

1 Appendix

A Robot Control Table

Table 1: Comparison of Biped Locomotion Control Methods

Method	Model-Based	Adaptability	Complexity	Robustness to Disturbances
Zero Moment Point (ZMP)	Yes	Low	Low	Moderate
Model Predictive Control (MPC)	Yes	Moderate	High	High
Whole-Body Control (WBC)	Yes	Moderate	High	High
Hybrid Zero Dynamics (HZD)	Yes	Moderate	Moderate	High
RL	No	High	Very High	Very High

B Training Parameters

Table 2: Capsules and meshes Franka training parameters

Parameter	Value
num_timesteps	1_000_000
episode_length	2000
action_repeat	1
num_envs	8
num_evals	5
max_devices_per_host	null
num_eval_envs	1
learning_rate	0.0001
entropy_cost	0
discounting	0.98
seed	42
unroll_length	32
batch_size	2
num_minibatches	4
num_updates_per_batch	5
num_resets_per_eval	1
normalize_observations	True
reward_scaling	0.1
clipping_epsilon	0.2
gae_lambda	0.95
deterministic_eval	False
normalize_advantage	True
eval_env	null
randomization_fn	null
max_grad_norm	0.1
restore_checkpoint_path	null

Table 3: *Regularization validation Franka training parameters.*

Parameter	Value
num_timesteps	100_000_000
episode_length	2000
action_repeat	1
num_envs	8
num_evals	15
max_devices_per_host	null
num_eval_envs	128
learning_rate	0.0001
entropy_cost	0
discounting	0.98
seed	42
unroll_length	32
batch_size	512
num_minibatches	16
num_updates_per_batch	5
num_resets_per_eval	1
normalize_observations	True
reward_scaling	0.1
clipping_epsilon	0.2
gae_lambda	0.95
deterministic_eval	False
normalize_advantage	True
eval_env	null
randomization_fn	null
max_grad_norm	0.1
restore_checkpoint_path	null

Table 4: *Talos fixed base parameters.*

Parameter	Value
num_timesteps	100_000_000
episode_length	1000
action_repeat	1
num_envs	2048
num_evals	1
max_devices_per_host	null
num_eval_envs	128
learning_rate	0.0001
entropy_cost	0
discounting	0.98
seed	0
unroll_length	32
batch_size	512
num_minibatches	4
num_updates_per_batch	5
num_resets_per_eval	1
normalize_observations	True
reward_scaling	0.1
clipping_epsilon	0.2
gae_lambda	0.95
deterministic_eval	False
normalize_advantage	True
eval_env	null
randomization_fn	null
max_grad_norm	0.1
restore_checkpoint_path	null

Table 5: Kangaroo training parameters.

Parameter	Value
num_timesteps	100_000_000
episode_length	1500
action_repeat	1
num_envs	2048
num_evals	8
max_devices_per_host	null
learning_rate	0.0003
entropy_cost	0.02
discounting	0.98
seed	0
unroll_length	32
batch_size	512
num_minibatches	4
num_updates_per_batch	5
num_resets_per_eval	1
normalize_observations	True
reward_scaling	0.1
clipping_epsilon	0.2
gae_lambda	0.95
deterministic_eval	False
normalize_advantage	True
eval_env	null
randomization_fn	null
max_grad_norm	0.1
restore_checkpoint_path	null

C Training Graphs

C.1 Exploding Gradient

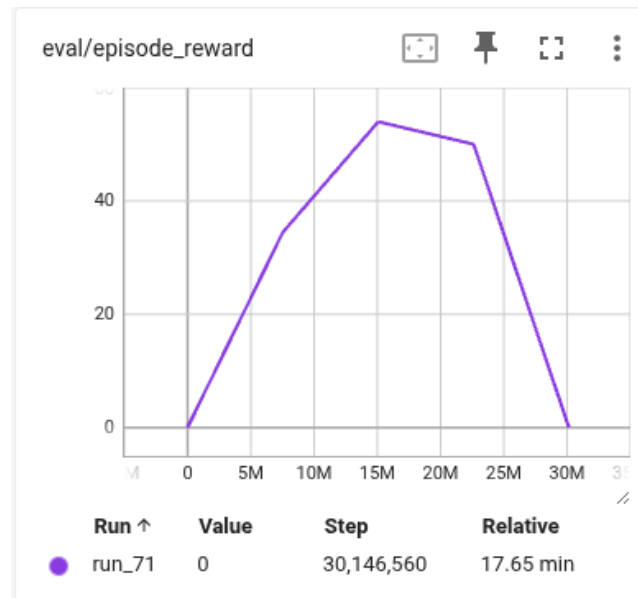


Figure 1: Exploding gradient training reward.

C.2 Talos

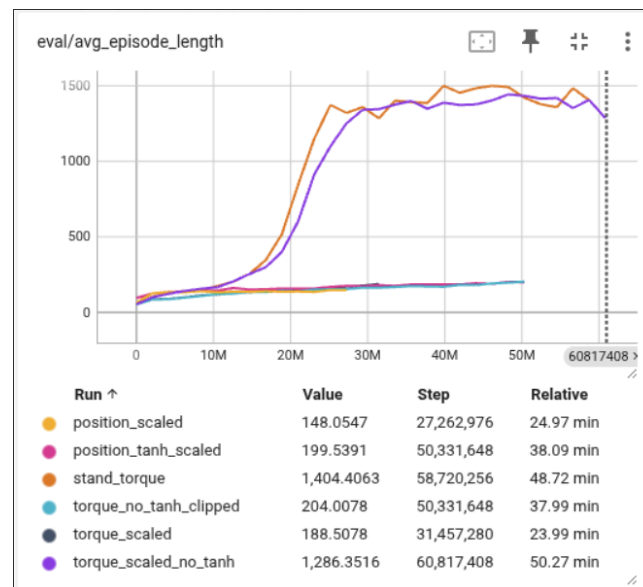


Figure 2: Talos learning episode length

C.3 Kangaroo

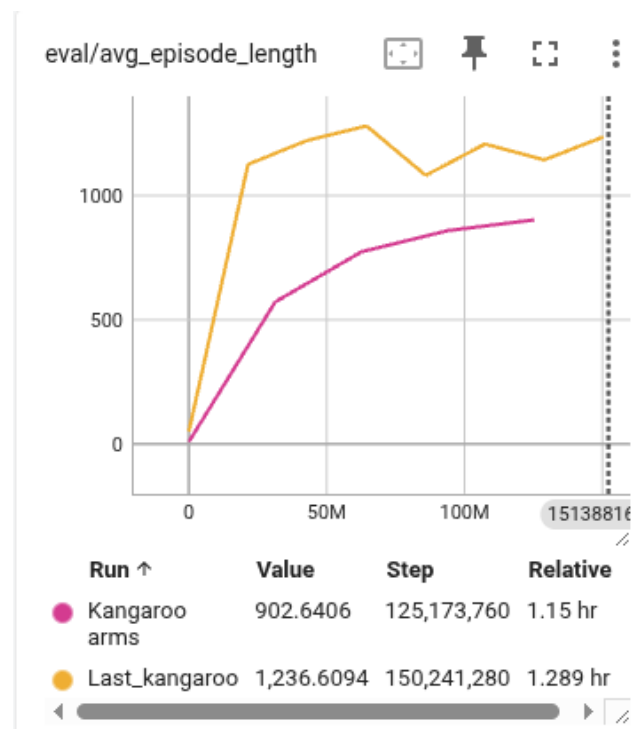


Figure 3: Kangaroo learning episode length

D Network Architecture

