Contact-Staged Stand-up Learning for Humanoids under Shifted Center-of-Mass Conditions

Sangbeom Park, Seunghyun Lee, Paul Kim, Daewan Kim, Haechang Chung, Juwon Kim, Jaehyuk Lee, Minho Oh, Euigon Jung, Dongha Chung, Byeongho Yu, Dongkyu Lee†

***Abstract*— Humanoid robot stand-up recovery is a contact-rich whole-body control problem involving support formation, contact transition, and coordinated weight transfer. The challenge becomes more severe when customized hardware shifts the center of mass (CoM) and alters the nominal mass distribution. To address this, we propose a contact-staged stand-up framework for humanoid robots under biased CoM conditions. The proposed method decomposes the stand-up motion into four sequential stages defined by contact and body-state conditions. It further assigns dedicated critics to distinct reward groups and adopts a two-phase curriculum to improve motion smoothness and transferability. Experiments on a customized Unitree G1 with a modified head module and a rear payload demonstrate that the proposed framework outperforms ablated variants in both simulation and real-world experiments.**

I. Introduction

Recent advances in robot learning have led to significant progress in humanoid control, across bipedal locomotion [1], [2] and whole-body behaviors [3], [4]. However, falls are a potential and recurring occurrence in real-world deployment, making autonomous fall recovery [5], [6] an essential capability. In particular, stand-up from ground poses a significant challenge [7], [8], as it requires complex contact transitions and coordinated weight transfer throughout the whole-body motions.

Moreover, customized hardware configurations further complicate stand-up by altering the nominal mass distribution of the platform. For example, mounting a rear-facing depth camera on the Unitree G1 for bidirectional locomotion [9] shifts the nominal mass distribution. Similarly, we consider a Unitree G1 equipped with a customized head module and a rear bag payload, resulting in a rear-shifted center of mass (CoM). This rearward bias makes the robot more vulnerable to balance loss while rising. In such cases, unstable contact can easily lead to failure during stand-up.

In this work, we propose a contact-staged stand-up framework for humanoid robots under shifted CoM conditions. Because stable contact is critical for successful stand-up in this setting, we decompose the motion into four sequential stages: *Supine*, *Contacting*, *Lifting*, and *Standing* defined by explicit contact and body-state conditions, as shown in Fig. 1. We further incorporate stage information as privileged observations during training.

The authors are with URobotics Corp., Seoul, Republic of Korea (email: {sangbeom.park, seunghyun, paulk, daewan.kim, haechang.chung, marimo117, jaehyuk, minoh.oh, j.eg, chungdongha, bhyu, dklee}@urobotics.ai).

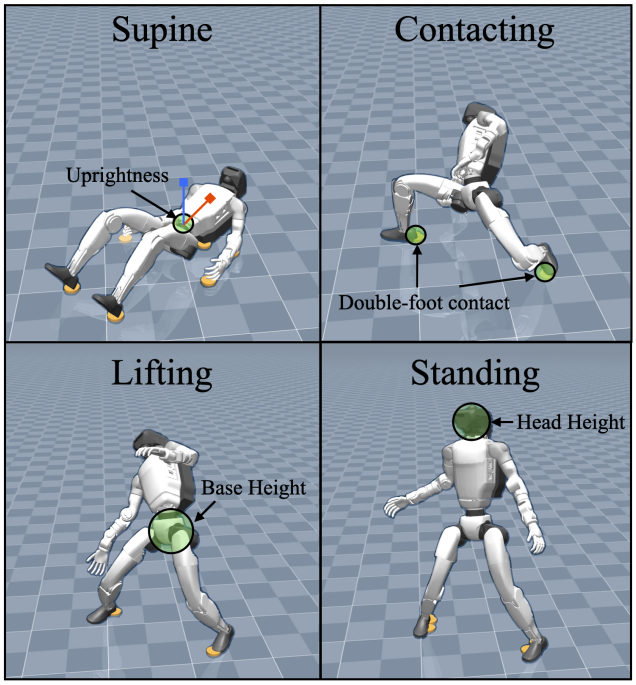

Fig. 1: Visualization of the four stages in the proposed stand-up framework. Each stage is defined by key contact and body-state conditions.

We design a categorical reward function to encourage safe and natural stand-up motion, organized into five groups: *task* for body elevation and uprightness, *contact* for stable foot support formation, *style* for penalizing unnatural joint configurations and postural deviation, *target* for post-stand posture refinement, and *regularization* for suppressing unstable behaviors. To optimize these distinct objectives jointly, we adopt a multi-critic reinforcement learning framework, assigning a dedicated critic to each group [10].

While the first phase allows the policy to broadly explore feasible stand-up strategies, the resulting motions tend to be excessively fast and aggressive. However, these motions often lack the temporal smoothness required for stable real-world execution, a gap that domain randomization alone does not address. Therefore, we introduce a second phase in which the discovered motions are temporally scaled and used as imitation targets. The policy is then trained with stronger regularization to reproduce the same trajectories at a slower and more controlled pace.

TABLE I: Stage transition conditions for stand-up behavior.

Stage	Double-foot contact	h_{base}	Uprightness (u)
Supine	–	–	–
Contacting	✓	$\geq 0.35\,\text{m}$	–
Lifting	✓	$\geq 0.50\,\text{m}$	> 0.7
Standing	✓	$\geq 0.65\,\text{m}$	> 0.9

We evaluate the proposed framework against several ablated variants under rear-biased mass distribution conditions, as summarized in Section IV. Sim-to-Sim experiments in MuJoCo [11] show that our method achieves a 100% success rate, while removing the second training phase leads to complete failure regardless of critic architecture. The single-critic variant achieves only 70%, indicating the importance of the multi-critic design. Additionally, we successfully deploy the proposed method on the customized humanoid platform, validating its practical applicability.

II. Problem Formulation

A. Contact-staged Stand-up

We focus on stand-up under shifted CoM conditions, where rear-biased mass distribution makes stable foot contact crucial during the rising motion. Therefore, we define stand-up as a sequence of four stages defined by contact and body-state conditions. Reward terms are activated stage-wise rather than uniformly throughout the episode [12], [13]. The stage definitions and their transition conditions are summarized in Table I.

The *Supine* stage corresponds to the initial ground posture. The remaining stages are defined by combinations of base height h_{base} and an uprightness score $u \in [-1, 1]$ computed from the projected gravity vector, where $u = 1$ indicates a fully upright configuration. All stages from *Contacting* onward require stable double-foot contact to be maintained.

B. Partially Observable Markov Decision Process

We formulate humanoid stand-up as a finite-horizon partially observable Markov decision process (POMDP) [14], defined by the tuple $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{R}, \gamma \rangle$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{O}$ is the observation space, $\gamma \in [0, 1]$ is the discount factor, and $\mathcal{R}$ is the reward function.

At each timestep t, the actor receives proprioceptive observation $o_t = [\omega_t, g_t, q_t, \dot{q}_t, a_{t-1}]$, where ω_t is the base angular velocity, g_t the projected gravity vector, q_t the joint positions, $\dot{q}_t$ the joint velocities, and a_{t-1} the previous action. The policy $\pi_\theta(a_t \mid o_t)$ outputs action $a_t \in \mathbb{R}^{29}$, which serves as target joint positions tracked by a low-level PD controller. The environment transitions to s_{t+1} under the dynamics, and the agent receives reward r_t. The objective is to find an optimal policy maximizing the expected discounted return $\mathbb{E}_{\pi_\theta}\left[\sum_{t=0}^{T-1} \gamma^t r_t\right]$, where T is the episode horizon.

To address partial observability, we adopt an asymmetric actor-critic architecture [15] in which the critic additionally accesses privileged observation $p_t = [c_t, z_t, h_t^{\text{base}}, h_t^{\text{head}}]$ during training, where c_t denotes the binary foot-contact state, z_t the one-hot stage indicator, and h_t^{base}, h_t^{head} the base and head heights, respectively. This asymmetric design ensures that the actor relies only on observations available at deployment, while the critic leverages additional task-relevant signals during training.

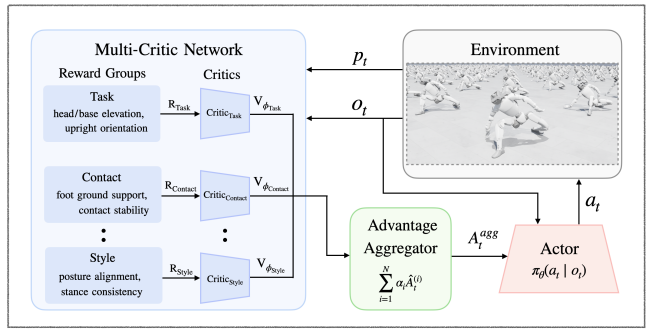

Fig. 2: Overview of the proposed framework. Reward-group-specific critics estimate separate value functions from observations and privileged observations, and their weighted advantages are aggregated for policy optimization.

III. Proposed Method

A. Multi-Critic Actor Learning

The reward groups serve distinct objectives, and optimizing them jointly with a single critic may lead to inaccurate return estimation [16]. As seen in Fig. 2, we assign an independent value function V_{ϕ_i} to each reward group $i \in \{\textit{task}, \textit{contact}, \textit{target}, \textit{style}, \textit{regularization}\}$, treating each as a separate learning objective. Each critic is optimized independently via

$$\mathcal{L}(\phi_i) = \mathbb{E}\left[\left(r_t^{(i)} + \gamma \bar{V}_{\phi_i}(o_{t+1}, p_{t+1}) - V_{\phi_i}(o_t, p_t)\right)^2\right], \tag{1}$$

where $r_t^{(i)}$ is the reward of group i and $\bar{V}_{\phi_i}$ is the target value function. For each critic, the advantage is independently estimated via GAE [17], and the per-group advantages are combined into a single aggregate advantage:

$$A_t^{\text{agg}} = \sum_i \alpha_i \hat{A}_t^{(i)}, \tag{2}$$

where α_i is the weight for reward group i. These weights are adjusted across training phases: the first phase emphasizes task and contact rewards, while the second phase increases the regularization weight to encourage smoother motion. The actor is then updated via PPO [18]:

$$\mathcal{L}_{\text{PPO}}(\theta) = \mathbb{E}[\min(\rho_t A_t^{\text{agg}},\ \text{clip}(\rho_t,\ 1-\epsilon,\ 1+\epsilon)\, A_t^{\text{agg}})]\,, \tag{3}$$

where ρ_t denotes the probability ratio between the current and old policy, and ϵ is the clipping parameter. To improve optimization stability in the multi-critic setting, we apply

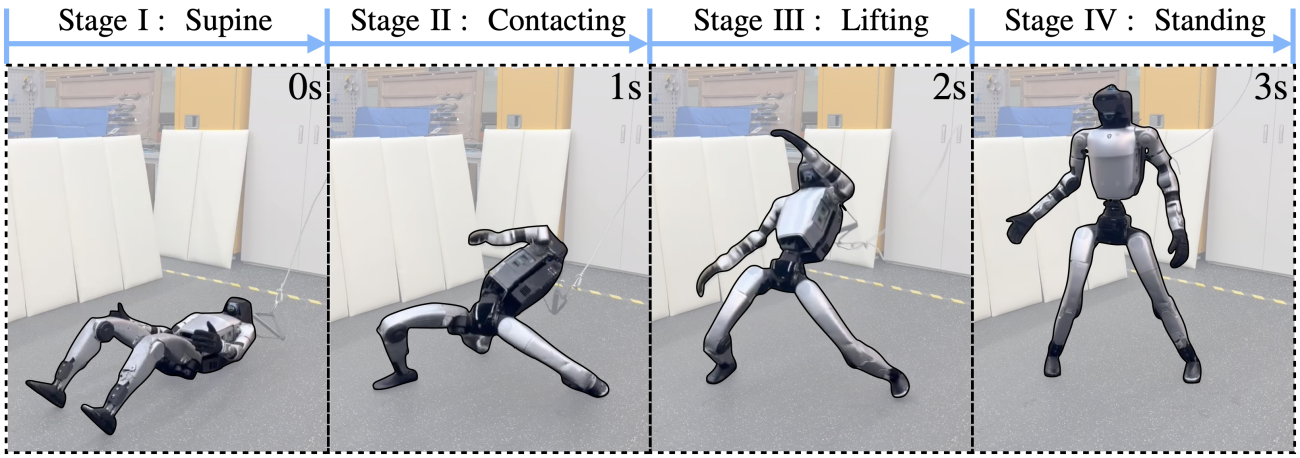

Fig. 3: Snapshots of the real-world stand-up demonstration on the customized Unitree G1. The robot progresses through the four sequential stages of the proposed policy: Supine, Contacting, Lifting, and Standing.

TABLE II: Success rates of different methods for stand-up in Sim-to-Sim transfer.

Method	Success Rate
Ours	**10/10**
Multi-Critic w/o Phase 2	0/10
Single-Critic	7/10
Single-Critic w/o Phase 2	0/10

L2C2 regularization [19], which encourages local smoothness of the policy and value functions. This regularization helps reduce instability arising from multiple value estimates across reward groups. The total objective is:

$$\mathcal{L}_{\text{total}}(\theta) = \mathcal{L}_{\text{PPO}}(\theta) + \mathcal{L}^{\pi}_{\text{smooth}} + \mathcal{L}^{V}_{\text{smooth}}. \tag{4}$$

B. Two-Phase Training Curriculum

We adopt a two-phase curriculum learning scheme to improve both the feasibility and reliability of the learned stand-up motion. The first phase prioritizes feasibility by assigning larger weights to the task and contact critics in the aggregated advantage, encouraging exploration of viable stand-up strategies.

The second phase improves motion reliability by temporally stretching the learned reference motions by $1.5\times$ through frame interpolation under stronger regularization, producing slower and more controlled behaviors. This reduces fast and aggressive motions, leading to more stable execution and higher stand-up success.

IV. Experiments

We evaluate the proposed stand-up policy under a rear-biased mass distribution against several ablated variants in a Sim-to-Sim transfer setting, as summarized in Table II. Specifically, policies are trained in Isaac Sim [20] and directly deployed in MuJoCo without additional fine-tuning to assess their transferability across simulation environments. This evaluation examines whether the learned policy remains effective in a different simulation environment. We use success rate as the evaluation metric, measured over 10 trials in the simulation environment. The proposed method succeeds in all 10 trials, while the multi-critic variant without the second phase fails in all trials. The single-critic variant achieves 70% success rate, and removing the second phase from this variant again leads to complete failure.

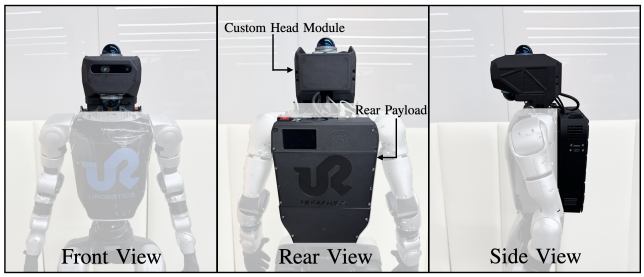

Fig. 4: Experimental Settings: Hardware configuration of the customized Unitree G1 with a modified head module and a rear payload.

The results consistently show that the second training phase improves transfer performance by making the learned motions more reliable. Furthermore, the proposed method outperforms the single-critic variant, suggesting that multiple critics more effectively capture and preserve multiple reward objectives throughout the sequential stand-up process.

To assess real-world applicability, we further demonstrate the proposed method on a customized Unitree G1 humanoid robot with a 1.5 kg modified head module and a 3 kg rear payload, as shown in Fig. 4. These modifications shift the mass distribution rearward and induce shifted CoM conditions. Figure 3 shows that the learned policy successfully performs stand-up recovery on the real robot from supine posture to upright standing. The ablated variants were not deployed on the real robot, as they failed to achieve reliable stand-up performance in the Sim-to-Sim evaluation.

V. Conclusion

In this work, we addressed humanoid stand-up under rear-shifted mass distribution conditions induced by external payload attachment. To this end, we proposed a contact-staged stand-up framework consisting of a multi-critic architecture and a two-phase curriculum. The first phase focuses on feasibility, while the second improves the reliability of the learned motion through slower and more controlled execution. Experimental results validate the effectiveness of each component, demonstrating robust stand-up performance in both Sim-to-Sim and real-world settings.

However, our approach has several limitations that should be addressed in future work. One notable limitation is the reliance on predefined reward groups, which requires substantial human engineering effort and task-specific design. This dependency may limit scalability to more diverse stand-up behaviors and hardware settings. In addition, the current policy is limited to a supine initial posture. Extending the framework to a wider range of initial postures and terrain conditions remains an important direction for future research.

References

[1] J. Long, J. Ren, M. Shi, Z. Wang, T. Huang, P. Luo, and J. Pang, "Learning humanoid locomotion with perceptive internal model," in *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, pp. 9997–10 003.

[2] A. Tang, T. Hiraoka, N. Hiraoka, F. Shi, K. Kawaharazuka, K. Kojima, K. Okada, and M. Inaba, "Humanmimic: Learning natural locomotion and transitions for humanoid robot via wasserstein adversarial imitation," in *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 13 107–13 114.

[3] Q. Liao, T. E. Truong, X. Huang, Y. Gao, G. Tevet, K. Sreenath, and C. K. Liu, "Beyondmimic: From motion tracking to versatile humanoid control via guided diffusion," *arXiv preprint arXiv:2508.08241*, 2025.

[4] L. Yang, X. Huang, Z. Wu, A. Kanazawa, P. Abbeel, C. Sferrazza, C. K. Liu, R. Duan, and G. Shi, "Omniretarget: Interaction-preserving data generation for humanoid whole-body loco-manipulation and scene interaction," *arXiv preprint arXiv:2509.26633*, 2025.

[5] P. Chen, Y. Wang, C. Luo, W. Cai, and M. Zhao, "Hifar: Multi-stage curriculum learning for high-dynamics humanoid fall recovery," in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2025, pp. 2908–2915.

[6] Z. Zhuang and H. Zhao, "Embrace collisions: Humanoid shadowing for deployable contact-agnostic motions," *arXiv preprint arXiv:2502.01465*, 2025.

[7] T. Huang, J. Ren, H. Wang, Z. Wang, Q. Ben, M. Wen, X. Chen, J. Li, and J. Pang, "Learning humanoid standing-up control across diverse postures," in *Robotics: Science and Systems (RSS)*, 2025.

[8] X. He, R. Dong, Z. Chen, and S. Gupta, "Learning getting-up policies for real-world humanoid robots," in *Robotics: Science and Systems (RSS)*, 2025.

[9] Y. Zhang, Y. Seo, J. Chen, Y. Yuan, K. Sreenath, P. Abbeel, C. Sferrazza, K. Liu, R. Duan, and G. Shi, "Rpl: Learning robust humanoid perceptive locomotion on challenging terrains," *arXiv preprint arXiv:2602.03002*, 2026.

[10] A. E. Vijayan, A. Cramariuc, M. Risiglione, C. Gehring, and M. Hutter, "Multi-critic learning for whole-body end-effector twist tracking," in *Proc. of Conference on Robot Learning (CoRL)*. PMLR, 2025, pp. 1470–1485.

[11] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2012, pp. 5026–5033.

[12] D. Kim, H. Kwon, J. Kim, G. Lee, and S. Oh, "Stage-wise reward shaping for acrobatic robots: A constrained multi-objective reinforcement learning approach," in *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, pp. 10 268–10 274.

[13] C. Liu, L. Jiang, Y. Wang, K. Yao, J. Fu, and X. Ren, "Humanoid whole-body badminton via multi-stage reinforcement learning," *arXiv preprint arXiv:2511.11218*, 2025.

[14] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial Intelligence*, vol. 101, no. 1–2, pp. 99–134, 1998.

[15] L. Pinto, M. Andrychowicz, P. Welinder, W. Zaremba, and P. Abbeel, "Asymmetric actor critic for image-based robot learning," in *Robotics: Science and Systems (RSS)*, 2018.

[16] S. Mysore, G. Cheng, Y. Zhao, K. Saenko, and M. Wu, "Multi-critic actor learning: Teaching rl policies to act with style," in *Proc. of International Conference on Learning Representations (ICLR)*, 2022.

[17] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, "High-dimensional continuous control using generalized advantage estimation," in *Proc. of International Conference on Learning Representations (ICLR)*, 2016.

[18] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[19] T. Kobayashi, "L2c2: Locally lipschitz continuous constraint towards stable and smooth reinforcement learning," in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 4032–4039.

[20] NVIDIA, "Isaac sim." [Online]. Available: https://github.com/isaac-sim/IsaacSim