iPPG 2 cPPG: Reconstructing contact from imaging photoplethysmographic signals using U-Net architectures

Frédéric Bousefsaf, Djamaleddine Djeldjli, Yassine Ouzar, Choubeila Maaoui, Alain Pruski

iPPG 2 cPPG: reconstructing contact from imaging photoplethysmographic signals using U-Net architectures

Frédéric Bousefsaf*, Djamaleddine Djeldjli, Yassine Ouzar, Choubeila Maaoui and Alain Pruski

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

Abstract

Imaging photoplethysmography (iPPG) is an optical technique dedicated to the assessment of several vital functions using a simple camera. Significant efforts have been made to reliably estimate heart and respiratory rates. Currently, research is focusing on the remote estimation of oxygen saturation and blood pressure (BP). The limited number of publicly available data tends to restrict the advancements related to BP estimation. To overcome this limit, we propose to split the problem in a two-stage processing chain: (i) converting iPPG to contact PPG (cPPG) signals using available video dataset and (ii) estimate BP from converted cPPG signals by exploiting large existing databases (e.g. MIMIC). This article presents the first developments where a method for converting iPPG signals measured using a camera into cPPG signals measured by contact sensors is proposed. Real and imaginary parts of the continuous wavelet transform (CWT) of cPPG and iPPG signals are passed to various deep pre-trained U-shaped architectures. Conventional metrics and specific waveform estimators have been implemented to validate the relevance of the predictions. The results exhibit good agreements towards a large portion of metrics, showing that the neural architectures properly estimated cPPG from iPPG signals through their CWT representations. The performance indicates that BP estimation from iPPG signals converted to cPPG signals can now be envisaged. Consequently, future work will focus on the integration of models dedicated to BP estimation trained on MIMIC. This is the first demonstration of a method for accurate reconstruction of cPPG from iPPG signals satisfying pulse waveform criteria.

Keywords: imaging photoplethysmography, U-Net, blood volume pulse,

pulse waveform

1. Introduction

In the recent years, research on contactless technologies dedicated to physiological signals measurement have made significant progress [1]. Photoplethysmography (PPG) can be remotely measured by observing the subtle fluctuations of skin color. These fluctuations reflect complex light-tissue interactions, from which their origin is not fully agreed [2]. The simplest cameras (webcams) to the most advanced ones (professional, laboratory or industrial cameras) can be used to reliably measure PPG signals [3]. Different regions of interest (ROI) have been studied over time but the face remains the most frequently observed area [4].

The field is booming and supported by several significant studies. Computer vision, image processing and artificial intelligence (AI) methods have been used or developed specifically to reliably transform input video into biomedical parameters [4]. Numerous studies have shown that pulse rate and its variability can be estimated with high robustness. In this context, artificial intelligence is playing an increasingly important role [5] where the most efficient pulse rate measurement methods are now based on deep neural models [6]. These architectures are often based on convolutional layers [7] and can be trained with synthetic data [8] reinforced by real data [9].

Current research in this field is now directed towards the measurement of new physiological parameters such as oxygen saturation [10] and blood pressure [11]. They impact the amplitude and waveform of PPG signals over

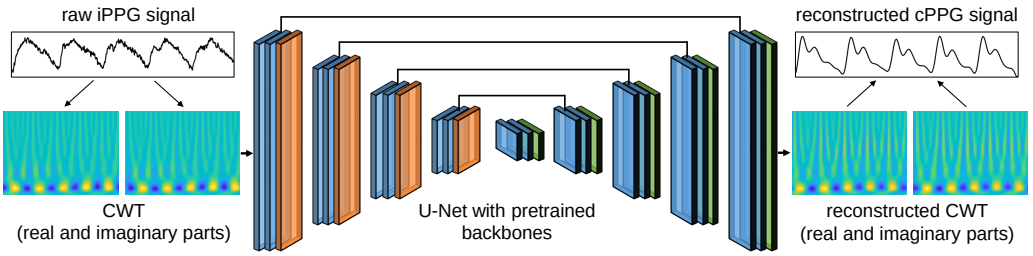


Figure 1: General overview of the method.

different wavelength ranges. Blood pressure estimation based on video analysis is complex and very few works show its feasibility. Two research directions are considered. First, measurement of the pulse transit time (PTT) on single [12] or several [13] ROI. PTT is a parameter considered to be correlated with blood pressure. Secondly, analysis of the PPG signal waveform [11]. To our knowledge, deep learning techniques based on video analysis have not been considered for the estimation of blood pressure yet.

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. Djeldjli et al. recently showed that temporal, derivative and area features computed from imaging PPG (iPPG) waveform and contact sensor (placed on the finger or the ear) evolve similarly [14]. This point is important because it motivates the present study. We envisage estimating BP with a two-stage processing chain. A model dedicated to the conversion of iPPG signals to contact PPG (cPPG) signals using available video dataset corresponds to the first part of the processing chain. The second stage consists in constituting a deep learning model dedicated to blood pressure estimation from these converted signals by exploiting large existing databases (e.g. MIMIC [15]).

The developments related to the first stage are presented in this study. To add more details, we propose to train a deep U-shaped neural architecture (U-Net) dedicated to the conversion of contact PPG signals from imaging PPG signals simultaneously measured on the face by conventional video analysis. Continuous wavelet representation of the signals is employed to take advantage of transfer learning through pretrained backbones on large databases. To the best of our knowledge, this is the first demonstration of a method for accurate reconstruction of cPPG from iPPG signals.

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 metrics and results of the proposed approach are presented and discussed in section 4. We present the future works and a summary of the contributions in sections 5 and 6, respectively.

2. Related works

This section reviews the studies that exploit deep learning for iPPG analysis as well as conventional and deep learning approaches for blood pressure assessment from both iPPG and cPPG.

2.1. Deep Learning for iPPG signal and pulse rate estimation

Relevant surveys in the imaging PPG field of research have been proposed the last past years [1, 3, 4]. They cover conventional techniques that generally include both image and signal processing approaches to improve PPG signal-to-noise ratio and therefore the estimation of biomedical parameters like pulse and breathing rates. Video and image processing operations like face detection, tracking of region(s) of interest and skin segmentation have been employed [16, 17, 18]. Constituting an iPPG signal from a sequence of frames is usually carried out with a spatial averaging operation [19]). Standard signal processing techniques include blind source separation approaches [20], Fourier and Wavelet transforms [21]. The impact of color space on pulse rate assessment has also been investigated in previous research [22, 17].

The most recent studies present artificial intelligence through deep learning methods to automatically estimate the pulse signal or directly the pulse rate. These approaches currently deliver the best performances and present root mean squared errors between 2.7 and 3.8 beats per minute [5] on public datasets like UBFC-RPPG [23], MAHNOB-HCI [24] and PURE [25]. Both hybrid and end-to-end approaches have been investigated. Hybrid strategies take either processed frames or iPPG signals as input and output the biomedical parameters of interest. End-to-end models takes a video (sequence of frames) as input and output the biomedical parameters.

Hybrid strategies combine conventional with deep learning methods. For instance, Qiu et al. developed a three-stage pipeline including face tracking, features extraction and finally pulse rate estimation based on a convolutional neural network (CNN) [26]. Hsu et al. proposed a deep CNN trained to predict pulse rate based on the time–frequency representation of processed iPPG signals [27]. Chen et al. proposed DeepPhys [28] and DeepMag [29], deep CNN trained to respectively predict pulse wave and magnify color variations produced by the periodic changes in blood flow. Inputs are transformed using a skin reflection model while the convolutional layers are guided using attention masks to ensure the robust estimation of PPG signals under lighting fluctuation and motion. They used a modified version of VGG, a model dedicated to object recognition in images [30].

End-to-end strategies were recently investigated through different neural architectures: CNN-based extractor and estimator [31], 3D CNN [8, 32, 33], combination of CNN and long short-term memory [32, 34], CNN and gated recurrent unit [35], Siamese network including two branches with identical structure that analyze two different facial regions [36] and temporal difference

convolution [6]. These models have been trained with synthetic data [8] reinforced by real data [9, 33]. They estimate the pulse signal [32] or directly the pulse rate from a sequence of images.

Few studies investigated the interpretability and behavior of the models to understand the representations learned by the features. Zhan et al. studied this aspect by analyzing that CNN properly learn PPG during training [7]. They conclude that color variations produced by blood flow fluctuations are correctly exploited by the neural networks.

2.2. Blood pressure assessment from iPPG

Both systolic and diastolic blood pressures (BP) have been estimated using the propagation time of pulse waves from two different skin areas (typically hand and face) in video recordings [37, 38]. The positional of the two skin areas must be maintained during the measurement. This approach is therefore very restrictive. The scientific literature covers few studies dedicated to the estimation of BP from a single facial region [39, 12, 40, 41]. To the best of our knowledge, only the seminal work from Luo et al. [11] presents a pipeline that includes an artificial intelligence model. They feed a multilayer perceptron with 155 features (reduced to 30 after principal component analysis) computed from iPPG waves. Their results show that PPG waveform extracted from video exhibits information that relates to BP. All these studies pointed out the feasibility of remote BP monitoring from facial video but showed that there is still room for improvements.

2.3. Blood pressure assessment from cPPG

Based on the current literature, there is clear evidence that the fluctuations in BP are reflected in cPPG signals [42, 43] even if estimating absolute BP values from cPPG remains a challenging problem. The changes in morphological contours due to interaction of other physiological systems make the extraction of features, and thus the estimation of BP, challenging but achievable [44]. Exploration of deep learning techniques is here particularly interesting because it allows overriding of handcrafted features [45]. These features are somewhat restricted because the cPPG waveform fluctuates from subject to subject and also because the filtering procedure can change its morphology [46].

Several recent studies show that deep learning frameworks can effectively be deployed to translate BP from cPPG signals. Tanveer and Hasan proposed to associate artificial neural network (ANN) with long short-term memory

for BP estimation [47]. A similar network structure was proposed by Panwar et al. in 2020 [48]. 1D CNN replace the ANN part from Tanveer and Hasan architecture. The network concurrently estimates diastolic BP, systolic BP and heart rate from a single cPPG signal. Chowdhury et al. then proposed to employ machine learning algorithms dedicated to BP estimation using cPPG signal and demographic features (e.g. weight and height) [49]. Time, frequency and time-frequency features were extracted 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. Slapnicar et al. introduced a similar framework but with a deep neural network architecture with residual connections [50]. A part of the network is dedicated to the analysis of the signal spectral representation using gated recurrent units. Ibtehaz et Raman employed a deep learning based method that manages to predict the continuous BP waveform from cPPG signals. 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 U-Net [51].

3. Methods

3.1. Database and experimental protocol

The data used to learn the neural models (section 3.3) have been presented in a previously published article [14]. 12 volunteers aged between 20 and 35 years participated to the study. The experiments were conducted in a dark room where the only source of light was two Neewer LED panels (NL480) set to 2700 lux / m with a color temperature of 3750 K (neutral white light). During the experiments, they were asked to seat at approximately 1 meter from a fast camera (16mm C Series Lens mounted on a EO-2223C Color camera from Edmund Optics). The recorded sequences of RGB images were save without compression at resolution 640 $\times$ 480 pixels (24 bits per pixel) and with a frame rate of 125 frames per second. Autoexposure and white balance have been disabled.

The ground truth cPPG signals were recorded using approved contact probes (BVP-Flex / Pro. By Thought Technologies Ltd.) placed on the finger and the ear. Two 60-second videos were recorded for every participant. First video: participants were asked to stay calm and breathe normally. Second video: participants were asked to hold their breath as much as possible, the

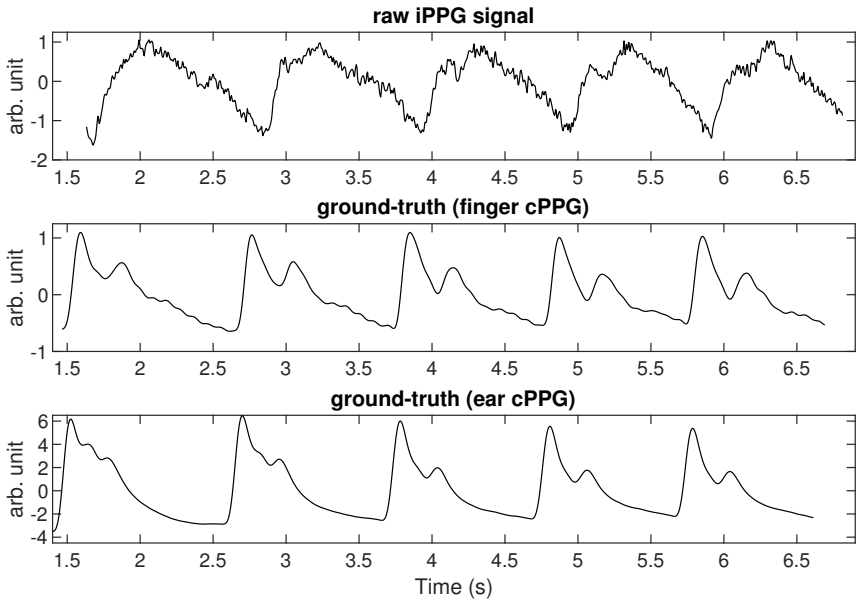


Figure 2: Excerpts of participant #1 (collected during breath holding experiment). Top figure: raw iPPG signals computed with a spatial averaging operation over the forehead region [19]. Video recordings have been collected using a fast camera (125 frames per second). Reference cPPG signals have been recorded with contact probes placed on the finger (middle figure) and the ear (bottom figure).

objective being to cause physiological variations that modify blood pressure and impact the recorded PPG signals. We refer the reader to the original publication for more details concerning the procedure and the material used [14].

The database contains 724 signals. Each of them contains 5 PPG waves (more details in section 3.2) defined over 256 values. About 80% of the data (600 signals) were reserved for training and 20% (124 signals) for testing. The sets contain a balanced portfolio of the different participants and tasks. We evaluated the models relevance through k-fold cross-validation (k=5). A fold contains 120 signals that are reserved for validation. The 4 remaining folds include 480 signals that are employed for training the neural models.

3.2. Image and signal processing

The forehead corresponds to a relevant area of interest in terms of signal-to-noise ratio [17]. This region has been automatically detected with a model

composed of 68 points positioned on the main shapes of the face [52]. These different points are tracked along the video. Some of them are used to find the position of the forehead. In practice, algorithms for face and facial landmarks detection included in OpenCV [1] and Dlib [2] libraries have been employed.

iPPG signals are computed by averaging all the forehead pixels from the green channel. This technique has been used since the very first publications related to the measurement of contactless PPG signals by camera [19]. The raw iPPG signals are then detrended using a specific low-pass filter [53] based on a smoothness priors that attenuates low frequencies [20]. We then robustly detect the valleys to extract each PPG signal wave. Each signal is ultimately sampled over 256 points and contains 5 successive iPPG waves. An excerpt is presented in figure 2. The ground truth cPPG signals measured at the finger and the ear are also presented in this figure. All the signals have been standardize ($\mu = 0$ and $\sigma = 1$).

In this article, we propose to exploit the wavelet representation of PPG signals to train the different neural architectures presented in section 3.3 (figure 1). The continuous wavelet transform (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 [54].

[1] https://opencv.org/
[2] http://dlib.net/

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 PPG signal in the frequency range [0.6, 4.5] Hz, which corresponds to the physiological range of the human heart rate [4]. Wavelet representations of dimension 256×256 will be used to train the neural architectures presented in section 3.3.

Typical iPPG signal, cPPG signal and their respective wavelet representations (real, imaginary and absolute part) are presented in figure 3. A typical difference in shape between both signals and in phase between their wavelet representations can be noted: the real part of the iPPG signal starts with a series of low intensity coefficients (blue pseudo-ellipse) while the real part of the cPPG signal starts with strong intensity coefficients (yellow pseudo-

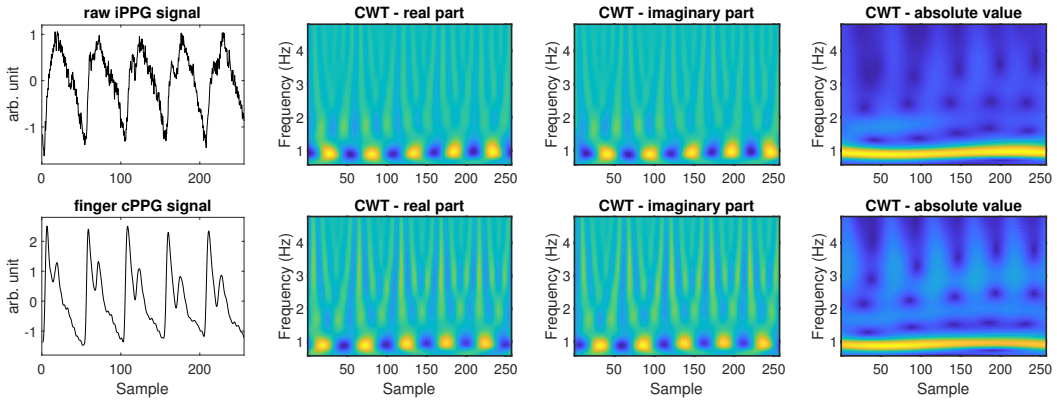


Figure 3: The continuous wavelet transform of the iPPG signal (top figure) and cPPG signal (ear or finger, see bottom figure for a finger cPPG signal) is computed in the frequency range [0.6, 4.5] Hz. The wavelet representation of the iPPG signal (a complex image with a real and imaginary part) serves as input for training the neural networks presented in section 3.3. The absolute of the continuous wavelet transform is depicted for information and is not learned by the model.

ellipse). The neural network will learn this specificity during the training phase.

3.3. Neural architectures

The U-Net neural architecture was initially proposed by Ronneberger et al. [51]. This network has been used for segmentation of medical images [55]. 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 deconvolution layers connected to the convolutions of the descending branch. Connections help to restore the spatial information. A schematic representation of the network is given in figure 1. In this study, we employ the U-Net1 version proposed by Leclerc et al. [55]. The model hyperparameters vary slightly compared to the original version proposed by Ronneberger et al. Details are presented in table 1. The number of filters is given for the first and for the last convolutional block as well as at the center of the network, where the spatial information is most compressed. Each convolutional layer integrates a core $(3, 3)$ coupled to a Rectified Linear Unit (ReLU) activation function.

A Backbone (e.g. VGG16) can be integrated into the encoder part of the U-Net network (figure 4). Its internal parameter are blocked during training (the weights of the network remain fixed). In practice, a backbone correspond to a model subpart pre-trained on ImageNet, a database deployed for object

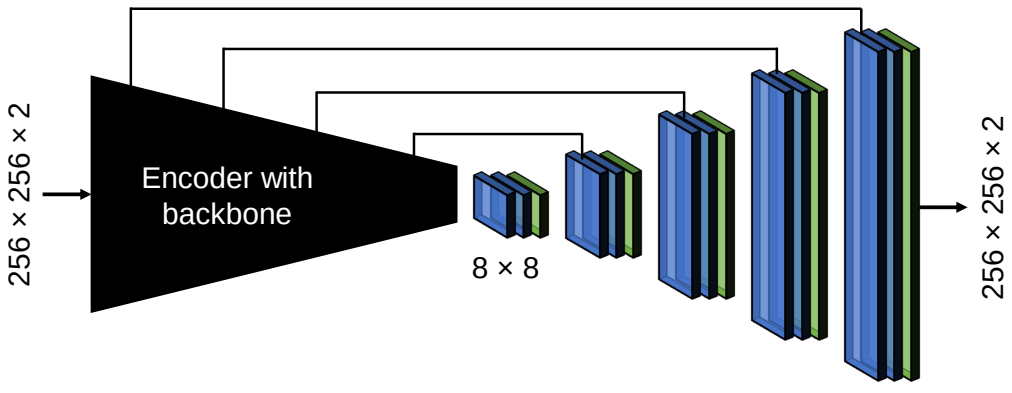


Figure 4: A backbone corresponds to a pre-trained network included in the encoder part of U-Net.

recognition tasks in images [61]. Training a U-Net network supported by a backbone consists in optimizing the internal parameters of the decoder part. This approach can be associated to a transfer learning strategy.

The various backbones tested and their main characteristics are summarized in table 1. VGG [30] is a model composed of $(3, 3)$ convolutional layers and pooling layers. The 16-layer version (VGG16) was used in this study. ResNet [56] are neural modules nested in a larger network (network-in-network) through residual units composed of convolutional filters. The architecture is about 8 times deeper than VGG. ResNet models at different depth levels (18, 34, 50, 101 and 152 layers) were trained on the ImageNet database but only the 101 layers was used in this study. DenseNet networks [60] include Dense blocks that are densely connected together: each layer is directly connected with the following ones. Thus, the input vector of a given layer integrates all the characteristics of those that precede it. The 201-layer version was chosen. Inception networks [59] contain modules composed of convolution and pooling layers of different sizes. The InceptionV3 and InceptionResNetV2 versions (with residual connections) were used in this work.

Conventional regularization techniques (e.g. dropout) have not been introduced while a normalization scheme (i.e. batch normalization) is used in networks having a backbone. These details are summarized in table 1. No output 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. The number of variables to be trained

Network	**Number of conv. filters**	**Lowest resolution**	**Normalization**	**Number of parameters**
U-Net1 [55]	$32 \downarrow 128 \uparrow 16$	8×8	$\varnothing$	2M
U-Net$_{\text{VGG16}}$ [30]	$64 \downarrow 512 \uparrow 16$	8×8	BatchNorm	9M
U-Net$_{\text{VGG19}}$ [30]	$64 \downarrow 512 \uparrow 16$	8×8	BatchNorm	9M
U-Net$_{\text{ResNet101}}$ [56]	$64 \downarrow 2048 \uparrow 16$	8×8	BatchNorm	9M
U-Net$_{\text{ResNeXt101}}$ [57]	$64 \downarrow 2048 \uparrow 16$	8×8	BatchNorm	9M
U-Net$_{\text{SE-ResNet101}}$ [58]	$64 \downarrow 2048 \uparrow 16$	8×8	BatchNorm	9M
U-Net$_{\text{SE-ResNeXt101}}$ [58]	$64 \downarrow 2048 \uparrow 16$	8×8	BatchNorm	9M
U-Net$_{\text{InceptionResNetV2}}$ [59]	$32 \downarrow 2080 \uparrow 16$	8×8	BatchNorm	7.5M
U-Net$_{\text{InceptionV3}}$ [59]	$32 \downarrow 448 \uparrow 16$	8×8	BatchNorm	8M
U-Net$_{\text{DenseNet201}}$ [60]	$64 \downarrow 128 \uparrow 16$	8×8	BatchNorm	8.5M

Table 1: Main properties of the U-Net networks used in this study.

(weights and biases) is comprised between 2 and 9 million (table 1).

The input dimensions of networks with backbones are fixed by the data used for their training (256×256 pixels RGB images from the ImageNet database). The inputs being in our case two-channels wavelet representations, it is necessary to introduce an adaptation strategy. An additional 2D convolutional layer with a $(1, 1)$ kernel has therefore 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 [62]. Biases are initialized to zero. The Mean Squared Error (MSE) has been selected as loss for training all the models:

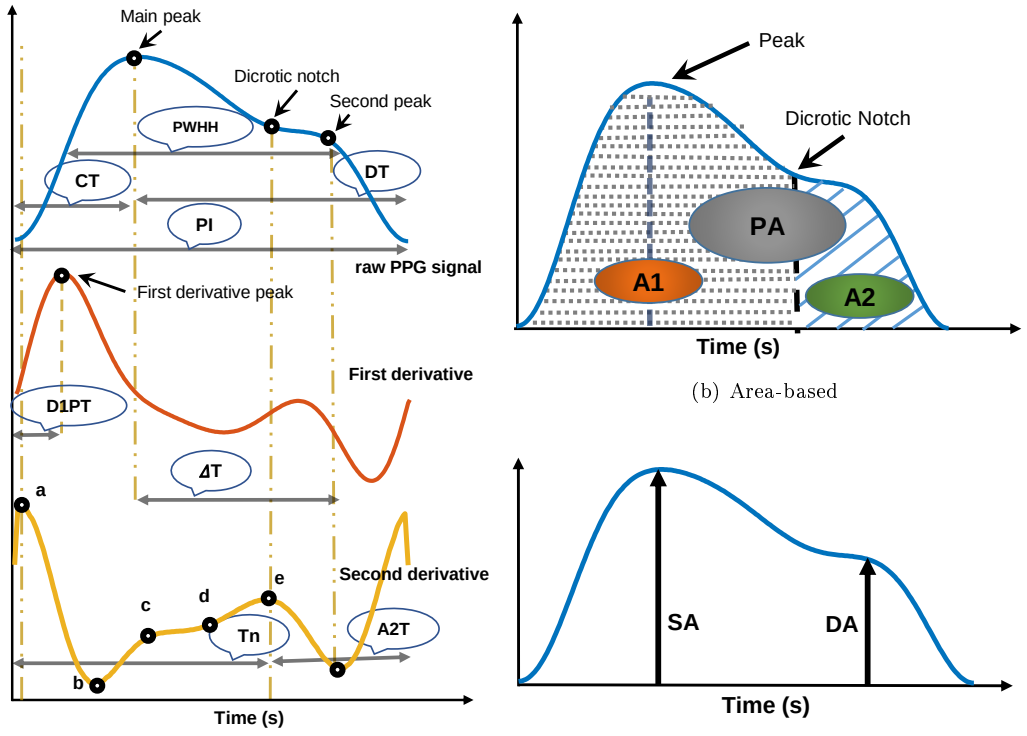
$$MSE = \frac{1}{n} \sum_{i,j} \left(CWT_{i,j} - \widehat{CWT}_{i,j} \right)^2 \tag{5}$$

CWT corresponds to the wavelet transform (see section 3.2) of the ground truth cPPG 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 [63] proposed by P. Yakubovskiy was used to develop the neural networks presented in table 1. The training sessions were launched over 5000 epochs through batches of 16 images. We used, in this study, the Adam optimization algorithm [64] with a learning rate of 0.0001. A dedicated computer equipped with a dual Intel Xeon Silver 4114 and two Nvidia Quadro P6000s was used to carry out network learning.

3.4. Waveform estimators

Different features have been proposed to characterize the waveform of a PPG signal [42]. In order to validate the predictions of the neural architectures presented in the previous section, we propose to compare the estimates of the most commonly observed waveform features [42] [43] between the reconstructed PPG signal (computed using the inverse transform of the predicted wavelet representation) and the ground truth cPPG signal. It has recently been shown that some of these features can properly be estimated on iPPG signals [14], the contact and contactless waveform features evolving in a same way.



(a) Temporal and derivative-based

(b) Area-based

(c) Amplitude-based

Figure 5: Presentation of the features computed from a PPG wave. These parameters have been categorized in four groups. Temporal: Pulse Interval (PI), Crest Time (CT), Diastolic Time (DT), time between the main peak and the secondary peak (ΔT), Dicrotic Notch Time (Tn), Pulse Width at Half Height (PWHH), time between the dicrotic notch and the end of the wave (A2T) and First Derivative Peak Time (D1PT). Derivatives: a, b, c, d and e correspond to specific points that are detected on the second derivative. Area: Pulse Area (PA) and area computed between the start of the wave and the inflection point (A1) and between the inflection point and the end of the wave (A2). Amplitude: Systolic Amplitude (SA) and Diastolic Amplitude (DA).

Waveform features can be categorized into 4 families: temporal, amplitude-based, area-based, and (first and second) derivative-based. All features are presented in figure 5. We refer the reader to the article of Elgendi et al. [42] which details the PPG waveform features and their physiological interpretation.

3.4.1. Temporal features

The Pulse Interval (PI) corresponds to the total time of the wave, which is measured between two successive valleys. This feature is used to estimate the pulse rate. The Crest Time (CT) corresponds to the time between the start (first valley) and the main peak of the wave. The Diastolic Time (DT) corresponds to the time between the main peak and the end of the wave. ΔT corresponds to the time between the main peak and the secondary peak. Dicrotic Notch Time (Tn) is the time between the start of the wave and the dicrotic notch. A2T corresponds to the time between the dicrotic notch and the end of the wave. Pulse Width at Half Height (PWHH) is the time equal to the width of the wave at half height. The First Derivative Peak Time (D1PT) parameter corresponds to the time between the start of the wave and its first derivative peak.

3.4.2. Features based on first and second derivatives

The points a, b, c, d and e (figure 5a) are detected on the second derivative of the PPG signal. These points reflect the wave inflections. They are used to compute all the ratios presented in figures 9, 10 and 11. These ratios change with age and reflect arterial stiffness [43].

3.4.3. Area-based features

The area-based features are shown in figure 5b. The Pulse Area (PA) parameter corresponds to the total area of the PPG wave. Area 1 (A1) is computed between the start of the wave and the inflection point (systolic phase). Area 2 (A2) is computed between the inflection point and the end of the wave (diastolic phase). The Inflection Point Area ratio (IPA) corresponds to the ratio between A2 and A1.

3.4.4. Amplitude-based features

The systolic (SA) and diastolic (DA) amplitudes are calculated from the main and the secondary peaks (figure 5c). The Reflection Index (RI) is the ratio between DA and SA while the Augmentation Index (AI) is the difference between SA and DA divided by SA.

3.5. Metrics

In this section, we detail the different metrics employed for evaluating the performances of the models. The Root Mean Squared Error ($RMSE$, equation 6) has been computed between the PPG traces obtained after inverse

wavelet transform (equation 3). Because the amplitudes are arbitrary and normalized, we also propose the Mean Absolute Percentage Error ($MAPE$, see equation 7). Both metrics along with scatter plots and Pearson correlation coefficients have been used to quantify the level of agreement between the predicted ($\widehat{PPG}$) and the ground truth signals (PPG).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i}\left(\widehat{PPG_i} - PPG_i\right)^2} \tag{6}$$

$$MAPE = \frac{1}{n}\sum_{i}\left|\frac{\widehat{PPG_i} - PPG_i}{PPG_i}\right| \tag{7}$$

4. Results and discussion

4.1. Learning performance

k-fold cross-validation results for each model are presented in table 2. The MSE correspond to the minimum validation loss (equation 5) observed during training. Each value presented in the table corresponds to the average and standard deviation computed for a specific U-Net network from the lowest MSE of each fold.

Network	MSE_{finger}	MSE_{ear}
U-Net1	0.382 ± 0.054	0.266 ± 0.024
U-Net$_{\text{VGG16}}$	0.319 ± 0.029	0.224 ± 0.032
U-Net$_{\text{VGG19}}$	0.322 ± 0.033	0.232 ± 0.031
U-Net$_{\text{ResNet101}}$	0.341 ± 0.037	0.244 ± 0.022
U-Net$_{\text{ResNeXt101}}$	**0.316 ± 0.036**	**0.222 ± 0.022**
U-Net$_{\text{SE-ResNet101}}$	0.367 ± 0.031	0.249 ± 0.021
U-Net$_{\text{SE-ResNeXt101}}$	0.368 ± 0.042	0.259 ± 0.024
U-Net$_{\text{InceptionResNetV2}}$	0.385 ± 0.041	0.268 ± 0.030
U-Net$_{\text{InceptionV3}}$	0.386 ± 0.036	0.271 ± 0.026
U-Net$_{\text{DenseNet201}}$	0.317 ± 0.036	0.234 ± 0.027

Table 2: k-fold cross-validation results for each model presented in table 2. The MSE (see equation 5) is computed between predicted and ground truth CWT transforms (real and imaginary parts). U-Net1 corresponds to the neural network proposed by Leclerc et al. [55], which does not include a pre-trained backbone. All the other neural networks are U-shaped architectures supported by a backbone.

Independently of the measurement site, the network supported by ResNeXt101 presents the lowest MSE, thus indicating the best performance in terms of wavelet transform reconstruction (real and imaginary parts). We note that performances of architectures supported by VGG16 and DenseNet101 are close from ResNeXt101. Backbones based on ResNet and ResNeXt structure with squeeze and excitation are less efficient. U-Net1 presents higher MSE values than the other models. This observation probably reflects the fact that the network contains between 4 to 5 times less trainable parameters. Models supported by a backbone performed generally better. This translates a real impact of pre-trained convolutional layers on very large databases. As a reminder, the backbone layers are blocked during the training phase. Inception-based backbones also present degraded performances.

Regarding the two sites, ear measurements deliver better general performances (lower MSE) than finger measurements. We assume that this gap

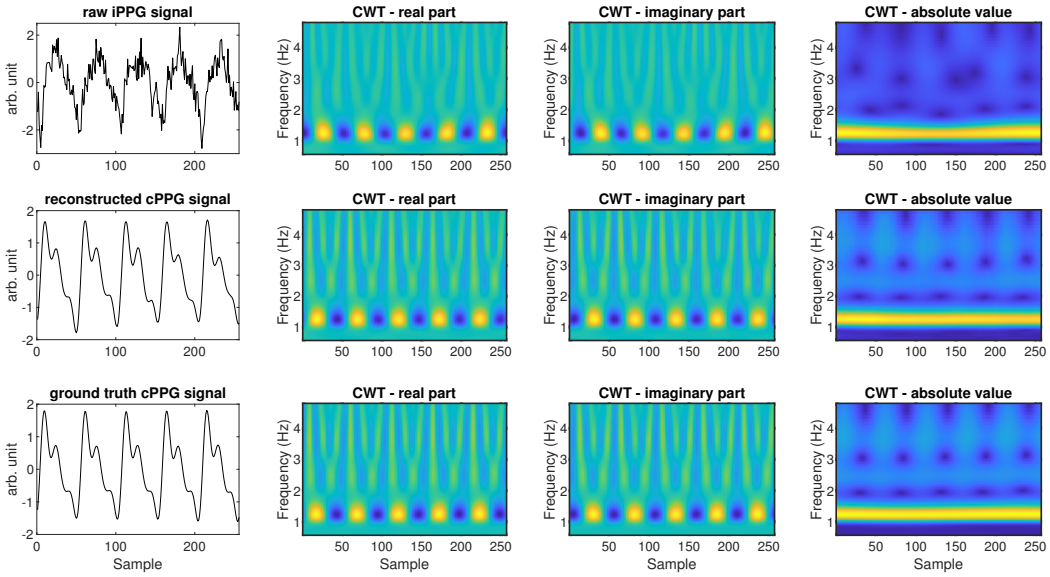


Figure 6: The real and imaginary parts of the reconstructed wavelet transform by the U-Net network supported by ResNeXt101 (middle figures) are similar to those computed from the finger ground truth signal (bottom figures). We can notice a small phase difference in the wavelet representations of the raw iPPG signal (top figures) and the ground truth cPPG signal (bottom figures). The neural network learned this specificity, the reconstructed wavelet transform being in phase with the ground truth one. The absolute representations are depicted for information.

reflects the differences between signal waveform: a PPG signal measured at the forehead surface is generally closer to a PPG signal measured at the ear than measured at the finger [65].

4.2. Point-to-point validation of reconstructed PPG signals

This section is dedicated to the evaluation of PPG signals produced by the neural architectures presented in table 1.

The trained neural models deliver a two-channel wavelet representation (a real part and an imaginary part). The temporal PPG signal is then reconstructed from the inverse transform (equation 3). An example is presented in figure 6, where we can appreciate the prediction quality of the real and imaginary parts of the wavelet transform produced by the U-Net$_{\text{ResNeXt101}}$ network. The phase has been properly recovered. We can also observe that the dicrotic notch is well reproduced whereas it was almost absent on the raw iPPG signal. The reconstructed PPG signal is smooth and its width is smaller. This shows that the network properly corrects the high frequency

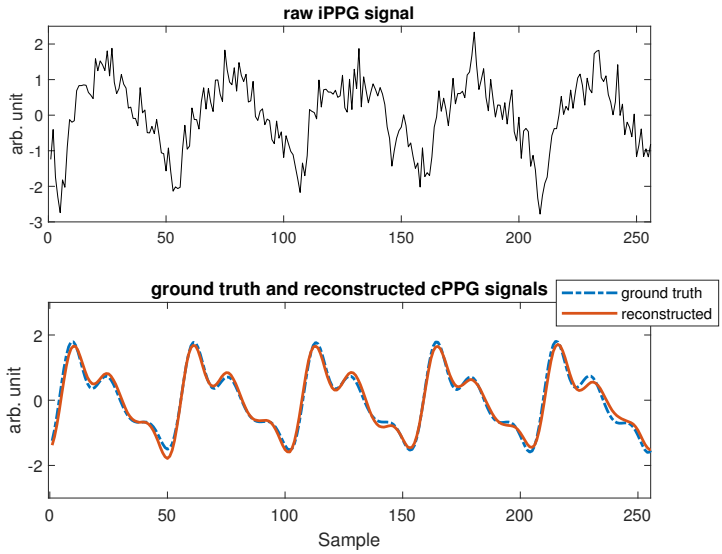


Figure 7: PPG signal prediction (bottom figure) from an iPPG signal (top figure). U-Net supported by ResNeXt101 and trained on finger cPPG signals produced wavelet coefficients that gave, after inverse transform, the reconstructed PPG signal. Ground truth and reconstructed signals are quite similar even if small discrepancies can be noticed.

coefficients, which transcribe the noise, as well as the central frequency coefficients, which determine the pulse signal.

In order to better appreciate the quality of the reconstruction, we present, in figure 7, a superposition of a reference finger cPPG signal and the PPG signal predicted by the U-Net network supported by ResNeXt101 (after computation of the inverse wavelet transform). The $RMSE$ and $MAPE$ have been computed between the two signals (equations 6 and 7). The results after cross-validation on k-fold are presented in table 3. The predictions delivered by the neural models present good overall performance.

The error on the U-Net network supported by ResNeXt101 is slightly lower, which is consistent with the results presented in section 4.1 and table 2. This particular network was therefore selected for further analysis. Table 4 presents the same results but across the test set. Additional comparisons, in particular raw iPPG against ground truth cPPG signals, are presented for information. The errors are here much more important, the $MAPE$ being higher than 50%. The last row of the table is given for comparison and indicates the error between the cPPG signals recorded on the two measurement sites.

Network	$\mathbf{cPPG_{finger}}$ ***vs*** $\widehat{\mathbf{cPPG}}_{\mathbf{finger}}$		$\mathbf{cPPG_{ear}}$ ***vs*** $\widehat{\mathbf{cPPG}}_{\mathbf{ear}}$	
	$RMSE$	$MAPE$	$RMSE$	$MAPE$
U-Net1	0.260 ± 0.018	0.064 ± 0.010	0.210 ± 0.010	0.033 ± 0.007
$\text{U-Net}_{\text{VGG16}}$	0.245 ± 0.013	0.053 ± 0.014	0.196 ± 0.014	0.031 ± 0.010
$\text{U-Net}_{\text{VGG19}}$	0.248 ± 0.011	0.055 ± 0.012	0.197 ± 0.014	0.034 ± 0.006
$\text{U-Net}_{\text{ResNet101}}$	0.251 ± 0.013	0.058 ± 0.009	0.205 ± 0.010	0.032 ± 0.008
$\mathbf{U\text{-}Net_{ResNeXt101}}$	**0.244 ± 0.014**	**0.045 ± 0.008**	**0.196 ± 0.009**	**0.032 ± 0.009**
$\text{U-Net}_{\text{SE-ResNet101}}$	0.260 ± 0.010	0.058 ± 0.008	0.207 ± 0.012	0.032 ± 0.005
$\text{U-Net}_{\text{SE-ResNeXt101}}$	0.261 ± 0.014	0.060 ± 0.003	0.211 ± 0.012	0.037 ± 0.009
$\text{U-Net}_{\text{InceptionResNetV2}}$	0.265 ± 0.012	0.063 ± 0.008	0.213 ± 0.013	0.038 ± 0.007
$\text{U-Net}_{\text{InceptionV3}}$	0.266 ± 0.011	0.061 ± 0.010	0.213 ± 0.011	0.032 ± 0.004
$\text{U-Net}_{\text{DenseNet101}}$	0.245 ± 0.012	0.052 ± 0.007	0.201 ± 0.012	0.033 ± 0.004

Table 3: k-fold cross-validation for $RMSE$ and $MAPE$ (see equations 6 and 7) computed between reconstructed PPG signals and ground truth cPPG signals. $\text{cPPG}_{\text{finger}}$ and cPPG_{ear} correspond to ground truth cPPG signals measured at finger and ear respectively (see signal depicted in blue in figure 7 for a typical example). $\widehat{\text{cPPG}}_{\text{finger}}$ and $\widehat{\text{cPPG}}_{\text{ear}}$ correspond to reconstructed PPG signals computed by inverse transform on the CWT predicted by the different neural architectures (see signal depicted in orange in figure 7 for a typical example).

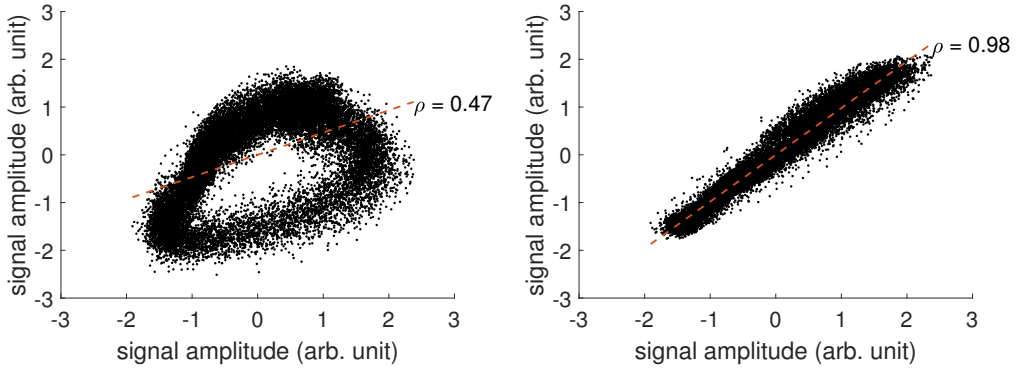
Comparison	$RMSE$	$MAPE$	ρ
$\text{cPPG}_{\text{finger}}$ *vs* $\widehat{\text{cPPG}}_{\text{finger}}$	0.219	0.045	0.97
cPPG_{ear} *vs* $\widehat{\text{cPPG}}_{\text{ear}}$	0.185	0.0187	0.98
$\text{cPPG}_{\text{finger}}$ *vs* iPPG	0.985	0.534	0.47
cPPG_{ear} *vs* iPPG	0.994	0.543	0.46
$\text{cPPG}_{\text{finger}}$ *vs* cPPG_{ear}	0.198	0.020	0.98

Table 4: $RMSE$, $MAPE$ and Pearson correlation (ρ) computed across samples included in the test set for ground truth cPPG signals, predicted cPPG signals and raw iPPG signals. An illustration of an iPPG signal is presented in black in figure 7. Predicted signals ($\widehat{\text{cPPG}}$) are produced by the selected $\text{U-Net}_{\text{ResNeXt101}}$ model (see signal depicted in orange in figure 7 for a typical example). All correlations presented p-values lower than 0.001.

Figure 8 presents scatter plots coupled with Pearson correlation coefficients. These representations aim to assess and compare the amplitudes of iPPG, ground truth cPPG and reconstructed cPPG signals over the test set. The graph representing cPPG_{ear} against iPPG signals is not presented in this figure because of its close similarity with the graph presented in figure 8a. The concentric shape of the points distribution reflects the natural waveform difference between raw iPPG signals and cPPG signals. This specificity is mainly due to the dicrotic notch which is generally prominent on cPPG signals and, in contrast, not perceptible on iPPG signals (see figure 2 for a typical example). The inherent pulse width difference between cPPG and iPPG signals also impacts the scatter plot representation presented in figure 8a. Figure 8b depicts finger and ear cPPG measurements and is provided for information.

Figures 8c and 8d illustrate the quality of cPPG signal reconstruction by the $\text{U-Net}_{\text{ResNeXt101}}$ network on the test set. The Pearson correlations coupled with the statistical results presented in table 4 (in particular the low $MAPE$) show that the PPG waveform is suitably reconstructed through its wavelet representation. This conclusion is valid for both finger (figure 8c) and ear (figure 8d) cPPG signals.

We propose, in the next subsection, an in-depth analysis of these results by studding pulse waveform features, whose values are originally very different between iPPG and cPPG signals.



(a) $\text{cPPG}_{\text{finger}}$ *vs* iPPG

(b) $\text{cPPG}_{\text{finger}}$ *vs* cPPG_{ear}

(c) $\text{cPPG}_{\text{finger}}$ *vs* $\widehat{\text{cPPG}}_{\text{finger}}$

(d) cPPG_{ear} *vs* $\widehat{\text{cPPG}}_{\text{ear}}$

Figure 8: Scatter plots along with their respective Pearson correlation (ρ). All the p-values are lower than 0.001. The concentric shape observed in figure (a) reflects the natural waveform difference between raw iPPG signals and cPPG signals. Figure (b) depicts finger and ear cPPG measurements. Bottom row figures present the cPPG signals reconstructed by the $\text{U-Net}_{\text{ResNeXt101}}$ network for both finger (c) and ear (d) measurement sites.

4.3. Waveform features

The point-to-point evaluation presented in the previous subsection provides an overall vision of the predictions quality made by the neural architecture presented in table 1. Here, we propose an evaluation of the reconstructed PPG waves through specific waveform features across the test set. The studied features have briefly been presented in section 3.4. They are divided into 4 categories: temporal, area-based, amplitude-based and based on first and second derivatives.

Scatter plots along with their correlation coefficients are presented for each feature in figures 9 and 10. We focus this specific evaluation on the U-Net$_{\text{ResNeXt101}}$ network. A good general performance on each feature can be observed on each subfigure, showing that the neural network (that take as input CWT of iPPG waves) reliably recovered the shape of finger and ear cPPG waves. As a reminder, iPPG signals computed from video on the forehead region are quite noisy, include artifacts and present a signature that is very different from cPPG signals measured on other sites [65] (see figure 2).

Several temporal features like PI (total width of the pulse wave) show high correlations. PI directly reflects the pulse rate, a parameter estimated from iPPG signals with reliability and precision. Crest time (CT) presents better correlation than DT (diastole time), which seems to be in accordance with studies focusing on arterial pressure estimation based on PPG waveform analysis [45]. In contrast, the temporal parameter ΔT exhibits low correlation. We assume that the specific points associated with the detection of ΔT, in particular the secondary peak, are less accurately recovered. Its estimation is therefore potentially less reliable. It is however interesting to note that this weak correlation is also observed in figure 11 that presents a scatter plot computed between finger cPPG and ear cPPG signals for each waveform feature.

The parameters related to the amplitudes (SA, DA, RI and AI) present more or less high scores. The arbitrary nature of the PPG signals amplitudes makes their estimation very complex. The amplitude of cPPG signals is mainly modulated by the pressure applied between the sensor and the measurement site, by the light absorption of the tissues as well as by the optical properties of the skin. The iPPG signal amplitude also depends on the emitted and reflected quantity of light, the distance as well as internal camera parameters. In general, the predictions produced from finger cPPG signals (figure 9) exhibit higher correlations for the amplitude features than for the predictions computed from ear cPPG signals (figure 10).

Waveform features related to areas and derivatives are relatively well transcribed by the neural model. The correlations presented in figures 9 and 10 are close to the correlations between finger cPPG and ear cPPG signals presented in figure 11.

Overall, the reconstructions of cPPG signals measured on the ear (figure 10) exhibit features that are slightly better correlated with the corresponding ground truth than those measured on the finger (figure 9). This conclusion

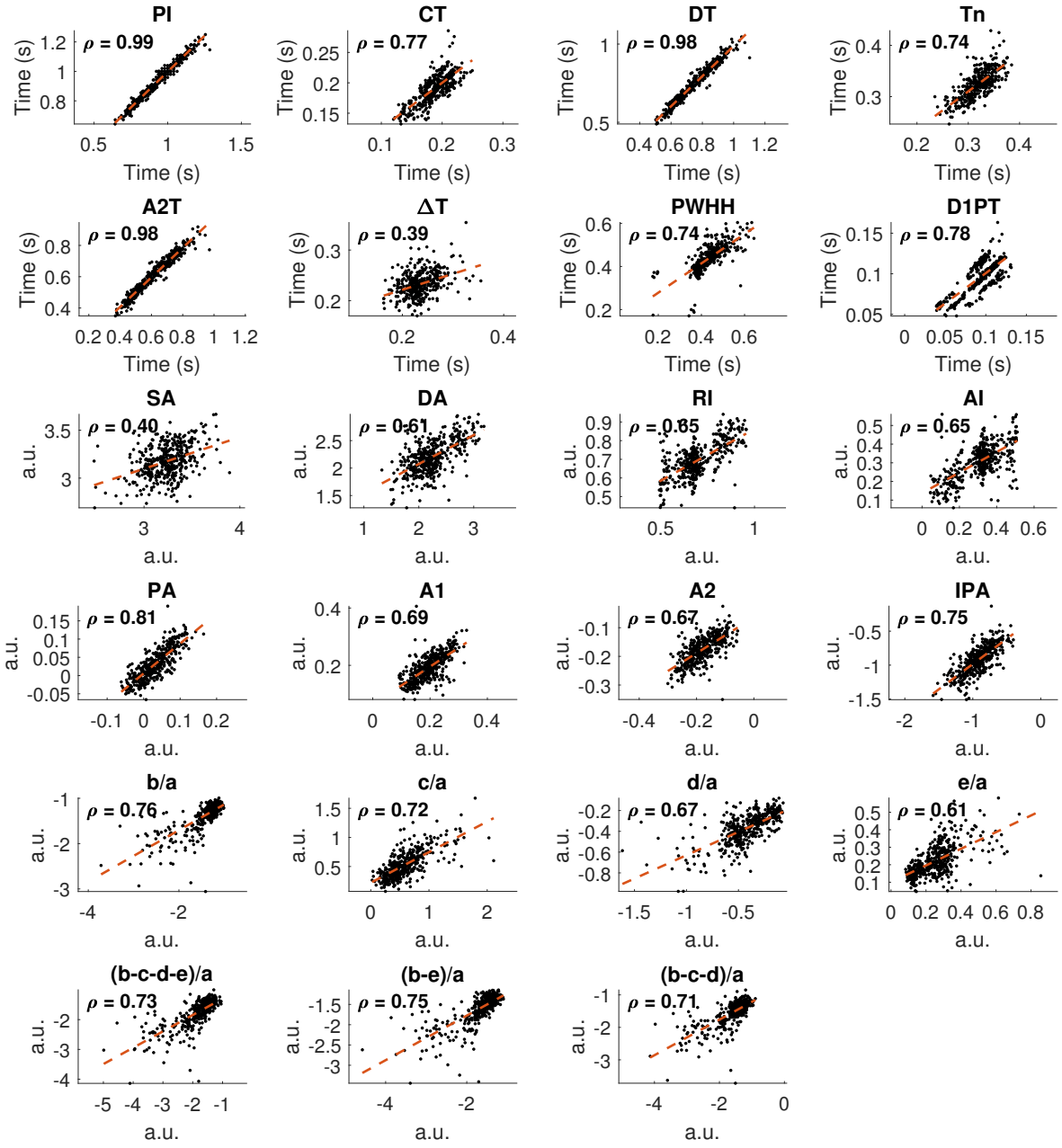


Figure 9: Scatter plots showing the different waveform features computed from ground truth finger cPPG signals ($\text{cPPG}_{\text{finger}}$, x-axis) against the waveform features computed from signals reconstructed by U-$\text{Net}_{\text{ResNeXt101}}$ network ($\widehat{\text{cPPG}}_{\text{finger}}$, y-axis). Associated Pearson correlation coefficients are presented for each feature (on each sub-figure). p-values are all lower than 0.001.

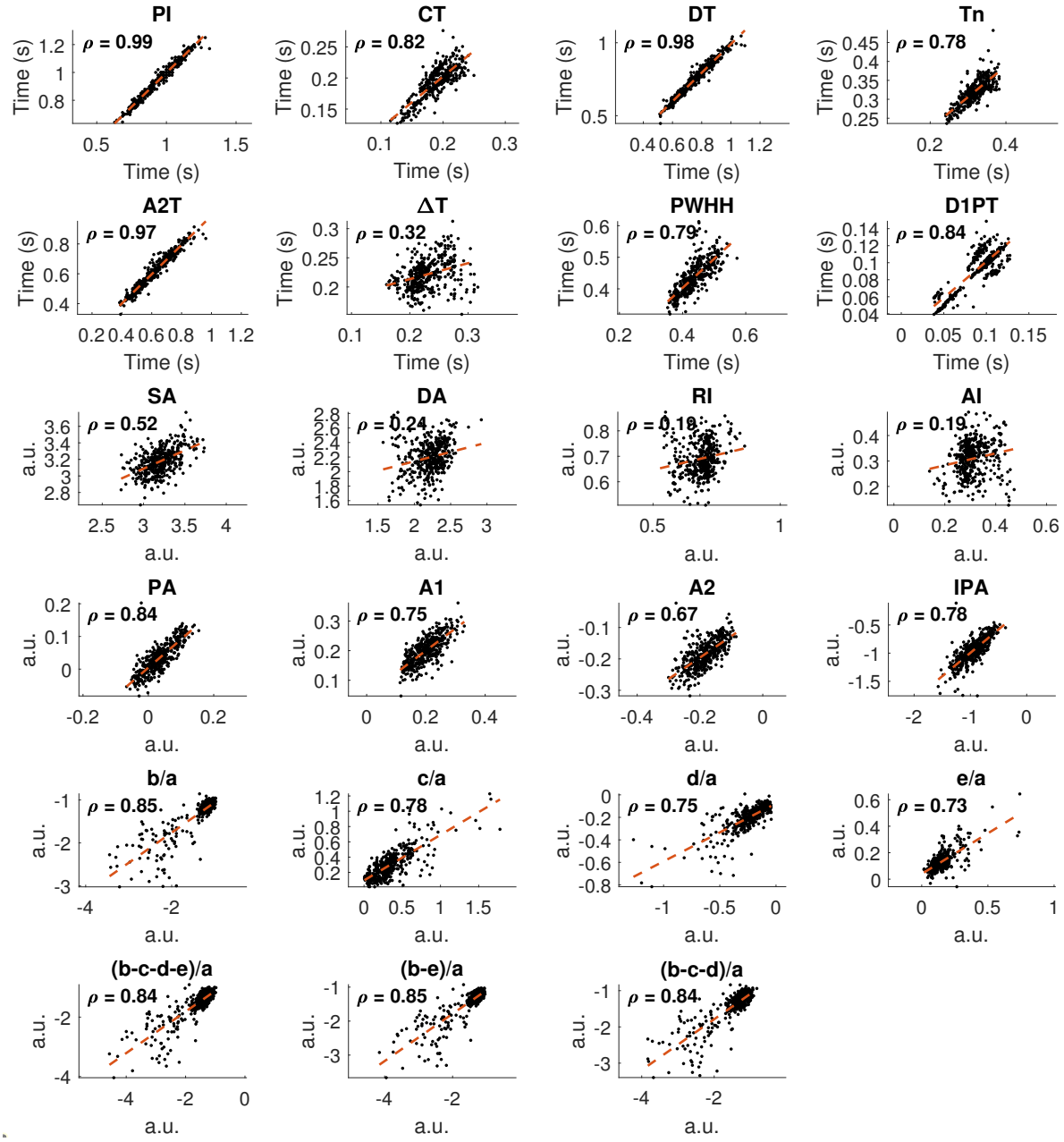


Figure 10: Scatter plots showing the different waveform features computed from ground truth ear cPPG signals (cPPG_{ear}, x-axis) against the waveform features computed from signals reconstructed by U-$\text{Net}_{\text{ResNeXt101}}$ network ($\widehat{\text{cPPG}}_{\text{ear}}$, y-axis). Associated Pearson correlation coefficients are presented for each feature (on each sub-figure). p-values are all lower than 0.001.

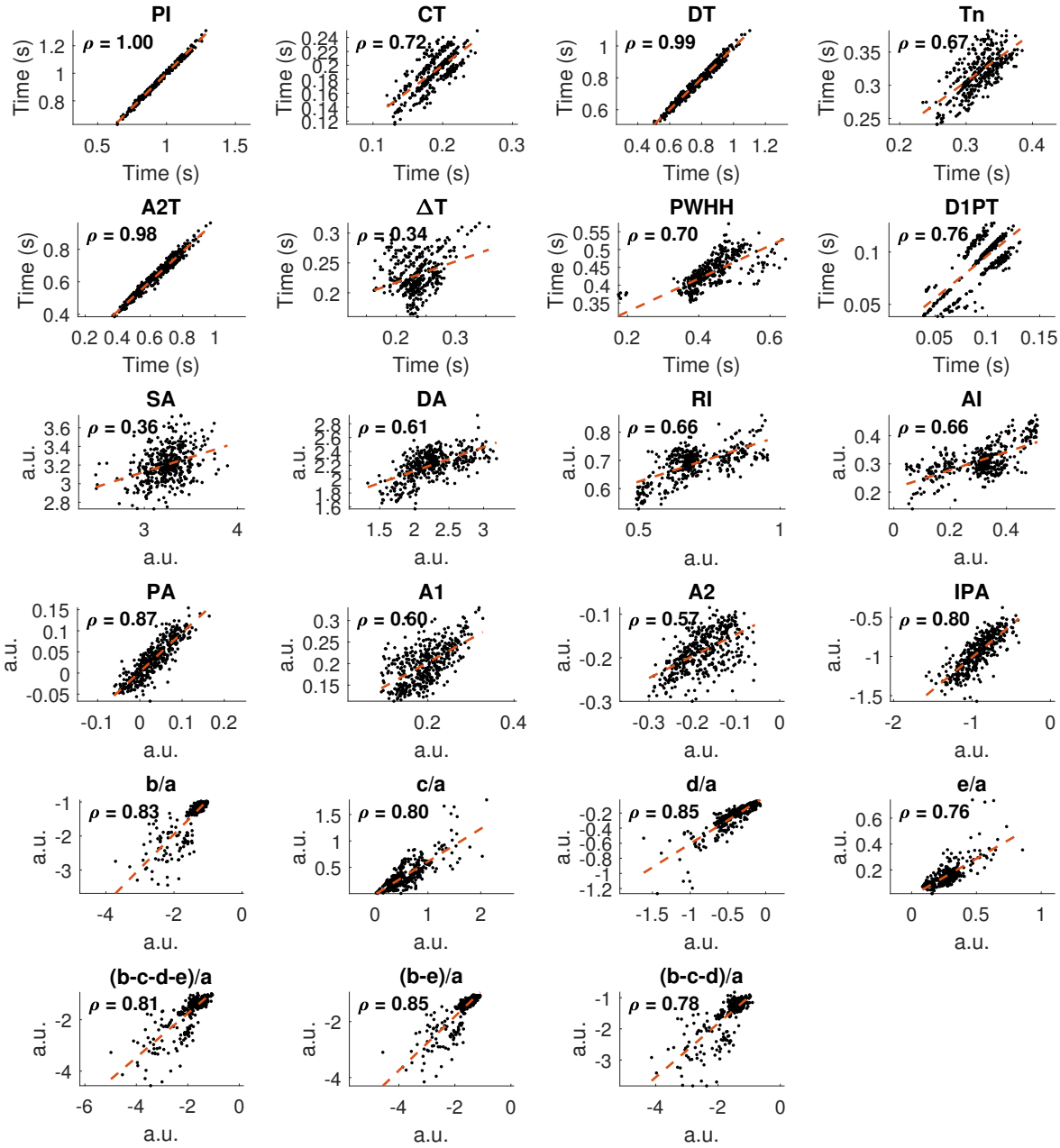


Figure 11: Scatter plots showing the different waveform features computed from ground truth finger cPPG signals ($\text{cPPG}_{\text{finger}}$, x-axis) against the waveform features computed from ground truth ear cPPG signals (cPPG_{ear}, y-axis). Associated Pearson correlation coefficients are presented for each feature (on each sub-figure). p-values are all lower than 0.001.

is in accordance with what we presented in sections 4.1 and 4.2, in particular in tables 2 and 3. We assume that this difference in performance is due to the recovering of the dicrotic notch and the secondary peak that characterize PPG signals. The notch is much more prominent on finger cPPG signals than on ear cPPG signals. It directly impacts the profile of the wave by considerably modifying the inflections and therefore the features linked to the second derivatives. The neural models trained on the wavelet representations computed from finger cPPG signals must therefore recover the coefficients describing the dicrotic notch with more difficulty because this trait is rarely apparent on raw iPPG signals. The top illustration presented in figure 12 shows a prediction of lesser quality where the successive dicrotic notches are approximately reconstructed by the model. The bottom illustration exhibits phase discrepancies. These differences do not systematically impact the shape of the waves but can create unwanted fluctuations in several temporal features, the number one factor being the pulse interval.

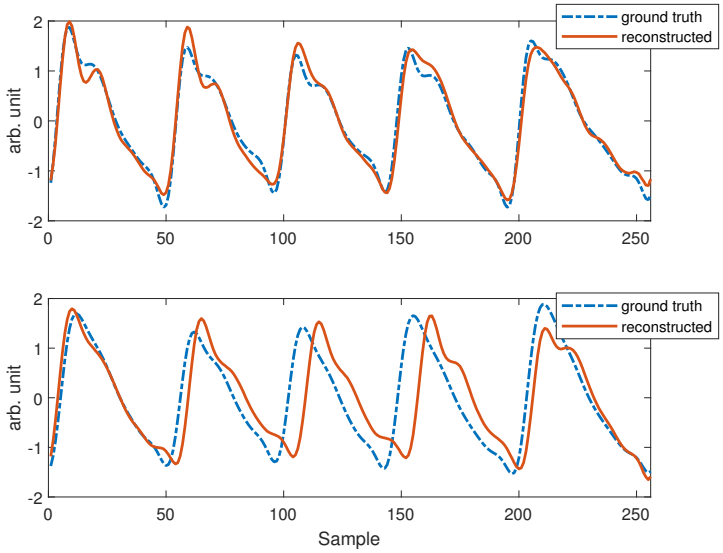


Figure 12: Predictions of lesser quality include approximate dicrotic notch reconstruction (top figure) or phase discrepancies (bottom figure). The signals presented in the two subfigures correspond to finger PPG signals.

5. Limitations and future works

The principal limitation of this study corresponds to the small number of volunteers that participated to the experiments. First validation of the concept on well-formed signals validated this choice. Thus, the videos we employed present a high frame rate which, after processing, results in highly sampled iPPG signals. These signals do not completely reflect those constituted from frames delivered by conventional cameras or webcams. In addition, participants were asked to remain still even during the breath holding experiment.

Several ways of improvement for this work are therefore considered. We first propose expanding the currently limited database by increasing the number of recordings and participants. We also envisage studying the impact of skin color, which directly affects the quality of PPG signals, on performances by assessing the evolution of waveform features against skin phototype.

Continuous wavelet transform using Morlet's wavelet has been employed in this work. 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. Modification of the internal parameters of the U-Net architectures (e.g. the number of layers and number of neurons by layer) will also be assessed. Moreover, we propose to study the impact of convolutional attention networks [28] and temporal difference convolution [6] on performances. Currently, the wavelet transform of 5 consecutive waves sampled over 256 values are inputted to the neural network. We envisage varying the number of consecutive waves but with particular consideration for small values (e.g. a single wave) that can produce inconsistencies in the time-frequency representations.

As stated at the beginning of this section, the videos used in this research were acquired by a fast (125 fps) camera. We plan to study in future work iPPG signals computed from recordings delivered by conventional (30 fps) cameras. The waves present, in this context, less details and are therefore more complex to analyze. Training models with larger volume of data can however be envisaged because many databases dedicated to the study of PPG signals measured by conventional cameras are now publicly available.

Inputting video in an U-Net architecture rather than time-frequency representation will be the subject of long-term research. We propose to test 3D U-Net architectures coupled with custom loss function that will constrain reconstruction of cPPG signals through their waveform features. This specific

loss function will be directly integrated into the training phase. The neural network will thus try to minimize an overall error regarding the shape of the pulse waves. Compliance with these criteria could thus allow high quality reconstruction of cPPG from iPPG waves.

6. Summary of contributions

We proposed, in this article, neural architectures that allow accurate recovering of cPPG signals from iPPG signals estimated in video recordings. The reconstruction is carried out using the time-frequency representation of the signals via the continuous wavelet transform. The proposed neural networks correspond to U-Net architectures supported by specific backbones. The recovered signals present waveform features close to those computed on ground truth finger and ear cPPG signals. To the best of our knowledge, this is the first demonstration of a method for accurate reconstruction of cPPG from iPPG signals.

The main motivation behind this work corresponds to the possibility of proposing an estimation of arterial blood pressure from video by analyzing iPPG signals. The next step towards this direction is therefore the integration of the recovered cPPG signals into AI models dedicated to the estimation of blood pressure using contact signals collected from large public databases [50, 48, 45].

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] A. Al-Naji, K. Gibson, S.-H. Lee, J. Chahl, Monitoring of Cardiorespiratory Signal: Principles of Remote Measurements and Review of Methods, IEEE Access (2017).

[2] M. V. Volkov, N. B. Margaryants, A. V. Potemkin, M. A. Volynsky, I. P. Gurov, O. V. Mamontov, A. A. Kamshilin, Video capillaroscopy clarifies mechanism of the photoplethysmographic waveform appearance, Scientific reports 7 (2017) 13298.

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

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

[5] 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.

[6] 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.

[7] Q. Zhan, W. Wang, G. de Haan, Analysis of CNN-based remote-PPG to understand limitations and sensitivities, arXiv preprint arXiv:1911.02736 (2019).

[8] F. Bousefsaf, A. Pruski, C. Maaoui, 3D Convolutional Neural Networks for Remote Pulse Rate Measurement and Mapping from Facial Video, Applied Sciences 9 (2019) 4364. URL: https://www.mdpi.com/2076-3417/9/20/4364. doi:10.3390/app9204364.

[9] X. Niu, H. Han, S. Shan, X. Chen, Synrhythm: Learning a deep heart rate estimator from general to specific, in: 2018 24th International Conference on Pattern Recognition (ICPR), IEEE, 2018, pp. 3580–3585.

[10] A. Moço, W. Verkruysse, Pulse oximetry based on photoplethysmography imaging with red and green light, Journal of Clinical Monitoring and Computing (2020) 1–11. Publisher: Springer.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, H. E. Stanley, Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals, Circulation 101 (2000) e215–e220.

[16] L.-M. Po, L. Feng, Y. Li, X. Xu, T. C.-H. Cheung, K.-W. Cheung, Block-based adaptive ROI for remote photoplethysmography, Multimedia Tools and Applications (2017) 1–27.

[17] F. Bousefsaf, C. Maaoui, A. Pruski, Automatic Selection of Webcam Photoplethysmographic Pixels Based on Lightness Criteria, Journal of Medical and Biological Engineering 37 (2017) 374–385.

[18] S. Bobbia, D. Luguern, Y. Benezeth, K. Nakamura, R. Gomez, J. Dubois, Real-Time Temporal Superpixels for Unsupervised Remote Photoplethysmography, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 1341–1348.

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

[20] 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.

[21] 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.

[22] 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.

[23] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, J. Dubois, Unsupervised skin tissue segmentation for remote photoplethysmography, Pattern Recognition Letters (2017).

[24] M. Soleymani, J. Lichtenauer, T. Pun, M. Pantic, A multimodal database for affect recognition and implicit tagging, IEEE Transactions on Affective Computing 3 (2012) 42–55.

[25] R. Stricker, S. Müller, H.-M. Gross, Non-contact video-based pulse rate measurement on a mobile service robot, in: Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on, IEEE, 2014, pp. 1056–1062.

[26] Y. Qiu, Y. Liu, J. Arteaga-Falconi, H. Dong, A. El Saddik, EVM-CNN: Real-time contactless heart rate estimation from facial video, IEEE Transactions on Multimedia 21 (2018) 1778–1787. Publisher: IEEE.

[27] G.-S. Hsu, A. Ambikapathi, M.-S. Chen, Deep learning with time-frequency representation for pulse estimation from facial videos, in: Biometrics (IJCB), 2017 IEEE International Joint Conference on, IEEE, 2017, pp. 383–389.

[28] W. Chen, D. McDuff, Deepphys: Video-based physiological measurement using convolutional attention networks, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 349–365.

[29] W. Chen, D. McDuff, DeepMag: Source Specific Motion Magnification Using Gradient Ascent, arXiv preprint arXiv:1808.03338 (2018).

[30] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, in: Y. Bengio, Y. LeCun (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San

Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015. URL: http://arxiv.org/abs/1409.1556.

[31] R. Špetlík, V. Franc, J. Matas, Visual Heart Rate Estimation with Convolutional Neural Network, in: The British Machine Vision Conference (BMVC), 2018.

[32] Z. Yu, X. Li, G. Zhao, Remote Photoplethysmograph Signal Measurement from Facial Videos Using Spatio-Temporal Networks, in: BMVC, 2019.

[33] O. Perepelkina, M. Artemyev, M. Churikova, M. Grinenko, HeartTrack: Convolutional Neural Network for Remote Video-Based Heart Rate Monitoring, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 288–289.

[34] E. Lee, E. Chen, C.-Y. Lee, Meta-rppg: Remote heart rate estimation using a transductive meta-learner, in: European Conference on Computer Vision, Springer, 2020, pp. 392–409.

[35] 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 (2019). Publisher: IEEE.

[36] Y.-Y. Tsou, Y.-A. Lee, C.-T. Hsu, S.-H. Chang, Siamese-rPPG network: remote photoplethysmography signal estimation from face videos, in: Proceedings of the 35th Annual ACM Symposium on Applied Computing, 2020, pp. 2066–2073.

[37] 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.

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

[39] 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.

[40] 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.

[41] 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.

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

[43] 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.

[44] S. S. Mousavi, M. Firouzmand, M. Charmi, 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 (2019) 196–206.

[45] 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).

[46] M. Elgendi, R. Fletcher, Y. Liang, N. Howard, N. H. Lovell, D. Abbott, K. Lim, R. Ward, The use of photoplethysmography for assessing hypertension, npj Digital Medicine 2 (2019). URL: `http://www.nature.com/articles/s41746-019-0136-7`. doi:`10.1038/s41746-019-0136-7`.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] V. Kazemi, J. Sullivan, One millisecond face alignment with an ensemble of regression trees, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1867–1874.

[53] M. Tarvainen, P. Ranta-aho, P. Karjalainen, An advanced detrending method with application to HRV analysis, IEEE Transactions on Biomedical Engineering 49 (2002) 172–175. URL: http://ieeexplore.ieee.org/document/979357/. doi:10.1109/10.979357.

[54] 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.

[55] 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).

[56] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[57] 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.

[58] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.

[59] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 31, 2017.

[60] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.

[61] 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.

[62] 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.

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

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

[65] V. Hartmann, H. Liu, F. Chen, Q. Qiu, S. Hughes, D. Zheng, Quantitative Comparison of Photoplethysmographic Waveform Characteristics: Effect of Measurement Site, Frontiers in physiology 10 (2019).