Generating Multitrack Music using Deep Learning

Hao-Wen Dong

University of California San Diego

About me

Hi, I'm Herman.
I do Music x AI research.
I love music and movies!

National Taiwan University
B.S. in Electrical Engineering
2013 – 2017

2017 – 2019
Academia Sinica
Research Assistant
Music and AI Lab, Research Center for IT Innovation

Summer 2019
Yamaha
Research Intern
AI Group, Research and Development Division

University of California San Diego
M.S. in Computer Science
2019 – 2021

Summer 2021
Dolby
Deep Learning Audio Intern
Applied AI Team, Advanced Technology Group

Summer 2022
Sony
Student Intern
Tokyo Laboratory 30, R&D Center

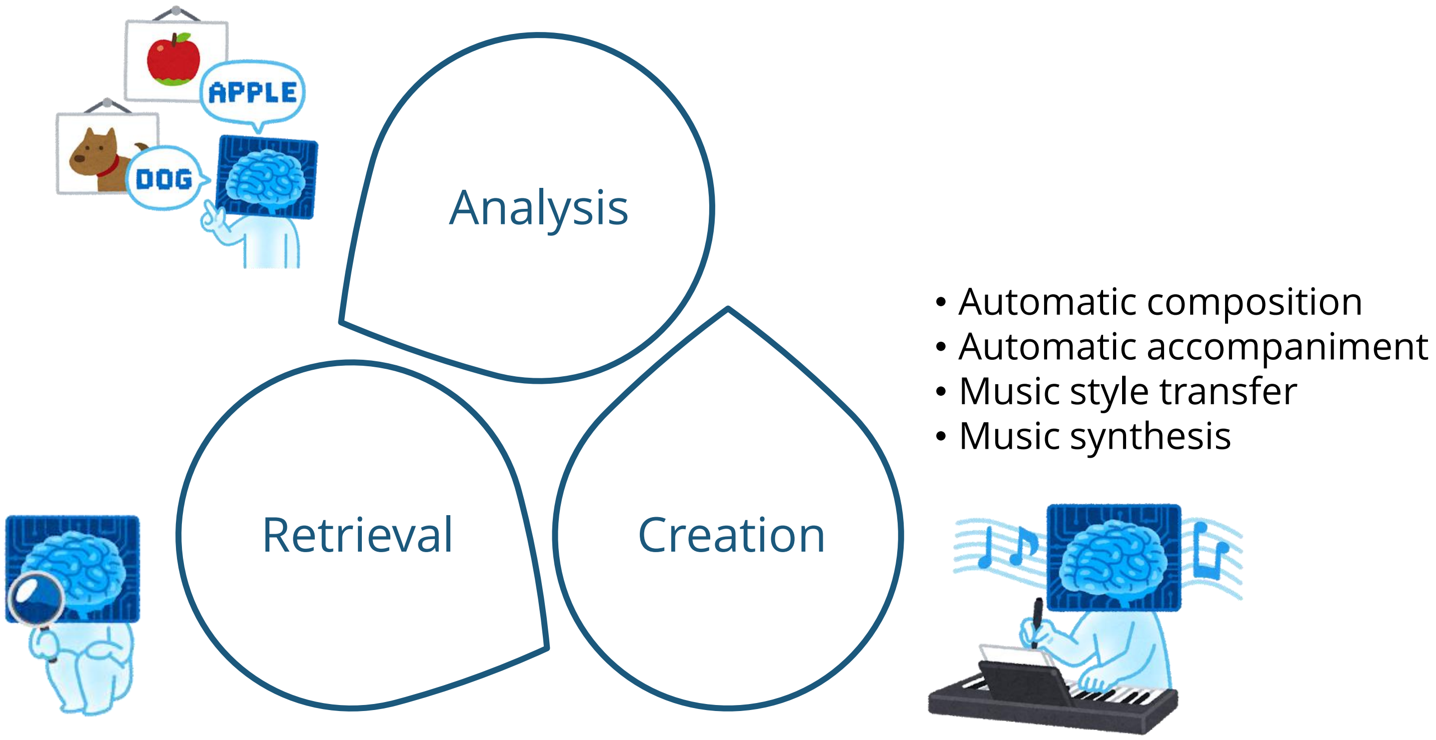
(Fall 2022)
Amazon
Applied Science Intern

University of California San Diego
Ph.D. in Computer Science (in progress)
UC San Diego
2019 – present

Music Information Research

Music information research (MIR)

- Intelligent ways to analyze, retrieve and create music (Yang 2018)



Yang, "Music Information Research," *SNHCC*, TIGP, lecture notes, April 2018.

Outlines

- MuseGAN for multitrack music generation (AAAI 2018)
- Arranger for automatic instrumentation (ISMIR 2021)
- Multitrack Music Transformer for multitrack music generation

MuseGAN

Generating multitrack music using convolutional GANs

(AAAI 2018)

Wen-Yi Hsiao

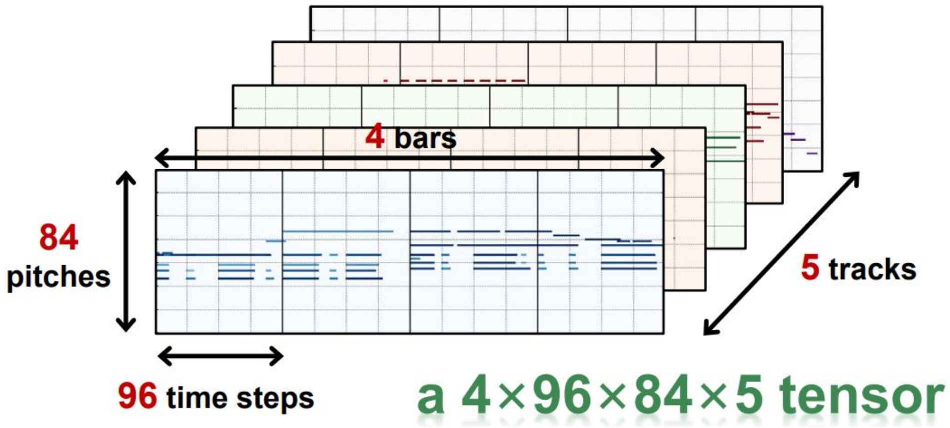
Li-Chia Yang

Yi-Hsuan Yang

Overview

Generate pop music

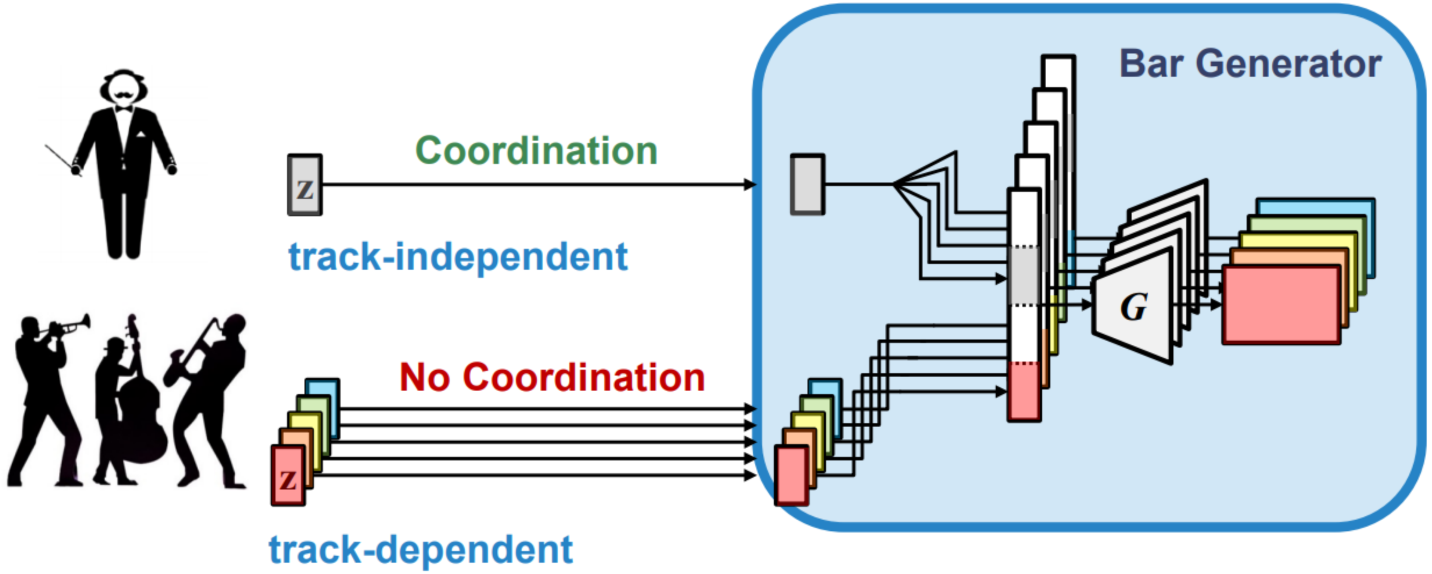
- of five polyphonic tracks
- in the piano-roll format
- using convolutional GANs (generative adversarial networks)
- on the Lakh MIDI Dataset



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,” *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, 2018.

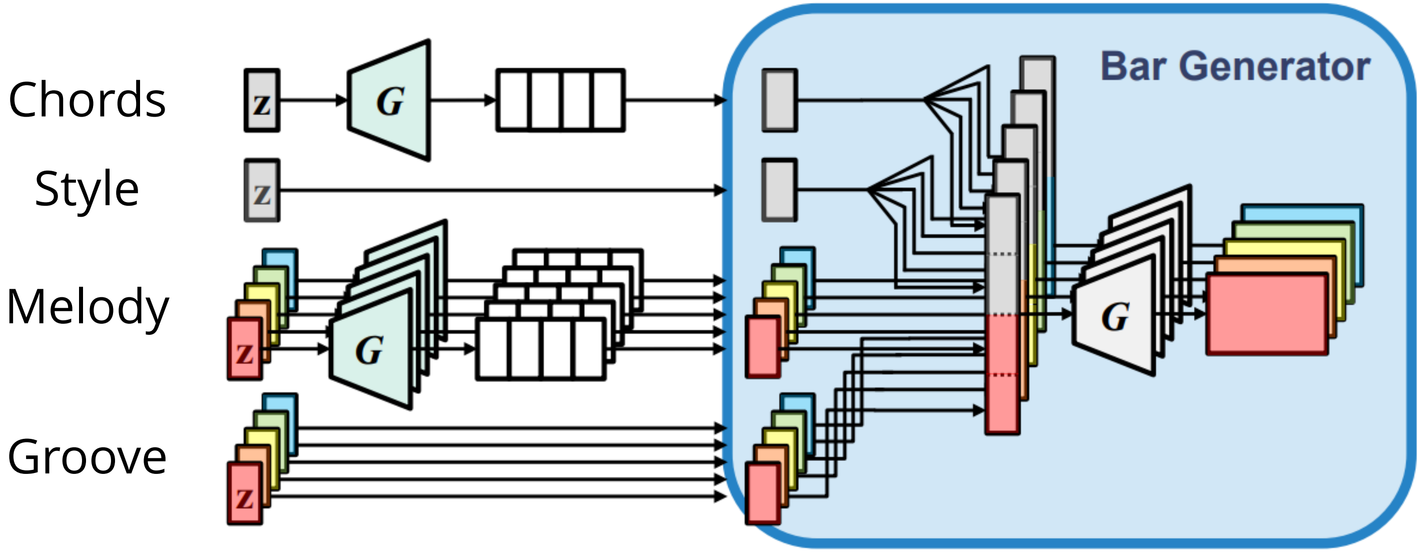
MuseGAN – Model

- Each track takes a shared and a private random vectors as inputs



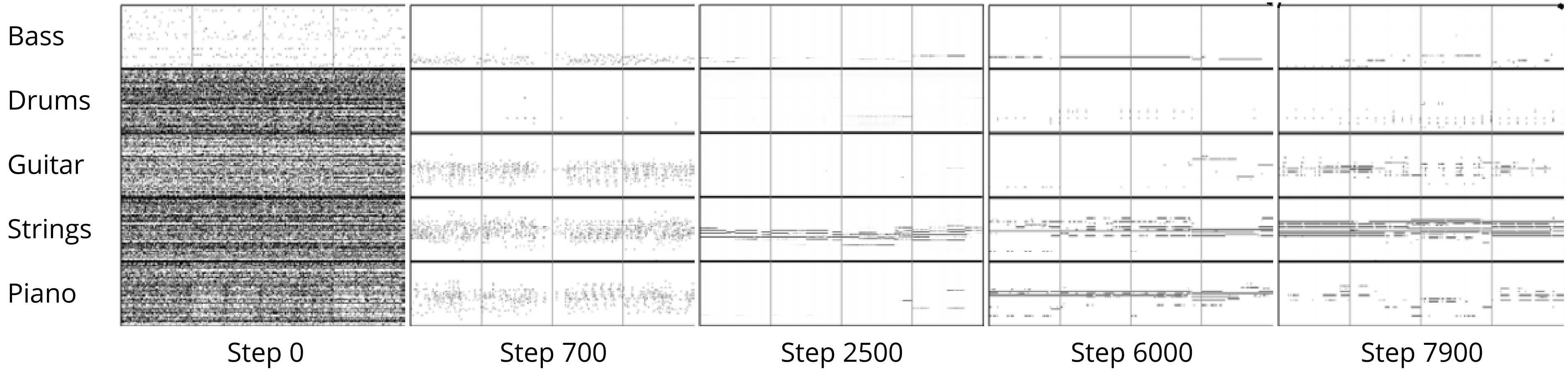
MuseGAN – Model

- Each random vector inputs corresponds to different aspects of music
  - Offer better controllability than one single random vector input



Demo

Unconditional generation samples



Training progress

More examples are available at salu133445.github.io/musegan/ .

Summary

- Proposed the first deep learning model for generating music consisting of multiple polyphonic tracks
- Proposed the shared and private latent variables to enhance the controllability
- Showed that the proposed model can learn basic musical concepts

Arranger

Approaching automatic instrumentation by learning to separate parts

(ISMIR 2021)

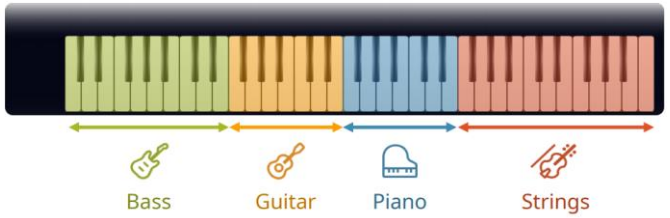


Chris Donahue Taylor Berg-Kirkpatrick Julian McAuley

Overview

Dynamically assign instruments to notes in solo music

- by learning to separate parts from a mixture
- using LSTMs and transformers
- on four diverse datasets (Bach chorales, string quartets, game music, pop music)



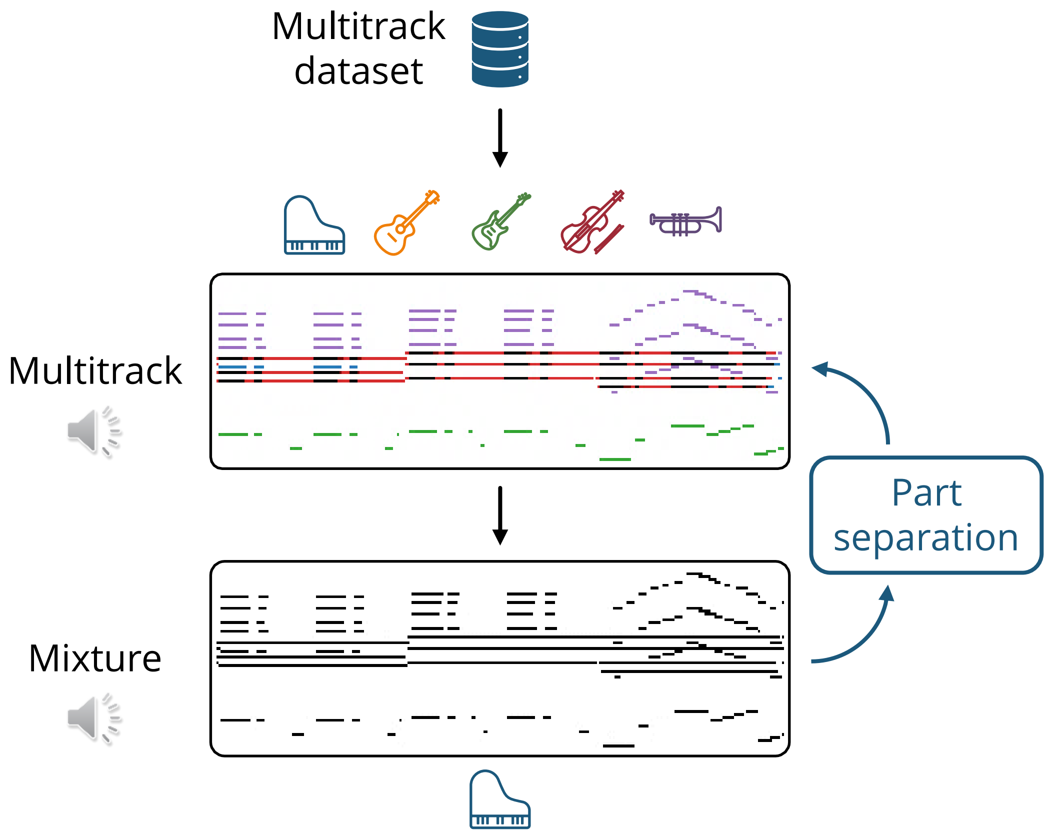
Intelligent keyboard

Assistive composing tools

Hao-Wen Dong, Chris Donahue, Taylor Berg-Kirkpatrick and Julian McAuley, "Towards Automatic Instrumentation by Learning to Separate Parts in Symbolic Multitrack Music," *Proceedings of the 22nd International Society for Music Information Retrieval Conference (ISMIR)*, 2021.

Pipeline

- Downmix multitracks into single-track mixtures (to acquire paired data)
- Train the model to predict the part label for each note in a mixture
- Treat input from a keyboard player as a downmixed mixture and separate out the relevant parts (to perform automatic instrumentation)



Arranger – Data

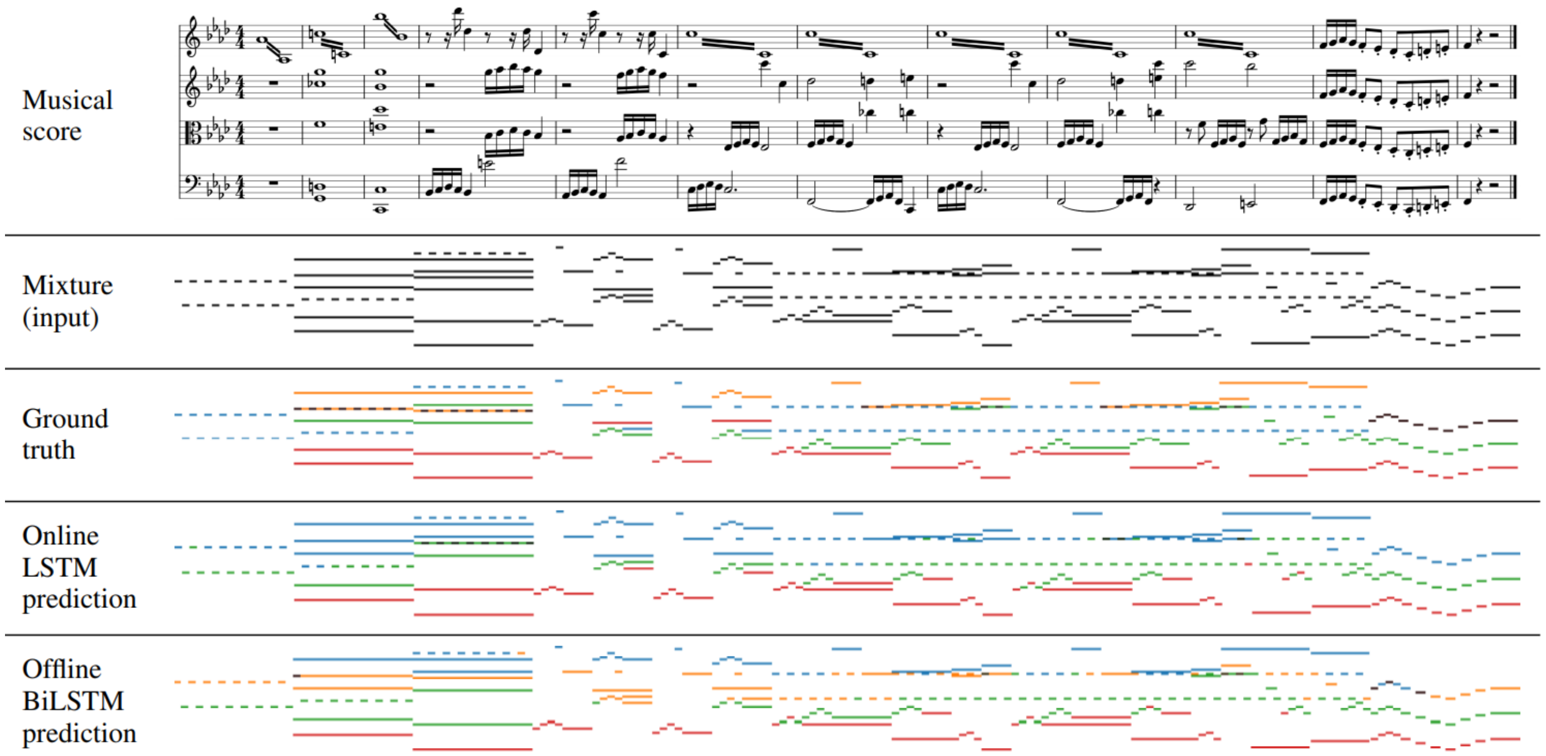
- Four datasets of diverse genres and ensembles

Dataset	Hours	Files	Notes	Parts	Ensemble	Most common label
Bach chorales [31]	3.23	409	96.6K	4	soprano, alto, tenor, bass	bass (27.05%)
String quartets [32]	6.31	57	226K	4	first violin, second violin, viola, cello	first violin (38.72%)
Game music [33]	45.05	4.61K	2.46M	3	pulse wave I, pulse wave II, triangle wave	pulse wave II (39.35%)
Pop music [34]	1.02K	16.2K	63.6M	5	piano, guitar, bass, strings, brass	guitar (42.50%)

A challenging example

Beethoven's String Quartet No. 11 in F minor

(op. 95, movement 1, measures 72–83)



(Audio available.[1] Colors: first violin, second violin, viola, cello.)

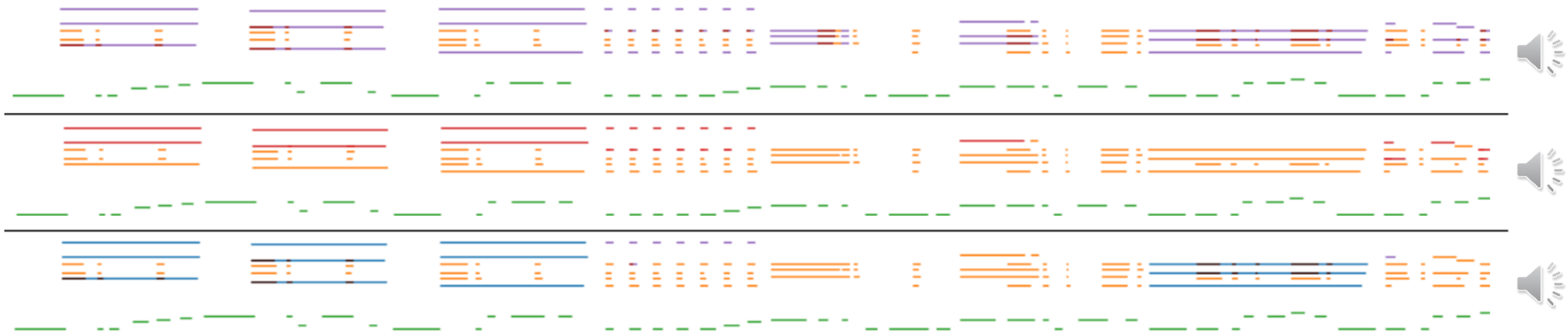
Demo

- The proposed models can produce alternative convincing instrumentations for an existing arrangement

Original

LSTM without entry hints

BiLSTM with entry hints



(Audio available. [1] Colors: piano, guitar, bass, strings, brass.)

More examples are available at salu133445.github.io/arranger/ .

Summary

- Approached automatic instrumentation by learning to separate parts
- Showed that our proposed models outperform various baselines
- Produced alternative convincing instrumentations for an existing arrangement

Multitrack Music Transformer

Generating multitrack music using transformers

Ke Chen

Shlomo Dubnov

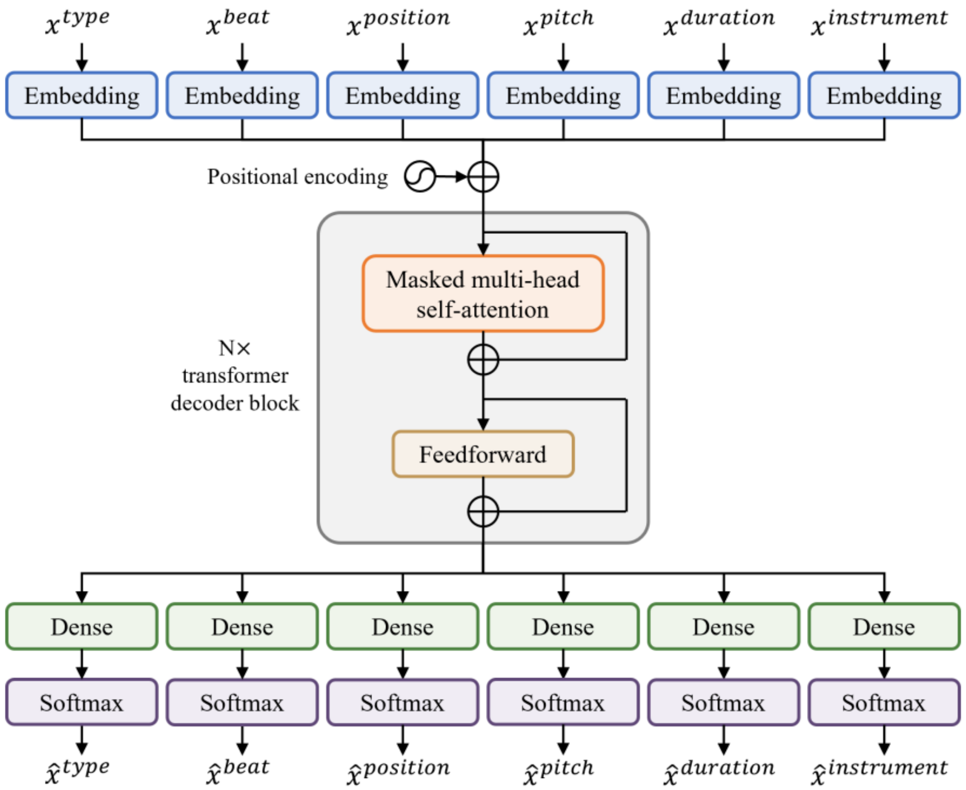
Julian McAuley

Taylor Berg-Kirkpatrick

Overview

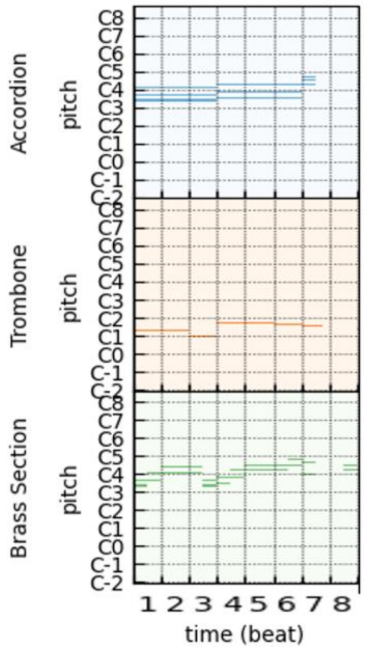
Generate music

- of diverse instruments
- with a multi-dimensional transformer
- using a new compact representation
- on pop and orchestral music datasets



Representation

- Represent 2-4x longer music within the same sequence length (compared to existing representations)



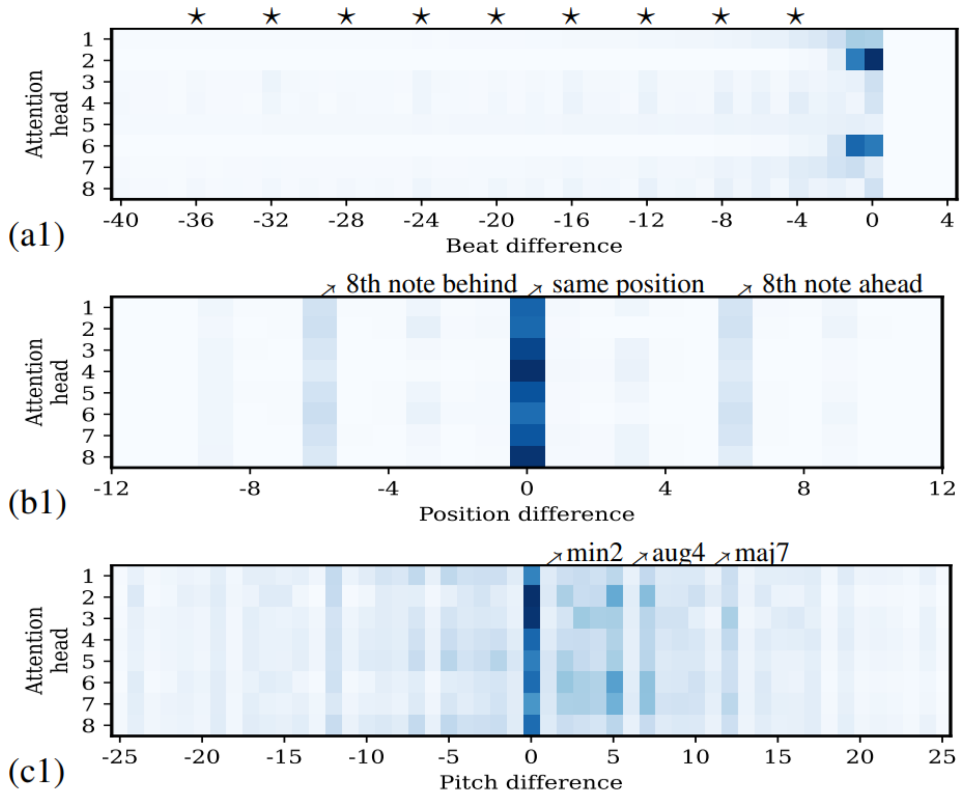

```
(0, 0,  0,  0,  0,  0)   Start of song
(1, 0,  0,  0,  0, 15)   Instrument: accordion
(1, 0,  0,  0,  0, 36)   Instrument: trombone
(1, 0,  0,  0,  0, 39)   Instrument: brasses
(2, 0,  0,  0,  0,  0)   Start of notes
(3, 1,  1, 41, 15, 36)   Note: beat=1, position=1,  pitch=E2, duration=48, instrument=trombone
(3, 1,  1, 65,  4, 39)   Note: beat=1, position=1,  pitch=E4, duration=12, instrument=brasses
(3, 1,  1, 65, 17, 15)   Note: beat=1, position=1,  pitch=E4, duration=72, instrument=accordion
(3, 1,  1, 68,  4, 39)   Note: beat=1, position=1,  pitch=G4, duration=12, instrument=brasses
(3, 1,  1, 68, 17, 15)   Note: beat=1, position=1,  pitch=G4, duration=72, instrument=accordion
(3, 1,  1, 73, 17, 15)   Note: beat=1, position=1,  pitch=C5, duration=72, instrument=accordion
(3, 1, 13, 68,  4, 39)   Note: beat=1, position=13, pitch=G4, duration=12, instrument=brasses
(3, 1, 13, 73,  4, 39)   Note: beat=1, position=13, pitch=C5, duration=12, instrument=brasses
(3, 2,  1, 73, 12, 39)   Note: beat=2, position=1,  pitch=C5, duration=36, instrument=brasses
(3, 2,  1, 77, 12, 39)   Note: beat=2, position=1,  pitch=E5, duration=36, instrument=brasses
         ...             ...
(4, 0,  0,  0,  0,  0)   End of song
```

Example results

Unconditional generation 1

Unconditional generation 2

4-beat continuation



Attention visualization

More examples are available at https://dezimynona.github.io/mtmt/ .

Summary

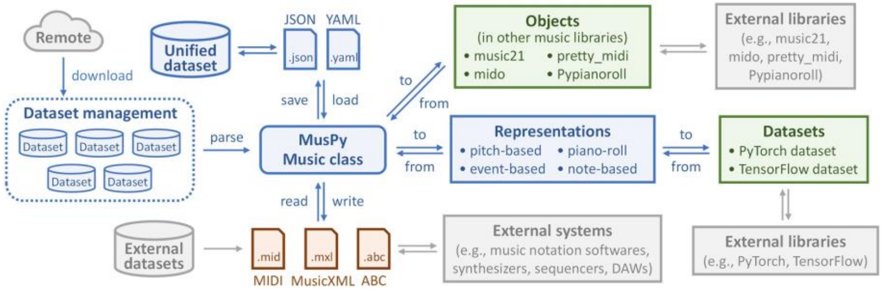
- Proposed a new representation that can represent 2-4x longer multitrack music within the same sequence length (compared to existing representations)
- Showed that the proposed model can achieve competitive quality against two baseline models (of similar sizes)
- Showed that the model can generate 2-3x more notes in the same inference time (compared to the two baseline models)

What's next?

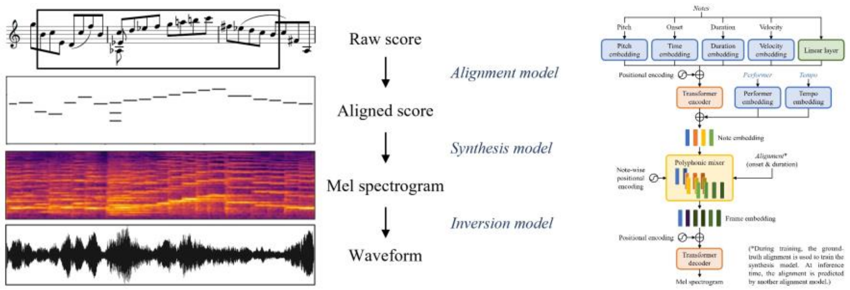
- Scaling up music generation models → MuseScore dataset (1.5M songs)
- Improving controllability of music generation systems

Some other projects

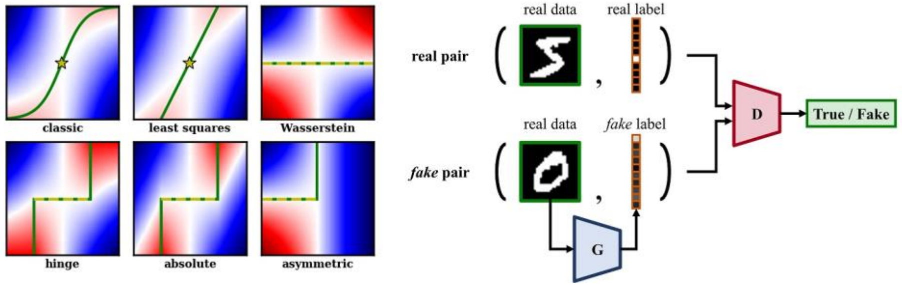
MusPy
A toolkit for symbolic music generation



DeepPerformer
Score-to-audio music performance synthesis



On Output Activation Functions for Adversarial Losses



Acknowledgment


Julian McAuley


Taylor Berg-Kirkpatrick


Shlomo Dubnov


Yi-Hsuan Yang


Ke Chen


Chris Donahue

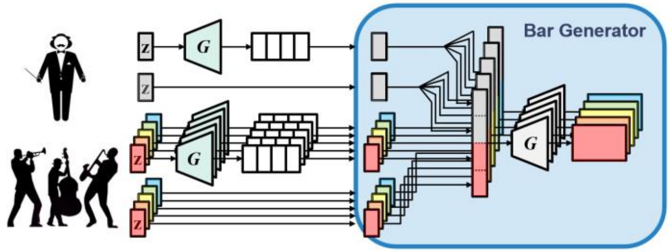

Wen-Yi Hsiao


Li-Chia Yang

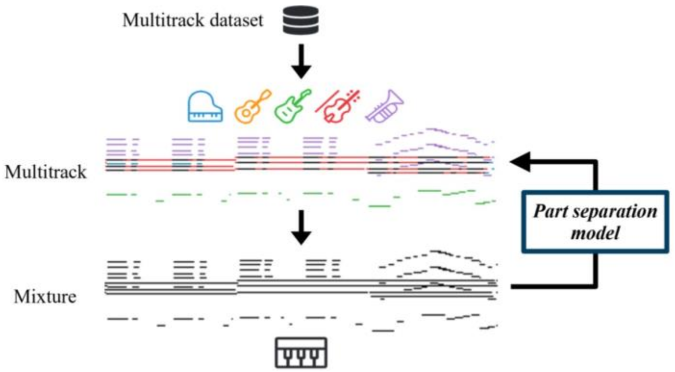
I would like to thank **J. Yang and Family Foundation** and
Taiwan Ministry of Education for supporting my PhD study.

Thank you!

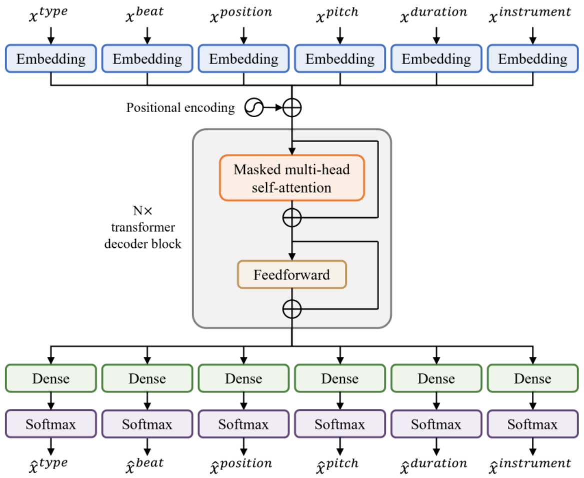
MuseGAN



Arranger



Multitrack Music Transformer



Learn more about my projects at salu133445.github.io .