ME-GAN: Learning Panoptic Electrocardio Representations for Multi-view ECG Synthesis Conditioned on Heart Diseases

Jintai Chen [* 1] Kuanlun Liao [* 1] Kun Wei [2] Haochao Ying [3 4] Danny Z. Chen [5] Jian Wu [6]

Abstract

Electrocardiogram (ECG) is a widely used non-invasive diagnostic tool for heart diseases. Many studies have devised ECG analysis models (e.g., classifiers) to assist diagnosis. As an upstream task, researches have built generative models to synthesize ECG data, which are beneficial to providing training samples, privacy protection, and annotation reduction. However, previous generative methods for ECG often neither synthesized multi-view data, nor dealt with heart disease conditions. In this paper, we propose a novel disease-aware **g**enerative **a**dversarial **n**etwork for **m**ulti-view **E**CG synthesis called ME-GAN, which attains panoptic electrocardio representations conditioned on heart diseases and projects the representations onto multiple standard views to yield ECG signals. Since ECG manifestations of heart diseases are often localized in specific waveforms, we propose a new *mixup normalization* to inject disease information precisely into suitable locations. In addition, we propose a *view discriminator* to revert disordered ECG views into a predetermined order, supervising the generator to obtain ECG representing correct view characteristics. Besides, a new metric, *rFID*, is presented to assess the quality of the synthesized ECG signals. Comprehensive experiments verify that our ME-GAN performs well on multi-view ECG signal synthesis with trusty morbid manifestations.

*Equal contribution [1]College of Computer Science and Technology, Zhejiang University, Hangzhou 310058, China; [2]School of Electronic Engineering, Xidian University, Xi'an 710071, China; [3]School of Public Health, Zhejiang University, Hangzhou, China; [4]The Key Laboratory of Intelligent Preventive Medicine of Zhejiang Province, Hangzhou, Zhejiang 310058, China; [5]Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA [6]The Second Affiliated Hospital School of Medicine, School of Public Health, and Institute of Wenzhou, Zhejiang University, Hangzhou 310058, China. Correspondence to: Haochao Ying <haochaoying@zju.edu.cn>.

1. Introduction

Heart diseases are major threats to global health (Abubakar et al., 2015; Roth et al., 2018). Electrocardiogram (ECG), a non-invasive diagnostic tool for heart diseases, is widely-used in clinical practice (Holst et al., 1999), putting a heavy burden on cardiologists. To alleviate this stress, many research efforts were put into developing ECG classifiers (Bian et al., 2022; Golany et al., 2021; Kiranyaz et al., 2015) for automatic ECG analysis, yielding considerable performances on limited tests. But, there were still concerns on developing automatic ECG classifiers, such as patient privacy protection (Hossain et al., 2021; Hazra & Byun, 2020) and absence of well-annotated ECG signals (Golany et al., 2020). Thus, a key upstream task (Golany et al., 2020; Zhang & Babaeizadeh, 2021) is to synthesize ECG signals in order to increase the diversity and quantity of training samples, which can also help reduce annotation needs if ECG signals are synthesized conditioned on heart diseases.

There are two main issues in the known ECG synthesis methods. **(1) Trusty multi-view ECG synthesis.** In clinical practice, ECG signals from different views are clinically useful (Graybiel et al., 1946; Case et al., 1979), representing heartbeat signals from various viewpoints (see Fig. 1(a)). However, previous ECG synthesis models often yielded single-view ECG (Golany & Radinsky, 2019; Golany et al., 2020; Hossain et al., 2021; Hazra & Byun, 2020; Zhu et al., 2019) or independently synthesized different views without explicitly considering view correlations (Kuznetsov et al., 2020; McSharry et al., 2003; Thambawita et al., 2021). Such synthesized ECG signals cannot provide trusty representations of heartbeats, resulting in limited applications. **(2) ECG synthesis conditioned on heart diseases.** Unlike some global conditions used in image synthesis (Karras et al., 2019) (e.g., the gender impacts synthesized face pictures globally), ECG manifestations of some heart diseases are often localized in specific waveforms (Frank, 1956). For example, the *left bundle block abnormality* (*LBBB*) often presents notches on/near R waves (R wave is a typical waveform in ECG signals; see Fig. 1(b)). But, most ECG generative methods either omitted disease information (Hazra & Byun, 2020; Zhu et al., 2019) or synthesized ECG with different heart diseases by developing different models (Zhang

(a) ECG views represent heartbeat signals (i.e., vectorcardiogram) from various viewpoints.

(b) Abnormal waveform with *LBBB*.
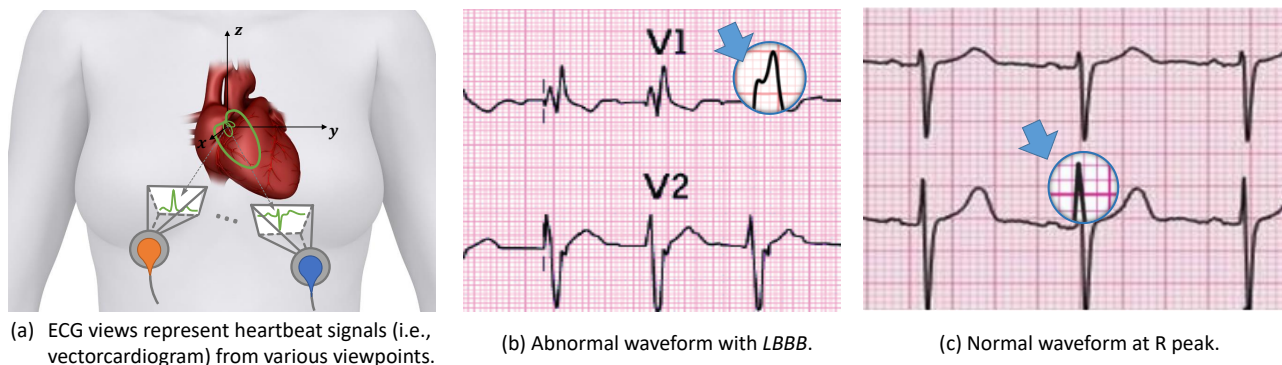
(c) Normal waveform at R peak.

Figure 1. (a) Illustrating the principle of multi-view ECG signal recording by projecting heartbeat signals along different viewpoints. A viewpoint is represented by two angles (θ, ϕ) (the polar and azimuthal angles; defined in Sec. 2). Comparing (b) and (c), the waveform changes caused by diseases can be localized (e.g., *left bundle branch block abnormality* (*LBBB*) can lead to notches on R waves).

& Babaeizadeh, 2021; Golany et al., 2020), which do not address the general heart disease embedding problem.

To address these two issues, this paper develops a novel **GAN** for **M**ulti-view **E**CG synthesis conditioned on heart diseases called ME-GAN, which consists of a generator, a major discriminator to distinguish real ECG and to predict the disease categories, and an auxiliary discriminator called *view discriminator* to ensure the synthesized ECG representing correct view characteristics.

To synthesize highly trusty multi-view ECG signals, we devise the generator following the principle of ECG signal recording. As shown in Fig. 1(a), ECG signals in different views actually record electrocardio signals (representing heartbeats) from different viewpoints (Frank, 1956). A study (Chen et al., 2021) has verified that a neural network can predict ECG views based on electrocardio representations with query viewpoints (represented by angles). Motivated by 3D-aware GANs (Zhou et al., 2021; Tian et al., 2018), our generator performs in two stages, which synthesizes new "stereo" electrocardio representations of heartbeat signals in stage one, and employs the Nef-Net decoder (Chen et al., 2021) to yield the standard ECG views based on the synthesized representations in stage two. To emphasize view correlations among the synthesized views, we randomly shuffle the view order and utilize a *view discriminator* to revert the views into the pre-determined order. Thus, the *view discriminator* supervises the generator to synthesize signals representing correct and differentiable view characteristics.

To inject disease information, we present a novel *mixup normalization* layer in the generator to precisely control the degrees and locations of disease information embedding in the synthesized signals, since morbid manifestations are localized in specific waveforms. Different from the self-modulation (Chen et al., 2019) and AdaIN (Karras et al., 2019) that provided channel-wise normalization with global conditions, our *mixup normalization* seeks the correct lo-

cations and performs length-wise normalization to inject disease information in the synthesized signals.

In addition, we present a pre-trained 1D Inception (Szegedy et al., 2016) with a metric called *relative Fréchet Inception Distance (rFID)* to assess the quality of synthesized multi-view ECG signals. The *rFID* is analogous to the *FID* metric for synthesized images (Heusel et al., 2017).

The main contributions of this paper are as follows.

(A) We propose a novel GAN model called ME-GAN, which is the first model that is able to synthesize standard 12-view ECG signals and simultaneously considers heart diseases. Following the 3D-aware GAN strategy, ME-GAN first synthesizes the representations of heartbeat signals and then projects them into standard views to represent ECG signals. Our design follows the ECG recording principle and essentially ensures representations of ECG views to be consistent.

(B) We propose a novel *mixup normalization* to precisely determine the locations and degrees to inject disease information into the synthesized ECG signals, based on the ECG characteristics of morbid manifestations.

(C) We present a novel *view discriminator*, which reverts randomly shuffled ECG views into a pre-determined order, by processing heavily masked ECG signals. Thus, any locations on the synthesized ECG signals are ensured to represent correct view characteristics.

(D) We introduce a pre-trained 1D Inception with a new metric called *rFID* to assess the quality of synthesized multi-view ECG signals. Comprehensive experiments verify that the synthesized ECG signals by our ME-GAN are highly trusty and are useful for improving ECG classifiers.

2. Backgrounds

Electrocardiogram (ECG) and ECG GANs. Clinical studies (Graybiel et al., 1946; Case et al., 1979; Grant, 1950) indicated that single-view ECG signals cannot provide

enough information for diagnosis and surgery, and ECG views (12 views as default) essentially represent the projections of heartbeat signals along various viewpoints (see Fig. 1(a)). But, most previous ECG synthesis methods (Hossain et al., 2021; Hazra & Byun, 2020; Zhu et al., 2019; Kuznetsov et al., 2020; Thambawita et al., 2021; Kuznetsov et al., 2021) either yielded a single ECG view or independently synthesized multi-view ECG signals following various GAN models (Arjovsky et al., 2017) while omitting view correlations. Besides, most of the known methods (Zhu et al., 2019; Hazra & Byun, 2020) ignored the presence of heart diseases, while only a few (Zhang & Babaeizadeh, 2021; Golany et al., 2020) synthesized ECG signals belonging to different diseases using different models.

A Review of Electrocardio Panorama and Nef-Net. It was empirically proved that multi-view ECG signals are projections of electrocardio signals along corresponding viewpoints (Chen et al., 2021). Researchers often used a spherical coordinate system with the origin point at the central electric terminal of the heart, and the *x*-axis, *y*-axis, and *z*-axis are defined by the anatomical sagittal axis, inverse frontal axis, and vertical axis, respectively. In the spherical coordinate system, the viewpoints of ECG views are represented by two angles, a polar angle θ and an azimuthal angle ϕ. Based on this system, an auto-encoder, Nef-Net, was proposed (Chen et al., 2021) in which the encoder predicts the "stereo" electrocardio representations of a known ECG case and the decoder yields a new ECG view by projecting the representations along the query viewpoints. In this paper, we also utilize the spherical coordinate system (see Fig. 1(a)), represent viewpoints by (θ, ϕ), and use the decoder as a part of our generator. Unlike Nef-Net which synthesizes new ECG views of known cases, our ME-GAN is developed to synthesize multi-view ECG of new cases.

Related Generative Adversarial Networks. The generative adversarial network (GAN) (Goodfellow et al., 2014) was first proposed for image synthesis, and was then applied in many fields (Sutskever et al., 2014; De Cao & Kipf, 2018). Various studies (Sutskever et al., 2014; Ramsundar et al., 2015) showed that the condition information contributed to synthesis effects. Typically, conditions were provided to generators by concatenation with random noises (Mirza & Osindero, 2014). Later, additional supervisions were used in discriminators to promote the effects of condition embedding (Odena et al., 2017; Chen et al., 2016; Odena, 2016), which were widely used in medical data synthesis and obtained competitive performance (Yi et al., 2019). Recently, conditions were also introduced into GANs indirectly, by style transfer (Huang & Belongie, 2017; Karras et al., 2020), linear interpolation (Radford et al., 2015), and latent vector modification (Agrawal et al., 2021). However, these methods did not give strict definitions and bounds of the embedded "conditions", and were not suitable to biomedical fields. Other related work is on multi-view image synthesis. Following Nerf (Mildenhall et al., 2020; Martin-Brualla et al., 2021), GANs employed 3D-aware generators to synthesize 3D representations (Zhou et al., 2021; Tian et al., 2018; Gu et al., 2021; Chan et al., 2021), which were further projected along query viewpoints into images. These methods were verified to be efficient to synthesize multi-view images with consistent 2D representations of 3D objects.

3. Methodology

In this paper, we propose a novel generative adversarial network (GAN) for **M**ulti-view **E**CG synthesis conditioned on heart diseases, called ME-GAN, which is composed of a two-stage generator and two discriminators (a major discriminator and a *view discriminator*). Extending the idea of 3D-aware generators to ECG synthesis, our generator processes Gaussian noises $\mathbf{z} \in \mathbb{R}^d$ to synthesize the "stereo" electrocardio representations conditioned on heart diseases $\mathbf{c} \in \mathbb{R}^k$ (one of k possible heart disease categories represented by a one-hot vector), and then employs the Nef-Net decoder (Chen et al., 2021) to project the synthesized electrocardio representations into multi-view standard ECG signals. The major discriminator distinguishes real and synthesized ECG signals, and predicts the disease categories. The *view discriminator* learns to revert randomly shuffled ECG views into a pre-determined order, which supervises the generator to synthesize ECG views representing proper view characteristics. Below we present the designs of these three key components of ME-GAN respectively.

3.1. Generator

THE OVERALL ARCHITECTURE

The generator architecture of ME-GAN has two stages (see Fig. 2). The first stage synthesizes new "stereo" electrocardio representations using a ladder-shaped model; the second stage synthesizes multi-view ECG signals with a shallow 1D convolutional decoder of Nef-Net which takes the synthesized panoptic electrocardio representations and query viewpoint angles as input. Note that the input branch for deflection representation of the original Nef-Net decoder is not used here. The generator is performed end-to-end.

The ladder-shaped model of the first stage contains two model paths and some skip connections. The major path with 1D convolution blocks (marked in blue) processes and up-samples Gaussian noises $\mathbf{z} \in \mathbb{R}^d$ in steps, which obtains latent representations $\mathbf{z}_i \in \mathbb{R}^{d_i \times l_i}$ at the i-th level (d_i and l_i denote the feature channel number and feature length, respectively). All $\mathbf{z}_i$ were up-sampled to the final resolution and added together for the Nef-Net decoder. The other path (marked in orange) and two additional skip connections (marked by green dashed arrows) jointly yield the location-
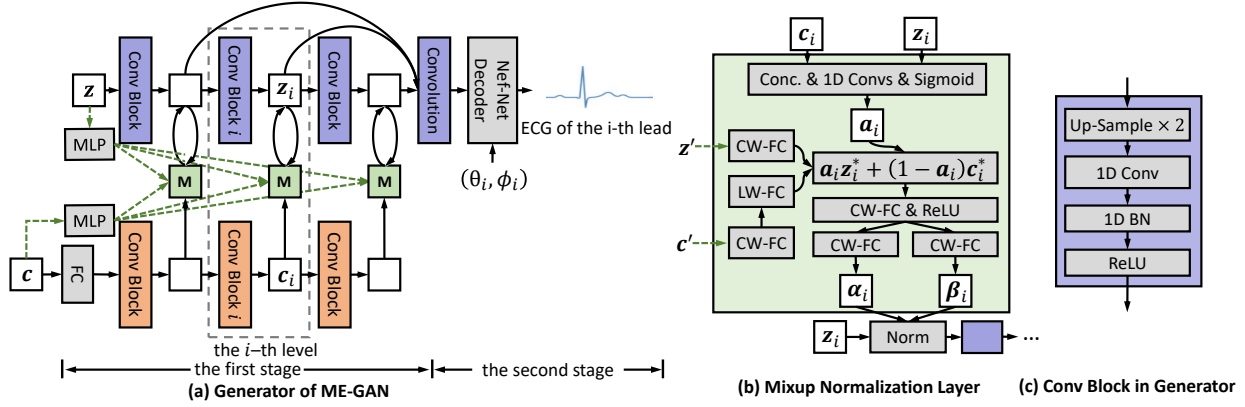
Figure 2. Illustrating the ME-GAN generator (a), in which "M" denotes a *Mixup Normalization* layer as shown in (b) and the design of "Conv Block" is shown in (c). "CW-FC" and "LW-FC" denote a "channel-wise fully connected layer" and a "length-wise fully connected layer", respectively, and "**z**" and "**c**" represent Gaussian noises and disease condition, respectively. Notably, we build four levels in the first stage in our implementation.

aware disease embedding and inject it into $\mathbf{z}_i$ via our proposed *mixup normalization* layers (*MixupNorm*), since the manifestations of heart diseases in ECG signals are localized in specific waveforms. Concretely, a disease condition $\mathbf{c} \in \mathbb{R}^k$ is first embedded into a latent code by a fully connected layer and is further processed and up-sampled by 1D convolution blocks into latent representations $\mathbf{c}_i \in \mathbb{R}^{k_i \times l_i}$ (k_i indicates the channel size of $\mathbf{c}_i$). Apart from this, two multi-layer perceptrons (containing four fully connected layers with ReLU activations in between) respectively transform $\mathbf{z} \in \mathbb{R}^d$ and $\mathbf{c} \in \mathbb{R}^k$ into $\mathbf{z}' \in \mathbb{R}^{d' \times 1}$ and $\mathbf{c}' \in \mathbb{R}^{k' \times 1}$ (d' and k' indicate the channel sizes of $\mathbf{z}'$ and $\mathbf{c}'$, respectively), and feed them into *MixupNorm* layers at all levels (see the green dashed arrows in Fig. 2(a)). *MixupNorm* uses $\mathbf{z}_i$ and $\mathbf{c}_i$ to compute a spatial attention to determine where and to what degree the disease condition (from $\mathbf{c}'$) impacts the representation $\mathbf{z}_i$. As shown in Fig. 2(c), a convolution block in the ladder-shaped model consists of a linear up-sample layer, a 1D convolution, a 1D batch normalization layer (Ioffe & Szegedy, 2015), and a ReLU activation.

Mixup Normalization Operation

Without loss of generality, here we describe the operation of *MixupNorm* at the i-th level (see Fig. 2), since the operations at the other levels are similar. Similar to AdaIN (Karras et al., 2019) and the self-modulation (Chen et al., 2019), we inject condition information into the synthesized latent representation $\mathbf{z}_i$ by normalization operations.

MixupNorm first concatenates the representation $\mathbf{z}_i \in \mathbb{R}^{d_i \times l_i}$ (which conveys the location information of the synthesized waveform) and $\mathbf{c}_i \in \mathbb{R}^{k_i \times l_i}$ (which conveys disease information) into a comprehensive representation of size $(d_i + k_i) \times l_i$, which is processed to attain an attention $\mathbf{a}_i \in \mathbb{R}^{d_a \times l_i}$ (d_a is the channel size of $\mathbf{a}_i$). Intuitively, by fusing the information of the waveform location and disease information, the attention $\mathbf{a}_i$ can specify the degree to in-

ject disease information at each location. Formally, $\mathbf{a}_i$ is computed by:

$$\mathbf{a}_i = \text{sigmoid}(\text{Conv}(\text{ReLU}(\text{BN}(\text{Conv}([\mathbf{z}_i, \mathbf{c}_i]))))), \quad (1)$$

where "$[\cdot, \cdot]$" denotes channel-wise concatenation, and "Conv" and "BN" denote a 1D convolution layer and a 1D batch normalization layer, respectively. On the other hand, in *MixupNorm*, the latent representation $\mathbf{z}'$ is transformed by a channel-wise fully connected layer into $\mathbf{z}_i^* \in \mathbb{R}^{d_h \times 1}$, and $\mathbf{c}'$ is transformed into $\mathbf{c}_i^* \in \mathbb{R}^{d_h \times l_i}$ by a channel-wise fully connected layer and a length-wise fully connected layer (with a ReLU in between), by:

$$\mathbf{z}_i^* = \mathbf{z}'^T U_i \,, \; \mathbf{c}_i^* = (\max(\mathbf{c}' V_i + b_i, 0))^T W_i, \quad (2)$$

where $U_i \in \mathbb{R}^{d' \times d_h}$, $V_i, b_i \in \mathbb{R}^{1 \times l_i}$, and $W_i \in \mathbb{R}^{k' \times d_h}$ are learnable weights, and d_h indicates the channel sizes of $\mathbf{z}_i^*$ and $\mathbf{c}_i^*$. Then, $\mathbf{z}_i^*$ is extended into size $(d_h \times l_i)$ and is incorporated with $\mathbf{c}_i^*$ according to attention $\mathbf{a}_i$, by:

$$\mathbf{h}_i^* = \mathbf{a}_i \mathbf{c}_i^* + (\mathbf{1} - \mathbf{a}_i)\mathbf{z}_i^*, \quad (3)$$

where $\mathbf{h}_i^* \in \mathbb{R}^{d_h \times l_i}$. A channel-wise fully connected layer then processes the representation $\mathbf{h}_i^*$ to synthesize the parameters $\alpha_i, \beta_i \in \mathbb{R}^{1 \times l_i}$ for normalization. Finally, α_i and β_i are applied to the representation $\mathbf{z}_i$ in the major path, by:

$$\widetilde{\mathbf{z}}_i = \alpha_i \odot \frac{\mathbf{z}_i - \mu}{\sigma} + \beta_i, \quad (4)$$

where $\odot$ denotes point-wise multiplication, and $\mu \in \mathbb{R}^{1 \times l_i}$ and $\sigma \in \mathbb{R}^{1 \times l_i}$ are the mean and standard deviation of $\mathbf{z}_i$. The output $\widetilde{\mathbf{z}}_i$ goes forward to the next level.

Discussion. In *MixupNorm*, we use attention (in Eq. (1)) to represent the locations and degrees for disease information injection. Note that $\mathbf{z}_i^* \in \mathbb{R}^{d_h \times 1}$ conveys the intrinsic style of noises similar to the self-modulation (Chen et al., 2019), but $\mathbf{c}_i^* \in \mathbb{R}^{d_h \times l_i}$ represents different values along
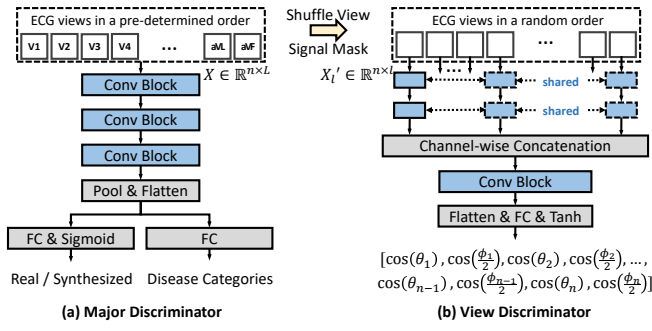
Figure 3. Illustrating the two ME-GAN discriminators. A blue rectangle denotes a convolution block consisting of a 1D convolution layer, an instance normalization, and a leaky ReLU. The input of the major discriminator is a tensor of size $n \times L$ containing all the n ECG views, while the ECG signal input to the view discriminator is views shuffled and highly masked.

the length dimension. Unlike self-modulation, our *Mixup-Norm* focuses on learning different disease embeddings at different locations in synthesized ECG signals explicitly.

3.2. Discriminators

Our ME-GAN employs two discriminators (see Fig. 3(a)). The major discriminator takes ECG signals as input to determine whether they are real signals and their disease categories simultaneously, by employing LSGAN (Mao et al., 2017) with an auxiliary classifier as in ACGAN (Odena et al., 2017). The objective of the major discriminator D and its supervision for generator G are defined by:

$$
\begin{aligned}
\min_D \max_G & \mathcal{V}(D, G) \\
= & \mathbb{E}_{X \sim P_{\text{data}}}[(D_{\text{r/f}}(X) - a)^2 + \text{CE}(D_c(X), \mathbf{c})] \\
& + \mathbb{E}_{X \sim P_{\text{data}}; \mathbf{z} \sim P(\mathbf{z})}[(D_{\text{r/f}}(G(\mathbf{z}|\mathbf{c})) - b)^2 \\
& + \text{BCE}(D_c^{\text{su}}(G(\mathbf{z}|\mathbf{c})), \mathbf{c})],
\end{aligned} \tag{5}
$$

where we set $a = 1$, $b = 0$ as default (Mao et al., 2017). $D_{\text{r/f}}$ indicates the major discriminator with the classifier head to predict whether the input is real, while D_c indicates the major discriminator with the auxiliary classifier for disease classification (the superscript "su" denotes "stop updation"). In Eq. (5), "BCE" denotes the binary cross-entropy loss function (since the heart disease classification is often a multi-class classification task), the tensor $X \in \mathbb{R}^{n \times L}$ contains ECG signals of length L in all the n views sampled from a dataset P_{data}, and $\mathbf{z}$ is sampled from Gaussian distribution $P_{\mathbf{z}}(\mathbf{z})$. The major discriminator is built with three convolution blocks (see Fig. 3), and each block is built with a 1D convolution layer, a 1D instance normalization layer, and a leaky ReLU.

VIEW DISCRIMINATOR

Unlike the major discriminator that takes all the ECG views organized in a pre-determined order as input, the input of

the auxiliary discriminator (*view discriminator*) contains all the ECG views given in a random order. In training, we randomly sample a permutation order of the views, and shuffle the ECG views and the corresponding target viewpoints (angles) according to the random order by the function s_P:

$$
X' = s_P(X) = (X^T P)^T, \ \Theta' = s_P(\Theta) = (\Theta^T P)^T, \tag{6}
$$

where the n views in $X \in \mathbb{R}^{n \times L}$ are organized in a pre-determined order. The cosines of the corresponding viewpoints are loaded in $\Theta \in \mathbb{R}^{n \times 2}$ by $\Theta[p, 1] = \cos(\theta_p) \in (-1, 1]$ and $\Theta[p, 2] = \cos(\phi_p/2) \in (-1, 1]$ ($p = 1, 2, \ldots, n$, the polar angle $\theta \in [0, \pi)$, and the azimuthal angle $\phi \in [0, 2\pi)$). The transition matrix P of size $n \times n$ is a random doubly-stochastic matrix whose element values are 0 or 1. By multiplication with a random P, the orders of views and viewpoints in X and Θ are shuffled into X' and Θ'. In our setting, the task of the *view discriminator* is to take X' as input and predict the corresponding viewpoints Θ'. Since P is random in each training epoch, correct viewpoint order predictions on synthesized ECG signals indicate that the synthesized ECG signals follow proper characteristics of the views. The *view discriminator* employs a sub-network to iteratively process ECG signals in each view and concatenate the obtained representations to predict $2n$ values ranged from -1 to 1 (via a "tanh" operation on top of the model) to estimate the $2n$ values in Θ'. Instead of predicting the permutation matrix P directly, we predict the angle information to avoid iterative processes in attaining the doubly-stochastic matrix P.

An obvious training strategy for the *view discriminator* is to take X' as input and predict Θ' directly. If the *view discriminator* performs well on both real and synthesize ECG cases and obtains similar performances, then it means that the synthesized ECG views contain similar view characteristics as real ECG views. Yet, we need to further consider two issues. (1) The classification strategy can only verify that the view characteristics exist in the ECG views, but cannot prove that all the locations of the ECG signals represent correct view characteristics. (2) View characteristics in different locations vary. To address these two issues, we only feed a piece of ECG signals to the *view discriminator*, and compare the predictions on the synthesized ECG signals and real ECG signals obtained by the *view discriminator*. Formally, a piece of signals of a fixed length l ($l < L$, where L is the length of the original ECG signals) is selected for the *view discriminator* by the function t_ϵ, as:

$$
X'_l = t_\epsilon(X') = X'[:, \epsilon : \epsilon + l], \tag{7}
$$

where ϵ is randomly selected (subject to $\epsilon + l < L$) in each epoch, and $X'_l \in \mathbb{R}^{n \times l}$. In addition, we conduct positional encoding (following the design in (Vaswani et al., 2017)) on X' before it is processed by t_ϵ. In training the *view discriminator*, the *view discriminator* processes real

ECG signals X_l' to predict Θ' under the guidance of mean square error (MSE). Meanwhile, it is required to enlarge the predictions of real and synthesized ECG signals, in order to push the *view discriminator* to seek for some "error view characteristics" in the synthesized ECG views. In training the generator, the generator learns to reduce the prediction gap between real and synthesized ECG signals. The objective function for the *view discriminator* V and its supervision for the generator G are defined by:

$$
\begin{aligned}
&\min_V \max_G \mathcal{V}(V, G) \\
&= \mathbb{E}_{X \sim P_{\text{data}}; \mathbf{z} \sim P_{\mathbf{z}}(\mathbf{z})}[\text{MSE}(V(t_\epsilon(s_P(X))), \Theta') \\
&\quad - \text{MSE}(V(t_\epsilon(s_P(X))|\mathbf{c}), V(t_\epsilon(s_P(G(\mathbf{z}|\mathbf{c})))))].
\end{aligned} \quad (8)
$$

4. Relative Fréchet Inception Distance

To measure the synthesis quality of GANs for ECG, we propose a new metric called *relative Fréchet Inception Distance (rFID)*, inspired by the *Fréchet Inception Distance (FID)* (Heusel et al., 2017). The *FID* metric is used for imaging GANs to compute the distribution distance between a set of real images and a set of synthesized images, whose distributions are typically estimated by a pre-trained 2D Inception (Szegedy et al., 2016). According to this, we need a pre-trained model to estimate the distribution of multi-view ECG signals. A direct solution for this is to train a 1D Inception on a large-scale ECG dataset. However, we found that the open source ECG datasets are not enough to train a strongly generalized model in terms of the amount of data and the number of disease categories. For example, the MIT-BIH arrhythmia dataset has only 4 categories and 2 ECG views, while the Tianchi ECG dataset[1] has 34 categories.

Standard ECG data contain 12 views, and a length of 512 can well represent the waveform details. In this work, we set $n = 12$ and $L = 512$ ($X \in \mathbb{R}^{12 \times 512}$) as default, and use a 1D Inception pre-trained on ImageNet (Deng et al., 2009). Using the idea of Vision Transformer (Dosovitskiy et al., 2020), we resize images into 46×46 and divide each image into 529 patches of size $2 \times 2 \times 3$ (3 is the number of color channels). Thus, an image can be reorganized into size 12×529. Then, we drop the last 17 patches and obtain 12×512. In building the 1D Inception, we change 2D operations of the 2D Inception into the corresponding 1D operations. For a 2D convolution layer, we drop the last dimension of the kernel size (e.g., replace a 7×1 kernel by a 1D kernel of size 7). After pre-training the 1D Inception on the images, we use it to process ECG signals and obtain a latent representation of length 2048 (similar to the design of 2D Inception). To eliminate the incompatibility caused by using a model pre-trained on images to access the distribution of

ECG data, we define a relative version of *FID* (*rFID*) by:

$$
rFID(\{G(\mathbf{z})\}) = \frac{FID(\{G(\mathbf{z}|\mathbf{c})|\mathbf{c} \in C_X\}, \{X\}_2)}{FID(\{X\}_1, \{X\}_2)}, \quad (9)
$$

where the samples in a test set are divided into two groups $\{X\}_1$ and $\{X\}_2$ with identical distributions (e.g., the data amount of each category is identical), and G is a trained GAN generator that synthesizes multi-view ECG signals with $\mathbf{z} \sim P_{\mathbf{z}}(\mathbf{z})$ and query conditions in C_X that is identical to the condition distributions of $\{X\}_1$ and $\{X\}_2$. The denominator $FID(\{X\}_1, \{X\}_2)$ is the *FID* score between $\{X\}_1$ and $\{X\}_2$, which is used to reduce the gap between images and ECG. The effect of *rFID* is verified in Sec. 5.6.

5. Experiments

5.1. Experimental Setup

Datasets. We conduct experiments on the Tianchi ECG dataset and PTB dataset (Bousseljot et al., 1995) to evaluate our proposed ME-GAN. The Tianchi dataset contains 31,779 12-view ECG signals recorded at a frequency of 500 Hertz. The PTB dataset contains 549 12-view ECG signals recorded at a frequency of 1000 Hertz. On each dataset, we randomly take 80% of ECG signals for training and the rest 20% for test. We partition an ECG record into several cardiac cycles (representing heartbeats) using the partition annotations provided in (Chen et al., 2021). Since an official heart disease annotation is for each ECG record containing several cardiac cycles, we use only some heart diseases that theoretically are present on all the cardiac cycles if occurred. Thus, for the Tianchi ECG dataset, we test our ME-GAN on 3 heart disease categories in two configurations: (1) with three disease conditions: the *left axis deviation (LAD)*, the *right axis deviation (RAD)*, and the *right bundle branch block (RBBB)*; (2) without considering diseases. For the PTB dataset, we only test without using the heart disease conditions. Note that we use all the three selected diseases as conditions simultaneously in the tests on the Tianchi ECG dataset. For the configuration (1), we also use all the training data in the training phase, and the ECG data out of the three disease conditions were synthesized with input condition as a vector with zeros. All the cardiac cycles are used as independent samples, which are de-noised with the Python package (Kachuee et al., 2018), interpolated to 500 Hertz, padded to length of 512, and are linearly scaled to 0–1 in the pre-processing. We follow the viewpoint angle definitions given in (Chen et al., 2021) (12 views): $(\theta, \phi) \in \{(\frac{\pi}{2}, \frac{\pi}{2}), (\frac{5\pi}{6}, \frac{\pi}{2}), (\frac{5\pi}{6}, -\frac{\pi}{2}), (\frac{\pi}{3}, -\frac{\pi}{2}), (\frac{\pi}{3}, \frac{\pi}{2}), (\pi, \frac{\pi}{2}), (\frac{\pi}{2}, -\frac{\pi}{18}), (\frac{\pi}{2}, \frac{\pi}{18}), (\frac{19\pi}{36}, \frac{\pi}{12}), (\frac{11\pi}{20}, \frac{\pi}{6}), (\frac{8\pi}{15}, \frac{\pi}{3}), (\frac{8\pi}{15}, \frac{\pi}{2})\}$.

Implementation. Our method is implemented with PyTorch 1.7 on an RTX2080Ti GPU. In training, the batch size is 16. For the generator, major discriminator, and *view discriminator*, we use the learning rate 10^{-4} and the Adam optimizer

---

[1] https://tianchi.aliyun.com/competition/entrance/231754/introduction

Table 1. **Synthesis performances of various GAN models**. The lower *rFID* score is the better, and the accuracy score of 1-NN classifier (1NNC) is better if it is closer to 0.5. The best performances are marked in **bold**.

Methods	Tianchi w/ diseases								Tianchi w/o diseases		PTB w/o diseases	
	overall		LAD		RBBB		RAD		overall		overall	
	1NNC	rFID	1NNC	rFID	1NNC	rFID	1NNC	rFID	1NNC	rFID	1NNC	rFID
WGAN-GP	0.722	7.534	0.697	5.359	0.683	3.509	0.676	6.892	0.640	5.494	0.656	42.046
ACGAN	0.699	15.944	0.668	7.464	0.672	10.942	0.684	13.144	–	–	–	–
LSGAN	0.870	13.427	0.795	6.151	0.676	8.984	0.746	11.367	0.775	17.341	0.673	43.481
CGAN	0.757	12.479	0.585	8.495	**0.607**	5.489	**0.594**	10.849	–	–	–	–
SMDCGAN	0.998	38.619	0.934	23.985	0.823	18.370	0.924	31.408	0.617	2.618	0.623	20.182
BC-GAN	0.832	12.701	0.723	8.299	0.617	5.975	0.634	10.685	0.983	42.545	0.997	158.698
CBL-GAN	0.826	6.990	0.673	5.322	0.739	3.471	0.694	5.573	0.611	6.173	0.713	65.750
ME-GAN (Ours)	**0.643**	**3.722**	**0.567**	**2.662**	0.663	**1.491**	0.712	**3.433**	**0.582**	**2.343**	**0.618**	**15.282**

Table 2. **Classification performances of 1D ResNet-34 trained on augmented training sets.**

Method	Diseases (PR-AUC)			
	LAD	RBBB	RAD	Mean
1D ResNet-34 (baseline)	0.925	0.801	0.911	0.879
WGAN-GP	0.916	0.826	**0.922**	0.889
ACGAN	0.912	0.817	0.913	0.881
LSGAN	0.918	0.858	0.910	0.895
CGAN	0.906	0.867	0.904	0.892
SMDCGAN	0.913	0.832	0.921	0.889
BC-GAN	0.896	0.820	0.898	0.871
CBL-GAN	0.910	0.857	0.896	0.888
ME-GAN (Ours)	**0.927**	**0.870**	0.908	**0.902**

with $\beta_1 = 0.5$, $\beta_2 = 0.999$. In the test without disease conditions, we replace the condition c by a constant vector, and drop the auxiliary classifier in the major discriminator.

Comparison Baselines. We compare our ME-GAN with common GAN methods including conditional GAN (CGAN) (Mirza & Osindero, 2014), WGAN-GP (Gulrajani et al., 2017), ACGAN (Odena et al., 2017), LSGAN (Mao et al., 2017), and self-modulation DCGAN (SMDCGAN) with hinge loss (Chen et al., 2019). Besides, we also compare the performances with state-of-the-art ECG synthesis methods, including BiLSTM-CNN GAN (BC-GAN) (Zhu et al., 2019) and CNN-BiLSTM-GAN (CBL-GAN) (Delaney et al., 2019). We use the GAN implementations in the projects[2,3] with few modifications to fit the data formats and to directly synthesize 12-view ECG signals. For the models that do not specify how conditions are used, we

---

[2] https://github.com/znxlwm/
pytorch-generative-model-collections

[3] https://github.com/MikhailMurashov/
ecgGAN

utilize the same strategy as in conditional GAN (Mirza & Osindero, 2014). For the experimental settings without disease conditions, the input disease conditions and the related classifier (if a model has one) are not used. All the methods are trained in 100,000 iterations.

5.2. Synthesis Performances

To verify the quality of synthesized ECG signals, we report the *rFID* scores and 1-NN classifier accuracy scores (1NNC) of various GAN methods in Table 1. One can see that the *rFID* scores of our ME-GAN are much lower than those of the other methods on the two datasets, no matter with or without disease conditions. Specifically, comparing the performances of SMDCGAN and our method, ME-GAN outperforms SMDCGAN just a little if the disease conditions are not used (on both the Tianchi and PTB datasets). But, our method outperforms SMDCGAN with a clear margin if ECG syntheses are conditioned on the heart diseases. This might be because our proposed *MixNorm* can better deal with the heart disease conditions than the self-modulation. We can also obtain the similar conclusion on 1-NN classifier accuracy scores.

Further, we evaluate the synthesis quality with classification performances (based on PR-AUC, i.e., the area under the precision-recall curve). We train 1D ResNet-34 models on the training sets augmented by the synthesized ECG signals, respectively. In training, we add synthesized samples to the original training set to double the size of the training data belonging to those three disease categories. The classification performances are reported in Table 2. One can see that, compared to the performances obtained by an 1D ResNet-34 trained on the original training set (see the third row in Table 2, mean PR-AUC = 0.879), synthesized ECG signals promote the classification performances except for BiLSTM-CNN GAN. Among them, the synthesized ECG signals provided by our ME-GAN promote the classification

Table 3. **Pairwise distance comparison of view signals within the same ECG cases.** For all the methods, 1000 samples were synthesized for the average pairwise distance computing.

Methods	Real ECG (baseline)	WGAN-GP	AC-GAN	LSGAN	CGAN	SMDCGAN	BC-GAN	CBL-GAN	ME-GAN
Distances	0.259	0.327	0.482	0.435	0.423	0.843	0.405	0.355	**0.286**

Table 4. **Ablation studies for the key components of our proposed ME-GAN.**

	View Discriminator	Condition Injection	Auxiliary Classifier	Two-Stage Generator	PR-AUC	*rFID*
(1)	✓	*MixNorm*	✓	✓	0.902	3.722
(2)	✓	Self-modulation	✓	✓	0.858	5.432
(3)		*MixNorm*	✓	✓	0.867	5.212
(4)	✓	*MixNorm*		✓	0.895	4.580
(5)	✓	*MixNorm*	✓		0.884	4.585

performances by the most clear margins.

5.3. Multi-view Consistency Validation

To examine the multi-view consistency of synthesized ECG views, we compute and compare the pairwise distances of signals within an ECG case. We first train an 1D ResNet-34 (without the final fully connection layer) on the Tianchi ECG dataset with disease conditions supervised by the triplet loss (Schroff et al., 2015), which independently processes each ECG signal of a view to attain an embedding, and learns to shorten the pairwise distances of the embedding from the same ECG case while to enlarge those distances from different cases. After training, the 1D ResNet is utilized to process each ECG signal on a view independently and obtains the corresponding embedding. We compute and report the average pairwise distances within ECG cases as in Table 3. The average pairwise distance computed on 1000 randomly selected cases from test sets is 0.259, while our ME-GAN obtains 0.286 which is pretty close to real ECG data (0.259). The performances of other methods were worse than ME-GAN, suggesting that ME-GAN can synthesize ECG signals whose waveforms on individual views were more consistent.

5.4. Ablation Studies

We conduct ablation studies to examine the effects of the key components of our ME-GAN on the Tianchi ECG dataset (with disease conditions), including the *MixNorm*, *view discriminator*, the auxiliary classifier used in the major discriminator, and the two-stage generator (with Nef-Net decoder). We evaluate *MixNorm* by replacing it with the self-modulation (Chen et al., 2019), and evaluate the two-stage generator by removing the Nef-Net decoder and using only the first stage part to synthesize 12 ECG views directly. The performances (based on *rFID* and PR-AUC) are reported in Table 4, and row (1) is ME-GAN with all the three components. It is evident that all the components provide performance gains for ME-GAN, and the *view discriminator*, two-stage generator design, and *MixNorm* are critically

important. Besides, one can see that the performance ranks according to the mean PR-AUC and *rFID* are the same, which suggests that the proposed *rFID* is reliable.

5.5. Multi-view ECG Synthesis Visualization

To intuitively show the synthesized ECG quality of ME-GAN, we visualize two cases of 12-view ECG signals conditioned on *RBBB*, synthesized by ME-GAN and WGAN-GP for comparison (see Fig. 4). It can be seen that our ME-GAN can synthesize smooth ECG signals, which represent clear waveforms (e.g., P, Q, R, S, T waves, as labeled by "view=I" in Fig. 4(a)). In contrast, the conditional WGAN-GP does not synthesize clear waveforms, with noisy signals. Similar performances are seen by the other methods. There are two key manifestations of *RBBB* in ECG signals: (1) a notch around R wave in the "V2" view; (2) the S waves in the "V6" and "I" views are of great duration, representing "obtuse" wave shapes (see Fig. 4(a)). The synthesized ECG signals by our ME-GAN represent the manifestations (marked by orange circles) of both (1) and (2). While the ECG signals in the "V2" view synthesized by WGAN-GP (in Fig. 4(b)) represent a notch around R wave (satisfying (1)), the S waves in the "V1" and "V6" views are of short duration (representing "narrow" and "small" S waves), which do not satisfy (2). Besides, we find that the signals in different views synthesized by conditional WGAN-GP look similar (e.g., in the "V4", "V5", and "V6" views), but similar phenomena do not occur in the synthesized ECG signals by our ME-GAN. We think that this might be because our two-stage major generator can learn the panoptic electrocardio representations for better multi-view ECG signal synthesis.

5.6. Effects of the Metric *rFID*

Since the computation of *rFID* in Eq. (9) needs to divide the test set into two parts ($\{X\}_1, \{X\}_2$) with identical distribution to calculate the denominator (i.e., $FID(\{X\}_1, \{X\}_2)$), we repeat computing the denominator with 10 different random divisions. We find that the values thus obtained are stable, with a standard deviation of 0.843. The same com-
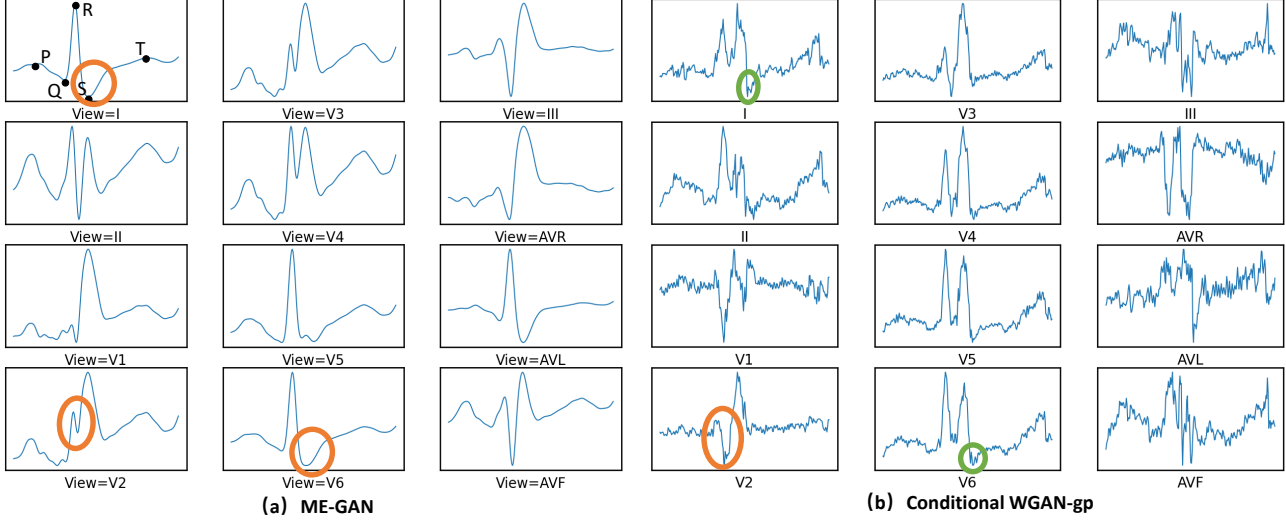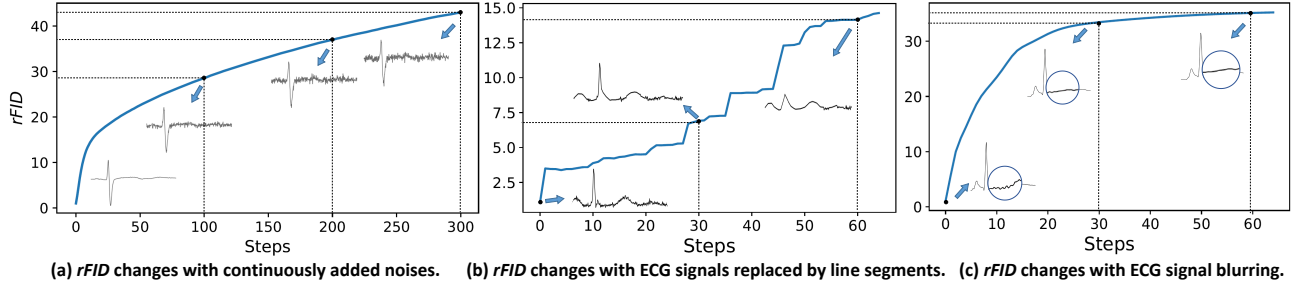
Figure 4. **Visualization comparison of synthesized 12-view ECG signals by ME-GAN and conditional WGAN-GP, conditioned on the *RBBB* disease**. Orange circles mark some correct morbid manifestations of *RBBB*, while green circles mark error manifestations.

(a) *rFID* changes with continuously added noises.   (b) *rFID* changes with ECG signals replaced by line segments.   (c) *rFID* changes with ECG signal blurring.

Figure 5. **Illustrating *rFID* changes during a process in which noise is continually added onto ECG signals.** The "V2" view of an ECG signal is presented for visually showing the signal quality.

putation on the PTB dataset yields a standard deviation of 0.212. This suggests that the denominator of *rFID* is stable for a given dataset, and hence it can contribute to the consistency evaluation of ECG syntheses. Further, we observe the changes of $rFID(\{X_{\text{train}}\})$, when we use some normalized real ECG data to replace $G(z)$ in Eq. (9) and continually add various data perturbation. We test *rFID* with three perturbation settings: in each step, (a) we add noise $\varepsilon \sim U(0, 0.0002)$ onto the real ECG signals (normalized into 0–1); (b) we randomly erase 30 sequential sampling points on the ECG signal curves, and connect the onset and end breakpoints with a line segment; (c) we blur the ECG signals with an average blurring filter with 2×2 kernel. As shown in Fig. 5, *rFID* scores increase with the perturbations continually added on, which indicates that *rFID* is sensitive in assessing the quality of ECG signals.

6. Conclusions

In this paper, we proposed a novel disease-aware GAN, ME-GAN, which learns to obtain new panoptic electrocardio representations for multi-view ECG synthesis, applying the idea of 3D-aware generator to ECG synthesis and ensuring representation consistency among synthesized views. To

make the synthesized ECG signals trusty, we also developed a novel *view discriminator* which pushes synthesized views to represent proper view characteristics. A novel *Mixup Normalization* layer injects disease information into synthesized waveforms adaptively, which contributes to synthesizing correct morbid manifestations. Besides, to directly assess the quality of synthesized ECG signals, we presented a pre-trained 1D Inception with a new metric *rFID*. Comprehensive experiments showed that our designs are effective in multi-view ECG synthesis conditioned on heart diseases.

7. Acknowledgements

References

Abubakar, I., Tillmann, T., and Banerjee, A. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: A systematic analysis for the global burden of disease study 2013. *The Lancet*, 2015.

Agrawal, S., Venkitachalam, S., et al. Directional GAN: A novel conditioning strategy for generative networks. *ArXiv Preprint ArXiv:2105.05712*, 2021.

Arjovsky, M., Chintala, S., and Bottou, L. Wasserstein generative adversarial networks. In *ICML*, 2017.

Bian, Y., Chen, J., et al. Identifying electrocardiogram abnormalities using a handcrafted-rule-enhanced neural network. *TCBB*, 2022.

Bousseljot, R., Kreiseler, D., and Schnabel, A. Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet. *Biomedical Engineering / Biomedizinische Technik*, 1995.

Case, R. B. et al. A sequential angular lead presentation. *Journal of Electrocardiology*, 1979.

Chan, E. R., Monteiro, M., et al. Pi-GAN: Periodic implicit generative adversarial networks for 3D-aware image synthesis. In *CVPR*, 2021.

Chen, J., Zheng, X., et al. Electrocardio panorama: Synthesizing new ECG views with self-supervision. In *IJCAI*, 2021.

Chen, T., Lucic, M., et al. On self modulation for generative adversarial networks. In *ICLR*, 2019.

Chen, X., Duan, Y., et al. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. In *NeurIPS*, 2016.

De Cao, N. and Kipf, T. MolGAN: An implicit generative model for small molecular graphs. *ArXiv Preprint ArXiv:1805.11973*, 2018.

Delaney, A. M., Brophy, E., et al. Synthesis of realistic ECG using generative adversarial networks. *ArXiv Preprint ArXiv:1909.09150*, 2019.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Li, F.-F. ImageNet: A large-scale hierarchical image database. In *CVPR*, 2009.

Dosovitskiy, A., Beyer, L., et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2020.

Frank, E. An accurate, clinically practical system for spatial vectorcardiography. *Circulation*, 1956.

Golany, T. and Radinsky, K. PGANs: Personalized generative adversarial networks for ECG synthesis to improve patient-specific deep ECG classification. In *AAAI*, 2019.

Golany, T., Radinsky, K., and Freedman, D. SimGANs: Simulator-based generative adversarial networks for ECG synthesis to improve deep ECG classification. In *ICML*, 2020.

Golany, T., Radinsky, K., et al. 12-lead ECG reconstruction via Koopman operators. In *ICML*, 2021.

Goodfellow, I., Pouget-Abadie, J., et al. Generative adversarial nets. *NeurIPS*, 2014.

Grant, R. P. Spatial vector electrocardiography: A method for calculating the spatial electrical vectors of the heart from conventional leads. *Circulation*, 1950.

Graybiel, A., White, P. D., et al. *Electrocardiography in practice*. Saunders, 1946.

Gu, J., Liu, L., et al. StyleNerf: A style-based 3D-aware generator for high-resolution image synthesis. *ArXiv Preprint ArXiv:2110.08985*, 2021.

Gulrajani, I., Ahmed, F., et al. Improved training of wasserstein GANs. *ArXiv Preprint ArXiv:1704.00028*, 2017.

Hazra, D. and Byun, Y.-C. SynSigGAN: Generative adversarial networks for synthetic biomedical signal generation. *Biology*, 2020.

Heusel, M., Ramsauer, H., et al. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. *NeurIPS*, 2017.

Holst, H., Ohlsson, M., et al. A confident decision support system for interpreting electrocardiograms. *Clinical Physiology*, 1999.

Hossain, K. F., Kamran, S. A., et al. ECG-Adv-GAN: Detecting ECG adversarial examples with conditional generative adversarial networks. *ArXiv Preprint ArXiv:2107.07677*, 2021.

Huang, X. and Belongie, S. Arbitrary style transfer in real-time with adaptive instance normalization. In *ICCV*, 2017.

Ioffe, S. and Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, 2015.

Kachuee, M., Fazeli, S., and Sarrafzadeh, M. ECG heartbeat classification: A deep transferable representation. In *ICHI*, 2018.

Karras, T., Laine, S., and Aila, T. A style-based generator architecture for generative adversarial networks. In *CVPR*, 2019.

Karras, T., Laine, S., et al. Analyzing and improving the image quality of StyleGAN. In *CVPR*, 2020.

Kiranyaz, S., Ince, T., and Gabbouj, M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, 2015.

Kuznetsov, V., Moskalenko, V., and Zolotykh, N. Y. Electrocardiogram generation and feature extraction using a variational autoencoder. *ArXiv Preprint ArXiv:2002.00254*, 2020.

Kuznetsov, V., Moskalenko, V., et al. Interpretable feature generation in ECG using a variational autoencoder. *Frontiers in Genetics*, 2021.

Mao, X., Li, Q., et al. Least squares generative adversarial networks. In *ICCV*, 2017.

Martin-Brualla, R., Radwan, N., et al. NeRF in the wild: Neural radiance fields for unconstrained photo collections. In *CVPR*, 2021.

McSharry, P. E., Clifford, G. D., et al. A dynamical model for generating synthetic electrocardiogram signals. *IEEE Transactions on Biomedical Engineering*, 2003.

Mildenhall, B., Srinivasan, P. P., et al. NeRF: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020.

Mirza, M. and Osindero, S. Conditional generative adversarial nets. *ArXiv Preprint ArXiv:1411.1784*, 2014.

Odena, A. Semi-supervised learning with generative adversarial networks. *ArXiv Preprint ArXiv:1606.01583*, 2016.

Odena, A., Olah, C., and Shlens, J. Conditional image synthesis with auxiliary classifier GANs. In *ICML*, 2017.

Radford, A., Metz, L., and Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *ArXiv Preprint ArXiv:1511.06434*, 2015.

Ramsundar, B., Kearnes, S., et al. Massively multi-task networks for drug discovery. *ArXiv Preprint ArXiv:1502.02072*, 2015.

Roth, G. A., Abate, D., et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: A systematic analysis for the global burden of disease study 2017. *The Lancet*, 2018.

Schroff, F., Kalenichenko, D., and Philbin, J. FaceNet: A unified embedding for face recognition and clustering. In *CVPR*, 2015.

Sutskever, I., Vinyals, O., and Le, Q. V. Sequence to sequence learning with neural networks. In *NeurIPS*, 2014.

Szegedy, C., Vanhoucke, V., et al. Rethinking the Inception architecture for computer vision. In *CVPR*, 2016.

Thambawita, V., Isaksen, J. L., et al. DeepFake electrocardiograms using generative adversarial networks are the beginning of the end for privacy issues in medicine. *Scientific Reports*, 2021.

Tian, Y., Peng, X., et al. CR-GAN: Learning complete representations for multi-view generation. *ArXiv Preprint ArXiv:1806.11191*, 2018.

Vaswani, A., Shazeer, N., Parmar, N., et al. Attention is all you need. *NeurIPS*, 2017.

Yi, X., Walia, E., and Babyn, P. Generative adversarial network in medical imaging: A review. *Medical Image Analysis*, 2019.

Zhang, Y.-H. and Babaeizadeh, S. Synthesis of standard 12-lead electrocardiograms using two dimensional generative adversarial network. *ArXiv Preprint ArXiv:2106.03701*, 2021.

Zhou, P., Xie, L., et al. CIPS-3D: A 3D-aware generator of GANs based on conditionally-independent pixel synthesis. *ArXiv Preprint ArXiv:2110.09788*, 2021.

Zhu, F., Ye, F., et al. Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network. *Scientific Reports*, 2019.