BachBot - Bridging Classical Elegance with AI Innovation

Kyuil Lee kyuil@stanford.edu
Dezhi Yu dezhiy@stanford.edu
Yongkang Huang yka@stanford.edu

Summary

Johann Sebastian Bach (1685-1750)

In recent years, the field of music generation has witnessed growing interest, particularly within the realm of generative models. Johann Sebastian Bach, an iconic figure in classical music composition, has left a lasting legacy, influencing generations of musicians.

We focused on curating a dataset sourced from Bach's time-sequenced MIDI files and explored the music generation using various generative models, including Long Short-Term Memory (LSTM), Attention Mechanism, Variational Autoencoder (VAE), and Generative Adversarial Network (GAN) architectures,

Relevant Background Information

We explore various ways we can learn to generate discrete representations of music while also learning to mimic Bach's composition style. The fact that our input and output are discrete meant that we'd have to modify common generative architectures (VAE, GANs) to effectively handle discrete sequences of data. We explored using LSTMs with attention[1], variants of VAE (VAE with hierarchical decoder, e.g. MusicVAE[6], VQVAE[7]), and GANs (W-GAN[5] with LSTMs)

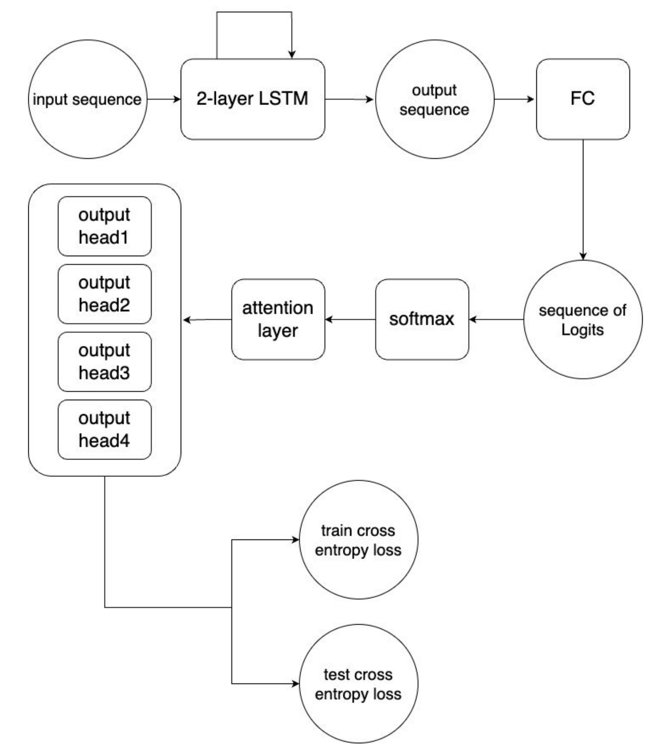
Dataset

We scraped all of Bach's works from the web in MIDI format, and converted them into a sequence of multi-hot vectors to represent notes being played at each timestep. The timestep was chosen per piece as the most common note duration that occurred in the piece. We preprocessed the sequences by removing long pauses, long repeated notes, etc. and customize the dataset to fit each of our models.

Technical methods

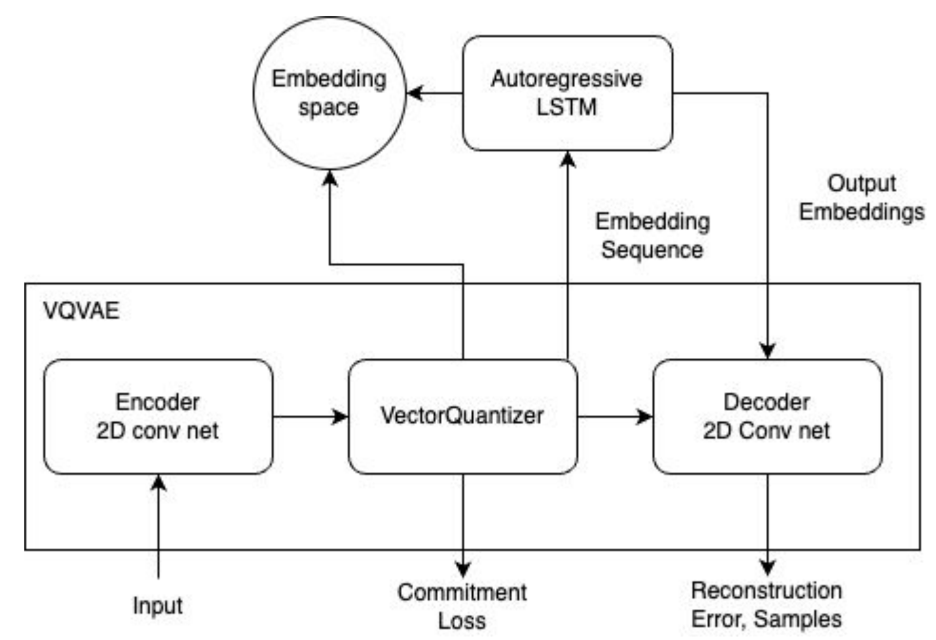
Autoregressive LSTM with Attention

- LSTM networks are ideal for sequence prediction problems like music generation because they can remember long-term dependencies in data.
- 2 LSTM + attention focus on specific parts of the input sequence, improving quality of generation
- Each head is responsible for predicting a different aspect of different 'voices'. The final output from each head, which represent different components.



VAE (VAE-LSTM, VQ-VAE)

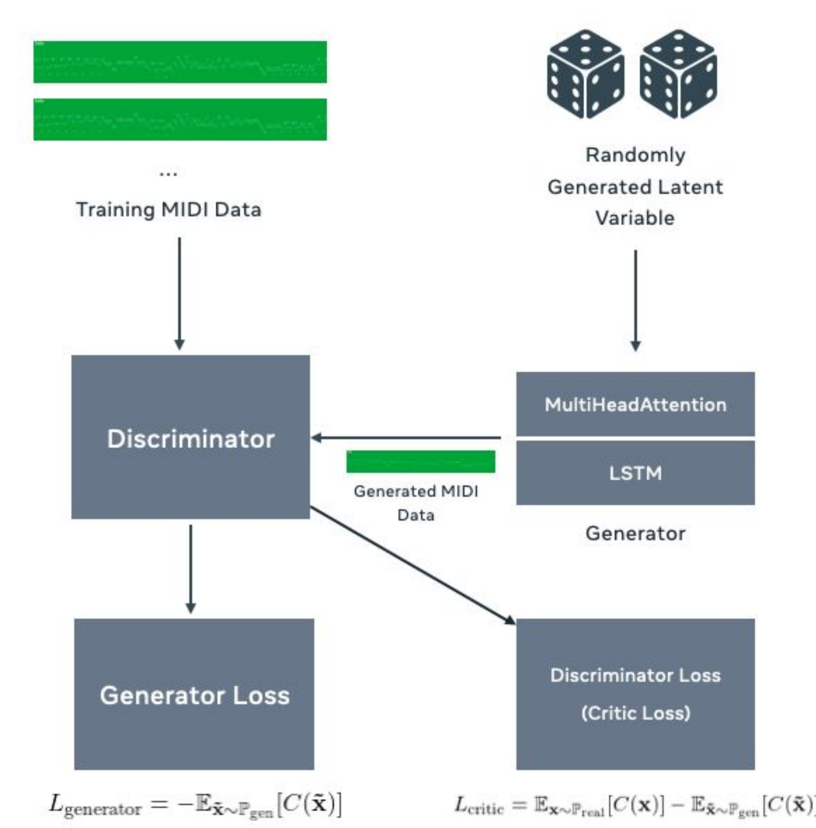
- VAE-LSTM
  - Modified VAE implementation with LSTMs
  - Tried multiple decoder architectures (hierarchical)
- VQ-VAE
  - Learned embeddings to represent groups of four consecutive notes
  - A 2-layer LSTM was autoregressively trained over the embedding space to generate samples



$$L = \log p(x|z_q(x)) + \beta||z_e(x) - \mathrm{sg}[e]||_2^2$$

GAN

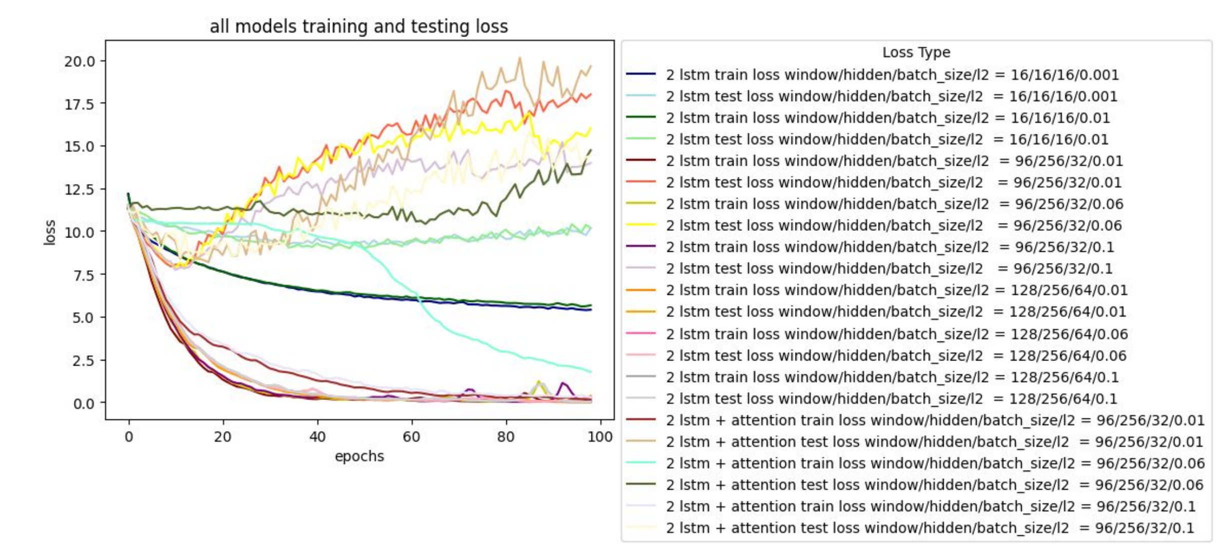
- Experimented with Generative Adversarial Networks (GANs) using various configurations of generators and discriminators.
- Selected combinations of generators and discriminators for optimal performance.
- Applied Minmax Loss and Wasserstein Loss
- Utilized multiple training heuristics to enhance the optimization process, based on sampled results



Evaluations

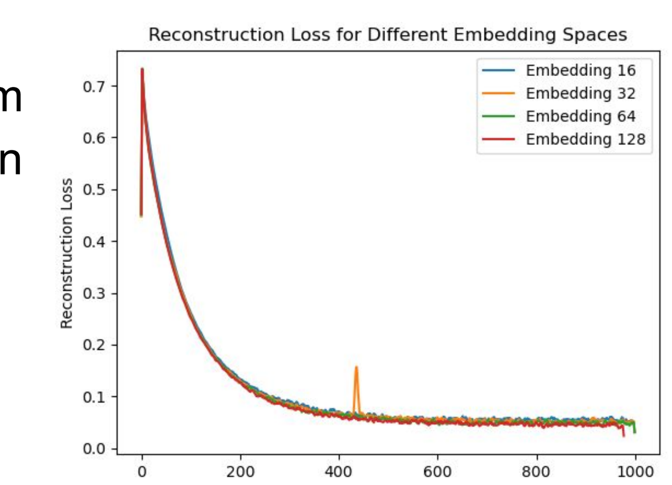
The model's performance was evaluated based on its ability to generate music that is stylistically similar to Bach's compositions:

- LSTM with attention used train/test loss as metric with different hyperparameter. 2 lstm + attention with 96 window_size, 256 hidden_size,32 batch_size, 0.01 L2 has the lowest loss and the best effect. Successfully generated realistic samples.



Autoregressive Loss Function Hyper Parameters Grid Search

- VAE
  - The VAE-LSTM models suffered from posterior collapse very early on, even with the hierarchical decoding
  - Avoided posterior collapse by quantizing the latent space, using Vector-Quantized VAEs
  - Successfully learned good 4-note patterns



- GANs rely on latent representations for their functioning, and thus, the evaluation of their generated results' similarity to the training dataset is predominantly based on subjective judgment.
  - Applied Wasserstein Loss to reduce the probability to gradient explosion; Experimenting with different generator and discriminator to resolve the mode collapse and style generalization.

Other information

1. Generating Original Classical Music with an LSTM Neural Network and Attention
2. Generating Long-Term Structure in Songs and Stories
3. Neural Machine Translation by Jointly Learning to Align and Translate
4. S-LSTM-GAN: Shared Recurrent Neural Networks with Adversarial Training
5. Wasserstein GAN
6. A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music
7. Neural Discrete Representation Learning