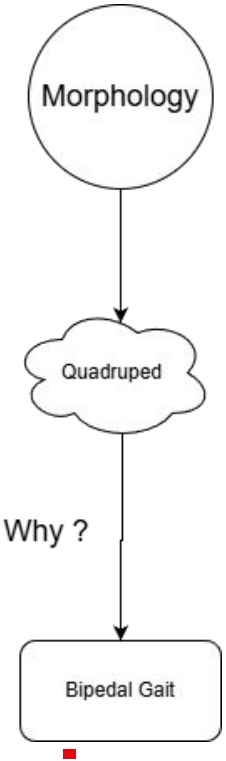
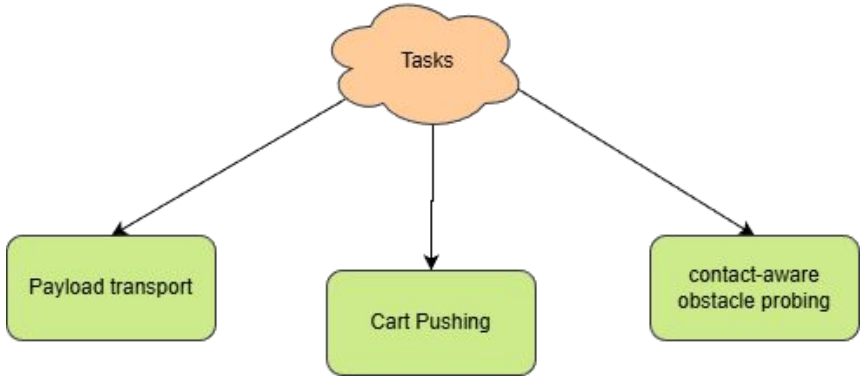
EPFL

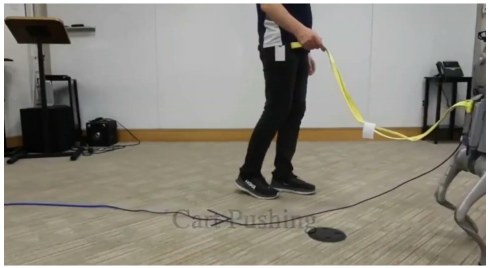
Bipedalism for Quadrupedal Robots: Versatile Loco-Manipulation through Risk-Adaptive Reinforcement Learning

Salah Slaoui Hasnaoui
Tom Herrmann
Alexandros Dellios

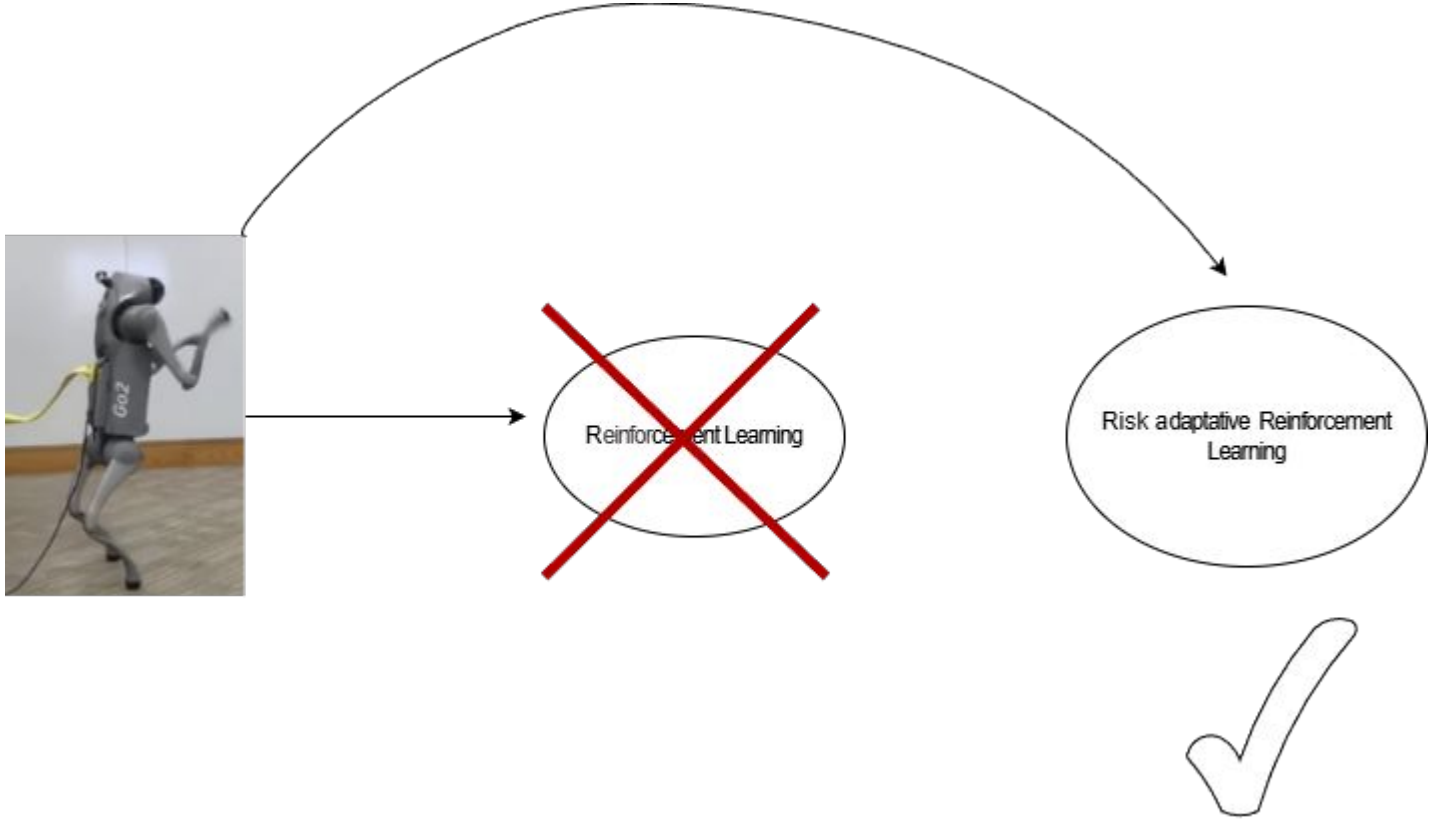
Key aspects 1 : Morphology, Gait and Tasks



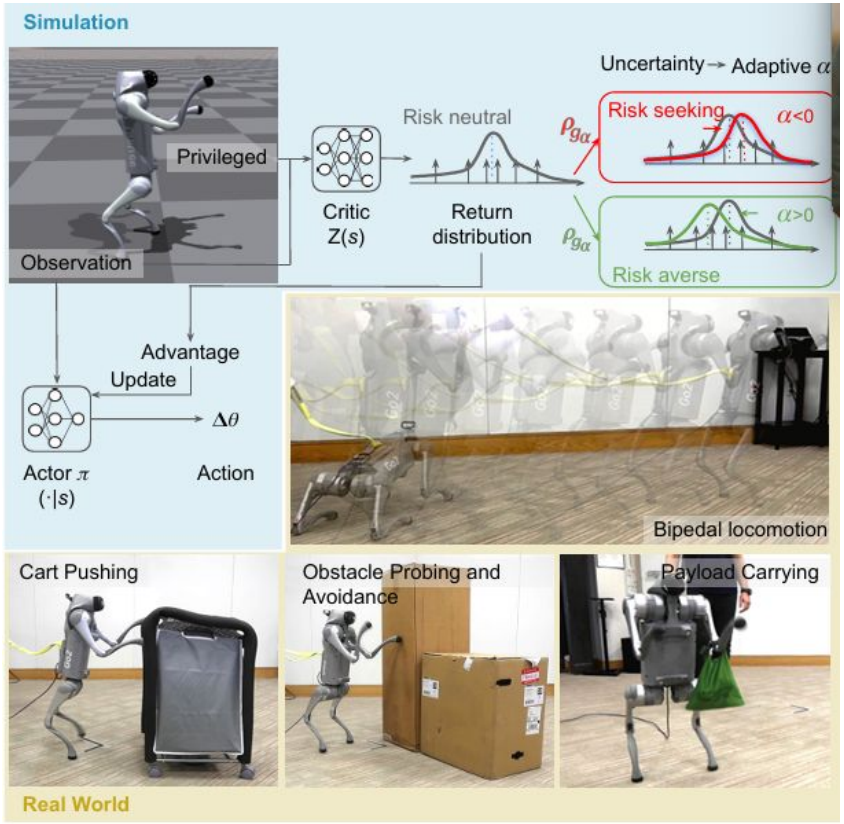




Introduction : Main idea



Key Aspect 2 : Learning Method



- CV measure : allows to **measure** the **uncertainty** of the return distribution and then adapt the risk level

$$CV_{Z_\theta^\pi(x)} = \frac{\sqrt{\mathrm{Var}(Z_\theta^\pi(x))}}{\mathbb{E}Z_\theta^\pi(x)} = \frac{\sigma}{\mu}.$$

$$\alpha_t = (\alpha_0 - \alpha_T)e^{-\frac{t/T}{CV_t}} + \alpha_T,$$

$\alpha > 0 \rightarrow$ Risk-averse

$\alpha < 0 \rightarrow$ Risk-seeking

Distorted return distribution

Reward functions

Upright and stable posture

Follow the tracking command

TABLE I

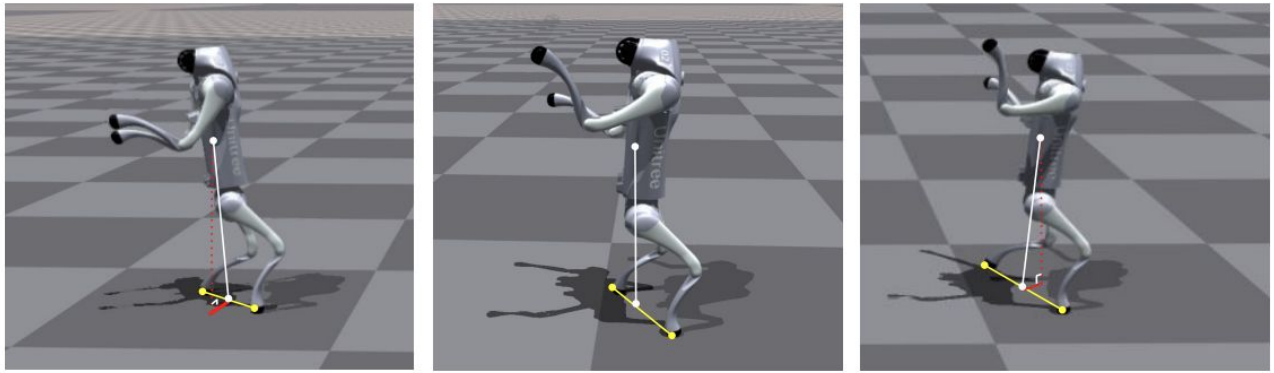
TASK REWARD FUNCTIONS

Base Height	$-(z - z^c)^2$
Base Pitch	$-\cos(p^c - p)$
Upright Balance	$\exp(-v_z^2/\sigma) + \exp(-\dot{p}^2/\sigma_{\text{yaw}})$ if is upright, else 0
Linear Tracking	$\exp(-\lvert v_{xy} - v_{xy}^c\rvert^2/\sigma)$ if is upright, else 0
Angular Tracking	$\exp(-\lvert w_{\text{yaw}} - w_{\text{yaw}}^c\rvert^2\sigma_{\text{yaw}})$ if is upright, else 0
Support Polygon	$-\lvert v_x^c\rvert^2 \left(\frac{\pi}{2} - \lvert \arctan(\frac{\Delta x_b}{\Delta z_b})\rvert\right)^2$ if $\arctan(\frac{\Delta x_b}{\Delta z_b})v_x^c < 0$

Source : Zhao Y., Corcodel R., Zhao D. – Bipedalism for Quadrupedal Robots: Versatile Loco-Manipulation through Risk-Adaptive Reinforcement Learning

Reward functions

- Area determined by the contact points of the legs
- Allows continuous balance control



(a) Accelerate (b) Neutral (c) Decelerate

Source : Zhao Y., Corcodel R., Zhao D. – Bipedalism for Quadrupedal Robots: Versatile Loco-Manipulation through Risk-Adaptive Reinforcement Learning

Results

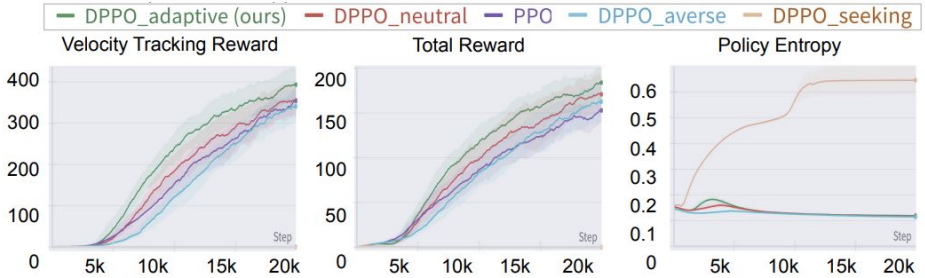


Fig. 3. Learning curves of proposed method (*DPPO_adaptive*) against baselines listed in Table II. The rewards are averaged over three seeds, and the shaded region represents the standard error.

High success rate and lowest tracking error compared to others

-0.25 m/s -> performs not as well (going backwards)

Risk-seeking fails to learn a policy in 20k steps

Risk-averse second to last (doesn't explore)

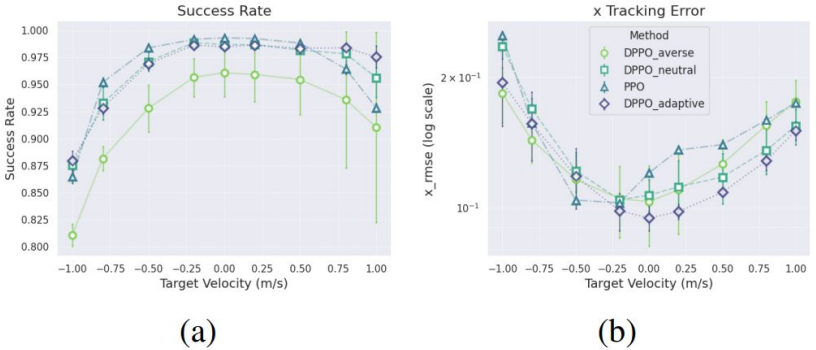
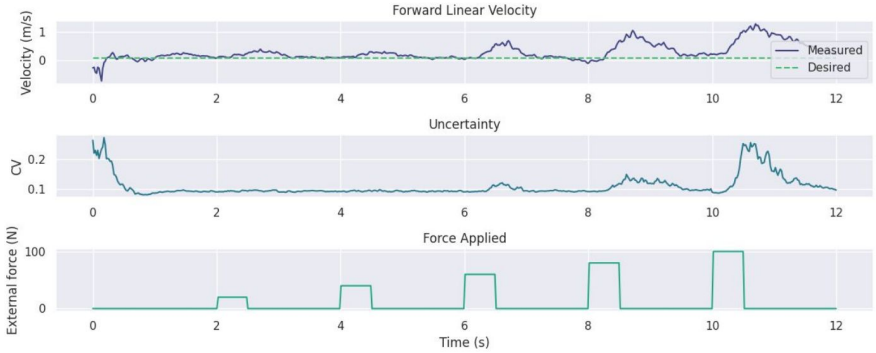


Fig. 5. Success Rate (a) and X Tracking Error (b) across target velocities ranging from -1.0 m/s to 1.0 m/s. Comparison of *DPPO_adaptive* with three baseline methods.

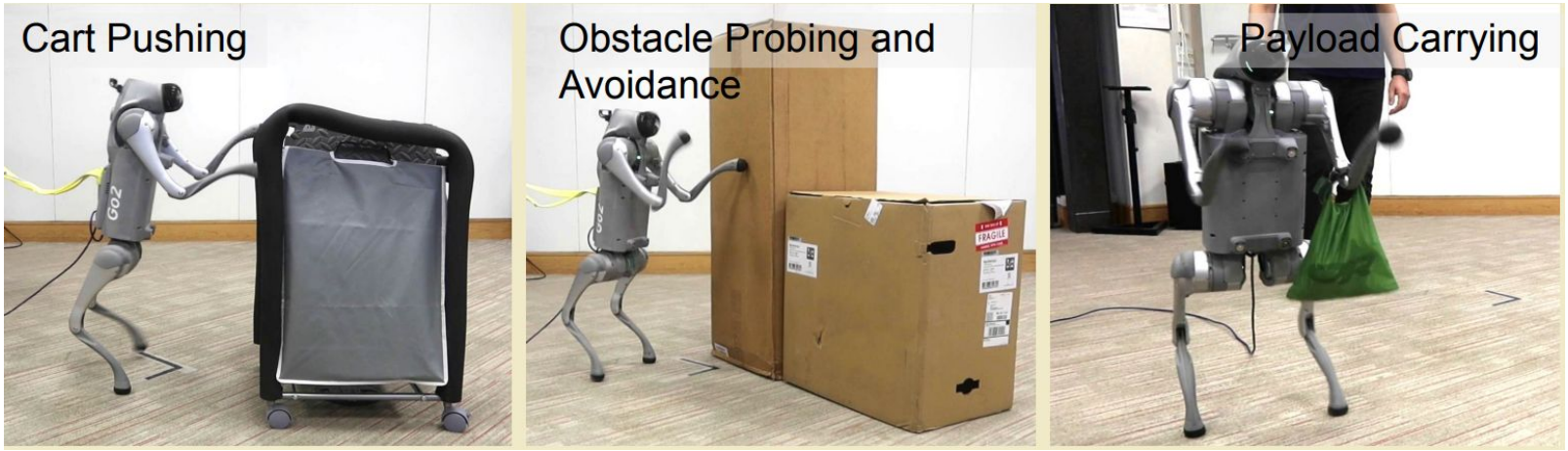
Uncertainty



The model correctly identifies higher-risk situation caused by external perturbations

t=0 → less explored state → more uncertainty

<< Balancing **conservatism** in high-uncertainty states with **optimism** in well-explored states>>



single bipedal locomotion policy -> the robot manages to accomplish multiple tasks **without requiring specific task training** => robustness

Pros and cons

- More efficient training
- Robust to variety of tasks
- Doesn’t require manual risk tuning
- Not destabilized as much from sim-to-real transfer

- Backward walking
- Front legs not able to complete complex tasks
- 0 Citations -> still too new ?