ReActor: Reinforcement Learning for Physics-Aware Motion Retargeting

Supplementary Material

DAVID MÜLLER, Disney Research, Switzerland
AGON SERIFI, Disney Research, Switzerland
SAMMY CHRISTEN, Disney Research, Switzerland
RUBEN GRANDIA, Disney Research, Switzerland
ESPEN KNOOP, Disney Research, Switzerland
MORITZ BÄCHER, Disney Research, Switzerland

Table 1. **Reward Terms for Downstream RL Training.** The root position is given by $\mathbf{x}^{\mathrm{rt}}$ and the root height is z^{rt}. The root orientation matrix is $\mathbf{R}^{\mathrm{rt}}$, the root's linear and angular velocities are $\mathbf{v}^{\mathrm{rt}}$ and $\boldsymbol{\omega}^{\mathrm{rt}}$, respectively. We denote rigid body positions as $\mathbf{x}^{b}$ and rigid body orientations as $\mathbf{R}^{b}$. The terms $\boldsymbol{\tau}_t^{\mathrm{jts}}$ and $\ddot{\mathbf{q}}_t$ are joint torques and accelerations. The policy actions are given by $\mathbf{a}_t^{\mathrm{jts}}$. Note that in this case $\mathbf{g}_t$ the retargeted trajectory and $\mathbf{s}_t$ denotes the simulation state of the downstream RL policy.

Name	Reward Term	Weight G1	Weight Lima
	Motion Tracking		
Root position xy	$-\lVert \mathbf{x}^{\mathrm{rt}}_{\mathbf{g}_t} - \mathbf{x}^{\mathrm{rt}}_{\mathbf{s}_t} \rVert_2^2$	5.0	5.0
Root height	$-(z^{\mathrm{rt}}_{\mathbf{g}_t} - z^{\mathrm{rt}}_{\mathbf{s}_t})^2$	5.0	5.0
Root orientation	$-\lVert \mathrm{Log}((\mathbf{R}^{\mathrm{rt}}_{\mathbf{s}_t})^T \mathbf{R}^{\mathrm{rt}}_{\mathbf{g}_t}) \rVert_2^2$	3.0	3.0
Root lin vel.	$-\lVert \mathbf{v}^{\mathrm{rt}}_{\mathbf{g}_t} - \mathbf{v}^{\mathrm{rt}}_{\mathbf{s}_t} \rVert_2^2$	0.5	0.5
Root ang. vel.	$-\lVert \boldsymbol{\omega}^{\mathrm{rt}}_{\mathbf{g}_t} - \boldsymbol{\omega}^{\mathrm{rt}}_{\mathbf{s}_t} \rVert_2^2$	0.5	0.5
Rbs position	$-\lVert \mathbf{x}^{b}_{\mathbf{g}_t} - \mathbf{x}^{b}_{\mathbf{s}_t} \rVert_2^2$	5.0	5.0
Rbs orientation	$-\lVert \mathrm{Log}((\mathbf{R}^{b}_{\mathbf{s}_t})^T \mathbf{R}^{b}_{\mathbf{g}_t}) \rVert_2^2$	2.5	2.5
Survival	1.0	10.0	1.0
	Regularization		
Joint torques	$-\lVert \boldsymbol{\tau}_t^{\mathrm{jts}} \rVert_2^2$	$1.0 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$
Joint acc.	$-\lVert \ddot{\mathbf{q}}_t \rVert_2^2$	$2.5 \cdot 10^{-8}$	$2.5 \cdot 10^{-6}$
Joint action rate	$-\lVert \mathbf{a}_t^{\mathrm{jts}} - \mathbf{a}_{t-1}^{\mathrm{jts}} \rVert_2^2$	0.15	3.0
Joint action acc.	$-\lVert \mathbf{a}_t^{\mathrm{jts}} - 2\mathbf{a}_{t-1}^{\mathrm{jts}} + \mathbf{a}_{t-2}^{\mathrm{jts}} \rVert_2^2$	$1.0 \cdot 10^{-2}$	1.0

Table 2. **Performance Variability (Lima).** Best and worst case (min/max of the per-motion mean) for the kinematic metrics reported in the main paper.

Metric	ReActor	OmniRetarget	GMR
Ground Pen. [cm]	**0.0** / **0.0**	0.0 / 0.0	0.0 / 9.3
Self Pen. [cm]	**0.0** / **0.0**	0.0 / 17.2	0.0 / 15.3
Foot Slide [cm/s]	**0.0** / 52.3	0.0 / 98.6	0.0 / 95.5
Foot Float [cm]	**0.0** / 9.4	0.0 / 21.4	0.0 / 32.4

A Downstream RL Policy Details

Following prior work, we evaluate the retargeting methods on a downstream tracking task by training RL policies on the retargeted motions, with tracking performance serving as a proxy for motion quality [Liao et al. 2025; Yang et al. 2025]. We train the RL policies following the DeepMimic framework [Peng et al. 2018], where the policy is conditioned on a motion reference and optimized using explicit tracking rewards. The reward terms used for training are detailed in Tab. 1.

We measure the success rate based on the training termination criteria, as proposed in [Yang et al. 2025]. A trial is considered a failure if the robot's root deviates by more than 1 m from the target root position or if the geodesic distance between the current and target root orientation exceeds 45°.

B Performance Variability

Tab. 2 reports best- and worst-case results (min/max of the per-motion mean) for Lima, complementing the mean and standard deviation reported in the main paper.

References

Qiayuan Liao, Takara E. Truong, Xiaoyu Huang, Yuman Gao, Guy Tevet, Koushil Sreenath, and C. Karen Liu. 2025. BeyondMimic: From Motion Tracking to Versatile Humanoid Control via Guided Diffusion. doi:10.48550/arXiv.2508.08241

Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel van de Panne. 2018. DeepMimic: example-guided deep reinforcement learning of physics-based character skills. *ACM Trans. Graph.* 37, 4 (2018). doi:10.1145/3197517.3201311

Lujie Yang, Xiaoyu Huang, Zhen Wu, Angjoo Kanazawa, Pieter Abbeel, Carmelo Sferrazza, C. Karen Liu, Rocky Duan, and Guanya Shi. 2025. OmniRetarget: Interaction-Preserving Data Generation for Humanoid Whole-Body Loco-Manipulation and Scene Interaction. doi:10.48550/arXiv.2509.26633

Authors' Contact Information: David Müller, david.cao@disneyresearch.com, Disney Research, Switzerland; Agon Serifi, agon.serifi@disneyresearch.com, Disney Research, Switzerland; Sammy Christen, sammy.christen@disneyresearch.com, Disney Research, Switzerland; Ruben Grandia, ruben.grandia@disneyresearch.com, Disney Research, Switzerland; Espen Knoop, espen.knoop@disneyresearch.com, Disney Research, Switzerland; Moritz Bächer, moritz.baecher@disneyresearch.com, Disney Research, Switzerland.