EUROPEAN PARTNERSHIP
Co-funded by the European Union
NCBR Narodowe Centrum Badań i Rozwoju

Towards Semantic Embeddings of Cardiological Signals with Diffusion Autoencoders

Bartosz Marcinkowski[1] Jakub Siuta[1] Ana Candela Celdran[2] Mena Nadum[2] Marek Wachnicki[1] Jerzy Orłowski[1]
MIM.AI, Warsaw, Poland[1] CSW Therapeutics AB, Stockholm, Sweden[2]

TL;DR

- **Problem**: Wearable cardiac devices lack sufficient training data
- **Solution**: Self-supervised diffusion autoencoder trained on mixed datasets
- **Innovation**: Linear Recurrent Units + Attention Reducer for embeddings
- **Results**: 76.3% macro F1 on 4-class ECG classification by a linear model
- **Impact**: ML development for new devices with minimal device data

Data scarcity in ML for heart monitoring devices

Cardiovascular diseases affect millions globally, driving need for remote monitoring. Novel wearable devices collect ECG, SCG, GCG, PPG but generate limited training data. Existing datasets have heterogeneous annotations, sensor placements, signal types. Supervised approaches require specific labeled datasets.

Self-Supervised Framework for cardiovascular signals

We extract single channel snippets of fixed length, frequency and scale from multiple datasets. We train a diffusion autoencoder to denoise corrupted signals using a hint from a semantic encoder that provides fixed-size embeddings of clean signals. Those learned representations are later used as features for downstream classification tasks.

Architecture

The **denoising model** is a U-Net, with blocks conditioned on the diffusion step and a semantic embedding provided by the **semantic encoder**. The **denoising model**'s job is to reverse Gaussian diffusion noising. The **semantic encoder**'s job is to encode the most useful fixed-size representation of a clean signal as hint for the **denoising model**.

Semantic Encoder: Linear Recurrent Unit (LRU)

The semantic encoder uses Linear Recurrent Units (LRU) to capture oscillatory patterns via complex-valued recurrence: $x_k = \Lambda x_{k-1} + Bu_k$. This operation supports efficient parallel training with scan operations and has a better inductive bias for cardiological signals compared to CNNs.

Semantic Encoder: Attention Reducer

We developed the Attention Reducer to obtain fixed-sized embeddings for any signal length. Given $X \in \mathbb{R}^{C \times L}$ with C channels and arbitrary length L, the reducer outputs a compact embedding $z \in \mathbb{R}^C$. With two trainable parameter matrices $W_a, W_v \in \mathbb{R}^{C \times C}$ we compute attention $A = W_a X$, per-time values $V = W_v X$, softmax weights $S_{ij} = \frac{e^{A_{ij}}}{\sum_{t=1}^{L} e^{A_{it}}}$ and the final embedding values $z_i = \sum_{t=1}^{L} S_{it} V_{it}$.

Training

Both the semantic encoder and the denoising U-Net are trained at the same time. The denoising model is given a noising level, the noised signal and the output of the semantic encoder on the original signal. The loss function is defined as as the mean square error between the output of the denoising model and the added noise. We chose the semantic embedding dimensionality to be 128. We used a training dataset of 356099 10-second single-channel signal snippets sampled at 300Hz collected from public datasets (PTB-XL, MIT-BIH) and trained the model for 8 days on a single NVidia A100 40 GB.

Results

The encoder was evaluated on a Challenge2017 classifcation problem (classes: normal, atrial fibrillation, other, noisy). The embeddings were used as features for a simple linear classifier, achieving 76.3% macro F1, which is below contest winners' 83% (including heuristics crafted for the particular challenge dataset), but competitive, scoring above the 73% mean submission.

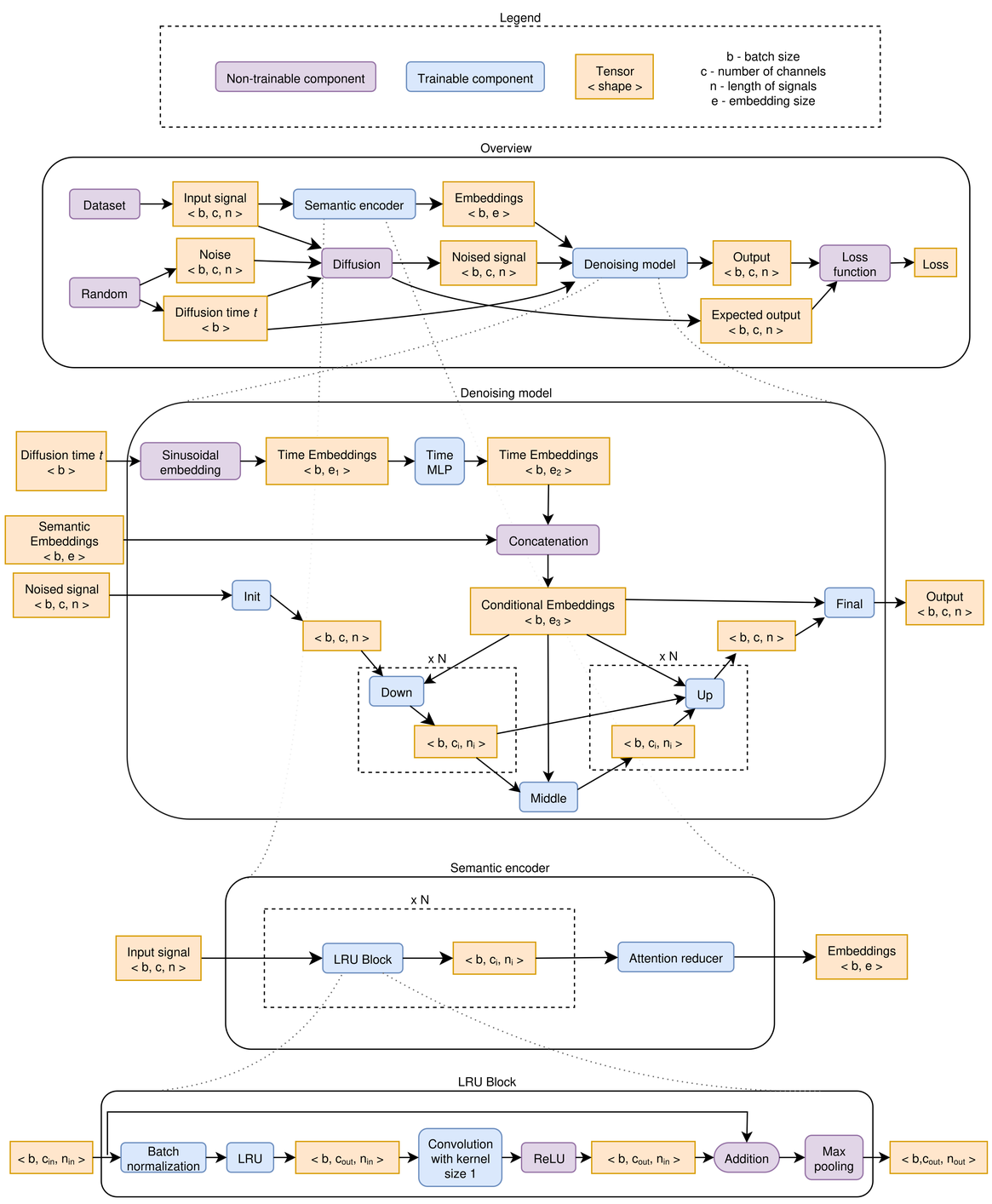

Figure 1: The model architecture, visualized on multiple diagrams, from a global overview to more detailed inspections of components. The overview visualizes how input signals are transformed by the model to obtain the loss, necessary for training. Then, two main components are inspected in more detail: the denoising model and the semantic encoder. Finally, the LRU block used in the semantic encoder is explained.

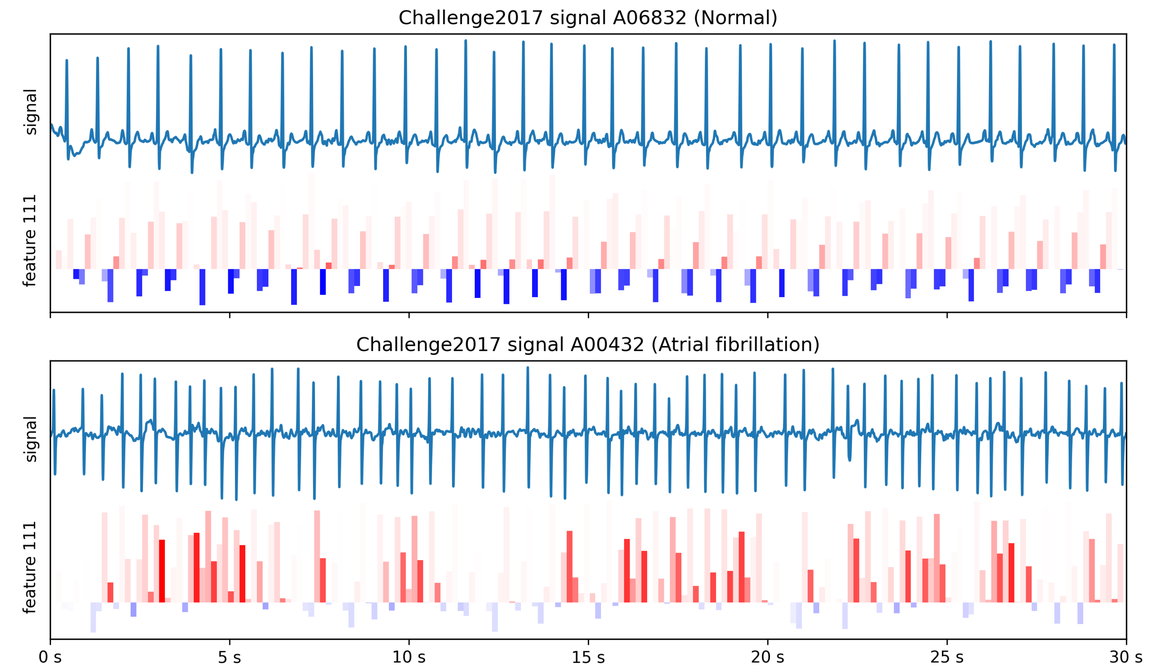

Figure 2: Visualization of how a single semantic embedding feature z_{111} is constructed by the Attention Reducer for two selected Challenge2017 signals: the top one labeled as "Normal" and the bottom as "Atrial fibrillation". The final value z_{111} is computed by the Attention Reducer as a weighted average of values $V_{111,t}$ (visualized by bar heights) with weights $S_{111,t}$ (visualized by bar color intensities). The bars are red for positive values and blue for negative values. The 111^{th} feature was selected as the one most correlated with atrial fibrillation.

Acknowledgements

This work was part of the project titled "Remote treatment of cardiovascular disease using a non-invasive, AI-based cardiac shockwave therapy device to induce tissue regeneration in infarcted heart muscles" funded by the Eureka Eurostars program (Project No. 2569) through the National Centre for Research and Development (NCBiR) in Poland (InnovativeSMEs/3/28/2023).