InsightNet: non-contact blood pressure measuring network based on face video

Jialiang Zhuang\textsuperscript{a}, Bin Li\textsuperscript{b}, Yun Zhang\textsuperscript{c}, Xiujuan Zheng\textsuperscript{a,\ast}

\textsuperscript{a}College of Electrical Engineering, Sichuan University, Chengdu, Sichuan, China
\textsuperscript{b}School of Computer Science, Northwest University, Xi’an, China
\textsuperscript{c}School of Information Science and Technology, Xi’an Jiaotong University, Xi’an, China

Abstract

Blood pressure indicates cardiac function and peripheral vascular resistance and is critical for disease diagnosis. Traditionally, blood pressure data are mainly acquired through contact sensors, which require high maintenance and may be inconvenient and unfriendly to some people (e.g., burn patients). In this paper, an efficient non-contact blood pressure measurement network based on face videos is proposed for the first time. An innovative oversampling training strategy is proposed to handle the unbalanced data distribution. The input video sequences are first normalized and converted to our proposed YUVT color space. Then, the Spatio-temporal slicer encodes it into a multi-domain Spatio-temporal mapping. Finally, the neural network computation module, used for high-dimensional feature extraction of the multi-domain spatial feature mapping, after which the extracted high-dimensional features are used to enhance the time-domain feature association using LSTM, is computed by the blood pressure classifier to obtain the blood pressure measurement intervals. Combining the output of feature extraction and the result after classification, the blood pressure calculator, calculates the blood pressure measurement values. The solution uses a blood pressure classifier to calculate blood pressure intervals, which can help the neural network distinguish between the high-dimensional features of different blood pressure intervals and alleviate the overfitting phenomenon.

\textsuperscript{\ast}Corresponding author xiujuanzheng@scu.edu.cn
It can also locate the blood pressure intervals, correct the final blood pressure values and improve the network performance. Experimental results on two datasets show that the network outperforms existing state-of-the-art methods. 

**Keywords:** non-contact, blood pressure measurement, oversampling training strategy, face video

### 1. Introduction

Blood pressure is a primary physiological parameter of the human body and an essential basis for disease diagnosis. Cardiovascular diseases are the leading cause of death globally. Most of them are highly related with hypertension. In this context, convenient ordinary monitoring of blood pressure in daily life becomes critical.

Many outstanding scientists [1, 2, 3, 4] have devoted themselves to the study of calculating blood pressure by extracting the waveform shape characteristics of the single point pulse wave and obtained good results, which means that single point pulse wave can effectively reflect the change characteristics of blood flow in the human cardiovascular system in different heart cycles. However, they all need the blood volume pulse wave obtained by special sensors, which greatly limits the application range of blood pressure calculation. Now there are many excellent works that can accurately simulate pulse waves based on face video and accurately calculate heart rate and heart rate variability. RthymNet [5] can calculate accurate average heart rate by feature extraction of Spatio-temporal map, which shows that Spatio-temporal map of current face review can perfectly simulate blood volume pulse wave and blood flow state in blood vessels. Cheng [6] also proposed a method to convert blood volume pulse wave into pressure pulse wave, which proves that there is an essential relationship between single-point blood volume pulse wave and pressure pulse wave, enabling us to believe that the Spatio-temporal map of single point face video can calculate accurate blood pressure value.

Still, there are many challenges to indirect blood pressure measurements; the
accuracy of blood pressure measurement is easily affected by many factors [7], including the professional abilities of the conductors, the height, weight, and age of the subjects, and the seasons and environments. For example, head motion creates difficulties for the region cutting algorithm, causes residual motion image, and degrades the image quality. In addition, the weak remote BVP signal reflecting changes in hemoglobin concentration in the circulatory system can easily be contaminated and submerged by changing ambient light. Moreover, because of difficulties of blood pressure data collection and the privacy issues of face video, there is very few large-scale blood pressure database with face videos. The lack of data greatly hinders the rapid development in the field of non-contact measurement of blood pressure.

Therefore, this study is proposing the first end-to-end non-contact blood pressure measurement network based on face video in this study. To reduce the interference caused by motion, multiple effective regions of the face are recombined into a new feature sequence. Then a spatial-temporal slicer to adopted to crate a multi-time domain spatial mapping to help the blood pressure estimator generate BVP signals effectively. Finally, a blood pressure estimator is used to extract the high-level spatial-temporal features and calculate the systolic and diastolic blood pressure, respectively. The main contributions are summarized as follows:

1. Based on the face video, this paper proposes for the first time a non-contact end-to-end blood pressure measurement network named Insight-net.

2. Only a 15-second video sequence is needed to measure diastolic and systolic blood pressure, which has higher detection efficiency than traditional blood pressure measurement methods.

3. An oversampling training scheme for blood pressure monitoring task is proposed for the first time, which effectively addresses the uneven distribution of blood pressure data set in the training process.

4. A private database for non-contact blood pressure estimation was estab-
lished, containing 124 people’s physiological data and corresponding face videos.

2. Related work

2.1. Blood pressure measurement method

Until now, two categories of methods have been used to measure blood pressure: direct and indirect measurements. Direct blood pressure measurements are made with medical equipment to directly monitor human biological signals [8, 9, 10, 11], including two main technical approaches. One route is to measure the pressure change directly with a pressure transducer, for example, arterial manometry [12]. A mathematical model of the transducer artery system is proposed to design a new sensor and replace arterial manometry for long-term blood pressure monitoring without calibration [13]. For patients with white coat hypertension, ambulatory blood pressure monitoring using this technology route is particularly important. Nevertheless, this sensor is still an invasive instrument and susceptible to displacement and pressure [14], making it challenging to apply on a large scale.

The other technical approach is to control the pressure applied to the measured area by an external device and then use the information related to the Coriolis sound for blood pressure measurement. The most classic apparatus is the mercury sphygmomanometer, inexpensive and with high accuracy of measurement. However, due to safety and environment concerns, it has been gradually replaced by liquid-free sphygmomanometers [11]. Nevertheless, the defects of manometers, cuffs, and piping systems make it challenging to maintain them regularly, and it is difficult to ensure a high level of professionalism among the users, both of which may lead to inaccuracy in practice [15, 7, 16]. In recent years, oscillometric monitors have also used in blood pressure measurements, which use the cuff to record the pressure pulse caused by the arterial volume pulse and then calculates the mean arterial pressure by calculating the pulse wave amplitude, area, and other parameters [8, 10, 9]. However, the measure-
ment accuracy of the oscillographic devices is easily affected by calibration. Influenced by the perimeter of the measured area as well as other parameters, the wide popularization of calibration in practical application is limited, which calls for more straightforward and more effective blood pressure measurement solutions. The blood pressure measurement approaches mentioned above, relying on invasive or contact instruments are uncomfortable for continuous measurements in daily life. As a result, indirect measurement methods have been gaining popularity in recent years.

Indirect blood pressure measurements usually calculate the systolic and diastolic blood pressure based on the features of measured blood volume pulse (BVP) signals using Photoplethysmography (PPG) technology\[16\]. Pulse wave velocity (PWV) refers to the pressure wave velocity that propagates along the wall of the great artery during each cardiac ejection. It is a simple, effective, and economic index for the noninvasive evaluation of arterial stiffness\[17, 18\]. Moreover, it is an independent predictor of cardiovascular events closely related to blood pressure\[19, 20, 21, 22\]. A series of simulation experiments on humans and animals confirmed that pulse wave transit time (PWT) is closely related to diastolic and systolic blood pressures. In this context, a novel indirect method, which differed from the traditional direct methods using cuff or oscilloscope devices, was proposed to measure blood pressure. First, the noninvasive sensors were placed on the skin to collect the blood volume pulse (BVP), and then the arterial blood pressure value was calculated based on the extracted features of BVP signals\[1, 2\]. A continuous blood pressure measurement method was proposed based on optical plethysmography (PPG) technology\[2\]. After detecting the blood volume changes of the cardiac cycle with infrared light, the time delay of systolic and diastolic peak values was calculated to estimate the blood pressures. Another study added 1/2 pulse width and 2/3 pulse width as the characteristics of BVP signal together with systolic upstroke time and diastolic time\[1\]. However, in some previous studies, the relationship between blood pressure and BVP signals was assumed as nonlinear. Then, artificial neural networks were adopted to get more accurate models between the blood pressure
and BVP signals [3, 4].

In order to obtain more characteristics related to blood pressure, some researchers [23] proposed that electrodes and instruments could collect ECG signals while PPG signals were obtained. At the same time, others designed a new sensor integrating PPG signal acquisition and ECG signal acquisition [24] and then eliminated baseline drift noise by wavelet change, which significantly improved the measured results. Similar to the development route of blood pressure detection method with only PPG, some papers [25] also used artificial neural network ANN to train the network model to predict blood pressure according to the sample experiment [25, 26]. Based on previous work, demographic information (including weight, height, and other factors) was added to the network for training. The results showed that the performance of the model obtained generated this way was much better than that of the multiple regression method. In addition to the use of artificial neural networks and other network models, traditional machine learning method [18, 27, 28] was also used in many studies to optimize the feature extraction process.

2.2. **Deep convolution neural network**

With the rapid development of deep learning, many network structures with strong feature extraction abilities have emerged and evolved for diversified applications. The most representative ResNet [29], which first solved the degradation problem of a deep network via introducing a residual structure. It dramatically increases the number of convolution layers and consequently the fitting ability of neural networks took a huge leap forward. Then, Huang et al. proposed forward DenseNet, which comprehensively promoted the idea of residuals and constructed an extensive range of residuals structure [30]. In this case, the input and final output were able to directly realize gradient propagation, which could reduce the gradient disappearance, strengthens the transmission of features and the reuse of features, and reduces the parameter sequence to a certain extent.

Furthermore, on the premise of ensuring the quality of the model, the network could be widened, the number of parameters reduced, and high-dimensional
features extracted \cite{12}. As a result, ResNext is popular with a wide range of developers, as they do not need to design complex details of the induction structure manually but use the same topology for each branch. Its essence is to control the number of groups through cardinality, which further improves the accuracy in some specific tasks.

In the field of physiological measurement from face video, Rhythmnet \cite{5} uses 2D convolutional neural networks and LSTM with a robust feature extraction ability to extract high-dimensional features from feature maps and fit heart rate values. PhysNet \cite{31} uses 2D residual network to encode each video frame and combines with LSTM to extract the BVP signals.

3. Methodology

This study proposes a non-contact blood pressure measurement network, which aims to estimate the blood pressure from the input face videos by a designed robust neural network. There are five main steps in the proposed method, and the overall workflow is shown in Figure 1. Firstly, the ROI, which contains the blood pressure signal from the input facial video, needs to be defined. Secondly, the data augmentation (DA) module is proposed to cope with position drift and face motion artifacts in adjacent frames. Thirdly, in order to reduce interferences caused by illumination variations, the Color Space Conversion (CSC) module is proposed to transform RGB color space into YUV color space. Then, to reduce interferences caused by sudden non-rigid motions, the ROI sequences are sliced to generate the multi-domain spatial-temporal mapping. Finally, the Blood Pressure Estimator module estimates systolic and diastolic blood pressure. Details of each module are explained in the following subsections.

3.1. ROI Definition

In order to reduce the interference of background information and invalid face regions (such as eyes, nose, hair, and mouth), four regions of interest (ROIs)
Figure 1: Overview of the InsightNet. First, Seetaface6 is used to detect human face and localize 68 landmarks. They are then fed into the Data augmentation (DA) module and the color space conversion (CSC) module in turn, from which spatial-temporal feature mapping is generated by the spatial-temporal slicer (STS) module. Finally, a blood pressure estimator is constructed to calculate the blood pressure value.

are defined by selecting the forehead and cheek with rich vascular information. First, the authors use SeetaFace6 to detect face and localize 68 landmarks and select 7 points to delineate the ROIs in each frame, as shown in Figure 1, total of \( N \) ROIs are selected. Next, we get the average pixel values of each ROI to represent Physiological information. noted as:

\[
f_{roi}(t, n) = \bigcup_{n=1}^{N}(R_n) \tag{1}
\]

where \( R_n \) represents four ROI regions at frame \( t \) as shown in Figure 2.

3.2. Data Augmentation

The change of external illumination easily interferes with the color information of the face video, which will affect the accuracy of non-contact blood pressure detection. Therefore, a data augmentation strategy is proposed to simulate the problem of signal contamination caused by external illumination changes thus improving the algorithm’s robustness. First, the authors randomly select the segment with the length of \( V \) in the \( T \) frame sequence \( f_{roi}(v, n), v \in V \), as shown in Figure 1. For each frame, we randomly select a region with the size of \( p \times q \) \( f_{roi}^{p \times q}(v, n), v \in V \), which is then masked and generate a new feature sequence \( f'_{roi}(v, n), v \in V: f'_{roi} (v, n) = f_{roi}^{p \times q} (v, n) \times 0 + 1 \).
The authors splice the frame sequences, $f'_{\text{roi}}(v, n) \in V$ and $f_{\text{roi}}(t, n), t \notin V$ in the time dimension and form a new feature sequence recoded as $F'_{\text{roi}}(t, n), t \in T$. Finally, YUV color space is proposed here to extract information related to blood volume pulse wave, which gives more attention to brightness dimension features in color space. The new color space transformation can be formulated as:

$$
\begin{bmatrix}
Y_t(x, y) \\
U_t(x, y) \\
V_t(x, y)
\end{bmatrix}
= \begin{bmatrix}
0.299 & 0.587 & 0.114 \\
-0.169 & -0.331 & 0.5 \\
0.5 & -0.419 & -0.081
\end{bmatrix}
\begin{bmatrix}
R_t(x, y) \\
G_t(x, y) \\
B_t(x, y)
\end{bmatrix}
$$

(2)

where $t$ indicates video frame index, $R_t(x, y), G_t(x, y), B_t(x, y) \in F'_{\text{roi}}(t, r_i)$, $R_t(x, y), G_t(x, y), B_t(x, y)$ represents the pixel value in RGB color space, while $Y_t(x, y), U_t(x, y), V_t(x, y)$ are the pixel value of the YUVT color channel. All of them constitute the video sequence after data enhancement together. The enhanced video sequence is noted as $F_{\text{roi}}(t, n)$.

3.3. Spatial-Temporal Slicer

Fig. 2: Architecture of the spatial-temporal slicer. Firstly, four ROIs are obtained by using detection points, and then each region is divided into $N$ sub-ROIs. After all the sub-ROIs are expanded and sliced in time domain, $M$ spatial-temporal feature maps $\{TS_1, TS_2, \ldots, TS_M\}$ are generated.

For $F_{\text{roi}}(t, n)$, each subsets of ROI is pooled to get the average of each ROI pixel $AP(n, t)$, which can be stated as:

$$
AP(n, t) = \frac{\sum_{x,y \in \text{ROI}, n, t} V(x, y, t)}{\text{NUM}}
$$

(3)
Where \( V(x,y,t) \) represents the pixel value of the \( n \)th subsets position \((x, y)\) of frame \( t \) of the face video sequence, \( \text{NUM} \) represents the number of pixels in the selected ROI region. The average value of each subsets pixel is flattened and spliced to form the feature vector corresponding to the \( t \) frame, which can be written as:

\[
\text{SS}(t) = \text{CAT}\left(\emptyset \left( \bigcup_{n=1}^{N} \text{AP}(n, t) \right) \right)
\]

where \( \emptyset \) represents the extracting non empty subsets operation, and \( \text{CAT}(\cdot) \) represents the feature vector splicing operation. We combine the eigenvectors of all frames into a spatial-temporal feature map, and divide them into two parts by sliding windows with step length \( C_1 \). \( M \) separate time-space fragments \( \{\text{TS}_1, \text{TS}_2, \ldots, \text{TS}_M\} \) are constructed as multi-domain spatial-temporal feature mapping \( \text{STS}(m) \), which can be defined as:

\[
\text{STS}(m) = \bigcup_{t=N \times cl}^{(N+1) \times cl} \text{SS}(t)
\]

where \( m \) is the index of spatial-temporal slicer. In order to use video information effectively and help network learning, I normalized the pixel values of all video frames to the range of \([0,1]\).

### 3.4. Blood Pressure Estimator

In this experiment, the residual convolution neural network [31] is used as the backbone network, and the size of all convolution kernels is \( 3 \times 3 \). The step is 1. Firstly, the backbone network is used to extract the depth features of spatiotemporal slices, and then the cascaded LSTM enhances the temporal correlation again. Next the classifier is used to classify the blood pressure interval. This operation can establish a benchmark for the features and effectively reduce the over fitting phenomenon of the network. Finally, integrating the classifier results with the output of the feature extractor can effectively help the classifier training on the one hand, and calculate the specific value of blood pressure on the other hand.
We express the whole BPE process as the following formula:

\[ F = \text{CAT} \left( \bigcup \text{FE}(\text{STS}(m)) \right) \]  

(6)

\[ F_{\text{cla}} = \text{CLA}(F) \]  

(7)

\[ R_{\text{reg}} = \text{REG} \left( \text{CAT}(F, F_{\text{cla}}) \right) \]  

(8)

\[ R_{\text{cla}} = \text{SOFTMAX}(F_{\text{cla}}) \]  

(9)

where FE represents the feature extraction network composed of depth residual network and LSTM, CAT(·) represents dimension splicing operation, CLA represents blood pressure classifier, REG represents blood pressure calculator, R_{cla} represents the output features of classifier, R_{reg} represents the output results of classifier and calculator respectively, and SOFTMAX(·) represents softmax operation. We use the cross entropy loss function to train the classifier and MAE loss function to train the calculator. The two loss functions are combined to train the proposed model, which are formulated as follows:

\[ \text{Loss}_{\text{cla}} = \frac{1}{N} \sum_i \left[ y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i) \right] \]  

(10)

\[ \text{Loss}_{\text{reg}} = \frac{\sum_{i=1}^{N} |f_i - yf_i|}{N} \]  

(11)

\[ \text{Loss} = \text{Loss}_{\text{cla}} + \text{Loss}_{\text{reg}} \]  

(12)

where \( p_i \) represents the result of classifier, \( f_i \) represents the result of calculator, \( y_i \) represents the category label of blood pressure truth value, \( yf_i \) represents the truth value of blood pressure. The final output result needs to be calculated with the classifier result as the reference value and the deviation calibration of the calculator

\[ R = \alpha \cdot \text{STA}(R_{\text{cla}}) + \beta \cdot R_{\text{reg}} \]  

(13)

where STA represents the dynamic reference value of each blood pressure interval and the deviation weight, and \( \alpha \) represents the weight of the calculator results, \( \beta \) represents the weight of the classifier results.
3.5. **Oversampling training strategy**

In order to make effective use of limited training data more effectively and solve the problem of uneven distribution of data in different blood pressure intervals in datasets, an oversampling training strategy is proposed for the first time. First, according to the distribution and level of blood pressure, different grouping strategies are designed for systolic and diastolic pressure data, which is then divided into four big groups based on the value of blood pressure \{G_1, G_2, G_3, G_4\}. In the five-fold cross validation experiment, each big group was divided into five small groups equally again, four of them are as training set and the other one is as the validation set. For each batch, the ratio of samples from each large group is 1:1:1:1. Data is extracted from each group in order. If the quantity is insufficient, the sample will be selected from the beginning. The sampling strategy to construct training dataset group ST(d,c) and validation dataset group SV(d,c) can be respectively computed as:

\[
\text{ST}(d, c) = \bigcup_{d=1}^{4} (G_{dc})
\]

\[
\text{SV}(d, c) = \bigcup_{d=1}^{4} \left( \bigcup_{d=1}^{c} G_{dc} + \bigcup_{d=c+1}^{5} G_{dc} \right)
\]

where \(G_{dc}\) represents a small dataset group, and \(d\) and \(c\) indicates the times of cross-validation and the index of small group in each big group.

4. **Experiments**

4.1. **Dataset**

The performance of the proposed network was evaluated on a public dataset and a private dataset. MMSE-HR [32] is a public non-contact heart rate and blood pressure estimation database composed of 102 face videos from 40 subjects and recorded at 25 frames per second (fps). A biological data acquisition system was used to collect the average HR and BP values.
A private database for non-contact blood pressure estimation was created, named as Advanced Sensor Physiological Dataset (ASPD). In this dataset, the physiological data and corresponding face videos of 124 people were collected by using a multi-channel physiological signal acquisition system (Biopac M160), a blood pressure monitor (OMRON HEM-1020), and a mobile phone camera (Huawei Mate30), respectively. The face video for each subject was collected for about 1 minute. The frame rate of the video is 30fps, and the resolution is $1920 \times 1080$. The distributions of systolic and diastolic blood pressures of ASPD dataset are shown in Figure 3 and Figure 4. Moreover, the devices and setup is illuminated in Figure 5.

Three metrics were used to evaluate the performance of the network including the standard deviation (SD), the root mean square error (RMSE), and the mean absolute error (MAE).

In ASPD cross validation, all face videos were sampled to 30fps. Training was done with Nvidia-v100 and Pytorch. The authors designed independent oversampling learning strategy for systolic and diastolic blood pressure, and conducted independent cross validation and cross dataset testing. Firstly, a five-fold cross validation was performed on ASPD dataset, and comparative experiments were conducted to verify various key module parameters of the network and the optimal combination scheme. These parameters included the
number of intercepted small ROIs, the type of loss function, the type and depth of backbone network, the color space, and the frame size. After that, ASPD was used to train and test cross datasets on MMSE-HR. Finally, a five-fold cross-validation was performed on ASPD, and ablation experiments were done on the oversampling training scheme to verify its effectiveness.

4.2. Cross-dataset testing

A recent study provide a solution for blood pressure estimation based on rPPG. It used thousands of data for training, and 50 samples for test [33]. For
Table 1: Results of BP measurements

|       | Method | SD (mmHg) | RMSE (mmHg) | MAE (mmHg) |
|-------|--------|-----------|-------------|------------|
| SBP   | NIBPP  | –         | –           | 13.6       |
|       | ours   | 16.02     | 16.55       | 12.35      |
| DBP   | NIBPP  | –         | –           | 10.3       |
|       | ours   | 11.98     | 12.22       | 9.54       |

Our InsightNet, 125 data samples are used for training and MMSE-HR data are used for testing. The sample ratio tends to be close to 1:1. After three fine-tuning of the model, our InsightNet obtained the best performance for the test is that the systolic blood pressure MAE is 13.6mmHg and the diastolic blood pressure MAE is 10.3mmHg. The comparative results are shown in Table 1. From the results, it can be seen that, our method get better performance than the recent study [33]. In order to verify the effectiveness of InsightNet, the Bland-Altman plots as showed in Figure 6 show a good consistency between the estimated blood pressure and the ground truth.

Figure 6: The Bland-Altman plots demonstrating the agreement of the estimated SBP and the ground truth for InsightNet.
4.3. Ablation study

4.3.1. The effectiveness of spatial-temporal slicer

In order to verify the effectiveness of the combination of the spatiotemporal slicer and LSTM, we use the oversampling training scheme to train systolic and diastolic blood pressure, respectively. The number of sub-ROIs in each ROI and the loss function used in training of SBP and DBP were set to 6, L1 and 4, L2 respectively. At the same time, they all used ResNet 50-512 as the backbone network and YUVT as the color space. It was predicted that the specific length of the spatiotemporal slicer would greatly affect the final training results, and the optimal scheme of spatiotemporal slicing could be found. Clip length was tested for 90, 150, 225 and 450.

From Table 2, spatiotemporal slicer with LSTM can help the network obtain the detailed characteristics of physiological information more effectively, and the slice length of 150 frames can maximize the effect of this network combination. For example, as shown in Table 2, MAE of systolic blood pressure and diastolic blood pressure reached 8.07mmHg and 6.78mmHg, respectively. Compared with other slice length schemes, MAE of systolic blood pressure and diastolic blood pressure increased by 0.37mmHg and 0.55mmHg, respectively, which was much better than the network scheme without temporal and spa-
Table 2: Performance comparison of different network.

|       | Method    | SD(mmHg) | RMSE (mmHg) | MAE(mmHg) |
|-------|-----------|----------|-------------|-----------|
| **SBP** | LSTM-90   | 10.06    | 10.33       | 8.44      |
|       | LSTM-150  | 9.81     | 9.94        | 8.07      |
|       | LSTM-225  | 10.09    | 10.16       | 8.32      |
|       | Without LSTM | 10.49 | 17.85       | 15.78     |
| **DBP** | LSTM-90   | 8.35     | 9.12        | 7.31      |
|       | LSTM-150  | 8.28     | 8.45        | 6.78      |
|       | LSTM-225  | 8.33     | 9.05        | 7.33      |
|       | Without LSTM | 9.28 | 17.06       | 15.04     |

Notes: Red indicates the best performance of different network.

4.3.2. The effectiveness of oversampling strategy (OSS)

Table 3: Performance comparison of different networks.

|       | Method   | SD(mmHg) | RMSE (mmHg) | MAE(mmHg) |
|-------|----------|----------|-------------|-----------|
| **SBP** | With OSS | 9.81     | 9.94        | 8.07      |
|       | Without OSS | 10.12 | 10.14       | 8.27      |
| **DBP** | With OSS | 8.28     | 8.45        | 6.78      |
|       | Without OSS | 8.35 | 8.6         | 6.92      |

Notes: Red indicates the best performance of different training strategy.
Then the effectiveness of the oversampling training scheme was tested, with the standard data sampling training scheme as a comparison (in each epoch, training samples were imported into the network for training). The above results show that the combination of spatiotemporal slice and LSTM is the best when clip length is 150. Therefore, in this test, LSTM-150 was used as the primary network structure. Table 3 shows that the training scheme enabled the same network model with more vital fitting ability and robustness. MAE of systolic blood pressure and diastolic blood pressure decreased from 8.27mmHg and 6.92mmHg to 8.07mmHg and 6.78mmHg. RMSE also decreased by 0.2mmHg and 0.15mmHg, respectively. This shows that for the blood pressure task with few training samples, appropriate oversampling training scheme is helpful to further enhance the network potential.

5. Conclusion

This paper presents the first non-contact end-to-end blood pressure measurement network, which only uses face video information to quickly calculate diastolic and systolic blood pressure in 15 seconds. Tests have been run on the public dataset MMSE-HR and the self-collected data set ASPD, and the results show that the proposed InsightNet can effectively and efficiently measure blood pressure.

Acknowledgments

This work is supported by the Sichuan Science and Technology Program under Grant 22ZDYF0946 to Xiujuan Zheng, the Shaanxi Provincial Natural Science Basic Research Program under Grant 2021JQ-455 to Bin Li. The authors would like to thank the engineers of Xi’an Singularity Fusion Information Technology Co. Ltd for their supports of data collection and experimental procedures.
Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] X. Teng, Y. Zhang, Continuous and noninvasive estimation of arterial blood pressure using a photoplethysmographic approach, in: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439), Vol. 4, 2003, pp. 3153–3156 Vol.4. doi:10.1109/IEMBS.2003.1280811

[2] R. Samria, R. Jain, A. Jha, S. Saini, S. R. Chowdhury, Noninvasive cuffless estimation of blood pressure using photoplethysmography without electrocardiograph measurement, in: 2014 IEEE REGION 10 SYMPOSIUM, 2014, pp. 254–257. doi:10.1109/TENCONSpring.2014.6863037

[3] Y. Kurylyak, F. Lamonaca, D. Grimaldi, A neural network-based method for continuous blood pressure estimation from a ppg signal, in: 2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2013, pp. 280–283. doi:10.1109/I2MTC.2013.6555424

[4] L. Wang, W. Zhou, Y. Xing, X. Zhou, A novel neural network model for blood pressure estimation using photoplethysmography without electrocardiogram (2018) 1–9doi:10.1155/2018/7804243

[5] X. Niu, S. Shan, H. Han, X. Chen, Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation, IEEE Transactions on Image Processing 29 (2020) 2409–2423. doi:10.1109/TIP.2019.2947204
[6] J. Cheng, Y. Xu, R. Song, Y. Liu, C. Li, X. Chen, Prediction of arterial blood pressure waveforms from photoplethysmogram signals via fully convolutional neural networks, Computers in Biology and Medicine 138 (2021) 104877. doi:10.1016/j.compbiomed.2021.104877

[7] N. Markandu, F. D. Whitcher, A. Arnold, C. Carney, The mercury sphygmomanometer should be abandoned before it is proscribed, Journal of Human Hypertension 14 (2000) 31–36.

[8] G. L. Young P, A surrogate arm for evaluating the accuracy of instruments for indirect measurement of blood pressure, Biomed Instrum Technol 24 (1990) 130–135.

[9] K. Kario, Home blood pressure monitoring: Current status and new developments, Am J Hypertens 34 (2021) 783–794.

[10] M. A. Ulrich Tholl, Klaus Forstner, Measuring blood pressure: pitfalls and recommendations, Nephrol Dial Transplant 19 (2004) 766–70.

[11] G. Beevers, G. Lip, E. O’Brien, Abc of hypertension: Blood pressure measurement. part ii-conventional sphygmomanometry: technique of auscultatory blood pressure measurement, BMJ (Clinical research ed.) 322 (2001) 1043–7.

[12] G. L. Pressman, P. M. Newgard, A transducer for the continuous external measurement of arterial blood pressure, IEEE Transactions on Bio-medical Electronics 10 (2) (1963) 73–81. doi:10.1109/TBME.1963.4322794

[13] M. Gorostidi, E. Vinyoles, J. R. Banegas, A. de la Sierra, Prevalence of white-coat and masked hypertension in national and international registries, Hypertension Research 38 (2015) 1–7.

[14] G. M. Drzewiecki, J. Melbin, A. Noordergraaf, Arterial tonometry: Review and analysis, Journal of Biomechanics 16 (2) (1983) 141–152.
[15] T. M. Krass ME, La Barber LP, Blood pressure measurement in family physician offices, Family Med 20 (1988) 25–27.

[16] T. Knight, F. Leech, A. Jones, L. Walker, R. Wickramasinghe, S. Angris, P. Rolfe, Sphygmomanometers in use in general practice: an overlooked aspect of quality in patient care, Journal of Human Hypertension 15 (2001) 681–684.

[17] B. Gribbin, A. Steptoe, P. Sleight, Pulse wave velocity as a measure of blood pressure change, Psychophysiology 13 (1976) 86–90.

[18] S. S. Mousavi, M. Firouzmand, M. Charni, M. Hemmati, M. Moghadam, Y. Ghorbani, Blood pressure estimation from appropriate and inappropriate ppg signals using a whole-based method, Biomedical Signal Processing and Control 47 (2006) 196–206.

[19] M. Landowne, A method using induced waves to study pressure propagation in human arteries, Circulation Research 5 (1957) 594–601.

[20] R. A. Payne, C. N. Symeonides, D. J. Webb, S. R. J. Maxwell, Pulse transit time measured from the ecg: An unreliable marker of beat-to-beat blood pressure, Journal of Applied Physiology 100 (2006) 136–141. doi:10.1152/japplphysiol.00657.2005

[21] L. A. Geddes, M. H. Voelz, C. F. Babbs, J. D. Bourland, W. A. Tacker, Pulse transit time as an indicator of arterial blood pressure, Psychophysiology 18 (1) 71–74. doi:10.1111/j.1469-8986.1981.tb01545.x

[22] X. Fan, Q. Ye, X. Yang, S. Choudhury, Robust blood pressure estimation using an rgb camera, Journal of Ambient Intelligence and Humanized Computing 11. doi:10.1007/s12652-018-1026-6

[23] Z. Shen, F. Miao, Q. Meng, Y. Li, Cuffless and continuous blood pressure estimation based on multiple regression analysis, in: 2015 5th International Conference on Information Science and Technology (ICIST), 2015, pp. 117–120. doi:10.1109/ICIST.2015.7288952
[24] D. Han, J. Zhang, S. Shan, Leveraging auxiliary tasks for height and weight estimation by multi task learning, in: 2020 IEEE International Joint Conference on Biometrics (IJCB), 2020, pp. 1–7. doi:10.1109/IJCB48548.2020.9304855.

[25] J. Y. Kim, B. H. Cho, S. M. Im, M. J. Jeon, I. Y. Kim, S. Kim, Comparative study on artificial neural network with multiple regressions for continuous estimation of blood pressure, in: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, 2005, pp. 6942–6945. doi:10.1109/IEMBS.2005.1616102.

[26] N. Maher, G. Elsheikh, W. Anis, T. Emara, Non-invasive Calibration-Free Blood Pressure Estimation Based on Artificial Neural Network, 2020, pp. 701–711. doi:10.1007/978-3-030-14118-9_69.

[27] R. He, Z.-P. Huang, L.-Y. Ji, J.-K. Wu, H. Li, Z.-Q. Zhang, Beat-to-beat ambulatory blood pressure estimation based on random forest, in: 2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN), 2016, pp. 194–198. doi:10.1109/BSN.2016.7516258.

[28] Y. Zhang, Z. Feng, A svm method for continuous blood pressure estimation from a ppg signal, Proceedings of the 9th International Conference on Machine Learning and Computing.

[29] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778. doi:10.1109/CVPR.2016.90.

[30] S. Xie, R. Girshick, P. Dollar, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, 2017, pp. 5987–5995. doi:10.1109/CVPR.2017.634.

[31] Z. Yu, X. Li, G. Zhao, Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks, in: 30th British Machine
[32] S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J. F. Cohn, N. Sebe, Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2396–2404. doi:10.1109/CVPR.2016.263.

[33] F. Schrumpf, P. Frenzel, C. Aust, G. Osterhoff, M. Fuchs, Assessment of non-invasive blood pressure prediction from ppg and rppg signals using deep learning, Sensors 21 (18).