

DribbleBot: Dynamic Legged Manipulation in the Wild

Yandong Ji*, Gabriel B. Margolis*, and Pulkit Agrawal

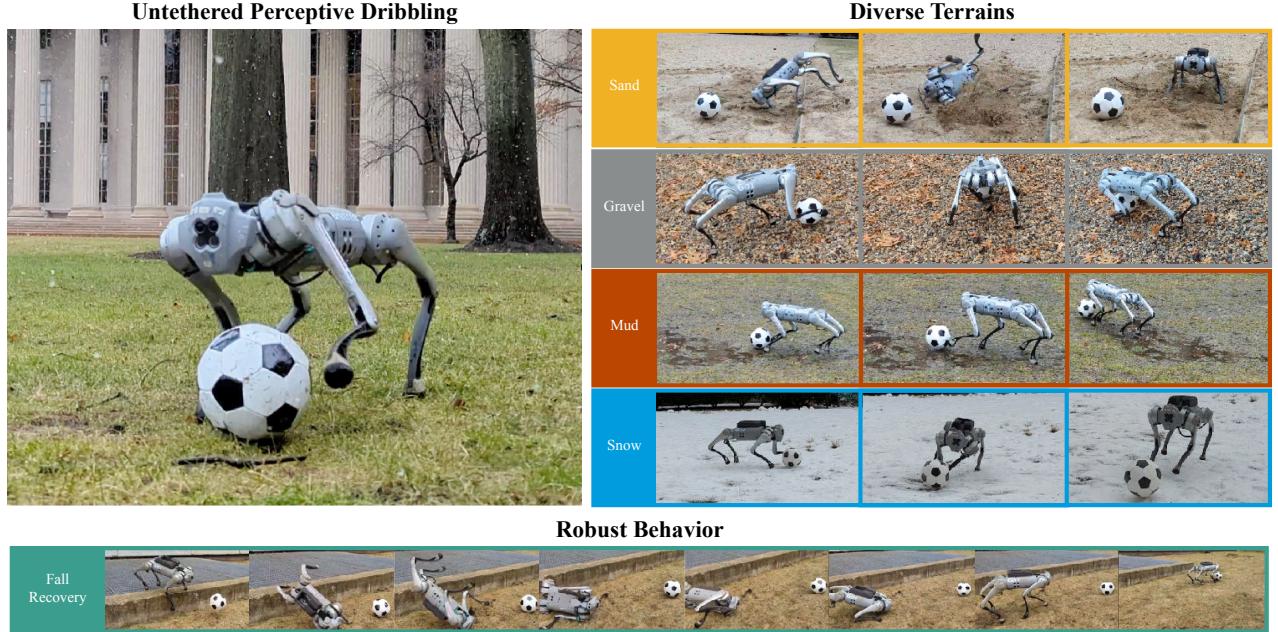


Fig. 1: **In-the-wild** dribbling on diverse natural terrains including sand, gravel, mud, and snow using onboard sensing and computing. Our system trained using reinforcement learning successfully adapts to varying ball dynamics on different terrains, and can get up and recover the ball after falling down from an extreme perturbation.

Abstract—**DribbleBot (Dexterous Ball Manipulation with a Legged Robot)** is a legged robotic system that can dribble a soccer ball under the same real-world conditions as humans (i.e., *in-the-wild*). We adopt the paradigm of training policies in simulation using reinforcement learning and transferring them into the real world. We overcome critical challenges of accounting for variable ball motion dynamics on different terrains and perceiving the ball using body-mounted cameras under the constraints of onboard computing. Our results provide evidence that current quadruped platforms are well-suited for studying dynamic control problems involving simultaneous locomotion and manipulation directly from sensory observations, such as soccer play. Video and code are available at <https://gmargo11.github.io/dribblebot>.

I. INTRODUCTION

Consider dynamic mobile manipulation, a family of compound tasks requiring tight integration of perception, dynamic locomotion, and object manipulation. In applications such as delivery and search-and-rescue, a robot is often required to move quickly while carrying an object. Instead of the object being held at a fixed position in the robot’s frame of reference, often the object must be manipulated while the robot is in motion to achieve the desired performance

under time constraints. An impressive demonstration to this effect was recently made by Boston Dynamics. It showcased the humanoid Atlas robot picking up, running with, and throwing heavy objects in a construction site [1]. The study of dynamic mobile manipulation has historically required expensive, specialized hardware and complex system architecture, so despite its relevance, it has received less attention than the individual sub-problems of locomotion and manipulation. However, in the past few years, the use of reinforcement learning has simplified the control stack for these problems, and advances in hardware have made agile, well-sensorized robots more accessible. In particular, training control policies using reinforcement learning in simulation and deploying them zero-shot in the real world has enabled fast, robust legged locomotion across challenging terrains like stairs, hiking trails, sand, and mud [2]–[11]. The same approach has also succeeded at dynamic object manipulation using dexterous hands [12]–[15]. What challenges remain in extending these successes to dynamic mobile manipulation?

As a case study, we investigate the task of soccer ball dribbling in the wild. Our aim is to develop a system that, like a human athlete, operates from onboard perception and dynamically controls a ball across a wide variety of natural terrains including grass, mud, snow, and pavement. In

* indicates equal contribution.

All authors are with the Improbable AI Lab, Massachusetts Institute of Technology, USA. Correspondence to: {yandong, gmargo}@mit.edu

contrast, previous studies on robot soccer [16]–[22] assumed a restricted setting: (i) the playing surface was flat and smooth; (ii) external perception was used instead of onboard perception and (iii) interactions largely consisted of static dribbling, where the ball comes to rest before each kick.

Bringing soccer dribbling from the laboratory into the wild is not merely a matter of synthesizing techniques from locomotion and manipulation but presents several unique challenges. One is adapting to the ball-terrain dynamics, which varies independently from the robot-terrain dynamics due to the much lighter ball and the different nature of rolling contact. On pavement, the ball may roll away faster than the robot can run; on grass, the ball will slow down quickly and requires more frequent and stronger kicks. We overcome this challenge by introducing a custom ball drag model to train a policy that can adapt to such variation. Another consideration is the limited precision and range of onboard perception systems. When a small robot is dribbling close to its body, it is hard to localize the ball using a conventional body-mounted camera due to its narrow field of view. Instead, we use observations from multiple wide-angle fisheye cameras, which introduces additional challenges that we overcome. Finally, the robot can lose control of the ball due to failure in locomotion, especially when traversing challenging terrains. To address these scenarios, we integrate a recovery policy that enables the robot to stand up autonomously after falling down. We find this controller can successfully regain control of the ball and continue to dribble.

The resulting system, DribbleBot (Dexterous Ball Manipulation with a Legged Robot), demonstrates dynamic real-world dribbling maneuvers across a variety of terrains. By providing evidence that existing hardware and sensors are capable of successful behavior, we hope to motivate more work in both robot soccer and more generally on the problem of dynamic mobile manipulation.

II. MATERIALS

Hardware: We use the Unitree Go1 robot [23] and a size 3 soccer ball for all experiments. This small robot quadruped stands 40 cm tall. We use two onboard 210° field-of-view fisheye cameras to capture images, one facing forward and one facing downward. All computation is performed on two onboard NVIDIA Jetson Xavier NX units. Due to the computation, communication bandwidth, and electrical power limitations of the robot, we critically process full-resolution images locally on each board and send only the ball location estimates to the policy board. We accelerate perception inference using TensorRT. This allows the system to process 400 × 480 resolution images comfortably at 30 Hz.

Simulator: We simulate the Unitree Go1 robot in Isaac Gym [24] using the manufacturer-provided URDF model. Simulation and training run on a single NVIDIA RTX 3090.

Pretrained Perception Module: We obtain the YOLOv7 [25], [26] model weights from the internet, pretrained on the COCO dataset [27], to perform our own fine-tuning as described in section III-B.1.

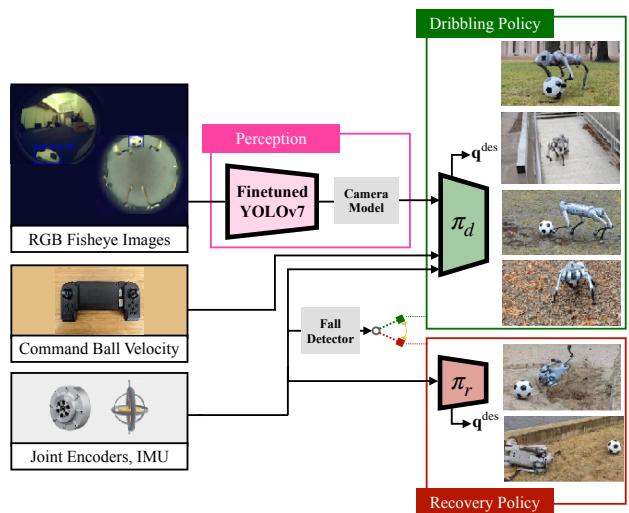


Fig. 2: **System architecture for DribbleBot.** π_d and π_r are multilayer perceptrons trained using reinforcement learning in simulation. YOLOv7 is an object detection network [26] that we finetune on images from our domain using supervised learning.

III. METHOD

Overview: We train control policy $\mathbf{a}_t = \pi_d(\mathbf{o}_t, \mathbf{c}_t)$ in simulation and transfer it to the real world. The observation \mathbf{o}_t consists of the proprioceptive sensory data and the ball position \mathbf{b}_t . The ball velocity command \mathbf{c}_t is provided as input (e.g., by a human user during deployment). We choose not to train our policy end-to-end on camera images, because of the numerous challenges including slow simulation speed, poor sample efficiency in training from high-dimensional visual observations, and the sim-to-real gap in image observations. Instead, we train the policy using ball position that is easily available in simulation. For real-world deployment a separately trained object detection model (\mathbf{Y}) predicts the ball position from images (\mathbf{o}_t^v) captured by on-board cameras: $\hat{\mathbf{b}}_t = \mathbf{Y}(\mathbf{o}_t^v)$. The policy π_d outputs actions \mathbf{a}_t , which are the joint position targets of twelve motors (three motors per leg) at 50 Hz. π_d is trained using reinforcement learning algorithm, Proximal Policy Optimization (PPO) [28].

A. Training the Dribbling Policy

1) *Environment Design:* We train the robot in simulation to dribble the ball on flat ground with physical parameters randomly varied as detailed in Section VIII. At the start of every episode, the robot's yaw orientation is randomly initialized, and its initial leg positions are randomized around a nominal pose. The soccer ball is initialized at a random position within 2 m of the robot. The target ball velocity is also uniformly randomized. These considerations ensure that the robot learns omnidirectional locomotion and dribbling. The episode length is 40 s and the control timestep is 50 Hz.

2) *Control Interface for Dribbling in the Wild:* Successful dribbling involves adjusting the leg swings to apply targeted forces while the robot moves, balances itself, and orients its position relative to a moving ball. Previous works in sim-to-real legged locomotion commonly use body velocity

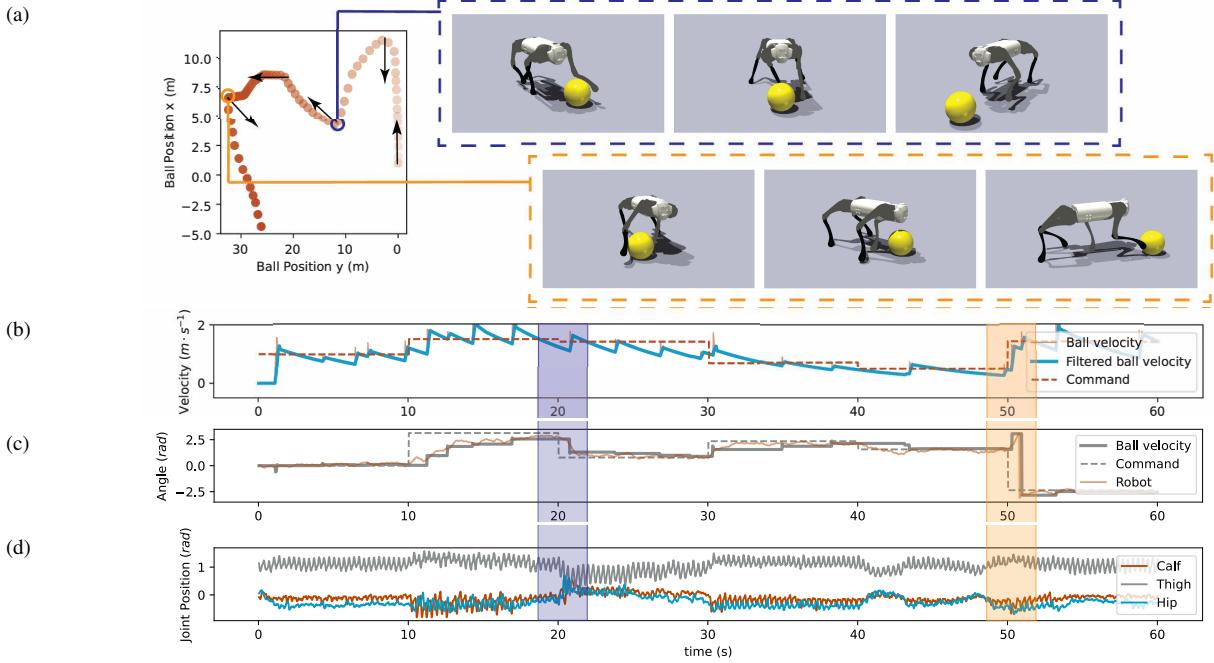


Fig. 3: **Simulated dribbling performance evaluation.** (a) Ball position in world frame and turning moment snapshots. The red points indicate the ball position and darken as time elapses. The dark arrows represent the approximate direction of the commanded velocity. (b), (c) Ball velocity tracking performance in polar coordinates. Here, the robot is first commanded to dribble the ball forward at 1 m/s, and then to execute a sequence of turns at various speeds. (d) We illustrate the joint position of the front right leg, which is typically used for dribbling and for executing left turns. The blue highlight around 20 s corresponds to the left turn visualized in the first row of images above. The orange highlight around 50 s corresponds to the right turn visualized in the second row.

command or position [29] as the control interface, with a few exceptions that allow the user to tweak gait [9], [30] or foot placement [31] parameters. The dribbling skill is not easily expressed just with gait or body-level control commands, and issuing such commands would be unintuitive for a human user deploying the robot. Instead, we directly command the ball’s linear velocity in the 2D ground plane expressed in the global reference frame. Because the ball has full rotational symmetry and the robot’s orientation can vary rapidly during a kicking maneuver, the local reference frames of both the robot and the ball are constantly changing and less useful for a human operator. We choose to command the robot in the global frame as it does not change with the robot’s motion making it much easier for the human to control the robot. The local body frame of the robot at the first time step serves as the global frame of reference.

3) Observation and Action Space: The input to the policy π_d is \mathbf{o}_t consisting of the 15-step history of command \mathbf{c}_t , ball position \mathbf{b}_t , joint positions and velocities $\mathbf{q}_t, \dot{\mathbf{q}}_t$, gravity unit vector in the body frame \mathbf{g}_t , global body yaw ψ_t , and timing reference variables θ_t^{cmd} as defined in [9]. The commands \mathbf{c}_t consist of the target ball velocities $\mathbf{v}_x^{\text{cmd}}, \mathbf{v}_y^{\text{cmd}}$ in the global frame (Sec. III-A.2). For the robot to process commands in the global frame, it must know its orientation in the global frame for which we provide the global body yaw ψ_t as input to the policy. The action space \mathbf{a}_t is the twelve target joint positions that are tracked using a PD controller

with $k_p = 20.0$, $k_d = 0.5$ [9].

4) Reward Model: Table III (Appendix) provides the task reward terms used for ball dribbling: reward for tracking the commanded ball velocity in the global reference frame and a reward shaping term incentivizing the robot to be close to the ball. Since the camera has limited range, to promote ball visibility in the camera, the robot is rewarded for facing towards the ball. Another set of gait reward terms [9], [30] encourage the robot to adopt a consistent ground contact schedule, generating well-formed gaits without the constraints of a full reference trajectory. Additional standard safety reward terms [8], [9], [32], [33] are included to penalize dangerous commands and facilitate sim-to-real transfer.

5) Policy Architecture and Optimization: We use Proximal Policy Optimization (PPO) [28] to train our soccer dribbling policy, an MLP with hidden layer sizes [512, 256, 128]. The policy converges after 7 billion timesteps, or about 48 hours of training. For visualization and debugging, it is ideal to have access to state variables such as the robot’s body velocity, ball velocity, and ball-terrain drag force coefficient. A regression model represented as two-layer MLP with [256, 128] units takes \mathbf{o}_t (Sec. III-A.3) as input and is trained using supervised learning to predict these parameters in simulation [33]. Additionally, these predicted parameters are also input to the policy network. While prior works have found concurrent state estimation useful for sim-to-real transfer of running policies [33], we did not evaluate its effect on sim-to-real performance on the dribbling task.

Tile			Grass		
Full	4/4		Full	4/4	
-R	4/4		-R	4/4	
-Y	0/4		-Y	0/4	
-D	4/4		-D	4/4	
Sand			Snow		
Full	4/4		Full	3/4	
-R	4/4		-R	3/4	
-Y	0/4		-Y	0/4	
-D	2/4		-D	-	
Curb Step-Down			Ramp		
Full	2/4		Full	0/4	
-R	1/4		-R	0/4	
-Y	-		-Y	-	
-D	-		-D	-	

TABLE I: **Real-world dribbling performance evaluation.** The robot executes a fixed dribbling trajectory on each trial. We test each scenario with full system design (**Full**) and with ablations: No recovery controller (**-R**); No YOLO fine-tuning (**-Y**); No drag model during training (**-D**)

B. Measures to Mitigate the Sim-to-Real Gap

1) *Perception in Fisheye Images:* To make dribbling feasible, the robot needs to localize a size-3 soccer ball (diameter 18 cm) using a body-mounted camera when the ball is as close as 10 cm to the body. This mandates a wide field of view. We found the visible workspace of common cameras like RealSense (field of view 105°) much too small for this application. Our robot is equipped with ultrawide forward-facing and downward-facing fisheye cameras, each with a field of view 210°. With such a wide view, rectification results in substantial warping of the spherical soccer ball, making it infeasible to apply off-the-shelf object detection networks that are trained on datasets of narrow-view rectified images taken from the internet. To recover good detection performance in the entire field of view, we fine-tune the YOLOv7 [26] network on 254 hand-labeled images of soccer balls from our robot’s camera, including images with the ball at the edge. Specifically, we fine-tune the YOLOv7 model that was pretrained on the COCO dataset [27]. The learned model accurately detects the ball in cluttered environments.

While object detection outputs a bounding box, our control policy takes as input the ball position. We obtain ball position through an approximate application of the equidistant fisheye lens model. Given the ball pixel coordinates, we first compute the angle Ψ from the camera principal axis to the ball center using the equidistant model $r = f\Psi$. Ignoring warping effects, observing the size of the ball in the image, and knowing the actual ball radius, we use the perspective projection ratio to compute the distance between the ball and the camera. The observed ball positions in the front and bottom cameras are transformed into the body frame using the known camera extrinsics. Finally, if the ball is detected in both cameras, we must fuse this information into a single

position estimate. We found it effective to select the detection with a higher confidence value output by YOLOv7.

2) *Perception Noise Model:* While in simulation accurate ball position estimates are available, in the real world, the ball position estimates are noisy. To ensure that the policy is robust to perception noise, we add noise sampled from a uniform distribution to ball position during training. Further, to emulate large changes to ball position that might happen due to a human kicking the ball or the ball going outside the robot’s field of view, we also randomly teleport the ball in the ground plane. Finally, because the data rate of the camera is limited, we simulate camera communication delay. Noise model details are provided in Section VIII (Appendix).

3) *Robot System Identification:* To mitigate the sim-to-real gap in robot dynamics, we employ two standard and effective system identification measures: (a) Train an actuator network on real-world torque data to account for the non-ideal motor dynamics [9], [15], [34]; and (b) Identify and model the lag between the time the observation is measured and the time the action is applied [9], [15], [35].

4) *Ball-Terrain Interaction Model:* Human soccer players can quickly adapt to variations in ball dynamics, due to physical factors like air pressure and size as well as external perturbations to the ball due to uneven terrain or opposing players. We embed similar robustness in DribbleBot by implementing a custom ball drag model and random perturbation scheme in our training environment. Our ball drag model applies drag force proportional to the square of the velocity: $F_D = C_D v^2$, following the standard equation for aerodynamic drag. Different values of C_D serve to emulate terrains with various resistance forces, such as a field with tall grass (high C_D) or pavement (low C_D). In addition, we randomize the ball mass and apply random changes in ball velocity during training. Ball velocity randomization simulates external perturbations to the ball such as human intervention or contact with uneven terrain. The randomization range details are given in Table IV (Appendix).

5) *Fall Recovery Controller:* If the robot encounters a harsh perturbation such as a shove or a steep curb that results in locomotion failure, it will also cause loss of ball control. In this scenario, we would like the robot to get up from its fall and resume dribbling. Similar to prior works [34], [36], we train a dedicated recovery policy that enables the robot to return to a standing position from diverse fall scenarios. We first generate a set of 1000 initial fall configurations by randomly dropping the robot from different orientations, then train a policy with rewards for body orientation, base height, and action smoothness. The details of the reward function and training procedure for the recovery policy are provided in Table III (Appendix). To transition between dribbling and recovery policy, we define a finite state machine with transitions based on the body orientation. When the roll or pitch angle is larger than 1.0 rad, indicating locomotion failure, the recovery policy executes a return to the standing pose. When roll and pitch are smaller than 0.5 rad, the dribbling policy is reactivated.

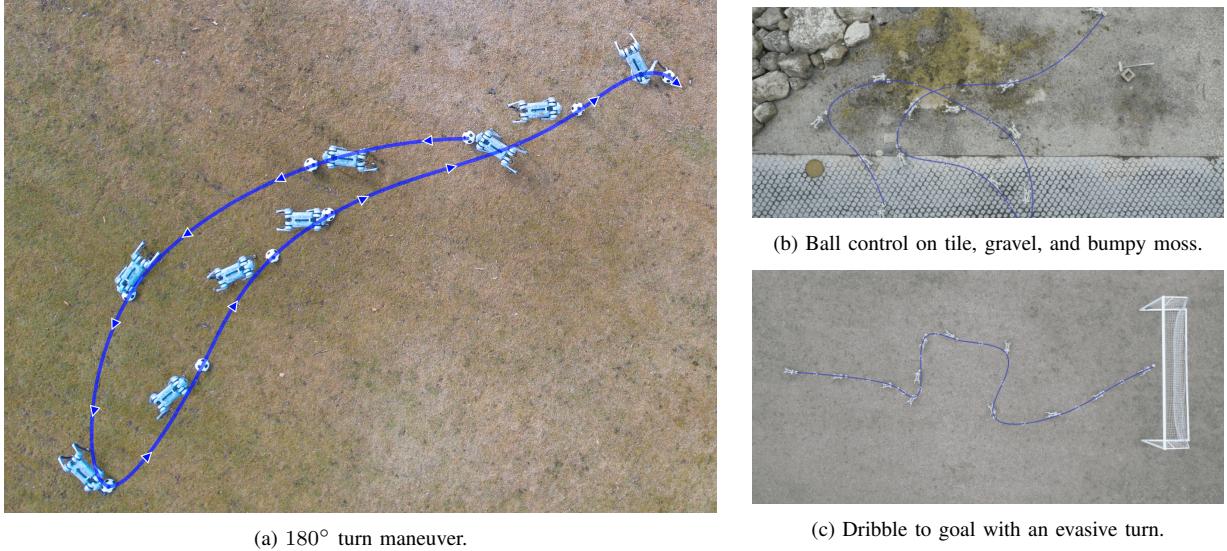


Fig. 4: Overhead images of DribbleBot during real-world deployment.

IV. RESULTS

A. Simulation Performance

1) Dribbling Control: We first evaluate our dribbling policy in simulation under the same conditions experienced during training. If the robot is dribbling well, we can characterize its performance by how closely it tracks the ball velocity command, how fast it can dribble, and how sharply it can turn. Figure 3 visualizes a long 60 s run, and shows the dribbling path, command tracking performance, and joint motion. The corresponding behavior is shown in Video S1.

We observe that the robot is able to track dribbling speeds up to 1.5 m/s and the entire range of instantaneous changes in command direction up to 180°. Unlike in ordinary locomotion, the task of turning the ball may extend across a long time horizon, during which the robot establishes control of the ball and executes multiple kicks. This incurs a delay between the change in command and the change in ball state, sometimes as long as several seconds (Figure 3b, 3c).

Dynamically dribbling a soccer ball requires task-oriented coordination of all the robot’s joints (*whole-body control*). Figure 3d shows that dribbling a soccer ball involves different participation of the right front leg during left and right turns. This leg is used to kick the ball during a left turn (blue highlight, Figure 3a upper images), and its motion is substantially changed during this maneuver. Later, when the robot makes a right turn (orange highlight, Figure 3a lower images), the left front leg is used to kick the ball, but the motion of the right leg also adjusts to stabilize the maneuver.

B. Real-world Deployment

1) Qualitative Results on Diverse Terrains: We qualitatively evaluate our dribbling controller under teleoperation on a number of terrains with different ball-terrain dynamics. A locomotion controller must adapt when the terrain causes the feet to slip or stumble. Dribbling additionally requires

that running and kicking adjust depending on how the ball interacts with the terrain, which may be very different due to the ball’s lighter mass and rolling surface contact. For example, on grass, high drag tends to slow down the rolling ball; on pavement, low drag may cause it to speed away from the robot; on gravel, the ball tends to roll but the robot slips; on snow or sand, both ball and robot slip; on bumpy dirt, the ball changes direction unpredictably as it impacts the terrain surface. DribbleBot is able to execute dribbling and turning motions on each terrain, tracking ball velocity commands from a human. These test terrains are illustrated in Figure 1, and the behavior is fully shown in the supplementary video.

Because our system operates without a tether or external sensing, it is capable of manipulating the soccer ball across large outdoor spaces. To illustrate this, we collect drone footage of the robot’s entire path across (a) a 180° turn, (b) a series of diverse terrains including tile, gravel, and bumpy moss, and (c) a 10 m run towards a soccer goal, with an evasive turning maneuver. Figure 4 shows stitched overhead photos to illustrate the real-world dribbling performance.

2) Quantitative Results and Ablation: We quantitatively evaluate the fully autonomous behavior of DribbleBot while executing a scripted trajectory across diverse terrains. The robot is commanded with a predetermined trajectory: dribble forward at 1.5 m/s for 10 s, then stop the ball for 5 s, then return towards the starting line at 1.5 m/s until the line is reached. We count a trial as a failure if the robot loses control of the ball, although if the loss is due to the robot falling, we allow it to autonomously recover and continue the attempt.

As shown in Table I, the robot executed four consecutive successful maneuvers using the full system design on tile, grass, and sand. Snow, step-down, and ramp are progressively more challenging and yield lower performance. Because the system was never exposed to steps or ramps during training, they are examples of out-of-distribution terrain for our policy. In the step-down task, the robot once fell in a pile of

snow and failed to recover, and once knocked the ball far away as it fell, and could not perceive it upon recovery. The ramp was traversed successfully under teleoperation, but the robot did not make substantial forward progress in the standard experiment when the dribbling commands were pre-specified. The ball repeatedly rolled behind the robot which turned around to recover it, remaining near the bottom.

We also conduct an ablation study to quantify the impact of our design choices on real-world performance. We evaluate the ablated configurations under the same methodology as above: No recovery controller ($-R$); No YOLO fine-tuning ($-Y$); No custom ball drag model during training ($-D$). Ablation results show that YOLO fine-tuning ($-Y$) is critical to performance, and recovery policy ($-R$) improves one run in the challenging curb environment. The system without ball drag model ($-D$) maintains control on both grass and tile, but the video supplement shows that it dribbles substantially slower on grass despite the equal velocity command. On sand, the policy without drag model fails twice during the turning maneuver after missing a kick on the ball. This suggests that the additional robustness from the drag model may also improve response to unexpected ball trajectories.

3) *Playing with a Human and Emergent Behaviors:* A real soccer match is not played alone, but with another agent who is seeking to control the ball. To understand the robot's behavior under this scenario, we explore the setting where the robot interactively plays with a human partner on both grassy field and flat ground (Video B1). Unlike the typical locomotion task, the task of controlling ball velocity affords the robot a high degree of freedom in its behavior, even when the user is not changing the command. Successful dribbling is not a monolithic skill: it often involves extended aperiodic movements to seek the ball, orient the body for a kick, and double back if the ball has been lost due to an unexpected perturbation, temporary perception or control failure.

V. RELATED WORK

A. Soccer Skills for Legged Robots

Soccer has long been an area of interest for roboticists. The RoboCup competition, this year in its 26th season, has attracted thousands of annual participants. RoboCup teams have implemented effective rule-based approaches to kicking, passing, and shooting in the past [16]–[18].

Recently, some works have applied learning to legged ball manipulation tasks in simulation [19], [37] and in externally instrumented indoor settings [21], [22], [38]. [38] demonstrated that a quadruped lying on its back can control and reorient a ball with its legs. A number of soccer skills such as dribbling [39] and juggling [37] have been demonstrated for physically simulated characters using reinforcement learning. [20] used imitation learning to perform static dribbling in the real-world indoor setting assisted by motion capture. [21] applied a hierarchical framework to the soccer shooting task in the real world, selecting the front right foot Bézier curve parameters as the low-level command inputs and leveraging a real-world fine-tuning stage to improve the shooting accuracy. [22] trained a control policy for jumping to block an

oncoming ball in an instrumented laboratory setting using sim-to-real reinforcement learning.

In addition to low-level skill learning, some work has focused on learning high-level soccer play end-to-end. Notably, [40] approaches the problems of muscle level control and long-horizon decision-making by first pretraining low-level skills using human soccer players' motion-capture video clips and then finding solutions for the multi-agent coordination goal in the low-level control space using reinforcement learning. To learn low-level skills like dribbling, [40] relies on motion capture data of human soccer players, which is not available for the quadruped form factor.

B. Dynamic Object Manipulation

Prior work has explored manipulating objects dynamically using a fixed or fully actuated base. [41] controlled a robotic arm to blindly perform ball juggling using an open-loop policy. Another work on robotic table tennis [42] estimated the ball state using an extended Kalman Filter which internally leverages a model of flight and bouncing behaviors. [43] learned a residue physics model to randomly pick up and throw a rigid object into a box. [44] bootstrapped a human behavior model and trained on both simulated and real data to learn a control policy for a table tennis-playing robot.

Another relevant line of work has investigated manipulating objects using a quadruped with mounted arm. [45] manipulated objects with a quadruped-mounted arm, coordinating the body and arm motion through learned estimation module. [46] implemented a model-based controller to manipulate objects with a quadruped-mounted arm in standing pose. [47] trained an end-to-end controller using reinforcement learning to perform coordinated manipulation with a quadruped-mounted arm under teleoperation and demonstrate vision-guided reaching using AprilTags. These works investigate complementary problems in the space of dynamic mobile manipulation.

VI. DISCUSSION

DribbleBot has a number of limitations which we hope to explore and improve upon in future work. We enumerate several here, with videos of failure cases available on the project website. *Slow turning response:* our system can execute sharp turns of the ball, but there is lag between the command onset and the actual turn (Figure 3). *Perception sensitivity to lighting:* We found that the perception module can perform poorly in bright, direct sunlight that produces glare from reflection on the ball and cameras. Fine-tuning the perception network with a more diverse set of outdoor images may resolve this problem. *Imprecision at high speeds:* If the ball is moving too fast on low-drag terrain, or a sharp instantaneous turn is commanded at high speed, a missed attempt to stop the ball can fail with the robot losing sight. *Lack of geometry awareness:* While the robot can dribble on slippery and uneven terrains, it cannot traverse larger obstacles like steep slopes and staircases with good consistency. Moreover, it is not aware of objects in the environment like poles and walls. Future work could incorporate more information about the

environment geometry into the controller to improve ball control in cluttered and harsh settings.

We believe there are many exciting frontiers to explore with a strong baseline for in-the-wild dribbling as a starting point. Dribbling is just one component of soccer. In particular, a combination of shooting [21] and goalkeeping [22] skills as well as high-level gameplay and awareness of other agents will be required to play a competitive game. Similarly, applying dynamic mobile manipulation for practical tasks like delivery and emergency response will require diverse skills, high-level planning, and rich world understanding. In-the-wild soccer may further be an interesting context in which to study human-robot interaction. While direct physical interaction with a legged robot is typically limited, interaction through the soccer ball as a shared medium proves rich and fun. Future work could explore how the robot is perceived by humans during play.

ACKNOWLEDGMENT

We thank the members of the Improbable AI lab for helpful discussions and feedback. We are grateful to MIT Supercloud and the Lincoln Laboratory Supercomputing Center for providing HPC resources. This research was supported by the DARPA Machine Common Sense Program, the MIT-IBM Watson AI Lab, and the National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, <http://iaifi.org/>). This research was also sponsored by the United States Air Force Research Laboratory and the United States Air Force Artificial Intelligence Accelerator and was accomplished under Cooperative Agreement Number FA8750-19-2-1000. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the United States Air Force or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes, notwithstanding any copyright notation herein.

AUTHOR CONTRIBUTIONS

- **Yandong Ji** contributed to ideation and implementation of the entire system, experimental evaluation, and writing.
- **Gabriel B. Margolis** contributed to ideation and implementation of the entire system, experimental evaluation, and writing.
- **Pulkit Agrawal** advised the project and contributed to its development, experimental design, and writing.

REFERENCES

- [1] R. Deits and T. Koolen, “Picking up momentum,” Jan 2023. [Online]. Available: <https://www.bostondynamics.com/resources/blog/picking-momentum>
- [2] J. Tan, T. Zhang, E. Coumans, A. Iscen, Y. Bai, D. Hafner, S. Bohez, and V. Vanhoucke, “Sim-to-real: Learning agile locomotion for quadruped robots,” in *Proc. Robot.: Sci. and Syst. (RSS)*, Pittsburgh, Pennsylvania, USA, June 2018, pp. 1–9.
- [3] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter, “Learning agile and dynamic motor skills for legged robots,” *Sci. Robot.*, vol. 4, no. 26, p. aau5872, Jan. 2019.
- [4] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, “Learning quadrupedal locomotion over challenging terrain,” *Sci. Robot.*, vol. 5, no. 47, p. eabc5986, Oct. 2020.
- [5] A. Kumar, Z. Fu, D. Pathak, and J. Malik, “RMA: Rapid motor adaptation for legged robots,” in *Proc. Robot.: Sci. and Syst. (RSS)*, Virtual, July 2021.
- [6] J. Siekmann, K. Green, J. Warila, A. Fern, and J. Hurst, “Blind bipedal stair traversal via sim-to-real reinforcement learning,” in *Proc. Robot.: Sci. and Syst. (RSS)*, Virtual, July 2021.
- [7] T. Miki, J. Lee, J. Hwanbo, L. Wellhausen, V. Koltun, and M. Hutter, “Learning robust perceptive locomotion for quadrupedal robots in the wild,” *Sci. Robot.*, vol. 7, no. 62, p. abk2822, Jan. 2022.
- [8] G. B. Margolis, G. Yang, K. Paigwar, T. Chen, and P. Agrawal, “Rapid locomotion via reinforcement learning,” in *Proc. Robot.: Sci. and Syst. (RSS)*, June 2022.
- [9] G. B. Margolis and P. Agrawal, “Walk these ways: Tuning robot control for generalization with multiplicity of behavior,” in *Proc. Conf. Robot Learn. (CoRL)*, Auckland, New Zealand, Dec. 2022.
- [10] A. Agarwal, A. Kumar, J. Malik, and D. Pathak, “Legged locomotion in challenging terrains using egocentric vision,” in *Proc. Conf. Robot Learn. (CoRL)*, Auckland, New Zealand, Dec. 2022.
- [11] S. Choi, G. Ji, J. Park, H. Kim, J. Mun, J. H. Lee, and J. Hwangbo, “Learning locomotion on deformable terrain,” *Sci. Robot.*, vol. 8, no. 74, p. eade2256, 2023.
- [12] O. M. Andrychowicz, B. Baker, M. Chociej, R. Jozefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray, et al., “Learning dexterous in-hand manipulation,” *Int. J. Robot. Res. (IJRR)*, vol. 39, no. 1, pp. 3–20, 2020.
- [13] OpenAI, I. Akkaya, M. Andrychowicz, M. Chociej, M. Litwin, B. McGrew, A. Petron, A. Paino, M. Plappert, G. Powell, R. Ribas, J. Schneider, N. Tezak, J. Tworek, P. Welinder, L. Weng, Q. Yuan, W. Zaremba, and L. Zhang, “Solving Rubik’s cube with a robot hand,” *arXiv preprint*, 2019.
- [14] T. Chen, J. Xu, and P. Agrawal, “A system for general in-hand object re-orientation,” in *Proc. Conf. Robot Learn. (CoRL)*, London, UK, Nov. 2021, pp. 297–307.
- [15] T. Chen, M. Tippur, S. Wu, V. Kumar, E. Adelson, and P. Agrawal, “Visual dexterity: In-hand dexterous manipulation from depth,” *arXiv preprint arXiv:2211.11744*, 2022.
- [16] M. Veloso, W. Uther, M. Fijita, M. Asada, and H. Kitano, “Playing soccer with legged robots,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, vol. 1. IEEE, 1998, pp. 437–442.
- [17] P. Stone, “Intelligent autonomous robotics: A robot soccer case study,” *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 1, no. 1, pp. 1–155, 2007.
- [18] M. Friedmann, J. Kiener, S. Petters, D. Thomas, O. Von Stryk, and H. Sakamoto, “Versatile, high-quality motions and behavior control of a humanoid soccer robot,” *International Journal of Humanoid Robotics*, vol. 5, no. 03, pp. 417–436, 2008.
- [19] A. F. Muzio, M. R. Maximo, and T. Yoneyama, “Deep reinforcement learning for humanoid robot behaviors,” *Journal of Intelligent & Robotic Systems*, vol. 105, no. 1, pp. 1–16, 2022.
- [20] S. Bohez, S. Tunyasuvunakool, P. Brakel, F. Sadeghi, L. Hasenclever, Y. Tassa, E. Parisotto, J. Humprik, T. Haarnoja, R. Hafner, M. Wulfmeier, M. Neunert, B. Moran, N. Siegel, A. Huber, F. Romano, N. Batchelor, F. Casarini, J. Merel, R. Hadsell, and N. Heess, “Imitate and repurpose: Learning reusable robot movement skills from human and animal behaviors,” *arXiv preprint arXiv:2203.17138*, 2022.
- [21] Y. Ji, Z. Li, Y. Sun, X. B. Peng, S. Levine, G. Berseth, and K. Sreenath, “Hierarchical reinforcement learning for precise soccer shooting skills using a quadrupedal robot,” *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, Oct. 2022.
- [22] X. Huang, Z. Li, Y. Xiang, Y. Ni, Y. Chi, Y. Li, L. Yang, X. B. Peng, and K. Sreenath, “Creating a Dynamic Quadrupedal Robotic Goalkeeper with Reinforcement Learning,” *arXiv preprint arXiv:2210.04435*, 2022.
- [23] Unitree Robotics, Go1, 2023, <https://www.unitree.com/go1>, [Online; accessed Jan. 2023].
- [24] V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa, and G. State, “Isaac Gym: High performance GPU-based physics simulation for robot learning,” *arXiv preprint*, 2021.
- [25] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [26] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” *arXiv preprint arXiv:2207.02696*, 2022.
- [27] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft coco: Common objects in context,” in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13*. Springer, 2014, pp. 740–755.

- [28] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” *arXiv preprint*, 2017.
- [29] N. Rudin, D. Hoeller, M. Bjelonic, and M. Hutter, “Advanced Skills by Learning Locomotion and Local Navigation End-to-End,” in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 2497–2503.
- [30] J. Siekmann, Y. Godse, A. Fern, and J. Hurst, “Sim-to-real learning of all common bipedal gaits via periodic reward composition,” in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, Xi'an, China, June 2021, pp. 7309–7315.
- [31] H. Duan, A. Malik, J. Dao, A. Saxena, K. Green, J. Siekmann, A. Fern, and J. Hurst, “Sim-to-real learning of footstep-constrained bipedal dynamic walking,” in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, 2022, pp. 10428–10434.
- [32] N. Rudin, D. Hoeller, P. Reist, and M. Hutter, “Learning to walk in minutes using massively parallel deep reinforcement learning,” in *Proc. Conf. Robot Learn. (CoRL)*, London, UK, Nov. 2021, pp. 91–100.
- [33] G. Ji, J. Mun, H. Kim, and J. Hwangbo, “Concurrent training of a control policy and a state estimator for dynamic and robust legged locomotion,” *IEEE Robot. Automat. Lett. (RA-L)*, vol. 7, no. 2, pp. 4630 – 4637, Apr. 2022.
- [34] J. Lee, J. Hwangbo, and M. Hutter, “Robust recovery controller for a quadrupedal robot using deep reinforcement learning,” *arXiv preprint arXiv:1901.07517*, 2019.
- [35] Z. Xie, X. Da, M. van de Panne, B. Babich, and A. Garg, “Dynamics randomization revisited: A case study for quadrupedal locomotion,” in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, Virtual, May 2021, pp. 4955–4961.
- [36] L. Smith, J. C. Kew, X. B. Peng, S. Ha, J. Tan, and S. Levine, “Legged robots that keep on learning: Fine-tuning locomotion policies in the real world,” in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*. IEEE, 2022, pp. 1593–1599.
- [37] Z. Xie, S. Starke, H. Y. Ling, and M. van de Panne, “Learning soccer juggling skills with layer-wise mixture-of-experts,” in *ACM SIGGRAPH 2022 Conference Proceedings*, 2022, pp. 1–9.
- [38] F. Shi, T. Homberger, J. Lee, T. Miki, M. Zhao, F. Farshidian, K. Okada, M. Inaba, and M. Hutter, “Circus anymal: A quadruped learning dexterous manipulation with its limbs,” in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*. IEEE, 2021, pp. 2316–2323.
- [39] X. B. Peng, Z. Ma, P. Abbeel, S. Levine, and A. Kanazawa, “AMP: Adversarial motion priors for stylized physics-based character control,” *ACM Transactions on Graphics (TOG)*, vol. 40, no. 4, pp. 1–20, 2021.
- [40] S. Liu, G. Lever, Z. Wang, J. Merel, S. M. A. Eslami, D. Hennes, W. M. Czarnecki, Y. Tassa, S. Omidshafiei, A. Abdolmaleki, N. Y. Siegel, L. Hasenclever, L. Marris, S. Tunyasuvunakool, H. F. Song, M. Wulfmeier, P. Muller, T. Haarnoja, B. Tracey, K. Tuyls, T. Graepel, and N. Heess, “From motor control to team play in simulated humanoid football,” *Science Robotics*, vol. 7, no. 69, p. eab0235, 2022.
- [41] K. Ploeger, M. Lutter, and J. Peters, “High acceleration reinforcement learning for real-world juggling with binary rewards,” *arXiv preprint arXiv:2010.13483*, 2020.
- [42] K. Mülling, J. Kober, O. Kroemer, and J. Peters, “Learning to select and generalize striking movements in robot table tennis,” *Int. J. Robot. Res. (IJRR)*, vol. 32, no. 3, pp. 263–279, 2013.
- [43] A. Zeng, S. Song, J. Lee, A. Rodriguez, and T. Funkhouser, “Tossing-bot: Learning to throw arbitrary objects with residual physics,” *IEEE Trans. Robot. (T-RO)*, vol. 36, no. 4, pp. 1307–1319, 2020.
- [44] S. Abeyruwan, L. Graesser, D. B. D'Ambrosio, A. Singh, A. Shankar, A. Bewley, and P. R. Sanketi, “i-sim2real: Reinforcement learning of robotic policies in tight human-robot interaction loops,” *arXiv preprint arXiv:2207.06572*, 2022.
- [45] M. Mittal, D. Hoeller, F. Farshidian, M. Hutter, and A. Garg, “Articulated object interaction in unknown scenes with whole-body mobile manipulation,” in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 1647–1654.
- [46] Y. Ma, F. Farshidian, T. Miki, J. Lee, and M. Hutter, “Combining learning-based locomotion policy with model-based manipulation for legged mobile manipulators,” *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2377–2384, 2022.
- [47] Z. Fu, X. Cheng, and D. Pathak, “Deep whole-body control: learning a unified policy for manipulation and locomotion,” *arXiv preprint arXiv:2210.10044*, 2022.

TABLE II: Notation.

Parameter	Definition	Units	Dimension
<i>System Components</i>			
π_d	Dribbling Policy	-	-
π_r	Recovery Policy	-	-
\mathbf{Y}	YOLOv7 Network [26]	-	-
<i>Robot State</i>			
\mathbf{q}	Joint Angles	rad	12
$\dot{\mathbf{q}}$	Joint Velocities	rad/s	12
$\ddot{\mathbf{q}}$	Joint Accelerations	rad/s	12
$\boldsymbol{\tau}$	Joint Torques	rad/s	12
\mathbf{g}	Gravity Unit Vector, Body Frame	m/s ²	3
ψ_t	Body Yaw, Global Frame	rad	1
\mathbf{q}_{des}	Joint Position Targets	rad	12
$\boldsymbol{\theta}^{\text{cmd}}$	Timing Reference Variables [9]	-	4
$\mathbf{p}_{\text{FRHip}}$	Front Right Hip Position	m	3
<i>Ball State</i>			
\mathbf{o}^v	Fisheye Camera Image	-	400 × 480
\mathbf{b}	Ball Position, Body Frame	m	3
$\hat{\mathbf{b}}$	Estimated Ball Position, Body Frame	m	3
\mathbf{v}^b	Ball Velocity, Global Frame	m/s	2
\mathbf{v}^{cmd}	Command Ball Velocity, Global Frame	m/s	2
ψ_b	Direction of Ball Velocity	rad	1
ψ_{cmd}	Direction of Command Ball Velocity	rad	1
<i>Control Policy</i>			
\mathbf{o}	Policy Observation	-	37 × 15
\mathbf{a}	Policy Action	-	12
\mathbf{c}	Command	-	2

APPENDIX

VII. REWARD STRUCTURE

Dribbling Policy: Table III provides the reward terms for learning soccer dribbling. \mathbf{v}^b and \mathbf{v}^{cmd} are the desired and commanded ball velocity in the global reference frame. ψ_b and ψ_{cmd} are the desired and commanded direction of the ball velocity in the global reference frame, which can be directly computed from \mathbf{v}^b and \mathbf{v}^{cmd} . \mathbf{b} is the ball position in the body frame. $\mathbf{p}_{\text{FRHip}}$ is the front right hip position in the body frame. e_{rbcmd} represents the angle difference between the robot ball vector and ψ_b . e_{rbbase} is the angle difference between the base yaw angle and ψ_b . κ is the target contact state $\mathbf{C}_{\text{foot}}^{\text{cmd}}(\boldsymbol{\theta}^{\text{cmd}}, t)$, where, as in prior work [9], \mathbf{f}^{foot} is foot contact force in stance phase, \mathbf{v}^{foot} is foot velocity in swing phase, $\mathbf{C}_{\text{foot}}^{\text{cmd}}$ denotes desired foot contact sequence, and $\boldsymbol{\theta}^{\text{cmd}}$ denotes timing reference variables. $\boldsymbol{\tau}$, \mathbf{q} , $\dot{\mathbf{q}}$, $\ddot{\mathbf{q}}$ denote joint torque, position, velocity and acceleration. i is the index of joints. \mathbf{g}_t^{xy} denotes the gravity unit vector projected onto the robot transverse plane. \mathbf{a}_t denotes the action at timestep t . $|\cdot|$ denotes l^2 -norm. The total reward at timestep t is represented as $r_t = r_t^{\text{pos}} \exp(r_t^{\text{neg}})$, where r_t^{pos} and r_t^{neg} represent positive task reward and negative penalizing reward respectively [33].

Recovery Policy: Table III provides the reward terms for learning the recovery policy, inspired by [34], [36]. Here, \mathbf{g}_z is the vertical component of the gravity unit vector in the body frame. When the robot is perfectly upright, $\mathbf{g}_z = -1$. \mathcal{I} is shorthand for an indicator variable $\mathbb{1}_{\mathbf{g}_z < -0.6}$ which denotes that the body height, body pose, and foot height rewards are only activated when the robot's body is nearly upright. clamp clamps the value of its input between 0 and 1.

VIII. NOISE MODEL

Robot Physics Randomization: We randomize the robot's payload mass, motor strength, joint calibration, foot friction, foot

TABLE III: Reward terms for ball dribbling and recovery policies.

Ball Dribbling Policy		
Term	Expression	Weight
Projected Ball Velocity	$\exp\{-\delta_v \mathbf{v}^b - \mathbf{v}^{\text{cmd}} ^2\}$	0.5
Robot Ball Distance	$\exp\{-\delta_p \mathbf{b} - \mathbf{p}_{\text{FRHip}} ^2\}$	4.0
Yaw Alignment	$\exp\{-\delta_\psi (e_{\text{rbcmd}}^2 + e_{\text{rbbase}}^2)\}$	4.0
Ball Velocity Norm	$\exp\{-\delta_n (\mathbf{v}^{\text{cmd}} - \mathbf{v}^b)^2\}$	4.0
Ball Velocity Angle	$1 - (\psi_b - \psi_{\text{cmd}})^2 / \pi^2$	4.0
Swing Phase Schedule	$[1 - \kappa] \exp\{-\delta_{\text{cf}} \mathbf{f}^{\text{foot}} ^2\}$	4.0
Stance Phase Schedule	$\kappa \exp\{-\delta_{\text{cv}} \mathbf{v}_{\text{xy}}^{\text{foot}} ^2\}$	4.0
Joint Limit Violation	$\mathbb{1}_{q_i > q_{\max}} \mathbb{1}_{q_i < q_{\min}}$	-10.0
Joint Torque	$ \boldsymbol{\tau} ^2$	-0.0001
Joint Velocity	$ \dot{\mathbf{q}} $	-0.0001
Joint Acceleration	$ \ddot{\mathbf{q}} $	-2.5e-7
Hip/Thigh Collision	$\mathbb{1}_{\text{collision}}$	-5.0
Projected Gravity	$ \mathbf{g}_{\text{xy}} ^2$	-5.0
Action Smoothing	$ \mathbf{a}_{t-1} - \mathbf{a}_t ^2$	-0.1
Action Smoothing, 2nd Order	$ \mathbf{a}_{t-2} - 2\mathbf{a}_{t-1} + \mathbf{a}_t ^2$	-0.1
Recovery Policy		
Term	Expression	Weight
Body Orientation	$(0.5 - 0.5\mathbf{g}_z)^2$	1.0
Body Height	$\mathcal{I}(1.0 - \text{clamp}((h_{\text{target}}^{\text{body}} - h_{\text{body}}^{\text{body}})/h_{\text{target}}^{\text{body}})^2)$	1.0
Body Pose	$\mathcal{I}(1.0 - \text{clamp}(q - q_{\text{standing}} ^2/20.0))$	1.0
Foot Height	$\mathcal{I}(\exp(-10 \sum_i (h_i^{\text{foot}})^2))$	1.0
Action	$ \mathbf{a}_t ^2$	-1e-3
Joint Torque	$ \boldsymbol{\tau} ^2$	-1e-5

restitution, and center of mass displacement. Table IV provides the ranges of randomized parameters.

Ball Physics Randomization: We randomize the ball mass and ball-terrain drag coefficient. Table IV provides the ranges of randomized parameters.

Ball Teleportation: We teleport the ball to a uniformly sampled random location within 1.0 m at regular intervals of 7.0 s.

Camera Delay: We model the arrival time of the next observation as a Poisson distribution with mean arrival time randomized each episode between 20ms and 60ms.

IX. POLICY OPTIMIZATION

We used the same set of PPO hyperparameters for training the dribbling and recovery policies. Table V provides these hyperparameter values. They are the same settings used in prior work for training locomotion policies on this robot [9].

TABLE IV: Randomization ranges for robot dynamics, ball dynamics, and commands during training.

Dynamics Parameter	Range	Units
<i>Robot Dynamics</i>		
Payload Mass	[−1.0, 3.0]	kg
Motor Strength	[90, 110]	%
Joint Calibration	[−0.02, 0.02]	rad
Robot-Terrain Friction	[0.40, 1.00]	–
Robot-Terrain Restitution	[0.00, 1.00]	–
Robot Center of Mass Displacement	[−0.15, 0.15]	m
<i>Ball Dynamics</i>		
Mass	[0.159, 0.254]	kg
Perception Arrival Rate	[0.3, 0.7]	–
Teleporting Position	[0.0, 1.0]	m
Perturbation Velocity	[0.0, 0.3]	m/s
Ball-Terrain Drag Coefficient	[0.0, 1.5]	–
<i>Command</i>		
$\mathbf{v}_x^{\text{cmd}}$	[−1.5, 1.5]	m/s
$\mathbf{v}_y^{\text{cmd}}$	[−1.5, 1.5]	m/s

Hyperparameter	Value
Discount factor	0.99
GAE parameter	0.95
Timesteps per rollout	21
Epochs per rollout	5
Minibatches per epoch	4
Entropy bonus (α_2)	0.01
Value loss coefficient (α_1)	1.0
Clip range	0.2
Reward normalization	yes
Learning rate	1e−3
# Environments	4096
# Total timesteps	7B
Optimizer	Adam

TABLE V: PPO hyperparameters.