P2E-WGAN: ECG Waveform Synthesis from PPG with Conditional Wasserstein Generative Adversarial Networks

Khuong Vo
Emad Kasaeyan Naeini
khuongav@uci.edu
ekasaeya@uci.edu
Department of Computer Science,
University of California, Irvine

Amir Naderi
Daniel Jilani
amnaderi@uci.edu
djilani@uci.edu
Department of Electrical Engineering and Computer Science,
University of California, Irvine

Amir M. Rahmani
a.rahmani@uci.edu
School of Nursing,
Department of Computer Science,
University of California, Irvine

Nikil Dutt
dutt@uci.edu
Department of Computer Science,
Department of Electrical Engineering and Computer Science,
University of California, Irvine

Hung Cao
hungcao@uci.edu
Department of Electrical Engineering and Computer Science,
Department of Biomedical Engineering,
University of California, Irvine

ABSTRACT

Electrocardiogram (ECG) is routinely used to identify key cardiac events such as changes in ECG intervals (PR, ST, QT, etc.), as well as capture critical vital signs such as heart rate (HR) and heart rate variability (HRV). The gold standard ECG requires clinical measurement, limiting the ability to capture ECG in everyday settings. Photoplethysmography (PPG) offers an out-of-clinic alternative for non-invasive, low-cost optical capture of cardiac physiological measurement in everyday settings, and is increasingly used for health monitoring in many clinical and commercial wearable devices. Although ECG and PPG are highly correlated, PPG does not provide much information for clinical diagnosis. Recent work has applied machine learning algorithms to generate ECG signals from PPG, but requires expert domain knowledge and heavy feature crafting to achieve good accuracy. We propose **P2E-WGAN**: a pure end-to-end, generalizable deep learning model using a conditional Wasserstein generative adversarial network to synthesize ECG waveforms from PPG. Our generative model is capable of augmenting the training data to alleviate the data-hungry problem of machine learning methods. Our model trained in the subject independent mode can achieve the average root mean square error of 0.162, Fréchet distance of 0.375, and Pearson's correlation of 0.835 on a normalized real-world dataset, demonstrating the effectiveness of our approach.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**;

SAC '21, March 22–26, 2021, Virtual Event, Republic of Korea

ACM ISBN 978-1-4503-8104-8/21/03.
https://doi.org/10.1145/3412841.3441979

KEYWORDS

generative adversarial networks, electrocardiogram, photoplethysmogram

ACM Reference Format:
Khuong Vo, Emad Kasaeyan Naeini, Amir Naderi, Daniel Jilani, Amir M. Rahmani, Nikil Dutt, and Hung Cao. 2021. P2E-WGAN: ECG Waveform Synthesis from PPG with Conditional Wasserstein Generative Adversarial Networks. In *The 36th ACM/SIGAPP Symposium on Applied Computing (SAC '21), March 22–26, 2021, Virtual Event, Republic of Korea.* ACM, New York, NY, USA, Article 4, 7 pages. https://doi.org/10.1145/3412841.3441979

1 INTRODUCTION

Electrocardiogram (ECG) is the measure of electrical activity produced by one's heart that generally requires multiple electrodes placed on the body. ECG is the gold standard for monitoring vital signs as well as diagnosing, controlling, and preventing cardiovascular diseases (CVDs) [1, 2]. According to the Global Burden of Disease reports, CVDs have become a prominent source of human death – about 32% of all deaths worldwide in 2017 [3]. Further, it has been shown that young people, especially athletes, are more prone to sudden cardiac arrests than ever before [4]. Monitoring ECG regularly is proven beneficial for the early detection of CVDs [5]. Recent advances in electronics, wearable technologies, and data science have allowed for simpler and more reliable ECG recording, as well as efficient analysis of large-scale data. However, collecting high-quality ECG data continuously and over a long period of time, especially in the daily life, still remains a challenge. The clinical gold standard, namely 12-lead ECG, and the simpler versions (i.e., Holter ECG) are bulky and inconvenient since placing multiple electrodes on the body usually causes discomfort and the signals get deteriorated over time as the skin-electrode impedance changes [2]. A potential solution is to mathematically derive ECG from an alternative, highly-correlated, non-invasive signal such as the photoplethysmogram (PPG) that is simple and easily acquired by a number of popular wearable devices (e.g., fitness trackers,

smart rings, etc.). PPG optically detects blood volume changes in the vascular bed of tissue, and has been widely used for heart rate and heart rate variability monitoring [6]. In a PPG device, the tissue (usually a peripheral organ like a fingertip or a wrist) is illuminated by a light-emitting diode, and the light absorption is measured by a photodetector.

ECG and PPG are inherently correlated. For the heart to contract and pump blood, a series of coordinated electrical signals are generated. These heart contractions cause variations in peripheral blood volume, which are measured by the PPG. ECG with PQRST waves provides a lot more information, especially for diagnosis in clinical practice. However, PPG is more convenient, economical, and easier to use, especially in daily life. For these reasons, PPG has been recently widely used in consumer electronics in wearable devices that deliver continuous and long-term monitoring capabilities. PPG devices can be attached to the earlobes, fingertips, wrists, and other body areas. Although PPG has become popular for health monitoring [7], ECG still plays the role as the gold standard for both medical diagnosis as well as vital sign monitoring. While standards in interpreting ECG have been long established, the vast use of PPG is still limited, thus it is mostly utilized for monitoring of heartrate and pulse oximetry. The close correlation between ECG and PPG can be exploited to develop an efficient method to synthesize ECG from PPG waveform, allowing an economical and user-friendly ECG screening for continuous and long-term monitoring. This approach takes advantage of both the rich clinical knowledge base of ECG in many instances, ranging from well-established ECG-based arrhythmia detection and abnormalities of heartbeat in unhealthy patients, focal caused of atrial fibrillation, and ventricular tachycardias, to name a few, as well as the convenience and accessibility of PPG in daily settings. Obviously, if ECG can be synthesized successfully from PPG signals acquired through nowadays' wearable devices, clinical diagnoses of cardiac diseases and anomalies could be done in real time, 24/7, anywhere, ideally with the aids from machine learning and cloud computing.

Nevertheless, there is a limited amount of research on ECG reconstruction. In [8], a machine learning-based approach was proposed to estimate ECG parameters and the ranges of RR, PR, QRS, and QT intervals based on the extracted time and frequency domain features from a fingertip PPG signal. Furthermore, the authors in [9] and [10] proposed the models to reconstruct the entire ECG from PPG in the frequency domain. However, those works require extensive preprocessing on the raw data, which could potentially lead to unexpected biases. Other researchers attempted to synthesize realistic ECG [11, 12] for data augmentation in order to solve the data-hungry problem of machine learning methods. Nonetheless, those had very little control over the synthesized ECG, which is not applicable when synthetic signals need to be conditioned on PPG signals.

In this paper, we propose an end-to-end deep learning model with a game-theoretic training approach to accurately estimate ECG waveforms from raw PPG. The contributions of this work are three-fold:

- We present **P2E-WGAN**, a conditional Generative Adversarial Network (GAN) model that works directly on 1-D time-series data to reconstruct/synthesize realistic ECG from PPG, and can be generalized effortlessly to any related signal pairs.
- We provide a novel method for ECG data augmentation based on PPG with a novel loss function that takes into account the key attributes of the generated ECG. To the best of our knowledge, this is the first study on conditional ECG synthesis using GANs.
- Our method requires no significant data preprocessing or feature engineering while achieving high morphological similarity to the reference ECG signals.

The implementation is released as open source on GitHub [1]. The subsequent sections are organized as follows. Section 2 describes the experimental data and our proposed method based on the GAN model. Section 3 presents the experimental results and analysis of synthetic ECG signals. Section 4 concludes the paper.

2 MATERIALS AND METHODS

2.1 Experimental Data

The Multi-parameter Intelligent Monitoring in Intensive Care (MIMIC) II database [13] was used for the experiments. The database contains thousands of recordings collected from patients at various hospitals. Each session has multiple physiological signals, including PPG and ECG signals, sampled at a frequency of 125 Hz. From the published preprocessed dataset [14], we randomly picked the records of 276 different patients having a record duration of at least 8 minutes. The first 80% of each section are selected for training and validation, and the rest are for testing. All the signals are scaled to the range $[-1, 1]$ and split into 3-second segments.

2.2 ECG Synthesis with Generative Adversarial Networks

The GAN is a generative model trained by a pair of neural networks in a game-theoretic approach [15]. In the GAN, a discriminator neural network D is trained to distinguish real from synthetic ECG signals, while a generator neural network G is trained to generate ECG from a latent space with the goal of making them indistinguishable by the discriminator. With ECG signals y drawn from data generating distribution $p_{data}(y)$ and signals z drawn from the noise prior p_z, G and D jointly optimize a non-artificial objective:

$$\begin{aligned}\mathcal{L}_{GAN}(G, D) =& \mathbb{E}_{y\sim p_{\text{data}}(y)}[\log D(y)] + \\ & \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z))]\end{aligned} \tag{1}$$

The discriminator is expected to output a high probability for a valid ECG signal and a low probability for a synthesized one, corresponding to the values of $\log D(y)$ and $\log(1 - D(G(z))$, respectively. G and D are trained simultaneously until G is able to successfully fool D.

2.3 Conditional training on PPG

Synthesizing ECG signals from the noise distribution yields stochasticity in the outputs of the generator, which means we barely control the characteristics of the synthesized signals. Nevertheless, we aim

[1] https://github.com/khuongav/P2E-WGAN-ecg-ppg-reconstruction

for mapping from an observed PPG signal to an output ECG signal. Therefore, we apply the objective function in [16] so that G and D directly observe signals x drawn from PPG distribution $p_{data}(x)$:

$$\begin{aligned}\mathcal{L}_{cGAN}(G,D) =& \mathbb{E}_{x,y\sim p_{\text{data}}(x,y)}[\log D(x,y)]+ \\ & \mathbb{E}_{x\sim p_{data}(x),z\sim p_z(z)}[\log(1-D(x,G(x,z))]\end{aligned} \tag{2}$$

In our model, the noise is expressed through the dropout [17] that is applied on several layers of the generator at both training and testing time.

2.4 ECG feature-based Wasserstein Loss Function

Instead of minimizing the Jensen-Shannon divergence [18] between the distribution of real data and the synthetic data distribution as in Equation 2, we shift the objective to minimizing the Wasserstein distance [18] between them. Optimizing the Wasserstein loss enhances the gradient propagation from the discriminator to the generator, especially when the discriminator can easily distinguish synthetic from real samples, which makes the gradient of the divergency eventually diminishes. Moreover, so as to satisfy the constraint of a 1-Lipschitz discriminator function [18], the gradient penalty is enforced between the two distributions [19]. Accordingly, the objective becomes:

$$\begin{aligned}\mathcal{L}_{cWGAN-GP}(G,D) =& -\mathbb{E}_{x,y\sim p_{\text{data}}(x,y)}[D(x,y)]+ \\ & \mathbb{E}_{x\sim p_{data}(x),z\sim p_z(z)}[D(x,G(x,z))]+ \\ & \delta\mathbb{E}_{\hat{y}\sim p_{\hat{y}}}\left[\left(\left\|\nabla_{\hat{y}}D(\hat{y})\right\|_2-1\right)^2\right]\end{aligned} \tag{3}$$

where $\hat{y}$ sampled from $\tilde{y}$ and y with t uniformly sampled between 0 and 1: $\hat{y}=t\tilde{y}+(1-t)y$ with $0\le t\le 1$. D is the set of 1-Lipschitz functions. The gradient penalty is weighted by a factor δ.

With this objective, the discriminator outputs an unbounded score that corresponds to the validity of a sample instead of the probability of real or synthetic.

We also add the traditional loss function that measures the $L2$ distances between synthesized signals and the ground truth:

$$\mathcal{L}_{L2}(G)=\mathbb{E}_{x,y\sim p_{\text{data}}(x,y),z\sim p_z(z)}[\|y-G(x,z)\|_2]. \tag{4}$$

This loss encourages the generator not only to deceive the discriminator but also synthesize signals that are morphologically similar to the real signals.

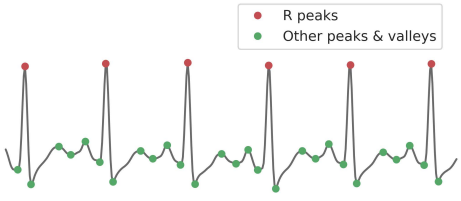


Figure 1: Detected peaks and valleys of an ECG segment.

For the purposes of data augmentation, it is not desired to reconstruct all samples to be as close as possible to the ground truth. Instead, the objective is to force the synthesized signals to only keep their main features. The peaks and valleys of ECG signals are chosen as the robust features that characterize both normal and abnormal heart rhythms (arrhythmia), as shown in Figure 1. The Hamilton-Tompkins algorithm [20] is applied to detect R peaks, and the *find_peaks* method implemented in Scipy library [21] is used to detect other peaks and valleys. In this case, the additional loss can be expressed as:

$$\mathcal{L}_{extrema}(G)=\mathbb{E}_{x,y\sim p_{\text{data}}(x,y),z\sim p_z(z)}[\|y_{extrema}-G(x,z)_{extrema}\|_2]. \tag{5}$$

Hence, the total objective is the weighted sum of the above loss functions:

$$\mathcal{G}^*=\arg\min_G\max_D \mathcal{L}_{cWGAN-GP}(G,D)+\gamma\mathcal{L}_{L2}(G)+ \tag{6}$$

$$\lambda_R\mathcal{L}_{extrema_R}(G)+\lambda_O\mathcal{L}_{extrema_O}(G) \tag{7}$$

where $\gamma, \lambda_R, \lambda_O$ denote the weights of per-sample loss, R-peaks loss, and other peaks-and-valleys loss, respectively.

2.5 Network architectures and hyperparameters

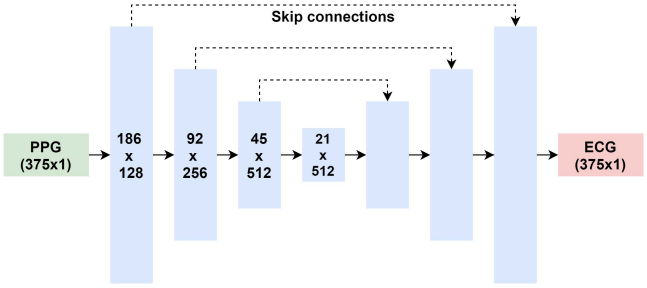


Figure 2: The architecture of the "U-Net" encoder-decoder generator with skip connections. The sizes of the feature maps (width x #channels) indicated inside the rectangles. 1-D convolution with width 4 and stride 2.

As shown in Figure 2, our generator network consists of a "U-Net" encoder-decoder [22] architecture based on 1-D convolution. Note that due to being inherently parallelizable, the convolutional networks run much faster than recurrent networks while achieving superior or competitive performance [23, 24]. The input is a 3-second PPG segment which has 375 samples, and the output is an ECG segment of the same length as the input. A 4-point moving average filter smoothes the ECG output to alleviate small artifacts presenting in the training signals. The input is compressed in down-sampled layers until a bottleneck layer, at which point it is expanded in up-sampled layers. The convolutional layers with replicate padding in the encoder and decoder consist of the kernels having width 4 and stride 2. The decoder uses transposed, or fractionally-strided, convolutions to increase the length of the sequence till the last layer having the *Tanh* activation function. Both the encoder and decoder components have $L=4$ layers. The output of layer i in the encoder is connected with the output of layer $L-i$

in the decoder by the skip connections, in order to consistently retain information in the down-sampled layers.

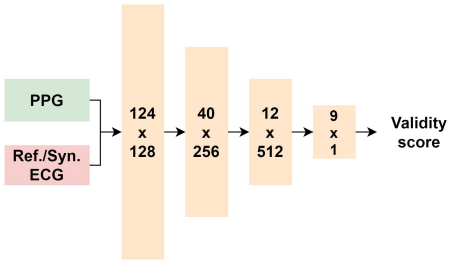


Figure 3: The architecture of the of the PatchGAN discriminator. The model receives concatenated PPG-ECG pairs. The sizes of the feature maps (width x #channels) indicated inside the rectangles. 1-D convolution with width 5 and stride 3, except the last convolution having stride 1.

As shown in Figure 3, the discriminator mirrors the encoder component of the generator. The input PPG segment with either the valid output ECG segment or the synthesized one is depth-wise concatenated to be fed to the discriminator. The model uses standard convolutional layers with filter length 5 and stride 3, except at the final layer of stride 1. The final layer uses a linear activation function. In addition, to encourage the discriminator to model the high-frequency structure of the heartbeats, we restrict the attention of the model to the structure of local signal patches with the discriminator architecture patchGAN [16]. The patchGAN architecture has fewer parameters, computationally efficient, and can be applied on arbitrarily long sequences. This discriminator attempts to score the validity of each 110-sample sequence, roughly an ECG cycle, as valid or invalid. The receptive field, which is the association of an output feature to an input region, is determined through the following formula [25]:

$$r_0 = \sum_{l=1}^{L} \left((k_l - 1) \prod_{i=1}^{l-1} s_i \right) + 1 \tag{8}$$

where r_0 is the receptive field size at the input layer, k_l is the kernel size at the l-th layer, and s_i is the stride at the i-th layer.

Instance normalization [26] and *ReLU* activation functions [27] are applied on the layers of G and D. *ReLUs* are leaky with the slope 0.2, to stimulate easier gradient flow. The encoder's layers have the dropout rates of 50%.

The GAN model was trained by alternating updates for the discriminator D and the generator G. The rate at which D learns is set to be three times more than that of G. The weight of the gradient penalty δ is set at 10. The model was trained with Adam optimizer [28] with the initial learning rate of 0.0002, the exponential decay rates $\beta_1 = 0.5$ and $\beta_2 = 0.999$, and the constant for numerical stability = 1e−8. All weights are initialized using a zero-centered Gaussian distribution with a standard deviation of 0.02. The method was implemented in PyTorch, and the training process ran for a total of 6000 epochs with a batch size of 192 on three NVIDIA Quadro RTX 5000 GPUs.

3 EVALUATION

3.1 Evaluation metrics

To determine the success of each model, we used several metrics to quantitatively measure the relation of each ECG signal (or PPG signals in the case of the *E2P-WGAN model*) to its synthesis.

Pearson's correlation coefficient (ρ) measures how much an ECG signal and its reconstruction co-vary. ρ ranges from -1 to 1, where the magnitude of ρ indicates the strength of the correlation and the sign of ρ determines whether the correlation is positive or negative.

$$\rho = \frac{\left(y_{ECG} - \overline{y}_{ECG}\right)^T \left(y_{rECG} - \overline{y}_{rECG}\right)}{\|y_{ECG} - \overline{y}_{ECG}\|_2 \|y_{rECG} - \overline{y}_{rECG}\|_2} \tag{9}$$

where y_{ECG} and y_{rECG} represent the ECG and the reconstruction respectively. The function $\| * \|_2$ denotes the Euclidean distance.

Root Mean Squared Error (RMSE) measures the differences between the values of an ECG signal and its reconstruction, called errors, and aggregates the magnitudes of these errors. The closer the RMSE is to zero, the more accurate the reconstruction.

$$RMSE = \|y_{ECG} - y_{rECG}\|_2 \tag{10}$$

Fréchet Distance (FD) examines the location and order of points on an ECG waveform and its synthesis as curves in order to measure the similarity of the signals. The closer FD is to zero, the more similar the reference ECG and its synthesis, as well as the more diverse the synthesis will be.

$$FD = \min \left(\max_{i \in Q} \left(d\left(y_{ECG_i}, y_{rECG_i}\right) \right) \right), Q = [1, m] \tag{11}$$

where the function $d(*)$ denotes the Euclidean distance between two corresponding points on the ECG and synthesis curves. FD is the shortest Euclidean distance between any of the corresponding points on the ECG and synthesis curves.

3.2 Experimental models

We developed four synthesis models as follows.

P2E-WGAN-γ model: Our model with $\gamma = 50$ to reconstruct all the samples of the ECG signals, while λ_R and λ_0 are set to 0.

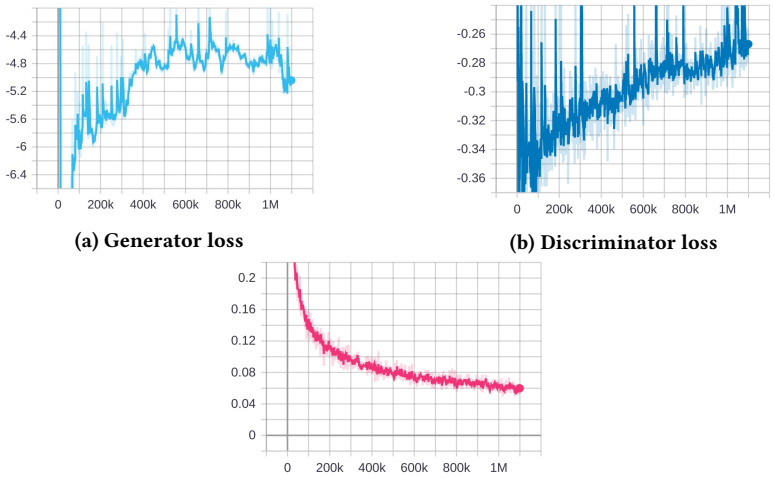
E2P-WGAN-γ model: Similar to *P2E-WGAN-γ* model but translating from ECG to PPG.

P2E-WGAN-λ model: Our model with $\lambda_R = 80$ and $\lambda_0 = 20$ to synthesize ECG signals that remain only important features to serve the data augmentation purpose.

P2E-GAN-λ model: Similar to *P2E-WGAN-λ* but without Wasserstein loss and gradient penalty. We applied the one-sided label smoothing technique, as suggested in [29], as well as introduced noisy labels with a 5% probability in order to stabilize the training of this model.

DCT-CNN model: Based on the work of Zhu *et al.* [9], we implemented a model that learns non-linear mappings from PPG signals to ECG signals in the frequency domain as follows.

The segments are converted from the time domain to the frequency domain through Discrete Cosine Transform (DCT). The first 50 DCT coefficients of the PPG signals and the first 125 DCT coefficients of the ECG signals are retained as these coefficients represent the highest signal energies.



(a) Generator loss

(b) Discriminator loss

(c) Peaks and valleys loss

Figure 4: The dynamic behavior of the Wasserstein loss with gradient penalty when the training optimization process is constrained with important ECG features only.

Each transformed PPG is first fed into a convolutional layer having 16 filters with kernel sizes of 7. The output is then average-pooled before being inputted into another convolutional layer having 32 filters with kernel sizes of 7. A flattening layer transforms the convolutional layer's output into a vector, which is inputted into a multi-layer perceptron (MLP). The MLP consists of two dense hidden layers of sizes 922 and 700 in order, and a dense output layer of size 125. Taken together, these layers predict the coefficients of the corresponding ECG signals. All convolution and dense layers are characterized by the *ReLU* activation function except for the dense output layer. The neural network optimizes the mean square error of the predicted DCT coefficients of the ECG signals to the reference DCT coefficients.

$$W^* = \underset{W}{\operatorname{argmin}} \|X_{PPG}W - X_{ECG}\|_F^2 \tag{12}$$

where W represents the weights of the network that determine the reconstruction of the DCT coefficients of ECG signals, X_{ECG}, from the DCT coefficients of reference PPG signals, X_{PPG}. $\|*\|_F^2$ denotes the Frobenius norm of a matrix.

After the model outputs the predicted DCT coefficients of the ECG signals and subsequently padded, the ECG waveforms are reconstructed by converting from frequency-domain representation into time-domain representation through inverse DCT.

3.3 Results

Table 1 compares the performance of the four models in terms of the means and standard deviations of RMSE, FD, and ρ. The correlation of the signals generated by the *P2E-WGAN-γ* model and the reference signals are statistically strong with the ρ value of 0.835. Also, low values of FD (0.357) and RMSE (0.162) measurements show the strong similarities between them and reference ECG signals.

In the second row of the Table 1, results from the *DCT-CNN* model can be seen. This model is based on the approach in the state-of-the-art [9], and we also enhanced the learning algorithm by using the deep neural network model instead of the linear regression. However, as it can be understood from the table, the *P2E-WGAN-γ* model outperforms it by large margins. That can be explained by the fact that the DCT model summarizes the signal after selecting limited DCT coefficients while in the *P2E-WGAN-γ* model, all of the samples represented in the time domain are used. The reconstructions of ECG from PPG in the related works are cycle-based. Cycle wise segmentation of ECG signals could reduce the accuracy and robustness of the frameworks because of the dependence on cumbersome algorithms for the alignment of PPG and ECG cycles as well as peak and onset point detection. In our work, 3-second segments were used to avoid the mentioned problems and have a more recurrent reconstruction context.

Table 1: Comparison of reconstruction performance of different models - μ and σ are mean and standard deviation, respectively

Models	RMSE		FD		ρ	
	μ	σ	μ	σ	μ	σ
P2E-WGAN-γ	**0.162**	0.059	**0.375**	0.185	**0.835**	0.128
DCT-CNN	0.373	0.121	0.594	0.225	0.711	0.191
E2P-WGAN-γ	*0.113*	0.046	*0.227*	0.118	*0.958*	0.052

Table 2: Comparison of synthesis performance of the two models on the data augmentation task - μ and σ are mean and standard deviation, respectively

Models	RMSE		FD		ρ	
	μ	σ	μ	σ	μ	σ
P2E-WGAN-λ	**0.238**	0.063	**0.582**	0.193	**0.771**	0.132
P2E-GAN-λ	0.329	0.084	0.590	0.174	0.656	0.177

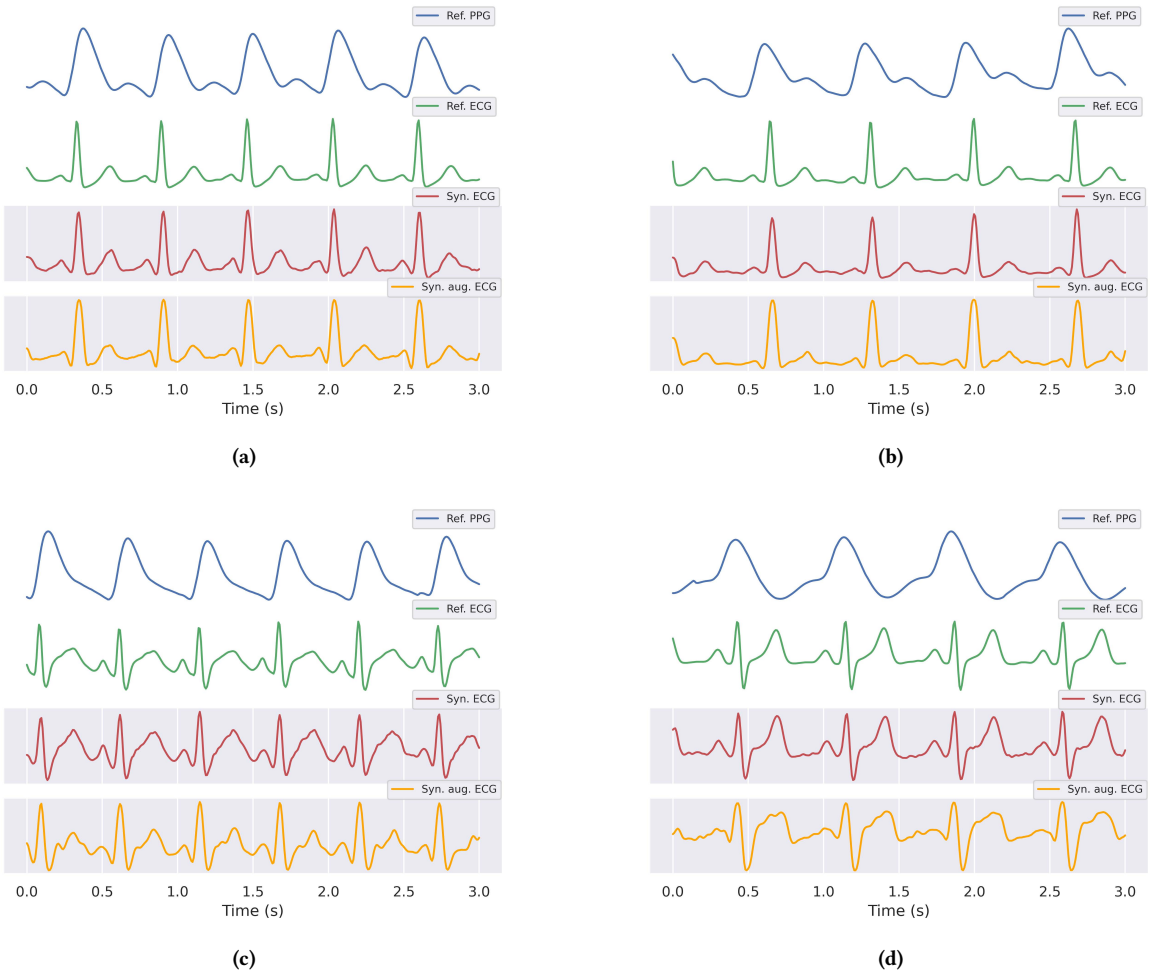


Figure 5: Examples of the synthetic ECG signals. In each example, from top to bottom: reference PPG, reference ECG, synthetic ECG by *P2E-WGAN-γ* model, and synthetic ECG by *P2E-WGAN-λ* model.

The last row of Table 1 is the results of the *E2P-WGAN-γ* model. This model shows better performance comparing to the *P2E-WGAN-γ* model, which can be explained due to the physiological differences between PPG and ECG signals. High-frequency structures are more significant in ECG signals than PPG.

Table 2 compares the performance of the two models for data augmentation. The signals generated by the *P2E-WGAN-λ* model is strongly correlated with the ρ value of 0.771. The low values of FD (0.582) and RMSE (0.238) also support the diversity in the generated signals and strong similarities between them and reference ECG signals. The *P2E-GAN-λ* model performed more poorly than *P2E-WGAN-λ*. We attribute this to the generator being provided with stronger gradients from the discriminator in the *P2E-WGAN-λ* model. Interestingly, although the *P2E-WGAN-λ* model is trained with only peaks and valleys, its reconstruction performance for the entire waveforms is satisfactory. It also successfully approximated the reference ECG distribution. The significance of this work as compared to other works on ECG synthesis for data augmentation is that the resulting ECG signals in our work correspond to input PPG signals, which makes the data more realistic and informative.

In Figure 4, we can observe how the Wasserstein loss function with the gradient penalty evolves over time during the training of the *P2E-WGAN-λ* model. The loss values of both the discriminator and generator are highly stable and convergent. Moreover, this demonstrates that one of the notable benefits of using the Wasserstein loss function is that its convergence reflects the quality of the synthetic samples [18].

In Figure 5, synthetic ECG waveforms are plotted with respect to the reference PPG waveforms. We can see that the models can closely reconstruct various ECG waveforms and maintain their essential features.

4 CONCLUSION AND FUTURE WORKS

In this work, we present, **P2E-WGAN**, a novel conditional GAN-based approach to accurately synthesize ECG waveforms with PPG signals as inputs. The results demonstrate our model's potential

for providing a paradigm shift in telemedicine by bringing ECG-based clinical diagnoses of cardiovascular disease to individuals via simple PPG assessment by wearables. Our model trained in the subject independent mode achieves an average root mean square error of 0.162, Fréchet distance of 0.375, and Pearson's correlation of 0.835 on a normalized real-world dataset, demonstrating the efficacy of our approach. In our future work, we plan to further validate the proposed method with other ECG and PPG datasets that contain noisy PPG signals where the source of noise is from real-life activities. Besides, we aim to validate the generalizability of the model with other cardiac physiological signals as inputs. Our approach holds promise to enable screening and early detection of heart diseases in the home setting, saving cost and labor, and supporting society in unusual pandemic scenarios such as the current COVID-19.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support from the NIH grant R44OD024874 (H.C.).

REFERENCES

[1] Paul Kligfield et al. "Recommendations for the standardization and interpretation of the electrocardiogram: part I: the electrocardiogram and its technology a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society endorsed by the International Society for Computerized Electrocardiology". In: *Journal of the American College of Cardiology* 49.10 (2007), pp. 1109–1127.

[2] Tai Le et al. "Electrocardiogram: Acquisition and Analysis for Biological Investigations and Health Monitoring". In: *Interfacing Bioelectronics and Biomedical Sensing*. Springer, 2020, pp. 117–142.

[3] J. Allen. "Photoplethysmography and its application in clinical physiological measurement". In: *Physiol Meas.* 28.3 (2007), pp. 1–39.

[4] sudden. *Sudden death in young people: Heart problems often blamed*. https://www.mayoclinic.org/diseases-conditions/sudden-cardiac-arrest/in-depth/sudden-death/art-20047571. Retrieved on September 2020.

[5] Anna Rosiek and Krzysztof Leksowski. "The risk factors and prevention of cardiovascular disease: the importance of electrocardiogram in the diagnosis and treatment of acute coronary syndrome". In: *Therapeutics and clinical risk management* 12 (2016), p. 1223.

[6] Andrew Reisner et al. "Utility of the photoplethysmogram in circulatory monitoring". In: *Anesthesiology: The Journal of the American Society of Anesthesiologists* 108.5 (2008), pp. 950–958.

[7] Denisse Castaneda et al. "A review on wearable photoplethysmography sensors and their potential future applications in health care". In: *International journal of biosensors & bioelectronics* 4.4 (2018), p. 195.

[8] Rohan Banerjee et al. "PhotoECG: Photoplethysmographyto estimate ECG parameters". In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2014, pp. 4404–4408.

[9] Qiang Zhu et al. "ECG reconstruction via PPG: A pilot study". In: *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE. 2019, pp. 1–4.

[10] Xin Tian et al. "Cross-Domain Joint Dictionary Learning for ECG Reconstruction from PPG". In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2020, pp. 936–940.

[11] Fei Zhu et al. "Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network". In: *Scientific reports* 9.1 (2019), pp. 1–11.

[12] Tomer Golany and Kira Radinsky. "PGANs: Personalized generative adversarial networks for ECG synthesis to improve patient-specific deep ECG classification". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 2019, pp. 557–564.

[13] Ary L Goldberger et al. "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals". In: *circulation* 101.23 (2000), e215–e220.

[14] Mohamad Kachuee et al. "Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time". In: *2015 IEEE international symposium on circuits and systems (ISCAS)*. IEEE. 2015, pp. 1006–1009.

[15] Ian Goodfellow et al. "Generative adversarial nets". In: *Advances in neural information processing systems*. 2014, pp. 2672–2680.

[16] Phillip Isola et al. "Image-to-image translation with conditional adversarial networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1125–1134.

[17] Nitish Srivastava et al. "Dropout: a simple way to prevent neural networks from overfitting". In: *The journal of machine learning research* 15.1 (2014), pp. 1929–1958.

[18] Martin Arjovsky, Soumith Chintala, and Léon Bottou. "Wasserstein gan". In: *arXiv preprint arXiv:1701.07875* (2017).

[19] Ishaan Gulrajani et al. "Improved training of wasserstein gans". In: *Advances in neural information processing systems*. 2017, pp. 5767–5777.

[20] Patrick S Hamilton and Willis J Tompkins. "Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database". In: *IEEE transactions on biomedical engineering* 12 (1986), pp. 1157–1165.

[21] Pauli Virtanen et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python". In: *Nature methods* 17.3 (2020), pp. 261–272.

[22] Olaf Ronneberger, Philipp Fischer, and Thomas 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.

[23] Maha Elbayad, Laurent Besacier, and Jakob Verbeek. "Pervasive attention: 2d convolutional neural networks for sequence-to-sequence prediction". In: *arXiv preprint arXiv:1808.03867* (2018).

[24] Colin Lea et al. "Temporal convolutional networks: A unified approach to action segmentation". In: *European Conference on Computer Vision*. Springer. 2016, pp. 47–54.

[25] André Araujo, Wade Norris, and Jack Sim. "Computing Receptive Fields of Convolutional Neural Networks". In: *Distill* (2019). https://distill.pub/2019/computing-receptive-fields. DOI: 10.23915/distill.00021.

[26] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. "Instance normalization: The missing ingredient for fast stylization". In: *arXiv preprint arXiv:1607.08022* (2016).

[27] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. "Deep sparse rectifier neural networks". In: *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. 2011, pp. 315–323.

[28] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[29] Tim Salimans et al. "Improved techniques for training gans". In: *Advances in neural information processing systems*. 2016, pp. 2234–2242.