# Learning Goal-Following Locomotion Controllers for Humanoids Using Demonstration and Reinforcement Learning

Kishor Kumar[1], Kameshwar Rao[1], Somdeb Saha[1], Vighnesh Vatsal[1] and Kaushik Das[1]

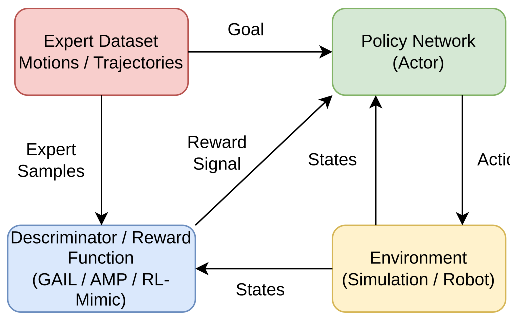


Fig. 1. Training pipeline combining imitation learning from AMASS and reinforcement learning in LocoMuJoCo for goal-conditioned whole-body control.

***Abstract*—Humanoid robots must coordinate locomotion with upper-body motion while responding to high-level goals. This work presents a goal-conditioned controller trained through Archive of motion capture as surface shapes (AMASS)-based imitation learning and reinforcement learning (RL) in the LocoMuJoCo framework. A policy is first pretrained on AMASS trajectories to acquire humanlike gait dynamics and coordinated arm–leg motion, then fine-tuned with RL to track target root velocities and hand poses using a lightweight DeepMimic-inspired reward.**

**Using a Unitree H1–scale model, we find that AMASS-initialized RL converges faster and yields higher stability, smoother motion, and more accurate goal tracking than RL-from-scratch. These results demonstrate an effective and scalable strategy for developing natural whole-body humanoid control suitable for future loco-manipulation tasks.**

***Index Terms*—Humanoid Locomotion, Imitation Learning, Reinforcement Learning, Motion Capture (AMASS), Goal-Conditioned Control**

## I. Related Work

Learning humanoid control from motion capture has gained significant traction in recent years. The AMASS dataset [1] provides high–quality human motion sequences that have been widely used to build expressive motion priors for physics-based control. Several imitation-learning frameworks leverage such data to initialize policies with humanlike coordination before reinforcement learning (RL) refinement.

Benchmark systems such as LocoMuJoCo [2] enable scalable imitation and RL experiments for locomotion, offering consistent evaluation settings across controllers. Beyond datasets and benchmarks, adversarial imitation approaches have demonstrated that motion discriminators can guide humanoids toward natural whole-body behaviors [5]. Recent advances further explore bi-level optimization [6] and latent motion representations [7] to bridge the gap between mocap data and robot dynamics.

Complementary to imitation learning, robust RL has achieved impressive results in locomotion and transfer to real hardware [8]. These works collectively motivate our approach, which integrates AMASS-based imitation with goal-conditioned RL for unified locomotion and upper-body control.

[1]TCS Research, Tata Consultancy Services Ltd., Bengaluru, Karnataka - 560066, India. *Corresponding author, e-mail: k.kishor3@tcs.com

## II. Method

Our approach trains a goal-conditioned humanoid controller using a two-stage pipeline: (i) AMASS-based imitation learning to acquire natural locomotion patterns, and (ii) reinforcement learning (RL) in the LocoMuJoCo simulator to enable goal-aware whole-body control.

### A. Demonstration Pretraining

AMASS motion clips are retargeted to the humanoid model and used to initialize the policy $\pi_\theta(a|s)$, parameterized by $\theta$, where $s$ represents the robot state and $a$ the joint action (torque commands).

Given reference motion from AMASS at time step $t$, the policy is trained to match: $q_t^{\text{ref}}$ reference joint angles, $e_t^{\text{ref}}$ reference end-effector (hand and foot) positions, $r_t^{\text{ref}}$ reference root orientation.

The imitation loss is defined as:

$$\mathcal{L}_{\text{imit}} = w_q\|q_t - q_t^{\text{ref}}\|^2 + w_e\|e_t - e_t^{\text{ref}}\|^2 + w_r\|r_t - r_t^{\text{ref}}\|^2, \quad (1)$$

where: $q_t$ current joint angles of the humanoid, $e_t$ current end-effector positions, $r_t$ current root orientation, $w_q, w_e, w_r$ scalar weights balancing the importance of each tracking term.

This stage biases the policy toward stable gaits, correct posture, and humanlike coordination.

### B. Goal-Conditioned Reinforcement Learning

After pretraining, the policy is fine-tuned to follow task-specific goals:

$$g_t = (v_x^*, v_y^*, \dot{\psi}^*, p_L^*, p_R^*), \quad (2)$$

where: $v_x^*, v_y^*$ desired root linear velocities in the horizontal plane, $\dot{\psi}^*$ desired yaw (turning) angular velocity, $p_L^*, p_R^*$ desired left and right hand positions relative to the pelvis.

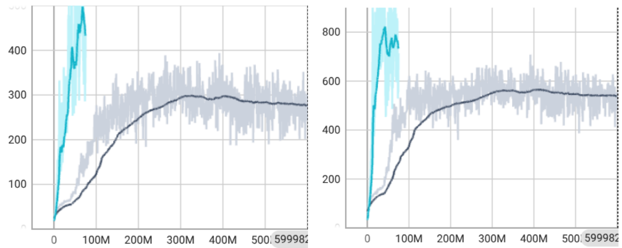


Fig. 2. Comparison of imitation-pretrained learning (blue) and RL-from-scratch (black). Left: mean episode return. Right: episode length. Imitation learning accelerates convergence and yields significantly higher returns.

The RL objective is to maximize the expected discounted return:

$$J(\theta) = \mathbb{E}_{\pi_\theta}\left[\sum_{t=0}^{T} \gamma^t r_t\right], \tag{3}$$

where: $\gamma \in (0, 1)$ discount factor, $r_t$ reward at timestep $t$.

The reward function is defined as:

$$r_t = -\lambda_v \|v_t - v_t^*\|^2 - \lambda_h \|p_{h,t} - p_{h,t}^*\|^2 - \lambda_s \|a_t\|^2 + \lambda_{\text{prior}} \phi(s_t, s_t^{\text{ref}}), \tag{4}$$

where: $v_t$ current measured root velocity, $p_{h,t}$ current hand position vector, $a_t$ action vector (joint torques), $\lambda_v, \lambda_h, \lambda_s$ weights controlling tracking and smoothness penalties, $\phi(s_t, s_t^{\text{ref}})$ motion prior term encouraging similarity to reference motion, $\lambda_{\text{prior}}$ weight for human-motion regularization.

Policy optimization is performed using Proximal Policy Optimization (PPO) with parallel rollouts enabled by MJX for efficient simulation.

### *C. Integrated Training Pipeline*

The imitation-trained policy provides a strong initialization that reduces exploration complexity during RL. Fine-tuning then adapts the controller to dynamic goal variations while preserving smooth and physically plausible motion. This two-stage approach results in stable, humanlike, and goal-responsive whole-body control, significantly outperforming controllers trained purely with reinforcement learning.

## III. Experiments and Results

We evaluate our method using the LocoMuJoCo pipeline, which provides AMASS retargeting, MJX-based physics simulation, and large-batch PPO training. The robot model is a Unitree-H1–scale humanoid (22-DoF). AMASS walking and upper-body motion clips are retargeted to the robot and used to pretrain the policy before goal-conditioned RL fine-tuning. During RL, the agent receives commands $g_t = (v_x^*, v_y^*, \dot{\psi}^*, p_L^*, p_R^*)$ that change every 1–2 s to evaluate responsiveness and stability.

Figure 2 shows the training curves for imitation-pretrained learning compared to RL-from-scratch. The AMASS-initialized policy rapidly improves within the first 50–80M steps and reaches substantially higher episode returns. In contrast, RL-from-scratch requires several hundred million steps to reach moderate performance and exhibits larger variance due to unstable early-phase exploration.

The episode-length curve further highlights the benefit of motion priors. Imitation-trained policies achieve long, uninterrupted episodes early in training, indicating stable balance and coherent whole-body motion. RL-only policies show shorter and inconsistent episodes, reflecting frequent falls and unstable transitions.

Qualitatively, imitation-pretrained policies produce smoother gaits, reduced foot slippage, and more consistent arm–leg coordination while following changing velocity and hand-pose goals. Across tasks such as forward walking, sidestepping, and turning while reaching, the AMASS+RL controller maintains stability and natural motion patterns long before RL-from-scratch becomes reliable.

Overall, these results show that initializing RL with AMASS-based imitation dramatically improves data efficiency, training stability, and the resulting whole-body motion quality.

## IV. Conclusion

We presented a goal-conditioned humanoid controller that combines AMASS-based imitation learning with reinforcement learning. Experiments show that imitation pretraining provides strong motion priors, leading to faster convergence, higher returns, and more stable rollouts compared to RL-from-scratch. The approach enables natural and responsive whole-body behaviors for a wide range of velocity and hand-pose goals. Future work will focus on hardware transfer and vision-conditioned goal generation.

## References


[1] Mahmood, Naureen, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. "AMASS: Archive of motion capture as surface shapes." In Proceedings of the IEEE/CVF international conference on computer vision, pp. 5442-5451. 2019.

[2] Al-Hafez, Firas, Guoping Zhao, Jan Peters, and Davide Tateo. "Locomujoco: A comprehensive imitation learning benchmark for locomotion." arXiv preprint arXiv:2311.02496 (2023).

[3] Peng, Xue Bin, Pieter Abbeel, Sergey Levine, and Michiel van de Panne. "DeepMimic: Example-guided deep reinforcement learning of physics-based character skills." ACM Transactions on Graphics (TOG), vol. 37, no. 4, pp. 1–14, 2018.

[4] Rudin, Nikita, David Hoeller, Philipp Reist, and Marco Hutter. "Learning to walk in minutes using massively parallel deep reinforcement learning." In Proceedings of the 5th Conference on Robot Learning (CoRL), pp. 91–100, 2022.

[5] Peng, Xue Bin, Ze Ma, Pieter Abbeel, Sergey Levine, and Angjoo Kanazawa. "AMP: Adversarial motion priors for physics-based character control." ACM Transactions on Graphics (TOG), vol. 40, no. 4, pp. 1–20, 2021.

[6] Zhao, Wenshuai, Yi Zhao, Joni Pajarinen, and Michael Muehlebach. "Bi-level Motion Imitation for Humanoid Robots." In Proceedings of the Conference on Robot Learning (CoRL), PMLR, 2025.

[7] Luo, Zhengyi, Jinkun Cao, Josh Merel, Alexander Winkler, Jing Huang, Kris Kitani, and Weipeng Xu. "Universal Humanoid Motion Representations for Physics-Based Control." In International Conference on Learning Representations (ICLR), 2024.

[8] Radosavovic, Ilija, Tete Xiao, Bike Zhang, Trevor Darrell, Jitendra Malik, and Koushil Sreenath. "Real-World Humanoid Locomotion with Reinforcement Learning." Science Robotics, 2024.