Revised Selected Papers

Accademia Musicale Studio Musica
Michele Della Ventura, *editor*

2019

Proceedings of the International Conference on New Music Concepts and Inspired Education

Vol. 6

ACCADEMIA MUSICALE STUDIO MUSICA

Accademia Musicale Studio Musica

International Conference on New Music Concepts and Inspired Education

Proceeding Book
Vol. 6

Accademia Musicale Studio Musica
Michele Della Ventura
Editor

Printed in Italy
First edition: April 2019

www.studiomusicatreviso.it
Accademia Musicale Studio Musica – Treviso (Italy)
ISBN: 978-88-944350-0-9

Preface

This volume of proceedings from the conference provides an opportunity for readers to engage with a selection of refereed papers that were presented during the International Conference on New Music Concepts and Inspired Education. The reader will sample here reports of research on topics ranging from mathematical models in music to pattern recognition in music; symbolic music processing; music synthesis and transformation; learning and conceptual change; teaching strategies; e-learning and innovative learning. This book is meant to be a *textbook* that is suitable for courses at the advanced undergraduate and beginning master level. By mixing theory and practice, the book provides both profound technological knowledge as well as a comprehensive treatment of music processing applications.
The goals of the Conference are to foster international research collaborations in the fields of Music Studies and Education as well as to provide a forum to present current research results in the forms of technical sessions, round table discussions during the conference period in a relax and enjoyable atmosphere.
36 papers from 16 countries were received. All the submissions were reviewed on the basis of their significance, novelty, technical quality, and practical impact. After careful reviews by at least three experts in the relevant areas for each paper, 12 papers from 10 countries were accepted for presentation or poster display at the conference.

I want to take this opportunity to thank all participants who have worked hard to make this conference a success. Thanks are also due to the staff of "Studio Musica" for their help with producing the proceedings. I am also grateful to all members of Organizing Committee, Local Arrangement Committee and Program Committee as well as all participants who have worked hard to make this conference a success.
Finally I want to appreciate all authors for their excellent papers to this conference.

April 2019 Michele Della Ventura

Contents

MGTGAN:Cycle-Consistent Adversarial Networks for Symbolic Multi-track Music Genre Transfer

YenLun Peng, Haitao Zheng

Tsinghua university. Open Faculty for Innovation, Education, Science, Technology and Art.
Allenpeng0209@gmail.com

Abstract. Transferring music from one genre to another has a few differences from transferring image from one domain to other domain. Color digital images are made of pixels, and pixels are made of combinations of primary colors represented by a series of code. However, Music is consist of multiple sound tracks, in symbolic music representation like MIDI, we tend to use different instruments with their temporal dynamics to represent to each track. In order to discover the difference between these two types of representation for channels, we proposed MGTGAN to analyze how different channel affect to the result of music genre transfer. Following the recent research from MuseGAN and CycleGAN, we inspired by MuseGAN's three type of model, and chose two of them, which are jamming model, composer model. By combining these two models with CycleGAN. We successfully transfer multi-track music from one genre to another. We also investigate applying different loss function like WGAN in CycleGAN and adding two distinct discriminators to get better convergence. Due to the complexity of the multi-track music is hard to train and evaluate, we also propose a Desert Camel MIDI dataset to simplify the experiment. Each song is well defined in multiple genre, which means it is a paired data and have a correct answer. The evaluation can more focus on judging the certain chord for invariant content and the rhythm style for genre transfer. Our result show that jamming model can transfer each track better than composer model. However, subjective evaluation from human gave composer model higher harmonic score. The code and dataset are available at https://github.com/AllenPeng0209/MGTGAN

Keywords. Music Genre Transfer, Deep Learning, Generative Adversarial Networks

1 Introduction

In this paper, we propose two different models to solve the multi-track music genre transfer problem, the basic idea is to use CycleGAN to transfer music genre, first mode is jamming model with CycleGAN, each track transfer its own music score, the advantage of this model is that each track can adjust its own parameter and get better convergence. However, each track is transfer without considering other track, so the result would be not that harmonic. The second one is composer model with CycleGAN, the model take all tracks as single music score, and try to learn the harmonization between in each track.

As we know that GAN is hard to train, and easy to collapse, we also try different loss function like WGAN-GP, LSGAN and using additional discriminator help the model learn better. We made music genre dataset by iReal Pro software, which consist of 129117 music phrases in five different genre, including rock, funk, bassanova, R&B, soul. To evaluate the model, we use both objective and subjective way. In objective way, we train separate music genre classifier to classify transferred music. In subjective way, we invited 73 users to evaluate the music evaluation metric for transferred music and origin music. The result shows our model can successfully transfer multi-track music genre.

2 Related works

The concept of the style transfer stem from Gatys[3], they use pre-trained CNN to merge the style from one image and content image from another image. Since then, not only the image translation task emerged and improved rapidly, but the idea of music style(genre) transfer also been proposed by different researchers. Malik et al.[4] proposed StyleNet to capture musical style through dynamics. StyleNet is able to synthesize and inject dynamics into MIDI. This would be extremely beneficial for musicians. Lu et al.[5] proposed using recurrent neural network and autoregressive models to transfer style in homophonic music. The result shows that combining LSTM with autoregressive model, the transferred music can preserve the music content and also change the rhythm to another genre while keeping the harmonic structure. Brunners et al.[6] introduced MIDI-VAE, a multi-task Variational Autoencoder model with shared latent space that is capable of changing the style of complete compositions. This is the first proposed architecture that transfer the whole music composition, the model not only consider pitch at timestep as factor, but also velocities, note duration and instrumentation. Zhao[1] using CycleGAN to transfer symbolic music genre, by adding additional discriminators and noise, this method can prevent generator collapse or fail to learn music structure from origin domain.

3 Proposed model

Most of architecture follow the origin CycleGAN, we will introduce which part had been changed compared to origin CycleGAN. The goal of CycleGAN is to learn mapping function between two domain X and Y in the absence of paired data. CycleGAN consist of two sets of GANs. The task of one generator G is to transfer $X \rightarrow Y$ such that G(X) is indistinguishable from distribution Y using adversarial loss. However, without constrained, the generator may not keep the content from domain A and randomly generate data like Y domain to fool discriminator. To solve this problem, another generator F would transfer $Y \rightarrow X$ back, and use cycle consistency loss to enforce $F(G(x)) \approx X$(and vice versa). The whole model shown in Fig. 1

Adversarial Loss:

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim P_{data(y)}}[log D_Y(Y)] + E_{y \sim P_{data(x)}}[log(1 - D_Y(G_x(x))]$$

Cycle Consistency loss:

$$L_{cyc}(G, F) = E_{y \sim P_{data(x)}}[\|F(G(x)) - x\|] + E_{y \sim P_{data(x)}}[\|G(F(x)) - y\|]$$

Full Object Function

$$L_{all}(G, F, D_x, D_y) = L_{GAN}(G, D_{y,} X, Y) + L_{GAN}(F, D_x, X, Y) + L_{cyc}(G, F)$$

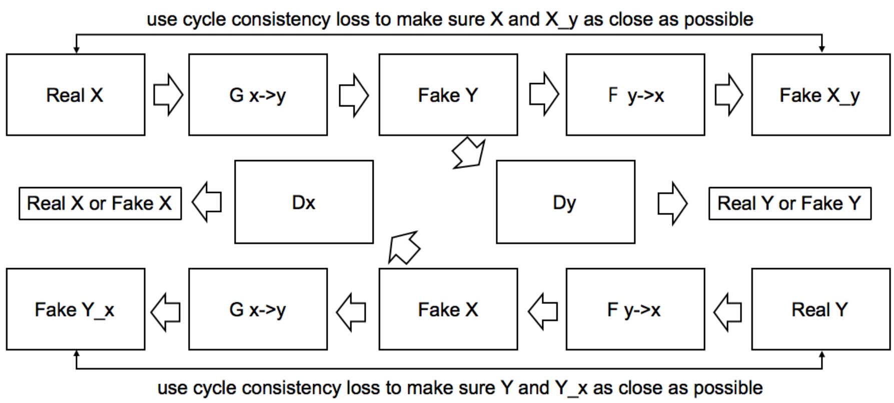

Fig. 1. Architecture of our base model.

- Jamming Model with CycleGAN
  Each generator work independently and transfer its own track from given data X_m, where M denote the number of track. Each discriminator also discriminates its own track. Figure 2 show the general idea of the Jamming model. To generate music of M track, we need M generators and M discriminators.

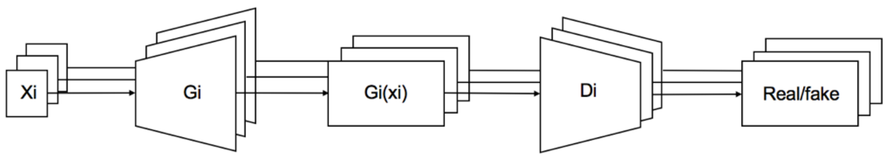

Fig. 2. Jamming model with CycleGAN.

- Composer Model with CycleGAN
  Composer model handle all track simultaneously, each track from given data X_m is stacked and directly feed into generator. Also, discriminator also discriminates the stacked data. Figure 3 show the general idea of the Composer model. To generate music of M track, we only need one generator and one discriminator.

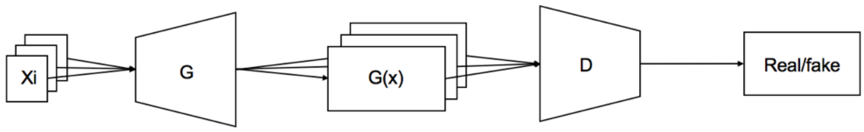

Fig. 3. Composer model with CycleGAN.

4 Result

The evaluation of the model performance can be divided into two parts, the first part is content preservation measurement, and the second part is transferred music genre classification. For the content preservation, we simply do the human evaluation and measure the similarity between two songs. For the music genre classification, we train separated classifier to measure how well the music had been transferred.

Genre Classifier

In order to get objective evaluation, we train another genre classifier to classify whether the transferred data is actually transferred, also compare the effect of jamming model and composer model. we build binary classifier $C_{A,B}$ outputs a probability distribution over domain A and B. The classifier architecture is shown in Table.

The classifier is train on real data, to evaluate the ability of the classifier, we use the 90/10 train/test split. Due to our dataset is created by iReal Pro, it has strong rule for certain music style, so the classifier accuracy on testset is all above to 99.99%.

After training the classifier, we apply the classifier to distinguish the transferred data.

The result show that Jamming model can get 89.56% accuracy which is better than 85.31% from Composer model. Jamming model transferred each track independently, so the detail of the music can be transferred more clearly. However, Composer model have to consider each track, somehow have to sacrifice the detail in each track.

TABLE.1. CLASSIFICATION ACCURACY ON FAKE DATA GENERATED BY JAMMING MODEL AND COMPOSER MODEL.

	Jamming model	Composer model
Fake A vs B	89.56%	85.31%

Human Evaluation

We conduct a listening test of 73 subjects recruited from the Internet via our social circles. 22 of them are deemed ‘pro user,’ according to a simple questionnaire probing their musical background. Each subject has to listen to five music clips in random order. Each clip consists of two four-bar phrases generated by one of the proposed models. The subject rates the clips in terms of whether they 1) chord progression similarity, 2) music genre, 3) harmonic score, 4) the overall rating, in a 5-point Likert scale.

TABLE 4.2 : RESULT OF USER STUDY (CP: CHORD PROGRESSION, MG: MUSIC GENRE, HS: HARMONIC SCORE, OR: OVERALL RATING)

		CP	MG	HS	OR
Jamming	non pro	3.2	3.3	3.0	**3.4**
	pro	3.1	2.9	2.7	3.0
Composer	non pro	**3.3**	3.3	3.1	3.3
	pro	3.2	**3.4**	**3.3**	3.1

From the result shown in Table 3, the Jamming model has a higher score in terms of chord progression. In our case, it means that the content is preserved better than composer model. However, the composer model gets higher score for music genre transfer and harmonic score. It can be interpreted that transferring tracks separately can have better content preservation but will sacrifice harmony at the same time.

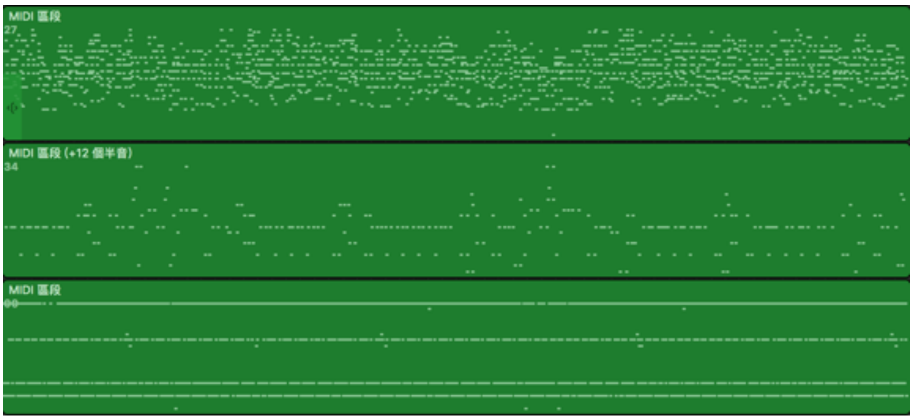

Fig.5. Samples transferred from Rock to Bossanova generated by Jamming model.

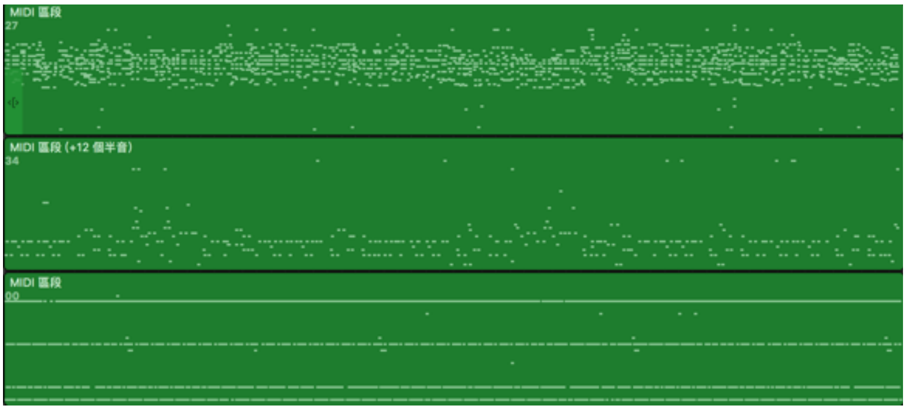

Fig.6. Samples transferred from Rock to Bossanova generated by composer model.

5 Conclusion

In this paper, we extend the Zhao[1] et al. works by exploring the possibility of using CycleGAN to transfer multi-track symbolic music genre. We combine the idea of modeling the multi-track interdependency from MuseGAN[2]. Also, we use iReal Pro to make music transfer dataset, the music genre in this dataset is defined clearly so that our evaluation process can be more objective. Our result show that jamming model can transfer genre in each track better than composer model. However, subjective evaluation from human gave composer model higher harmonic score.
In the future, we plan to incorporate other feature such as instruments information, volume of each note, and velocity. Also, modeling music not in a 4 bars phrase, but in a complete song would be a good way to go.

References

[1] Brunner, Gino, et al. "Symbolic music genre transfer with cyclegan." 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE, 2018.

[2] Dong, Hao-Wen, et al. "MuseGAN: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment." *Proc. AAAI Conf. Artificial Intelligence*. 2018

[3] Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "Image style transfer using convolutional neural networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.

[4] Malik, Iman, and Carl Henrik Ek. "Neural translation of musical style." *arXiv preprint arXiv:1708.03535* (2017).

[5] Lu, Wei-Tsung, and Li Su. "TRANSFERRING THE STYLE OF HOMOPHONIC MUSIC USING RECURRENT NEURAL NETWORKS AND AUTOREGRESSIVE MODELS."
[6] Brunner, Gino, et al. "MIDI-VAE: Modeling dynamics and instrumentation of music with applications to style transfer." *arXiv preprint arXiv:1809.07600* (2018).

This book presents a collection of selected papers that present the current variety of all aspect of music research, development and education, at a high level. The respective chapters address a diverse range of theoretical, empirical and practical aspects underpinning the music science and teaching and learning, as well as their pedagogical implications. The book meets the growing demand of practitioners, researchers, scientists, educators and students for a comprehensive introduction to key topics in these fields. The volume focuses on easy-to-understand examples and a guide to additional literature.

Michele Della Ventura, editor

New Music Concepts and Inspired Education

Revised Selected Papers

ISBN: 978-88-944350-0-9

9 788894 435009

www.studiomusicatreviso.it