EECS 545 Project Progress Report:
The Little Drummer: drum beats generated from other instruments

Congni Shi, Li Ye, Muye Jia, Jerry Cheng
{congni, yeliqd ,muyej, chengjry}@umich.edu

January 4, 2022

1 Problem Statement

In this project, the group would like to investigate some of the existing Deep Neural Networks (DNN) for music composition/generation. More particularly, the group would like to produce a drum beats generator model. The group is also expecting to make some improvements on the existing models, so that the training of the model would be attainable with the limited hardware resources for this course project.

2 Significance

Comparing to the copious deep learning researches in vision, acoustic generative networks are less studied. However, the demand of music composition and professional music composers are expected to go up in the next few years. Music generating neural network would be one of the most lucrative fields in the near future.

3 Related Works/Novelty

Although the field of music generation is not as well developed as the field of image generation, there are still some rather impressive works that can generate drum beats or music in general. The model that the group is most interested in is the MuseGAN [1] model. It is a multi-track sequential generative adversarial network. It is a state-of-the-art temporal model for symbolic music generation, and is one of the only known models for generating polyphonic music.

This paper has suggested that the convolutional neural networks (CNN) could offer optimal performance in recognizing local, translation-invariant patterns. Such observation was also confirmed in an internet article [3] that discussed generation of drum beats for non-drum musics. It was specifically addressed that CNN could usually out-perform LSTM while longer input length is offered during training. Beyond that, training the CNN models would be more attainable for this course project comparing to the computation equipment needed for training LSTM models.

Another important implementation details about the MuseGAN is the way this model uses to segment the input music into fixed size matrices for CNNs, namely the multi-track piano-roll representation. This representation has made possible for this course project to fine tune some GAN or CNN with different architecture and compare their performance against the existing models.

4 Proposed Method/approach

Inspired by the related reading materials, the group has decided that the generative networks of the project's model should also be a temporal model, which it would have information related to both the entire track and the local segments of the music for pattern recognition. A course architecture of the model is demonstrated in Figure 1.

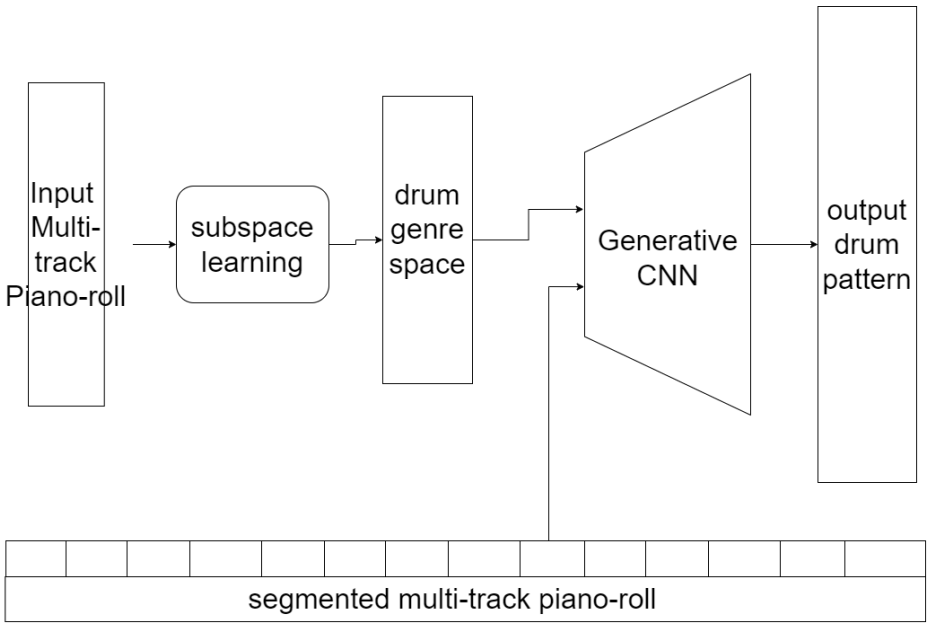

Figure 1: Course Architecture of the model

5 Evaluations

References

[1] H.-W. Dong, W.-Y. Hsiao, L.-C. Yang, and Y.-H. Yang. Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment, 2017.

[2] H.-W. Dong and Y.-H. Yang. Convolutional generative adversarial networks with binary neurons for polyphonic music generation, 2018.

[3] T. Huang. Neural networks generated lamb of god drum tracks, Mar 2019.

[4] S. Nikolov. Neuralbeats: Generative techno with recurrent neural networks, Apr 2016.