Learning quadrupedal locomotion over challenging terrain

(Lee et al, 2020)

Legged Robots
Micro-502

Niklas Schröder
Julien Thévenoz

03.12.2024

EPFL

Learning quadrupedal locomotion over challenging terrain

EPFL

Learning quadrupedal locomotion over challenging terrain

- A controller for quadruped robots using a **reinforcement learning** policy driven **exclusively by proprioceptive sensors**

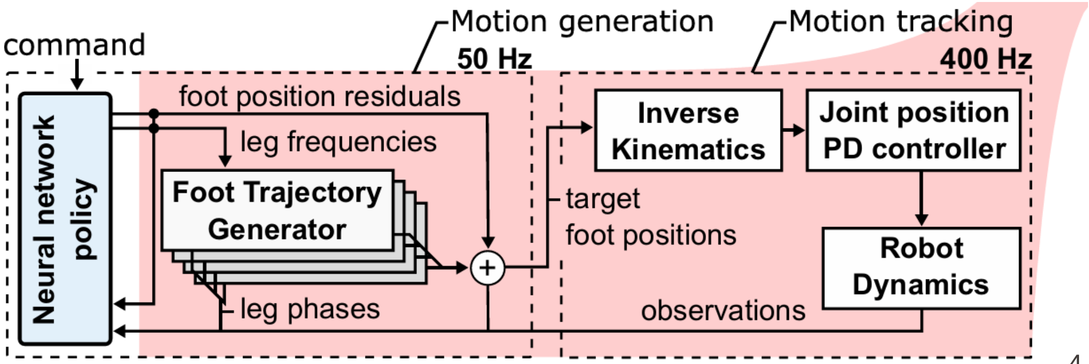
- The neural network policy is trained through Reinforcement Learning in simulation with a combination of two methods :
  - **Student – Teacher** learning
  - **Adaptative Curriculum**

Learning quadrupedal locomotion over challenging terrain

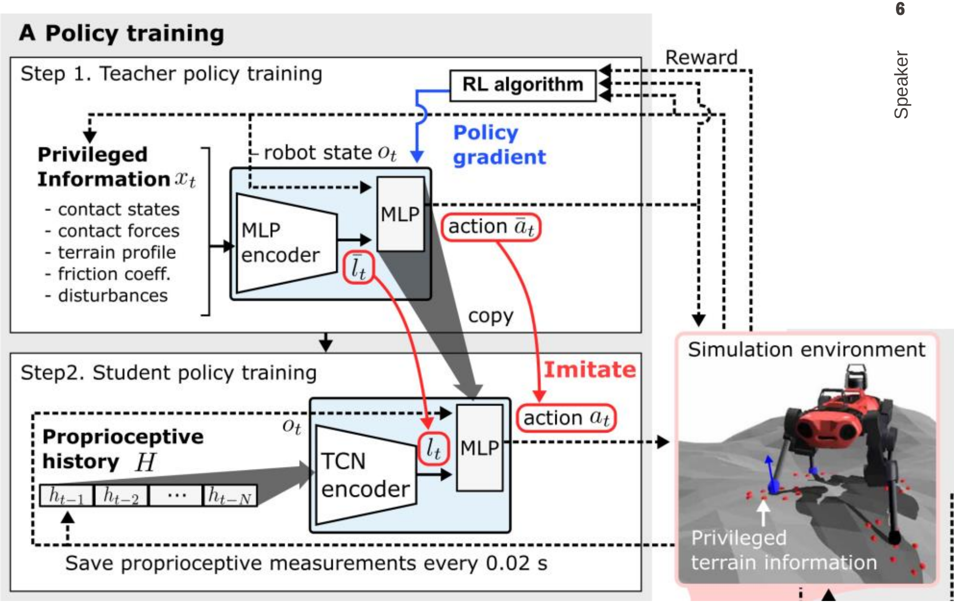
- Resulting control policy is **exceptionally robust**, capable of traversing terrains that no previous controller had ever achieved (mud, snow, thick vegetation, running water)
- **Zero-shot generalization** from simulation to natural environments
- Adapts to terrains on which it has never been trained in simulation

Executive summary

- Robot : **ANYmal** quadruped
- Gait : **Trot** only (but authors say other gaits could be developed)
- Design method : **RL** w/ teacher-student learning and adaptative curriculum
- Sensors : only **proprioceptive** ones (joint encoders and IMU)
- Steering/control input : Direction to move in and orientation of robot
- Control : **joint position PD**



Student - Teacher Method



Concept

Teacher:

- Receives privileged information that is not available in real world
- Able to learn a good policy fast
- Only works in simulation

Student:

- Recieves more limited, realistic information
- Tries to imitate teacher behaviour using limited information
- Works in simulation and real world

Perfect environment information

Limited environment information

$$\mathcal{L} := (\bar{a}_t(I_t)) - a_t(i_t)))^2 + \left(\bar{\ell}_t(I_t) - \ell_t(i_t)\right)^2$$

Teacher's action

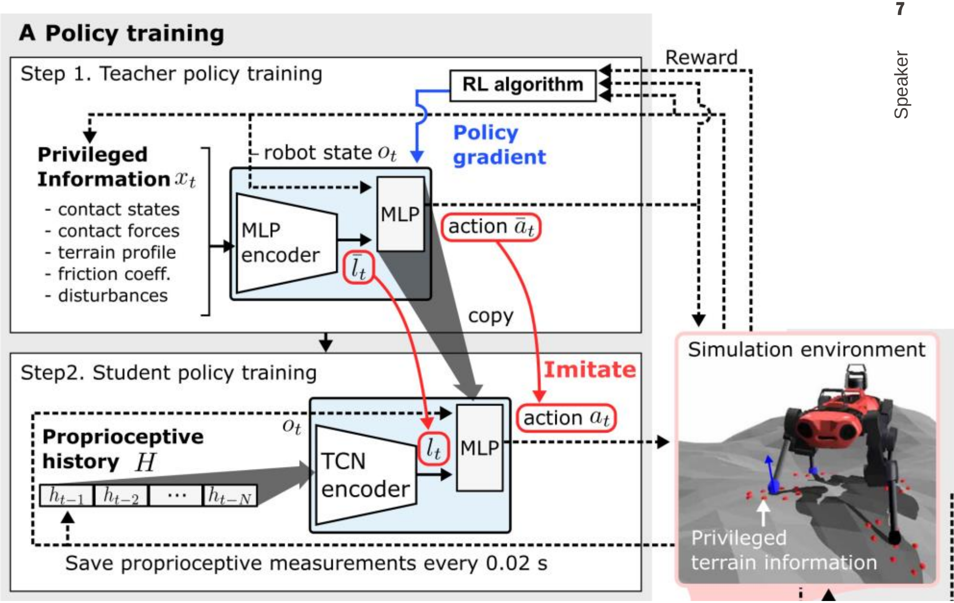
Student's action

Teacher's latent space representation

Student's latent space representation

Student - Teacher Method

Implementation



Inputs

- **Direction** to walk in
- **Priveleged info** (terrain shape, contact forces, friction, etc)

Teacher NN

Outputs

-Foot position residuals

-Foot frequencies

Teacher goal : move efficiently

- **Direction** to walk in
- **Proprioceptive history** of past 2 seconds (joint angles & speeds, base orientation & twist)

Student NN

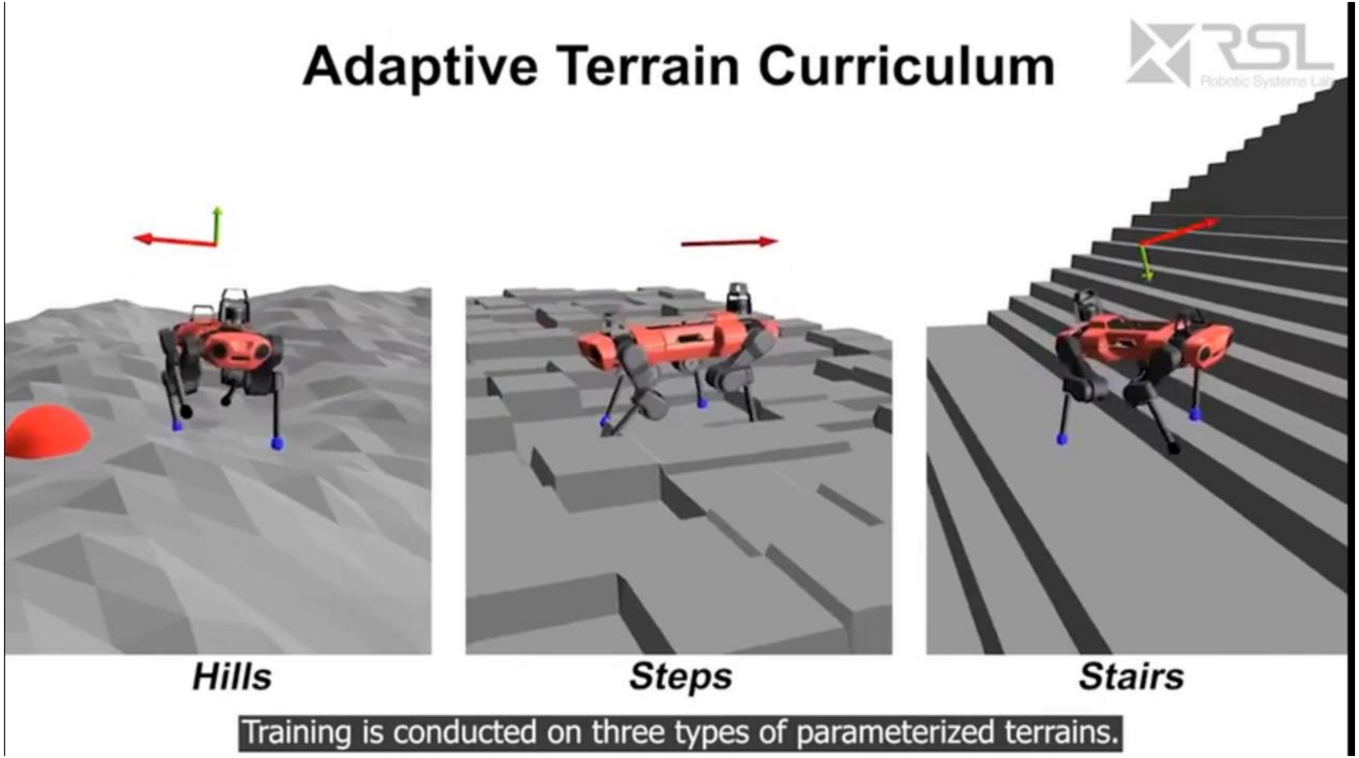
-Foot position residuals

-Foot frequencies

Student goal : Imitate teacher's output

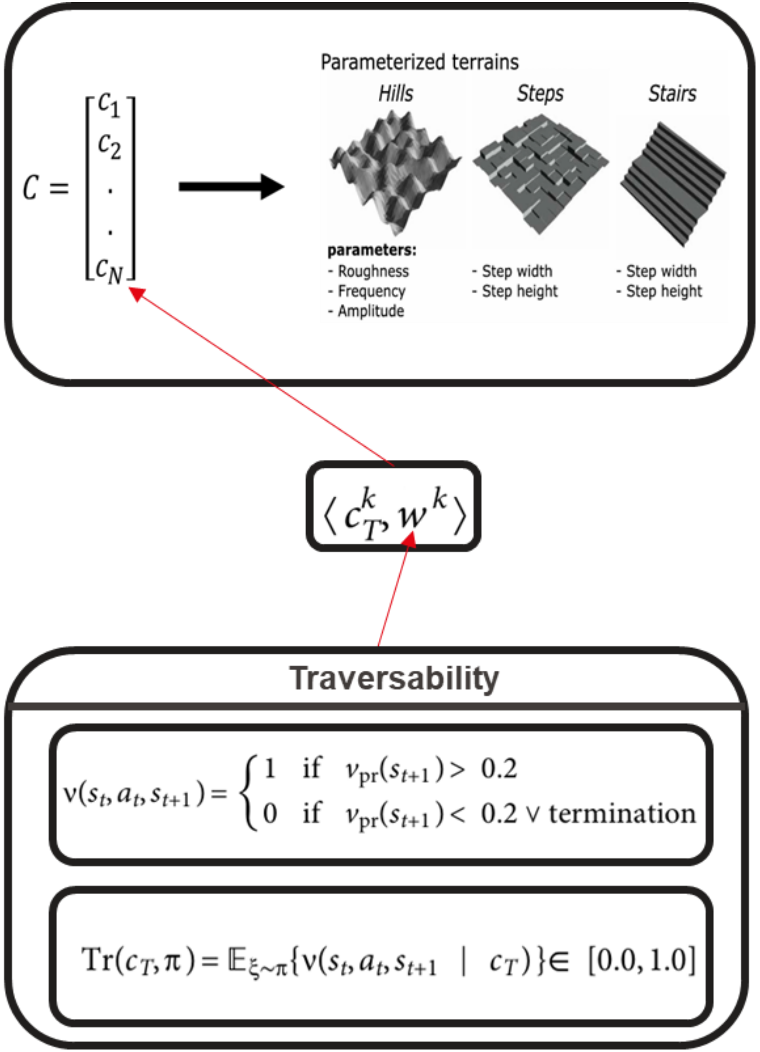
Speaker

Adaptive Terrain Curriculum



EPFL Adaptive Terrain Curriculum

- Environments should be neither to easy nor too challenging
- Approximate distribution of desirable environment values with particle filter
- Environment variables are evaluated by using traversability of generated terrains
- **Traversability:** Succes rate of traversing a terrain



Performance comparison vs state of the art baseline

Lee et al.’s controller is tested against a «traditional» controller :
it explicitely tries to detect slippage or contact and triggers predefined «reflex» motions accordingly.

Design process of a traditional controller

1. State estimator
2. Leg trajectories
3. Contact/Slip estimator
4. Reflex motions
5. Tune

Design process of a RL controller

1. Simulate
2. Train
3. Enjoy !

Baseline :
Dynamic locomotion on slippery ground
By F. Jenelten, J. Hwangbo, F. Tresoldi, C. D. Bellicoso, M. Hutter,
IEEE Robot. Autom. Lett. 4, 4170–4176 (2019).

EPFL

Performance comparison vs a state of the art baseline

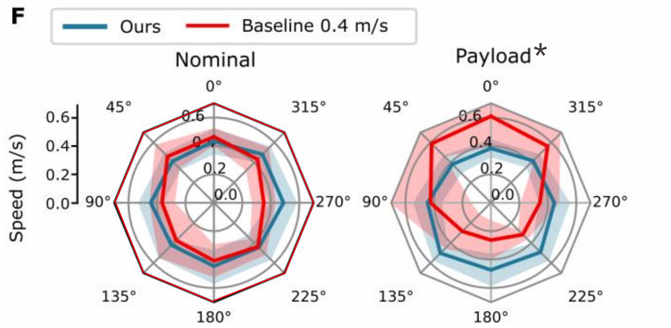
$$COT = \sum_{12\text{ actuators}} \frac{[\tau\dot{\theta}]^{+}}{mgv}$$

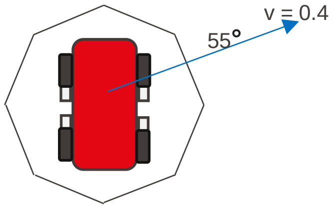
Table 1. Comparison of locomotion performance in natural environments. The mechanical COT is computed using positive mechanical power exerted by the actuators.

Quantity	Controller	Terrain		
		Moss	Mud	Vegetation
Average speed (m/s)	Ours	**0.452**	**0.338**	**0.248**
	Baseline	0.199	0.197	–
Average mechanical COT	Ours	**0.423**	**0.692**	**1.23**
	Baseline	0.625	0.931	–

The NN policy manages to walk over moss, mud, vegetation and snow with **100% success**, while the baseline regularly experienced catastrophic failure.
At the same time, the baseline was about two times slower, and 50% worse in terms of energy efficiency.

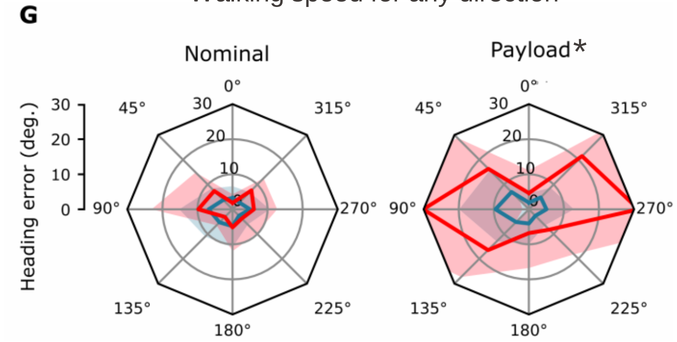
Performance comparison vs a state of the art baseline

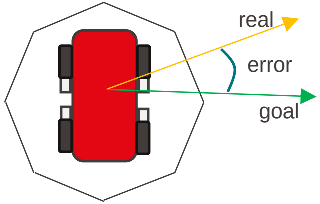




Walking speed for any direction

The NN policy manages to move at a constant given speed in any direction





Error for any direction, walking in straight line

The NN policy manages a smaller direction error, especially with a payload

*Payload : additional 10kg weight on the robot (not trained in simulation or included in controller model)

Citations

676 citations (top 1% paper according to SCOPUS)

Many of the citations are praise/acknowledgement but the paper has also inspired some researchers (mostly in robotics, self-driving cars and RL communities, but also in the medical field)

Some researchers have reused the idea of terrain curriculum : Kumar et al (”Enhancing Efficiency of Quadrupedal Locomotion Over Challenging Terrains with Extensible Feet”, 2023) or Rudin et al (“Learning to Walk in Minutes Using Massively Parallel Deep Reinforcement Learning”, 2021)

Other have been inspired by the teacher-student learning method, like Ao et al (“SafeRPlan: Safe deep reinforcement learning for intraoperative planning of pedicle screw placement”) or Chen et al (“Learning to drive from a world on rails“, 2021).

Pros and Cons

+ More robustness and all-terrain capabilities than any previous controller ?

+ “Simple” simulation but able to walk through complex environments

+ No need for “expert demonstration”

+ No sim-to-real gap

- No exteroception (which means there is a lot of potential if someone wants to add it !)

- Conservative speed because of blind locomotion

Exam Questions

What is the concept of an adaptive terrain curriculum?

Adapt the diffucilty of the terrain according to the current skill of the robot. The complexity of an environment is evaluated by the traversability, which is defined by the succes rate of traversing a terrain. To always keep track of good training terrains, the distribution of desirable environments is approximated using a particle filter.

Why can't the Teacher policy be used in the real world?

The Teacher policy relies on privileged observations as input, which are only available in simulations and not accessible in real-world scenarios. Consequently, only the Student policy can be utilized in the real world.