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)

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

MuseGAN architecture

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

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

Results

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

User study

<table>
<thead>
<tr>
<th>H</th>
<th>R</th>
<th>MS</th>
<th>C</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>harmonious</td>
<td>rhythmic</td>
<td>musical structure</td>
<td>coherent</td>
<td>overall rating</td>
</tr>
</tbody>
</table>

Table 1: Result of user study

• 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