Supplementary Materials

Anonymous Authors
https://dribble-master.github.io

I. REWARDS DESIGN

We have divided all the reward functions to two parts: locomotion rewards and dribbling rewards. In the 1st training stage, locomotion rewards are given higher weights to encourage stable walking and ball approaching, while in the 2nd training stage dribbling rewards are giving high weights for precise and agile dribbling skill. We also design active sensing reward to promote active perception. Table I summarizes the task rewards used for the two-stage learning of agile dribbling skills.

II. ROBOT PLATFORM

The simulated robot is constructed from the URDF model of the Booster T1 humanoid robot (Fig. 1), featuring 23 degrees of freedom (DOFs) distributed throughout the body. To simplify control and focus on locomotion and perception, we actuate 12 joints in the legs and 2 in the head, which are critical for walking and ball tracking. The remaining upper body joints, including those in the arms and torso, are fixed in their default positions to reduce complexity and computational cost. We also deploy trained policy to real Booster T1 humanoid robot for real-world experiments.
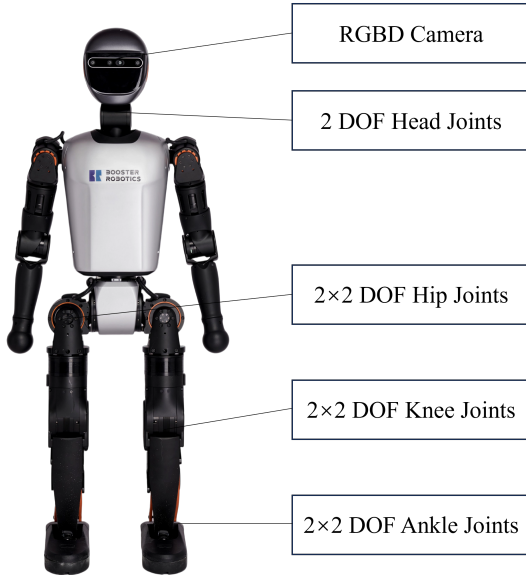

Fig. 1: Booster T1 Humanoid Robot

Category	Reward	Definition	1st Stage Scale	2nd Stage Scale
Locomotion Rewards	Base Orientation	$\exp\{-|(\theta^b_{pitch})^2 + (\theta^b_{roll})^2|\}$	1.0	1.0
	Feet Orientation	$\exp\{-|(\theta^f_{pitch})^2 + (\theta^f_{roll})^2|\}$	1.0	1.0
	Feet Distance	$|\|P^f_{xy,L} - P^f_{xy,R}\|_2 - D_t|$	1.0	1.0
	Feet Clearance	$\exp\{-2\|P^f_z - P^f_{z,t}\|_2\}$	2.0	2.0
	Termination	$n_{termination}$	-10.0	-10.0
	Reference Joint Position	$\exp\{-|\mathbf{q_t} - \mathbf{q}|^2\}$	1.2	0.0
	Symmetric Action	$\|\|\Sigma_t^{t+T}|\boldsymbol{\tau}_L|dt\|^2 - \|\Sigma_t^{t+T}|\boldsymbol{\tau}_R|dt\|^2|$	-0.001	-0.001
	Joint Torque	$\|\boldsymbol{\tau}\|^2$	-0.015	-0.015
	Joint Speed	$\|\dot{\mathbf{q}}\|$	-0.0001	-0.0001
	Action Smoothness	$\|a_t - 2a_{t-1} + a_{t-2}\|_2$	-0.01	-0.01
	Active Sensing	$\mathbf{1}(\textbf{ball in view})$	0.2	0.2
	Chasing	$\exp\{-2|\mathbf{b} - \mathbf{p}_{\text{com}}|^2\}$	2.0	1.0
Dribbling Rewards	Projected Ball Velocity	$\exp\{-|\mathbf{v}^b - \mathbf{v}^{\text{cmd}}|^2\}$	0.0	1.5
	Yaw Alignment	$\exp\{-2((\theta^{base-cmd}_{yaw})^2 + (\theta^{base-ball}_{yaw})^2)\}$	0.2	0.2
	Yaw Alignment (No ball in view)	$|\dot{\theta}_{yaw} - \dot{\theta}_{yaw,max} sign(\theta^{base-ball}_{yaw})|$	-0.2	-0.2

TABLE I: Reward functions and scales at each curriculum learning phase.