近年来研究发现传统心率检测设备价格昂贵，舒适度差，适用场景少，无法实现实时监测。随着现代科技的发展，基于面部视频的非接触式心率检测成为了一个备受关注的研究方向。这种方法不用任何传感器或设备接触身体，仅通过对面部视频信号进行分析，就能够准确地检测出人体的心率，具有便捷、无损伤、实时性强等优点，因此被广泛应用于医学诊断、健康监测以及情感识别等领域。

基于面部视频的非接触式心率检测在国外已经得到了广泛的研究。在早期的研究中，研究者主要利用颜色变化、脉冲波形^[1]和心跳峰值等特征来提取心率信息。但是由于图像噪声等因素的影响，这些方法并不能很好地抑制干扰，并且对于光线、摄像头质量等环境条件也有较高的要求。

近年来，随着深度学习等技术的发展，基于面部视频的非接触式心率检测在国外取得了很好的研究成果。例如，Somekh 等人使用了深度学习算法^[2]，通过对大量数据的训练，成功地提取出了面部皮肤颜色变化与心率之间的关系，从而实现了高精度的心率检测。另一方面，Xu 等人将深度卷积神经网络应用于面部视频信号的处理，不仅能够有效地消除光线、噪声等因素带来的影响，还能够对多种情况下的心率进行准确检测。

基于 rPPG 信号的心率计算在国内的研究起步较晚，但也取得了很多成功。张等使用 CHROM 算法提取出 rPPG 信号并生成时特征图，利用 ResNet-18 从该特征图中预测出心率值。胡等人将池化层从主干网络中独立出来，形成注意力机制，并利用 ConvLSTM 对时域特征进行提取，从人脸图像序列中得到了 rPPG 信

号并预测出了心率值。陆等人使用红外视频对 3 个感兴趣区域使用独立成分分析与多变量经验分解算法，得到了较为精确的心率值。目前基于视频提取仍旧在 ROI 选取、光照强度变化、人体肤色、头部运动等方面存在技术缺点，对面部的分割与跟踪，光照影响的干扰都亟待解决。

本文从面部视频估计心率的角度出发，利用 MUSIC 算法在去噪方向的高效性，分析此算法在心率估计领域的应用，最后进行仿真实验和相关性分析。通过实验结果和其他文献结果进行对比，来验证 MUSIC 算法的可靠性和准确性。

2 原理分析

2.1 传统 MUSIC 算法

传统 MUSIC 算法主要用于声源定位，由 Schmidt 等人于 1979 年提出，是一种基于子空间分解的算法，它利用信号子空间和噪声子空间的正交性，构建空间谱函数，通过谱峰搜索，估计信号的参数。对于声源定位来说，需要估计信号的 DOA。MUSIC 算法对 DOA 的估计有很高的分辨率，且对麦克风阵列的形状没有特殊要求，因此应用十分广泛。

假设由信号源发出的信号为 $y(n)$ ，其形式为：

$$y(n) = \sum_{k=1}^K \alpha_k e^{j\omega_k n} \quad (1)$$

其中 $\alpha_k = |\alpha_k|e^{j\varphi_k}$ 与 ω_k 分别是信号复幅度与角频率。初始相位 φ_k 是在 $[0, 2\pi]$ 均匀分布的独立随机变量。假设接收到的信号为 $x(n)$ ，则：

$$x(n) = y(n) + e(n) \quad (2)$$

其中 $e(n)$ 是与信号独立、均值为 0、方差为 σ_e^2 的高斯白噪声。

对给定的正整数 M ，记：

$$\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-M+1)]^T \quad (3)$$

$$\mathbf{a}(\omega) = [1, e^{-j\omega}, \dots, e^{-j(M-1)\omega}]^T \quad (4)$$

$$\mathbf{A} = [\mathbf{a}(\omega_1), \mathbf{a}(\omega_2), \dots, \mathbf{a}(\omega_K)] \in \mathbb{C}^{M \times K} \quad (5)$$

则信号的协方差矩阵 $\mathbf{R} \in \mathbb{C}^{M \times M}$ 为：

$$\mathbf{R} = E[\mathbf{x}(n)\mathbf{x}^H(n)] = \mathbf{A}\mathbf{A}^H + \sigma_e^2 \mathbf{I} \quad (6)$$

其中：

$$\mathbf{P} = \text{diag}[|\alpha_1|^2, |\alpha_2|^2, \dots, |\alpha_K|^2] \sigma_e^2 \quad (7)$$

由于矩阵 \mathbf{A} 的秩 $r(\mathbf{A})=K$ ，易知 $r(\mathbf{A}\mathbf{P}\mathbf{A}^H) = K$ ，故 $\mathbf{A}\mathbf{P}\mathbf{A}^H$ 有 K 个非零特征值。记 $\mathbf{A}\mathbf{P}\mathbf{A}^H$ 的所有特征值为 $\mu_1, \mu_2, \dots, \mu_M$ ，对应的正交归一化特征向量为 $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M$ 。不妨设非零特征值为 $\mu_1, \mu_2, \dots, \mu_K$ 。

事实上，协方差矩阵 \mathbf{R} 的 M 个特征值与 $\mathbf{A}\mathbf{P}\mathbf{A}^H$ 的 M 个特征值之间只相差一个 σ_e^2 ，记 \mathbf{R} 的 M 个特征值为 $\lambda_1, \lambda_2, \dots, \lambda_M$ ，则有：

$$\lambda_i = \begin{cases} \mu_i + \sigma_e^2, & i = 1, 2, \dots, K \\ \sigma_e^2, & i = K + 1, \dots, M \end{cases} \quad (8)$$

且 $\lambda_1, \lambda_2, \dots, \lambda_K$ 对应的特征向量同样为 $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M$ 。

由此可知，协方差矩阵 \mathbf{R} 的 M 个特征值中，前 K 个与信号 $\mathbf{y}(n)$ ，噪声 $\mathbf{e}(n)$ 都有关，后 $M - K$ 个只与噪声有关。

定义信号子空间 E_S 是由 $\lambda_1, \lambda_2, \dots, \lambda_K$ 对应的特征向量张成的空间，即：

$$E_S = \text{span}\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_K\} \quad (9)$$

同理，噪声子空间 E_N 是由 $\lambda_{K+1}, \dots, \lambda_M$ 对应的特征向量张成的空间，即：

$$E_N = \text{span}\{\mathbf{p}_{K+1}, \dots, \mathbf{p}_M\} \quad (10)$$

有如下结果：

$$\mathbf{a}^H(\omega_k) \mathbf{p}_i = 0, k = 1, 2, \dots, K, i = K + 1, \dots, M \quad (11)$$

即噪声子空间特征向量与信号频率向量正交。对任意给定的正整数 k ，有：

$$|\mathbf{a}^H(\omega_k) \mathbf{p}_i|^2 = 0, i = K + 1, \dots, M \quad (12)$$

对上式两边关于 i 求和，则有：

$$\sum_{i=K+1}^M |\mathbf{a}^H(\omega_k) \mathbf{p}_i|^2 = 0 \quad (13)$$

记以噪声子空间特征向量 $\mathbf{p}_{K+1}, \dots, \mathbf{p}_M$ 为列向量的矩阵为 \mathbf{N} ，即：

$$\mathbf{N} = [\mathbf{p}_{K+1}, \dots, \mathbf{p}_M] \in \mathbb{C}^{M \times (M-K)} \quad (14)$$

就可以把求和式改写为：

$$\mathbf{a}^H(\omega_k) \mathbf{N} \mathbf{N}^H \mathbf{a}(\omega_k) = 0 \quad (15)$$

由于在实际应用中，协方差矩阵的估计值 $\hat{\mathbf{R}}$ 与理论值不严格相等，故 $\mathbf{a}^H(\omega_k) \mathbf{N} \mathbf{N}^H \mathbf{a}(\omega_k)$ 并不严格等于 0，而是十分接近 0。据此，可以构造谱函数

$$\hat{P}_{MUSIC}(\omega) = \frac{1}{\mathbf{a}^H(\omega) \mathbf{N} \mathbf{N}^H \mathbf{a}(\omega)}, \omega \in [-\pi, \pi] \quad (16)$$

在计算时，令 ω 取遍 $[-\pi, \pi]$ 之间的所有值，当 $\hat{P}_{MUSIC}(\omega)$ 取到极大值时，对应的 ω 就是所求频率。

2.2 心率估计中的 MUSIC 算法

基于面部视频的心率估计总体步骤如下：

- 1.通过对视频中面部颜色变化获得周期性的光电容积脉搏波(PPG)信号。
- 2.对 PPG 信号进行过滤处理，减小不良血液灌流和运动伪影对数据影响。
- 3.从数据中删除平均和线性趋势
- 4.应用 MUSIC 算法：

由于在测量心率时，接收到的信号不是完全精确，要用以下方式估计协方差矩阵：

假设接收到的心率信号长度为 L ，即 $\mathbf{x} = [x(1), x(2), \dots, x(L)]$ ，现取长度为 $M \leq L$ 的区间，并记：

$$\mathbf{x}_i = [x(i), x(i+1), \dots, x(i+M-1)]^T \quad (17)$$

则可用以下公式估计心率信号 \mathbf{x} 的协方差矩阵 $\hat{\mathbf{R}}$ ：

$$\hat{\mathbf{R}} = \frac{1}{L-M+1} \sum_{i=1}^{L-M+1} \mathbf{x}_i \mathbf{x}_i^T \quad (18)$$

其中 M 为适当选取的参数。

再计算协方差矩阵的特征值 λ_i 和特征向量，得到 \mathbf{R} 的特征值分为 K 个较大特征值 μ_i 和 $M-K$ 个较小的接近的特征值 φ_i ，即：

$$\lambda_i = \begin{cases} \mu_i, & i = 1, 2, \dots, K \\ \varphi_i, & i = 1, 2, \dots, M-K \end{cases} \quad (19)$$

其中， $\mu_1, \mu_2, \dots, \mu_K$ 对应的特征向量为 $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_K$ ，而 $\varphi_1, \varphi_2, \dots, \varphi_{M-K}$ 对应的特征向量为 $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_{M-K}$

在估算心率时，仅考虑有意义心率限值内的频率。搜索频率向量定义为：

$$\mathbf{a}(\omega) = [1, e^{-j\omega}, \dots, e^{-j(M-1)\omega}]^T \quad (20)$$

利用形成的信号或噪声子空间的伪谱：

$$\hat{P}_{MUSIC}(\omega) = \frac{1}{\mathbf{a}^H(\omega) \mathbf{N} \mathbf{N}^H \mathbf{a}(\omega)}, \omega \in [-\pi, \pi] \quad (21)$$

找到峰值即可。找出 MUSIC 的伪谱峰值即为心率估计值。其中 $\mathbf{N} =$

$[q_1, q_2, \dots, q_{M-K}] \in \mathbb{C}^{M \times (M-K)}$ 是以噪声子空间特征向量 q_1, q_2, \dots, q_{M-K} 为列向量的矩阵。

3 实验仿真

3.1 传统 MUSIC 仿真

为验证理论的正确性、完备性，本实验基于 Matlab 进行仿真，通过预先设置变量以及信号预定义的方式来进行仿真，观察仿真结果，判断波达角估计的准确性。

由于 MUSIC 算法对于不同的协方差矩阵处理方式不同，因此本实验采用两种方法(特征值分解 EVD 与奇异值分解 SVD)分别进行实验。特征值分解主要针对方阵，而奇异值分解主要针对非方阵。下图展示了两种方式的仿真结果。

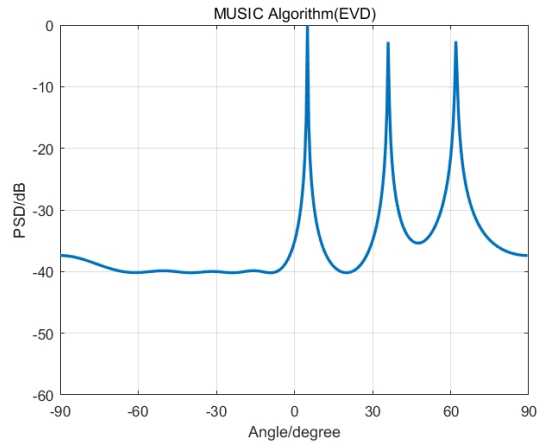


图 1 EVD 实现的 MUSIC 算法仿真结果

Fig.1 Simulation results of MUSIC algorithm implemented by EVD

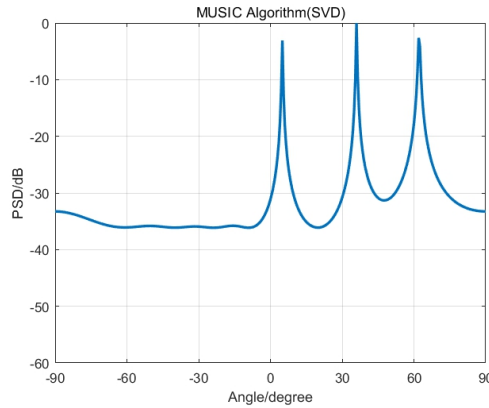


图 2 SVD 实现的 MUSIC 算法仿真结果

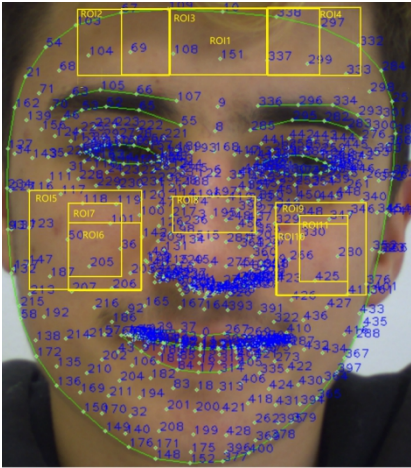
Fig.2 Simulation results of MUSIC algorithm implemented by SVD

本实验预设有 3 个信号源，波达角角度分别为 5° ， 36° 和 62° ，阵元有 8 个，假设阵元间距为 0.5，采样点数为 500，信噪比为 10dB，通过加入高斯白噪声的方式来模拟较为真实的情况。

由图 1 和图 2 可得，两种方式的估计结果均较为准确，说明 MUSIC 算法具有一定可行性。

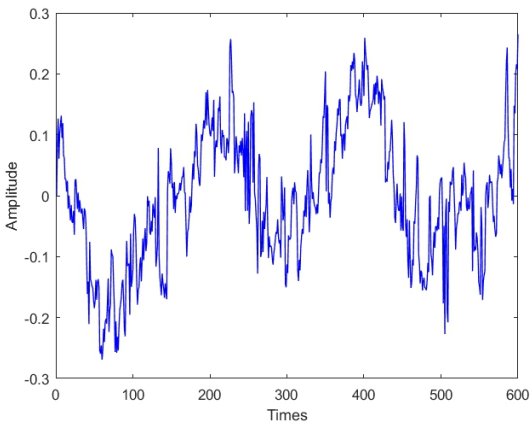
3.2 基于心率估计的 MUSIC 算法仿真

本实验利用 UBFC-rPPG 数据集集中的样本，先对特定的一帧图片进行 ROI 划分，可以得到已经滤波后的 rPPG 信号。



(a) ROI 区域(黄色框)

(a) The area of ROI



(b) 时域图

(b) Time-domain diagram

图 3 rPPG 信号提取

Fig.3 rPPG signal extraction

由图 3 (a)，本实验共划分 11 个 ROI 区域，每个 ROI 区域中包含有 rPPG 信号，可得到时域图 (b)。最后带入 MUSIC 算法中即可得到频率，换算为心率即为最终结果。但是值得注意的是，此 ROI 没有经过滤波处理，会导致最后的估计结果不准确。

4 相关性分析

在得到一段视频中若干个 ROI 的原始信号后，应该对其进行相关性分析，

以此判断 ROI 信号的有效性，并决定某个 ROI 是否要参与心率估计。

假设在一段视频中共提取 m 个 ROI，每个 ROI 对应一个长度为 N 的时间序列，记第 k 个 ROI 对应的时间序列为：

$$\mathbf{x}_k = [x_1^{(k)}, x_2^{(k)}, \dots, x_N^{(k)}] \quad (22)$$

记 $E(\mathbf{x}_k)$ 是 \mathbf{x}_k 的各个分量的均值，即：

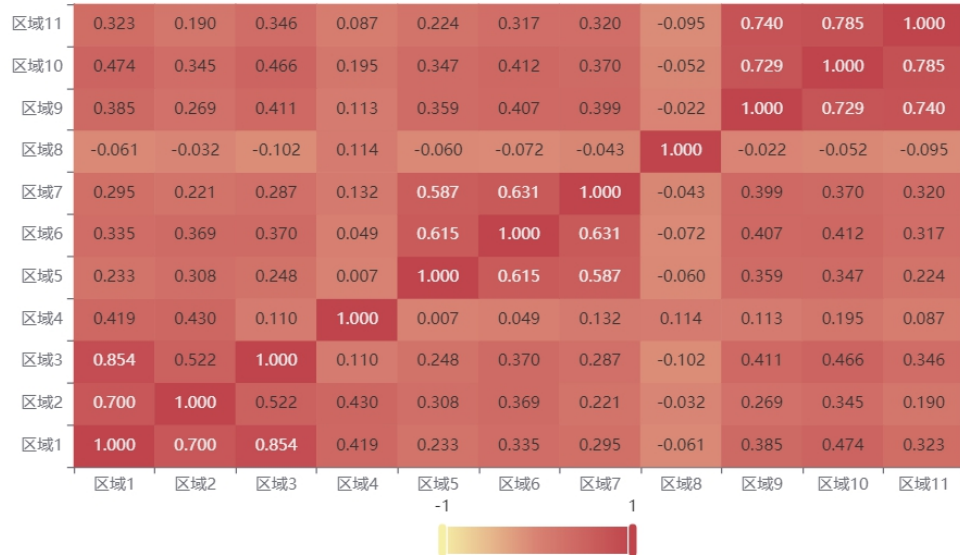
$$E(\mathbf{x}_k) = \frac{1}{N} \sum_{i=1}^N x_i^{(k)} \quad (23)$$

记第 p 个 ROI 和第 q 个 ROI 时间序列的皮尔逊相关系数为 $\rho_{p,q}$ ：

$$\rho_{p,q} = \frac{\sum_{i=1}^N x_i^{(p)} x_i^{(q)} - N \times E(\mathbf{x}_p) E(\mathbf{x}_q)}{\sqrt{\sum_{i=1}^N (x_i^{(p)})^2 - N(E(\mathbf{x}_p))^2} \sqrt{\sum_{i=1}^N (x_i^{(q)})^2 - N(E(\mathbf{x}_q))^2}} \quad (24)$$

这样，第 k 个 ROI 可以算出 $m-1$ 个相关系数。若进行异常值检测，可用如下方法：将这 $m-1$ 个相关系数取平均，均值记为 ρ_k 将 $\rho_1, \rho_2, \dots, \rho_m$ 降序排列，若最小的小于 0.7，（比如 ρ_6 ）则删除第六个 ROI 的所有相关系数，再重新计算均值，直到最小的均值也大于 0.7，余下的 ROI 即为有效区域。

在实验中共从数据集中获取三组数据，以下为 SPSSPRO 绘制的相关系数热力图：



(a) 数据集 1

(a) The first dataset



(b) 数据集 2

(b) The second dataset



(c) 数据集 3

(c) The third dataset

图 4 相关系数图

Fig.4 Correlation coefficient graph

由图 4，方格颜色越深，表示两个区域相关系数越大，也就是正相关性越强。由图可以看出，数据集 1 的相关性整体弱于数据集 2 和 3，且数据集 1 的 4 号、8 号 ROI，数据集 2，3 的 8 号 ROI 相关性表现较差，即该区域表现不佳，在后

续心率计算中可考虑去除。

5 结语

本文详细介绍了 MUSIC 算法在心率检测中的数学原理，该算法是一种经典的信号处理算法，可以用于定位信号源和估计信号参数等。本文基于该算法的原理进行了模型仿真，并比较了两种特征值求解算法的结果，以及针对不同信噪比和发射信号源等情况进行了仿真，结果表明该仿真模型是正确且准确的。并且本文对各种不同噪声下生产的信号进行了相关性分析处理，在使用 ROI 提取出的 11 个区域中两个区域数据与其他区域数据的相关性存在明显的差异，意味着该数据不准确，会对心率检测结果有影响。剔除该组数据后数据整体的相关性较强，这对利用 MUSIC 算法进行心率检测将具有积极作用，可以提高心率检测的准确性和可靠性。

由于心率接收信号与 MUSIC 算法本身之间存在差异，本文指出，对于基于人脸的 ROI 信号提取进行的 G 通道信号的相关性处理和 MUSIC 算法的应用将是未来的研究重点。研究人员还需要深入研究如何将 MUSIC 算法应用于心率检测，并考虑如何处理各种信号和噪声条件下的情况，以提高算法的鲁棒性和可靠性。这些研究将为未来开发更先进的心率检测技术提供重要的参考和指导。

参考文献

- [1] Poh M-Z, McDuff D J, Picard R W. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation[J]. Optics express, 2010, 18(10): 10762-10774.
- [2] JA Crowe, NM Gibson, MS Woolfson, Michael Geoffrey Somekh Journal of biomedical engineering 14 (3), 268-272, 1992
- [3] Chen W, McDuff D. DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 349-365.
- [4] Wang F, Jiang M, Qian C, Yang S, Li C, Zhang H, et al. Residual attention network

for image classification[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern

Recognition. 2017: 3156–3164.

[5] Stricker R, Müller S, Gross H M. Non-contact video-based pulse rate measurement on a mobile service robot[C]. The 23rd IEEE International Symposium on Robot and Human Interactive Communication. IEEE, 2014: 1056–1062.

[6] S. H. Fouladi, I. Balasingham, T. A. Ramstad and K. Kansanen, "Accurate heart rate estimation from camera recording via MUSIC algorithm," 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 2015, pp. 7454–7457, doi: 10.1109/EMBC.2015.7320115.

[7] Bousefsaf F, Pruski A, Maaoui C. 3D Convolutional Neural Networks for Remote Pulse Rate Measurement and Mapping from Facial Video[J]. Applied Sciences, 2019, 9 (20).

[8] Špetlík R, Franc V, Matas J. Visual heart rate estimation with convolutional neural network[C] Proceedings of the British Machine Vision Conference, Newcastle, UK. 2018: 3–6.

[9] Perepelkina O, Artemyev M, Churikova M, Grinenko M. HeartTrack: Convolutional neural network for remote video-based heart rate monitoring[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020: 288–289.

[10] Lee E, Chen E, Lee C Y. Meta-rppg: Remote heart rate estimation using a transductive meta learner[C]. European Conference on Computer Vision. Springer, Cham, 2020: 392–409.