Learning-Based Locomotion and Model-Based Jumping for Quadrupedal Robots

Xinqi Li† Rui Huang† Tianyang Tang† Peng Lu*
Adaptive Robotic Controls Lab, Department of Mechanical Engineering, The University of Hong Kong
†Equal contribution *Corresponding author: lupeng@hku.hk

1. MOTIVATION AND CONTROLLER OVERVIEW

MOTIVATION AND CONTRIBUTIONS

Construction inspection can require a mobile robot to traverse **uneven approaches, stairs, ramps, and restricted openings**. Continuous locomotion and a prescribed constrained jump place different demands on planning and feedback control.

This poster reports two independent ArcDog case studies:

- a history-based **PPO locomotion policy** trained in 4,096 randomized environments, with simulation and hardware sequences for forward motion, turning, and stair-to-ramp traversal;
- a published **TO–NMPC–WBC** controller and hardware experiment for jumping through a constrained opening.

EVIDENCE BOUNDARY

The controllers are **not integrated**; the jumping experiment assumes known geometry and a prescribed contact sequence.

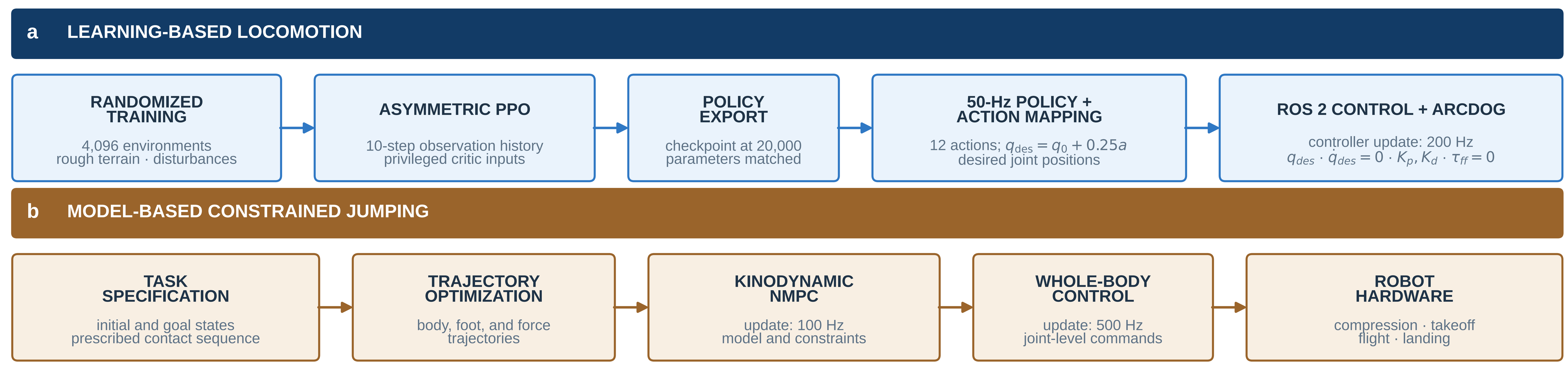


Fig. 1. Controller overview. (a) Randomized simulation and asymmetric PPO produce a velocity-commanded locomotion policy deployed through ROS 2. **(b)** The separate constrained-jumping controller combines trajectory optimization, 100-Hz kinodynamic NMPC, and 500-Hz whole-body control.

2. LEARNING-BASED LOCOMOTION

PPO TRAINING AND SIMULATED TERRAIN TRAVERSAL

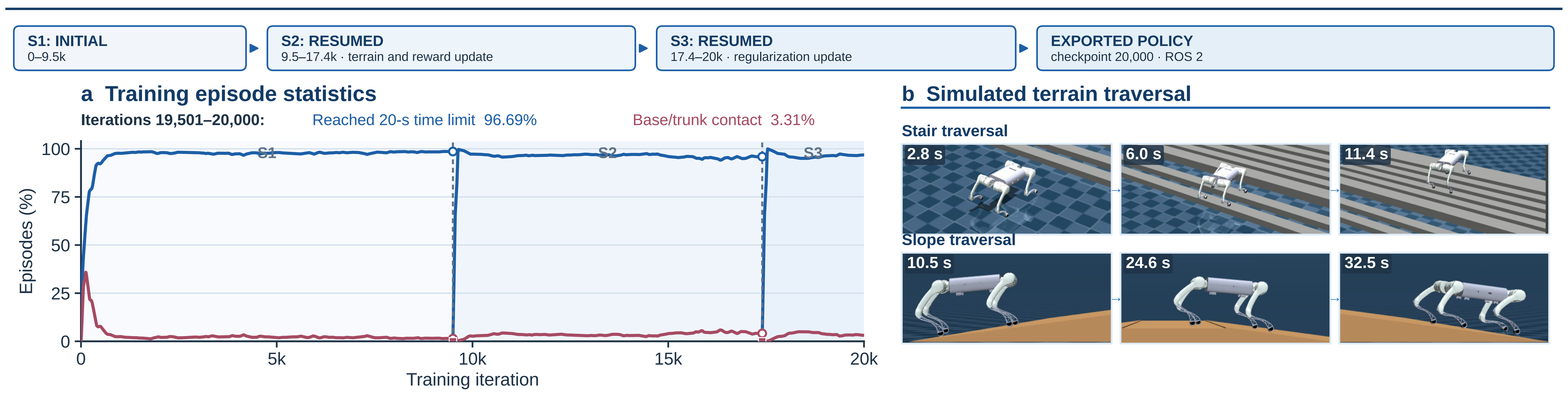


Fig. 2. Training and simulation. Curves are 100-iteration moving averages for one run continued through three stages. Training resumed with configuration updates at iterations 9,500 and 17,400. The sequences show stair and continuous-slope traversal.

HARDWARE EXPERIMENTS

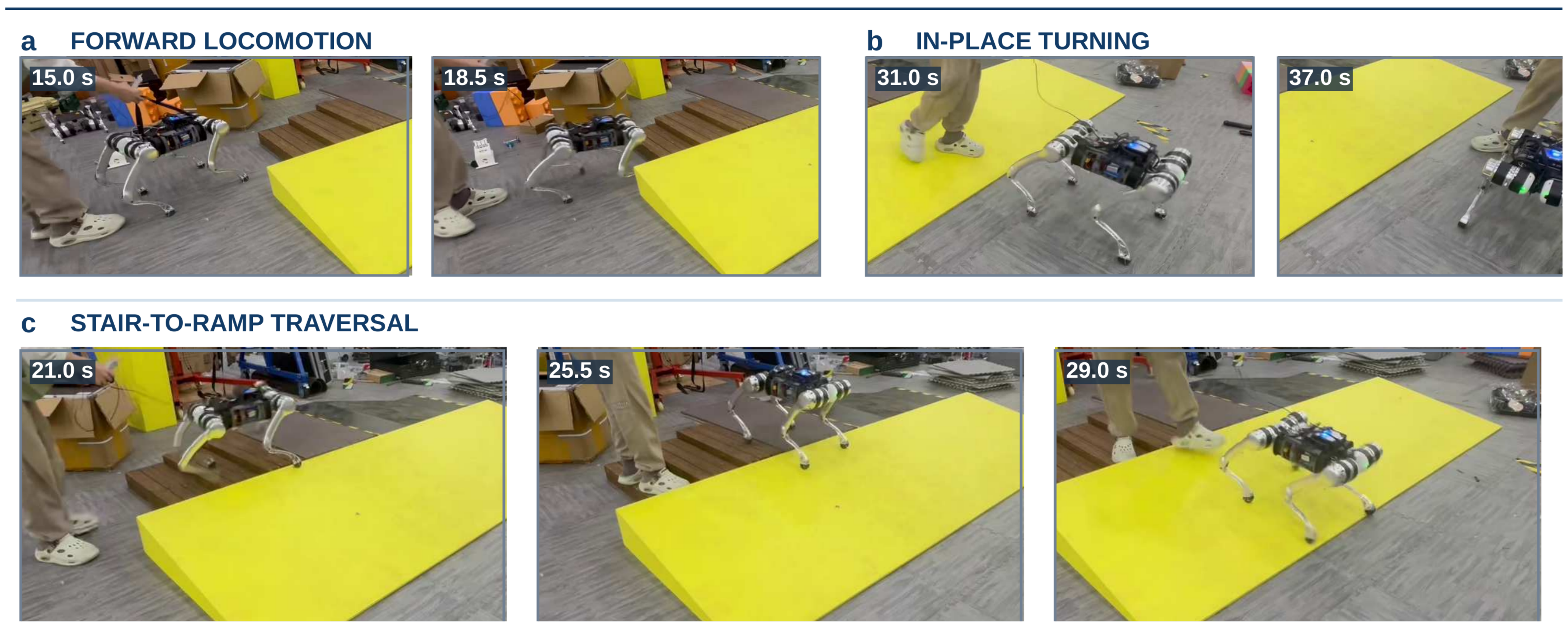


Fig. 3. Real-robot locomotion. The iteration-20,000 policy is shown in forward locomotion, in-place turning, and stair-to-ramp traversal. The ramp is approximately 3.0 m long and 1.0 m wide, with a 0.19 m rise; the brown portion is a low stair section.

POLICY AND TRAINING

The actor receives a **10-step history** of 45-dimensional observations: base angular velocity, projected gravity, velocity commands, 12 joint positions, 12 joint velocities, and the previous action. A 512–256–128 ELU network maps the 450-dimensional history to 12 actions. The critic uses a 150-dimensional current state that additionally includes base linear velocity and a 102-point terrain-height scan. Terrain heights, explicit contact state, and gait phase are excluded from the actor input.

Training uses **4,096 Isaac Lab environments**, 200-Hz physics, and a 50-Hz policy. Each update collects 98,304 transitions. A single run was continued to iteration 20,000 across the three stages shown in Fig. 2.

TERRAIN, COMMANDS, AND RANDOMIZATION

Terrain. Stairs, random grids, rough height fields, and slopes.

Commands. Resampled every 5–10 s over $v_x \in [-1.0, 1.0]$ m/s, $v_y \in [-0.8, 0.8]$ m/s, and $\omega_z \in [-1.0, 1.0]$ rad/s, with a planar-velocity curriculum.

Objective. Velocity and yaw tracking with gait, clearance, posture, contact, smoothing, joint, and power terms.

Surface. Friction coefficient **0.4–1.2**.	**Robot model.** Base mass **0.85–1.15**×; CoM offset ±0.02 **m/axis**.
Actuation. Stiffness and damping **0.8–1.2**×; joint-friction increment **0.05–0.20**.	**Disturbance.** Planar velocity perturbation ±0.5 **m/s every 5–10 s**.

TRAINING RESULTS

96.69%	3.31%	984.34
reached the 20-s episode limit	ended after base or trunk contact	mean steps out of 1,000

Statistics cover iterations 19,501–20,000. A full-length episode runs for 20 s without early termination. Terrain-boundary terminations were at most 0.0244% in stage S1 and absent in S2 and S3.

DEPLOYMENT AND HARDWARE

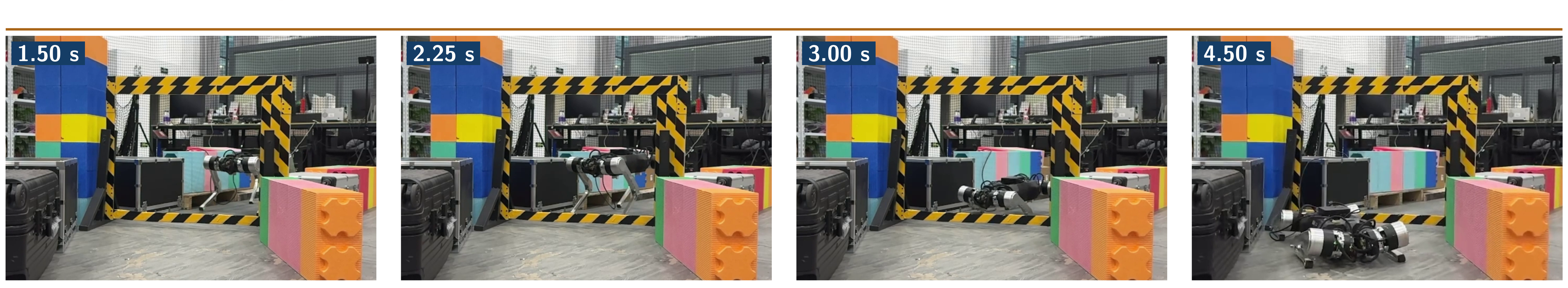
The final actor has **396,684 parameters** and runs at 50 Hz. Actions map to desired joint positions, $q_{des} = q_0 + 0.25a$. A 200-Hz ROS 2 controller sends the latest targets with fixed gains, zero desired velocity, and zero feedforward torque. The same exported actor was used in all three Fig. 3 sequences. These are **qualitative demonstrations**; repeated-trial success rates were not measured.

3. MODEL-BASED CONSTRAINED JUMPING

JUMPING THROUGH A CONSTRAINED OPENING

Planning visualization and hardware sequence. Optimized body and foot trajectories define the prescribed maneuver; the multi-pose image shows the corresponding passage through the frame.

REAL-ROBOT SEQUENCE



Constrained-opening jump. Selected frames show approach, entry, frame passage, and recovery. Task geometry and the contact sequence are prescribed.

1 TASK SPECIFICATION

Initial and goal states, obstacle constraints, and a prescribed contact sequence define the maneuver.

2 TRAJECTORY OPTIMIZATION

Centroidal dynamics produce body, foot, and contact-force trajectories subject to motion and contact constraints.

3 ONLINE FEEDBACK

100-Hz kinodynamic NMPC updates the motion; 500-Hz whole-body control generates joint-level commands.

CONTROLLER AND HARDWARE EXPERIMENT

The method follows Huang et al. Known initial and goal states, opening geometry, and a prescribed contact sequence define the task. Centroidal-dynamics trajectory optimization computes body, foot, and contact-force trajectories; a full-body mapper supplies base and joint references.

During execution, **100-Hz kinodynamic NMPC** updates motion and contact-force commands over a 0.6-s, 60-node horizon, while **500-Hz whole-body control** generates joint-level commands. Joint states and leg kinematics remain in the online problem to enforce foot-motion and whole-body constraints.

WHAT THE HARDWARE RESULT SHOWS

The robot executes a known-geometry constrained jump with the prescribed contact sequence. This controller was developed independently of the PPO policy and does not share its runtime interface.

CONSTRUCTION RELEVANCE AND OPEN WORK

PPO locomotion addresses continuous travel over stairs and ramps; the jumping experiment addresses a restricted opening with known geometry. These case studies address access mobility, not the inspection task itself.

Before construction deployment, the system still requires:

- onboard perception of terrain and opening geometry;
- controller integration, switching conditions, and safe recovery;
- evaluation with inspection payloads and around workers;
- repeated trials under representative site conditions.

This poster does not report construction-site tests, perception-driven jumping, or integrated locomotion-to-jump operation.

SELECTED REFERENCES

[1] J. Tan et al., Sim-to-real: Learning agile locomotion for quadruped robots, *RSS*, 2018.
[2] N. Rudin et al., Learning to walk in minutes using massively parallel deep reinforcement learning, *CoRL*, 2022.
[3] R. Huang et al., An efficient optimal trajectory planning and control framework for quadruped robot jumping through constrained obstacles, *IEEE TIE*, 2026. doi: 10.1109/TIE.2026.3692851.