A DENOISING VAE FOR INTRACARDIAC TIME SERIES IN ISCHEMIC CARDIOMYOPATHY

Samuel Ruipérez-Campillo *
Dept. of Computer Science
ETH Zurich
samuel.ruiperezcampillo@inf.ethz.ch

Alain Ryser, Thomas M. Sutter,
Dept. of Computer Science
ETH Zurich
{alain.ryser,thomas.sutter}@inf.ethz.ch

Ruibin Feng, Prasanth Ganesan, Brototo Deb, Kelly A. Brennan, Maxime Pedron, Albert J. Rogers
Dept. of Medicine
Stanford University
{ruibin, prash030, bdeb7, kbrenn, mpedron, rogersaj}@stanford.edu

Maarten Z.H. Kolk, Fleur V.Y. Tjong,
Dept. of Cardiology
Amsterdam University
{m.kolk,f.v.tjong}@amsterdamumc.nl

Sanjiv M. Narayan†**,**
Dept. of Medicine
Stanford University
sanjiv1@stanford.edu

Julia E. Vogt†
Dept. of Computer Science
ETH Zurich
julia.vogt@inf.ethz.ch

ABSTRACT

In the field of cardiac electrophysiology (EP), effectively reducing noise in intracardiac signals is crucial for the accurate diagnosis and treatment of arrhythmias and cardiomyopathies. However, traditional noise reduction techniques fall short in addressing the diverse noise patterns from various sources, often non-linear and non-stationary, present in these signals. This work introduces a Variational Autoencoder (VAE) model, aimed at improving the quality of intra-ventricular monophasic action potential (MAP) signal recordings. By constructing representations of *clean* signals from a dataset of 5706 time series from 42 patients diagnosed with ischemic cardiomyopathy, our approach demonstrates superior denoising performance when compared to conventional filtering methods commonly employed in clinical settings. We assess the effectiveness of our VAE model using various metrics, indicating its superior capability to denoise signals across different noise types, including time-varying non-linear noise frequently found in clinical settings. These results reveal that VAEs can eliminate diverse sources of noise in single beats, outperforming state-of-the-art denoising techniques and potentially improving treatment efficacy in cardiac EP.

1 INTRODUCTION

Arrhythmias and cardiomyopathies pose significant challenges in cardiac healthcare. These conditions not only impact patient well-being but also complicate the clinical approach to treatment (Benjamin et al., 2018; Gopinathannair et al., 2015). Ablation therapy, a key intervention for many arrhythmias, heavily relies on the accurate interpretation of intracardiac signals (Huang & Miller, 2019). Additionally, cellular mechanisms reflected on these intracardiac recordings may reveal risk for cardiomyopathies (Rogers et al., 2021). Nevertheless, the availability of databases containing intracardiac signals for these conditions remains exceedingly scarce. Additionally, the complexity of these signals and their sensitivity to noise present substantial obstacles, frequently hindering effective treatment (Starreveld et al., 2020).

Common traditional denoising methods such as template matching (Houben et al., 2006), beat averaging (Ng et al., 2006) and bandpass filtering (Venkatachalam et al., 2011), fall short when dealing

* First Author's correspondence address: ETH Zürich, CAB G15.2, Universitätstrasse 6, 8092, Switzerland.
† Co-senior authors.

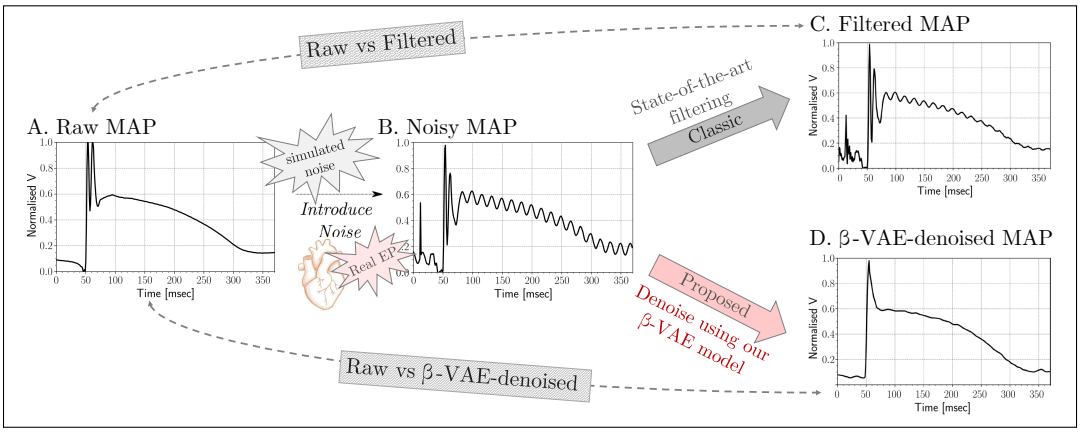

Figure 1: Denoising Pipeline. **A.** Raw intracardiac MAP recording from a patient with ischemic cardiomyopathy. **B.** Noisy counterpart of (A), following the introduction of various sources of simulated noise and EP noise. **C.** De-noised MAP time series obtained after applying the best-performing filter. **D.** Denoised MAP time series using our proposed β-VAE method.

with the various noise sources (De Groot et al., 2022; Stevenson & Soejima, 2005). In particular, electrophysiological (EP) noise, which arises from various factors like patient movement, electronic interference from medical devices, and physiological variations, becomes especially challenging due to its non-linear and non-stationary nature, rendering conventional denoising methods ineffective and demanding the use of an alternative approach (Starreveld et al., 2020). To address this challenge, we propose the use of Variational Autoencoders (VAEs) (Kingma & Welling, 2013). We hypothesize that VAEs have the potential to learn and interpret physiological morphologies of signals in this context, providing a novel solution to eliminate EP noise outperforming existing filtering techniques.

We develop a model capable of identifying and mitigating EP noise effects in easily verifiable clinically, yet challenging-to-obtain signal types. Our model, along with a unique dataset, presents a versatile denoising method that can be applied to less clinically verifiable signals, potentially advancing cardiac care and improving the outcomes of ablation therapy.

2 Methods

2.1 Denoising VAE Model

Denoising Autoencoders (AEs) (Vincent et al., 2008) are a type of encoder-decoder neural network that removes noise from corrupted or noisy data by learning to reconstruct the initial data $\mathbf{x}$ from its noisy counterpart $\tilde{\mathbf{x}}$. With g as the encoding function and h as the decoding counterpart, the primary objective becomes $\mathcal{L}_{DAE} = \|\mathbf{x} - h_{\varphi}\left(g_{\theta}(\tilde{\mathbf{x}})\right)\|^2$.

VAEs (Kingma & Welling, 2013) are a class of generative models designed to learn a probabilistic mapping between observed data and latent variables. Differing from traditional AEs, VAEs explicitly model the underlying data distribution using a generative model $p(\mathbf{x}, \mathbf{z})$, where $\mathbf{z}$ is the latent representation of the input. The ultimate goal of VAEs is to maximize the marginal likelihood $p_{\theta}(\mathbf{x})$, which quantifies the model's effectiveness in explaining the observed data. To implement the VAE denoising model, we aim to maximize the tractable evidence lower bound (ELBO), decomposed in the Kullback-Leibler ($\mathbb{KL}$) divergence and the expected log-likelihood. The $\mathbb{KL}$ divergence quantifies the loss of information when approximating a posterior distribution to a simpler prior, typically a standard normal distribution. Additionally, it acts as a regularization term by encouraging similarity between these distributions (Kingma & Welling (2013)). On the other hand, when assuming a Gaussian likelihood with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}^2$, the expected log-likelihood reduces to the expectation of the mean squared error (MSE) between $\mathbf{x}$ and $\mu(\mathbf{z})$ scaled up by a constant. We use Monte Carlo sampling to approximate such expectation. Lastly, following Higgins et al. (2016), we augment the original VAE framework with a β-weighting component to the $\mathbb{KL}$ term that modulates

the learning constraints applied to the model. The ELBO becomes:

$$\mathbf{ELBO} = \mathbb{E}_{q_\varphi(\mathbf{z}|\boldsymbol{x})}\left[\log p_\theta(\mathbf{x} \mid \mathbf{z})\right] - \beta \cdot \mathbb{KL}\left(q_\varphi(\mathbf{z} \mid \boldsymbol{x}) \| p_\theta(\mathbf{z})\right). \tag{1}$$

For further details on the architecture of the VAE and its formulation, we refer to Appendices B.2 to B.4.

2.2 Noise Library

Given the lack of ground truth clean MAP signals for noisy recordings, we conducted simulations to replicate prevalent forms of interference present in intracardiac signals. These include white noise, baseline wander, powerline interference, spike and truncation artifacts, as well as *semi-synthetic* EP noise extracted from real physiological recordings. These interference patterns were later introduced into the recognizable MAPs – see fig. 1.A.

We modeled Gaussian noise as a stationary and ergodic random process with zero mean and an autocorrelation function solely dependent on time lag. For baseline wander we combined a set of sinusoidal waves spanning a randomized range of low frequencies, from 0.01 to 0.3 Hz. To simulate powerline interference, we employed a single sinusoidal time series at 50Hz, adjusting the amplitude to mimic interference generated by various electrical devices and inadequate shielding. We introduced spike artifacts as finite discrete Dirac δ-functions of varying amplitude, occurring within the initial segment of the MAP morphology – before the upstroke. Additionally, we approximated truncation artifacts as multiplicative noise, using a square function with values set to one over the portion of the signal intended for retention while zeroing out the remainder.

The computation of EP noise involved extracting noise from clinical recordings, operating on the premise that there is minimal variability in the true MAP morphology within a single patient. This is based on the understanding that MAP signals deviating significantly from the average recorded signal are likely to encompass noise stemming from diverse sources. As a result, we constructed a library of physiological noise time series extracted from patients' signals. These were subsequently merged to generate distinctive additive EP noise time series – see an illustration in fig. 1.B.

3 Experimental Settings

Subject Recruitment and Study Protocol The study involved 53 individuals with low ventricular ejection fraction ($\leq 40\%$) and coronary artery complications, undergoing ventricular stimulation. After exclusions, 42 patients remained. Conscious sedation was administered, and MAP intracardiac signals were recorded. Signals were pre-processed, resulting in 5706 individual MAP time series. The analysis focused on voltage-time series of cardiac beats within 370 ms windows, following post-alignment and artifact removal Alhusseini et al. (2020). Further details, including summarized baseline characteristics for the patient population, are described in appendix B.1.

Baselines There is a notable lack of established benchmarks for denoising intracardiac signals, particularly MAPs, in both laboratory and clinical settings. We have explored various classic signal processing filtering techniques commonly used in clinical practice (Starreveld et al., 2020; De Groot et al., 2022). Among these techniques, the 5th order Butterworth filter (Yusuf et al., 2020) is often considered the standard choice (De Groot et al., 2022; Venkatachalam et al., 2011), providing a solid baseline with potential efficacy in our specific dataset. The N-th order continuous-time Butterworth filter is characterized by its transfer function $H(s) = \prod_{k=1}^{N}\left(1 - \frac{s}{p_k}\right)^{-1}$ with poles p_k located on the negative half-plane on a circle of radius ω_c, as described by $p_k \triangleq \omega_c e^{j\frac{\pi(2k+N-1)}{2N}}$.

Evaluation Metrics We assessed the model's effectiveness from diverse angles. This included using Pearson's Correlation Coefficient (PCC) to measure the linear relationship between the original and denoised signals (Eq. 9); calculating the Root Mean Squared Error (RMSE) to assess time domain alignment (Eq. 10); and measuring the Power Signal to Noise Ratio (PSNR) (Eq. 11) to contrast signal power with noise power, as quantified by the Mean Squared Error (MSE) – see appendix B.5.

4 RESULTS

Noise poses a challenge in understanding physiological time series and is not effectively removed by current techniques in the clinic De Groot et al. (2022); Starreveld et al. (2020), commonly negatively impacting treatment outcomes in cardiomyopathies or arrhythmias (de Bakker, 2019).

Table 1: De-noising Results to All Noise Types

Set	Labels	RMSE (x10-3) ↓	PCC ↑	PSNR ↑
Train	Noisy	14.98 ± 0.40	0.867 ± 0.007	20.52 ± 0.13
	Filtered	13.67 ± 0.31	0.883 ± 0.007	20.40 ± 0.08
	Ours	**2.68 ± 0.43**	**0.990 ± 0.001**	**27.79 ± 0.72**
Test	Noisy	15.41 ± 0.60	0.864 ± 0.024	20.33 ± 0.24
	Filtered	14.45 ± 0.55	0.879 ± 0.022	20.21 ± 0.19
	Ours	**7.05 ± 0.89**	**0.967 ± 0.009**	**22.91 ± 0.51**

Our study leverages a VAE model to mitigate diverse sources of noise in physiological signals of the highly prevalent ischemic cardiomyopathies.

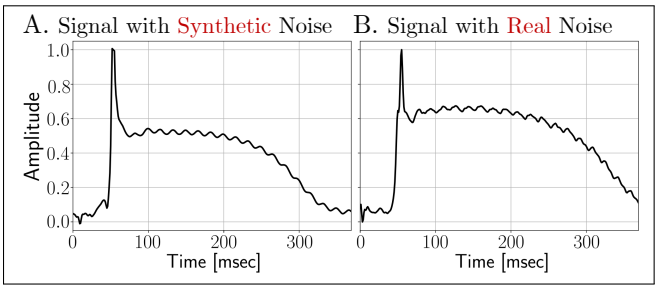

Figure 2: Synthetic Noise Reliability. **A.** Shows a raw MAP recording from an AF patient with added synthetic powerline interference noise. **B.** Presents a noisy recording of an MAP from a different patient, containing real powerline interference noise.

Table 2 (in Appendix A) provides an overview of the evaluation metrics for denoising all noise types except EP noise, and table 1 includes EP noise among the other types. These metrics compare the denoised MAP signals using our proposed β-VAE model, the benchmark (filtered) signals, and the noisy signals against the original clean ones. We find that our model successfully reduces different types of noise in intracardiac signals. Notably, it is effective in reducing clinical EP noise, which is usually difficult to filter out using traditional methods (de Bakker, 2019; Starreveld et al., 2020). The model's ability to lessen the impact of EP noise is confirmed by comparing the denoising performance to all noise types but EP noise (see supplementary table 2) against the denoising performance to all noise types, including EP noise (see table 1), observing only a minor decrease in performance.

One innovative aspect of our β-VAE model is how it denoises medical time series by learning representations of clean signals (and cellular processes) and uses this knowledge to identify and remove variations caused by noise. In our study, we focus on using MAPs, which are easier to visually verify by clinicians compared to other intracardiac signals (Franz, 1991). Compared to traditional filtering methods, the β-VAE model exhibits better denoising performance, especially in handling EP noise, which is challenging for standard filters given its non-linear and non-stationary nature. Figure 1 illustrates an example of a clinically interpretable MAP signal, as recognized by experts, alongside its noisy counterpart affected by various synthetic noise types. Figure 1.C-D illustrates the effectiveness of the proposed model against the best-performing clinical filter, showing that the VAE-reconstructed MAP signal not only removes synthetic noise but also eliminates non-physiological features from the original recording, such as fractionated activation upstrokes. Additionally, our study employed both synthetic and real noise in testing the model's capabilities. The differentiation between these two types of noise is crucial, as they present varying levels of complexity and realism in the context of EP signal interference.

Looking ahead, our research highlights several areas for further exploration. Bridging the domain gap between synthetic and real noise is important to explore further – yet, fig. 2 indicates a strong visual resemblance between a synthetic noisy MAP and actual recordings with real noise. On the ethical aspect, while our denoising method holds promise for improving diagnostic accuracy and treatment outcomes, we must pay careful attention to potential risks and biases introduced. Developing clinical metrics to assess denoising effectiveness would be key in this regard. Additionally, validating the effectiveness of such models across different conditions and patient groups may be key for a broader applicability, especially in underserved communities. Moreover, exploring the latent space of these models will offer deeper insights into their functioning and potential improvements. Ultimately, our study sets the stage for crucial clinical validation in real-time settings, assessing the

impact on patient outcomes and transitioning from theoretical models to practical medical solutions centered on patients. However, the responsible integration of AI technologies into healthcare workflows necessitates ongoing evaluation and validation to ensure patient safety and equitable access to care, something that will need to be taken into consideration moving forward.

5 CONCLUSIONS

Our study presents a VAE model designed to denoise intracardiac MAP signals, tackling the drawbacks of usual filtering methods used in the EP laboratory and clinical settings. This model, trained on a complex and elusive dataset mimicking clinical diagnostic scenarios, displays superior performance and resilience against different types of noise. Its skill in recreating important relevant features suggests it could be a big step forward for integration in real-time heart care settings, marking a significant advancement in cardiac care and improving the outcomes of EP therapies.

REFERENCES

Mahmood I Alhusseini, Firas Abuzaid, Albert J Rogers, Junaid AB Zaman, Tina Baykaner, Paul Clopton, Peter Bailis, Matei Zaharia, Paul J Wang, Wouter-Jan Rappel, et al. Machine learning to classify intracardiac electrical patterns during atrial fibrillation: machine learning of atrial fibrillation. *Circulation: Arrhythmia and Electrophysiology*, 13(8):e008160, 2020.

Emelia J Benjamin, Salim S Virani, Clifton W Callaway, Alanna M Chamberlain, Alexander R Chang, Susan Cheng, Stephanie E Chiuve, Mary Cushman, Francesca N Delling, Rajat Deo, et al. Heart disease and stroke statistics—2018 update: a report from the american heart association. *Circulation*, 137(12):e67–e492, 2018.

Jacques MT de Bakker. Electrogram recording and analyzing techniques to optimize selection of target sites for ablation of cardiac arrhythmias. *Pacing and Clinical Electrophysiology*, 42(12): 1503–1516, 2019.

Natasja MS De Groot, Dipen Shah, Patrick M Boyle, Elad Anter, Gari D Clifford, Isabel Deisenhofer, Thomas Deneke, Pascal Van Dessel, Olaf Doessel, Polychronis Dilaveris, et al. Critical appraisal of technologies to assess electrical activity during atrial fibrillation: a position paper from the european heart rhythm association and european society of cardiology working group on ecardiology in collaboration with the heart rhythm society, asia pacific heart rhythm society, latin american heart rhythm society and computing in cardiology. *EP Europace*, 24(2):313–330, 2022.

Michael R Franz. Method and theory of monophasic action potential recording. *Progress in cardiovascular diseases*, 33(6):347–368, 1991.

Rakesh Gopinathannair, Susan P Etheridge, Francis E Marchlinski, Francis G Spinale, Dhanunjaya Lakkireddy, and Brian Olshansky. Arrhythmia-induced cardiomyopathies: mechanisms, recognition, and management. *Journal of the American College of Cardiology*, 66(15):1714–1728, 2015.

Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. In *International conference on learning representations*, 2016.

Richard PM Houben, Natasja MS de Groot, Fred W Lindemans, and Maurits A Allessie. Automatic mapping of human atrial fibrillation by template matching. *Heart Rhythm*, 3(10):1221–1228, 2006.

Shoei K Stephen Huang and John M Miller. *Catheter Ablation of Cardiac Arrhythmias E-Book: Catheter Ablation of Cardiac Arrhythmias E-Book*. Elsevier Health Sciences, 2019.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Jason Ng, Alan H Kadish, and Jeffrey J Goldberger. Effect of electrogram characteristics on the relationship of dominant frequency to atrial activation rate in atrial fibrillation. *Heart Rhythm*, 3 (11):1295–1305, 2006.

Albert J Rogers, Anojan Selvalingam, Mahmood I Alhusseini, David E Krummen, Cesare Corrado, Firas Abuzaid, Tina Baykaner, Christian Meyer, Paul Clopton, Wayne Giles, et al. Machine learned cellular phenotypes in cardiomyopathy predict sudden death. *Circulation Research*, 128 (2):172–184, 2021.

Roeliene Starreveld, Paul Knops, Maarten Roos-Serote, Charles Kik, Ad JJC Bogers, Bianca JJM Brundel, and Natasja MS de Groot. The impact of filter settings on morphology of unipolar fibrillation potentials. *Journal of Cardiovascular Translational Research*, 13:953–964, 2020.

William G Stevenson and Kyoko Soejima. Recording techniques for clinical electrophysiology. *Journal of cardiovascular electrophysiology*, 16(9):1017–1022, 2005.

KL Venkatachalam, Joel E Herbrandson, and Samuel J Asirvatham. Signals and signal processing for the electrophysiologist: part ii: signal processing and artifact. *Circulation: Arrhythmia and Electrophysiology*, 4(6):974–981, 2011.

Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, pp. 1096–1103, 2008.

Samson D Yusuf, Francis C Maduakolam, Ibrahim Umar, and Abdulmumini Z Loko. Analysis of butterworth filter for electrocardiogram de-noising using daubechies wavelets. *SSRG International Journal of Electronics and Communication Engineering*, 7(4):8–13, 2020.

A SUPPLEMENTARY RESULTS

Additionally, Table 2 offers an overview of the evaluation metrics for denoising all noise types except EP noise, complementing the information provided in Table 1.

Table 2: De-noising results to all noise sources but EP

Set	Labels	RMSE (x10-3) ↓	PCC ↑	PSNR ↑
Train	β-VAE	**2.18 ± 0.44**	**0.992 ± 0.001**	**28.61 ± 0.86**
	Filtered	11.76 ± 0.27	0.899 ± 0.005	21.09 ± 0.13
	Noisy	12.88 ± 0.20	0.884 ± 0.006	21.34 ± 0.13
Test	β-VAE	**6.25 ± 0.86**	**0.970 ± 0.009**	**23.39 ± 0.59**
	Filtered	12.09 ± 0.63	0.899 ± 0.017	21.04 ± 0.21
	Noisy	12.91 ± 0.80	0.885 ± 0.018	21.30 ± 0.29

This comparison allows for an assessment of the model's generalizability in denoising EP noise.

In addition, Figure 2 illustrates the strong visual resemblance between a synthetically constructed MAP signal with added synthetic noise, and a noisy signal as recorded directly from the device, featuring real noise.

B SUPPLEMENTARY METHODS

B.1 EXPANDED SUBJECT RECRUITMENT AND STUDY PROTOCOL

The research study protocol was approved by the Ethics Committee of Stanford University. The research protocol encompassed a cohort of 53 individuals who exhibited low ventricular ejection fraction ($\leq$40%) along with coronary artery complications, scheduled for ventricular stimulation. Those individuals with a history of prolonged ventricular arrhythmias, sudden cardiac arrest, or fewer than 50 reliable monophasic action potential (MAP) recordings due to technical inconveniences were excluded from the study, reducing the population to 42 patients. The study was conducted under conscious sedation, utilizing midazolam and fentanyl. Specialized intracardiac signals were recorded using a dedicated 7F MAP catheter from Boston Scientific, MA, in conjunction with standard catheters. The high-fidelity MAP catheter was introduced either transvenously or through the aorta into the ventricles, with signals recorded using a physiological recorder from Bard-Boston Scientific in Marlborough, MA. These signals were filtered at various frequencies and digitized at a rate of 1 kHz. In brief, the study employed a total of 5706 individual MAP EGM time series exported at a 16-bit digital resolution. The analysis centered on voltage-time series of cardiac beats, which were partitioned within 370 ms windows obtained following alignment and the removal of artifactual deviations. Table 3 outlines the baseline characteristics of the entire patient population.

Table 3: Summary Baselines Characteristics of the Patient Population

Variable	All Subjects (n=42)
Age, y	64.7 ± 13.0
LVEF a, %	27.0 ± 7.6
QRS Duration, ms	126 ± 33
Hypertension, % (n)	19.0 (8)
Diabetes Mellitus, % (n)	14.3 (6)
BNP b, pg/ml (median, IQR)	341 (157-999)

B.2 NETWORK ARCHITECTURE

The encoder module processes input data (batch size = 32, input dim = 370) through six 1D convolutional layers, followed by flattening and two linear layers to produce mean and log variance of the

latent distribution, with Leaky-ReLU activations after each layer. The decoder module takes a latent representation (batch size = 32, latent size = 32) and reconstructs the original input using an initial linear layer and a series of 1D transposed convolutional layers, mirroring the encoder's architecture, with similar post-layer activations.

B.3 Denoising VAE Method

VAEs (Kingma & Welling (2013)) are a class of generative models designed to learn a probabilistic mapping between observed data and latent variables. Differing from traditional AEs, VAEs explicitly model the underlying data distribution using a generative model $p(\mathbf{x}, \mathbf{z})$. The ultimate goal of VAEs is to maximize the marginal likelihood $p_\theta(\mathbf{x})$, which quantifies the model's effectiveness in explaining the observed data. To implement the VAE denoising model, we aim to maximize the tractable evidence lower bound (ELBO), decomposed in the Kullback-Leibler ($\mathbb{KL}$) divergence and the expected log-likelihood. The $\mathbb{KL}$ divergence quantifies the loss of information when approximating a posterior distribution to a simpler prior, typically a standard normal distribution. Additionally, it acts as a regularization term by encouraging similarity between these distributions (Kingma & Welling (2013)). Assuming $q_\varphi\left(\mathbf{z} \mid \mathbf{x}^{(i)}\right) = \mathcal{N}\left(\mu, \operatorname{diag}\left\{\sigma_1^2, \ldots, \sigma_d^2\right\}\right)$ and a standard normal distribution for the prior, the $\mathbb{KL}$ divergence is derived as:

$$\mathbb{KL}\left(q_\varphi(\boldsymbol{z} \mid \boldsymbol{x}) \| p_\theta(\boldsymbol{z})\right) = \frac{\sum_{d=1}^{D}\left(1 - \mu_d^2 + \log\left({\sigma_d}^2\right) - \sigma_d^2\right)}{2}, \tag{2}$$

see Appendix B.4 for further detail.

Assuming the likelihood is Gaussian with mean μ and variance σ^2, the expected log-likelihood is expressed as:

$$\log p_\theta(\boldsymbol{x} \mid \boldsymbol{z}) = -\frac{\|\mathbf{x} - \mu(\mathbf{z})\|^2 + \sigma^2 D \log\left(2\pi\sigma^2\right)}{2\sigma^2}, \tag{3}$$

which reduces to the expectation of the mean squared error (MSE) between $\mathbf{x}$ and $\mu(\mathbf{x})$ scaled up by a constant.

Using Monte Carlo sampling to approximate such expectation, sampling Λ latent variables $\mathbf{z}(\lambda)$ from $q_\varphi(\boldsymbol{z} \mid \boldsymbol{x})$, we find that:

$$\mathbb{E}_{q_\varphi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x} \mid \mathbf{z})\right] \approx \frac{\sum_{\lambda=1}^{\Lambda} \log p_\theta\left(\mathbf{x} \mid \mathbf{z}^\lambda\right)}{\Lambda} \tag{4}$$

Lastly, following Higgins et al. (2016), we augment the original VAE framework with a β-weighting component to the $\mathbb{KL}$ term that modulates the learning constraints applied to the model. Hence, we define the objective as:

$$\mathbf{ELBO} = \mathbb{E}_{q_\varphi(\mathbf{z}|\boldsymbol{x})}\left[\log p_\theta(\mathbf{x} \mid \mathbf{z})\right] - \beta \cdot \mathbb{KL}\left(q_\varphi(\mathbf{z} \mid \boldsymbol{x}) \| p_\theta(\mathbf{z})\right). \tag{5}$$

B.4 KL Divergence between two Gaussians and Specifics for our Prior

Given two Gaussian distributions

$$q_\phi(\mathbf{z}|\mathbf{x}^{(i)}) = \mathcal{N}\left(\mu, \operatorname{diag}\left\{\sigma_1^2, \ldots, \sigma_D^2\right\}\right)$$

and

$$p_\theta(\mathbf{z}) = \mathcal{N}(0, \mathbf{I}),$$

we want to compute the $\mathbb{KL}$ divergence between them.

The general formula for the $\mathbb{KL}$ divergence between two Gaussian distributions $\mathcal{N}(\mu_1, \Sigma_1)$ and $\mathcal{N}(\mu_2, \Sigma_2)$ is:

$$\mathbb{KL}\left(\mathcal{N}(\mu_1, \Sigma_1) \,||\, \mathcal{N}(\mu_2, \Sigma_2)\right) = \frac{1}{2}\Bigg(\text{Tr}(\Sigma_2^{-1}\Sigma_1) + (\mu_2 - \mu_1)^T \Sigma_2^{-1}(\mu_2 - \mu_1) - D - \log\frac{\det \Sigma_1}{\det \Sigma_2}\Bigg) \tag{6}$$

For our specific case, we have $\mu_1 = \mu$, $\Sigma_1 = \text{diag}\{\sigma_1^2, \ldots, \sigma_D^2\}$, $\mu_2 = 0$, and $\Sigma_2 = \mathbf{I}$.

Substituting these into the formula, we get:

$$\begin{aligned}\mathbb{KL}\left(q_\phi(\mathbf{z}|\mathbf{x}^{(i)}) \,||\, p_\theta(\mathbf{z})\right) &= \frac{1}{2}\Bigg(\text{Tr}(\mathbf{I}^{-1}\,\text{diag}\{\sigma_1^2, \ldots, \sigma_D^2\}) \\ &\quad + (0-\mu)^T \mathbf{I}^{-1}(0-\mu) \\ &\quad - D - \log\frac{\det \text{diag}\{\sigma_1^2, \ldots, \sigma_D^2\}}{\det \mathbf{I}}\Bigg) \qquad (7) \\ &= \frac{1}{2}\sum_{d=1}^{D}\left(1 + \log\left((\sigma_d)^2\right) - (\mu_d)^2 - (\sigma_d)^2\right) \qquad (8)\end{aligned}$$

B.5 Evaluation Metrics

The implementation of the evaluation metrics was performed as described by the equations below, including PCC to measure the linear relationship between the original and denoised signals:

$$\text{PCC}_{s_1,s_2} = \frac{\sum_{i=1}^{n}\left(s_{1_i} - \overline{s_1}\right)\left(s_{2_i} - \overline{s_2}\right)}{\sqrt{\sum_{i=1}^{n}\left(s_{1_i} - \overline{s_1}\right)^2}\sqrt{\sum_{i=1}^{n}\left(s_{2_i} - \overline{s_2}\right)^2}} \tag{9}$$

where s_1 and s_2 represent the raw MAP and its denoised counterpart, n denotes the number of data points in the signals, i refers to the sample number within the respective signals, and $\overline{s_1}$ and $\overline{s_2}$ represent the mean values of s_1 and s_2 respectively.

The RMSE allowed us to assess the time domain alignment and sample-wise similarity of both the raw signal and denoised counterpart, implemented as:

$$\text{RMSE}_{s_1,s_2} = \sqrt{\frac{\sum_{i=1}^{n}\left(s_{1_i} - s_{2_i}\right)^2}{n}} \tag{10}$$

where s_1 and s_2 represent the signals being compared, n denotes the number of data points in the signals, and s_{1_i} and s_{2_i} represent individual data points within the respective signals.

Lastly, the PSNR permitted contrasting signal power to noise power, the latter quantified by the MSE:

$$\text{PSNR} = 20 \cdot \log_{10}\left(\text{MAX}_s\right) - \log_{10}(\text{MSE}) \tag{11}$$

where MAX_s denotes the maximum possible pixel value of the signal (e.g., for images, this would typically be 255 for 8-bit images).