Music Generation using GANs

NC STATE UNIVERSITY

Authors: **Akanksha Bhattacharya, Varsha, Nagarajan, Vandan Phadke**
Advisor: Dr. Edgar Lobaton

Course Number: ECE 542
Team ID: 542-006

Instructor: Dr. Edgar Lobaton

Motivation and Goals

Generating **music** is notably different from generating images, texts or videos. The implicit complexity of modeling the sequence of notes and their corresponding frequencies to generate frequencies that are melodious, is a challenge of its own. While images and videos require extensive exploration of local patterns, each musical track embeds its own **temporal dynamics**, with a variety of notes, pitches, tones and other nuances that need to be accurately modelled. Many researchers have worked around this problem by using discrete symbolic representation of music which does not essentially capture such nuances. We take our cue from the success of **Generative Adversarial Networks (GANs**) and apply it to a **continuous representation** of music to generate melodies. In this project, we evaluate the performance of using different GAN architectures such as **FCN-RNN and C-RNN GANs** for music generation.

Overview

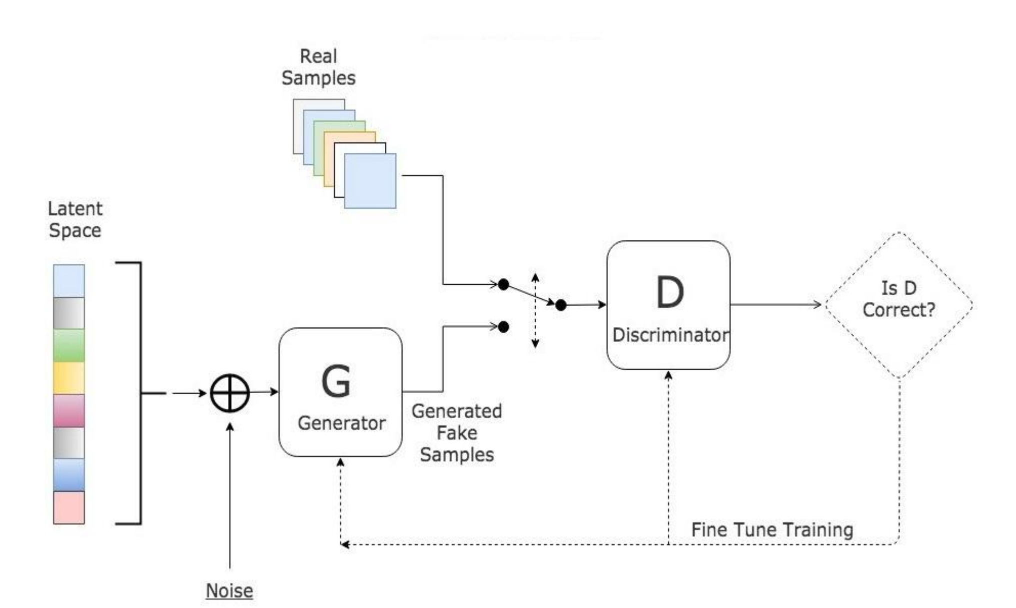
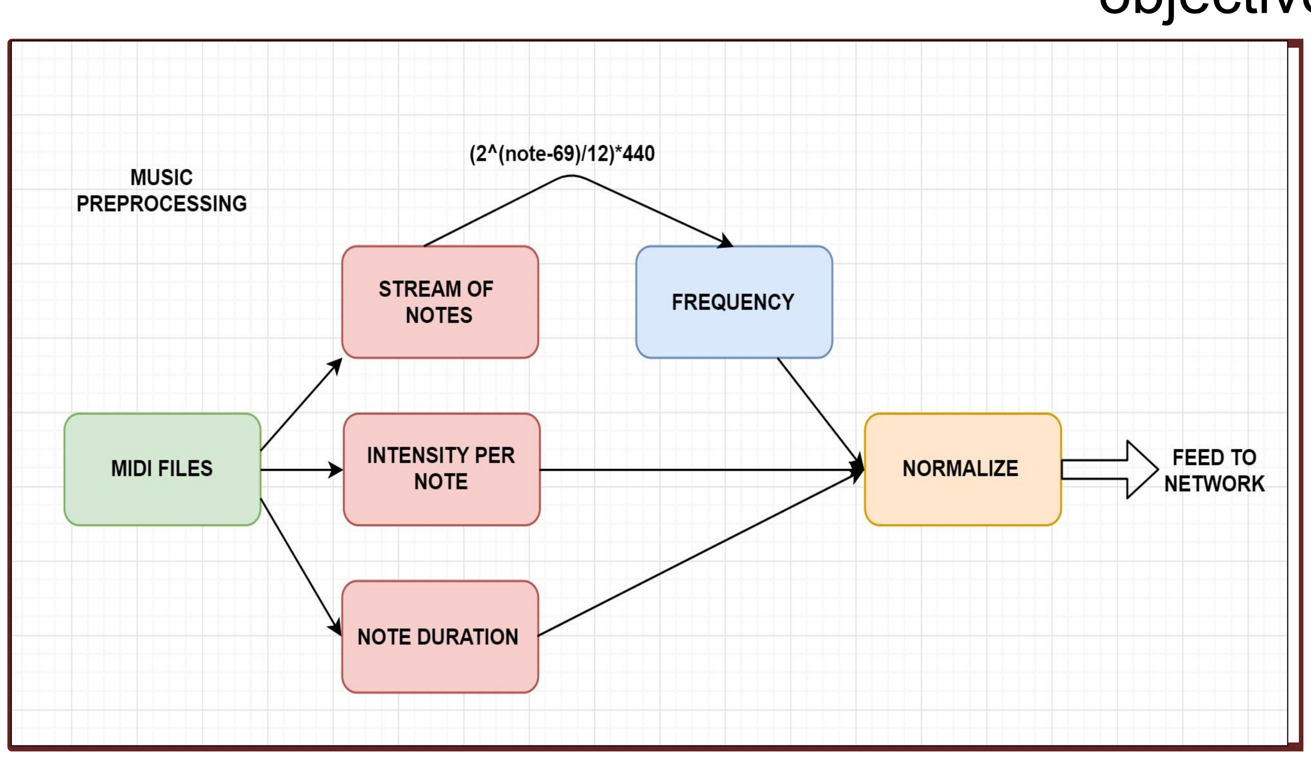


Fig 1: GAN Architecture

Generative Adversarial Networks (GANs) are a class of generative models that are known to work really well for text, images and video generation. The core of our proposed model is a generative adversarial network which aims at learning a discriminator to distinguish between real and generated data samples, and a generator that is trained to generate data samples indistinguishable from the real ones. Since objective evaluation of generative models is difficult, we use statistical metrics such as tone span, scale consistency, repetitions and tone uniqueness to evaluate the generated music.

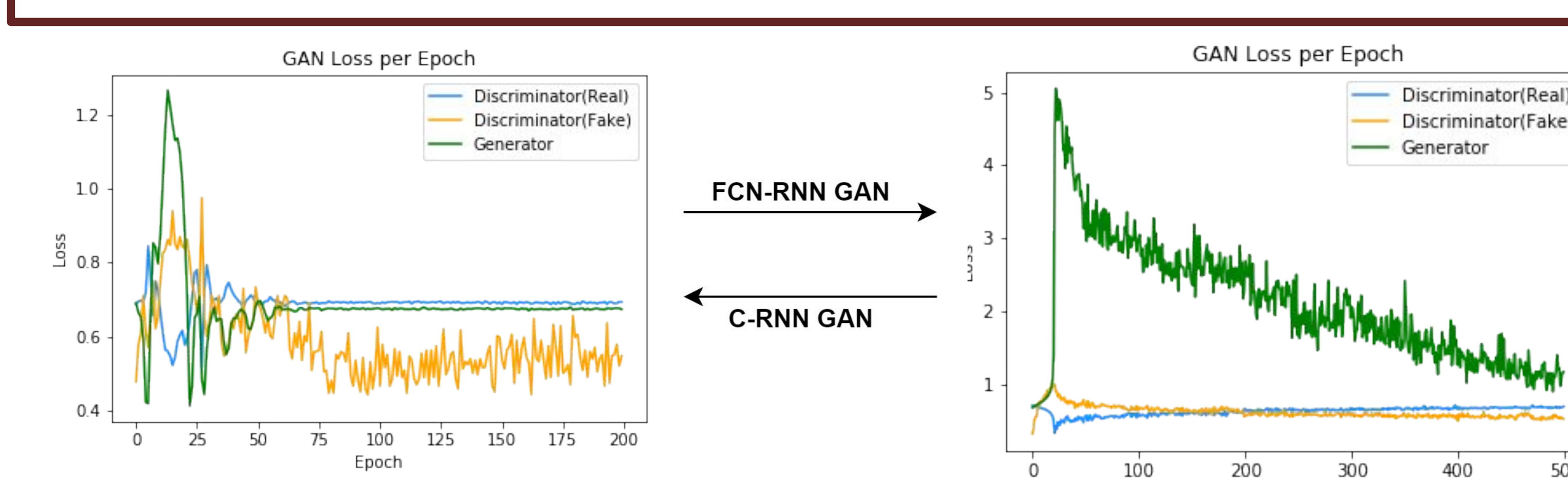
Dataset

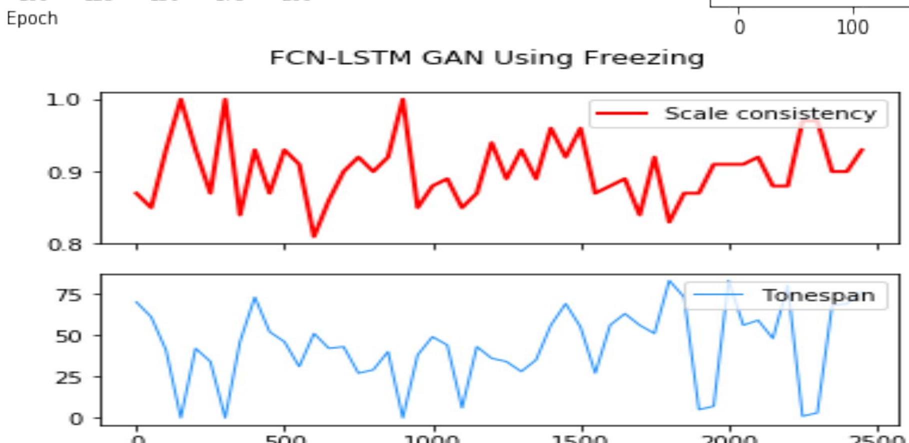
- Subset of samples of classical music taken from **LAKH MIDI** dataset.
- **2001** songs with **182041** notes with frequency range: **8Hz - 12543Hz**
- Feature-wise normalization

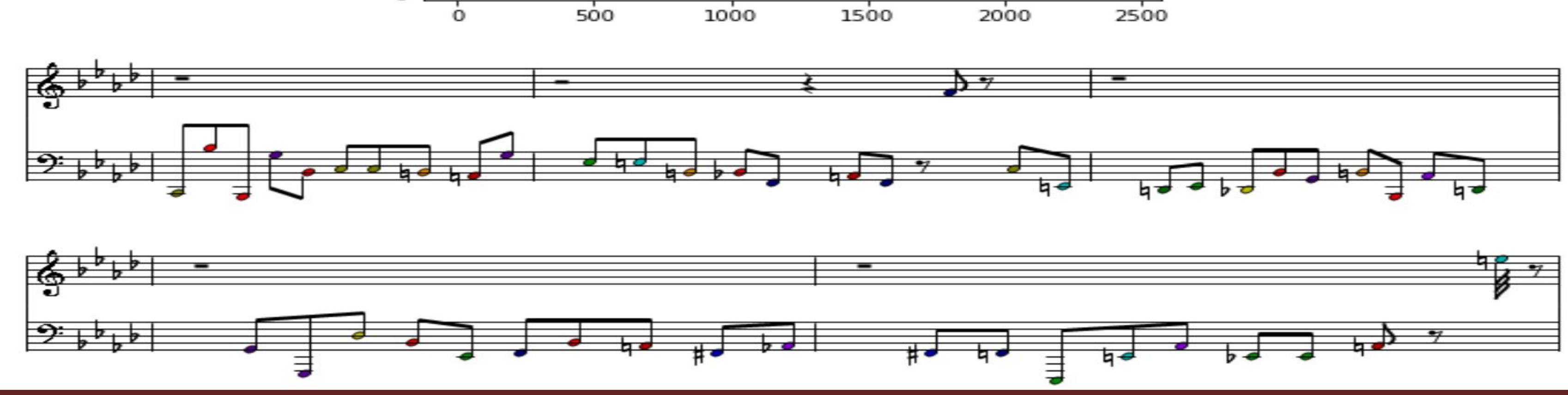


Model Architecture & Methodology

1. ***Baseline:*** A Simple **RNN** with multiple **LSTMs** stacked and connected to an **FCN.** An input sequence to the model produces the most probable next note to be generated.
2. ***FCN-RNN GAN***: The generator consists of a fully-connected layer which up-samples a randomly generated latent feature vector into a sequence of notes. The discriminator has multiple stacked LSTM layers followed by FCN which compute real or fake probability using sigmoid activation.
3. ***C-RNN GAN***: Both the generator and discriminator networks have stacked LSTM layers feeding input to FCNs.







Evaluation and Results

Metrics used for statistical evaluation of generated music:

- **Scale Consistency:** Fraction of tones part of a standard scale.
- **Repetitions:** Number of repeated short subsequences where a score is given based on the amount of recurrence.
- **Tone span:** Measure of the number of half-tone steps between the lowest and the highest tone in a sample.
- **Unique Tones:** Number of unique tones generated.

	Baseline	FCN-RNN	C-RNN
Scale Consistency	0.90	0.93	0.95
Repetitions	{**2**: 17,**3**: 8 **4**: 5}	{**2**: 4, **3**: 1}	{**2**: 7, **3**: 2, **4**:1}
Tone Span	37	55	62
Unique Tones	29	37	39

Conclusion and Extensions

1. We observed that adversarial training helps generate unique music with lots of embedded variations introduced by the combination of good variety of notes played across different intensity spans.
2. The scale consistency shows that our generated music shares a lot of notes that are part of the standard scale.
3. However, the generated music cannot yet compare to real music and the reasons for this remain to be explored.
4. We also intend to try "feature matching", an approach to encourage greater variance in generator, and avoid overfitting to the current discriminator.

Reference

1. O. Mogren, "C-rnn-gan: Continuous recurrent neural networks with adversarial training," 2014.
2. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley,S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," inAdvances in neural information processing systems, pp. 2672–2680, 2014.