

Electrocardio Panorama: Synthesizing New ECG Views with Self-supervision

Jintai Chen¹, Xiangshang Zheng¹, Hongyun Yu¹, Danny Z. Chen², Jian Wu^{3*}

¹College of Computer Science and Technology, Zhejiang University, Hangzhou, China

²Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, USA

³The First Affiliated Hospital, and Department of Public Health, Zhejiang University School of Medicine, Hangzhou, China

jtchen721@gmail.com, {xszheng,yuhongyun777,wujian2000}@zju.edu.cn, dchen@nd.edu

Abstract

Multi-lead electrocardiogram (ECG) provides clinical information of heartbeats from several fixed viewpoints determined by the lead positioning. However, it is often not satisfactory to visualize ECG signals in these fixed and limited views, as some clinically useful information is represented only from a few specific ECG viewpoints. For the first time, we propose a new concept, Electrocardio Panorama, which allows visualizing ECG signals from any queried viewpoints. To build Electrocardio Panorama, we assume that an underlying electrocardio field exists, representing locations, magnitudes, and directions of ECG signals. We present a Neural electrocardio field Network (Nef-Net), which first predicts the electrocardio field representation by using a sparse set of one or few input ECG views and then synthesizes Electrocardio Panorama based on the predicted representations. Specially, to better disentangle electrocardio field information from viewpoint biases, a new *Angular Encoding* is proposed to process viewpoint angles. Also, we propose a self-supervised learning approach called *Standin Learning*, which helps model the electrocardio field without direct supervision. Further, with very few modifications, Nef-Net can also synthesize ECG signals from scratch. Experiments verify that our Nef-Net performs well on Electrocardio Panorama synthesis, and outperforms the previous work on the auxiliary tasks (ECG view transformation and ECG synthesis from scratch). The codes and the division labels of cardiac cycles and ECG deflections on Tianchi ECG and PTB datasets are available at <https://github.com/WhatAShot/Electrocardio-Panorama>.

1 Introduction

Electrocardiogram (ECG) has been instrumental in saving millions of lives since it emerged. Multi-lead ECG visualizes heartbeat signals in several views (see Fig. 1(a)), and the corresponding viewpoints are physically determined by

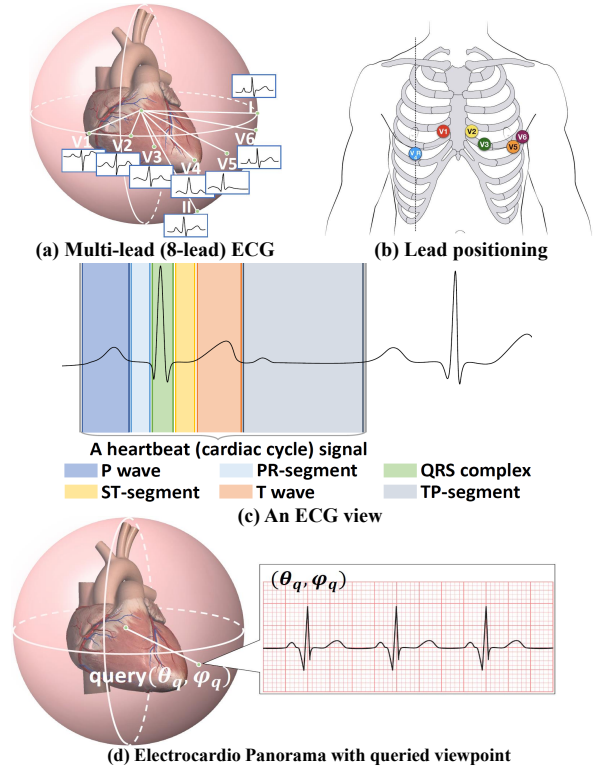


Figure 1: (a) A conventional multi-lead (e.g., 8-lead) ECG represents multi-view ECG signals from viewpoints determined by the lead positioning. (b) Illustrating the conventional lead positioning. (c) A continuous ECG signal can be divided into several cardiac cycles, and each cardiac cycle has 6 types of non-overlapping deflections, which are colored for better viewing. (d) Electrocardio Panorama can provide ECG signals from any queried viewpoints.

the lead positioning (Fig. 1(b)¹). In the past years, ECG visualization approaches have continued to evolve quickly. On one hand, more clinical information was recorded as the number of views increased from 3 to 8, 12, and even 18, with each view recording the signals (see Fig. 1(c)) from one specific viewpoint. For example, the increase of viewpoints helps determine the exact origin of the arrhythmia locating

*The corresponding author.

¹From <https://litfl.com/right-ventricular-infarction-ecg-library/>.

accuracy for the radiofrequency catheter ablation (RFCA) surgery. On the other hand, the visualization form has become increasingly more read-friendly [Graybiel *et al.*, 1946; Case *et al.*, 1979]. Still, current ECG visualization approaches do not meet all the needs, as some clinical information is only intuitively represented in certain specific views (possibly outside of the conventional views).

In this paper, we propose a novel concept, Electrocardio Panorama, which allows real-time querying of any ECG views, as illustrated in Fig. 1(d). Conventional ECG viewpoints are physically determined by the leads, while Electrocardio Panorama seeks to decouple the ECG views from the leads. Such panoramic representation increases the value of ECG monitoring, and thus is potentially life-saving.

Some literature assumed an ECG view as the recording of an underlying electrocardio field from the corresponding viewpoint [Grant, 1950], as illustrated in Fig. 1(a). Following this assumption, it is possible to synthesize Electrocardio Panorama based on the underlying electrocardio field. Motivated by NeRF [Mildenhall *et al.*, 2020], we present a new neural network, called **Neural electrocardio field Network** (Nef-Net), which encodes a sparse set of one or few input ECG views to predict the electrocardio field representation, and a reverse model (decoder) utilizes this representation to synthesize new ECG views. Specifically, we model the electrocardio field by using a basic representation and a set of deflection representations for separately modeling 6 ECG signal deflections (such 6 deflections are shown in Fig. 1(c)), as ECG deflections contain comprehensive personalized or morbid information. Since there is no direct supervision for electrocardio field modeling, we propose a novel *Angular Encoding* approach to process viewpoint information (represented by angles), and a novel self-supervised learning approach called *Standin Learning* to help disentangle electrocardio field information from viewpoint biases.

Motivated by representation interpolations [Karras *et al.*, 2019], our Nef-Net can also be used to synthesize ECG signals from scratch by mixing the extracted electrocardio field representations without explicitly modeling the latent distributions. This work has five major **contributions**:

(A) To our best knowledge, we are the first to propose the concept of Electrocardio Panorama, which decouples ECG signal recordings (views) from physical lead positioning and can provide ECG signals from any desired viewpoints.

(B) We present a new *Angular Encoding* to construct a high-dimensional angle space, which is helpful to disentangling electrocardio field information from viewpoint biases.

(C) We propose a novel self-supervised learning approach, *Standin Learning*, for electrocardio field information purification, which we verify to be effective.

(D) Our proposed Nef-Net can synthesize the Electrocardio Panorama using sparse input ECG views, and can from scratch synthesize multi-lead ECG signals.

(E) We publish cardiac cycle division and deflection division labels (e.g., see Fig. 1(c)) for two large open ECG datasets, contributing to the ECG research and community.

2 Background and Related Work

2.1 Electrocardiogram (ECG)

ECG signals record the heartbeats in multiple views, which are determined by the physical lead positioning [Anderson *et al.*, 1994] (see Fig. 1(a)-(b)). Currently, the most widely used ECG recording method is the standard 12-lead system, utilizing 12 ECG views (see Fig. 1(a)). A signal in a ECG view can be divided into several cardiac cycles (each represents a heartbeat), and each cardiac cycle signal can be divided into 6 non-overlapping deflections (containing rich personalized and morbid information): P wave, PR-segment, QRS complex, ST-segment, T wave, and TP-segment (see Fig. 1(c)). Besides, there is a U wave in the TP-segment, but it is often ignored in practice due to its minimal deflection.

2.2 ECG Visualization

Grant [1950] indicated that the P wave and QRS complex were clinically useful and suggested to interpret ECG signals based on them. Interestingly, doctors summarized the correlations of various ECG views and proposed many visualization methods for multi-lead ECG in [Graybiel *et al.*, 1946; Case *et al.*, 1979; Anderson *et al.*, 1994] based on their ECG reading preferences. These studies [Case *et al.*, 1979; Anderson *et al.*, 1994] suggested that the limited and fixed ECG views were far from satisfactory, as each doctor has personal habit in ECG reading. Also, Anderson *et al.* [1994] pointed out that some clinical information was more evident in some of the views. Hence, we propose a new concept, Electrocardio Panorama, which allows to provide ECG signals in any views.

2.3 ECG Synthesis and View Transformation

Some prior work [Golany *et al.*, 2020a; Golany *et al.*, 2020b] employed conditional generative adversarial networks (conditional GANs) [Mirza and Osindero, 2014] to synthesize personalized or morbid ECG signals. But, these methods could only synthesize signals of one view and did not consider the correlations among various views, thus giving only limited assistance in clinical practice. In [Lee *et al.*, 2019], V-lead ECG signals (“V” is a kind of ECG lead) were synthesized using ECG signals recorded by other leads. Earlier work [Edenbrandt and Pahlm, 1988a; Kors *et al.*, 1990] explored ECG view transformation between the Frank lead [Frank, 1956] and standard lead system², and was meticulously compared in [dos Santos Silva *et al.*, 2020]. Recently, some neural network based regression models [Sohn *et al.*, 2020; Matyschik *et al.*, 2020] were proposed for ECG view transformation. However, these methods were confined to the known ECG view transformation and could not synthesize new ECG views.

3 Neural Electrocardio Field Network

3.1 Architecture

Here we introduce the proposed **Neural electrocardio field Network** (Nef-Net). Nef-Net is based on an encoder-decoder

²The Frank is another ECG systems.

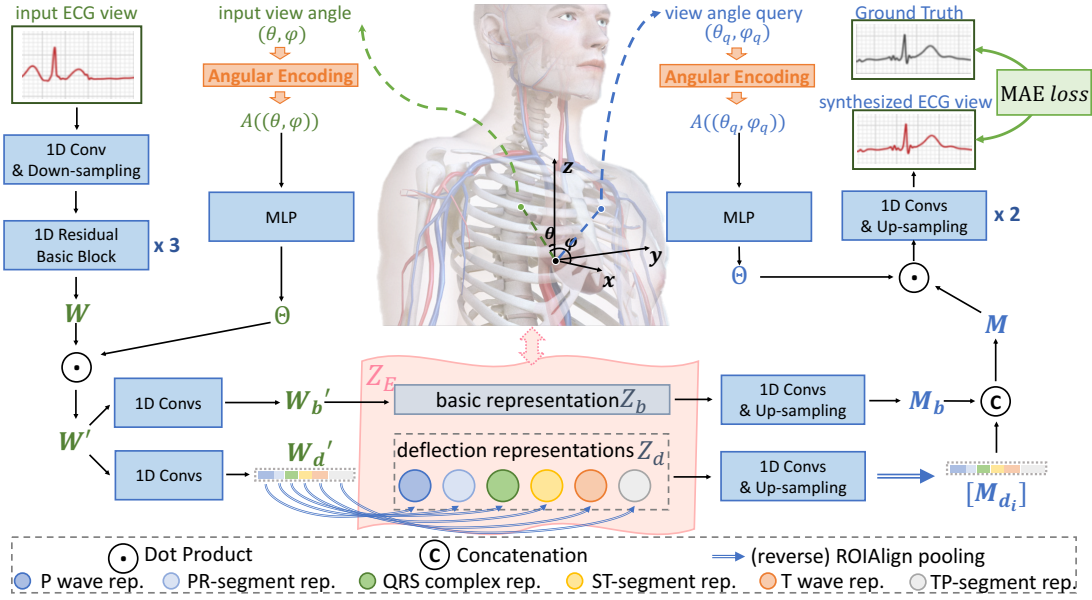


Figure 2: **Our proposed Nef-Net architecture for Electrocardio Panorama synthesis (illustrated using one input ECG view).** Nef-Net extracts the electrocardio field representation Z_E from the input ECG views, and synthesizes ECG signals of a new view based on the learned representation Z_E and the queried viewpoint (θ_q, φ_q) . In the legend, “rep.” means “representation”.

architecture, as shown in Fig. 2. In **encoding**, Nef-Net uses ECG signals in one or few views with the corresponding viewpoints to predict the electrocardio field representation; in **synthesizing**, a reverse model (decoder) uses this representation to predict new ECG views conditioned on the queried viewpoints. To explore viewpoint information, a new **Angular Encoding** is used in both the encoding and synthesizing processes.

Input Structure

In our paper, a tuple containing one cardiac cycle view and its viewpoint is treated as one sample. We build a spherical coordinate system, with its origin point at the central electric terminal, the anatomical sagittal axis as the x -axis, the inverse frontal axis as the y -axis, and the vertical axis as the z -axis (see Fig. 2). Note that the radius is viewed as independent of ECG signals. Thus, we define a viewpoint by a polar angle θ and an azimuthal angle φ . In the following, we first depict the processing of a single input ECG view and then multiple input views. A single-view signal is denoted by x of length t (in time dimension) and the viewpoint is denoted by (θ, φ) . A cardiac cycle signal with L views is denoted by $\{x^{(l)}\}$ ($l \in \{1, 2, \dots, L\}$). In the decoder, a new view is synthesized conditioned on a queried viewpoint (θ^q, φ^q) .

Encoding

Given a cardiac cycle signal x and the corresponding viewpoint (θ, φ) , we first extract a time-aligned feature volume $W = f(x)$ by a 1D convolution module f , and extract a high dimensional angle feature $\Theta = g(A(\theta, \varphi))$ by a multi-layer perceptron (MLP) g and a new **Angular Encoding** process A , where $W \in \mathbb{R}^{c \times t'}$, $\Theta \in \mathbb{R}^c$, c is the channel size, and t' is the time dimension size. Then, we remove the viewpoint information from the feature volume W by $W'_{(c \times t')}$ =

$W_{(c \times t')} \odot \Theta'_{(c \times t')}$, where $\Theta'_{(c \times t')}$ is the stacking of Θ by t' times and \odot is the dot product.

Since the deflections contain rich diagnostic features, we learn the electrocardio field representation, Z_E , using a basic representation Z_b and a set of deflection representations Z_{d_i} ($i \in \{1, 2, 3, 4, 5, 6\}$ indexing the deflection types). The basic representation Z_b learns the global information of the underlying electrocardio field, and each deflection representation Z_{d_i} separately models each type of deflection. 1D convolution modules are then used to project W' to predict W'_b and W'_d for Z_b and Z_d , respectively. We project features of the i -th deflection d_i on W'_d into the representation space Z_{d_i} using ROIAlign pooling [He *et al.*, 2017], as the lengths of the same type of deflections vary among cardiac cycle signals but the size of Z_{d_i} is fixed. The range (in t' -dimension) of the deflection d_i 's features on W'_d is denoted by $\tau_i - \tau_{i-1}$, where $\tau_i = \frac{T_w}{T_x} \times D_i$, and D_i is the demarcation point of the $(i-1)$ -th and i -th deflections in the original signal x (D_i is given). T_x, T_w are the lengths in time-dimension of x and W'_d , respectively. Specifically, $D_0 = 0, D_6 = T_x$, and τ_i is a value as in [Girshick, 2015] that varies among cardiac cycles.

Synthesizing

New view synthesizing can be regarded as a reverse process of encoding, using Z_E to predict ECG views. Each deflection representation Z_{d_i} is first processed by a 1D convolution module (not shared among deflection representations) to obtain features M'_{d_i} , and then is projected into features M_{d_i} to restore the length proportions of the deflections using a reverse ROIAlign pooling (reROI):

$$M_d[:, \tau_{i-1} : \tau_i] = M_{d_i} = \text{reROI}(M'_{d_i}) \quad (1)$$

where τ_i is computed as in encoding, and the exact grid value in M_d is computed by linear interpolation like ROIAlign. We

set M_d and M_b (from Z_b) to have an identical size. Then M_b and M_d are stacked along the channel dimension to form M . A queried viewpoint is processed by *Angular Encoding* and an MLP, and is then multiplied by M to add the queried viewpoint information. In this way, the features are processed to predict new views (as shown in Fig. 2).

Angular Encoding

As mentioned above, *Angular Encoding* is used to handle the viewpoint angles in both encoding and decoding. Nef-Net learns to remove from or add viewpoint information to the electrocardio field (see Fig. 2), but there is no direct supervision for the electrocardio field representation Z_E . Hence, ultimately, extracting viewpoint information is vital. It is intuitive to treat the viewpoint (angles) as a condition, and an ECG signal from a viewpoint (θ, φ) is defined as $x = h(Z_E | (\theta, \varphi))$, where h is a function (e.g., a decoder). Previous conditional GANs fed such conditions to a model directly. But, directly feeding raw angles may ignore the spatial relationships and thus result in poor performance.

We propose a new *Angular Encoding* to process viewpoint angles. We define a triplet as $\pi(\rho) = [\rho, \sin(\rho), \cos(\rho)]$ for an angle ρ . We define the *Angular Encoding* A by:

$$A((\theta, \varphi)) = A((\tilde{\theta}, \tilde{\varphi})) = [\pi(\tilde{\theta}), \pi(\tilde{\varphi}), \pi(\tilde{\theta} + \tilde{\varphi}), \pi(\tilde{\theta} - \tilde{\varphi})] \quad (2)$$

where $\tilde{\rho} = \rho + \epsilon$ and ϵ is a random variable sampled from a Gaussian distribution $\mathcal{N}(0, \frac{\pi}{50})$. *Angular Encoding* transforms two angles into a 12-element tensor, which is then processed by an MLP. Since we treat the recorded ECG views as projections of the electrocardio field, some spatial projection related expressions (e.g., $\sin(\theta)\cos(\varphi)$) should be considered. However, it is hard to fit such expressions (especially multiplication) by using a finite MLP with raw angles, as an MLP can be regarded as an affine transformation if we ignore the activations. With $\sin(\theta \pm \varphi)$ and $\cos(\theta \pm \varphi)$, it is possible to represent the spatial projection related expressions via sum-to-product formulas. Thus, various relationships of angles (e.g., addition, cosine, multiplication) can be expressed with the *Angular Encoding* and a finite MLP. We also add a perturbation ϵ to help avoid over-fitting the given angles, as Nef-Net is used to synthesize new views (of unknown patterns).

3.2 Multi-View ECG Signal Processing

Given a cardiac cycle signal $\{x^{(l)}\}$ ($l \in \{1, 2, \dots, L\}$) with L views, we separately deal with these ECG views in parallel by group convolutions, instead of fusing them together as in the previous work [Kachuee *et al.*, 2018; Chen *et al.*, 2020b]. Formally, we predict L basic representations and L groups of deflection representations, and then compute the averaged basic representations Z_b and averaged deflection representations Z_d over the L views as the input of the decoder.

3.3 Learning

To synthesize Electrocardio Panorama, we train Nef-Net under ECG view transformation supervision and use it for new view synthesis. A set of ECG data is partitioned for training and testing, and all kinds of views are divided into three

groups, called the *input group* (IG), *reconstruction group* (RG), and *synthesis group* (SG). In training, Nef-Net takes the views in IG of an ECG in the training set to reconstruct its views in RG, under the guidance of the Mean Absolute Error (MAE). In testing, Nef-Net synthesizes views in SG of an ECG in the testing set by using its views in IG.

Standin Learning

As mentioned above, Nef-Net is trained to synthesize new ECG views based on the electrocardio field. However, there is no direct supervision for the electrocardio field representation learning, and the input and output are related to the viewpoints. To deal with this, we perform self-supervision tricks [Wei *et al.*, 2019; Feng *et al.*, 2020; He *et al.*, 2020] to eliminate the biases. To capture information from multiple views, a previous self-supervision work [Tian *et al.*, 2019] gathered view representations and implicitly built a scene embedding, by iteratively anchoring a view as the temporary target. But, this is ineffective as each view is somewhat biased and the intrinsic information is not directly extracted and fully used.

We propose a new self-supervised learning approach called *Standin Learning* for electrocardio field prediction, anchoring at the averaged representations as temporary targets to neutralize viewpoint biases. Given a cardiac cycle signal with L views, $\{x^{(l)}\}$ ($l \in \{1, 2, \dots, L\}$), the corresponding basic representations and deflection representations are $\{Z_b^{(l)}\}$ and $\{Z_d^{(l)}\}$, and $Z_b = \sum_{l=1}^L Z_b^{(l)} / L$ and $Z_d = \sum_{l=1}^L Z_d^{(l)} / L$ are the averaged basic representation and averaged deflection representations, respectively. We define the contrastive loss for *Standin Loss* $\mathcal{L}_Z = \frac{1}{2}(\mathcal{L}_{Z_b} + \mathcal{L}_{Z_d})$, where \mathcal{L}_{Z_b} and \mathcal{L}_{Z_d} are defined as:

$$\begin{cases} \mathcal{L}_{Z_b} = \sum_{l=1}^L \mathcal{L}_{\text{MAE}}(h_{\text{sg}}(Z_b, Z_d | (\theta_q, \varphi_q)), h(Z_b^{(l)}, Z_d | (\theta_q, \varphi_q))) / L \\ \mathcal{L}_{Z_d} = \sum_{l=1}^L \mathcal{L}_{\text{MAE}}(h_{\text{sg}}(Z_b, Z_d | (\theta_q, \varphi_q)), h(Z_b, Z_d^{(l)} | (\theta_q, \varphi_q))) / L \end{cases} \quad (3)$$

where h denotes the decoder and “sg” means “stop gradient”. “ $\mathcal{L}_{\text{MAE}}(\text{target}, \text{prediction})$ ” denotes the Mean Absolute Error. \mathcal{L}_Z pushes the reconstruction based on the basic and deflection representations of one view close to that based on the averaged representations (fusing information of multiple views). We optimize the output rather than the representations to avoid changing the representation distribution and causing information loss. Hence, the electrocardio field representations are optimized iteratively: better single view representations improve the averaged representations, which further guide the single view representations to learn better.

Losses

We supervise the Nef-Net training using the reconstruction loss MAE (\mathcal{L}_{MAE}) and contrastive loss \mathcal{L}_Z as:

$$\mathcal{L} = \mathcal{L}_{\text{MAE}}(x_q, \hat{x}_q) + \mathcal{L}_Z \quad (4)$$

where x_q and $\hat{x}_q = h(Z_b, Z_d | (\theta_q, \varphi_q))$ indicate the ground truth and predicted ECG signals of the queried view.

4 Multi-Lead ECG Synthesis from Scratch

Our Nef-Net can also be utilized in ECG synthesis from scratch, without directly modeling the distributions as Generative Adversarial Nets (GAN) [Goodfellow *et al.*, 2014]

and Variational Auto-Encoder (VAE) [Kingma and Welling, 2013; Rezende *et al.*, 2014]. Since ECG signal features are subtle, a simple representation distribution assumption (e.g., Gaussian distribution as in VAE) or a discriminator-guided distribution may not be suitable.

Motivated by representation interpolations [Karras *et al.*, 2019] and Mixup [Zhang *et al.*, 2018; Chen *et al.*, 2020b; Chen *et al.*, 2020a], we synthesize electrocardio field representations of a new ECG by mixing the extracted electrocardio field representations of the same categories (e.g., diseases). As the range of each deflection type varies and the features of different types of deflections should not be mixed for new data synthesis, for this task we train a Nef-Net such that the electrocardio field representation is only predicted by the deflection representation learning branch (i.e., removing the branch for the basic representation). After training on a dataset, we introduce a memory bank \mathbf{B} to store all the averaged deflection representations Z_d for inference. Given the electrocardio field representations $Z_d^{(p)}, Z_d^{(q)} \in \mathbf{B}$ of randomly selected cardiac cycles p and q , the representation of a new cardiac cycle n is:

$$Z_d^{(n)} = aZ_d^{(p)} + (1 - a)Z_d^{(q)} \quad (5)$$

where $a \sim \text{Beta}(1.0, 1.0)$ as default. The new electrocardio field representation is then fed to the trained decoder h_{sg} to synthesize new ECG data. As the deflection lengths T_{d_i} of p and q may be different (used to compute τ_i for reverse ROIAlign pooling (Eq. 1)), we use the weighted average length for n , with $T_{d_i}^{(n)} = a \times T_{d_i}^{(p)} + (1 - a) \times T_{d_i}^{(q)}$.

5 Experiments

5.1 Questions

We conduct evaluations for five major questions in experiments. **(A)** Can Nef-Net synthesize Electrocardio Panorama, especially the ECG views that Nef-Net has not seen before? **(B)** As Nef-Net is trained with the view transformation supervision, can it outperform previous work in view transformation? **(C)** Are our proposed approaches (e.g., *Angular Encoding*) helpful? **(D)** Is Nef-Net good at synthesizing ECG signals from scratch? **(E)** The model complexity of Nef-Net and its potential value in ECG monitoring?

5.2 Datasets

We conduct experiments using the MIT-BIH dataset [Moody *et al.*, 2001], PTB dataset [Bousseljot *et al.*, 1995], and Tianchi ECG dataset³. The MIT-BIH dataset contains 48 half-hour ECG signals recorded at a frequency of 360 Hertz, and each ECG signal is divided into several cardiac cycles (heartbeats). Each signal has 2 views, and the disease ground truth for each cardiac cycle is provided. The PTB dataset contains 549 12-lead ECG signals recorded at a frequency of 1,000 Hertz. The Tianchi dataset contains 31,779 12-lead ECG signals recorded at a frequency of 500 Hertz. Different from the MIT-BIH dataset, the PTB and Tianchi ECG

datasets provide signals in 12 views. However, both these datasets do not provide the cardiac cycle division and the disease ground truth for cardiac cycles, and we annotate and publish the cardiac cycle division and deflection division ground truth for the PTB and Tianchi ECG datasets. For MIT-BIH, we divide the deflections using the package [Makowski *et al.*, 2021]. The PTB and Tianchi datasets are randomly partitioned into a training set and a test set with probabilities 0.8 and 0.2, respectively. The training-test partition for the MIT-BIH dataset follows [Golany *et al.*, 2020a], with 51,020 training samples and 49,711 test samples. We perform pre-processing, including de-noising (using the package [Makowski *et al.*, 2021]), interpolation to 500 Hertz, and linear scaling normalization to 0–1. To our best knowledge, no prior work is known on angle quantification for conventional ECG leads. Thus, we measure angles on 30 body CTs and compute the average angles for 12 leads (in lead-angle pair form), as: I- $(\frac{\pi}{2}, \frac{\pi}{2})$, II- $(\frac{5\pi}{6}, \frac{\pi}{2})$, III- $(\frac{5\pi}{6}, -\frac{\pi}{2})$, aVR- $(\frac{\pi}{3}, -\frac{\pi}{2})$, aVL- $(\frac{\pi}{3}, \frac{\pi}{2})$, aVF- $(\pi, \frac{\pi}{2})$, V1- $(\frac{\pi}{2}, -\frac{\pi}{18})$, V2- $(\frac{\pi}{2}, \frac{\pi}{18})$, V3- $(\frac{\pi}{2}, \frac{\pi}{18})$, V4- $(\frac{11\pi}{20}, \frac{\pi}{6})$, V5- $(\frac{8\pi}{15}, \frac{\pi}{3})$, V6- $(\frac{8\pi}{15}, \frac{\pi}{2})$.

5.3 Experimental Setups

We use PyTorch 1.7.1 to implement Nef-Net. In training, the batch size is 32. Nef-Net is run 150 epochs in training. The learning rate is initialized to 0.1, and is reduced by $10\times$ at the 50-th and 100-th epoch. We use SGD as the optimizer with momentum 0.9. We compute the structural similarity index measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR) for evaluating the performances of new view synthesis and view transformation. Though SSIM is originally for 2D images, we regard ECG signal as a special case of images. We report the means and standard deviations over 3 runs with an RTX2080Ti GPU for all the experiments.

5.4 Performances

View Transformation Performances

To examine the view transformation performances, we compare our Nef-Net with the known state-of-the-art models: conventional ECG lead transformation methods, including the Kors inverse matrix method (KIM) [Kors *et al.*, 1990], Kors quasi-orthogonal method (KQO) [Kors *et al.*, 1990], and Dower inverse matrix method [Edenbrandt and Pahlm, 1988b]; neural network based methods, including ensemble LSTM (E-LSTM) [Sohn *et al.*, 2020] and VAE-CNN [Matyschik *et al.*, 2020]. The Dower, KQO, and KIM methods were originally for Frank vectorcardiogram reconstruction. We use them for view transformation with the procedure that we first transform the standard ECG signals to Frank vectorcardiogram, and then transform the vectorcardiogram back to the standard ECG signals. For fair comparison, the types of views used for input and the transformation (reconstruction) follow the previous work, and the numbers of views are orderly listed in the parentheses. As shown in Table 1, it is clear that our Nef-Net significantly outperforms the previous work on view transformation tasks, **by about 6–10 in PSNR and about 0.05–0.2 in SSIM.**

³<https://tianchi.aliyun.com/competition/entrance/2317544/information?lang=en-us>

Methods		Tianchi		PTB	
		PSNR	SSIM	PSNR	SSIM
Dower	(12, 12)	21.88	0.837	19.99	0.844
Nef-Net (ours)	(12, 12)	33.84 \pm 0.30	0.973	30.74 \pm 0.18	0.983
KQO	(3, 12)	21.55	0.776	19.86	0.810
Nef-Net (ours)	(3, 12)	34.25 \pm 0.31	0.972	31.53 \pm 0.33	0.975
KIM	(8, 12)	22.30	0.835	21.91	0.862
Nef-Net (ours)	(8, 12)	33.24 \pm 0.29	0.973	30.46 \pm 0.32	0.982
VAE-CNN	(1, 11)	26.40 \pm 0.21	0.918	23.99 \pm 0.09	0.912
Nef-Net (ours)	(1, 11)	32.94 \pm 0.37	0.968	30.82 \pm 0.07	0.972
E-LSTM	(3, 9)	22.88 \pm 0.33	0.830	20.20 \pm 0.14	0.821
Nef-Net (ours)	(3, 9)	32.95 \pm 0.22	0.971	30.48 \pm 0.22	0.971

Table 1: The performances (mean \pm std) for view transformation tasks on the Tianchi and PTB datasets. The conventional methods (Dower, KQO, and KIM) did not show std. We report std only for PSNR, since the std of SSIM is too small. The numbers of views for input and transformation (reconstruction) are orderly listed in the parentheses, and the better performances are marked in **bold**.

Group		Tianchi		PTB	
		PSNR	SSIM	PSNR	SSIM
(1) Nef-Net	(1, 9, <u>2</u>)	29.60 \pm 0.35	0.949	28.95 \pm 0.65	0.951
	(1, 9+ <u>2</u> , 0)	31.74 \pm 0.21	0.953	29.77 \pm 0.06	0.952
(2) Nef-Net	(3, 7, <u>2</u>)	29.19 \pm 0.19	0.943	29.27 \pm 0.26	0.955
	(3, 7+ <u>2</u> , 0)	31.99 \pm 0.37	0.953	30.54 \pm 0.24	0.954
(3) Nef-Net	(5, 5, <u>2</u>)	30.77 \pm 0.25	0.950	29.78 \pm 0.40	0.958
	(5, 5+ <u>2</u> , 0)	32.57 \pm 0.27	0.964	30.62 \pm 0.37	0.960
(4) Nef-Net	(1, 10, <u>1</u>)	29.97 \pm 0.36	0.938	29.31 \pm 0.05	0.954
	(1, 10+ <u>1</u> , 0)	32.14 \pm 0.03	0.956	30.68 \pm 0.07	0.954
(5) Nef-Net	(3, 8, <u>1</u>)	30.51 \pm 0.45	0.946	29.43 \pm 0.25	0.951
	(3, 8+ <u>1</u> , 0)	32.24 \pm 0.23	0.968	30.89 \pm 0.22	0.961
(6) Nef-Net	(5, 6, <u>1</u>)	30.58 \pm 0.35	0.949	29.53 \pm 0.44	0.959
	(5, 6+ <u>1</u> , 0)	32.35 \pm 0.12	0.974	31.28 \pm 0.34	0.962

Table 2: The performances for view synthesis on the Tianchi and PTB datasets. The numbers of views for input, transformation (reconstruction), and synthesis are orderly listed in parentheses. The part of views used to compute SSIM and PSNR is underlined.

New View Synthesis Performances

Since we are the first to propose the Electrocardio Panorama, we compare the synthesized results with the transformed results of Nef-Net on the same types of views. As there are a large amount of combinations, we can only report some cases for references in Table 2. In Group (1) of Table 2, we train two Nef-Nets: one is trained with the 1 view input and with the 9 views as supervision, and synthesizes 2 views in testing; the other one is trained with the 1 input view, but with all the rest (11 views) as supervision, and predicts the corresponding 2 views (in test sets) in inference. The other groups are run following similar ways. We compare SSIM and PSNR on the 2 synthesized views of the test samples, and the key difference of these two models is whether the 2 types of views for comparison are used in training. We use only 1 or 2 views for synthesis, making the supervision similar. As shown in Table 2, the qualities of the synthesized views are comparable to the transformed views (in PSNR and SSIM). These results suggest that our synthesized views have high authenticity.

Besides, comparing among the group of (1), (2), and (3), and among the group of (4), (5), and (6), one can see that with more input views, the synthesis performances often get better. Also, one may note that with one input view (groups (1) and

(4)), the synthesis performances of Nef-Net are comparable to its transformation performances, and largely outperform the transformations of previous work (see Table 1). Thus, our Nef-Net is potentially useful in dynamic ECG applications (possibly with only one lead).

To further verify the authenticity of the synthesized views, we also present one case of panoramic representations where the viewpoint angles θ and φ are evenly distributed in the angle ranges with an interval $\frac{\pi}{3}$, as shown in Fig. 3. The synthesized ECG views are though to be highly trusty.

Ablation Study

Based on the benchmark Nef-Net (4, 4, 4) (denoted by (a) in Table 3), we evaluate the effects of our proposed approaches, including (b) *Angular Encoding* (AE) and the perturbation ϵ used in AE, (c) deflection modeling by independent representations Z_d , and (d) contrastive losses \mathcal{L}_{Z_b} and \mathcal{L}_{Z_d} . Besides, to verify (e) the effectiveness of using averaged representations as the temporary targets for the contrastive loss, we use \mathcal{L}'_{Z_b} and \mathcal{L}'_{Z_d} to replace \mathcal{L}_{Z_b} and \mathcal{L}_{Z_d} in Eq. (3), respectively. For \mathcal{L}'_{Z_b} , we replace the averaged basic representation Z_b in \mathcal{L}_{Z_b} with the basic representation of a random view (sampled once per iteration). For \mathcal{L}'_{Z_d} , we replace the averaged deflection representations Z_d in \mathcal{L}_{Z_d} with the deflection rep-

	ϵ	AE	Z_d	\mathcal{L}_{Z_b}	\mathcal{L}_{Z_d}	\mathcal{L}'_{Z_b}	\mathcal{L}'_{Z_d}	synthesis		transformation	
								PSNR	SSIM	PSNR	SSIM
(a)	✓	✓	✓	✓	✓			29.34	0.957	32.32	0.964
(b)		✓	✓	✓	✓			29.18	0.945	31.90	0.955
	✓		✓	✓	✓			28.44	0.941	32.35	0.964
(c)	✓	✓		✓				28.64	0.946	31.77	0.959
(d)	✓	✓	✓	✓				29.16	0.946	32.31	0.956
	✓	✓	✓		✓			28.65	0.943	31.13	0.961
	✓	✓	✓					28.50	0.945	30.00	0.959
(e)	✓	✓	✓		✓	✓		27.02	0.936	27.20	0.936
	✓	✓	✓	✓			✓	28.91	0.934	32.07	0.948
	✓	✓	✓			✓	✓	25.97	0.922	26.44	0.929

Table 3: Ablation study using the Tianchi ECG dataset.

Diseases	VGAN	DCGAN	SimVGAN	SimDCGAN	Nef-Net (ours)
SVEB	0.768±0.06	0.685±0.04	0.765±0.07	0.724±0.05	0.782±0.05
VEB	0.982±0.01	0.981±0.00	0.979±0.01	0.980±0.01	0.983±0.00
FUSION	0.811±0.08	0.823±0.09	0.778±0.05	0.827±0.09	0.855±0.07

 Table 4: Classification performances (ROC-AUC) on the MIT-BIH dataset augmented by synthesized data. The best results are in **bold**.

representations of a random view.

As shown in Table 3, all our proposed approaches are helpful in synthesis. One may notice that the models without *Angular Encoding* or deflection representations can attain comparable performances for the view transformation tasks, but are not effective for new view synthesis. The results suggest that our *Angular Encoding* and deflection modeling can help enhance generalization of models. As shown in (d) and (e) of Table 3, \mathcal{L}_Z (especially \mathcal{L}_{Z_b}) and *Standin Learning* help Nef-Net considerably on both the view transformation and view synthesis tasks.

ECG Synthesis from Scratch

To evaluate Nef-Net’s capability to synthesize ECG signals from scratch, we compare its classification performances (with ROC-AUC) on morbid ECG signal synthesis with the state-of-the-art methods, simDCGAN and simVGAN [Golany *et al.*, 2020b], DCGAN [Radford *et al.*, 2015], and vanilla GAN [Goodfellow *et al.*, 2014], on the MIT-BIH dataset. The performances of these known methods are obtained by running the open source [Golany *et al.*, 2020b], and the classifier for evaluating the synthesized samples is implemented as in [Kachuee *et al.*, 2018]. All the other settings follow [Golany *et al.*, 2020b] for comparison. As shown in Table 4, our Nef-Net outperforms the previous work, our method outperforms the previous work by clear margins on those performance-unsaturated diseases: SVEB (1.4%) and FUSION (2.8%). For VEB, since the performances were close to 100% in the previous work, there is not much room for improvement. Such performances reveal that our Nef-Net captures the key information of ECG.

A visualization of multi-lead ECG data synthesis from scratch by Nef-Net is given in Fig. 4, which is considered as having a high degree of authenticity. As shown in Fig. 4, Nef-Net can synthesize multi-lead ECG signals from scratch, presenting very detailed waves and segments (e.g., the key waves are marked in the top-left sub-figure of Fig. 4(a)),

while SimGAN can synthesize only single-lead ECG whose key waves are lost (e.g., P waves and T waves; see Fig. 4(b)). Also, the signals synthesized by SimGAN have considerable noise (e.g., as highlighted in the orange rectangles), while those synthesized by ours are nearly noise-free. The three II-lead cases in Fig. 4(b) of SimGAN are actually different but they look similar, which shows that SimGAN might cause the mode collapses. In contrast, the synthesized ECG data of our Nef-Net are of diversity.

Model Complexity

In inference, the fps (frames per second) of our Nef-Net is 13.9, which can be used for real-time Electrocardio Panorama synthesis. The Nef-Net model size is 7.18 MB, and is feasible to be used on some mobile devices. Thus, Nef-Net is potentially helpful to medical doctors (e.g., for dynamic ECG (holter ECG) and home-ECG monitoring).

6 Conclusions and Future Work

In this paper, we proposed a new concept, Electrocardio Panorama, which decouples the ECG views from the physical lead positioning and can provide ECG signals of any queried views in real time. We presented a versatile neural network called Nef-Net, which can transform ECG views, synthesize Electrocardio Panorama, and synthesize ECG signals from scratch. Experiments verified that Nef-Net performs well on these tasks. To extract the electrocardio field representations without supervision, we proposed a new *Angular Encoding* for angle processing and an effective self-supervised learning approach (*Standin Learning*) to help extract and purify electrocardio field representations. Ablation study showed that our new approaches are helpful. Finally, the cardiac cycle and deflection division labels for the PTB and Tianchi ECG datasets are available for reference.

There are some potential improvements for future research on Electrocardio Panorama. First, ECG signals are currently

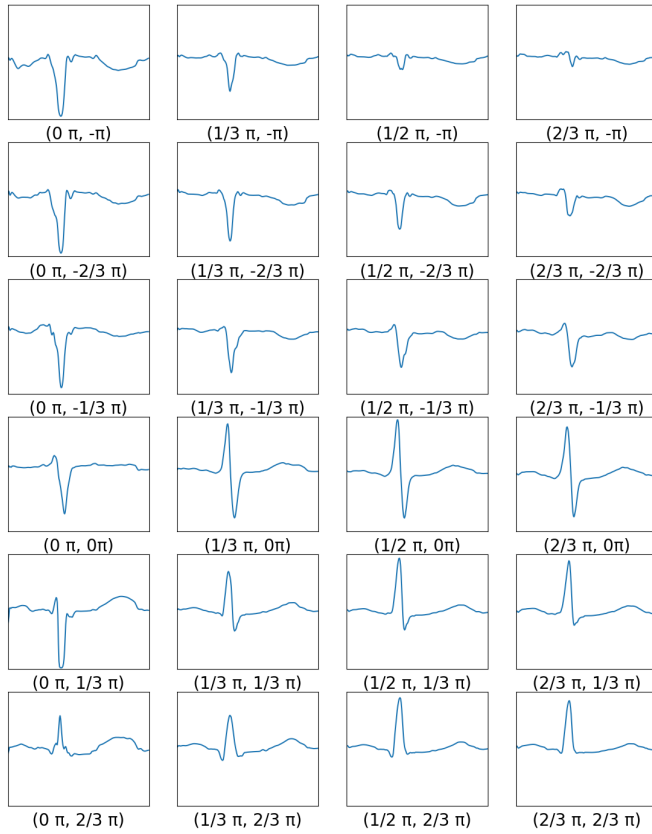


Figure 3: A case of Electrocardio Panorama.

normalized before feeding into Nef-Net. However, normalization might neglect absolute amplitudes, which are important in practice. Second, an interesting issue is to consider ECG synthesis from scratch. Our Nef-Net can synthesize ECG data from scratch and significantly outperform the previous work. It is worth exploring to improve the GAN and VAE methods for ECG by incorporating Nef-Net. Third, although the datasets used in our study contain some ECG signals of patients with heart diseases, we did not explicitly utilize some specific operations for processing morbid ECG signals. Thus, it is important to investigate improving Electrocardio Panorama with respect to morbid ECG signals. Finally, developing a modern visualization system for Electrocardio Panorama is still an open research topic.

Acknowledgments

This research was partially supported by the National Key Research and Development Program of China under grant No. 2019YFC0118802, the National Natural Science Foundation of China under grant No. 61672453, the Zhejiang University Education Foundation under grants No. K18-511120-004, and No. K17-511120-017, the Zhejiang public welfare technology research project under grant No. LGF20F020013, and the Key Laboratory of Medical Neurobiology of Zhejiang Province. D.Z. Chen's research was supported in part by NSF Grant CCF-1617735. Jintai Chen, Xiangshang Zheng, and Hongyun Yu are co-first authors.

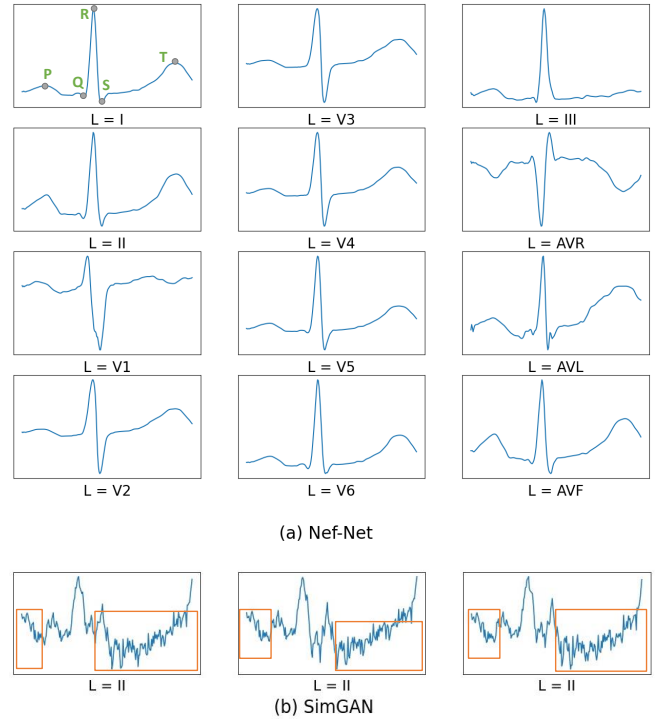


Figure 4: Examples of a multi-lead ECG synthesized from scratch by Nef-Net, and 3 single-lead ECG signals synthesized by SimGAN.

References

- [Anderson *et al.*, 1994] Stanley T Anderson, Olle Pahlm, et al. Panoramic display of the orderly sequenced 12-lead ECG. *Journal of electrocardiology*, 1994.
- [Bousseljot *et al.*, 1995] R Bousseljot, D Kreiseler, and A Schnabel. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. *Biomedizinische Technik/ Biomedical Engineering*, 1995.
- [Case *et al.*, 1979] Robert B Case, William A Tansey, and Allen H Mogtader. A sequential angular lead presentation. *Journal of Electrocardiology*, 1979.
- [Chen *et al.*, 2020a] Jintai Chen, Haochao Ying, et al. A transfer learning based super-resolution microscopy for biopsy slice images: the joint methods perspective. *TCBB*, 2020.
- [Chen *et al.*, 2020b] Jintai Chen, Hongyun Yu, et al. Flow-Mixup: Classifying multi-labeled medical images with corrupted labels. In *BIBM*, 2020.
- [dos Santos Silva *et al.*, 2020] Ittalo dos Santos Silva, José Raimundo Barbosa, et al. Comparison of spatial temporal representations of the vectorcardiogram using digital image processing. *Journal of Electrocardiology*, 2020.
- [Edenbrandt and Pahlm, 1988a] Lars Edenbrandt and Olle Pahlm. Vectorcardiogram synthesized from a 12-lead ECG: Superiority of the inverse Dower matrix. *Journal of Electrocardiology*, 1988.

- [Edenbrandt and Pahlm, 1988b] Lars Edenbrandt and Olle Pahlm. Vectorcardiogram synthesized from a 12-lead ECG: superiority of the inverse Dower matrix. *Journal of Electrocardiology*, 1988.
- [Feng *et al.*, 2020] Hao-Zhe Feng, Zhaoyang You, et al. KD3A: Unsupervised multi-source decentralized domain adaptation via knowledge distillation. *arXiv preprint arXiv:2011.09757*, 2020.
- [Frank, 1956] Ernest Frank. An accurate, clinically practical system for spatial vectorcardiography. *Circulation*, 1956.
- [Girshick, 2015] Ross Girshick. Fast R-CNN. In *ICCV*, 2015.
- [Golany *et al.*, 2020a] Tomer Golany, Gal Lavee, et al. Improving ECG classification using generative adversarial networks. In *AAAI*, 2020.
- [Golany *et al.*, 2020b] Tomer Golany, Kira Radinsky, and Daniel Freedman. SimGANs: Simulator-based generative adversarial networks for ECG synthesis to improve deep ECG classification. In *ICML*, 2020.
- [Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, et al. Generative adversarial nets. In *NeurIPS*, 2014.
- [Grant, 1950] Robert P Grant. Spatial vector electrocardiography: A method for calculating the spatial electrical vectors of the heart from conventional leads. *Circulation*, 1950.
- [Graybiel *et al.*, 1946] Ashton Graybiel, Paul Dudley White, et al. *Electrocardiography in practice*. Saunders, 1946.
- [He *et al.*, 2017] Kaiming He, Georgia Gkioxari, et al. Mask R-CNN. In *ICCV*, 2017.
- [He *et al.*, 2020] Kaiming He, Haoqi Fan, et al. Momentum contrast for unsupervised visual representation learning. In *CVPR*, 2020.
- [Kachuee *et al.*, 2018] Mohammad Kachuee, Shayan Fazeli, and Majid Sarrafzadeh. ECG heartbeat classification: A deep transferable representation. In *ICHI*, 2018.
- [Karras *et al.*, 2019] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, 2019.
- [Kingma and Welling, 2013] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In *ICLR*, 2013.
- [Kors *et al.*, 1990] JA Kors, G Van Herpen, et al. Reconstruction of the Frank vectorcardiogram from standard electrocardiographic leads: Diagnostic comparison of different methods. *European Heart Journal*, 1990.
- [Lee *et al.*, 2019] JeeEun Lee, KyeongTaek Oh, et al. Synthesis of electrocardiogram V-lead signals from limb-lead measurement using R-peak aligned generative adversarial network. *JBHI*, 2019.
- [Makowski *et al.*, 2021] Dominique Makowski, Tam Pham, Zen J. Lau, et al. Neurokit2: A python toolbox for neurophysiological signal processing. *Behavior Research Methods*, 2021.
- [Matyschik *et al.*, 2020] Maksymilian Matyschik, Henry Mauranen, et al. Feasibility of ECG reconstruction from minimal lead sets using convolutional neural networks. In *CINC*, 2020.
- [Mildenhall *et al.*, 2020] Ben Mildenhall, Pratul P Srinivasan, et al. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020.
- [Mirza and Osindero, 2014] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.
- [Moody *et al.*, 2001] George B Moody, Roger G Mark, and Ary L Goldberger. PhysioNet: A web-based resource for the study of physiologic signals. *IEEE Engineering in Medicine and Biology Magazine*, 2001.
- [Radford *et al.*, 2015] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- [Rezende *et al.*, 2014] Danilo Jimenez Rezende, Shakir Mohamed, et al. Stochastic backpropagation and variational inference in deep latent Gaussian models. In *ICML*, 2014.
- [Sohn *et al.*, 2020] Jangjay Sohn, Seungman Yang, Joonnyong Lee, et al. Reconstruction of 12-lead electrocardiogram from a three-lead patch-type device using a LSTM network. *Sensors*, 2020.
- [Tian *et al.*, 2019] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. *arXiv preprint arXiv:1906.05849*, 2019.
- [Wei *et al.*, 2019] Kun Wei, Muli Yang, et al. Adversarial fine-grained composition learning for unseen attribute-object recognition. In *ICCV*, 2019.
- [Zhang *et al.*, 2018] Hongyi Zhang, Moustapha Cisse, et al. Mixup: Beyond empirical risk minimization. In *ICLR*, 2018.