Unsupervised motion retargeting for human-robot imitation

Louis Annabi, Ziqi Ma, Sao Mai Nguyen
U2IS, ENSTA Paris, Institut Polytechnique de Paris, France

Goal: Imitation by retargeting motions from *any* source (e.g. human demonstrator) performer to *any* target performer (e.g. robot)
Challenge: Learn without access to *paired* motion data
Approach: Adapt unsupervised domain-to-domain translation deep learning algorithms to skeleton motion data
Take away message: This approach is not yet competitive with naive unpaired retargeting baselines

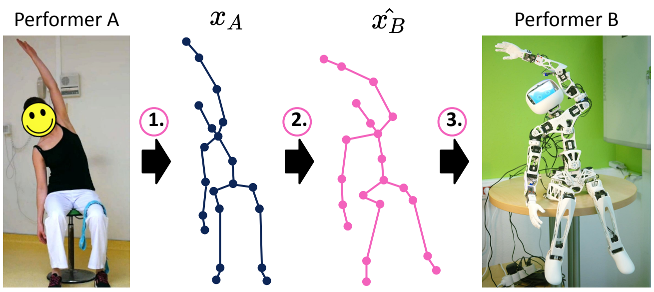
Context

This work is part of the Keraal project [1], aiming to develop a robot coach for physical rehabilitation of patients suffering from low back pain. In order to provide personalized feedback, and to learn new exercises, the robot needs to accurately and quickly **imitate demonstrated motions**. Because the demonstrators and the robot have different morphologies, it is necessary to **retarget** the human motion into a space of motions achievable by the robot.

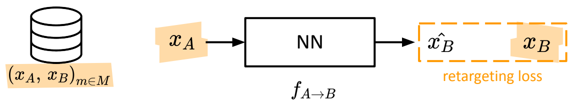
The imitation procedure is composed of three steps:

1. Pose estimation
2. Retargeting
3. Robot control

In this work, we focus on the retargeting step, and explore deep learning methods allowing to perform this retargeting without paired motions.

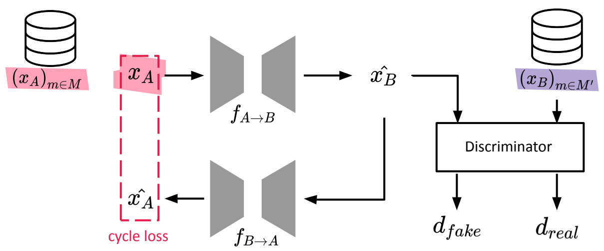


Supervised learning for **paired** motions

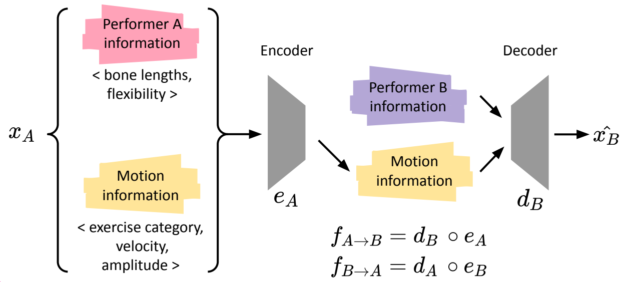


Methods

Unsupervised learning for **unpaired** motions



Retargeting between **any** pair of performers



Results

We have experimented with two models working on unpaired data:

- CycleGAN model [2]
- UNIT model [3, 4]

We use Mixamo [5], a dataset of short character animated motions. Training is performed on a set of **unpaired** motions from Mixamo:

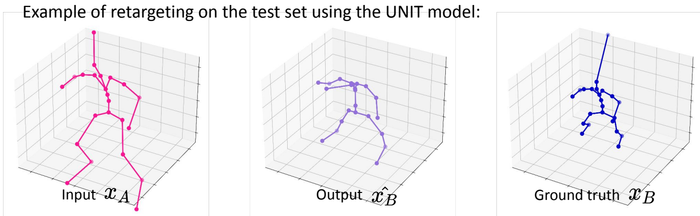
- 25 characters
- 83 ≠ motions per character on average

Evaluation is performed on a test set of **paired** motions from Mixamo:

- 4 characters
- 106 same motions

Model	Reconstruction loss (train) $d(x_A, \hat{x_A})$	Reconstruction loss (test) $d(x_A, \hat{x_A})$	Retargeting loss (test) $d(x_B, \hat{x_B})$
Position copy	0 mm	0 mm	195 mm
Rotation copy	0 mm	0 mm	**79 mm**
CycleGAN	70 mm	182 mm	243 mm
UNIT	48 mm	164 mm	209 mm

Example of retargeting on the test set using the UNIT model:



Conclusion

Our models seem to perform similarly well for reconstruction (A->A) and retargeting (A->B) but still fall short compared to simpler baselines.

Future work will focus on improving the decoder neural network, and apply our models to new data: motion capture & pose estimation, robot motion

References

[1] Blanchard, A., Nguyen, S.M., Devanne, M., Simonnet, M., Goff-Pronost, L., & Rémy-Néris, O. (2022). Technical Feasibility of Supervision of Stretching Exercises by a Humanoid Robot Coach for Chronic Low Back Pain: The R-COOL Randomized Trial. BioMed research international, 2022.
[2] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223-2232).
[3] Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, *30*.
[4] Aberman, K., Li, P., Lischinski, D., Sorkine-Hornung, O., Cohen-Or, D., & Chen, B. (2020). Skeleton-aware networks for deep motion retargeting. *ACM Transactions on Graphics (TOG)*, *39*(4), 62-1.
[5] Adobe Systems Inc. (2018). Mixamo. https://www.mixamo.com.