Push-and-Step: From RL-Based Balance Recovery to Physical Simulation of Dense Crowds

Anonymous CVPR submission

Paper ID

Figure 1. Evolution of a dense crowd as a push propagates through physically simulated agents maintaining balance via steps and contacts.

1. Introduction

Crowd simulation has traditionally modeled collective behavior through local interaction rules in moderately dense environments, representing individuals as simplified 2D geometries such as disks or particles [5, 8]. This paradigm suffices when interactions remain social and non-contact. In highly dense settings—packed subways, concert pits, or emergency evacuations—interactions become predominantly physical: contact is unavoidable, and a character recovering balance inevitably exerts force on neighbors, propagating and potentially amplifying disturbances through the crowd. Falls, trampling, and crush injuries are real-world consequences that 2D representations are ill-equipped to model [16, 20].

We investigate full-body, physics-based simulation of humanoid agents maintaining balance under external perturbations in dense crowds. The core challenge is threefold: (i) physically simulated bipeds are notoriously difficult to control under large impulsive loads; (ii) in multi-agent settings, every attempted recovery can trigger secondary pushes; and (iii) social norms constrain how agents interact—brief shoulder contacts are acceptable, full-body collisions are not [6, 15].

To address this, we employ reinforcement learning (RL) to train a balance-recovery policy that allows simulated characters to maintain or regain balance through socially aware interactions when subjected to unpredictable external pushes. Our training pipeline consists of two stages: a single character first learns isolated balance recovery via motion imitation, before being adapted to a multi-agent, dense-crowd setting.

The code is available on GitHub, and the main contributions are:

- **RL for omnidirectional balance recovery.** We introduce an RL framework combining adversarial imitation learning with physics-based balance rewards, enabling agents to generalize from a compact reference dataset to pushes of varying strength, direction, and duration.
- **Socially aware multi-agent interactions.** An online hand-contact heuristic guides agents toward neighbors' shoulders for mechanically efficient and socially plausible balance dissipation, enabling a single-character policy to extend naturally to multi-agent crowds.
- **Scalable dense-crowd simulation.** The learned policy scales to large numbers of physically simulated agents, opening new avenues for studying safety and collective behavior in high-density settings.

2. Related Work

Dense crowd simulation. Prior work has focused on navigation and social behaviors, using particle-based 2D models that can capture push-propagation waves but miss the physical dynamics of limb-level interactions [8, 11, 13, 20]. To our knowledge, full-body physical simulation of dense crowds has not been previously attempted.

Human balance strategies. Maintaining bipedal balance relies on ankle and hip torques for small perturbations, with stepping to expand the Base of Support (BoS) for larger ones [2, 3]. In dense environments, hands can further assist recovery by exerting forces on neighbors [3, 6]. Social context shapes these interactions: pushing is tolerated when necessary for stability but must be directed at socially appropriate body locations such as the shoulder [1, 15, 19].

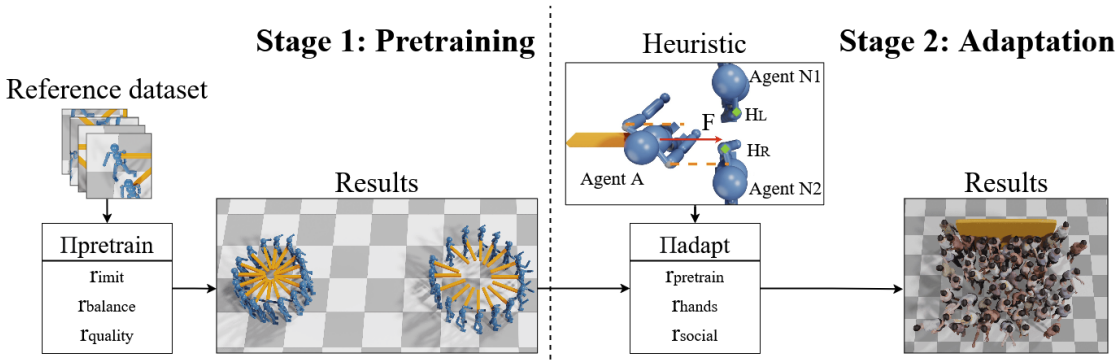


Figure 2. Two-stage training: pretraining (π_{pretrain}) for omnidirectional stepping, and adaptation (π_{adapt}) for multi-agent hand-shoulder interaction guided by the online heuristic.

Physics-based character control. Expert-designed heuristics [4, 17, 26] and trajectory optimization [22, 27] have been the traditional tools for physics-based balance. Learning-based methods have since demonstrated strong motion imitation [12, 23] and adaptation [10, 25]. Multi-agent physics-based scenarios have been explored for sports and locomotion [9, 21, 24], but dense physical crowd interaction with balance recovery remains unaddressed.

3. Method

3.1. Overview

All agents are physically simulated humanoids controlled via proportional-derivative (PD) servos [18]. The control policy $\pi(\mathbf{a}_t \mid \mathbf{s}_t)$ takes the agent's state $\mathbf{s}_t$ (joint poses and velocities over the past four frames) as input and outputs target joint angles $\mathbf{a}_t$. Training proceeds in two stages (Figure 2): pretraining (π_{pretrain}) for isolated balance recovery, and adaptation (π_{adapt}) for multi-agent interactions.

3.2. Stage 1: Adaptive Stepping Response

The first stage produces a policy that mimics human stepping strategies in response to external perturbations.

Balancing dataset. We collected a motion capture dataset of a single subject pushed horizontally on the upper back at eight evenly spaced angles. A wall placed 1 m ahead prompted natural arm-raising motions, which are important for the adaptation stage. Pushes range from 70 N to 200 N, matching experimental observations [7].

Training reward. We follow the GAN-like ICCGAN framework [23] with PPO [14] and an overall reward:

$$r_{\text{pretrain}} = w_i \tfrac{1}{2}(r_{\text{imit}} + 1) + w_g\, r_{\text{balance}} + w_q\, r_{\text{quality}}, \quad (1)$$

with weights w_i=0.6, w_g=0.2, w_q=0.2. The imitation reward r_{imit} is the discriminator output (hinge loss in $[-1, 1]$) measuring similarity to reference motions. The balance reward penalizes deviations from physically stable postures:

$$r_{\text{balance}} = e^{-\|\text{CoM}-\text{CoM}_{\text{target}}\|} + e^{-\|\text{CoP}-\text{CoP}_{\text{target}}\|}. \quad (2)$$

$\text{CoM}_{\text{target}}$ is the center of the Base of Support at resting height; $\text{CoP}_{\text{target}}$ is derived from momentum regulation [22], shifting from ankle pressure adjustment to foot placement when a step is needed:

$$\text{CoP}_{\text{target},x} = \text{CoM}_x + \tfrac{d_l p_x}{f_z}\text{CoM}_z + \tfrac{d_h L_y}{f_z}, \quad (3)$$

with damping factors d_l=4, d_h=6 and an analogous expression for the y component. The quality reward r_{quality} suppresses artifacts when generalizing beyond the reference dataset: it penalizes foot sliding (r_{foot}), heading deviation (r_{heading}), and unnecessary kinetic energy (r_{effort}). During training, push angles, magnitudes (70–200 N), and durations (0.7–1.3 s) are sampled uniformly, and small noise is added to initial joint configurations to improve robustness.

3.3. Stage 2: Adaptation to Multi-Agent Interaction

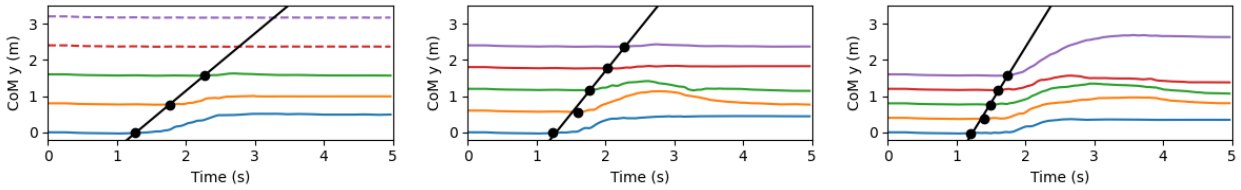
The second stage adapts π_{pretrain} to multi-agent settings using the AdaptNet architecture [25], which freezes the pretrained policy and learns new contact behaviors through latent-space injection and parallel adapter layers.

Hand-contact heuristic. Each agent's hand targets are computed online by an algorithm that considers candidate shoulders of neighbors within 5 m. Targets are appended to the state vector $\mathbf{s}_t$, and two objectives drive the selection:

1. *Collision avoidance*: the agent's shoulder trajectory over a one-second horizon (estimated from current linear momentum) is compared to all candidate shoulders. Any shoulder within δ=0.25 m of this trajectory is flagged as a collision risk and selected as a hand target.
2. *Fall prevention*: for any hand without a collision-risk target, reachable shoulders (distance to the predicted position in $[0.1, 0.6]$ m) are evaluated by collinearity with the agent's linear momentum [19], and the most energy-efficient one is selected.

Adaptation reward. The final reward extends the pretraining reward with two new terms:

$$r_{\text{adapt}} = w_p\, r_{\text{pretrain}} + w_h\, r_{\text{hands}} + w_s\, r_{\text{social}}, \quad (4)$$



0.8 m 0.6 m 0.4 m

Figure 3. CoM trajectories in the line experiment for three interpersonal distances. Dashed lines indicate agents not yet reached by the push; dots mark contact events.

with w_p=0.5, w_h=0.2, w_s=0.3. The hand placement reward r_{hands} penalizes position and orientation errors relative to the heuristic targets. The cumulative social reward r_{social} accumulates contact-force penalties and maximum hand-placement errors over time, rewarding brief, well-located contacts while penalizing prolonged or misplaced pushes.

Training episodes include three scenario types: (1) three agents in column or line-abreast formation, with only the central agent trained; (2) single-agent omnidirectional push scenarios (as in Stage 1); and (3) a no-perturbation baseline.

4. Results

4.1. Pretraining

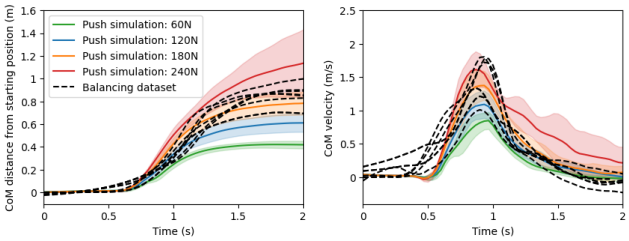


Figure 4. CoM trajectory (left) and velocity (right) for the reference (dashed) and simulated pushes at 60, 120, 180, and 240 N.

The policy π_{pretrain} generalizes beyond the reference dataset: Figure 4 shows CoM trajectories and velocities that match the reference data across push strengths from 60 to 240 N. Stronger pushes naturally produce longer steps and, at the extremes, falls—consistent with human behavior.

Ablation study. Removing r_{balance} reduces the maximum sustainable push from ≈240 N to 21 N, demonstrating its critical role. Table 1 summarizes the effects of ablating r_{imit} (increased heading deviation) and r_{quality} (foot sliding and excess kinetic energy).

	π_{pretrain}	No r_{imit}	No r_{quality}
Heading dev.↓	**5.93°**	44.81°	13.19°
Foot sliding↓	**22 cm**	24 cm	49 cm
Kinetic energy↓	**933 J**	2381 J	1444 J

Table 1. Pretraining ablation over 80 trials (16 directions × 5 force levels). Lower is better.

4.2. Adaptation

The policy π_{adapt} produces socially appropriate motions: agents raise hands to shoulder height only when a nearby agent obstructs their recovery path, and return hands to a neutral position afterward.

Ablation study. Table 2 reports average hand height and transmitted impulse over 90 trials (9 directions × 2 formations × 5 force levels). Without r_{hands}, agents keep hands raised unnecessarily after a push (final height 1.16 m vs. neutral 0.81 m). Without r_{social}, hands remain at the sides, leading to unmitigated head and torso collisions and nearly double the transmitted impulse.

	π_{adapt}	No r_{hands}	No r_{social}
Final hand height↓	**0.81 m**	1.16 m	0.81 m
Max. hand height	0.85 m	1.16 m	0.81 m
Impulse transm.↓	**40 Ns**	75 Ns	74 Ns

Table 2. Adaptation ablation over 90 trials. Neutral hand height is 0.81 m. Lower is better except max. hand height (a raised hand is desired during contact).

4.3. Applications

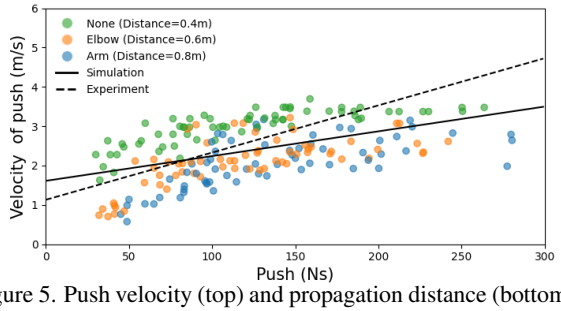


Figure 5. Push velocity (top) and propagation distance (bottom) as a function of impulse for varying interpersonal distances, consistent with empirical trends [6].

Push propagation in a line. We simulate a line of agents pushed from one end, varying interpersonal distances (0.4, 0.6, 0.8 m). Figure 3 shows CoM trajectories over time; black dots mark hand-contact events, and the black line is a linear regression fit of the push-propagation wavefront. At 0.4 m spacing, pushes cascade rapidly down the line with amplification; at 0.8 m, propagation stalls as agents have room to step and absorb energy. These trends are consistent with empirical studies of crowd push propagation [16, 20].

Dense crowd. We simulate large crowds using π_{adapt} and the hand-contact heuristic, generalizing well beyond the three-agent training configurations. Figure 1 shows a moving-wall force propagating through a crowd, with agents adapting individually. At extreme density, small perturbations trigger rapid chaotic dispersion, reproducing amplification dynamics observed in real incidents. An additional ablation confirms that π_{pretrain} alone fails in multi-agent settings, underscoring the necessity of the adaptation stage.

5. Conclusion

We presented Push-and-Step, a physics-based framework for simulating full-body humanoid agents recovering balance through stepping and socially aware hand-shoulder contacts in dense crowds. The two-stage RL pipeline—adversarial imitation learning for omnidirectional stepping, followed by AdaptNet-based finetuning for multi-agent interactions—produces agents that generalize well beyond their training scenarios and scale to large populations. Simulations reproduce empirically documented push-propagation trends, demonstrating the framework's validity as a tool for studying crowd safety.

Current limitations include a small single-subject dataset that constrains motion diversity, an engineered contact heuristic that could be replaced by a learned module given richer data, and a focus on static scenarios that precludes dynamic crowd scenarios involving locomotion. Extending the framework to navigating crowds, incorporating larger multi-person interaction datasets, and replacing heuristic components with learned contact policies remain promising directions. As a principled first step toward physically grounded dense-crowd simulation, this work opens avenues for applications in architecture, event planning, and crowd-safety analysis.

References

[1] J. Adrian, A. Seyfried, and A. Sieben. Crowds in front of bottlenecks at entrances from the perspective of physics and social psychology. *Journal of the Royal Society Interface*, 17(165):20190871, 2020. 1

[2] Z. Aftab, T. Robert, and P.-B. Wieber. Ankle, hip and stepping strategies for humanoid balance recovery with a single model predictive control scheme. In *IEEE-RAS International Conference on Humanoid Robots*, pages 159–164, 2012. 1

[3] T. Chatagnon, S. Feldmann, J. Adrian, A.-H. Olivier, C. Pontonnier, L. Hoyet, and J. Pettré. Standing balance recovery strategies of young adults in a densely populated environment following external perturbations. *Safety Science*, 177:106601, 2024. 1

[4] S. Coros, P. Beaudoin, and M. Van de Panne. Generalized biped walking control. *ACM Transactions on Graphics*, 29(4):1–9, 2010. 2

[5] H.-T. Dang, B. Gaudou, and N. Verstaevel. A literature review of dense crowd simulation. *Simulation Modelling Practice and Theory*, 134:102955, 2024. 1

[6] S. Feldmann and J. Adrian. Forward propagation of a push through a row of people. *Safety science*, 164:106173, 2023. 1, 3

[7] S. Feldmann, T. Chatagnon, J. Adrian, J. Pettré, and A. Seyfried. Temporal segmentation of motion propagation in response to an external impulse. *Safety Science*, 175:106512, 2024. 2

[8] D. Helbing, I. Farkas, and T. Vicsek. Simulating dynamical features of escape panic. *Nature*, 407(6803):487–490, 2000. 1

[9] M. Kim, E. Jung, and Y. Lee. Physicsfc: Learning user-controlled skills for a physics-based football player controller. *ACM Transactions on Graphics*, 44(4):1–21, 2025. 2

[10] Y. Li, M. Lin, Z. Lin, Y. Deng, Y. Cao, and L. Yi. Learning physics-based full-body human reaching and grasping from brief walking references. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27673–27682, 2025. 2

[11] N. Pelechano, J. M. Allbeck, M. Kapadia, and N. I. Badler. *Simulating heterogeneous crowds with interactive behaviors*. CRC Press, USA, 2016. 1

[12] X. B. Peng, P. Abbeel, S. Levine, and M. Van de Panne. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions on Graphics*, 37(4):1–14, 2018. 2

[13] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras. Human motion trajectory prediction: A survey. *The International Journal of Robotics Research*, 39(8):895–935, 2020. 1

[14] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. 2

[15] A. Sieben, J. Schumann, and A. Seyfried. Collective phenomena in crowds—where pedestrian dynamics need social psychology. *PLoS one*, 12(6):e0177328, 2017. 1

[16] C. Son, D.-H. Ham, S. Jin, and T. Park. 158 deaths at halloween night: An accimap analysis of 2022 itaewon crowd crush in south korea. *Safety Science*, 184:106741, 2025. 1, 3

[17] B. J. Stephens and C. G. Atkeson. Push recovery by stepping for humanoid robots with force controlled joints. In *IEEE-RAS International Conference on Humanoid Robots*, pages 52–59, 2010. 2

[18] J. Tan, K. Liu, and G. Turk. Stable proportional-derivative controllers. *IEEE Computer Graphics and Applications*, 31(4):34–44, 2011. 2

[19] S. Tonneau, J. Pettré, and F. Multon. Using task efficient contact configurations to animate creatures in arbitrary environments. *Computers & Graphics*, 45:40–50, 2014. 1, 2

[20] W. Van Toll, T. Chatagnon, C. Braga, B. Solenthaler, and J. Pettré. SPH crowds: Agent-based crowd simulation up to extreme densities using fluid dynamics. *Computers & Graphics*, 98:306–321, 2021. 1, 3

[21] J. Won, D. Gopinath, and J. Hodgins. Control strategies for physically simulated characters performing two-player competitive sports. *ACM Transactions on Graphics*, 40(4):1–11, 2021. 2

[22] C.-C. Wu and V. Zordan. Goal-directed stepping with momentum control. In *ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 113–118, 2010. 2

[23] P. Xu and I. Karamouzas. A GAN-like approach for physics-based imitation learning and interactive control. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 4(3), 2021. 2

[24] P. Xu, Z. Wu, R. Wang, V. Sarukkai, K. Fatahalian, I. Karamouzas, V. Zordan, and C. K. Liu. Learning to ball: Composing policies for long-horizon basketball moves. *ACM Transactions on Graphics*, 44(6), 2024. 2

[25] P. Xu, K. Xie, S. Andrews, P. G. Kry, M. Neff, M. McGuire, I. Karamouzas, and V. Zordan. AdaptNet: Policy adaptation for physics-based character control. *ACM Transactions on Graphics*, 42(6), 2023. 2

[26] K. Yin, K. Loken, and M. Van de Panne. Simbicon: Simple biped locomotion control. *ACM Transactions on Graphics*, 26(3):105–es, 2007. 2

[27] V. Zordan. Angular momentum control in coordinated behaviors. In *International Conference on Motion in Games*, pages 109–120. Springer, 2010. 2