LoComposition: Terrain-Adaptive Energy-Efficient Quadruped Locomotion without Gait Priors

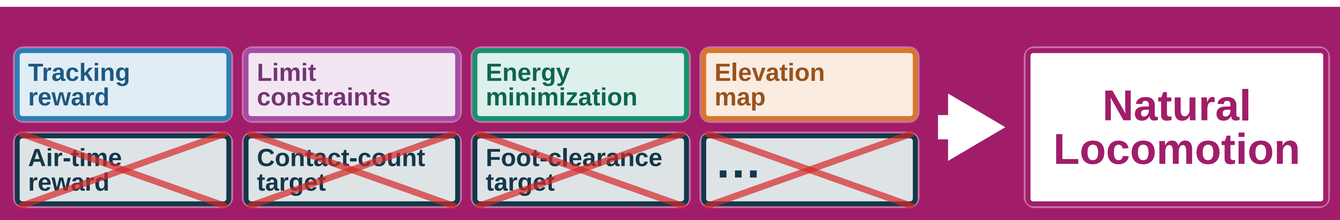
Loukas Kordos[1,2] Leonard T. Franz[3,2] Simon Rappenecker[2] Oliver Hausdörfer[1] Angela P. Schoellig[1] Pavel Kolev[2] Georg Martius[2,4]

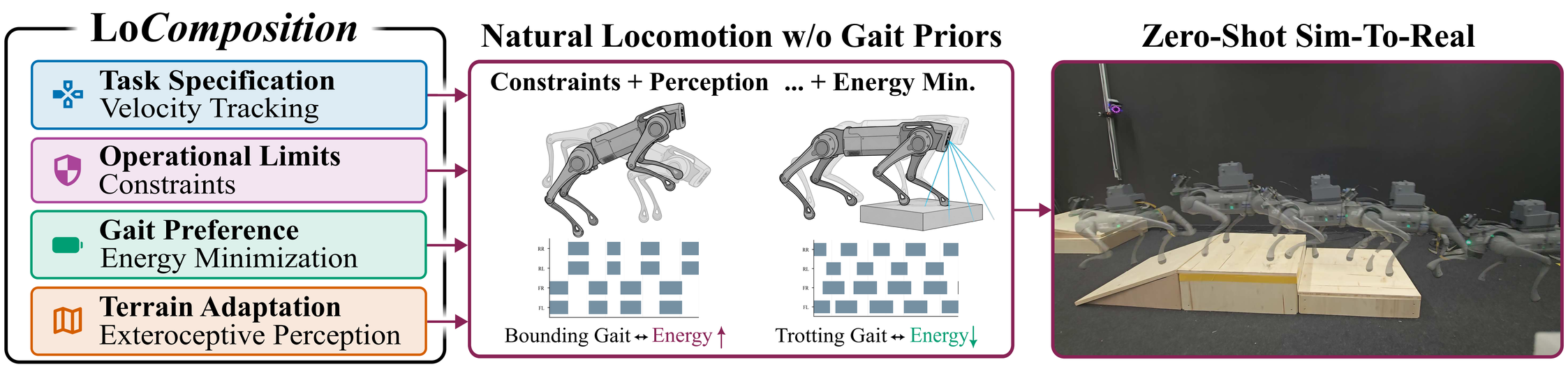
[1]Technical University of Munich · [2]University of Tübingen · [3]Hertie Institute for Clinical Brain Research & Center for Integrative Neuroscience · [4]MPI for Intelligent Systems, Tübingen



A minimal training recipe

Instead of over-specifying the gait in a complex reward function, we give each of our training components (Task, **L**imits, **E**nergy, **P**erception) a clear role. This is enough for an efficient, terrain-adaptive trot to emerge **without explicit gait priors**.





Problem

Learning-based locomotion relies on **complex reward formulations** that **entangle distinct functions** such as task specification, operational limits, gait preference, and terrain adaptation within a single objective. The entanglement **obscures** the **relationships** between the resulting **behavior** and **training components**, making style shaping necessary and robot transfer difficult.

One mechanism per function

LoComposition (LEP) separates them:

Task → Reward
velocity tracking

Operational Limits → CaT Constraints (L)
as in [2], for operational limits only, not for gait style

Gait preference → Energy (E)
mechanical-power penalty

Terrain → Perception (P)
body-centric elevation map

… and no explicit gait priors!

Four hypotheses

H1 Gait priors are unnecessary
*Explicit **gait priors** are **unnecessary** and can **harm** rough-terrain locomotion by biasing the learned gait.*

H2 Energy selects the gait
***Energy minimization** acts as a **gait-preference signal** by changing the motion selected among feasible behaviors, rather than merely improving the cost of transport.*

H3 Perception directs energy
***Terrain perception** makes this energy preference compatible with rough-terrain traversal by allowing the policy to **spend energy selectively** where the terrain requires it.*

H4 Limits need constraints
***Operational limits** are essential for real-robot deployment and **require termination-based constraints**. Simple reward penalty formulations are not sufficient.*

Formulation

Reward = velocity tracking − mechanical power

Constraints = Constraints as Terminations [2]:

$$G := \sum_t \left(\prod_{k=0}^{t} \gamma(s_k, a_k)\right) r_t$$

$$\gamma(s,a) := \gamma\,(1-\delta(s,a)), \quad \delta \sim \text{violation of } \mathcal{C}$$

$\mathcal{C} = \{$torque, joint velocity, joint acceleration, action rate, base orientation$\}$

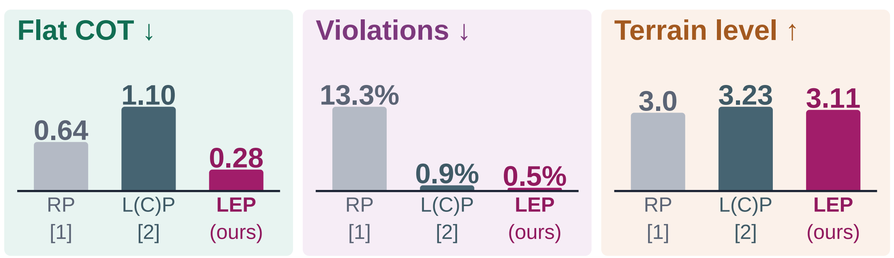
Minimality test (ablations)

Change one component, and something **breaks**:

Change	What breaks	Evidence
Add Gait-prior constraints (**H1**, LCEP)	**Efficiency, terrain level**	0.28 → 0.43 COT ↓; 3.11 → 2.65 Terrain ↑
Remove **E**nergy (**H2**, LP)	**Efficiency, deployability**	0.28 → 1.18 COT ↓; 4/5 → 0/5 HW ramp ↑
Remove **P**erception (**H3**, LE)	**Traversability**	3.11 → 1.4 Terrain level ↑
Remove **L**imits (**H4**, EP)	**Deployability**	0.5% → 53.4% Violations ↓
Simplify **L**imits to **P**enalties (**H4**, PEP)	**Limit compliance, efficiency**	0.5% → 1.9% Violations ↓; 0.28 → 0.34 COT ↓

Baseline comparison

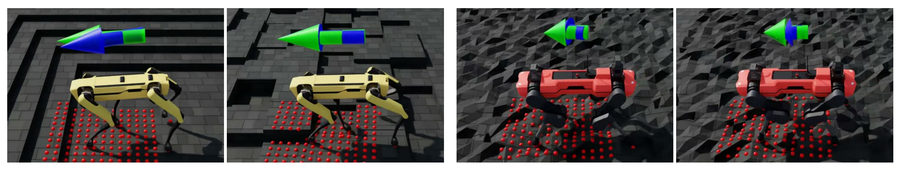
LEP against the complex reward RP [1] and CaT's full recipe L(C)P [2], in simulation:



Same recipe, other robots

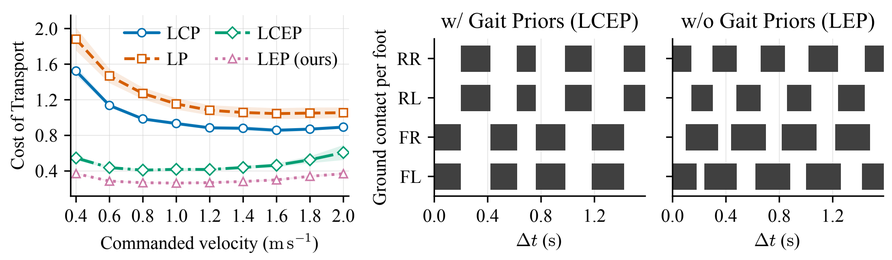
We **retrain LEP from scratch** on Spot and ANYmal C in simulation. Only actuator limits and a mass-scaled energy weight change: **no reward retuning**, no gait priors.

Spot COT 0.29 ANYmal C COT 0.33



Both learn an efficient trot.

Natural gait w/o gait priors H1 H2

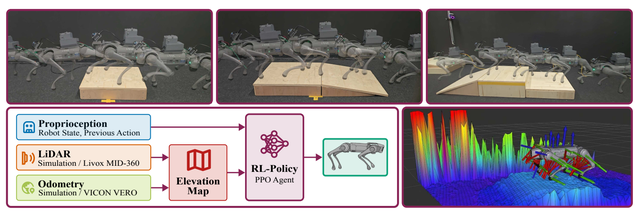


Without gait priors, energy minimization selects a trot with the lowest COT. LCEP has the same power penalty but stays bounding.

Take-home

- **No gait priors needed:** an efficient trot emerges.
- **Every component counts:** removing one exposes a distinct failure.
- **Zero-shot** to a real Go2, no retraining.
- **Same recipe, other robots:** retrained on Spot and ANYmal C in simulation, both trot efficiently without reward retuning.

Sim-to-real transfer



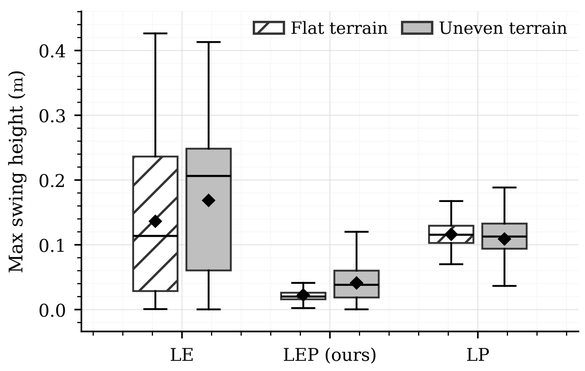
Zero-shot to Unitree Go2, using an online LiDAR elevation map matching the simulated policy observations.

	LP (no E)		LEP (ours)	
Scenario	Success	Viol. ↓	Success	Viol. ↓
Flat	5/5	0.7%	**5/5**	**0.53%**
Down ramp	0/5	14%	**4/5**	**1.2%**
Up ramp	3/5	15.7%	**5/5**	**0.9%**

Energy minimization improves hardware robustness, although LP reaches the higher terrain level in simulation. The likely reason is dynamism: LP is airborne 22–26% of the time; LEP always keeps a foot down.

Selective energy use H3

With the map, LEP raises swing height from 2.28 cm on flat to 5.72 cm on uneven terrain; LP steps high everywhere.



Contact
loukas.kordos@tum.de
leonard-tobias.franz
@uni-tuebingen.de

Video arXiv LinkedIn (Leonard Loukas)

Prof. Dr. Angela Schoellig Learning Systems & Robotics Lab · Eberhard Karls Universität Tübingen · Autonomous Learning · Hertie Institute for Clinical Brain Research · Max Planck Institute for Intelligent Systems · imprs-is

[1] Rudin et al., Learning to walk in minutes using massively parallel deep RL, CoRL 2022. [2] Chane-Sane et al., CaT: Constraints as terminations for legged locomotion RL, IROS 2024.

Kordos et al., arXiv:2606.15896v2. CoRL 2026. Figures from the paper (CC BY 4.0).
Support: EXC 2064/1 (390727645), ERC 101045454 REAL-RL, BMBF 01IS18039A, BMFTR 01GQ2508, IMPRS-IS.