

Reinforcement Learning-based Control Policy Optimization for Transtibial Prosthesis via Physics-Informed Simulation

Jonathan (Jintong) He
Mechanical Engineering
Carnegie Mellon University
Pittsburgh, USA
jintongh@andrew.cmu.edu

Shawn Krishnan
Mechanical Engineering
Carnegie Mellon University
Pittsburgh, USA
shawnakk@andrew.cmu.edu

Manuel Lancaster
Mechanical Engineering
Carnegie Mellon University
Pittsburgh, USA
manuella@andrew.cmu.edu

Eric Zhao
Mechanical Engineering
Carnegie Mellon University
Pittsburgh, USA
ericzhao@andrew.cmu.edu

Abstract—Individuals with lower extremity impairments often face significant mobility challenges, as current control strategies for transtibial prostheses and exoskeletons may not be adequately adapted to varied environments and user-specific needs. Traditional control methods, such as finite-state machines and impedance control, can be limited in their ability to handle the dynamic and complex nature of human locomotion. To address these limitations, our research explores the application of deep learning techniques to control transtibial prostheses and exoskeletons to improve mobility for individuals with lower extremity limitations. We compare imitation learning (IL) and deep reinforcement learning (RL) approaches to develop adaptive control policies for ankle assistive devices, and evaluate their produced locomotion performance against expert human locomotion models. Using the MuJoCo physics engine and Loco-MuJoCo framework, we implemented a two-agent imitation learning framework: a Variational Adversarial Imitation Learning (VAIL) agent that controls the humanoid body and a separate agent that manages control of the ankle-level assistive device. We compared the performance of this IL ankle controller to a Proximal Policy Optimization (PPO) based RL model controller of the same ankle-level assistive device. We achieved promising results with both approaches: both the IL and RL frameworks successfully produce ankle controller outputs for a transtibial prosthesis that reproduce normal walking gait behavior. However, the IL framework produced much better results, with the humanoid model able to walk much longer compared to the RL model controller. The imitation learning controller was also able to mimic the expert human gait cycle with more accuracy than the reinforcement learning model, even with decreasing input data. This work demonstrates the potential of imitation and deep reinforcement learning approaches for developing adaptable, user-compatible prosthetic controllers that can function effectively across diverse environments and user conditions.

I. INTRODUCTION

Approximately 150,000 people undergo a lower extremity amputation each year [1], and in people over 40 years of age with both diagnosed diabetes and lower extremity disease, 33% reported difficulty walking a quarter mile and climbing 10 steps without rest [2]. The application of machine learning for problem-solving has been revolutionized over the last decade. These advancements have led to groundbreaking developments in a multitude of fields, especially in robotics and controls.

Particularly, advancements in reinforcement learning [3] have produced highly effective algorithms to solve complex control problems without the need for explicit dynamic models, which can be difficult to implement. We intend to use these advances to address mobility issues in patients with limited lower body control.

There are many modern approaches to this challenge that use traditional control methods, such as adaptive whole-body dynamics with joint torque output [4]. These strategies have proven to be highly effective. However, they lack one key requirement: adaptability and compatibility between different users and environments. Recently, an increasing number of approaches have been published that utilize reinforcement learning as the primary adaptive control algorithm to learn weights and parameters for a controller to input torques into a prosthetic or an exoskeletal device. This enables the devices to apply appropriate torques and maintain weight balance to mimic ankle kinematics [5].

Our work builds on recent advances by comparing two custom control pipelines: one that uses imitation learning (IL) and another using deep reinforcement learning (RL) to train policies to understand ankle kinematics and human gait control. We first use IL to efficiently bootstrap the agent's walking behavior from expert demonstrations, providing a structured prior that avoids unstable or random exploration during early training. This structured prior is then used as the base model to train an ankle controller. One ankle controller model is trained using IL to learn the behavior of the ankle from the expert controller. Separately, we train another controller using RL with a custom reward function that emphasizes survival. Finally, we developed a comprehensive set of metrics to evaluate the kinematics, kinetics and gait cycles of the deployed IL and RL ankle controllers against the baseline expert human model.

Our results show that training an ankle actuator using imitation learning produced the best results, as the agent is able to walk for much longer than the reinforcement learning agent in most cases. Even with decreasing number of state inputs to the agent, the IL ankle controller still produced

effective walking behavior. However, we did observe a trend of decreasing number of steps and increasing error in ankle torque and angles as the number of state inputs decreased. Additionally, we compared the walking gait cycles of the IL and RL agents against the expert human baseline using simultaneous rollouts and plotting of ankle angles. From this, we observed high similarity in the walking behavior of the IL agents with the expert behavior, and less agreement between the RL agent and the expert model.

II. RELATED WORK

A. Modern Control Methods for Assistive Devices

Current assistive devices like prosthetics and exoskeletons rely on actuators to provide external forces to assist or mimic the function of an individual's limbs in everyday tasks. This means that advanced control methods have to be applied in order to increase the usability of these devices in a wide range of scenarios, from sitting and standing to running. As a result, a variety of control methods have been proposed for achieving these desired functionalities.

1) Model-based Control: Model-based control methods employ mathematical models of the system to create and test control strategies. Impedance control is widely used in exoskeletons to regulate the dynamic relationship between the device and user [6]. This approach enables compliant human-robot interactions that prioritize safety and comfort. Model predictive control (MPC) has shown effectiveness in prosthetics by anticipating future states and optimizing trajectories while respecting constraints. Manchola et al. [7] demonstrated an MPC framework for powered ankle prostheses that balances performance with power consumption. These approaches perform well when accurate models are available but struggle with the inherent complexity and variability of human movement and unpredictable environments.

2) Model-free Control: Model-free control methods overcome the limitations of explicit modeling by establishing direct mappings between sensor inputs and control outputs. Electromyography (EMG)-based control is particularly prevalent, using muscle activity signals to predict user intent. Woodward and Hargrove [8] developed a real-time EMG pattern recognition system for lower limb prostheses that classifies locomotion modes with high accuracy. Finite state machines offer another approach, as demonstrated by Young et al. [9], who implemented an intent recognition system for powered lower limb prostheses using mechanical sensors to transition between activity modes. Sensor fusion techniques that combine multiple data sources have further improved robustness and adaptability, as shown by Li et al. [10] in their real-time adaptive assistance system for exoskeletons. The drawbacks of these approaches is that they are often fine tuned to individual users and specific conditions, making them less adaptable to variations across different people and dynamic environments as well as a difficulty scaling these solutions.

3) Deep Learning Methods: Deep learning approaches represent the cutting edge in prosthetic and exoskeleton control, offering powerful tools for handling complex, high-dimensional data without explicit modeling. Reinforcement learning (RL) has emerged as particularly promising for developing adaptive controllers. Wen et al. [11] demonstrated how deep RL can learn to generate appropriate torque commands for a powered prosthetic leg across different walking conditions. Imitation learning combines aspects of supervised learning and RL to develop controllers that mimic expert behavior. Idzikowski et al. [12] used this approach to train neural network policies that reproduce natural ankle prosthesis behavior. Recent work has explored integration of these methods. Chen et al. [13] proposed a model-based reinforcement learning framework that learns dynamics models while optimizing control policies. This hybrid approach combines sample efficiency with adaptability. Zhang et al. [14] pushed this further with a personalized assistance policy that continuously adapts to user fatigue and environmental conditions using meta-learning techniques, enabling rapid adaptation to new users with minimal calibration. Unlike traditional model-free control methods that rely on reactive mappings from sensor data to actions, these learning-based approaches can generalize across users and environments by optimizing over long-term outcomes. Deep reinforcement learning methods are also more sample efficient, as data is collected from simulation as opposed to physical data collection with the user. By combining the structure of imitation learning with the adaptability of reinforcement learning, they offer greater robustness and personalization without requiring extensive manual tuning or per-user calibration.

Despite these advances, challenges remain in developing practical control systems. Sample efficiency is critical for reinforcement learning approaches, as collecting real-world human interaction data is time-consuming. Sim-to-real transfer techniques, as explored by Song et al. [15], offer solutions by pre-training policies in simulation. Additionally, personalization remains essential due to individual differences in anatomy and movement patterns. Our work builds upon these advances by comparing two approaches: one using imitation learning and a second using deep reinforcement learning specifically for ankle kinematics and human gait locomotion control. We want to validate the training sample efficiency and environment adaptability of these methods, which are features that address the key limitations of existing methods. A comparison of these two approaches will provide insight into the most effective approach in creating robust, effective prosthesis controllers.

III. METHODS

The aim of this study is to develop and evaluate an ankle prosthesis with a single degree of freedom (DoF) that enables stable bipedal locomotion through reinforcement learning (RL) and imitation learning (IL). A central challenge is that, while a reference humanoid agent observes the complete state of all joints, the prosthetic ankle must rely on a restricted subset of states.

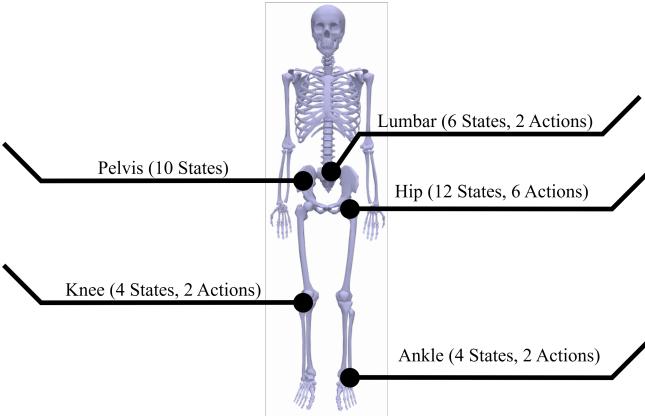


Fig. 1. Visualization of where the 36 states and 12 actions are located on the humanoid model.

A. Environment Selection

All simulations were carried out in *MuJoCo* (Multi-Joint dynamics with Contact) [16], a physics engine optimized for articulated-body dynamics, accurate contact handling, and fast numerical integration.

B. Humanoid Model Selection

The humanoid model employed is the torque-based humanoid from Loco-Mujoco [17]. Loco-Mujoco is a physics-based reinforcement learning framework designed for simulating locomotion in humanoid and legged robots. Loco-Mujoco provides realistic control mechanisms and allows for the application of torque-based actuation, which simplifies the complexity of muscle excitations while maintaining biologically plausible joint structures. The humanoid model used in this study consists of 36 state variables that represent the kinematic properties of the agent and 12 action outputs that govern the joint torques. As illustrated in Fig. 1, different states and actions are located in various joints in the humanoid model, which focuses on the joints of the lower body in particular. Detailed representations of state and action are illustrated in Tables I and IV. Torque-driven control was chosen over muscle-driven models in this study because the primary focus is on the plausibility of gait kinematics rather than muscle activations. Additionally, muscle-driven models are computationally expensive, making them impractical for our current framework.

C. Prosthetic Leg Model

To demonstrate algorithm robustness with respect to limb properties, we incorporated an open-source prosthesis from the University of Michigan Neurobionics Lab¹. The original knee actuator was replaced by a rigid socket so that only the ankle is powered. The final mass distribution—0.4 kg socket, 0.6 kg motor, 0.5 kg foot—is illustrated in Fig. 2.

¹<https://neurobionics.robtics.umich.edu/research/wearable-robotics/open-source-leg>

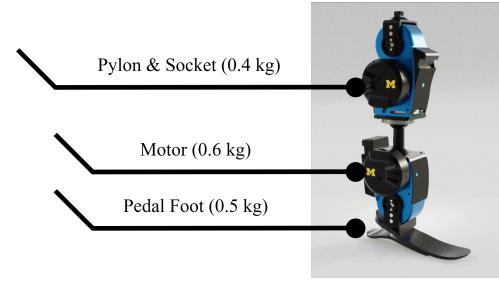


Fig. 2. Mass distribution the open source leg developed by the University of Michigan Neurobionics Lab. In this study, the knee actuator of this leg is eliminated since we only focuses on the ankle actuator.

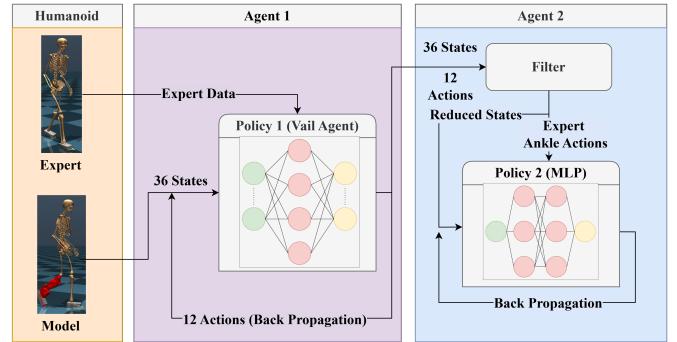


Fig. 3. Flowchart illustrating the two-agent system used for training the humanoid model with an ankle prosthesis. The VAIL agent (Agent 1) learns from expert data to control the humanoid, while the MLP-based prosthesis agent (Agent 2) learns to predict ankle actions based on filtered state information.

D. Baseline via Variational Adversarial Imitation Learning (VAIL)

A baseline expert walking agent was first trained with Variational Adversarial Imitation Learning (VAIL) on the public Loco-MuJoCo dataset. In this reference agent the ankle torque is produced directly by the VAIL policy; in the prosthesis experiments it is supplied by a separate IL/RL controller constrained to the reduced state space.

E. Imitation Learning Approach

A multilayer perceptron (MLP) was trained to replicate the expert ankle torque by minimizing the mean-squared error between its predictions and the expert output (Fig. 3). During inference (Fig. 4) the VAIL expert generates 12 torques, after which the MLP output overwrites the right-ankle torque.

F. Reinforcement Learning Implementation

Building upon the IL implementation, three Proximal Policy Optimization (PPO) variants were examined:

- 1) **Vanilla policy gradient:** unstable in continuous control.
- 2) **Basic PPO:** gradient clipping improved stability but remained sensitive to sparse rewards.

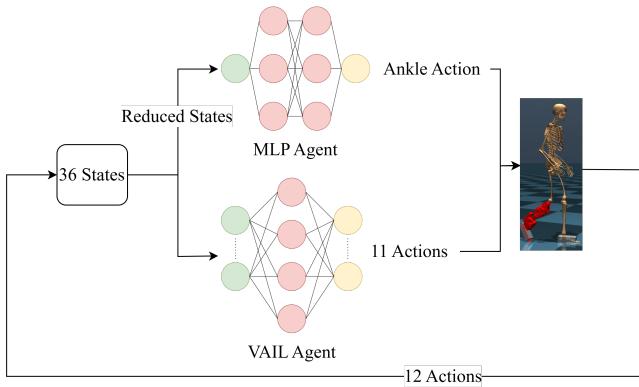


Fig. 4. Diagram showing the evaluation of the system, demonstrating the mapping from 36 state inputs into 12 action outputs by combining the action space of the two agents implemented.

- 3) **Enhanced PPO:** added generalized-advantage estimation (GAE), randomized batch sampling, and policy reuse, yielding the most consistent learning.

The reward function was

$$R = \begin{cases} +1, & \text{for each successful step,} \\ -100, & \text{when } \texttt{has_fallen}() \text{ is triggered.} \end{cases} \quad (1)$$

The `_has_fallen()` predicate flags excessive pelvic tilt or height loss.

Qualitative assessment showed that the enhanced PPO agent achieved sustained, upright walking with human-like gait patterns.

IV. EXPERIMENT

We compared ankle kinematics (angle) and kinetics (torque) for the expert, IL, and RL agents. For each controller three roll-outs were executed with an identical random seed. As summarized in Fig. 5, raw signals were low-pass filtered, segmented into ten gait cycles, and time-normalized to a single gait period. Mean absolute error (MAE) versus the expert profile was leveraged to quantify performance.

Experiments covered three observation sets (36, 22, and 16 states) and both the original and prosthetic limb properties. The reduced state lists are provided in Tables II and III.

V. RESULTS

Fig. 6 reports kinematic and kinetic MAE across all conditions; Fig. 7 shows the mean number of simulation steps completed before a fall.

The MLP-based IL controller consistently outperformed the PPO agent, achieving kinematic MAE of $2.29^\circ - 4.59^\circ$ and torque MAE of $4.07 - 10.19$ Nm, all within 3% of the expert trajectories. Performance was insensitive to state-space size and limb properties. In contrast, PPO results deteriorated with both reduced observability and prosthetic dynamics, although

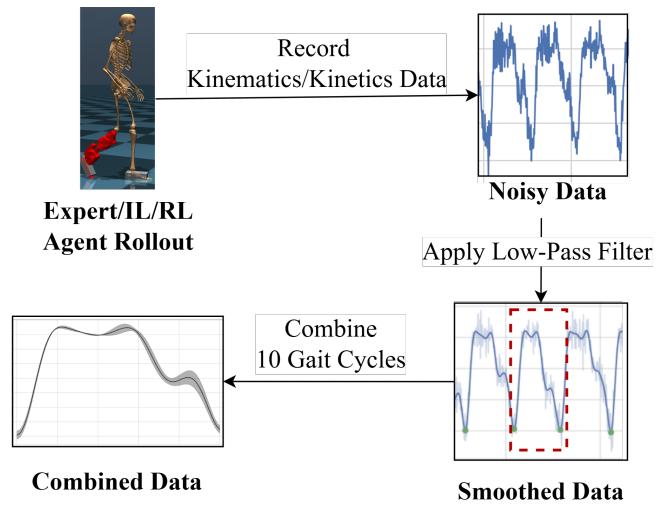


Fig. 5. Flowchart illustrating how kinematics/kinetics data are collected, processed, and compared in the experiment setup.

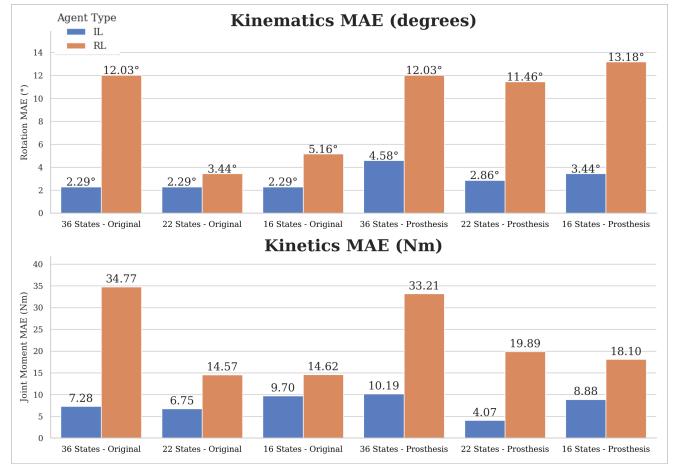


Fig. 6. Comparison of IL and RL agents regarding kinematics (top) and kinetics (bottom) mean absolute error (MAE) between original and prosthesis conditions across different input state dimensions (36, 22, and 16 states).

the agent maintained over 2000 steps before failure at full observability.

Overall, both controllers generalized to the prosthetic configuration, but IL provided superior tracking accuracy and stability.

VI. DISCUSSION AND CONCLUSION

The performance results of our experiments provide valuable insights into the comparative effectiveness of imitation learning (IL) versus reinforcement learning (RL) approaches for ankle prosthesis control. Our findings demonstrate that imitation learning significantly outperforms reinforcement learning in terms of both kinematic and kinetic accuracy, as well as overall walking stability.

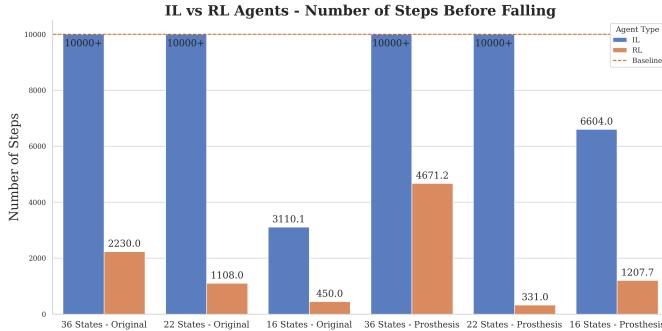


Fig. 7. Comparison of IL and RL agents in terms of the number of steps walked before falling, across original and prosthesis conditions with varying input state dimensions (36, 22, and 16 states).

A. Kinematic and Kinetic Performance Analysis

IL agents consistently achieved lower mean absolute error (MAE) values for ankle angle (2.29° to 4.59°) and joint torque (4.07 to 10.19 Nm), representing deviations of less than 3% from expert trajectories. In contrast, RL agents exhibited substantially higher kinematic MAE (3.44° to 13.18°) and kinetic MAE (14.57 to 34.77 Nm). This superior performance of IL was maintained across all state space dimensions and in both original and prosthetic limb configurations.

Our detailed gait cycle analysis revealed that IL agents produced smooth, consistent ankle angle trajectories that closely matched the expert data across the full gait cycle. This was particularly evident in our phase plots, where we visualized the relationship between ankle angle and angular velocity throughout the stride. The IL agents maintained near-identical phase relationships to the expert model, indicating biomechanically appropriate control patterns that preserve natural ankle dynamics.

The IL approach showed remarkable resilience to state space reduction, with only minimal performance degradation as the input dimension was reduced from 36 to 22 states (increasing from 2.29° to 2.86° MAE in the prosthetic condition). However, further reduction to 16 states resulted in more noticeable accuracy losses (3.44° MAE), particularly in the prosthetic configuration. This suggests that while some state information is redundant, certain proprioceptive signals remain critical for accurate ankle control, especially when adapting to altered limb dynamics.

B. Stability and Locomotion Endurance

The stability analysis further confirms the superiority of the IL approach, with IL agents consistently maintaining significantly longer walking durations before falling compared to their RL counterparts. In the original body configuration with 36 states, the IL agent achieved the maximum simulation length of 10,000 steps (1 step = 0.01 seconds) without falling, while the RL agent managed only 2,230 steps (see Fig. 7). This performance gap widened further in the prosthetic con-

figuration, where the 22-state IL agent still achieved 10,000 steps compared to just 331 steps for the RL counterpart.

Our time-normalized gait cycle visualizations revealed that RL agents struggled to maintain consistent ankle behavior across multiple consecutive strides, with increasing variability over time that eventually led to instability and falling. This was particularly evident in the 16-state condition, where the RL agent exhibited oscillatory behavior that diverged significantly from the reference trajectory, causing premature termination of walking episodes.

C. Model Architecture and Learning Approach Comparison

RL agents demonstrated greater sensitivity to both state space dimensionality and physical property changes. When limited to 16-state observations, RL controllers not only showed increased kinematic and kinetic MAE values but also exhibited qualitatively different ankle behavior patterns, sometimes producing counterproductive movements as seen in the phase plots. This suggests that the sparse reward signal (+1 for each successful step, -100 for falling) provides insufficient guidance for learning optimal ankle control in isolation, particularly with reduced state information.

A notable finding is that simpler multilayer perceptron (MLP) architectures consistently outperformed more complex networks such as transformers in both IL and RL implementations. This aligns with previous research suggesting that for continuous control problems with well-structured state spaces, the inductive bias of simpler feed-forward networks often proves advantageous. The more complex architectures likely suffer from overfitting to the training data without capturing the fundamental biomechanical principles governing ankle function during locomotion.

D. Live Plotting and Comparative Analysis

Our live plotting methodology, seen in Figures 8 and 9, provided critical insights into the real-time performance of the controllers. By simultaneously visualizing the ankle joint angles and torques of both the expert and learned controllers, we could directly compare their kinematic and kinetic profiles throughout the gait cycle. This approach revealed that IL agents maintained remarkably consistent phase relationships with expert data, preserving the characteristic plantar flexion during push-off and dorsiflexion during swing phase, which are essential for natural locomotion.

Our visualization showed that RL controllers, particularly with reduced state inputs, struggled to maintain consistent ankle behavior across consecutive strides. Their trajectory plots showed increasing deviation from the expert pattern over time, with some configurations even exhibiting reverse-phase behavior (moving in opposition to the desired direction). This was particularly evident in the 16-state RL controller for the prosthetic configuration, where the phase relationship between angle and angular velocity exhibited a completely different pattern from the expert, resulting in biomechanically inappropriate motion that quickly led to instability.

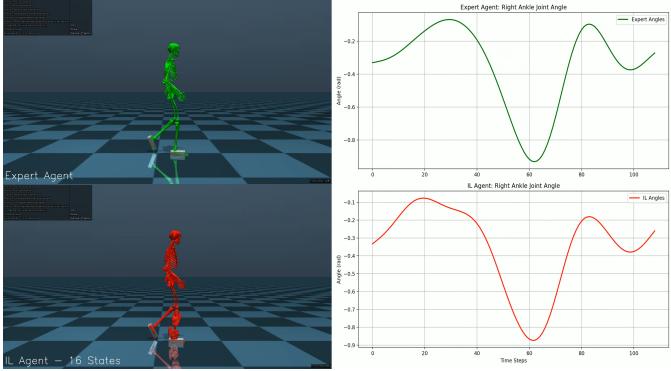


Fig. 8. Comparison of IL agent (red) and expert baseline (green) gait cycle using ankle angular position, in flat ground, single speed walking. The 16 state IL agent demonstrates consistent phase relationships with the expert gait cycle despite fewer state inputs.

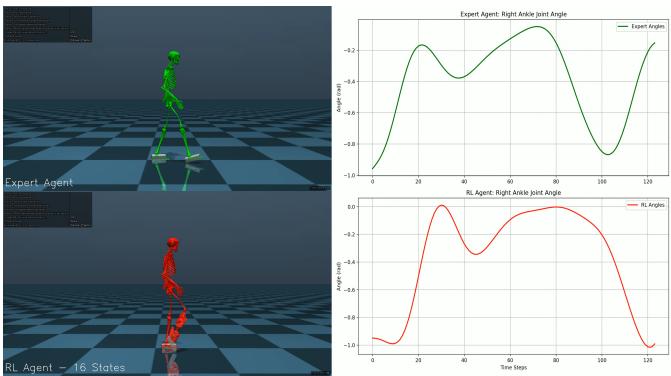


Fig. 9. Comparison of RL agent (red) and expert baseline (green) gait cycle using ankle angular position, in flat ground, single speed walking. The 16 state RL agent shows inconsistent phase relationships with the expert gait cycle and discrepancies due to imperfect ankle controller outputs.

E. Implications for Prosthetic Control

Our results demonstrate a clear advantage of imitation learning over reinforcement learning for this specific ankle control task. IL provides superior tracking accuracy and stability when tasked with reproducing expert behavior. This suggests that for applications where mimicking human-like gait patterns is the primary objective, IL offers a more direct and effective approach.

The RL approach, while less successful in our current implementation, may still offer potential benefits in scenarios requiring adaptation to novel environments or user-specific needs that go beyond the expert demonstrations available. However, our findings indicate that significant improvements to the reward function, state representation, or training methodology would be necessary to make RL competitive with IL for basic walking tasks.

The overall results demonstrate that both IL and RL-based controllers can generalize to prosthetic configurations with altered physical properties (e.g., reduced inertia), though with

varying degrees of success. This generalization capability is essential for practical prosthetic applications where the mechanical properties of the device will inevitably differ from those of a biological limb. The superior performance of the IL approach in this challenging transfer scenario further highlights its potential value for real-world prosthetic controller development.

APPENDIX

A. Humanoid Model State and Action Definitions

The following tables provide definitions for the state and action variables used in our Loco-MuJoCo torque-driven humanoid model. These variables represent the kinematic and control properties of the humanoid agent during simulation.

Each state variable is carefully bounded according to the physical limitations of human biomechanics, with angular positions having specific ranges (e.g., knee angle is limited between -2.094 and 0.175 radians) while velocities are generally unbounded. The state vector provides sufficient information for the neural network controllers to understand the humanoid's current configuration and motion state, enabling effective learning of gait patterns and balance control for the prosthetic ankle experiments described in the research.

TABLE I
COMPLETE STATE OBSERVATION - 36 STATES

Idx	Description	Min	Max	Unit
0	Joint pelvis_ty	$-\infty$	∞	m
1	Joint pelvis_tilt	$-\infty$	∞	rad
2	Joint pelvis_list	$-\infty$	∞	rad
3	Joint pelvis_rotation	$-\infty$	∞	rad
4	Joint hip flexion_r	-0.787	0.787	rad
5	Joint hip_adduction_r	-0.524	0.524	rad
6	Joint hip_rotation_r	-2.094	2.094	rad
7	Joint knee_angle_r	-2.094	0.175	rad
8-16	Other position joints	var.	var.	rad
17-19	Pelvis velocities	$-\infty$	∞	m/s
20-22	Pelvis angular vel.	$-\infty$	∞	rad/s
23-35	Joint velocities	$-\infty$	∞	rad/s

Table I shows the full 36-state representation of the humanoid model. These 36 states describe the positional and velocity information of the relevant lower body joints of the humanoid model. The first 17 dimensions (indices 0-16) correspond to position values, including the pelvis position and orientation, hip, knee, and ankle joint angles for both legs, and lumbar spine positions. The remaining dimensions (indices 17-35) represent velocity information, including linear and angular velocities of the pelvis and joint angular velocities. This is the default observation space that is available to the humanoid torque model.

TABLE II
SUBSET OF 22 STATES USED IN EXPERIMENT

Idx	Description	Min	Max	Unit
0	Joint pelvis_ty	$-\infty$	∞	m
1	Joint pelvis_tilt	$-\infty$	∞	rad
2	Joint pelvis_list	$-\infty$	∞	rad
3-9	Other position joints	var.	var.	rad
17-22	Pelvis velocities	var.	var.	var.
23-28	Right-leg joint vel.	$-\infty$	∞	rad/s

Table II presents a reduced 22-state representation of the humanoid model. This subset focuses primarily on the pelvis and right leg states, maintaining key positional and velocity information while reducing dimensionality. This representation was tested to determine how a reduced state space affects the performance of learning algorithms and the quality of the learned gait.

TABLE III
SUBSET OF 16 STATES USED IN EXPERIMENT

Idx	Description	Min	Max	Unit
0	Joint pelvis_ty	$-\infty$	∞	m
1	Joint pelvis_tilt	$-\infty$	∞	rad
2	Joint pelvis_list	$-\infty$	∞	rad
4-8	Other position joints	var.	var.	rad
17-19	Pelvis velocities	$-\infty$	∞	m/s
23-27	Right-leg joint vel.	$-\infty$	∞	rad/s

Table III shows a further reduced state representation with only 16 dimensions. This minimal state vector focuses exclusively on the pelvis and right leg, eliminating all information about the left leg. This significant reduction in state dimensions was evaluated to test the lower bounds of state information needed for effective control policy learning in our prosthesis experiments.

TABLE IV
NORMALIZED ACTION VECTOR (12 TORQUES)

Idx	Action	Min	Max
0	mot_lumbar_ext	-1	1
1	mot_lumbar_bend	-1	1
2	mot_lumbar_rot	-1	1
3-7	Right leg motors	-1	1
8-12	Left leg motors	-1	1

Table IV defines the normalized action vector that controls the humanoid's joint torques. This 12-dimensional action space represents the control inputs for the lumbar, hip, knee, and ankle joints. All actions are normalized to the range [-1, 1], which are then mapped to appropriate torque values through the actuator gear ratios defined in the MuJoCo simulation model. These normalized values provide a standardized interface for the reinforcement learning controller.

ACKNOWLEDGMENT

The authors thank Dr. Inseung Kang, Assistant Professor in the Department of Mechanical Engineering at Carnegie Mellon University, and the Director of the MetaMobility Lab at CMU. Dr. Kang provided invaluable guidance throughout the course of the project, sharing his expert insight in the fields of robotics, biomechanics, and deep learning methods.

REFERENCES

- [1] C.S. Molina and J.B. Faulk. *Lower Extremity Amputation*. StatPearls Publishing, Treasure Island, FL, 2022. Updated 2022 Aug 22.
- [2] Centers for Disease Control and Prevention. Mobility limitation among persons aged ≥ 40 years with and without diagnosed diabetes and lower extremity disease — united states, 1999–2002. *Morbidity and Mortality Weekly Report*, 54(46):1183–1186, 2005.
- [3] Ding Wang, Ning Gao, Derong Liu, Jinna Li, and Frank L. Lewis. Recent progress in reinforcement learning and adaptive dynamic programming for advanced control applications. *IEEE/CAA Journal of Automatica Sinica*, 11(1):18–36, 2024.
- [4] Ryan R. Posh, Jonathan A. Tittle, David J. Kelly, James P. Schmiedeler, and Patrick M. Wensing. Hybrid volitional control of a robotic transtibial prosthesis using a phase variable impedance controller. *arxiv*, 2023.
- [5] L. Huang, J. Zheng, Y. Gao, et al. A lower limb exoskeleton adaptive control method based on model-free reinforcement learning and improved dynamic movement primitives. *Journal of Intelligent & Robotic Systems*, 111:24, 2025.
- [6] Michael R Tucker, Jeremy Olivier, Anna Pagel, Hannes Bleuler, Mohamed Bouri, Olivier Lambery, José del R Millán, Robert Riener, Heike Vallery, and Roger Gassert. Control strategies for active lower extremity prosthetics and orthotics: a review. *Journal of neuroengineering and rehabilitation*, 12(1):1–30, 2015.
- [7] Martha Manchola, Lisa R Mayag, Marcela Munera, Carlos A Garcia, and Julian D Colorado. Model-based strategy for gait rehabilitation using a robotic ankle exoskeleton. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(5):972–981, 2019.
- [8] Richard B Woodward and Levi J Hargrove. Real-time out-of-distribution detection in lower-limb prostheses. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(11):2229–2237, 2019.
- [9] Aaron J Young, Ann M Simon, and Levi J Hargrove. Intent recognition in a powered lower limb prosthesis using time history information. *Annals of biomedical engineering*, 42(3):631–641, 2014.
- [10] Juanjuan Li, Binghan Zhong, Carly Pieringer, Mengjia Liang, Hualei Zeng, Seyha Balashov, Peter Brown, Krystyna Kim, Andy Fok, Lisa Lu, et al. Human-in-the-loop optimization of exoskeleton assistance during walking. *Science*, 356(6344):1280–1284, 2019.
- [11] Yue Wen, Jennie Si, Xiangyang Gao, Sharon Huang, and He Helen Huang. Reinforcement learning control of a powered prosthetic leg with deep neural network state representation. *IEEE Transactions on Control Systems Technology*, 28(5):2007–2017, 2019.
- [12] Mathew Idzikowski, Madison Mitchell, Claudia Mancinelli, Suraj Nair, and Levi Hargrove. Imitation learning-based framework for learning 3d gait on a powered transfemoral prosthetic leg. *IEEE Transactions on Biomedical Engineering*, 67(12):3370–3380, 2020.
- [13] Ruohan Chen, Jyothir Desai, Ramesh Raskar, and Greg Yang. Model-based reinforcement learning for closed-loop dynamic control of soft robotic manipulators. *IEEE Transactions on Robotics*, 36(5):1510–1526, 2020.
- [14] Tianyu Zhang, Huiqi Ma, Ashish Deshpande, and Luis Sentis. Personalized adaptive assistance policies for lower-limb exoskeletons using deep reinforcement learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:1598–1608, 2023.
- [15] Simiao Song, John Zhai, Yilin Yang, Wanli Pang, Mingze Jia, Geraint Rees, and Neil Burgess. Sim-to-real transfer of robotic control with dynamics randomization. *IEEE International Conference on Robotics and Automation (ICRA)*, pages 8583–8589, 2020.
- [16] Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5026–5033, 2012.
- [17] F. Al-Hafez, G. Zhao, J. Peters, and D. Tateo. Locomujoco: A comprehensive imitation learning benchmark for locomotion. *arXiv.org*, 2023. Accessed: Feb. 27, 2025.