

You're your own best teacher: A Self-Supervised Learning Approach For Expressive Representations in Time Series Generation

Johan Vik Mathisen

June 8, 2024

TODO: How the order of things should be is not clear to me yet. Would like some pointers.

0.1 Acknowledgements

This thesis is the culmination of 6 years of studies in Trondheim. There are many people that deserve big thank you for making these years so memorable. If you are reading this, chances are quite high that your name should be here.

I would firstly like to thank my supervisors, Erlend Aune and Daesoo Lee, for introducing me to this fascinating research field, and being open minded and supportive throughout the process. Erlend Lokna, a dear friend and great collaborator, for your indispensable contributions on this project. Developing advanced machine learning models is both challenging and time consuming, and synchronously banging our heads made it all a little easier. From the many late nights of writing, experimentation and discussion, to trash talking over the ping pong table, this project would not be the same without you. It was sweet to end on a 9-1 victory, which once and for all cemented me as the top dog, and nothing (except a rematch) could change that.

I would too like to thank my family, mom, dad, Otto, Andreas and Tiril for being a continuous source of inspiration and motivation. You move so graciously through life, sharing both of wisdom and skill, paving the way for the youngest in the flock. Erik, Ludvig and Sivert, my day ones. You have been there through all the ups and downs, with unwavering support just a small phone call away. My roommates, Pernille and Eldrun, for being devilishly funny and handling me so well as I go full goblin mode. Preben, Elias and the rest of the Matteland mafia, despite placing bets on whether I would last one or two weeks into my bachelors, for their endless curiosity and fearlessness in face of difficult mathematics, pushing me to learn way more than I ought to have.

Finally, Tindeggruppa, the university climbing group, my crew. You have such a special place in my heart. You reintroduced the joy of movement to my life, after abruptly retiring as professional handball player. You provided an arena to play, connect with nature and explore. Someone once told me, to have a great time as a student in Trondheim, you need to find your little community, be it on Samfundet or some part of NTNUI. I can wholeheartedly say I found mine, and to all my climber friends, such loving, crazy and interesting people, thank you. See you on top of some remote peak very soon.

0.2 Motivation

In this thesis, we are interested in representation learning for improved generative modelling of time series. Time series, as the name kinda spoils, is data with a time component. With the ever forward pointing arrow of time, it is really no surprise that these types of data are everywhere. Time series data is ubiquitous across various domains, ranging from finance and healthcare to environmental monitoring and industrial processes. Its importance lies in its ability to capture sequential observations over time, offering insights into underlying patterns, trends, and behaviors. Analyzing time series data can enable us to make informed decisions, predict future trends, and understand the dynamics of complex systems. However, traditional statistical models often struggle to capture the intricate structures present in time series data. These models typically assume linear relationships and stationary behavior, failing to adequately represent the complexities inherent in real-world time series. As a result, there is a pressing need for more sophisticated modeling approaches that can capture the non-linear, non-stationary, and high-dimensional nature of time series data.

Time series generation (TSG), a branch of machine learning focused on synthesizing time series with similar characteristics as some observed data. TSG is applicable across numerous fields, whether it's generating synthetic data for training machine learning models, simulating realistic scenarios for predictive analytics, or augmenting limited datasets for robust analysis. The ability to generate plausible and diverse time series is highly valuable, but it is also notoriously difficult.

Unsupervised representation learning offers a promising path for enhancing time series generation, as it has for natural language and image data. By automatically discovering and extracting meaningful features from raw data, representation learning can assist the modeling of complex relationships and structures without the need for explicit supervision. This approach is particularly appealing for time series data, where the underlying patterns may be latent and not easily found through manual inspection. The ability to learn expressive representations of time series data holds immense potential for advancing research and solving real-world problems.

0.3 What is this thesis about?

The overarching theme of this thesis is machine learning, with a focus on two specialized areas: generative modeling and representation learning.

For the reader unfamiliar with the subject, we attempt to give a gentle introduction to the content of this thesis. Machine learning algorithms are fundamentally pattern finders, designed to pick up on patterns in data and utilize these patterns for various tasks. These tasks may include distinguishing images of dogs

from cats, identifying fraudulent bank transactions from legitimate ones, or predicting tomorrow's weather. This is known as *predictive modeling*, which aims to use patterns in existing data to make predictions about unseen data. In *generative modeling*, the objective shifts to recognizing patterns that enable the creation of data that resembles the training data. A now familiar example of this is ChatGPT, a generative model capable of producing text similar to its training input.

The second major component of this thesis is representation learning, which involves viewing information from different perspectives. Just as diverse perspectives on an issue can highlight various aspects and be useful in different contexts, different data representations can reveal distinct patterns. In machine learning, representation learning is about finding perspectives that are computationally useful. For instance, consider the sequence of numbers 1, 3, 7, 15, 31, ... One approach to finding a generating formula might involve examining the differences between consecutive numbers. However, a change in perspective, such as expressing the numbers in binary (i.e., 1, 11, 111, 1111, 11111), makes the pattern more apparent—simply add another 1 to each subsequent number. While this is a simple example, it illustrates how changing perspectives can clarify patterns, a process that is crucial in complex scenarios commonly encountered in modern machine learning.

In this thesis we attempt, and in many ways succeed, to find a better perspective on time series data, such that we can create new ones that resemble, but not completely mimic, the training data. We do this by squishing the data into a smaller space, in a way that preserves the most important information. In this compressed space, we too push similar looking time series to the same regions, which in a sense organizes it. Finally, using a method similar to the one used by ChatGPT, we create new data in the compressed space, and decompress them to get new time series. The primary contribution of this thesis lies in this compressed but structured perspective on time series data.

Technically speaking, we investigate possible enhancements to the TimeVQVAE model presented in [TimeVQVAE] by introducing a non-contrastive self-supervised loss to the tokenization model. We specifically examine if the representations learned are more expressive, in the sense that they improve the downstream classification accuracy, while simultaneously enable high quality reconstruction, and investigate how the learned representations affect the quality of the synthetic samples.

0.4 Research questions

We aim to leverage non-contrastive self-supervised learning to enhance the expressiveness of the discrete latent representations in TimeVQVAE. Non-contrastive SSL methods leverage augmented data as a supervisory signal during training,

and the particular choice of augmentations typically have great influence on the learned representations. Hence, our research revolves around the following four questions:

- RQ1:** Will a self supervised learning approach enhance downstream classification, while simultaneously enable high quality reconstruction?
- RQ2:** How does different augmentations influence reconstruction quality and downstream classification accuracy?
- RQ3:** Will more expressive representations improve synthetic sample quality?
- RQ4:** How does different augmentations influence synthetic sample quality?

0.5 Sustainability impact

With respect for the incredible work done around the globe to address the UN sustainable development goals (SDG) [Agenda2023], I will make no claim that sustainability and considerations regarding environmental impact have been a part of our work. But, as new policy from NTNU demand I reflect around the works relation to the SDGs, I will do so.

The field of AI and machine learning is in rapid development, and the fear of being left in the dust in the gold rush push large actors to train ever larger models. Though the modern generative models such as ChatGPT and Stable Diffusion are very impressive, it is hard to ignore their colossal environmental impact. These large models are both power hungry and water thirsty. According to an article in Nature magazine [AI_energy], Kate Crawford suggests that large AI systems will soon likely need energy on the scale of entire nations. Furthermore in [li2023making] they make the claim that the global demand for AI could account for freshwater withdrawals nearly half the volume of the United Kingdom by 2027. The current trajectory puts the AI field at odds with SDG 6, which aims to ensure the availability and sustainable management of water and sanitation for all.

While we believe that the generative AI has valuable applications and can be deployed without conflicting with the SDGs, it necessitates careful considerations. Our work on time series offers advantages because it is less resource-intensive than image and text generation. Additionally, work on time series has the benefit of being less flashy than image and text generation, and we consider TSG less likely to be used excessively, more easily aligning it with sustainable practices.

0.6 Collaboration and AI Declaration

During the work on our theses, Erlend Lokna and myself have collaborated extensively. This has been declared from the beginning and talked about continuously with our supervisors. Erlend, with his developer experience, has taken lead on

code development for our model, and should get credit for his high level of programming skills. I have, in addition to code contribution and being a source of ideas, taken the lead on data processing and visualization. We have exchanged ideas, possible paths forward and the overall structure of the theses. Additionally have conducted the same experiments, hence our works will have many similarities. All writing though, is done independently. The collaboration has been very fruitful, and working closely with Erlend has made the process quite enjoyable, despite him being lousy at ping pong.

Content of this thesis have been to some degree been made with the assistance of LLMs. As a non-native english speaker, ChatGPT has been helpful in improving grammar and readability. It has provided assistance in producing and debugging LaTeX and Python code throughout.

0.7 Thesis Overview

TODO: Under construction

We begin by building up the theoretical background for this thesis. There we introduce neural networks, and relevant architectures for our work, such as convolutional neural networks and transformers. We give an introduction to representation learning, highlighting standard evaluation protocols, and give an introduction to self-supervised learning (SSL), with emphasis on siamese architecture-based SSL. We present the Vector Quantized Variational Autoencoder, trying to highlight aspects brushed over in [VQVAE], building it up from Autoencoders. Lastly we go through the evaluation metrics used to assess our models.

In the related works we present models relevant to our work, including TimeVQVAE [TimeVQVAE] and MaskGIT [chang2022maskgit], as well as the non-contrastive SSL methods Barlow Twins [zbontar2021barlow] and VibCReg [lee2024computer].

The methodology section includes the introduction of the proposed model NC-VQVAE.

Experiments include details regarding implementations, hyperparameters and datasets used.

Results and discussion presents all results from our experiments. We discuss the results and answer our research questions.