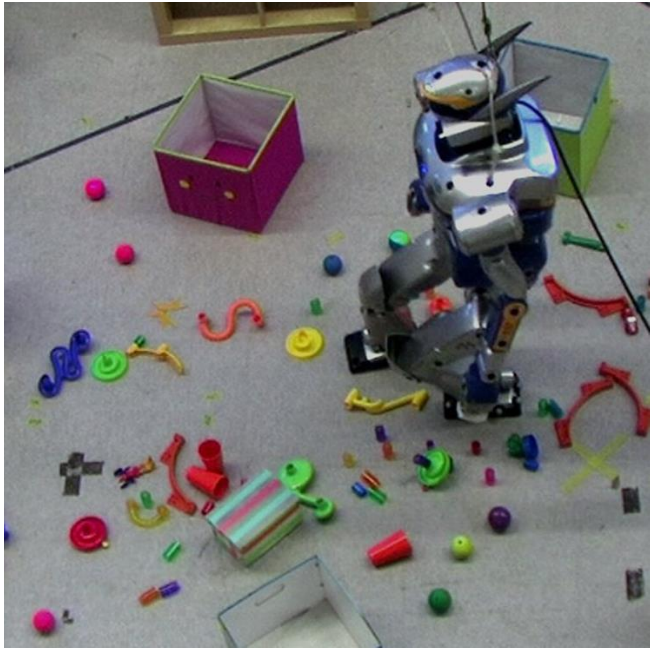
Agile Legged Locomotion and Visual Imitation

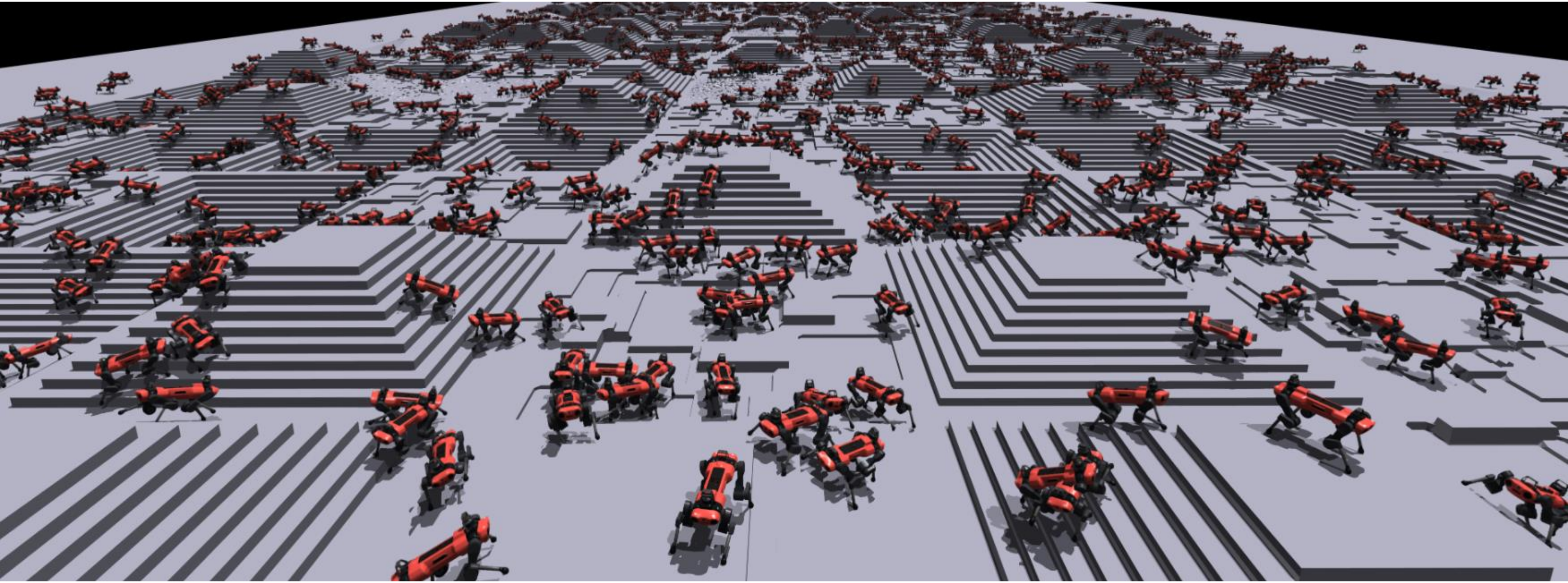
Elliot Chane-Sane
LAAS-CNRS

Legged Locomotion



Robots that can **go anywhere.**

Reinforcement Learning



Train policies in physics simulators, then transfer on real robots.

Isaac Gym: High Performance GPU-Based Physics Simulation For Robot Learning, Makoviychuk et al., 2021
Learning to Walk in Minutes Using Massively Parallel Deep Reinforcement Learning, Rudin et al., CoRL 2021

Challenges

Agile but safe

Visual understanding

Animal-like behaviors

Constrained Reinforcement Learning

CaT: Constraints as Terminations for Legged Locomotion Reinforcement Learning

Elliot Chane-Sane*, Pierre-Alexandre Leziart*, Thomas Flayols, Olivier Stasse, Philippe Souères, Nicolas Mansard

IROS 2024

Visuomotor Control

SoloParkour: Constrained Reinforcement Learning for Visual Locomotion from Privileged Experience

Elliot Chane-Sane*, Joseph Amigo*, Thomas Flayols,Ludovic Righetti, Nicolas Mansard

CoRL 2024

Visual Imitation from Internet Videos

Reinforcement Learning from Wild Animal Videos

Elliot Chane-Sane, Constant Roux, Olivier Stasse, Nicolas Mansard

LocoLearn Workshop at CoRL 2024 - **Best Paper Award**

RL for Quadruped Locomotion

Learn a policy **maximizing** the **discounted** sum of future **rewards**

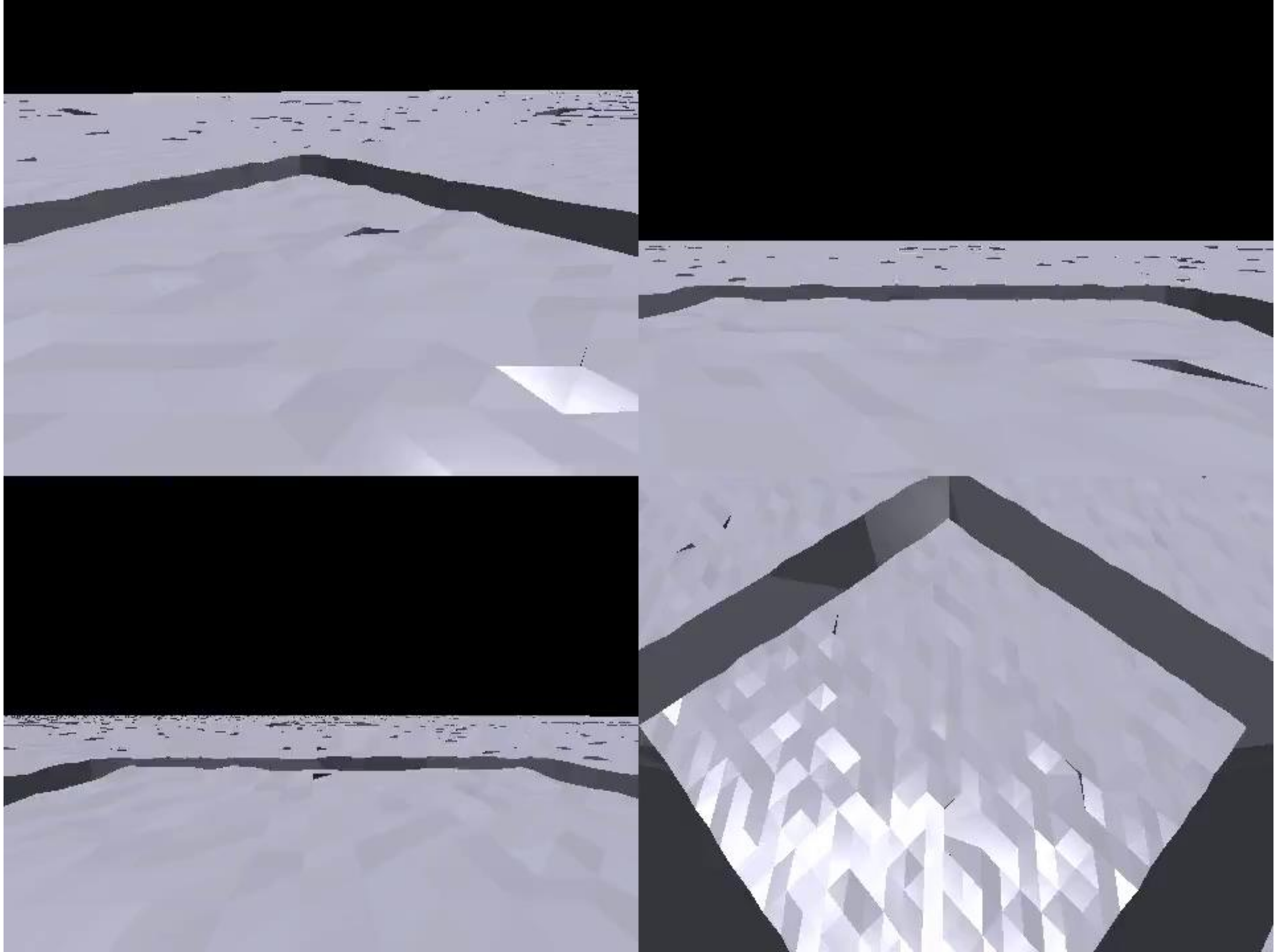
$$\pi^* = \arg\max_{\pi} \mathbb{E}_{\tau \sim \pi, \mathcal{T}} \left[\sum_{t=0}^{\infty} \underbrace{\gamma^t}_{\text{Discount factor}} \underbrace{r(s_t, a_t)}_{\text{Rewards}} \right].$$

Reward Engineering is Hard

Reward engineering is a **challenging** problem in RL.

Ex: in legged locomotion, this velocity tracking reward looks appropriate

$$r_{\text{velocity}} = \left\langle v^{\text{robot}}, d^{\text{command}} \right\rangle$$



Reward Engineering is Hard

1 new hyperparameter

$$r = r_{\text{velocity}} + \alpha_{\text{ori}} \cdot r_{\text{ori}}$$

Stabilize the base of the robot

Number of terms: 2

Reward Engineering is Hard

12 new hyperparameters

$$r = r_{\text{velocity}} + \alpha_{\text{ori}} \cdot r_{\text{ori}} + \sum_{\text{joint } i=1}^{12} \alpha_{\text{torque, i}} \cdot r_{\text{torque, i}}$$

Limit the torques applied to each joint

$$r_{\text{torque, i}} = -\text{torque}_i^2$$

<u>Number of terms</u>: 14

Reward Engineering is Hard

$$r = r_{\text{velocity}} + \alpha_{\text{ori}} \cdot r_{\text{ori}} + \sum_{\text{joint } i=1}^{12} \alpha_{\text{torque, i}} \cdot r_{\text{torque, i}} + \alpha_{\text{vel, i}} \cdot r_{\text{vel, i}}$$

Number of terms: 26

Limit the joint velocities

Reward Engineering is Hard

$$r = r_{\text{velocity}} + \alpha_{\text{ori}} \cdot r_{\text{ori}} + \sum_{\text{joint } i=1}^{12} \alpha_{\text{torque, i}} \cdot r_{\text{torque, i}} + \alpha_{\text{vel, i}} \cdot r_{\text{vel, i}} + \alpha_{\text{lim, i}} \cdot r_{\text{lim, i}}$$

Number of terms: 38

Limit the joint positions

Reward Engineering is Hard

$$r = r_{\text{velocity}} + \alpha_{\text{ori}} \cdot r_{\text{ori}} + \sum_{\text{joint } i=1}^{12} \alpha_{\text{torque, i}} \cdot r_{\text{torque, i}} + \alpha_{\text{vel, i}} \cdot r_{\text{vel, i}} + \alpha_{\text{lim, i}} \cdot r_{\text{lim, i}}$$

Hard to tune

High-variance in the learned behaviors

Weird to add penalties to the reward

Constrained RL

Maximize rewards: $$\pi^* = \arg\max_{\pi} \mathbb{E}_{\tau\sim\pi,\mathcal{T}}\left[\sum_{t=0}^{\infty}\gamma^t r(s_t, a_t)\right].$$

While satisfying constraints:

$$\left.\begin{array}{c} -3\ \mathrm{N.m} \leq \mathrm{torque}_i \leq 3\ \mathrm{N.m} \\ -16\ \mathrm{rad/s} \leq \mathrm{joint\ velocity} \leq 16\ \mathrm{rad/s} \\ \dots \end{array}\right\}$$

Involve **limits** that often have **physical meanings**

Facilitate reward engineering
Maximize the capabilities of the robot
A more **natural problem** formulation

CaT 🐈 Constraints as Terminations

Discount factor

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{\tau \sim \pi, \mathcal{T}} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \right].$$

Modify the discount factor at each step such that:

- All constraints are satisfied ⟶ the discount factor is not changed
- The bigger the constraint violation ("dangerous") ⟶ the more the discount factor is reduced ("lower life expectancy")

Task formulation: through rewards (Option A)	
Reward function	$r = e^{-\frac{\lVert v_{xy}^{\text{des}} - v_{xy} \rVert_2^2}{0.25}} + \frac{1}{2} e^{-\frac{\lvert \omega_z^{\text{des}} - \omega_z \rvert^2}{0.25}}$
Task formulation: through soft constraints (Option B)	
Reward function	$r = 1$
Linear velocity tracking	$c_{\text{lin vel}} = \lVert v_{xy}^{\text{des}} - v_{xy} \rVert_2 - \epsilon_{\text{track}}$
Angular velocity tracking	$c_{\text{ang vel}} = \lvert \omega_z^{\text{des}} - \omega_z \rvert - \epsilon_{\text{track}}$
Hard constraints for safety	
Knee or base collision	$c_{\text{knee/base contact}} = 1_{\text{knee/base contact}}$
Foot contact force	$c_{\text{foot contact}_j} = \lVert f^{\text{foot}_j} \rVert_2 - f^{\text{lim}}$
Soft constraints for safety ($\forall k \in 1..12$)	
Torque limits	$c_{\text{torque}_k} = \lvert \tau_k \rvert - \tau^{\text{lim}}$
Joint velocity limits	$c_{\text{joint velocity}_k} = \lvert \dot{q}_k \rvert - \dot{q}^{\text{lim}}$
Joint acceleration limits	$c_{\text{joint acceleration}_k} = \lvert \ddot{q}_k \rvert - \ddot{q}^{\text{lim}}$
Action rate limits	$c_{\text{action rate}_k} = \frac{\lvert \Delta q_{t,k}^{\text{des}} - \Delta q_{t-1,k}^{\text{des}} \rvert}{dt} - \dot{q}^{\text{des lim}}$
Soft constraints for style (Active on flat terrains only, $\forall j \in 1..4$)	
Base orientation	$c_{\text{ori}} = \lVert \text{base ori}_{\text{xy}} \rVert_2 - \text{base}^{\text{lim}}$
Hip orientation	$c_{\text{hip}_j} = \lvert \text{hip ori}_j \rvert - \text{hip}^{\text{lim}}$
Foot air time	$c_{\text{air time}_j} = t_{\text{air time}}^{\text{des}} - t_{\text{air time}_j}$
Number of foot contacts	$c_{\text{n foot contacts}} = \lvert n_{\text{foot contact}} - n_{\text{foot contact}}^{\text{des}} \rvert$
Stand still if $v^{\text{des}} = 0$	$c_{\text{still}} = (\lVert q - q^{\star} \rVert_2 - \epsilon_{\text{still}}) \times 1_{v^{\text{des}}=0}$

We can **stack** as **many constraints** as we need (> 100)

Constrained Reinforcement Learning

CaT: Constraints as Terminations for Legged Locomotion Reinforcement Learning

Elliot Chane-Sane*, Pierre-Alexandre Leziart*, Thomas Flayols, Olivier Stasse, Philippe Souères, Nicolas Mansard

IROS 2024

Visuomotor Control

SoloParkour: Constrained Reinforcement Learning for Visual Locomotion from Privileged Experience

Elliot Chane-Sane*, Joseph Amigo*, Thomas Flayols,Ludovic Righetti, Nicolas Mansard

CoRL 2024

Visual Imitation from Internet Videos

Reinforcement Learning from Wild Animal Videos

Elliot Chane-Sane, Constant Roux, Olivier Stasse, Nicolas Mansard

LocoLearn Workshop at CoRL 2024 - **Best Paper Award**

CoRL 2024

SoloParkour: Constrained Reinforcement Learning for Visual Locomotion from Privileged Experience

Elliot Chane-Sane*, Joseph Amigo*, Thomas Flayols, Ludovic Righetti, Nicolas Mansard

Gepetto
LAAS-CNRS

NYU

Constrained Reinforcement Learning

CaT: Constraints as Terminations for Legged Locomotion Reinforcement Learning

Elliot Chane-Sane*, Pierre-Alexandre Leziart*, Thomas Flayols, Olivier Stasse, Philippe Souères, Nicolas Mansard
IROS 2024

Visuomotor Control

SoloParkour: Constrained Reinforcement Learning for Visual Locomotion from Privileged Experience

Elliot Chane-Sane*, Joseph Amigo*, Thomas Flayols, Ludovic Righetti, Nicolas Mansard
CoRL 2024

Visual Imitation from Internet Videos

Reinforcement Learning from Wild Animal Videos

Elliot Chane-Sane, Constant Roux, Olivier Stasse, Nicolas Mansard
LocoLearn Workshop at CoRL 2024 - **Best Paper Award**

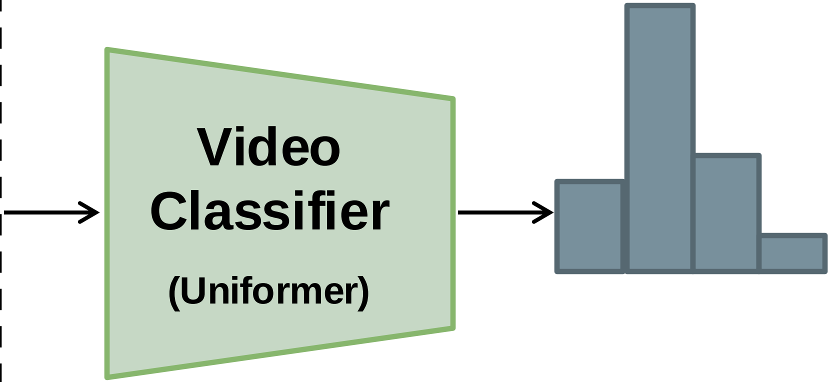
Can robots learn by watching thoushands of wild animal videos from the internet ?

Cross-Embodiement Visual Imitation

Humans/animals can visually imitate each other, even with significant differences in physical embodiements.

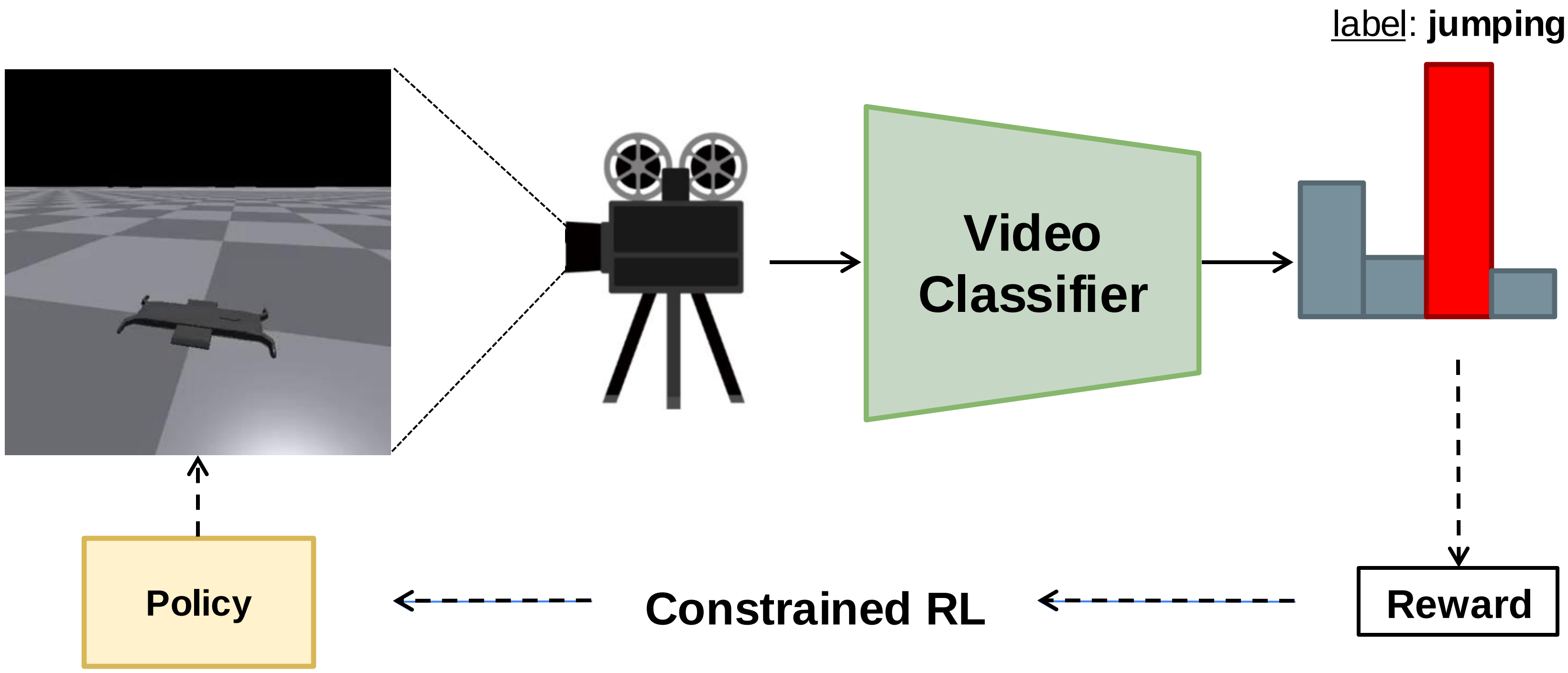
Video-Based Reward Function

Learning motion concepts by watching thoushands of wild animal videos



Animal Kingdom: A Large and Diverse Dataset for Animal Behavior Understanding, Ng et al., CVPR 2022
UniFormer: Unified Transformer for Efficient Spatiotemporal Representation Learning, Li et al., ICLR 2022

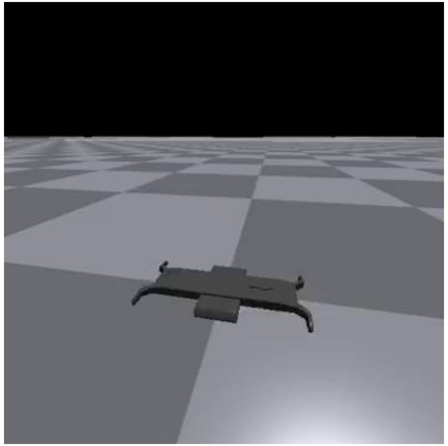
Reinforcement Learning in a Physics simulator



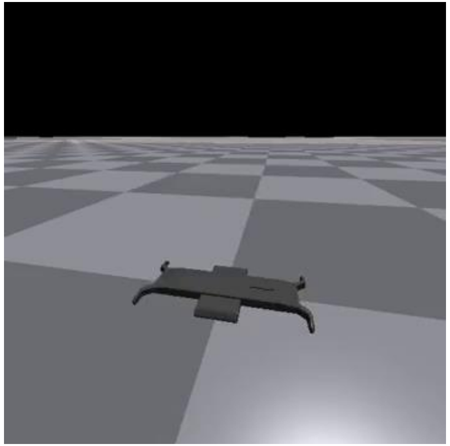
Grounding these motion concepts into physical robotic skills

Multi-skill Locomotion Policy

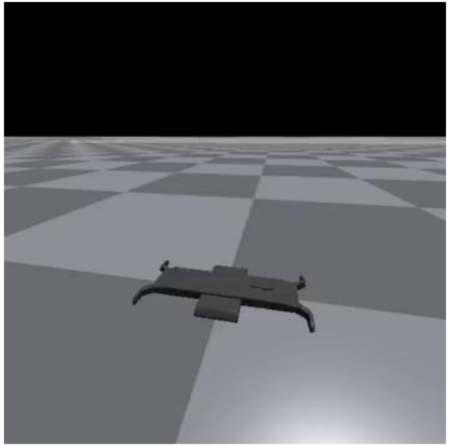
Keeping Still



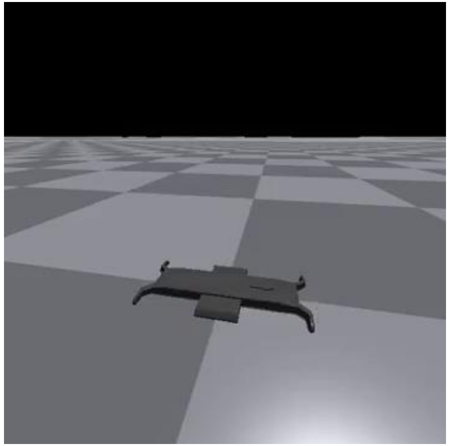
Walking



Running



Jumping




Walking

Constrained Reinforcement Learning

CaT: Constraints as Terminations for Legged Locomotion Reinforcement Learning

Elliot Chane-Sane*, Pierre-Alexandre Leziart*, Thomas Flayols, Olivier Stasse, Philippe Souères, Nicolas Mansard

IROS 2024

Visuomotor Control

SoloParkour: Constrained Reinforcement Learning for Visual Locomotion from Privileged Experience

Elliot Chane-Sane*, Joseph Amigo*, Thomas Flayols,Ludovic Righetti, Nicolas Mansard

CoRL 2024

Visual Imitation from Internet Videos

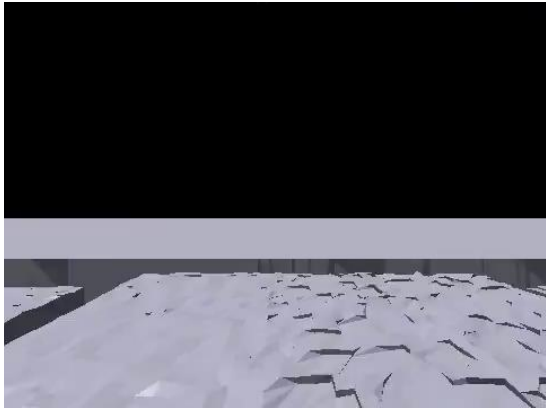
Reinforcement Learning from Wild Animal Videos

Elliot Chane-Sane, Constant Roux, Olivier Stasse, Nicolas Mansard

LocoLearn Workshop at CoRL 2024 - **Best Paper Award**

Towards general locomotion ?

RL in Physics Simulators
(SoloParkour)



Narrow motion intelligence
Physically grounded

Fine for parkour but
not yet for general phycisal intelligence

Internet-Scale Video Understanding
(Kling AI video editing)

General motion intelligence
No physical grounding

Fine for generating nice videos
but not yet for robotics

Towards general locomotion ?

Large-Scale Visual Imitation of Internet Videos

Action conditioning

Language conditioning

“do the shuffle dance”

ANITI

LAAS CNRS

Nicolas Mansard

P-A Leziart

Thomas Flayols

NYU

Ludovic Righetti

Joseph Amigo

Olivier Stasse

Constant Roux