

Ablation	$r_{v_{x,y}^{cmd}}$	$r_{\omega_z^{cmd}}$	$r_{c_f^{cmd}}$	$r_{c_v^{cmd}}$
Trotting	0.92 ± 0.01	0.70 ± 0.04	0.98 ± 0.00	0.95 ± 0.00
Pronking	0.85 ± 0.01	0.64 ± 0.05	0.98 ± 0.00	0.94 ± 0.00
Pacing	0.83 ± 0.02	0.66 ± 0.04	0.96 ± 0.01	0.94 ± 0.01
Bounding	0.88 ± 0.01	0.63 ± 0.05	0.96 ± 0.01	0.95 ± 0.00
Gait-free Baseline	0.94 ± 0.02	0.76 ± 0.01	—	—

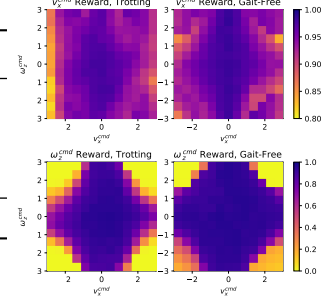


Table 5: Removing gait constraints results in improved velocity tracking task performance on flat ground. Heat maps (right) break down the mean task reward for each velocity command, revealing that the gait-free approach is most beneficial for combinations of high linear and angular velocity.

Term	Minimum	Maximum	Units
Payload Mass	-1.0	3.0	kg
Motor Strength	90	110	%
Joint Calibration	-0.02	0.02	rad
Ground Friction	0.40	1.00	—
Ground Restitution	0.00	1.00	—
Gravity Offset	-1.0	1.0	m/s ²
v_x^{cmd}	—	—	m/s
v_y^{cmd}	-0.6	0.6	m/s
ω_z^{cmd}	—	—	m/s
f^{cmd}	1.5	4.0	Hz
$\theta_1^{cmd}, \theta_2^{cmd}, \theta_3^{cmd}$	0.0	1.0	—
h_z^{cmd}	0.10	0.45	m
ϕ^{cmd}	-0.4	0.4	rad
s_y^{cmd}	0.05	0.45	m
$h_z^{f^{cmd}}$	0.03	0.25	m

Table 6: Randomization ranges for dynamics parameters (top) and commands (bottom) during training. $v_x^{cmd}, \omega_z^{cmd}$ are adapted according to a curriculum.

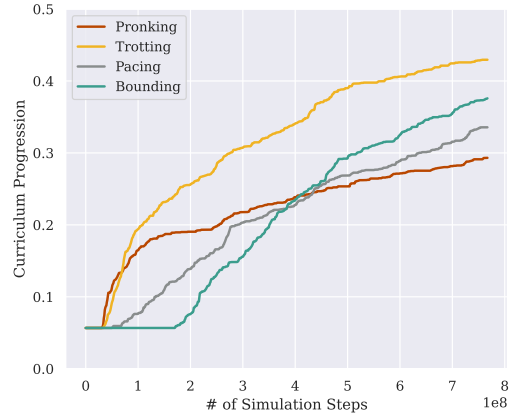


Figure 4: Pronking and trotting gaits are easier to learn and tend to dominate pacing and bounding early in training. However, when discovered, pacing and bounding gaits can yield good performance and later become preferred for some downstream tasks (Section 4.2).

A Training Details

The ranges used for domain and command randomization are provided in Table 6. The hyperparameters used for PPO are provided in Table 7. The hyperparameters used in the curriculum are provided in Table 8.

Figure 6 illustrates data flow during training. The curriculum engine first samples from a Gaussian distribution centered at one of the four main gaits (trotting, pronking, bounding, pacing); then it samples velocity commands from a grid distribution according to the method of [3]; then finally it samples the stepping frequency and body height uniformly. The policy and simulator are rolled out, and the reward is computed as a function of the gait parameters. Then the curriculum is updated if the episodic reward meets the thresholds given in Table 8.

B Teleoperation Interface

Figure 5 illustrates the mapping from remote control inputs to gait parameters used during teleoperation. The front bumpers on the top toggle between control modes to accommodate mixing our large number of gait parameters. Preprogrammed sequences such as dancing and leaping can be assigned to the rear bumpers.

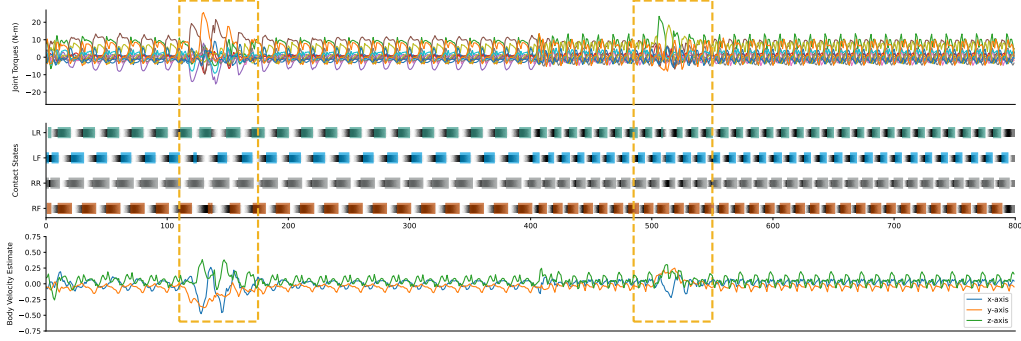


Figure 7: **Shove Robustness Test.** Joint torques (top), contact states (middle), and velocity estimate (bottom) during trotting in the laboratory setting. Dashed boxes indicate shove events. The robot was first shoved to the right during trotting at low frequency, and then shoved to the left during trotting at high frequency. The learned state estimator correctly infers the change in lateral velocity and adjusts the joint torques and foot contacts to stabilize the robot.

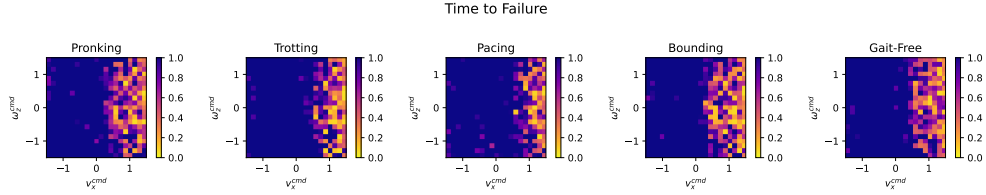


Figure 8: **Forward vs Backward Walking on Platforms.** Time to failure for different gaits and velocities in the random platforms environment (zero-shot test). The temperature bar unit is the mean fraction of a 20s episode elapsed before failure, between zero and one. The gait frequency is 3Hz. The pacing gait boasts the longest mean survival time, possibly due to higher footswings or more practice with recovery strategies during training. Interestingly, almost all failures occur during forward locomotion, and the robot is much more robust when moving backward. Possible explanations include (i) recovery strategies that are dependent on knee orientation and (ii) the weight distribution of the robot.

$$[t^{\text{FR}}, t^{\text{FL}}, t^{\text{RR}}, t^{\text{RL}}] = \text{clip}([t + \theta_2^{\text{cmd}} + \theta_3^{\text{cmd}}, t + \theta_1^{\text{cmd}} + \theta_3^{\text{cmd}}, t + \theta_1^{\text{cmd}}, t + \theta_2^{\text{cmd}}], 0, 1)$$

From these, we can directly compute the desired contact states:

$$C_{\text{foot}}^{\text{cmd}}(t^{\text{foot}}(\theta^{\text{cmd}}, t)) = \Phi(t^{\text{foot}}, \sigma) * (1 - \Phi(t^{\text{foot}} - 0.5, \sigma)) + \Phi(t^{\text{foot}} - 1, \sigma) * (1 - \Phi(t^{\text{foot}} - 1.5, \sigma))$$

where $\Phi(x; \sigma)$ is the cumulative density function of the normal distribution:

$$\Phi(x; \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x}{\sigma})^2}$$

which is an approximation to the Von Mises distribution used in [8] to form a smooth transition between stance and swing.

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	2.58B
optimizer	Adam

Table 7: PPO hyperparameters.

Parameter	Value	Units
$\mathbf{v}_x^{\text{cmd}}$ initial	[-1.0, 1.0]	m/s
ω_z^{cmd} initial	[-1.0, 1.0]	rad/s
$\mathbf{v}_x^{\text{cmd}}$ max	[-3.0, 3.0]	m/s
ω_z^{cmd} max	[-5.0, 5.0]	rad/s
$\mathbf{v}_x^{\text{cmd}}$ bin size	0.5	m/s
ω_z^{cmd} bin size	0.5	rad/s
$r_{v_x^{\text{cmd}}}$ threshold	0.8	–
$r_{\omega_z^{\text{cmd}}}$ threshold	0.7	–
$r_{c_f^{\text{cmd}}}$ threshold	0.95	–
$r_{c_v^{\text{cmd}}}$ threshold	0.95	–

Table 8: Curriculum parameters.

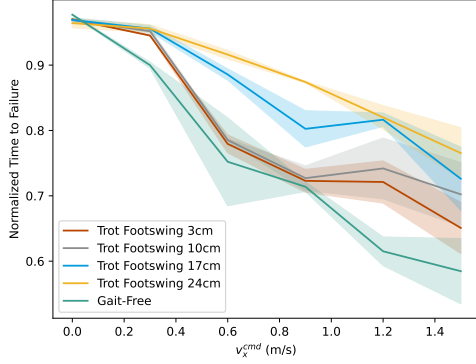


Figure 9: **Footswing Height vs Robustness:** Impact of footswing height on time to failure on the platform terrain (Section 4.2). Increased footswing height yields better generalization to uneven terrain from flat terrain compared to the gait-free policy.

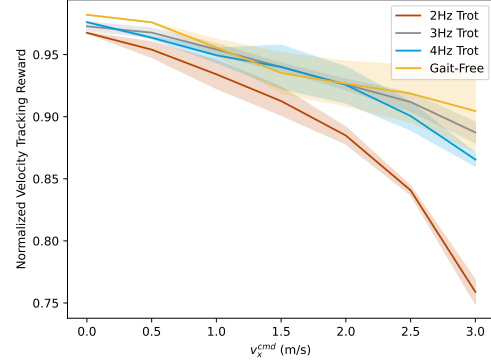


Figure 10: **Frequency vs Speed:** Impact of trotting frequency on flat-ground velocity tracking reward across speeds (Section 4.2). Enforcing low frequency (2Hz) makes high speeds less attainable. The gait-free policy offers slightly better performance at the lowest and highest speeds.

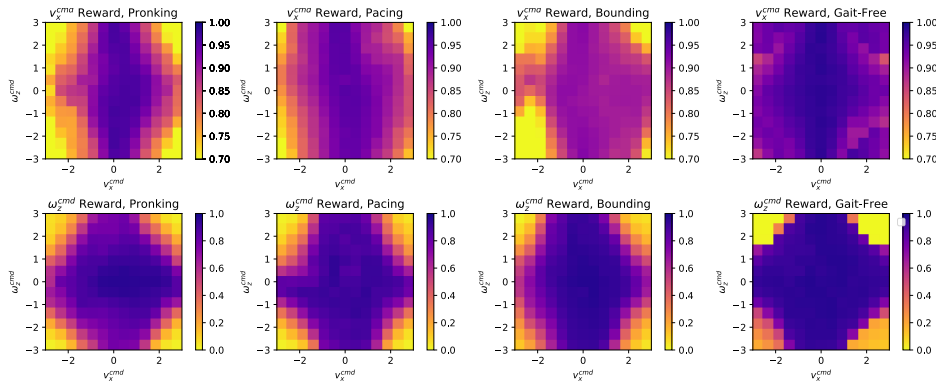


Figure 11: **Flat Ground Velocity Tracking Heatmaps:** We provide heatmaps as in Table 5 for the other major gaits: pronting, pacing, and bounding. In all cases, the policy forgoes performance on the training task of flat-ground velocity tracking to achieve diversity that will help accomplish new tasks. However, in-distribution task performance is maintained well for the lower range of speeds.