Animation videos

Evaluation of Body Parts-based Latent Representations for Skeletal Human Motion Reconstruction

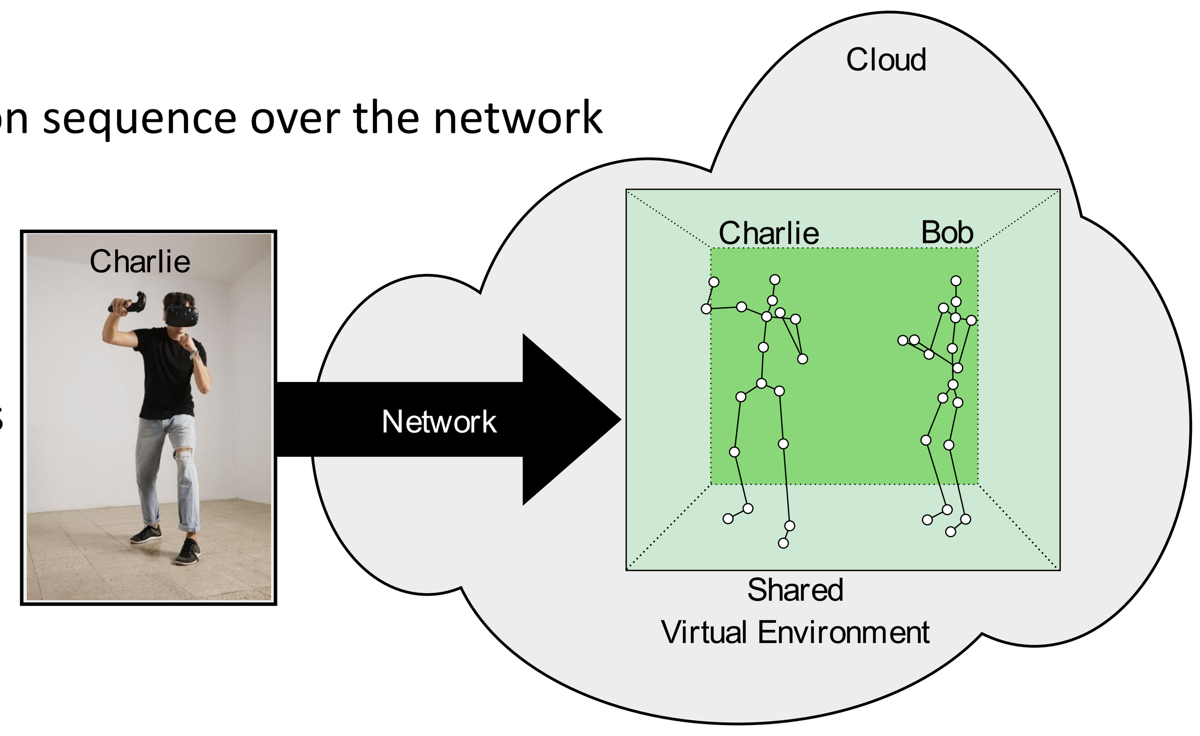
MIG 2024 Arlington

Philippe de Clermont Gallerande[1,2], Ludovic Hoyet[1], Ferran Argelaguet[1], Quentin Avril[2], Philippe-Henri Gosselin[2]
[1]Inria, Univ. Rennes, CNRS, IRISA [2]InterDigital

Context

Transmission of a body motion sequence over the network

Constraints
- **motion** quality
- distinct **noise** distributions
- **semantic** information (interoperability)

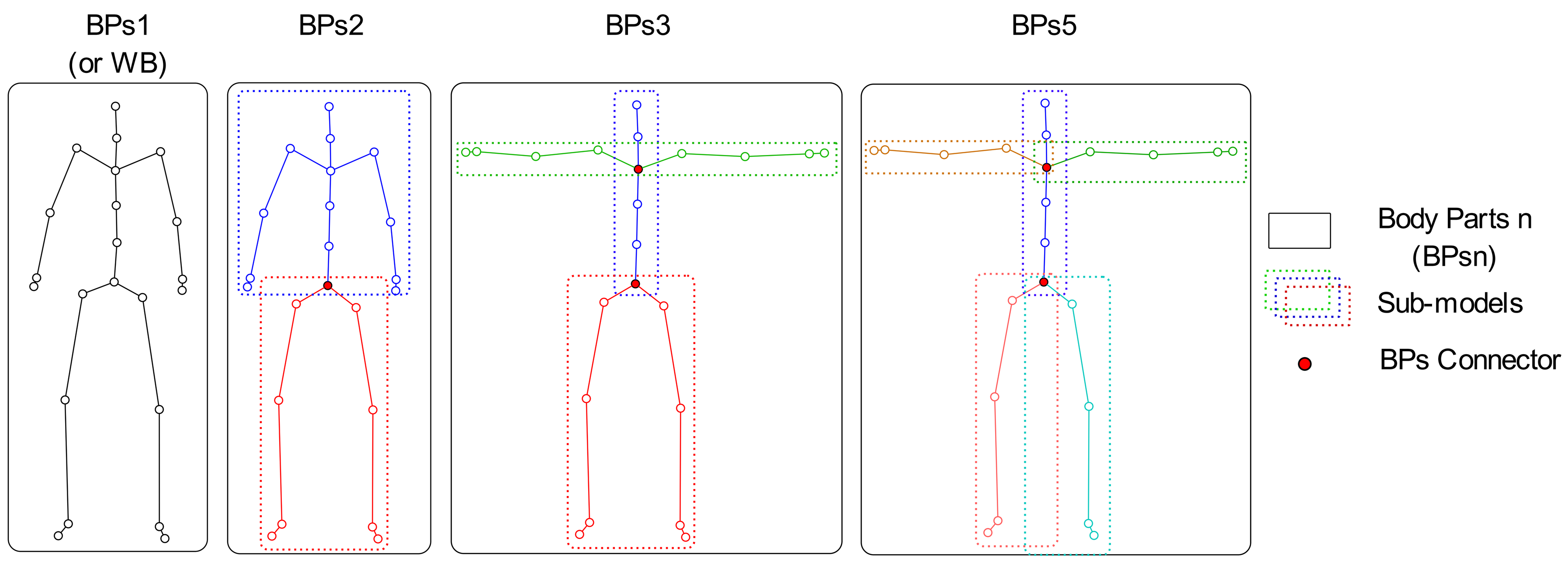


Proposal
Evaluating explicitly partitioned representation (body parts approach [2, 3] in contrast to a whole-body one [1]) to perform a motion sequence **reconstruction** task

Main Contributions:
1. **A novel framework for human motion modelling featuring high modularity**
2. **An evaluation of the impact of the body decomposition granularity**

Body Parts Model

A Body Parts (BPs) model wraps $n \in \mathbb{N}^*$ **neural networks**, each learning the motion of a part of the body **independently** of its neighbours



We obtain n **latent representations**
The transmitted latent data is **semantically partitioned**

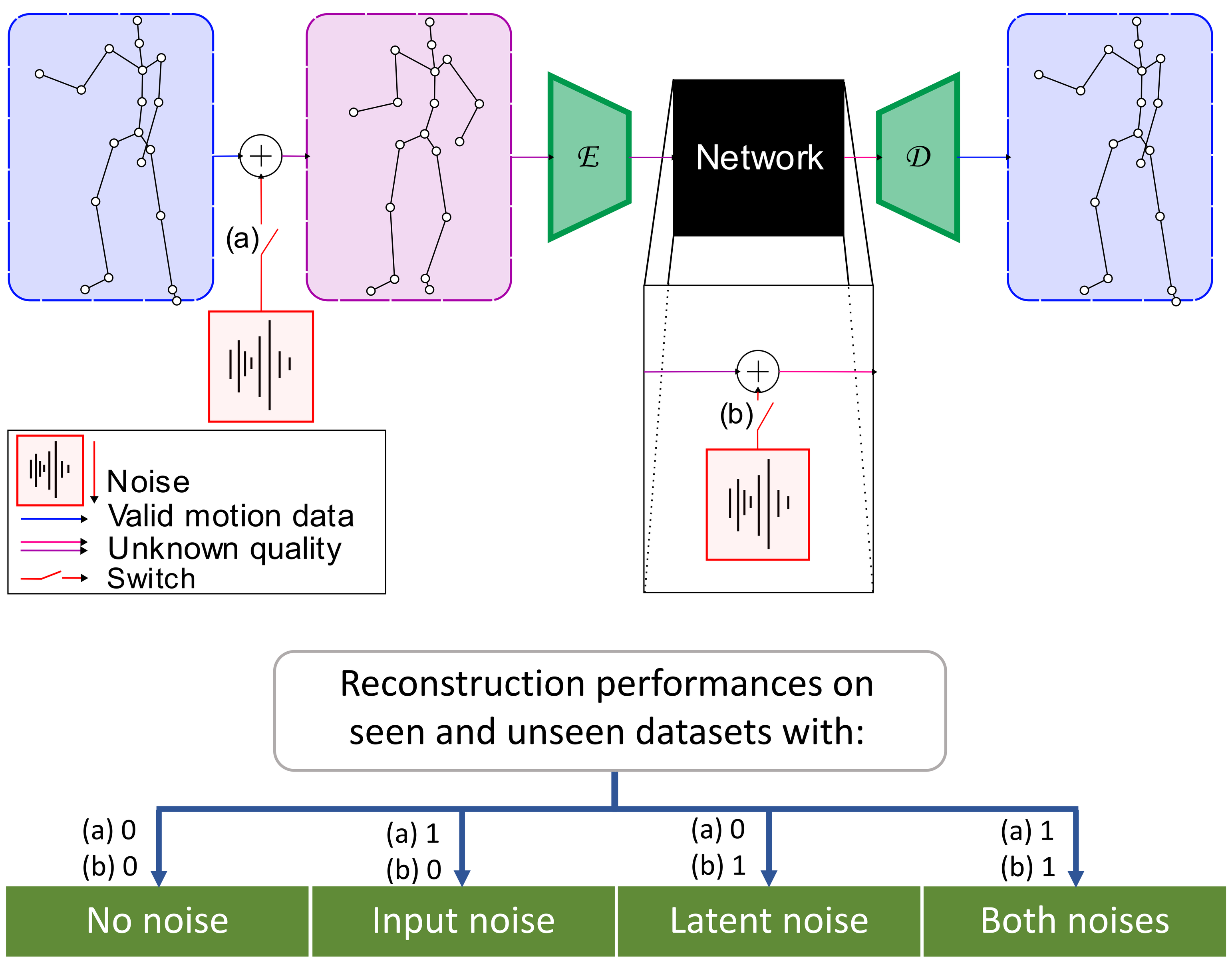
Evaluation

Simulate real-time motion streaming use cases
Assert the impact of the **granularity** of the body decompositions on the performances

Skeletal motion representation

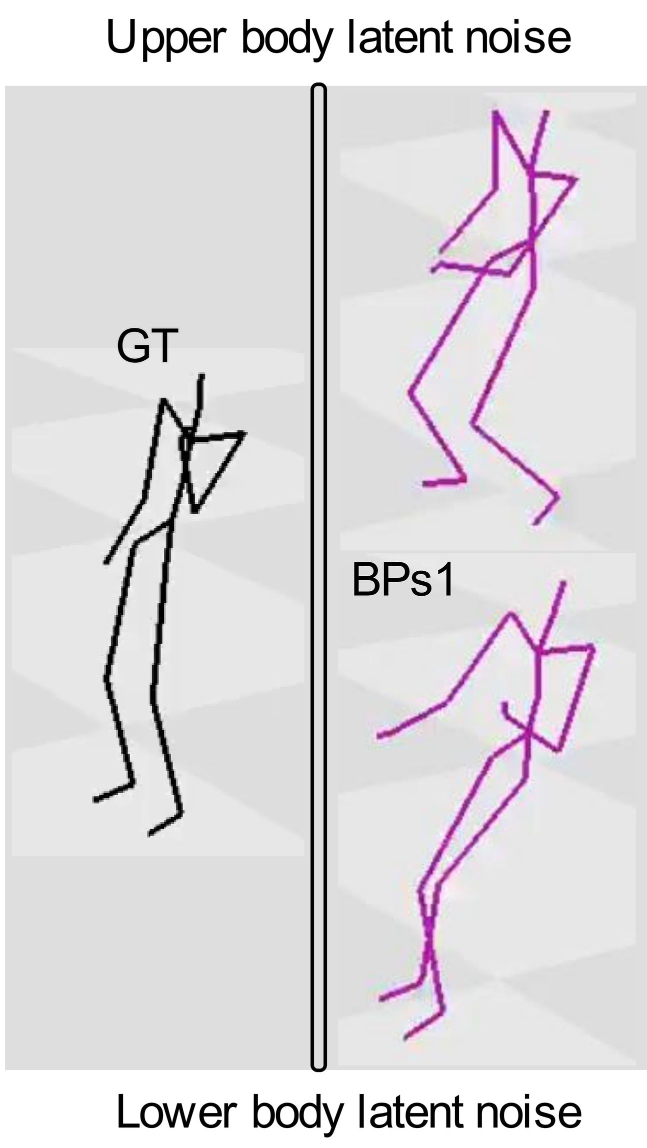
Robustness to noise

Assert reconstruction performance



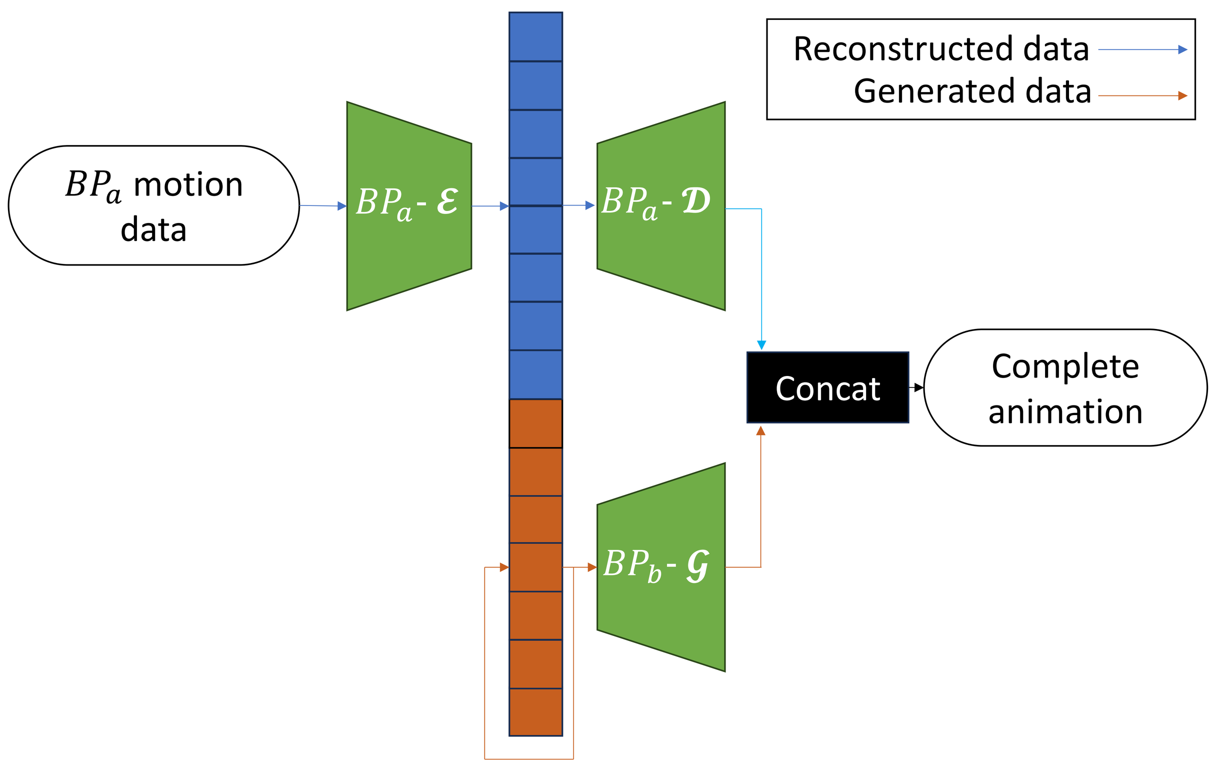
Interpretability

Assert noise spatial coherency



Controllability

Assert framework flexibility in other use cases



- Semantic-based animations for novel user-created topologies
- Seated immersive videoconferences (e.g., occlusion)

Results

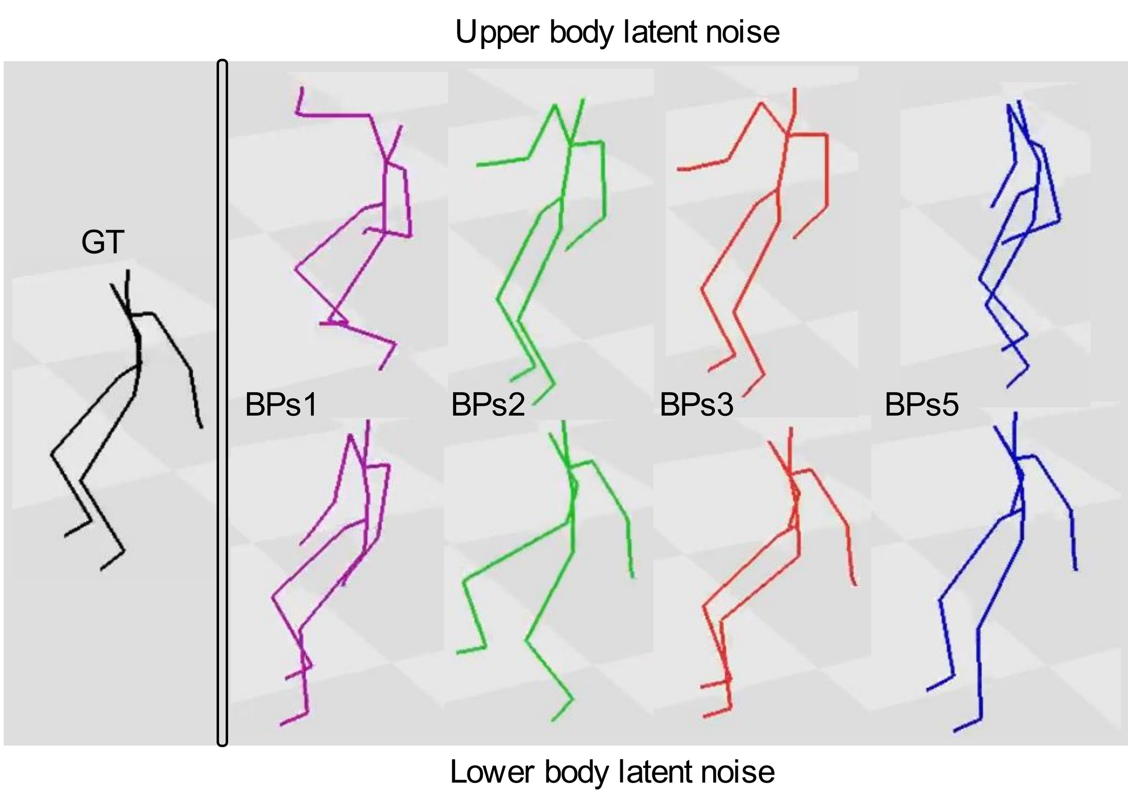
Relevant reconstruction results

Reconstruction metric MPJPE [cm]

	BPs1	BPs2	BPs3	BPs5
CMU_test	**5.50 ± 0.65**	<u>6.75 ± 0.48</u>	7.09 ± 0.94	7.28 ± 0.69
MHAD	13.43 ± 0.60	12.84 ± 0.44	**12.09 ± 0.73**	<u>12.19 ± 0.47</u>
Edin_loco	**8.13 ± 0.62**	9.96 ± 0.44	<u>9.78 ± 0.70</u>	10.00 ± 0.61
Edin_fight	**11.25 ± 0.81**	12.74 ± 0.51	12.39 ± 1.00	<u>11.72 ± 0.83</u>

(**Bold** = best, <u>underlined</u> = second)

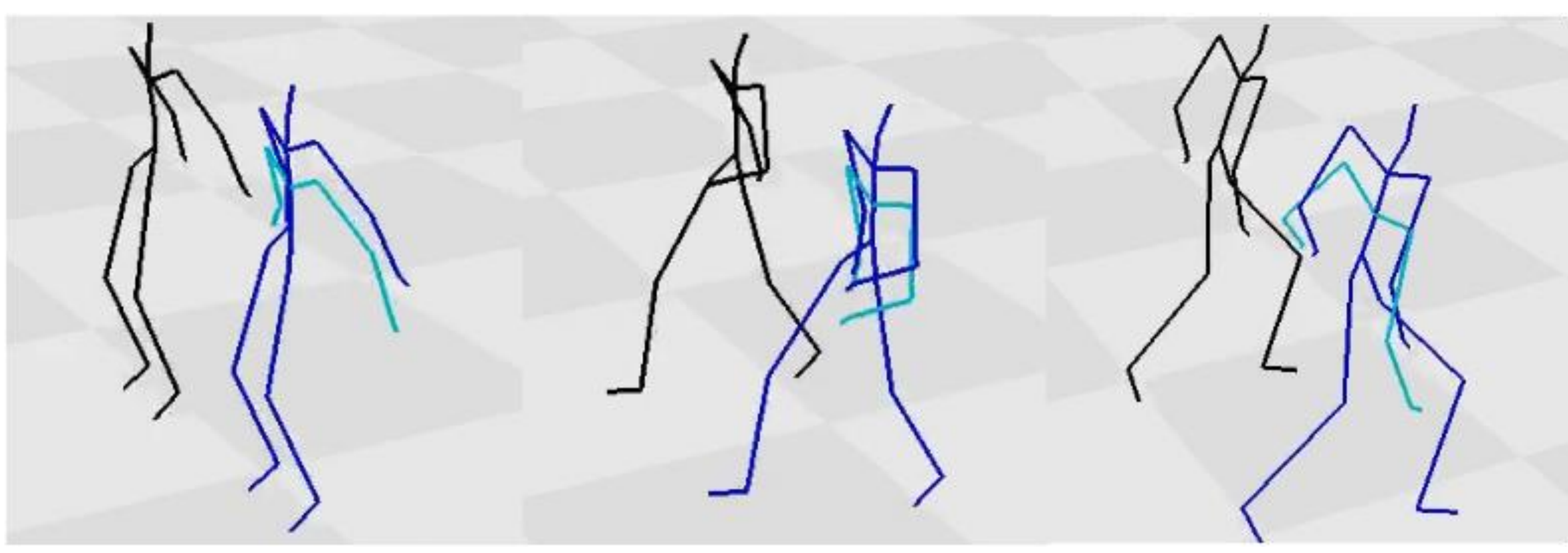
Semantically partitioned latent representation



Conclusion & Take Away

Flexible sub-models	Better reconstruction than expected
Parallel training	Training on different datasets

Coarse animation of a user-created topology (here, a Chimera)



(Black = GT, dark blue = reconstructed, light blue = generated)

[1] Daniel Holden, Jun Saito, Taku Komura, and Thomas Joyce. 2015. Learning motion manifolds with convolutional autoencoders. In SIGGRAPH Asia 2015 Technical Briefs. ACM, 1–4. https://doi.org/10.1145/2820903.2820918
[2] Ashesh Jain, Amir R. Zamir, Silvio Savarese, and Ashutosh Saxena. 2016. Structural-RNN: Deep Learning on Spatio-Temporal Graphs. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 5308–5317. https://doi.org/10.1109/CVPR.2016.573
[3] Yingying Wang and Michael Neff. 2015. Deep signatures for indexing and retrieval in large motion databases. In 8th ACM SIGGRAPH Conference on Motion in Games (MIG '15). ACM, New York, NY, USA, 37–45. https://doi.org/10.1145/2822013.2822024

Université de Rennes · NENO•AI LAB by interdigital & Inria · YS•AI · Inria · interdigital