



HAL
open science

Long Short-Term Memory Deep-Filter in Remote Photoplethysmography

D. Botina-Monsalve, Yannick Benezeth, R. Macwan, P. Pierrart, F. Parra, K. Nakamura, R. Gomez, J. Mitéran

► **To cite this version:**

D. Botina-Monsalve, Yannick Benezeth, R. Macwan, P. Pierrart, F. Parra, et al.. Long Short-Term Memory Deep-Filter in Remote Photoplethysmography. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jun 2020, Seattle (virtual), United States. pp.1242-1249, 10.1109/CVPRW50498.2020.00161 . hal-03110208

HAL Id: hal-03110208

<https://u-bourgogne.hal.science/hal-03110208>

Submitted on 14 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Long Short-Term Memory Deep-Filter in Remote Photoplethysmography

D. Botina-Monsalve¹, Y. Benezeth¹, R. Macwan³, P. Pierrart¹, F. Parra¹, K. Nakamura², R. Gomez², J. Miteran¹

¹Univ. Bourgogne Franche-Comté, ImViA EA7535, France

²Honda Research Institute Japan Co., Honcho, Wako-shi, Saitama, Japan

³LPL, CNRS-UMR7538, Univ. Paris 13, France

Deivid-Johan.Botina-Monsalve@u-bourgogne.fr

Abstract

Remote photoplethysmography (rPPG) is a recent technique for estimating heart rate by analyzing subtle skin color variations using regular cameras. As multiple noise sources can pollute the estimated signal, post-processing techniques, such as bandpass filtering, are generally used. However, it is often possible to see alterations in the filtered signal that have not been suppressed, although an experienced eye can easily identify them. From this observation, we propose in this work to use an LSTM network to filter the rPPG signal. The network is able to learn the characteristic shape of the rPPG signal and especially its temporal structure, which is not possible with the usual signal processing-based filtering methods. The results of this study, obtained on a public database, have demonstrated that the proposed deep-learning-based filtering method outperforms the regular post-processing ones in terms of signal quality and accuracy of heart rate estimation.

1. Introduction

Heart rate (HR) measurement has made it possible to monitor the behavior of the heart in humans, and analyze physiological conditions. There are mainly two ways to measure HR, by means of electrocardiography (ECG) and photoplethysmography (PPG). ECG measures the electrical field induced by the heart activity in the chest, and PPG measures reflection and absorption of light on skin tissues, which indirectly captures the blood volume variations caused by the periodic beating of the heart. During cardiac contraction (systole), the volume of blood increases, and during the cardiac relaxation (diastole), the volume of blood decreases. These contact-based HR measurement methods may cause discomfort, hygiene issue, or even be impossible on fragile skins [22].

Verkruysse *et al.* [26] demonstrated that these perfusion variations, associated with heartbeats are not noticeable by the naked eye, but can be measured from a standard video

camera to measure PPG signals remotely, using ambient light as an illumination source. This remote Photoplethysmography (rPPG) technique is a low-cost and noninvasive way to measure HR. RPPG aims to estimate color changes in the frames. It is also worth mentioning here that there is another technique to obtain the HR remotely using a camera, by Ballistography (BCG) such as in [21]. BCG measures the periodic head movements due to ejections of blood into great vessels along with each heartbeat and, that way extracts HR signals.

With PPG or rPPG measurement, several biomedical parameters can potentially be computed: vascular occlusion, peripheral vasomotor activity, breathing rate, blood pressure by pulse transit time estimation, blood oxygen level, heart rate variability (HRV), and obviously heart rate [1]. As a consequence, the applications are numerous, including control of vital signs in the elderly and newborns, mixed reality, lie detection in criminals, physiological measurements of drivers, face anti-spoofing, and automatic skin detection, to mention a few [6, 4, 1]. First of all, it can be noted that the advent of non-contact measurements has opened the way to many new applications. Secondly, it is also important to note here that many of these applications use a fairly advanced analysis of the temporal signal (e.g. detection of systolic and diastolic peaks [17]), which is considerably more complex than frequency analysis over long periods (usually ten to twenty seconds to measure heart rate for example).

In the literature, multiple ways of rPPG estimation can be found. Initially, only the green channel was used [27]. Later approaches can be categorised in two groups: the first ones are based on Blind Source Separation techniques (BSS): ICA [12], PCA [14], PVM [15], EVM [30]. The second ones are based on a light interaction model to determine a projection vector: CHROM [9], POS [29], PBV [10]. Readers interested in learning more about these techniques can refer to the state of the art reviews presented in [16, 11, 24]. End-to-end approaches based on deep learning have also been used recently [7, 22, 6, 4, 32]. One of

the main advantages of these CNN-based measurements is that it allows achieving good results without the need for the designer to analyze the problem in depth [33]. In addition, it is no longer necessary to use a pipeline-based framework where regions of interest (ROI) are first detected and tracked over frames, RGB channels are then combined to estimate the pulse signal, which is filtered and analyzed to extract physiological parameters such as heart rate or respiration rate. As a consequence, CNN-based approaches are less prone to error propagation in the pipeline. However, as noted in [33], recent works in this field have focused on performance rather than understanding. Consequently, it is often hard to predict the limitations of the system, and it is well known that the training dataset used is critical.

RPPG signals estimated from a video usually contain some noise due to motion, illumination variations, internal noise of the digital camera, and also due to the estimation technique. This generally calls for post-processing steps to smooth the signal and remove unnecessary information, such as frequencies out of the regular physiological range of interest. This smoothing operation is habitually performed by a bandpass filter (*e.g.* in [18]). Some works also used wavelet filters [3, 15] or Singular Spectrum Analysis (SSA) [28]. Although these filtering methods do smooth the signals, they do not necessarily eliminate particular signal alteration, which can, however, be easily identified by experts. An example of one rPPG signal extracted and filtered by a band pass filter and a wavelet filter is presented in figure 1. RPPG signals filtered by both methods show significant alterations compared with their PPG ground truth signal. This kind of noise may alter the accuracy of heart rate measurement but more gravely prevent further advanced analysis of the rPPG signals that are based on pulse shape characteristics or peak detection on the temporal signal. For example, it has been shown that even a single heart period artifact, within a 2-minutes recording, can lead to errors of heart rate variability features that are considerably larger than typical effect size in psychological studies [2]. Therefore, in this work, we want to benefit from the advantages of deep learning-based methods to improve the filtering of rPPG signals while relying on a handcrafted pipeline thanks to which it is possible to control the system limits better.

Recurrent Neural Networks (RNNs) have been used in numerous applications with sequential information as inputs, like Natural Language Processing (NLP) and speech processing. RNNs are relevant in applications where the structure embedded in the data sequence transfers valuable knowledge, much like an expert would do to identify the wrong peaks in an rPPG signal. However, these networks show a problem during backpropagation: they are sensitive to vanishing and exploding gradients. This implies that the gradients can fade or explode as a consequence of the multiplication of many small or large derivatives during training.

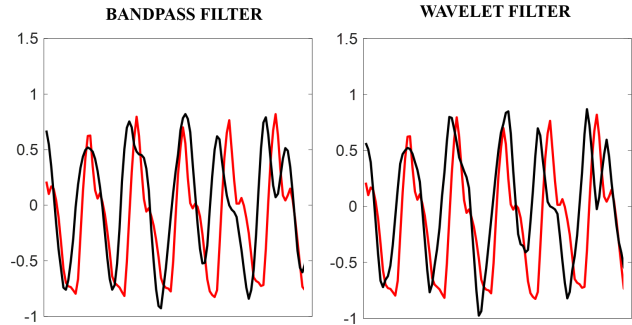


Figure 1. Example of noise in two remote photoplethysmographic signals after bandpass and wavelet filtering. The PPG ground truth signal is presented in red while the filtered rPPG signal is presented in black.

As a result, although their goal is to learn long-term dependencies, literature shows that RNNs cannot store information for very long. Long Short-Term Memory networks (LSTMs), on the other hand, provide memory blocks in its recurrent connections, storing more information for a longer period of time and avoiding the problem presented in RNN networks [13, 4]. This kind of networks have already been used for biomedical signal analysis. For example, Oh *et al.* [19] proposed a CNN-LSTM model to detect five different classes of arrhythmias. Yildirim [31] a bidirectional LSTM to classify ECG signals. Tan *et al.* [25] used convolutional neuronal networks (CNNs) and LSTMs to classify normal versus coronary artery disease ECG signals, the CNN is applied to reduce the number of data points in 5-second ECG signals, and the LSTM is applied to extract the temporal features of the signals. Thus, RNNs are indeed capable of handling of heartbeat-related signals.

To the best of our knowledge, no deep learning approach has been used to filter biomedical signals and especially to filter rPPG signals. In this article we propose an LSTM-based deep-filter to reduce noise in rPPG signals, to improve signal quality and in consequence, improve HR or HRV measurement. We validate our methodology using the public database Multimodal Spontaneous Emotion Corpus - Heart Rate database (MMSE-HR) [34].

This paper is organized as follows: In section 2 an overview of the methods used are presented, the Periodic Variance Maximization (PVM) used to acquire rPPG signals from video, and the Long Short-Term Deep-Filter (LSTM-DF). In section 3 the experiments implemented are explained, with the specification of the LSTM-DF architecture and its hyperparameters. In section 4 a comparison is made between band pass filtering, wavelet filtering and LSTM-DF filtering. Finally, conclusions and perspectives are presented in section 5.

2. Method

The proposed method aims to filter rPPG signals with the implemented LSTM-DF filter. The workflow of the procedure is presented in Figure 2. The first step consists of the rPPG signal estimation from a video sequence. As a second step, the rPPG signals are filtered with our LSTM-DF filter. The main components of the framework are explained in the following subsections.

2.1. rPPG signal estimation

The first step of the rPPG signal extraction procedure consists in the selection of the Region Of Interest (ROI). For each video frame, face is detected and tracked using the *reduced ResNet-10 model* and the *MedianFlow Tracker* of the OpenCV library [5]. Skin detection, as formulated by Conaire et al. [8], is then performed to select the skin pixels, which subsequently are spatially averaged to obtain a triplet of RGB values per frame and concatenated to obtain the RGB temporal trace. The RGB temporal traces are then preprocessed to remove the DC component of the signals dividing samples by their mean over a temporal interval. DC-normalized RGB signals are then band-pass filtered using Butterworth filter (with cut-off frequencies of 0.7 and 3.5Hz).

Then, we decided to use the recently proposed PVM algorithm to estimate the rPPG signals because of its efficiency and good performances. The PVM method has been extensively described and evaluated in [15]; we preset below the main concepts. The algorithm extracts the rPPG signal by estimating the time period τ^* that maximizes the periodic information embedded in the temporal mixture \mathbf{x} , of size $3 \times M$, obtained by spatial averaging of RGB video frames over a temporal window of M samples. τ^* is in turn found by searching for the optimum 3×3 weighting matrix \mathbf{W} that maximizes periodicity over the range of different lags τ_i corresponding to the human heart rate in the range [7-3] Hz. The ideal weighting matrix \mathbf{W}_{τ_i} at each τ_i is estimated by performing GEVD on the pair of the covariance and lagged covariance matrices $\mathbf{C}_x = \mathbf{x}\mathbf{x}^T$ and $\mathbf{P}_x = \mathbf{x}\mathbf{x}_\tau^T$ respectively. Here \mathbf{x}_τ is the signal mixture \mathbf{x} lagged by τ samples, representing the original signal mixture \mathbf{x} evolved over time. The 3×1 generalized eigenvector \mathbf{w} corresponding to the strongest generalized eigenvalue is used to extract the temporal signal $y(t)$, represented as the vector \mathbf{y} of size $1 \times M$, using $\mathbf{y} = \mathbf{w}^T \mathbf{x}$.

This formulation is based on the fact that the similarity between the quasi-periodic \mathbf{y} and its lagged version \mathbf{y}_τ is strongly related to its periodicity and can be expressed in terms of a periodicity metric given by

$$\mathbb{P}(\tau, \mathbf{w}) \doteq \frac{\mathbf{y}\mathbf{y}_\tau^T}{\mathbf{y}\mathbf{y}^T} = \frac{\mathbf{w}^T \mathbf{P}_x \mathbf{w}}{\mathbf{w}^T \mathbf{C}_x \mathbf{w}}. \quad (1)$$

2.2. Long Short-Term Deep-Filter (LSTM-DF)

In order to train the LSTM-DF network that will filter rPPG signals, we have created a data matrix \mathbf{Y} composed of a set of rPPG signals with a fixed length. The same process is applied on the corresponding PPG signals to create the ground truth data matrix $\hat{\mathbf{Y}}$. \mathbf{y}_i is the rPPG signal of the video i with length T_i , where $i \in [1, N]$ and N is the number of videos in the database, and therefore the number of rPPG signals. \mathbf{Y} and $\hat{\mathbf{Y}}$ are created with a simple sliding window procedure; the algorithm is presented in Algorithm 1, with L being the length of the sliding window and s being the step size.

Algorithm 1: RPPG training dataset building

```

count ← 1;
N ← 79;
L ← 5;
s ← 0.04;
for i ← 1 : N do
    for t ← 1 : s : Ti - L do
        Y(count) ← yi,train(t : t + L);
        Ŷ(count) ← ŷi,train(t : t + L);
        count = count + 1;
    end
end

```

A filtered rPPG signal $\tilde{\mathbf{y}}$ is predicted by the LSTM-DF from an rPPG signal \mathbf{y} , this is carried out as a regression problem. An LSTM network is composed of an input, a memory block, and an output layer. The input is the rPPG signal \mathbf{y} . The memory block consists of: one *input gate* that learns which information should be stored in the memory block, a *forget gate* that learns how much information must be forgotten or withheld from the memory block, and an *output gate* that takes care of learning when the collected information can be used. The architecture of the LSTM-DF is shown in Figure 3 where b is the batch size, L is the length of \mathbf{y} , and Return Sequence (RS) is an argument that decides whether a layer outputs each time step or its final time step. The size of the input signal is defined to be $b \times L \times 1$. The output size of the first two layers remains $b \times L \times 125$ due to their number of units and their return sequence argument. The third and fourth layers decrease the output to $b \times L \times 1$ representing, as a result, the predicted filtered rPPG signal $\tilde{\mathbf{y}}$. A dropout step is done on the first three layers to avoid the overfitting problem. The *mean squared error* (MSE) metric is chosen as the loss function for measuring the difference between $\tilde{\mathbf{y}}$ and $\hat{\mathbf{y}}$. The batch size b and the length of the signals L are discussed in the experimental section.

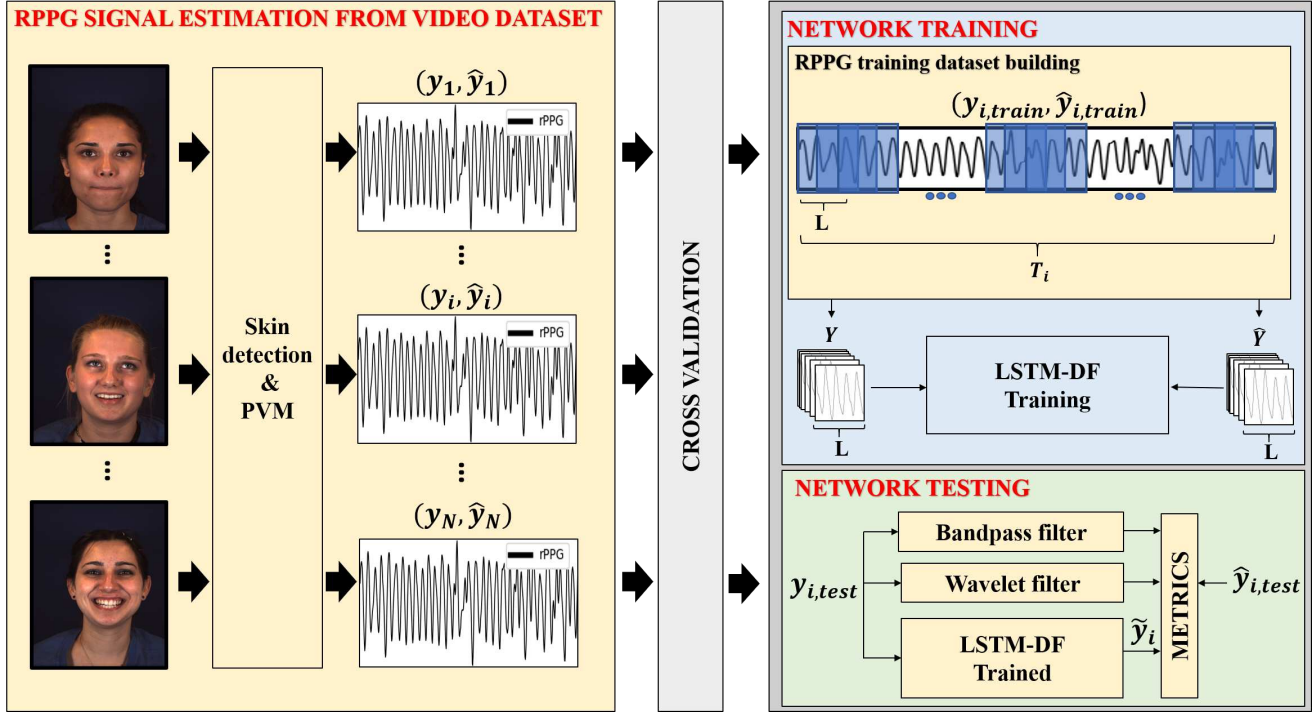


Figure 2. System framework: rPPG signal estimation is done in a sequence of videos using face and skin detection followed by the PVM algorithm [15]. Then an rPPG data set for training LSTM-DF network is built, concatenating fixed-length rPPG signal windows. Finally, LSTM-DF is compared with bandpass and wavelet filters in rPPG filtering application. A cross-validation procedure is used to assess the performance of the LSTM-DF.

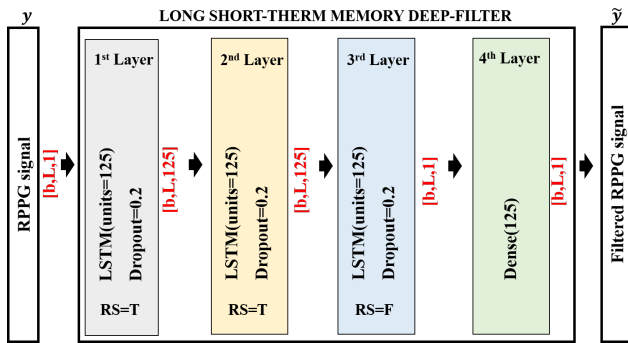


Figure 3. LSTM-DF architecture. Where numbers in red represent the size of each data. RS=Return sequences, T=True, F=False.

3. Experiments

3.1. Database

The LSTM-DF was trained and tested using the database Multimodal Spontaneous Emotion Corpus - Heart Rate database, which was made with emotion recognition in mind, so the videos include a large number of facial movements and expressions. The length of each video is between 30 to 60 seconds with a resolution of 1040×1392 pixels

and a frame rate of 25 fps. The ground truth signal is the PPG signal with 1 kHz of sampling rate, obtained by the BIOPAC 150 data acquisition system [34]. $N = 79$ videos were used in our experiments with a 5-fold cross validation procedure.

3.2. Metrics

We compare our proposed filtering method with two classical approaches, namely a bandpass (BP) filtering with cutoff frequencies in 0.7 and 3.5 Hz, and a wavelet (WV) filtering technique using the same parameters as in [15]. For all the experiments, we evaluate the performance of our algorithm in terms of the quality of the heart rate estimation and also in terms of the quality of the estimated rPPG signal. Therefore, we use Mean Absolute Error (MAE) in beats per minute (bpm) and Pearson's correlation coefficient (r) to evaluate the heart rate estimation performance and Signal to Noise Ratio (SNR) and Template Matching Correlation (TMC) [20] to evaluate the quality of the estimated signal. Metrics are calculated using heart rates estimated from the PPG signal and the filtered rPPG signals by the three filtering methods. 15-second window length was used to compute HR.

3.3. LSTM-DF setup

The proposed method was implemented on a Windows PC equipped with Intel Xeon 2.4 GHz CPU, with a RAM of 16GB, and with a NVIDIA GeForce RTX 2070 GPU, using the Keras (*v.2.2.4*) library with the Tensorflow (*v.2.0*) backend. LSTM layers were set with default hyperbolic tangent activation function, and the dense layer with a linear activation function. The first two LSTM layers had a dropout equal to 0.2, the loss function was mean squared error, and the Adam optimizer was set with a learning rate of 0.002. The random weights initialization were made with the Glorot uniform initializer, also called Xavier uniform initializer with seed=1.

First, we conducted an experiment to determine the best hyperparameters of our network varying the batch sizes and the number of epochs. Table 1 presents the results of this experiment using the architecture proposed in Figure 3 with a 5-fold cross validation procedure. Results given in the table are the average of the 5 folds. For these experiments, it is possible to observe that the obtained results are quite stable. Therefore, in all the following experiments, we decided to use a batch size of 32 and a number of epochs of 100.

Batch size	No. epochs	HR-MAE	SNR	TMC
8	30	1.50	9.69	0.71
32	30	1.51	9.59	0.70
32	100	1.32	9.44	0.72
64	100	1.32	9.50	0.70

Table 1. Results obtained varying the batch size and number of epochs of the LSTM network.

3.4. LSTM-DF filtering

A re-sampling procedure on \hat{y} from 1 KHz to 25 Hz was required in order to match with y , albeit without any interpolation or perturbation in the signal shape, as 25 Hz is a multiple of 1 KHz. Each signal was also normalized between -1 and 1. Then, Algorithm 1 is applied with $y_{train,i}$ and $\hat{y}_{train,i}$ signals to create the matrices \mathbf{Y} and $\hat{\mathbf{Y}}$, with a $L = 5$ seconds and $s = 0.04$ seconds (*c.f.* the *rPPG training dataset building* section in Figure 2). For each of the 5 folds a \mathbf{Y} and $\hat{\mathbf{Y}}$ was used to train the LSTM-DF network. Then, filtered rPPG signals $\tilde{y}_{test,i}$ are obtained by filtering $y_{test,i}$ with the deep filter. The number of rPPG signals used for training and testing in fold number 1 was 64 and 15 respectively, while for folds from 2 through 5, it was 63 and 16 respectively.

4. Results and discussion

We compare in Table 2 our filtering method with a bandpass filter (BP) and a wavelet filter (WV). The results are

presented as the average of each measure for the 5 folds, the last row of the table is the average of all values in the 5 folds. The best result for each metric and each fold are depicted in bold.

As mentioned previously, we use *MAE* to evaluate the heart rate estimation performance along with *SNR* and Template Matching Correlation (*TMC*) to evaluate the signal quality. First, it is possible to observe that the wavelet and bandpass filtering methods show very similar results in terms of *MAE* and that these results are already quite good, with *MAE* below *2bpm* in average. Interestingly, the proposed method presents the lowest *MAE* values for all the folds as well as in average. This observation confirms that our LSTM-DF improves the accuracy of the heart rate measurements. This is probably due to the removal of false peaks in the rPPG signals (as seen in Figure 5), allowing the FFT method to find the dominant frequency more precisely.

Figure 4 shows the correlation plots of the HR estimations for the three filtering methods. Pearson correlation metric r is also presented in the figure. Similar conclusions can be drawn with these correlation plots. Even if the distribution of the HR estimations of BP and WV are already quite good, the results of LSTM-DF is even better with a fitting line closer to the 45° line as compared to BP and WV. The Pearson correlation coefficient is also improved with LSTM-DF from 0.88 and 0.89 for BP and WV to 0.96.

With respect to the evaluation of the signal quality, Table 2 shows the *SNR* and the *TMC* metrics. Regarding to *SNR*, the wavelet filter shows a slight signal quality improvement over the bandpass filter. The LSTM-DF filter, on the other hand, presents a considerable improvement in the *SNR* metric with respect to the other two methods for all the folds as well as their average. The same behavior is present in the results of the *TMC* metric, where the proposed method surpasses the other filtering methodologies in all the folds, and therefore in their average.

With the proposed method of filtering, rPPG signals are smoothed in a different way than with wavelet or bandpass filters. To be more specific, in Figure 5, three filtered rPPG signals are depicted, the first one was filtered by BP, the second one by WV, and the last one by our LSTM-DF. WV and BP filtering methods, give a filtered rPPG signal with a good approximation of the number of peaks present in the PPG signal. However, it can be seen that these methods also remove the characteristic shape of the PPG signal, specifically, its dichrotic notch [23]. Consequently, rPPG signals filtered by BP and WV filters appear to be sinusoidal in shape. On the other hand, LSTM-DF allows not only to preserve the frequency and the number of peaks of the PPG signal, but also preserves the characteristic shape of the ground truth PPG signal, which is also quantified by the *TMC* metric in Table 2. In Figure 5, it is possible to see that the characteristic *dichrotic notch* is preserved with

Fold	HR measurement			Signal Quality					
	HR-MAE [bpm]			SNR [dB]			TMC [%]		
	BP	WV	LSTM-DF	BP	WV	LSTM-DF	BP	WV	LSTM-DF
1	1.53	1.52	0.53	7.41	7.44	10.47	0.49	0.50	0.82
2	2.82	2.61	2.24	5.54	5.63	8.42	0.49	0.50	0.64
3	1.80	1.79	1.43	6.52	6.61	9.95	0.32	0.33	0.71
4	0.99	1.00	0.84	7.04	7.20	9.59	0.64	0.65	0.75
5	2.31	2.40	1.38	4.88	4.92	9.00	0.43	0.44	0.65
Average	1.88	1.85	1.31	6.27	6.36	9.44	0.48	0.49	0.71

Table 2. 5-fold cross validation comparison in HR-measurement and signal quality improvement in rPPG signals filtered by: Bandpass filter (BP), Wavelet filter (WV), and the method proposed (LSTMD-F).

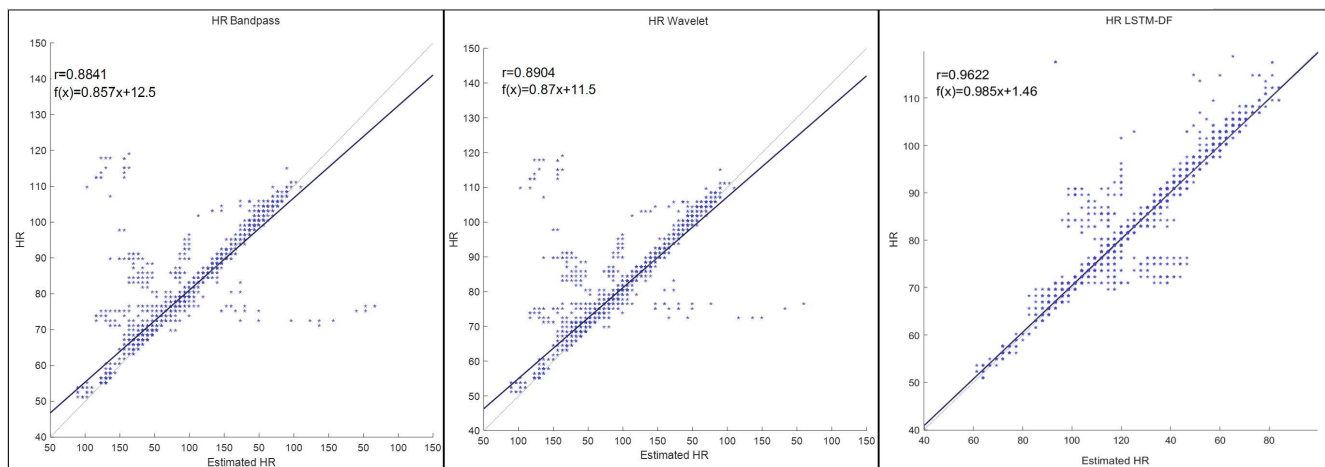


Figure 4. Correlation plots of the HR estimation obtained with (from left to right) bandpass filtering, wavelet filtering and the proposed LSTMD-DF filtering.

the the LSTM-DF filtering (*c.f.* the blue circles).

5. Conclusions and perspectives

The Long Short-Term Memory Deep-Filer (LSTM-DF) was presented in this paper to filter rPPG signals as an alternative to conventional signal processing techniques that cannot encapsulate the characteristic shape of the PPG signal neither its temporal structure. We have demonstrated, on a public database, that the LSTM-DF increases the accuracy of heart rate measurements as well as the quality of the estimated signal. Interestingly, the LSTM-DF filter preserves the characteristic shape of a PPG signal and paves the road to an advanced analysis of the temporal signal. It is important to note that the same methodology can be applied to the filtering of other temporal signals with characteristic shapes (*e.g.* ECG signals).

However, there is still a lot of experimentation to be car-

ried out to validate our methodology further. For example, it would be interesting to validate our work on other databases and to evaluate the robustness during a cross-database validation. Finally, even if we postulate that our filtering better preserves the shape of the PPG signal and thus allows to estimate more easily advanced temporal characteristics, it now seems necessary to validate this assertion experimentally.

References

- [1] J. Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological measurement*, 28(3), 2007.
- [2] G. Berntson and J.R. Stowell. Ecg artifacts and heart period variability: don't miss a beat! *Psychophysiology*, 35(1):127–132, 1998.

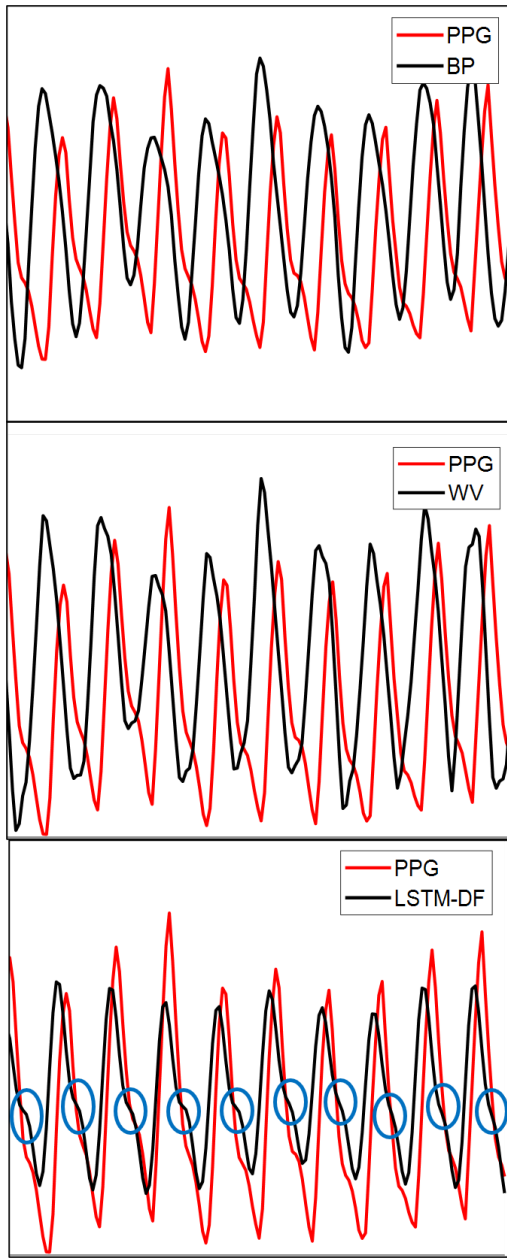


Figure 5. Signal quality and correspondence with ground truth PPG signals, of rPPG signals filtered by: bandpass (BP), wavelet (WV) and LSTM-DF filters. Blue circles show the remaining characteristic *hitch* presented in filtered rPPG signals.

[3] F. Bousefsaf, C. Maoui, and A. Pruski. Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate. *Biomedical Signal Processing and Control*, 8(6):568–574, 2013.

[4] F. Bousefsaf, A. Pruski, and C. Maoui. 3d convolutional neural networks for remote pulse rate measurement and mapping from facial video. *Applied Sciences*, 9(20), 2019.

[5] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of*

Software Tools, 2000.

[6] S. Chaichulee, M. Villarroel, J. Jorge, C. Arteta, K. McCormick, A. Zisserman, and L. Tarassenko. Cardio-respiratory signal extraction from video camera data for continuous non-contact vital sign monitoring using deep learning. *Physiological measurement*, 40(11), 2019.

[7] W. Chen and D. McDuff. Deepphys: Video-based physiological measurement using convolutional attention networks. In *The European Conference on Computer Vision*, 2018.

[8] Ciarán Ó Conaire, Noel E. O'Connor, and Alan F. Smeaton. Detector adaptation by maximising agreement between independent data sources. *CVPR*, 2007.

[9] G. De Haan and V. Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Trans. on Biomedical Engineering*, 60(10):2878–2886, 2013.

[10] G. De Haan and A. Van Leest. Improved motion robustness of remote-ppg by using the blood volume pulse signature. *Physiological measurement*, 35(9), 2014.

[11] M.A. Hassan, A.S. Malik, D. Fofi, N. Saad, B. Karasfi, Y.S. Ali, and F. Meriaudeau. Heart rate estimation using facial video: A review. *Biomedical Signal Processing and Control*, 38:346–360, 2017.

[12] A. Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE trans. on Neural Networks*, 10(3):626–634, 1999.

[13] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[14] M. Lewandowska, J. Rumiński, T. Kocejko, and J. Nowak. Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity. In *federated conference on computer science and information systems*, pages 405–410, 2011.

[15] R. Macwan, S. Bobbia, Y. Benezeth, J. Dubois, and A. Mansouri. Periodic variance maximization using generalized eigenvalue decomposition applied to remote photoplethysmography estimation. In *IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, 2018.

[16] D.J. McDuff, J.R. Estep, A.M. Piasecki, and E.B. Blackford. A survey of remote optical photoplethysmographic imaging methods. In *IEEE International Engineering in Medicine and Biology Conference*, pages 6398–6404, 2015.

[17] D. McDuff, S. Gontarek, and R. Picard. Remote detection of photoplethysmographic systolic and diastolic peaks using a digital camera. *IEEE Trans. on Biomedical Engineering*, 61(12):2948–2954, 2014.

[18] X. Niu, S. Shan, H. Han, and X. Chen. Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation. *IEEE Trans. on Image Processing*, 2019.

[19] S.L. Oh, E. Ng, R. San Tan, and U.R. Acharya. Automated diagnosis of arrhythmia using combination of cnn and lstm techniques with variable length heart beats. *Computers in biology and medicine*, 102:278–287, 2018.

[20] C. Orphanidou, T. Bonnici, P. Charlton, D. Clifton, D. Valance, and L. Tarassenko. Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring. *IEEE journal of biomedical and health informatics*, 19(3):832–838, 2014.

- [21] E.J. Pino, J. Chávez, and P. Aqueveque. Bcg algorithm for unobtrusive heart rate monitoring. In *IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT)*, pages 180–183, 2017.
- [22] Ying Qiu, Yang Liu, Juan Arteaga-Falconi, Haiwei Dong, and Abdulmotaleb El Saddik. Evm-cnn: Real-time contactless heart rate estimation from facial video. *IEEE Transactions on Multimedia*, 21(7):1778–1787, 2018.
- [23] Rakesh Sahni. Noninvasive monitoring by photoplethysmography. *Clinics in perinatology*, 39(3):573–583, 2012.
- [24] Y. Sun and N. Thakor. Photoplethysmography revisited: from contact to noncontact, from point to imaging. *IEEE Trans. on Biomedical Engineering*, 63(3):463–477, 2015.
- [25] J.H. Tan, , Y. Hagiwara, W. Pang, I. Lim, S.L. Oh, M. Adam, et al. Application of stacked convolutional and long short-term memory network for accurate identification of cad ecg signals. *Computers in biology and medicine*, 94:19–26, 2018.
- [26] Wim Verkrusse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Optics express*, 16(26):21434–21445, 2008.
- [27] Wim Verkrusse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Optics express*, 16(26):21434–21445, 2008.
- [28] W. Wang, B. Balmaekers, and G. de Haan. Quality metric for camera-based pulse rate monitoring in fitness exercise. *IEEE International Conference on Image Processing*, 2016.
- [29] Wenjin Wang, Albertus C den Brinker, Sander Stuijk, and Gerard de Haan. Algorithmic principles of remote ppg. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2016.
- [30] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM transactions on graphics (TOG)*, 31(4):1–8, 2012.
- [31] O. Yildirim. A novel wavelet sequence based on deep bidirectional lstm network model for ecg signal classification. *Computers in biology and medicine*, 96:189–202, 2018.
- [32] Zitong Yu, Xiaobai Li, and Guoying Zhao. Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks. In *Proc. BMVC*, pages 1–12, 2019.
- [33] Q. Zhan, W. Wang, and G. de Haan. Analysis of cnn-based remote-ppg to understand limitations and sensitivities. *Biomedical Optics Express*, 11(3):1268–1283, 2020.
- [34] Z. Zhang, J.M. Girard, Y. Wu, X. Zhang, et al. Multimodal spontaneous emotion corpus for human behavior analysis. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 3438–3446, 2016.