Estimation of blood pressure waveform from facial video using a deep U-shaped network and the wavelet representation of imaging photoplethysmographic signals

Frédéric Bousefsaf, Théo Desquins, Djamaleddine Djeldjli, Choubeila Maaoui, Alain Pruski

Estimation of blood pressure waveform from facial video using a deep U-shaped network and the wavelet representation of imaging photoplethysmographic signals

Frédéric Bousefsaf[*,1], Théo Desquins[1,2], Djamaleddine Djeldjli[1], Yassine Ouzar[1], Choubeila Maaoui[1], Alain Pruski[1]

1. Université de Lorraine, LCOMS, F-57000 Metz, France

2. i-Virtual, F-57000 Metz, France

Abstract

BACKGROUND. The remote measurement of physiological signals from video has gained a particular attention over the last past years. Estimating cardiovascular parameters like oxygen saturation and arterial blood pressure (BP) is covered by a limited volume of studies and remain a very challenging issue. Recent attempts demonstrated that BP can be estimated from facial video but under very controlled scenarios or with moderate performances. The data used in these works have not been publicly released or were gathered in a clinical setting. **METHODS**. We, in contrast, propose a framework for estimating BP from publicly available data in order to allow replication and to facilitate fair comparison. We developed and trained a deep U-shaped neural network to recover the blood pressure waveform from its imaging photoplethysmographic (iPPG) signal counterpart. The model predicts the continuous wavelet transform (CWT) representation of a BP signal from the CWT of an iPPG signal. Inverse CWT transform is ultimately computed to recover the BP time series. **RESULTS**. The proposed framework has been evaluated on 57 participants using international standards developed by the AAMI and the BHS. Results exhibit close agreement with ground truth BP values. The method satisfies all standards in the estimation of mean and diastolic BP (grade A) and nearly all standards in the estimation of systolic BP (grade B). **CONCLUSIONS**. This is, to the best of our knowledge, the first demonstration of a deep learning-oriented framework that manages to

predict the continuous blood pressure waveform from facial video analysis. Codes developed during the study are publicly available (`https://github.com/frederic-bousefsaf/ippg2bp`).

Keywords: imaging photoplethysmography, blood pressure, continuous wavelet transform, deep learning, U-Net

---

1. Introduction

Research on the remote measurement of physiological signals and cardiovascular parameters from facial video has made significant progress the last past years. The field is booming and supported by several significant studies [1]. The principle, termed imaging (or remote) photoplethysmography (iPPG), consists in measuring the subtle fluctuations of skin color. These fluctuations reflect complex light-tissue interactions. The simplest cameras (webcams) to the most advanced ones (professional, laboratory or industrial cameras) can be employed to reliably recover iPPG signals. Different regions of interest (ROI) have been studied over time but the face remains the most frequently observed area [2]. Several studies demonstrated that pulse rate and its variability can be robustly and precisely estimated with conventional image processing techniques and, more recently, with deep learning solutions [3, 4].

Current research in this field is now directed towards the measurement of new physiological parameters such as oxygen saturation [5] and blood pressure [6]. Estimating arterial blood pressure (BP) from video is covered by a limited volume of studies and remain a very challenging issue.

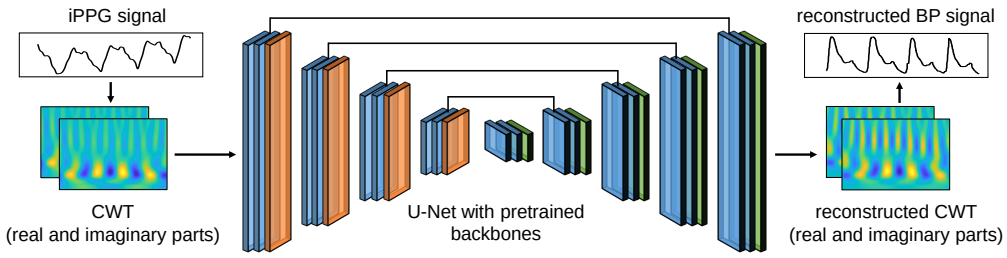

Figure 1: General overview of the method.

Two research directions are considered. First, measurement of the pulse transit time (PTT) on single [7] or several [8] ROI. PTT is a parameter considered to be correlated with blood pressure. Secondly, analysis of the iPPG signal waveform [6, 9]. To our knowledge, deep learning techniques have only been considered by Schrumpf et al. for the estimation of blood pressure from iPPG signals [10]. The model includes 5 layers and exhibit moderate performances, i.e. high mean average error and no compliance with international standards. These recent attempts demonstrated that BP can be estimated from facial video but under very controlled scenarios or with moderate performances. In addition, the data used in these works have not been publicly released or were gathered in a clinical setting. Only Schrumpf et al. released a sub-part of the data employed in their study. At the time of writing, this subset includes small excerpts of iPPG signals and discrete BP values from 17 over 50 participants (see `https://github.com/Fabian-Sc85/non-invasive-bp-estimation-using-deep-learning`). To conclude on this point, training an artificial neural network that accurately estimates blood pressure from video is constrained by the amount of available data because few public databases exist.

We propose, in this article, a framework for estimating BP from publicly available data. The dataset, namely BP4D+, includes video streams of moving participants. Video analysis dedicated to remote physiological sensing is therefore very challenging. A deep learning-oriented method (see figure 1) has been specifically developed to recover the blood pressure waveform from its imaging photoplethysmographic (iPPG) signal counterpart. The deep U-shaped model presented in this work has already been applied for translating iPPG to contact PPG signals in a previous work [11]. The full pipeline includes several stages. Skin pixels are first extracted using a recent segmentation techniques that relies on fully convolutional networks. iPPG signal is computed by averaging all the skin pixels from the green channel. We then employed the continuous wavelet transform (CWT) of iPPG (and respectively BP) signals to train the aforementioned neural architecture. The model therefore predicts a CWT representation of a BP signal from the CWT of an iPPG signal. Inverse CWT transform is ultimately computed to recover the BP time series.

The article includes five additional sections. Section 2 presents the background and related works. Section 3 introduces the used data and the developed methodologies. The full processing pipeline is detailed in this section. The metrics and results of the proposed approach are presented and discussed

in sections 4 and 5, respectively. We present the future works and a summary of the contributions in section 6.

This is, to the best of our knowledge, the first demonstration of a deep learning-oriented framework that manages to predict the continuous blood pressure waveform from iPPG signals computed using publicly released data. Several avenues of interest are envisaged to improve this research that, in its current state, exhibits very encouraging results. Two out of three estimated measures (i.e. diastolic BP and mean BP) already satisfy metrics defined by international standards.

2. Related works

A survey related to blood pressure estimation from video has recently been proposed by Lu et al. [12]. Several studies of interest have nevertheless been proposed since its publication. We therefore, and in the two first subsections, propose to review the studies that exploit iPPG for blood pressure assessment using both conventional and deep learning approaches. The estimation of blood pressure from contact PPG is closely related to this topic. We therefore dedicate the last subsection to this part.

2.1. iPPG for blood pressure estimation from propagation time

Systolic and diastolic blood pressures have been estimated using the propagation time of pulse waves from two different skin areas (typically hand and face) in video recordings [13, 14, 15, 8]. The positional of the two skin areas must be maintained during the measurement. This approach is therefore very restrictive. In this context, the time delay must be robustly assessed. Dedicated techniques were proposed for this purpose the last past years. Shao et al. compared peak locations from iPPG signals measured from two sites [16]. To improve accuracy, the peaks were estimated with two linear curves fitted on the edges of the rising and falling parts of the signal. Fan and Tjahjadib [17] analyzed the wave peaks with a custom signal quality index. Peaks of low confidence are removed using a Kalman filter to improve performances. Sugita et al. proposed to analyze videos of human hands recorded at different heights from the heart [18]. They analyze the difference in amplitude of iPPG pulse waves to build a model that estimates SBP.

2.2. iPPG for blood pressure estimation from single facial region

The estimation of BP from a single facial region is covered by very few studies in the scientific literature. The general approach, inspired from the

contact PPG field [19, 20], consists in computing waveform features that are correlated to BP. In this direction, Djeldjli et al. recently showed that temporal, derivative and area features computed from iPPG and cPPG waveform evolve similarly [21].

Jain et al. developed a simple regression framework that analyzes 21 waveform features computed on the iPPG signal to estimate BP [22]. Sugita et al. proposed to quantify the degree of distortion of iPPG signals [7]. They showed that this quantity exhibits correlation with BP close to correlations computed between BP and propagation times. Viejo et al. estimated BP from video using handcrafted features and machine learning models [23]. They studied the evolution of BP using a shallow neural network in the context of food sensory responses but no direct BP assessment is presented in their article.

The seminal work from Luo et al. [6] presents for the first time a pipeline that includes an artificial intelligence model. A multilayer perceptron has been fed with 30 features computed from iPPG waves. Their results show that iPPG waveform extracted from video exhibits information that are correlated to BP. Combining handcrafted features from iPPG signals with a machine learning approach to estimate systolic and diastolic BP has also been investigated by Rong and Li [9]. Deep learning architectures were recently studied by Schrumpf et al. [10]. The authors fine-tuned a network that integrates convolutional, long short-term memory and dense layers. They conclude that iPPG signals computed from standard RGB video streams may not be suitable to reliably estimate BP. All these studies pointed out the feasibility of remote BP monitoring from facial video but showed that there is still room for improvements and that the estimation remains a very challenging issue. A synthetic overview of the existing studies is presented in table 1. An important disparity in the number of subjects as well as overall low performances can be observed from this table. In addition, all the results presented in these studies have been tested on data that has not been released. To the best of our knowledge, no research dedicated to the estimation of blood pressure from iPPG has yet been conducted with public datasets.

2.3. Blood pressure estimation from contact PPG

Estimating absolute BP values from contact PPG (cPPG) remains a challenging problem even if there is clear evidence that the fluctuations in BP are reflected in cPPG signals [19, 20].

Deep learning techniques have recently been investigated [26] and recent developments show that these frameworks can effectively be deployed to convert BP waveform from cPPG signals. Different type of artificial neural architectures have been proposed the last past years. They combine fully connected [27] or convolutional layers [28] with long short-term memory. Simultaneous estimation of systolic and diastolic BP is ensured by these networks. Demographic features (e.g. weight and height) have additionally been included in machine learning algorithms to improve BP estimation from cPPG signals [29]. Time, frequency and time-frequency features were computed from the PPG and their derivative signals. Feature selection techniques were used for reducing the computational complexity and simultaneously decreasing the chance of over-fitting the machine learning algorithms.

Number of subjects	**Sampling freq. (fps)**	**iPPG signal extraction**	**Features**	**Model**	**Performances**		**Ref.**
					SBP	**DBP**	
17	140	Green	T_{BH} index	–	-0.6†	–	[7]
45	50	PCA	21 time and frequency features	regression	3.90‡ $\pm$ 5.37	3.72‡ $\pm$ 5.08	[22]
45	15	Green	amplitude, freq. and pulse rate)	shallow ANN	–	–	[23]
1328	30	TOI	155 features (30 after PCA)	ANN (MLP)	0.67† 0.39* $\pm$ 7.30	0.63† -0.2* $\pm$ 6.00	[6]
189	30	Green	26 features (16 after feature selection)	SVR	9.97‡ 2.1* $\pm$ 3.35	7.59‡ 0.79* $\pm$ 2.58	[9]
25	32	POS	–	CNN-LSTM-Dense (transfer learning using MIMIC III)	13.6‡	10.3‡	[10]

Table 1: Overview of the existing studies in the field of BP estimation from single facial region in video streams.

*: bias

†: correlation coefficient

‡: Mean Absolute Error (MAE)

ANN: Artificial Neural Network

CNN: Convolutional Neural Network

Green: iPPG signal formed using only the green channel [24]

LSTM: Long Short-Term Memory

MLP: MultiLayer Perceptron

PCA: Principal Component Analysis

POS: Plane-Orthogonal-to-Skin method [25]

SVR: Support Vector Regression

TOI: Transdermal Optical Imaging [6]

A similar framework but with a deep architecture with residual connections has been proposed by Slapnicar et al. [30]. A part of the network is dedicated to the analysis of the spectral representation of the signal using gated recurrent units. Deep learning networks that manage to predict the continuous BP waveform from cPPG signals have recently been proposed [26]. An approximation network learns a rough approximation of the BP waveform while a refinement network further enhances the preliminary estimate. The approximation and refinement networks are based on a U-Net architecture [31].

3. Methods

3.1. Database

BP4D+ is a multimodal dataset publicly available to the research community[1]. The database initially includes the physiological, thermal, 2D video, 3D and different metadata and annotations of 140 participants [32]. Ten tasks were proposed to elicit different emotions in a lab environment.

Because of the nature of the tasks, strong motion artifacts are present alongside an ensemble of videos, leading to difficult iPPG signal extraction. Video analysis for remote physiological sensing is therefore very challenging. We conducted a first selection process where only videos presenting clear iPPG signals have been kept. The procedure relies on a conventional signal-to-noise ratio (SNR). The index is defined using the Fourier transform of iPPG signal in 15-second windowed intervals so that sub-parts of partially impacted videos can be selected. The SNR has already been used in the field of iPPG [33, 34]. All the selected video parts have been manually controlled after this first automatic preselection. A subset of 57 subjects (21 females, 36 males), leading to a total of 157 videos, has been built. We additionally removed samples where the reference continuous blood pressure signal was improperly constituted or flawed (negative values). Details about the selected participants and tasks are available on a dedicated file in the website hosting the project (`https://github.com/frederic-bousefsaf/ippg2bp`). This subset has been employed for training and testing the neural architecture presented in this study.

[1] `http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html`

Each signal (for each participant and for each task) has been processed using the different techniques detailed in section 3.2. Each full-length signal has been split in excerpts of 2.56 seconds defined over 256 values. This constituted a dataset of 4123 portions of signal. About 70% of the data (2887 randomly selected excerpts) has been reserved for training, 15% (618 randomly selected excerpts) for validation and the remaining 15% (618 randomly selected excerpts) for the testing phase. The different sets contain a balanced portfolio of the participants and tasks.

We computed systolic BP (SBP) by averaging the intensities of the max peaks over the entire excerpt. Diastolic BP (DBP) has been computed with a similar strategy but using the min peaks intensities instead of the max ones. Mean arterial pressure (MAP) is the average value computed over all the excerpt samples. The distribution of SBP, DBP and MAP values for the training, validation and test sets are presented in figure 2. The distributions share similar properties and ranges.

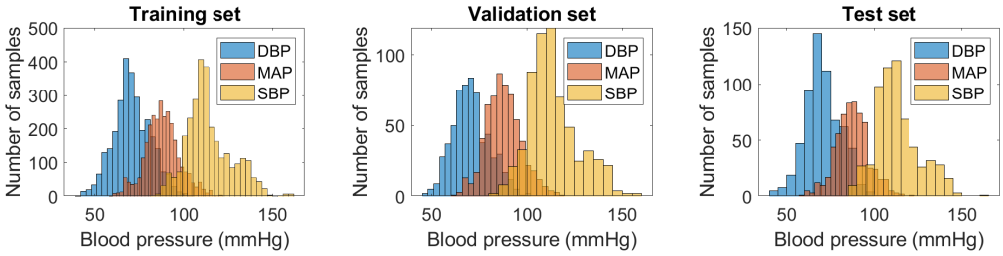

Figure 2: Distribution of DBP, MAP and SBP for the different sets. All the samples were extracted from the BP4D+ dataset.

3.2. iPPG signal constitution

The overall processing pipeline is quite similar to the one presented in [11]. This method (called iPPG 2 cPPG) consists in employing the continuous wavelet representation (real and imaginary parts) of an iPPG signal to reconstruct the wavelet representation of a contact PPG (cPPG) signal. Inverse transform is then computed to recover the cPPG time series.

First, we employed a recent face segmentation technique that relies on fully convolutional networks [37]. The approach robustly removed the background and non-skin areas. The method has recently been employed in the field of imaging photoplethysmography [38].

iPPG signal has been computed by averaging all the remaining skin pixels from the green channel. Figure 3a exhibits a raw iPPG signal computed from one of the BP4D+ video stream. Raw iPPG signals are then interpolated at a sampling frequency of 100 Hz and detrended using a specific low-pass filter [35] based on a smoothness priors that attenuates low frequencies [36].

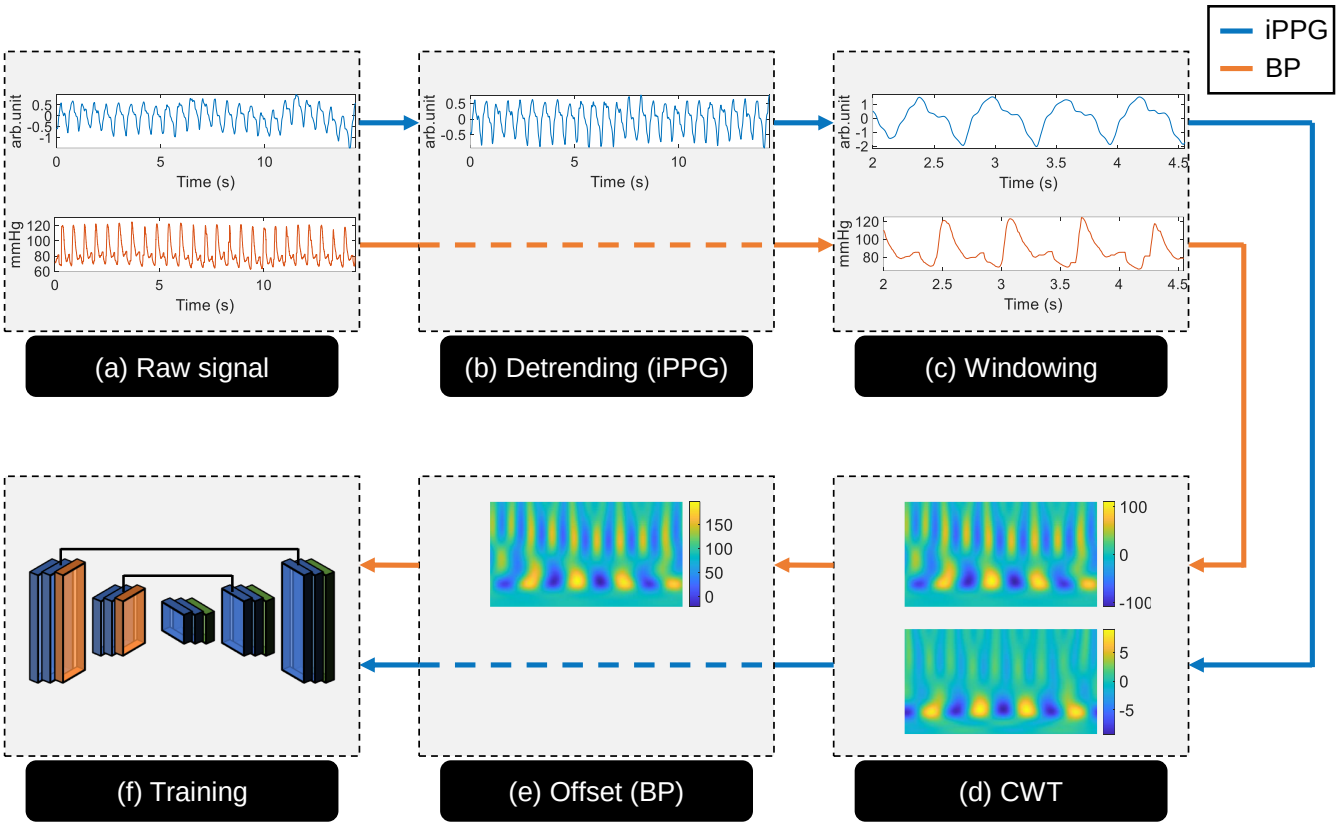

Figure 3: Signal processing before CWT computation. (a) Example of a raw iPPG signal that contains noise and trends (top illustration) and of a BP signal that has been simultaneously recorded using a continuous non-invasive sensor (bottom illustration). (b) iPPG trends removal is ensured by a method [35] that has already been used in this field [36]. (c) Small excerpts of 2.56 seconds are extracted for further processing. (d) The CWT (real part) of both iPPG and BP signals is computed in the frequency range [0.6, 4.5] Hz. (e) The average value is lost when computing the CWT in the aforementioned frequency range. This information is therefore directly encoded in the CWT of the BP signal by adding the mean value to every CWT coefficient. See the the difference in the ranges of the colorbars between subfigures (d) and (e). (f) The CWT (real and imaginary parts) are used for training the neural architecture presented in section 3.3.

Figure 3b shows the impact of the detrending operation on the iPPG signal. We then extract small excerpts for both the iPPG and the ground truth BP signals (see figure 3c for a typical example). An overlapping sliding window scheme has been selected to increase the volume of data employed during training. The sampling frequency of the interpolated iPPG signal being set to 100 Hz, 2.56 seconds are necessary to form time-frequency representations of 256 pixels in width. The window length has therefore been set to 2.56 seconds with an empirically defined step size of 0.5 seconds (50 samples). All the iPPG excerpts have been standardized using the z-score formula (so that $\mu = 0$ and $\sigma = 1$). Training, validation and testing sets were then constituted from this ensemble of excerpts (see section 3.1).

Like in [11], we employed the continuous wavelet transform (CWT) representation to train the neural architecture presented in section 3.3. The global approach is depicted in figure 3. The CWT (equation 1) of a signal $x(t)$ corresponds to a time-frequency representation computed from a prototype function commonly called mother wavelet. Unlike the Fourier transform, the wavelet transform can detect abrupt changes in frequency using a family of wavelets $\psi_{\tau,s}$ (equation 2) computed from the mother wavelet ψ.

$$CWT_x^{\psi}(\tau, s) = \int_{-\infty}^{\infty} x(t)\, \psi_{\tau,s}(t) dt \tag{1}$$

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{|s|}} \psi\left(\frac{t-\tau}{s}\right) \tag{2}$$

$\psi_{\tau,s}$ corresponds to the mother wavelet dilated by s and translated by τ. Dilating the wavelet allows the transform to analyze larger portions of signal in the time domain, thus covering lower frequencies. Different mother wavelets have been developed and the choice depends mainly on the application and the properties of the signal. The Morlet mother wavelet used in this study was already used in previous work related to the analysis of PPG signals by camera [39, 40, 11].

The original signal $x(t)$ can be reconstructed by the inverse transform:

$$x(t) = \frac{1}{C_{\psi}} \int_0^{\infty} \int_{-\infty}^{\infty} \frac{1}{s^2} CWT_x^{\psi}(\tau, s) \frac{1}{\sqrt{|s|}} \psi\left(\frac{t-\tau}{s}\right) d\tau\, ds \tag{3}$$

$$C_{\psi} = \int_0^{\infty} \frac{\left|\hat{\psi}(\zeta)\right|^2}{|\zeta|} d\zeta < \infty \tag{4}$$

C_ψ is the admissibility condition and $\hat{\psi}$ is the Fourier transform of ψ.

The continuous wavelet transform was computed on each iPPG and BP signal in the frequency range [0.6, 4.5] Hz, which corresponds to the physiological range of the human heart rate [2]. Typical iPPG signal, BP signal and their respective wavelet representations (real part) are presented in figure 4. As it was presented before, the iPPG signals have been standardized ($\mu = 0$ and $\sigma = 1$, see top-left illustration in figure 4 for a typical example). This type of process has not been applied to the BP signals because we need to recover both the average, systolic and diastolic values (see top-mid illustration in figure 4). The average value being lost when computing the CWT in the frequency range [0.6, 4.5] Hz, we chose to directly encode this information in the CWT of BP signals by adding the mean value to every CWT coefficients (see figure 3e):

$$CWT_{BP} = CWT_{BP} + \mu_{BP} \tag{5}$$

Here, μ_{BP} corresponds to the average value of a BP signal (top-mid illustration in figure 4 for a typical BP signal example) and CWT_{BP} to its

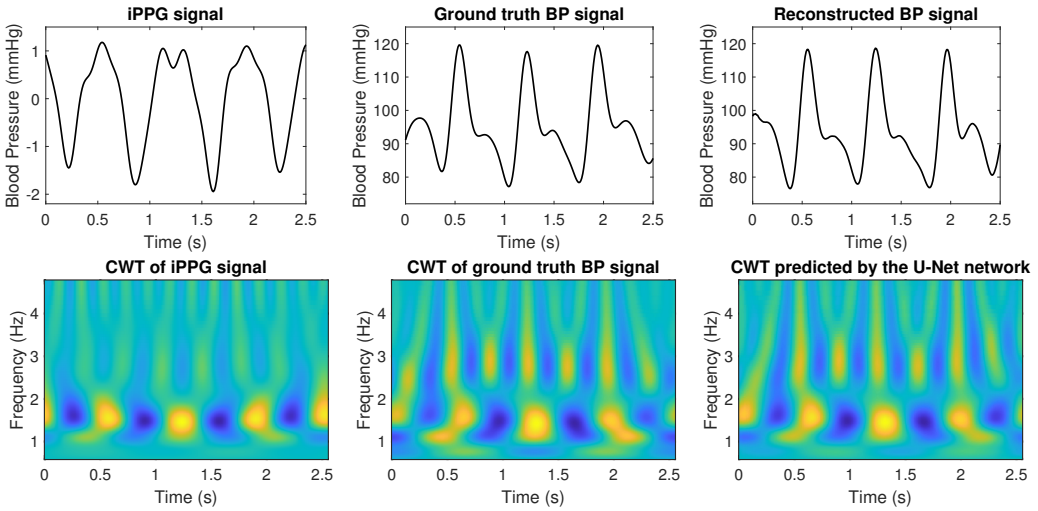

Figure 4: An iPPG and its corresponding ground truth BP are respectively presented in the bottom-left and bottom-mid figures. Their corresponding CWT (real part) are presented below. The transform (a complex image with a real and imaginary part) is computed in the frequency range [0.6, 4.5] Hz. Figures on the right present the CWT predicted by the neural network and the corresponding reconstructed BP signal, computed using the inverse CWT transform.

corresponding CWT (see bottom-mid illustration in figure 4).

The produced wavelet representations have a dimension of $256 \times 256 \times 2$ pixels. They are used to train the neural architectures (figure 3f) presented in the next section.

3.3. Neural architectures

The neural architecture has already been developed and tested in previous work [11]. Briefly, it consists in a U-Net architecture, which was initially proposed by Ronneberger et al. [31], enhanced by a backbone. This type of network has been widely used for segmentation of medical images [41]. Its architecture consists of a descending (encoder) branch completed by an ascending (decoder) branch, giving a U-shape to the network. The descending branch contains an ensemble of convolution and pooling layers. The ascending branch integrates upsampling layers connected to the convolutions of the descending branch. Connections help to restore the spatial information. A schematic representation of the network is provided in figure 5. Each convolutional layer are coupled with a Rectified Linear Unit (ReLU) activation function.

A Backbone (e.g. VGG16) can be integrated into the encoder part of the U-Net network. Its internal parameters can be blocked during training, meaning that the weights of the network remain the same. In practice, a backbone correspond to a model subpart pre-trained on ImageNet, a database deployed for object recognition tasks in images [42]. Training a U-Net network supported by a backbone consists, in this case, in optimizing the internal parameters of the decoder part. This approach can be associated to a transfer learning strategy. In this work, we initialized the U-Net architecture with a ResNeXt101 backbone [43]. The encoder parameters were not blocked during training, meaning that they were optimized during the learning phase. The number of variables to be trained (weights and biases) is 52 million. We chose ResNeXt101 because it performed better than other standard backbones on the reconstruction of contact PPG signals from non contact ones through their continuous wavelet representation, a problem that is in fact quite similar [11].

Conventional regularization techniques (e.g. dropout) have not been introduced while a normalization scheme (i.e. batch normalization) has been employed. Linear activation function was specified because the targeted task corresponds to a regression in the form of a pixel-to-pixel reconstruction of a two-channel wavelet representation.

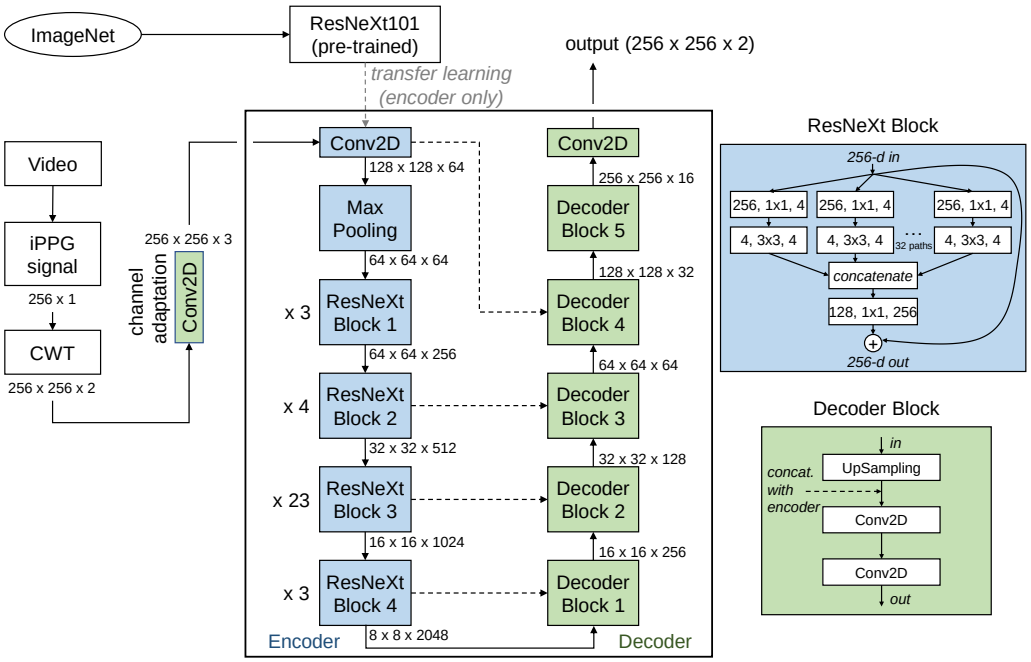

Figure 5: Overview of the U-Net [31] proposed in this study, which includes encoder (down-sampling) and decoder (upsampling) portions. The encoder is replaced by a ResNeXt101 backbone [43]. ResNeXt and decoder blocks are detailed on the right-side of the figure. The input of a ResNeXt block (256 dimensions in the example depicted in the figure) is split into 32 lower dimensional branches (or paths) that will next be merged through concatenation. This architecture exploits Inception's split-transform-merge strategy but with a uniform topology. The parameters of each stage inside this ResNeXt block example are respectively the number of input filters, the filter size and the number of output filters. Each ResNeXt block present different parameters. They are specified in [43].

The input dimensions of a U-Net network supported by a backbone are fixed by the data used for their training (256×256 pixels RGB images from the ImageNet database). The inputs being in our case a two-channels wavelet representation, an adaptation strategy must be introduced. We employed an additional 2D convolutional layer with a $(1, 1)$ kernel that has been placed between the input layer and the encoder part of the network. The neurons of this layer allow conversion of the input from N to 3 channels. The weights of all the networks have randomly been initialized by the method proposed by Glorot and Bengio [44]. Biases are initialized to zero. The Mean Squared Error (MSE) has been selected as loss for training all the models:

$$MSE = \frac{1}{n}\sum_{i,j}\left(CWT_{i,j} - \widehat{CWT}_{i,j}\right)^2 \tag{6}$$

CWT corresponds to the wavelet transform (see figure 3) of the ground truth BP signal. $\widehat{CWT}$ is the wavelet representation predicted by the neural network starting from the wavelet representation of the iPPG signal.

The architecture implementation was carried out under Python using Keras API and Tensorflow library. The Segmentation Models library [45] proposed by P. Yakubovskiy was used to develop the neural network. The training sessions were launched over 500 epochs through batches of 16 images. We used, in this study, the Adam optimization algorithm [46] with a learning rate of 0.001. A dedicated computer equipped with a dual Intel Xeon Silver 4114 and two Nvidia Quadro P6000s was used to carry out network learning.

4. Results

The proposed U-Net architecture transforms an iPPG signal to a continuous BP signal through their wavelet representation. Figure 4 illustrates a typical example of BP estimation (top-right figure) from an iPPG wave (top-left figure). The predicted waveform closely follows the ground truth BP wave presented in top-mid figure. The shape and magnitude, which were initially different, have been preserved. We can notice small phase differences in the wavelet representations of the iPPG signal (bottom-left figure)

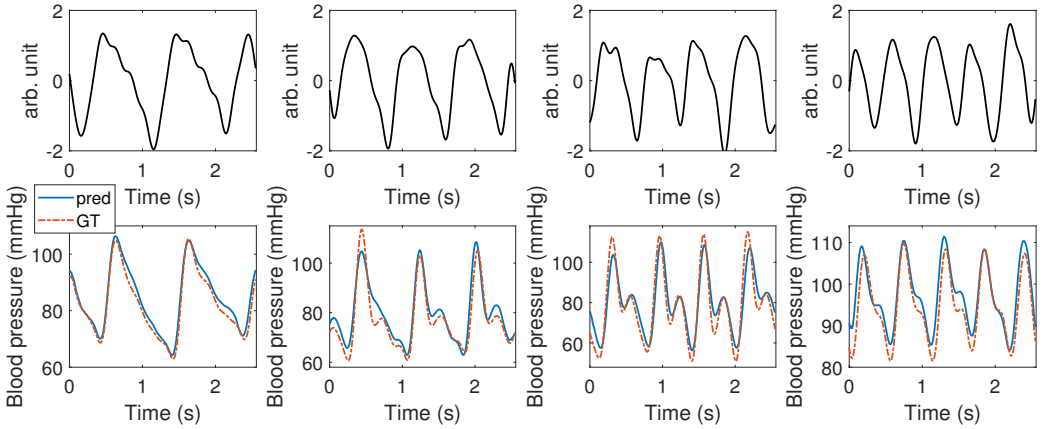

Figure 6: Typical BP signals reconstruction for several pulse rate values. Top figures: iPPG signals. Bottom figures: predicted and ground truth (GT) BP.

and the ground truth BP signal (bottom-mid figure). The neural network learned this specificity, the reconstructed wavelet representation (bottom-right figure) being in phase with the ground truth one (bottom-mid figure). The phase has therefore been properly recovered. This follows previous observations that we made when testing this U-Net to transform contact PPG to iPPG signals [11] and observations from other authors that employed deep learning to convert contact PPG to BP waves [26].

Figure 6 illustrates several examples of blood pressure estimation from iPPG signals. We evaluated the performances of the proposed technique with international standards [47, 48] from the Association for the Advancement of Medical Instrumentation (AAMI) and from the British Hypertension Society (BHS). We, however, emphasize that BP4D+ contains videos and physiological data that have not been recorded in a clinical setting. Also, the constituted subset integrates 57 participants while the AAMI recommends to evaluate BP estimation techniques on a minimum of 85 subjects.

4.1. General metrics and Bland-Altman plots

The Mean Absolute Error (MAE, equation 7) and the Root Mean Square Error ($RMSE$, equation 8) have been used to quantify the level of agreement between the predicted ($\widehat{BP}$) and the ground truth blood pressure (BP). We computed these metrics for DBP, MAP and SBP over all the test set (see section 3.1).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|BP_i - \widehat{BP_i}| \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(BP_i - \widehat{BP_i}\right)^2} \tag{8}$$

Table 2 presents a comparative analysis of results taken from similar works. Bland-Altman representations have been computed for DBP, MAP and SBP over all the test data. The average between the estimated and ground truth BP values is depicted on the x-axis while the differences between the estimated and ground truth BP values are depicted on the y-axis. The resulting plots are presented in figure 7. Means are represented by dash-dot lines and 95% limits of agreement ($\pm$ 1.96 SD) by dashed lines. The ranges of these limits are [-12.3 14.3], [-12.0 11.6] and [-19.6 16.6] for DBP, MAP and SBP respectively.

		MAE (mmHg)	RMSE (mmHg)
Rong and Li [9]	DBP	7.59	–
	SBP	9.97	–
Schrumpf et al. [10]	DBP	10.3	–
	SBP	13.6	–
iPPG2BP (our results)	DBP	5.1	6.85
	MAP	4.47	6.01
	SBP	6.73	9.34

Table 2: Blood pressure estimation errors. Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) have been computed between the estimated and ground truth DBP, MAP and SBP. Results from similar studies are also reported.

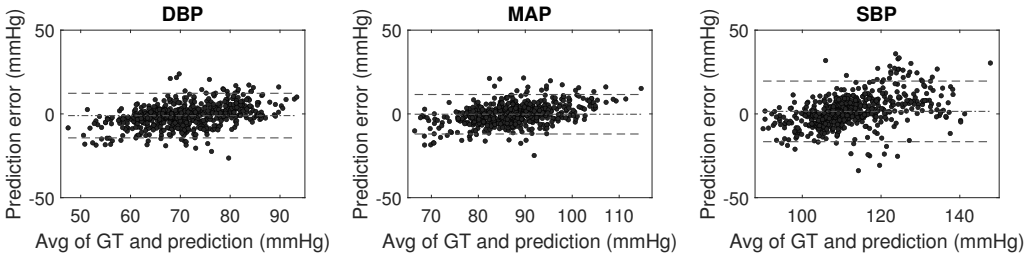

Figure 7: Bland-Altman plots for DBP, MAP and SBP prediction. Means are represented by dash-dot lines and 95% limits of agreement ($\pm$ 1.96 SD) by dashed lines.

4.2. BHS standards

The BHS assesses blood pressure estimation techniques by their cumulative percentage of errors [47]. Different grades are provided (see table 3) according to the percentage of the predictions on the test samples that fall under three empiric thresholds, i.e. 5, 10 and 15 mmHg.

Table 3 presents a comparative analysis of the BHS evaluation on our results. We reported the values provided by Rong and Li [9] as it appears to be the only study that computed BHS metrics. Our results exhibit good overall performances with more than 60%, 87% and 95% of the test samples having estimation errors less than, respectively, 5, 10 and 15 mmHg for both DBP and MAP (grade A). More than 50% and 79% of SBP predictions fall under 5 and 10 mmHg respectively (grade B) while 89.6% of SBP predictions fall under 15 mmHg, which is slightly under the 90% threshold.

The conclusions drawn from the analysis of the results presented in table 3 are graphically presented in figure 8.

		Cumulative Error Percentage		
		$\leq$ 5 mmHg	$\leq$ 10 mmHg	$\leq$ 15 mmHg
Rong and Li [9]	DBP	55.4%	85.7%	98.2%
	SBP	48.2%	78.6%	94.6%
iPPG2BP (our results)	DBP	60.2%	87.1%	95.8%
	MAP	66.8%	90.9%	96.4%
	SBP	50.2%	79.0%	89.6%
BHS	grade A	60%	85%	95%
	grade B	50%	75%	90%
	grade C	40%	65%	85%

Table 3: BHS metrics for DBP, MAP and SBP prediction.

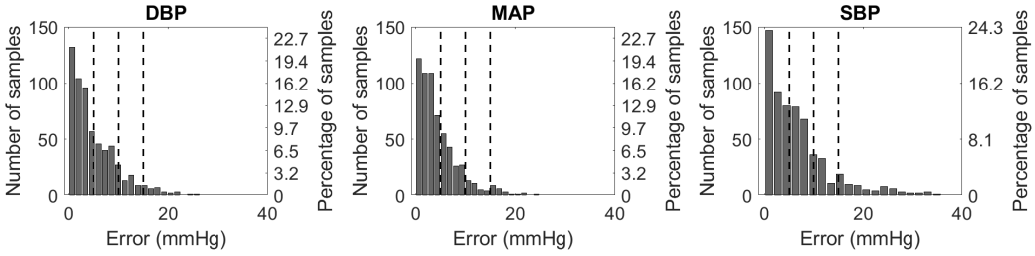

Figure 8: Absolute error in DBP, MAP and SBP predictions. Dashed lines represent the 5, 10 and 15 mmHg thresholds recommended by the BHS.

4.3. AAMI standards

The AAMI proposes to assess blood pressure estimation techniques by analyzing the mean error (ME) and the standard deviation of errors (SDE) on the test set [48]. The former must be lower than 5 mmHg while the latter must be lower than 8 mmHg to fully respect the recommendation.

Table 4 presents a comparative analysis of the AAMI evaluation on our results. We additionally reported the values provided by Luo et al. [6] and Rong and Li [9]. Our results exhibit good overall performances. Both DBP and MAP satisfy the AAMI standards. They exhibit a small ME and a SDE lower than 8 mmHg. Regarding SBP estimations, the ME condition is fulfilled but the SDE is a bit higher (1.2 mmHg over the 8 mmHg threshold defined by the AAMI).

The histograms of prediction errors for DBP, MAP and SBP are presented in figure 9. The spread of these histograms gives a graphical picture of the different SDE presented in 4 (narrower for MAP, wider for SBP).

		ME (mmHg)	SDE (mmHg)
Luo et al. [6]	DBP	-0.20	6.00
	SBP	0.39	7.30
Rong and Li [9]	DBP	0.79	2.58
	SBP	2.1	3.35
iPPG2BP (our results)	DBP	-1.001	6.781
	MAP	-0.205	6.007
	SBP	1.51	9.221
AAMI standard		≤ 5	≤ 8

Table 4: AAMI metrics for DBP, MAP and SBP prediction. ME: Mean Error; SDE: Standard Deviation of Errors.

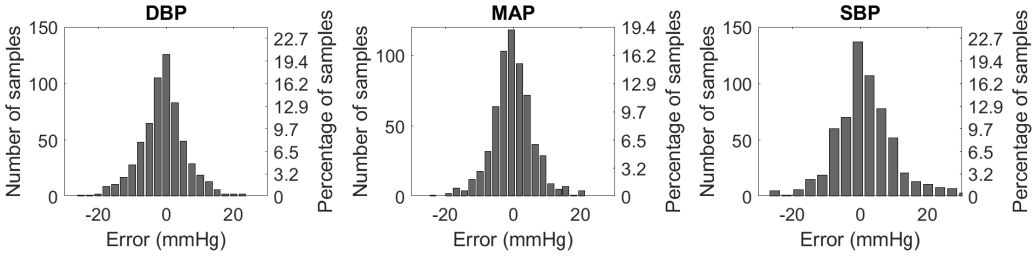

Figure 9: Error in DBP, MAP and DBP predictions.

5. Discussion

The method presented in this paper corresponds to one of the few proposals that relies on deep learning to estimate blood pressure from facial video. We propose, in the next subsection, to discuss and compare our results with related works. Section 5.2 presents the limitations of this study. We ultimately present and discuss the results of a leave-one-patient-out cross-validation procedure (section 5.3).

5.1. About the results presented in this study

Regarding previous works, and to the best of our knowledge, only Rong and Li presented Bland-Altman representations to assess their results. The technique proposed by the authors seems to underestimate low BP values and overestimate high BP values, both for DBP and SBP [9]. Our results depict a similar tendency but with lesser impact, the Bland-Altman plots presented in figure 7 being quite consistent across all the BP range. Table 2 presents a comparative analysis of results taken from similar works. The technique proposed in this study performs better than the other methods in

terms of MAE and RMSE. We, however, emphasize that the results reported from other studies were computed from data of different nature. To the best of our knowledge, these data are not publicly available.

Results presented in sections, 4.2 and 4.3 exhibit a relevant level of agreement between predicted and ground truth BP values. It can however be observed that several BP predictions exceed the 15 mmHg threshold, in particular for SBP (see table 3). We emphasize that no other techniques focusing on the analysis of BP from a single facial video have obtained grade B in SBP prediction, in particular from challenging data. Techniques dedicated to the conversion of contact PPG signals to the BP waveform [26] or from contact PPG signals to DBP and SBP values [30, 28, 29] also produce SBP estimations that are less relevant than DBP estimations. We do not report the AAMI and BHS analysis from Schrumpf et al. because none of their results seems to satisfy the requirements [10].

Integrating the wavelet representation of iPPG signals instead of raw iPPG signals in the network is a key-point of the method presented in this study. We here take advantage of transfer learning through a ResNeXt backbone pre-trained on large databases [11]. U-Nets have been widely used for segmentation of medical images and can be trained with a low volume of data [41].

5.2. Limitations

Figure 10 presents a prediction of lesser quality where the mean BP value is approximately estimated by the model. Apart from the mean error, DBP and SBP seem to be properly estimated. Adding more data during the learning phase of the network may solve, or at least minimize, this mean error. Balancing the distribution of ground truth BP values while varying the iPPG and BP waveform (shape of the signals) may be a relevant approach to tackle this issue.

All the presented results are limited by the current dataset: a low percentage of subjects (<85) has been used to derive the results presented in section 4. We point out that the reference blood pressure, gathered using a continuous non-invasive sensor, has not been recorded in a clinical setting. There might be irrelevant ground truth values, ultimately leading to improper BP learning by the U-Net model presented in section 3.3. We also emphasize that only videos presenting clear iPPG signals have been included in the dataset. Videos with motion can lead to iPPG signals that contain strong artifacts. This particular source of noise can negatively impact the

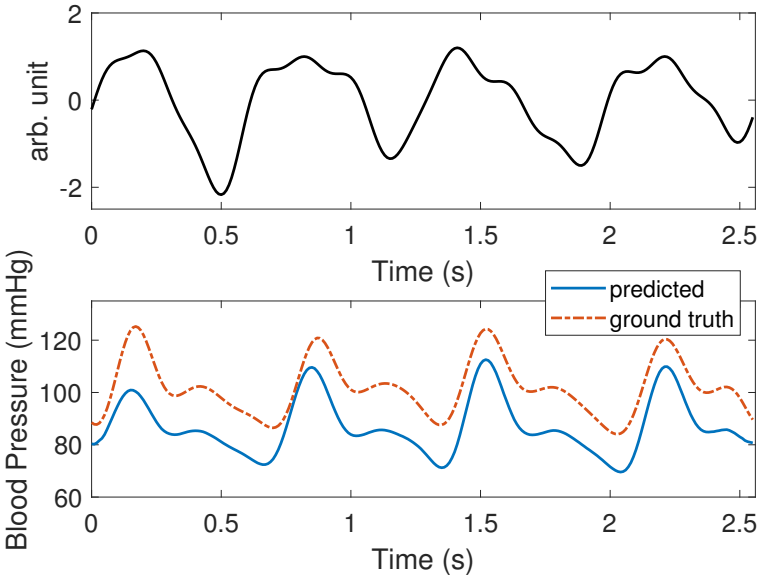

Figure 10: Predictions of lesser quality. Top figure: iPPG signal. Bottom figure: predicted and ground truth BP.

CWT coefficients. Including noisy iPPG signals into the dataset will be the objective of future works. Broaden the currently limited dataset is necessary so that all types of noise are represented.

The data distributions presented in figure 2 are not well-balanced across all the BP range. This can drastically impact training, in particular by negating the generalization power of the model (see next subsection). To tackle this issue, the development of a smart overlapping selection could be a potential approach. It would consist in automatically increasing the overlapping to produce more signals in the underrepresented BP ranges. We also emphasize that data augmentation strategies were recently proposed in the field of pulse rate estimation from video to improve the models performances [49]. These approaches are however not conceivable in the case of BP estimation because removing frames or augmenting the videos with conventional transformations may directly impact the shape of iPPG waveforms. Developing an augmentation strategy towards the wavelet representations, by for example adding random noise to the CWT coefficients, can here be an approach of interest.

The distributions of the training, validation and test sets presented in figure 2 contain a mix of all the participants data. In the next subsection,

we analyze the impact of a leave-one-patient-out cross-validation procedure on the performances.

5.3. Leave-one-patient-out cross-validation

Table 5 presents the assessment of BP using the method proposed in this study (section 3) but under a leave-one-patient-out cross-validation procedure (three folds). We can observe a decrease in performances over all the folds, even if some values are close from the international standards recommendations. The Bland-Altman representations for DBP, MAP and SBP over all data from the first fold are presented in figure 11. They exhibit wider point clouds than those computed from the randomly distributed subsets (see the Bland-Altman plots presented in figure 7) where each set includes a balanced portfolio of participants and tasks (details in section 3.1). We can also observe that SBP predictions depicted in figure 11 follow an inverse trend than those displayed in figure 7. Here, the trained model overestimates SBP in low BP values and underestimates SBP in high BP values. All these results exhibit a limitation in the generalization power of the network but are, in contrast, encouraging because the model has been trained with limited data.

It can also be observed, from table 5, that the model performed poorly for SBP estimations of fold 2. After a closer look on the iPPG signals and ground truth BP, we remarked that this decrease in performance was due to a patient who presents the highest SBP values. All these patient signals were included in the test and were therefore totally missing from the training set. We therefore believe that the network did not learn the features relative to these specific samples. As stated in the previous subsection, broadening the dataset is a necessary step to improve generalization.

6. Conclusion and future works

We proposed, in this article, a deep learning-oriented solution dedicated to the recovering of blood pressure from facial video. The reconstruction is carried out using a U-shaped network supported by a ResNeXt backbone from the time-frequency representation of the iPPG signal. To the best of our knowledge, this study presents the first demonstration of an automatic framework that manages to estimate the continuous BP waveform from facial video. The approach corresponds to an efficient way for predicting BP without a prior extraction of complicated hand-crafted waveform features from

Fold	Errors (mmHg)		BHS			AAMI (mmHg)		
	MAE	RMSE	≤ 5	≤ 10	≤ 15	ME	SDE	
1	8.28	11.78	49%	73%	83%	4.23	10.99	DBP
	7.52	10.66	50%	76%	86%	3.39	10.11	MAP
	9.79	12.64	33%	61%	76%	4.56	11.79	SBP
2	5.83	7.12	48%	85%	97%	0	7.12	DBP
	8.03	10.24	43%	65%	87%	3.97	9.45	MAP
	16.41	21.61	26%	46%	57%	12.99	17.27	SBP
3	11.43	14.12	28%	51%	68%	-4.77	13.29	DBP
	8.11	10.21	38%	69%	86%	-3.81	9.47	MAP
	8.87	11.33	38%	62%	81%	-4.77	10.28	SBP

Table 5: Assessment of the proposed solution under a leave-one-patient-out cross-validation procedure.

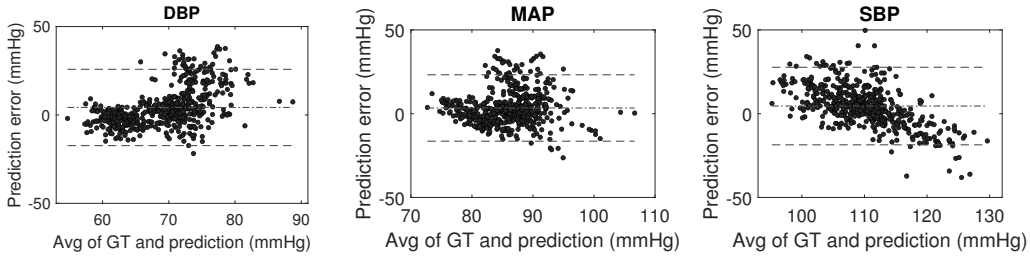

Figure 11: Bland-Altman plots for DBP, MAP and SBP prediction under a leave-one-patient-out, three folds, cross-validation procedure (only the results from the first fold are presented here).

the iPPG signal. Our extensive experiments showed the effectiveness of the proposed method, which achieves high accuracy and satisfies all international standards in the estimation of mean and diastolic BP (grade A) and nearly all international standards in the estimation of systolic BP (grade B).

Several ways of improvement for this work are considered. We first propose expanding the currently limited volume of data by increasing the number of included recordings and participants. We, in this study, conducted a manual selection of videos that presented well-defined iPPG signals. This step can be automatized using a quality index [17]. Also, it has recently been shown that data augmentation strategies can significantly improve the performances of deep learning models dedicated to pulse rate estimation from video [49]. Producing more overlapped signals in the range of low represented BP values might be a first considered approach for re-balancing the dataset distribution.

The Morlet wavelet has been used as a prototype function for the computation of the CWT. We propose evaluating the impact on performances with different mother wavelets as well as investigating different time-frequency representations like short-time Fourier and constant-Q transforms.

Inputting directly the video stream in an end-to-end architecture rather than the time-frequency representation of iPPG signal will be the subject of long-term research. We also envisage to extend this work in the context of blood oxygen saturation using a similar approach (inputting CWT representations of iPPG signals to a deep U-Net model).

7. Acknowledgments

This work has been partly funded by the Contrat Plan État Région (CPER) Innovations Technologiques, Modélisation et Médecine Personnalisée (IT2MP) and Fonds Européen de Développement Régional (FEDER).

References

[1] D. McDuff, Camera measurement of physiological vital signs, arXiv preprint arXiv:2111.11547 (2021).

[2] S. Zaunseder, A. Trumpp, D. Wedekind, H. Malberg, Cardiovascular assessment by imaging photoplethysmography–a review, Biomedical Engineering/Biomedizinische Technik (2018).

[3] A. Ni, A. Azarang, N. Kehtarnavaz, A Review of Deep Learning-Based Contactless Heart Rate Measurement Methods, Sensors 21 (2021) 3719. URL: https://www.mdpi.com/1424-8220/21/11/3719. doi:10.3390/s21113719.

[4] C.-H. Cheng, K.-L. Wong, J.-W. Chin, T.-T. Chan, R. H. Y. So, Deep Learning Methods for Remote Heart Rate Measurement: A Review and Future Research Agenda, Sensors 21 (2021) 6296. URL: https://www.mdpi.com/1424-8220/21/18/6296. doi:10.3390/s21186296.

[5] A. Al-Naji, G. A. Khalid, J. F. Mahdi, J. Chahl, Non-Contact SpO2 Prediction System Based on a Digital Camera, Applied Sciences 11 (2021) 4255. URL: https://www.mdpi.com/2076-3417/11/9/4255. doi:10.3390/app11094255.

[6] H. Luo, D. Yang, A. Barszczyk, N. Vempala, J. Wei, S. J. Wu, P. P. Zheng, G. Fu, K. Lee, Z.-P. Feng, Smartphone-based blood pressure measurement using transdermal optical imaging technology, Circulation: Cardiovascular Imaging 12 (2019) e008857.

[7] N. Sugita, M. Yoshizawa, M. Abe, A. Tanaka, N. Homma, T. Yambe, Contactless Technique for Measuring Blood-Pressure Variability from One Region in Video Plethysmography, Journal of Medical and Biological Engineering (2018) 1–10.

[8] X. Fan, Q. Ye, X. Yang, S. D. Choudhury, Robust blood pressure estimation using an RGB camera, Journal of Ambient Intelligence and Humanized Computing (2018) 1–8.

[9] M. Rong, K. Li, A Blood Pressure Prediction Method Based on Imaging Photoplethysmography in combination with Machine Learning, Biomedical Signal Processing and Control 64 (2021) 102328. URL: https://linkinghub.elsevier.com/retrieve/pii/S1746809420304444. doi:10.1016/j.bspc.2020.102328.

[10] 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 (2021) 6022. URL: https://www.mdpi.com/1424-8220/21/18/6022. doi:10.3390/s21186022.

[11] F. Bousefsaf, D. Djeldjli, Y. Ouzar, C. Maaoui, A. Pruski, iPPG 2 cPPG: reconstructing contact from imaging photoplethysmographic signals using U-Net architectures, Computers in Biology and Medicine 138 (2021) 104860. URL: https://linkinghub.elsevier.com/retrieve/pii/S0010482521006545. doi:10.1016/j.compbiomed.2021.104860.

[12] Y. Lu, C. Wang, M. Q.-H. Meng, Video-based Contactless Blood Pressure Estimation: A Review, in: 2020 IEEE International Conference on Real-time Computing and Robotics (RCAR), IEEE, Asahikawa, Japan, 2020, pp. 62–67. URL: https://ieeexplore.ieee.org/document/9303040/. doi:10.1109/RCAR49640.2020.9303040.

[13] N. Sugita, K. Obara, M. Yoshizawa, M. Abe, A. Tanaka, N. Homma, Techniques for estimating blood pressure variation using video images, in: Engineering in Medicine and Biology Society (EMBC), 2015 37th

Annual International Conference of the IEEE, IEEE, 2015, pp. 4218–4221.

[14] I. C. Jeong, J. Finkelstein, Introducing contactless blood pressure assessment using a high speed video camera, Journal of medical systems 40 (2016) 77.

[15] P.-W. Huang, C.-H. Lin, M.-L. Chung, T.-M. Lin, B.-F. Wu, Image based contactless blood pressure assessment using Pulse Transit Time, in: Automatic Control Conference (CACS), 2017 International, IEEE, 2017, pp. 1–6.

[16] D. Shao, Y. Yang, C. Liu, F. Tsow, H. Yu, N. Tao, Noncontact monitoring breathing pattern, exhalation flow rate and pulse transit time, IEEE Transactions on Biomedical Engineering 61 (2014) 2760–2767.

[17] X. Fan, T. Tjahjadi, Robust contactless pulse transit time estimation based on signal quality metric, Pattern Recognition Letters 137 (2020) 12–16.

[18] N. Sugita, T. Noro, M. Yoshizawa, K. Ichiji, S. Yamaki, N. Homma, Estimation of Absolute Blood Pressure Using Video Images Captured at Different Heights from the Heart, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 4458–4461.

[19] M. Elgendi, On the analysis of fingertip photoplethysmogram signals, Current cardiology reviews 8 (2012) 14–25.

[20] E. von Wowern, G. Östling, P. M. Nilsson, P. Olofsson, Digital photoplethysmography for assessment of arterial stiffness: repeatability and comparison with applanation tonometry, PloS one 10 (2015) e0135659.

[21] D. Djeldjli, F. Bousefsaf, C. Maaoui, F. Bereksi-Reguig, A. Pruski, Remote estimation of pulse wave features related to arterial stiffness and blood pressure using a camera, Biomedical Signal Processing and Control 64 (2021) 102242.

[22] M. Jain, S. Deb, A. Subramanyam, Face video based touchless blood pressure and heart rate estimation, in: Multimedia Signal Processing

(MMSP), 2016 IEEE 18th International Workshop on, IEEE, 2016, pp. 1–5.

[23] C. G. Viejo, S. Fuentes, D. D. Torrico, F. R. Dunshea, Non-Contact Heart Rate and Blood Pressure Estimations from Video Analysis and Machine Learning Modelling Applied to Food Sensory Responses: A Case Study for Chocolate, Sensors 18 (2018) 1802.

[24] W. Verkruysse, L. O. Svaasand, J. S. Nelson, Remote plethysmographic imaging using ambient light., Optics express 16 (2008) 21434–21445.

[25] W. Wang, A. C. den Brinker, S. Stuijk, G. de Haan, Algorithmic Principles of Remote PPG, IEEE Transactions on Biomedical Engineering 64 (2017) 1479–1491.

[26] N. Ibtehaz, M. S. Rahman, PPG2ABP: Translating Photoplethysmogram (PPG) Signals to Arterial Blood Pressure (ABP) Waveforms using Fully Convolutional Neural Networks, arXiv preprint arXiv:2005.01669 (2020).

[27] M. S. Tanveer, M. K. Hasan, Cuffless blood pressure estimation from electrocardiogram and photoplethysmogram using waveform based ANN-LSTM network, Biomedical Signal Processing and Control 51 (2019) 382–392.

[28] M. Panwar, A. Gautam, D. Biswas, A. Acharyya, PP-Net: A Deep Learning Framework for PPG based Blood Pressure and Heart Rate Estimation, IEEE Sensors Journal (2020). Publisher: IEEE.

[29] M. H. Chowdhury, M. N. I. Shuzan, M. E. Chowdhury, Z. B. Mahbub, M. M. Uddin, A. Khandakar, M. B. I. Reaz, Estimating Blood Pressure from the Photoplethysmogram Signal and Demographic Features Using Machine Learning Techniques, Sensors 20 (2020) 3127. Publisher: Multidisciplinary Digital Publishing Institute.

[30] G. Slapničar, N. Mlakar, M. Luštrek, Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network, Sensors 19 (2019) 3420. Publisher: Multidisciplinary Digital Publishing Institute.

[31] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.

[32] Z. Zhang, J. M. Girard, Y. Wu, X. Zhang, P. Liu, U. Ciftci, S. Canavan, M. Reale, A. Horowitz, H. Yang, others, Multimodal spontaneous emotion corpus for human behavior analysis, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3438–3446.

[33] G. de Haan, V. Jeanne, Robust pulse rate from chrominance-based rPPG, IEEE Transactions on Biomedical Engineering 60 (2013) 2878–2886.

[34] A. Hammer, M. Scherpf, M. Schmidt, H. Ernst, H. Malberg, K. Matschke, A. Dragu, J. Martin, O. Bota, Camera-based assessment of cutaneous perfusion strength in a clinical setting, Physiological Measurement (2022). URL: http://iopscience.iop.org/article/10.1088/1361-6579/ac557d.

[35] M. P. Tarvainen, P. O. Ranta-Aho, P. A. Karjalainen, An advanced detrending method with application to HRV analysis, IEEE transactions on biomedical engineering 49 (2002) 172–175. Publisher: IEEE.

[36] M.-Z. Poh, D. J. McDuff, R. W. Picard, Advancements in noncontact, multiparameter physiological measurements using a webcam, IEEE transactions on biomedical engineering 58 (2011) 7–11.

[37] Y. Nirkin, I. Masi, A. T. Tuan, T. Hassner, G. Medioni, On face segmentation, face swapping, and face perception, in: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), IEEE, 2018, pp. 98–105.

[38] Y. Ouzar, D. Djeldjli, F. Bousefsaf, C. Maaoui, LCOMS Lab's Approach to the Vision for Vitals (V4V) Challenge, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 2750–2754.

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

[40] F. Bousefsaf, C. Maaoui, A. Pruski, Peripheral vasomotor activity assessment using a continuous wavelet analysis on webcam photoplethysmographic signals, Bio-medical materials and engineering 27 (2016) 527–538.

[41] S. Leclerc, E. Smistad, J. Pedrosa, A. Østvik, F. Cervenansky, F. Espinosa, T. Espeland, E. A. R. Berg, P.-M. Jodoin, T. Grenier, others, Deep learning for segmentation using an open large-scale dataset in 2d echocardiography, IEEE transactions on medical imaging (2019).

[42] E. C. Too, L. Yujian, S. Njuki, L. Yingchun, A comparative study of fine-tuning deep learning models for plant disease identification, Computers and Electronics in Agriculture 161 (2019) 272–279. Publisher: Elsevier.

[43] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1492–1500.

[44] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the thirteenth international conference on artificial intelligence and statistics, 2010, pp. 249–256.

[45] P. Yakubovskiy, Segmentation Models, GitHub, 2019. URL: `https://github.com/qubvel/segmentation_models`, publication Title: GitHub repository.

[46] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[47] E. O'Brien, J. Petrie, W. Littler, M. de Swiet, P. L. Padfield, K. O'Malley, M. Jamieson, D. Altman, M. Bland, N. Atkins, The british hypertension society protocol for the evaluation of automated and semi-automated blood pressure measuring devices with special reference to ambulatory systems., Journal of hypertension 8 (1990) 607–619.

[48] G. S. Stergiou, B. Alpert, S. Mieke, R. Asmar, N. Atkins, S. Eckert, G. Frick, B. Friedman, T. Graßl, T. Ichikawa, others, A universal standard for the validation of blood pressure measuring devices: Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO) Collaboration Statement, Hypertension 71 (2018) 368–374. Publisher: Am Heart Assoc.

[49] Z. Yu, X. Li, X. Niu, J. Shi, G. Zhao, AutoHR: A Strong End-to-End Baseline for Remote Heart Rate Measurement With Neural Searching, IEEE Signal Processing Letters 27 (2020) 1245–1249. URL: https://ieeexplore.ieee.org/document/9133501/. doi:10.1109/LSP.2020.3007086.