LEAD SHEET GENERATION AND ARRANGEMENT VIA A HYBRID GENERATIVE MODEL

Hao-Min Liu*, **Meng-Hsuan Wu***, **Yi-Hsuan Yang**
MAC Lab, CITI, Academia Sinica, Taiwan
paul115236@citi.sinica.edu.tw, b04902068@ntu.edu.tw, yang@citi.sinica.edu.tw

ABSTRACT

Due to the popularity of using lead sheets and MIDIs to store symbolic music data, many existing models for symbolic music generation aim to generate lead sheets or multitrack pianorolls. In this demo paper, we bring up a new task called *lead sheet arrangement* to bridge the gap between this two data formats. Specifically, we present a new hybrid model that exploits harmonic features from unpaired lead sheets and MIDIs to learn to add arrangement to a lead sheet. Our model can generate lead sheets and their arrangements of eight-bar long.

1. INTRODUCTION

Broadly speaking, recent works on automatic music generation can be divided into two groups according to the target form of musical notation. The first group focuses on creating the lead sheet, which comprises a melody line and a sequence of accompanying chord labels. The second group aims to create music of the MIDI format, which indicates all the voicing and accompaniment of different instruments. An interesting and yet seldom-studied task is to generate something in the middle of the aforementioned two forms. We call it *lead sheet arrangement*.

We have proposed this task for the first time in a prior work and presented a fully convolutional model to approach it [4]. This demo paper extends directly from that work and attempts to use a hybrid model instead.

The task can be divided into two subtasks: *lead sheet generation* and *arrangement generation*. The first subtask has to generate a melody line and a chord sequence from scratch, whereas the second subtask assumes a lead sheet has been given and aims to generate its arrangement, as exemplified in Figure 1. While we use a convolutional generative adversarial model (GAN) [1] in our prior work [4] for the first subtask, we adopt a recurrent variational autoencoder (VAE) [2] model here instead, since both melody and chords can be modeled by recurrent neural networks.

*These authors contributed equally to this work.

Figure 1. A lead sheet of *Amazing Grace* with its arrangement generated by the proposed hybrid model.

On the other hand, as musical arrangement is polyphonic, we still use a convolutional GAN for the second subtask to generate multi-instrument accompaniment. Following [4], we extract harmonic features from the lead sheet and use these features to condition the generation for the second subtask. A system diagram of the proposed model is depicted in Figure 2. This new hybrid model leverages the advantages of both types of generative models (i.e., VAE and GAN) to achieve better generation result.

In this work, we use the TheoryTab dataset (TTD) [4] for the lead sheets and the Lakh pianoroll (LPD) dataset [1] for the multitrack pianorolls arrangement. An example generation result is shown in Figure 1. More example results can be found at https://liuhaumin.github.io/LeadsheetArrangement/.

2. PROPOSED MODEL

2.1 Lead Sheet Generation

Aiming at eight-bar lead sheet generation, we use a recurrent VAE network structure illustrated on the left hand side of Figure 2 to learn an eight-bar embedding space. We use a one-hot vector for melody representation and a chroma

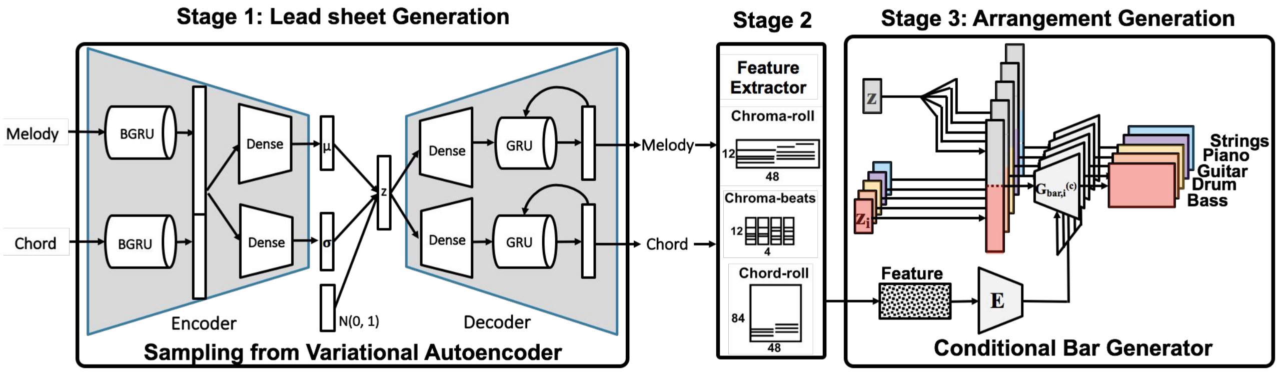

Figure 2. Architecture of the proposed hybrid model for lead sheet generation and arrangement. It uses recurrent variational autoencoder for lead sheet generation, and convolutional generative adversarial network for arrangement generation.

vector for chord representation. In the encoder part, eight-bar melody and chord sequences are fed into bidirectional gated recurrent units (BGRU) to learn the correlation between bars. The outputs of all GRU timesteps are then concatenated together and passed through a few dense (i.e., fully-connected) layers for calculating μ and σ. In the decoder part, a latent vector $\mathbf{z}$ is sampled from a normal distribution characterized by μ and σ and then gone through a few fully-connected layer to separately form the initial states of melody and chord. The outputs are finally generated by a unidirectional GRU with a sigmoid activation layer to form an eight-bar lead sheet.

2.2 Arrangement Generation

Here, we aim to generate the pianorolls of a number of instruments to accompany the melody of a given lead sheet. Therefore, it is a *conditional* generation task. For now, we only generate the arrangement bar-by-bar, but this can be improved in the future work. As depicted on the right hand side of Figure 2, we use a conditional bar generator $G_{\text{bar}}^{(c)}$ to generate five-track pianorolls of one bar long. As the bottom-right corner of Figure 2 shows, conditional generation is achieved by training a convolutional encoder E to embed the harmonic features extracted by the middle feature extractor to the same space as the output of the intermediate hidden layers of $G_{\text{bar}}^{(c)}$, and the discriminator $D^{(c)}$.

In order to project the lead sheets and MIDIs to the same feature space to make the conditional generation possible, we propose and experiment with the following three symbolic-domain harmonic features in [4], as shown in the 'second stage' marked in Figure 2:

- **Chroma pianoroll representation** compresses the pitch range into chroma (twelve pitch classes), leading to a 12×48 matrix per bar, assuming 4/4 time signature and 48 time steps per bar.
- **Chroma beats representation** further reduces the temporal resolution of chroma-roll by taking the average per beat, leading to a 12×4 matrix per bar.
- **Chord pianoroll representation** applies an existing chord recognition model [3] to for chord estimation to recognize major and minor chords for each beat and then uses a pianoroll to represent the chords, yielding a 84×48 matrix per bar.

A subjective test from 25 participants in [4] suggests that the last feature (i.e., chord pianoroll representation) performs much better than the other two in generating harmonic and rhythmic arrangement for existing lead sheets. Therefore, we also adopt the last feature in this demo.

3. CONCLUSION

In this paper, we have presented a novel hybrid conditional generative model to learn from unpaired lead sheets and MIDIs for lead sheet arrangement, a relatively new task in symbolic music generation. We used recurrent variational autoencoder for lead sheet generation, and convolutional generative adversarial network for arrangement generation. We have not systematically compared the performance of the new hybrid model and the previous fully convolutional model for lead sheet generation, leaving it as a subject of future work. We instead provide audio examples at the project website for subjective evaluation.

4. REFERENCES

[1] Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, and Yi-Hsuan Yang. MuseGAN: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment. In *Proc. AAAI*, 2018. [Online] `https://salu133445.github.io/lakh-pianoroll-dataset/`.

[2] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[3] Filip Korzeniowski and Gerhard Widmer. Feature learning for chord recognition the deep chroma extractor. In *Proc. ISMIR*, pages 37–43, 2016.

[4] Hao-Min Liu and Yi-Hsuan Yang. Lead sheet generation and arrangement by conditional generative adversarial network. In *Proc. ICMLA*, 2018. [Online] `https://arxiv.org/abs/1807.11161`.