SIGGRAPH 2021

ChoreoMaster : Choreography-Oriented Music-Driven Dance Synthesis

Keywords: Style Transfer, Deep Learning

Paper:

https://dl.acm.org/doi/abs/10.1145/3450626.3459932

https://netease-gameai.github.io/ChoreoMaster/Paper.pdf

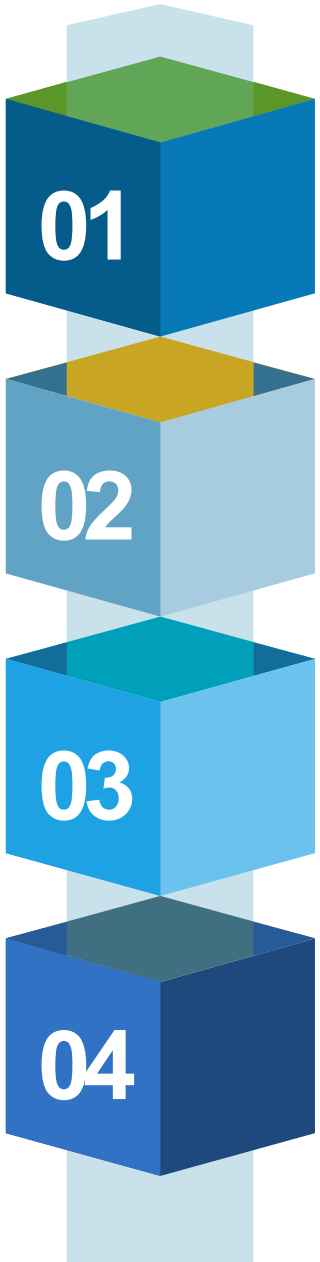
Introduction:

https://blog.siggraph.org/2021/09/how-choreomaster-combines-cutting-edge-ai-and-graphics-technologies.html/

https://netease-gameai.github.io/ChoreoMaster/

IPHD YuanFu Yang

Outline



Background

Art Statement/Method

Result

Connection

Background -

网易互娱 AI Lab

RL Team in Netease Fuxi AI Lab
Reinforcement Learning Team in Netease Fuxi AI Lab
Hangzhou, China https://fuxi.163.com/en/about.html FuxiRL@163.com

網易互娛AI Lab的Athena AI在國際強化學習頂級賽中奪冠

2021-12-16 由 万俟傲白 釋出於 科技

近日，在NeurIPS會議上舉辦的MineRL 2021 Diamond Competition落下帷幕，來自網易互娛AI Lab的Athena AI憑藉高超的挖鑽技巧，在以《我的世界》遊戲為競技環境的比賽中拿下Intro賽道的冠軍以及Research賽道的亞軍。這是AI第一次在《我的世界》中挖掘到鑽石。

據悉，該比賽由CMU、OpenAI、DeepMind、Microsoft Research等機構聯合舉辦，是強化學習方向最負盛名的比賽之一。比賽自2019年起，每年在機器學習和計算神經科學領域頂級學術會議NeurIPS上舉辦，今年為第三屆。近年來該比賽吸引了包括騰訊AI Lab以及清華、北大、斯坦福大學等在內的眾多工業界和學術界的相關研究人員。

網易互娛AI Lab奪得ICCV 2021人體重識別競賽冠軍

2021-10-20 13:09:27 來源:品玩 舉報

O

分享至

品玩10月20日訊，近日，由計算機視覺領域的頂級會議ICCV 2021（International Conference on Computer Vision）舉辦的VIPriors挑戰賽成績正式揭曉。網易互娛AI Lab從螞蟻集團、美團、加州大學伯克利分校、復旦大學等來自工業界和學術界的強隊中脫穎而出，一舉斬獲人體重識別（Person Re-ID）競賽的冠軍。這是網易互娛AI Lab在3D視覺、語音、自然語言處理、遊戲AI等領域奪得多項國際冠軍後，再次登頂國際AI競賽，彰顯了網易互娛AI Lab在人工智能領域的綜合實力。

網易互娛AI Lab奪取CVPR2021 3D人臉重建挑戰賽冠軍

科技網 · 2021/06/30 11:30

近日，第三十四屆國際計算機視覺頂級會議CVPR 2021 (Conference on Compute
舉辦的多角度圖片3D人臉重建競賽公佈了最終成績，網易互娛AI Lab提出的多視角
一，擊敗了來自OPPO研究院，虎牙AI Lab，清華大學，北京大學，中科院自動化所
隊，一舉斬獲大賽冠軍。

近日，計算機視覺頂級會議CVPR 2021舉辦的PIC Challenge AI挑戰賽成績出爐。
穎而出，獲得了多視角圖片3D 人臉重建競賽的冠軍。這也是網易互娛AI Lab繼IJCAI

AAAI 2021最「嚴」一屆放榜:錄取率僅21%,網易伏羲9篇論文入選

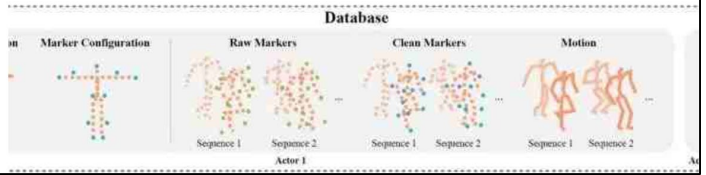
2020-12-10 中國青年網

近日，國際人工智慧頂級會議AAAI 2021公布論文錄取結果。網易伏羲實驗室再創佳績，共有9篇論文入選，研究方向涉及強化學習、虛擬人、自然語言處理(NLP)、圖像動畫、用戶畫像等領域。科研成果的集中爆發，充分顯示網易伏羲在人工智慧的多個領域已經具備國際頂尖的技術創新能力。

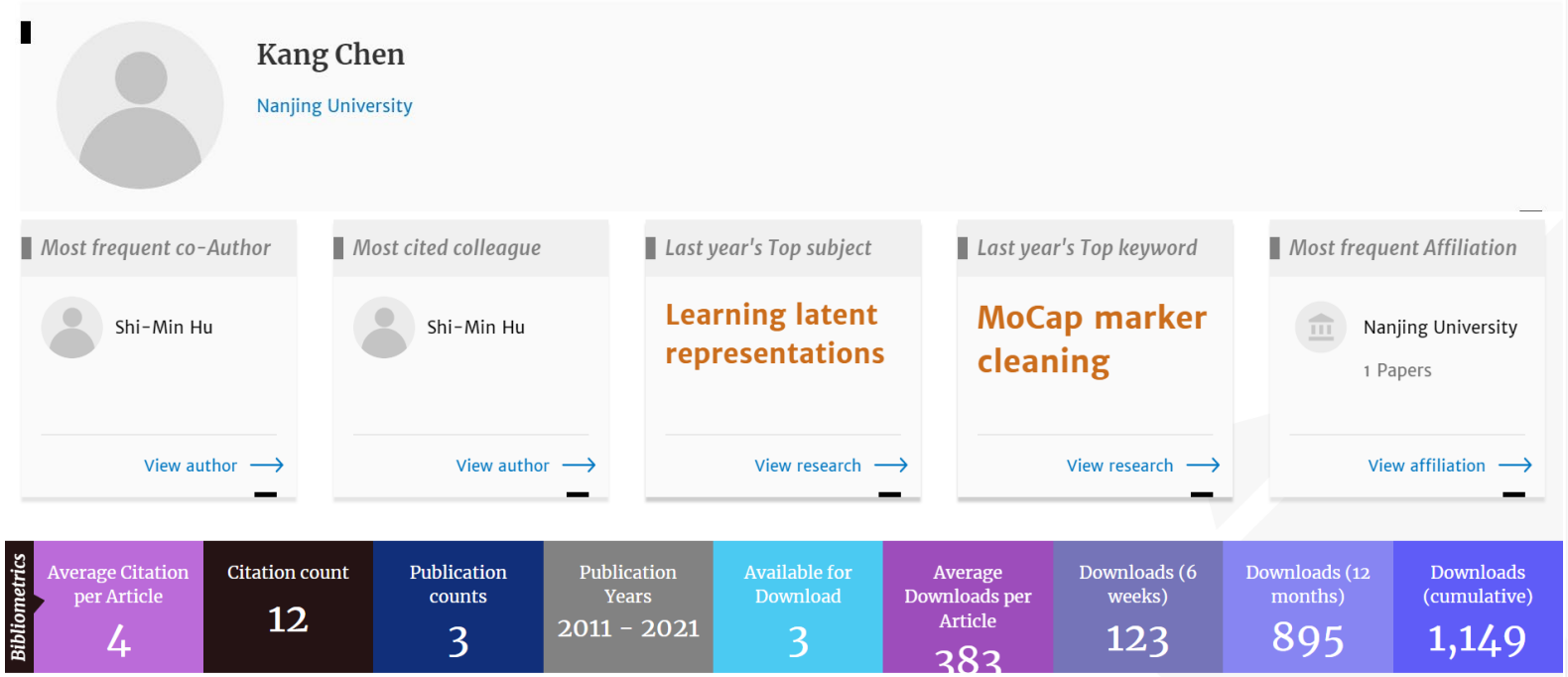
AAAI(Association for the Advance of Artificial Intelligence)是美國人工智慧協會主辦的年會，是人工智慧領域中歷史最悠久、涵蓋內容最廣泛的的國際頂級學術會議之一。在中國計算機學會的國際學術會議排名中，AAAI被列爲人工智慧領域的A類頂級會議。

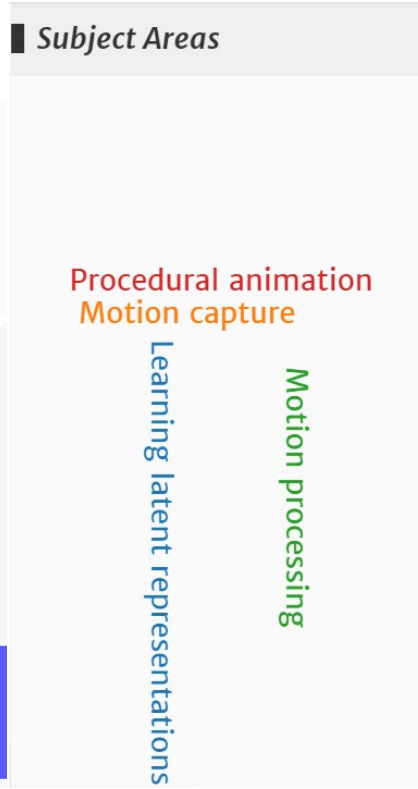
AI 賦能遊戲工業化，網易互娛AI Lab動捕去噪新方法入選SIGGRAPH 2021

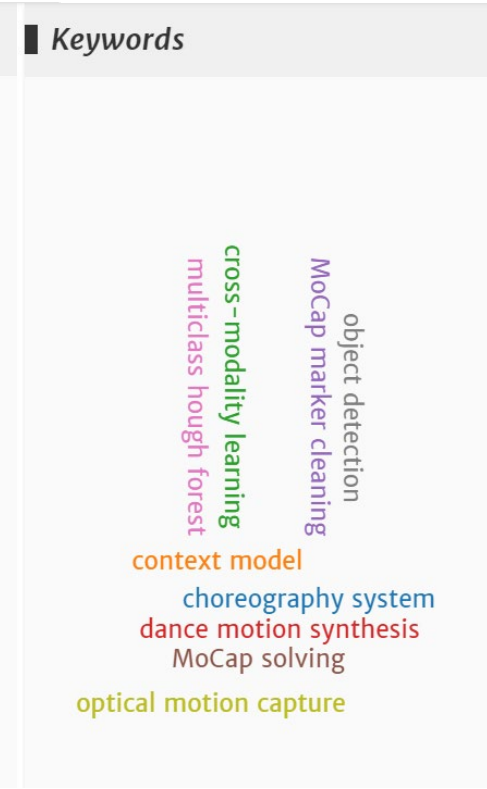
AI科技評論 微信號：aitechtalk 發表2021年08月09日



Background - Author







Author's Latest Publications

RESEARCH-ARTICLE

ChoreoMaster: choreography-oriented music-driven dance synthesis

Kang Chen, Zhipeng Tan, Jin Lei, Song-Hai Zhang, Yuan-Chen Guo, + 2

July 2021 • ACM Transactions on Graphics, Volume 40, Issue 4 • https://doi.org/10.1145/3450626.3459932

RESEARCH-ARTICLE

MoCap-solver: a neural solver for optical motion capture data

Kang Chen, Yupan Wang, Song-Hai Zhang, Sen-Zhe Xu, + 2

July 2021 • ACM Transactions on Graphics, Volume 40, Issue 4 • https://doi.org/10.1145/3450626.3459681

SHORT-PAPER

Multiclass object detection by combining local appearances and context

LiMin Wang, Yirui Wu, Tong Lu, Kang Chen

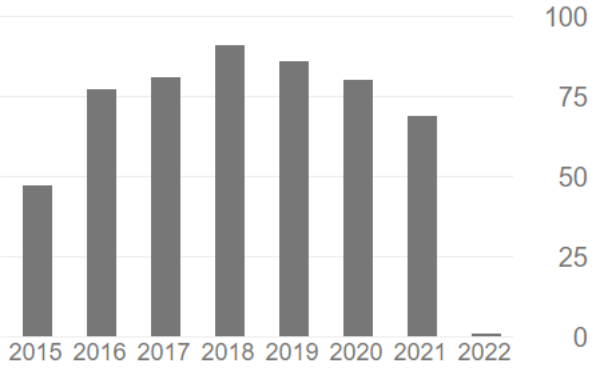
November 2011 • MM '11: Proceedings of the 19th ACM international conference on Multimedia • https://doi.org/10.1145/2072298.2071964

Background - Author

标题	引用次数	年份
Sketch2Scene: Sketch-based co-retrieval and co-placement of 3D models K Xu, K Chen, H Fu, WL Sun, SM Hu ACM Transactions on Graphics (TOG) 32 (4), 1-15	198	2013
Structure recovery by part assembly CH Shen, H Fu, K Chen, SM Hu ACM Transactions on Graphics (TOG) 31 (6), 1-11	137	2012
Automatic semantic modeling of indoor scenes from low-quality RGB-D data using contextual information K Chen, YK Lai, YX Wu, R Martin, SM Hu ACM Transactions on Graphics 33 (6)	103	2014
3D indoor scene modeling from RGB-D data: a survey K Chen, YK Lai, SM Hu Computational Visual Media 1 (4), 267-278	77	2015
Magic decorator: automatic material suggestion for indoor digital scenes K Chen, K Xu, Y Yu, TY Wang, SM Hu ACM Transactions on graphics (TOG) 34 (6), 1-11	48	2015
Multiclass object detection by combining local appearances and context LM Wang, Y Wu, T Lu, K Chen Proceedings of the 19th ACM international conference on Multimedia, 1161-1164	17	2011
View suggestion for interactive segmentation of indoor scenes S Yang, J Xu, K Chen, H Fu Computational Visual Media 3 (2), 131-146	9	2017
Choreomaster: choreography-oriented music-driven dance synthesis K Chen, Z Tan, J Lei, SH Zhang, YC Guo, W Zhang, SM Hu ACM Transactions on Graphics (TOG) 40 (4), 1-13	4	2021

引用次数 查看全部

	总计	2017 年至今
引用	596	409
h 指数	7	7
i10 指数	6	5



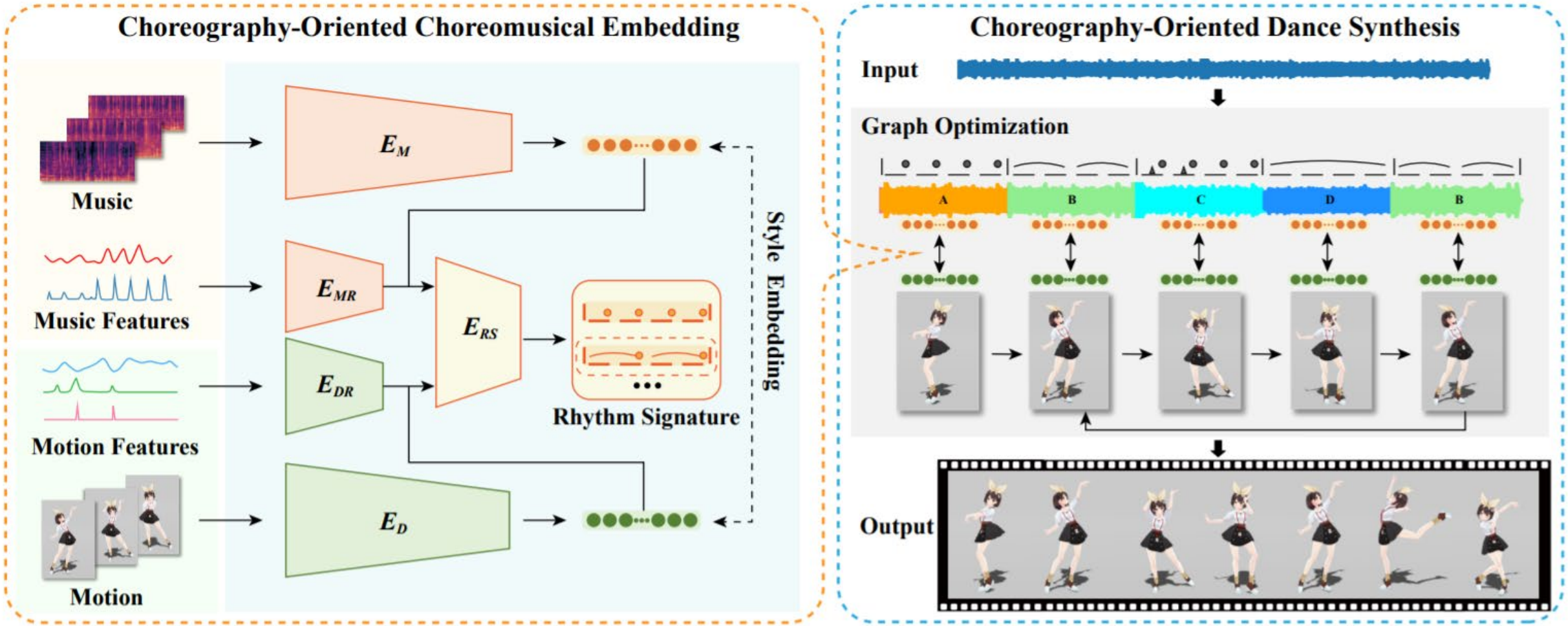
Art Statement

- The style of music and body movements should be consistent, conveying similar mood and tone,
- Each synchronized dance and music segment should present the same rhythmic pattern, while rhythmic patterns in dance phrases appear with great regularity,
- The organization of a dance should be coordinated with the structure of the corresponding music, e.g., repeated musical phrases (verse and chorus) are typically associated with re-peated movements, while identical meters in a phrase often correspond to symmetrical movements.

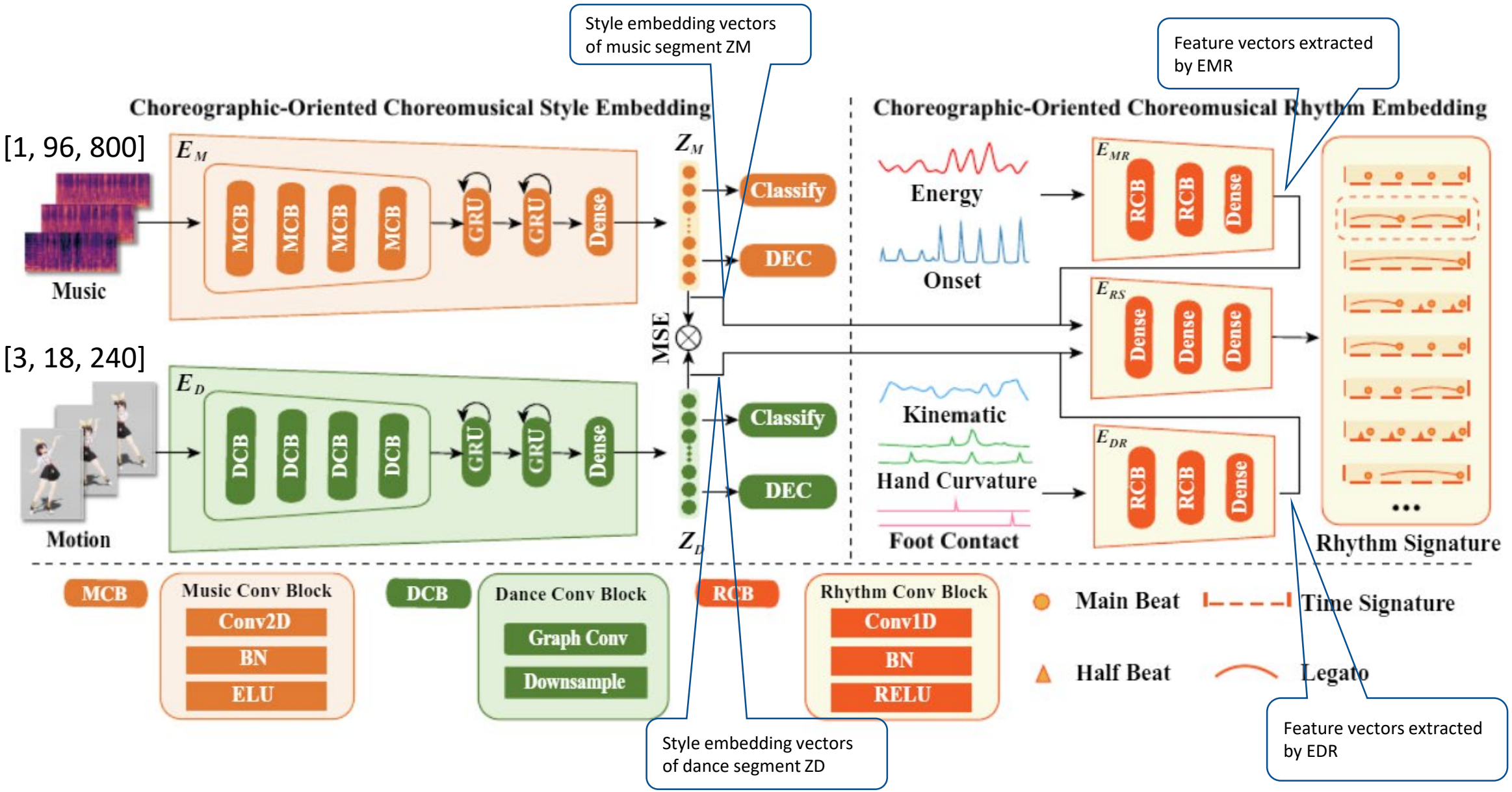
ChoreoMaster: Choreography-Oriented Music-Driven Dance Synthesis

KANG CHEN, NetEase Games AI LAB, China
ZHIPENG TAN, NetEase Games AI LAB, China
JIN LEI, NetEase Games AI LAB, China
SONG-HAI ZHANG*, Tsinghua University, China
YUAN-CHEN GUO, Tsinghua University, China
WEIDONG ZHANG, NetEase Games AI LAB, China
SHI-MIN HU, Tsinghua University, China

Art Statement



Method


Style embedding vectors of music segment ZM
Feature vectors extracted by EMR
Choreographic-Oriented Choreomusical Style Embedding
Choreographic-Oriented Choreomusical Rhythm Embedding
[1, 96, 800]
Music
[3, 18, 240]
Motion
MCB
GRU
Dense
Classify
DEC
MSE
DCB
Energy
Onset
Kinematic
Hand Curvature
Foot Contact
RCB
Rhythm Signature
Music Conv Block
Conv2D
BN
ELU
Dance Conv Block
Graph Conv
Downsample
Rhythm Conv Block
Conv1D
BN
RELU
Main Beat
Time Signature
Half Beat
Legato
Style embedding vectors of dance segment ZD
Feature vectors extracted by EDR

Method

- MFCC (Mel-Frequency Cepstral Coefficients)

Step1: Pre-emphasis

$s2(n) = s(n) - a*s(n-1)$

a: 0~1

Step2: Signal Framing

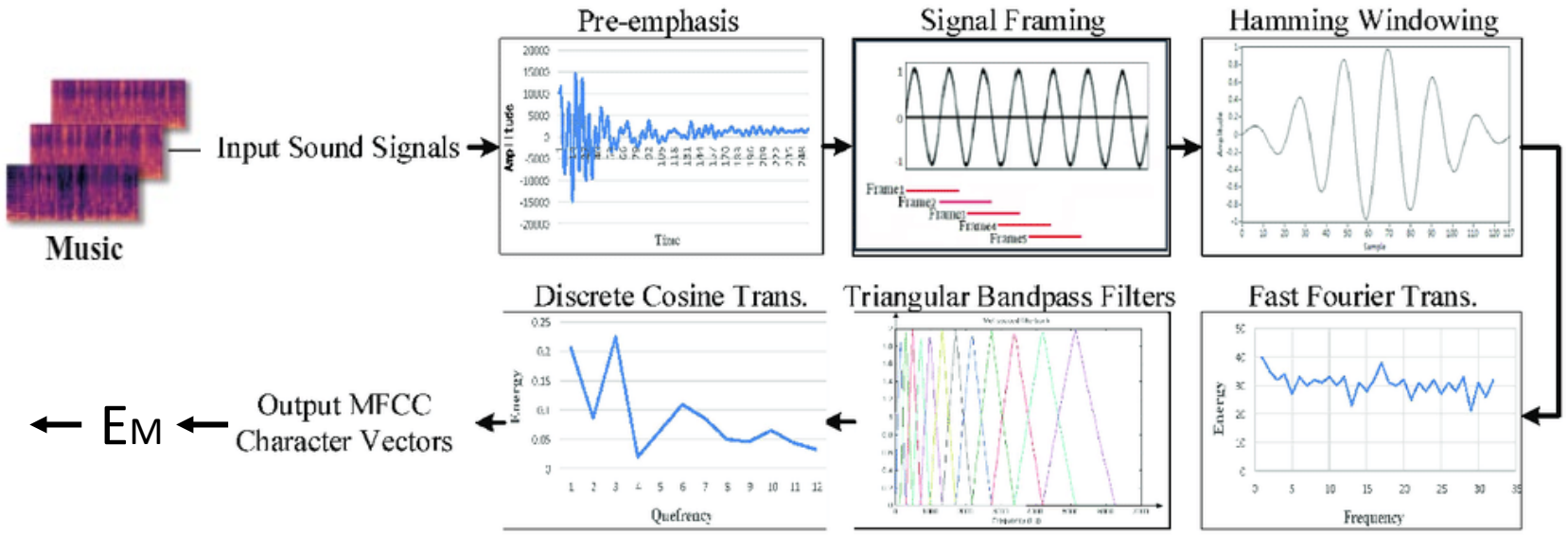
N: 256 or 512 (set of sample)

M: N/3 (set of overlap sample)

Step3: Hamming Windowing

$W(n, a) = (1 - a) - a \cos(2pn/(N-1))$,

$0 \leqq n \leqq N-1$



Step6: Discrete Cosine Trans.

$C_m=\Sigma_{k=1}^{N}\cos[m*(k-0.5)*\pi/N]*E_k$,

m=1,2, ..., L

Step5: Triangular Bandpass Filters

$mel(f)=2595*\log10(1+f/700)$

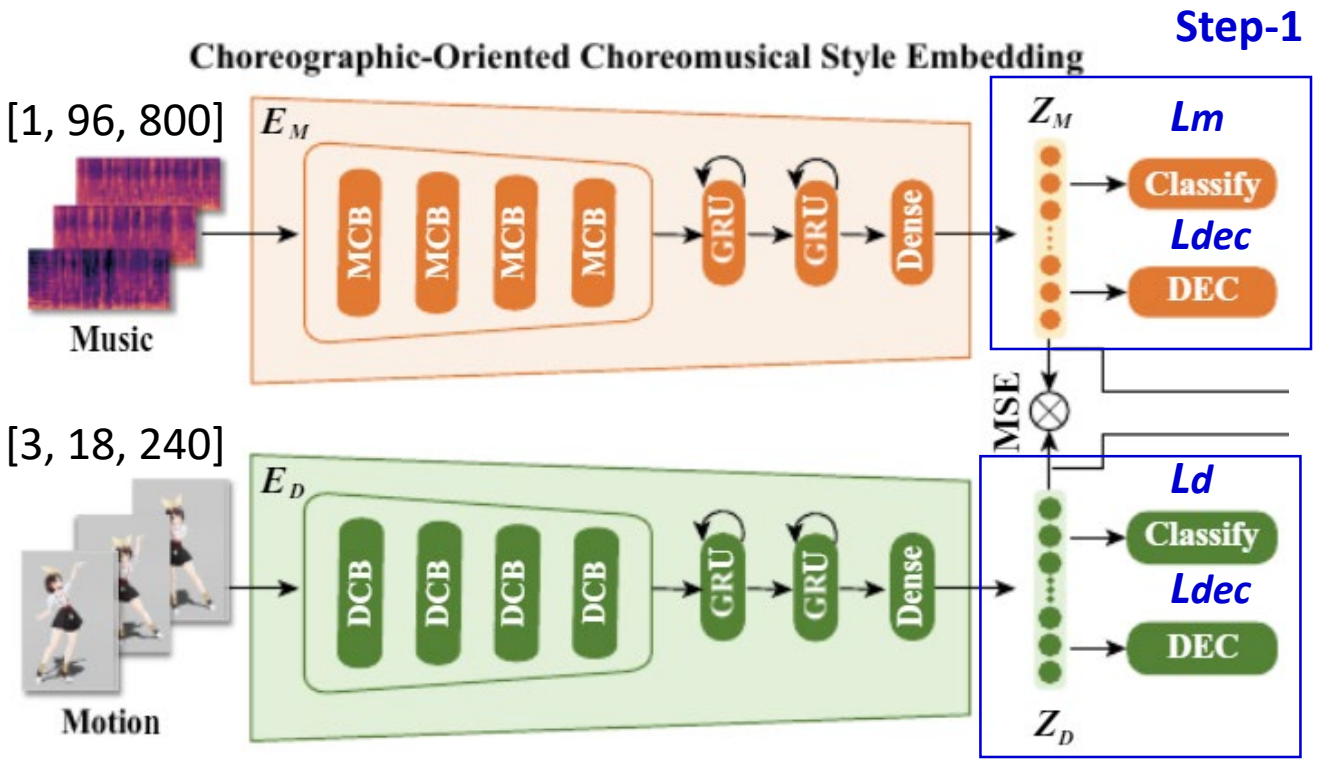
Step4: Fast Fourier Transform

$F(x) = \sum_{n-0}^{N-1} f(n)e^{-i2\pi(x\frac{n}{N})}$

$f(n) = \frac{1}{N}\sum_{n-0}^{N-1} F(x)e^{i2\pi(x\frac{n}{N})}$

Method

- Loss Function



$$\mathcal{L}_m = \lambda_1 L_m + \lambda_2 L_{\text{dec}}$$

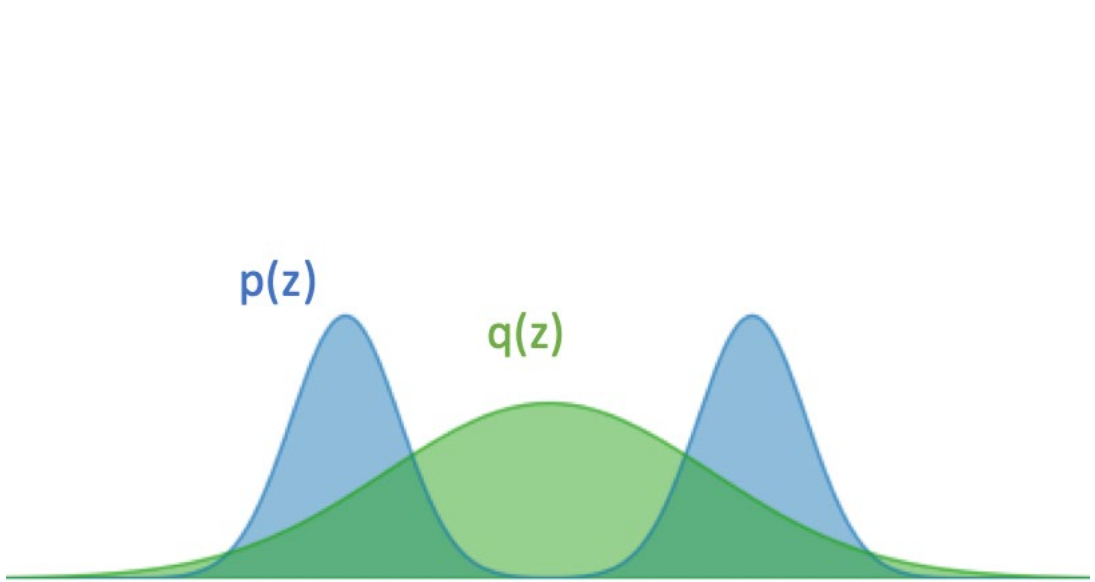
$$\mathcal{L}_d = \lambda_1 L_d + \lambda_2 L_{\text{dec}}$$

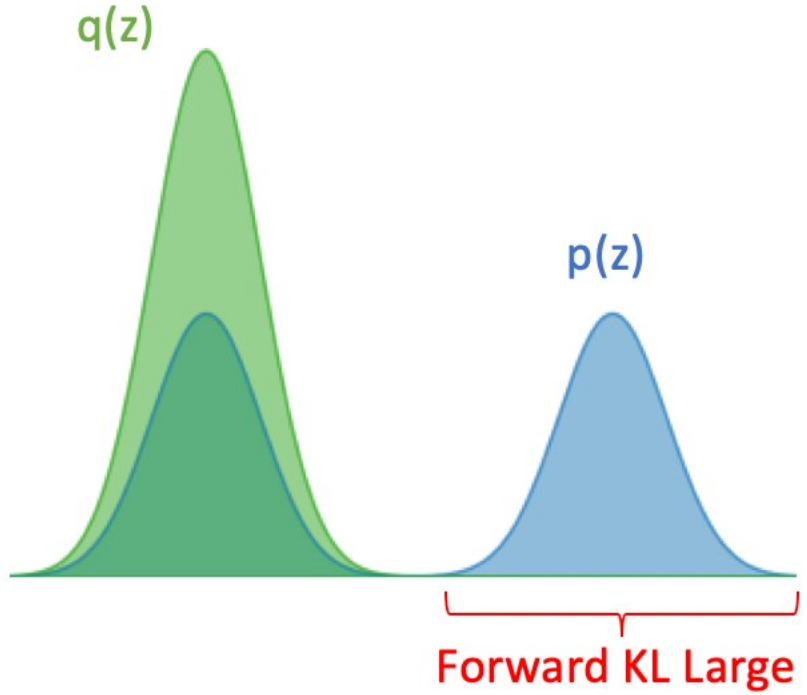
where Lm, Ld are the classification losses (i.e., cross-entropy loss) of music and dance respectively, Ldec is the the KL divergence loss defined in Equation (2) of [Xie et al. 2016], and $\lambda 1$ and $\lambda 2$ are balancing weights.

Method

- Loss Function

$$
\begin{aligned}
\textit{Ldec} &= -E_{x\sim p}[\ln q(x) - \ln p(x)] = -E_{x\sim p}\left[\ln \tfrac{q(x)}{p(x)}\right] \\
&= E_{x\sim p}[-\ln q(x)] - E_{x\sim p}[-\ln p(x)] \\
&= H(p,q) - H(p)
\end{aligned}
$$





Method

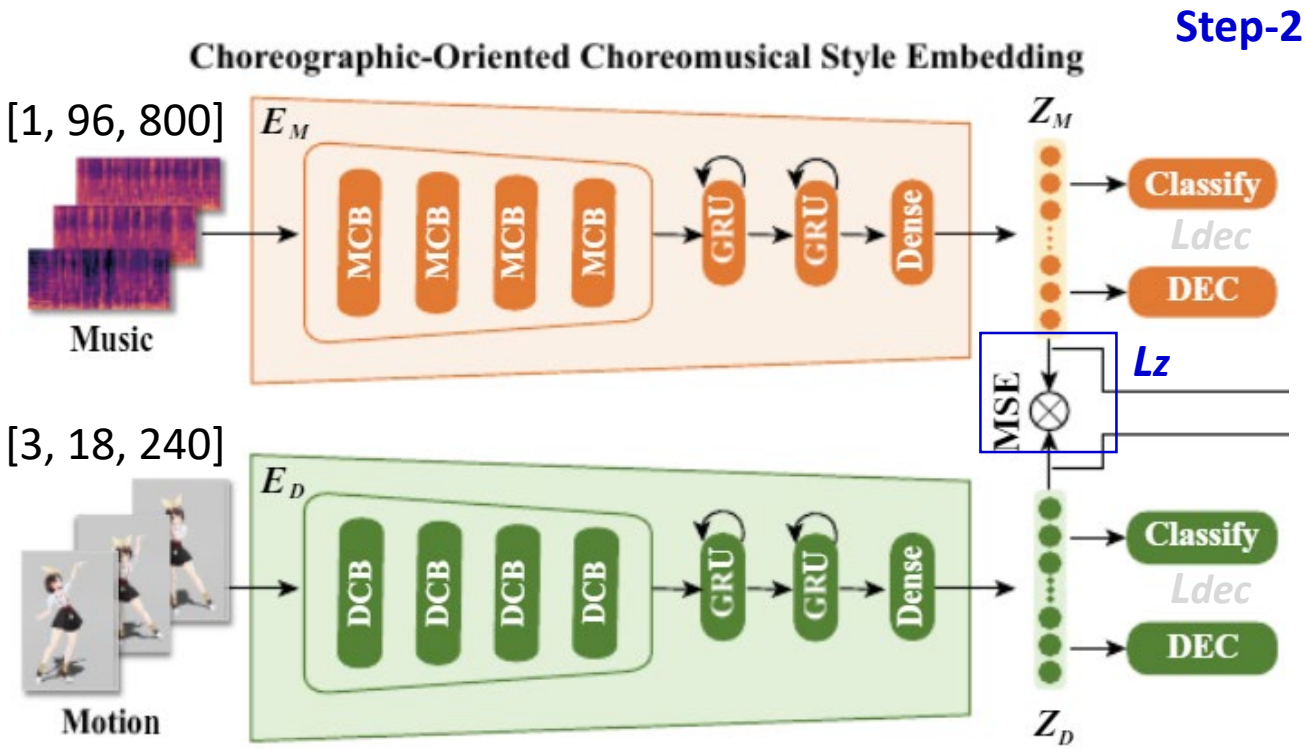
- GridSearch – Hyperparameters turning

$$\mathcal{L}_m = \lambda_1 L_m + \lambda_2 L_{\text{dec}}$$
$$\mathcal{L}_d = \lambda_1 L_d + \lambda_2 L_{\text{dec}}$$

λ_1		1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
λ_2		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Motion	Accuracy 1 (%)	72.5	72.4	72.4	72.4	72.3	71.7	70.0	68.0	65.1	60.9
	Accuracy 2 (%)	67.5	67.4	67.4	67.4	67.3	66.9	64.5	62.9	59.3	53.0
Music	Accuracy 1 (%)	95.2	95.2	95.1	95.1	94.8	94.0	91.2	86.6	82.1	78.9
	Accuracy 2 (%)	78.4	78.3	78.3	78.3	78.2	77.5	74.0	70.3	68.2	64.1

Method

- Loss Function



$$\mathcal{L}_{\text{style}} = \lambda_3 L_d + \lambda_4 L_m + \lambda_5 L_z$$

where Lm and Ld are the classification losses, Lz is the MSE loss between ZM and ZD , and $\lambda 3, \ldots, \lambda 5$ are weights. Through these two phases of training, we can map any music and dance segments into a unified choreomusical embedding space, where style consistency between music and dance can be measured by the Euclidean distance between the corresponding embedding vectors.

Method

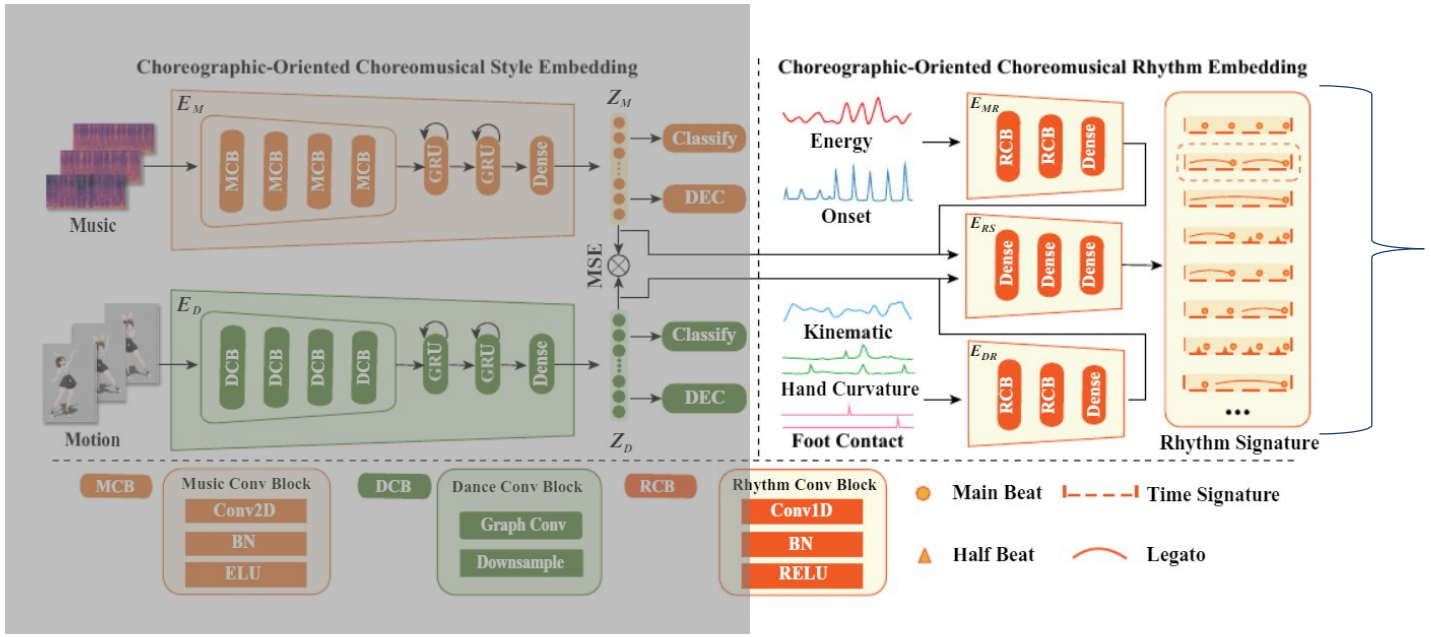
- GridSearch – Hyperparameters turning

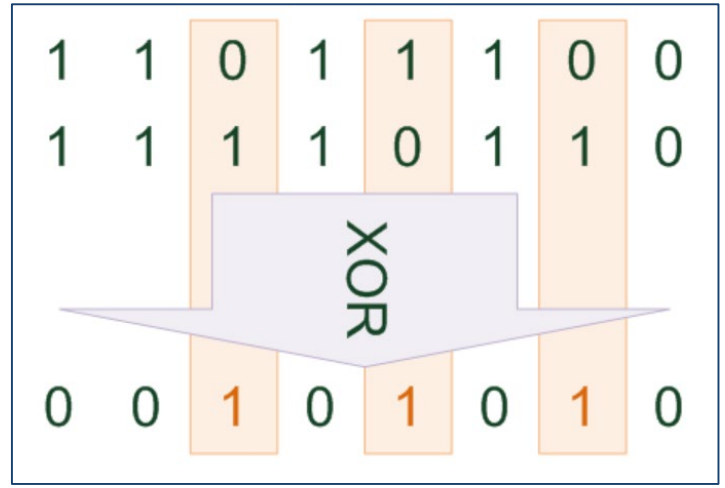
$$\mathcal{L}_{\text{style}} = \lambda_3 L_d + \lambda_4 L_m + \lambda_5 L_z$$

$\lambda_3, \lambda_4 = 0.5$	λ_5	1	2	3	4	5	6	7	8	9
Motion	Accuracy 1 (%)	70.8	70.6	70.7	70.6	70.5	70.2	69.6	69.4	69.1
	Accuracy 2 (%)	66.1	65.9	65.8	65.8	65.7	65.5	64.9	64.2	64.3
Music	Accuracy 1 (%)	91.0	90.9	90.9	90.8	90.8	90.7	90.7	90.3	90.1
	Accuracy 2 (%)	77.2	77.2	77.1	77.2	77.1	77.1	77.0	69.9	69.9
Choreomusical Embedding Distance		0.51	0.43	0.36	0.32	0.32	0.31	0.31	0.31	0.31
$\lambda_5 = 5$	λ_3	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
	λ_4	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Motion	Accuracy 1 (%)	70.6	70.6	70.6	70.5	70.5	70.5	70.3	70.3	70.0
	Accuracy 2 (%)	65.8	65.8	65.7	65.7	65.7	65.7	65.5	65.5	64.3
Music	Accuracy 1 (%)	89.9	89.1	90.3	90.6	90.8	90.8	90.9	90.9	90.9
	Accuracy 2 (%)	76.4	77.0	77.1	77.1	77.1	77.2	77.2	77.2	77.2
Choreomusical Embedding Distance		0.33	0.33	0.32	0.32	0.32	0.32	0.32	0.32	0.33

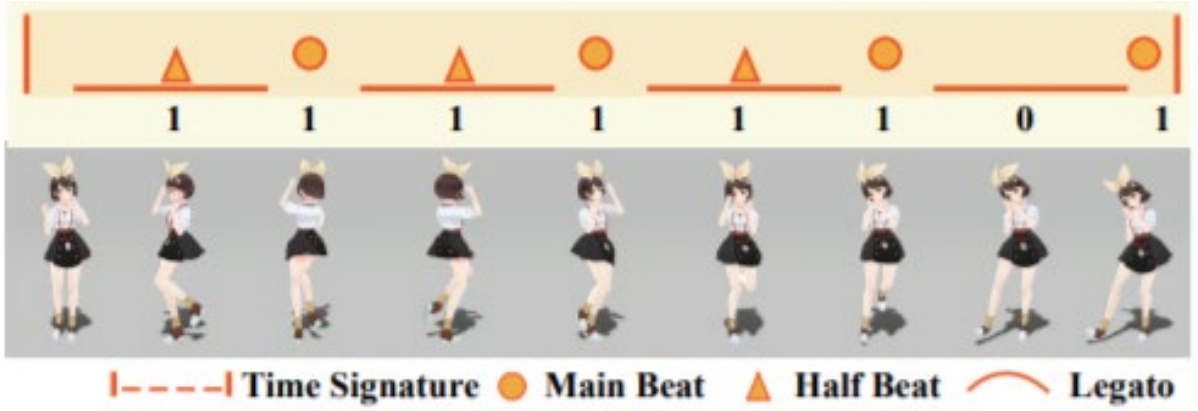
Art Statement

- Rhythm Embedding





Hamming distance = 3

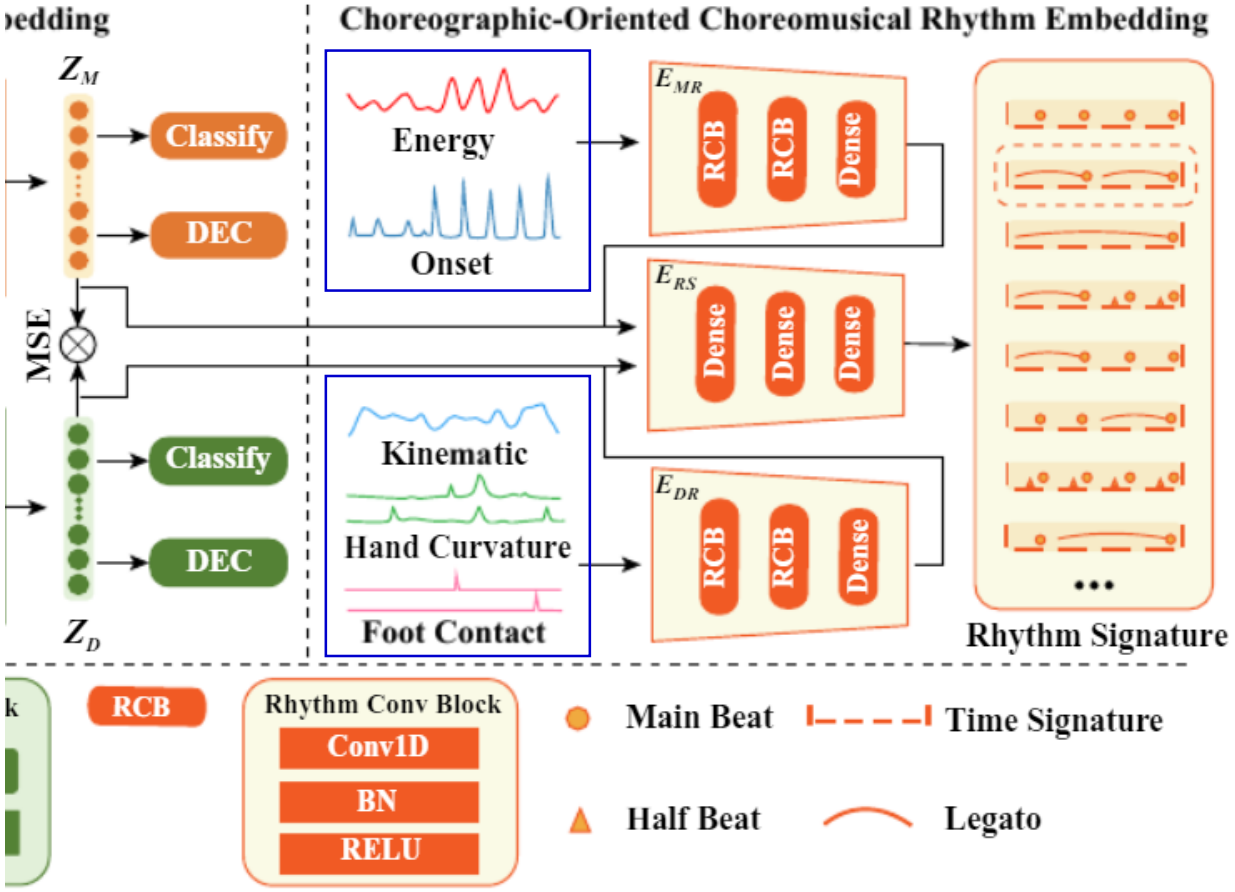


Art Statement

- Feature Extraction

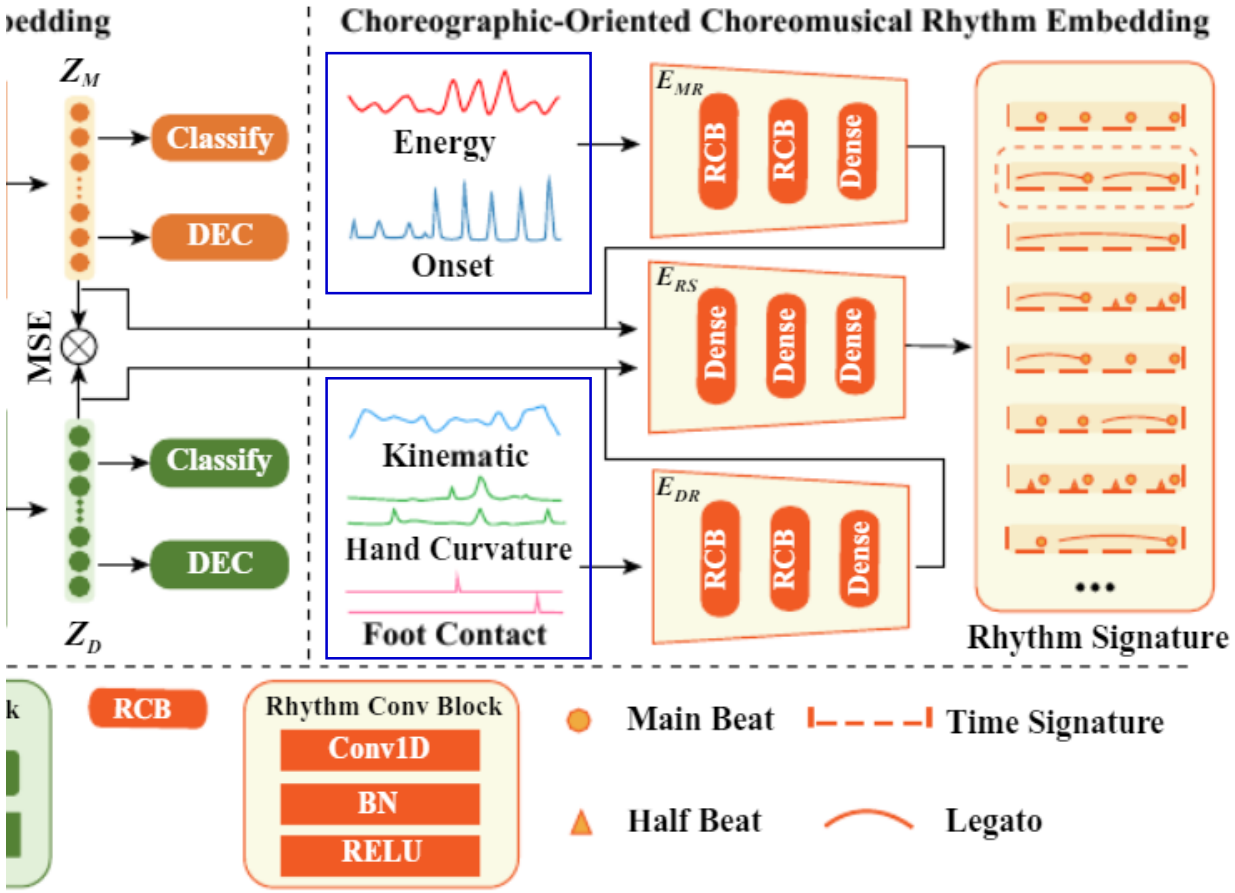
This study feed the network with certain extracted features. Specifically, the spectral onset strength curve [Böck and Widmer 2013] and RMS energy curve for music (dimension: [2, 200]); the motion kinematic curve, two hand trajectory curvature curves and two foot contact curves for dance (dimension: [5, 60]).

The motion kinematic curve is computed using the weighted angular velocity function proposed by Shiratori et al. [2006].
The hand trajectory curvature curve records the curvatures of the trajectories of the two wrist joints.
The foot contact curve records contact information between both feet and the floor.



Art Statement

- Loss Function



$$\mathcal{L}_{\text{rhythm}} = \lambda_6 L_{dr} + \lambda_7 L_{mr}$$

where Ldr and Lmr are the classification loss for dance and music respectively, and $\lambda 6$ and $\lambda 7$ are weights. To penalize large prediction errors, as well as conventional cross-entropy loss, we add the weighted Hamming distance between the predicted rhythm signature and the ground truth rhythm signature when calculating Ldr and Lmr.

Method

- GridSearch – Hyperparameters turning

$$\mathcal{L}_{\text{rhythm}} = \lambda_6 L_{dr} + \lambda_7 L_{mr}$$

λ_6		1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
λ_7		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Motion	Top-3 (%)	67.8	70.1	71.4	71.9	72.2	73.2	73.8	73.8	73.3	73.0	-
	Top-1 (%)	40.9	42.3	43.7	44.8	45.3	46.0	47.1	47.1	47.1	46.7	-
Music	Top-3 (%)	-	81.2	81.2	81.8	82.1	82.4	82.3	81.3	80.6	79.9	76.7
	Top-1 (%)	-	58.1	58.2	58.8	59.1	59.1	59.1	58.9	57.7	56.1	54.3

Method

- Motion Graph Construction

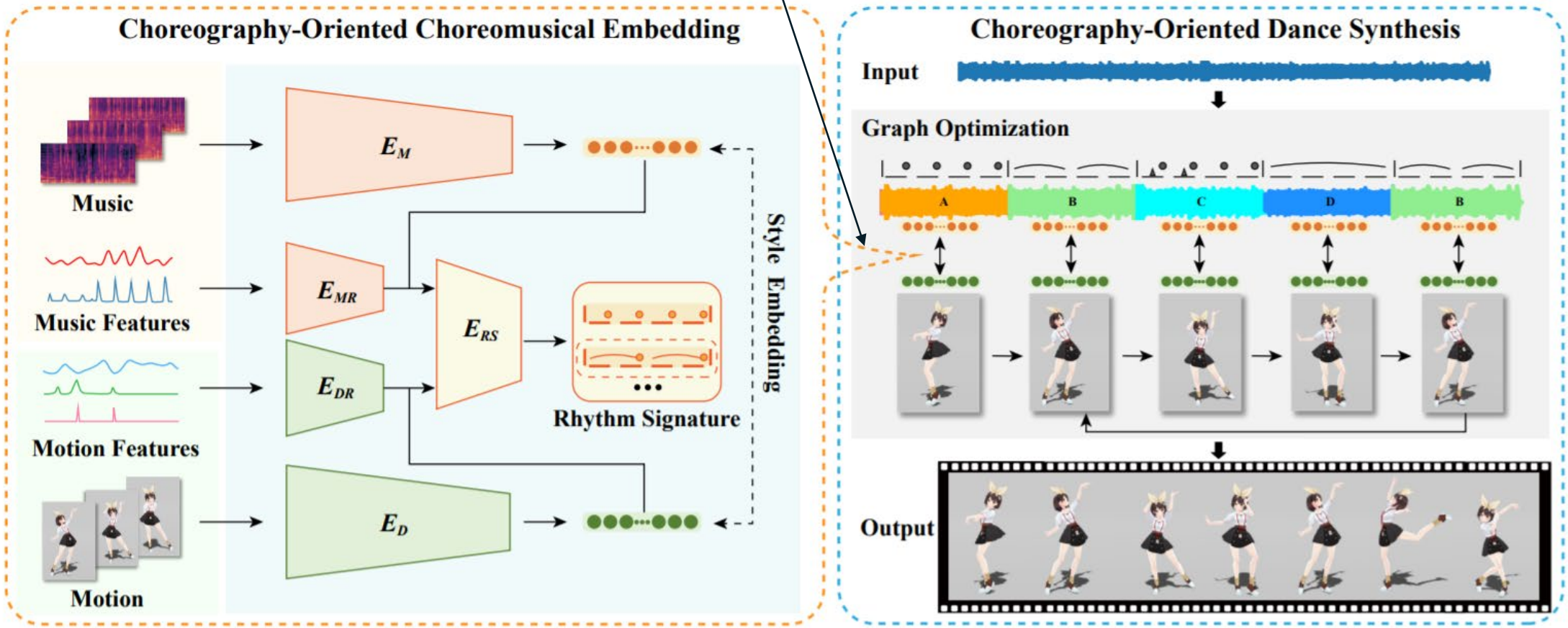
Fig. 5. Without considering style compatibility, improper edges may appear, resulting in a lovely motion switching to a sexy or cool motion as illustrated in the figure.

Fig. 6. Dance motions in our database are augmented with **mirroring, blending and reshuffling**.

Method

The learned style embedding vector and labeled rhythm signature are also attached to each graph node.

$$T(D_p, D_q) = \lambda_8 T_d + \lambda_9 T_z$$



Result

Quantitative Evaluation. They adopt several evaluation metrics to quantitatively compare these methods, as shown in Table 3. These metrics are:

1. **FID score**. Fréchet inception distance (FID) [Heusel et al. 2017] was used to measure how close the distribution of generated dances is to that of the real ones.
2. **Beat accuracy.** This measures how accurately the motion beats are aligned to the music beats, represented by the **ratio of aligned beats to all music beats.**
3. **Diversity.** They follow [Lee et al. 2019] to evaluate the average feature distance between generated dances for different music inputs.

Method	FID	Beat Accuracy	Diversity
Real Dance	2.7	92.6%	83.5
Lee et al. [2013]	24.5	38.4%	75.1
Yan et al. [2019]	94.6	8.2%	56.2
Sun et al. [2020]	87.4	12.7%	64.1
Ours (w/o EC)	20.5	85.2%	72.4
Ours (w/o RC)	17.9	58.3%	76.5
Ours (w/o SC)	18.5	83.8%	78.3
Ours	16.8	88.4%	77.9

Result

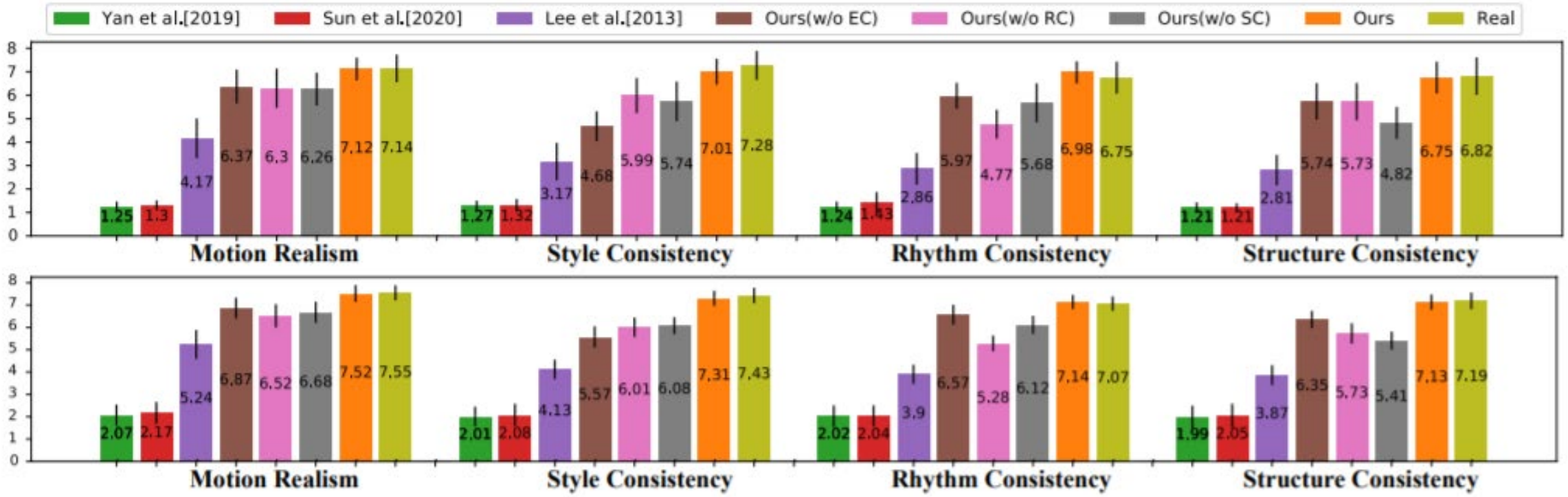


Fig. 9. User study results for ordinary users (above) and for choreographers and artists (below). Our method achieves higher scores than Lee et al. [2013], Lee et al.[2019], Yan et al. [2019] and Sun et al. [2020].

Result

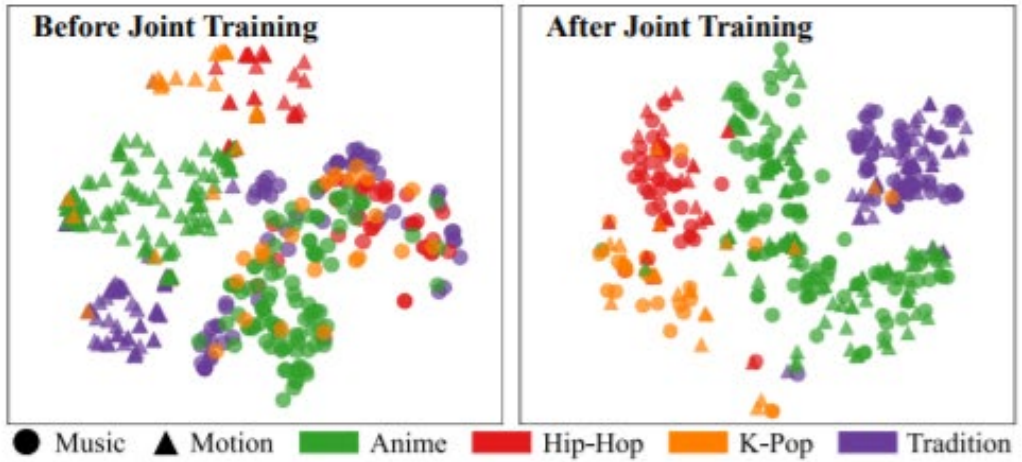


Fig. 10. T-SNE visualization of the choreomusical style embedding of musics and dances from the test set. Left, right: results before and after joint training. The embedding of music-dance pairs becomes much closer after joint training

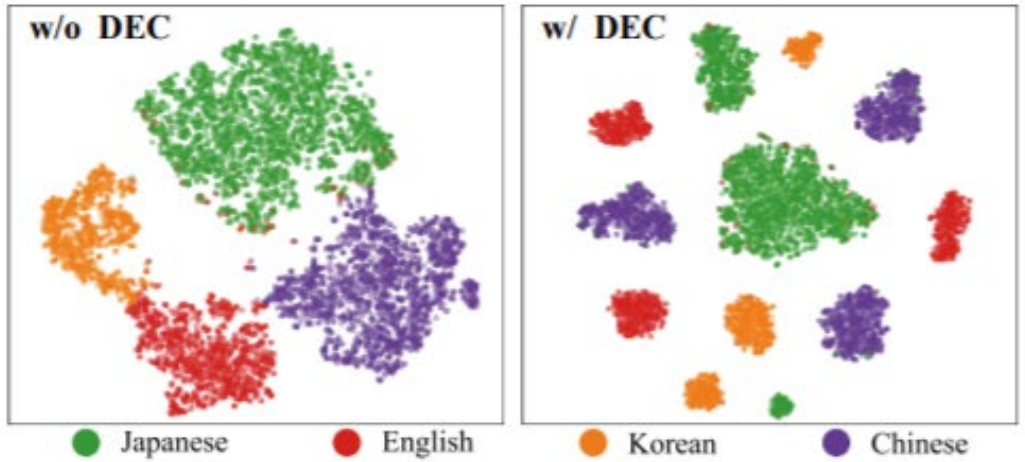


Fig. 11. T-SNE visualization of music distributions in the feature space of the classification network trained without DEC loss (left) and with DEC loss (right).

Result


Dance Clip:
clip034@dance1062
Rhythm Signature:
01 01 00 01
Music by Room97 on audiostock.net

Connection

- This study introduce three rules from choreography theory, which greatly facilitate music-driven dance motion synthesis;
- They develop a cross-domain embedding framework, incorporating the introduced rules, to correctly and effectively characterize complex choreomusical relationships from limited available high-quality music/motion data, which successfully casts qualitative choreographic knowledge into computable metrics;
- This study present the first production-ready dance motion synthesis system, which can robustly generate high-quality dance motions in a highly controllable way;