EPFL

Motion Priors Reimagined: Adapting Flat-Terrain Skills for Complex Quadruped Mobility

Group 01:
Rayan Gauderon
Yann Thomé
Mouhamad Rawas

- MICRO-507: LEGGED ROBOTS

11.11.2025

EPFL

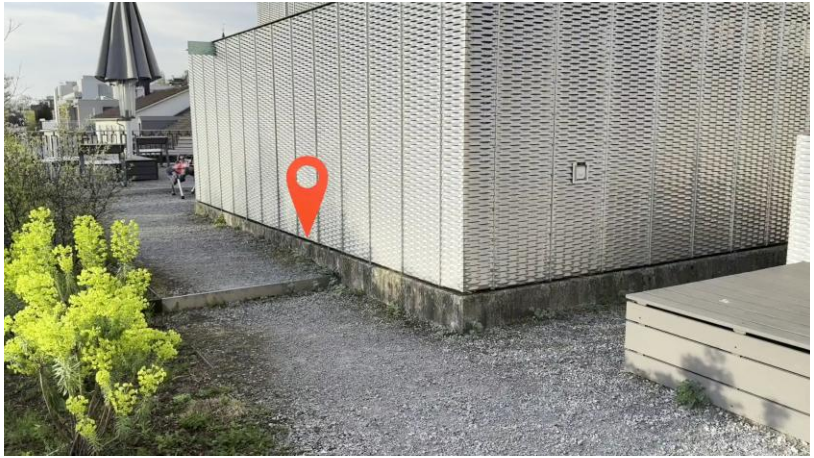
Introduction

Topic: Hierarchical RL for Animal-like Locomotion on Complex Terrain

Main Idea: Learn motion priors from animal data on flat ground, then add residual corrections to handle rough terrain.

Contribution: Simplifies reward engineering while achieving natural gaits and robust navigation across diverse environments.

Example Videos



Executive Summary

Robot's type:

- Quadruped (ANYmal-D platform)
- Dog-sized (approx. height and weight similar to a dog) with 12 actuated DoFs (3 per legs)

Control Methods:

- Position-based control with residual correction
- Low level: motion imitation policy (frozen after pre-trained)
- High level: task policy with latent commands + joint residuals

Design Methods:

- Reinforcement learning (PPO) with motion retargeting from animal mocap data
- Two-stage training: flat terrain → complex terrain
- Teacher-student distillation for sim-to-real

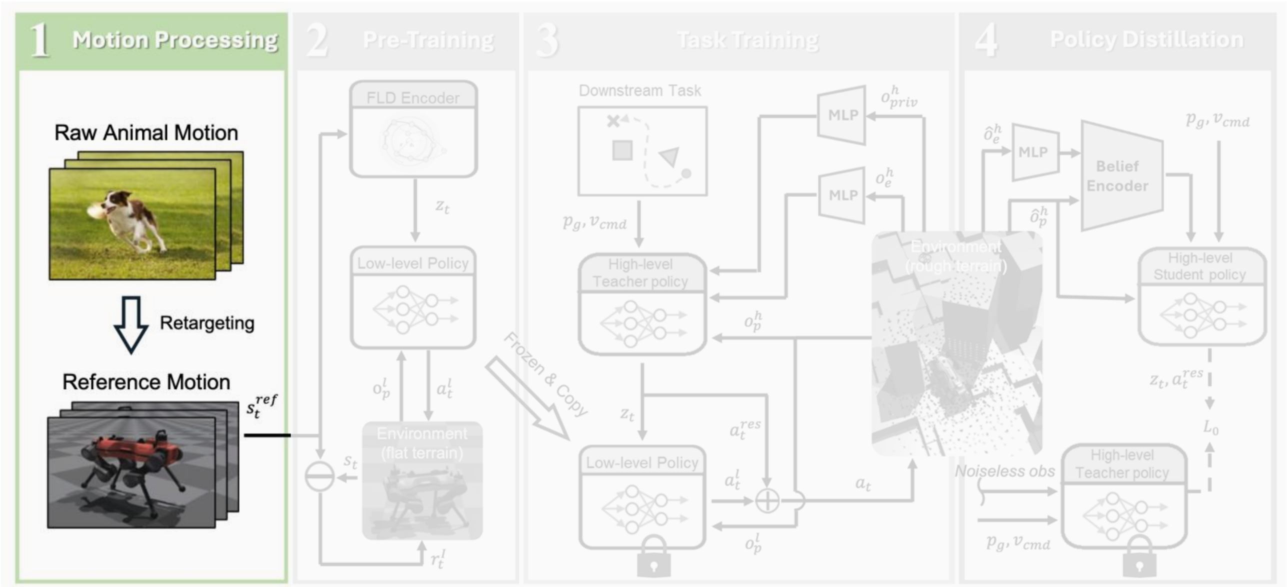
Gait types:

- Walking, Pacing, Cantering (from animal motion data)

Sensors:

- IMU, joint encoders (proprioceptive) and LiDAR, elevation scans (exteroceptive)

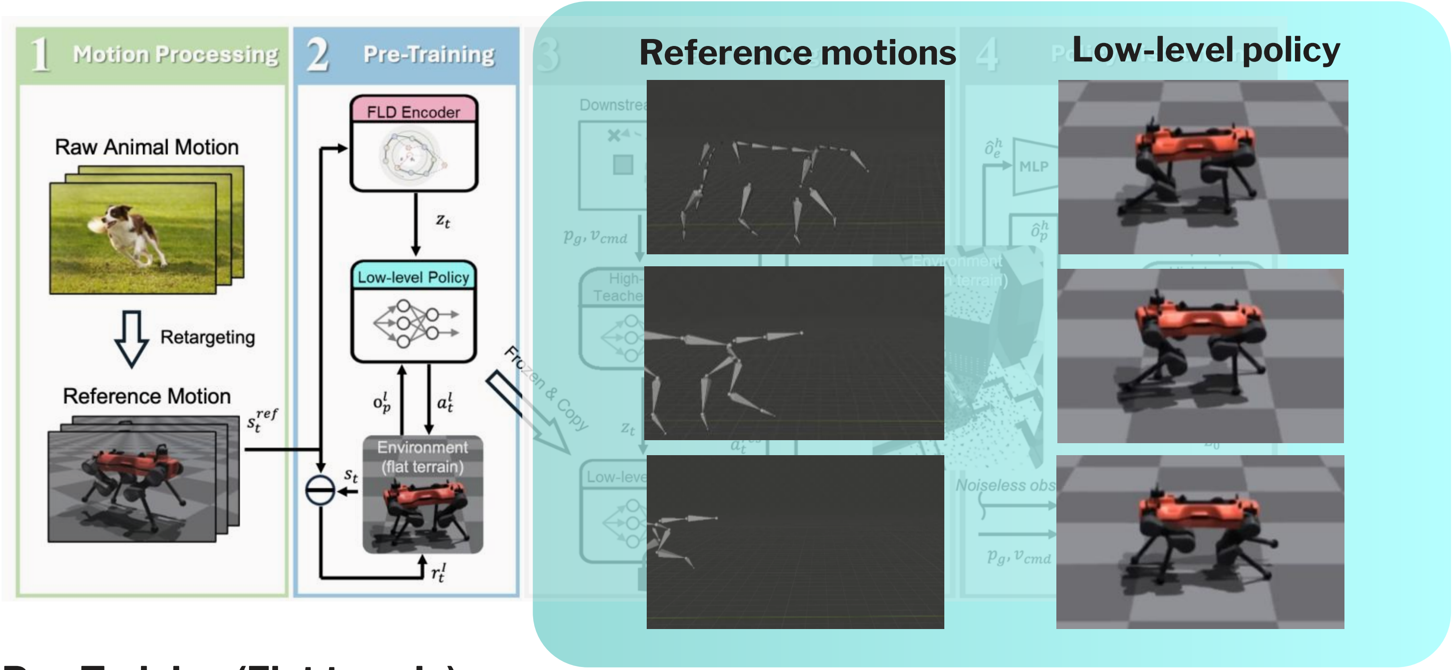
Four-Stage Training Framework



Stage 1: Motion Processing

- Retarget animal motion capture to robot kinematics
- Source: walking, pacing, cantering on flat ground only at various speed

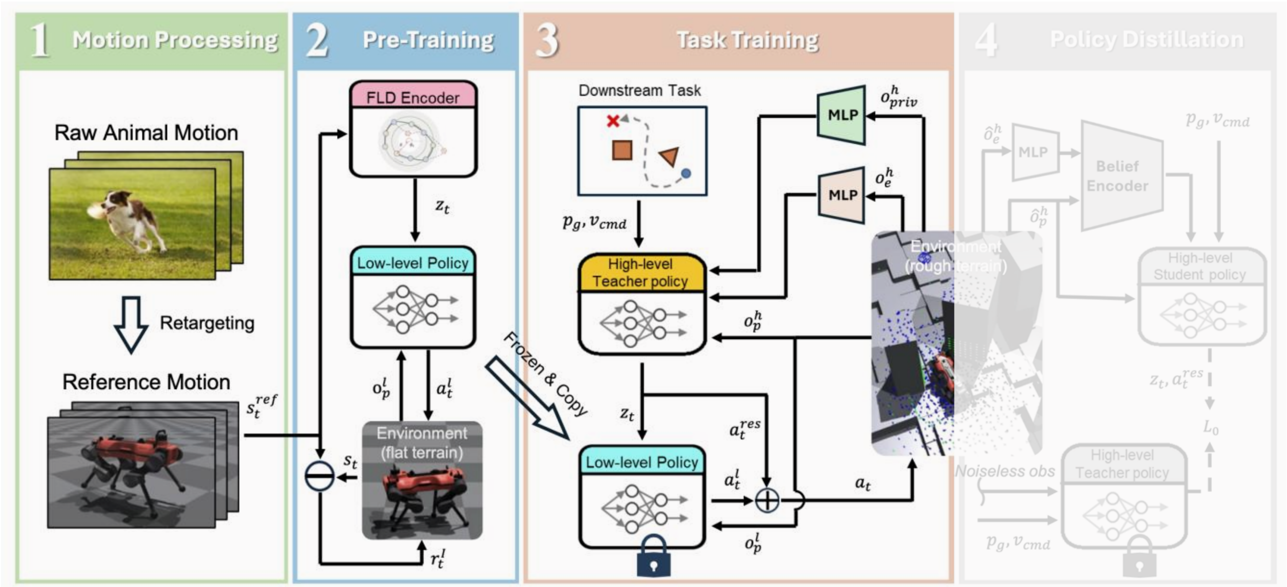
Four-Stage Training Framework



Stage 2: Pre-Training (Flat terrain)

- Train FLD encoder: compress motions into latent commands
- Train low-level policy: follow latent commands
- Reward: imitation & regularization
- Policy FROZEN after this stage ❄️

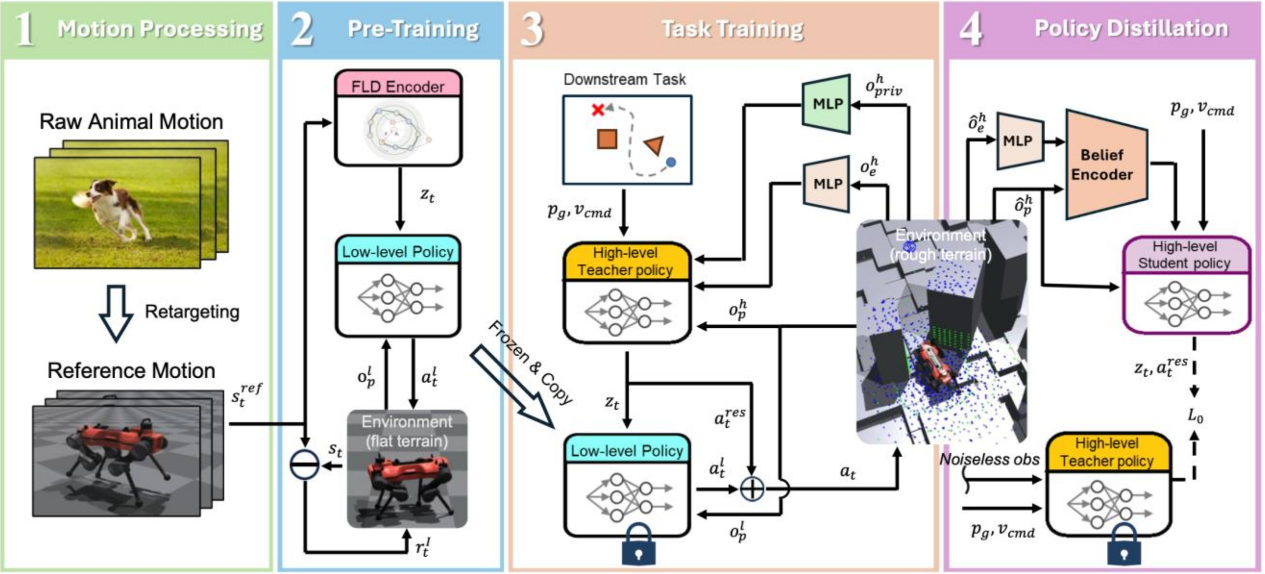
Four-Stage Training Framework



Stage 3: Task Training (Rough terrain)

- Learn goal navigation & obstacle avoidance while preserving natural gait
- High-level policy outputs:
  - 16D latent commands (which motion to perform)
  - 12D joint residuals (terrain corrections)

Four-Stage Training Framework



Stage 4: Policy Distillation

- Teacher: privileged information (perfect sensing)
- Student: realistic noisy sensors + belief encoder

Results

Low-Level Policy Performance

- Excellent tracking of reference joint trajectories
- Smooth gait transitions & maintains periodic structure

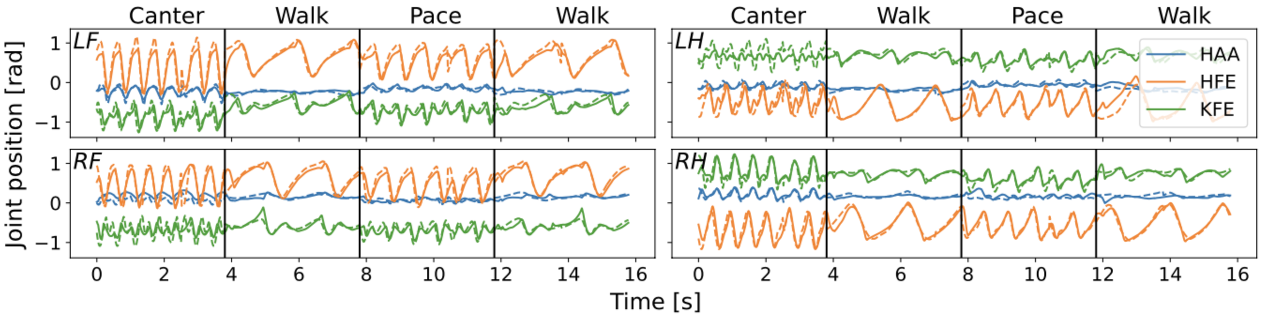


Figure 3: Actual vs. reference joint positions for the low-level policy on flat ground. Dashed lines are reference trajectories; solid lines are the actual trajectories. The plot (canter → walk → pace → walk) shows close alignment between reference and actual motions.

Results

Low-Level Policy Performance

- Excellent tracking of reference joint trajectories
- Smooth gait transitions & maintains periodic structure

High-Level Policy Performance

- Success rates: 75-95% across terrain types

Table 8: Success Rate of High-Level Student Policy in Simulation

Terrain Type	**Success Rate**
0.25m Stairs (Up)	84/100
0.25m Stairs (Down)	85/100
24-deg Slope (Up)	90/100
24-deg Slope (Down)	95/100
Random Boxes	81/100
Random Boxes with High Obstacles	75/100
Flat ground with High Obstacles	90/100

Results

Low-Level Policy Performance

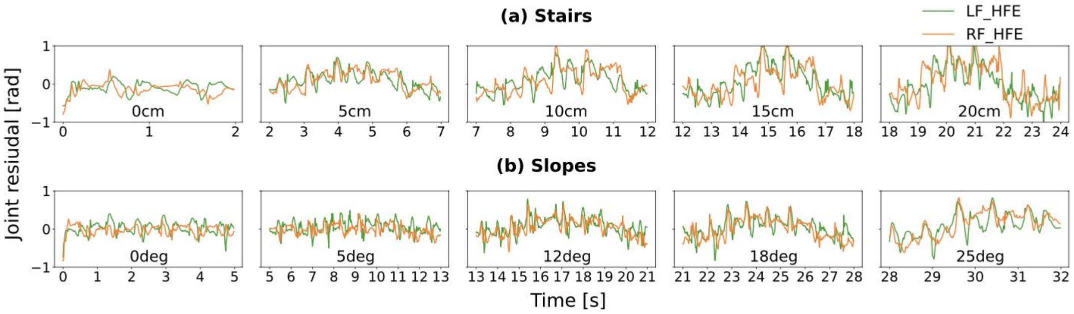
- Excellent tracking of reference joint trajectories
- Smooth gait transitions & maintains periodic structure

High-Level Policy Performance

- Success rates: 75-95% across terrain types

Residual Adaptation

- Modify low-level actions for terrain adaptation
- Residual variance grows with terrain difficulty



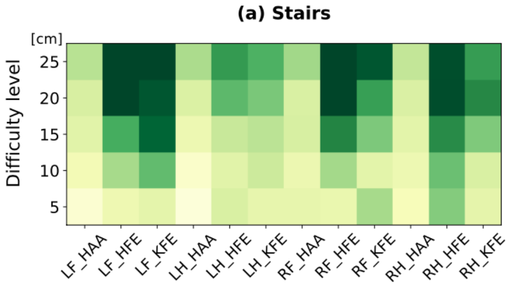


Figure 4: Comparison of high-level joint residuals for the HFE joint on the left front (LF) and right front (RF) leg across two terrain types: (a) pyramid stairs and (b) pyramid slopes with a rugged surface (see Fig. 11). Each terrain type is divided into five difficulty levels (displayed at the bottom of each subplot), with difficulty increasing from left (easy) to right (hard).

Results

Mouhamad Rawas

Low-Level Policy Performance

- Excellent tracking of reference joint trajectories
- Smooth gait transitions & maintains periodic structure

High-Level Policy Performance

- Success rates: 75-95% across terrain types

Residual Adaptation

- Modify low-level actions for terrain adaptation
- Residual variance grows with terrain difficulty

Residual Penalty

- Trade-off between adaptation and style preservation (best performance with balanced weight)

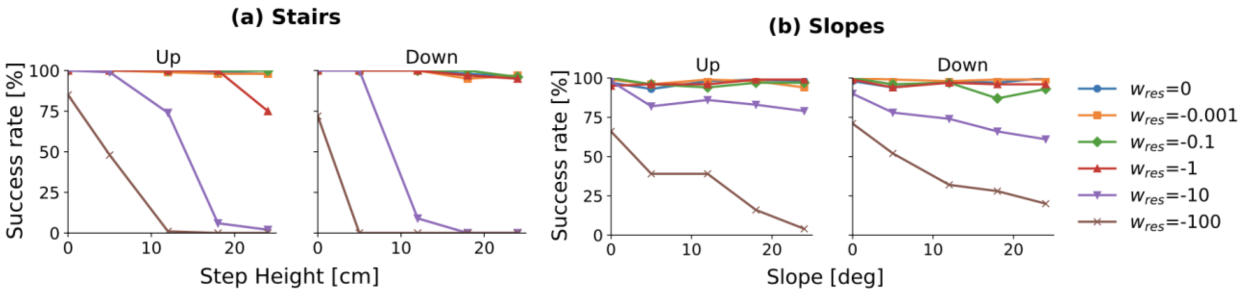


Figure 6: Goal-reaching success rate on stairs and sloped terrains across different difficulty levels, with the left subplots representing ascent and the right ones representing descent. Each terrain type is evaluated over 100 trials, with randomized initial robot poses in each experiment.

Results

Low-Level Policy Performance

- Excellent tracking of reference joint trajectories
- Smooth gait transitions & maintains periodic structure

High-Level Policy Performance

- Success rates: 75-95% across terrain types

Residual Adaptation

- Modify low-level actions for terrain adaptation
- Residual variance grows with terrain difficulty

Residual Penalty

- Trade-off between adaptation and style preservation (best performance with balanced weight)

Real World Deployement

- Robot autonomously climbs stairs and avoids obstacles
- Maintains animal-like gait under perception noise

Article Impact

Publication status

- ArXiv preprint (August 2025), presented at CoRL 2025 in Korea
- 1 citation: Survey by Li, Hutter & Krause (2025) on learning from demonstrations

Discussions surrounding the article

- Adresses a fundamental limitation of feature-based imitation method: adaptability
- Part of emerging trend toward «structured motion representations» bridging feature-based and GAN-based approaches

Pros and Cons

Mouhamad Rawas

Pros

- Simplified reward engineering
- Natural animal-like gaits and smooth locomotion
- Unified locomotion & navigation for rough terrains

Cons

- Mode collapse, relying on a single low-level gait
- Limited terrain coverage (no gaps, stepping stones or overhanging obstacles)
- Limited to demonstrated motion style

Conclusion

Core Achievement: Hierarchical RL framework enabling natural, animal-like locomotion on complex terrains using motion data collected only on flat ground.

Future Directions: Extend to more complex motor skills (jumping, manipulation) and scale to even more challenging environments (gaps, overhangs)

Mouhamad Rawas

Questions & Feedback ?

Group 01:
Rayan Gauderon
Yann Thomé
Mouhamad Rawas

Appendix – RL Training Details

Fourier Latent Dynamic (FLD) Model Training

- Autoencoder-like architecture to model the latent dynamics, defined as:

$$\mathbf{z}_t = (\theta_t, \phi_t) = \mathbf{enc}(\mathbf{s}_t), \quad \hat{\mathbf{z}}_{t+i} = (\theta_t, \phi_t + i f_t \Delta t),$$

$$\hat{\mathbf{s}}_{t+i} = \mathbf{dec}(\hat{\mathbf{z}}_{t+i}), \quad L^{FLD} = \sum^{N} \mathrm{MSE}(\mathbf{s}_{t+i}, \hat{\mathbf{s}}_{t+i}).$$

- State space: base's linear and angular velocities, gravity and joint positions

Table 5: Network Architecture for FLD

Network	Layer	Output Size	Activation
Encoder	Conv1d	64×31	ELU
	Conv1d	64×31	ELU
	Conv1d	4×31	ELU
Phase Encoder	Linear	4×2	Atan2
Decoder	Conv1d	64×31	ELU
	Conv1d	64×31	ELU
	Conv1d	27×31	ELU

Appendix – RL Training Details

Low-Level Policy

- Observation space: base linear & angular velocities, gravity, joint positions, latent encodings

Table 1: Reward Equations for Low-Level Policy

Name	Equation
Linear Velocity Tracking	$2\exp\left(-\left\lVert \mathbf{v}_t^b - \mathbf{v}_t^{b,ref} \right\rVert^2\right)$
Angular Velocity Tracking	$0.8\exp\left(-0.8\left\lVert \mathbf{w}_t^b - \mathbf{w}_t^{b,ref} \right\rVert^2\right)$
Joint Position Tracking	$1.4\exp\left(-2\sum_{i=1}^{12}(q_{t,i} - q_{t,i}^{ref})^2\right)$
Projected Gravity Tracking	$0.8\exp\left(-3\left\lVert \mathbf{g}_t - \mathbf{g}_t^{ref} \right\rVert^2\right)$
Action Rate	$-0.005\sum_{i=1}^{12}(a_{t,i} - a_{t-1,i})^2$
Collision	$-\sum_{k\in\text{thigh, shanks}} c_k$
Torque Limits	$-0.2\sum_{i=1}^{12}\max(\tau_{t,i} - \tau_{lim}, 0)$
Torques	$-0.00002\sum_{i=1}^{12}\tau_{t,i}^2$
Joint Acceleration	$-0.00007\sum_{i=1}^{12}\ddot{q}_{t,i}^2$
Feet Acceleration	$-0.0001\sum_{i=1}^{4}\left\lVert \mathbf{v}_{t,i}^f - \mathbf{v}_{t-1,i}^f \right\rVert^2$
Contact Forces	$-0.005\sum_{i=1}^{4} {F_{t,i}^f}^2$

Appendix – RL Training Details

High-Level Policy

- Observation space: low-level observations, leg contact, friction, external disturbances, elevation scans, LiDAR

Table 2: Reward Equations for High-Level Policy

Name	Equation
Position Tracking	$15 r_{reach}$
Heading Velocity	$5 r_{vel}$
Joint Residual	$-0.1 \sum_{i=1}^{12} (a_{t,i}^{res})^2$
Action Rate	$-0.005 \sum_{i=1}^{12} (a_{t,i}^{res} - a_{t-1,i}^{res} + a_{t,i} - a_{t-1,i})^2$
Collision	$-\sum_{k \in \text{thigh, shanks}} c_k$
Stand Still	$-2.5 \left\lVert \mathbf{v}_t^b \right\rVert^2 - \left\lVert \mathbf{w}_t^b \right\rVert^2$
Stand Pose	$-0.2 \sum_{i=1}^{12} (q_{t,i} - q_i^*)^2 - 5(g_{x,t}^2 + g_{y,t}^2)$
Torque Limits	$-0.2 \sum_{i=1}^{12} \max(\tau_{t,i} - \tau_{lim}, 0)$
Termination	-200
Torques	$-0.00002 \sum_{i=1}^{12} \tau_{t,i}^2$
Joint Acceleration	$-0.00007 \sum_{i=1}^{12} \ddot{q}_{t,i}^2$
Feet Acceleration	$-0.0001 \sum_{i=1}^{4} \left\lVert \mathbf{v}_{t,i}^f - \mathbf{v}_{t-1,i}^f \right\rVert^2$
Contact Forces	$-0.005 \sum_{i=1}^{4} {F_{t,i}^f}^2$

$$r_{reach} = \frac{1}{T_r}\left(1 - \frac{\lVert \mathbf{d} \rVert_2}{2}\right) \quad \text{if } t > T - T_r \text{ and } \lVert \mathbf{d} \rVert_2 < 2; \text{else } 0,$$

$$r_{vel} = \min(v_{cmd}, \langle \mathbf{v}, \mathbf{d} \rangle) \quad \text{if } \lVert \mathbf{d} \rVert_2 > 0.15; \text{else } 0,$$

Appendix – RL Training Details

Training Environment

- Boxes, staires, rugged slopes and high obstacles

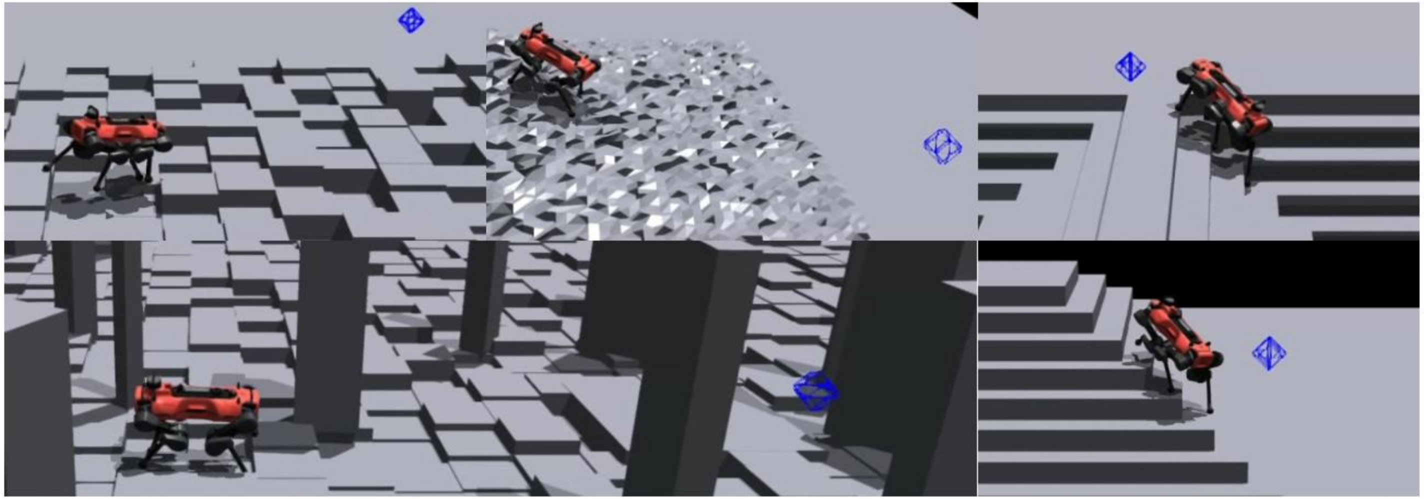


Figure 9: Overview of the training environment and terrain configuration for high-level policy. The setup includes diverse terrain types, boxes, stairs, rugged slopes, and high obstacles with the blue marker indicating the goal position.