On the Transferability of Multi-Agent DRL architecture for a Physics-based Model of a Lower-Limb Amputee Across Varied Locomotion Environments

Lorenza Cotugno[1,*], Matthia Sabatelli[2], Roberta Siciliano[3] and Raffaella Carloni[2]

[1] *Department of Physics, University of Naples Federico II, Italy*
[2] *Department of Electrical Engineering and Information Technology, University of Naples Federico II, Italy*
[3] *Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, The Netherlands*

Abstract

This work investigates whether a multi-agent architecture trained on level-ground walking can facilitate learning and improve performance in more challenging locomotion environments. Specifically, the musculoskeletal model is trained using a multi-agent variant of Proximal Policy Optimization, combined with imitation learning, where the healthy part and prosthesis are modeled as separate agents that learn under a shared reward. The level-ground walking policy serves as the starting point for transfer. We fine-tune it across three more challenging environments, uneven terrain, ramps, and stairs, each equipped with task-specific reward functions and imitation data tailored to their respective locomotion objectives. We compare the fine-tuned policy against a baseline that is trained from scratch. Preliminary results suggest that transfer learning improves initial performance and yields higher rewards throughout the evaluation. When quantified using the area ratio metric $\mathcal{R}$, which compares the area under the learning curve of the pre-trained model to that of the baseline, transfer learning demonstrates a benefit of at least $\mathcal{R} = 0.50$ across all tested environments. Ongoing work explores bidirectional transfer by rotating the source environment to study how task complexity and inter-environment similarity affect performance. Future work will focus on reducing reliance on imitation data and designing more generalizable reward functions to support autonomous, robust adaptation across varied real-world conditions.

Keywords

Reinforcement Learning, Transfer Learning, Multi-Agent Systems, Biomechanical Simulation

1. Introduction

Artificial Intelligence (AI) has become increasingly relevant in the field of lower-limb rehabilitation, offering new opportunities for enhancing motor recovery and personalized assistance. A comprehensive literature review [1] highlights the growing integration of AI techniques, ranging from neural networks to reinforcement learning, in exoskeleton-assisted gait rehabilitation. At the same time, there is a rising emphasis on the need for intelligent systems to adapt to diverse terrain conditions. In this regard, [2] demonstrates that deep learning models can effectively

Workshop on Social Robotics for Human-Centered Assistive and Rehabilitation AI (a Fit4MedRob event) - ICSR 2025
*Corresponding author.
lorenza.cotugno@unina.it (L. Cotugno); m.sabatelli@rug.nl (M. Sabatelli); roberta.siciliano@unina.it (R. Siciliano); r.carloni@rug.nl (R. Carloni)

classify ground types and estimate parameters such as ramp inclination or stair height, using data collected from wearable sensors. Among the various AI approaches, reinforcement learning (RL) stands out as particularly promising for prosthetic control, enabling systems to learn motor behaviors through trial-and-error interaction with the environment.

Recent studies in assistive mobility have explored RL in various directions, including automatic tuning of prosthetic control parameters and optimization of state-based control strategies [3, 4, 5]. RL has also been used in simulated bipedal locomotion tasks, providing a foundation for the development of algorithms aimed at future prosthetic control and personalization.

Within musculoskeletal simulations for gait modeling, most prior approaches rely on a single-agent formulation, where a unified policy governs both the healthy part and the prosthetic limbs. Although this setup can be effective in simulation, it is less suitable for real-world deployment. In [6], a notable multi-agent alternative was introduced, modeling the human and the prosthesis as separate agents trained concurrently but independently, thereby fostering collaboration to achieve adaptive and natural gait patterns. However, this approach explicitly imposed symmetry between the prosthetic and contralateral limbs.

Building on this perspective, previous work [7] proposed a multi-agent reinforcement learning (MARL) architecture in which the human and the prosthesis are treated as independent agents. Each agent is trained using Independent Proximal Policy Optimization (IPPO) combined with imitation learning, and their coordination is guided by a shared reward. The agents operate in distinct observation spaces and, unlike previous approaches, no symmetry is enforced. Instead, coordination and gait patterns are learned through interaction and imitation.

By reusing the neural network weights of a policy previously trained to achieve stable locomotion on level ground, this work investigates whether transfer learning can facilitate training and enhance performance in more complex locomotion environments, such as stairs, ramps, and uneven terrain. Each target environment features its own task-specific reward function and imitation dataset. We compare the pretrained model against a baseline trained from scratch in each setting, focusing on initial rewards and convergence speed.

2. Physics-based Model

This section presents the physics-based musculoskeletal model of the transfemoral amputee, which is used as the agent of the MARL, and the contact model, which has been created to perfom walking on complex terrains.

Musculoskeletal Model

We adopt the *gait1415+2* musculoskeletal model [8], developed in OpenSim 4.2, to simulate a transfemoral amputee. The model features 14 degrees of freedom (DOFs). The intact limb is driven by 15 Hill-type musculotendon units, while the prosthetic side is actuated by ideal torque generators in the knee and ankle joints [9], using first-order activation dynamics. This setup captures key biomechanical features of the human and prosthesis, and allows realistic simulation of neuromuscular control strategies.

Contact Model

To enable realistic interaction with structured and irregular terrains, we implemented a custom contact model inspired by [10]. Although the musculoskeletal model includes anatomical foot segments, it lacks the explicit contact geometry necessary to simulate physical interactions with the environment. Therefore, spherical meshes were added to the feet to define discrete contact regions, enabling the computation of ground reaction forces. Furthermore, OpenSim's default Hunt–Crossley model was replaced with an elastic foundation formulation to ensure stable and biomechanically plausible contact dynamics [10]. The geometry of the foot meshes is shown in Figure 1. In addition, Figure 3 shows the full model interacting with stair, ramp, and uneven terrain surfaces.

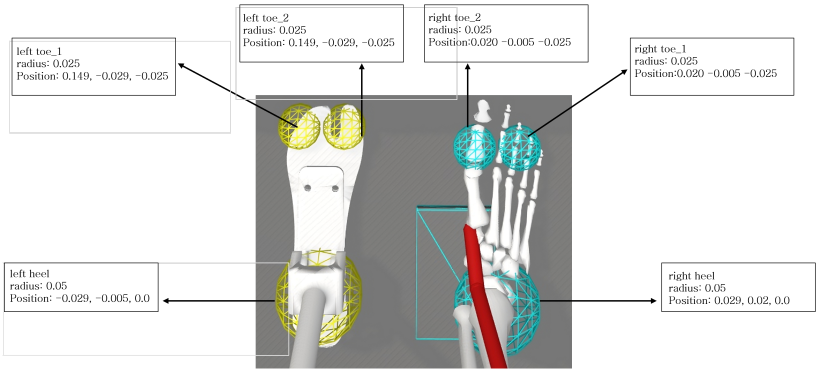
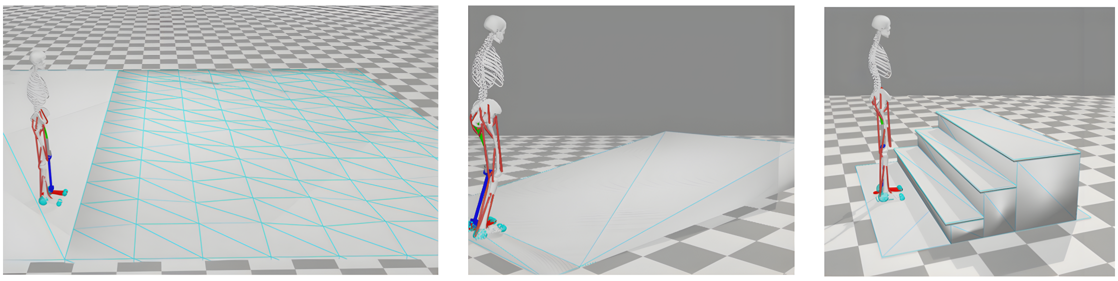


Figure 1: Bottom view of the spherical contact meshes for the heels and toes with respect to the bones in the feet.



(a) Uneven terrain (b) Ramp ascent (c) Stair ascent

Figure 2: Simulation environments used in the transfer learning experiments.

3. The Proposed Methodology

This section outlines the proposed methodology, which employs a collaborative multi-agent reinforcement learning architecture. The *gait1415+2* musculoskeletal model [8] is split into two agents (i.e., the prosthesis and the contralateral healthy human part) which interact within a shared physics-based simulation. Rather than retraining level-ground walking, we initialize each target task (i.e, uneven terrain, ramp, and stairs ascent) with the pretrained policy from [7] and compare its performance against policies trained from scratch.

Multi-Agent Reinforcement Learning Architecture

This work builds on the MARL architecture introduced in [7], which applies Independent Proximal Policy Optimization (IPPO + imitation data). The healthy human agent receives information about the full state of the body, including joint positions and velocities of the pelvis and lower limbs, ground reaction forces (GRFs), and muscle-related variables such as activation, fiber length, and force. In contrast, the prosthesis agent operates under more realistic sensing constraints and observes only signals available from on-board sensors: joint angles and velocities in the prosthetic knee and ankle, GRFs under the prosthetic foot, and internal actuator states (torque, power, control signal, velocity, activation). This asymmetry in observations reflects the limited sensory capabilities typically available in robotic prosthetic systems.
The agents also differ in their action spaces:

- The human agent outputs a 15-dimensional vector corresponding to the activation levels of the musculotendon units on the intact limb.
- The prosthesis agent controls two actuators (in the prosthetic knee and ankle) via discrete activation commands.

The total reward at each timestep integrates both the imitation and task-level objectives.

$$r_t = 0.9 \cdot r_{\text{imit},t} + 0.1 \cdot r_{\text{goal},t} \tag{1}$$

The imitation component encourages the reproduction of a reference trajectory and consists of two terms: one penalizing deviations in joint positions, the other in joint velocities. Together, these terms promote coordinated and biomechanically plausible movements.
The task-related component, in parallel, reinforces the successful execution of the locomotion objective. Its formulation is environment-specific: for elevation-based tasks, such as stairs and ramps, the reward emphasizes pelvis velocity to reflect dynamic performance; for level and irregular ground, it prioritizes pelvis position accuracy.
By combining these components, the reward function guides the agent toward realistic and functionally effective motion strategies.

PPO and Imitation Dataset

We used motion capture data to guide imitation learning across tasks. Flat-ground walking was modeled using a public pediatric gait dataset with kinematic and kinetic data from typically developing children aged 10–12 [11]. Stairs and ramp ascent trajectories were obtained from the CMU Motion Capture Database [12], using the same trials as in [10]. Specifically, we used lower-body motion data from subject 14 (trial 22) for stairs ascent and subject 74 (trial 19) for ramp ascent. These trials were selected for their completeness, marker quality, and task consistency. All signals were temporally normalized and scaled to match the OpenSim model.

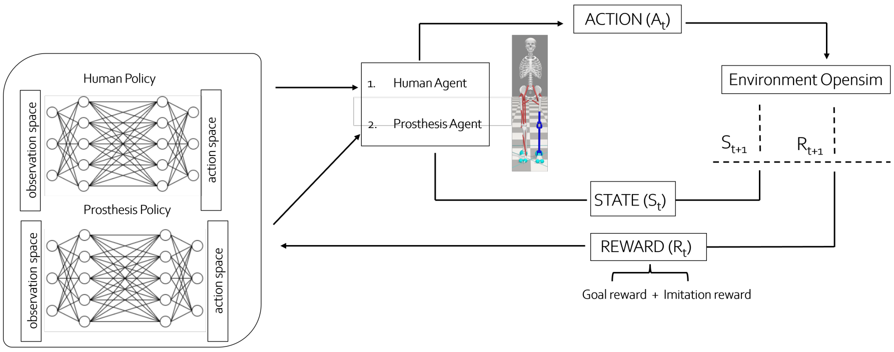


Figure 3: Overview of the reinforcement learning framework in the OpenSim simulation environment. At each timestep t, the environment provides the current state S_t to both the human and prosthesis agents. Each agent generates an action based on its policy; the agents' actions are applied in the simulator, which returns a reward R_t and the next state S_{t+1}. The reward, composed of a goal-oriented term and an imitation term, is used to improve the agents' policies toward more optimal behavior.

Transfer Learning

We compare two training strategies: (i) a baseline approach, in which agents are trained from scratch with randomly initialized neural network weights on each $\mathcal{E}_T$; and (ii) a transfer learning approach, in which the policy is initialized using weights trained on a simpler *source environment* $\mathcal{E}_S$, corresponding to level-ground walking. In both conditions, training is conducted using task-specific reward functions and imitation data tailored to each environment. The results suggest that transferring a well-trained policy from $\mathcal{E}_S$ improves initial performance and yields higher rewards throughout the evaluation.

Evaluation Metric

To quantitatively assess the benefit of transfer learning, we compute the area under the learning curve (AUC) for each training run and define the *area ratio metric* $\mathcal{R}$, following the approach in [13], as follows:

$$\mathcal{R} = \frac{\mathrm{AUC}_{\mathcal{E}_S \to \mathcal{E}_T} - \mathrm{AUC}_{\mathcal{E}_T}}{\mathrm{AUC}_{\mathcal{E}_T}} \tag{2}$$

Here, the area under the curve is approximated via trapezoidal numerical integration:

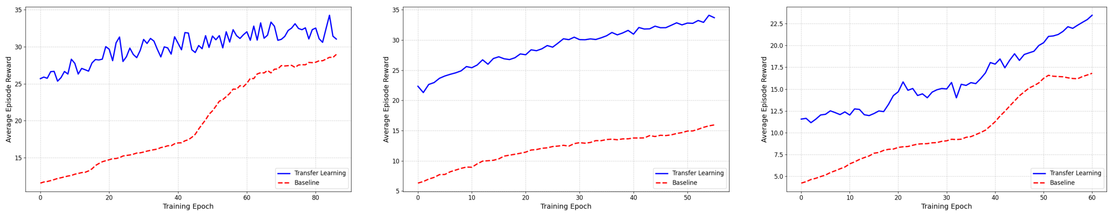
$$\int_a^b f(x)\,dx \approx (b-a) \cdot \frac{1}{2}\left(f(a) + f(b)\right) \tag{3}$$

with a and b corresponding to the first and last training epochs, respectively. A value of $\mathcal{R} > 0$ indicates a performance gain due to transfer learning.

4. Results

This section presents preliminary results evaluating the effectiveness of the transfer learning strategy in three challenging target environments denoted $\mathcal{E}_T$: uneven terrain, ramp, and stairs.

This effect is illustrated in Figures 4a, 4b, and 4c, which show the episode reward curves for both training strategies across all three $\mathcal{E}_T$ settings. In an uneven terrain environment, the initialized transfer scenario begins with approximately 25 reward points, more than twice the initial score of the baseline of about 11, and maintains this advantage throughout training. In the ramp environment, the transfer case starts near 22, compared to just over 6 for the baseline. After 50 epochs, it achieves nearly 30 reward points, while the baseline reaches just under 16. In the stairs environment, the baseline starts at approximately 4.5 and the transfer at 11, again showing an initial benefit. The computed ratio $\mathcal{R}$ confirms the advantage of transfer learning across environments: 0.516 for uneven terrain, 1.421 for ramps, and 0.560 for stairs. All agents were trained for an equal duration of six hours.



(a) *Uneven* ($\mathcal{R}$ = 0.516) (b) *Ramp* ($\mathcal{R}$ = 1.421) (c) *Stairs* ($\mathcal{R}$ = 0.560)

Figure 4: Episode reward curves for the three target environments. Transfer learning improves both convergence speed and cumulative reward across all settings, as reflected by the ratio $\mathcal{R}$.

5. Conclusions and Future Work

This study demonstrates the potential of transfer learning to improve efficiency and early performance in training prosthetic control policies across diverse locomotion environments. By leveraging a policy trained on level-ground walking as a source, we observed consistent gains across ramp, uneven terrain, and stairs tasks. These preliminary results are promising, suggesting that transfer from a simpler source environment can significantly accelerate learning and yield more stable training dynamics. Ongoing work explores bidirectional transfer by rotating the source and target environments (e.g., from stairs to ramp and vice versa), allowing us to assess how task complexity and inter-environment similarity influence transfer effectiveness. In particular, we aim to investigate whether transfer is symmetric or whether certain source-target pairs yield more generalizable behavior. A more extensive evaluation, including longer training runs and additional terrain combinations, will be presented at the conference.
In the current setup, imitation learning plays a key role. Future research will focus on strategies to gradually reduce the influence of imitation during fine-tuning, fostering more autonomous and robust behavior. Promoting such autonomy is essential to generalize across broader locomotion tasks, including variations in ramp steepness, stair height, or surface irregularity. Taken together, these findings lay the foundation for environment-aware transfer learning frameworks for embodied prosthetic agents.

Acknowledgments

The work of Lorenza Cotugno and Roberta Siciliano was supported by the Italian Ministry of Research, under the complementary actions to the NRRP "Fit4MedRob - Fit for Medical Robotics" Grant PNC0000007, (CUP: B53C22006990001). The work of Raffaella Carloni was supported by the European Commission's Horizon 2020 Programme as part of the project MyLeg under grant no. 780871.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to check grammar and spelling, paraphrase and reword. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] O. Coser, C. Tamantini, P. Soda, L. Zollo, Ai-based methodologies for exoskeleton-assisted rehabilitation of the lower limb: a review, Frontiers in Robotics and AI 11 (2024) 1341580.

[2] O. Coser, C. Tamantini, M. Tortora, L. Furia, R. Sicilia, L. Zollo, P. Soda, Deep learning for human locomotion analysis in lower-limb exoskeletons: A comparative study, arXiv preprint arXiv:2503.16904 (2025).

[3] Y. Wen, M. Li, J. Si, H. Huang, Wearer-prosthesis interaction for symmetrical gait: A study enabled by reinforcement learning prosthesis control, IEEE Transactions on Neural Systems and Rehabilitation Engineering 28 (2020) 904–913.

[4] X. Gao, J. Si, Y. Wen, M. Li, H. H. Huang, Knowledge-guided reinforcement learning control for robotic lower limb prosthesis, in: IEEE International Conference on Robotics and Automation (ICRA), 2020, pp. 754–760.

[5] R. Wu, M. Li, Z. Yao, J. Si, H. Huang, Reinforcement learning enabled automatic impedance control of a robotic knee prosthesis to mimic the intact knee motion in a co-adapting environment, arXiv preprint arXiv:2101.03487 (2021).

[6] R. Wu, J. Zhong, B. Wallace, X. Gao, H. Huang, J. Si, Human–robotic prosthesis as collaborating agents for symmetrical walking, in: Advances in Neural Information Processing Systems (NeurIPS) 35, 2022, pp. 27306–27320.

[7] D. Petrillo, R. Siciliano, , M. Sabatelli, R. Carloni, Multi-agent reinforcement learning control for an osseointegrated transfemoral amputee physics-based model, 2025. In review.

[8] R. Carloni, R. Luinge, V. Raveendranathan, The gait1415+2 opensim musculoskeletal model of transfemoral amputees with a generic bone-anchored prosthesis, Medical Engineering & Physics 123 (2024) 104091.

[9] V. Raveendranathan, V. G. Kooiman, R. Carloni, Musculoskeletal model of osseointegrated transfemoral amputees in opensim, PLOS ONE 18 (2023) e0288864.

[10] A. J. C. Adriaenssens, V. Raveendranathan, R. Carloni, Learning to ascend stairs and ramps: Deep reinforcement learning for a physics-based human musculoskeletal model, Sensors 22 (2022) 8479. doi:10.3390/s22218479.

[11] M. H. Schwartz, A. Rozumalski, J. P. Trost, The effect of walking speed on the gait of typically developing children, Journal of Biomechanics 41 (2008) 1639–1650.

[12] CMU Graphics Lab, Cmu graphics lab motion capture database, http://mocap.cs.cmu.edu/, 2005–2006. Motion capture recordings from over 140 subjects; data used in many biomechanics and AI studies. If published, please cite source and send notification to jkh%2Bmocap@cs.cmu.edu.

[13] M. Sabatelli, Contributions to Deep Transfer Learning: From Supervised to Reinforcement Learning, Ph.D. thesis, University of Liège, Faculté des Sciences Appliquées, 2022.