Learning to Fuse Music Genres with Generative Adversarial Dual Learning

Zhiqian Chen, Chih-Wei Wu, Yen-Cheng Lu, Alexander Lerch, Chang-Tien Lu

Department of Computer Science, Virginia Tech
Center for Music Technology, Georgia Institute of Technology

Paper

Presentation Agenda

1

Motivation

How to be creative?

Examples

2

Method

GAN extension

Domain fusion

3

Evaluation

Quantitative study

Listening test

4

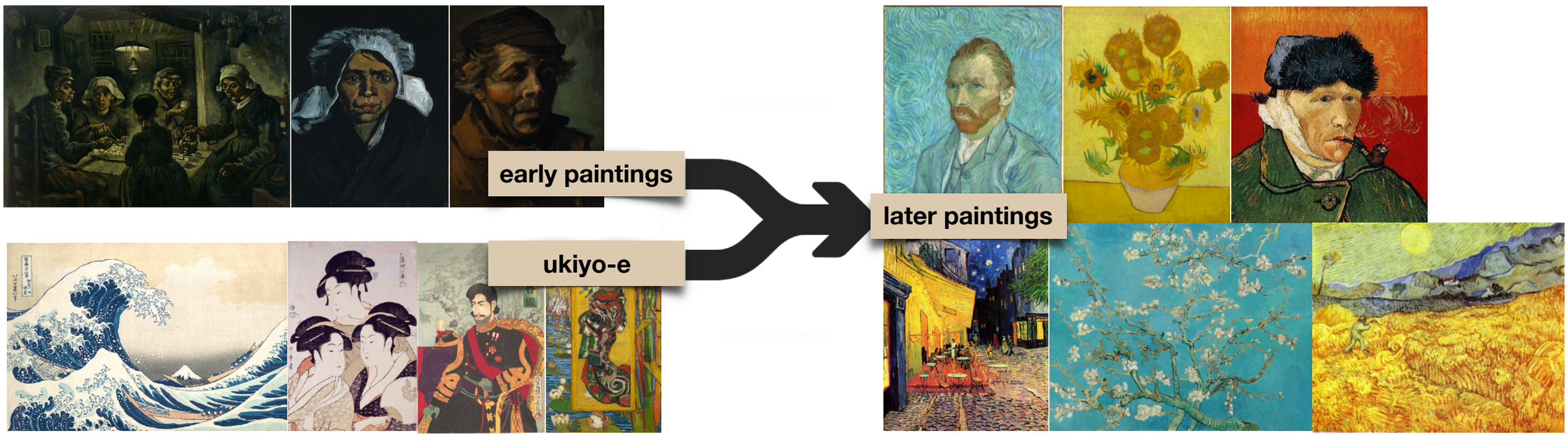
Demo

Q & A

Motivation

How to be creative?

- **Fuse ideas**
  Combining things that don't normally go together.
- **Art workers**
  Drawing a picture by imitation w/ fusion
- **Example**
  Vincent Van Gogh imitating Japanese art (ukiyo-e)

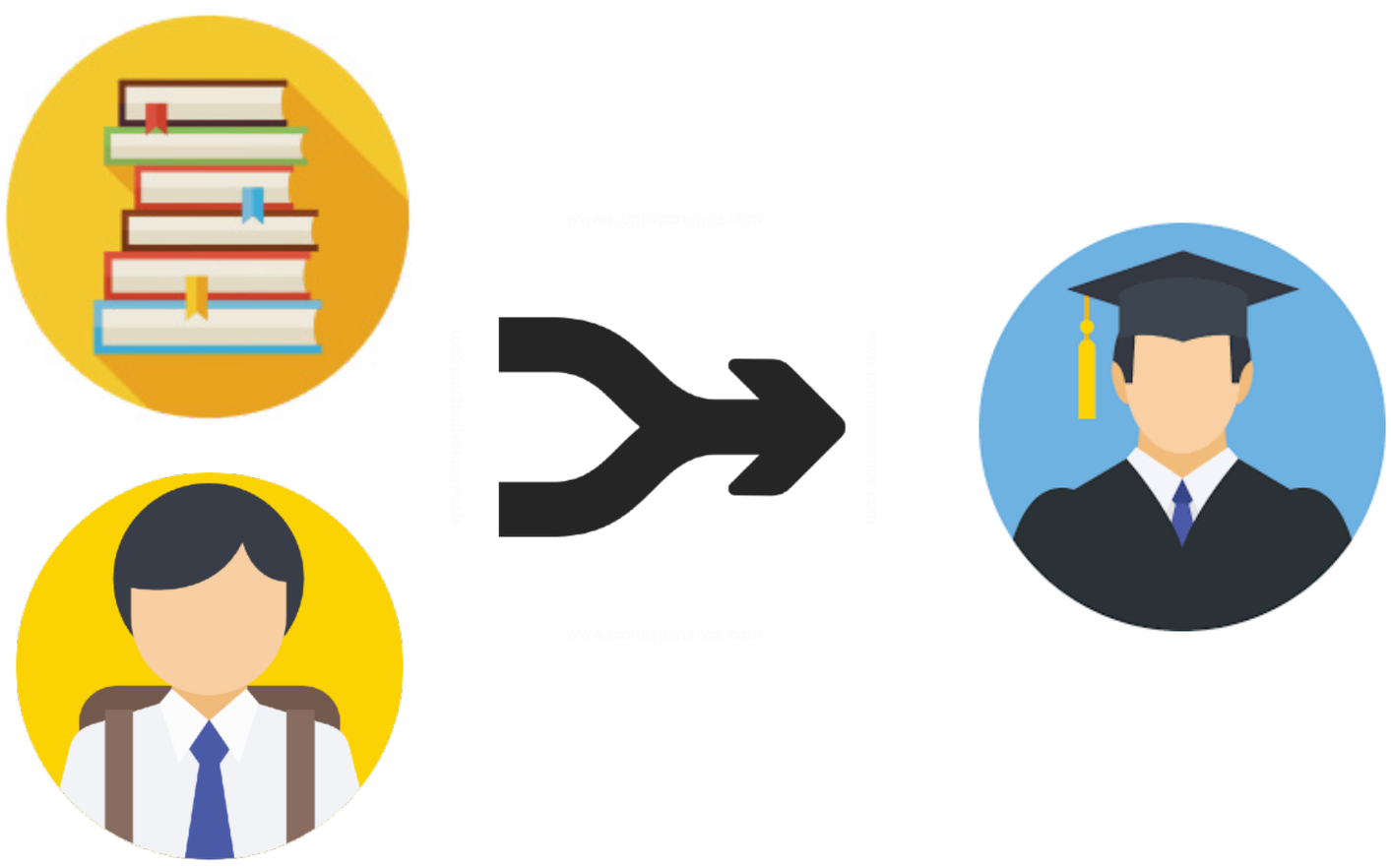


How to be creative?

- **Fuse ideas**
  Combining things that don't normally go together.

- **Researcher**
  Writing a paper by imitation w/ fusion

- **Ph.D. student**
  Get an idea after reading a lot of papers



Motivation

Music generation problem

Problem

Create new music after listening to music of different genres.

Challenge

How to combine the ideas of listened music in **an unsupervised fashion**?

Related work

Domain transferring, rather than combining

Solution

Generative adversarial networks (GAN) extension

Method

Fusion GAN

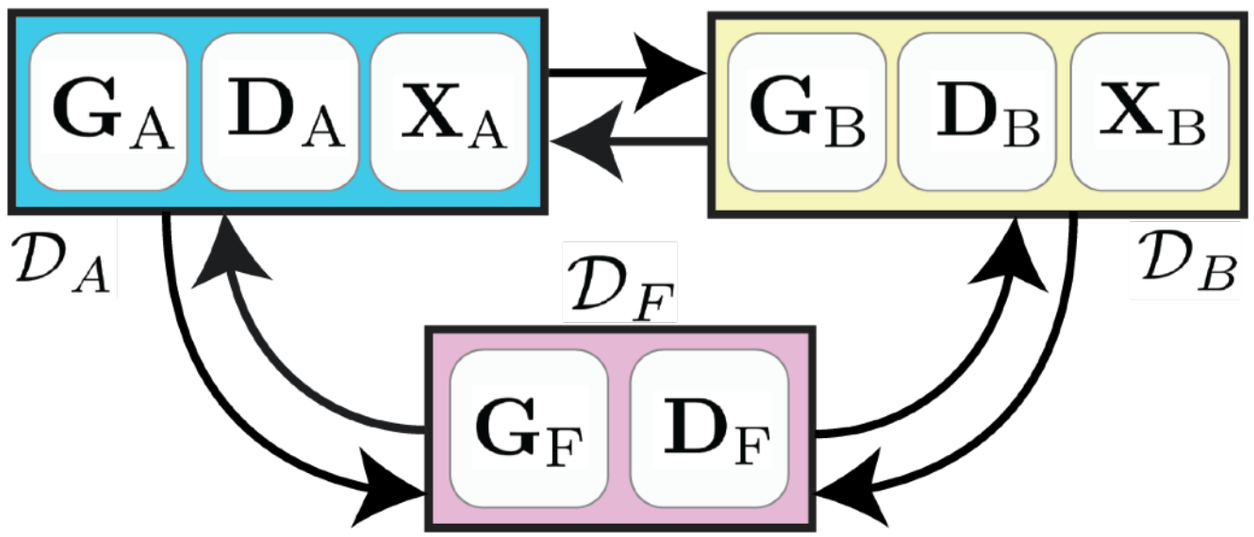
Problem settings

Three domains:

- Domain A/B: given music
- Domain F: generated music

Components of GAN

- G/D:Generator / Discriminator
- X_a/X_b: Data



Task

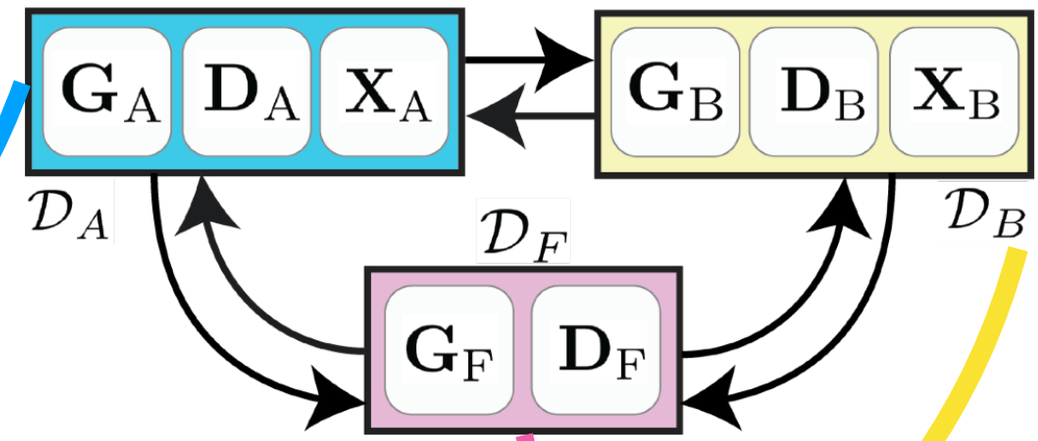
Learn a distribution from X_a/X_b

Balanced distribution between A and B

Method: Fusion GAN

Global loss function

From the perspectives of each D



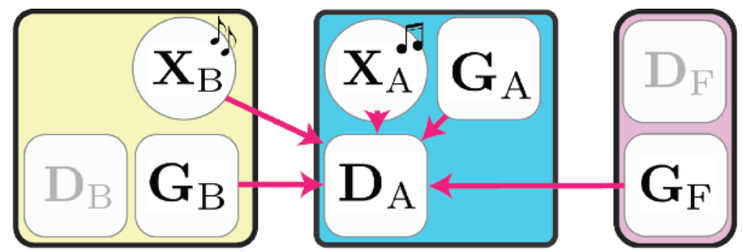
$$
\begin{aligned}
\mathcal{L} = \quad & W\left(\mathbb{P}_{\mathbf{X}_\mathrm{A}}, \mathbb{P}_{\mathbf{X}_\mathrm{B}}, \mathbb{P}_\theta\right) \\
= \quad & \mathbb{E}_{x\sim\mathbb{P}_{\mathbf{X}_\mathrm{A}}}\left[\mathbf{D}_\mathrm{A}\left(x\right)\right] - \mathbb{E}_{x\sim\mathbb{P}_{\mathbf{X}_\mathrm{B}}}\left[\mathbf{D}_\mathrm{A}\left(x\right)\right] - \\
& \mathbb{E}_{z\sim p(z)}\left[\mathbf{D}_\mathrm{A}\left(\mathbf{G}_\mathrm{A}\left(z\right)\right)\right] - \mathbb{E}_{z\sim p(z)}\left[\mathbf{D}_\mathrm{A}\left(\mathbf{G}_\mathrm{B}\left(z\right)\right)\right] - \\
& \mathbb{E}_{z\sim p(z)}\left[\mathbf{D}_\mathrm{A}\left(\mathbf{G}_\mathrm{F}\left(z\right)\right)\right] - \\
& \mathbb{E}_{x\sim\mathbb{P}_{\mathbf{X}_\mathrm{A}}}\left[\mathbf{D}_\mathrm{B}\left(x\right)\right] + \mathbb{E}_{x\sim\mathbb{P}_{\mathbf{X}_\mathrm{B}}}\left[\mathbf{D}_\mathrm{B}\left(x\right)\right] - \\
& \mathbb{E}_{z\sim p(z)}\left[\mathbf{D}_\mathrm{B}\left(\mathbf{G}_\mathrm{A}\left(z\right)\right)\right] - \mathbb{E}_{z\sim p(z)}\left[\mathbf{D}_\mathrm{B}\left(\mathbf{G}_\mathrm{B}\left(z\right)\right)\right] - \\
& \mathbb{E}_{z\sim p(z)}\left[\mathbf{D}_\mathrm{B}\left(\mathbf{G}_\mathrm{F}\left(z\right)\right)\right] + \\
& \mathbb{E}_{x\sim\mathbb{P}_{\mathbf{X}_\mathrm{A}}}\left[\mathbf{D}_\mathrm{F}\left(x\right)\right] + \mathbb{E}_{x\sim\mathbb{P}_{\mathbf{X}_\mathrm{B}}}\left[\mathbf{D}_\mathrm{F}\left(x\right)\right] + \\
& \mathbb{E}_{z\sim p(z)}\left[\mathbf{D}_\mathrm{F}\left(\mathbf{G}_\mathrm{A}\left(z\right)\right)\right] + \mathbb{E}_{z\sim p(z)}\left[\mathbf{D}_\mathrm{F}\left(\mathbf{G}_\mathrm{B}\left(z\right)\right)\right] - \\
& \mathbb{E}_{z\sim p(z)}\left[\mathbf{D}_\mathrm{F}\left(\mathbf{G}_\mathrm{F}\left(z\right)\right)\right],
\end{aligned}
$$

Method: Fusion GAN

- **Balance constraint: perfect mix is half-half**

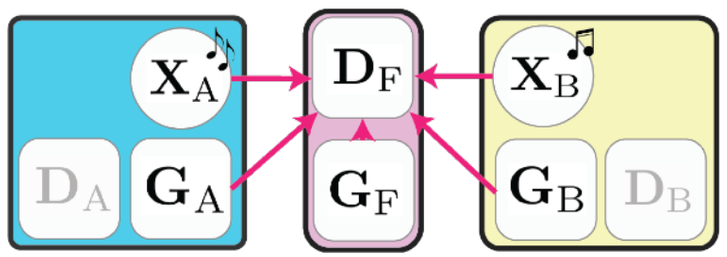
- Domain A $\mathbf{D}_{\mathrm{A}}\left(\mathbf{X}_{\mathrm{A}}\right) \geq \mathbf{D}_{\mathrm{A}}\left(\mathbf{G}_{\mathrm{F}}\left(z\right)\right) \geq \mathbf{D}_{\mathrm{A}}\left(\mathbf{X}_{\mathrm{B}}\right)$

$$\mathcal{L}_{A-bal} = \left\| \mathbf{D}_{\mathrm{A}}\left(\mathbf{X}_{\mathrm{A}}\right) - \mathbf{D}_{\mathrm{A}}\left(\mathbf{G}_{\mathrm{F}}\left(z\right)\right) \right\| + \left\| \mathbf{D}_{\mathrm{A}}\left(\mathbf{G}_{\mathrm{F}}\left(z\right)\right) - \mathbf{D}_{\mathrm{A}}\left(\mathbf{X}_{\mathrm{B}}\right) \right\|$$



- Domain F： keep same distance from A and B

$$\mathcal{L}_{F-bal} = \left\| \mathbb{E}_{z\sim p(z)}\left[\mathbf{D}_{\mathrm{F}}\left(\mathbf{G}_{\mathrm{A}}\left(z\right)\right)\right] - \mathbb{E}_{z\sim p(z)}\left[\mathbf{D}_{\mathrm{F}}\left(\mathbf{G}_{\mathrm{B}}\left(z\right)\right)\right] \right\| + \left\| \mathbb{E}_{x\sim \mathbb{P}_{\mathbf{X}_{\mathrm{A}}}}\left[\mathbf{D}_{\mathrm{F}}\left(x\right)\right] - \mathbb{E}_{x\sim \mathbb{P}_{\mathbf{X}_{\mathrm{B}}}}\left[\mathbf{D}_{\mathrm{F}}\left(x\right)\right] \right\|.$$



- FusionGAN Optimality Theorem

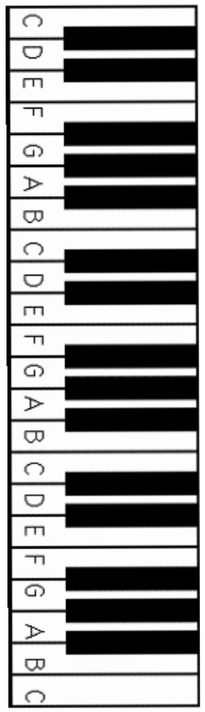
  - convergence guarantee

Evaluation

Quantitative Study

Music Genres Intrinsic property

Define intrinsic properties of music genre



Normalized Pitch Distribution (NPD)

88 keys into C/C#/D/D#/ E/F/F#/G/G#/A/A#/B

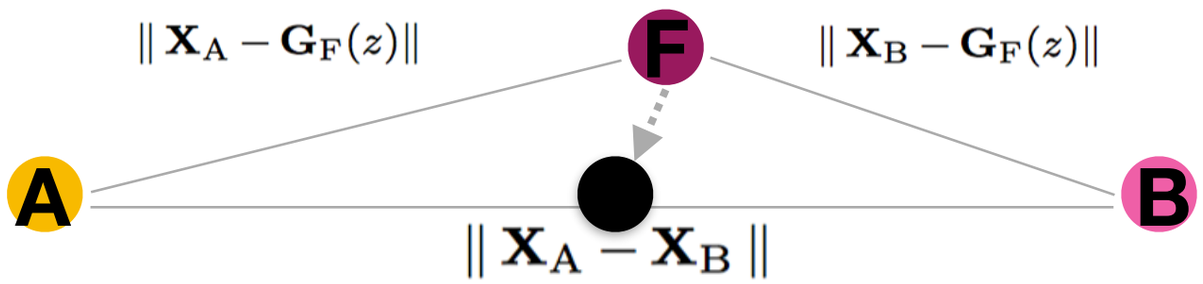
American / German note names	British note names	Note symbols	Note value
Whole note	**Semibreve**	𝅝	4 beats
Half note	**Minim**	𝅗𝅥	2 beats
Quarter note	**Crotchet**	♩	1 beat
Eighth note	**Quaver**	♪	1/2 of a beat
Sixteenth note	**Semiquaver**	𝅘𝅥𝅯	1/4 of a beat

Duration Distribution (DD)

Each note has a duration

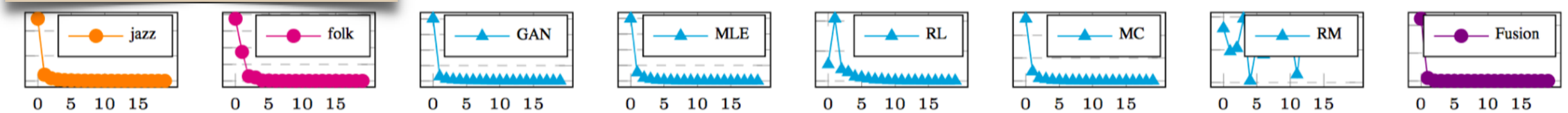
Evaluation

Quantitative Study

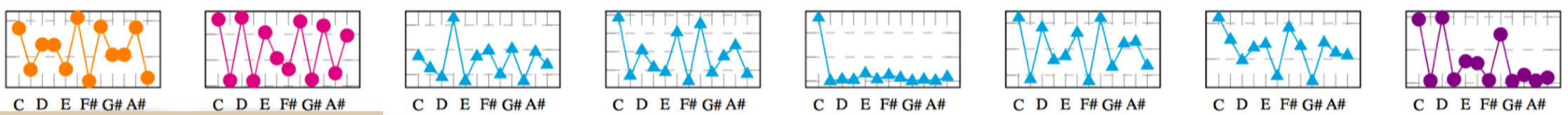


	EUD		EM	
	Diff	Ratio	Diff	Ratio
		DD		
RM	39742.2	1.375	2757.6	1.461
MLE	24546.4	1.231	2005.2	1.335
GAN	24765.2	1.233	2064.3	1.345
RL	37971.2	1.358	2629.0	1.439
MC	13988.7	1.132	1289.2	1.215
Fusion	19452.6	1.183	1831.9	1.306
		NPD		
RM	19586.6	1.647	5231.0	1.643
MLE	15921.4	1.526	4147.5	1.510
GAN	16807.1	1.555	4098.0	1.504
RL	16927.1	1.559	4399.2	1.541
MC	11175.0	1.369	2660.2	1.327
Fusion	11564.6	1.382	3182.5	1.391

Duration Distribution (DD)



Normalized Pitch Distribution



Evaluation

Listening test via user survey

amazon mechanical turk

Fusion Level

How much it is thought as fusion music

Given fused music, choose better fusion

(A) pure jazz

(B) pure folk

(C) mixture of jazz and folk

(D) neither

Rate the level of its composer

(A) expert

(B) newbie

(C) robot

Fusion recognition

	jazz	*folk*	*mixture*	*neither*	**FL**
RM	25.0%	22.5%	12.5%	40%	57.5%
MLE	43.6%	9.1%	30.9%	16.4%	49.1%
GAN	34.0%	17.0%	26.0%	14%	69%
RL	20.1%	28.3%	20.8%	30.8%	61%
MC	32.0%	2.0%	14.0%	52%	16%
Fusion	35.9%	25.0%	20.0%	19.1%	70%

Musicality

	expert	*newbie*	*robot*
RM	22.5%	50.0%	27.5%
MLE	45.5%	21.8%	32.7%
GAN	42.0%	32.0%	26.0%
RL	32.1%	37.7%	30.2%
MC	30.0%	36.0%	34.0%
Fusion	43.8%	28.1%	28.1%

Demo

Thank you!

Q&A

Code has been released!

https://github.com/aquastar/fusion_gan

fusion gan