



Generate
Drum sample
with

Presentation by
Hyeonsu Nam



Groove Midi Dataset



Table of Content



- Magenta Package를 활용하여 4마디 드럼 샘플을 생성하는 방법을 소개

MAGENTA PACKAGE 소개

MUSICVAE 논문

GROOVE MIDI DATASET

4마디 드럼 샘플 생성

CONCLUSION



Magenta 소개



- Google 에서 개발한 음악 및 예술 창작을 위한 오픈소스 라이브러리
- 딥러닝 알고리즘을 활용하여 음악 생성, 이미지 생성 등 다양한 창작 분야에서 적용됨

● MusicVAE 논문

A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music

Adam Roberts¹ Jesse Engel¹ Colin Raffel¹ Curtis Hawthorne¹ Douglas Eck¹

Abstract

The Variational Autoencoder (VAE) has proven to be an effective model for producing semantically meaningful latent representations for natural data. However, it has thus far seen limited application to sequential data, and, as we demonstrate, existing recurrent VAE models have difficulty modeling sequences with long-term structure. To address this issue, we propose the use of a *hierarchical* decoder, which first outputs embeddings for subsequences of the input and then uses these embeddings to generate each subsequence independently. This structure encourages the model to utilize its latent code, thereby avoiding the “posterior collapse” problem, which remains an issue for recurrent VAEs. We apply this architecture to modeling sequences of musical notes and find that it exhibits dramatically better sampling, interpolation, and reconstruction performance than a “flat” baseline model. An implementation of our “MusicVAE” is available online.²

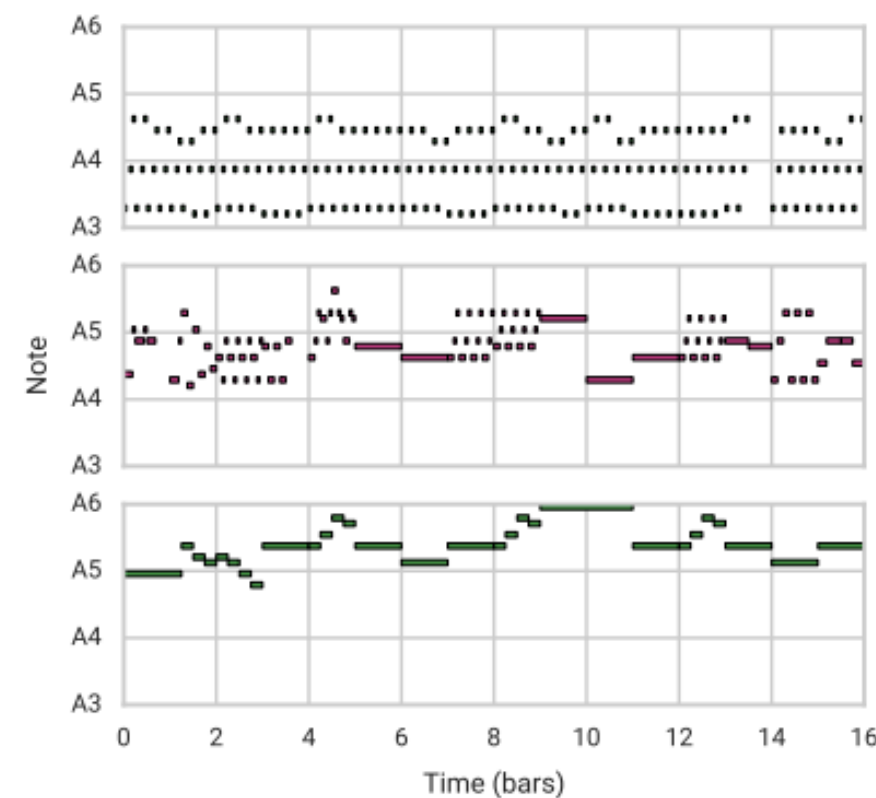


Figure 1. Demonstration of latent-space averaging using MusicVAE. The latent codes for the top and bottom sequences are averaged and decoded by our model to produce the middle se-

- 기존의 RNN 구조는 직전 데이터의 영향을 많이 받고 오래 전 데이터의 영향은 갈수록 줄어드는 문제점이 있다.
 - => 긴 Sequence에 대한 modeling 문제 극복을 원함
- 이를 위해 hierarchical recurrent decoder를 도입
- 기본적으로 본 모델을 AutoEncoder
 - But! 새로운 샘플을 생성하고 Latent space에서 공간 및 속성 벡터 연산을 수행할 수 있는 능력도 원했기에 VAE 프레임워크를 사용

MusicVAE의 구조

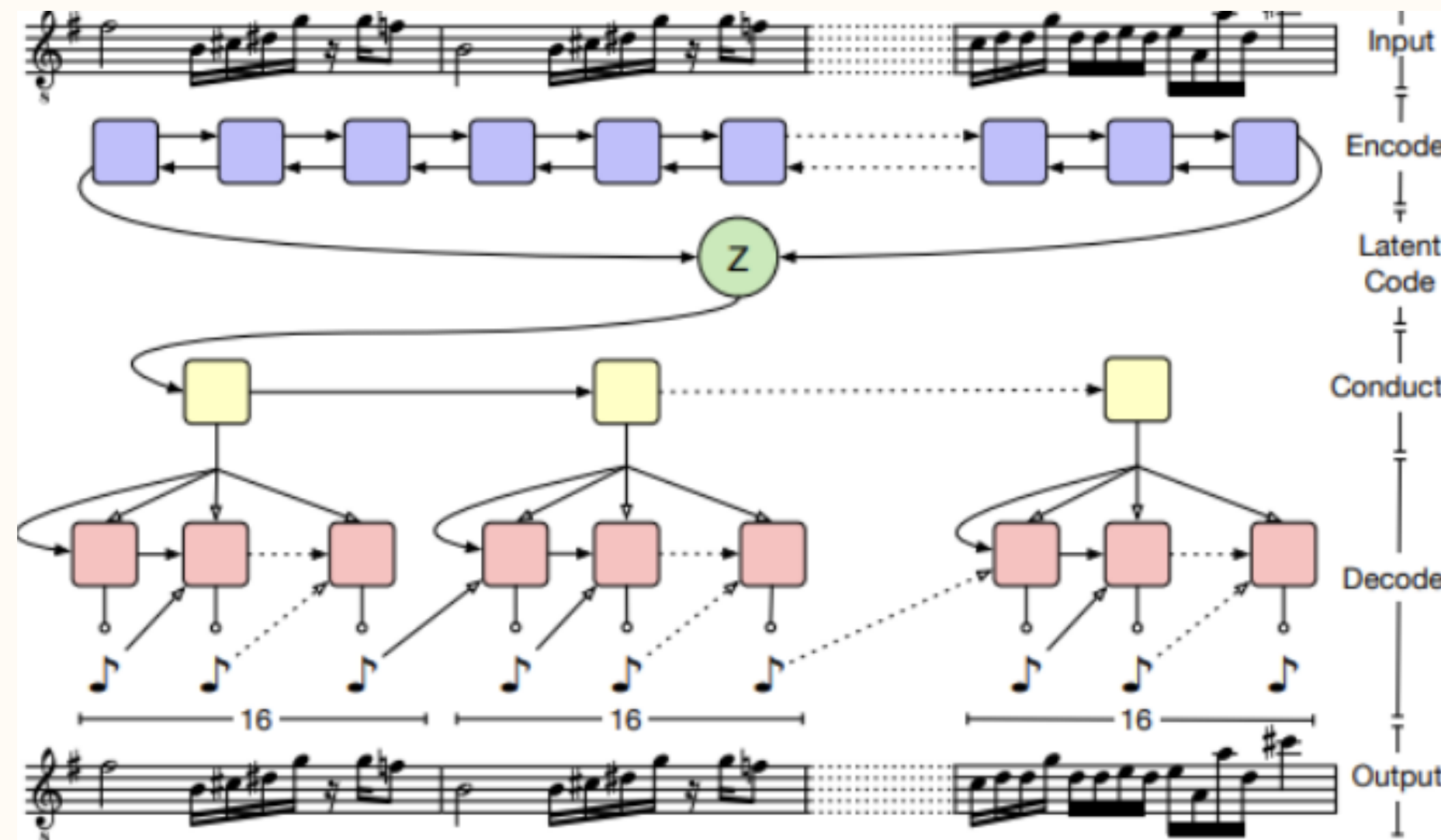


Figure 2. Schematic of our hierarchical recurrent Variational Autoencoder model, MusicVAE.

• 모델의 구조

- Encoder: Bi-LSTM(I.512, H.2048)
- Conductor: LSTM(I.512, H.1024)
- Decoder: LSTM(I.1024(512+512), H.1024)
- Optimizer: Adam

● Groove Midi Dataset

Field	Description
drummer	An anonymous string ID for the drummer of the performance.
session	A string ID for the recording session (unique per drummer).
id	A unique string ID for the performance.
style	A string style for the performance formatted as "<primary>/<secondary>". The primary style comes from the Genre List below.
bpm	An integer tempo in beats per minute for the performance.
beat_type	Either "beat" or "fill"
time_signature	The time signature for the performance formatted as "<numerator>-<denominator>".
midi_filename	Relative path to the MIDI file.
audio_filename	Relative path to the WAV file (if present).
duration	The float duration in seconds (of the MIDI).
split	The predefined split the performance is a part of. One of "train", "validation", or "test".

Genre List: afrobeat, afrocuban, blues, country, dance, funk, gospel, highlife, hiphop, jazz, latin, middleeastern, neworleans, pop, punk, reggae, rock, soul

● Groove MidiDataset

- 인간이 연주한 템포에 맞춘 표현력 있는 드럼 연주의 **13.6시간**에 해당하는 MIDI 와 (합성된) 오디오로 구성되어 있다.
- 이 데이터셋에는**1,150개의 MIDI 파일**과 **22,000개 이상의 드럼 연주 마디**가 포함되어 있다.

● 4마디 샘플 생성

..
0_groovae_4-4_sample.mid
1_groovae_4-4_sample.mid
2_groovae_4-4_sample.mid
3_groovae_4-4_sample.mid
4_groovae_4-4_sample.mid
5_groovae_4-4_sample.mid
6_groovae_4-4_sample.mid
7_groovae_4-4_sample.mid

- [Github 주소]

https://github.com/RedRabbit0314/MusicVAE_PZA/tree/main/generated_midi

● 4마디 샘플 생성



..



0_groovae_4-4_sample.mid



1_groovae_4-4_sample.mid



2_groovae_4-4_sample.mid



3_groovae_4-4_sample.mid



4_groovae_4-4_sample.mid



5_groovae_4-4_sample.mid



6_groovae_4-4_sample.mid



7_groovae_4-4_sample.mid

- [Github 주소]

https://github.com/RedRabbit0314/MusicVAE_PZA/tree/main/generated_midi

● Conclusion

- magenta 패키지를 활용하는데 있어서 환경 설정 관련 dependency 문제를 해결하기 위해 많은 시간이 소요되었었다.
- tensorflow 환경 문제로 인해서 학습이후에 생성을 만드는데 에러가 발생하여 결과물을 만들어내는 데 많은 어려움이 있었다. 모델을 import하는 과정에서 생긴 pylance error 문제, PATH 경로 설정 문제, Kernel 설정 문제 등 많은 오류가 복합적으로 생기다보니 결과물을 도출하는 과정에 있어서 많은 시간이 소요되었다.
- GPU를 사용하지 않아 학습을 하는 과정에서 많은 epoch 을 주지 못했다. 나중에 GCP를 이용하거나 Colab을 이용하여 모델을 구현할 예정
- 다른 dataset을 이용함과 동시에 다른 모델을 적용하여 새로운 모델을 만들어볼 예정



Thank you
presentation by Hyeonsu Nam