MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment

Hao-Wen Dong*, Wen-Yi Hsiao*, Li-Chia Yang, Yi-Hsuan Yang

Research Center of IT Innovation, Academia Sinica

*these authors contributed equally to this work

Outlines

- Goals & Challenges
- Data
- Proposed Model
- Results & Evaluation
- Recent Works
- Future Works

Source Code https://github.com/salu133445/musegan
Demo Page https://salu133445.github.io/musegan/

Goals & Challenges

Goals

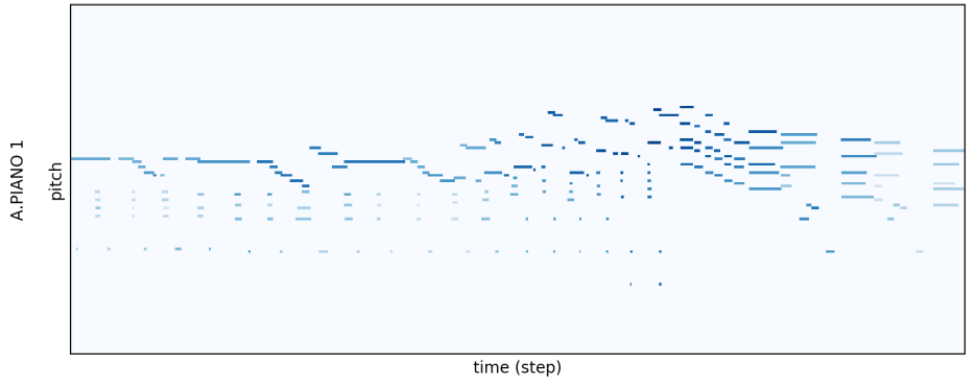
[Source Code]
https://github.com/salu133445/musegan

[Demo Page]
https://salu133445.github.io/musegan/

Generate pop music

- of **multiple tracks**

- in **piano-roll** format

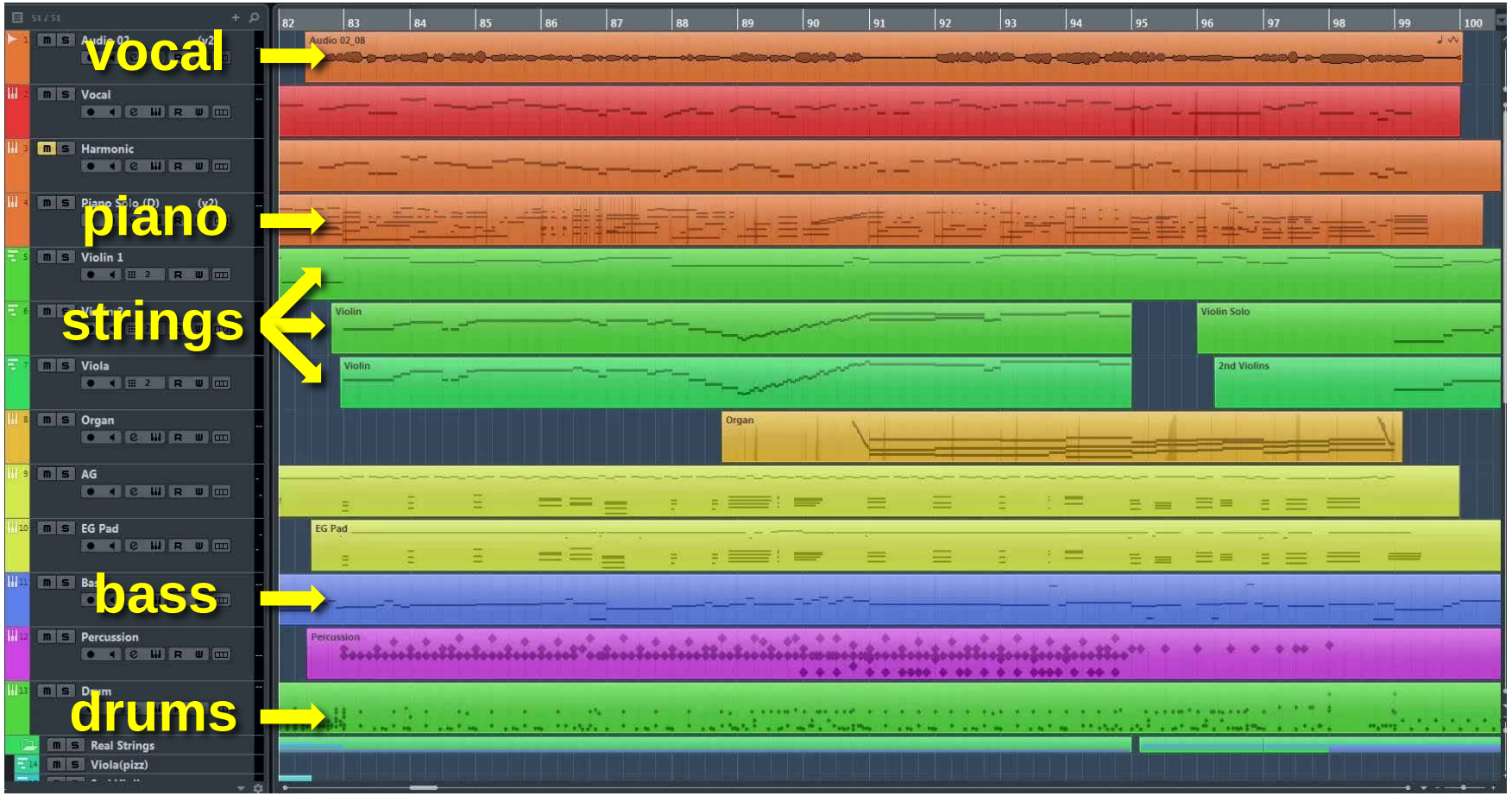


- using **GAN** with **CNNs**

Challenge I

Multi-track GAN

Multitrack Interdependency

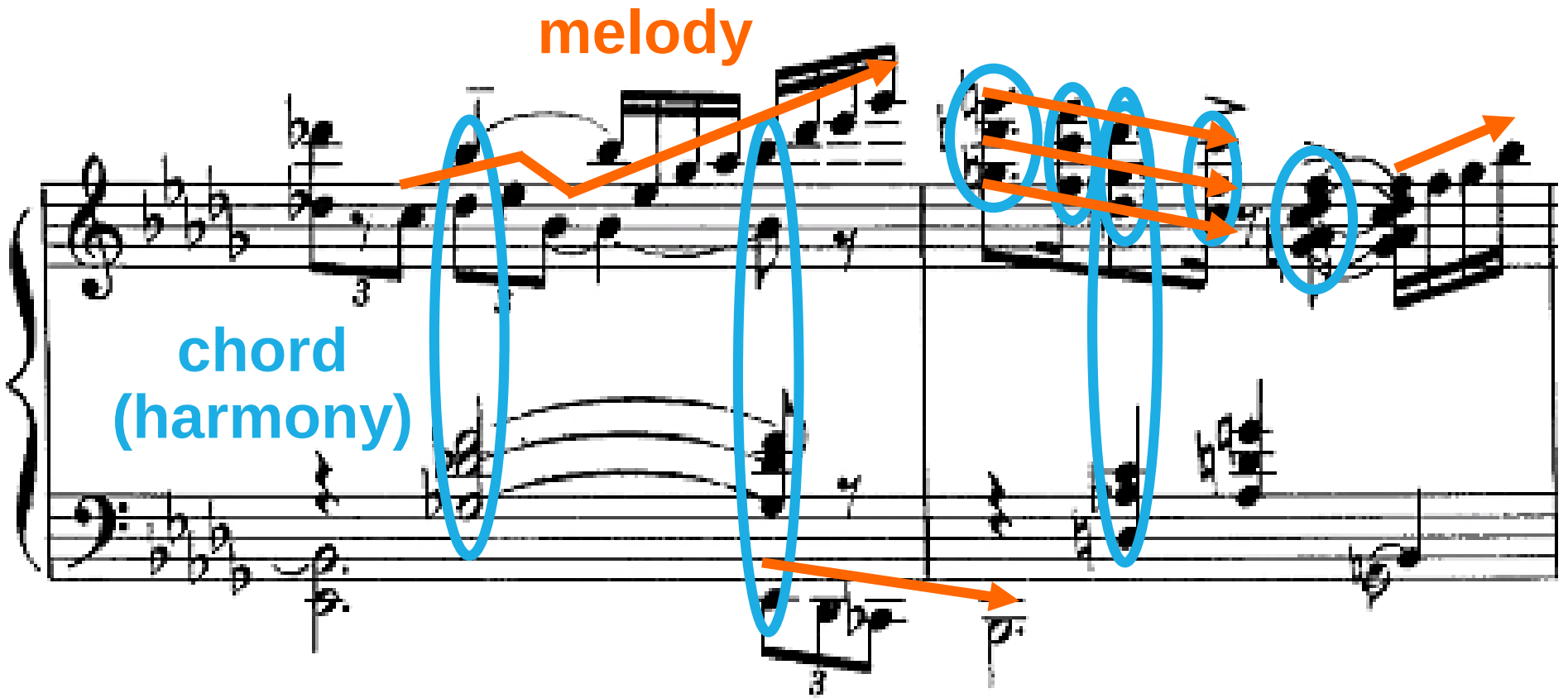


music & clip by
phycause

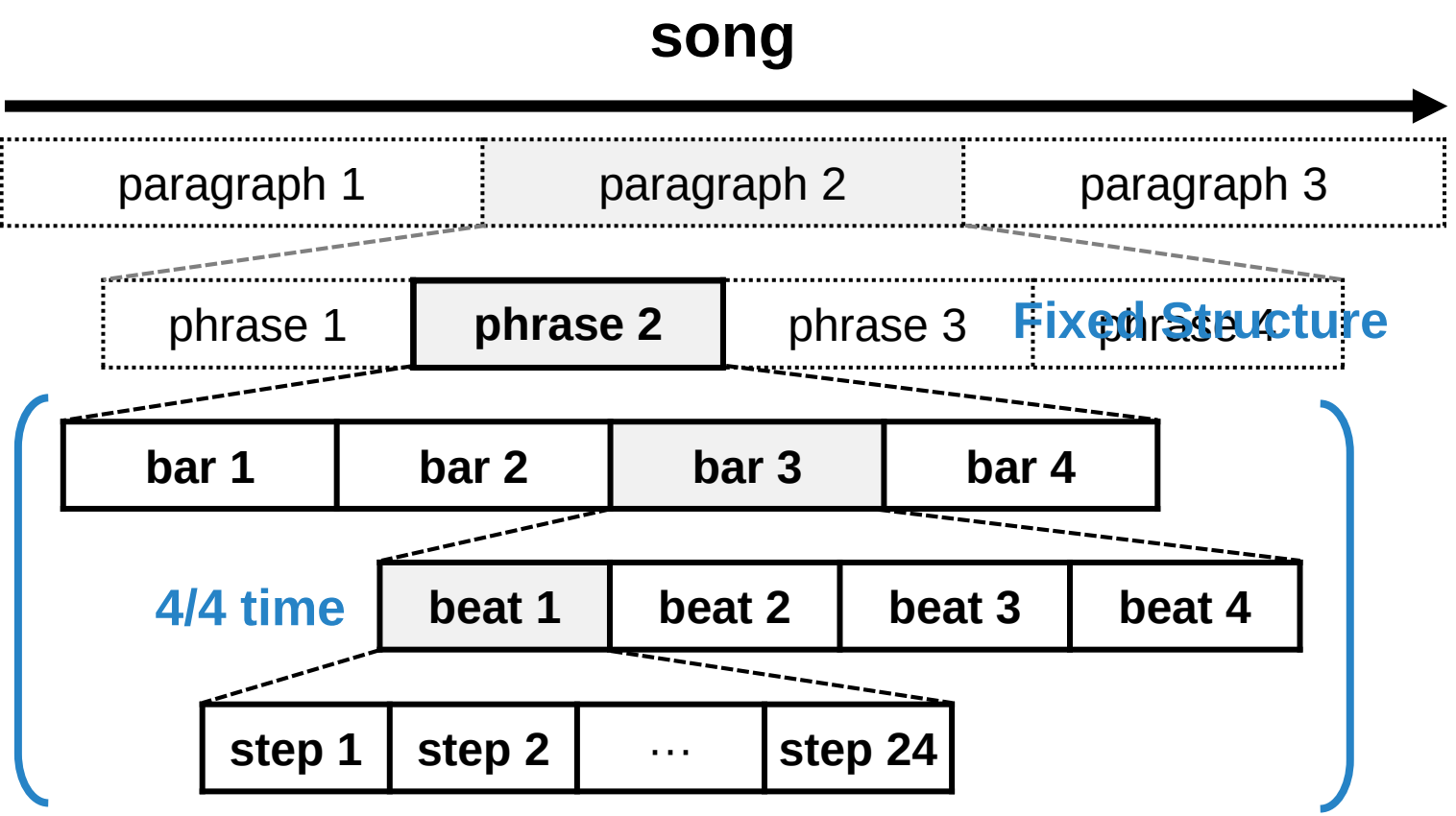
Challenge II

Convolutional Neural Networks

Music Texture



Temporal Structure



Data

Data Representation

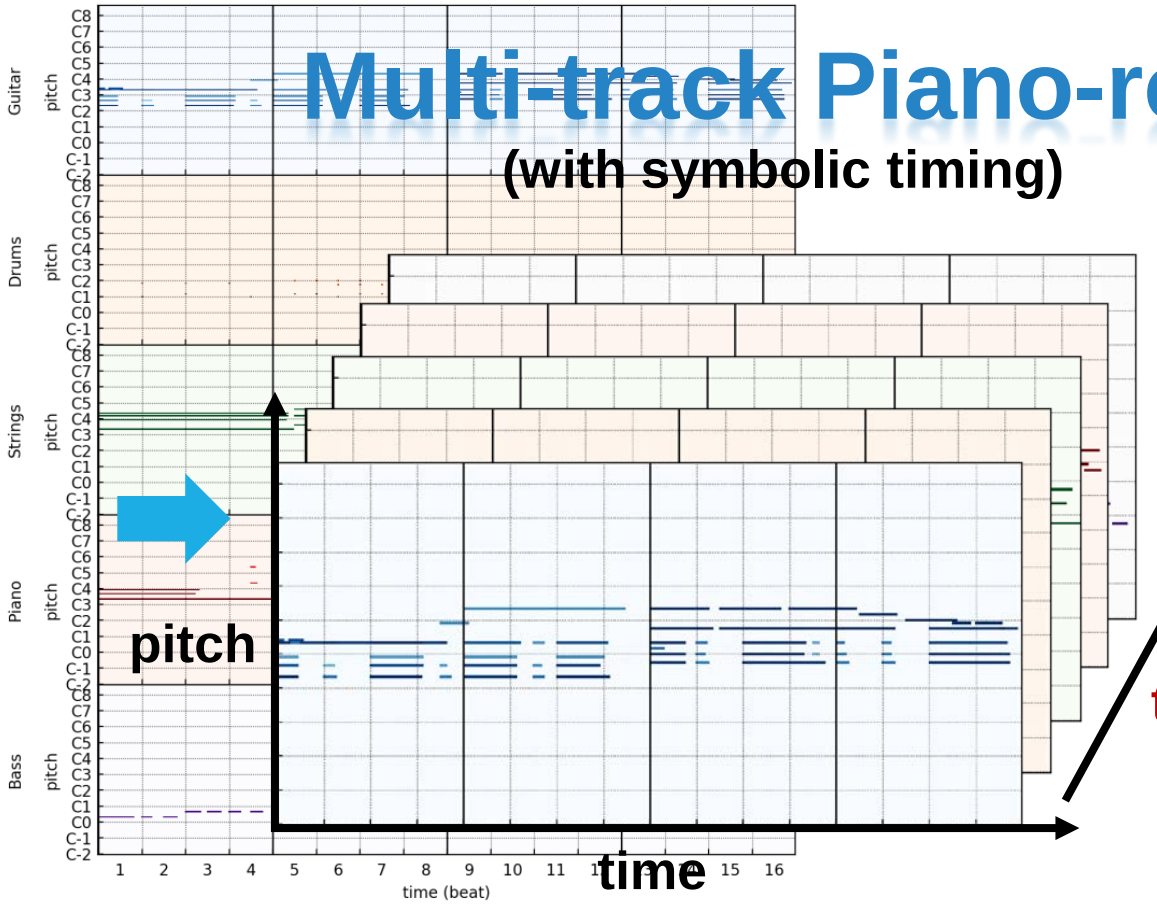
Piano-roll

(with symbolic timing)

polyphonic ✓

multi-track ✗

Data Representation



Data Representation

a 4×96×84×5 tensor

Data

[Dataset]
https://salu133445.github.io/lakh-pianoroll-dataset

[Pypianoroll]
https://salu133445.github.io/pypianoroll/

LPD (Lakh Pianoroll Dataset)

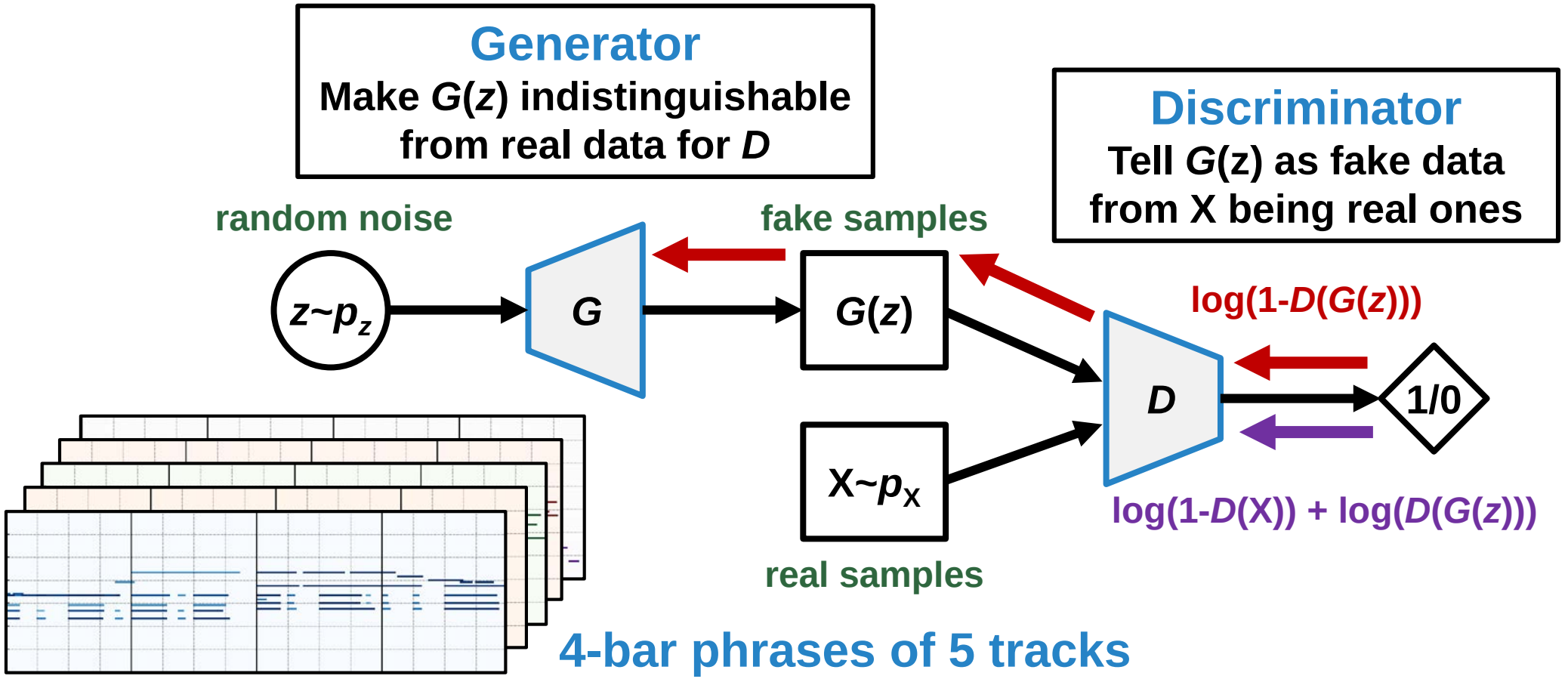
- **>170,000** multi-track piano-rolls
- Derived from Lakh MIDI Dataset
- Mainly pop songs

Pypianoroll (Python package)

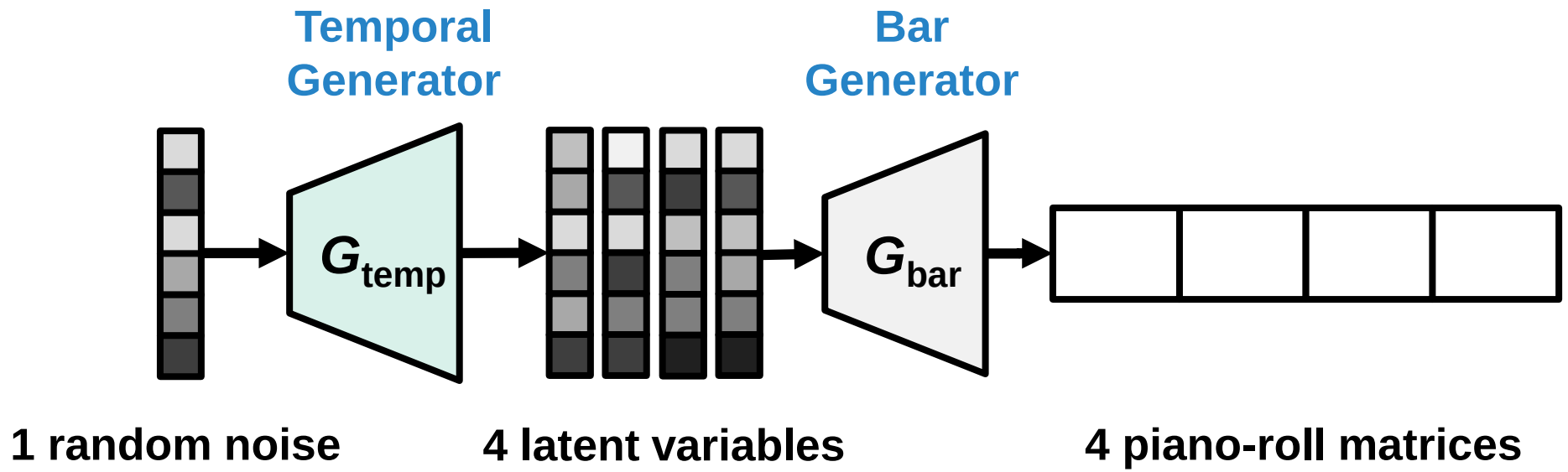
- Manipulation & Visualization
- Efficient I/O
- Parse/Write MIDI files
- On **PYPI** (pip installable)

Proposed Model

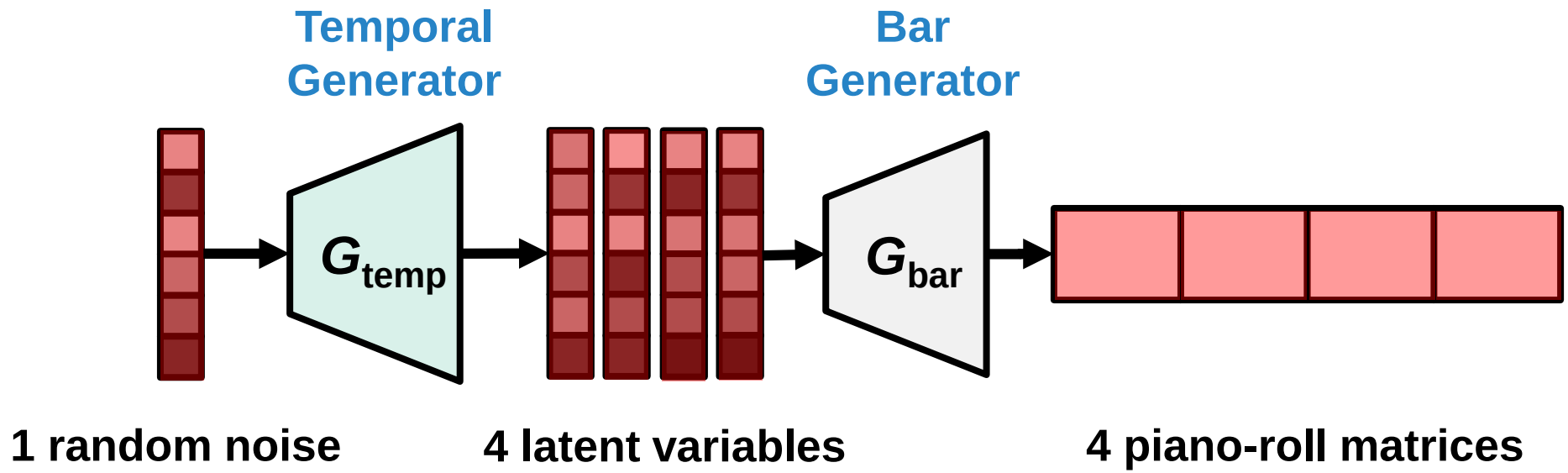
Generative Adversarial Networks



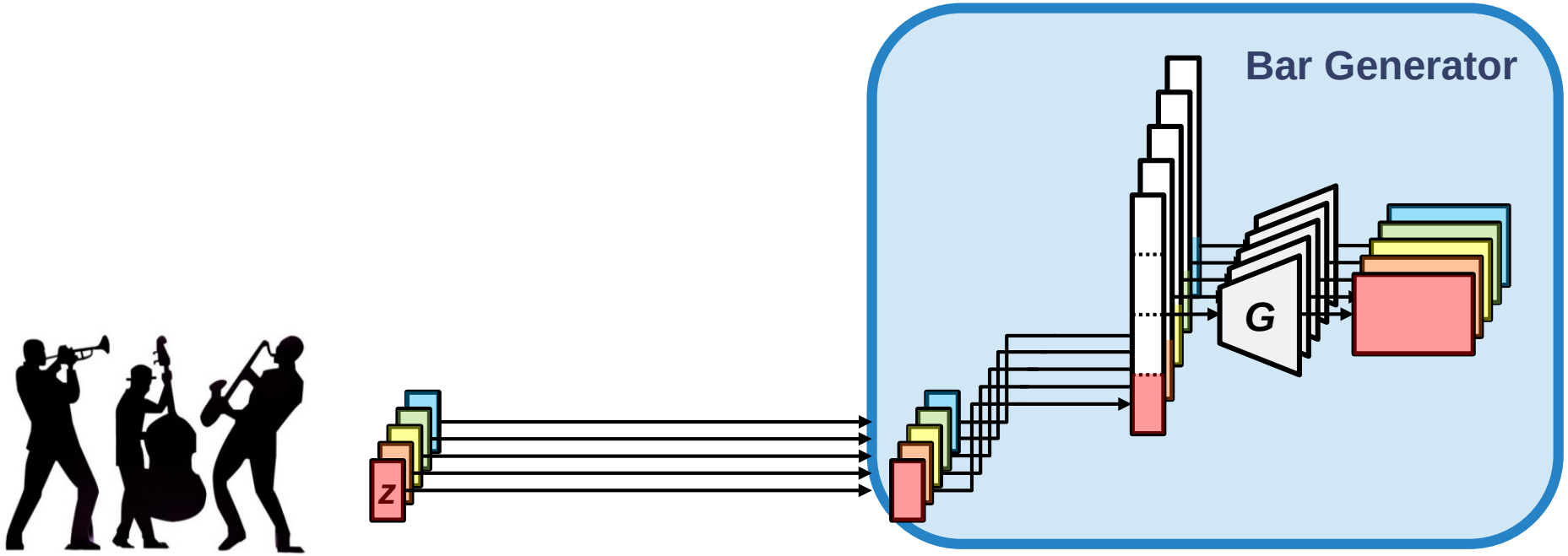
MuseGAN – An Overview



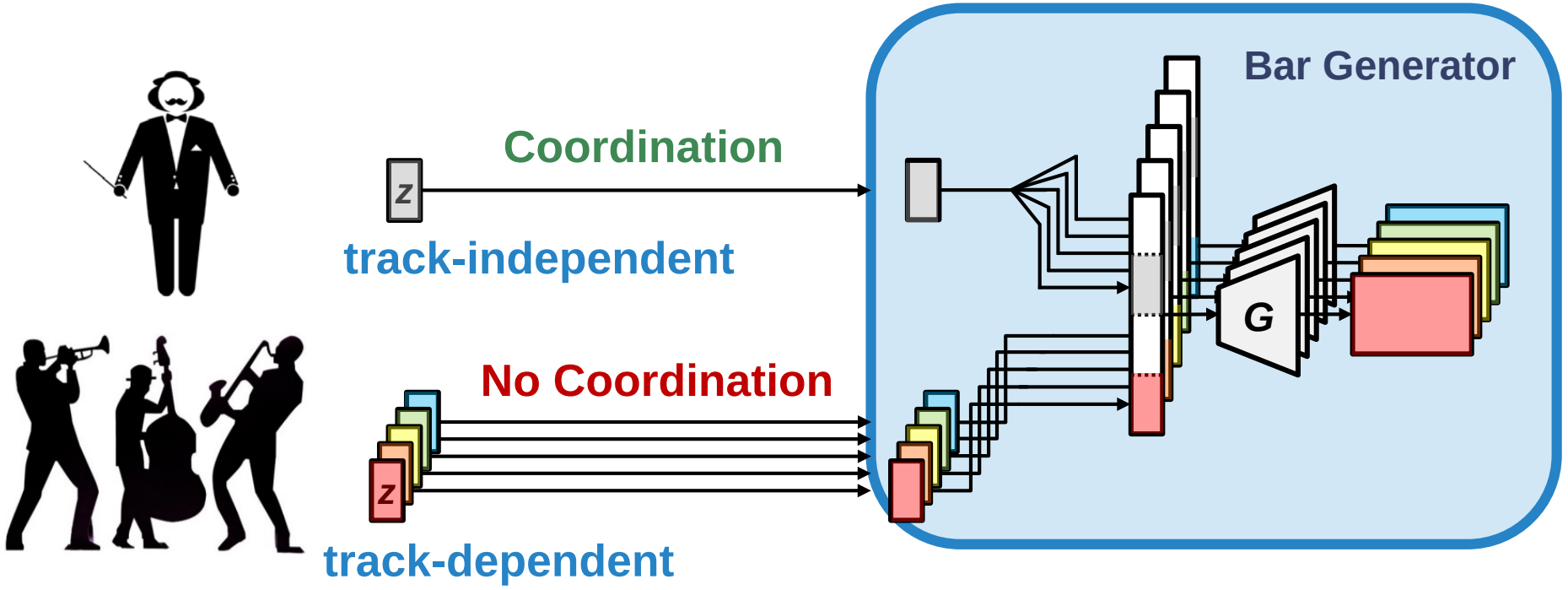
MuseGAN – An Overview



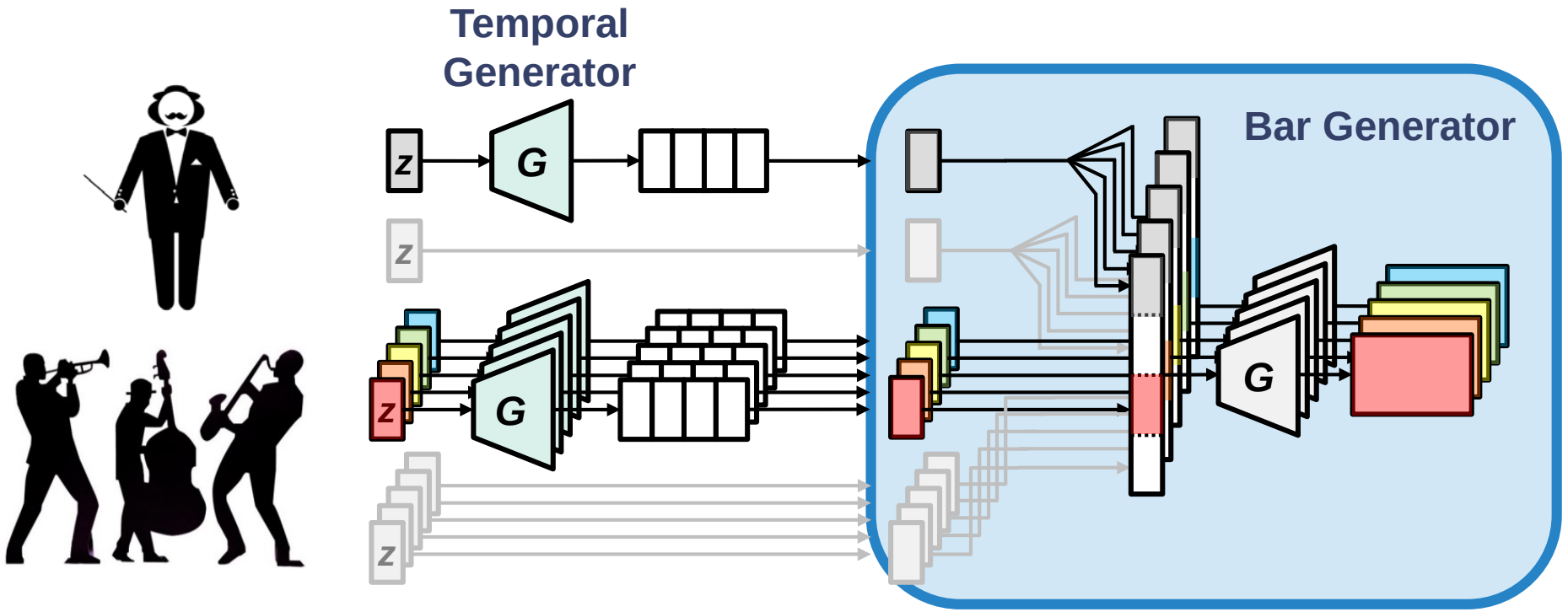
Generator



Generator

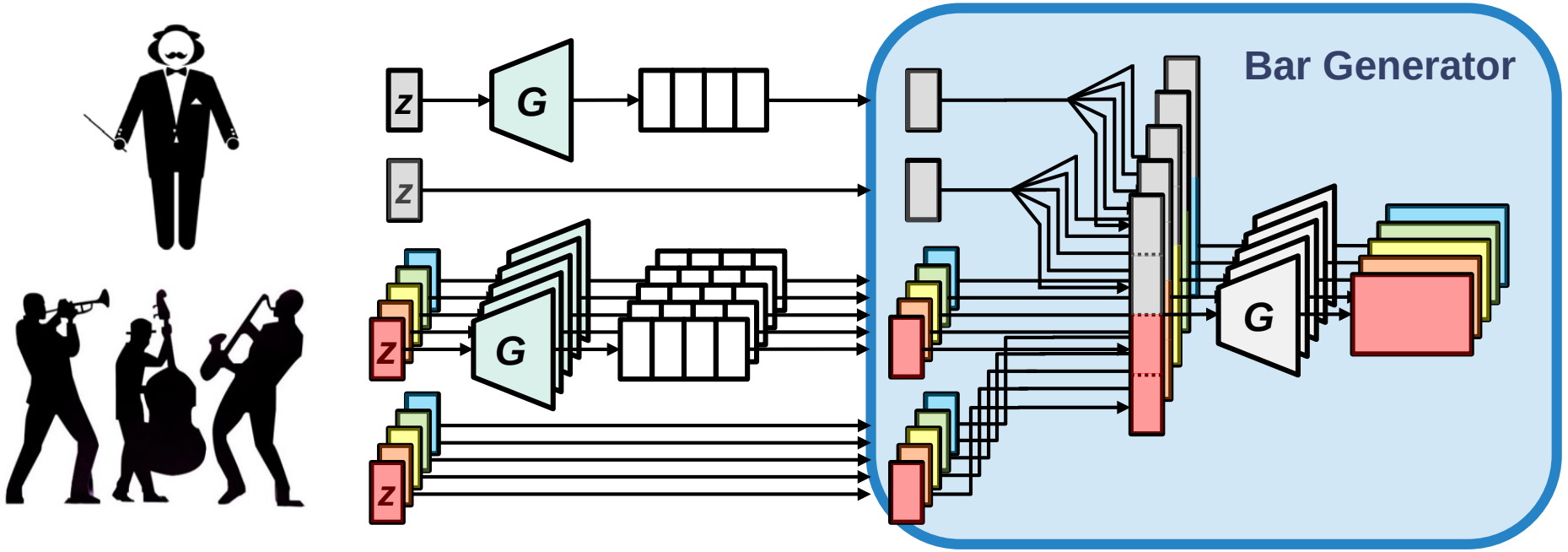

Coordination
z
track-independent
No Coordination
z
track-dependent
Bar Generator
G

Generator

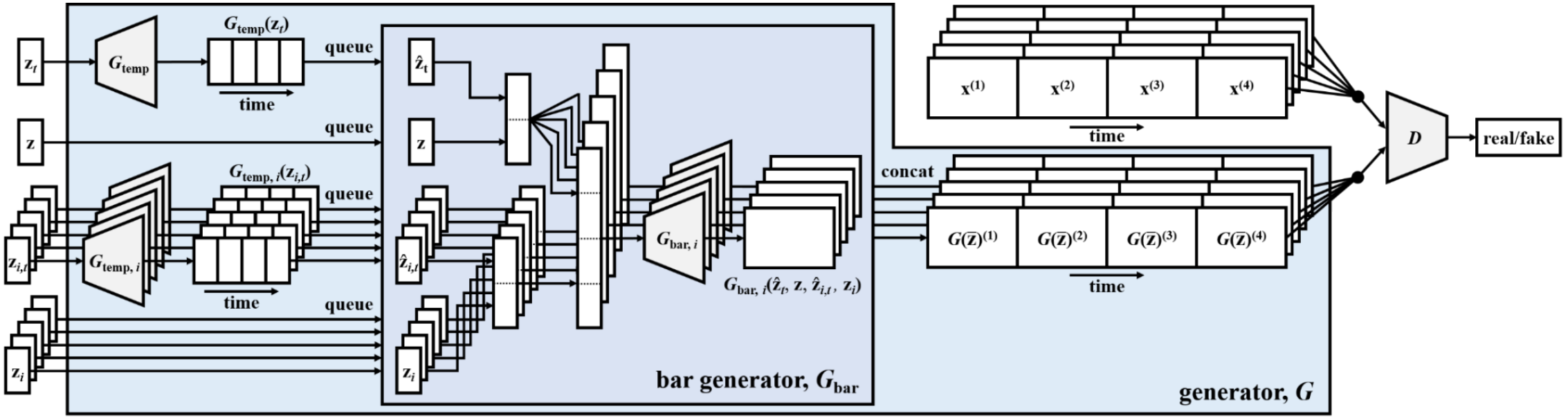


Generator

		Time	
		Dependent	Independent
Track	Dependent	**Melody**	**Groove**
	Independent	**Chords**	**Style**



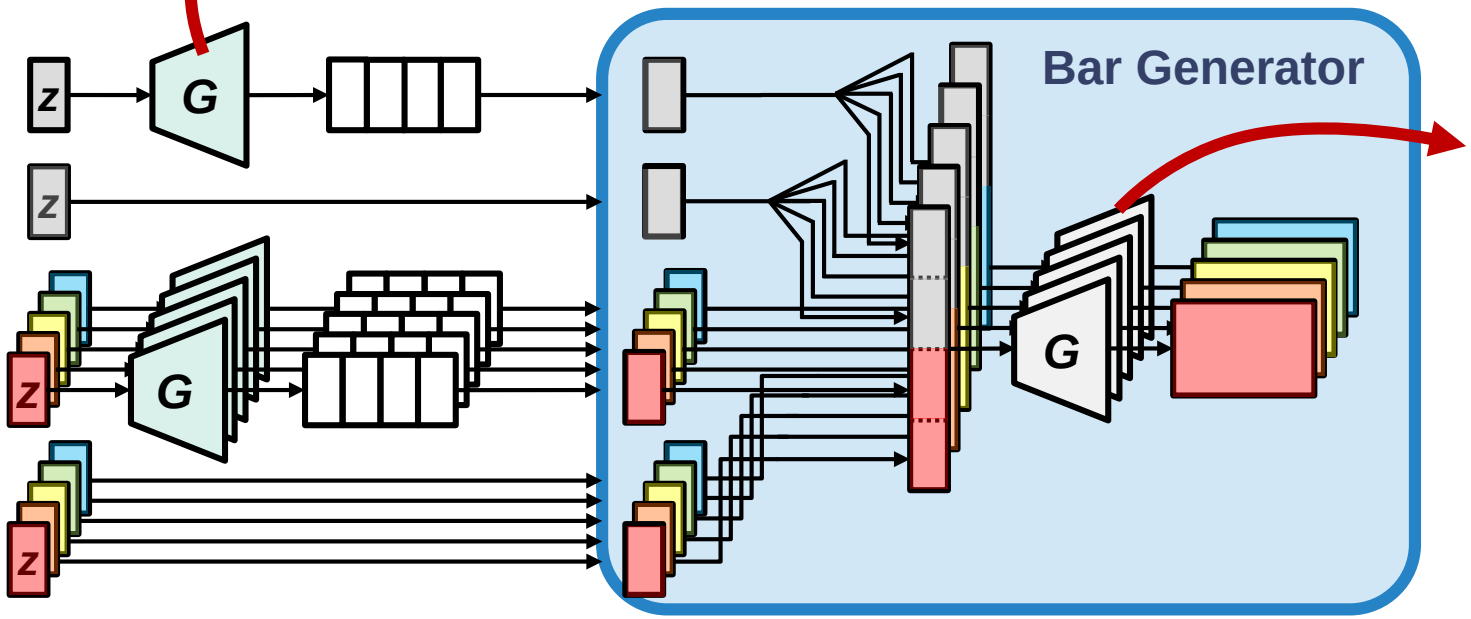
MuseGAN


z_t
G_{temp}
$G_{temp}(z_t)$
queue
time
z
queue
$z_{i,t}$
$G_{temp,\,i}$
$G_{temp,\,i}(z_{i,t})$
queue
time
z_i
queue
$\hat{z}_t$
z
$\hat{z}_{i,t}$
z_i
$G_{bar,\,i}$
$G_{bar,\,i}(\hat{z}_t, z, \hat{z}_{i,t}, z_i)$
bar generator, G_{bar}
concat
$x^{(1)}$ $x^{(2)}$ $x^{(3)}$ $x^{(4)}$
time
$G(\overline{z})^{(1)}$ $G(\overline{z})^{(2)}$ $G(\overline{z})^{(3)}$ $G(\overline{z})^{(4)}$
time
generator, G
D
real/fake

Network Architectures

Temporal Generator

Input	32			
transconv	1024	2	1	$(3, 1024)$
transconv	32	3	1	$(4, 32)$
Output	32			



Input	32			
dense	1024			
reshape to $(2, 1) \times 512$ *channels*				$(2, 1, 512)$
transconv	512	2×1	$(2, 1)$	$(4, 1, 512)$
transconv	256	2×1	$(2, 1)$	$(8, 1, 256)$
transconv	256	2×1	$(2, 1)$	$(16, 1, 256)$
transconv	128	2×1	$(2, 1)$	$(32, 1, 128)$
transconv	128	3×1	$(3, 1)$	$(96, 1, 128)$
transconv	64	1×7	$(1, 7)$	$(96, 7, 64)$
transconv	M	1×12	$(1, 12)$	$(96, 84, M)$
Output	$(96, 84) \times M$ channels			

Results

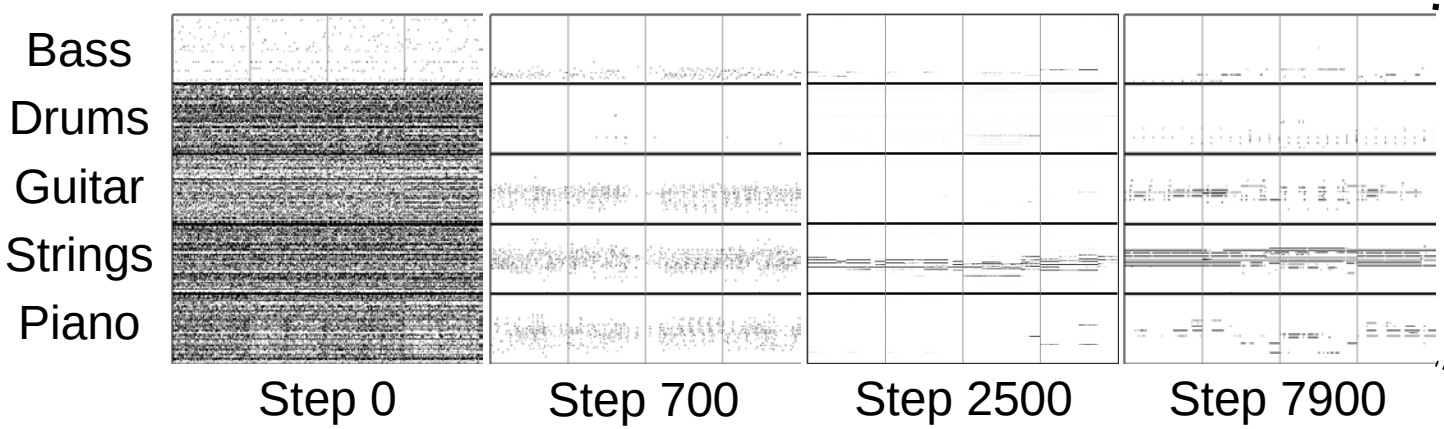
Results

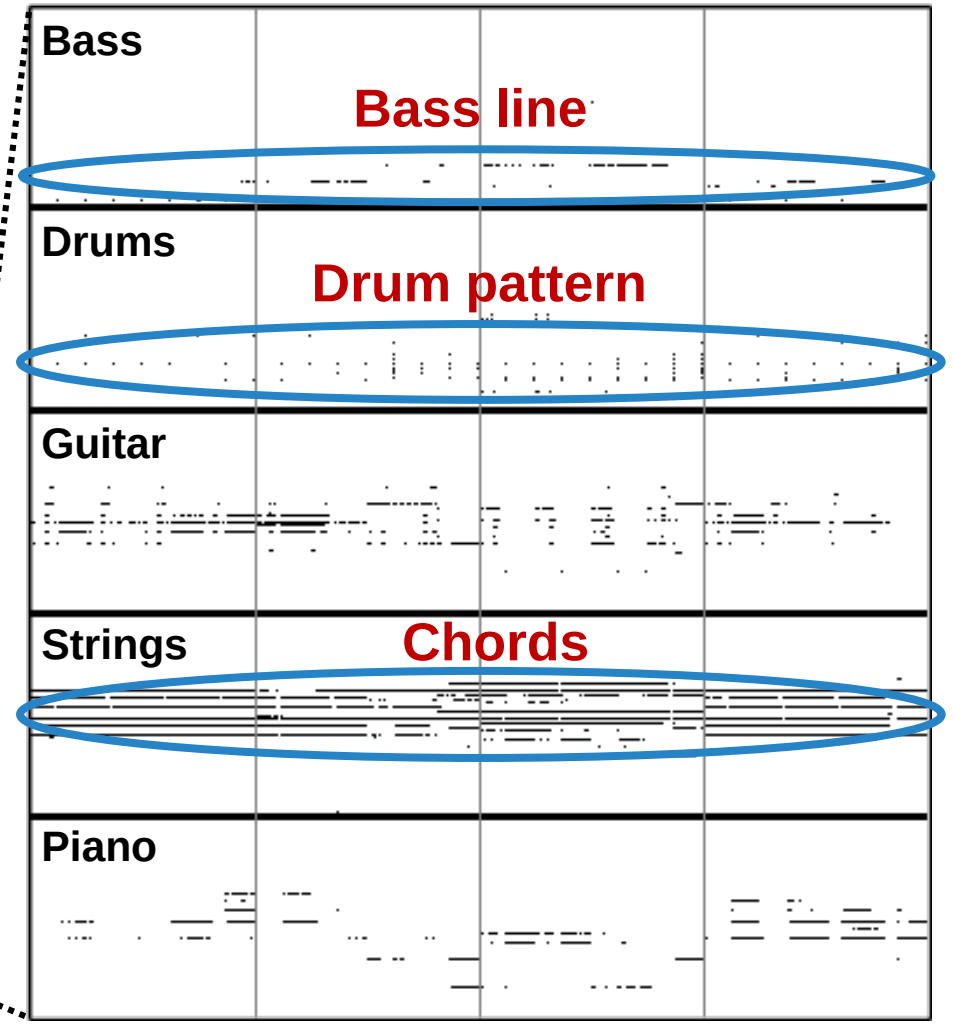
Sample 1

Sample 2

More samples available on demo page

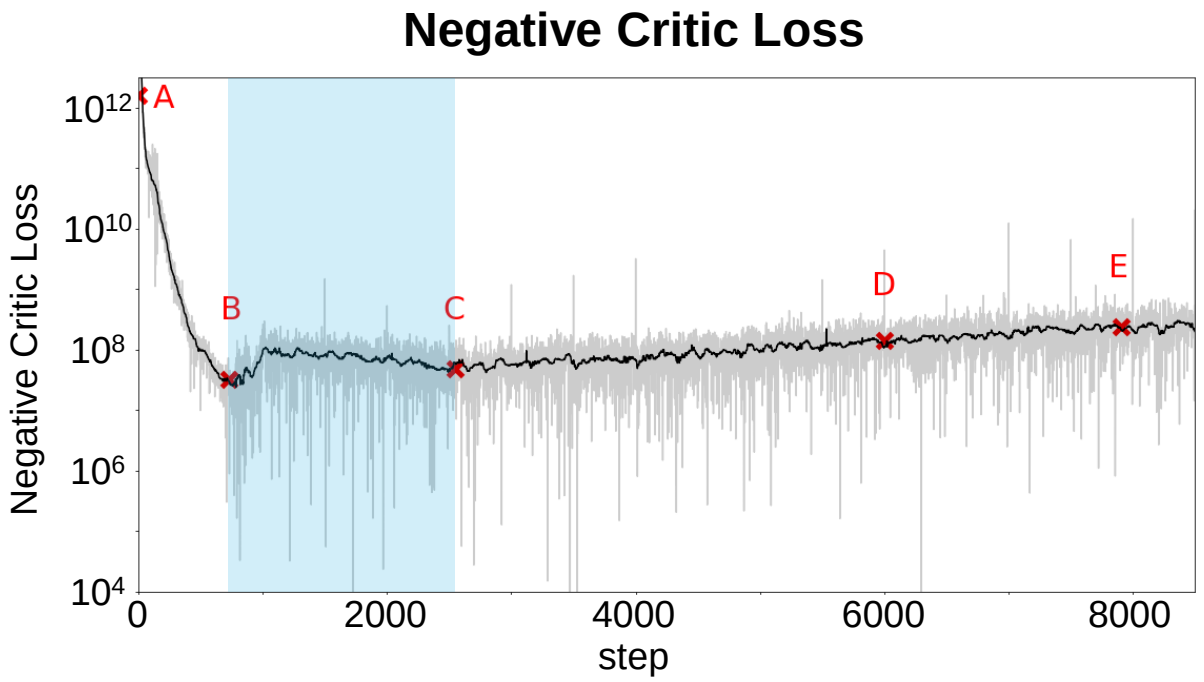
https://salu133445.github.io/musegan/

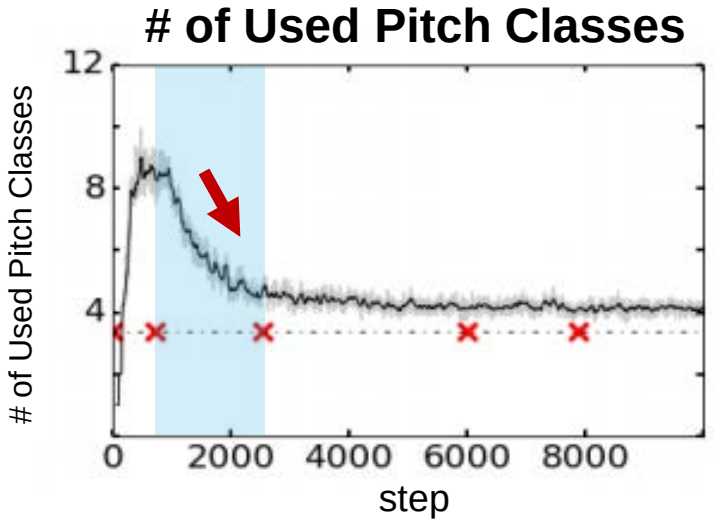


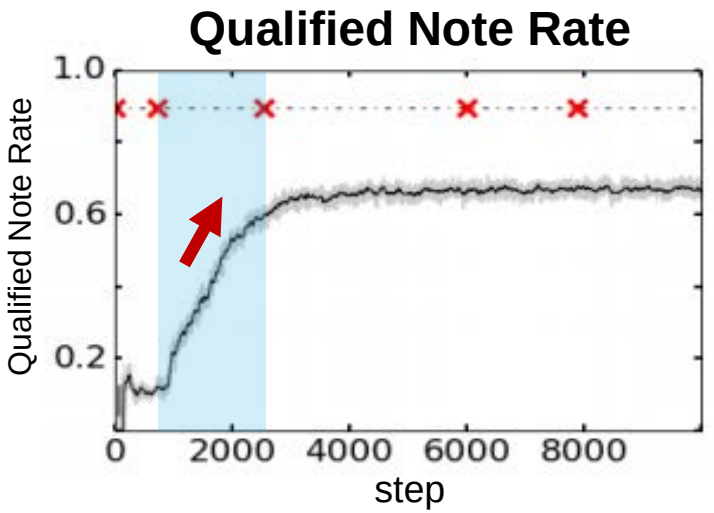


Monitor the Training

Objective Metrics







User Study

from scratch		H	R	MS	C	OR
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**

H: harmonious

R: rhythmic

MS: musically structured

C: coherent

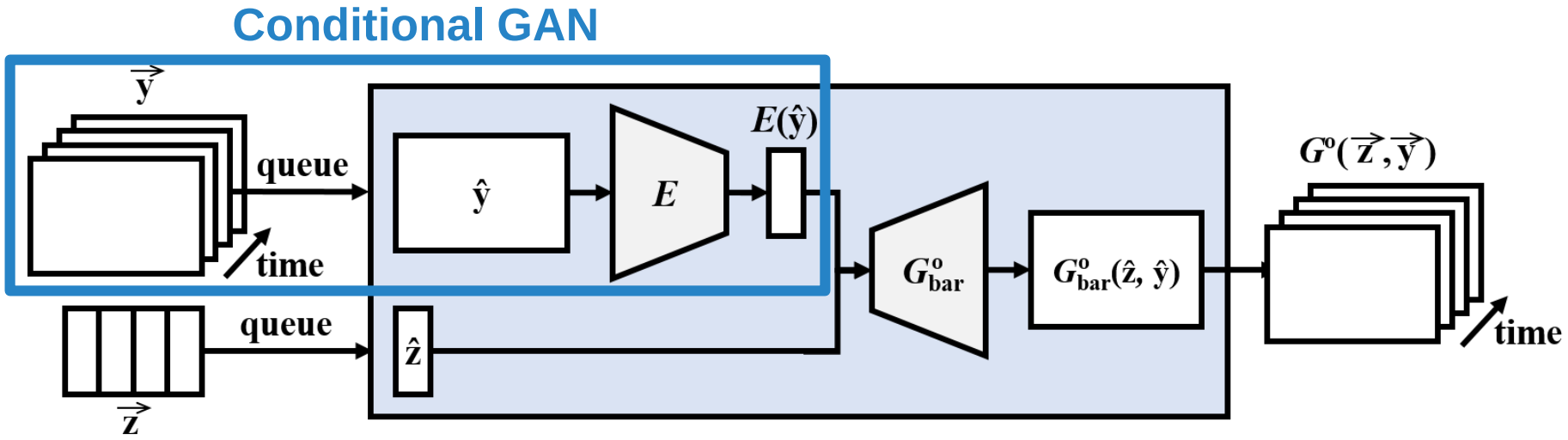
OR: overall rating

composer

jamming

hybrid

Accompaniment System



Generation from Scratch	**nothing → 5-track**
Accompaniment System	**single-track → 5-track**

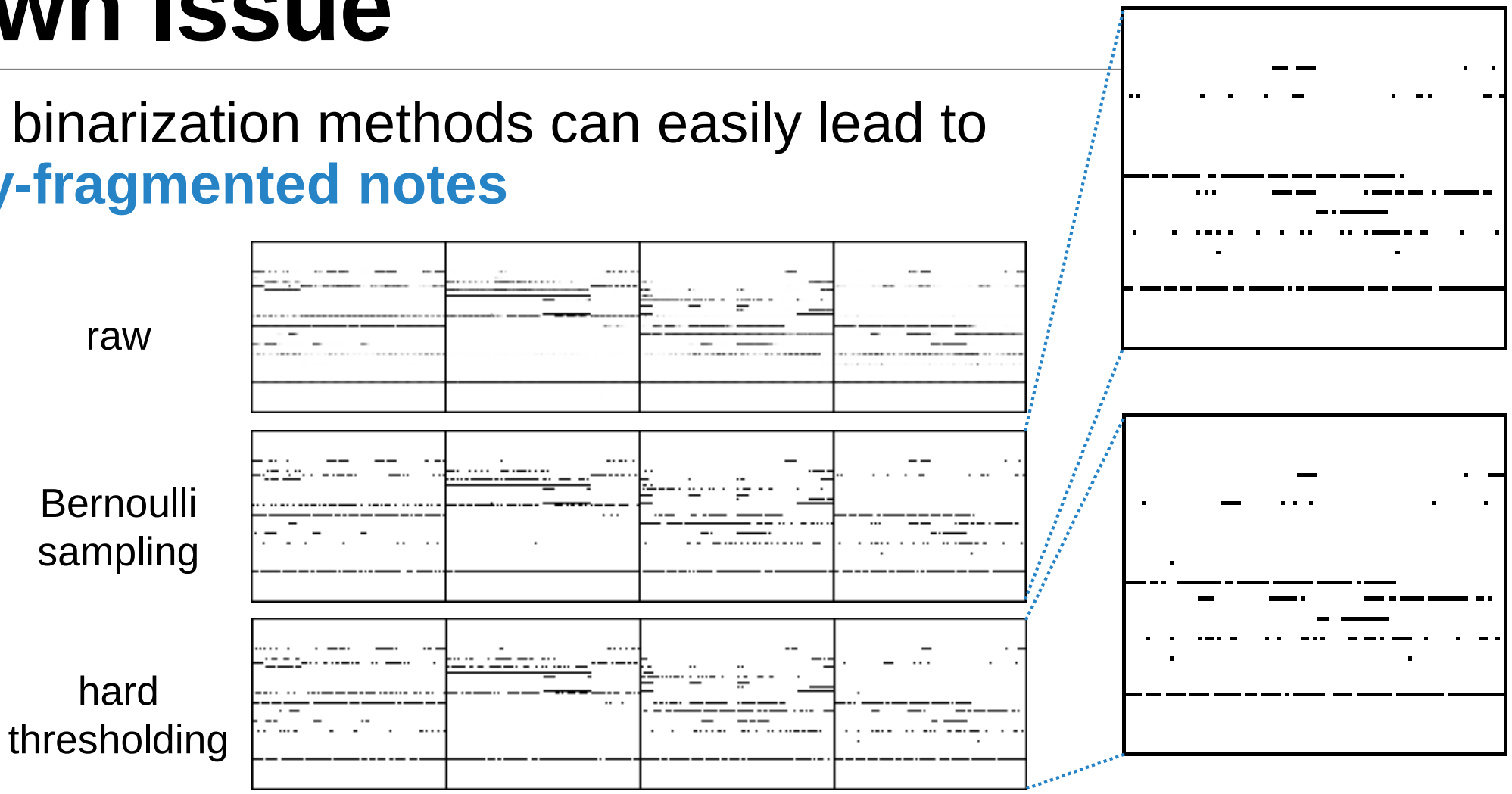
Summary

- **MuseGAN**
  - a novel GAN for **multi-track sequence** generation
  - **multi-track**, **polyphonic** music
  - **human-AI cooperative** scenario
- **Lakh Pianoroll Dataset (LPD)** (**new dataset**)
- **Pypianoroll** (**new Python package**)

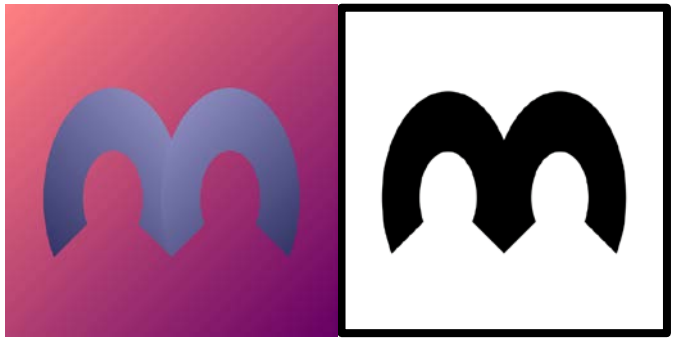
Recent Works

Known Issue

- Naïve binarization methods can easily lead to **overly-fragmented notes**



BinaryMuseGAN

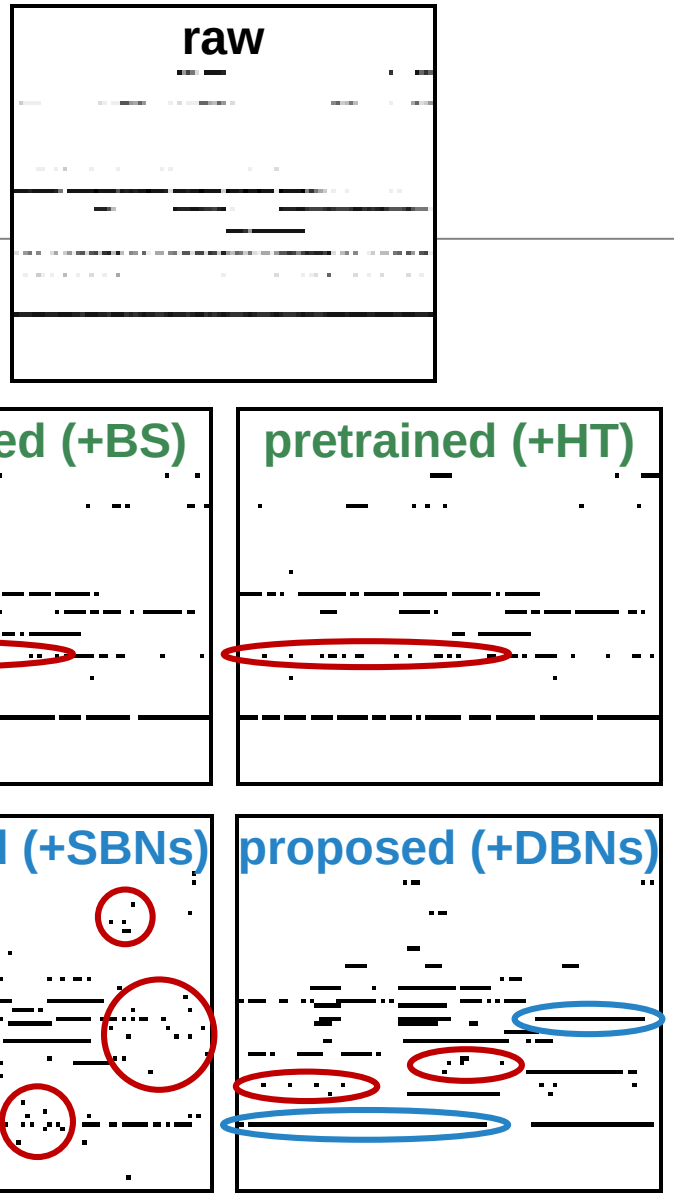


- use **binary neurons** at the output layer of the generator
- use **straight-through estimator** to estimate the gradients for the binary neurons (which involves nondifferentiable operation)

	Generator's outputs	Real data
MuseGAN	real-valued	binary-valued
BinaryMuseGAN	binary-valued	binary-valued

Hao-Wen Dong and Yi-Hsuan Yang, "Convolutional Generative Adversarial Networks with Binary Neurons for Polyphonic Music Generation," to appear at ISMIR, 2018.

Qualitative Comparison



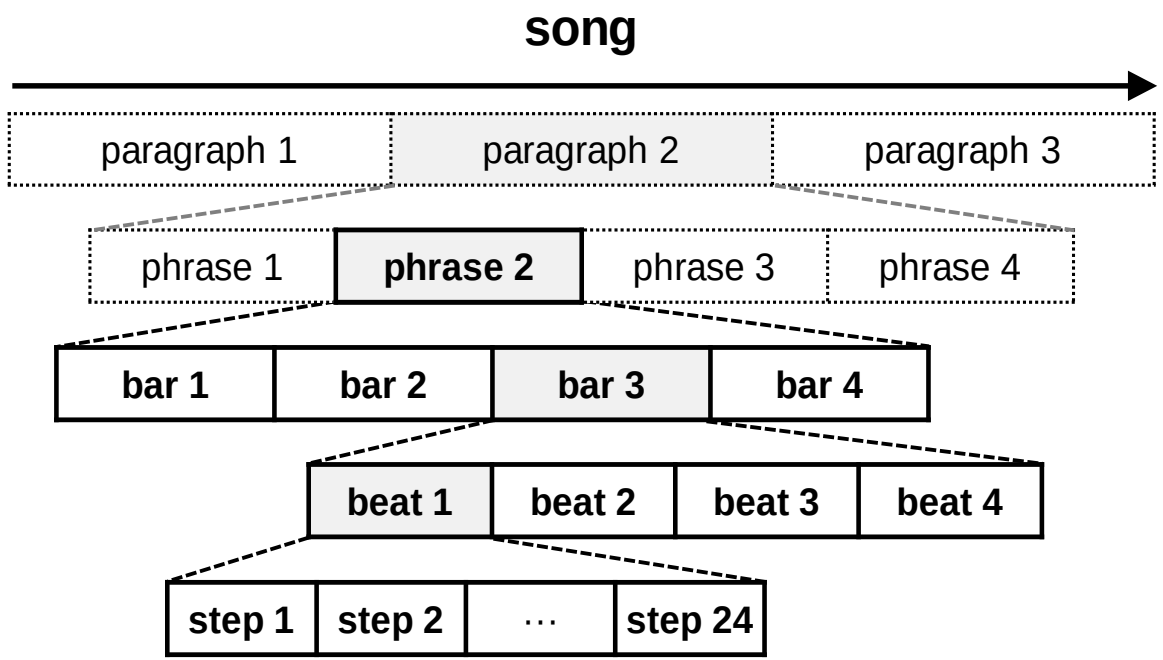
Future Works

Future Works

Full Song Generation

Challenges

- hierarchical temporal structure
- variable-length sequence generation



Future Works

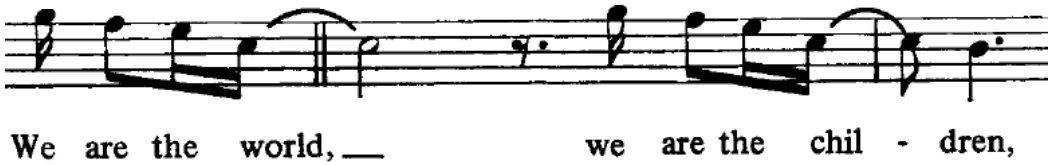
Cross-modal Generation

Challenge

- cross-modal temporal interdependency

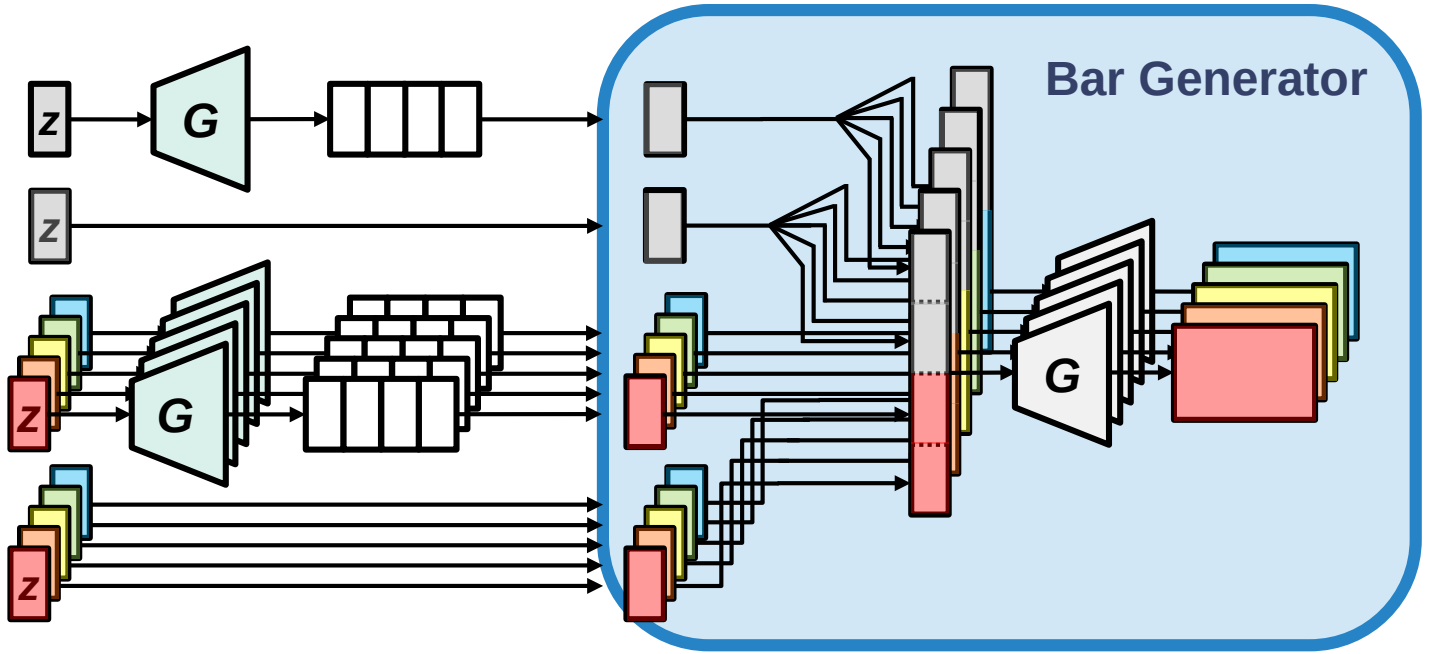
Applications in Music

- music + lyrics



- music + video

Source Code https://github.com/salu133445/musegan
Demo Page https://salu133445.github.io/musegan/



MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment
Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, Yi-Hsuan Yang

Convolutional Generative Adversarial Networks with Binary Neurons for Polyphonic Music Generation
Hao-Wen Dong and Yi-Hsuan Yang

Q&A