George Smith-Kolff

Data-309 Project Brief

SOUND GENERATION WITH VARIATIONAL AUTO-ENCODERS.

MOTIVATION:

Sound has always been an interest of George's. Ever since he was young, the characteristics of a "sound" is something that was frequently questioned. This developed into a love of sound design with respect to electronic music. After undertaking a Data Science degree, he realized that he also has a passion for neural networks and their characteristics. This led him to pose the question, is there any way a neural network can generate a sound for them? And the answer lies in VAE'S.

DOMAIN:

This project will be an amalgamation of machine learning & audio engineering. There is no specific organization that is affiliated with this project. Instead, a research-based approach will be taken. Special challenges will consist of making sure the architecture of the network is well thought out and that the code is developed in such a way that further expansion of this project is undertaken with ease. (E.g., Employing a prompt-based architecture once the network can generate audio).

ARCHITECTURE:

VAE'S are an extension of Variational Encoders. VAE'S contain two key components. The first component is the encoder, which shares similarities with CNN'S (convolutional neural networks). The encoders job is to learn effective data encoding from a specific data set, which then gets passed into the bottleneck layer. The other component is the decoder which uses the concept of latent space to regenerate an observation similar to the dataset. These results are then backpropagated with KL-loss. After this, the inverse operation is applied to turn the complex number matrix back into a raw .WAV file.

George Smith-Kolff

Data-309 Project Brief

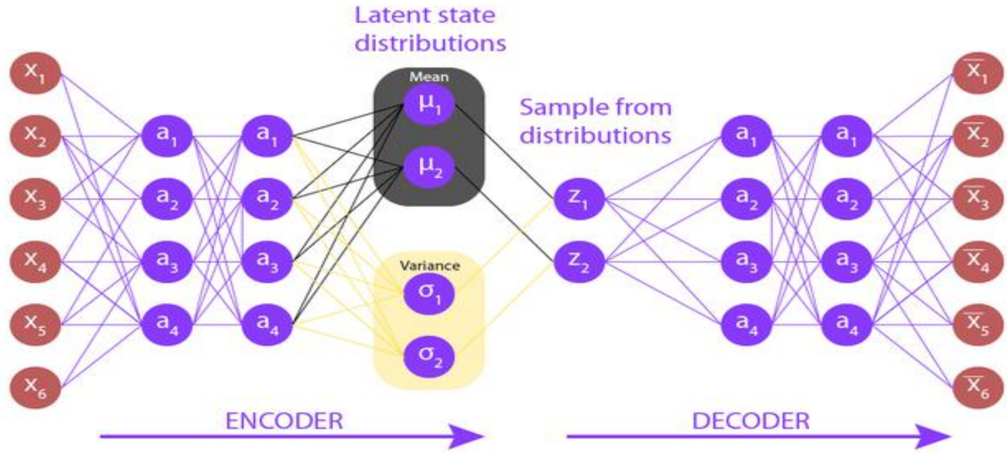
Latent state distributions
Mean
μ_1
μ_2
Variance
σ_1
σ_2
Sample from distributions
x_1 x_2 x_3 x_4 x_5 x_6
a_1 a_2 a_3 a_4
z_1 z_2
$\bar{x}_1$ $\bar{x}_2$ $\bar{x}_3$ $\bar{x}_4$ $\bar{x}_5$ $\bar{x}_6$
ENCODER
DECODER

Fig 1: Architecture of a simple VAE.

Source: https://www.geeksforgeeks.org/variational-autoencoders/

GOAL:

The goal of this project in its simplest implementation is to have a functioning VAE neural network that can generate unheard audio. We start with one constant style of sound (a folder full of snare drum samples) The data is then fed into a series of functions which turns the raw .WAV files into a three-dimensional complex number array, this is achieved through a short time inverse Fourier transformation. The data will then be stored in that state. Padding the arrays may be necessary to ensure size is preserved. This is the final state of the data before being passed into the network for training. The network will then be trained and optimized for the generation of new audio.
A series of new audio files will be sampled together and then compiled. These audio samples will then be used to display the findings of this project.

DATA:

The data used in this project will be a collection of lossless audio samples of different instruments, eg: snare samples, kick drum samples, bass samples. The samples will also be a collection of live & synthesized sounds. This will help keep the dataset diverse and hopefully aid in the creation of some interesting sounds. George is the sole owner of the data used for this project. Over the course of his music making career, he has built up an extensive collection of royalty free samples. But he is the only person who owns the rights to the data he has accessible on his computer. The data was produced by a collection of individuals who either: recorded the sounds or synthesized the sounds.

PLANNING:

The steps necessary to complete this project will loosely be broken into these steps:

- Research VAE'S and understand the math and architecture that is required to implement this network. (Currently underway)
- Curate the dataset that will be used to train the network, he will start off with just snare samples, and then once the network is trained successfully, he will try and open the framework up to allow for multiple different instruments. (Currently underway)
- Create his data loading functions that will turn the lossless .WAV into a complex numbered matrix that will then be stored locally. (70% complete)
- Start building the VAE for training (This will most likely be the most time-consuming component. George estimate 5 weeks to have this built as a lot of research may be needed as this section of the project progresses).
- Draft project report (he estimates 10-14 days for this).
- Draft project slides (he estimates 10-14 days for this).
- Train the network on the curated dataset (he estimates a week for this).
- Perform diagnostics and optimize learning rate if possible (This section will be time dependent).
- Start final copy of report.
- Tidy up any loose ends for the project, if time permits look at possible ways of extending the project to more than one type of audio sample (e.g., kick drum).
- Commit final project state to GITHUB.
- Convert code to Julia. (If time permits, as I do not know Julia).

RESOURCES:

For this project George will be using Python as it is at the forefront of machine learning. He will need access to GPU'S for the training section of this project. Additionally, it would be useful to have access to an instance of the math portal that will allow me to install external libraries and dependencies on my given IDE to get this project working. He plans on building the VAE in either PyTorch or TensorFlow, so these libraries will need to be installed. Access to research papers will be required, if this is an issue, George will notify Phil for access.

CONSTRAINTS:

The biggest constraint of this project George can see is time. This is purely because he will be doing this project on his own, so time management will be import as well as consistent progress reports to his supervisor Giulio. An additional constraint will be that the data used to train this network cannot be offered up as public data. This is because the royalties for these audio samples are for the users who have bought the royalties to use these audio samples.

References:

- GeeksforGeeks. (2022). Variational AutoEncoders. *GeeksforGeeks*. https://www.geeksforgeeks.org/variational-autoencoders/
- Rocca, J. (2021, December 11). Understanding Variational Autoencoders (VAES) - towards data science. *Medium*. https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73