Irregular heartbeats detection using deep generative modelling

Simone Maggi

University of Torino

simone.maggi738@edu.unito.it

April 12, 2021

Motivations

- Why should It be useful?

Motivations

- Why should It be useful?
- A doctor can't scrutinize carefully a very long ECG record.

Motivations

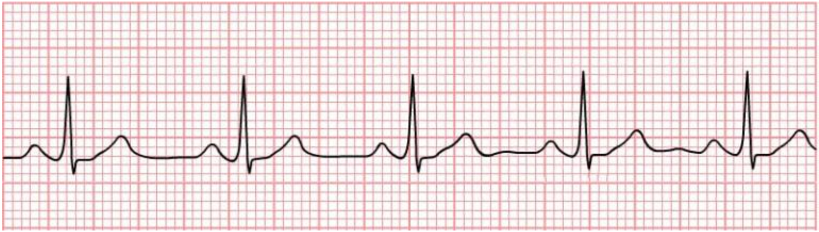
- Why should It be useful?
- A doctor can't scrutinize carefully a very long ECG record.
- The early and correct diagnosis of cardiac abnormalities can increase the chances of successful treatments.[1]

Motivations

- Why should It be useful?
- A doctor can't scrutinize carefully a very long ECG record.
- The early and correct diagnosis of cardiac abnormalities can increase the chances of successful treatments.[1]
- In this context an algorithm capable of spotting abnormalities is useful.

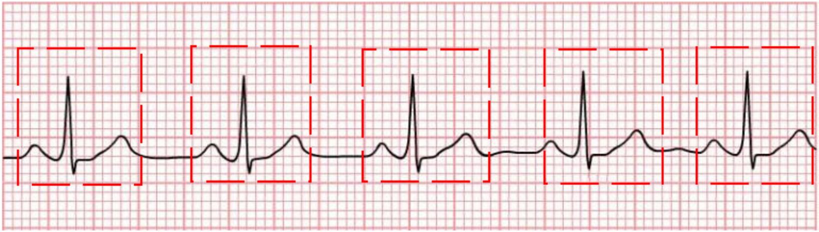
The ECG

- The electrocardiogram (ECG) is the electric signal obtained measuring heart activity through electrodes placed on the skin.



The ECG

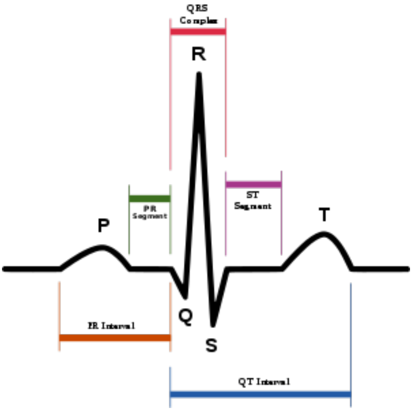
- The electrocardiogram (ECG) is the electric signal obtained measuring heart activity through electrodes placed on the skin.
- It is composed by a sequence of heartbeats of about 0.8 seconds long.



The Sinus heartbeat

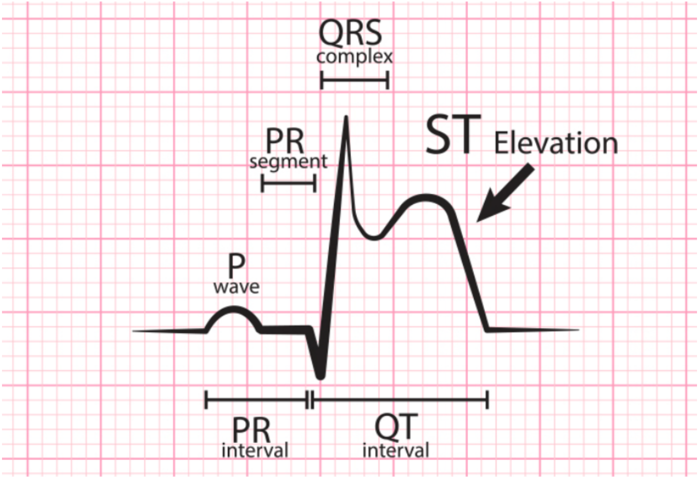
Heart cycle

1. P wave: Atria activation
2. QRS complex: Ventricles depolarization
3. T wave: Ventricles re-polarization



Abnormal heartbeats

- An irregular rhythm could be symptom of some disorder.



Thesis aim

- Find abnormal heartbeats inside the ECG using modern Data-Driven techniques.

Thesis aim

- Find abnormal heartbeats inside the ECG using modern Data-Driven techniques.
- The possible anomalies that could affect the ECG are plenty. Some are rare [5] and difficult to observe in a ECG record. On the other hand, the Sinus heartbeat is very common.

Thesis aim

- Find abnormal heartbeats inside the ECG using modern Data-Driven techniques.
- The possible anomalies that could affect the ECG are plenty. Some are rare [5] and difficult to observe in a ECG record. On the other hand, the Sinus heartbeat is very common.
- Supervised approach are limited by the data unbalance [1]

Thesis aim

- Find abnormal heartbeats inside the ECG using modern Data-Driven techniques.
- The possible anomalies that could affect the ECG are plenty. Some are rare [5] and difficult to observe in a ECG record. On the other hand, the Sinus heartbeat is very common.
- Supervised approach are limited by the data unbalance [1]
- We aim to teach to the model the normal ECG and then make it use the acquired knowledge to asses the abnormality of future data.

Semi-supervised anomaly detection

Semi-supervised anomaly detection $\Rightarrow$ $\begin{cases} \uparrow \text{Enough labelled normal data} \\ \downarrow \text{abnormal data scarce or unknown} \end{cases}$

Problem

Problem

- Let $\mathcal{B} = (\mathbf{x}_1, ..., \mathbf{x}_n)$, $\mathbf{x}_i \in R^{L \times N}$ be a collection of **normal** heartbeats, where N is the length and L is the number of leads.
- Given an **unseen** heartbeat $\mathbf{x}_0 \in \mathbb{R}^{L \times N}$
- identify if $\mathbf{x}_0$ has a behaviour which deviate significantly from $\mathcal{B}$.

Out of Distribution

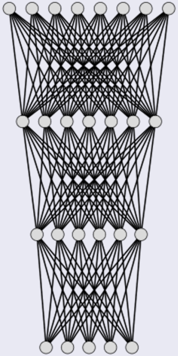
Problem

- Let $\mathcal{B} = (\mathbf{x}_1, ..., \mathbf{x}_n) \sim p_{normal}(\mathbf{x})$ be a random sample from the unknown probability measure on $\mathbb{R}^{L \times N}$ of normal heartbeats.
- **Estimate** $p_{normal}(\mathbf{x})$ given $\mathcal{B}$.
- Let $\mathbf{x}_0 \sim p_{heartbeats}(\mathbf{x})$ a sample from the distribution of all possible heartbeats.
- **Asses** if $\mathbf{x}_0$ comes from $p_{normal}(\mathbf{x})$.

Deep Generative Models

Deep

They make use of deep neural networks

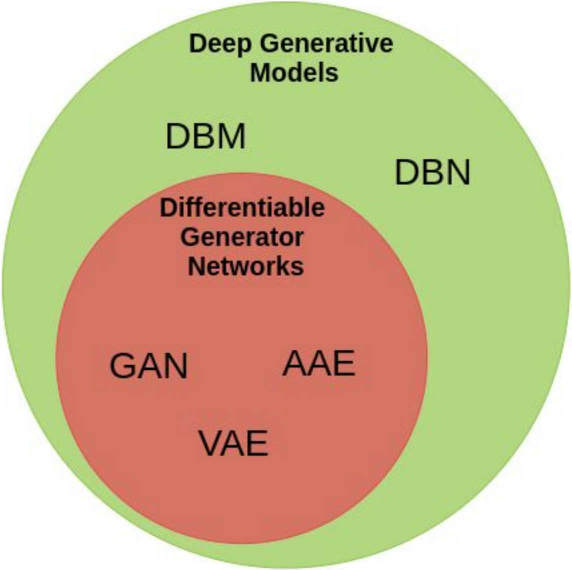


Generative

They aim to capture the data distribution.

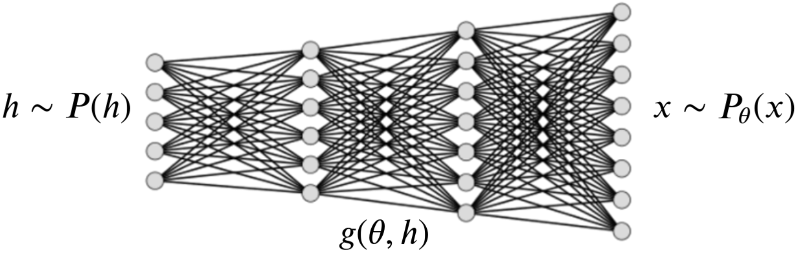
P(x)

Deep Generative Models



Differentiable generator networks

- Assume: input $\mathbf{x}$ is determined by hidden variables $\mathbf{h} \sim p(h)$
- The neural network $g(\theta, h)$ provides a non-linear change of variable which transfroms $P(h)$ into $P_\theta(x)$.



Differentiable generator networks

- **Change of variable**: the output distribution $P_{\boldsymbol{\theta}}(\mathbf{x})$

$$P_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{P_{\mathbf{h}}(g^{-1}(\mathbf{x}, \boldsymbol{\theta}))}{|det(\frac{\delta \mathbf{g}}{\delta \mathbf{h}})|}$$

Differentiable generator networks

- **Change of variable**: the output distribution $P_{\boldsymbol{\theta}}(\mathbf{x})$

$$P_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{P_{\mathbf{h}}(g^{-1}(\mathbf{x}, \boldsymbol{\theta}))}{|det(\frac{\delta \mathbf{g}}{\delta \mathbf{h}})|}$$

- **Objective**: Minimize a divergence $D(P_x(x)||P_{\boldsymbol{\theta}}(x))$.

Differentiable generator networks

- **Change of variable**: the output distribution $P_{\boldsymbol{\theta}}(\mathbf{x})$

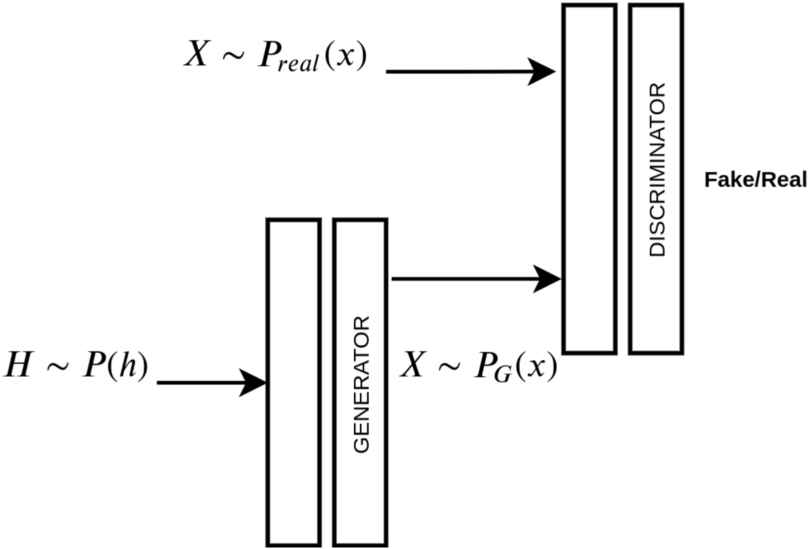
$$P_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{P_{\mathbf{h}}(g^{-1}(\mathbf{x}, \boldsymbol{\theta}))}{|det(\frac{\delta \mathbf{g}}{\delta \mathbf{h}})|}$$

- **Objective**: Minimize a divergence $D(P_x(x)||P_{\boldsymbol{\theta}}(x))$.
- **Optimization**: Gradient Descend.

GAN

- The quintessential example is the Generative Adversarial network.

GAN



$X \sim P_{real}(x)$
DISCRIMINATOR
Fake/Real
GENERATOR
$H \sim P(h)$
$X \sim P_G(x)$

GAN

It is a **mini-max optimization problem**[2]

$$\min_{\omega} \max_{\theta} V(\theta, \omega),$$

with

$$V(\theta, \omega) = \mathbb{E}_{x \sim P_x(x)} \left[log(D_\omega(x)) \right] + \mathbb{E}_{h \sim P_h(h)} \left[log(1 - D_\omega(G_\theta(h))) \right]$$

GAN

- **Objective**: assuming perfect discriminator can be shown that It minimizes the **Jensen-Shannon divergence**

$$D_{JS}(P_x(x)||P_{\boldsymbol{\theta}}(x)) =$$
$$= D_{KL}\left(P_x(x)||\frac{P_x(x)+P_{\boldsymbol{\theta}}(x)}{2}\right) + D_{KL}\left(P_{\boldsymbol{\theta}}(x)||\frac{P_x(x)+P_{\boldsymbol{\theta}}(x)}{2}\right)$$

- **Convergence**; $V(\theta, \omega)$ is minimum if and only if $P_{\boldsymbol{\theta}}(x) = P_x(x)$

WGAN

- The WGAN ameliorates the standard GAN framework
- It minimizes the **Wasserstein-1 distance**

$$d_w(P_x(x), P_{\boldsymbol{\theta}}(x))) = \sup_{f \in Lip_1(\mathcal{X})} \left(\mathbb{E}_{x \sim P_x(x)}\left[f(x)\right] - \mathbb{E}_{x \sim P_{\boldsymbol{\theta}}(x)}\left[f(x)\right]\right)$$

- Better convergence properties

Related works

Semi-supervised anomaly detection

Standard ML

PCA
OCSVM
KDE

Deep

AutoEncoder
AnoGAN
AnoBeat
BeatGAN

Contributions

AAECG

- We make use of an Adversarial Auto-Encoder.
- Additional **gender information** will be used in the modelling.

Contributions

AAECG

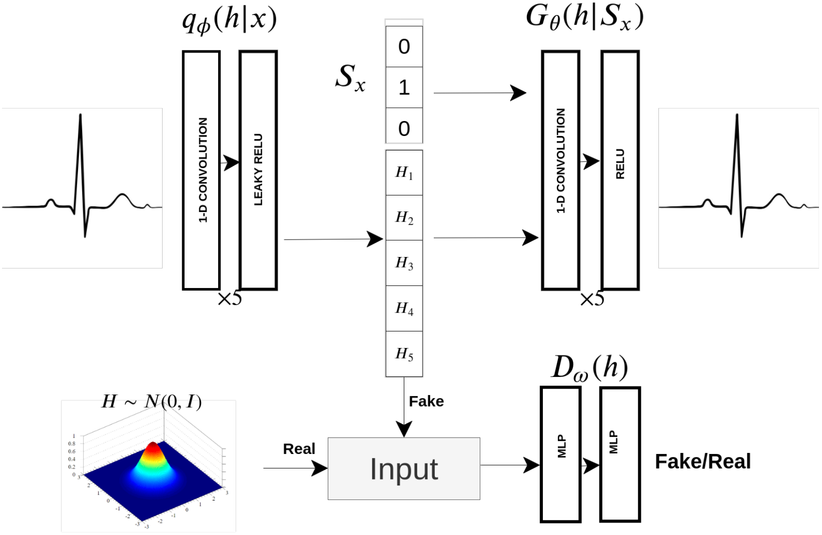
- We make use of an Adversarial Auto-Encoder.
- Additional **gender information** will be used in the modelling.
- The patient's gender influences the Sinus heartbeat wave-forms [3].

Contributions

AAECG

- We make use of an Adversarial Auto-Encoder.
- Additional **gender information** will be used in the modelling.
- The patient's gender influences the Sinus heartbeat wave-forms [3].
- A recent study showed that a neural network is capable of retrieving the patient's age and sex from its ECG record [4].

Proposed framework: AAECG



Proposed Framework: AAECG

- The three networks are trained with gradient descend by minimizing the objective function $\mathcal{L}(\theta, \phi, \omega; \mathbf{x})$

Objective

Mean Square Error

Both Generator and Encoder network are trained to minimize the Mean Squared Error of the Reconstructed data

$$\mathcal{L}_{REC}(\theta, \phi; \mathbf{x}) = \mathbb{E}_{\mathbf{h} \sim q_\phi(\mathbf{h}|\mathbf{x})} \left[\sum_{i=0}^{N \times L} \left(x_{(i)} - G_\theta(\mathbf{h}|S_\mathbf{x})_{(i)} \right)^2 \right]$$

Objective

Total variation penalty

The Generator additionally minimizes

$$\mathcal{L}_{TV}(\theta; \mathbf{x}) = \mathbb{E}_{\mathbf{h} \sim q_\phi} \left[TV(G_{\boldsymbol{\theta}}(\mathbf{h} | S_{\mathbf{x}})) \right]$$

Objective

Encoder objective

The Encoder is requested to match the aggregated distribution

$$q_\phi(\mathbf{h}) = \int q_\phi(\mathbf{h}|\mathbf{x}) P_x(\mathbf{x}) d\mathbf{x},$$

to the imposed prior by **fooling the discriminator** maximizing

$$\mathcal{L}_{ADV}(\phi; \mathbf{x}) = \mathbb{E}_{\mathbf{h} \sim q_\phi(\mathbf{h})} \left[D_{\boldsymbol{\omega}}(\mathbf{h}) \right]$$

Objective

Discriminator objective

The discriminator is trained to estimate the Wasserstein-1 distance between the prior and encoder distributions

$$\mathcal{L}_D(\omega; \mathbf{x}) = \mathbb{E}_{\mathbf{h} \sim p_h(\mathbf{h})} [D_{\boldsymbol{\omega}}(\mathbf{h}))] - \mathbb{E}_{\mathbf{h} \sim q_\phi(\mathbf{h})} [D_{\boldsymbol{\omega}}(\mathbf{h}))] + \lambda_{GP} \, \mathbb{E}_{\hat{\mathbf{x}} \sim p_{\hat{x}}} \left[(||\nabla_{\hat{\mathbf{x}}} D_\omega(\mathbf{x})||_2 - 1)^2 \right]$$

Objective

Objective

$$\mathcal{L}(\phi, \theta, \omega; \mathbf{x}) = \lambda_1 \mathcal{L}_{REC}(\theta, \phi; \mathbf{x})$$

Objective

Objective

$$\mathcal{L}(\phi, \theta, \omega; \mathbf{x}) = \lambda_1 \mathcal{L}_{REC}(\theta, \phi; \mathbf{x}) - \lambda_2 \mathcal{L}_{ADV}(\phi; \mathbf{x})$$

Objective

Objective

$$\begin{aligned}\mathcal{L}(\phi, \theta, \omega; \mathbf{x}) &= \lambda_1 \mathcal{L}_{REC}(\theta, \phi; \mathbf{x}) \\ &- \lambda_2 \mathcal{L}_{ADV}(\phi; \mathbf{x}) \\ &+ \lambda_3 \mathcal{L}_{TV}(\theta; \mathbf{x})\end{aligned}$$

Objective

Objective

$$
\begin{aligned}
\mathcal{L}(\phi, \theta, \omega; \mathbf{x}) = {} & \lambda_1 \mathcal{L}_{REC}(\theta, \phi; \mathbf{x}) \\
& - \lambda_2 \mathcal{L}_{ADV}(\phi; \mathbf{x}) \\
& + \lambda_3 \mathcal{L}_{TV}(\theta; \mathbf{x}) \\
& + \mathcal{L}_{D}(\omega; \mathbf{x})
\end{aligned}
$$

Anomaly Detection

Anomaly score

Given an unseen heartbeat $\mathbf{x}_0$ the model generates an anomaly score based on the **reconstruction error**.

$$A(\mathbf{x}_0) = \frac{1}{J} \sum_{j=1}^{J} MSE_j(\mathbf{x}_0, \hat{\mathbf{x}_0})$$

where,

$$MSE_j(\mathbf{x}_0, \hat{\mathbf{x}_0}) = \sum_{i=0}^{N \times L} \left(\mathbf{x}_{(i)} - G_{\boldsymbol{\theta}}(\mathbf{h}_j | S_{\mathbf{x}_0})_{(i)} \right)^2, \quad \mathbf{h}_j \sim q_{\hat{\phi}}(\mathbf{h}|\mathbf{x})$$

Experiment

Database

- MIT-BIH arrhythmia database, standard benchmark.

Experiment

Database

- MIT-BIH arrhythmia database, standard benchmark.
- It contains a set of half-hour excerpts of two-channel ambulatory ECG recordings.

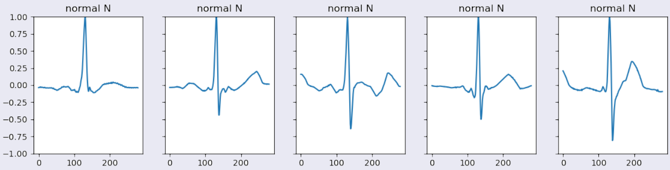
Experiment

Database

- MIT-BIH arrhythmia database, standard benchmark.
- It contains a set of half-hour excerpts of two-channel ambulatory ECG recordings.
- Total of 97,568 labelled heartbeats. 87,142 of them are normal heartbeats.

Experiment

Dataset samples



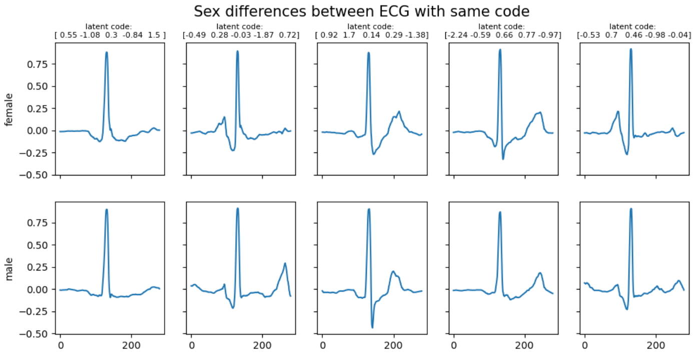
Results

Results obtained with 5-fold cross validation

Model	**PR-AUC**	**ROC-AUC**
PCA	0.7026 ± 0.0048	0.8352 ± 0.0020
Beat-AE	0.8731 ± 0.0043	0.9027 ± 0.0038
Beat-Fast AnoGAN	0.8304 ± 0.0113	0.8865 ± 0.0057
AnoBeat[1]	0.8799 ± 0.0038	0.9151 ± 0.0038
AAECG	$\mathbf{0.9204 \pm 0.0077}$	$\mathbf{0.9504 \pm 0.0045}$

[1] In the original paper[4] the results was comparable to our model: PR-AUC 0.9276 ± 0.0060 ROC-AUC 0.9576 ± 0.0035

Sex influence



Conclusion

- A novel semi-supervised heartbeat anomaly detection model was proposed.
- It can include additional external information in the modelling.
- AAECG could be used to continuously monitoring patient in intensive care alarming doctors when abnormalities are found.

Thanks for your attention

References

- Guidelines for the Early Management of Adults With Ischemic Stroke
  Circulation 2007; 115:e478–e534
- Misplaced ECG electrodes and the need for continuing training
- Sex Differences in the Electrocardiogram
- Age and Sex Estimation Using Artificial Intelligence From Standard 12-Lead ECGs
  Zachi I. Attia et al. 2019
- Antzelevitch C and Brugada P and Borggrefe M and Brugada J and Brugada R and Corrado D and Gussak I and LeMarec H and Nademanee K and Perez Riera AR and Shimizu W and Schulze-Bahr E and Tan H and Wilde A.
  Brugada syndrome: report of the second consensus conference. Heart Rhythm.

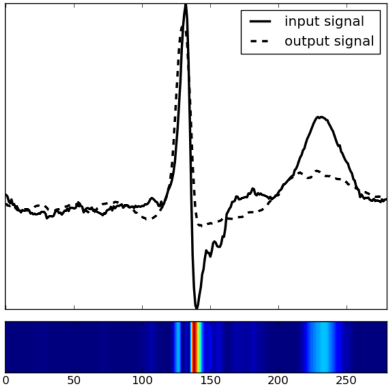
Li, Hongzu and Boulanger, Pierre
A Survey of Heart Anomaly Detection Using Ambulatory Electrocardiogram (ECG)

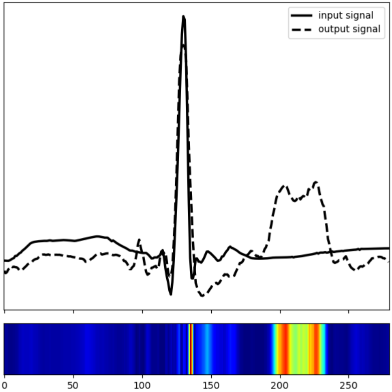
Generative Adversarial Networks
http://dx.doi.org/10.1007/978-1-4842-3679-6_8

f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks
Medical Image Analysis

Anobeat: Anomaly Detection for Electrocardiography Beat Signals
10.1109/DSC50466.2020.00029

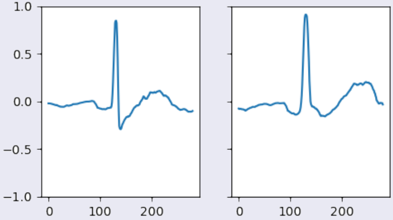
Interpretability



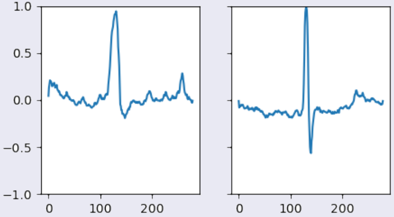


Total Variation penalty

Samples from AAECG



Samples from the AnoGAN



Proposed Framework: AAECG

- The heartbeats are represented by a vector $\mathbf{x} \in \mathbb{R}^{L \times N}$
- The Generator receives as input the five-dimensional hidden vector distributed as $\mathbf{h} \sim N(0, I))$ and the geneder information S_x.
- The encoder is stochastic with Gaussian conditional

$$q_\phi(\mathbf{h}|\mathbf{x}) = dN(\mathbf{h}; \mu_\phi(\mathbf{x}), \sigma^2_\phi(\mathbf{x})\mathbf{I})$$