Pre-IK Residual RL for Robust Humanoid Loco-Manipulation under Interaction Wrenches

Baekseok Kim, Dr.Paul Oh

This material is based upon work supported by the NASA EPSCoR under Grant No. #80NSSC24M0108

Abstract

Humanoid loco-manipulation requires handling dynamic interaction wrenches.

MPC provides a stability prior but suffers from an execution gap due to unmodeled dynamics.

We propose a pre-IK task-space residual RL framework that injects corrections before inverse kinematics, preserving feasibility while improving robustness.

The method demonstrates stability under disturbances up to 160 N in both simulation and hardware.

Introduction

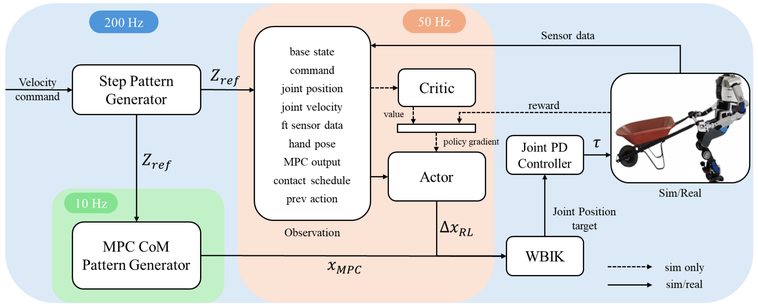
Humanoid loco-manipulation tasks, such as dragging hoses or carrying swinging payloads, introduce rapidly varying interaction wrenches that challenge balance and contact stability. While reduced-order MPC frameworks are widely used due to their computational efficiency and explicit constraint handling, they rely on simplified models and cannot fully capture fast execution-layer discrepancies. These discrepancies include early or delayed foot landing, compliance-induced drift, and contact uncertainty, which occur at time scales faster than MPC replanning. As a result, even physically consistent plans may fail during real-world execution. This motivates the need for an additional adaptation layer that can compensate for such unmodeled dynamics without sacrificing the structure and interpretability of MPC.

Methodology

We propose a hierarchical control framework combining an MPC-based locomotion planner with a residual reinforcement learning policy. The MPC generates nominal task-space references, including CoM trajectories and footstep plans, while the residual policy provides high-frequency corrections. The key idea is to inject residual actions in task space before inverse kinematics:

$$\mathbf{x_{cmd}(t) = x_{MPC}\,(t) + \lambda\,\Delta x_{RL}(t)}$$

The residual policy adjusts a compact set of task variables, including swing foot height, ankle orientation, and planar CoM offsets. This design allows the policy to directly influence contact outcomes while preserving kinematic feasibility through the IK layer.

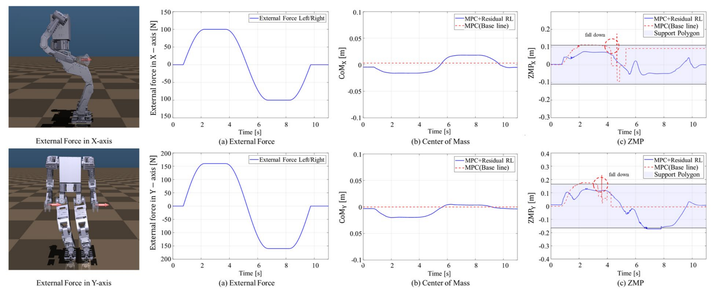


Overall multi rate control and training framework for simulation and deployment.

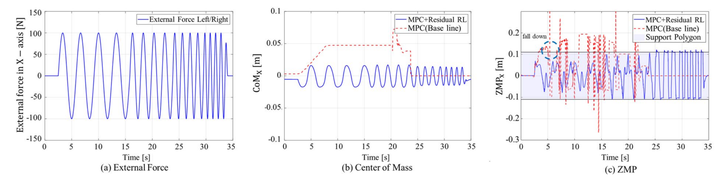
A step pattern generator and a reduced order MPC produce nominal walking references At 50 Hz, a PPO trained actor outputs a bounded pre-IK task space residual Δx_{RL} which is combined with Δx_{MPC} and sent to whole body IK to generate joint position targets for PD tracking.

Result

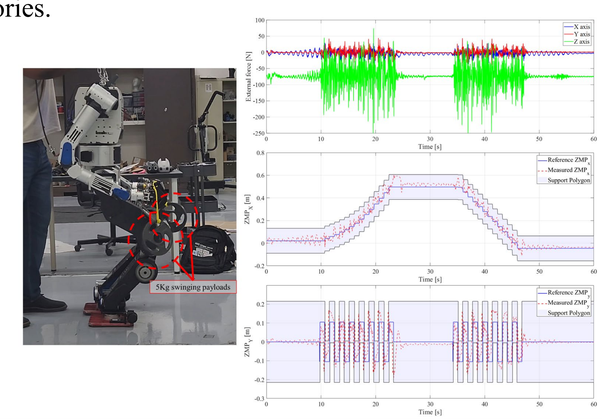
The proposed framework was evaluated in both simulation and hardware under significant interaction disturbances. In standing experiments, the controller maintained ZMP within the support polygon under disturbances up to ±160 N, exceeding the training range of ±100 N.



Under time-varying oscillatory disturbances, the MPC baseline exhibited instability, while the proposed method maintained stability for significantly longer durations. In locomotion tasks, the controller successfully sustained walking under lateral disturbances where the baseline failed.



Hardware experiments with a 10 kg freely swinging payload further validated the approach. Despite strong and unpredictable interaction forces, the robot maintained stable walking with bounded ZMP trajectories.



Conclusion

This work presents a pre-IK residual reinforcement learning framework that bridges the execution gap in humanoid loco-manipulation. By combining the stability prior of MPC with a high-bandwidth residual policy, the proposed method achieves robust performance under dynamic interaction disturbances while preserving feasibility. Future work will focus on incorporating active upper-body coordination to handle more complex manipulation tasks, as well as developing adaptive residual weighting strategies based on uncertainty. Extending the framework to diverse terrains and real-world scenarios remains an important direction toward general purpose humanoid autonomy.

Reference

[1] S. Kajita, F. Kanehiro, K. Kaneko, K. Fujiwara, K. Harada, K. Yokoi,and H. Hirukawa, “Biped walking pattern generation by using previewcontrol of zero-moment point,” in 2003 IEEE international conferenceon robotics and automation (Cat. No. 03CH37422), vol. 2, pp. 1620–1626, IEEE, 2003

[2] A. Herdt, H. Diedam, P.-B. Wieber, D. Dimitrov, K. Mombaur, andM. Diehl, “Online walking motion generation with automatic footstepplacement,” Advanced Robotics, vol. 24, no. 5-6, pp. 719–737, 2010.

[3] H. J. Lee, S. Hong, and S. Kim, “Integrating model-based footstepplanning with model-free reinforcement learning for dynamic leggedlocomotion,” arXiv preprint arXiv:2408.02662, 2024.

[4] S. H. Jeon, H. J. Lee, S. Hong, and S. Kim, “Residual mpc: Blendingreinforcement learning with gpu-parallelized model predictive con-trol,” arXiv preprint arXiv:2510.12717, 2025.