A Brand New Dance Partner: Music-Conditioned Pluralistic Dancing Controlled by Multiple Dance Genres

Jinwoo Kim[1], Heeseok Oh[2], Seongjean Kim[1], Hoseok Tong[1] and Sanghoon Lee[1]

[1]Department of Electrical and Electronic Engineering, Yonsei University, Seoul
[2]Department of Applied AI, Hansung University, Seoul

Motivation:
What if artificial intelligence could be used to help choreographers blend dance genres by suggesting various dances, and one that matches their choreographic style?

Goal & Contribution

- Is one-to-many 3D dance generation conditioned by music
  - To generate realistic and diverse dance motion
- Generate dance controlled by multiple dance genres
  - Scalable to multiple dance genres

Waacking

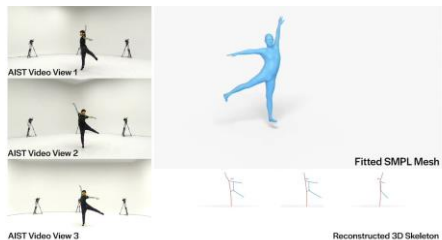Waacking   Break
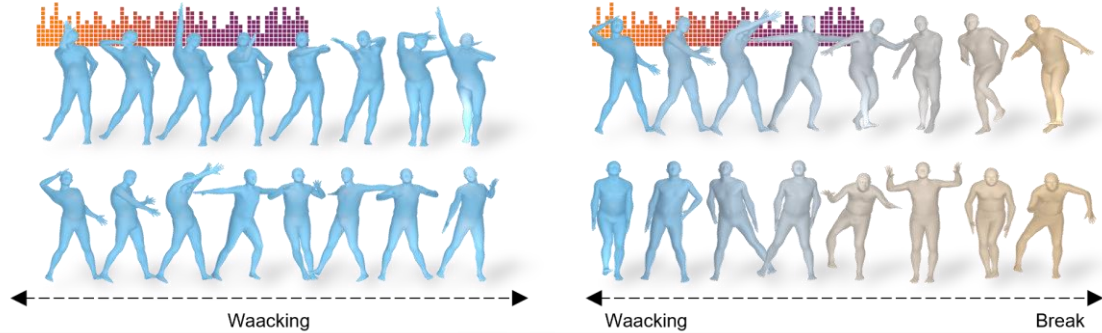
MNET: Music-conditioned transformer NETwork

- Non-autoregressive Transformer architecture (during training) [2]
- MNET autoregressively synthesizes realistic 3D dance sequences (during inference)
- Sequence-level generator via Transformer conditional GAN (mapping network)
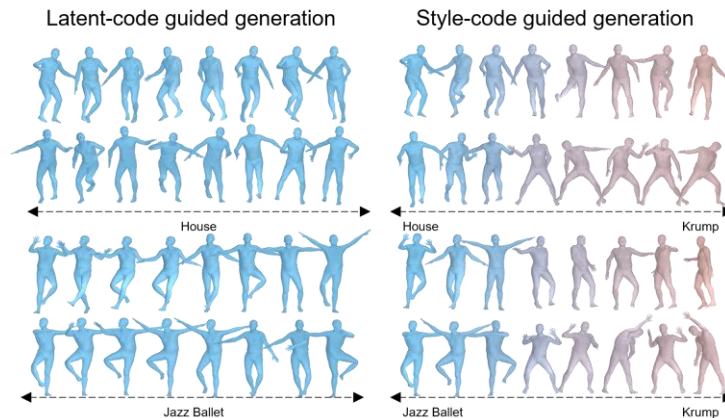- Allows to generate diverse 3D dance sequences with various dance genre styles
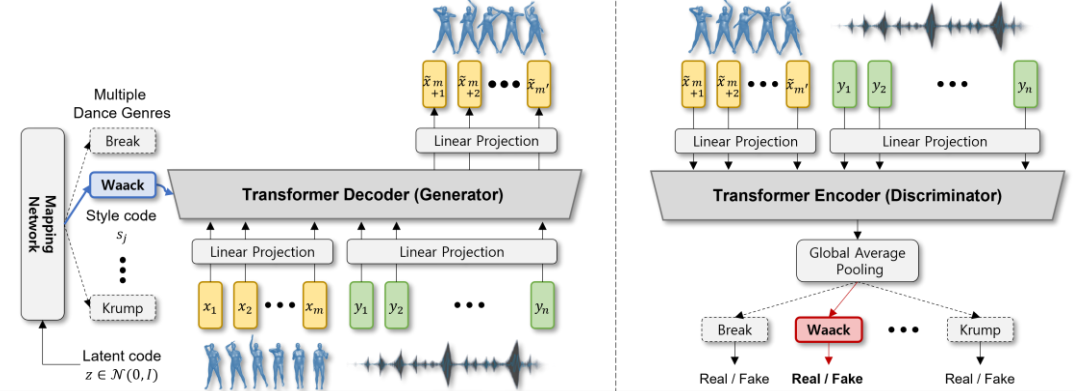
Training Data [1]
The AIST++ dance motion dataset

- Contains 1,408 sequences of 3D human dance
  - Represented as joint rotation along with root trajectory
- Provides 3D human key-point annotations and camera parameters
 - Covering 30 different dancers in 9 views

AIST Video View 1
AIST Video View 2
AIST Video View 3
Fitted SMPL Mesh
Reconstructed 3D Skeleton

Qualitative Results

Latent-code guided generation   Style-code guided generation

House   House   Krump

Jazz Ballet   Jazz Ballet   Krump

Comparisons with Previous works

- FID: Frechet Inception Distance
- DIST: Average Euclidean Distance

-- Kinematic beats   — Music beats   — Kinematic velocity

time
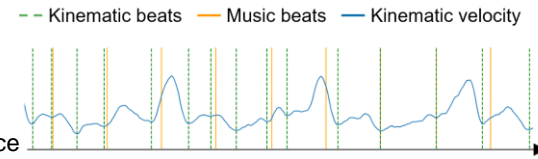
	Motion Plausibility				Generation Diversity			Motion-Music Cons.	User Study
	$FID_k \downarrow$	$FID_g \downarrow$	$FID_s \downarrow$	Acc. ↑	$Dist_{m,k} \uparrow$	$Dist_{m,g} \uparrow$	$Dist_{m,s} \uparrow$	BeatAlign ↑	MNET WinRate
AIST++	-	-	-	98.6	10.39	8.48	8.91	0.292	42.3%
Dancenet [66]	56.67	16.47	38.49	43.6	2.10	2.64	2.76	0.220	90.38 %
DanceRevolution [21]	42.93	14.85	26.53	72.9	3.82	3.31	2.45	0.215	84.17 %
FACT [38]	33.08	11.82	11.37	76.1	5.83	5.28	5.31	0.241	62.39 %
MNET (ours)	**29.52**	**9.36**	**7.90**	83.7	6.93	6.77	6.32	**0.246**	-

References
[1] Li, R., et al. "Ai choreographer: Music conditioned 3d dance generation with aist++." ICCV 2021
[2] Petrovich, M., et al. Li, R., et al. " Action-conditioned 3d human motion synthesis with transformer vae." ICCV 2021