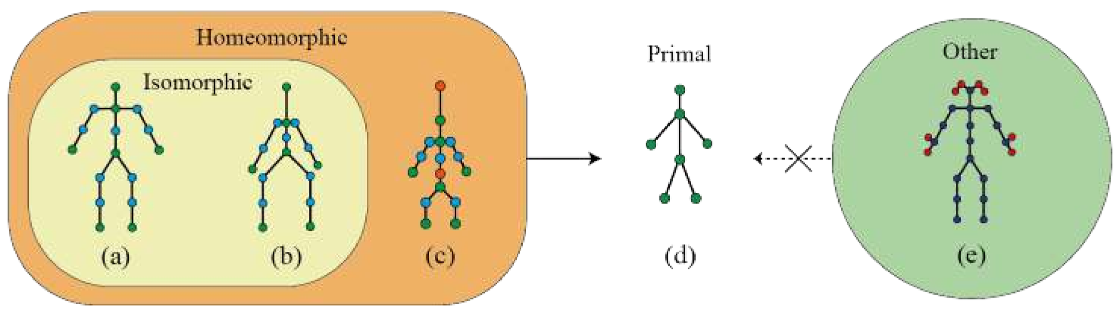
UNIVERSITY
OF TRENTO - Italy

MoMa: Motion retargeting using Masked pose modeling

Giulia Martinelli **Nicola Garau** **Niccolò Bisagno** **Nicola Conci**

University of Trento

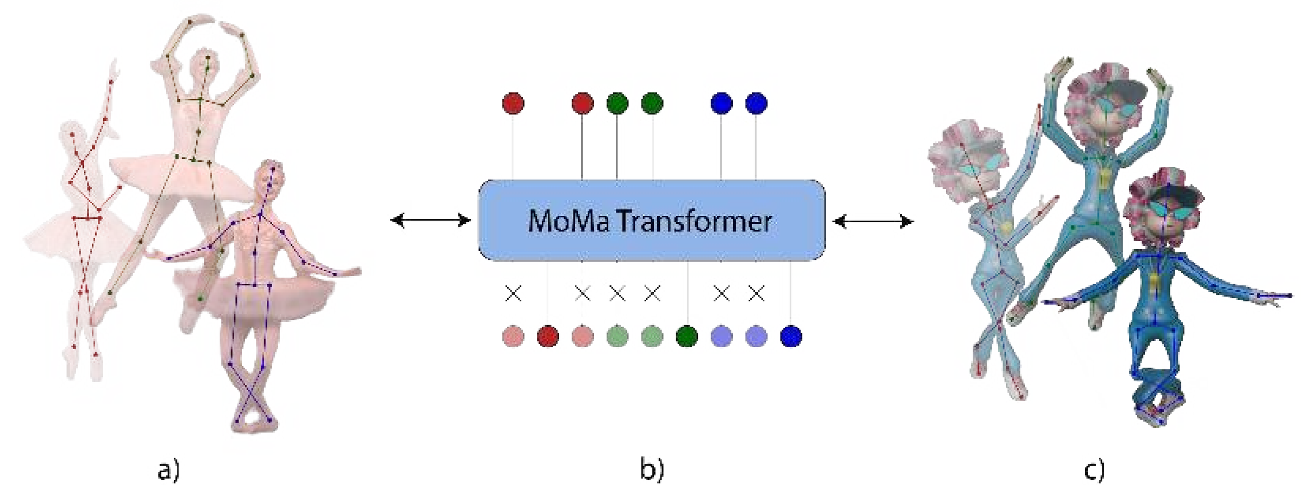
Context



The research in motion retargeting aims at transferring the body motion across different subjects and characters. State-of-the-art methods deal only with homeomorphic and isomorphic skeletons.

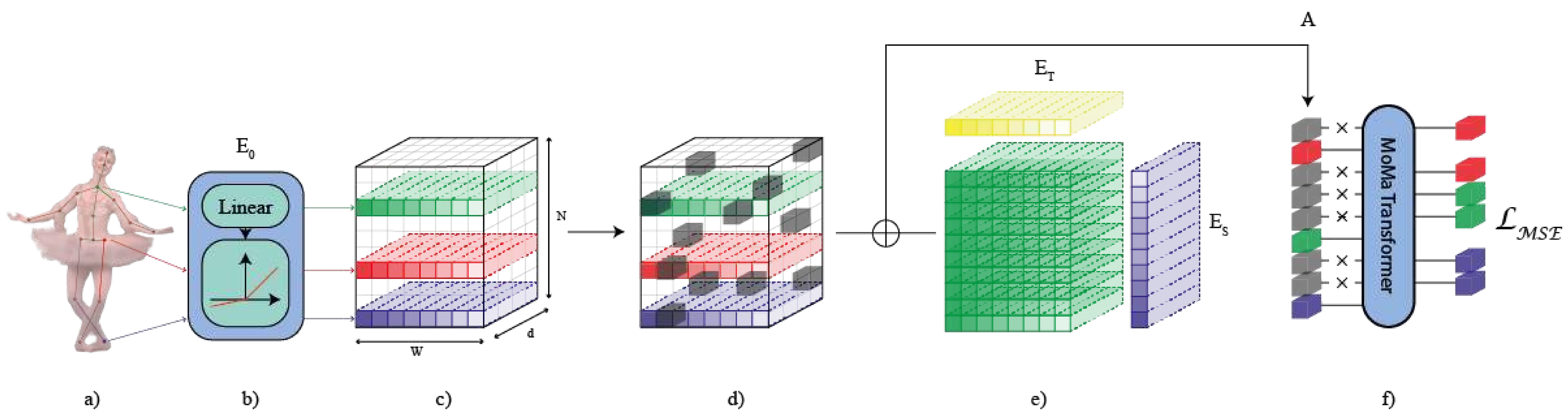
1. **Isomorphic**: same joints with varying bone lengths
2. **Homeomorphic**: different number of joints and varying bone lengths
3. **Other**: neither homeomorphic nor isomorphic

MoMa Transformer



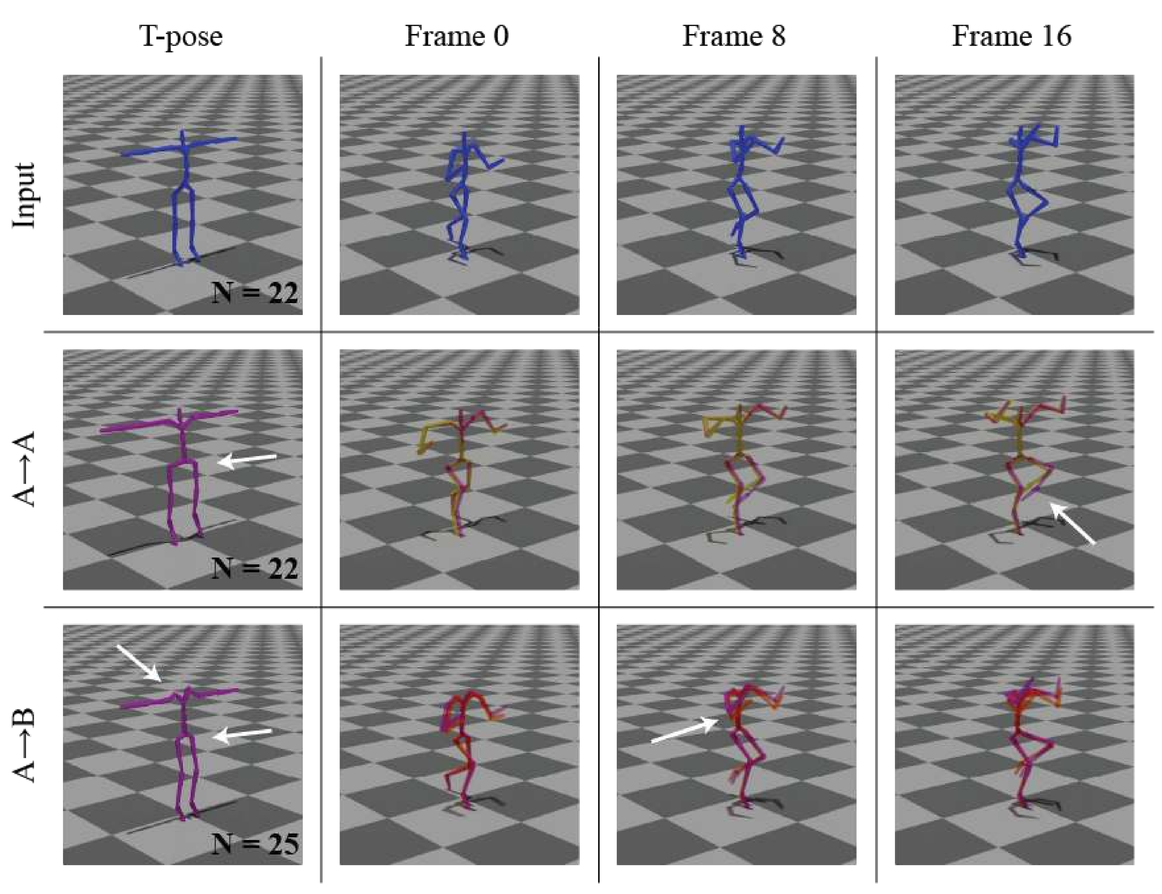
We show how we can transfer the artist's motion during the performance (a) to a different subject (c) and vice-versa, varying the skeleton shape. At each time frame, represented by RGB colors, we mask the differences between the two skeletons (b), represented by the symbol $\times$.

The architecture

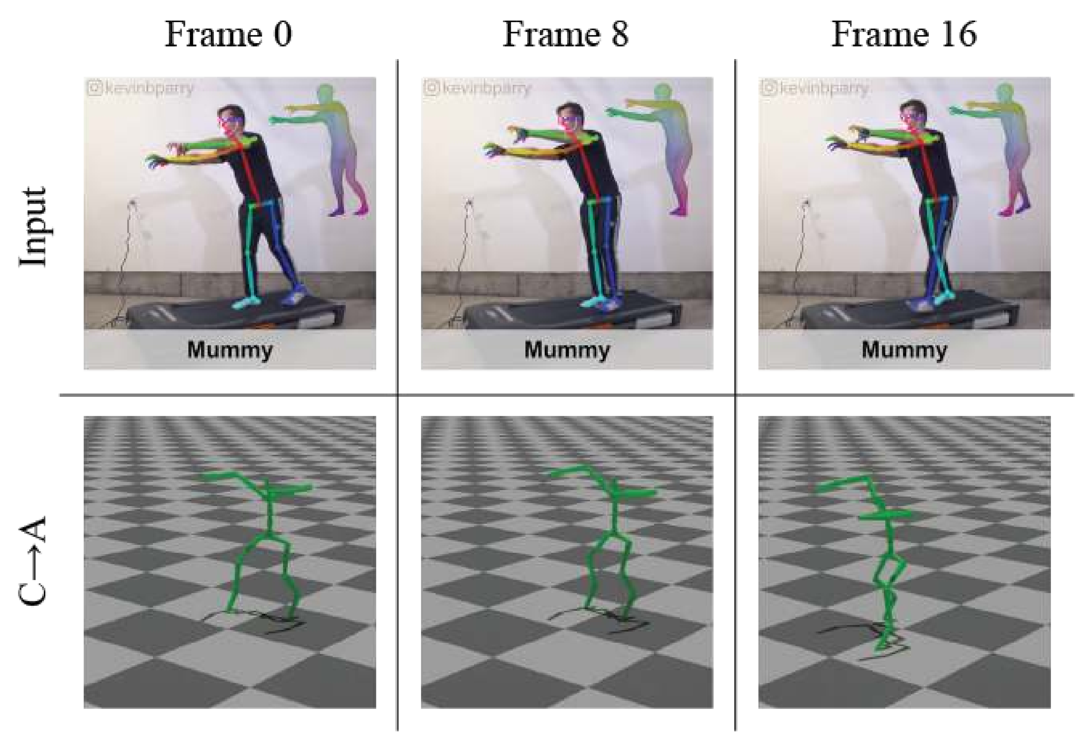


A motion sequence A of a character (a) is fed into the embedding E_0 (b), to obtain a 3D representation of size $W \times N \times d$ (c), where each colored vector represents the same joint over time. Then, we perform the random masking $\mathcal{M}$ of joints in space and time (d). The obtained representation is summed with the spatial-temporal embedding E_T and E_S (e), to get the embedded representation of the motion sequence. The vector is then fed into MoMa, which reconstructs the masked part of the input, on which we compute the loss $\mathcal{L}_{MSE}$.

Results

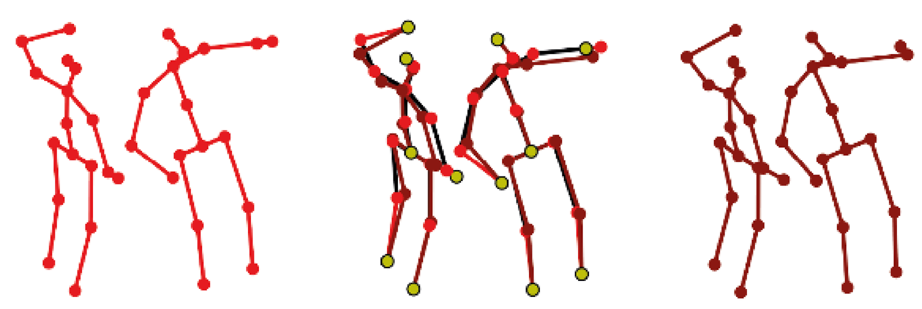


	Intra-Structural	Cross-Structural	Parameters
NKN	6.24	243	-
PMnet	5.72	-	-
Copy rotations	8.86	-	-
SA-noadv	0.47	3.81	35M
SA-full	2.76	2.25	35M
Ours	**0.2**	**0.49**	**5M**



Real world scenarios

Classic pose predictors (a) sometimes fail to detect some joints or entire limbs due to strong occlusions or out-of-image coordinates. Our method (b) can reconstruct an arbitrary number of joints in the incomplete pose even for imperfect predictions.



Examples of poses generated with all joints masked except the "leaf" ones highlighted in the figure. The GT is the red skeleton, while our network generated the dark red one.