No-contact Heart Rate monitoring based on Channel Attention Convolution Model

Wen Sun, Hao Wei, Xueen Li
Institute of Automation Chinese academy of Sciences Beijing, China
e-mail: {sunwen2017, weihao2016, xueen.li}@ia.ac.cn

ABSTRACT

No-contact heart rate monitoring based on remote Photoplethysmography(rPPG) via camera video has drawn more and more attention because of its promising use in patient nursing, telemedicine, fitness, trial. Many traditional signal processing methods (FFT, ICA, PCA) were proposed to solve this problem, but the results were still limited to interference of motion and lighting conditions. In facial RGB images, the signal-to-noise ratio of green channel is higher than that of the other two channels, and the heart rate can be measured more accurately by assigning different weights to three channels. In this paper we propose a novel deep convolution neural network model based on channel-attention mechanism to extract the heart rate information from each frame of the video. To get more accurate result of the heart rate in the condition of face moving, light change and other interference factors, the model was trained on the newly introduced public challenge ECG-Fitness database and the model's robustness was tested on this dataset. Testing results show that the model outperforms previous methods.

Keywords: Heart Rate, Channel Attention Mechanism, Convolution Neural Network, ECG-fitness Data-set

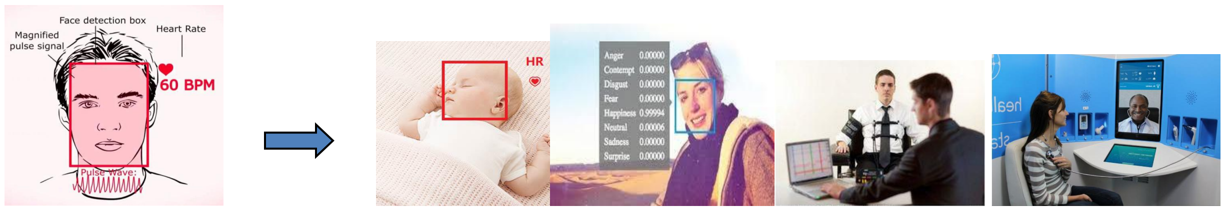
1. INTRODUCTION

Heart Rate (HR) is an important physiological parameter of human body, which can reflect people's health and even psychological activities. Traditional heart rate measurement methods usually rely on sensors that touch the human body, as well as professional operators. However non-contact HR estimation methods relying on a camera are more convenient, low-cost and non-invasive. The heart rate can be measured through the cardiovascular activities, therefore we can extract the heart rate value by analysing the color change of our face region captured by the video camera caused by the blood circulation. Researchers now are more interested in remote HR monitoring methods, and many valid methods have emerged [2], [3], [4], [5], [6], [7].

Published HR estimation methods rely on handcraft features and some signal processing methods, these methods are susceptible to light and motion, so researchers need to come up with more complex filtering methods to eliminate the effects of interference but without effective improvement. Recently, data driven methods especially deep-learning methods have led to a series of breakthrough for image classification, feature extraction, object detection and many other visual tasks in an end-to-end trained manner, results proved these deep-leaning methods were better than those handcraft methods in model building because of its no-linear characteristic. In the last year or two, some deep learning methods have been used to process face sequences and get heart rate values [8], [9], [10],[11], [12] and some self-collected database are also available to the public. Li et al. [8] firstly introduced the MAHNOB-HCI database, another challenge database called ECG-Fitness was introduced by [7], this database is designed to include light changes and subject's motion, so the trained model will be more robust and accurate.

This paper proposed a deep convolution neural network based on channel attention mechanism to selectively emphasize informative features(G-channel) and suppress less useful ones (R, B channels), the details of this channel attention model are shown in Figure 2.

In this paper, (i)Savitzky-Golay smoothing filter method has been applied to verify the rPPG principle and calculate the heart rate value. (ii) in order to get more accurate heart rate value, we developed a channel attention mechanism methods to reform the deep neural network model to emphasize green channel information and weaken other channel information because of the high signal- to- noise ratio of the G-channel, and displayed the corresponding test results on the ECG-fitness public database. The results showed a smaller estimation error result than some other related methods.



Remote heart rate monitoring   Sleep monitor   Emotion analysis   Identifying defection   Telemedicine

Figure 1. Heart Rate monitoring based on face video have a wide variety of practical applications

2. RELATED WORK

This section introduces some existing methods applied to remote HR estimation.

The basic approach for most existing remote HR monitoring methods is Photoplethysmography(PPG). Most of the measurement rely on specific narrowband wavelengths lighting conditions and sensors. The first non-contact remote HR estimation system using just a conventional camera (in daylight and normal office fluorescent lighting) was reported by Verkruysse et al. [13]. Paper [14] extract underlying source signals from R, G, B color bands using computer webcam, finger BVP sensor, and proposed FFT methods to extract HR value.

H.Rahman [3] shows that three different signal processing methods such as Fast Fourier Transform(FFT), Independent Component Analysis(ICA) and Principal Component Analysis(PCA) have been applied on color channels of the facial regions. Another attempt was done by Toshihiro [2] to located the facial ROI region including the cheeks and nose because the whole face image is subject to some interference such as eye blinks, mouth motion and hair disturbing, so the cheeks and nose are the most stable image areas, then the author applied a smoothing processing to extract the heart rate information, but all of the work was done only on G-channel.

The above work was conducted almost using traditional image filtering method, although some can remove some light and small motion interference, there is still much room to improve the accuracy and real time performance. Recently deep learning methods have emerged. A spatial-temporal representation conception was presented in paper [12], they use transfer learning strategy to train a deep heart rate estimator using a pre-training model. Chen et al. [10] showed that the green channel of a conventional camera provided the strongest PPG signal since hemoglobin light absorption is most sensitive to oxygenation changes for green light. In addition, the red and blue channels also contained some PPG information [17]. Paper [1] put up with a two-step convolutional neural network trained end-to-end by altering optimization and validated on the public available database, what's more a database contains light changes and facial motion was designed for research.

Recently, attention mechanism has been emerged to make the network modules achieve improved performance. The attention mechanism was inspired by the human attention mechanism. When people observe images, they don't look at each pixel of the whole image at a time, and most of them focus on specific parts of the image according to their needs. Moreover, humans will learn from the previously observed images to observe where the image attention should be concentrated in the future. Many spatial attention methods were proposed to solve visual tasks. After then model based channel attention emerged [15], Jie Hu put up with a squeeze-and -excitation(SE)[16] networks models, these model focus on the importance of different channels to the final effect, and this SE-Networks impose only a slight increase in model complexity and computational burden, and SE block can also be used to solve the problem of which channels have different effects. Inspired by the characteristics of the SE block, we proposed a convolutional neural network based on channel attention SE block. The rest of the paper is organized as follows: chapter 3 describes methods and materials. chapter 4 showed experiment procedure and results, chapter 5 makes a conclusion of the work and discusses future work.

3. METHODS

3.1 Channel Attention Convolution Model

As indicated in figure2, we utilized channel attention model to reform deep neural network. Some traditional methods only exploit the green channel to extract the heart information for this channel has high signal-to-noise ratio, indeed the other two channels also have useful information, by giving different weights to different channels we can get more accurate results. This attention model contains two operation: squeeze and excitation operation as follows:

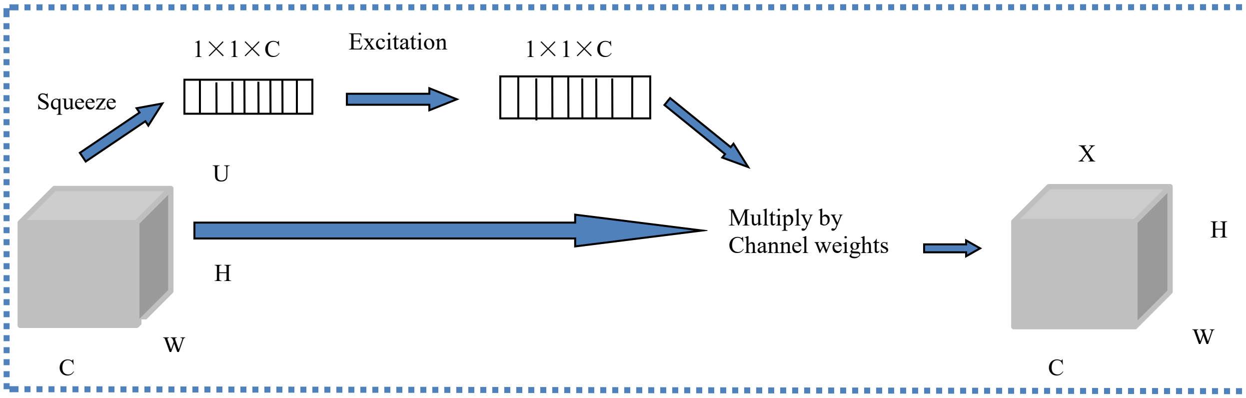


Figure 2. An illustration of channel attention model

Squeeze operation is achieved by using global average pooling to generate channel-wise descriptor , a descriptor z$\in R^c$, U's dimension changes from $H \times W \times C$ to $1 \times 1 \times C$ by decreasing it's spatial dimension, the process can be explained as follows:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(i,j) \tag{1}$$

Excitation operation aims to fully capture the channel-wise dependencies. After global average pooling operation, then two fully connected layers around a RELU and a sigmoid are applied to learn nonlinear interaction between channels and learn a non-mutually exclusive relationship.

$$s = F_{excitation}(z,W) = \sigma(g(z,W)) = \sigma(W_2\delta(W_1 z)) \tag{2}$$

δ refers to RELU function, $W_1 \in R^{\frac{C}{R}\times C}$,$W_2 \in R^{C\times\frac{C}{R}}$, σ refers to sigmoid function. The final output can be obtained by channel-wise multiplication.

$$X = s_c \cdot u_c \tag{3}$$

3.2 Channel Attention Convolution Model

The model was trained with two stage by end-to-end manner on ECG-fitness dataset, the loss of first stage is as follows:

$$l(\tau;\phi) = -\sum_{j=1}^{l} SNR(f^j, X^j;\phi) \tag{4}$$

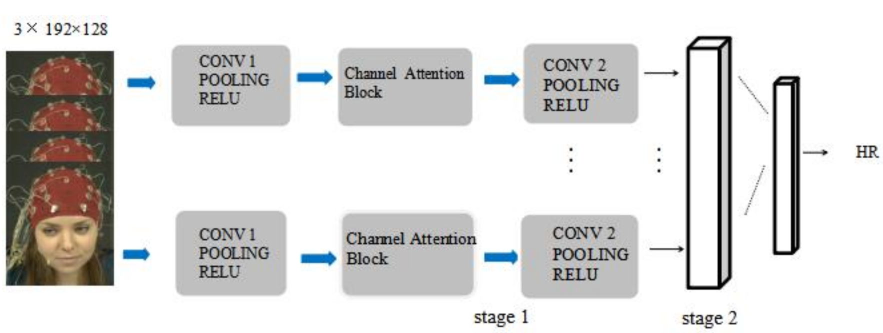


Figure 3. An illustration of channel Attention Convolution model

By minimizing the loss function we can get the parameter ϕ, the SNR means the signal-to- noise ratio, the concept of SNR was introduced by [16], in SNR the F^{+} means the frequency of the signal, $\mathrm{F} \backslash F^{+}$ denotes the noise frequency. The nominator of the SNR represents the strength of the true HR signal, the denominator illustrates the energy of the background noise. Given the true HR, the frequency of the heart rate can be calculated, given the frequency of the heart rate then the PSD function can be captured.

$$SNR(f, X; \phi) = \log_{10}\left(\sum_{\hat{f} \in F^{+}} PSD(f, \hat{X}; \phi) / \sum_{\hat{f} \in F \backslash F^{+}} PSD(\hat{f}, X; \phi)\right) \tag{5}$$

PSD denotes the power spectral density, h($\cdot$) means the output of the first stage cnn-network for every image and ϕ represents all convolutional filters parameters, X=($x_1, \ldots, x_N$) is a sequence of n facial images, and

$$PSD = \left(\sum_{n=0}^{N-1} h(x_n; \phi) \cdot \cos(2\pi \hat{f} \frac{n}{f_s})\right)^2 + \left(\sum_{n=0}^{N-1} h(x_n; \phi) \cdot \sin(2\pi \hat{f} \frac{n}{f_s})\right)^2 \tag{6}$$

The second training stage is another cnn-network which taking one-dimensional signal, and produces the HR value. Loss of this stage was defined as follows:

$$l(\tau; \theta) = \frac{1}{l} \sum_{j=1}^{l} \left| g(h(x_1; \phi), \ldots, h(x_N; \phi); \theta) - f^{j} \right| \tag{7}$$

f^{j} is the true HR value, while the g($h(x_1; \phi), \ldots, h(x_N; \phi); \theta$) is the predicted HR value, by minimizing the average loss between the predicted and true HR, we get the final parameter.

4. EXPERIMENTS AND RESULTS

4.1 The Validation of the RPPG Principle

Before using deep neural network to solve the problem, we proposed a time domain filtering algorithm called savitzky-golay to verify the principle of the RPPG. Firstly, we use common camera to capture the 4 second face video with 30fps, and the obtained ROI area is extracted frame by frame like Figure 3, after extracting the green channel of each ROI, calculate the average pixel value of the ROI. The mean value changes with time as shown in Figure 4, heart rate cannot be obtained from intensity curve due to the existence of motion and other unrelated interference, after applying the filtering algorithm, Figure 5 shows that about 3peaks for 3seconds. HR is calculated as:

$$HR = 60 / 4 * 4 = 60\text{bpm} \tag{8}$$

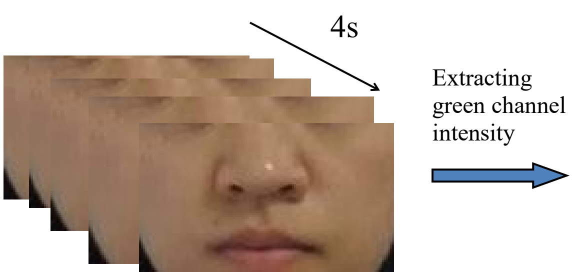


Figure 4. Time sequence of ROI images

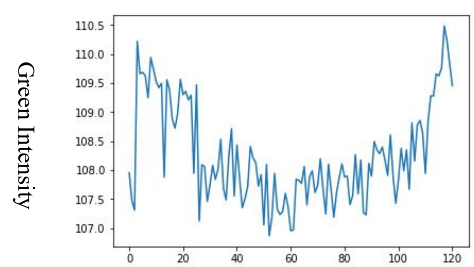


Figure 5. Green channel average intensity of ROI

Savitzky-Golay smoothing filter is a commonly used filtering method in spectral preprocessing. Its core idea is to perform k-order polynomial fitting on data points in a certain length window to obtain the fitted result. After discretizing it, S-G filtering is actually a weighted average algorithm for moving windows, but its weighting factor is not a simple constant window, but rather a least-squares-fit to a given high-order polynomial within a sliding window. The formula of the algorithm is as follows. C_i denotes the weight coefficient.

$$Y_j^* = \frac{\sum_{i=-m}^{m} C_i Y_{j+1}}{N} \tag{9}$$

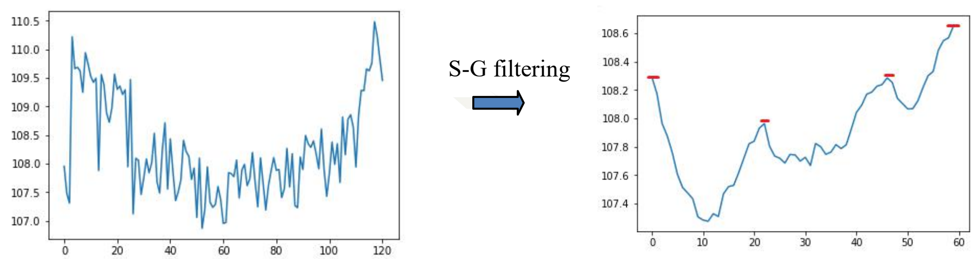


Figure 6. Curve of green channel average intensity of ROI

Figure 7. Curve of l average intensity of green channel after filtering

4.2 Training and Testing

We have trained our networks with adam optimizer utilizing Pytorch machine learning system running on Nvidia GPU with four TITAN XP. As shown in figure 6 the attention model was inserted in blocks consist of CONV, POLING, RELU, batch normalization was also used in the network. During the trainning stage, data augmentation methods such as random cropping, random mirror, random Gaussian blur were used to resist overfitting, the epoch and batch-size were designed as 100 and 300. In ECG-fitness about 17 testers are required to performing different activities (speaking, rowing, exercising) in different environment (lighting changes and interference etc.) in a one-minute video.

ECG-fitness	HR_{rmse} (bpm)	HR_{mae} (bpm)
2SR	52.86	21.39
CHROM	33.47	10.21
HR-CNN	19.15	7.25
(proposed method)	10	3.77

Table 1. An Illustration of the result of proposed method on ECG-fitness data-set comparing with other results.

faces were found by a face detector, and the ratio was designed about 3:2 to cut the whole face, then resized to 192 ×128. The other training procedure generally follows Špetlík R et al. [1]

During testing stage due to time constraints, we have no time to collect our own data sets for testing, so we used part of the ECG-fitness (The same lighting condition, no motion interference) to test the channel attention convolution network model's performance. Results on ECG-fitness show the effectiveness of the proposed method that about 10bpm error. The HR_{rmse} indicates root-mean-square error of the heart rate, HR_{rmse} and HR_{ame} are the criterion to measure the accuracy of the heart rate. Table 1 displays that our proposed method out-perform the network without channel attention model. What's more we also compared the training time of the model and model inserted with channel attention, result shows that attention model improved the performance of previous network without much increasing the computational complexity.

5. CONCLUSION

We proposed to use a channel attention model SE-net to reform the convolution network owing to different channels of image have different importance on inferring HR, after including the preference of channels, the model can learn the weights of channels better. And we will focus more attention on how to detect the special ROI region in the following work.

REFERENCES

[1] Špetlík R, Franc V, "Visual Heart Rate Estimation with Convolutional Neural Network[J]," in BMVC,2108.

[2] Kitajima T, Choi S O, Murakami E A Y. Heart rate estimation based on camera image[C]//2014 14th International Conference on Intelligent Systems Design and Applications. IEEE, 2014: 50-55.

[3] Rahman H, Ahmed M U, Begum S, et al. Real time heart rate monitoring from facial RGB color video using webcam[C]//The 29th Annual Workshop of the Swedish Artificial Intelligence Society (SAIS), 2–3 June 2016, Malmö, Sweden. Linköping University Electronic Press, 2016 (129).

[4] 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.

[5] Pursche T, Krajewski J, Moeller R. Video-based heart rate measurement from human faces[C]//2012 IEEE International Conference on Consumer Electronics (ICCE). IEEE, 2012: 544-545.

[6] Li X, Chen J, Zhao G, et al. Remote heart rate measurement from face videos under realistic situations[C]//Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2014: 4264-4271.

[7] Lam A, Kuno Y. Robust heart rate measurement from video using select random patches[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2015: 3640-3648.

[8] Hsu G S, Ambikapathi A M, Chen M S. Deep learning with time-frequency representation for pulse estimation from facial videos[C]//2017 IEEE International Joint Conference on Biometrics (IJCB). IEEE, 2017: 383-389.

[9] Tang C, Lu J, Liu J. Non-contact heart rate monitoring by combining convolutional neural network skin detection and remote photoplethysmography via a low-cost camera[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. 2018: 1309-1315.

[10] 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.

[11] Gotlibovych I, Crawford S, Goyal D, et al. End-to-end deep learning from raw sensor data: Atrial fibrillation detection using wearables[J]. arXiv preprint arXiv:1807.10707, 2018.

[12] Niu X, Han H, Shan S, et al. Synrhythm: Learning a deep heart rate estimator from general to specific[C]//2018 24th International Conference on Pattern Recognition (ICPR). IEEE, 2018: 3580-3585.

[13] W.Verkruyse, L.O. Svaasand, and J. S. Nelson, "Remote plethysmo-graphic imaging using ambient light. " Opt. Express, vol. 16, no. 26, pp.21434-21445, 2008

[14] De Haan G, Jeanne V. Robust pulse rate from chrominance-based rPPG[J]. IEEE Transactions on Biomedical Engineering, 2013, 60(10): 2878-2886.

[15] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 3-19.

[16] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition (CPVR). 2018: 7132-7141.

[17] G. de Haan and V. Jeanne. Robust Pulse Rate from Chrominance-Based rPPG. IEEE Transactions on Biomedical Engineering, 60(10):2878 – 2886, October 2013. ISSN 0018-9294. doi: 10.1109/TBME.2013.2266196