Learning quadrupedal locomotion over challenging terrain (2020)

Authors: Joonho LEE, Jemin HWANGBO, Lorenz WELLHAUSEN,
Vladlen KOLTUN, Marco HUTTER

Group 25: Christophe HUYNH, Adrian MOREL, Ryan TAN

EPFL

Legged Robots
MICRO-507

Video of the robot, source: https://www.youtube.com/watch?v=9j2a1oAHDL8

Type of robot and control

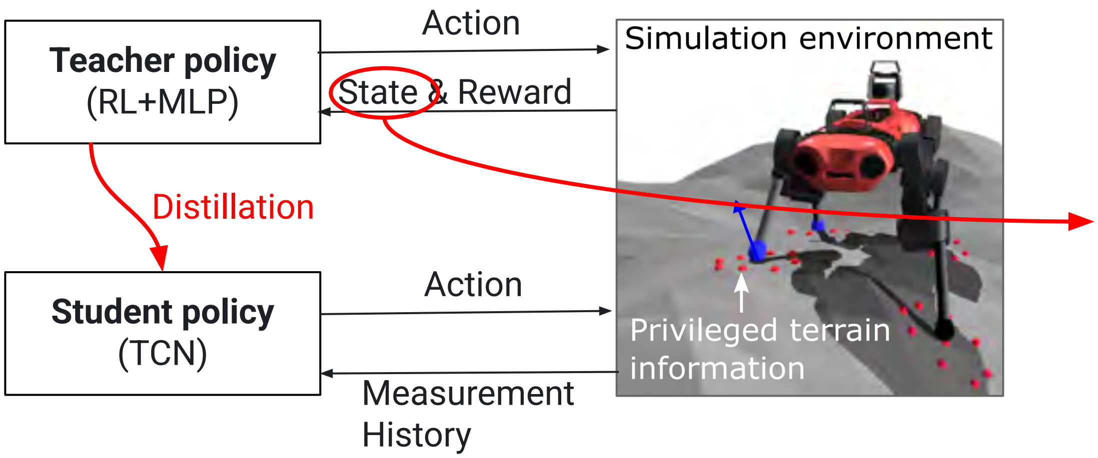
- The study utilizes quadrupedal robots.
- The robots are torque-controlled.
- The controller receives a direction from the locomotion policy (command), outputs desired foot positions
- Transforming the desired foot position in joint position using inverse kinematics.
- A PD controller tracks the desired joint positions.

Type of gait and sensors used

- The learned controller produces a natural trot gait.
- Diagonal pairs of legs move in sync.
- The controller uses only proprioceptive sensors: joint encoders (joint angles and velocities) and an IMU (body orientation and angular velocities).
- No exteroceptive sensors (cameras, LiDARs, or foot force sensors) are used.
- This design ensures high robustness against visual degradation, occlusions, and environmental factors.

Policy Training

- Training is done entirely in simulation and transferred to the physical robot without additional tuning.
- Reinforcement learning is used.
- Two-stage learning process.



1) Privileged Learning: MLP Teacher policy uses privileged information to learn locomotion via RL.

Privileged Informations:

- contact states & forces
- terrain profile
- friction coeff.
- disturbances

2) A teacher policy trains the TCN student policy through imitation using only proprioceptive measurements.

- Privileged training gives systematically better results than the baseline.

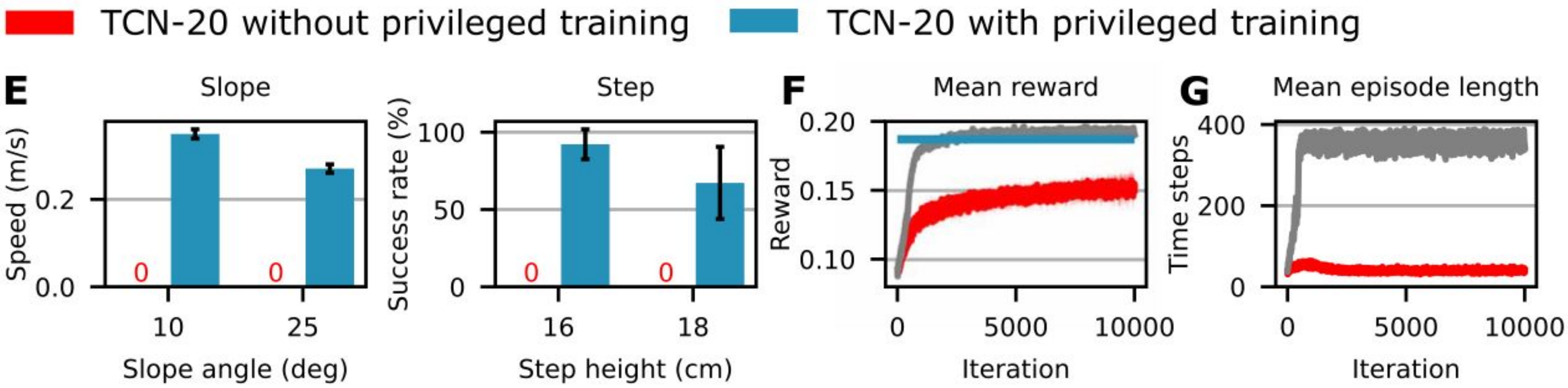
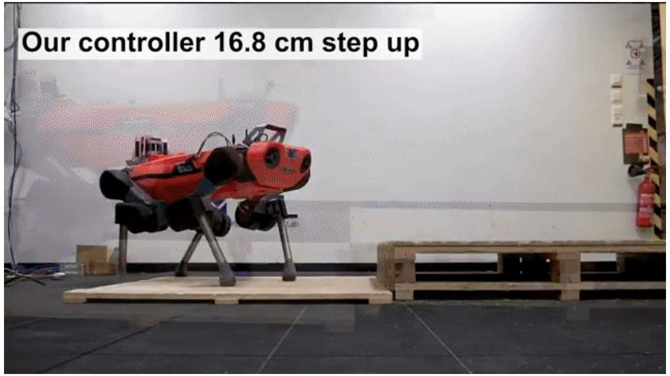


Fig 2, Ablation studies. (E to G) Importance of privileged training.

Temporal Convolutional Network (TCN)

- TCN uses temporal convolutions to process recent input sequences, enabling **short-term memory**.
- TCN learns about foot contact and slipping from proprioceptive history.



- Slope: small impact of memory length
- Step: controller with longer memory has better traversibilty and higher steps
- External force: more robust

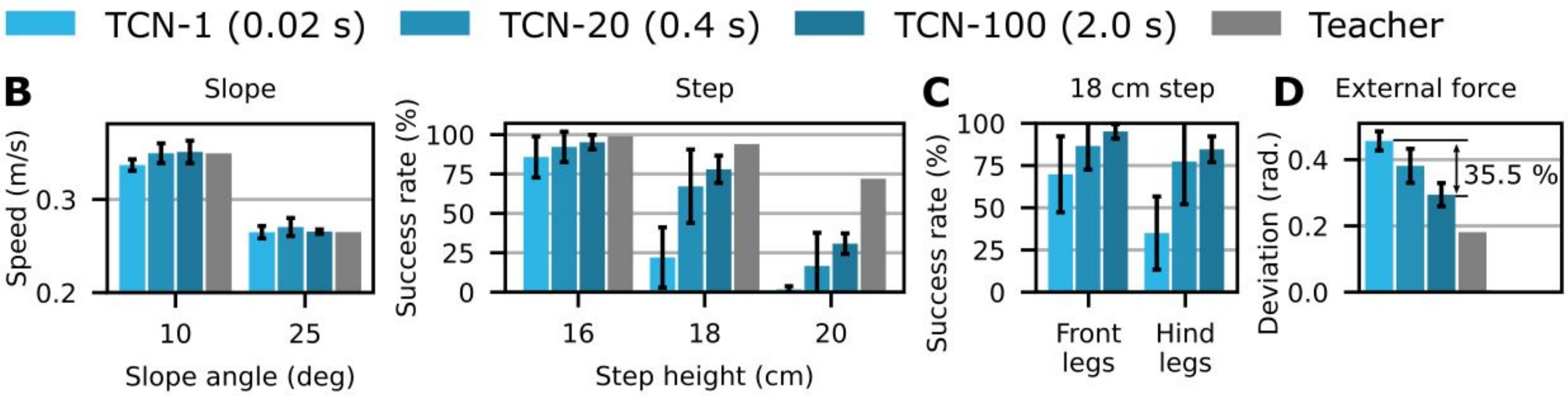
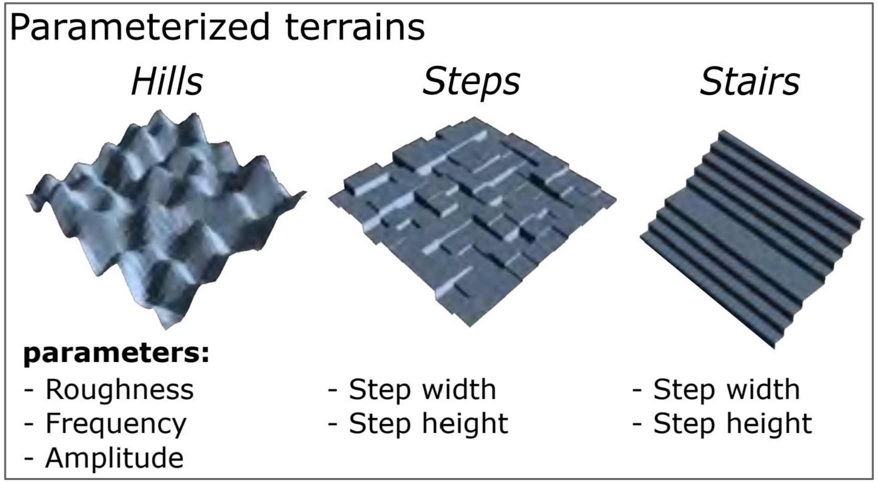


Fig 1, Ablation studies. (B to D) Importance of memory length N in the TCN-N encoder.

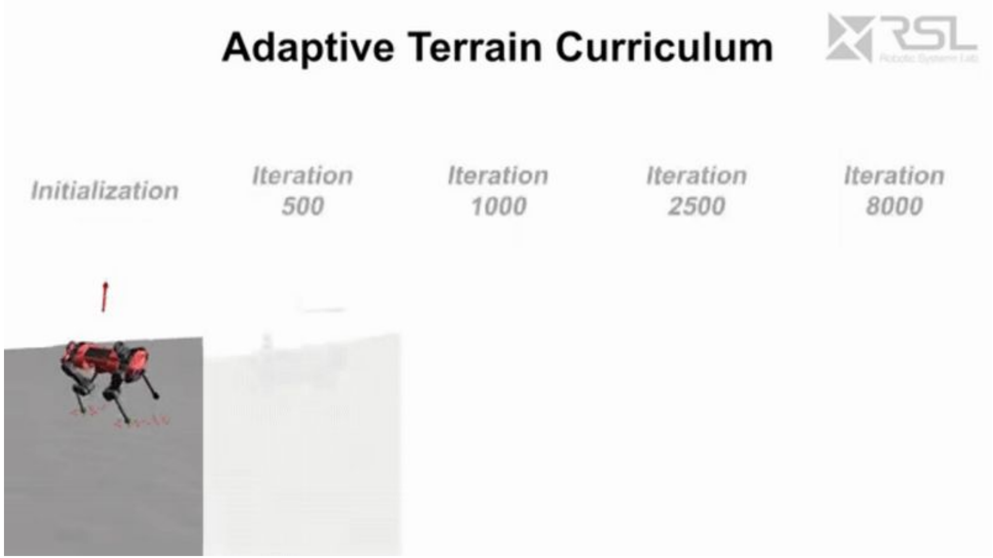
Automatic Terrain Curriculum

- To make the learning process efficient and robust, they introduce an **automatic terrain curriculum.**
- During training, the simulation continuously generates terrains (hills, steps and stairs) with various parameters.



Automatic Terrain Curriculum

- A measure of **traversability** is used to adjust the terrain difficulty accordingly to how well the current policy performs.
- This ensures that generated terrains are neither too easy nor too difficult.



- Untraversable terrains lead to early policy failure.
- The adaptive curriculum optimize learning per episode.

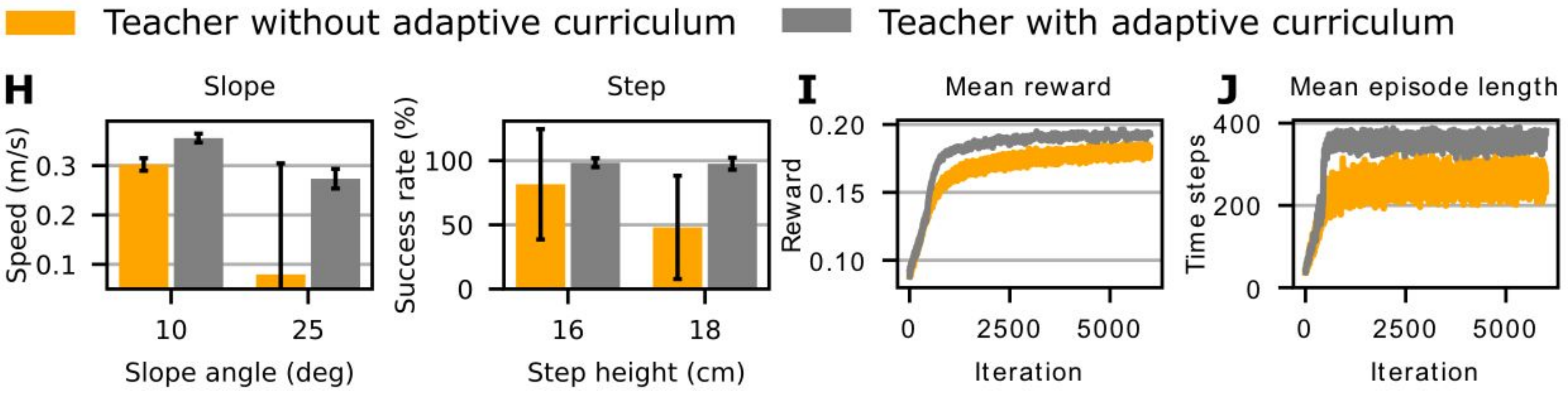


Fig 3, Ablation studies. (H to J) Importance of the adaptive curriculum.

Citation, influence and critics

*”Lee et al. [4] extended this approach and enabled rough-terrain locomotion by simulating challenging terrain in a privileged training setup with an adaptive curriculum.” “The presented approach achieves **substantial improvements over the state of the art in locomotion speed and obstacle traversability while maintaining exceptional robustness.**”*

Learning quadrupedal locomotion on deformable terrain (Choi et al., 2023)

Citation, influence and critics

- **Further studies build on limitations:**
  - Unable to simulate non-rigid terrains (mud or snow): "Learning quadrupedal locomotion on deformable terrain" (Choi et al., 2023)
  - Reliance on proprioception: "Learning robust perceptive locomotion for quadrupedal robots in the wild" (Miki et al., 2022)
- **Shift in approaches for controllers:** Handcrafted State Machines -> Learning-based Policies
- **Emphasis on Generalization and Robustness**

Citation, influence and critics

MIT CSAIL’s Mini Cheetah

Source: “Rapid Locomotion via Reinforcement Learning” (Margolis et al., 2022 / MIT CSAIL)

Advantages

- **Robustness and training strategy:** Demonstrate zero-shot generalization. Train with rigid terrains in simulation and robust in deformable terrains. The controller learns complex reflexes without being explicitly programmed. Can handle payload without being trained with.
- **Reduced sensor dependency:** No exteroceptive sensors that can't measure physical characteristics (e.g. friction, deformability).

Limitations

- **Blind Locomotion:** Can’t avoid hazards. The robot is forced to adopt a more conservative gait to feel out the environment with its body. Which limits its potential speed and efficiency in safer conditions.
- **Limited Gait Variety:** Only trot gait. The lack of proactive control limits the gait adaptivity.

References

- Kim et al., *Highly Dynamic Quadruped Locomotion via Whole-Body Impulse Control and Model Predictive Control* (2019).
- Lee et al., *Learning Quadrupedal Locomotion over Challenging Terrain* (2020).
- Margolis et al., *Rapid Locomotion via Reinforcement Learning* (MIT CSAIL, 2022).
- Miki et al., *Learning robust perceptive locomotion for quadrupedal robots in the wild* (2022).
- Choi et al., *Learning quadrupedal locomotion on deformable terrain* (2023).
- Shi et al., *Robust Quadrupedal Locomotion via Risk-Averse Policy Learning* (ICRA, 2024).

References

- YouTube. https://www.youtube.com/watch?v=tPixnjLbTvE
- YouTube. https://www.youtube.com/watch?v=aMPwB3t4idU
- YouTube. https://www.youtube.com/watch?v=9j2a1oAHDL8
- YouTube. https://youtu.be/-BqNl3AtPVw?si=UTGY0s-hv3t8-pYz