ICCV23

Demo & code available online

TMR: Text-to-Motion Retrieval Using Contrastive 3D Human Motion Synthesis

Mathis Petrovich[1,2] Michael J. Black[2] Gül Varol[1]

[1] IMAGINE, École des Ponts ParisTech, France

[2] Max Planck Institute for Intelligent Systems, Tübingen, Germany

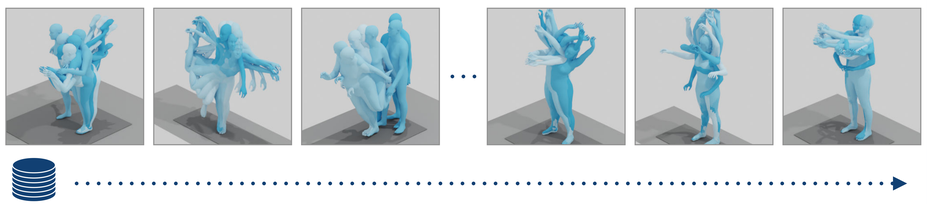
École des Ponts ParisTech

Introduction

Goal Searching for motions with natural language

Inputs Text query: *A person is stretching legs* — Gallery of motions

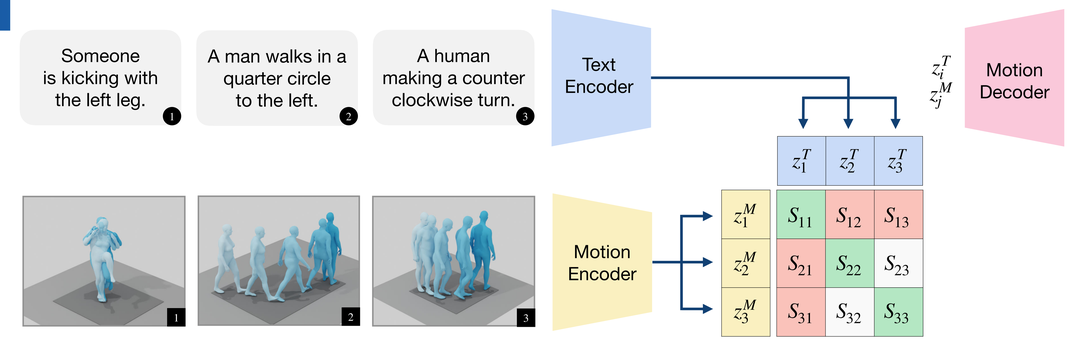
Output Ranking of gallery motions by similarity with the text query



TMR: Text-to-Motion Retrieval

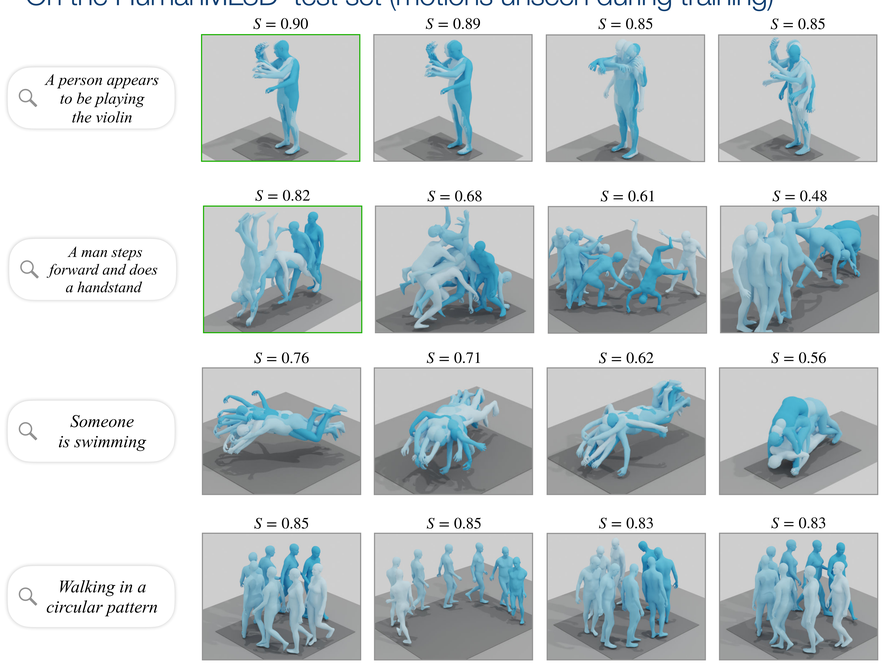
Key components

- Contrastive learning using InfoNCE
- Motion synthesis auxiliary loss
- Filtering negatives (when the texts are too similar)



Qualitative results

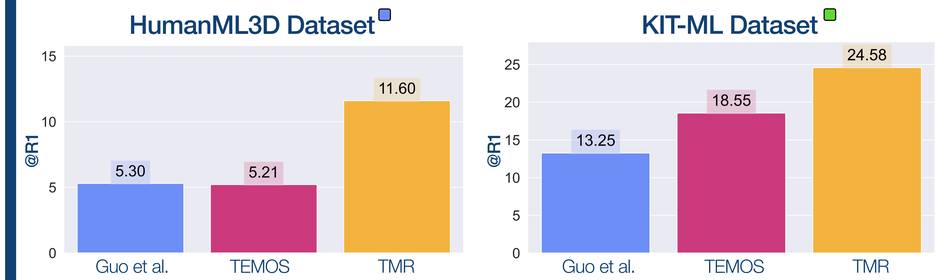
On the HumanML3D test set (motions unseen during training)



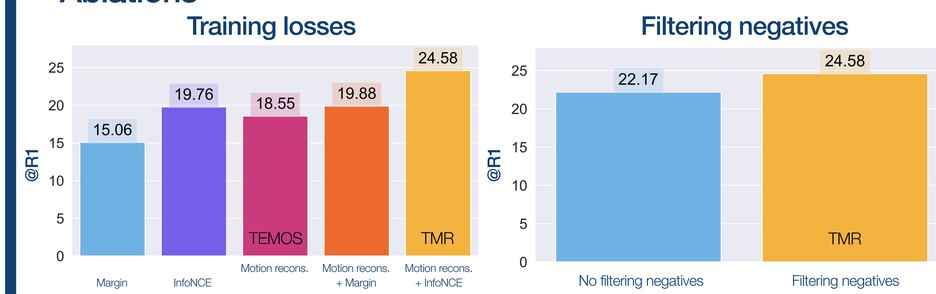
Quantitative results

Retrieval performance metric: @R1 ↑

Comparison to state of the art

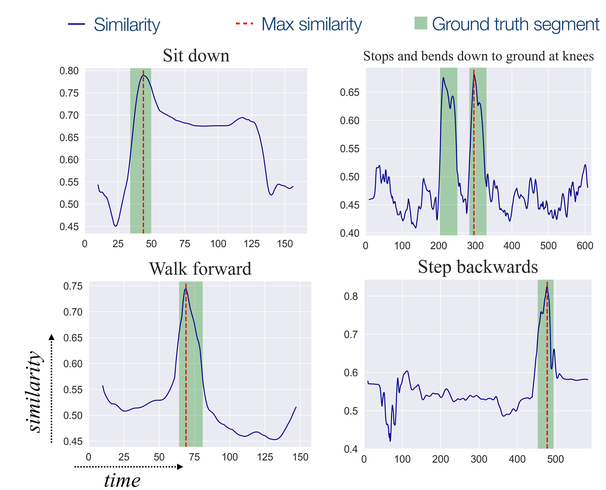


Ablations



Use case: localization

- Trained on HumanML3D motion-text pairs
- Tested on BABEL action sequences
- Similarity over a sliding window indicates that the localization ability emerges from our model.



References

- TEMOS: Petrovich et al. Generating diverse human motions from textual descriptions, ECCV 2022
- InfoNCE: van den Oord et al. Representation learning with contrastive predictive coding, arXiv 2018
- HumanML3D: Guo et al. Generating diverse and natural 3d human motions from text, CVPR 2022
- KIT-ML: Plappert et al. KIT Motion-Language dataset, Big Data 2016
- BABEL: Bodies, action and behavior with english labels, Punnakkal et al., CVPR 2021

mathis.petrovich@enpc.fr