Music Generation through Generative Adversarial Networks (GANs)

Hemangani Nagarajan
University of Massachusetts, Amherst
hemanganinag@umass.edu

Sudharshan Govindan
University of Massachusetts, Amherst
sgovindan@umass.edu

November 8, 2023

Abstract

This project explores the frontier of artificial music generation, harnessing the power of Generative Adversarial Networks (GANs) to produce original compositions. By leveraging the "lpd-17-cleansed" dataset, a subset of the comprehensive Lakh MIDI collection, we propose a novel GAN architecture tailored to understand and utilize the intricate patterns of music. Our objective is to develop a generator and discriminator that can process the sequential essence of musical tracks, aiming to create harmonically rich and stylistically coherent pieces.

1. Introduction

The intersection of artificial intelligence and music has opened a new avenue for exploring automated composition, challenging the traditional paradigms of creativity. Generative Adversarial Networks (GANs), known for their prowess in generating visual content, are now being tuned to the frequencies of music generation. Our project seeks to push the boundaries of this technology by focusing on the generation of multi-instrumental music tracks using the "lpd-17-cleansed" dataset, derived from the extensive Lakh MIDI dataset. This dataset's rich collection of MIDI files provides a fertile ground for training our GANs to understand and replicate the complex structure of music. By designing a specialized GAN architecture, we aim to create a system that not only mimics the sequential flow of music but also captures the emotive and dynamic aspects that make a composition compelling. The ultimate goal is to produce a model that can generate new, original music compositions, offering a tool that could serve as a source of inspiration and collaboration for musicians and composers, and potentially automating certain aspects of music production in the entertainment industry.

2. Problem Statement

The creation of music has always been an inherently human endeavor, intricately tied to emotional intelligence and creative expression. However, the advent of artificial intelligence presents a paradigm shift, offering tools to augment or even automate the compositional process. The challenge lies in developing an AI system that can not only understand the theoretical constructs of music but also encapsulate the emotional resonance that characterizes impactful compositions. Traditional algorithms have struggled to capture the complexity and creativity of music, often producing results that lack coherence and aesthetic appeal. Generative Adversarial Networks (GANs) present a potential solution, yet their application in music generation remains largely unexplored, particularly for multi-instrumental compositions.

The problem we address is twofold: firstly, to tailor a GAN architecture that can learn from the "lpd-17-cleansed" dataset, a derivative of the vast Lakh MIDI dataset, and secondly, to enable the generation of music that is not only structurally sound but also creatively inspiring. This involves overcoming the intricacies of sequential data processing and the subjective nature of musical quality, aiming to produce an AI capable of composing music that resonates with human listeners and supports artists in the creative process.

3. Technical Approach

In this section, we present our technical strategy for pioneering the use of Generative Adversarial Networks (GANs) in the realm of automated music composition. We detail the architecture of our model, the training process, and the dataset that informs our AI, setting the stage for a deeper dive into the mechanics of our approach and the future directions we intend to explore.

In the discriminator architecture, convolutional layers are employed, featuring parameterized rectified linear units (PReLU) to effectively process the input data. Following the convolutional processing, the output is channeled through a linear layer, which serves to produce the final discrimination score. To ensure stability during the learning process, batch normalization (BatchNorm2d) is applied to the convolutional neural network (CNN) layers before the activation function (ReLU) is executed.

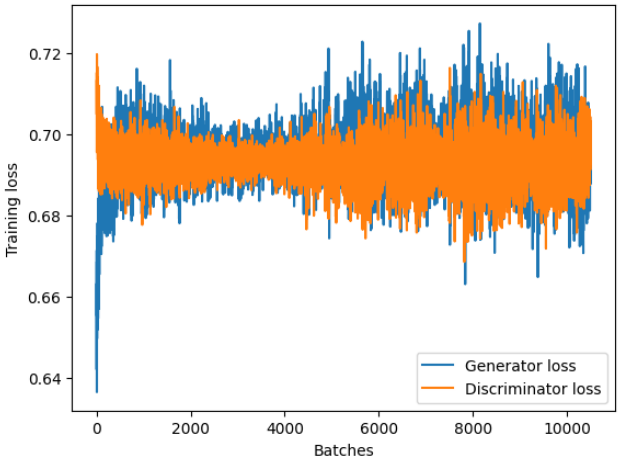

Figure 1. Generator and Discriminator Losses

The generator architecture is designed with transposed convolutional 2D layers (Transpose CNN 2D) that upscale the input to generate sequential data. Activation functions such as PReLU and sigmoid are integrated to introduce non-linearity and to map the outputs to the appropriate range, respectively. Similar to the discriminator, batch normalization is utilized within the generator's layers to promote smoother training dynamics and aid in network convergence.

For training and loss evaluation, the binary cross-entropy loss function is adopted to quantify the difference between the generated music and the actual data. The model undergoes training over the course of 100 epochs, utilizing 250 pianorolls from the "lpd-17-cleansed" dataset. Each pianoroll includes 17 channels, each representing a different instrumental track. These individual predicted channels are then synthesized to form a single, cohesive music track.

Looking ahead to future work, we plan to explore more sophisticated GAN architectures, such as Wasserstein GANs with gradient penalty (WGAN-GP). This approach is anticipated to further stabilize the training process and improve the overall quality of the music generated by our model.

4. Prelimniary Results

With limited resources and above mentioned training settings, we have found the above results. Figure 1. shows the loss curves for the generator and discriminator during training. The discriminator loss starts relatively high, which suggests that initially, it might be having difficulty distinguishing real data from fake data generated by the generator. As training progresses, its loss decreases quickly, which indicates it is learning to differentiate real data from the fake data more effectively. After this drop, the discriminator loss stabilizes but has fluctuations. These fluctuations suggests

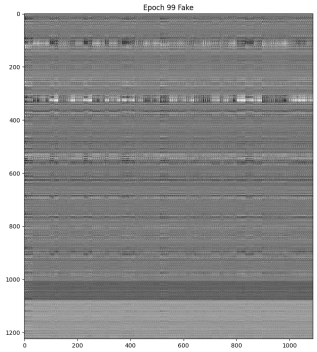

Figure 2. Generated Piano Rolls without threshold

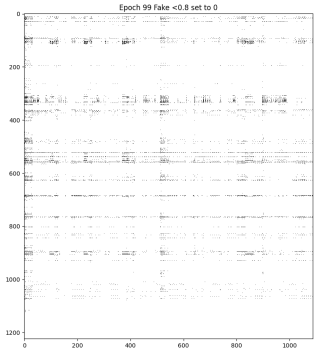

Figure 3. Generated Piano Rolls with threshold

that the discriminator is still learning but is facing a tougher challenge as the generator improves. The generator loss starts low and as training progresses, the loss for the generator increases, which indicates the discriminator is getting better at its job, forcing the generator loss to go up as it fails more often to fool the discriminator. It fluctuates as the generator constantly tries to adapt to the discriminator's feedback.

Figure 2. shows a continuous range of generated values from the generator represented in reversed grayscale, where pure black represents a value of 0 (note on) and pure white represents a value of 1 (note off), with various shades of gray indicating values in between. Figure 3. shows an image that is a binary thresholded version, where all activations below 0.8 have been set to 1. Only the values at or above 0.8 are left as black, representing a note being played.

5. References

1. Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, and Yi-Hsuan Yang, "**MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment**," in *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*

(AAAI), 2018.

2. Colin Raffel, "**Learning-Based Methods for Comparing Sequences, with Applications to Audio-to-MIDI Alignment and Matching**," *PhD Thesis*, 2016.

3. Yang, L.-C., Chou, S.-Y., & Yang, Y.-H. (2017). **MidiNet: A Convolutional Generative Adversarial Network for Symbolic-domain Music Generation**. arXiv preprint arXiv:1703.10847.

4. Arjovsky, Martin, Soumith Chintala, and Léon Bottou. "**Improved Training of Wasserstein GANs.**" In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017), 5767-5777.

5. S. Walter, G. Mougeot, Y. Sun, L. Jiang, K. -M. Chao and H. Cai, "**MidiPGAN: A Progressive GAN Approach to MIDI Generation**," 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Dalian, China, 2021, pp. 1166-1171, doi: 10.1109/CSCWD49262.2021.9437618.

References