Research Article Vol. 13, No. 11 / 1 Nov 2022 / Biomedical Optics Express 6035
Biomedical Optics EXPRESS

Blood pressure estimation by spatial pulse-wave dynamics in a facial video

KAITO IUCHI,[1,4,*] RYOGO MIYAZAKI,[1,4] GEORGE C. CARDOSO,[2] KEIKO OGAWA-OCHIAI,[3] AND NORIMICHI TSUMURA[1]

[1]Graduate School of Science and Engineering, Department of Imaging Science, Chiba University, Japan
[2]FFCLRP, Physics Department, University of São Paulo, Brazil
[3]Kampo Clinical Center, Department of General Medicine, Hiroshima University Hospital, Japan
[4]Equal contribution
[*]lvktcd@gmail.com

Abstract: We propose a remote method to estimate continuous blood pressure (BP) based on spatial information of a pulse-wave as a function of time. By setting regions of interest to cover a face in a mutually exclusive and collectively exhaustive manner, RGB facial video is converted into a spatial pulse-wave signal. The spatial pulse-wave signal is converted into spatial signals of contours of each segmented pulse beat and relationships of each segmented pulse beat. The spatial signal is represented as a time-continuous value based on a representation of a pulse contour in a time axis and a phase axis and an interpolation along with the time axis. A relationship between the spatial signals and BP is modeled by a convolutional neural network. A dataset was built to demonstrate the effectiveness of the proposed method. The dataset consists of continuous BP and facial RGB videos of ten healthy volunteers. The results show an adequate estimation of the performance of the proposed method when compared to the ground truth in mean BP, in both the correlation coefficient (0.85) and mean absolute error (5.4 mmHg). For comparison, the dataset was processed using conventional pulse features, and the estimation error produced by our method was significantly lower. To visualize the root source of the BP signals used by our method, we have visualized spatial-wise and channel-wise contributions to the estimation by the deep learning model. The result suggests the spatial-wise contribution pattern depends on the blood pressure, while the pattern of pulse contour-wise contribution pattern reflects the relationship between percussion wave and dicrotic wave.

1. Introduction

Blood pressure (BP) is an essential health status biomarker. Health risks due to hypertension include heart and kidney failure, and risks due to hypotension include a decline in metabolism and brain function. Moreover, instantaneous BP fluctuations contain critical information. For example, the reserve capacity can be observed by the resilience of BP to a postural change [1]. Thus, it is important to measure BP continuously, and high temporal resolution BP measurements can improve health management. Typically, continuous BP measurement requires the use of protrusive and expensive equipment, except for a few new ideas still under investigation, such as the use of finger oximeters for BP evaluation [2,3]. It is desirable to read BP using simple and non-contact equipment.

Recently, remote methods to estimate BP using RGB cameras have been intensively studied. For example, Jeong et al. [4], Fan et al. [5], and Huang et al. [6] focused on a correlation relationship between pulse transit time (PTT) measured remotely and their determination of BP assumes the Moens-Korteweg equation [7]. However, these methods require for simultaneous capturing of a face and a palm with an RGB camera, and very high temporal resolution of an RGB camera. Sugita et al. [8] and Buxi et al. [9] focused on the relationship between characteristics of pulse-wave contours and BP measured remotely, providing relaxation of the two

#473166 https://doi.org/10.1364/BOE.473166
Journal © 2022 Received 30 Aug 2022; revised 1 Oct 2022; accepted 14 Oct 2022; published 25 Oct 2022

Biomedical Optics EXPRESS

main requirements in PTT-based methods. Thus, these methods require imaging only a single body part, such as the face or the palm of the hand. Further, these methods require a relatively low temporal resolution, compatible with a typical RGB camera. Yet, the following limitations arise. First, the methods are based on a single pulse-wave acquired from the region of interest, which limit the amount of information acquired. Second, the response time to estimate BP is slow compared to the heart rate time frames. This long response time prevents the instantaneous tracking of pulse-waves characteristics and pulse-by-pulse BP measurements [10].

In this study, we propose a remote method that overcomes the two limitations mentioned above. We estimated continuous BP by continually tracking temporal and spatial information of facial pulse-waves. Our hypothesis is that the regions that the spatial and temporal distribution of blood pulsation on the face depend on blood pressure. We modeled a relationship between spatial information of facial pulse-waves and BP, based on a convolutional neural network (CNN). Estimations of continuous BP at a given time are based on a pseudo-continuous time variable built by an interpolation. To understand the physical source of the information involved in the BP estimation, we have conducted spatial-wise and channel-wise visual explanation of the estimation by the deep learning model based on Grad-CAM [11] and Gradient Explanation [12]. The main contribution of our study is the presentation of an algorithm and validation of a robust spatio-temporal imaging method to continually determine BP using RGB cameras. Key advantages of the algorithm over previous methods include improved BP measurement accuracy, and real-time BP estimation.

2. Methods

The proposed method consists of two main steps. The first step is a spatial description of the facial pulse-wave. The second step involves a CNN-based model of the relationship between the spatial information of facial pulse-waves and BP.

2.1. Step 1: face spatial pulse-wave

The goal in this step is to produce and extract pulse-wave features for later use in the CNN-based model. Figure 1 shows the data flow of this step.

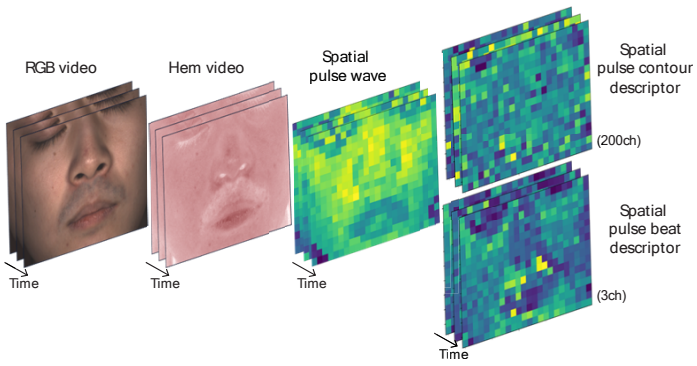

Fig. 1. The data flow for producing and extracting the spatial pulse-wave features from an RGB facial video, which are used in the CNN-based model.

To do so we rely on a video method to extract face pulse-waves [13]. This method separates a facial RGB image into intensity maps: melanin, hemoglobin, and shading (residual information). This method assumes that skin has two dominant types of pigments, melanin and hemoglobin, and the spatial distributions of those pigments in the skin tissue are uncorrelated. Using independent component analysis in logarithmic space of RGB pixel values, we separate the image into its

Biomedical Optics EXPRESS

relative intensities of melanin, hemoglobin, and shading. The hemoglobin component represents the blood volume of capillaries of skin tissue. Extracting the hemoglobin component removes deleterious effects of shading. In methods of non-contact pulse-wave extraction using an RGB camera, it is known that the effect of shading adds noise to the extracted pulse-wave. Specific components of shading that can add temporal-spatial noise due, for example, to changes in the relative position of the light source and the subject, and changes in the intensity of the light source, including flicker. By removing shading, the temporal-spatial data will better represent temporal-spatial hemoglobin pulse-wave.

Next we extract a pulse-wave from the generated hemoglobin map's frames by selecting one or more pixels windows in the facial RGB video (Fig. 2(a)-(c)).

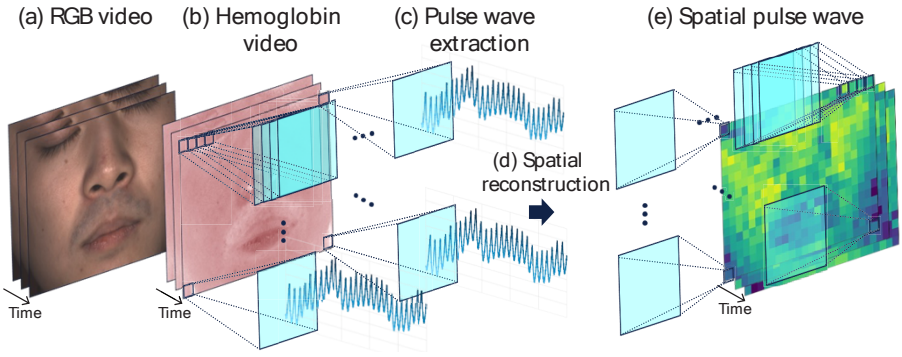

Fig. 2. Conceptual diagram of the construction of a *spatial pulse-wave*. The spatial wave comprises all macropixels, each representing a local pulse-wave.

To read pulse-waves from the whole face, we use multiple pixels windows (40 px × 40 px), that we shall call macropixels. The macropixels are used to improve signal-to-noise levels and cover the face in a mutually exclusive and collectively exhaustive manner. Each macropixel is located at position (x, y) where pulse-wave vectors are extracted (Fig. 2(c)), detrended and bandpass filtered, as in [13]. The distribution of pulse-wave vectors in all macropixels form a vector field we call *spatial pulse-wave* (Fig. 2(e)), which keeps both temporal and spatial information. The *spatial pulse-wave* has lower spatial resolution than the original image (macropixels resolution, instead of pixels, and shown in Fig. 2(e)) and is later used to extract individual pulse features.

To compare contours of different pulses, we construct a *contour descriptor* for the n-th pulse in the pulse-wave of each macropixel. First, pulses are identified by peak detection and extracted (Fig. 3(a)). Next, the start time of each pulse is defined as phase zero (Fig. 3(b)). Finally, we apply to each pulse a 200-point cubic spline interpolation to ensure the contours are indexed by an equal number of phase points, despite pulses being of different duration. Now, for each macropixel, the *contour descriptor* of the n-th pulse consists of its start time t(n) (or pulse start frame number), and a 200-channel phase index i (Fig. 3(b)). Since pulse contours may have various shapes, their amplitudes may be different for the same phase index (Fig. 3(b), cyan plane). The *contour descriptors* of all macropixels form a tensor field $I_{SCD}[(t(x,y,n), p(x, y, n, i)]$ that we call ***S**patial pulse **C**ontour **D**escriptor* (Figs. 1, 4).

Let us now consider a descriptor for pulse beat characteristics. First, we resample the *spatial pulse-waves* to 200 samples per second by a cubic spline interpolation. Then, for each macropixel's pulse-wave, we extract the following *pulse beat descriptors*:

- *Pulse delay* d(n): a vector containing the difference between the n-th pulse peak time and the mean peak time of the n-th pulse of all macropixels.

Biomedical Optics EXPRESS

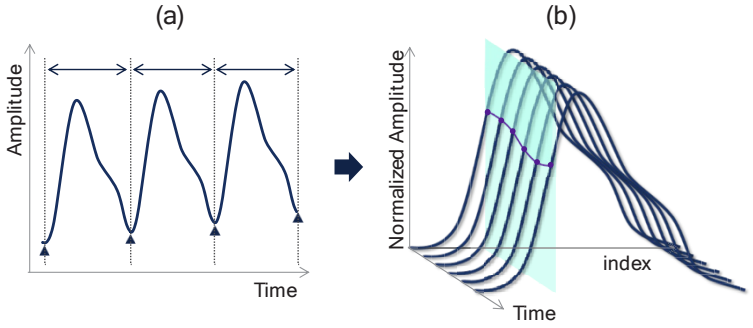

Fig. 3. *Contour descriptor* for a macropixel. (a) Extraction of pulses in the pulse-wave. (b) *Contour descriptor* construction, where pulses are rearranged to compare contours. Pulse amplitudes are normalized to one and pulse durations are normalized to 200 phase points. The cyan plane shows pulse contours varying in time.

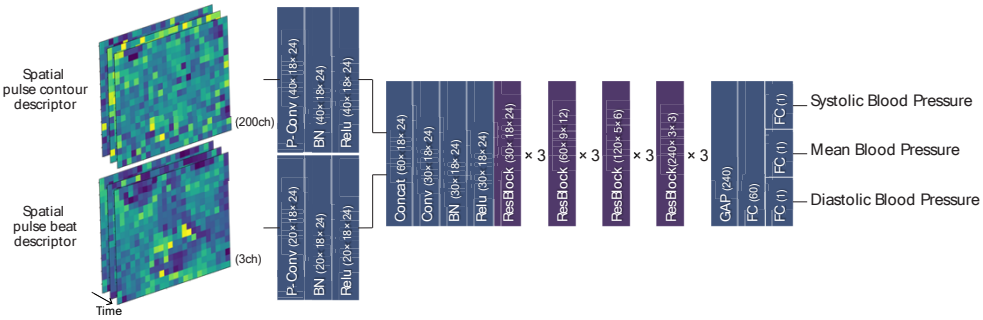

Fig. 4. Deep learning architecture. Abbreviation's guide: Conv (convolutional layer), P-Conv (point-wise convolutional layer), BN (batch normalization), Concat (concatenate), GAP (global average pooling), and FC (fully connected layer). The *spatial pulse contour descriptor* and the *spatial pulse beat descriptor* are independently encoded. Then, they are concatenated and encoded. The *spatial pulse descriptor* contains a *pulse phase*, *a pulse volume*, and a *pulse interval*.

- *Pulse volume* v(n): a vector containing the ratio between the AC and DC components of the pulse-wave. Here, the AC and DC components are calculated from the original pulse-waves, not from processed pulse-waves such as the ones of Fig. 3. The AC component is the difference between the n-th pulse's maximum and the average of its two lowest minima, at the start and at end of the pulse. The DC component is the average of one pulse.
- *Pulse interbeat interval* r(n): a vector containing the time difference between the peaks of the n-th pulse and its consecutive pulse.

Thus, the *pulse beat descriptors* of all macropixels form a tensor field T_{SBD}[d(x,y,n), v(x,y,n), r(x,y,n)] we call ***S**patial pulse **B**eat **D**escriptor* (Figs. 1, 4), where (x,y) is the location of a macropixel, n is the n-th pulse.

2.2. Step 2: CNN for blood pressure training

We use a deep-learning CNN architecture based on ResNet [14] and CBAM [15]. The *spatial pulse contour descriptor* and the *spatial pulse beat descriptor*, previously defined, are the inputs.

Biomedical Optics EXPRESS

Systolic BP (SBP), mean BP (MBP), and diastolic BP (DBP) are the outputs of the CNN (Fig. 4). Each module in the deep learning architecture is shown in Fig. 5.

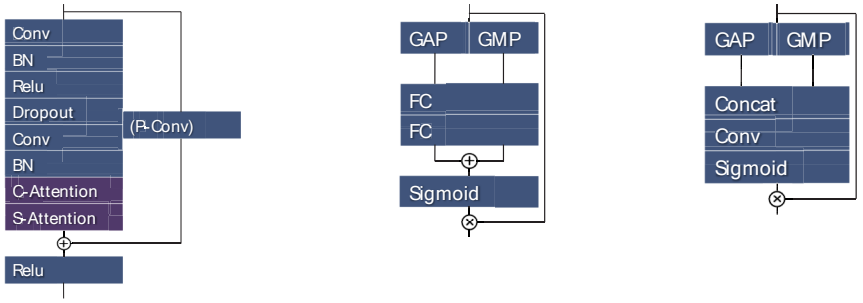

Fig. 5. Architecture of each block in the architecture of deep learning. Abbreviations guide: Conv (convolutional layer), P-Conv (point-wise convolutional layer), BN (batch normalization), Concat (concatenate), GAP (global average pooling), GMP (global max pooling), and FC (fully connected layer). (a) ResBlock (b) Channel attention block (C-Attention) (c) Spatial attention block (S-Attention)

3. Experiment

In what follows, first, we describe an experiment we conducted with volunteers to construct a dataset. Second, we describe a benchmarking design. Third, we describe the training of the CNN. Finally, study the accuracy of our proposed method by comparing predictions with the ground truth.

3.1. Dataset construction

To evaluate the effectiveness of the proposed method, we have recruited 10 healthy volunteers (9 males and 1 female), aged 23.3 ± 1.4 years old. To modify the BP of the volunteers during the data acquisition process, we run cycles consisting of 30 seconds of resting state, up to 60 seconds of breath-holding state, and 60 seconds of resting state. This protocol was performed 3.0 ± 0.5 times for each volunteer, which produced a total of 30 measurements for all volunteers. The variation in the number of times was due to data corruption caused by an equipment malfunction. We acquired BP data with a continuous monitor (Finometer MIDI, Finapres Medical Systems) attached to the left middle finger of the volunteers, and acquired video with an RGB camera (DFK33UX174, The Imaging Source) set at 160 fps and resolution 960 px $\times$ 740 px. The experimental environment is shown in Fig. 6. Boxplots of all BP measurements are shown in Fig. 7.

3.2. Benchmarks design

As a benchmark for BP estimation, we designed a multilayer perceptron to estimate BP from a set of conventional features extracted from facial pulse-waves. The multilayer perceptron is a standard model for a neural network and will not consider spatial information. The multilayer perceptron used has five intermediate layers. The number of channels of the input layer is the number of features used. The number of channels at the end of the intermediate layer is half of the number of channels in the input layer. The number of channels of the intermediate layers linearly decreases from the number of channels in the input layer to the number at the end of the intermediate layer, and so on. The channels of the output layer are systolic, mean, and diastolic

Biomedical Optics EXPRESS

Fig. 6. Experimental environment. Only one camera was used.

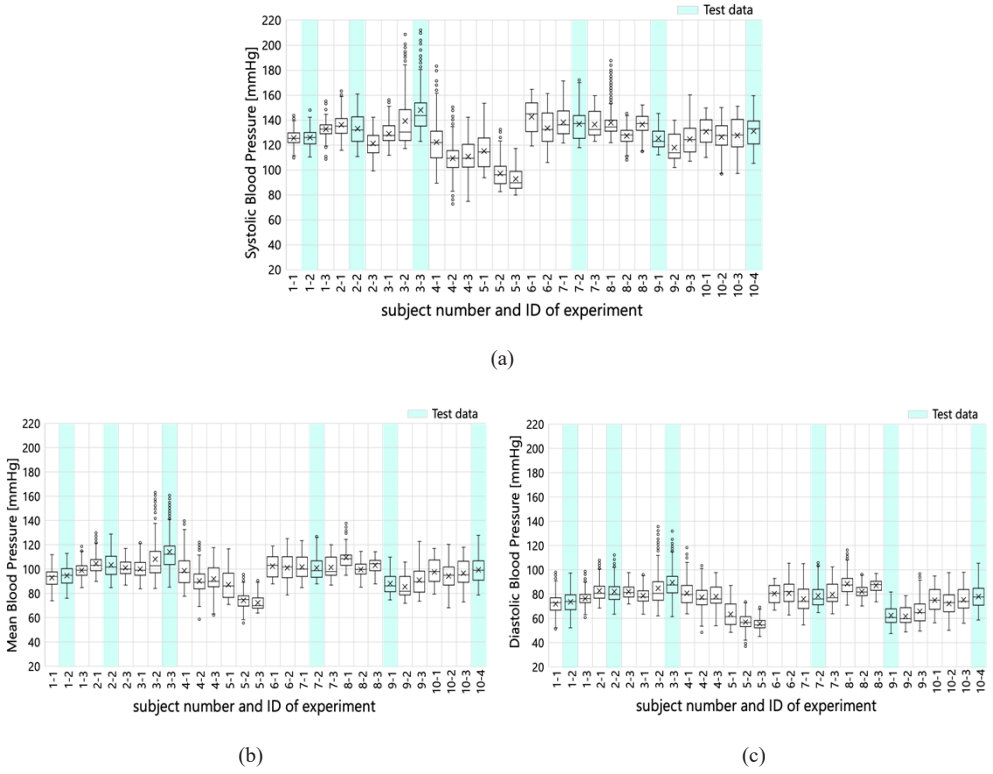

Fig. 7. Boxplots of BP of the constructed dataset. (a) Systolic BP (b) Mean BP (c) Diastolic BP. The test data (green bars) was not used for training.

Biomedical Optics EXPRESS

BP. As input, we chose pulse-wave features known to pertain to BP [3,16,17,18]: pulse-wave contour, its second derivative, and pulse beat, as shown in Table 1.

Table 1. Input features for BP benchmark using a multilayer perceptron.

Concept	Feature
Contour	Overall area of pulse contour
	Systolic area of pulse contour
	Diastolic area of pulse contour
	Phase of peak of pulse contour
	Diastolic area / Systolic area of pulse contour
	Width of pulse contour (10%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 90%)
Derivative	Phase of a-peak of 2nd derivative of pulse contour
	Magnitude of a-peak of 2nd derivative of pulse contour (a)
	Phase of b-peak of 2nd derivative of pulse contour
	Magnitude of b-peak of 2nd derivative of pulse contour (b)
	Phase of c-peak of 2nd derivative of pulse contour
	Magnitude of c-peak of 2nd derivative of pulse contour (c)
	Phase of d-peak of 2nd derivative of pulse contour
	Magnitude of d-peak of 2nd derivative of pulse contour (d)
	Phase of e-peak of 2nd derivative of pulse contour
	Magnitude of e-peak of 2nd derivative of pulse contour (e)
	b / a
	c / a
	d / a
	e / a
	(c - b) / a
	(d - b) / a
	(b – c – d – e) / a
Pulse beat	Interbeat interval of pulse-wave
	AC/DC ratio of pulse-wave

3.3. Training the neural networks

This section describes the training of the deep learning architecture for the evaluation of the proposed method. The total number of data points is 9000, [30 experiments × each experiment's duration (150 s) × sampling rate (2 s^{-1}). This sampling rate is the one from the *spatial pulse contour descriptor* and *spatial pulse beat descriptor*. We use 80% of the dataset (24 complete experiments) for training data and the remaining 20% of the dataset (6 experiments) for testing. In training, the number of epochs is 50. The batch size is 256. The loss function is defined as the sum of the mean squared error (MSE) of systolic, mean, and diastolic BP. Adam optimization is used. The learning rate starts at 0.01 and it is divided by 5 every 10 epochs.

3.4. Results and discussion

We start by comparing the results of the proposed method with benchmarks, in terms of the correlation coefficient and the mean absolute error, and then evaluate the proposed method based on its own merits. Compared to the naïve conventional features of a multilayer perceptron benchmark, the proposed method shows visibly better results (Fig. 8(a) vs Fig. 8(b)), with a MBP

Biomedical Optics EXPRESS

correlation coefficient improvement of 50% (to 0.85 from 0.55), and a MBP mean absolute error decrease of 35% (to 5.4 mmHg from 8.3 mmHg), Table 2 last line. Overall, for our cohort of volunteers, the proposed method is adequate on its own merits with a mean absolute error (MAE) of 5.4 mmHg, which is approximately 5% of the MBP.

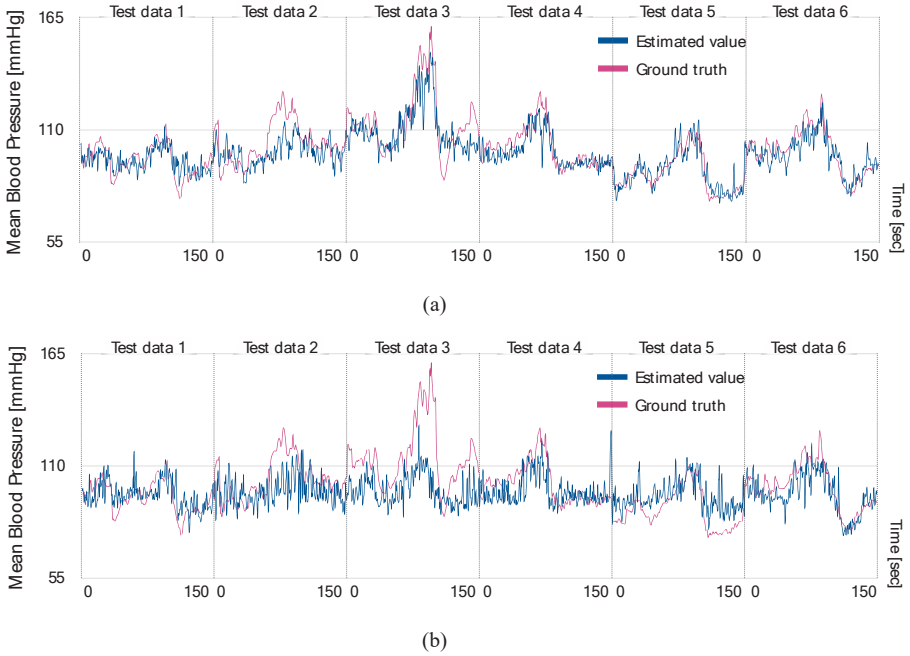

Fig. 8. Accuracy: BP ground truth vs. estimated value. (a) Proposed method (b) Benchmark (using all features).

Table 2. Estimation error between the proposed method and benchmarks

	Correlation coefficient			Mean absolute error [mmHg]		
	SBP	MBP	DBP	SBP	MBP	DBP
Benchmark (Pulse beat)	0.10	0.12	0.09	87	59	33
Benchmark (Derivative)	0.23	0.23	0.25	12	9.9	9.4
Benchmark (Contour)	0.37	0.36	0.38	11	10	9.6
Benchmark (All)	0.52	0.55	0.54	9.8	8.3	8.2
Proposed method	**0.81**	**0.85**	**0.84**	**6.7**	**5.4**	**5.4**

Table 3. Estimation error between the proposed method at 160 fps and 30 fps

	Correlation coefficient			Mean absolute error [mmHg]		
	SBP	MBP	DBP	SBP	MBP	DBP
Proposed method (160fps)	0.81	0.85	0.84	6.7	5.4	5.4
Proposed method (90fps)	0.50	0.56	0.61	9.9	8.1	8.1
Proposed method (60fps)	0.38	0.53	0.58	12	9.5	8.5
Proposed method (30fps)	0.15	0.24	0.31	14	12	11

Biomedical Optics EXPRESS

Next, we examined the impact of the camera frame rate on blood pressure estimation. Besides our original 160 fps capture rate, we have simulated frame rates of 30 fps, 60 fps, and 90 fps. Blood pressure estimation 30 fps or 60 fps, could made the proposed method appliable to consumer cameras. Specifically, we created the datasets have the 30, 60, or 90 fps by resampling the original video to have 180 fps and averaging every 6, 3, or 2 frames of these virtual frames, respectively. The lowered frame-rates videos were used to train and test the proposed deep learning model for blood pressure estimation. The correlation coefficients and MAE of the results show that the higher the fps, the higher the accuracy, as expected (Table 3). The results also indicate poor estimation for 30 fps and 60 fps, as the correlation coefficient is below 0.5. We believe the reason for the difficulty of estimation at lower fps is caused by the loss of the pulse contour features. Even contact-type phethysmographers typically use sampling rates of 100 s^{-1} or higher for pulse contour determination [1].

Here, we discuss the limitations of the experiment. First, the robustness found in this study cannot be generalized to conditions with large individual differences and varying imaging conditions are found. Specifically, we need to demonstrate two perspectives of reliability, which are the influence of individual differences and imaging conditions. The influence of individual differences on the proposed method was not sufficiently demonstrated in this study. Because in this study the training and testing data contained information from the same individuals (see Fig. 7), despite such data being collected under different conditions, and despite no duplicate data in training and testing data having been used. Another aspect is that it is well known that the quality of pulse waves obtained remotely depends on the characteristics of the camera, light source, skin, and movement of the subjects [10,13,19,20,21]. However, our data were acquired under nearly ideal conditions regarding these characteristics. A systematic study to understand the limitations caused by real-life scenarios in the quality of remotely obtained pulse waves requires further studies and is beyond the scope of this proof-of-concept paper. Finer details of the proposed method also deserve optimization, such as the precision and robustness of each part of the proposed process, such as peak detections. Finally, from the physiological point of view, we heuristically hypothesized that information about BP could be obtained from the pulse-wave spatio-temporal distribution on the face. Despite having successfully verified that such a hypothesis is plausible, a physiological explanation about the connection between the pulse-waves distribution and BP is left for further investigation.

4. Explanation of the deep learning model

In what follows, first, we describe a Grad- CAM (Gradient-weighted Class Activation Mapping) [11] visual explanation of the spatial-wise regions the deep learning model consider more relevant for the blood pressure evaluation. Second, we give an analogous visual explanation for the pulse contour regions most relevant for the deep learning model, according to Gradient Explanation [12].

4.1. Spatial-wise explanation

In this section, first we use Grad-CAM [11] to quantify and visualize the spatial distribution of contribution of the inputted face to the estimation by the deep learning model. Then we correlate the Grad-CAM of three different regions of the face with BP values for quantitative interpretation. Such analyses help how the relevant information (facial region, pulse contour, pulse transit time, heart rate, etc.) for efficient BP estimate may depend on BP itself.

For visualization of the spatial contribution, we used Grad-CAM, which calculates the contribution level of each coordinate downsampled from the original image inputted to the CNN and visualizes the calculated contribution level by coloring according to heatmap. Because we applied Grad-CAM onto the all test data, the visualized results were the superposition of 1800 images. The all results were broadly classified by the three pattern (Fig. 9). In the first pattern a

Biomedical Optics EXPRESS

weighted region is located in an upper cheek region in face (Fig. 9(a)), in the second pattern a weighted region is located in an entire cheek region (Fig. 9(b)), and in the third pattern a weighted region is located in an upper and a lower cheek region (Fig. 9(c)). Next, we study how each of these three saliencies is quantitatively relate to BP.

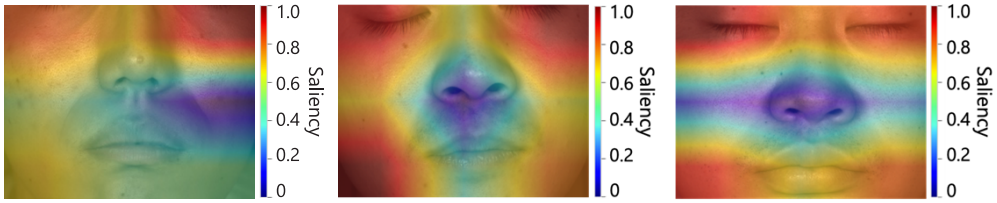

Fig. 9. The results of visualizing the region contributing on the estimation by the deep learning model based on Grad-CAM [11]. Red (blue) represents a stronger (weaker) dependence of the corresponding region for blood pressure information production by the neural network. (a) The region of salient weight is in the upper cheek region of the face. (b) The region of salient weight is over the entire cheek region. (c) The region of salient weight is in the upper and lower cheek regions.

Here, we verify how the spatial pattern of saliency of BP determination depends on BP itself. To simplify the analysis, we considered the sum of all Grad-CAM saliency maps and divided it into three vertical thirds of the face (top, middle, and bottom). We studied the correlation of mean saliency each one of these three thirds of the face with the mean BP. For normalization we took the ratio between Grad-CAM in the regions of interest (top, middle, or bottom) and the Grad-CAM of the entire face, and called this the ratio of Grad-CAM (Fig. 10). The results show that the upper part presents low positive correlation with BP ($r = 0.24$), the middle part presents a moderate positive correlation with BP ($r = 0.50$), and the lower part presents high negative correlation with BP ($r = -0.70$). From an absolute point of view, the upper part always shows a high contribution to the neural network BP estimation, but does not depend much on BP. The middle and bottom parts contribute less to BP, but depend more on BP. Overall, our results indicate that different regions of the face have different functional relationships with BP values, maintaining our driving hypothesis that spatial information improves BP estimation.

4.2. Channel-wise explanation

In this section, we describe the channel-wise explanation of the deep learning model based on Gradient Explanation [12]. Here, we visualize the degree of contribution of the spatial pulse contour descriptor to the estimation by the deep learning model. That is, we seek the relations between pulse contour phase and BP. Figure 11 shows the result of the visualization, which draws all pulse contours of an experimental data with two axes of phase and relative magnitude of the pulse contours by heatmap of degree of contribution. A heatmap shows the spatial-wise averaged spatial pulse contour descriptor degree of contribution to BP, as calculated by Gradient Explanation [12] (Fig. 11). The highest contributions to BP estimation by the deep learning model lie around the percussion and the dicrotic peaks of the plethysmographic pulse. From these finding, we hypothesize that the deep learning model we built uses on the relationship between a percussion wave and a dicrotic wave for estimating BP. A relationship between blood pressure, percussion wave, and dicrotic wave, is actually supported by the literatures [19,20].

Biomedical Optics EXPRESS

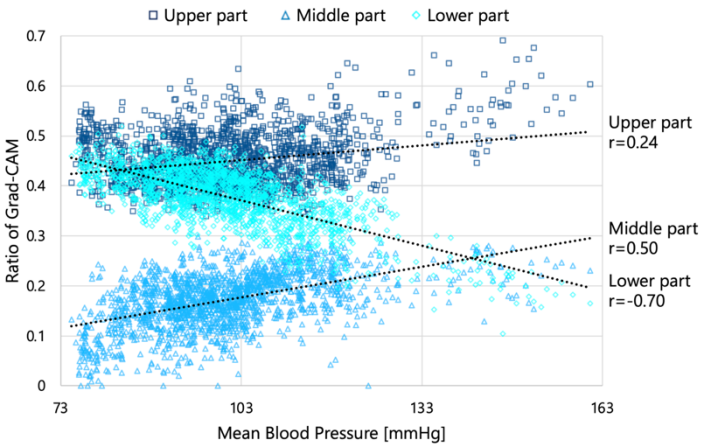

Fig. 10. Mean blood pressure vs. ratio of Grad-CAM to entire region, for in three vertical regions (upper, middle, and lower regions). This result shows the upper part has low correlation (r = 0.24), the middle part has moderate positive correlation (r = 0.50), and the lower part has high negative correlation (r = -0.70). The upper part of the face presents the most BP signal, while the lowest part of the face presents the most information on BP variation.

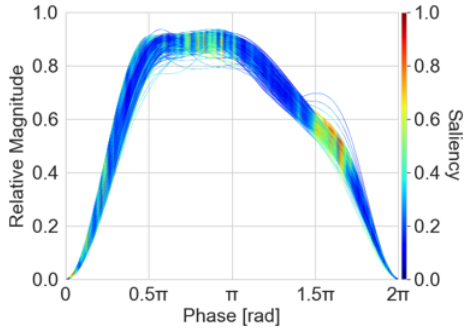

Fig. 11. The result of channel-wise explanation of the estimation by deep learning model, which are spatial-wise averaged *spatial pulse contour descriptor* colored by the degree of contribution calculated by Gradient Explanation according to heatmap.

Biomedical Optics EXPRESS

5. Conclusion and future work

We have proposed a remote method to estimate continuous BP based on time-space information of pulse-waves as observed on the face of a subject by an RGB camera. We modeled a relationship between spatio-temporal facial blood perfusion signals and BP using a convolutional neural network. A dataset constructed to validate the proposed method comprised continuous BP data and facial RGB video of 10 healthy volunteers. The error and the accuracy of the proposed method, compared to conventional methods for benchmarking, demonstrated the superior performance of the proposed method. Further, we determined spatial- and channel-wise weights to the contribution to the BP estimation by our method. The results indicate that the spatial-wise pattern contribution depends on BP, and the plethysmographic phase contribution pattern strongly weights the relationship between percussion and dicrotic waves.

This present study opens three main opportunities future research. Besides the validation of the method in a larger and more diverse cohort, there is a need to evaluate the robustness of the proposed method for noisy, non-restrictive environments. Finally, the method proposed can be further applied to measure other physiological quantities, such as stress and emotion recognition [21,22,23].

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. B. H. Shaw, D. Borrel, K. Sabbaghan, C. Kum, Y. Yang, S. N. Robinovitch, and V. E. Claydon, "Relationships between orthostatic hypotension, frailty, falling and mortality in elderly care home residents," BMC Geriatr. **19**(1), 1–14 (2019).
2. T. Nagasawa, K. Iuchi, R. Takahashi, M. Tsunomura, R. P. Souza, K. Ogawa-Ochiai, N. Tsumura, and G. C. Cardoso, "Blood pressure estimation by photoplethysmogram decomposition into hyperbolic secant waves," Appl. Sci. **12**(4), 1798 (2022).
3. S. G. Khalid, H. Liu, T. Zia, J. Zhang, F. Chen, and D. Zheng, "Cuffless blood pressure estimation using single channel photoplethysmography: a two-step method," IEEE Access **8**, 58146–58154 (2020).
4. I. C. Jeong and J. Finkelstein, "Introducing contactless blood pressure assessment using a high speed camera," J. Med. Syst. **40**(4), 77 (2016).
5. X. Fan, Q. Ye, and S. D. Choudhury, "Robust blood pressure estimation using an RGB camera," J. Ambient. Intell. Human Comput. **11**(11), 4329–4336 (2020).
6. P. W. Huang, C. H. Lin, M. L. Chung, T. M. Lin, and B. F. Wu, "Image based contactless blood pressure assessment using pulse transit time," In: *IEEE International Automatic Control Conference*, 1–6 (2017).
7. D. Buxi, J. M. Redoute, and M. R. Yuce, "A survey on signals and systems in ambulatory blood pressure monitoring using pulse transit time," Physiol. Meas. **36**(3), R1–R26 (2015).
8. N. Sugita, M. Yoshizawa, M. Abe, A. Tanaka, N. Homma, and T. Yambe, "Contactless technique for measuring blood-pressure variability from one region in video phethysmography," J. Med. Biol. Eng. **39**(1), 76–85 (2019).
9. M. Rong and K. Li, "A blood pressure prediction method based on imaging photoplethysmography in combination with machine learning," Biomed. Signal Process Control. **64**, 102328 (2021).
10. R. Takahashi, K. Ogawa-Ochiai, and N. Tsumura, "Non-contact method of blood pressure estimation using only facial video," Artif Life Robot **25**(3), 343–350 (2020).
11. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Bara, "Grad-CAM: visual explanations from deep networks via graident-based localization," *Proceedings of the IEEE International Conference on Computer Vision*, 618–626 (2017).
12. J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim, "Sanity checks for saliency maps," *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 9525–9536 (2018).
13. M. Fukunishi, K. Kurita, S. Yamamoto, and N. Tsumura, "Non-contact video-based estimation of heart rate variability spectrogram from hemoglobin composition," Artif Life Robot **22**(4), 457–463 (2017).
14. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," In: *Proceedings of the IEEE International Conference on Computer Vision*, 770–778 (2016).
15. S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," In: *Proceedings of the European Conference on Computer Vision (ECCV)*, 3–19 (2018).
16. M. Liu, L. M. Po, and H. Fu, "Cuffless blood pressure estimation based on photoplethysmography signal and its second derivative," International Journal of Computer Theory and Engineering **9**(3), 202–206 (2017).

Biomedical Optics EXPRESS

17. S. S. Mousave, M. Firouzmand, M. Charmi, M. Hemmati, M. Moghadam, and Y. Ghorban, "Blood pressure estimation from appropriate and inappropriate PPG signals using a whole based method," Biomed Signal Process Control. **47**, 196–206 (2019).
18. K. Matsumura, P. Rolfe, S. Toda, and T. Tamakoshi, "Cuffless blood pressure estimation using only a smartphone," Sci. Rep. **8**(1), 7298 (2018).
19. M. Huotari, A. Vehkaoja, K. Määttä, and J. Kostamovaara, "Photoplethysmography and its detailed pulse-waveform analysis for arterial stiffness," Journal of Structural Mechanics **44**(4), 345–362 (2011).
20. S. C. Millasseau, J. M. Ritter, K. Takazawa, and P. J. Chowienczyk, "Contour analysis of the photoplethysmograpic pulse measured at the finger," J. Hypertens. **24**(8), 1449–1456 (2006).
21. K. Iuchi, R. Mitsuhashi, T. Goto, A. Matsubara, T. Hirayama, H. Hashizume, and N. Tsumura, "Stress levels estimation from facial video based on non-contact measurement of pulse-wave," Artif Life Robot **25**(3), 335–342 (2020).
22. G. Okada, T. Yonezawa, K. Kurita, and N. Tsumura, "Monitoring emotion by remote measurement of physiological signals using an RGB camera," ITE Trans. Media Technol. Appl **6**(1), 131–137 (2018).
23. K. Masui, G. Okada, and N. Tsumura, "Measurement of advertisement effect based on multimodal emotional responses considering personality," ITE Trans. Media Technol. Appl **8**(1), 49–59 (2020).