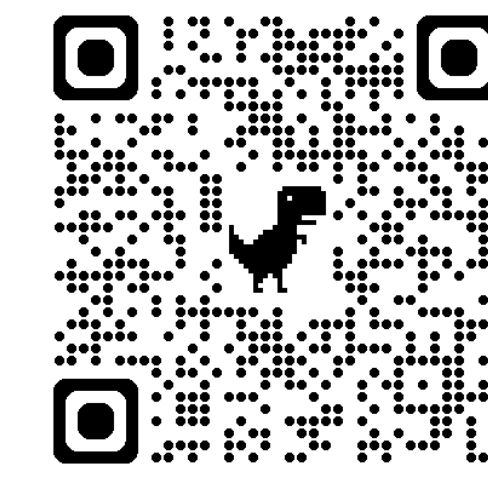




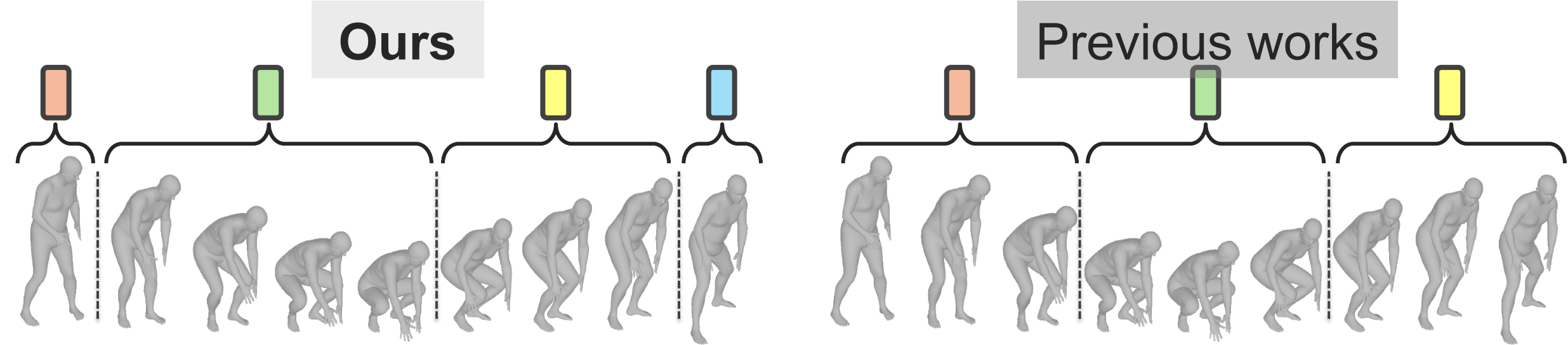
Action Motifs: Self-Supervised Hierarchical Representation of Human Body Movements

Genki Kinoshita¹ Shu Nakamura¹ Ryo Kawahara¹ Shohei Nobuhara² Yasutomo Kawanishi³ Ko Nishino¹
¹Kyoto University ²Kyoto Institute of Technology ³RIKEN

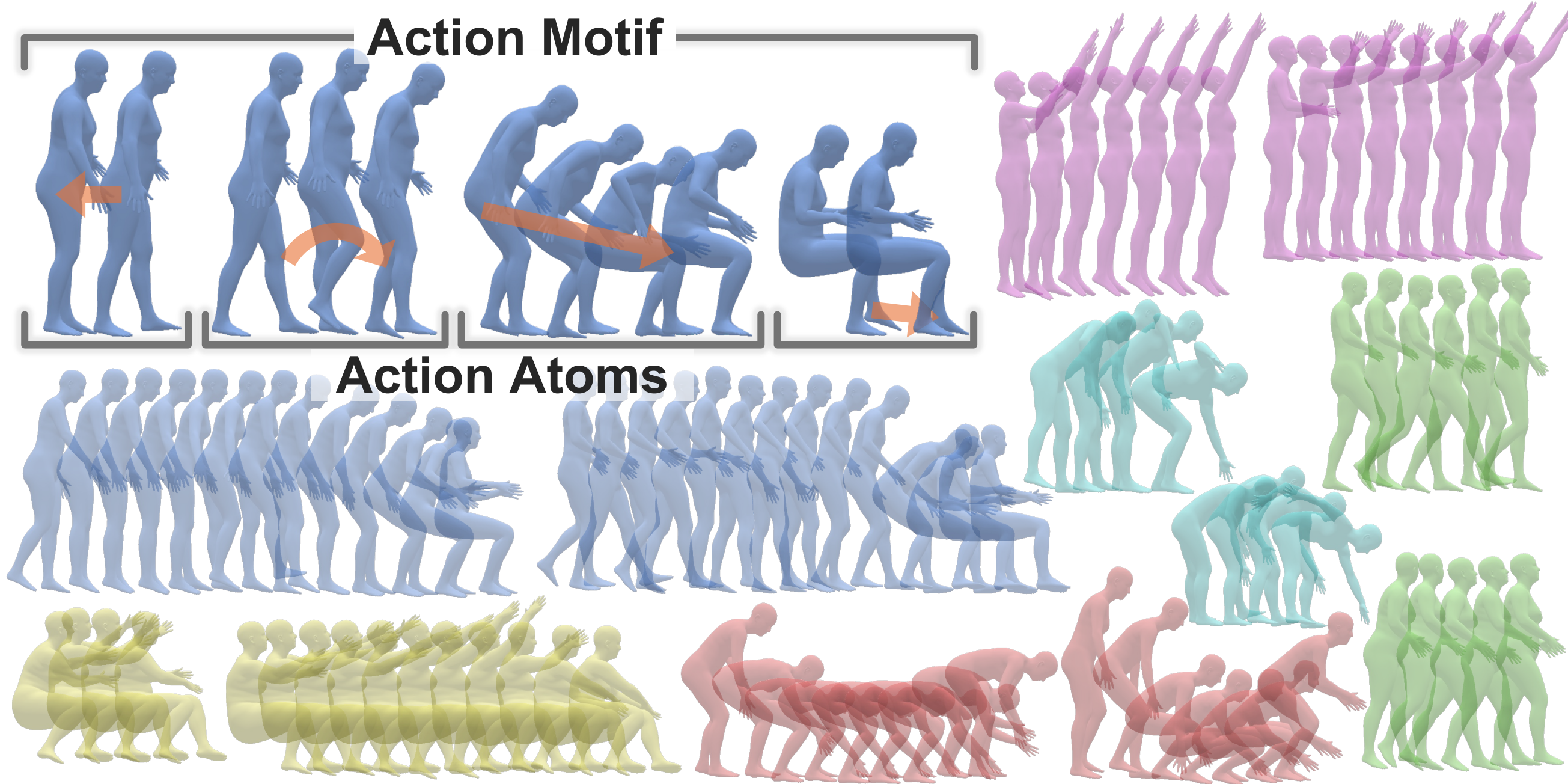


Overview

We learn human movement representations from 3D poses alone in **meaningful units**, not fixed-length units.



As meaningful units, we focus on **fine-grained movement patterns that recur in daily behaviors** (Action Motifs).

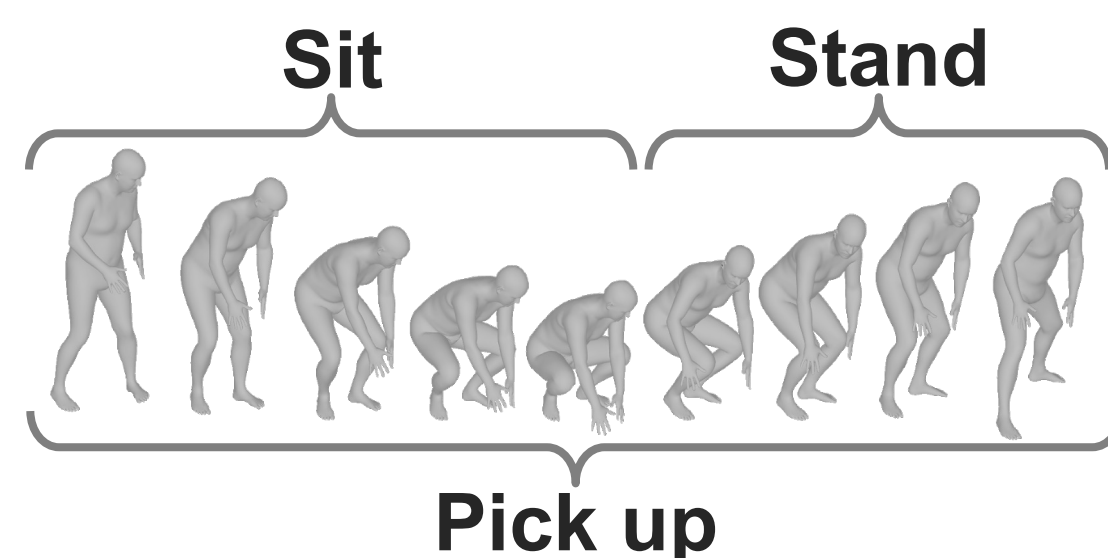


To learn Action Motif representations with only self-supervision, we also need to learn their temporal segments jointly.

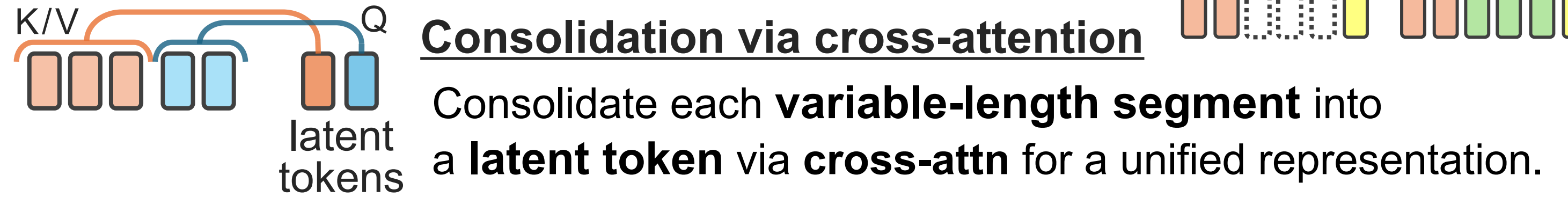
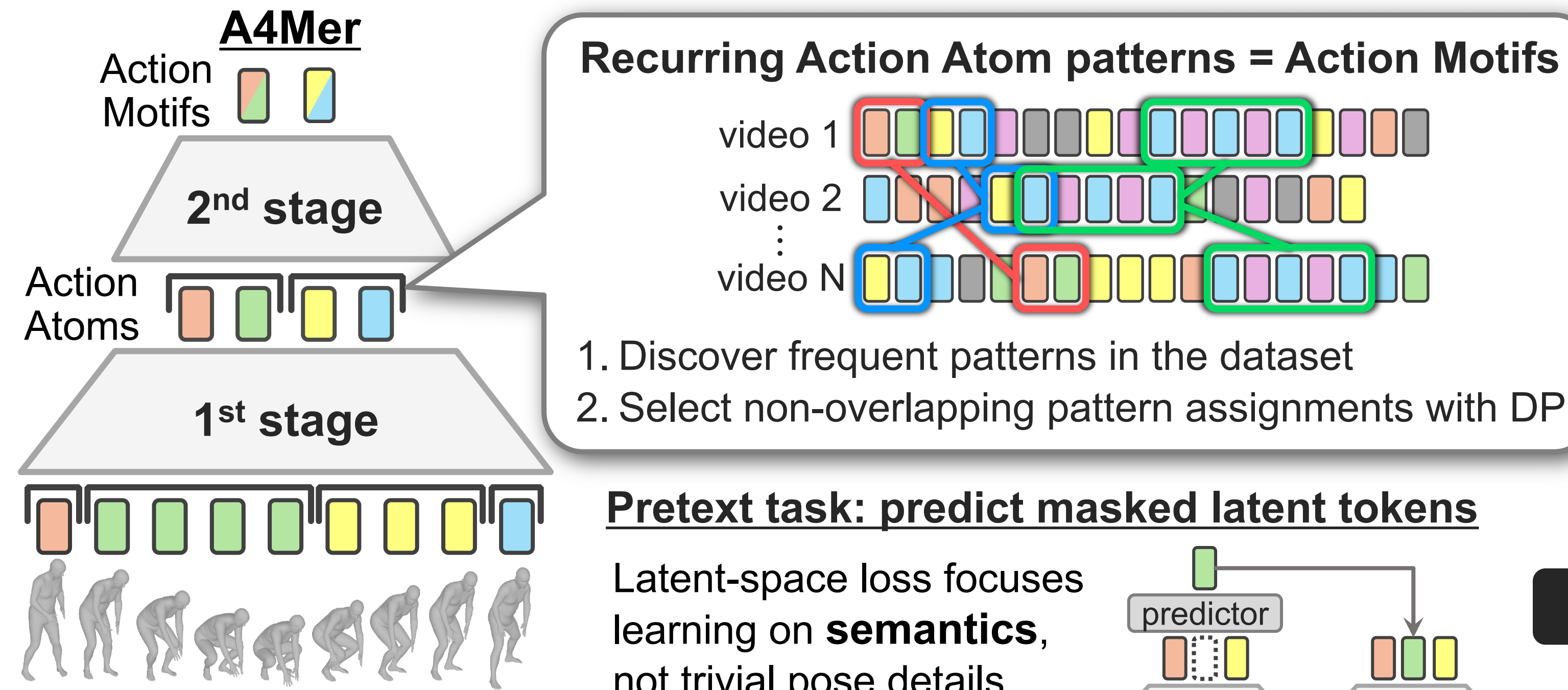
Key challenge

🐔 & 🥚 dependency between segmentation & representation semantics

- Different segmentations induce different semantic interpretations.
- Without understanding semantics, we can't determine the segments.



Let movement patterns emerge through bottom-up learning



Action Motif Dataset (AMD)

In a real house, occlusions make it hard to capture legs from exo views...

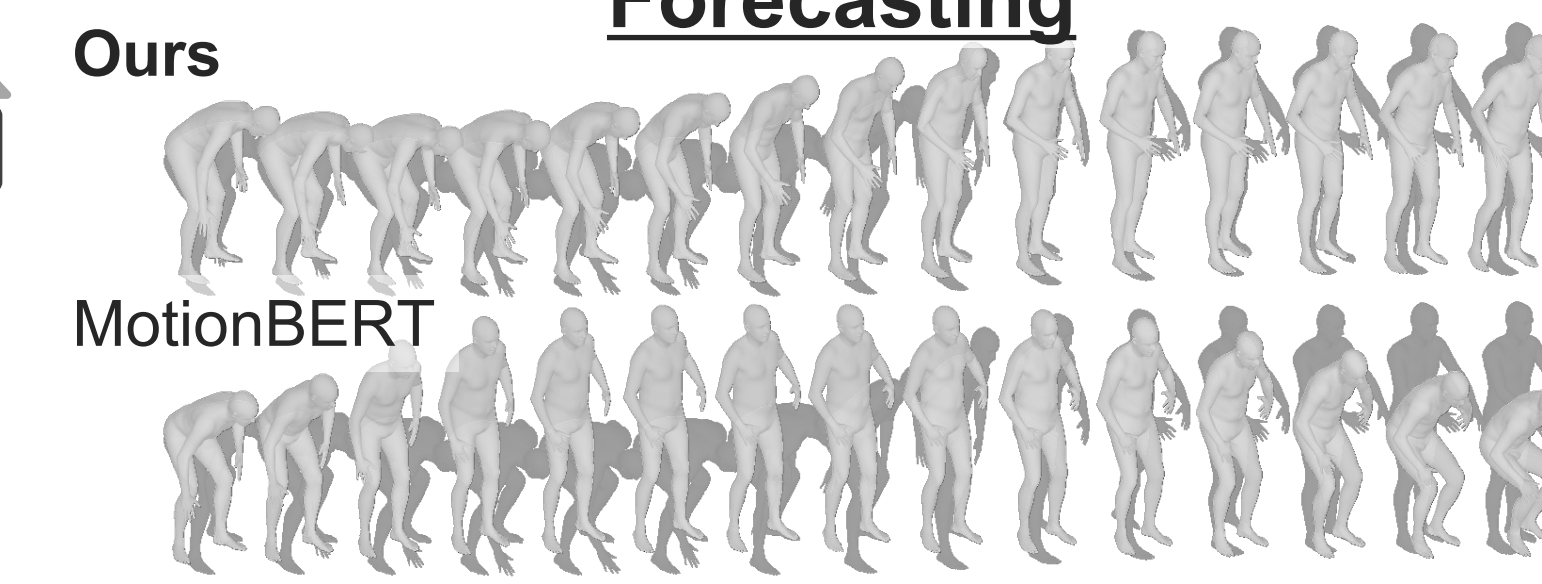


Action Motifs in AMD

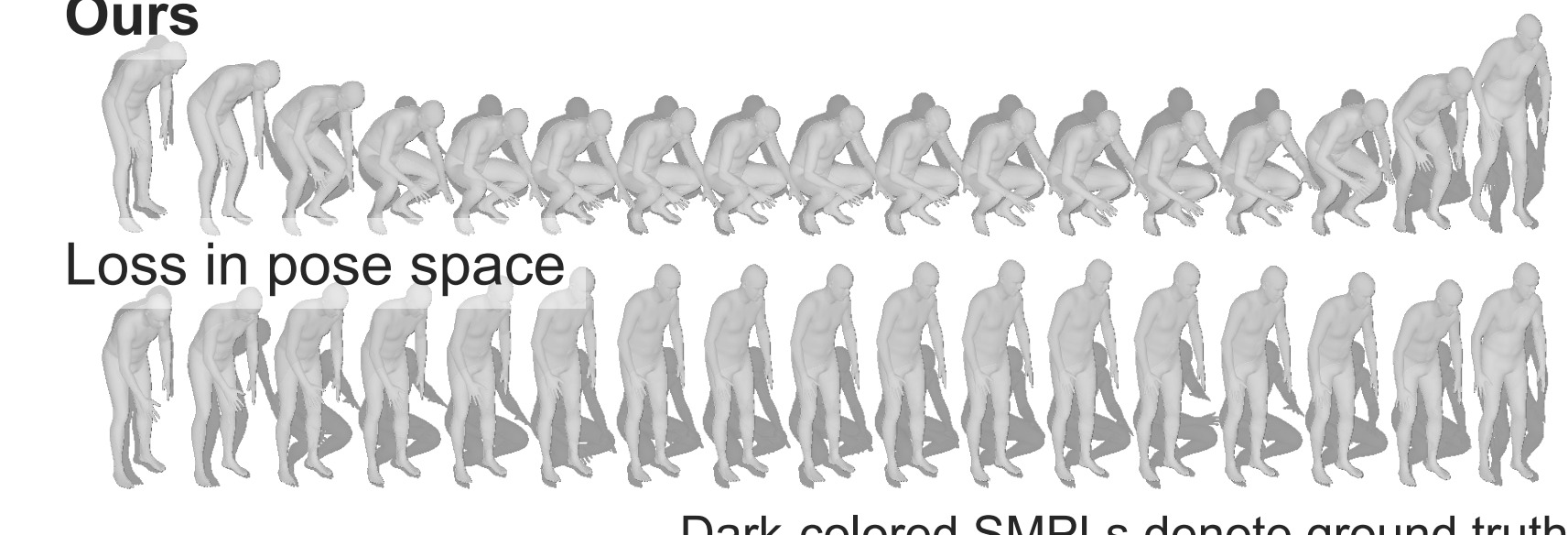
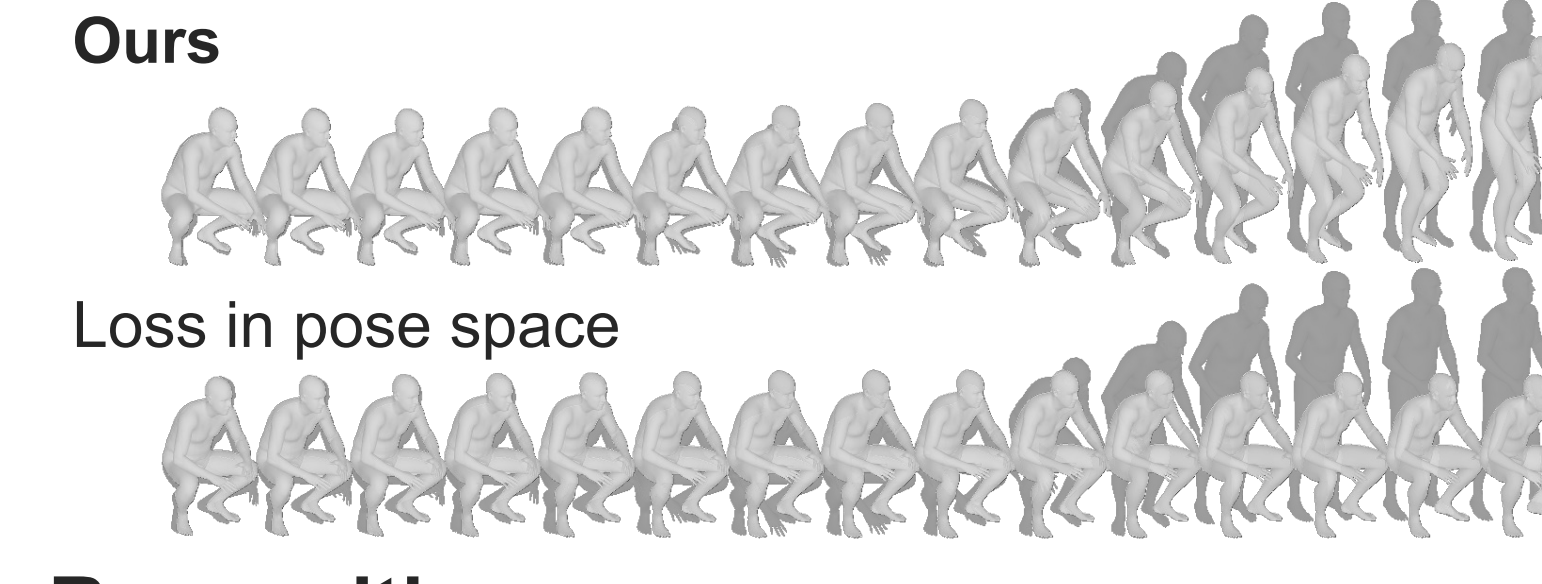
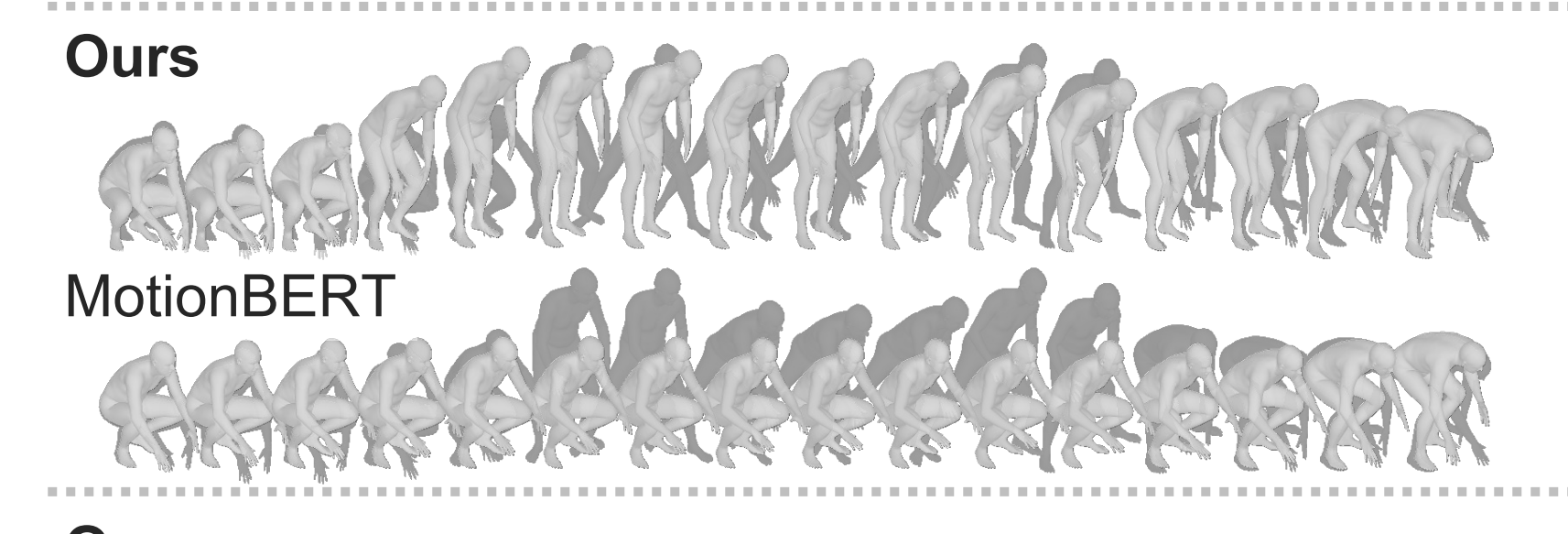
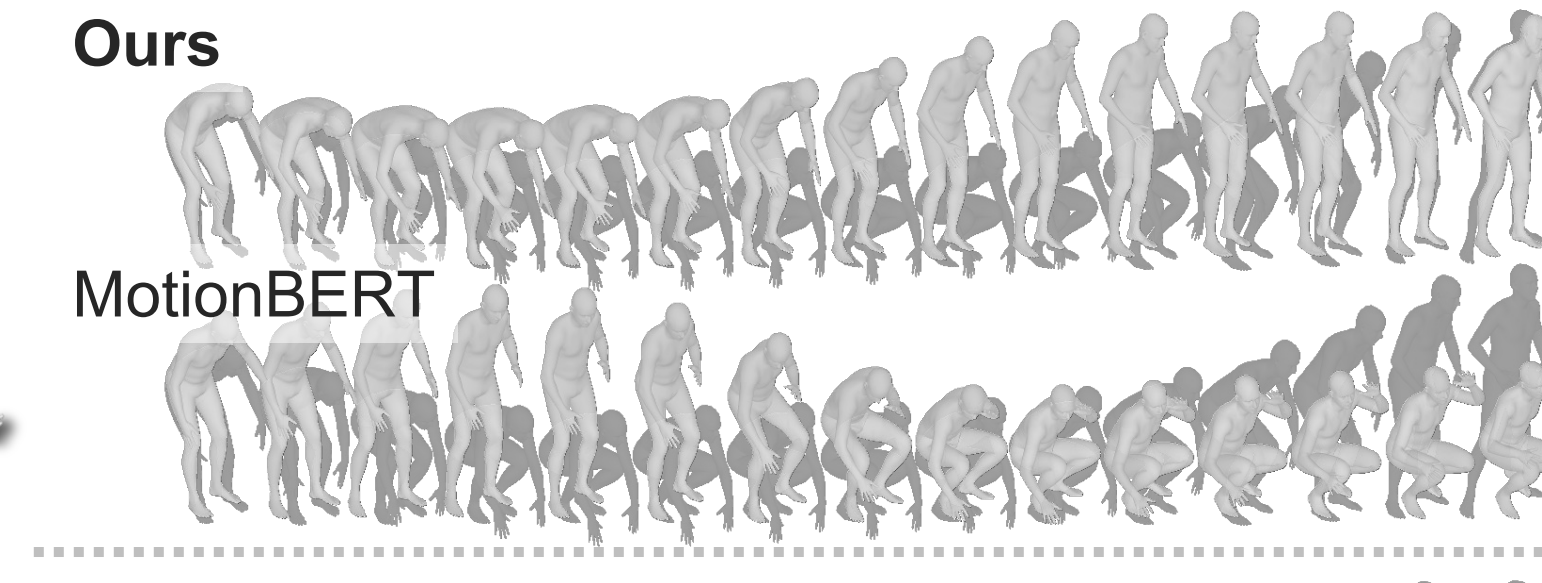
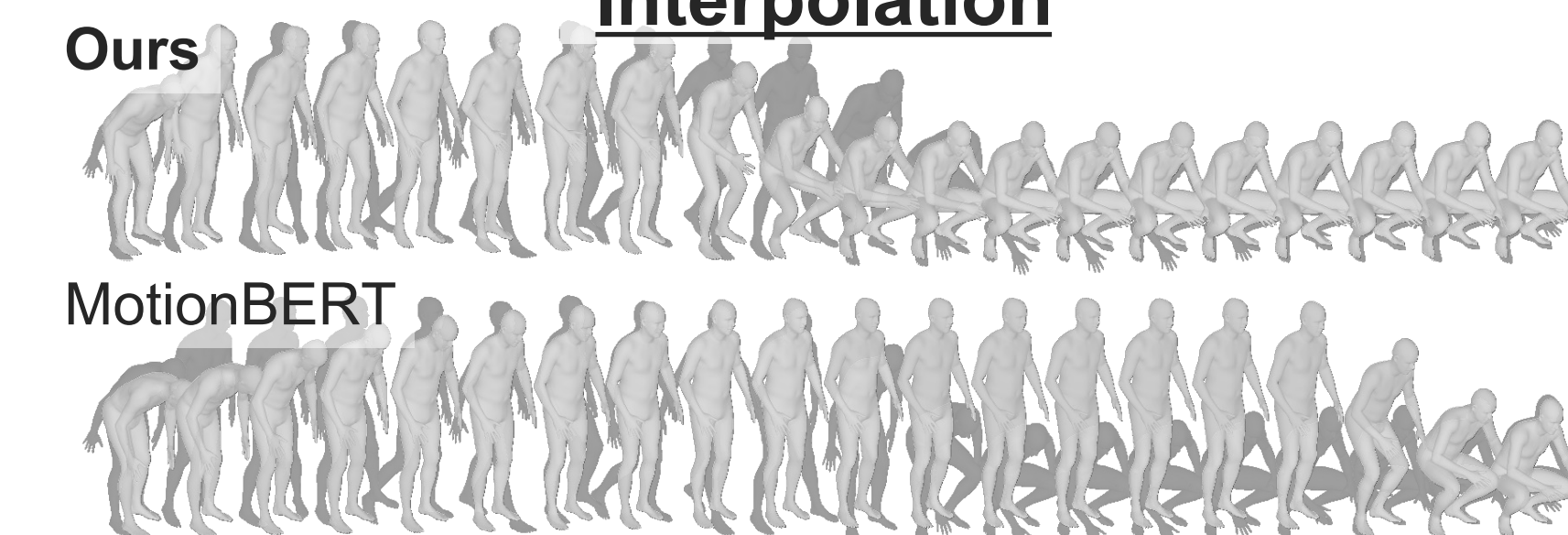


Downstream task results

Forecasting



Interpolation



Recognition

