An End-to-End and Accurate PPG-based Respiratory Rate Estimation Approach Using Cycle Generative Adversarial Networks

Seyed Amir Hossein Aqajari*,[1], Rui Cao[1], Amir Hosein Afandizadeh Zargari[1], and Amir M. Rahmani[1,2,3]

Abstract— **Respiratory rate (RR) is a clinical sign representing ventilation. An abnormal change in RR is often the first sign of health deterioration as the body attempts to maintain oxygen delivery to its tissues. There has been a growing interest in remotely monitoring of RR in everyday settings which has made photoplethysmography (PPG) monitoring wearable devices an attractive choice. PPG signals are useful sources for RR extraction due to the presence of respiration-induced modulations in them. The existing PPG-based RR estimation methods mainly rely on hand-crafted rules and manual parameters tuning. An end-to-end deep learning approach was recently proposed, however, despite its automatic nature, the performance of this method is not ideal using the real world data. In this paper, we present an end-to-end and accurate pipeline for RR estimation using Cycle Generative Adversarial Networks (CycleGAN) to reconstruct respiratory signals from raw PPG signals. Our results demonstrate a higher RR estimation accuracy of up to 2× (mean absolute error of 1.9±0.3 using five fold cross validation) compared to the state-of-th-art using a identical publicly available dataset. Our results suggest that CycleGAN can be a valuable method for RR estimation from raw PPG signals.**

I. INTRODUCTION

Respiratory rate (RR), often referred to as breathing rate, is the number of breaths a person takes per minute. A normal resting RR for adults ranges from 12 to 20 [1]. Abnormal changes in respiratory rate are an accurate indicator of physiological conditions such as anxiety, hypoxia, hypercapnia, metabolic and respiratory acidosis [2]. A diverse body of research studies has indicated the significance of respiration rate for forecasting events such as cardiac arrest, patient deterioration, and care escalation [3], [4], [5].

The importance of respiratory rate as one of the first indicators of health deterioration has attracted significant attention in RR's daily monitoring [6]. However, RR's reliable measurement devices are bulky and cumbersome, and are mainly used for inpatients. With the rapid development of wearable technologies, a change in an individual's physiological systems' functional state can be tracked and monitored in an everyday setting, for instance, by using photoplethysmography (PPG) [7]. PPG signals can easily be collected continuously and remotely using a wide range of inexpensive, convenient, and portable wearable devices (.e.g., smart watches, rings, etc.). The blood perfusion dynamics

[1] Department of Electrical Engineering and Computer Science, University of California Irvine, CA 92697, USA (*correspondence e-mail: saqajari@uci.edu)
[2] Department of Computer Science, University of California, Irvine, CA 92697, USA
[3] Institute for Future Health, University of California, Irvine, CA 92697, USA

are known to carry information on breathing, as respiration-induced modulations in PPG signals [8]. Hence, PPG signals are considered as a suitable source for respiratory rate extraction to forecast unexpected care admissions in a daily life setting .

The RR estimation from PPG signals has received remarkable attention in the literature [9]. Traditional RR estimation methods require several steps, including digital filtering, time/frequency domain analysis, extraction of signal components from composite signals, deriving respiratory surrogate waveforms and features using the fiducial points, signal quality estimations, and sensor fusion [10]. These techniques rely heavily on manual parameter tuning, optimization, and hand-crafted rules designed for specific target patient population. In contrast, Bian et al. [11] recently proposed an end-to-end deep learning approach in order to automatically and accurately estimate RR from raw PPG signals. Despite the automatic nature of their proposed model, the performance of this method is not ideal (mean absolute error (MAE) of 3.8 ± 0.5 bpm, which is about %25 inaccuracy considering 16 bpm as an average RR per minute).

This work proposes an automatic end-to-end generative deep learning approach using cycle generative adversarial networks (CGAN) [12] to reconstruct respiratory signals from raw PPG signals and estimate RR with a high accuracy. CGAN is a novel and powerful approach in the field of unsupervised learning, which targets learning the structure of two given data domains to translate an individual input from one domain to a desired output from the second domain. We also propose a novel loss function to be integrated in our CGAN model that takes into account the key attribute (i.e., RR) of the generated respiratory signals. Our results demonstrate that the proposed GAN-based approach estimates RR from raw PPG signals with 2× higher accuracy compared with the state-of-the-art approach [11] using real-world data. Furthermore, our method outperforms the classical RR estimations methods, despite utilizing the complete automatic end-to-end design.

In summary, this work makes the following key contributions:

- Proposing an end-to-end automatic approach based on CGAN which outperforms the performance of the classical RR estimation methods (using an identical setting and dataset).
- Proposing a novel loss function for our CGAN model that takes into account the RR of the generated respiratory signals.
- Demonstrating the performance of our approach using

the real-world data and comparing it against the state-of-the-art (using an identical setting and dataset).

The rest of this paper is organized as follows. Section II introduces the employed dataset and our proposed pipeline architecture. In Section III we summarize the result obtained by our proposed method. Section IV compares our result with the state-of-the-art in RR estimation from raw PPG signal. Finally, Section V concludes the paper.

II. MATERIAL AND METHODS

A. Dataset

We employed BIDMC PPG and Respiration Dataset [13] to evaluate our RR estimator method. This is a publicly available dataset which contains signals and numerics extracted from the much larger MIMIC II matched waveform Database, along with manual breath annotations made from two annotators, using the impedance respiratory signal.

In this dataset, PPG and impedence respiratory signals are collected from 53 adult patients for about 8-minute duration at sampling rate of 125 Hz. This dataset is widely used to evaluate the performance of different algorithms for estimating respiratory rate from PPG signals [9], [14], [15], [11].

B. RR Estimation Pipeline

Figure 1 shows our proposed pipeline for estimating respiratory rate from PPG signals. There are three different main stages in this pipeline: (1) Data Preparation, (2) PPG to Respiration Translator (PRT), and (3) RR Estimator. In the following subsections, we discuss each part in detail.

1) Data Preparation: The primary purpose of this stage is to prepare the data and pre-process it for the PPG-to-Respiration Translator (PRT) module. PPG signals are sampled at much higher frequency than required in the BIDMC dataset. Therefore, downsampling is done to save memory, and reduce processing time and computational complexity of our model while preserving the signals integrity. In this stage, first, raw PPG data are normalized to 0-1. Then, the signals are down-sampled to 30 Hz. Finally, 30-second windows of data are extracted from the signals to be used in the PRT module for translation.
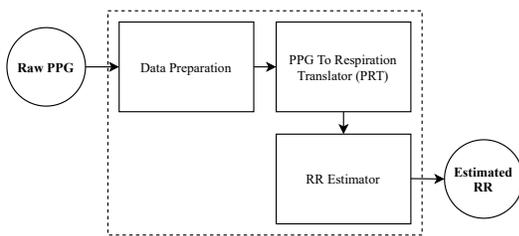

Fig. 1: RR Estimation Pipeline

2) PPG to Respiration Translator (PRT): In this module, the Cycle Generative Adversarial Networks (The Cycle GAN) are employed to reconstruct respiratory signals from raw PPG signals. The Generative Adversarial Networks belong to the field of unsupervised learning targeting to learn the structure of a given data in order to generate new unseen data. The GANs are composed of two models: a generator network and a discriminator network. The generator network starts at a point from a latent space as an input and aims to generate new data similar to the expected domain. The discriminator network on the other hand attempts to recognize if an input data is real (belongs to the original dataset) or fake (generated by the generator network).

The Cycle GAN is an extension of the Generative Adversarial Networks which was first proposed by Jun-Yan Zhu et al. [12]. The idea behind the Cycle GAN is to take an input from the first domain and generate an output of the second domain. In our case, the goal of Cycle GAN is to learn the mapping between PPG signals (domain *X*) and respiratory signals (domain *Y*). Each domain contains set of training samples $\{x_i\}_{i=1}^{N} \in X$ and $\{y_i\}_{i=1}^{N} \in Y$ used directly from BIDMC dataset. The model includes two generators with mapping functions as $G : X \to Y$ and $F : Y \to X$ and two discriminators D_X and D_Y. In the discriminator networks, D_X aims to distinguish between real PPG signals (x_i) and synthetic PPG signals ($F(y)$) while D_Y aims to discriminate between real respiratory signals (y_i) and synthetic respiratory signals ($G(x)$).

We indicate the distributions of our data as $x \sim p_{data}(x)$ and $y \sim p_{data}(y)$. Our objective loss function contains three terms: (1) adversarial losses [16], (2) cycle consistency losses, and (3) RR Loss.

Adversarial losses are employed for matching the distribution of generated synthetic signals to the data distribution of original signals. We apply adversarial loss function on both of our mapping functions. The objective function applied to the mapping function *G* is expressed as below:

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim p_{data}(y)}[log D_Y(y)] \\ + E_{x \sim p_{data}(x)}[log(1 - D_Y(G(x)))] \quad (1)$$

where G tries to generate respiratory signals ($G(x)$) that look similar to original respiratory signals collected from BIDMC dataset (domain Y), while D_Y aims to discriminate between synthetic respiratory signals ($G(x)$) and real samples (y). In a same way, adversarial loss for the mapping function F is expressed as $L_{GAN}(F, D_X, Y, X)$.

A mapping function trained only by adversarial loss as an objective function can map the same set of signals from the first domain to any random permutation of signals in the second domain. Therefore, cycle consistency losses are added to guarantee the mapping from an individual input (x_i) to a desired output (y_i) by considering learned mapping functions to be cycle consistent. This means that for each PPG signal x from domain X we must have $x \to G(x) \to F(G(x)) \approx x$ while for each respiratory signal y we have $y \to F(y) \to G(F(y)) \approx y$. This behaviour is indicated as:

$$L_{cyc}(G, F) = E_{x \sim p_{data}(x)}[||F(G(x)) - x||_1] \\ + E_{y \sim p_{data}(y)}[||G(F(y)) - y||_1] \quad (2)$$

In order to force the synthesized respiratory signals to only keep their main features, we define RR loss function which attempts to take into account the RR of the generated respiratory signals. The BioSPPy [17] public python library is used to calculate the respiration rate of the synthetic and original respiratory signals. This additional loss function can be expressed as:

$$L_{RR}(G) = E_{y \sim p_{data}(y)}[||G(F(y))_{RR} - y_{RR}||_1] \quad (3)$$

Therefore, the final objective is the weighted sum of the above loss functions:

$$\begin{aligned} L(G, F, D_X, D_Y) = &L_{GAN}(G, D_Y, X, Y) \\ &+ L_{GAN}(F, D_X, Y, X) \\ &+ \lambda_1 L_{cyc}(G, F) \\ &+ \lambda_2 L_{RR}(G) \end{aligned} \quad (4)$$

where λ_1 and λ_2 are the weights of cycle consistency loss and RR loss respectively (both are empirically selected as 10 in our work).

G and F attempt to minimize this objective against adversaries D_X and D_Y that try to maximize it. Hence, we aim to solve:

$$G^*, F^* = \arg\min_{G,F} \max_{D_X, D_Y} L(G, F, D_X, D_Y) \quad (5)$$

We use the CGAN architecture proposed by [12]. The architecture of generative networks is adopted from Johnson et al. [18]. This network contains two stride-2 convolutions, several residual blocks [19], and two fractionally-strided convolutions with stride 0.5. For the discriminator networks we use 70×70 PathGANs [20], [21], [22] which aims to classify whether the signals are fake or real.

3) RR Estimator: This module uses the BreathMetrics [23] for pre-processing, filtering, and calculating the respiratory rate of the signals. The BreathMetrics is a Matlab toolbox for analyzing respiratory recordings. Breathmetrics was designed to make analyzing respiratory recordings easier by automatically de-noising the data and extracting the many features embedded within respiratory recordings including respiratory rate. The methods used in this tool (de-noising and features extraction) have been validated, peer-reviewed, and published in Chemical Senses, a scientific journal.

C. Performance Metric

We calculate the mean absolute error (MAE) as the performance metric in our study in order to evaluate our RR estimation method. MAE is calculated by averaging the absolute differences between the values estimated by a model and the values observed.

MAE is defined as:

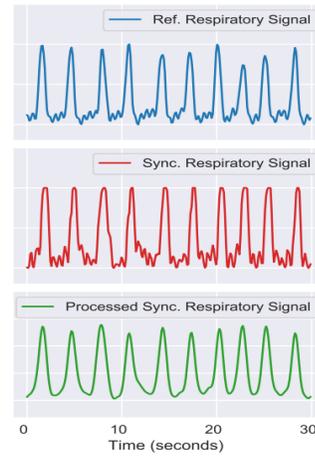$$MAE = \frac{1}{N} \sum_{i=1}^{N} |RR_e^i - RR_r^i| \quad (6)$$

Fig. 2: Example of the synthetic respiratory signal generated by our PRT module, from top to bottom: reference respiratory signal, synthetic respiratory signal, and processed synthetic respiratory signal using the BreathMetrics module.

Where N is total number of respiratory segments. RR_e^i and RR_r^i are estimated and reference respiratory rate for each 1 minute data respectively.

To evaluate the performance of our algorithm, we use 5-fold cross validation. We split the BIDMC dataset into five folds while making sure each subject's data is appeared in one fold only. We perform the training 5 times and each time 4 folds are for training and 1 fold is for testing. The average MAE of 5 runs is presented as the final performance results for our proposed model. We trained our models on Nvidia Quadro RTX 5000 GPU with 125 GB of RAM. For each training experiment we performed 100 epochs in which early stopping technique were employed to reduce over-fitting.

III. RESULTS

Figure 2 shows an example of a reconstructed respiration signal using our method alongside with its reference respiration signal for the same timing window of the same test subject. The blue signal represents the reference respiration signal from the BIDMC dataset. The red signal is the synthetic respiratory signal which is an output of our PRT module. The green signal shows the final respiratory signal after de-noising and pre-processing of the signal using the BreathMetrics tool. By comparing the reference respiratory signal and the processed synthetic respiratory signal it can be seen that our pipeline is capable of reconstructing the respiratory signal itself from the PPG signal with high precision.

Table I shows the summary of the MAE performance (average±std) of our method compared with the state of the arts. The same publicly available dataset (BIDMC) and the same performance metric (MAE using five fold cross validation) are used to fairly evaluate and compare the performance of our automatic approach.

As can be seen from the table, our CGAN-based method significantly (up to 2×) outperforms the state-of-the-art RR

TABLE I: The summary of MAE Performances in compare with state of the arts

RR Estimation Models	MAE
Bian et al. [11] DL method	3.8±0.5
Bian et al. [11] SmartQualityFusion method	2.6±0.4
Our proposed CycleGAN-based method	1.9±0.3

estimation methods from a raw PPG signals. In the next section, we discuss the results in more details.

IV. DISCUSSION

Existing PPG-based RR estimation methods heavily rely on features derived from various hand-crafted algorithms and tuning parameters for specific settings. In this work, we presented a novel fully end-to-end automatic approach for estimating RR from raw PPG signals.

A classical RR estimation method has been implemented in [11] which is referred as SmartQualityFusion, as a combination of Smart fusion [24] and Quality fusion [25]. The SmartQualityFusion algorithm achieved the MAE of 2.6±0.4 breaths/min (Table I). According to Table I, our proposed RR estimation pipeline demonstrate a significant improvement in accurately estimating RR compared with their algorithm (the MAE of 1.9 breaths/min), despite the complete end-to-end automatic environment of our method.

Bian et al. [11], also proposed an end-to-end learning approach based on deep learning to estimate RR from PPG. Their results demonstrate a clinically reasonable performance; however, the performance of their proposed automatic approach heavily depends to the availability of real world data. They used the mean absolute error (MAE) as a performance metric in their work. According to their results, their deep learning approach trained with real data could achieve the MAE of 3.8±0.5 breath/minutes which is substantially higher than the MAE error achieved by their classical algorithm (SmartQualityFusion). According to Table I, our proposed pipeline architecture provides about 2 times better performance of estimating respiratory rate compared with their proposed deep learning approach.

One existing limitation of our work, which is also present in the other state-of-the-art methods, is the lack of noisy PPG data in the real datasets used for training our models. Therefore, the performance of our proposed method might not be as expected on a noisy PPG data collected during daily life and physical activities. This limitation is also present in the other state-of-the-arts methods since the employed datasets are mostly consist of stationary data. However, as also mentioned in [11], this problem can easily be diminished by retraining the model using the new existing noisy PPG and respiratory signals.

In [11], the authors increased the performance of their proposed method, by augmenting the real datasets using the generated synthetic PPG data. This enhanced the MAE of their model from 3.8 brpm to 2.5 brpm. Thus, the data augmentation approach significantly enhanced the performance of their model. However, our proposed model still achieves a better performance (1.9 brpm), despite being trained only on

a real data. As a future work, we intend to further increase the performance of our proposed method by augmenting our dataset using synthetic data.

V. CONCLUSION

In summary, in this work we presented a novel pipeline architecture in order to estimate respiratory rate using PPG signals. We are the first one to use cycle adversarial networks in our model to reconstruct the respiratory signals from PPG signals. According to our results, our proposed pipeline architecture is able to estimate RR with the MAE of 1.9 which has the performance of 2.0x better than the state-of-the-art automatic RR estimation method.

REFERENCES

[1] "my.clevelandclinic.org," https://my.clevelandclinic.org/health/articles/, [Online; accessed Mar-2021].

[2] S. Rolfe, "The importance of respiratory rate monitoring," *BJN*, 2019.

[3] S. J. R. Cooper *et al.*, "Respiratory rate records: the repeated rate?" *JCN*, 2014.

[4] L. Dougherty *et al.*, *The Royal Marsden manual of clinical nursing procedures*. John Wiley & Sons, 2015.

[5] M. A. Cretikos *et al.*, "Respiratory rate: the neglected vital sign," *MJA*, 2008.

[6] A. Nicolò *et al.*, "The importance of respiratory rate monitoring: From healthcare to sport and exercise," *Sensors*, 2020.

[7] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological measurement*, 2007.

[8] M. Pirhonen *et al.*, "Acquiring respiration rate from photoplethysmographic signal by recursive bayesian tracking of intrinsic modes in time-frequency spectra," *Sensors*, 2018.

[9] P. H. Charlton *et al.*, "Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review," *IEEE RBME*, 2017.

[10] ——, "An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram," *Physiological measurement*, 2016.

[11] D. Bian *et al.*, "Respiratory rate estimation using ppg: A deep learning approach," in *2020 42nd Annual International Conference of the IEEE EMBC*, 2020.

[12] J.-Y. Zhu *et al.*, "Unpaired image-to-image translation using cycle-consistent adversarial networkss," in *2017 IEEE ICCV*, 2017.

[13] M. A. Pimentel *et al.*, "Toward a robust estimation of respiratory rate from pulse oximeters," *IEEE TBME*, 2016.

[14] D. A. Birrenkott *et al.*, "A robust fusion model for estimating respiratory rate from photoplethysmography and electrocardiography," *IEEE TBME*, 2017.

[15] D. Jarchi *et al.*, "Accelerometry-based estimation of respiratory rate for post-intensive care patient monitoring," *IEEE Sensors Journal*, 2018.

[16] I. J. Goodfellow *et al.*, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661*, 2014.

[17] C. Carreiras, "BioSPPy: Biosignal processing in Python," 2015–. [Online]. Available: https://github.com/PIA-Group/BioSPPy/

[18] J. Johnson *et al.*, "Perceptual losses for real-time style transfer and super-resolution," in *ECCV*. Springer, 2016.

[19] K. He *et al.*, "Deep residual learning for image recognition," in *Proceedings of the IEEE CVPR*, 2016.

[20] P. Isola *et al.*, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on CVPR*, 2017.

[21] C. Li *et al.*, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *ECCV*. Springer, 2016.

[22] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE CVPR*, 2017.

[23] T. Noto *et al.*, "Automated analysis of breathing waveforms using breathmetrics: a respiratory signal processing toolbox," *Chemical senses*, 2018.

[24] W. Karlen *et al.*, "Multiparameter respiratory rate estimation from the photoplethysmogram," *IEEE TBME*, 2013.

[25] A. M. Chan *et al.*, "Ambulatory respiratory rate detection using ecg and a triaxial accelerometer," in *2013 35th Annual International Conference of the IEEE EMBC*.