Arranging an Audio Track to other Genres by using CycleGAN-based Deep Learning Model *

Alex DongHyeon Seo
dseo22@wisc.edu

Hyecheol (Jerry) Jang
hyecheol.jang@wisc.edu

Stella Kim
ykim736@wisc.edu

Abstract

Changing the genre of a song is one of the methods used when compositing music. To the best of our knowledge, musicians usually add their new ideas to the song, while trying to keep most of the special characteristics of the original song when arranging music. Similar to other artistic tasks that require human creativity, converting the genre of a song takes a significant amount of time and effort. In this project, we propose a method to translate a music genre by using machine-learning, which can generate a new song with comparably less amount of time than humans. Specifically, we utilized cycleGAN based model to translate a soundtrack to another soundtrack.

Due to the complexity and difficulties of dealing with audio data, our model is able to handle files written with MIDI (Musical Instrument Digital Interface) specification only with specific characteristics. In the near future, we expect to expand our project to use regular audio files rather than MIDI to do our tasks to generalize our model. By doing so, we hope our model to be used for the general public without further modifications.

1. Introduction

In this project, our goal is to change the genre of a music track, given a set of user inputs containing song and the desired genre. To accomplish the goal, the main model that we are going to consider is CycleGAN [1]-based model. Though Isola et al. proposed this architecture for unpaired Image-to-Image translation, we are hoping that this model with a proper modification of the structure would accept relative information from audio tracks.

*Project proposal for Spring 2020, University of Wisconsin-Madison STAT453 Introduction to Deep Learning and Generative Models course (Instructor: Sebastian Raschka);
All authors contributed equally

2. Literature Review

2.1. Audio Files

2.2. Model Architecutures

2.3. Baseline Code

3. Proposed Method

Describe the method(s) you are proposing, developing, or using. I.e., details of the algorithms may be included here.

4. Experiments

Describe the experiments you performed. You may want to create separate subsections to further structure this section.

4.1. Dataset

Briefly describe your dataset in a separate subsection.

4.2. Software

Briefly list (and cite) software software you used.

4.3. Hardware

If relevant, list hardware resources you used.

5. Results and Discussion

Describe the results you obtained from the experiments and interpret them. Optionally, you could split "Results and Discussion" into two separate sections.

6. Conclusions

Describe your conclusions here. If there are any future directions, you can describe them here, or you can create a new section for future directions.

7. Acknowledgements

Our project is part of Spring 2020 semester's Statistics 453 [1] (Introduction to Deep Learning and Generative Models) course of the University of Wisconsin-Madison. We specially appreciate with Prof. Raschka for all of his effort toward lectures and support toward our project.

8. Contributions

Describe the contributions of each team member who worked on this project.

References

[1] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

[1] Course Website: `http://pages.stat.wisc.edu/~sraschka/teaching/stat453-ss2020/`
GitHub Repository: `https://github.com/rasbt/stat453-deep-learning-ss20`