Generative Human Motion Stylization in Latent Space

ICLR 2024

Chuan Guo*, Yuxuan Mu*, Xinxin Zuo, Peng Dai, Youliang Yan,
Juwei Lu, Li Cheng

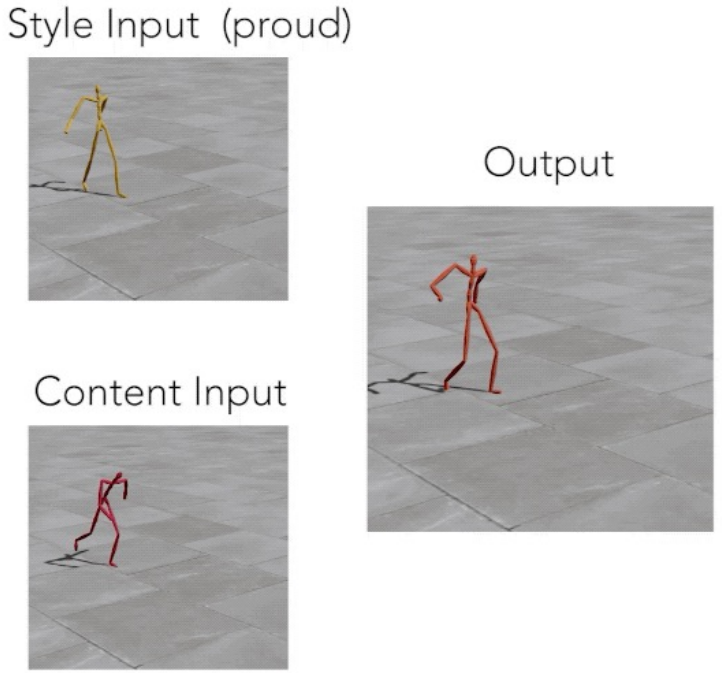
https://yxmu.foo/GenMoStyle/

Background

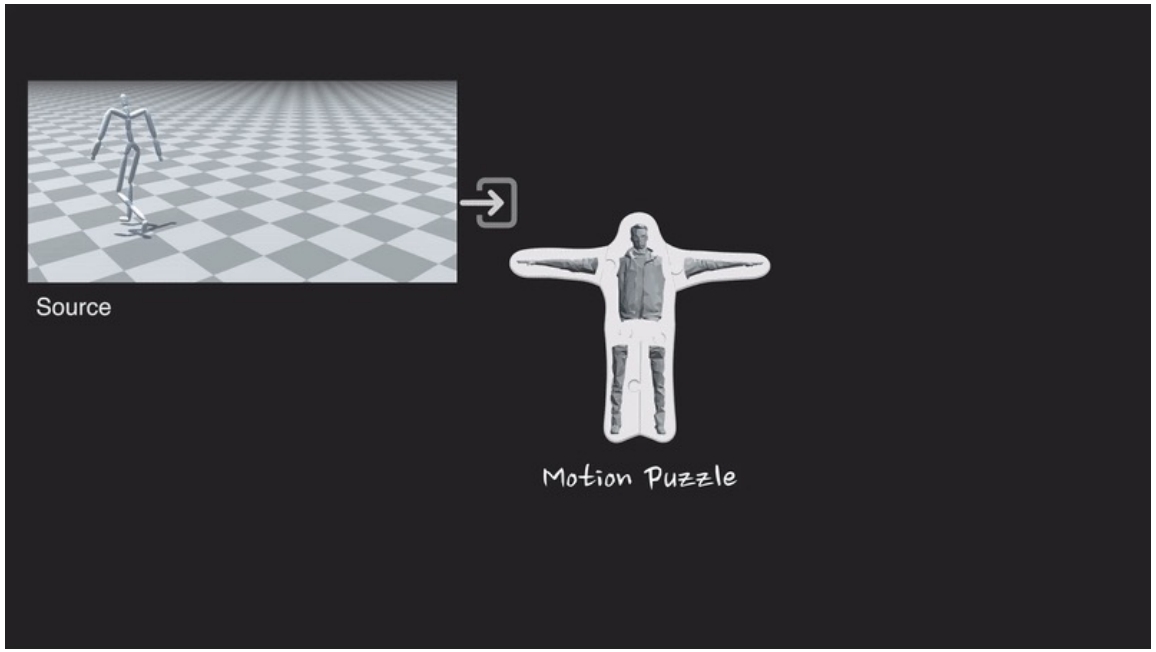
Existing motion stylization frameworks are limited in flexibility.

In particular:

1. Deterministic stylization
   - Given style, there is only way to stylize a motion.
2. Rigid designs
   - Either motion input, or label input.
3. Mandatory style input
   - Must explicitly specify style.



Aberman et al. 2020



Jang et al. 2022

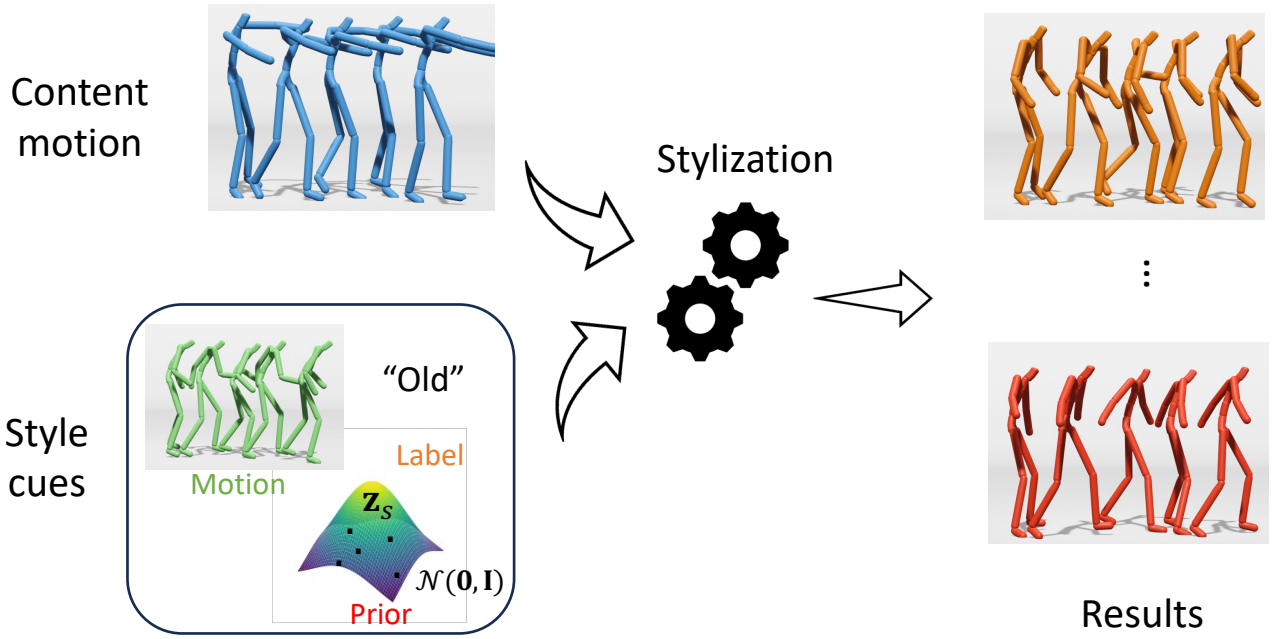
Goal

We propose a generative motion stylization framework.

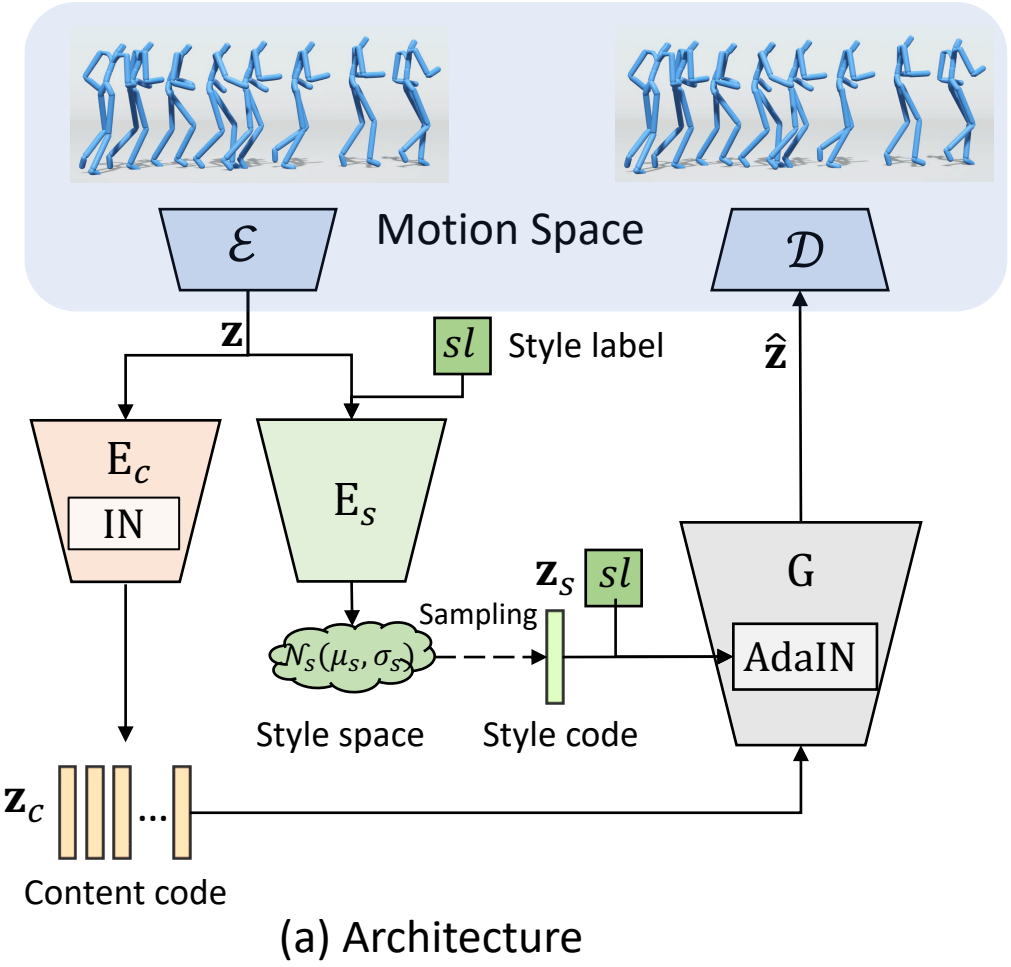
A motion can be stylized using the style cues from either of the following:

- ❑ Reference motion
- ❑ Style label
- ❑ Learned style prior

Moreover, we found stylization in latent space is more efficient than in pose space.



Approach

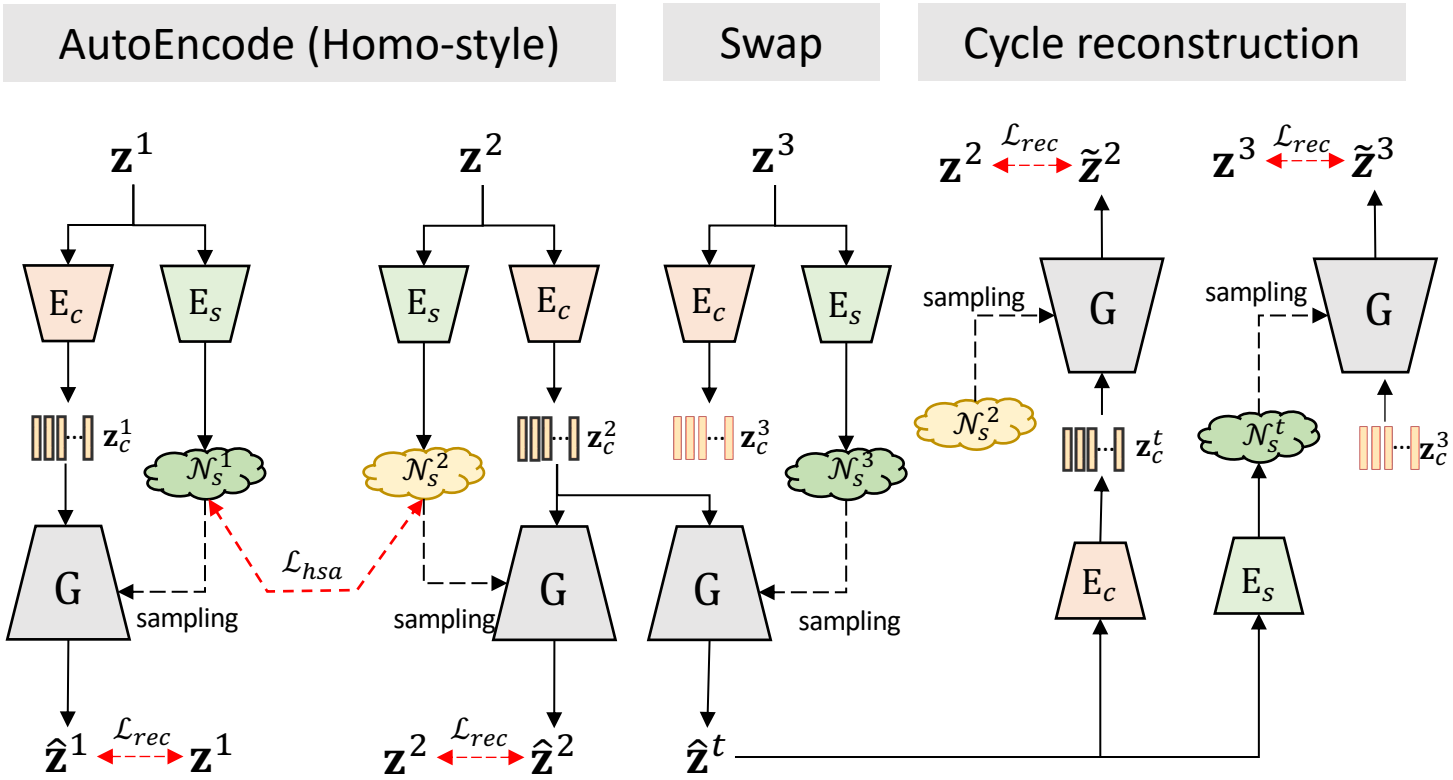


(a) Architecture

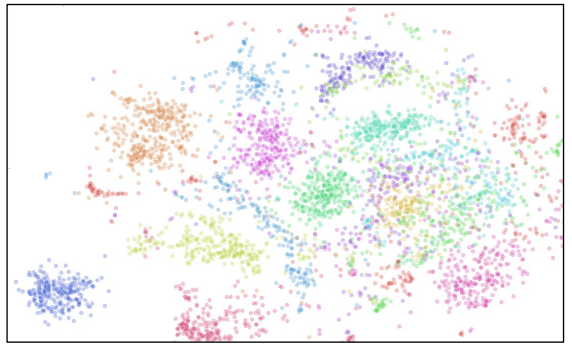
Model architecture:

- A pre-trained motion autoencoder.
- A content encoder.
- A style encoder.
- A motion generator.

Approach



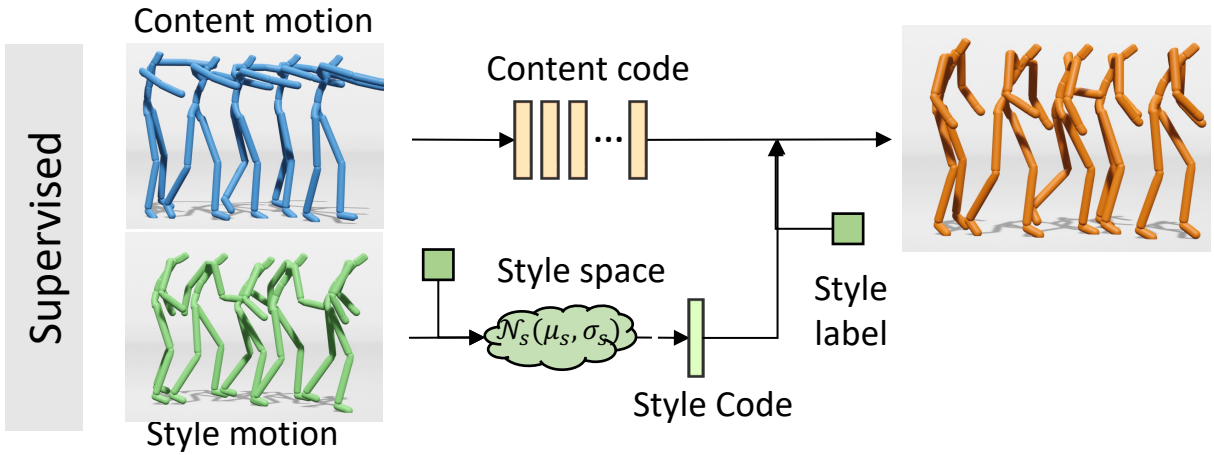
(b) Learning Scheme



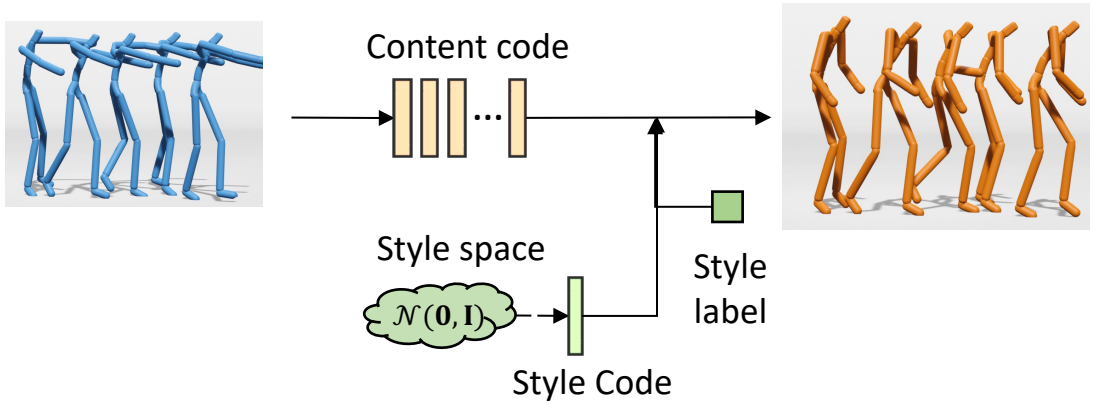
Style space visualization
(*Unsupervised*)

$\mathbf{z}^1$ and $\mathbf{z}^2$ are parts of the same long sequence.
Their styles are assumptively simar.

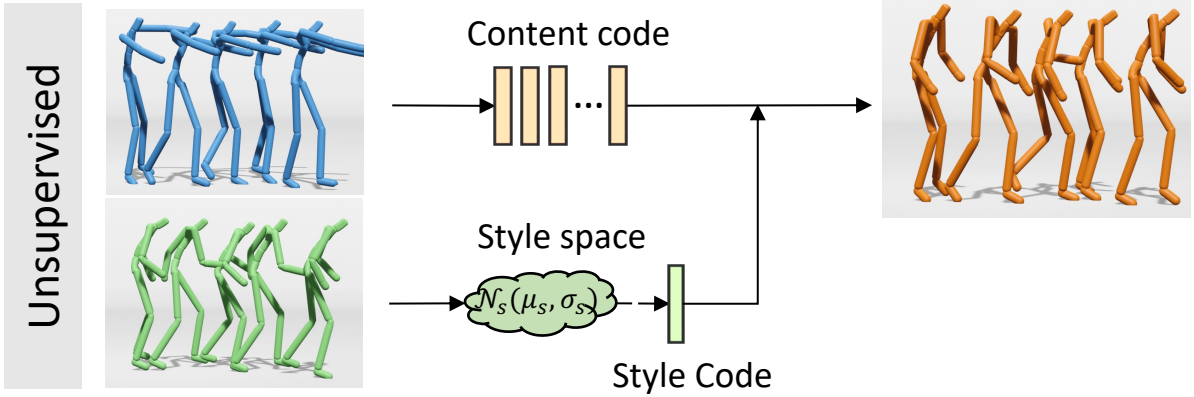
Inference



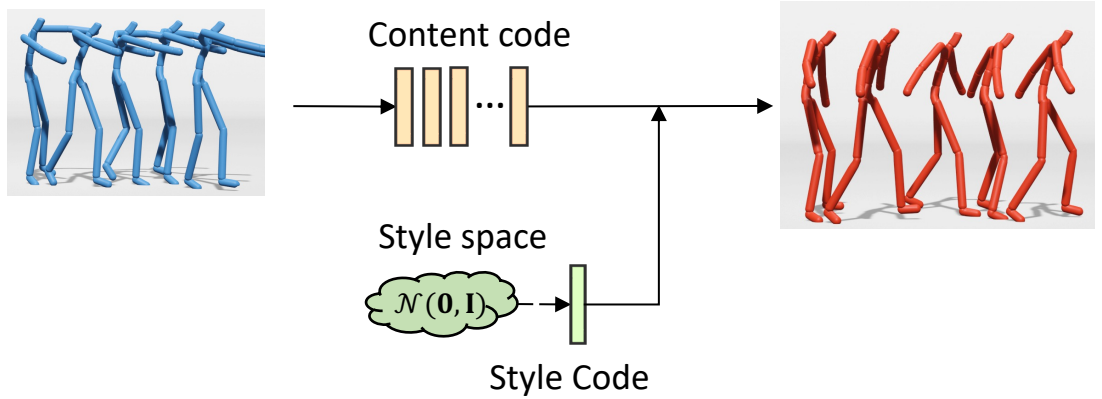
(a) Motion-based stylization



(b) Label-based stylization



(c) Motion-based stylization



(d) Prior-based stylization

Label-based Stylization

(*Diverse examples*)

** Ours (V) is used by default.*

Content input ***Our Results***

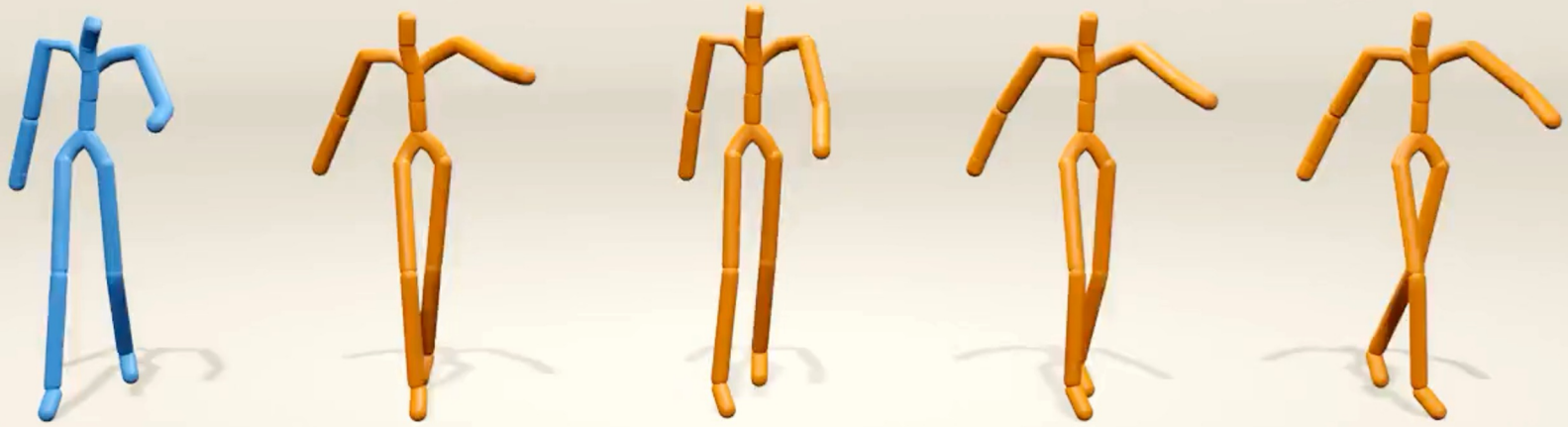
1 2 3 4

Style input: *Old*

* Different paces are made possible by post-*global motion prediction*.

Content input ***Our Results***

1 2 3 4



Style input: *FemaleModel*

* Different paces are made possible by post-*global motion prediction*.

Motion-based Stylization

** Ours (V) is used by default.*

Style input
(FemaleModel)

Content input	***Ours*** *(Supervised)*	***Ours*** *(Unsupervised)*
		

Style input
(Sneaky)

* Out-of-distribution content

Content input

Ours
(Supervised)

Ours
(Unsupervised)

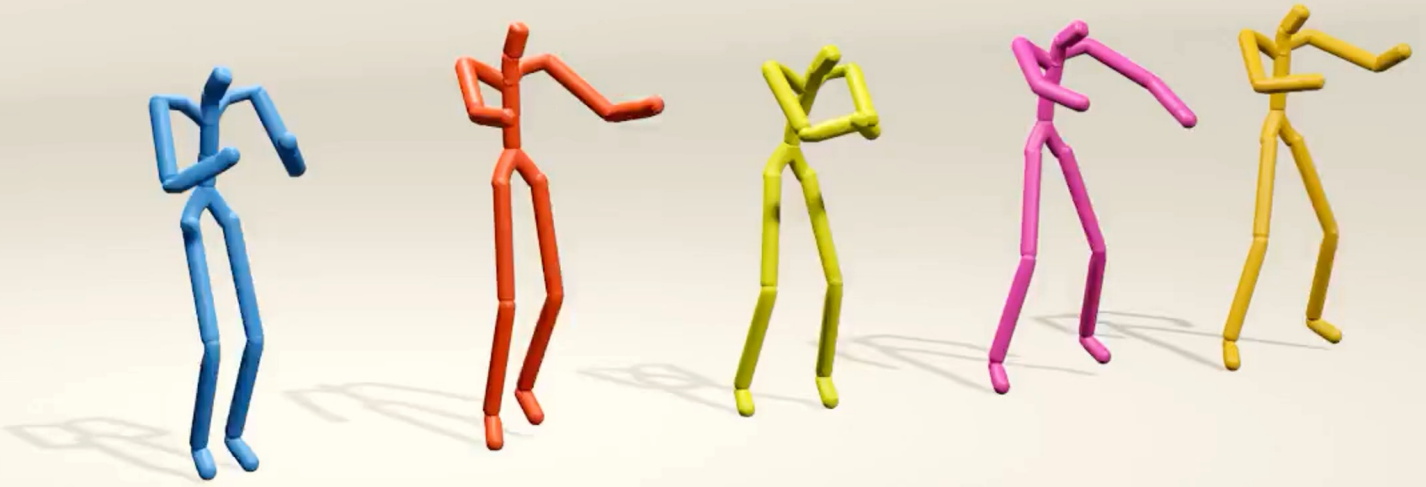
Prior-based Stylization

(*Diverse examples*)

** Ours (V) is used by default.*

Content ***Our Results***

1 **2** **3** **4**



* Different paces are made possible by post-*global motion prediction*.

Stylized Text2Motion

** Ours (V) is used by default.*
** Arbitrary content motion length.*

Style input
(FemaleModel)

Text input: *The person is doing a casual quick walk.*

* Out-of-distribution content

T2M Result

Stylized T2M Result

Thanks for watching!