EECS 545 Project Progress Report:
Tempo-Lite: Drum Beats Generated from Other Instruments

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

February 2, 2022

1 Problem Statement

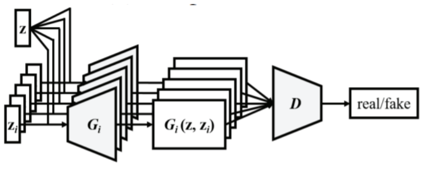
In this project, the group would like to investigate some of the existing Deep Neural Networks (DNN) for music composition/generation. More particularly, the group would like to produce a drum beats generator model, which that the model would take the complete, previously recorded tracks from the other musical instruments, and try to generate a drum track sequencially based on these provided other tracks. The group is expecting to make some improvements on an existing models, MuseGAN [1], so that it would be more suitable for the project's objective and more attainable in terms of training with limited computation resources.

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. Unlike professional composers, amateur composers may not be able to write out scores for all musical instruments — they are likely to be relatively more familiar to one or few musical instruments than others. In that case, our model can be a beneficial resource towards the amateur composers with their compositions. Although the focus of this course project is only on drum track generation, with properly labeled dataset and training, it should be able to work on other musical instruments as well. Moreover, since the aim of the group is to provide a computationally-light training model, it can be easily re-trained towards different musical instruments, even novel ones such as theremin or Kazoo with contemporary styles. Therefore, the group would argue that 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 the state-of-the-art temporal model at its time for symbolic music generation, and is one of the only known models for generating polyphonic music.



(c) Hybrid model

Figure 1: MuseGAN: Hybrid model

In this course project, we are expecting to utilize the Hybrid model in Figure 1 from MuseGAN. In this architecture, each segment GAN has one local Gaussian random vector input z_i and a global Gaussian random vector input z. z_i's are most likely different for each segment generator, and z would remain the same for each segment generators, serving as an artificial inductive bias for all generator to produce more style-aligned music for a single song. Then the GAN is expected to generate a segment of track using these two random vectors.

However, the goal of the MuseGAN model diverges from this course project, since it focuses on generating all tracks from scratch using a relatively large-scale Generative Adversarial Network (GAN). In this course project, the main focus was to generate drum tracks from other instruments by using a lighter model. To adapt this objective, the team would first like to perform a classification/encoding task on the existing tracks from other instruments, and serve this encoded vector as the inductive bias to further facilitate each segment GAN. Moreover, the team is also expecting to explore different loss implementation and discriminator architectures to further simplify the model while remaining the same performance.

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 coarse architecture of the model is demonstrated in Figure 2.

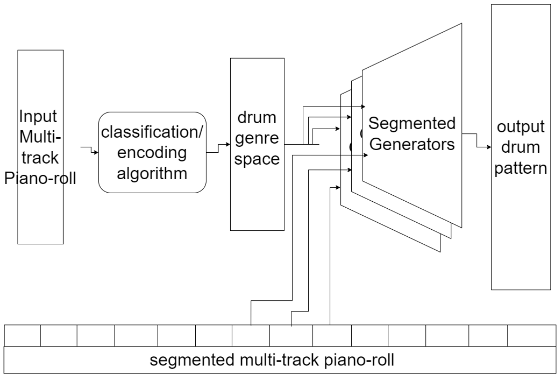

Input Multi-track Piano-roll
classification/ encoding algorithm
drum genre space
Segmented Generators
output drum pattern
segmented multi-track piano-roll

Figure 2: Course Architecture of the model

In the above diagram, first the completed tracks of other instruments are collected and feed into a classification/encoding algorithm. The algorithm would produce a drum genre space that will be utilized as the global inductive bias vector z for the segmented generators. The group has proposed three methods for the classification/encoding algorithm to implement this part of the network.

4.1 CNN Encoders

For the first option, the group would like to implement a conventional CNN encoder for the classification/encoding task. It is expected to be the most computationally intensive implementation but the most mature encoding structure available as of today. The group would probably inspect both the CNN encoder with freezed implementation during generation task or continuously fine-tuning the CNN encoder while training for the generation task.

4.2 Subspace Learning

For the second option, the group would like to test the classification task with the subspaces learning technique. Subspace learning should be much faster in inference comparing to the CNN encoder.

Because of the large scale of music data representation, doing training data Dimension Reduction (DR) and Principal Component Analysis(PCA) are essential to a lighter-load training process. Some related works use singular value decomposition(SVD) to find principal components to approximate the structure of the training data [**?**]. For the music genre classification task of this project, the team currently plans to first group the training data into different music genres, and then use SVD to obtain an orthonormal basis for each genre space. After having learned the genre subspaces, classification can be done by comparing the projections of one particular test data onto the orthogonal complement of the genre subspaces.

4.3 Perceiver

In the last option, one of the student in the group would like to implement an attention-based model, namely the Perceiver [**?**]. Since the output space of the perceiver model would be a logits vector, it can also serve as a traditional classifier/encoder structure. This part of the project would potentially be adopted from one of the student's EECS-542, Advanced Computer Vision course project.

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.