Journal of NeuroEngineering and Rehabilitation

RESEARCH

Open Access

Robust neural network controller for a lower limb exoskeleton with minimal sensor configuration: a deep reinforcement learning approach with policy distillation

Shuzhen Luo[1*], Ke Feng[1], Ghaith Androwis[2,3,4], Sergei Adamovich[2], Erick Nunez[2], Hao Su[5] and Xianlian Zhou[2*]

Abstract

Background Lower limb rehabilitation exoskeletons (LLREs) are increasingly used in clinical settings to improve mobility in individuals with neuromuscular impairments. Most LLREs employ controllers focused on trajectory tracking, which lack adaptability to user-specific variations or voluntary effort. Autonomous LLREs enable hands-free gait but often require a multitude of sensors such as IMU, encoders, and foot force sensors to enable user intent prediction, gait phase detection, and dynamic balance. This often introduces substantial challenges related to cost, complexity, reliability, and system scalability.

Methods In this study, we introduce a novel deep reinforcement learning (DRL) based approach for autonomous control of a custom-designed LLRE using a minimal sensor configuration, enabled through a privileged teacher–student policy distillation paradigm. Policies are trained in a physics-based simulation environment integrating a full-body musculoskeletal model and an exoskeleton interaction model. The privileged teacher control policy leverages privileged full-state information to learn stable walking behaviors, while the student control policy learns to replicate the privileged control policy's behavior via policy distillation. The student policy uses only proprioceptive signals derived from joint encoders, enabling direct deployment on physical hardware with minimal sensor requirements.

Results and conclusion We evaluate both privileged teacher and student control policies in simulated walking scenarios under external disturbances. Performance metrics such as gait symmetry and lateral stability confirm that the student policy, despite relying solely on encoder data, achieves comparable performance to the teacher policy and remains robust to disturbances. Further comparisons with sensor-rich configurations, including those incorporating IMU-based orientation and foot force sensor derived center-of-pressure (CoP), show minimal performance degradation under the joint encoder only configuration. These results highlight that robust LLRE control can be achieved with substantially reduced sensing demands. Our method supports seamless sim-to-real

*Correspondence:
Shuzhen Luo
LUOS@erau.edu
Xianlian Zhou
alexzhou@njit.edu

Full list of author information is available at the end of the article

transfer, simplifies hardware integration, reduces calibration and fault risks, and enhances the practicality of deploying autonomous exoskeletons in both clinical and real-world environments.

Keywords Autonomous walking control, Human–exoskeleton interactions, Deep reinforcement learning, Sim-to-real transfer

Introduction

Lower limb rehabilitation exoskeletons (LLREs) have emerged as promising assistive technologies for restoring mobility in individuals with neuromuscular disorders such as muscle weakness, paralysis, or spinal cord injury [1–8]. They have been increasingly used nowadays in rehabilitation clinics to support patients with mobility disorders by providing mechanical support to the lower limbs, which facilitates gait training, improves walking ability, and enhances the overall quality of life for patients undergoing rehabilitation [3, 4, 9–11]. Ensuring robustness and stability in LLREs for walking assistance is crucial for patient safety. Many current LLREs require additional support like crutches or supervision from a healthcare provider to prevent falls during locomotion. Some autonomous LLREs, such as Rex (Rex Bionics) [12] and Atalante (Wandercraft) [13], offer independent walking capabilities but often at the cost of slower speeds and increased weight and price. Enabling autonomous locomotion with LLREs can significantly boost patient confidence in clinical and home settings. Advanced controllers that robustly manage walking assistance under various human–exoskeleton interaction conditions are needed to achieve this goal.

Current LLRE controllers often prioritize trajectory tracking, which is essential in early-stage rehabilitation when patients have limited muscle strength. Devices such as Lokomat and ReWalk [14] utilize preprogrammed joint trajectories to guide the user's limbs through normative gait cycles, thereby enabling repetitive, task-specific training that promotes neuroplasticity. However, while effective in early rehabilitation in providing consistent motion patterns, these systems tend to lack responsiveness to real-time user interaction, residual muscular effort, or unique gait adaptations that emerge during recovery. The rigid nature of trajectory tracking may discourage active user participation, suppress muscle engagement, and impede progression to more natural, self-initiated walking.

More recently, LLREs have evolved to support increasingly autonomous and hands-free locomotion that is stable under voluntary or involuntary user interactions, enabled by sophisticated sensor integration and control algorithms. For example, Wandercraft's Atalante uses joint encoders, multiple IMUs, and foot force sensors to enable hands-free, autonomous walking. These sensors, with sophisticated fusion algorithms [15], support real-time posture estimation, gait phase detection, and dynamic balance control using model-based strategies [16–18]. Some exoskeletons, like Cyberdyne's HAL, also use surface EMG to detect voluntary muscle activation and user intent. This enables user-intent-driven assistance, particularly effective for rehabilitation in users with partial motor function [19]. While effective, these designs and control methods come with notable trade-offs: increased cost, higher system complexity, susceptibility to sensor failure, and longer calibration processes. These limitations are particularly restrictive in clinical or resource-limited settings, where affordability, reliability, and ease of use are critical. Furthermore, sensor degradation over time, such as IMU drift or foot sensor wear, can compromise safety and control fidelity.

Recent advancements in control strategies have introduced more robust and adaptable approaches for LLREs, particularly through the use of AI and learning-based methods [20–25]. For instance, Bingjing et al. [23] proposed a reinforcement learning (RL)-based admittance controller that adaptively tuned interaction parameters in response to human–exoskeleton dynamics during gait rehabilitation. Luo et al. [26–28] developed RL-based neural network (NN) controllers for squatting and walking motions, trained in a tightly coupled human–exoskeleton simulation environment. The resulting neural network-based control policy demonstrated strong robustness across a wide range of neuromuscular impairments, including quadriplegia, generalized muscle weakness, and hemiparesis. However, the controller utilized rich sensory inputs, including the user's lower limb motion state, exoskeleton proprioception, and foot center of pressure (CoP). The reliance on high-dimensional and difficult-to-measure human state information poses significant barriers to real-world deployment. Designing a sensor-efficient controller that achieves robust and autonomous walking assistance with minimal sensory input offers a promising solution to the limitations of current LLRE systems. Such a design enables easier and more robust transfer of the learned controller to physical hardware, removes the need for patient-specific calibration, and significantly improves adaptability in both clinical and home environments.

In this paper, we propose a deep reinforcement learning-based, privileged teacher–student learning control framework for a lower limb rehabilitation exoskeleton to achieve robust and autonomous walking locomotion using minimal sensor inputs. A privileged (teacher) control policy is first trained in simulation with access to

privileged information, such as reference trajectories, full-body kinematics, and center of pressure to learn robust and stable walking locomotion assistance. A student control policy is then trained to imitate the privileged control policy using only realistic onboard sensor inputs (e.g., joint encoders). Designed for deployment, the student control policy inherits the robustness of the privileged control policy while operating under partial observation. The student control policy can be directly deployed on the physical robot hardware.

Privileged learning has demonstrated strong robustness in legged robots such as quadrupeds and bipeds, particularly when navigating diverse and unstructured terrains [29–31]. This approach is especially advantageous for rehabilitation exoskeletons, as it enables stable and adaptive assistance using minimal sensing, making it better suited to accommodate user variability and complex human–exoskeleton interactions without requiring detailed physiological measurements. However, its application in LLREs remains limited. This is primarily due to the highly variable and unpredictable nature of human–exoskeleton interaction, the difficulty in modeling human musculoskeletal dynamics accurately, and the impracticality of obtaining privileged information from real users. Our work aims to bridge the gap between simulation-trained RL controllers and real-world deployment by introducing a deep reinforcement learning-based, privileged teacher–student learning control method that enables autonomous, crutch-free walking on a lightweight LLRE platform using only joint encoder data. This controller requires no manual parameter tuning or patient-specific calibration, facilitating seamless transfer to physical hardware and supporting deployment in both clinical and home rehabilitation settings.

Human–exoskeleton interaction modeling

The LLRE shown in Fig. 1 was developed in an earlier research [32] to support gait rehabilitation with a more comprehensive description of the exoskeleton design provided in our previous work [26]. We built a simulation environment for LLRE–human interaction on the open-source DART library [33]. The simulation environment

Fig. 1 The physical structure of the full lower-limb rehabilitation exoskeleton system and the corresponding human musculoskeletal–-exoskeleton model used in the simulation environment

operates at a time integration frequency of 600 Hz, with control inputs for both the exoskeleton and the human model updated at 30 Hz. The musculoskeletal model, shown in Fig. 1, is approximately 170 cm tall, weighs 72 kg, and includes 50 degrees of freedom (DoFs) and 284 musculotendon units.

Musculoskeletal modeling

A full-body human musculoskeletal model used in [27, 28, 34] is integrated with the LLRE to create realistic human–exoskeleton interaction forces and constraints. The Euler–Lagrangian equations governing the dynamics of the human musculoskeletal system, expressed in terms of generalized coordinates, are formulated as:

$$\boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{c}(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \boldsymbol{J}_m^T \boldsymbol{f}_m(\boldsymbol{a}) + \boldsymbol{J}_{ext}^T \boldsymbol{f}_{ext} \quad (1)$$

where $\boldsymbol{q}$ is the vector of joint angles, $\boldsymbol{f}_{ext}$ is the vector of external forces, and $\boldsymbol{f}_m$ is the vector of muscle forces which is a function of muscle activations $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$ for all muscles. For a subject with minimal or no active limb movement (e.g., a quadriplegic patient) as modeled in this study, all muscle activations can be set to zero. $\boldsymbol{M}(\boldsymbol{q})$ denotes the generalized mass matrix, and $\boldsymbol{c}(\boldsymbol{q}, \dot{\boldsymbol{q}})$ is Coriolis and gravitational forces. $\boldsymbol{J}_m$ and $\boldsymbol{J}_{ext}$ are Jacobian matrices that map the muscle and external forces to the joint space, respectively.

Modeling of the LLRE

Our LLRE system includes eight actuated degrees of freedom (DoFs), with each leg containing one DoF for hip flexion/extension, one for knee flexion/extension, and two for ankle motion. The joints are powered by Dynamixel Pro H54-200-S500-R smart servo motors. Both hip and knee joints can provide continuous torque output of 132 Nm and a rotational speed of $55°$/s. At the peak, these joints can deliver torque exceeding 220 Nm for a short duration. The hip joint offers a motion range from $-80°$ (extension) to $80°$ (flexion), while the knee joint ranges from full extension ($0°$) to $160°$ of flexion. Unlike many commercial exoskeletons that use passive or rigid ankle joints, this system includes a powered 2-DoF ankle module capable of both dorsiflexion/plantarflexion and inversion/eversion, generating torque above 160 Nm [27]. For the ankle joints, the plantarflexion/dorsiflexion axis allows a range of motion corresponding to $40°$ of plantarflexion and $15°$ of dorsiflexion. The inversion/eversion axis provides a symmetric range from $-15°$ to $15°$. These two ankle movements are achieved through distinct rotational axes and are driven by a closed-loop configuration involving two motors, with universal and screw joints. The exoskeleton weighs 20.4 kg, with most of the components 3D printed in Onyx (Markforged) and continous carbon fiber.

To model the physical coupling between the human musculoskeletal system and the lower limb exoskeleton, straps are placed around the pelvis, femur and tibia, as shown in Fig. 1. At each strap location, linear bushings [35] are used to simulate the interaction forces and torques. Each bushing connects a frame on the exoskeleton to a corresponding frame on the human body using direction-specific linear and rotational springs and dampers. These bushings generate forces and moments based on the relative displacement and velocity between the frames:

$$\begin{cases} f_i = k_i x_i + c_i \dot{x}_i \\ \tau_i = \alpha_i \theta_i + \beta_i \dot{\theta}_i \quad (i = x, y, z) \end{cases} \tag{2}$$

Directional stiffness and damping constants are tuned to reflect strap behavior. For example, the pelvic bushing only acts in the vertical direction with $k_y = 8000$, $c_y = 10$. At the straps, softer stiffness is used along the limb axis (e.g., $k_y = 500$) to allow natural sliding and rotation. The foot-to-exoskeleton interface is modeled as rigid due to tight attachment.

Methods

Privileged teacher–student control policy learning for autonomous gait control with minimal sensor inputs

An overview of our controller training framework is illustrated in Fig. 2. The framework follows a sim-to-real learning paradigm with policy distillation, where a privileged teacher control policy is first trained in simulation using privileged information, and then distilled into a deployable version that operates under minimal sensing. All training and testing procedures are performed on a desktop computer with an NVIDIA RTX 4080 GPU.

The initial training phase leverages reinforcement learning (RL) with access to rich privileged observations, including a reference joint trajectory (target motion), center of pressure (CoP), root (exoskeleton base) orientation, and historical joint states, to develop a robust and stable locomotion controller. This privileged teacher control policy, trained with this full-state information, enables efficient learning of human–exoskeleton interaction dynamics and generalized walking behaviors. The resulting control policy is distilled into a student control policy for the exoskeleton that relies solely on onboard

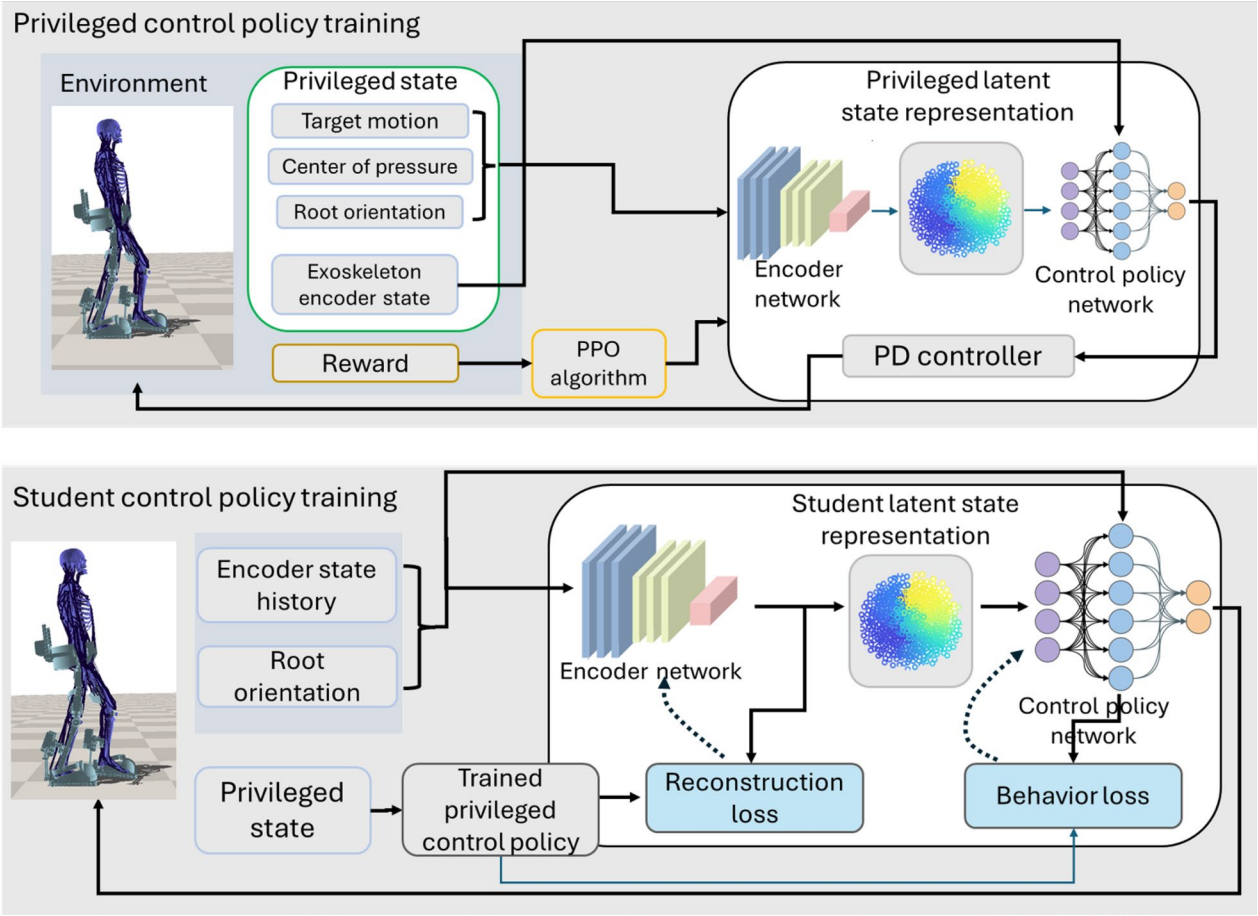

Fig. 2 Overview of the lower limb rehabilitation exoskeleton controller with privileged control policy and student control policy training

sensor signals, enabling deployment on physical hardware with minimal sensing requirements.

Privileged (teacher) control policy training

The privileged observation state is defined as $s_t = [q_t, \dot{q}_t, O_{\text{root}}, \text{CoP}_{\text{left}}, \text{CoP}_{\text{right}}, p_{\text{target}}]$, where privileged information such as exoskeleton root (base) orientation O_{root}, the positions of CoPs, target motion data p_{target}, as well as exoskeleton joint angles (q_t) and velocities ($\dot{q}_t$) are accessible to the privileged control policy to facilitate more efficient and stable training. The action space A_t of the privileged control policy network in Fig. 2 is an 8-dimensional vector representing the desired joint positions of the eight actuated DoFs.

The motor torques τ are computed using a proportional-derivative (PD) controller: $\tau = k_p(a_{s,t} - p_t) - k_v\dot{p}_t)$, where $a_{s,t}$ is the desired joint position, p_t is the current joint position, and $\dot{p}_t$ is the joint velocity. The control gains k_p and k_d are set to 400 and 28.28, respectively. The neural network configuration and interactions in Privileged control policy training are shown in Fig. 3. The teacher controller operates with privileged state information, including target motion, center of pressure, root orientation, and exoskeleton encoder states. These inputs are passed through a two-layer encoder network, and the extracted features are forwarded to both the policy and value networks. The three-layer policy network generates the control actions (reference joint positions) to drive the PD controller, while the three-layer value network predicts the scalar value function for each state. The Proximal Policy Optimization (PPO) algorithm updates both networks jointly, with the reward signal and value estimates contributing to stable policy learning. In this setup, the value network serves as a training baseline, whereas the policy network governs action generation. The privileged teacher control policy training process consists of the following components:

- Latent Encoder (policy_enc): A two-layer multilayer perceptron (MLP) that maps the privileged information (dimension $n_{\text{target}} = 84$) into a 20-dimensional latent embedding. The encoder uses ReLU activation functions.
- Control Policy Network (policy_net): A three-layer MLP that receives the concatenated latent embedding and the current system state (total input dimension = 50) and outputs the mean of a Gaussian distribution over actions (dimension $n_{\text{action}} = 8$). The network consists of an input layer with 50 neurons, two hidden layers with 128 neurons each, and an output layer with 8 neurons, corresponding to actions for the eight joints involved in the walking control task.
- Value Network: A separate three-layer MLP that receives the concatenated reference targets and system state and outputs a scalar value estimate of the expected future rewards. Its main role is to act as a critic that is crucial for calculating the advantage of actions and guiding the policy network to favor actions that lead to better than expected outcomes. This value network is trained jointly with the policy network under the Proximal Policy Optimization (PPO) framework, where it serves to approximate the state-value function $V(s)$. By providing a baseline for advantage estimation, the value network reduces the variance of gradient updates and stabilizes policy learning.
- Weight Initialization: All weights are initialized using the Xavier uniform initialization method [36], and all biases are initialized to zero.

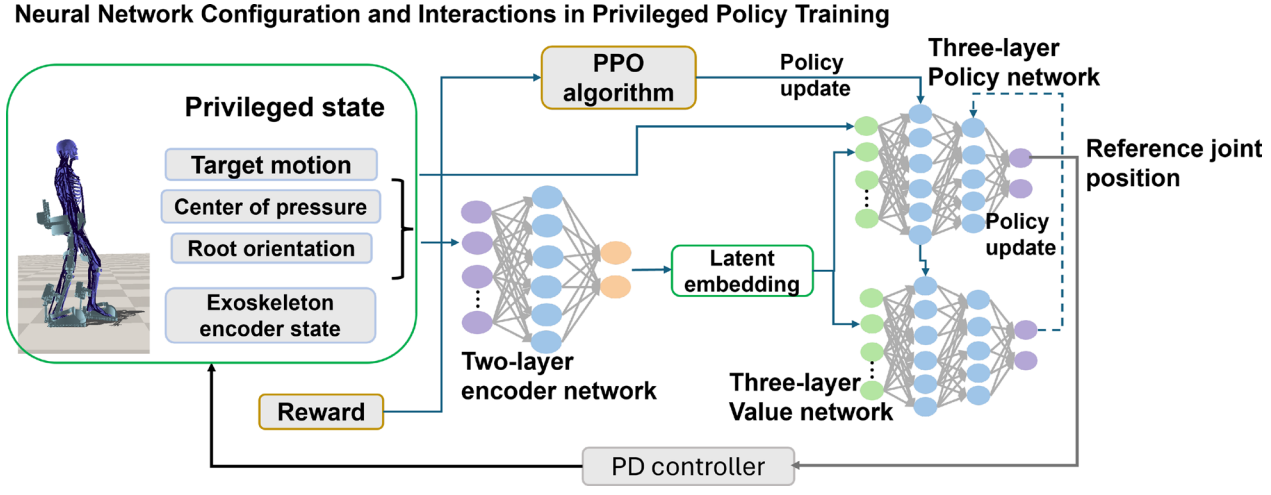

Fig. 3 Neural network configuration and interactions in privileged control policy training

- Action Filter: A custom action filter module is used to smooth action outputs and suppress high-frequency noise during deployment.

This privileged (teacher) control policy training process is formulated as an optimal control problem because the objective is to learn a policy that generates optimal actions that track desired joint trajectories while satisfying constraints related to stability, smoothness, and physical feasibility. To optimize the policy within this framework, we employ the Proximal Policy Optimization (PPO) algorithm [37]. PPO enables the privileged control policy to learn optimal actions based on privileged state information and maximize the reward function. The reward function is designed to encourage the privileged controller to autonomously learn stable and smooth walking locomotion under human–exoskeleton interaction and external disturbances, rather than simply mimicking a reference joint trajectory.

The reward r_t is defined as a weighted sum of several sub-rewards:

$$\begin{aligned} r_t =& w^{imitation} r_t^{imitation} + w^{root} r_t^{root} + w^{smooth} r_t^{smooth} \\ &+ w^{cop} r_t^{cop} + w^{fc} r_t^{fc} \end{aligned} \tag{3}$$

where $w^{imitation}$, w^{root}, w^{smooth}, w^{cop} and w^{fc} are the corresponding weights:

- Imitation reward $r^{imitation}$: Imitation reward is used to encourage the agent to reach the reference joint positions more accurately:

$$r_t^{imitation} = r_t^j * r_t^{ee} \tag{4}$$

$$r_t^j = exp[-\sigma_j \sum_i \left\| \hat{j}_t{}^i - j_t{}^i \right\|^2] \tag{5}$$

$$r_t^{ee} = exp[-\sigma_{ee} \sum_k \left\| \hat{ee}_t{}^i - ee_t{}^i \right\|^2] \tag{6}$$

where joint position reward r_j and end-effector reward r_{ee} are terms that increase exponentially with the squared differences between the current positions and the target positions. i is the joints index, and k is the end-effector positions index. The ( ) denotes the target positions.

- Root reward r^{root}: This term encourages the controller to maintain stability in the orientation and the position of the exoskeleton's base (that is strapped to the human pelvis), which is often crucial for robust walking locomotion tasks. It is described by an exponential function that increases with the squared differences in the orientation and the position of the exoskeleton's root joint:

$$r_t^{root} = exp[-\sigma_{root}(\left\| \hat{ro}_t{}^i - ro_t{}^i \right\|^2 - \left\| \hat{xz}_t{}^i - xz_t{}^i \right\|^2)] \tag{7}$$

where *ro* is the root orientation, and *xz* is the root position along the lateral and forward axes.

- Smoothness reward r^{smooth}: This sub-reward encourages the controller to produce smoother torques, contributing to more natural and robust walking locomotion. It is defined as the contribution of several smoothness-related penalties to the overall reward:

$$r_t^{smooth} = exp[-\sigma_{smooth} \left\| torque_t - 2torque_{t-1} + torque_{t-2} \right\|^2] \tag{8}$$

- r^{smooth} is the second-order finite differences between consecutive actions in three time steps.
- Stability reward r^{cop}: For CoP, the stable region S is defined as a rectangular area around the geometric center of the foot. The dimensions of this rectangle are configured with a width of 7 cm and a length of 11 cm. Notably, the width is narrower in the lateral direction compared to the forward direction. The $D(.,.)$ symbol denotes the Euclidean distance between the current CoP and the center of the stable region S.

$$r_t^{cop} = \begin{cases} exp[-\sigma_{cop} \left\| D(cop_t, S_{center}) \right\|^2], & \text{if } cop_t \in S \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

- Foot clearance reward r^{fc}: encourage the agent to maintain parallel alignment of the foot with the ground during walking by minimizing the roll and pitch angles Θ_X and Θ_Z along the lateral and forward axes. This helps prevent the foot from tripping.

$$r_t^{fc} = exp[-\sigma_{fc} \left\| sin(\Theta_X, \Theta_Z) \right\|^2] \tag{10}$$

Student control policy training

For effective sim-to-real transfer, the student policy is designed to rely on minimal sensor inputs, mimicking real-world deployment conditions. The LLRE provides joint encoder measurements, including joint angles q_t and angular velocities $\dot{q}_t$ for the hip, knee, and ankle joints. Additionally, a single inertial measurement unit (IMU) mounted on the base of the exoskeleton provides

its root (base) orientation, which is crucial for detecting balance during walking. The student observation o_t is defined as $o_t = [q_t, \dot{q}_t, O_{root}, A_{t-H:t}]$, where $A_{t-H:t}$ represents the history of previous actions in three time steps.

In contrast to the privileged teacher control policy, which has access to privileged information such as reference trajectories and full internal state available in the simulation, the student control policy relies solely on sensor-derived inputs. During training, the student control policy must learn to adapt its network parameters to replicate the privileged teacher's actions using only limited observations. Once trained, the student control policy can independently generate control commands with comparable accuracy to the teacher, without requiring access to the reference trajectory. The neural network configuration and interactions in student control policy training are shown in Fig. 4. The student controller receives only observable states, including the encoder state history and root orientation, as input. These states are encoded into compact feature representations by a two-layer encoder network and then passed through the three-layer policy network, which outputs actions converted to torque commands via a PD controller to the simulator. The student control policy is trained using a supervised learning approach, with the objective of minimizing the discrepancy between the latent vectors and output actions generated by the student and those produced by the privileged teacher control policy. The loss function is defined as the Mean Squared Error (MSE) between both the actions and the latent representations of the teacher control policy and the student:

$$L_{student} = \|\tilde{a}_t(r_t, o_t) - a_t(o_t)\|^2 + \left\|\tilde{l}_t(r_t, o_t) - l_t(o_t)\right\|^2 \quad (11)$$

where ( ) indicates quantities that are generated by the teacher control policy. Here the student controller shares a similar latent-conditioned policy structure with the privileged teacher controller, but removes the reference motion and center of pressure information. The student network architecture consists of the following modules:

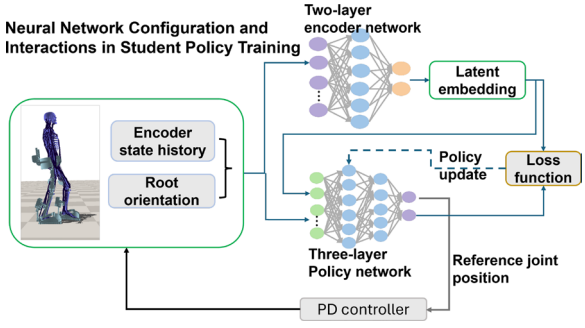

Fig. 4 Neural network configuration and interactions in student policy training

- State History Encoder (policy_enc): A two-layer MLP that encodes a concatenation of the system state history (e.g., 3-timestep history of a 22-dimensional state vector, resulting in 66 dimensions) into a 20-dimensional latent vector. The encoder includes an input layer with 66 neurons, a hidden layer with 256 neurons, and an output layer with 20 neurons. ReLU activations are used between layers.
- Student Policy Network: A three-layer MLP that takes the concatenated latent embedding and the current system state (22 dimensions) as input (total input dimension = 50) and outputs the mean of a Gaussian distribution over eight joint actions. The controller network comprises an input layer with 50 neurons, two hidden layers with 128 neurons each, and an output layer with 8 neurons, corresponding to the actions for the eight joints.
- Weight Initialization: All weights are initialized using Xavier uniform initialization, and all biases are initialized to zero.
- Action Filter: A second-order Butterworth low-pass filter is applied to smooth action outputs. The filter operates at 30 Hz with a cutoff frequency of 8 Hz to suppress high-frequency noise and improve control stability.

Pseudocode

The following algorithm outlines the two-stage training procedure. In Stage one, a privileged teacher control policy is trained using privileged information in a simulation via PPO. In Stage two, a student control policy is trained to imitate the teacher policy's behavior using only limited onboard observations, enabling real-world deployment without access to privileged inputs.

Initialize parameters of privileged teacher control policy π_t and student control policy π_s.
Stage 1: Train privileged teacher control policy
for each episode in environment $\mathscr{E}$ **do**
  Observe full state s_t from reference motion, encoders, center of pressure, and root orientation and joint angle and angular velocities (exoskeleton's hip, knee and ankle joints).
  Encode $z_t = \text{Encoder}_t(s_t)$.
  Generate action $a_t = \text{Controller}_t(z_t)$.
  Apply action a_t in simulation and receive reward r_t.
  Update privileged control policy π_t using PPO with reward r_t.
end for
Stage 2: Train Student Control Policy
for each sample $(s_t, a_t^{\text{privileged teacher control policy}})$ from the teacher control policy's trajectories **do**
  Observe limited input o_t (e.g., joint encoder and root orientation data history).
  Encode latent $\hat{z}_t = \text{Encoder}_s(o_t)$.
  Generate student action $\hat{a}_t = \text{Controller}_s(\hat{z}_t)$.
  Compute latent loss: $\|z_t - \hat{z}_t\|^2$.
  Compute behavior loss: $\|a_t^{\text{privileged control policy}} - \hat{a}_t\|^2$.
  Update student policy π_s to minimize total loss.
end for

Algorithm 1 Teacher–student control policy training

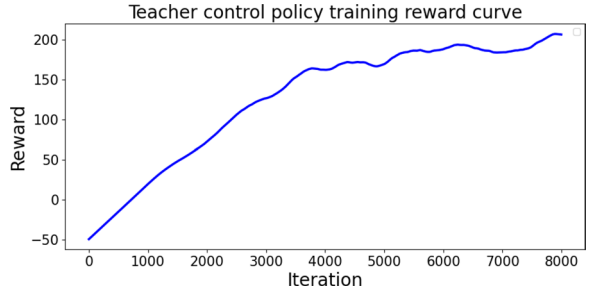

Fig. 5 Reward curve of the privileged control policy in the training process

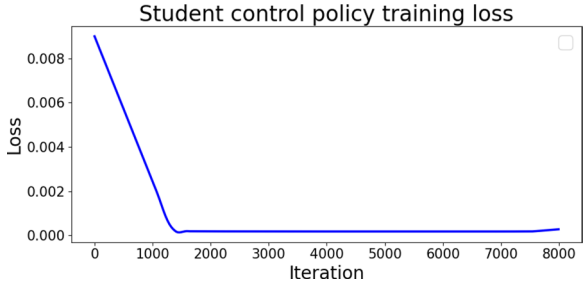

Fig. 6 Training loss value of student control policy

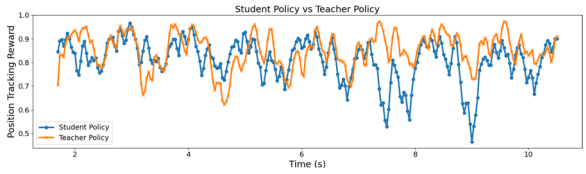

Fig. 7 Comparison of position tracking reward curves between the privileged control policy and student control policy

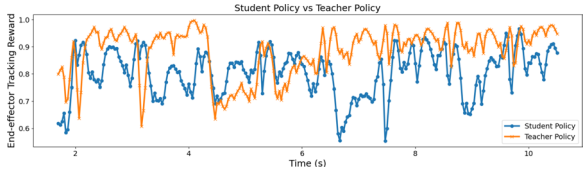

Fig. 8 Comparison of end-effector tracking reward curves between the privileged control policy and student control policy

Numerical experiments and results

Model simulation settings and RL-based controller settings

The simulation runs at 600 Hz, while the control frequency is set to 30 Hz. Sixteen agents are trained in parallel, each interacting with its own simulation environment to collect diverse training trajectories efficiently. The reinforcement learning framework is implemented using PyTorch [38], with neural networks initialized using the Xavier uniform initialization method [36]. Proximal Policy Optimization (PPO) is adopted as the learning algorithm. The privileged control policy network and the student policy network learn with a learning rate of 10^{-4}. The maximum number of training iterations is set to 120,000. Key hyperparameters are as follows: discount factor $\gamma = 0.99$; GAE factor $\lambda = 0.99$; batch size = 128; and entropy regularization weight $w_{\text{entropy}} = -0.001$. A buffer size of 2048 is used to store trajectory segments before policy updates.

These hyperparameter values (learning rate, γ, λ, batch size, entropy regularization weight w_{entropy}) were adopted from commonly used settings in reinforcement learning studies employing PPO for robot control tasks [29, 30, 39]. These values have been demonstrated to provide stable learning and robust convergence in locomotion control problems. The learning rate and entropy coefficient were further confirmed through preliminary sensitivity tests, where we varied each within a small range and observed that the chosen values yielded the most stable convergence without premature collapse or oscillatory behaviors. This combination of reference-based selection and empirical validation ensured that the adopted hyperparameters were well-suited to our exoskeleton locomotion training framework.

Results analysis

Training performance of the privileged teacher control policy and student control policy

Figure 5 shows the training reward curve of the privileged control policy using reinforcement learning. As training progresses, the average episode reward increases steadily, reaching a plateau after approximately 4000 iterations. This indicates that the privileged control policy effectively learns to generate optimal control actions that achieve the desired walking behavior. Figure 6 presents the training loss curve of the student control policy, which is supervised to imitate the privileged teacher control policy's behavior. The loss rapidly decreases within the first 1000 iterations and remains near zero throughout the training process. This demonstrates that the student policy successfully approximates the privileged control policy's output given the encoded state history and current minimal state input. These results confirm that (1) the privileged control policy can be reliably optimized using PPO, and (2) the student control policy for the exoskeleton can effectively mimic the privileged (teacher) control policy's behavior through latent-conditioned learning.

We also compared the position and end-effector tracking rewards of the student policy with that of the privileged control policy over time, as shown in Figs. 7 and 8. Both policies achieve consistently high tracking rewards, with the privileged control policy exhibiting slightly higher peak performance due to its direct optimization via reinforcement learning. The privileged control policy with the full privileged information generally achieves higher reward values, often maintaining values close to 1.0. The reward curves are also smoother, indicating more stable control and precise tracking performance. Despite having access to less sensory information, the

student control policy closely follows the overall reward trajectory, demonstrating that it effectively captures the underlying control behavior through supervised latent-conditioned imitation.

Figures 9, 10, 11, 12, 13, 14 and 15 illustrate the joint angle and torque trajectories generated by the privileged control policy and student control policy compared to the target motion. Despite relying on minimal sensor inputs and lacking direct access to the full reference motion, the student control policy is still able to produce joint trajectories that closely resemble those of the privileged control policy across hip flexion/extension, knee flexion/extension, ankle dorsiflexion/plantarflexion and ankle inversion/eversion. These results demonstrate that the student control policy can capture human motion intention and walking gait patterns over the gait cycle. While both the privileged control policy and student control policy show some deviation from the target motion, the discrepancies observed in the student policy are expected. This is because tracking the reference motion is not the primary objective in our multi-objective reward formulation. Instead, the reward function integrates multiple priorities, including locomotion stability, control smoothness, and foot clearance. As a result, the student control policy may intentionally deviate from the target trajectory to better satisfy gait balance maintenance and robustness. These findings suggest that even with limited sensory observations, the student policy is capable of learning a stable and autonomous walking control strategy, highlighting its potential for real-time deployment. Joint torque profiles across the hip, knee, and ankle joints (Figs. 13, 14, 15) demonstrate that the student policy is capable of generating torque control signals that are not only qualitatively similar in shape but also aligned in timing with those produced by the privileged control policy.

To assess the neural network controller's ability to generate autonomous torque assistance during walking locomotion, we analyzed its internal state representations using T-distributed Stochastic Neighbor Embedding (T-SNE) [40]. We visualized the learned representations by performing dimensionality reduction on the activations from each layer of the network (Figs. 16, 17). The X-axis is T-SNE Dimension 1 and the Y-axis is T-SNE Dimension 2. The axes are unitless and indicate relative distances in the reduced feature space. Axis limits and tick spacing are identical in both figures to enable direct comparison. The controller's internal states (input joint angles, joint angular velocities, and output joint torques) were used to generate the T-SNE embeddings. We then labeled the embeddings using commonly recognized gait phases. Each dot in the visualization corresponds to a single time step. As shown, the student control policy produces well-separated, cyclic clusters that closely mirror those of the privileged (teacher) policy. Distinct

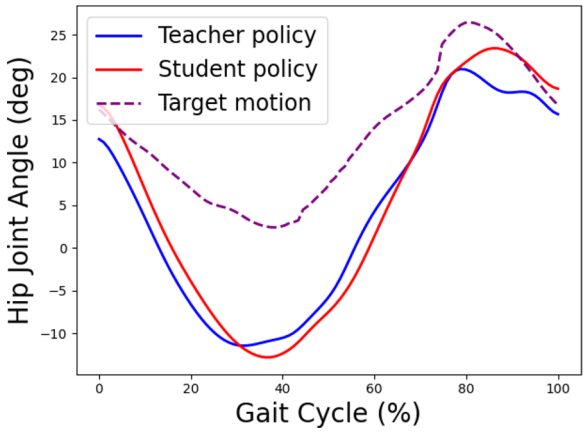

Fig. 9 Hip joint angle

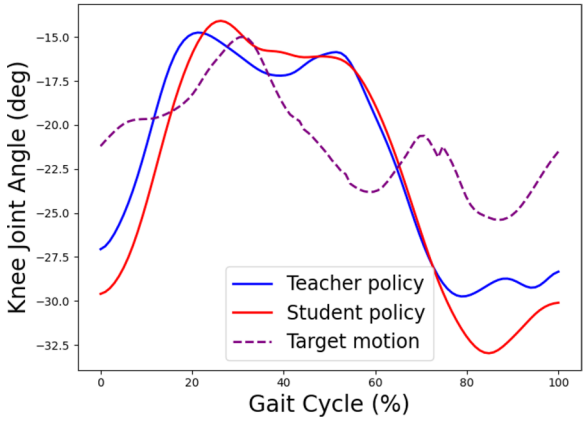

Fig. 10 Knee joint angle

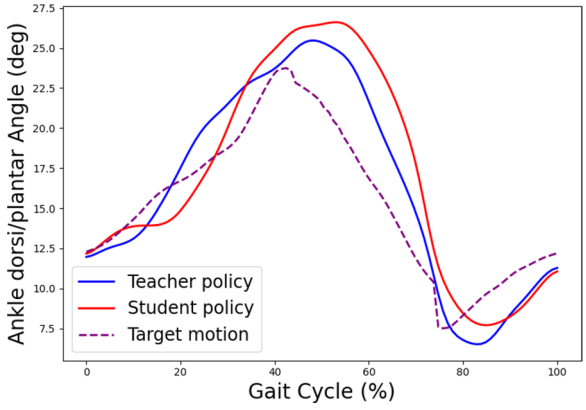

Fig. 11 Ankle dorsi/plantar flexion angle

phases (swing, heel strike, mid-stance, terminal stance, and push-off) are clearly identifiable, indicating that the student policy has internalized the gait-phase structure of human walking. These results show that the student policy not only approximates joint-level control outputs but also captures high-level gait features despite relying on minimal sensor input.

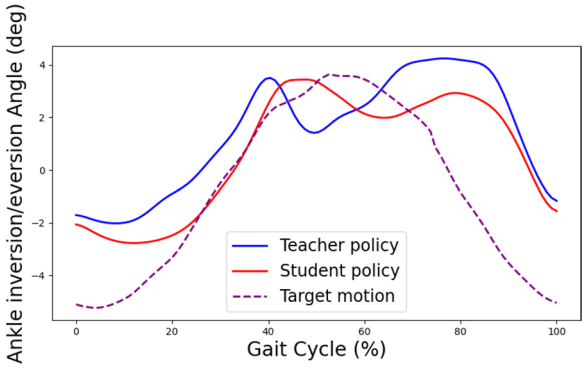

Fig. 12 Ankle inversion/eversion angle

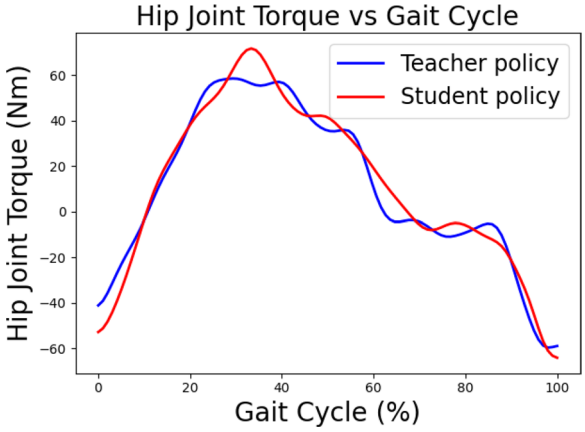

Fig. 13 Hip joint torque

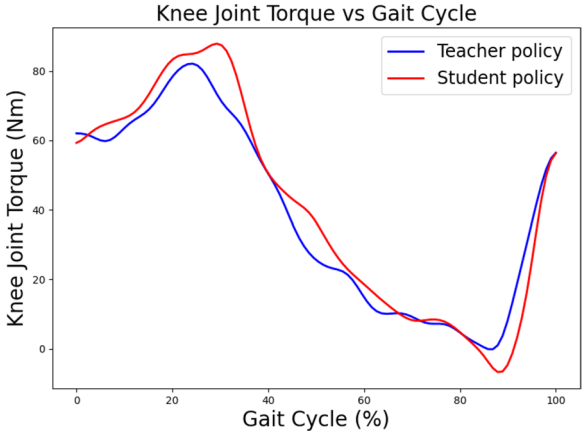

Fig. 14 Knee joint torque

Robustness analysis

We also evaluated the robustness of the trained privileged control policy and student control policy by applying an external force disturbance lasting one second. The summarized success rate results are shown in Fig. 18. As illustrated, both the privileged and student policies exhibit strong robustness under low to moderate disturbance levels (up to 40 N), achieving a 100% success rate under these conditions. At higher disturbance levels (e.g.,

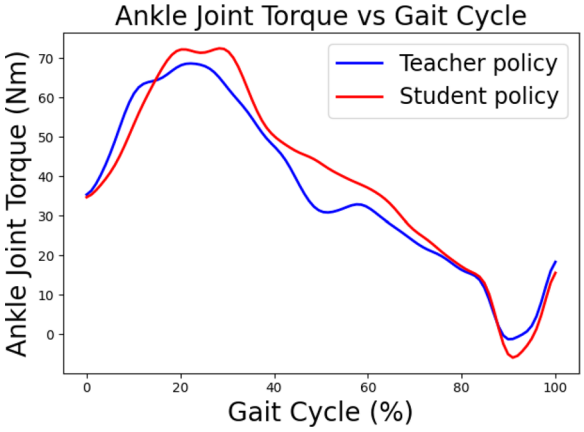

Fig. 15 Ankle dorsi/plantar joint torque

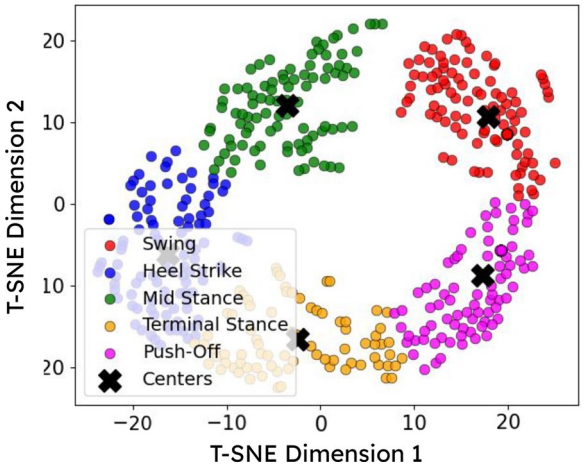

Fig. 16 T-SNE of the privileged teacher control policy

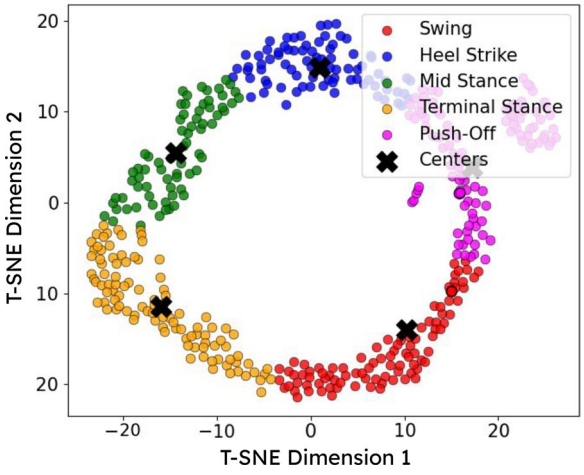

Fig. 17 T-SNE of the student control policy

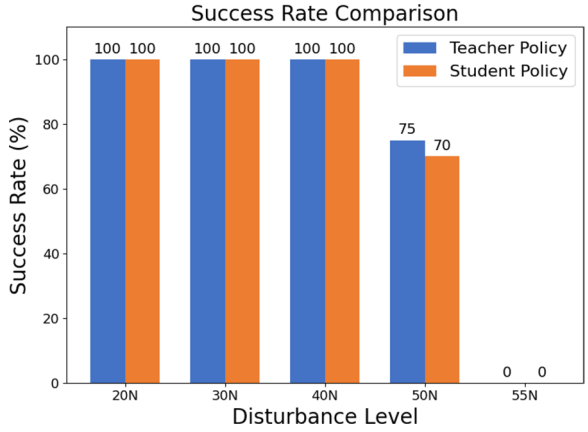

Fig. 18 Success rate under different external disturbances

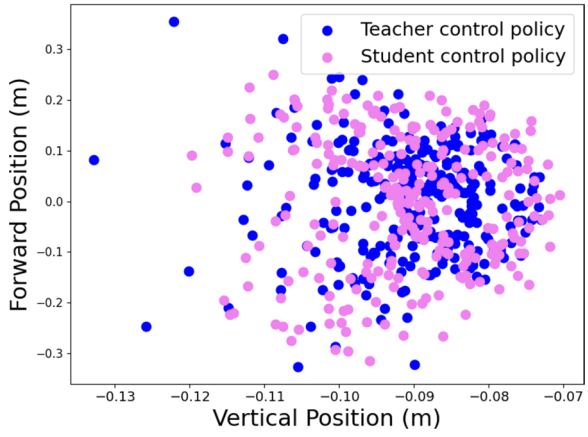

Fig. 19 The forward position and vertical position distribution of the center of mass (CoM) under the external disturbance at discrete time steps. Each dot represents the CoM position (vertical vs. forward) sampled during walking. Blue and violet markers correspond to the teacher and student control policies, respectively

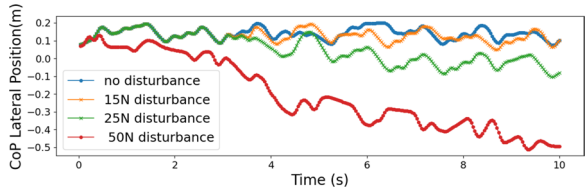

Fig. 20 CoP lateral position under different disturbances

50 N), the success rate decreases slightly, and both policies fail under even larger disturbance (55 N), indicating the limits of their robustness.

Figure 19 illustrates the forward position and vertical position distribution of the center of mass (CoM) during walking locomotion under a 50 N disturbance. As CoM forward and vertical positions are critical indicators of gait stability, especially under external disturbances, the ability of the student control policy to preserve a bounded CoM profile suggests that it successfully maintains dynamic balance during walking. Although the

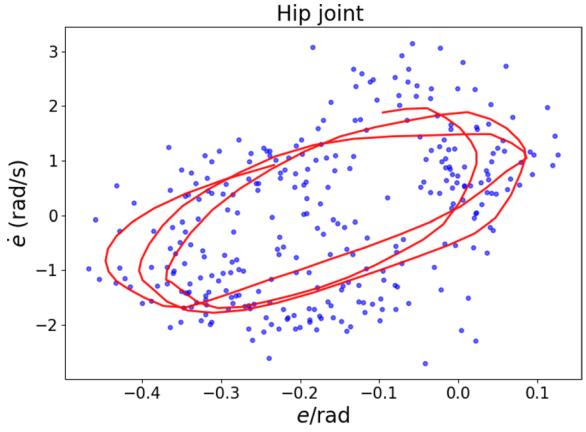

Fig. 21 Hip joint tracking error under external force disturbance

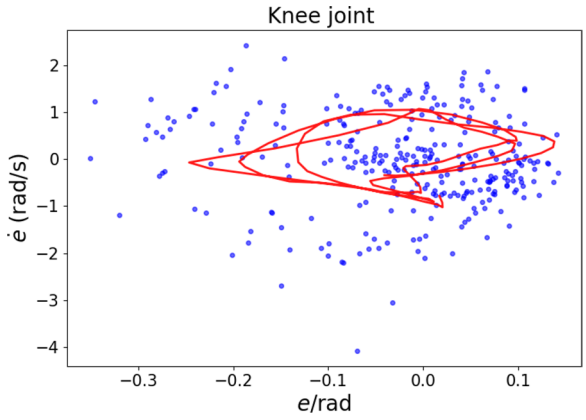

Fig. 22 Knee joint tracking error under external force disturbance

student control policy exhibits slightly greater variability, it avoids divergence and retains robustness in the presence of disturbances, further validating its effectiveness despite limited sensor inputs. These results conclude that the student control policy with minimal sensory input can achieve comparable levels of stability and robustness to external disturbances as the teacher policy. Figure 20 illustrates the lateral position of the center of pressure (CoP) of the right foot over time under varying levels of external disturbances (0 N–50 N). Under no disturbance and low disturbance levels (15 N and 25 N), the CoP trajectories remain relatively bounded and oscillate around a stable lateral region, indicating that the control policy can maintain lateral balance effectively in these conditions. Under a high disturbance level of 50 N, the CoP exhibits a clear and sustained lateral drift away from the baseline. This suggests that while the human altered the walking direction in response to the disturbance, the human–exoskeleton interaction system was still able to maintain dynamic walking stability.

The phase-space plots of joint error versus error rate for the hip, knee, and ankle joints (Figs. 21, 22, 23) provide further evidence of the system's stability and

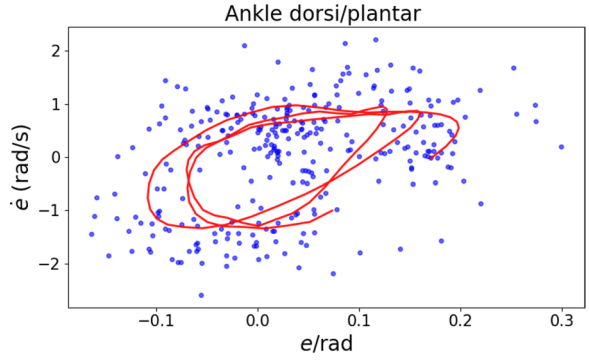

Fig. 23 Ankle joint tracking error under external force disturbance

robustness under external disturbance conditions. We plotted approximately ten seconds of walking data, where each dot represents a combination of the joint tracking error and its rate of change. The red solid line indicates the average trajectory of the joint tracking errors over time. As observed from these figures, the trajectories remain bounded and do not diverge, with most of the error points staying close to the origin and enclosed within clear elliptical regions. These bounded error dynamics indicate that the student policy is able to effectively regulate deviations in both joint position and velocity during perturbed locomotion. Together with the success rate and CoM results, these findings demonstrate that the student policy not only maintains walking stability but also ensures local joint-level tracking fidelity under significant disturbances. This highlights the controller's ability to generalize its learned policy to dynamic and uncertain environments even with limited observation input.

Comparative evaluation of sensor configurations

To determine whether a minimal sensor configuration can provide performance comparable to more sensor-rich alternatives, we conducted a set of comparative experiments using the student policy under different input conditions. The teacher control policy was not modified, as it inherently relies on privileged full-state information (e.g., reference trajectory, CoP, root orientation) that is not accessible in real-world deployment. Instead, new training sessions were performed for the student control policy (in Fig. 2) under progressively augmented sensor inputs, enabling a systematic evaluation of how sensing richness influences performance.

The student policy was retrained and evaluated under three sensor input configurations: (1) Input 1 (Minimal setting): joint angles and angular velocities obtained solely from onboard encoders. (2) Input 2: Input 1 plus exoskeleton root orientation from a single IMU mounted on the base. (3) Input 3: Input 1 plus center-of-pressure (CoP) information from bilateral foot force sensors. Each configuration was trained independently in simulation and subsequently evaluated during 10 s of walking. Gait trajectories were collected across multiple gait cycles within this period, and performance was quantified using symmetry and stability metrics. Specifically, we introduced exoskeleton root orientation from an IMU placed on its base and center-of-pressure (CoP) information derived from foot force sensors. We employed three indicators: (1) the Step Symmetry Index (SSI) to measure left–right gait symmetry; (2) the Pearson correlation coefficient of lateral CoP trajectories to assess bilateral foot coordination and lateral stability; and (3) the Euclidean distance between CoP centroids and the difference in lateral CoP variability to quantify stability of foot placement.

Table 1 demonstrates a quantitative comparison of gait symmetry and stability metrics in different sensory input configurations used for the control of the LLRE. As observed in Table 1, the gait symmetry index (SSI) [27] increases from 0.435 to 0.571 with additional CoP input, and the difference in lateral CoP variability between two feet decreases from 0.08 to 0.002. To assess lateral foot placement coordination, we also compute the Pearson correlation coefficient between the lateral CoP trajectories of the left and right feet, defined as: $r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$, where x_i and y_i are the lateral CoP values of the left and right foot at time step i, $\bar{x}$ and $\bar{y}$ are their respective means, and n is the total number of time steps in the gait cycle. A higher value of r (close to 1) indicates more synchronized lateral foot motion, serving as a measure of gait symmetry and indirectly reflecting lateral stability. Even under the minimal sensor condition (Input 1), the system achieves a high lateral CoP correlation (0.975) and maintains stable and coordinated gait patterns. These results suggest that

Table 1 Symmetry and stability performance comparison for three sensor configurations

Sensor input configuration	SSI	Pearson correlation (lateral CoP, %)	Euclidean distance between CoP centroids (cm) based on Eq. 9	Difference in standard deviation of lateral CoP
Input 1: Joint angle + angular velocity	0.435	0.975	0.380	0.080
Input 2: Input 1 + root orientation	0.492	0.969	0.259	0.100
Input 3: Input 1 + CoPs	0.571	0.906	0.311	0.002

the proposed method is robust to sensing limitations and capable of achieving strong gait stability even with minimum onboard sensors (joint encoders), highlighting its practical value for cost-efficient and deployable rehabilitation systems.

Discussion

This study presents a novel deep reinforcement learning-based, teacher–student distillation framework for the control of an LLRE, focusing on minimizing sensor requirements while ensuring robust locomotion assistance. The proposed privileged teacher–student learning architecture enables high-fidelity walking control using only proprioceptive signals, from just joint encoders. This represents a significant departure from traditional LLREs that rely on extensive sensor suites, including multiple IMUs, joint encoders, and foot force sensors, which introduce hardware complexity, increased cost, and maintenance challenges.

Our trained student control policy demonstrated high accuracy in generating smooth walking motion when compared to the teacher policy and maintained robust performance against disturbances. The comparative evaluation in Table 1 demonstrates that a minimal sensor configuration relying solely on encoder signals can already support stable and coordinated gait, with only marginal improvements observed when IMU or CoP inputs are added. This finding suggests that robust control can be achieved with substantially reduced sensing requirements, which has important implications for practical deployment. Specifically, minimizing sensor dependence lowers hardware cost, system complexity, and susceptibility to sensor failure, thereby facilitating clinical translation and large-scale rehabilitation use. Nonetheless, the incremental benefits of richer sensor configurations indicate that additional inputs may still be valuable for specific patient populations or under more demanding locomotion conditions, where enhanced stability or adaptability is required.

The privileged teacher–student policy distillation framework employed in this study offers several advantages. First, the privileged control policy, trained in simulation with full-state (privileged) information, provides a robust reference for stable gait generation. Second, the student policy, which is trained to imitate the privileged control policy using limited sensor data, demonstrates near-equivalent performance in terms of gait stability, symmetry, and robustness under disturbances, even without access to additional sensors or reference trajectories. In general, while the reference trajectory used by the teacher accelerates training and contributes to smoother, more natural motions, it poses practical challenges in deployment. Specifically, relying on a time-indexed reference trajectory introduces vulnerability to disruptions caused by unpredictable human–exoskeleton interactions. By eliminating this dependency, the student policy enables more robust and adaptable real-world control.

Compared to existing autonomous LLREs that rely on extensive sensing solutions, our approach offers a significantly simplified and more accessible solution. By leveraging a deep reinforcement learning-based teacher–student framework, we develop a controller that requires only joint encoder data and, optionally, a single extra base-mounted IMU. This minimal sensor configuration reduces system cost, integration complexity, and susceptibility to sensor failure. Moreover, we expect that the controller can generalize across a range of user conditions without the need for manual tuning or patient-specific modeling, as our trained controllers are able to assist a quadriplegic person during walking under disturbances. These features make the system particularly well-suited for deployment in clinical and home settings, where reliability, ease of setup, and operational efficiency are critical for both practitioners and users.

While the results demonstrated strong promise, several limitations should be noted: (1) The controller was only validated in a simulated environment. Although the simulation incorporates realistic musculoskeletal and exoskeleton models, real-world deployment and validation are necessary to confirm the effectiveness of sim-to-real transfer. (2) The study does not explicitly model or analyze the influence of common neuromuscular impairments, such as spasticity, contracture, or asymmetric weakness, on controller performance and human–exoskeleton interaction. While these effects can be explored using strategies from our prior work [27], they were not the focus of this investigation. (3) Although we explored minimal sensor configurations for exoskeleton control, the study does not systematically examine real-world sensor challenges such as noise, drift, and failure modes. Although domain randomization was used to improve policy robustness, explicit modeling of sensor degradation could be important for deployment in real-world environments. Overall, the proposed privileged teacher–student distillation framework demonstrates robust, stable gait control using only minimal onboard sensing. These capabilities not only validate the feasibility of sensor-efficient exoskeleton control in simulation but also establish a solid technical foundation for our future deployment on the lower limb exoskeletons [41–43].

Conclusion

In this study, we introduce a novel DRL-based approach for controlling a lower limb rehabilitation exoskeleton using minimal sensor configurations, enabled by a privileged teacher–student distillation framework that accounts for critical factors of sim-to-real transfer. Our approach leverages the pre-trained privileged control

policy, which takes advantage of privileged information in simulations, and learns a student control policy through distillation. The student policy only requires proprioceptive joint encoder sensor signals from the LLRE to ensure the feasibility of sim-to-real transfer. We validate the effectiveness and robustness of our DRL-based LLRE controller for simulated walking with a quadriplegic patient, evaluating key metrics such as stability and gait symmetry. Among the three proprioceptive sensor input configurations, we found that even the minimal configuration using only joint encoder signals demonstrated strong robustness. This finding provides critical insight into the minimal number of sensors required to ensure a robust controller. Our approach supports seamless deployment of trained controllers onto physical hardware through sim-to-real transfer, eliminating the need for patient-specific experimentation and parameter tuning. It represents a significant advancement in LLREs control methodology, promising enhanced functionality and adaptability for real-world applications.

Abbreviations

LLRE	Lower limb rehabilitation exoskeleton
RL	Reinforcement learning
CoP	Center of Pressure
CoM	Center of Mass

Acknowledgements
Not applicable.

Author Contributions
X.Z. and S.L. proposed the research idea and approach of the paper. The code implementation was done by S.L. and X.Z. Data analysis was performed by S.L. and K.F. The CAD design, development, and fabrication of the robotic exoskeleton were done by G.A. and E.N., and X.Z. created its multibody model. S.A., G.A., and H.S. provided valuable suggestions and feedback on the draft. All authors contributed to the article and approved the submitted version.

Funding
This work was partially supported by the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR) funded Rehabilitation Engineering Research Center Grant 90REGE0025-01-00, NSF Award #2524089, and by a New Jersey Health Foundation research grant PC 20-25. This work was also supported by the Embry-Riddle Aeronautical University's Faculty Innovative Research in Science and Technology (FIRST) Program.

Data availability
The datasets and code generated during and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Mechanical Engineering, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114, USA
[2]Department of Biomedical Engineering, New Jersey Institute of Technology, Newark, NJ 07102, USA
[3]The Center for Mobility and Rehabilitation Engineering Research, Kessler Foundation, West Orange, NJ 07052, USA
[4]Department of Physical Medicine and Rehabilitation, Rutgers New Jersey Medical School, Newark, NJ 07103, USA
[5]Tandon School of Engineering, New York University, New York, NY 10012, USA

Received: 2 July 2025 / Accepted: 12 December 2025
Published online: 31 December 2025

References

1. Pons JL. Rehabilitation exoskeletal robotics. IEEE Eng Med Biol Mag. 2010;29(3):57–63. https://doi.org/10.1109/MEMB.2010.936548.
2. Baud R, Manzoori AR, Ijspeert A, Bouri M. Review of control strategies for lower-limb exoskeletons to assist gait. J Neuroeng Rehabil. 2021;18(1):1–34.
3. Huo W, Mohammed S, Moreno JC, Amirat Y. Lower limb wearable robots for assistance and rehabilitation: a state of the art. IEEE Syst J. 2016;10(3):1068–81. https://doi.org/10.1109/JSYST.2014.2351491.
4. Banala SK, Kim SH, Agrawal SK, Scholz JP. Robot assisted gait training with active leg exoskeleton (alex). IEEE Trans Neural Syst Rehabil Eng. 2008;17(1):2–8.
5. Mergner T, Lippi V. Posture control—human-inspired approaches for humanoid robot benchmarking: conceptualizing tests, protocols and analyses. Front Neurorobot. 2018;12:21. https://doi.org/10.3389/fnbot.2018.00021.
6. Vouga T, Baud R, Fasola J, Bouri M, Bleuler H. Twiice—a lightweight lower-limb exoskeleton for complete paraplegics. In: 2017 International Conference on Rehabilitation Robotics (ICORR), 2017;pp. 1639–1645 . IEEE
7. Zhang T, Tran M, Huang H. Design and experimental verification of hip exoskeleton with balance capacities for walking assistance. IEEE/ASME Trans Mechatron. 2018;23(1):274–85. https://doi.org/10.1109/TMECH.2018.2790358.
8. Sun W, Lin J-W, Su S-F, Wang N, Er MJ. Reduced adaptive fuzzy decoupling control for lower limb exoskeleton. IEEE Trans Cybern. 2020;51(3):1099–10.
9. Deng M-Y, Ma Z-Y, Wang Y-N, Wang H-S, Zhao Y-B, Wei Q-X, et al. Fall preventive gait trajectory planning of a lower limb rehabilitation exoskeleton based on capture point theory. Front Inf Technol Electron Eng. 2019;20(10):1322–30.
10. Moreno JC, Figueiredo J, Pons JL. Chapter 7 - exoskeletons for lower-limb rehabilitation. In: Colombo, R., Sanguineti, V. (eds.) Rehabilitation Robotics, pp. 89–99. Academic Press, 2018. https://doi.org/10.1016/B978-0-12-811995-2.00008-4. https://www.sciencedirect.com/science/article/pii/B9780128119952000084
11. Yu S, Huang T-H, Yang X, Jiao C, Yang J, Chen Y, et al. Quasi-direct drive actuation for a lightweight hip exoskeleton with high backdrivability and high bandwidth. IEEE/ASME Trans Mechatron. 2020;25(4):1794–802.
12. Bionics R. Rex Technology. https://www.rexbionics.com/us/product-information/ Accessed April, 2020.
13. Wandercraft: Atalante. https://www.wandercraft.eu/en/exo/ Accessed April, 2020.
14. Esquenazi A, Talaty M, Packel A, Saulino M. The rewalk powered exoskeleton to restore ambulatory function to individuals with thoracic-level motor-complete spinal cord injury. Am J Phys Med Rehabil. 2012;91(11):911–2.
15. Xavier FE, Burger G, Pétriaux M, Deschaud J-E, Goulette F. Multi-imu proprioceptive state estimator for humanoid robots. In: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2023; pp. 10880–10887. IEEE.
16. Harib O, Hereid A, Agrawal A, Gurriet T, Finet S, Boeris G, et al. Feedback control of an exoskeleton for paraplegics: Toward robustly stable, hands-free dynamic walking. IEEE Control Syst Mag. 2018;38(6):61–87.
17. Vigne M, El Khoury A, Di Meglio F, Petit N. Improving low-level control of the exoskeleton atalante in single support by compensating joint flexibility. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020; pp. 3437–3443. IEEE.
18. Vigne M, El Khoury A, Di Meglio F, Petit N. State estimation for a legged robot with multiple flexibilities using IMUs: a kinematic approach. IEEE Robotics Autom Lett. 2019;5(1):195–202.

19. Sankai Y. Hal: Hybrid assistive limb based on cybernics. In: Robotics Research: The 13th International Symposium ISRR, 2010; pp. 25–34. Springer.
20. Coser O, Tamantini C, Soda P, Zollo L. Ai-based methodologies for exoskeleton-assisted rehabilitation of the lower limb: a review. Front Robot AI. 2024;11:1341580.
21. Sharifi M, Tripathi S, Chen Y., Zhang Q, Tavakoli M. Reinforcement learning methods for assistive and rehabilitation robotic systems: a survey. IEEE Trans Syst Man Cybern Syst. 2025.
22. Gao X, Si J, Wen Y, Li M, Huang H. Reinforcement learning control of robotic knee with human-in-the-loop by flexible policy iteration. IEEE Trans Neural Netw Learn Syst. 2021.
23. Guo B, Han J, Li X, Yan L. Human-robot interactive control based on reinforcement learning for gait rehabilitation training robot. Int J Adv Rob Syst. 2019;16(2):1729881419839584.
24. Peng Z, Luo R, Huang R, Hu J, Shi K, Cheng H, Ghosh BK. Data-driven reinforcement learning for walking assistance control of a lower limb exoskeleton with hemiplegic patients. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), 2020; pp. 9065–9071 . https://doi.org/10.1109/ICRA40945.2020.9197229
25. Huang R, Peng Z, Cheng H, Hu J, Qiu J, Zou C, Chen Q. Learning-based walking assistance control strategy for a lower limb exoskeleton with hemiplegia patients. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018; pp. 2280–2285 . https://doi.org/10.1109/IROS.2018.8594464
26. Luo S, Androwis G, Adamovich S, Su H, Nunez E, Zhou X. Reinforcement learning and control of a lower extremity exoskeleton for squat assistance. Frontiers in Robotics and AI. 2021;8:702845.
27. Luo S, Androwis G, Adamovich S, Nunez E, Su H, Zhou X. Robust walking control of a lower limb rehabilitation exoskeleton coupled with a musculoskeletal model via deep reinforcement learning. J Neuroeng Rehabil. 2023;20(1):34.
28. Luo S, Jiang M, Zhang S, Zhu J, Yu S, Dominguez Silva I, et al. Experiment-free exoskeleton assistance via learning in simulation. Nature. 2024;630(8016):353–9.
29. Lee J, Hwangbo J, Wellhausen L, Koltun V, Hutter M. Learning quadrupedal locomotion over challenging terrain. Sci Robot. 2020;5(47):5986.
30. Lee J, Bjelonic M, Reske A, Wellhausen L, Miki T, Hutter M. Learning robust autonomous navigation and locomotion for wheeled-legged robots. Sci Robot. 2024;9(89):9641.
31. Zhao Y, Wu K, Yi T, Xu Z, Che Z, Liu CH, Tang J. Efficient training of generalizable visuomotor policies via control-aware augmentation. In: Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems, 2025; pp. 2832–2834.
32. Androwis GJ, Karunakaran K, Nunez E, Michael P, Yue G, Foulds RA. Research and development of new generation robotic exoskeleton for over ground walking in individuals with mobility disorders (novel design and control). In: 2017 International Symposium on Wearable Robotics and Rehabilitation (WeRob), 2017; pp. 1–2. IEEE
33. Lee J, Grey M, Ha S, Kunz T, Jain S, Ye Y, et al. Dart Dynamic animation and robotics toolkit. J Open Source Softw. 2018;3(22):500.
34. Lee S, Park M, Lee K, Lee J. Scalable muscle-actuated human simulation and control. ACM Trans Graph (TOG). 2019;38(4):1–13.
35. Zhou X, Zheng L. Model-based comparison of passive and active assistance designs in an occupational upper limb exoskeleton for overhead lifting. IISE Trans Occup Ergon Hum Factor. 2021;1–19. https://doi.org/10.1080/24725838.2021.1954565.
36. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 2010; pp. 249–256.
37. Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347. 2017.
38. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S. Pytorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, 2019.
39. Hwangbo J, Lee J, Dosovitskiy A, Bellicoso D, Tsounis V, Koltun V, et al. Learning agile and dynamic motor skills for legged robots. Sci Rob. 2019;4(26):eaau5872.
40. Cai TT, Ma R. Theoretical foundations of t-sne for visualizing high-dimensional clustered data. J Mach Learn Res. 2022;23(301):1–54.
41. Huang T-H, Zhang S, Yu S, MacLean MK, Zhu J, Di Lallo A, et al. Modeling and stiffness-based continuous torque control of lightweight quasi-direct-drive knee exoskeletons for versatile walking assistance. IEEE Trans Rob. 2022;38(3):1442–59.
42. Rodríguez-Jorge D, Zhang S, Huang JS, Lopez-Sanchez I, Srinivasan N, Zhang Q, Zhou X, Su H. Biomechanics-informed mechatronics design of comfort-centered portable hip exoskeleton: actuator, wearable interface, controller. IEEE trans med robot bionics 2025.
43. Wang J, Li X, Huang T-H, Yu S, Li Y, Chen T, et al. Comfort-centered design of a lightweight and backdrivable knee exoskeleton. IEEE robot autom lett. 2018;3(4):4265–72.

Publisher's Note