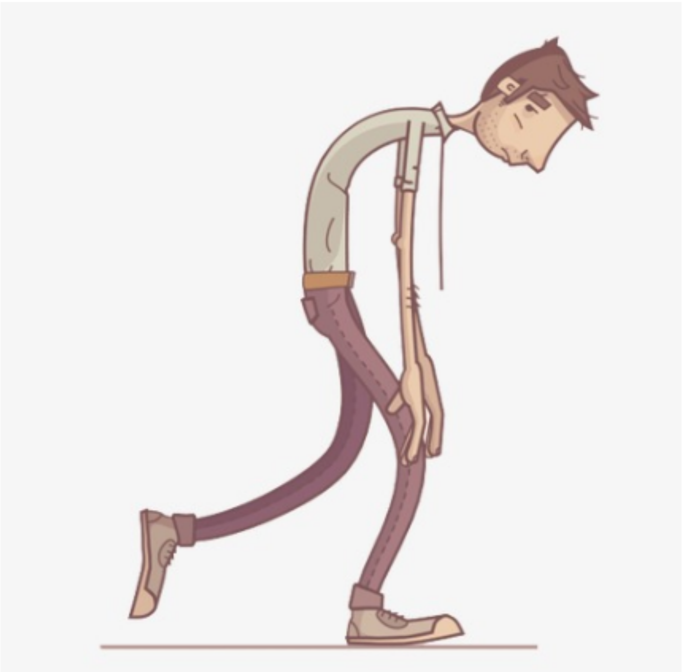
Efficient Neural Networks for Real-time Motion Style Transfer

Harrison Jesse Smith — University of California, Davis
Ph.D. Candidate

Chen Cao — Snap, Inc.
Michael Neff — University of California, Davis
Yingying Wang — Snap, Inc.

Motion Style

https://vimeo.com/26250920



https://tenor.com/view/walk-sad
-depressed-lonely-alone-gif-5011387

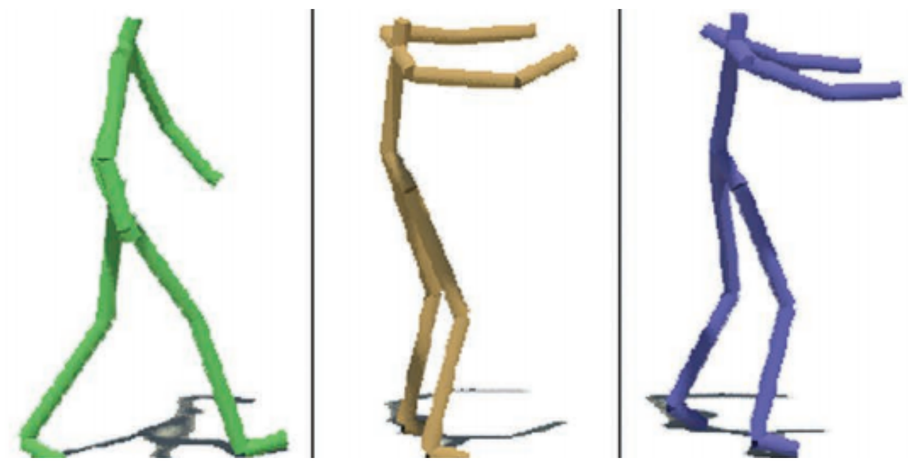
https://ramminanimation.tumblr.com/post/
136714032539/angry-walk-cycle

Goal

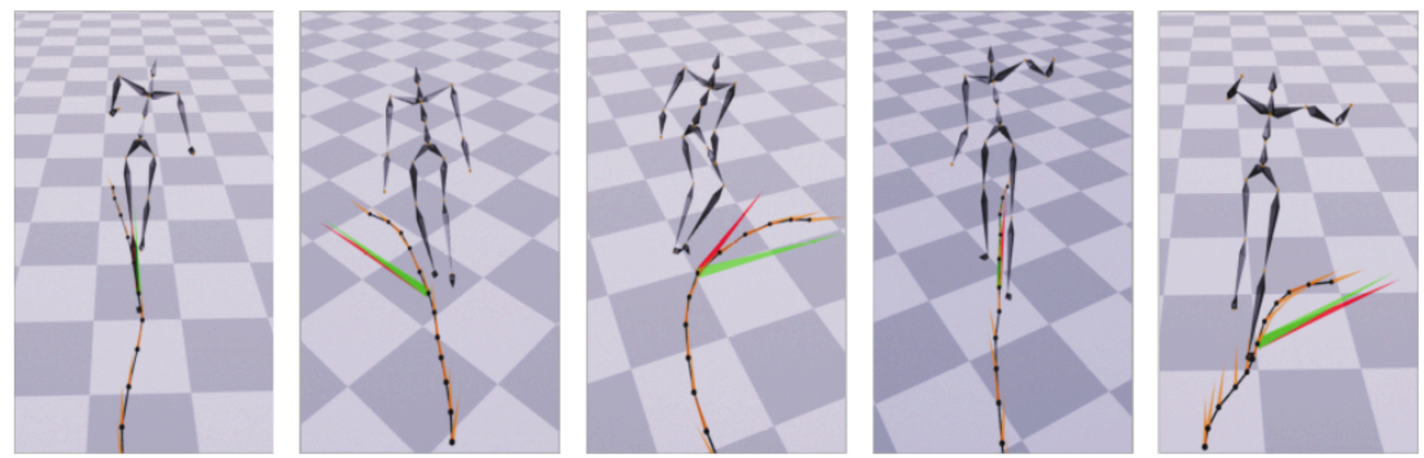
Style transfer that works:

- Unlabeled, heterogeneous motion sequences
- In real-time
- With user control
- In a computationally and memory efficient way

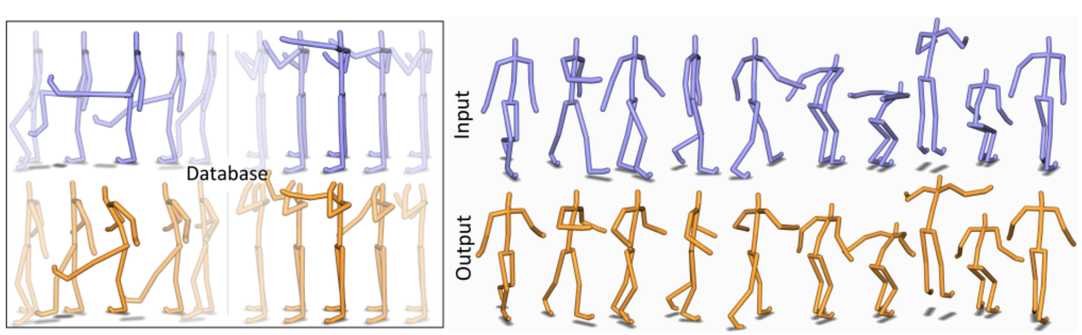
Existing Methods



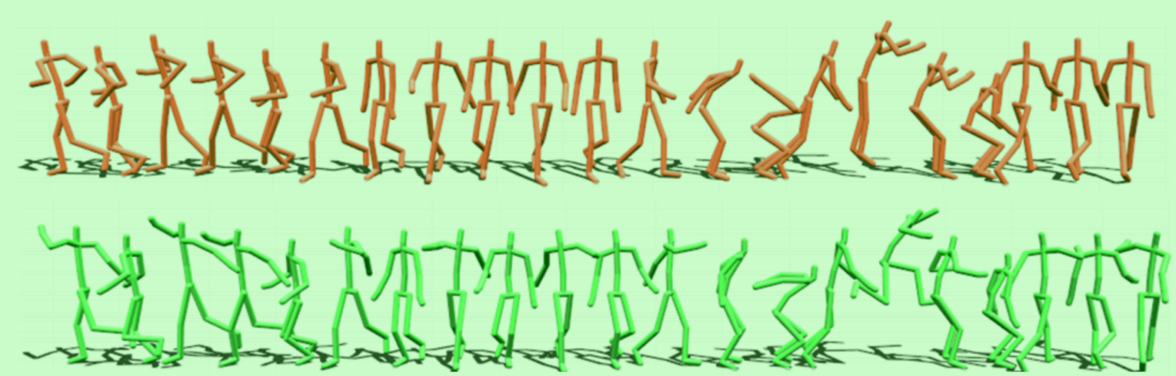
Holden, Daniel, et al. "Fast neural style transfer for motion data." *IEEE computer graphics and applications* 37.4 (2017): 42-49.



Mason, Ian, et al. "Few−shot Learning of Homogeneous Human Locomotion Styles." *Computer Graphics Forum*. Vol. 37. No. 7. 2018.

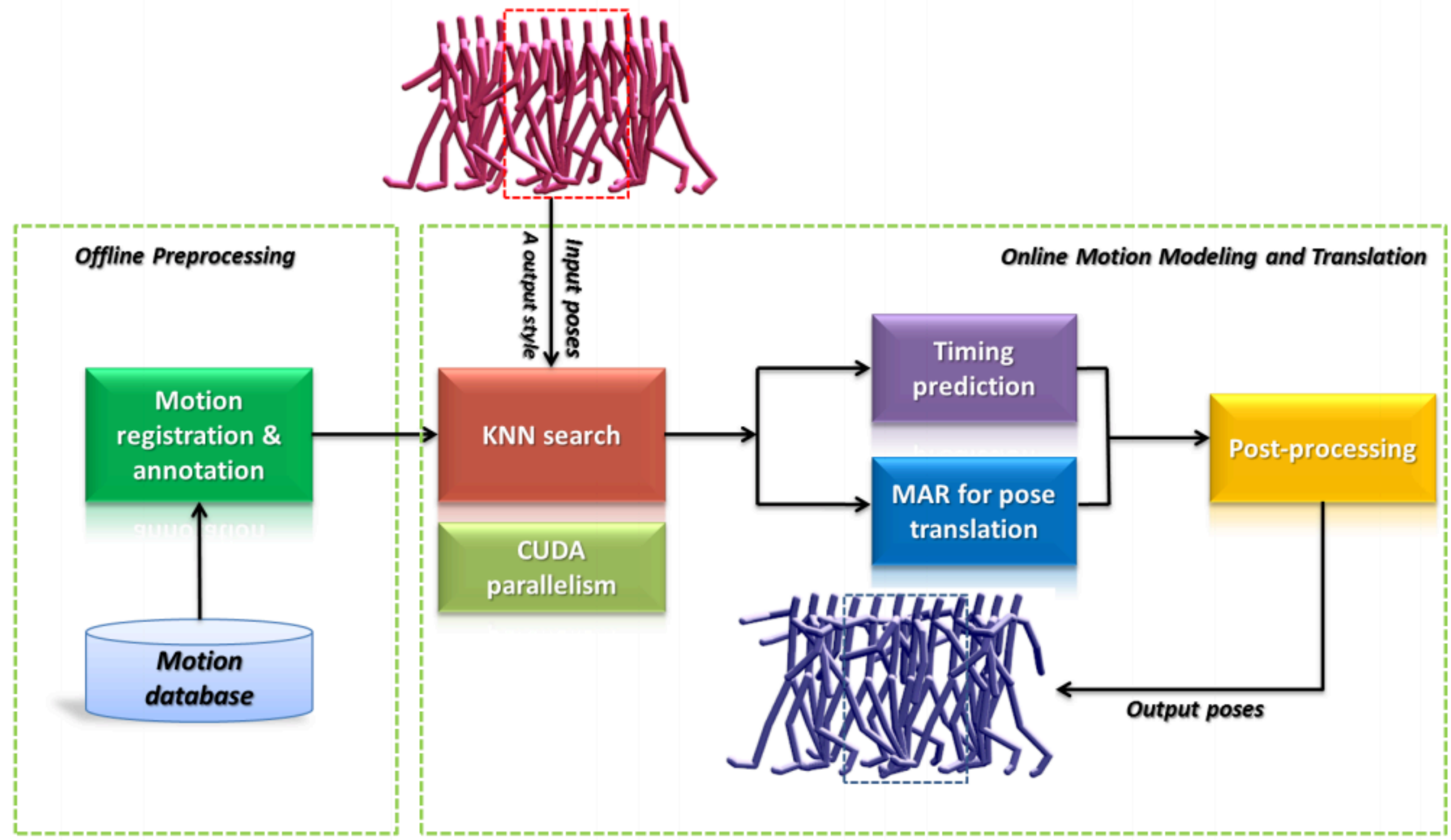


Yumer, M. Ersin, and Niloy J. Mitra. "Spectral style transfer for human motion between independent actions." *ACM Transactions on Graphics (TOG)* 35.4 (2016): 137.

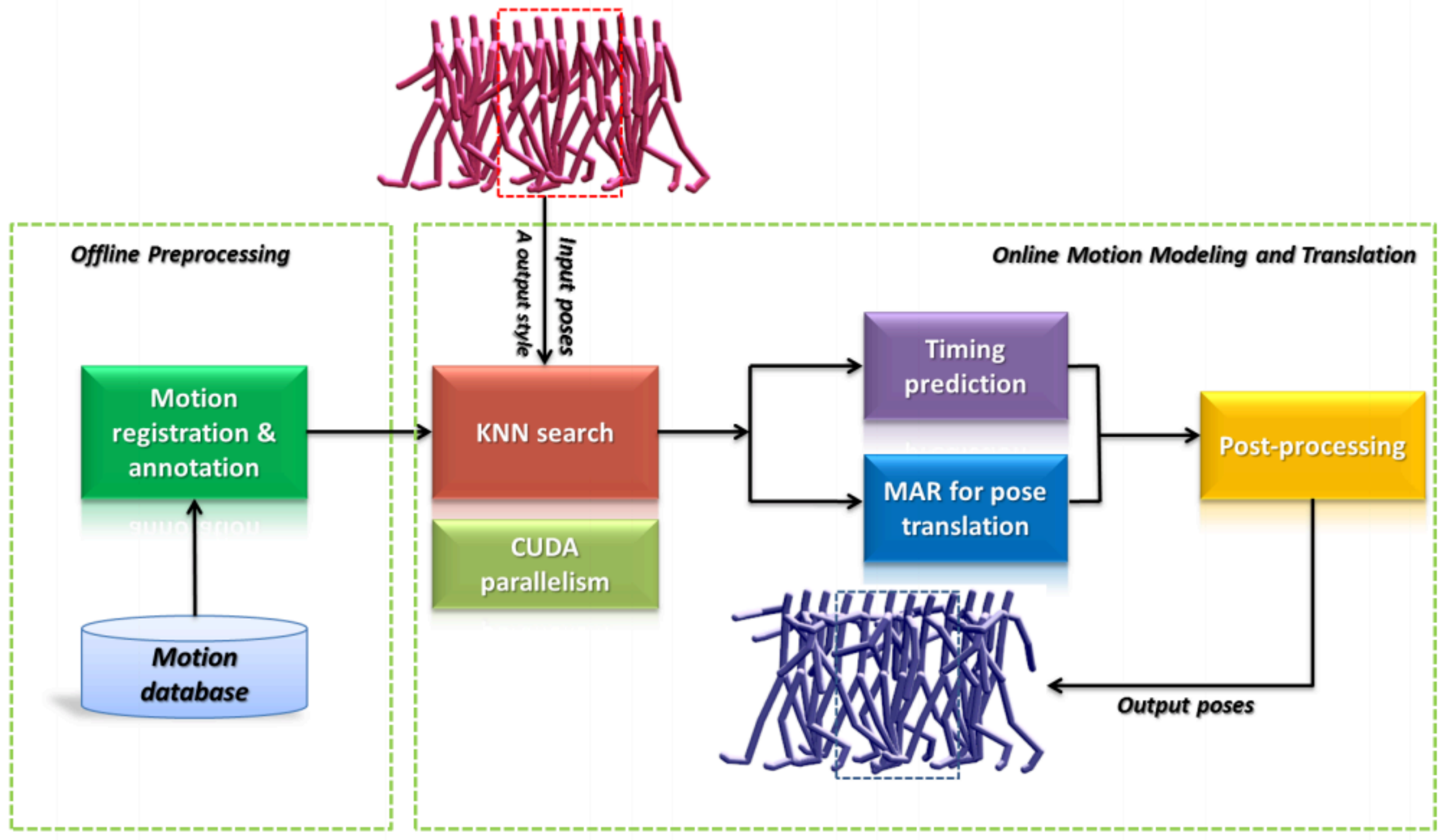


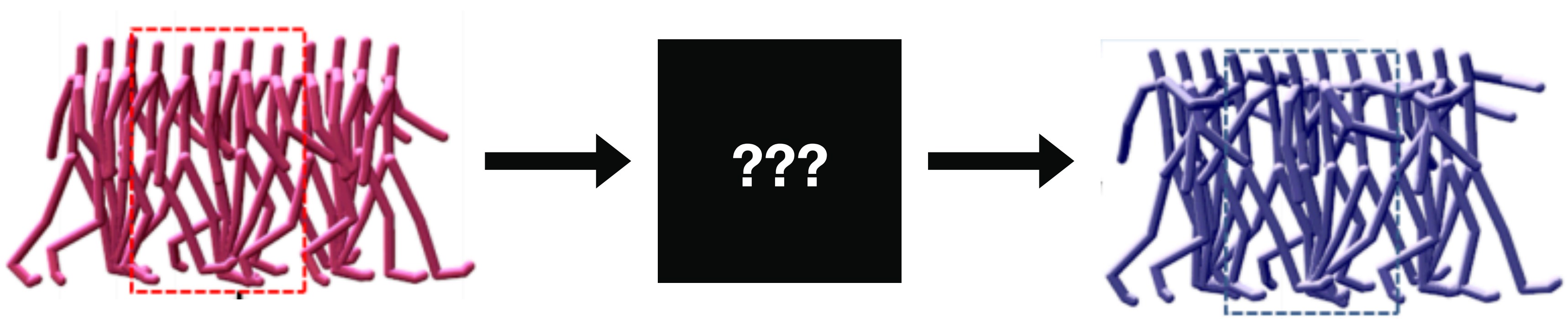
Xia, Shihong, et al. "Realtime style transfer for unlabeled heterogeneous human motion." *ACM Transactions on Graphics (TOG)* 34.4 (2015): 119.

Realtime Style Transfer for Unlabeled Heterogeneous Human Motion



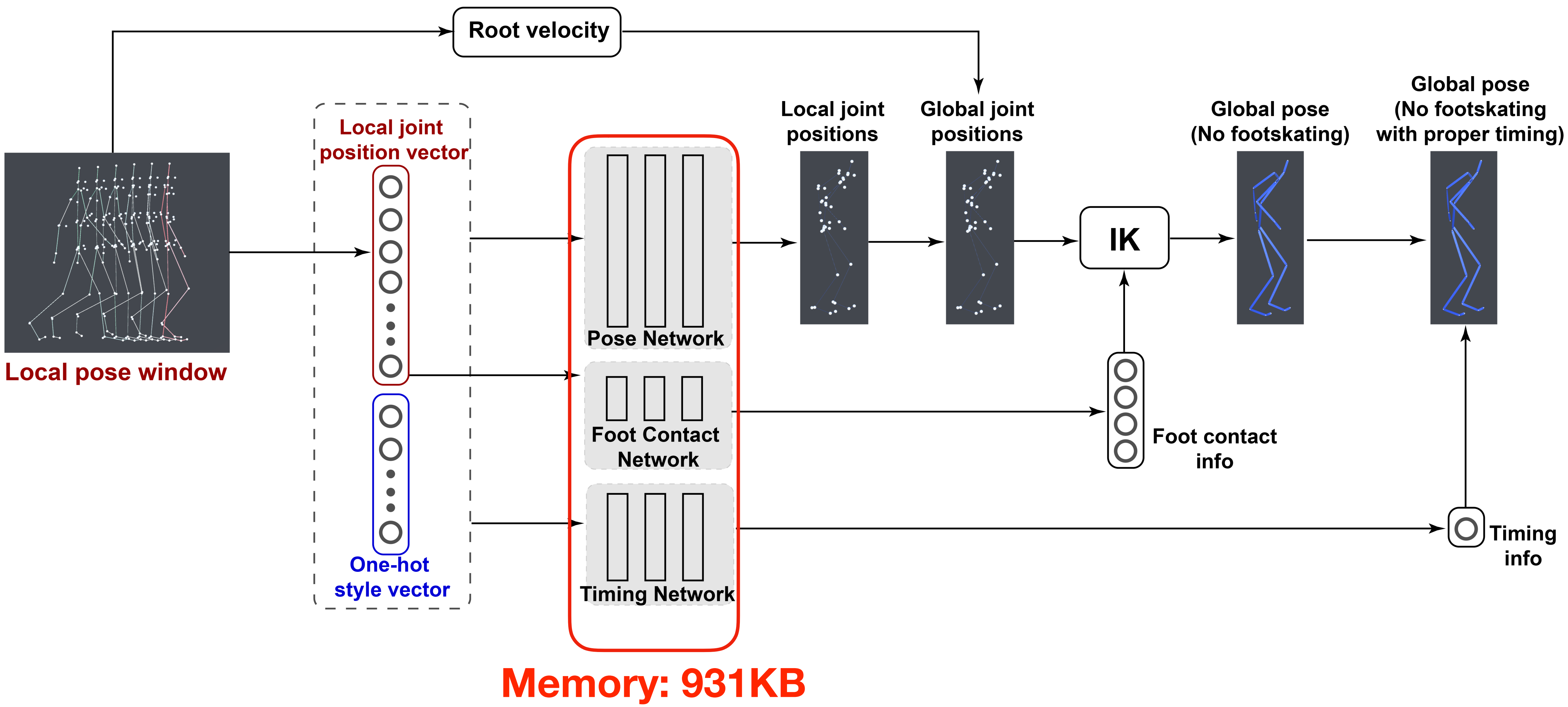
Xia, Shihong, et al. "Realtime style transfer for unlabeled heterogeneous human motion." *ACM Transactions on Graphics (TOG)* 34.4 (2015): 119.



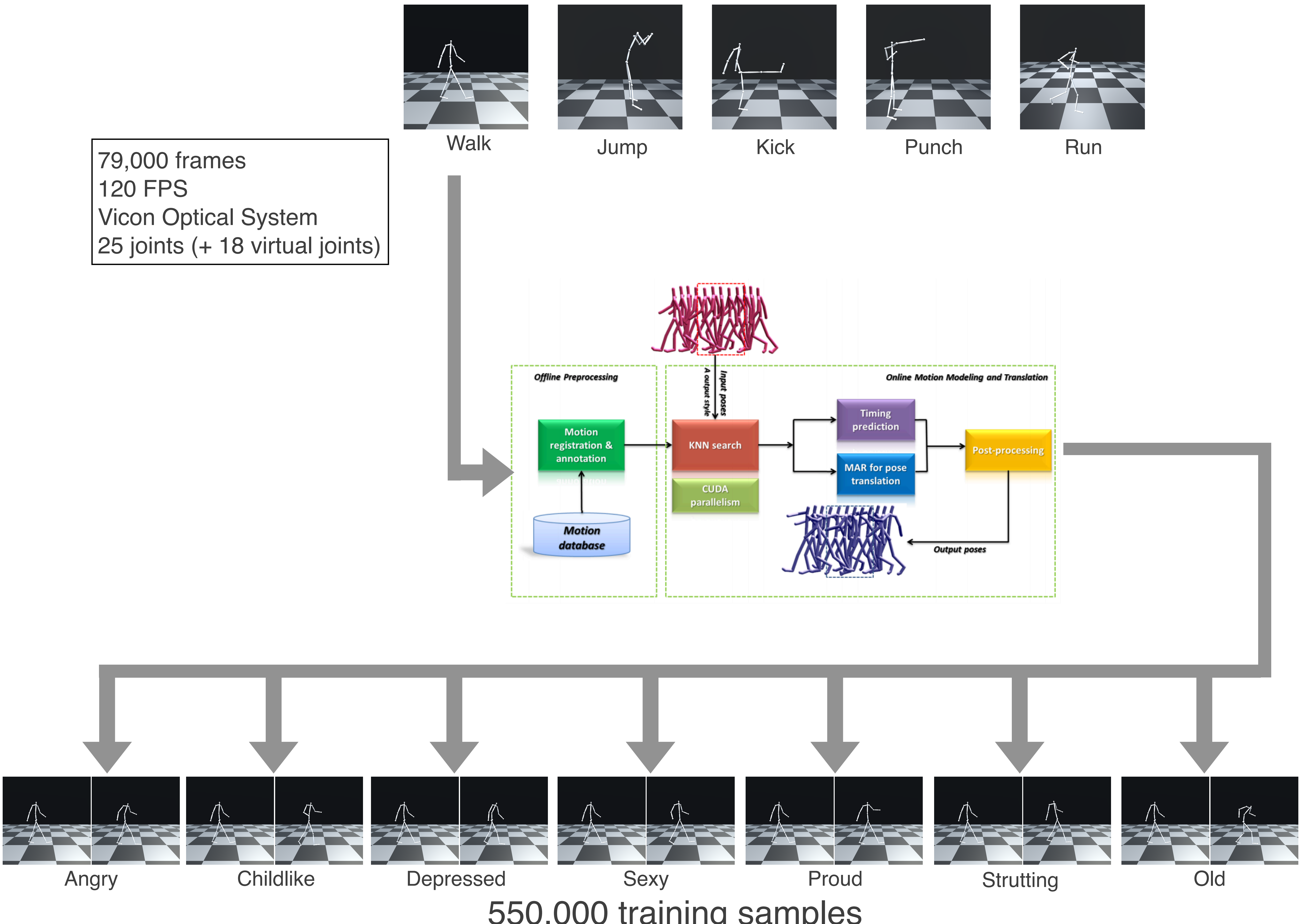


Xia, Shihong, et al. "Realtime style transfer for unlabeled heterogeneous human motion." *ACM Transactions on Graphics (TOG)* 34.4 (2015): 119.

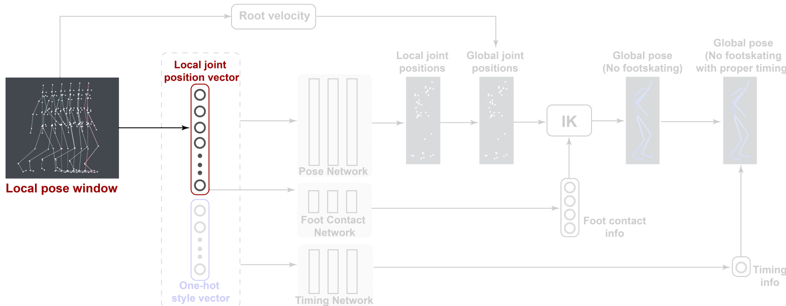
Overview



Root velocity
Local joint position vector
Local pose window
One-hot style vector
Pose Network
Foot Contact Network
Timing Network
Local joint positions
Global joint positions
IK
Global pose (No footskating)
Global pose (No footskating with proper timing)
Foot contact info
Timing info
Memory: 931KB



Walk
Jump
Kick
Punch
Run
79,000 frames
120 FPS
Vicon Optical System
25 joints (+ 18 virtual joints)
Offline Preprocessing
Motion registration & annotation
Motion database
Online Motion Modeling and Translation
A output style
Input poses
KNN search
CUDA parallelism
Timing prediction
MAR for pose translation
Post-processing
Output poses
Angry
Childlike
Depressed
Sexy
Proud
Strutting
Old
550,000 training samples



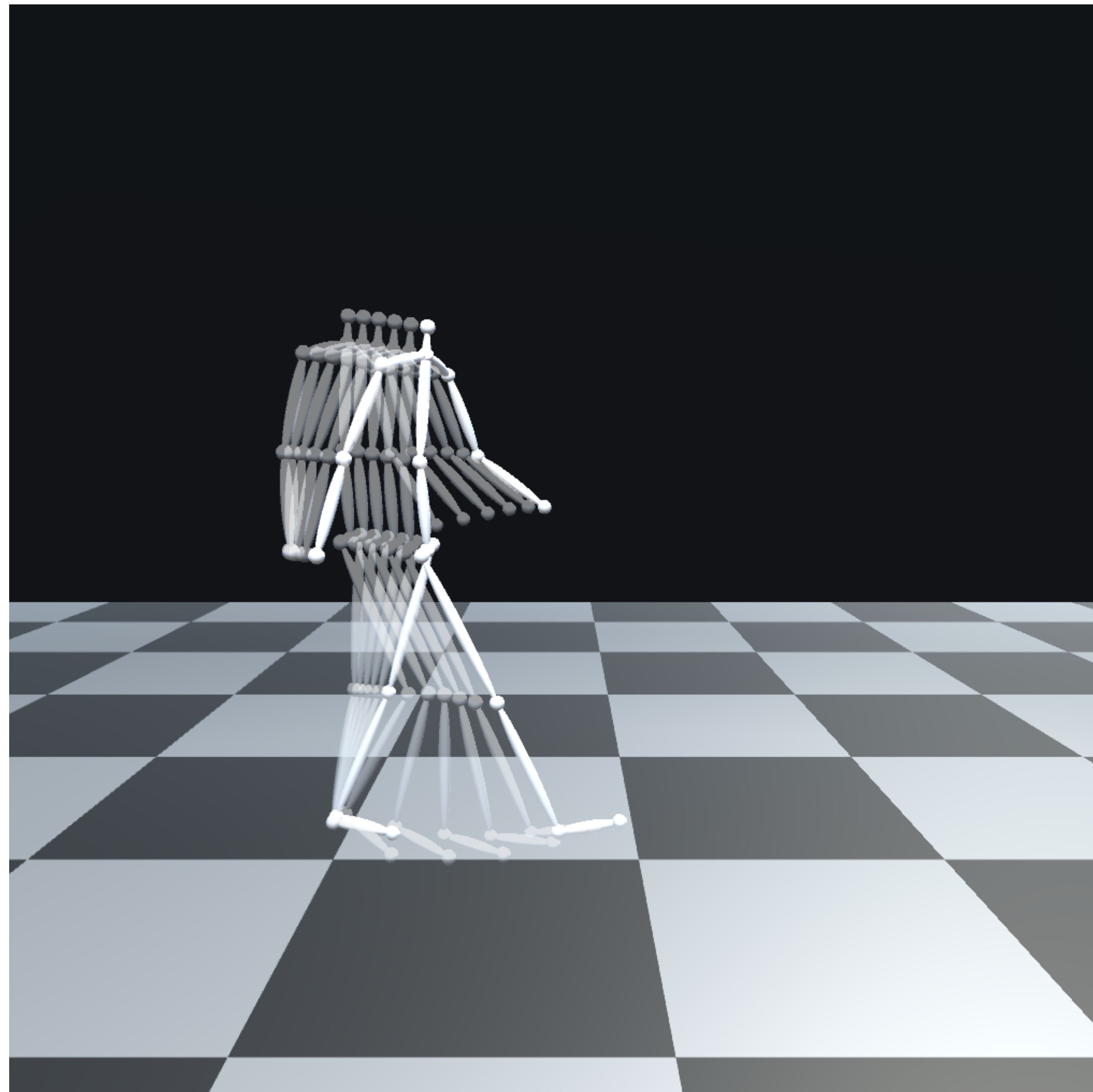
Input Pose Preprocessing

Input pose translated above origin

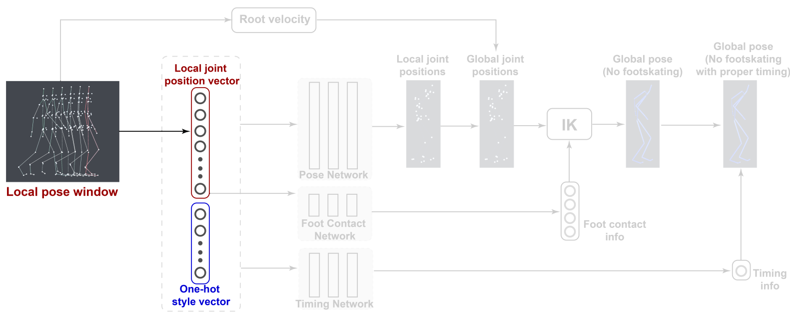
Rotated to face positive z axis

Concatenated with 5 previous poses, covering last 0.25 seconds

Gaussian normalization



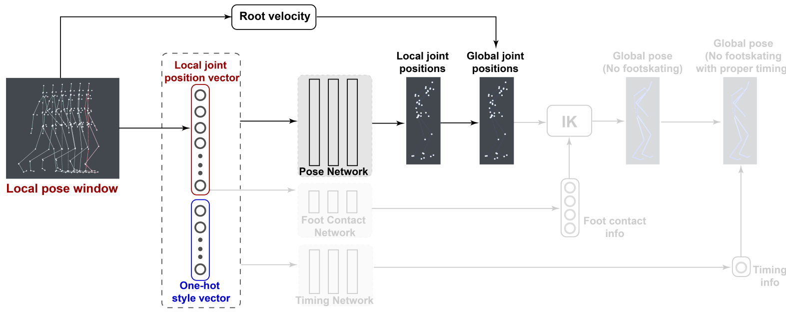
Input Style Vector



Old Style Vector

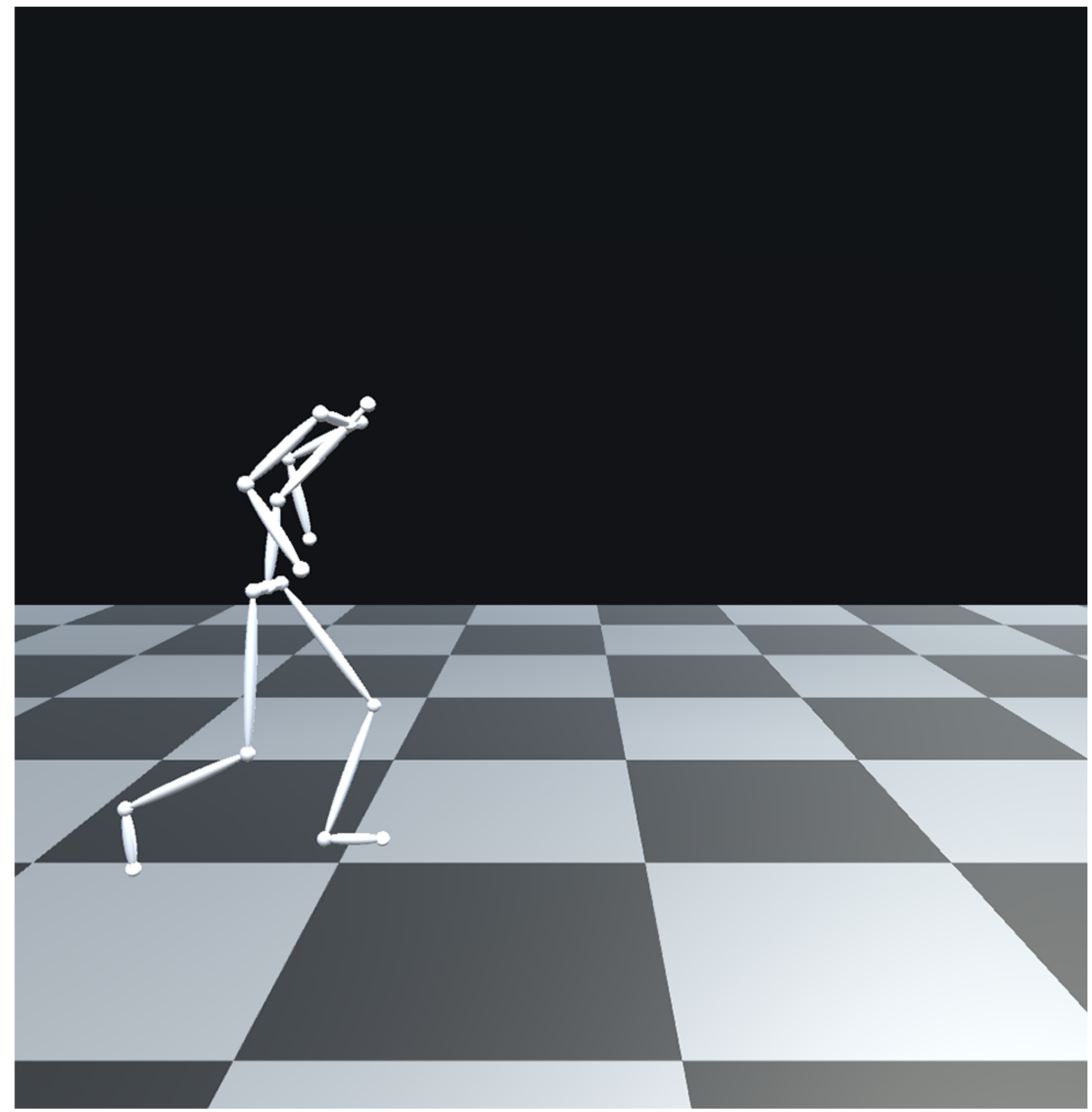
Angry Style Vector

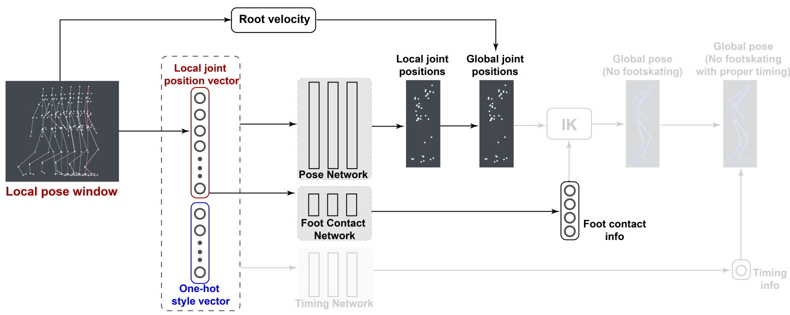
Strutting Style Vector



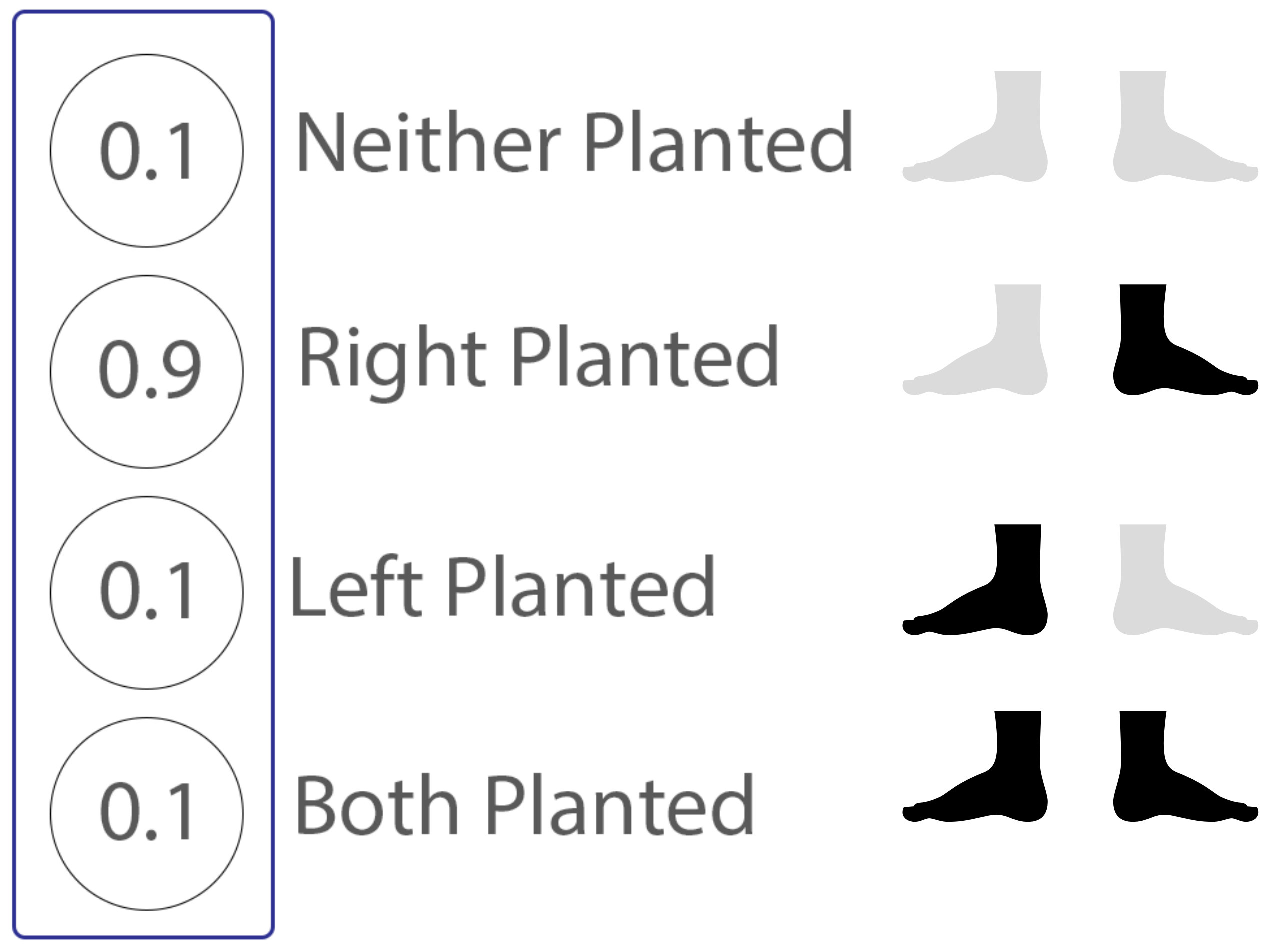
Root velocity
Local joint position vector
Local joint positions
Global joint positions
Global pose (No footskating)
Global pose (No footskating with proper timing)
IK
Local pose window
Pose Network
Foot Contact Network
Foot contact info
One-hot style vector
Timing Network
Timing info

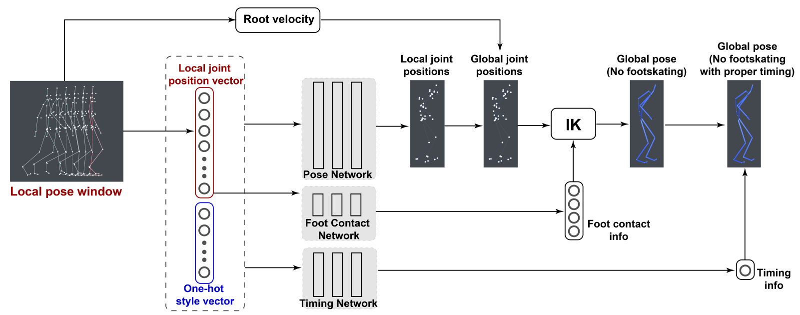
Output Pose



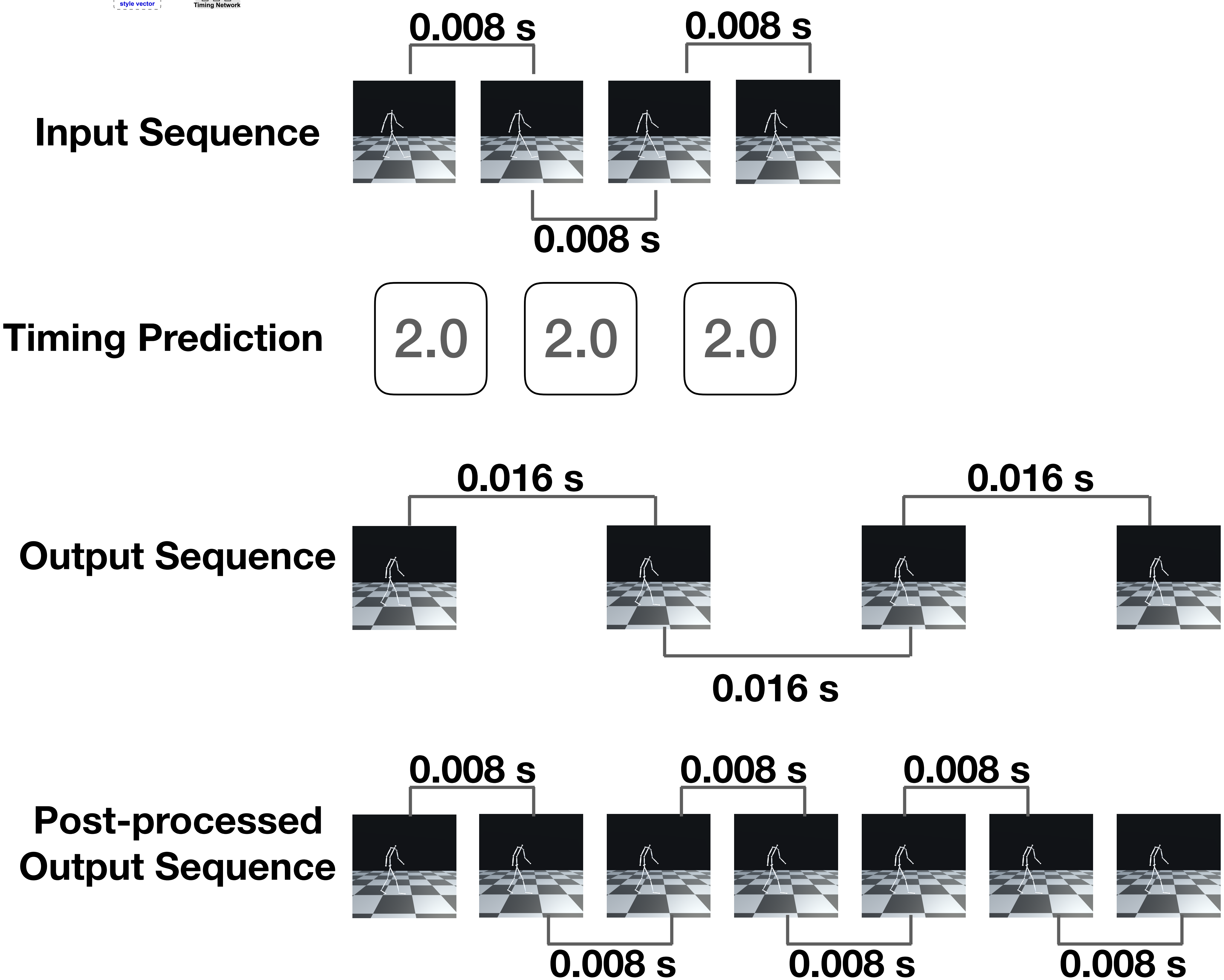


Output Foot Contact



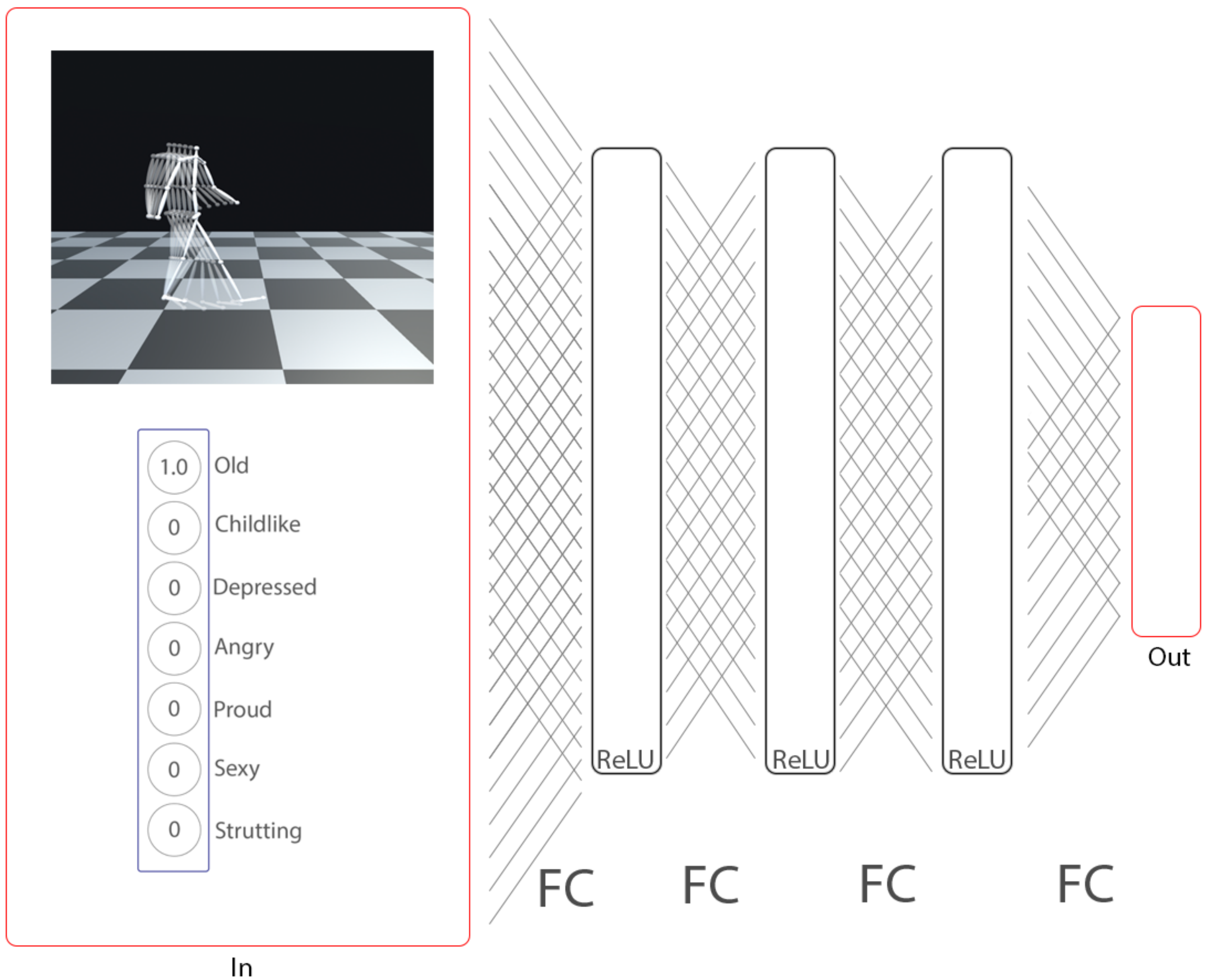


Timing Information



Networks and Training

3x



Pose Network

128 units per layer

Mean Squared Error loss

Timing Network

64 units per layer

Mean Squared Error loss

Foot Contact Network

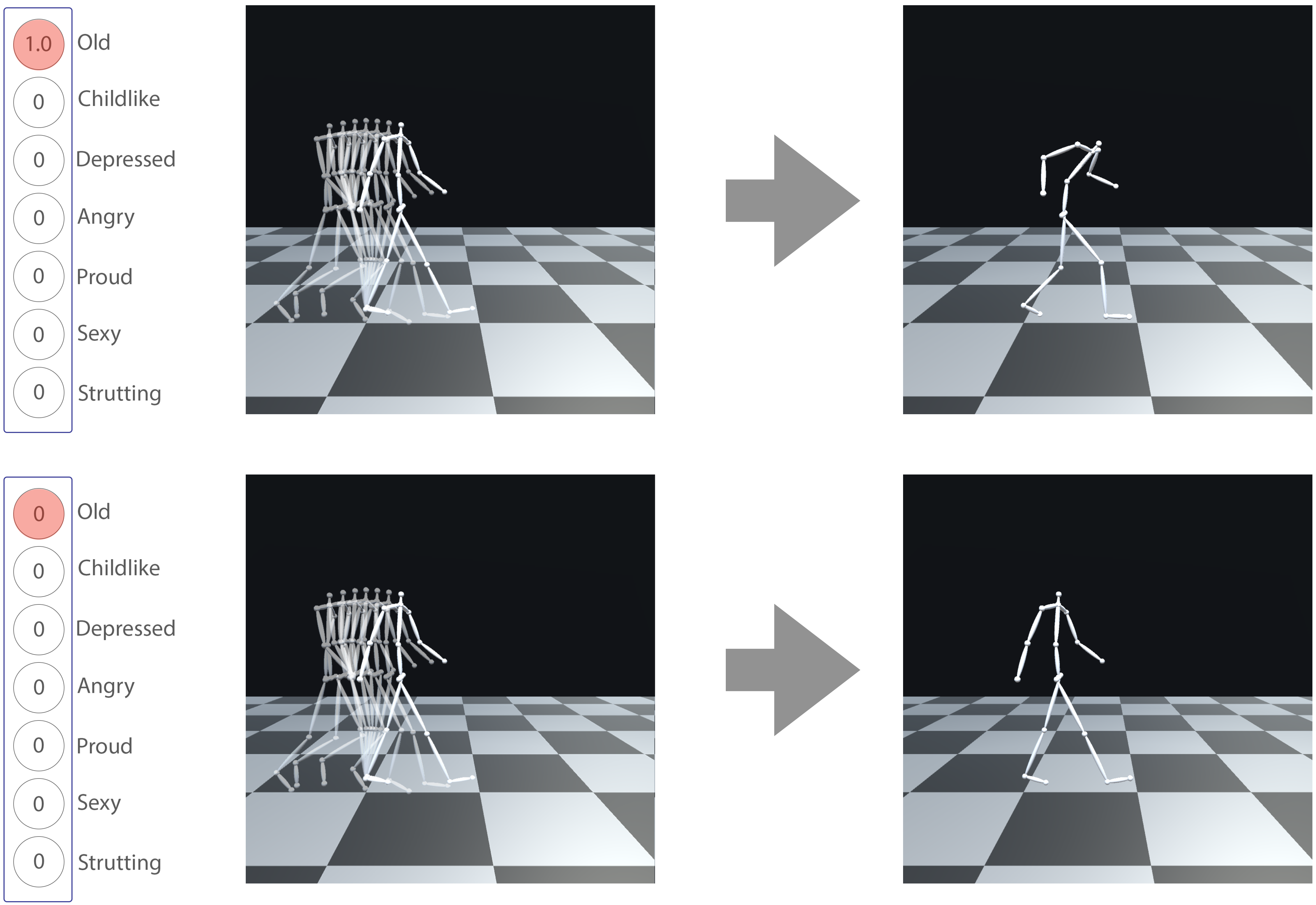
32 units per layer

Binary Cross-Entropy loss

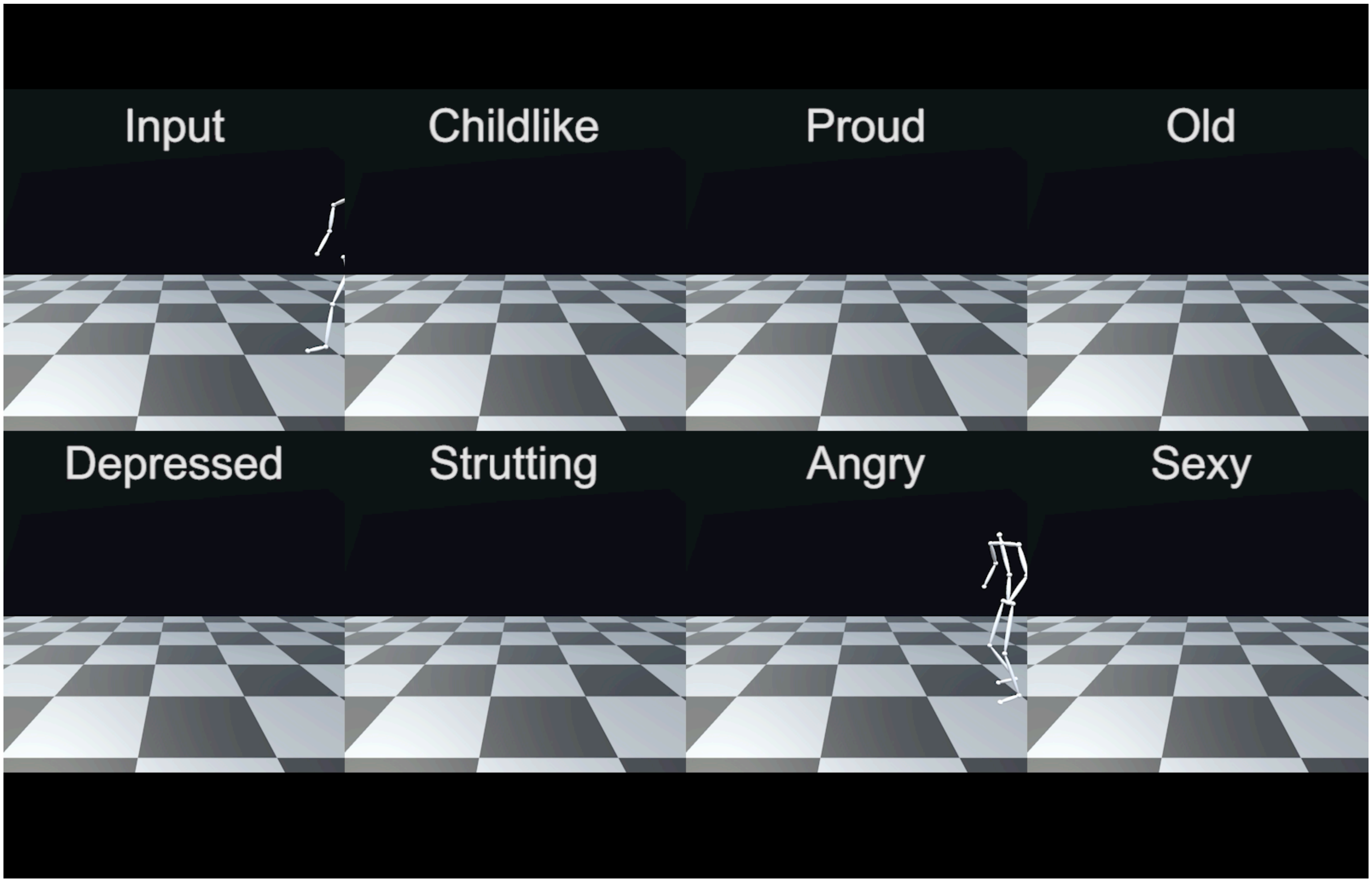
Optimizer: ADAM
Learning rate: 0.01
Epochs: 20
Minibatch Size: 64

Implemented in Pytorch
Training Time: 3 hours
i7 3.5GHz 4-core
with Geoforce GTX1070

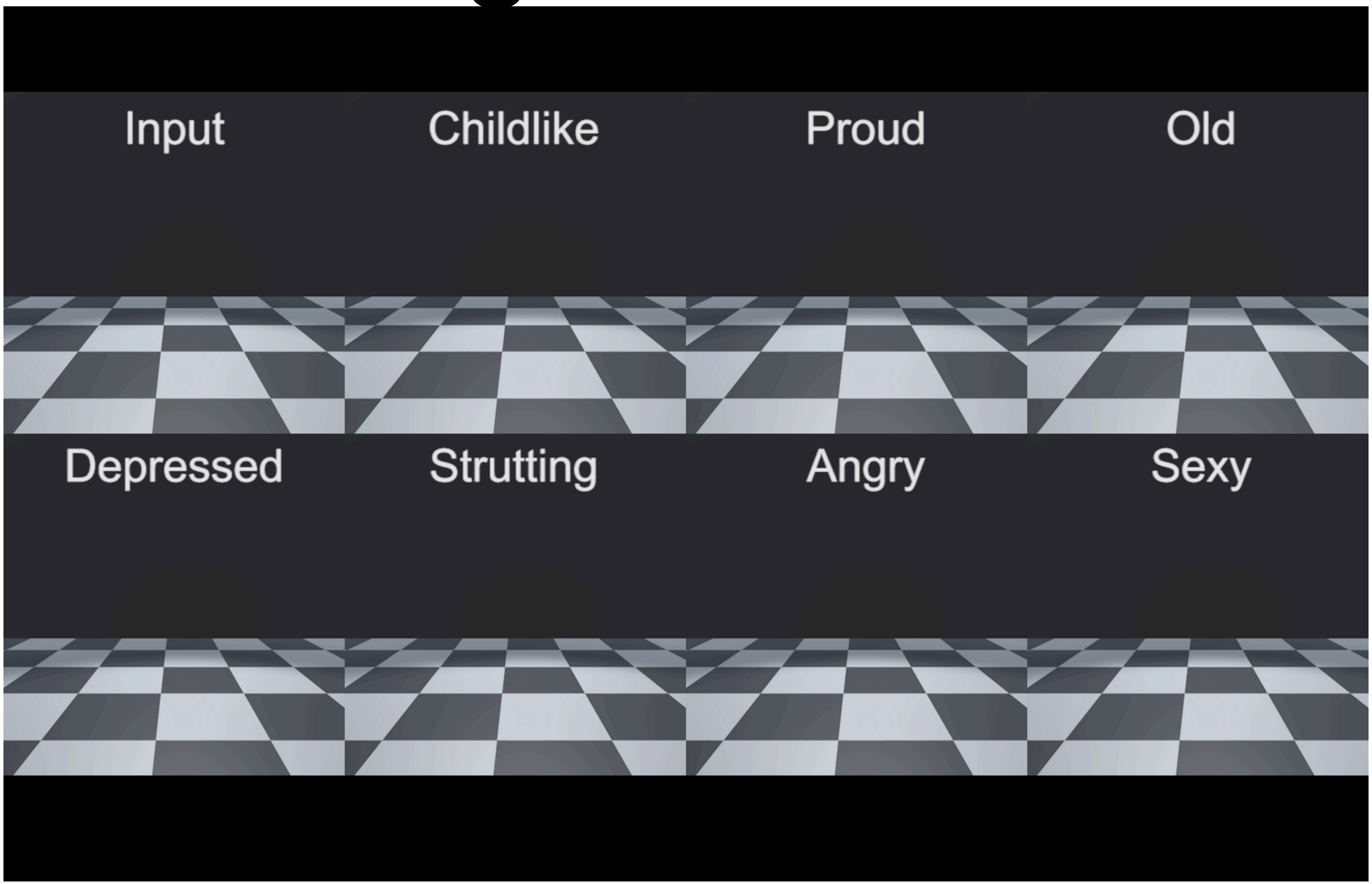
Predicting Input Poses



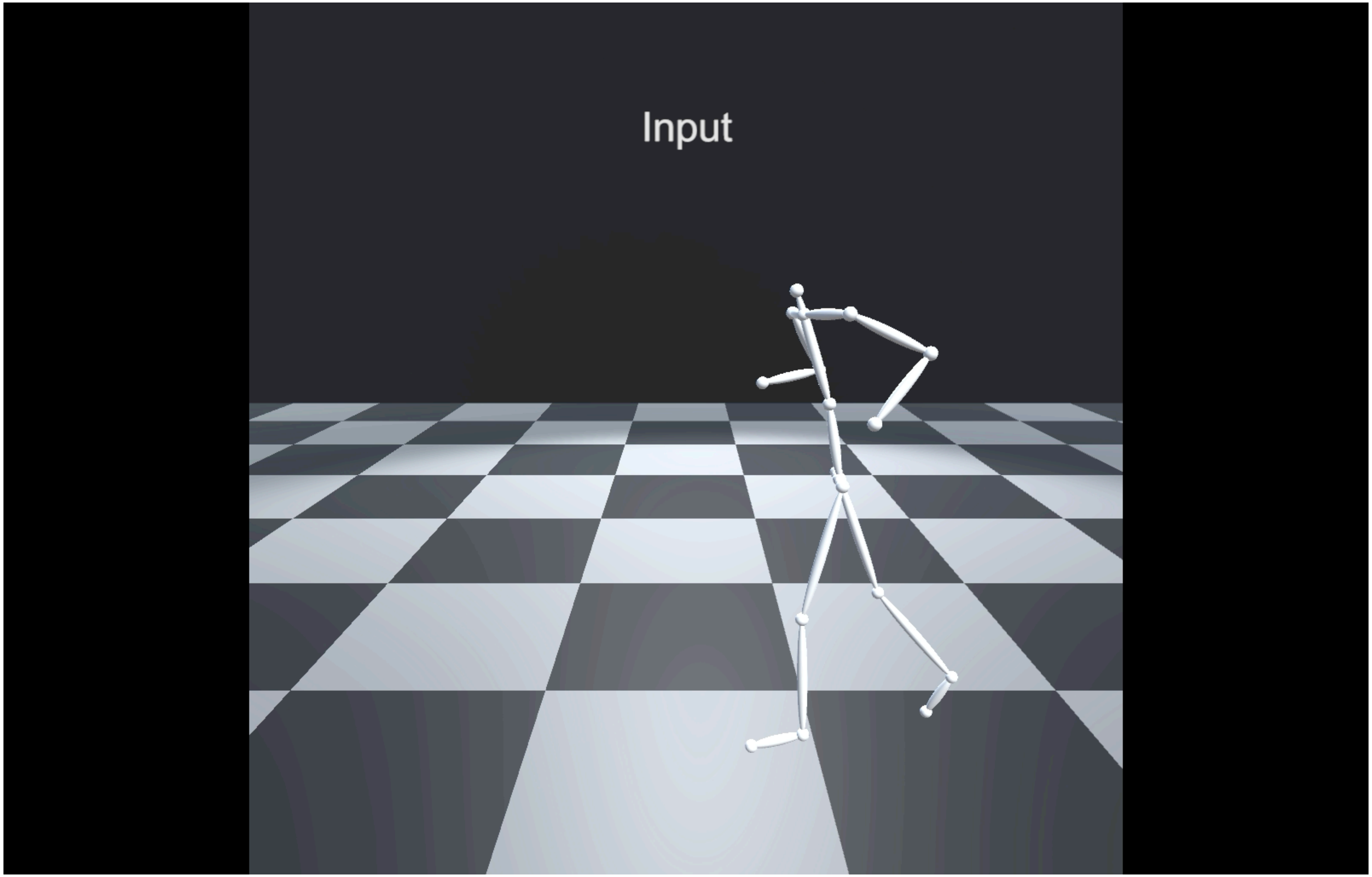
Results - Locomotion



Results - Heterogeneous Motion



Results - Comparison with Original

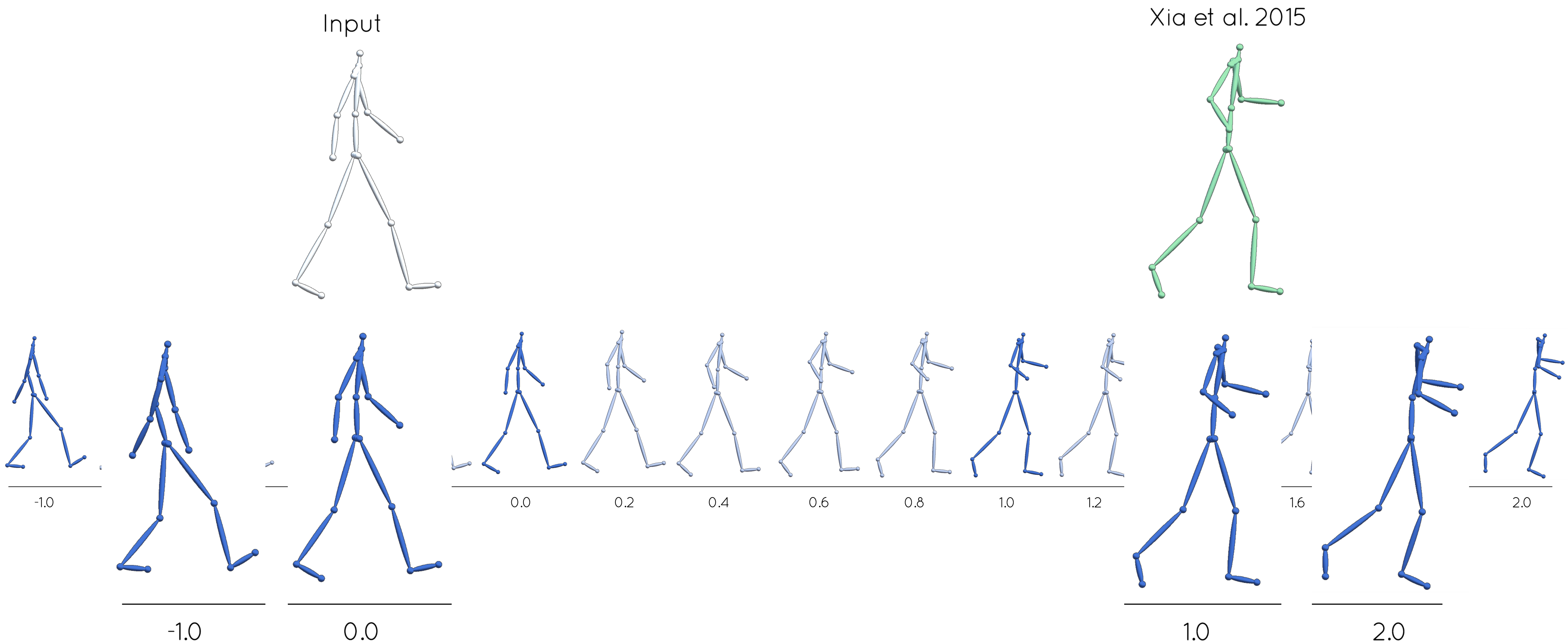


Run-time Performance

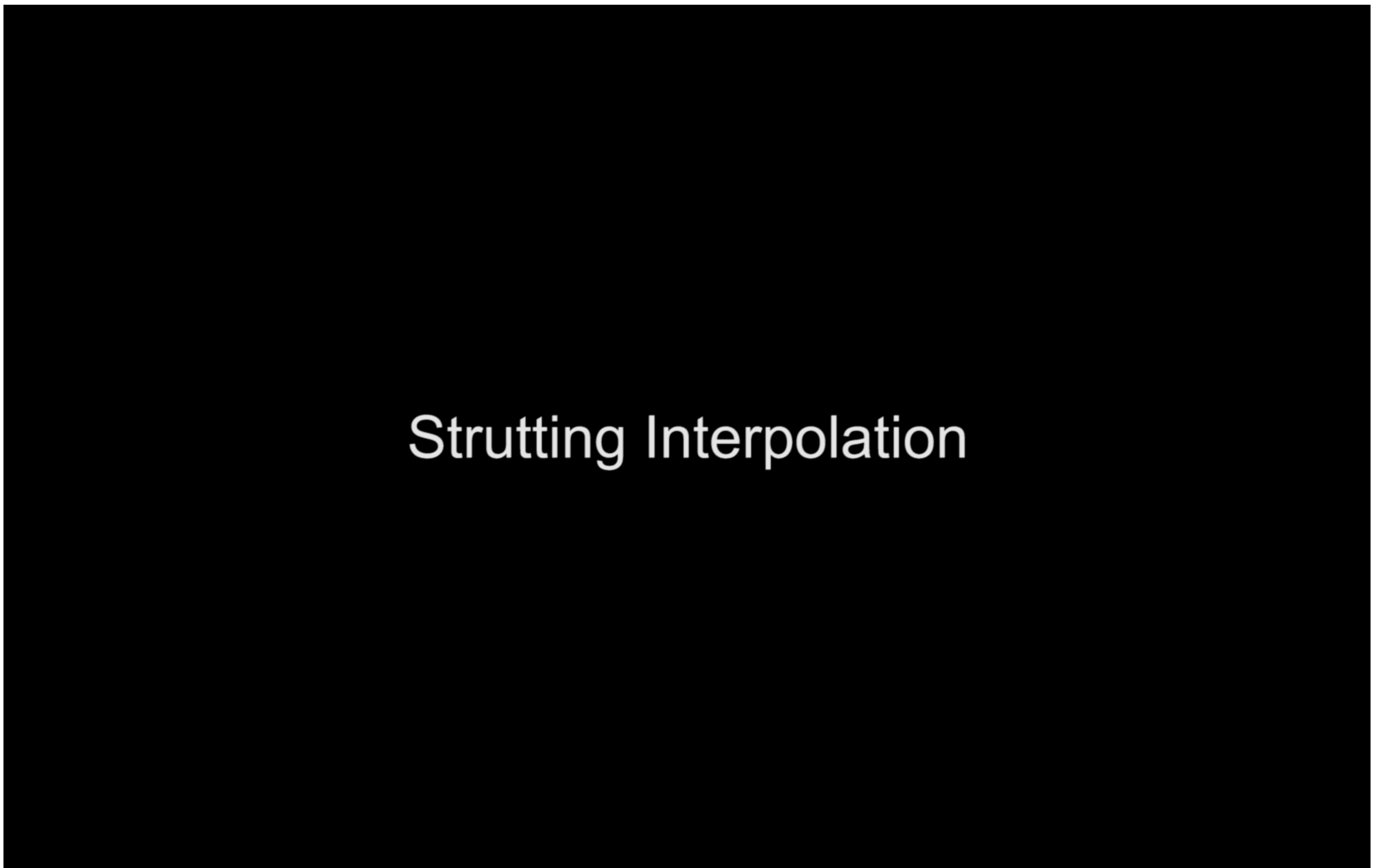
	Memory Footprint	FPS Achieved	
[Xia et al. 2015]	290 MB	55*	*CUDA + GPU
[Yumer and Mitra 2016]	Not Reported	50	
Our Method (PC)	931 KB	115	i7 3.5GHz CPU

	Memory Footprint	Neural Network Forward Pass	
Our Method (PC)	931 KB	0.0008s	i7 3.5GHz CPU
Our Method (Mobile)	931 KB	0.002s	iPhone 7 Plus
[Holden et al. 2017b] Cubic	10MB	0.0018s	
[Holden et al. 2017b] Constant	125MB	0.0008s	
[Mason et al. 2018]	126KB*	0.0011s	
[Zhang et al. 2018]	22MB	0.002s	

Interpolation and Extrapolation Example: Childlike



Strutting Interpolation

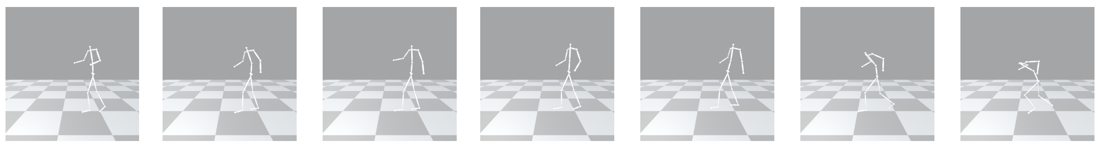


Blending Childlike to Old

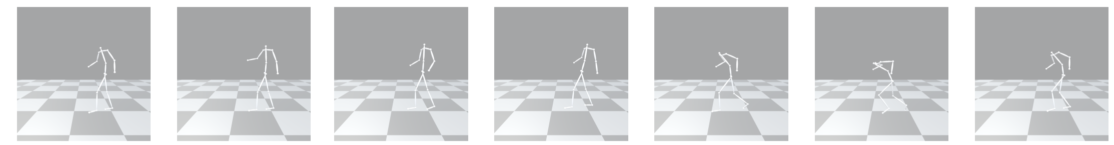
Combining Old and Strutting

User Study: Style Recognition

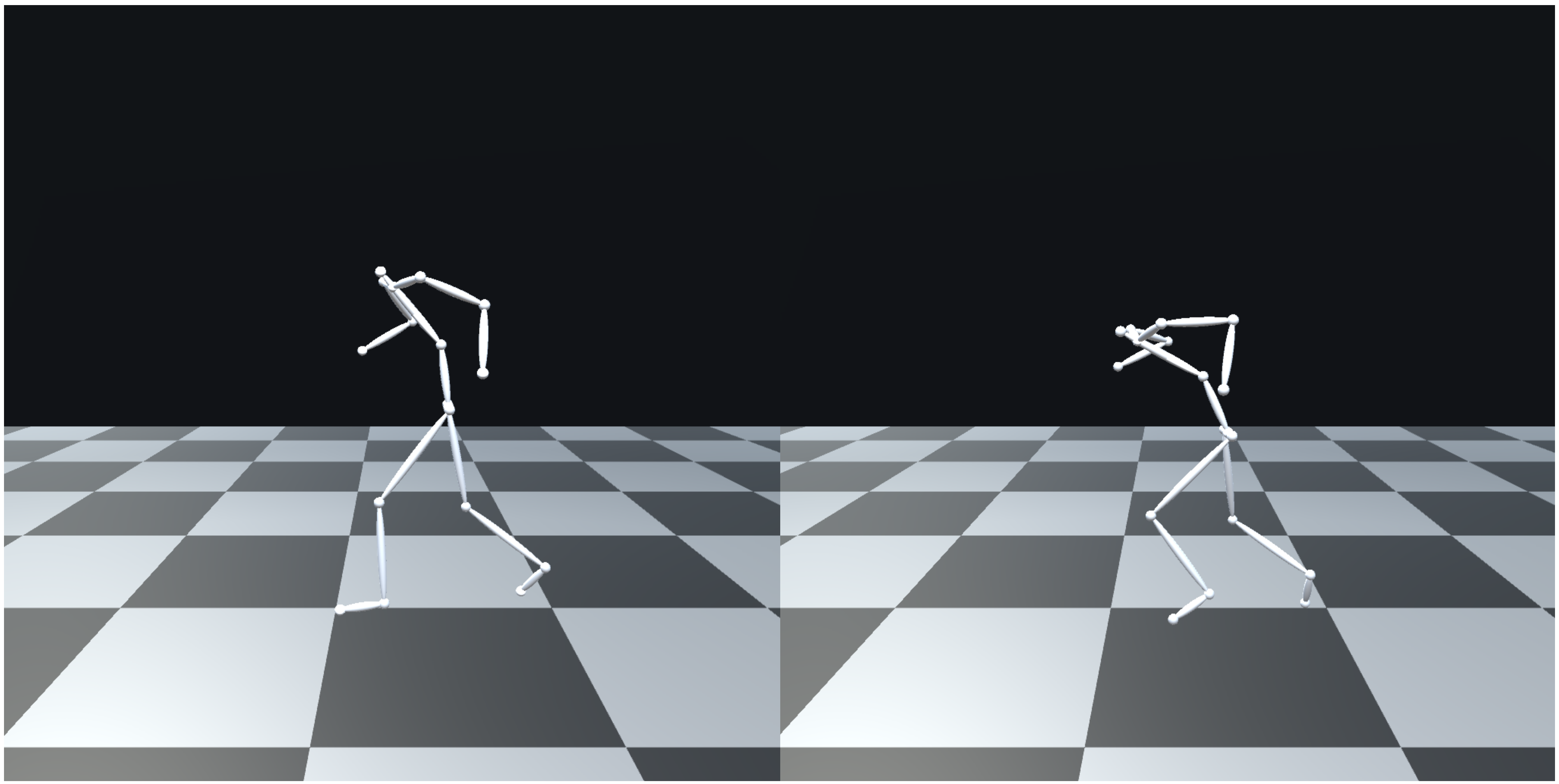
Ours



Xia’s



User Study: Exaggeration

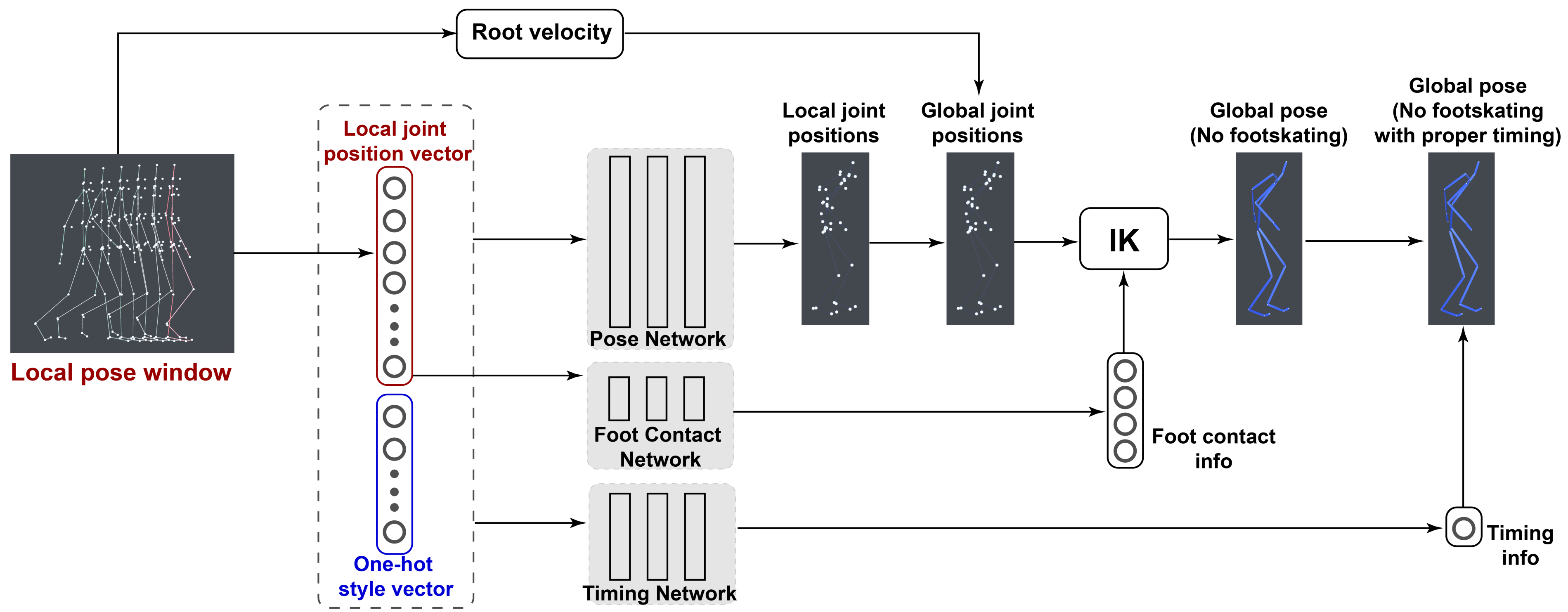


“Which Video is More Expressive of an ‘Old’ Style?”

User Study: Extrapolation

Style	Ours	Xia's	Don't Know
Angry	27 (63%)	15 (34%)	1 (3%)
Childlike	33 (77%)	6 (14%)	4 (9%)
Depressed	31 (72%)	12 (28%)	0
Old	37 (86%)	6 (14%)	0
Proud	3 (7%)	40 (93%)	0
Sexy	17 (39%)	25 (58%)	1 (3%)
Strutting	27 (63%)	15 (34%)	1 (3%)

In Summary



<u>A method for style transfer that works:</u>

Unlabeled, heterogeneous motion sequences

In a computationally and memory efficient way

In real-time

With user control

Questions?

Efficient Neural Networks for Real-time Motion Style Transfer

Harrison Jesse Smith — University of California, Davis
Ph.D. Candidate

Chen Cao — Snap, Inc.
Michael Neff — University of California, Davis
Yingying Wang — Snap, Inc.