Style Transfer–assisted Deep Learning Method for Photoplethysmogram Denoising

Sara Maria Pagotto[1], Federico Tognoni[1], Matteo Rossi[1], Dario Bovio[2], Caterina Salito[2], Luca Mainardi[1], Pietro Cerveri[1,3]

1 Department of Electronics, Information and Bioengineering, Politecnico di Milano, Italy
2 Biocubica srl, Milan, Italy
3 Department of Industrial and Information Engineering University of Pavia, Italy

Abstract

Wearable devices like photoplethysmogram (PPG) sensors are prone to motion artifacts, affecting the quality of cardiovascular data. Traditional denoising methods often degrade the signal, while AI-driven solutions like deep learning (DL) struggle with random and systematic distortions, requiring large datasets for training. To overcome these limitations, this study proposes a style transfer–assisted cycle-consistent generative adversarial network (stccGAN) for denoising 3-channel PPG signals (red, green, and infrared) acquired by the patented Soundi chest sensor. Two identical devices were used: one to collect the chest PPG signal (to be denoised) and another to obtain a synchronized finger PPG signal (reference signal). The proposed stccGAN uses style transfer with dual generators featuring U-Net, GRU, and LSTM layers to improve chest PPG quality. Validation with data from 30 subjects (20 for training, 10 for testing) showed an average 70% correlation with the reference signal, with a 15% improvement over raw chest PPG. This demonstrates effective signal restoration, enabling accurate cardiac assessment and blood pressure estimation from chest PPG signals.

1. Introduction

Physiological signals acquired from wearable devices, such as the photoplethysmogram (PPG) [1], face ongoing challenges due to substantial motion artifacts, demanding waveform denoising and restoration [2]. It has been extensively documented in the literature that ineffective procedures may often induce modifications or even the degradation of the underlying cardiovascular information extracted from the PPG, making wearable devices less reliable [3]. Current trends in biosignal processing focus on advanced data-driven artificial intelligence tools, as end-to-end deep learning (DL) networks [4]. They automatically encode the significant feature describing the intrinsic nature of the physiological waveforms and exploit that to restore the signal quality in the presence of artifacts [5]. Nonetheless, such methods often experience difficulties in denoising the inherent bodily signals due to many random and systematic in-band distortions. Besides, such AI-based techniques require large datasets to suitably estimate the model parameters evading bias issues, which may further compromise signal restoration. Finally, a supervised DL approach requires strict correspondence between input and labels in the training set, usually requiring simulated noise, of known statistics, added to the acquired signals. To address to above issues, in this paper, we developed a cycle-consistent generative adversarial network, named *stccGAN*, to denoise the PPG signal, without needing noise simulation. Signals were acquired by the patented (no. EP3248541A1) Soundi chest sensor [6], CE marked as a medical device (class II). It features a circular shape with a diameter not as much as 6 cm and a thickness of approximately 1 cm, with a weight not exceeding 40 grams. The device enables continuous signal recording for a maximum duration of approximately 24 hrs (see [7] for further technical details). To apply the CycleGAN, two identical devices were considered in the experiments, device #1 gathering the chest PPG signal (to be denoised) and device #2, properly synchronized with the first one, acquiring the finger PPG signal (reference signal).

2. Materials and Methods

2.1. Data acquisition and pre-processing

The wearable device used in this work acquired three PPG channels (red, green, and infrared). It was positioned on the chest surface close to the heart, on the 2^{nd}-3^{rd} intercostal space. A medically certified double-sided tape ensured that device #1 sticked securely to the chest. Simultaneously, the subjects were asked to place their left ring finger on the sensor of device #2. The adherence between the finger and the optical sensor was secured to avoid any movement or environmental light effects. Thirty healthy adult subjects were enrolled in the study, aged 26±8 years. The acquisition protocol included the following phases: a 5-minute rest period in sitting posture and a 5-minute session in standing posture. For each participant, 5 repetitions of the acquisition protocol were performed. Between each participant, proper sanitization of the utilized devices was carried out. Data was collected under the Ethical Committee approval Opinion 3/2019, Politecnico di

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.037

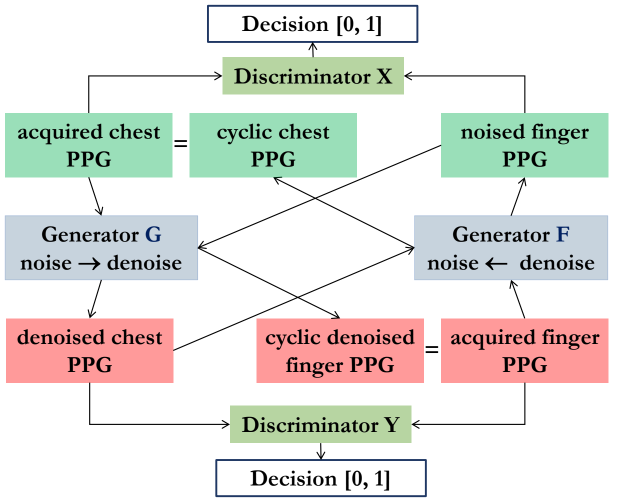


Figure 1. Training architecture of the denoising model stccGAN. Once trained, the focus is on the generator G.

Milano University. The acquired signals underwent uniform sampling at 400 Hz, filtering (pass-band Bessel model: 0.5 to 25 Hz), amplitude normalization, subdivision in 5-second chunks, and quality checking. In this last process, artifact finger PPG chunks were removed from the dataset along with the corresponding chest PPG chunks. Overall, 5140 valid chunks were obtained.

2.2. stccGAN Network

The developed stccGAN was based on a cyclic coherent GAN (Fig. 1). It comprised two main components: a) the generator, responsible for signal-to-signal translation, b) the discriminator enabling the signal classification. The interaction between these elements worked as follows. The generator G processed an acquired chest PPG and produced a corresponding denoised chest PPG. The discriminator Y received either the denoised chest signal or the acquired finger PPG signal, tasked with identifying whether the signal is original (finger PPG) or reconstructed (denoised chest PPG). It labeled original and reconstructed signals as "1" and "0", respectively. The goal of generator G was to produce a high-quality signal that can deceive the discriminator Y into mistaking reconstructed chest PPG for real finger PPG. Similarly, generator F took the acquired finger PPG and generated a noised version. Discriminator X distinguished between acquired chest PPG ("1") and noised finger PPG ("0"). This basic structure was extended with two additional cycles, giving the model its name, GAN with cyclic coherence. To reinforce the style-transfer process, the denoised chest PPG, generated by generator G, is fed into generator F, which outputs a "cyclic" chest PPG expected to match the acquired chest PPG. An identical process occurs for the noised version of the finger PPG, which is converted back into a “cyclic” denoised finger PPG to be compared with the original acquired finger PPG. Ultimately, the focus was on the performance of generator G and its ability to produce denoised chest PPG that closely resembles the acquired finger PPG. Both generators, implemented as UNet models, embedded three inception convolutional layers in the encoding path, a gated recurrent unit (GRU) layer on each skip connection. In the decoding path, the network implemented three deconvolutional layers. The output layer was composed of a bidirectional LSTM of the size 32 samples. The input of the generator was composed of a tridimensional array (red, green, and infrared) of 2000 samples, the acquired chest PPG to be denoised, while the output consisted of the denoised chest PPG signal. The discriminator was an encoder classification network, composed of three convolutional layers, with a sigmoidal neuron in the output.

2.3. Network training and evaluation metrics

The network architecture was developed to enact the elaboration in bundle of the three wavelength signals by means of channel-wise convolutions to take advantage of the intrinsic signal redundancy. Three main loss functions were implemented for each half-cycle, namely the adversarial loss L_{adv}, cycle loss L_{cycle} and identity loss L_{id}. The overall loss was computed by weighting the three loss functions as:

$$L_{tot} = L_{adv} + \lambda L_{cycle} + \beta L_{id}$$

where λ and β were set to 1 and 0.1, respectively. In the network training, the PPG signal acquired at the finger was utilized as a reference minimizing both the adversarial loss, the cycle coherence root mean squared error loss, and the transparency identify loss. Network training was performed by selecting the chunks from the first 20 participants, splitting the set into training (70%) and validation (30%) subsets. To reduce overfitting, the model was trained by monitoring the validation loss with a patience factor of 15 iteration steps. The evaluation metrics was computed on 1542 chunks of test set chunks (remaining 10 participants). Due to the intrinsic time lag of the finger and chest signals, cross-correlation was first computed to optimally aligning the reference finger PPG with the chest PPG, for both acquired and denoised versions as:

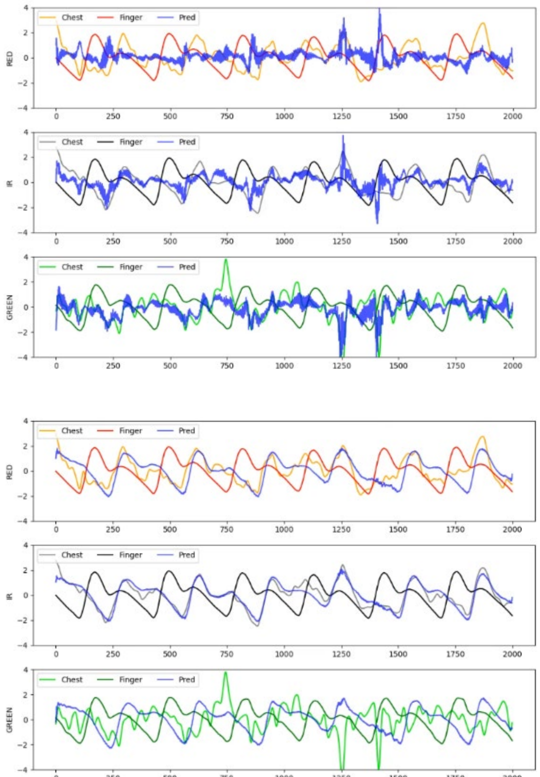
$$c_k = \sum_n s_{n+k} \cdot t_n$$

with the maximum correlation $\hat{c}_k$ found in correspondence of sample k minimizing the time lag.

3. Results

The experiments were performed using a 12th Gen Intel(R)

Core(TM) i7-12850HX, 2.10 GHz, featuring 32.0 GB of RAM, equipped with an NVDIA RTX A200 8GB video card. One training took, on average, about 15 minutes, while one denoising inference of a 5-second chunk was trey much instantaneous.



(a)

(b)

Figure 2. Example of network training of the three chest PPG channels: a) initial iteration step, b) last iteration step. In blue the denoised chest PPG, in lighter colors the acquired finger PPG, and in darker colors the acquired chest PPG.

Qualitative results (Fig. 2) showed that the denoised chest signal sensibly improved from the first training iteration (a, blue line) up to the convergence iteration (b, blue line). Especially the green channel, that was the less reliable, was nicely denoised recovering the typical PPG waveform, even modeling the dicrotic notch. Interestingly, at the beginning of the training the network output was very noisy but progressively approached the style of the finger PPG waveform. As expected, the network consistently neglected the time lag between the finger and chest signals, reconstructing nonetheless a time-consistent denoised waveform of the PPG. Quantitative results (Table 1) showed a chest signal correlation with the corresponding finger PPG signal in the range of about 70% with an improvement up to 15% for the green channel with respect to the correlation computed with the acquired chest PPG. The corresponding

Table 1. Correlation comparison between the acquired chest PPG and denoised PPG.

Channel	*acquired chest PPG*	*denoised chest PPG*
RED	0.50	0.66
IR	0.60	0.68
GREEN	0.58	0.70

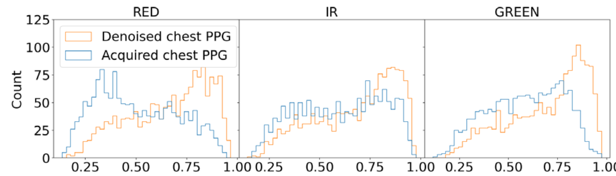


Figure 3. Histogram of cross-correlation of the test set computed using finger PPG as reference for both original and denoised chest PPG.

histogram distribution of the cross correlations, across all the test set chunks, highlighted (shift of the correlation toward 1) how the green and red channel benefit the most of the denoising process (Fig. 3). The frequency analysis (Figg. 4, 5) highlighted the ability of the model to preserving relevant low frequency content while removing undue noise especially in the green channel that, despite filter-based preprocessing, was carrying residual components higher than 25 Hz.

4. Discussion and conclusions

The results supported the hypothesis that the approach style transfer–assisted deep learning is feasible to denoise chest PPG signal considering the finger PPG as a waveform reference. We showed that the 3-channel signal improved up to 15%, with special gain in the green channel. In [4], the authors reported about a bidirectional recurrent denoising auto-encoder (BRDAE) trained and validated on a dataset with artificially augmented noise. The network learnt the simulated noise statistics, reporting an improvement of the signal-to-noise ratio of about 8 dB, while the generalization on real PPG signals was only indirectly computed through heart rate quality. The chest PPG signal is inherently noisier than finger PPG signals due to factors like respiratory movements, muscle contractions, and proximity to the heart, which may introduce additional low-frequency components. To denoise such a signal, we used the finger PPG signal as a reference because it is typically considered to be a high-quality signal with less susceptibility to motion artifacts compared to PPG signals from other locations like the chest. The assumption was that finger PPG provides a reliable baseline for comparison when denoising the chest PPG signal. The rationale of our work was meant to extend the usual denoising paradigm of removing high-frequency and in-band noise by considering the concept of style transfer from finger PPG to chest PPG, sparing however the inherent time lag.

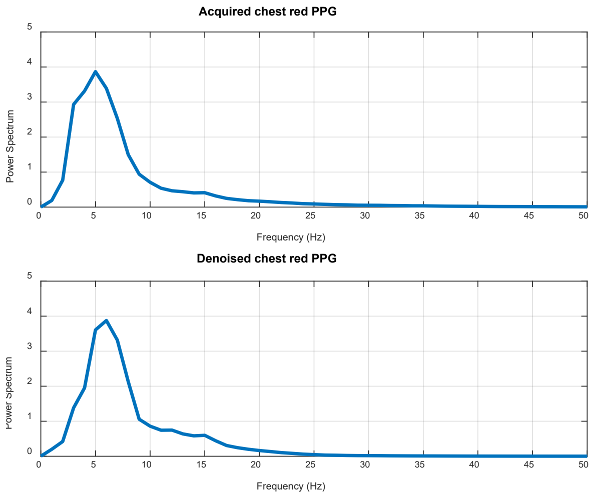


Figure 4. Power spectra of the acquired and denoised chest PPG (red channel), averaged across all the test set chunks.

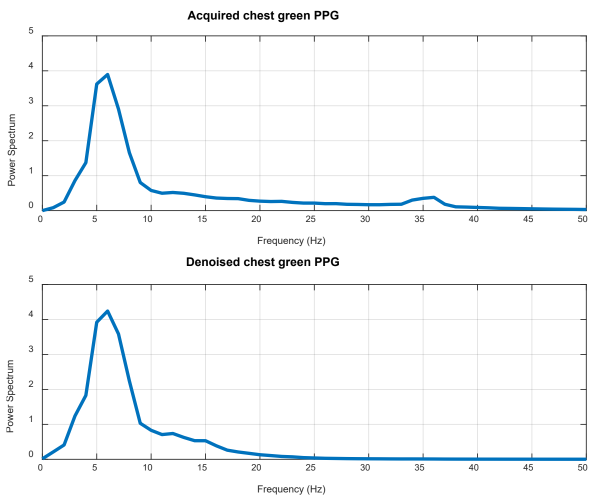


Figure 5. Power spectra of the acquired and denoised chest PPG (green channel), averaged across all the test set chunks.

We point out here that there are inherent differences between the two, such as variations in pulse wave morphology and respiratory modulation. This assumption may not fully account for these differences, leading to the possibility that the denoising process could inadvertently make the chest PPG signal resemble a finger PPG signal, potentially altering the physiological characteristics of the signal. This could be seen as a limitation if the goal is to preserve the unique characteristics of the chest PPG signal. On the other hand, it could be an advantage if the primary objective is to obtain a cleaner signal for specific analyses like blood pressure estimation, where the chest PPG signal might benefit from being more similar to a finger PPG signal. Chest PPG is not as widely used as finger PPG, primarily due to its susceptibility to noise. However, the method developed for chest PPG could be particularly valuable in expanding the capabilities of wearable devices, such as ECG chest patches, allowing them to provide additional physiological data like heart rate variability and blood pressure estimation. In additional, the developed methodology can be easily extended to wrist providing more reliable PPG signals. Overcoming the noise challenges associated with chest PPG could thus significantly enhance the utility of these devices in continuous health monitoring. While needing further methodological developments to optimize the network and extensive validation on larger datasets, the study could pave the way for more comprehensive cardiovascular monitoring solutions that integrate PPG with other modalities like ECG, enabling more accurate and holistic health assessments in a single device.

Acknowledgment

This work was supported in part by P.E. PE00000013-FUTURE ARTIFICIAL INTELLIGENCE RESEARCH (FAIR), Italian Ministry of Research and University, and UE-Horizon Europe-SMASH-HCM n. 101137115.

References

[1] M.A. Almarshad *et al.*, Diagnostic Features And Potential Applications Of PPG Signal In Healthcare: A Systematic Review. *Healthcare* (Basel). vol. 10, no. 3, pp. 547, 2022.

[2] I Garcia-Lopez and E. Rodriguez-Villegas, Characterization of Artifact Signals in Neck Photoplethysmography. IEEE Trans Biomed Eng. 2020;67(10):2849-2861.

[3] P. Martín-Escudero et al. Are Activity Wrist-Worn Devices Accurate for Determining Heart Rate during Intense Exercise?. Bioengineering (Basel). Vol. 10, no. 2 pp: 254, 2023.

[4] J. Lee *et al.*, Bidirectional Recurrent Auto-Encoder for Photoplethysmogram Denoising, *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2375-2385, 2018.

[5] Hu Q, Li M, Jiang L, Liu M. A Hybrid Denoising Approach For PPG Signals Utilizing Variational Mode Decomposition And Improved Wavelet Thresholding. *Technol Health Care*, vol. 32, no. 4, pp. 2793-2814, 2024.

[6] D. Marzorati *et al*., Hybrid Convolutional Networks For End-To-End Event Detection In Concurrent PPG And PCG Signals Affected By Motion Artifacts. *IEEE Transactions on Bio-Medical Engineering*, vol. 69, pp. 2512–2523, 2022.

[7] M. Rossi *et al.*, SLEEP-SEE-THROUGH: Explainable Deep Learning For Sleep Disorder Detection And Quantification From Wearable Somnography. *IEEE Journal of Biomedical Health Informatics*, vol. 7, no. 7, pp. 3129-3140, 2023.

Address for correspondence:
Department of Industrial and Information Engineering
University of Pavia
Via Adolfo Ferrata 5, 27100 Pavia, (IT)
pietro.cerveri@unipv.it