Lead sheet and Multi-track Piano-roll generation using MuseGAN

Hao-Min Liu, Hao-Wen Dong, Wen-Yi Hsiao, Yi-Hsuan Yang

Music and AI Lab, Research Centre for IT Innovation, Academia Sinica, Taipei, Taiwan

This work is based on our previous AAAI'18 paper.

Introduction

Challenges for music generation

- **Temporal dynamics:** music is an art of time with a hierarchical structure
- **Multi-track:** each track has its own temporal dynamics but collectively they unfold over time in an interdependent way

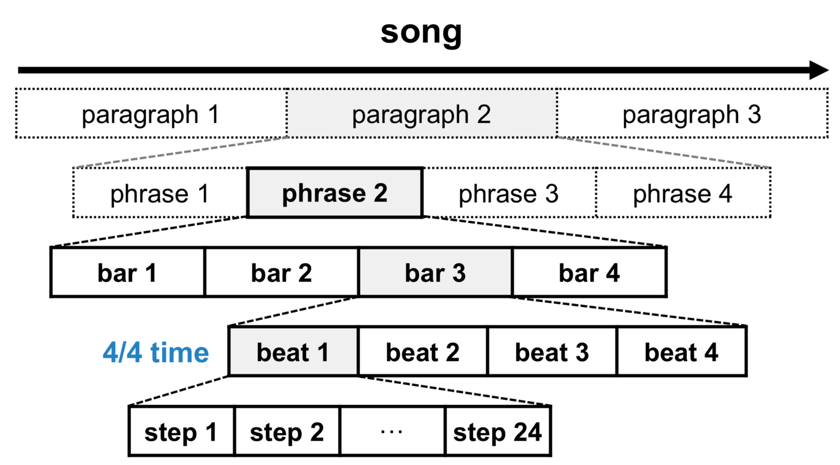

Figure 1. Hierarchical temporal structure of music

MuseGAN (multi-track sequential generative adversarial network) [1] aims to address these challenges altogether.
Key points:

- Use **GAN** (specifically WGAN-GP [2]) to support both "conditional generation" (e.g. following a prime melody) and "generating from scratch", following our previous MidiNet model.
- Use **convolutions** (instead of RNNs) for speed
- Learn from **MIDIs & Lead Sheet XMLs** (using piano-rolls)

Data

Dataset

The *matched* subset of the Lakh MIDI dataset

- Pop/rock, 4/4 time signature, C key
- Five tracks: bass, drums, guitar, piano, strings (others)
- Get 201,064 bars to form 4-bar phrases

Hooktheory XML dataset, after cleansing

- Pop/rock, 4/4 time signature, C key
- Two tracks: melody and chord
- Get 138,792 bars to form 8-bar phrases

Data representation

- **Notes:** 84 pitches (24-108)
- **Phrase:** 4 bars
- **Bar:** 96 time steps
- **Tracks:** 5 instruments

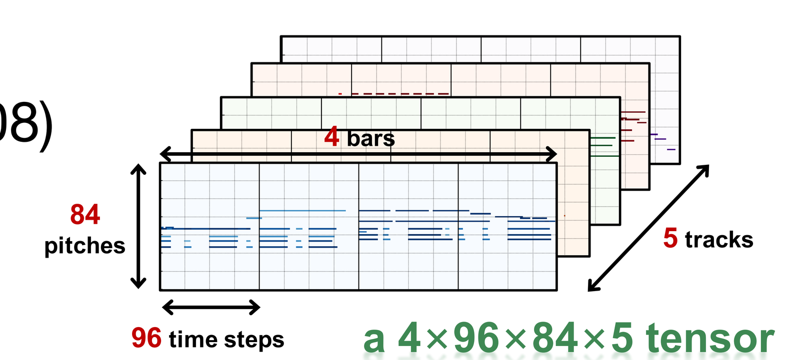

Figure 2. Multi-track piano-roll representation

Proposed Model

Modeling the multi-track interdependency

- Each track is generated independently by its own generator which takes a shared *inter-track* random vector and a private *intra-track* random vector as inputs; the result is evaluated by one single discriminator

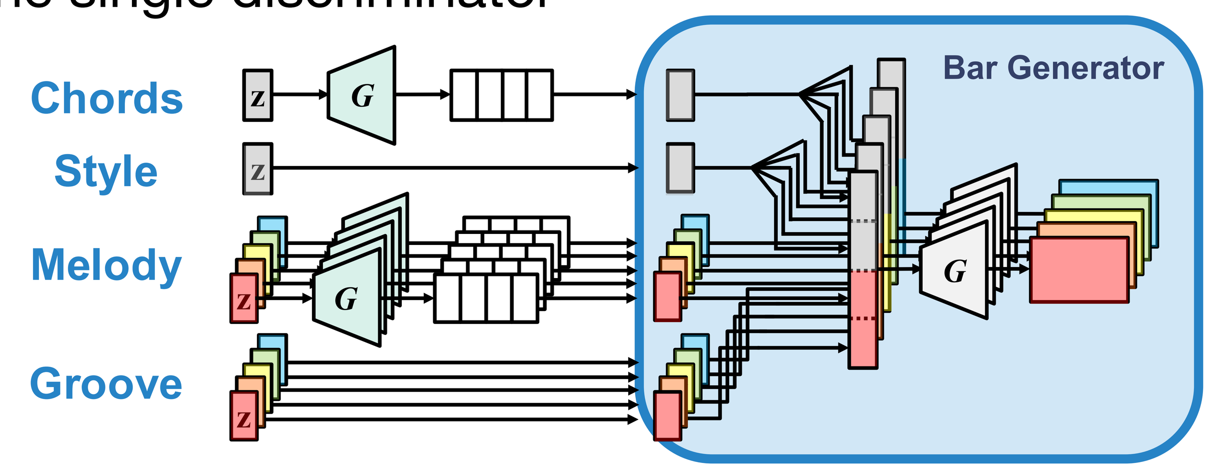

Figure 3. Hybrid model generator, combining the idea of jamming and composing

MuseGAN architecture

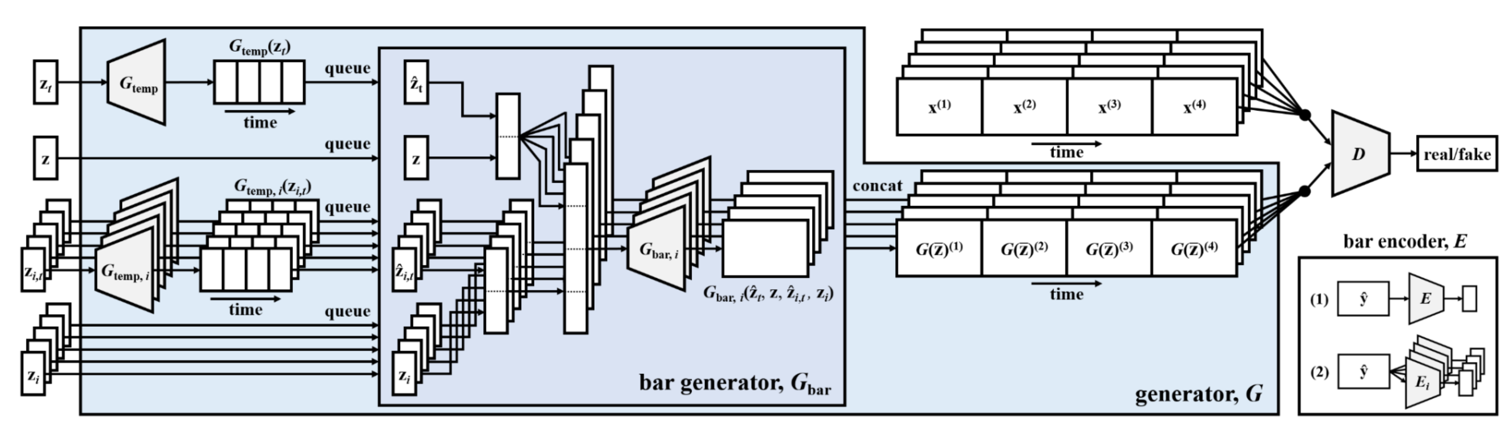

Figure 4. System diagram of the proposed MuseGAN model

Results

Training process

- The training time for each model is less than 24 hours with a Tesla K40m GPU.

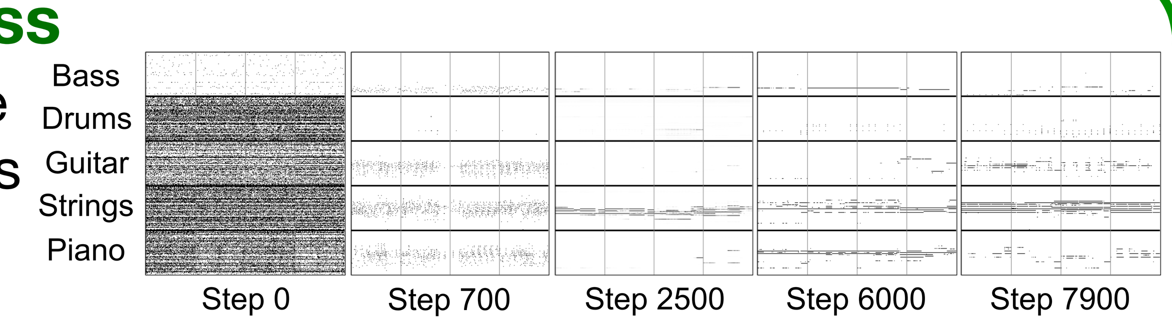

Figure 5: Evolution of a generated phrase

User study

- **H:** harmonious
- **R:** rhythmic
- **MS:** musical structure
- **C:** coherent
- **OR:** overall rating

Table 1: Result of user study

			H	R	MS	C	OR
from scratch	non-pro	jam.	2.83	3.29	2.88	2.84	2.88
		comp.	3.12	**3.36**	2.95	3.13	3.12
		hybrid	**3.15**	3.33	**3.09**	**3.30**	**3.16**
	pro	jam.	2.31	3.05	2.48	2.49	2.42
		comp.	2.66	3.13	2.68	2.63	2.73
		hybrid	**2.92**	**3.25**	**2.81**	**3.00**	**2.93**

Visualization

Lead sheet application

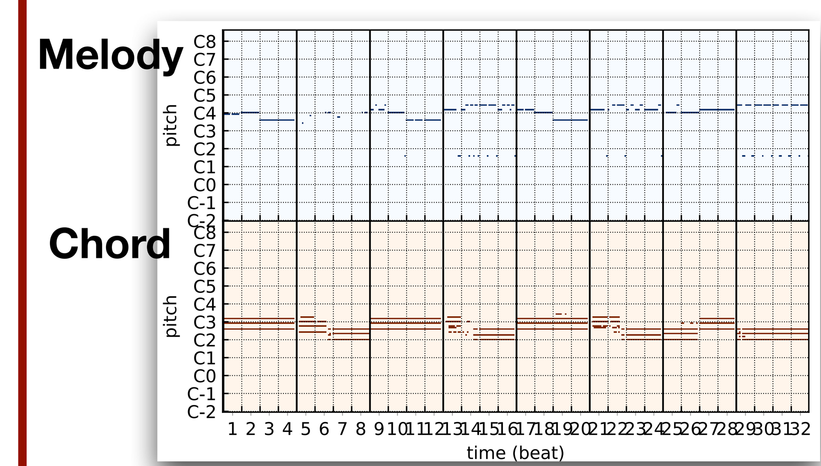

Figure 6. Lead sheet piano roll sample

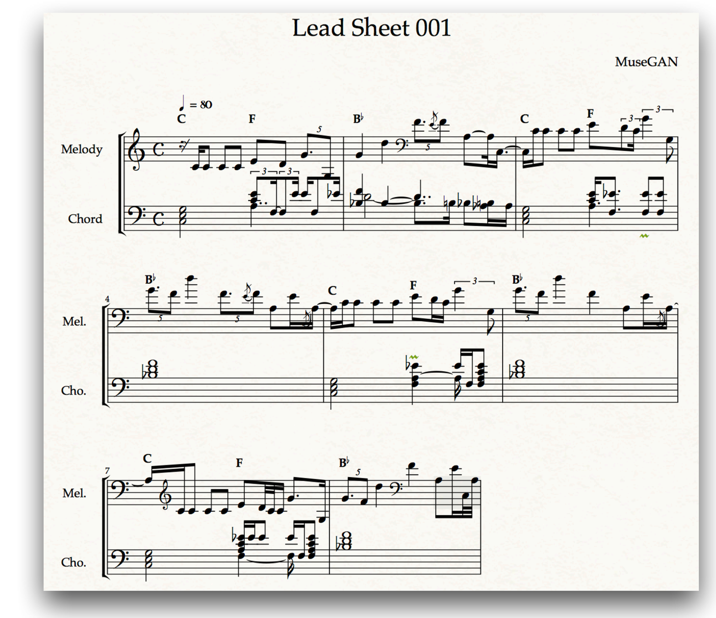

Figure 7. Lead sheet score sample

Interpolation

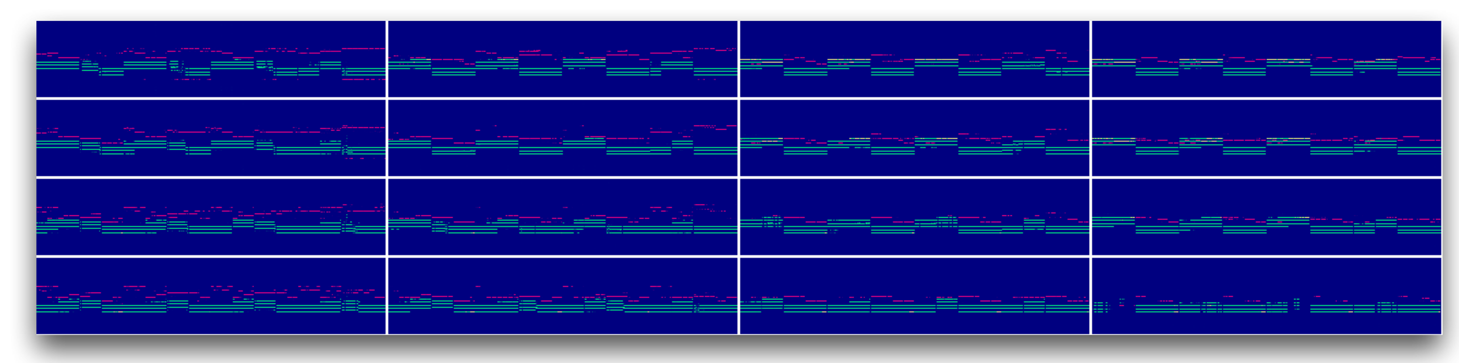

Figure 8. Spherical linear interpolation as a 4x4 matrix

Conclusion

- A new convolutional GAN model is proposed for creating multi-track sequences; we use it to generate pianorolls of pop/rock music by learning from a large set of MIDI and XML.
- Lead sheet generation using MuseGAN with piano-roll form could capture related transitions from chord to chord.

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 Conf. Artificial Intelligence (AAAI)*, 2018.

[2] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved training of Wasserstein GANs. In *NIPS*, 2017.