

VIRAL: Visual Sim-to-Real at Scale for Humanoid Loco-Manipulation

Supplementary Material

7. Training Details

7.1. Observation Details

Table 1 summarizes the observation terms and their corresponding dimensions.

State term	Dimensions
Base linear velocity	3
Base angular velocity	3
Projected gravity	3
Actions	31
Stage	5
Delta actions	11
DoF position	43
DoF velocity	43
Placement position	2
Table–pelvis transform	9
Finger-tip forces for hold_object	12
Hold_object transform	9
Hold_object–hand transform	9
Target pre-place position	3
Finger-tip forces for grasp_object	12
Grasp_object transform	9
Grasp_object–hand transform	9
Target lift position	3
HOMIE commands	7
Single-step total dim	226

Table 1. Observation dimensions for `teacher`.

Table 2 lists the observation terms and their corresponding dimensions. In addition to these state observations, we feed an RGB image of size 108×192 into the vision encoder. The resulting 128-dimensional visual feature is concatenated with the state observations and then passed to the policy head.

State term	Dimensions
Base angular velocity	3
Projected gravity	3
Actions	31
DoF position (w/o fingers)	29
DoF velocity (w/o fingers)	29
Delta actions	11
HOMIE commands	7
Single-step total dim	113

Table 2. Observation dimensions for `student`.

7.2. Reward Details

A single place–pickup cycle is decomposed into five stages: (1) walking toward the object; (2) moving the arm and hand to a pre-place pose; (3) placing the object; (4) grasping and lifting the next object; and (5) turning. Repeating this sequence produces a long-horizon loco-manipulation loop. At each step, the total reward is a stage-weighted sum

$$r_t = \sum_{i=0}^4 w_i \mathbb{1}[s_t = i] r_t^{(i)}, \quad w_i > 0,$$

and stage transitions are governed by stage-specific advancement and completion criteria. Table 3 instantiates $r^{(s)}$ with stage-dependent shaping terms for teacher policy.

7.3. Hyperparameters Details

Table 4 lists the PD gains used for the Unitree G1 robot equipped with 3-finger dexterous hands.

Table 5 lists the hyperparameters for teacher policy trained by PPO [59].

Table 6 lists the hyperparameters for student policy trained by the mixture of DAgger [56] and Behavior Cloning.

7.4. Domain Randomization

Table 7 summarizes all randomizations used during policy training, including image quality, dome lighting, materials, table properties, and camera extrinsics.

Term	Expression	Weight	Stage(s)
Termination / generic penalties			
Termination	$\mathbb{1}_{\{\text{termination}\}}$	-2000.0	0-4
Delta action rate	$\ \Delta a_t\ _2^2$	-0.01	0-4
DoF velocity	$\ \dot{\mathbf{q}}\ _2^2$	-0.5	0-4
DoF acceleration	$\ \ddot{\mathbf{q}}\ _2^2$	-3.0×10^{-6}	0-4
Torque limits	$\ \boldsymbol{\tau}\ _2^2$	-0.001	0-4
Output smoothness	$\ \pi_t - \pi_{t-1}\ _2^2$	-9.0	0-4
Finger primitive limits	$ \text{clip}(u_{\text{finger}}, [l, u]) - u_{\text{finger}} $	-20.0	0-4
Fast right-arm velocity	$\ \mathbf{q}_{\text{right arm}}\ _2^2$	-80.0	0-4
Finger qvel, when contacting ground with single-foot	$\ \dot{\mathbf{q}}_{\text{finger}}\ _2 \mathbb{1}_{\text{single-foot}}$	-3000.0	1-3
Arm qvel, when contacting ground with single-foot	$\ \dot{\mathbf{q}}_{\text{right arm}}\ _2 \mathbb{1}_{\text{single-foot}}$	-1300.0	1-3
Heading / command shaping			
Heading toward object	$((\psi_{\text{GraspObj}} - \psi_{\text{robot}})/\pi)^2$	-10000.0	0
Object in view	$\mathbb{1}[y_{\text{right hand}} > y_{\text{GraspObj}} - 0.1] + \mathbb{1}[y_{\text{left hand}} < y_{\text{GraspObj}} + 0.1]$	-1.0	0
Large linear v_x command	$\sum \max(0, v_x^{\text{cmd}} - 0.5)$	-20.0	0-4
Large linear v_y command	$\sum \max(0, v_y^{\text{cmd}} - 0.5)$	-20.0	0-4
Large angular ω command	$\sum \max(0, \omega^{\text{cmd}} - 0.5)$	-20.0	0-4
Large upper-body actions	$\sum \max(0, u_{\text{upper}} - 2\pi)$	-20.0	0-4
Zero linear v_x , linear v_y , angular ω cmd	$ v_x^{\text{cmd}} + v_y^{\text{cmd}} + \omega^{\text{cmd}} $	-12.0	1-3
Zero linear v_x , linear v_y cmd	$ v_x^{\text{cmd}} + v_y^{\text{cmd}} $	-4.0	4
Task / object-centric rewards			
Robot-Object distance	$\exp(-4(\ p_{\text{robot}} - p_{\text{GraspObj}}\ - 0.45)^2)$	2.0	0-4
Upper-body actions (pose)	$\ \mathbf{q}_{\text{right arm}}\ _2^2$	-1.0	0
Keep hand closed	$\exp(-4(u_{\text{finger}} - u_{\text{close}})^2)$	9.0	0-1, 3-4
Place objects when near tray	$-\ \mathbf{f}_{\text{placeObj}}\ * \mathbb{1}(\ p_{\text{placeObj}} - p_{\text{tray}}\ < 0.3)$	10.0	0-1
Holding object	$\exp(-4\ p_{\text{placeObj}} - p_{\text{hand}}\ _2)$	1.0	0-4
Hand-object distance	$\exp(-10 \max_k \ p_{\text{finger}}^{(k)} - p_{\text{GraspObj}}\ _2)$	20.0	3-4
Grasp based on obj-finger dir	$-\hat{\mathbf{d}}_{\text{thumb}}^\top \hat{\mathbf{d}}_{\text{index}}$	5.0	3-4
Grasp force	$\sum \ \mathbf{f}_{\text{GraspObj-hand}}\ $	1.0	3-4
Lift goal distance	$\exp(-10\ p_{\text{GraspObj}} - p_{\text{goal}}\ ^2)$	10.0	3-4
Lift z	$\min(h_{\text{GraspObj}} - h_{\text{table}}, 0.15)$	200.0	3-4
Turn around	$- y_{\text{robot}} - y_{\text{desired}} $	15.0	4
Right-arm qpos tracking (hold)	$\exp(-4\ \mathbf{q}_{\text{right arm}} - \mathbf{q}_{\text{place}}^*\ _2)$	5.0	0-2
Right-arm qpos tracking (front)	$\exp(-4\ \mathbf{q}_{\text{right arm}} - \mathbf{q}_{\text{Grasp}}^*\ _2)$	25.0	3-4
Finger qvel during right-arm qvel	$\exp(-6\ \dot{\mathbf{q}}_{\text{arm}}\ _2 \ \dot{\mathbf{q}}_{\text{finger}}\ _2)$	15.0	1-4
Object-table contact move	$\ \mathbf{v}_{\text{GraspObj},xy}\ \mathbb{1}_{\text{table-contact}}$	-1000.0	1-4
Object relative move (hand-obj v_z)	$ v_{\text{GraspObj}}^z - v_{\text{hand}}^z \mathbb{1}_{\text{in-grasp}}$	-3000.0	1-3
Object lean during pick	$ \phi_{\text{GraspObj}} + \theta_{\text{GraspObj}} $	-500.0	0-3
Object non- z velocity during pick	$\ \mathbf{v}_{\text{GraspObj},xy}\ _2$	-500.0	0-3

Table 3. Reward components, expressions, weights, and the stages (0-4) where each term is applied.

Joint	K_p [N·m/rad]	K_d [N·m·s/rad]
hip_yaw	150	2.0
hip_roll	150	2.0
hip_pitch	150	2.0
knee	200	4.0
ankle_pitch	40	2.0
ankle_roll	40	2.0
waist_yaw	250	5.0
waist_roll	250	5.0
waist_pitch	250	5.0
shoulder_pitch	100	5.0
shoulder_roll	100	5.0
shoulder_yaw	40	2.0
elbow	40	2.0
wrist_roll	20	2.0
wrist_pitch	20	2.0
wrist_yaw	20	2.0
hand_index	0.5	0.1
hand_middle	0.5	0.1
hand_thumb_1	0.5	0.1
hand_thumb_2	0.5	0.1
hand_thumb_0	2.0	0.1

Table 4. Joint-space PD gains (K_p , K_d) used in the low-level controller.

Hyperparameters	Values
Number of environments	32768 (2048*8GPUs*2Nodes)
Discount factor (γ)	0.998
Learning rate	0.00002
Entropy coefficient	0.01
Value loss coefficient	1
Init noise std (RL)	0.5
MLP size	[512, 256, 128]

Table 5. Hyperparameters for teacher policy.

Hyperparameters	Values
Number of environments	65535 (1024*8GPUs*8Nodes)
Number of steps per environment	1
Learning rate	0.0002

Table 6. Hyperparameters for student policy.

Table 7. Comprehensive domain randomization parameters during training

Parameter	Probability	Distribution
Image Augmentation		
Brightness	0.25	$\sim \mathcal{U}(0.7, 2)$
Contrast	0.25	$\sim \mathcal{U}(0.5, 1.5)$
Hue	0.5	$\sim \mathcal{U}(-0.1, 0.1)$
Saturation	0.25	$\sim \mathcal{U}(0.5, 2)$
Gaussian Noise Std	0.25	$\sim \mathcal{U}(0.0, 0.15)$
Gaussian Blur Kernel Size	0.25	$\sim \mathcal{U}(3, 5)$
Gaussian Blur Sigma	0.25	$\sim \mathcal{U}(0.1, 1.5)$
Lighting		
Dome Light Intensity	1.0	$\sim \mathcal{U}(800, 2000)$
Dome Light Yaw Rotation	1.0	$\sim \mathcal{U}(-\pi, \pi)$
Dome Light Texture Map	1.0	$\sim \mathcal{U}(\text{texture_maps})$ (Indoor, Clear, Cloudy, Night, Studio)
Material Randomization		
Robot Material - Roughness	1.0	$\sim \mathcal{U}(0.0, 0.8)$
Robot Material - Metallic	1.0	$\sim \mathcal{U}(0.0, 0.8)$
Robot Material - Specular	1.0	$\sim \mathcal{U}(0.0, 0.8)$
Floor Material Texture	1.0	$\sim \mathcal{U}(\text{texture_maps})$ (Wood, Carpet, Masonry, Metals, Natural, Plastics, Stone, Wall Board)
Table Material Texture	1.0	$\sim \mathcal{U}(\text{texture_maps})$ (Wood)
Object Material Texture	1.0	$\sim \mathcal{U}(\text{texture_maps})$ (All Base Materials)
Table Physical Properties		
Table Height (m)	1.0	$\sim \mathcal{U}(0.65, 0.6775)$
Table Depth (m)	1.0	$\sim \mathcal{U}(0.7, 0.75)$
Table Width (m)	1.0	$\sim \mathcal{U}(1.4, 1.6)$
Table Thickness (m)	1.0	$\sim \mathcal{U}(0.035, 0.04)$
Camera Extrinsic		
Position Noise - X (m)	1.0	$\sim \mathcal{U}(-0.02, 0.02)$
Position Noise - Y (m)	1.0	$\sim \mathcal{U}(-0.05, 0.05)$
Position Noise - Z (m)	1.0	$\sim \mathcal{U}(-0.02, 0.02)$
Rotation Noise - Roll (rad)	1.0	$\sim \mathcal{U}(-0.05, 0.05)$
Rotation Noise - Pitch (rad)	1.0	$\sim \mathcal{U}(-0.1, 0.1)$
Rotation Noise - Yaw (rad)	1.0	$\sim \mathcal{U}(-0.05, 0.05)$