Non-contact heart rate monitoring by combining convolutional neural network skin detection and remote photoplethysmography via a low-cost camera

Chuanxiang Tang[1] Jiwu Lu[1] Jie Liu[1]
jie_liu@hnu.edu.cn
[1]Hunan University
Changsha, China

Abstract

In this paper, we present a versatile methodology to accomplish the non-contact monitoring of heart rate signals in unconstrained environments, by combining the convolutional neural network (CNN) skin detection and the camera-based remote photoplethysmography (rPPG) methods. Compared to the widely-used three-step skin detection method (i.e., face detection, face tracking, and skin classification), the CNN method used here could enhance the monitoring robustness by achieving the skin detection in a single step. The proposed CNN-rPPG method has been tested in an unconstrained office environment to validate its applicability. Combined with the subsequent rPPG heart rate monitoring based on a low-cost camera, the method presented here is of practical interests for the large-scale deployment of the non-contact heart rate monitoring technologies.

1. Introduction

Due to its convenience to achieve harass-free continuous monitoring of the vital signs (e.g., the heart rate), the camera-based remote photoplethysmography (rPPG) methods have spurred intensive research interests in the past decade. In the recent years, a lot of rPPG algorithms have been proposed, e.g., the independent component analysis (ICA) method [11], the principal component analysis (PCA) [2], the sub-band rPPG method [14], etc., which are promising for the next-generation non-contact monitoring of vital signs in various unconstrained environments, like hospitals, homes, offices, gymnasiums, and public space, etc.

Though these camera-based rPPG algorithms achieve the non-contact vital signs monitoring by using different methodologies, they share a common step C the skin detection. Since the heart rate signals are extracted from the intensity of the electromagnetic waves (e.g., the visible light during daytime, the infrared light during nights) reflected from the skin region, a robust method to tell the skin region from the non-skin region in the video is the indispensible beginning step for all non-contact rPPG algorithms. The conventional skin detection methods typically consist of three consecutive steps: (1) the face detection, (2) the face tracking, and (3) the skin classification. These conventional skin detection methods suffer from two main drawbacks. Firstly, they are limited to the face skin detection. Secondly, their final detection results hinge on the success of each of the three steps in a consecutive manner.

These two drawbacks are noteworthy obstacles that hinder the potential large-scale deployment of the promising rPPG technologies. In a lot of application scenarios, only non-face skin regions, e.g., arms, legs, etc., are exposed to the camera. Though these non-face skin regions display highly valuable vital sign signals, the conventional three-step skin detection method cannot be applied. Furthermore, in the three-step skin detection method, the failure of the face detection step or the face tracking step will disable the subsequent rPPG functionality, imposing unnecessary constraints on the posture of the subjects being monitored.

As an early-stage attempt to remove these two inconvenient drawbacks, in this paper we apply the convolutional neural network (CNN) method to detect skin regions for the subsequent rPPG heart rate monitoring. Actually, recent studies have shown that the CNN is capable of achieving robust pixel-level skin detection [4, 8, 17, 19, 18],indicating promising application scenario in the vital sign rPPG monitoring [5, 6, 10, 13, 15, 16, 3].For instance, Chaichulee et al. applied multi-task CNN to accurately detect the baby skin, and proved that the CNN is highly adaptable to different skin tones, postures, and illumination conditions [3].

However, the multi-task CNN was tested using an expensive industry-level camera. To enable large-scale deployment, it deserves research efforts to test it using more affordable cameras. In addition, although it is known that the CNN could achieve the pixel-level skin detection, the direct combination of CNN and rPPG (CNN-rPPG) has not been thoroughly investigated, and it is still an open question that how well the CNN-based skin detection method could help

improving the non-contact rPPG heart rate monitoring.

In this paper, non-contact heart rate monitoring is tested based on the CNN-rPPG method, by using an inexpensive commodity-level camera (about 30 US dollars). We first use the CNN to accomplish skin detection. Then, the rPPG technologies are used to extract the heart rate signals from the detected skin regions. The results of the CNN-rPPG method are validated by comparing against the heart rate ground truth, in an unconstrained office environment.

In Section 2, the CNN-rPPG methodology is introduced. In Section 3, the results are shown and validated by comparing with the ground truth of a contact-based heart rate monitor in the office environment. In Section 4, the generality of the CNN-rPPG methodology is studied and verified. In Section 5, a conclusion is made.

2. Methodology

As shown in the schematic flowchart in Figure1, the CNN-rPPG methodology consists of four parts. Firstly, a video database containing both the skin regions and the non-skin environment is created. Secondly, the 8-layer CNN is trained using the database to tell the skin regions from the non-skin regions. Thirdly, the trained CNN is used to detect the skin regions in new videos. Fourthly, from the detected skin regions, the rPPG algorithms are applied to obtain the heart rate.

These four parts are introduced in the Section 2.1, Section 2.2, Section 2.3, and Section 2.4, respectively.

2.1. Video database creation

To prepare the skin region samples, a total number of 45 videos that contain the skin region are recorded, and each video is 1 minute long. From each recorded video, 10 frames are extracted at equal intervals. From the 45 videos, 30 videos are used for training the CNN model and 15 videos are used for testing the CNN-rPPG method.

To prepare the non-skin region samples, a total number of 407 frames are used. These frames contain non-skin indoor environment objects like walls, desks, chairs, cups, books, and windows, etc.

As shown in Figure 2, the video database is created in the typical indoor environment (e.g. offices, laboratories, and libraries, etc.) with ordinary ceiling illumination (e.g. 16 watts fluorescent lamp).10 male subjects and 5 female subjects (about 20 to 30 years old, yellow skin) are asked to sit still or exercise gently. The subjects are about 2 meters away from a low-cost camera (ELP-USB30W04MT-RL21 model, 640×480 pixels resolution, 30 frames per second, around 30 US dollars). During the video recording, the heart rate ground truth of these subjects is recorded by using a contact-based heart rate monitor, as shown in Figure 1.

2.2. CNN model training

The training of the CNN model is performed using the MatConvNet package [12]. To train the CNN model, the skin and non-skin regions are manually segmented from the video frames. The manually-segmented skin regions are used as positive samples. The non-skin video frames, as shown in the Figure 2(b), are used as negative samples.

The eight-layer CNN model structure is shown in Table 1. The first two layers of CNN are used for low-level feature extraction. The first layer is a convolution kernel with a size of 5×5, generating 20 feature maps. The pool layer of the second layer performs maximum down-sampling of 2×2 on the basis of the output of the previous convolution layer without overlapping. In the next two alternating layers, the convolution layer has a 10×10 kernel, and the max-pooling layer has a 2×2 kernel, which generates 200 feature maps for detecting higher-level image features. The next fifth convolution layer has 500 feature maps, and the kernel size is set to 10×10. The Relu layer follows immediately after the convolution layer. The remaining convolutional layer and Softmax layer implement the classification function. During the CNN model training, each sample image was normalized to a size of 64×64, and the learning rate was set to 0.0005, the batch size is set to 200, and the training time periods is 30.

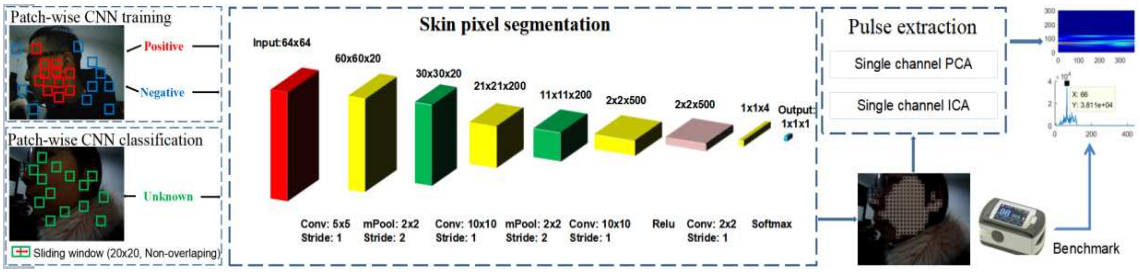
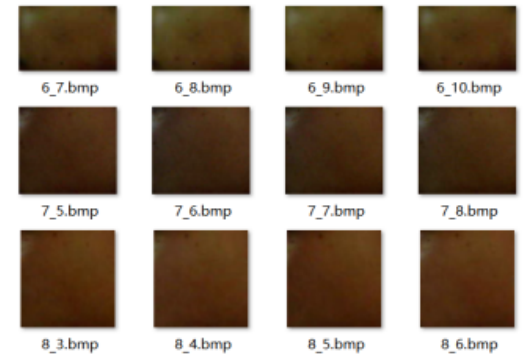
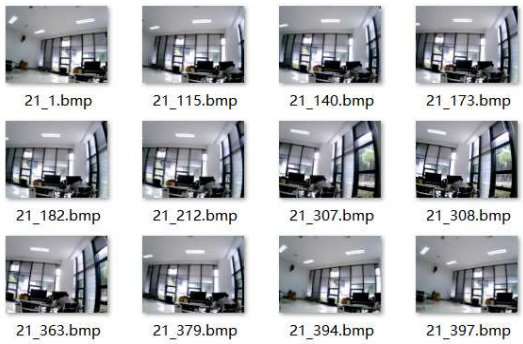


Figure 1. Schematic flowchart of the proposed methodology to achieve non-contact heart rate monitoring, by combining the convolutional neural network (CNN) skin detection and the rPPG technologies (CNN-rPPG). The CNN-rPPG results are validated by comparing against the ground truth of a contact-based heart rate monitor.



(a)



(b)

Figure 2. The selected positive-sample (skin) samples segmented from the video frames (a) and the selected negative (non-skin) video frames (b) used to train the CNN model. In the positive skin samples, no face detection or tracking is needed. The CNN-based skin detection is accomplished directly in the pixel-level.

CNN layer	Layer type	Number of Feature map	Input size
Layer 1	Conv.	20	64×64
Layer 2	mPool	20	60×60
Layer 3	Conv.	200	30×30
Layer 4	mPool	200	21×21
Layer 5	Conv.	500	11×11
Layer 6	Relu	500	2×2
Layer 7	Conv.	4	2×2
Layer 8	Softmax	1	1×1

Table 1. Parameters of the CNN model, where Conv. represents a convolutional layer, mPool represents a max-pooling layer, Relu represents an excitation layer, and softmax represents a classification layer.

2.3. Skin detection

Using the CNN mode trained in Section 2.2, the skin regions are detected from the 15 testing videos as mentioned in Section 2.1.

During skin detection, the sliding window size is 20×20 without overlapping. The positive training samples are the manually-segmented skin regions, which contain no clear edge contour information. So, the sliding movement goes through, without overlapping, all the 20×20 small equally cut windows from the full testing image. Only when both the texture information and color feature are close enough the ones trained by CNN models, the small window will be retained.

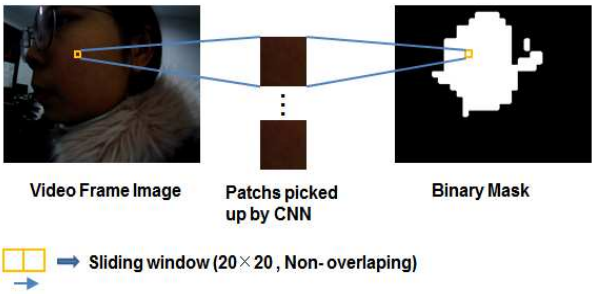


Figure 3. Schematic skin detection process using the CNN-based method.

2.4. Heart rate extraction

The cameras typically use an 8-bit integer to represent one channel of a pixel, i.e., v_i=0 to 255 where i=red, green, blue are the three channels. Due to the low-cost requirements for large-scale deployment, v_i is generally subject to a Δv_i=$\pm$1 or $\pm$2 noise introduced by the low-quality sensor, lens, or peripheral electrical circuits, etc.

In the CNN-based skin detection step, this noise has small impacts on the detection results. However, in the heart rate extraction step, the Δv_i has considerable impacts, since the heart beats could cause a change of only around $\pm$2 to $\pm$3 to the v_i.

Fortunately, the Δv_i typically has different magnitudes for different channels in a low-cost camera. Therefore, though multi-channel-based rPPG algorithms typically outperform the single-channel-based rPPG algorithms while using the expensive industry-level cameras [14], the rPPG algorithms based on single-channel video inputs could potentially offer more reliably heart rate extraction results by selecting the least noisy channel from the low-cost commodity-level cameras. Furthermore, the single-channel rPPG deserves special attention for versatile large-scale applications, since it is useful for heart rate extraction from the infrared cameras during the nights.

Therefore, here we use the green channel of the low-cost camera as input (Section 2.1), to extract HR signals using the well-established single-channel independent component analysis (ICA) method [7] and single-channel principal component analysis (PCA) method[2], in four consecutive steps. In the first step, the green channel of the pixels

in the skin regions detected using the CNN model (Section 2.3) is averaged into a single-channel time-dependent signal, x_t.

In the second step, by following the methodology used in [9], the delay matrix

$$X = \begin{pmatrix} x_t & x_{t+\tau} & \cdots & x_{t+N_\tau} \\ x_{t+\tau} & x_{t+2\tau} & \cdots & x_{t+(N+1)_\tau} \\ \vdots & \vdots & \ddots & \vdots \\ x_{t+(m-1)_\tau} & x_{t+m_\tau} & \cdots & x_{t+(N+m-1)_\tau} \end{pmatrix} \quad (1)$$

is constructed from x_t, where τ is the lag term which is set as $\frac{1}{f_s}$ ms, where f_s is the number of frames per second used to record the video.

In the third step, the single-channel ICA method or the single-channal PCA method are used, in order to process the raw input data to enhance the quality of the following HR extraction calculation. The single-channel ICA calculations are performed by using the FastICA package, which is based on the Hyvarinen's fixed-point algorithm [1]. The ICA is applied to decompose the delay matrix, $\mathbf{X}$, into a series of independent components, s_i. The p ($p < m$) independent components with the largest signal noise ratio (SNR)

$$SNR = \frac{\max_j |F_j|^2}{\sum_j |F_j|^2} \quad (2)$$

is selected, to construct the embedding matrix

$$B^i = a_i s_i^T, (i = 1, 2, ..., p) \quad (3)$$

in the measurement space. Here, F_j is the j^{th} component of the fast Fourier transform (FFT) of the independent component in the frequency-domain; and a_i is the corresponding column of the mixing matrix. Then, the output time series

$$y(t) = \frac{1}{m} \sum_{k=1}^{m} B^i_{k,t+k-1} \quad (4)$$

could be obtained by averaging the rows of the matrix B^i.

Alternatively, the single-channel PCA method can be used to generate the output time series, in the third step. In the single-channel PCA calculations, the singular value decomposition

$$\mathbf{X}^T\mathbf{X} = \mathbf{W}\hat{\mathbf{\Sigma}}^2\mathbf{W}^T \quad (5)$$

is used to replace the ICA to achieve decomposition, where $\mathbf{W}$ is the eigenvector matrix, and the singular values σ_k of $\mathbf{\Sigma}$ and the eigenvalues λ_k of $\mathbf{X}\mathbf{X}^T$ satisfy the relation $\sigma_k = \sqrt{\lambda_k}$.

In the fourth step, the output time series from the single-channel ICA/PCA calculation is used as the input, to run the FFT calculations. The peak of the FFT spectrum corresponds to the extracted HR value.

3. Results

In this section, the heart rate results are obtained. Firstly, the video quality required by the single-channel ICA and PCA rPPG algorithms is verified using a manually-selected rectangular reference skin region of interests (RoI), which is introduced in Section 3.1. Then, by combining the CNN-based skin detection and the single-channel ICA/PCA-based rPPG, the heart rate signals are extracted, as presented in Section 3.2. In Section 3.3, a comparison between the reference RoI method and the CNN-based RoI method is presented.

3.1. Reference RoI method

As a comparison, first the skin RoI is manually selected as shown in Figure 4 for the ICA and PCA rPPG heart rate extraction.

The results are shown in Table 2, which are in decent agreement with the ground truth measured by using the contact-based monitor.

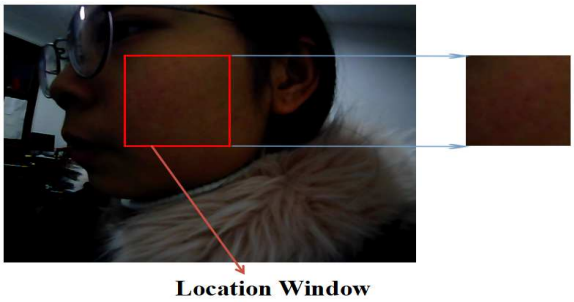


Figure 4. Schematic skin detection process using the manually-selected reference RoI method.

3.2. CNN-based RoI method

After verifying that the video quality of the low-cost camera could satisfy the rPPG requirement (Section 3.1), the CNN-rPPG method is applied in this section.

As shown in Figure 5, the CNN training errors decrease gradually. After a certain number of epochs, the training errors approach zero, indicating the success of the training of CNN models. Then, the trained CNN model is validated by using the 15 testing videos as introduced in Section 2.1. As shown in the Figure 6, the skin/non-skin pixels can be reliably segmented based on the CNN model.

From the skin pixels segmented by the CNN model, the ICA and PCA rPPG heart rate (HR) signals are extracted, as shown in Table 3. It can be seen that all rPPG HR results are within the range [R-3, R+3] beats per minute (bpm), where R is the ground truth value of the HR measured by using the contact-based monitor.

subject	True HR(bpm)	Reference RoI method			
		ICA		PCA	
		HR	err	HR	err
1	54	54.5	0.5	54.3	0.3
2	64	63.3	-0.7	63.1	-0.9
3	75	72.8	-2.2	72.5	-2.5
4	73	73.3	0.3	74.0	1.0
5	79	77.8	-1.2	78.5	-0.5
6	64	64.6	0.6	64.4	0.4
7	64	65.7	1.7	65.4	1.4
8	62	61.7	-0.3	61.5	-0.5
9	75	77.7	2.7	77.4	2.4
10	57	57.5	0.5	57.3	0.3
11	85	84.7	-0.3	84.4	-0.6
12	68	70.7	2.7	70.4	2.4
13	61	62.6	1.6	62.3	1.3
14	69	69.9	0.9	69.6	0.6
15	62	61.7	-0.3	62.4	0.4
Average error			1.10		1.03

Table 2. Heart rate (HR) results, in the unit of beats per minute (bpm), computed by using the manually-selected rectangular reference RoI skin detection and the ICA and PCA heart rate extraction rPPG algorithms.

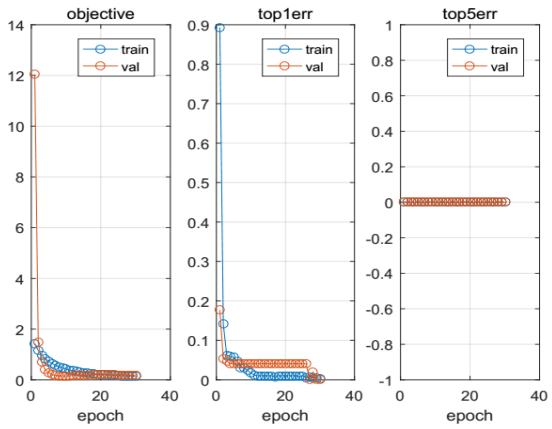


Figure 5. Error reduction process of the CNN model training process.

3.3. Comparison and discussion

As shown in Table 2 and Table 3, both RoI selection methods could offer reliable skin detection results for the subsequent single-channel rPPG extraction from the low cost camera. However, the CNN-based RoI skin detection is the only viable choice in the real application scenarios, since it is impractical to manually segment skin pixels from the widely deployed rPPG monitors.

While comparing the results in Table 2 and Table 3, it worth noting that the ground truth measurement device only has an accuracy of ± 1 bpm. Therefore, one should not infer that reference RoI method (average errors: 1.10 and 1.03) is better than the CNN-based RoI method (average errors: 1.23 and 1.08).

Figure 6. The skin detection results (top panel) and the binary skin mask (bottom panel) obtained by using the trained CNN model.

As shown in Table 3, both the ICA and the PCA could obtain accurate HR results with the help of the CNN-based skin detection. This indicates that the CNN model is compatible with different single-channel rPPG methods.

It worth noting that the above-mentioned testing of the CNN-rPPG method is done in unconstrained office environments with different illumination (e.g., day/night, bright/dark) and background (e.g., wall, windows, desks, chairs, etc.) conditions. Also, the CNN-rPPG works well to monitor subjects with various postures and true HR values.

As a continuation of the prior work of CNN-based skin segementation [3], which is used in a simple neonate incubator environment, this paper further proves the versatility of the CNN-rPPG methodology. The robustness (compared to the widely-used three-step skin detection), the adaptability (in various unconstrained environments), and the low-cost feature of the CNN-rPPG demonstrated in this paper pave a way for next-generation large-scale deployment of the promising camera-based rPPG technologies.

As an early-stage exploration of the CNN-rPPG applicaitons, this paper used the green channel camera signals to realize the single-channel ICA/PCA HR extraction, since the green channel is known to display strong rPPG signals. To explore more versatile applications (e.g., the totally dark nights), it deserve future attention to research the single-channel CNN-rPPG based on low-cost infrared cameras.

4. General applicability

To further verify that the proposed CNN-rPPG methodology is generally applicable in different unconstrained environments, two additional sets of video data are recorded,

subject	True HR(bpm)	CNN-based RoI method			
		ICA		PCA	
		HR	err	HR	err
1	54	54.5	0.5	54.2	0.2
2	64	61.6	-2.4	62.2	-1.8
3	75	72.7	-2.3	72.4	-2.6
4	73	73.7	0.7	73.4	0.4
5	79	78.6	-0.4	78.4	-0.6
6	64	64.6	0.6	64.3	0.3
7	64	65.6	1.6	65.3	1.3
8	62	61.6	-0.4	61.3	-0.7
9	75	77.7	2.7	77.3	2.3
10	57	77.7	2.7	77.3	2.3
11	85	84.8	-0.2	84.4	-0.6
12	68	70.5	2.5	70.3	2.3
13	61	62.6	1.6	62.3	1.3
14	69	70.6	1.6	70.3	1.3
15	62	61.67	-0.47	62.3	0.3
Average error			1.23		1.08

Table 3. Heart rate (HR) results, in the unit of beats per minute (bpm), computed by using the CNN-based RoI skin detection and the ICA and PCA heart rate extraction rPPG algorithms.

besides the testing video data as introduced in Section 2.1. Each set of additional video data sets consists of 8 different subjects in a totally different office environment, as shown in Figure 7 and Figure 8.

As shown in Figure 7 and Figure 8, the CNN skin detection model could reliably achieve the pixel-wise skin segmentation, which paves a way for the subsequent reliable rPPG HR extraction. As shown in Table 4 and Table 5, when applied on different subjects in different unconstrained environments, the proposed CNN-rPPG method could obtain HR results that match well with the ground truth HR.

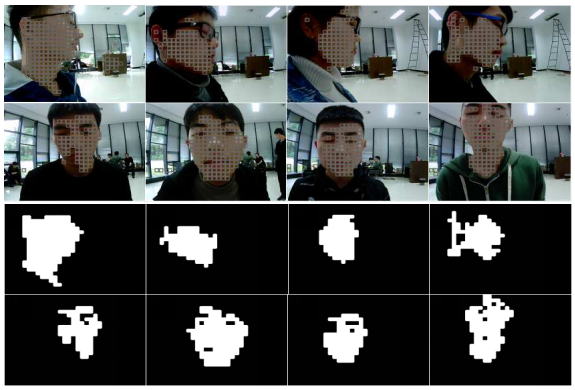


Figure 7. The snapshots (top panel) and the binary skin mask (bottom panel) of additional video data set #1, where the binary skin mask is obtained by using the trained CNN model as introduced in Section 3.2.

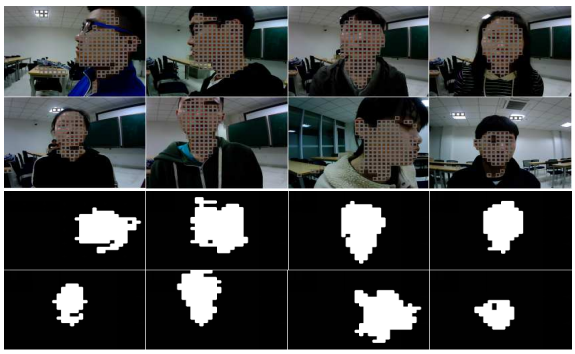


Figure 8. The snapshots (top panel) and the binary skin mask (bottom panel) of additional video data set #2, where the binary skin mask is obtained by using the trained CNN model as introduced in Section 3.2.

subject	True HR(bpm)	CNN-based RoI method			
		ICA		PCA	
		HR	err	HR	err
1	62	59.4	-2.6	62.3	0.3
2	76	77.6	1.6	77.5	1.5
3	62	61.4	-0.6	61.3	-0.7
4	78	81.0	3	76.2	-1.8
5	53	53.3	0.3	53.2	0.2
6	60	57.1	-2.9	57.2	-2.8
7	63	63.3	0.3	60.6	-2.4
8	59	57.5	-1.5	60.3	1.3
Average error			1.60		1.38

Table 4. Heart rate (HR) results of the additional video data set #1 computed by using the proposed CNN-rPPG method.

subject	True HR(bpm)	CNN-based RoI method			
		ICA		PCA	
		HR	err	HR	err
1	79	82.7	3.7	81.3	2.3
2	56	54.4	-1.6	54.5	-1.5
3	89	90.5	1.5	91.4	2.4
4	81	79.4	-1.6	79.3	-1.7
5	85	85.4	0.4	85.4	0.4
6	60	59.5	-0.5	61.3	1.3
7	72	73.6	1.6	72.3	0.3
8	80	78.7	-1.3	79.3	-0.7
Average error			1.53		1.33

Table 5. Heart rate (HR) results of the additional video data set #2 computed by using the proposed CNN-rPPG method.

5. Conclusion

In this paper, we achieved non-contact monitoring of the heart rate (HR) by combining the convolutional neural network (CNN) skin detection and the camera-based remote

photoplethysmography (rPPG) method. Though it is known that the CNN could achieve skin detection reliably, this is the first time according to our knowledge that the CNN skin detection and the rPPG technologies are directly combined to accomplish the non-contact HR monitoring.

The CNN-rPPG methodology presented here has several noteworthy merits for the potential large-scale deployment of the rPPG technologies. Firstly, due to the flexibility of the pixel-level skin detection of the CNN method, the CNN-rPPG method is capable of monitoring the HR in the unconstrained environments, by extracting the vital sign signals from any face skin region or non-face skin region.

Secondly, it is demonstrated that the CNN-rPPG method works well using a low-cost commodity-level camera, which is highly desirable for affordable large-scale applications.

Thirdly, it is shown that the CNN-rPPG method could accomplish the non-contact HR monitoring by using the single-channel video stream input, which is promising for applications during nights based on the gray scale infrared video input. Compared to the existing multi-channel rPPG algorithms, the single-channel CNN-rPPG method used here offers additional freedom to choose the least noisy channel of the inexpensive low-quality cameras and hardware.

Due to these merits, the CNN-rPPG method, as demonstrated in this concept-proof preliminary research, is promising to realize low-cost large-scale application of the next-generation non-contact vital sign monitoring.

References

[1] `http://www.cis.hut.fi/projects/ica/fastica/`.

[2] G. Balakrishnan, F. Durand, and J. Guttag. Detecting pulse from head motions in video. In *Computer Vision and Pattern Recognition*, pages 3430–3437, 2013.

[3] S. Chaichulee, M. Villarroel, J. Jorge, C. Arteta, G. Green, K. Mccormick, A. Zisserman, and L. Tarassenko. Multi-task convolutional neural network for patient detection and skin segmentation in continuous non-contact vital sign monitoring. pages 266–272, 2017.

[4] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *Computer Science*, (4):357–361, 2014.

[5] Z. Cui, J. Yang, and Y. Qiao. Brain mri segmentation with patch-based cnn approach. In *Control Conference*, pages 7026–7031, 2016.

[6] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical Image Analysis*, 36:61, 2016.

[7] M. Z. Poh, D. J. Mcduff, and R. W. Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering*, 58(1):7, 2011.

[8] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651, 2014.

[9] Y. Sun, S. Hu, V. Azorin-Peris, S. Greenwald, J. Chambers, and Y. Zhu. Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise. *Journal of Biomedical Optics*, 16(7):077010, 2011.

[10] Y. Sun and N. Thakor. Photoplethysmography revisited: from contact to noncontact, from point to imaging. *IEEE Trans Biomed Eng*, 63(3):463–477, 2016.

[11] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. Pugh. Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiological Measurement*, 35(5):807–831, 2014.

[12] A. Vedaldi and K. Lenc. Matconvnet: Convolutional neural networks for matlab. pages 689–692, 2014.

[13] W. Wang, A. C. D. Brinker, S. Stuijk, and G. D. Haan. Amplitude-selective filtering for remote-ppg. *Biomedical Optics Express*, 8(3):1965, 2017.

[14] W. Wang, A. C. den Brinker, S. Stuijk, and H. G. De. Robust heart rate from fitness videos. 38(6):1023, 2017.

[15] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan. Full video pulse extraction. *Biomedical Optics Express*, page 1, 2017.

[16] W. Wang, S. Stuijk, and G. D. Haan. A novel algorithm for remote photoplethysmography: Spatial subspace rotation. *IEEE Trans Biomed Eng*, 63(9):1974, 2015.

[17] Y. Wang, X. Ji, Z. Zhou, H. Wang, and Z. Li. Detecting faces using region-based fully convolutional networks. 2017.

[18] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr. Conditional random fields as recurrent neural networks. pages 1529–1537, 2015.

[19] H. Zuo, H. Fan, E. Blasch, and H. Ling. Combining convolutional and recurrent neural networks for human skin detection. *IEEE Signal Processing Letters*, PP(99):1–1, 2017.