GDC × NetEase Games
Passion of gamers

March 21-25,2022
San Francisco,CA

NeteaseGames AI Lab

AI Animator :
A Real Time Motion Completion System

Yinglin Duan
NetEase Games AI Lab

#GDC22
#NetEase Games

Overview

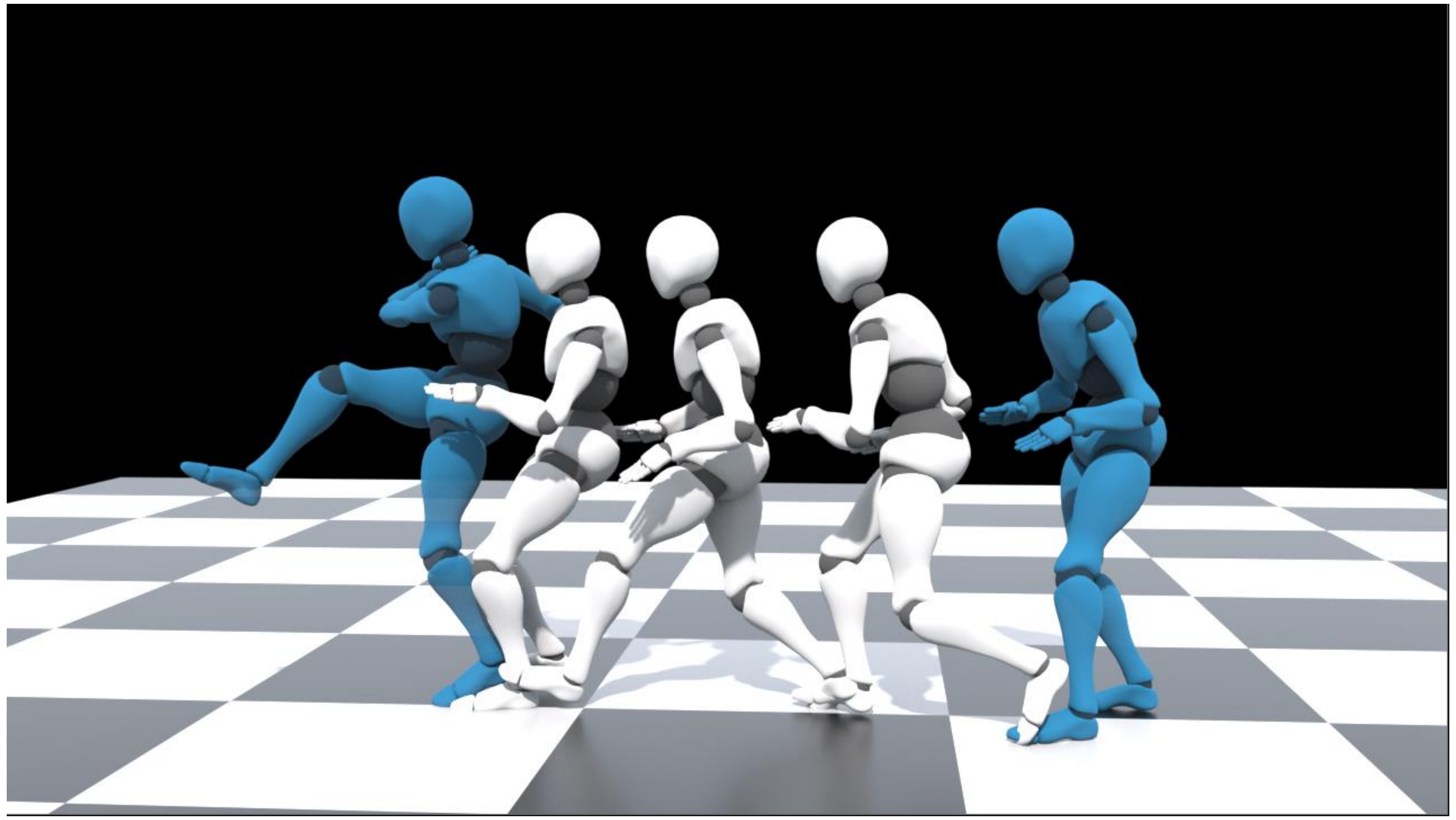
- Background
- Methodology
- Experiments
- Application
- Q & A

Why we are here?

- Discussion
  - Paper of motion completion state-of-art
  - Details and tips in our paper
- Take a guess
  - 3D computer animation
  - Application of deep learning

A Unified Framework for Real Time Motion Completion

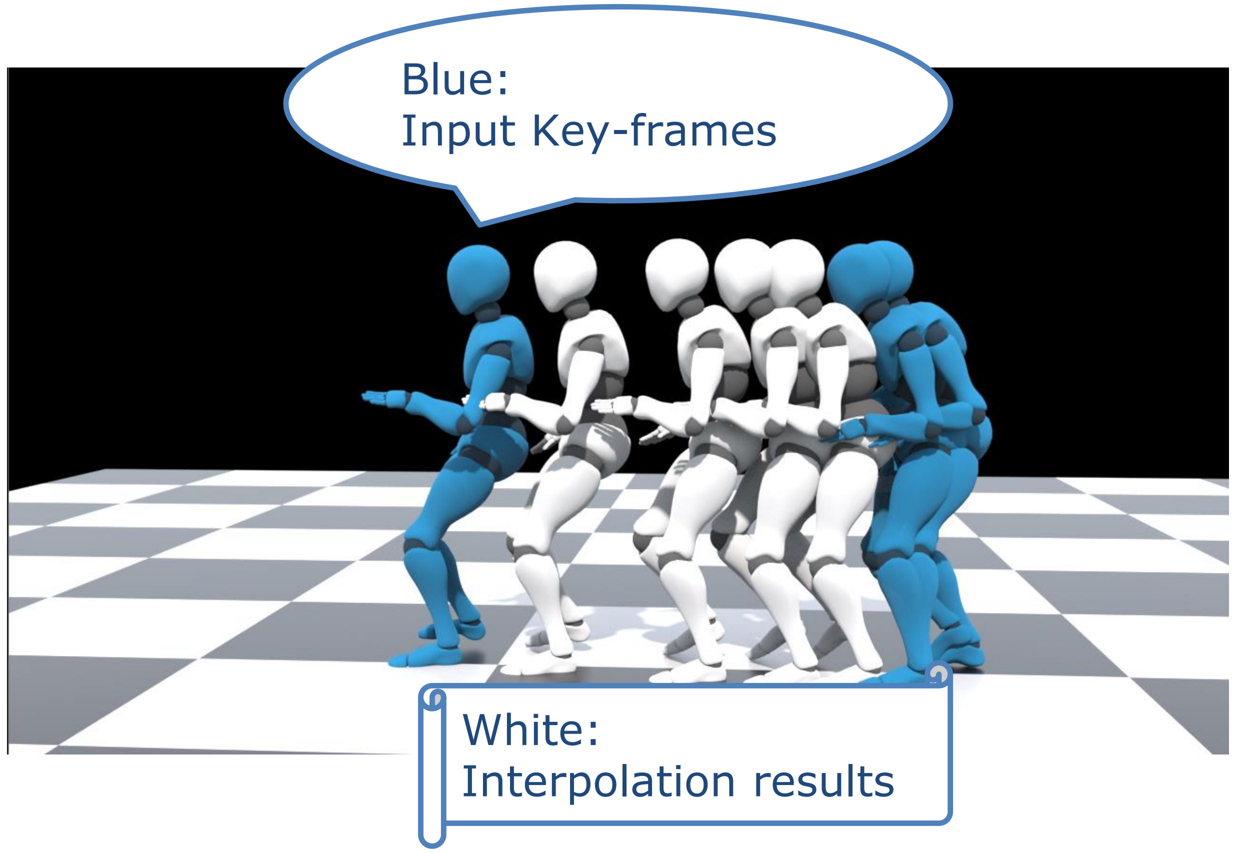
Yinglin Duan[1], Yue Lin[1], Zhengxia Zou[2*], Yi Yuan[3], Zhehui Qian[3], Bohan Zhang[3]
[1] NetEase Games AI Lab, Guangzhou, China,
[2] Beihang University,
[3] Netease, Inc., Hangzhou, China
{duanyinglin, gzlinyue, yuanyi, qianzhehui, hzzhangbohan}@corp.netease.com, zhengxiazou@buaa.edu.cn



Model from https://www.mixamo.com, only for illustrations

Traditional animation workflows

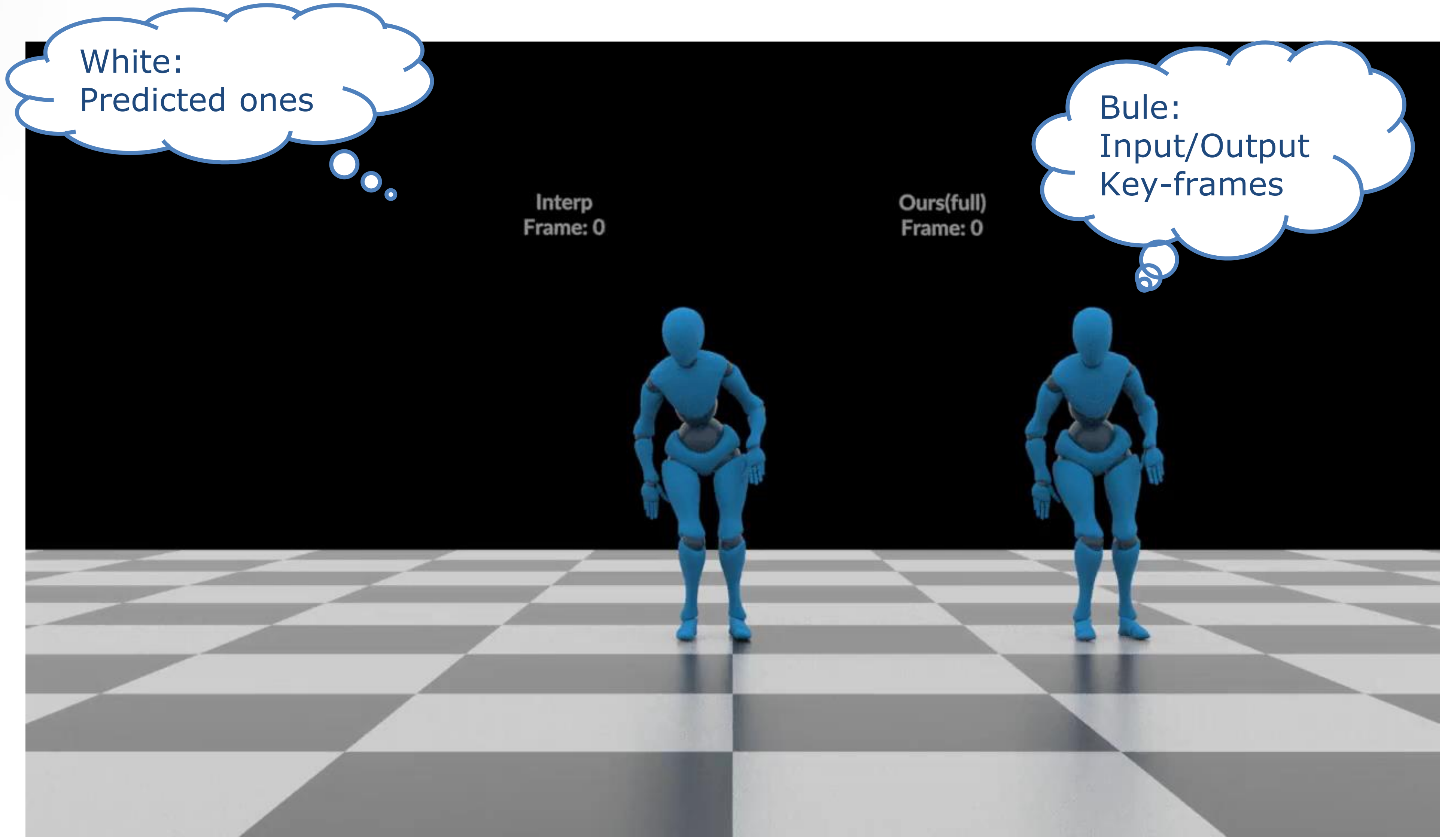
- Artists / animators
  - drawing key-frames are a burden on animators
- Linear interpolation, IK, searching-based
  - can't generate high-quality and long animations
- State machines
  - require a lot of manpower to build and maintain



Linear interpolation

Recently

- AI assisted workflow
  - RNN-based
  - Convolution-based
  - Ours



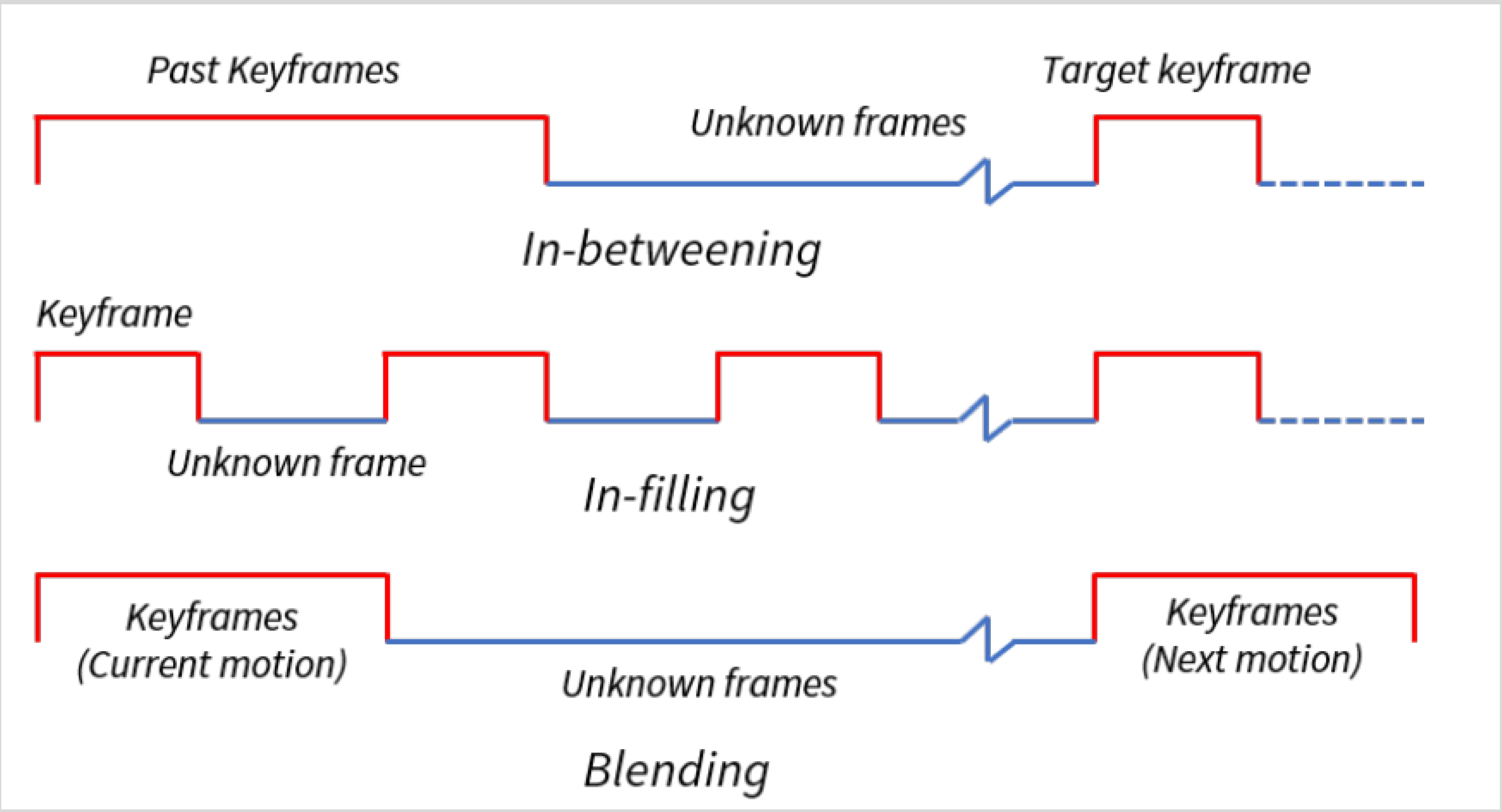
Linear interpolation & Our results

Our motivations

- We need a simple but effective method:
  - **familiar** to animators and enthusiasts
  - handling **a massive scope** of datasets
  - **real-time** high quality animation generation

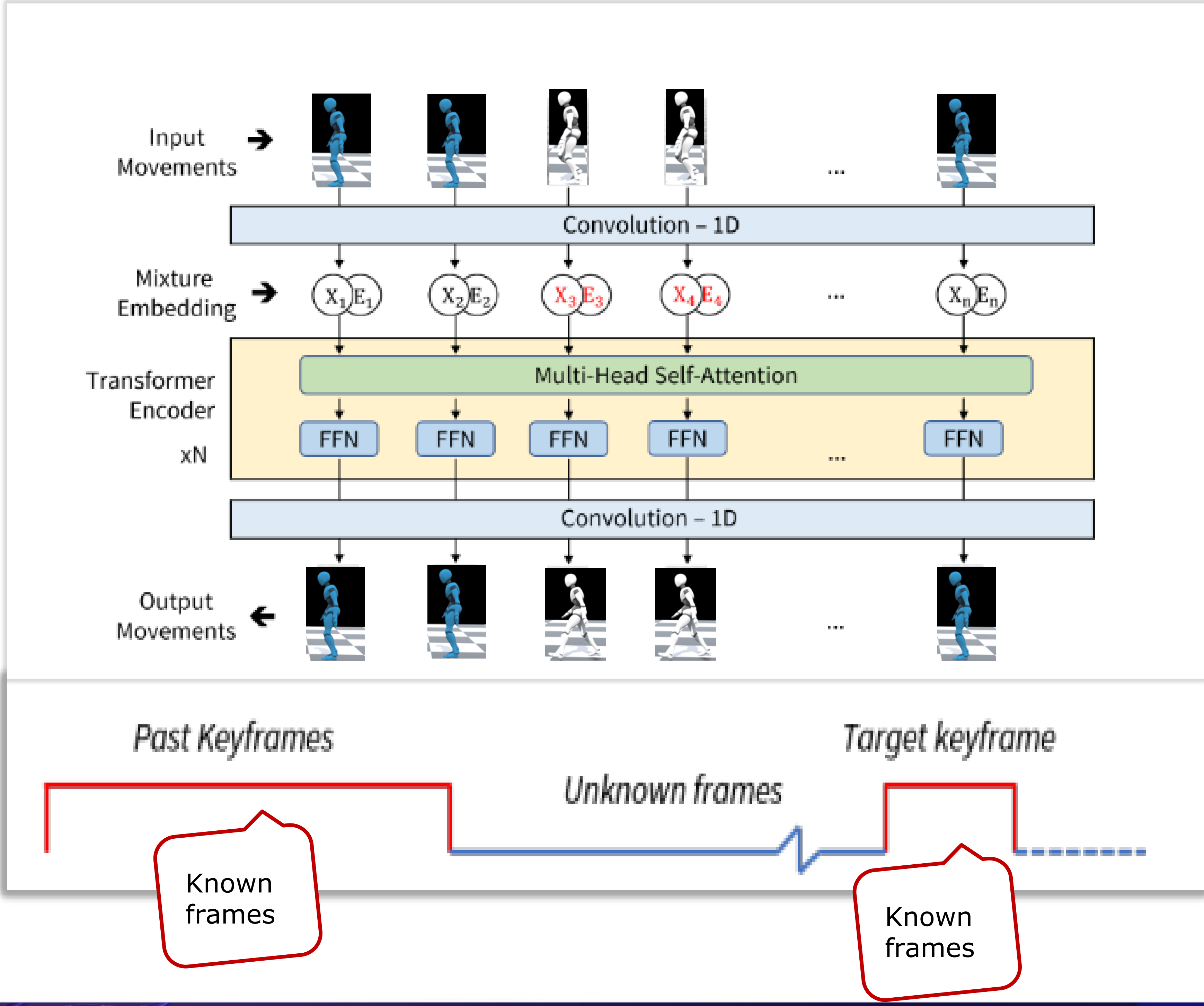
Our motivations

- A unified framework
  - In-betweening
  - In-filling
  - Blending



Architecture

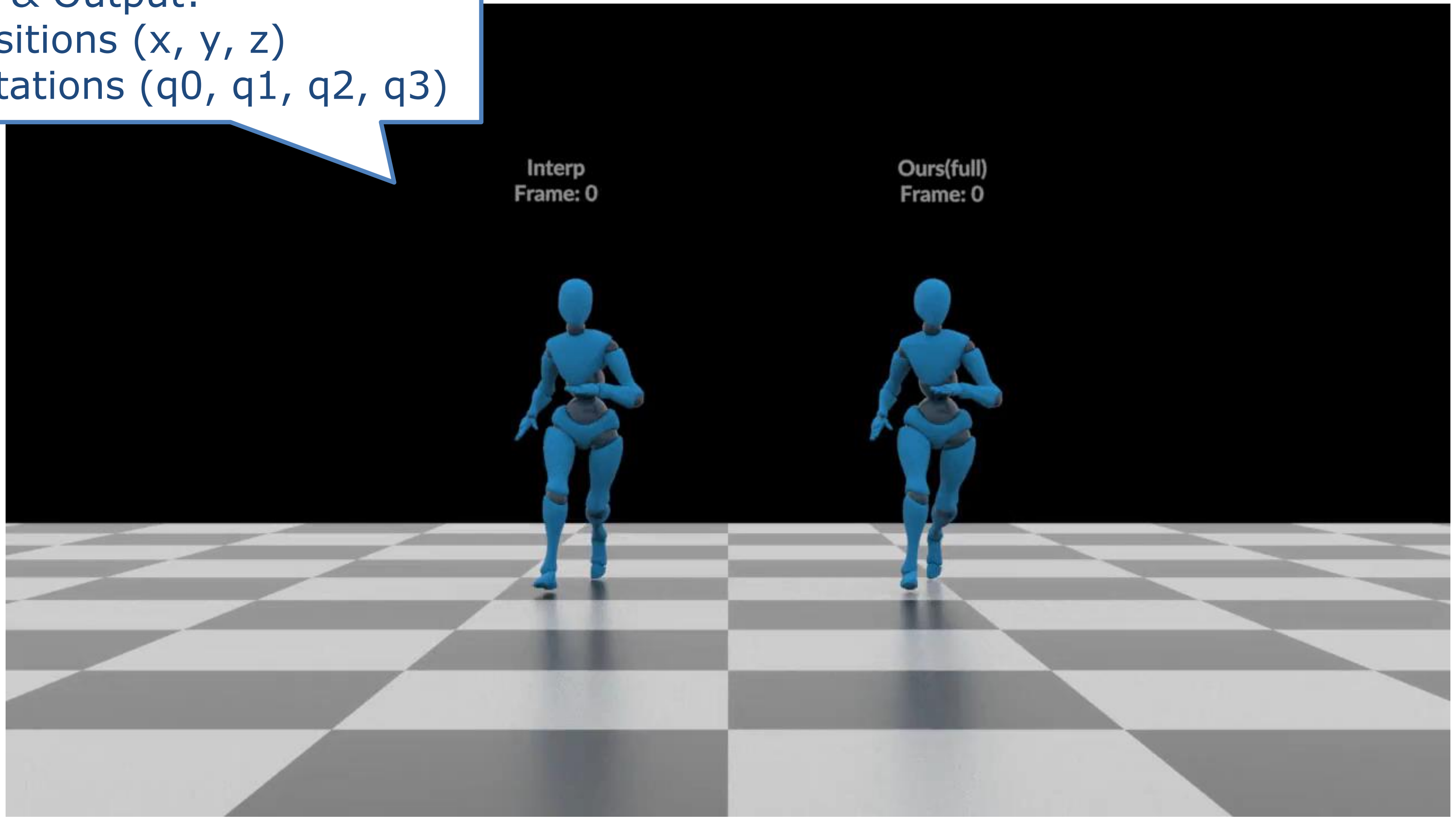
- Backbone:
  - Bidirectional Encoder Representation from Transformers
- Input:
  - known frames + unknown frames
- Output:
  - Predict ones



Frame in details

- Input :
  - Known frames: Key frame(s)
  - unknown frames : Interpolation
  - Target frames : Key frame(s)
- Output :
  - Generation frames
- Input & Output poses details:
  - joint positions - coordinate matrix
  - joint rotations - quaternion matrix

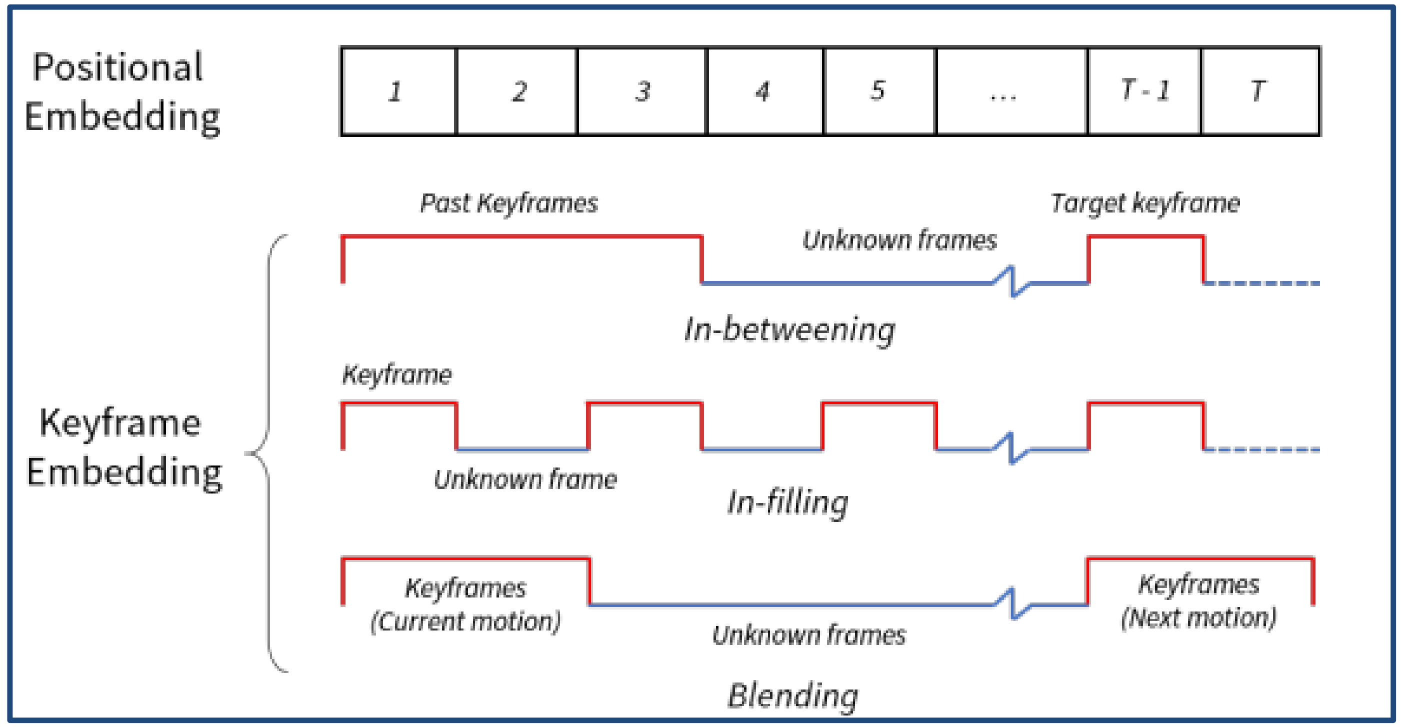
Input & Output:
P: positions (x, y, z)
Q: rotations (q0, q1, q2, q3)



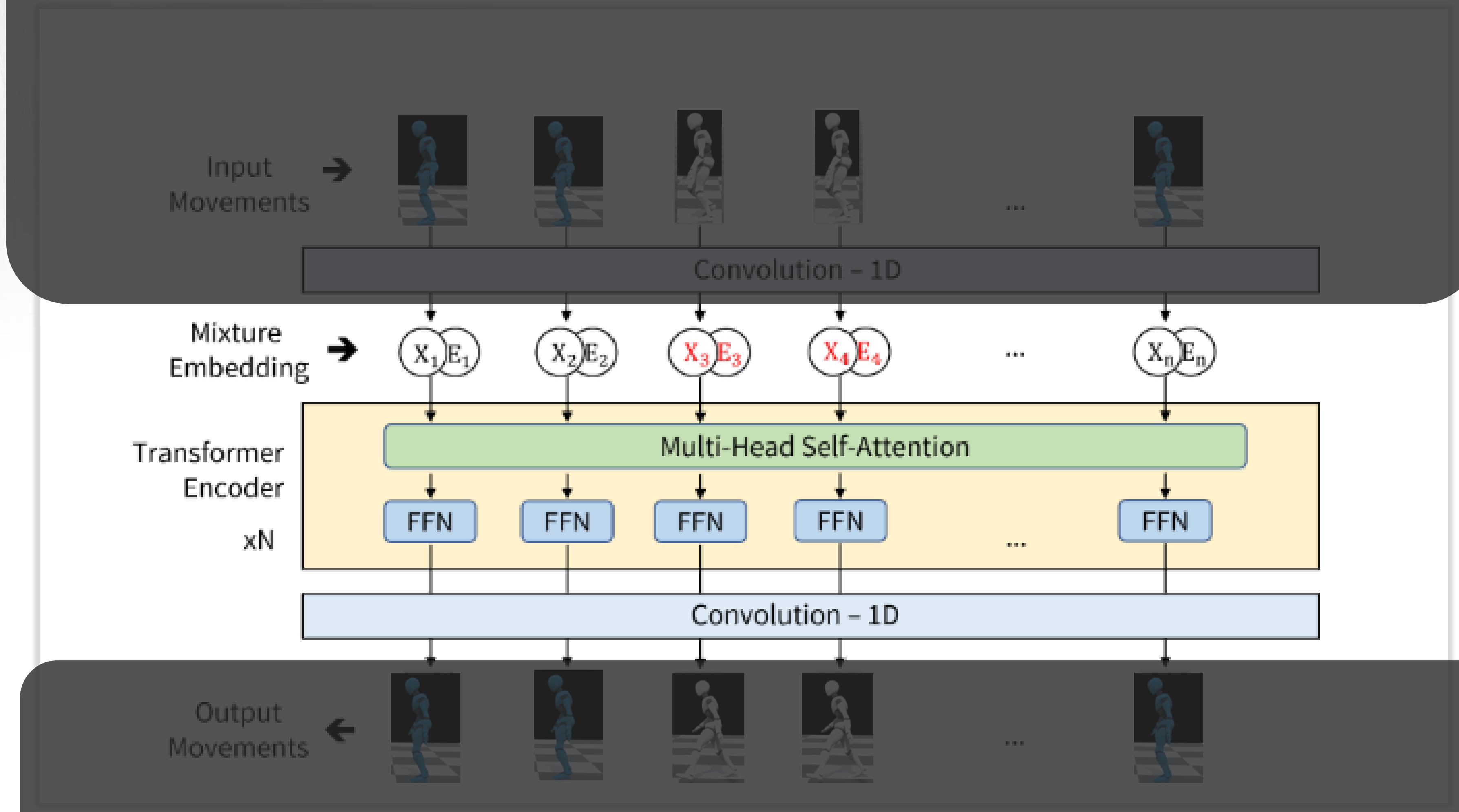
Linear interpolation & Our in-betweening results

Mixture embeddings

- Position embedding:
  - time sequence
  - each one for a single time step
- Keyframe embedding:
  - annotates a keyframe



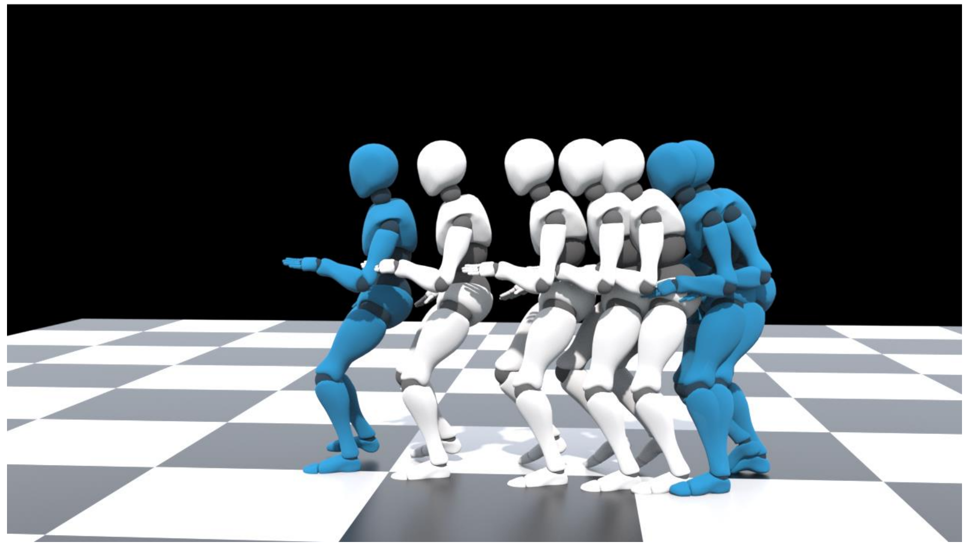
Transformer in details



Reconstruction Loss

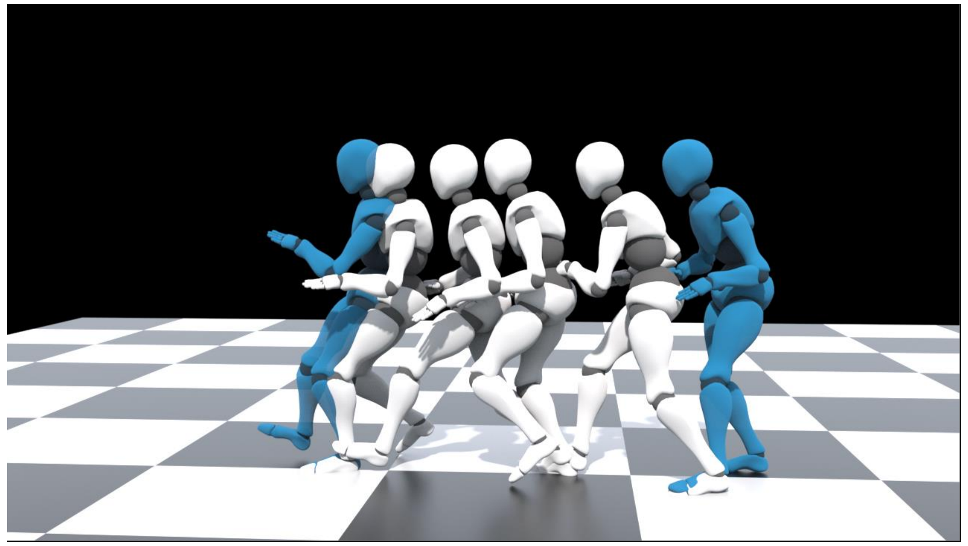
- Loss function
  - $\mathcal{L}_{\text{rec}} = \frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}(\|\boldsymbol{p}_n^t - \hat{\boldsymbol{p}}_n^t\| + \|\boldsymbol{p}_n^t - \hat{\boldsymbol{p}}_n^t\|)$

Known frames + Interpolation frames



Input

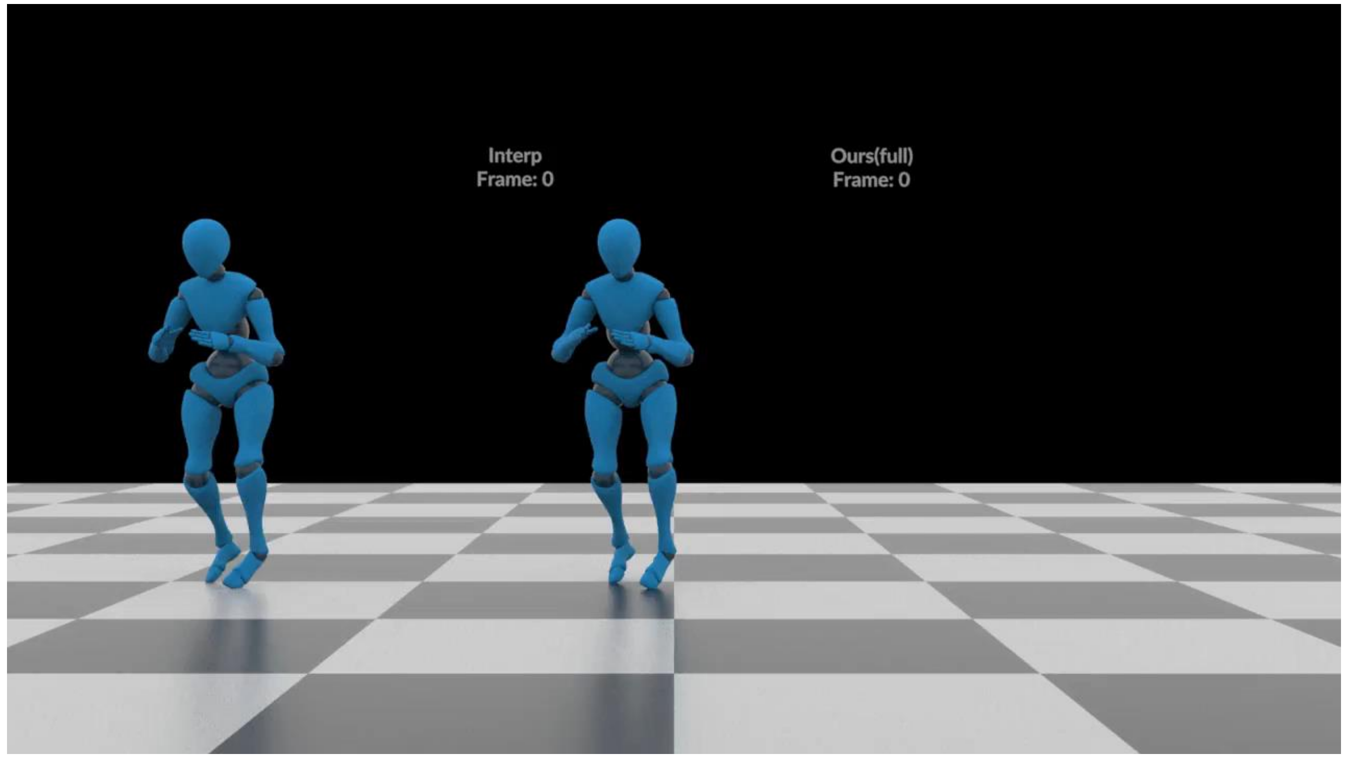
Reconstruction Frames



Output

Kinematics Loss

- Forward Kinematics loss
  - follow Harvey et al.[1]
  - in local coordinate system
- Inverse Kinematics loss
  - in global coordinate system

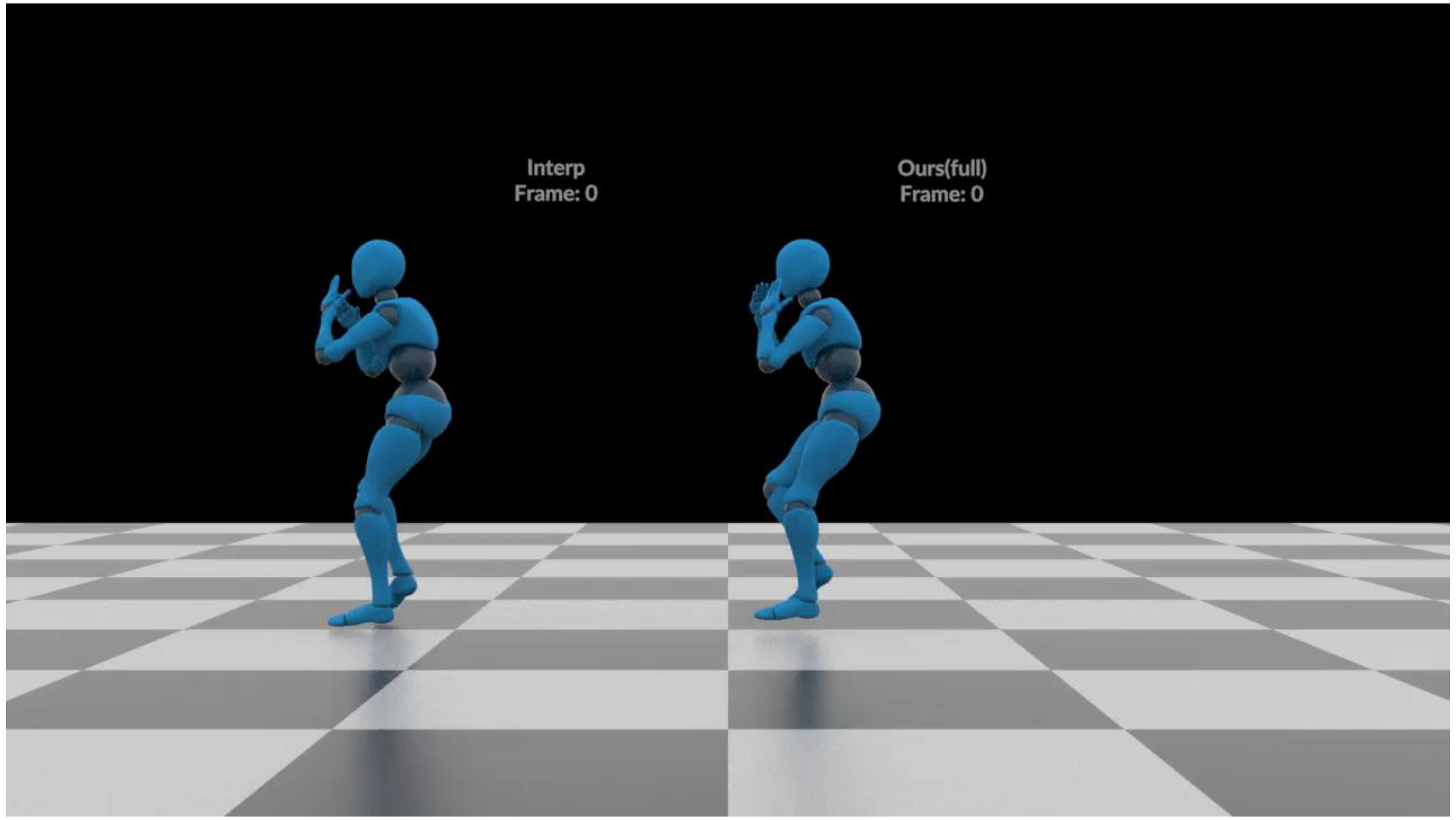


Our in-betweening results with IK loss

[1] Harvey F G, Yurick M, Nowrouzezahrai D, et al. Robust motion in-betweening[J]. ACM Transactions on Graphics (TOG), 2020, 39(4): 60: 1-60: 12.

Motion perceptual loss

- Ground contact constraints
  - reduce foot-skate in human motion
- Discrete wavelet transformation
  - capturing of high-frequency information



Our blending results with perceptual loss

Implementation Details

Item	Setting
Backbone	BERT
Depth	8 transformer layers
Bandwidth	256-dimension
Attention Heads	8
Optimizer	ADAM
Learning Rate	1e-3
Framework	PyTorch

Experiments

	L2Q			L2P			NPSS		
Length	5	15	30	5	15	30	5	15	30
Zero-Vel	0.56	1.10	1.51	1.52	3.69	6.60	0.0053	0.0522	0.2318
Interp	0.22	0.62	0.98	0.37	1.25	2.32	0.0023	0.0391	0.2013
Harvey *et al.* (LSTM-based)	0.17	0.42	0.69	0.23	0.65	1.28	0.0020	0.0258	0.1328
Kaufmann *et al.* (Conv-based)	0.49	0.60	0.78	0.84	1.07	1.53	0.0048	0.0345	0.1454
Ours (local w/ ME)	0.18	0.47	0.74	0.27	0.82	1.46	0.0020	0.0307	0.1487
Ours (local w/ ME & FK loss)	0.17	0.44	0.71	0.23	0.74	1.37	0.0019	0.0291	0.1430
Ours (global w/ ME)	0.14	0.36	0.61	**0.21**	0.57	1.11	0.0016	0.0238	0.1241
Ours (global w/ ME & IK loss)	**0.14**	**0.36**	0.61	0.22	0.56	1.10	**0.0016**	**0.0234**	0.1222
Ours (global transformer only)	0.16	0.37	0.63	0.24	0.61	1.16	0.0018	0.0243	0.1284
Ours (global full)	0.18	0.37	**0.61**	0.23	**0.56**	**1.06**	0.0018	0.0238	**0.1218**
Ours (noisy training data - 30 db)	0.19	0.39	0.63	0.27	0.59	1.11	0.0020	0.0248	0.1259

Experimental results on LaFAN1 dataset

Experiments

Length	L2P 5	L2P 15	L2P 30
Zero-Vel	2.34	5.12	6.73
Interp	0.94	3.24	4.68
Kaufmann *et al.*	3.57	3.69	3.93
Ours (full)	**0.84**	**1.46**	**1.64**

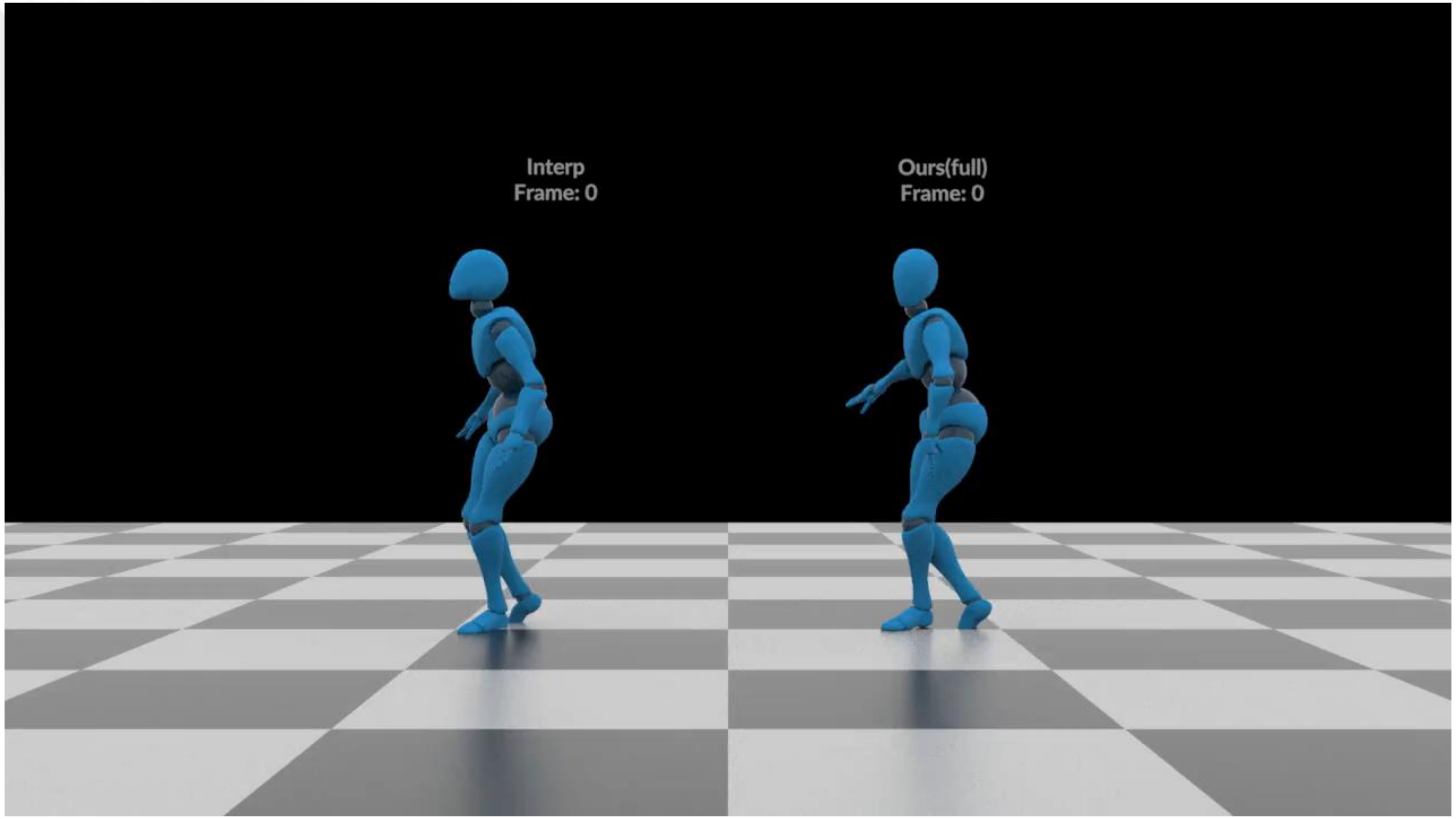
In-filling results on Anidance dataset

Experiments

	L2Q			L2P			NPSS		
Length	8	16	32	8	16	32	8	16	32
Zero-Vel	0.93	1.41	1.89	2.71	4.08	5.73	0.0221	0.1003	0.4875
Interp	0.40	0.92	1.44	1.16	2.34	3.70	0.0122	0.1095	0.4726
Kaufmann *et al.*	1.36	1.35	1.35	3.64	3.63	3.64	0.0392	0.1223	0.3944
Ours	**0.35**	**0.60**	**0.98**	**0.79**	**1.35**	**2.32**	**0.0122**	**0.0631**	**0.3124**
Ours (enhanced)	**0.29**	**0.51**	**0.89**	**0.62**	**1.12**	**2.12**	**0.0098**	**0.0529**	**0.2817**

Blending results of our new dance dataset

Experiments



Blending results of our new dance dataset

Application

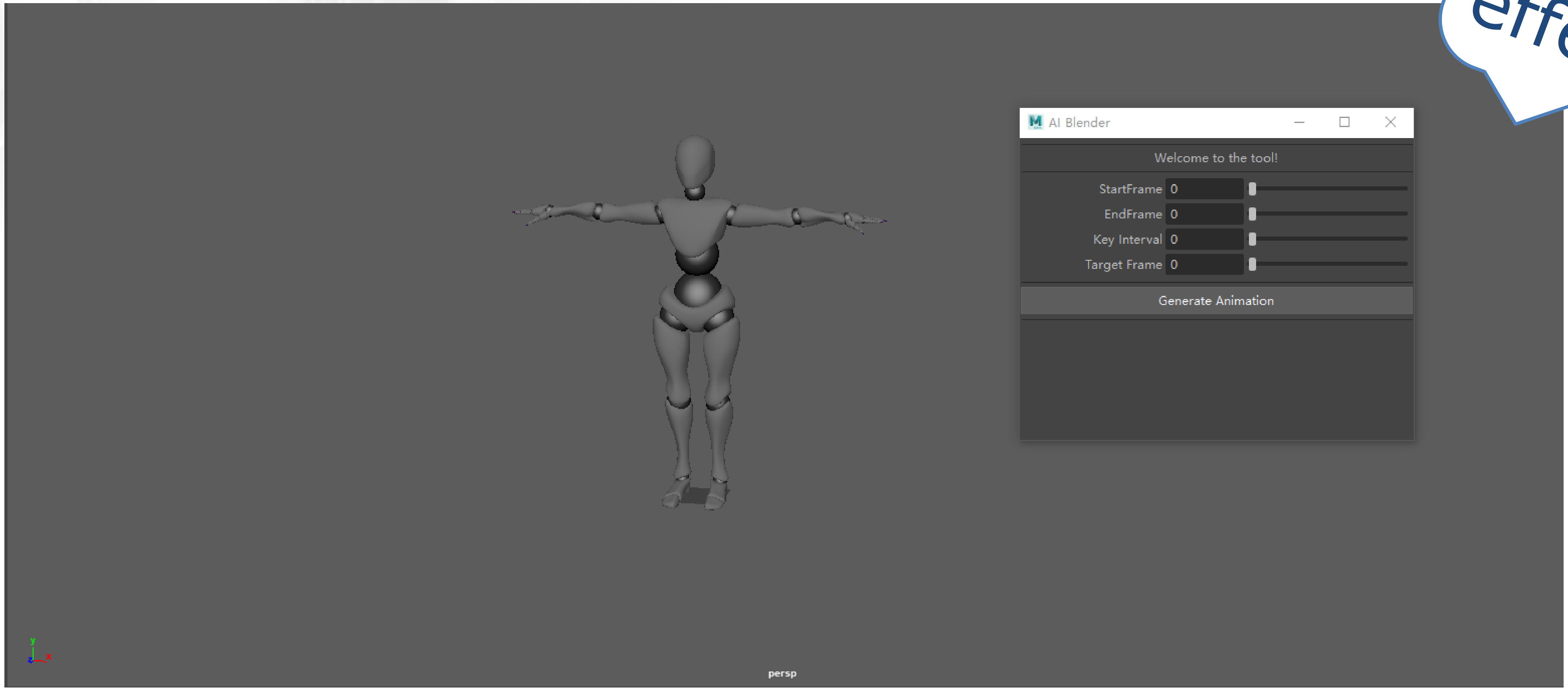
- Easy to use
  - Unified framework: mocap data / raw animation data / dance cards at any length
  - Embed in animation pipeline: polish and editing (e.g. Maya Plugin)
  - High performance: real-time inference on CPU

Method	1 x 30	10 x 30	CPU info
Harvey *et al.*	0.31s	0.40s	E5-1650
Kaufmann *et al.*	0.066s	0.33s	I7-8700K
Ours(global full)	**0.025s**	**0.083s**	I7-8700K

  - CPU inference time is* recorded in different batch sizes (1 & 10), where in-betweening length is set to 30 frames (i.e. 1 second).

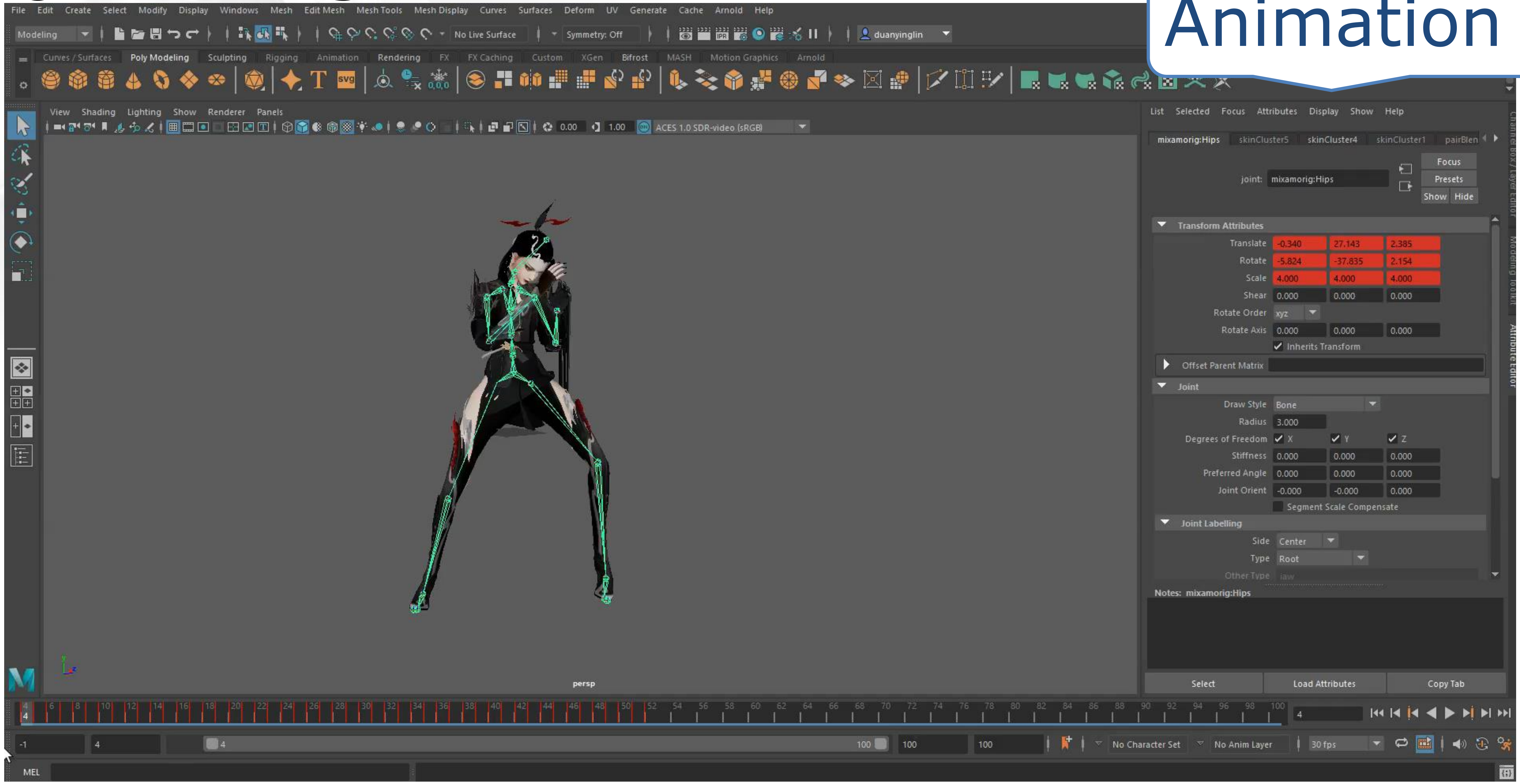
Maya plugin

simple but effective



Maya plugin

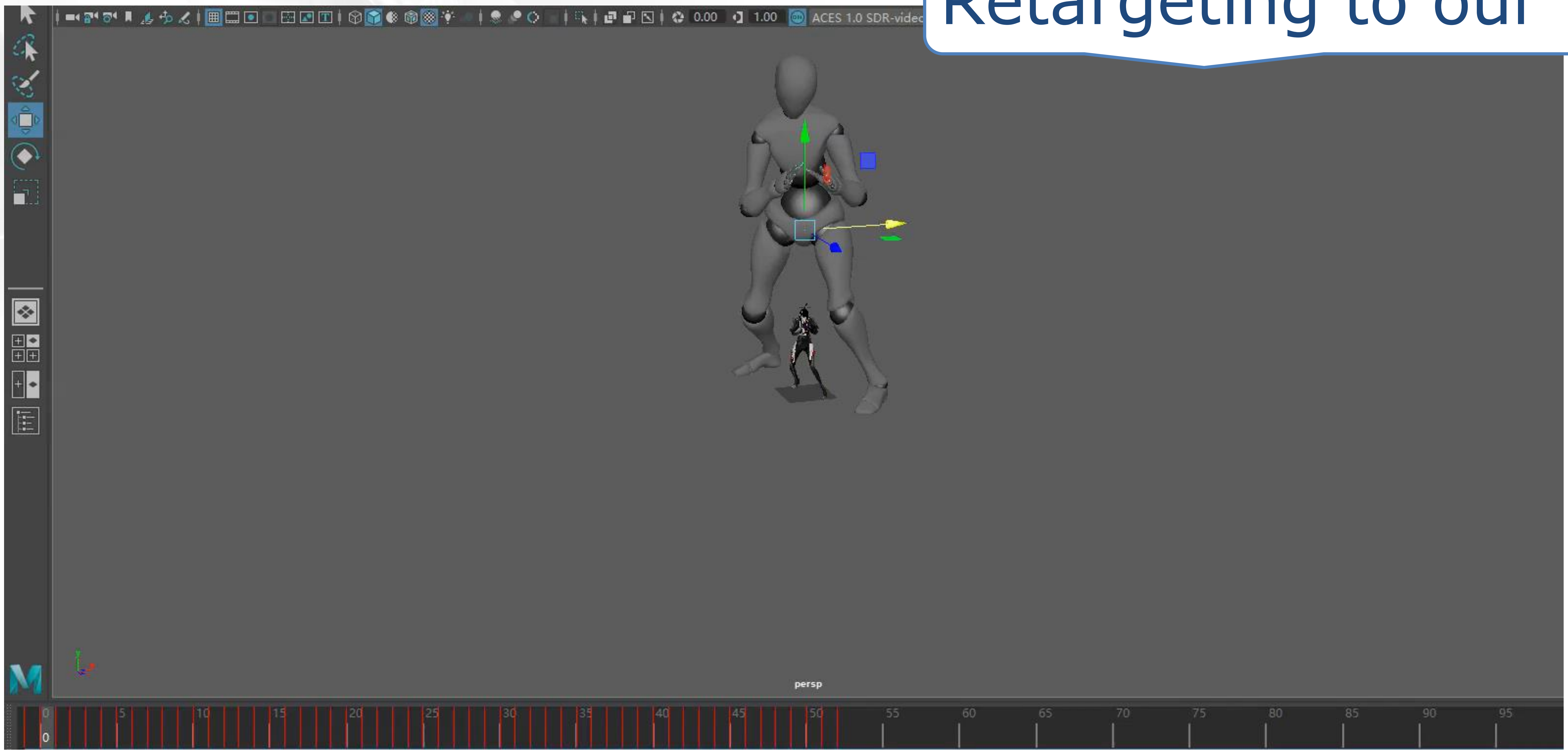
Animation Clip



Model from the Netease Game ***The World 3***

Maya plugin

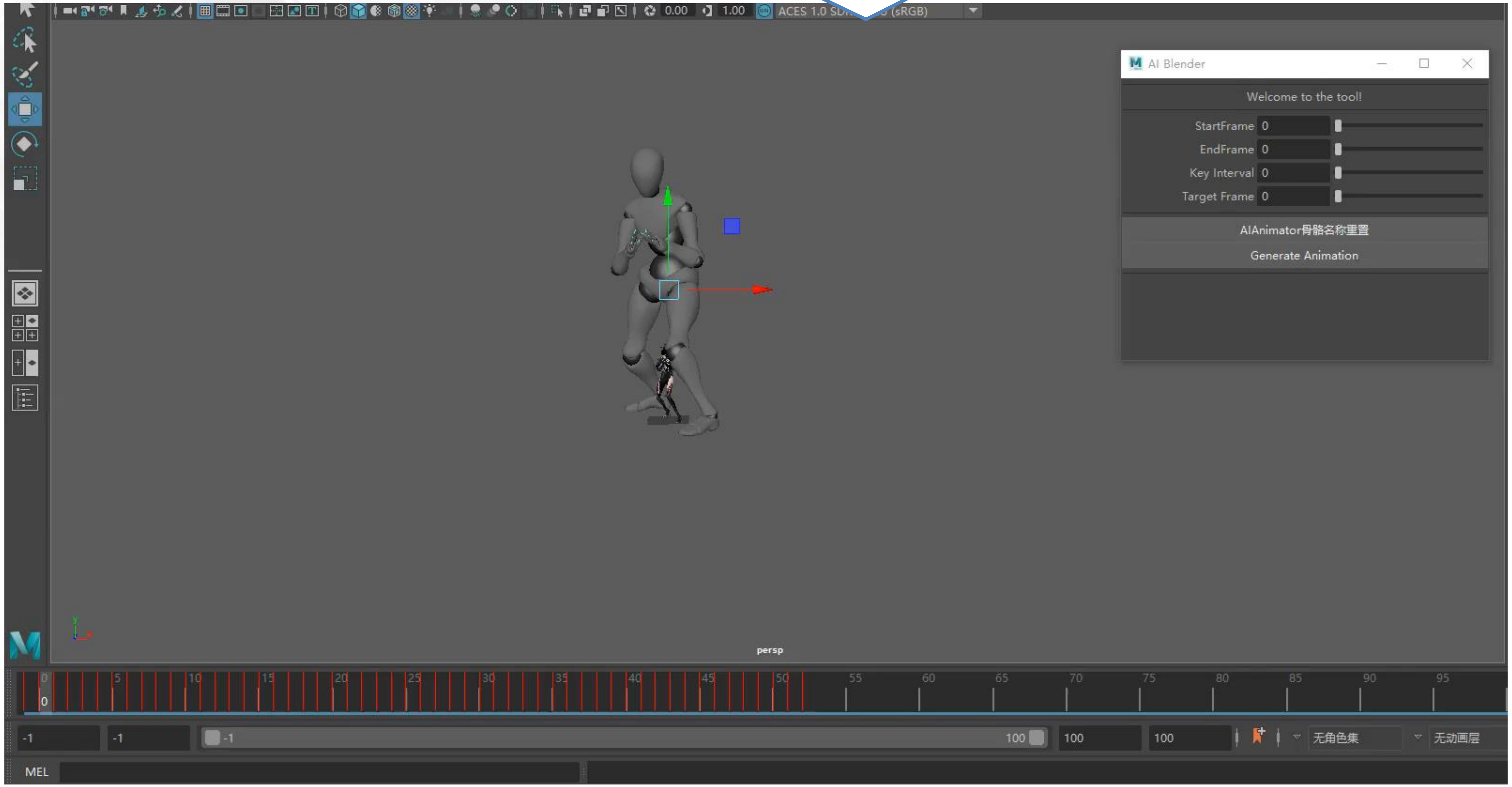
Retargeting to our model



Model from the Netease Game ***The World 3***

Maya plugin

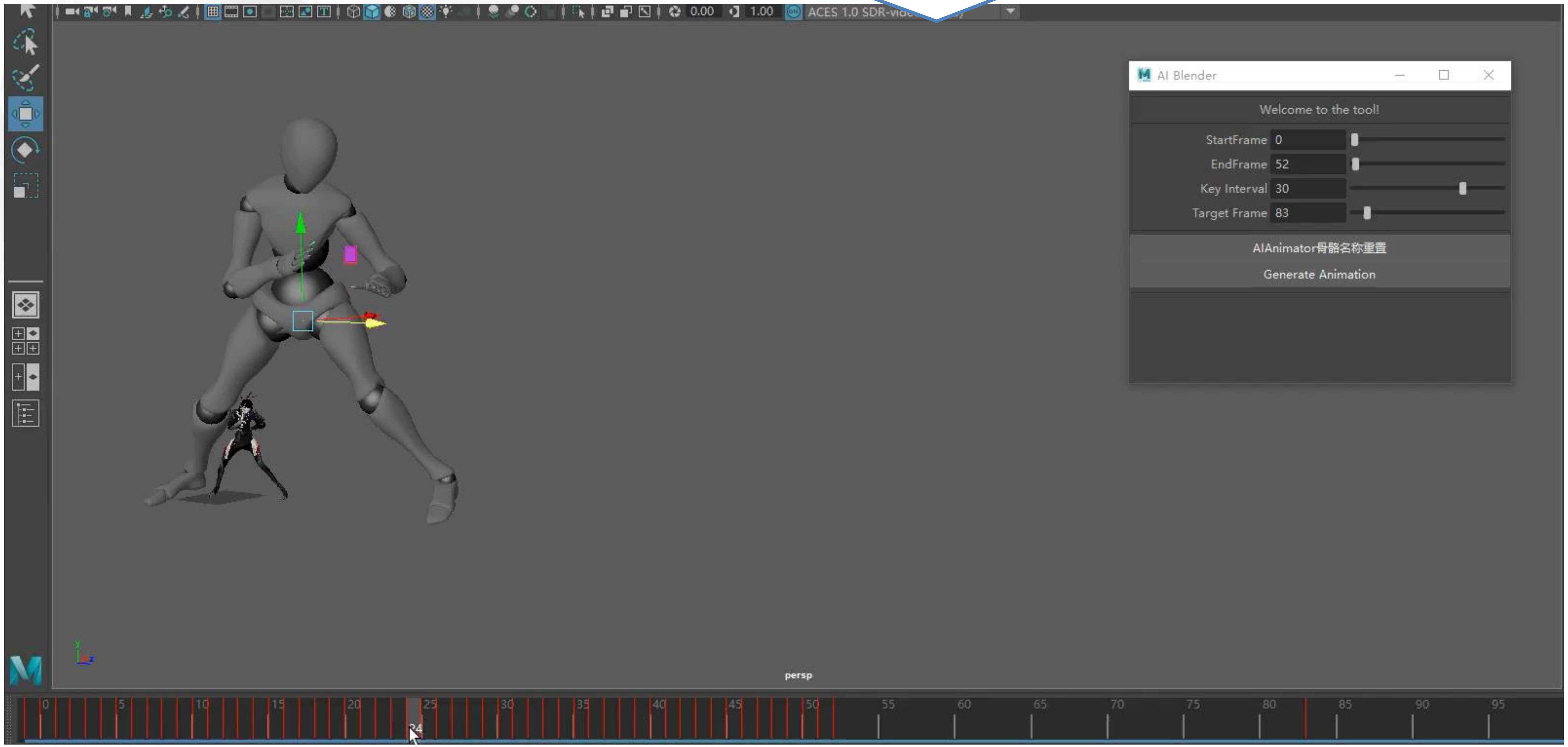
Polish and editing



Model from the Netease Game ***The World 3***

Maya plugin

Different missing frames



Model from the Netease Game ***The World 3***

THANKS FOR WATCHING

China: https://hr.game.163.com/recruit.html

Overseas: https://www.neteasegames.com/careers

Twitter Facebook Youtube

#GDC22
#NetEase Games