This ECG Does Not Exist: Using Variational Autoencoders To Generate Periodic ECG Signals.

Yonatan Oren*
Department of Computer Science
Stanford University
yonatano@cs.stanford.edu

Koren Gilbai*
Department of Computer Science
Stanford University
kgilbai@stanford.edu

Abstract

The use of synthetic clinical data, which has similar or identical statistical properties to real patient data, but is not sourced from real patients, is promising for enhancing clinical research and protecting patient privacy. We train a convolutional variational autoencoder on a dataset of 12-lead Electrocardiograms (ECGs) and demonstrate that samples generated by the model lack a global coherent structure–for example, the samples are not periodic, and are therefore inadmissible. We propose to modify the VAE architecture by using the decoder to predict a *filter*, which is then convolved with an impulse train to produce the output signal. This novel procedure guarantees that the signals sampled from the model are periodic.

1 Introduction

Among the most significant impediments to the application of machine learning in clinical settings are a lack of sufficient labeled training data and the desire to protect patient privacy. Generation of synthetic clinical data, which has similar or identical statistical properties to real patient data, but is not sourced from real patients, is promising for addressing both issues of scale and privacy–and by extension for enhancing clinical research. In this project, we implement a Variational Autoencoder (VAE)–a deep generative model–for use in generating synthetic clinical electrocardiograms (ECGs) to augment a small medical dataset. Such generated data could then be used in a down-stream classification task, such as detecting various cardiovascular conditions. The primary challenge is generating samples that adhere to certain structural properties such as the QRS complex and periodicity. As we will show, using a straightforward VAE does not produce samples with the requisite clinical structure. We therefore complement VAEs with classical signal processing techniques to produce periodic ECG samples.

2 Related Work

Hong et al. (2020) conducted a systematic review of the literature on deep learning methods for electrocardiogram data [5]. Common application tasks include ECG segmentation, disease detection, sleep staging, biometric human identification, denoising, and more.

The most notable results in applying deep learning methods to generate synthetic ECGs were achieved by Delaney et al. (2019) [1], Golany and Radinsky (2019) [3], and Zhu et al. (2019) [10]. All used different variants of generative adversarial networks (GANs), with the best results coming from Delaney et al. (2019) with Maximum Mean Discrepancy (MMD) metric of 1.05 × 10-3.

CS230: Deep Learning, Fall 2022, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

A more recent paper by Kuznetsov, VI V et al implements a simple VAE to generate ECGs corresponding to a single cardiac cycle [7]. The paper's primary drawback is the lack of complete ECG signal generation, and as such an absence of methodologies to address the challenge of synthetically producing global coherent structure.

3 An Overview of ECG and the Dataset

An electrocardiogram (ECG) is a diagnostic measure of the electrical activity of the heart, collected using electrodes placed on the body. It is one of the most important tools used by cardiologists to assess the function of the heart and detect potential pathologies. Tracings include different phases of a cardiac cycle, including the P, T, and U waves, as well as the QRS complex, which consists of the Q, R, and S peaks. The size, shape, and location of these components provide valuable information about the heart's function and the presence of certain diseases.

We use the dataset introduced in the ECG research paper by Zheng, J., Chu, H., Struppa, D. et al [9]. The dataset consists of 10,046 10-second 12-lead ECGs sampled at 500Hz. Each ECG is labeled with one of 12 *rhythms* (e.g. Sinus Rhythm, Atrial Fibrillation), and a subset of the 56 different *conditions* (e.g. myocardial infarction, U wave), along with the patient age and gender, and eleven other ECG characteristics. Our generative model is trained on and produces single-lead ECGs.

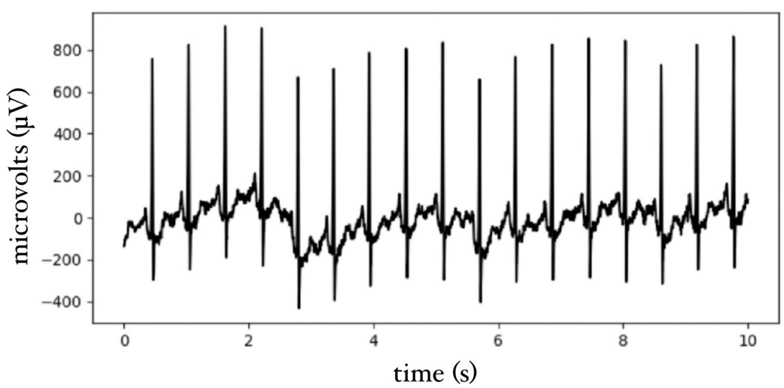

Figure 1: ECG sample from a 66 year old Female patient with Sinus Tachycardia

4 Convolutional Variational Autoencoder

In this section, we describe the VAE architecture used as the baseline model for our generative task (see Figure 2), which is the same model described in [6], whose paper was incredibly informative in understanding this architecture and its application to the setting of ECGs.

The model consists of two components: a probabilistic *encoder*, $q_\theta(z \mid x)$, which receives the raw signal x as input and predicts a latent vector z, and a probabilistic *decoder*, $p_\phi(x' \mid z)$, which takes z and produces a signal x'. The encoder and decoder are trained jointly to minimize the following loss:

$$l(\theta, \phi) = \sum_{i=1}^{N} -\mathbb{E}_{z \sim q_\theta(z|x_i)} \left[\log p_\phi(x'_i \mid z)\right] + D_{KL}(q_\theta(z \mid x_i) \mid\mid \mathcal{N}(0, I))$$

The first term is the reconstruction loss, and the second is a KL divergence term which encourages the marginal distribution of the latent vector produced by the encoder to be a standard multivariate Gaussian. We minimize the KL divergence in

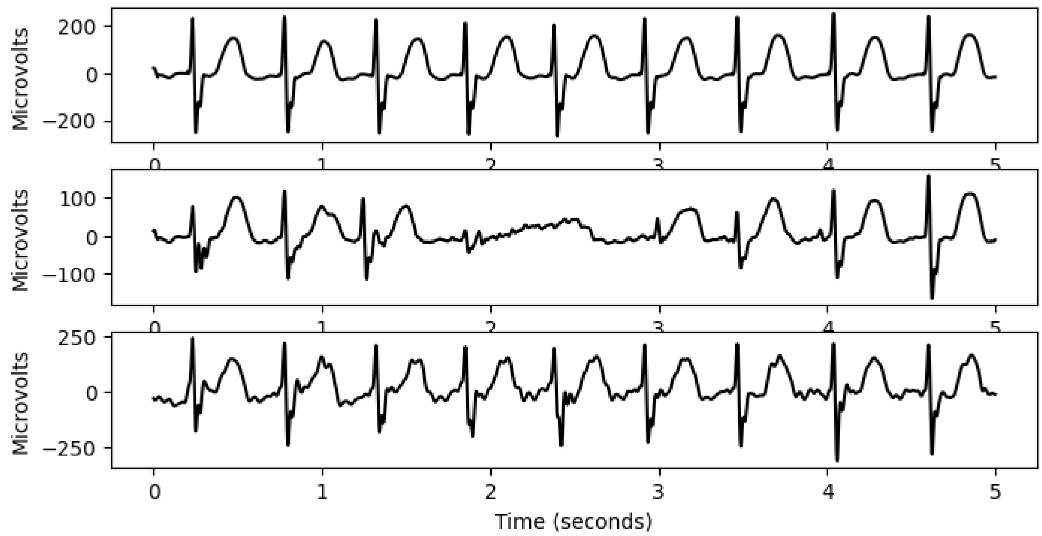

Figure 2: **Top:** a typical example from the dataset (single lead.) **Middle:** a sample from the VAE baseline, demonstrating missing periods. **Bottom:** a sample from our modified VAE architecture, generated using an impulse train extracted from the example in the top row. The sample from the modified VAE captures the periodic structure of the original.

the usual way by instead maximizing the Evidence Lower-Bound (ELBO) [2].

Encoder: The encoder is a neural network consisting of 10 residual blocks, followed by two separate linear layers, one for each parameter of $q_\theta(z \mid x)$ (the mean and variance.) Each block is composed of two convolutional layers, with each layer applying batch normalization and ReLU before a one-dimensional convolution with a kernel width of either 19 or 9. We employ skip connections in each residual block of the encoder by adding a copy of the input to the output of the convolutional layers. In some blocks, we downsample the signal by convolving with a 1×1 kernel with a horizontal stride of 2. Please see figure 5 for details.

The output is then flattened and passes through two (separate) linear layers to predict the mean and log-variance of the latent distribution $q_\theta(z \mid x)$. The "Re-parameterization Trick" is then used to sample a latent vector z. That is, we sample $\epsilon \sim \mathcal{N}(0, I)$ and let $z_i = \mu_i + \sigma_i \epsilon_i$. This enables gradients to flow through the latent distribution and to optimize the parameters of $q_\theta(z \mid x)$.

Decoder: The decoder is a neural network with 10 residual blocks, designed to be symmetric to the encoder. Each block is composed of two convolutional layers, with each layer applying batch normalization and ReLU before a one-dimensional *transpose* convolution with a kernel width of either 19 or 9. Here, we do not employ skip connections, and upsample the input instead of downsampling it.

Training: The network was trained to minimize the sum of the reconstruction error (mean-squared error) and the KL divergence term. Details on hyperparameters can be found in Figure 5.

Training the network, especially on the full dataset, was very difficult. We encountered two major problems when scaling up the VAE to the full dataset, both of which are well documented the literature: (1). the KL term diverges after a number of epochs, and (2). the KL term drops to zero after a number of epochs ("KL Vanishing"). We used both the annealing schedule described in [8], and also the cyclical annealing schedule described in [4], but after many attempts, were unsuccessful in achieving training stability on the full dataset. In order to proceed under limited time constraints, we opted to train the model on a small subset of the

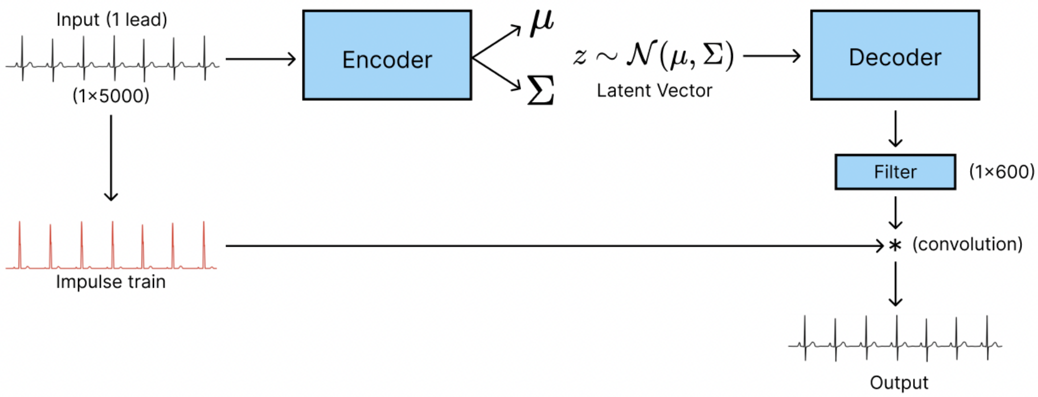

Figure 3: Our proposed VAE architecture for generating periodic signals.

training examples, truncated to 2048 time-steps per lead.

Alternative architectures: In addition to the convolutional VAE, we also implemented an LSTM-based sequence to sequence autoencoder trained with teacher forcing, but did not find this architecture to be succesful in modeling the data.

5 Modifying The Architecture to Generate Periodic Signals

In order to generate admissible synthetic data which is usable in a downstream task, it is crucial for the samples of the generative model to mimic the highly structured nature of ECG data. We have demonstrated that naively sampling from a VAE trained on this data can result in samples which are missing periods. To solve the issue of non-periodic samples, we propose an alternative architecture which forces the generated samples to be periodic.

Modified Architecture: Our proposed architecture (see Figure 3) differs from the standard VAE in two ways: (1). the input signals are processed to extract an *impulse train* from each signal, which captures its periodic structure, and (2). the decoder outputs a *filter*. To compute the final output, we convolve each impulse train with the corresponding filter, which produces a signal with the same periodic structure as the input.

Extracting the impulse trains: We apply a sequence of straightforward transformations to each input signal to produce a corresponding impulse train:

1. The signal is standardized by normalizing by twice its standard deviation.
2. The signal is raised to the power of four, which squashes the secondary local extrema while preserving the primary peak in each period. This also makes the signal non-negative, which is useful for ECGs with negative R-peaks.
3. A Hamming filter is applied to smooth the signal.
4. The function $\log(1 + x)$ is applied to the signal, which serves to restore the relative amplitudes of the peaks in the signal.

In our first attempt, we constructed impulse trains with a fixed period by estimating the period of each input signal as the second peak of its autocorrelation. However, we found that using an impulse train with a fixed period leads to poor results, since the period of the input signals is not fixed: the peaks of the reconstructed signal being even slightly off from the original signal would cause the reconstruction to incur a large MSE, even if the two signals look identical.

We attempted to use a number of peak-finding algorithms described in the literature, but did not find that these algorithms worked reliably on our dataset.

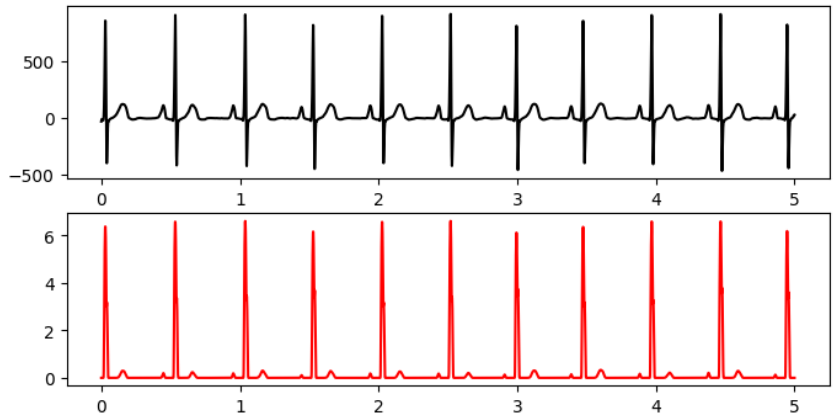

Figure 4: Example ECG signal and its corresponding impulse train.

We settled on the approach described above as it produces an impulse train which is sufficiently usable and whose peaks match the input exactly.

6 Conclusion/Future Work

Synthetic ECG generation presents a promising avenue towards the application of machine learning in clinical cardiology settings, especially given the ability to generate samples with global coherent structure. Pertinent future work on the basis of this project would include conditional generation of samples with certain pathologies, extrapolation of the model to include all 12 ECG leads.

7 Contributions

Both members jointly contributed to a literature review, dataset construction, model implementation and testing, and writing of the report.

8 Code

Code is available at: `https://github.com/yonatano/ecg_project`.

References

[1] Ward T. E. Delaney A. M. Brophy E. "Synthesis of Realistic ECG using Generative Adversarial Networks ". In: *arXiv* 1909.09150 (2019). DOI: https://arxiv.org/abs/1909.09150.

[2] Max Welling Diederik P Kingma. "Auto-Encoding Variational Bayes ". In: *arXiv* 1312.6114 (2013). DOI: https://doi.org/10.48550/arXiv.1312.6114.

[3] Radinsky K. Golany T. "PGANs: Personalized Generative Adversarial Networks for ECG Synthesis to Improve Patient-Specific Deep ECG Classification ". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 33(01).557-564 (2019). DOI: https://doi.org/10.1609/aaai.v33i01.3301557.

[4] Xiaodong Liu Jianfeng Gao Asli Celikyilmaz Lawrence Carin Hao Fu Chunyuan Li. "Cyclical Annealing Schedule: A Simple Approach to Mitigating KL Vanishing". In: *arXiv* 1903.10145 (2019). DOI: https://doi.org/10.48550/arXiv.1511.06349.

[5] Shang J. Xiao C. Sun J. Hong S. Zhou Y. "Opportunities and Challenges of Deep Learning Methods for Electrocardiogram Data: a Systematic Review." In: *Comput. Biol. Med* 122.103801 (2020). DOI: https://pubmed.ncbi.nlm.nih.gov/32658725/.

[6] Lim H-S Yoon D Jang J-H Kim TY. "Unsupervised feature learning for electrocardiogram data using the convolutional variational autoencoder". In: *PLoS ONE* 16(12).e0260612 (2021). DOI: https://doi.org/10.1371/journal.pone.0260612.

[7] V V et al. Kuznetsov. "Interpretable Feature Generation in ECG Using a Variational Autoencoder." In: *Frontiers in genetics* 12.638191 (2021). DOI: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8049433/.

[8] Oriol Vinyals-Andrew M. Dai Rafal Jozefowicz Samy Bengio Samuel R. Bowman Luke Vilnis. "Generating Sentences from a Continuous Space". In: *arXiv* 06349.1511 (2016). DOI: https://doi.org/10.48550/arXiv.1511.06349.

[9] Zhang J. Danioko S. et al. Zheng J. "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients". In: *Sci Data* 7.48 (2020). DOI: https://doi.org/10.1038/s41597-020-0386-x.

[10] Fu Y. Liu Q. Shen B. Zhu F. Ye F. "Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network". In: *Scientific reports* 9,1.6734 (2019). DOI: https://doi:10.1038/s41598-019-42516-z.

9 Appendix

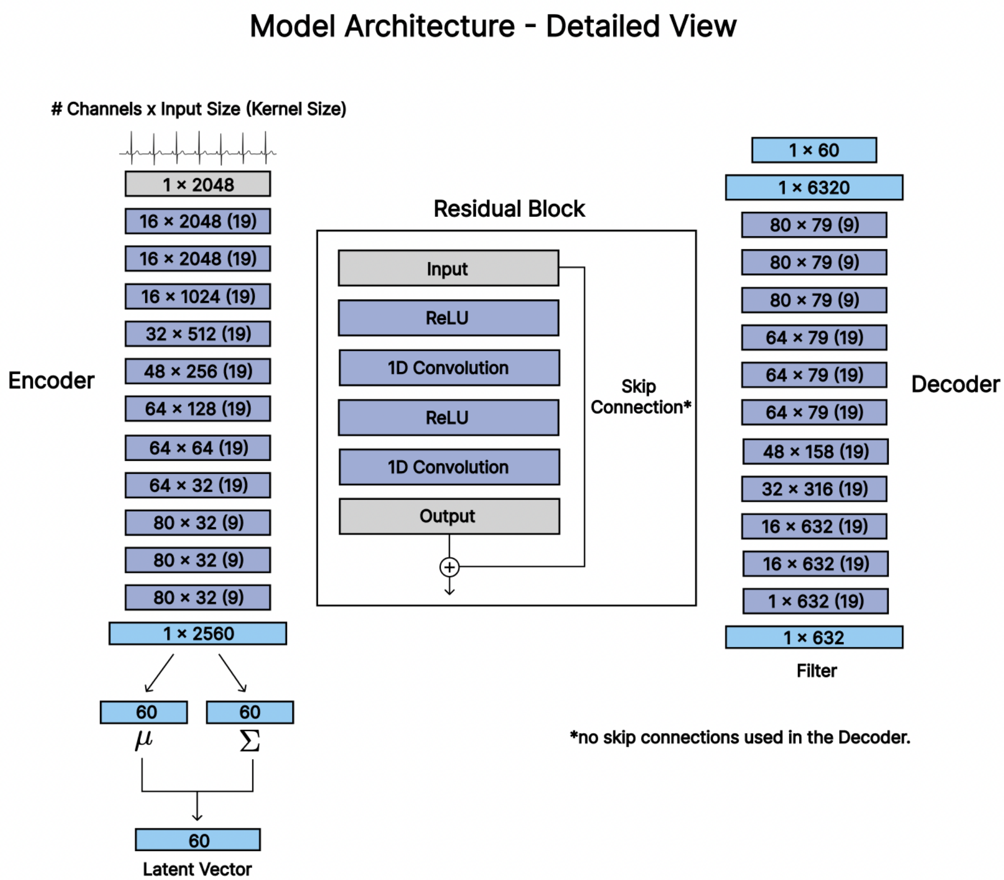

Figure 5: Detailed view of the modified VAE architecture.