



Continual **AI**



www.robots that dream.eu

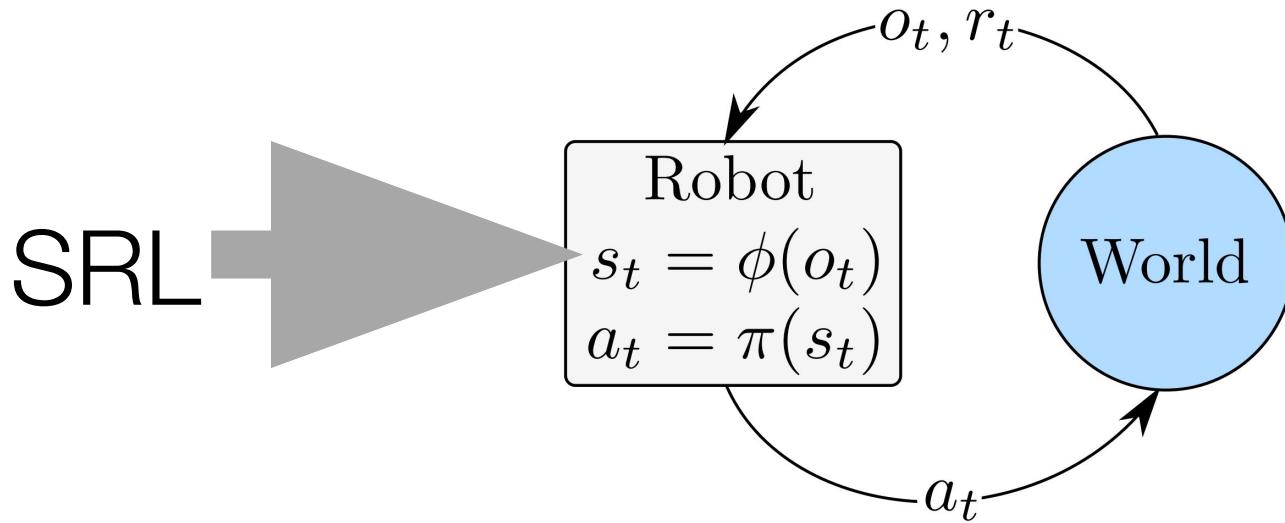
Decoupling Feature Extraction from Policy Learning: assessing benefits of state representation learning in goal based robotics

Natalia Díaz Rodríguez, PhD

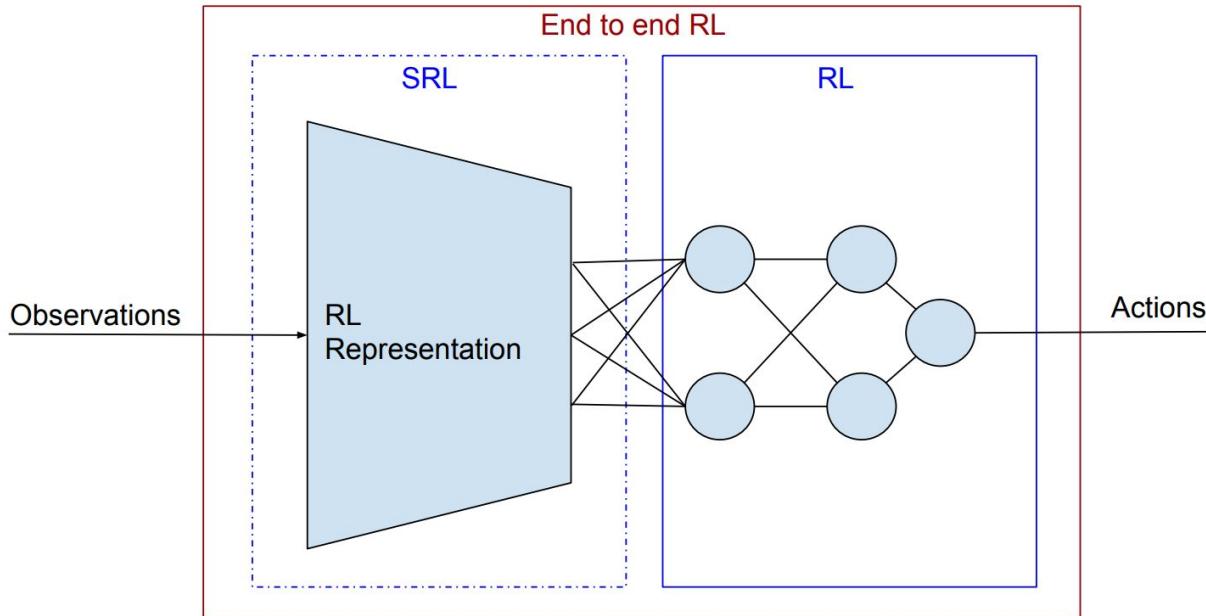


Antonin Raffin, Ashley Hill, René Traoré, Timothée Lesort, Natalia Díaz-Rodríguez, David Filliat
U2IS & INRIA FLOWERS Team, ENSTA ParisTech, Palaiseau, France

State Representation Learning (SRL) in RL context

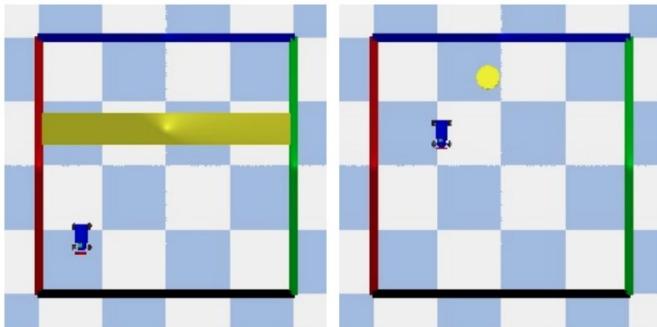


State Representation Learning (SRL) in RL context

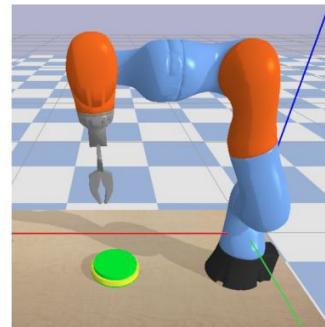


Goal based robotic tasks

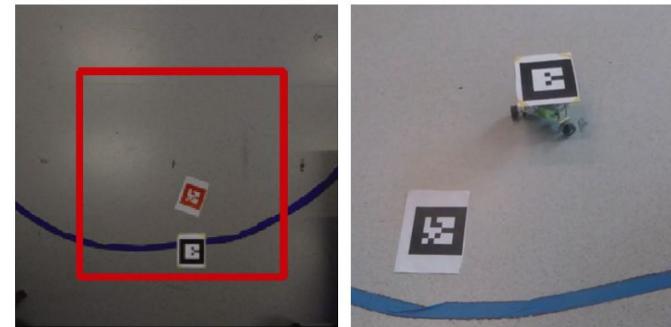
Mobile Navigation



Robotic Arm

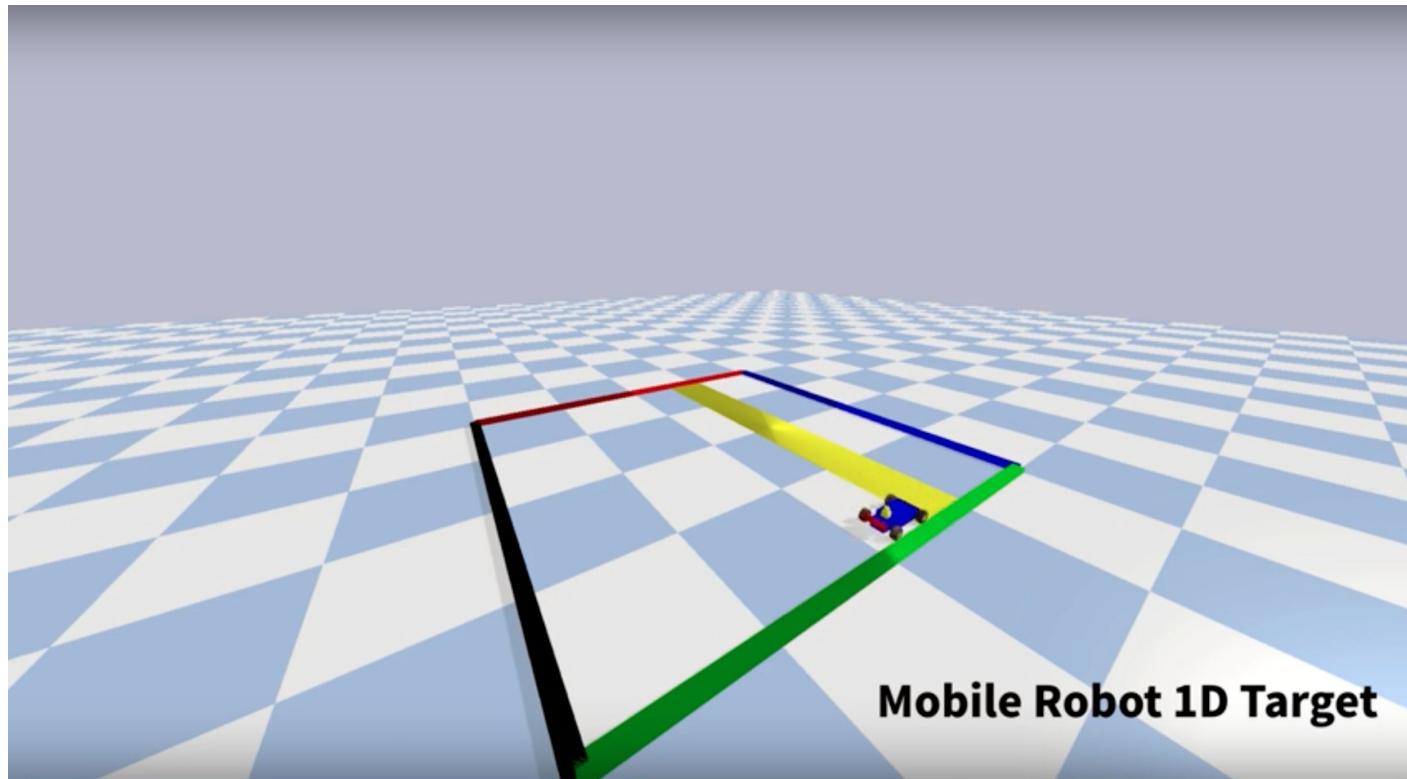


Simulated & Real Omnibot

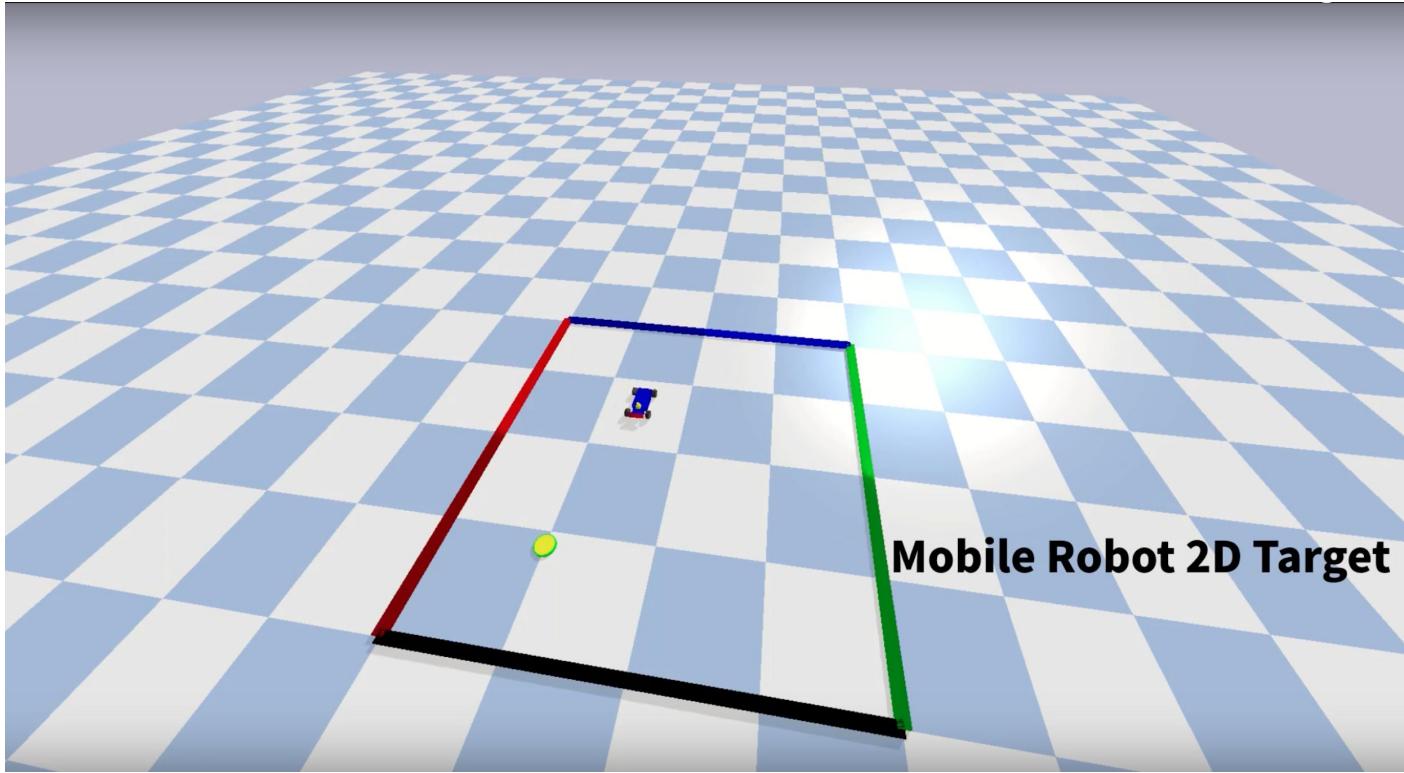


Environments, code and data are available at <https://github.com/araffin/robotics-rl-srl>.

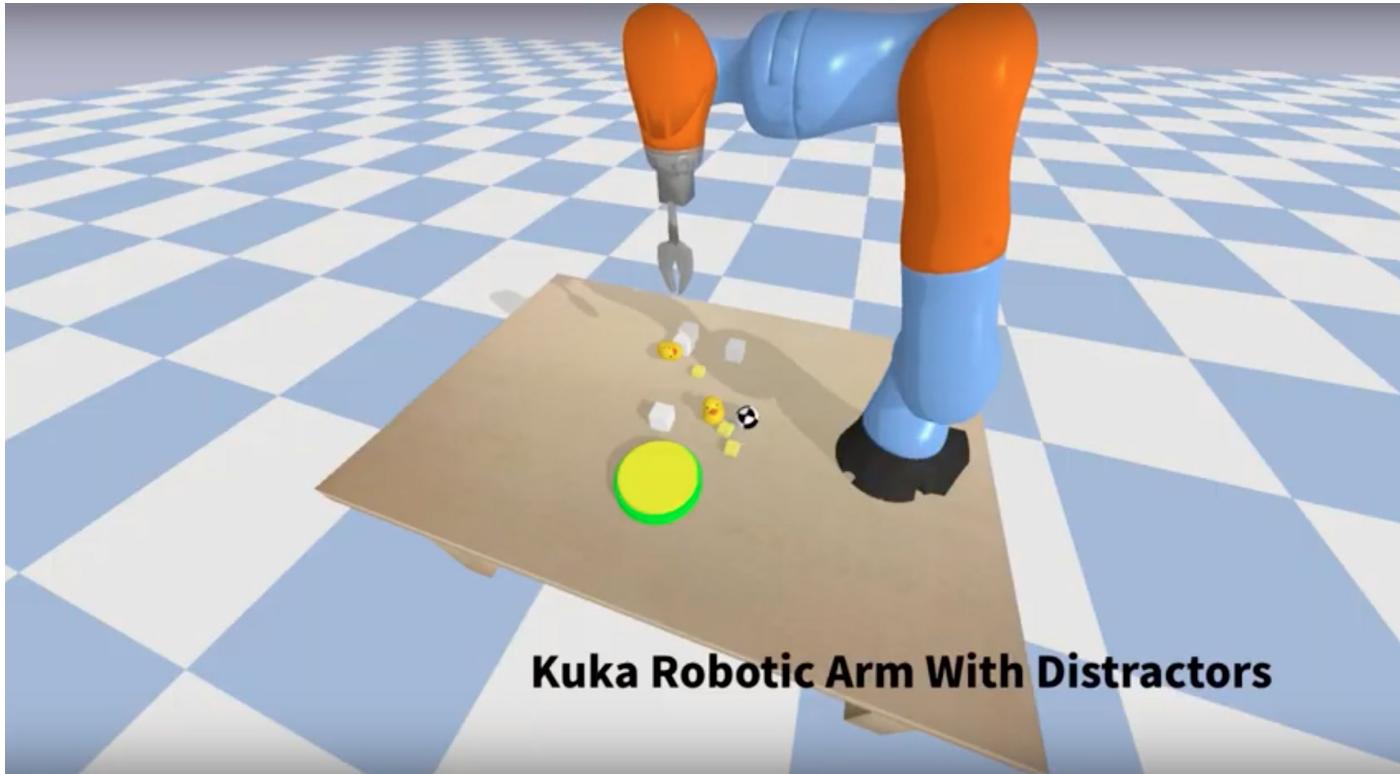
S-RL Toolbox



S-RL Toolbox

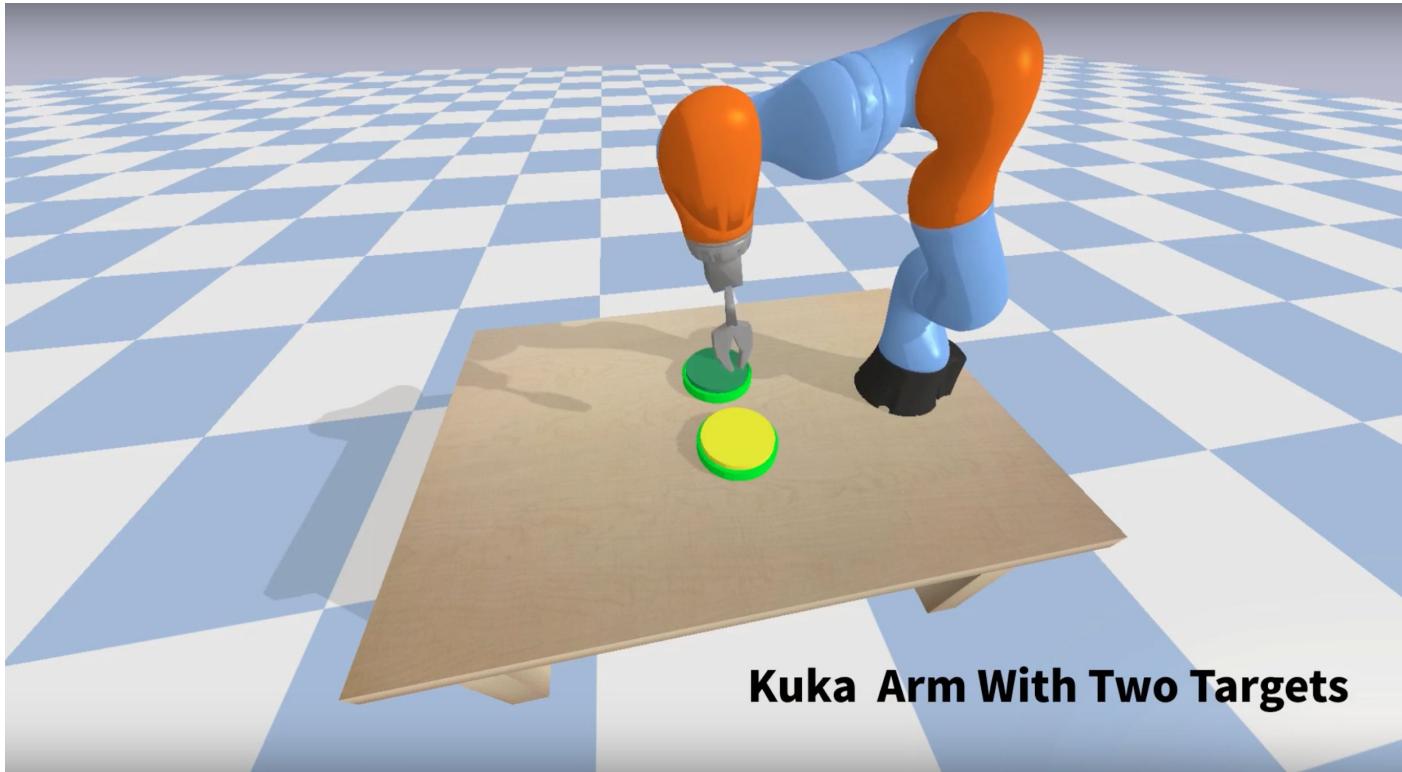


S-RL Toolbox



Kuka Robotic Arm With Distractors

S-RL Toolbox

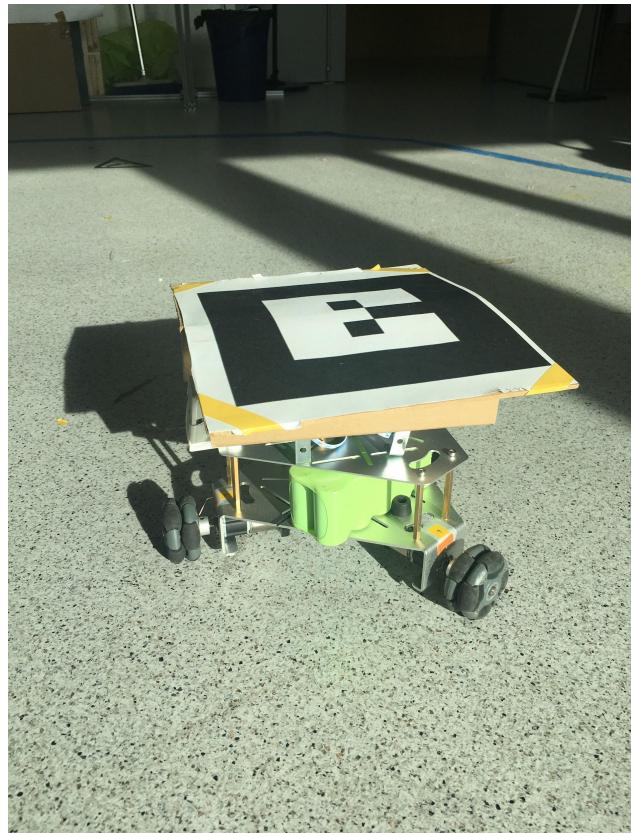


Kuka Arm With Two Targets

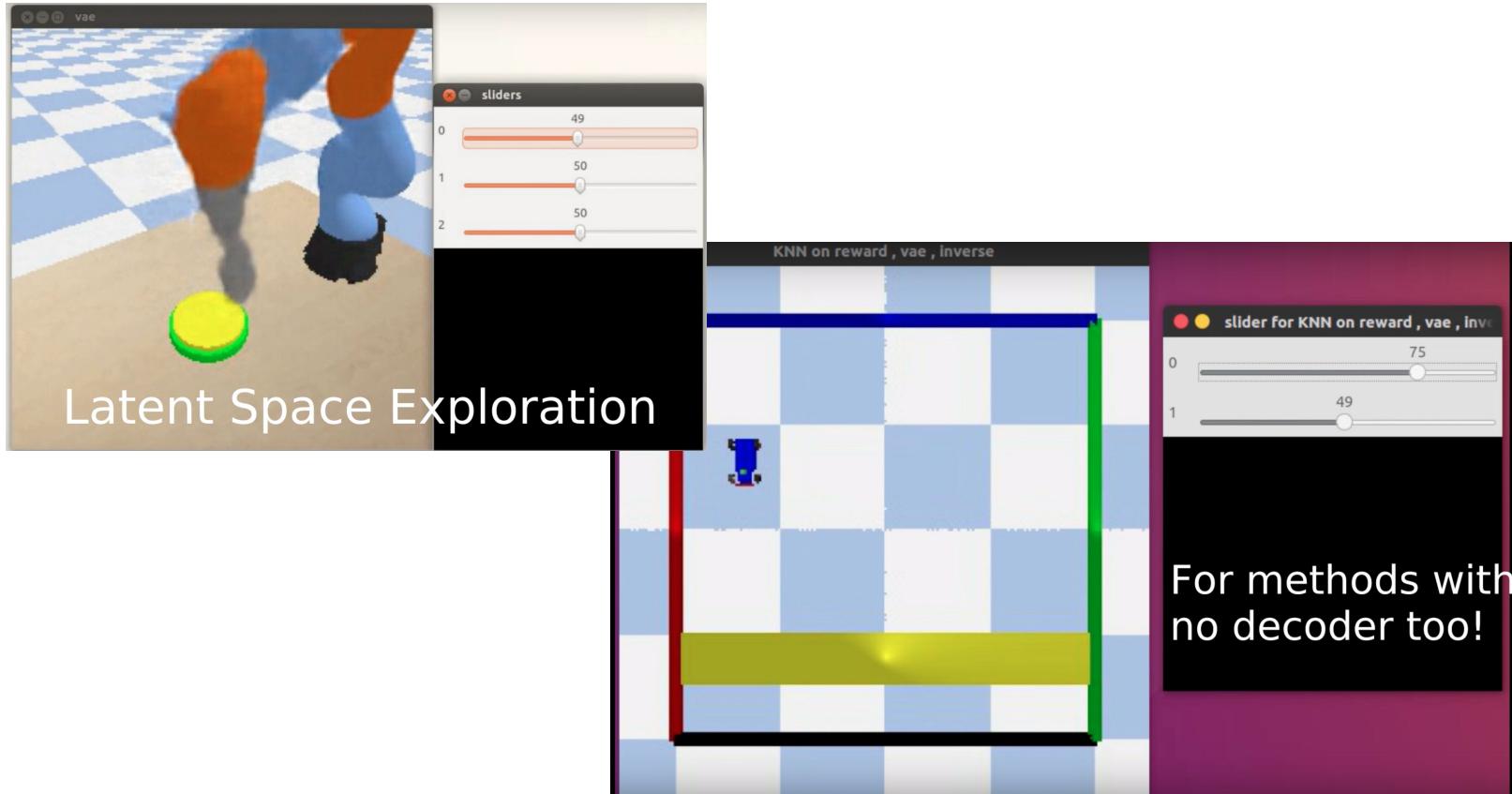
S-RL Toolbox



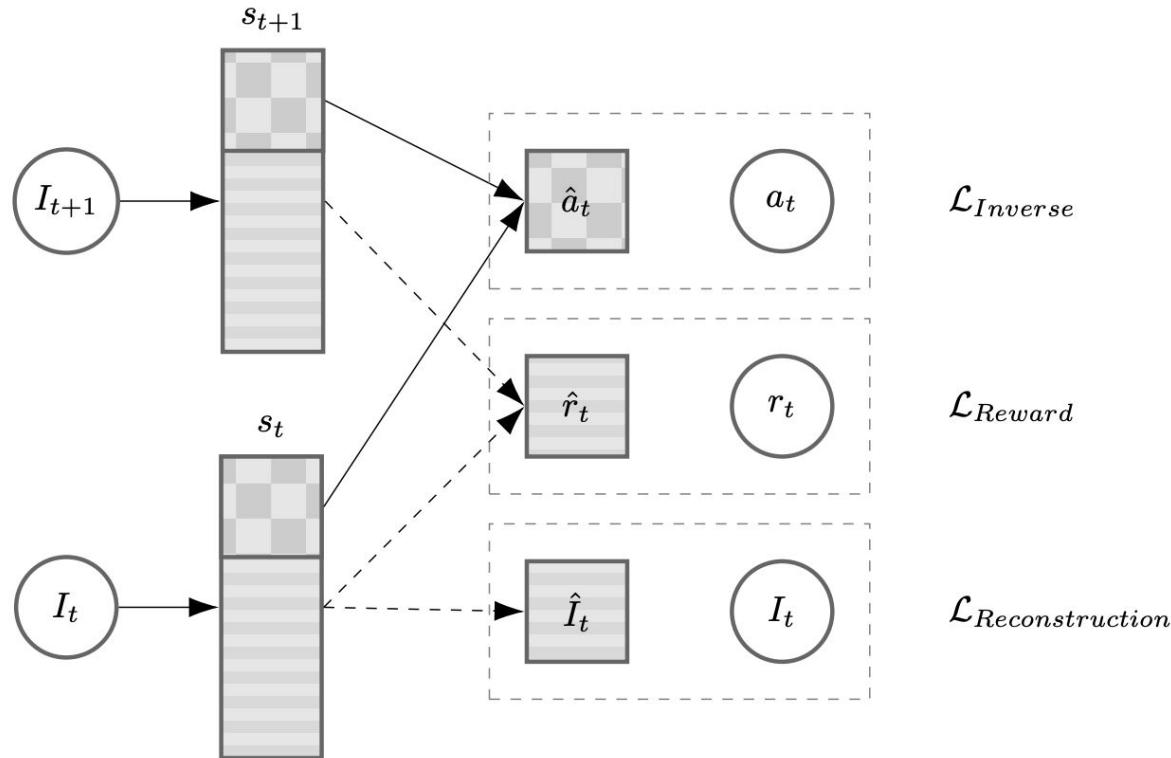
S-RL Toolbox



S-RL Toolbox



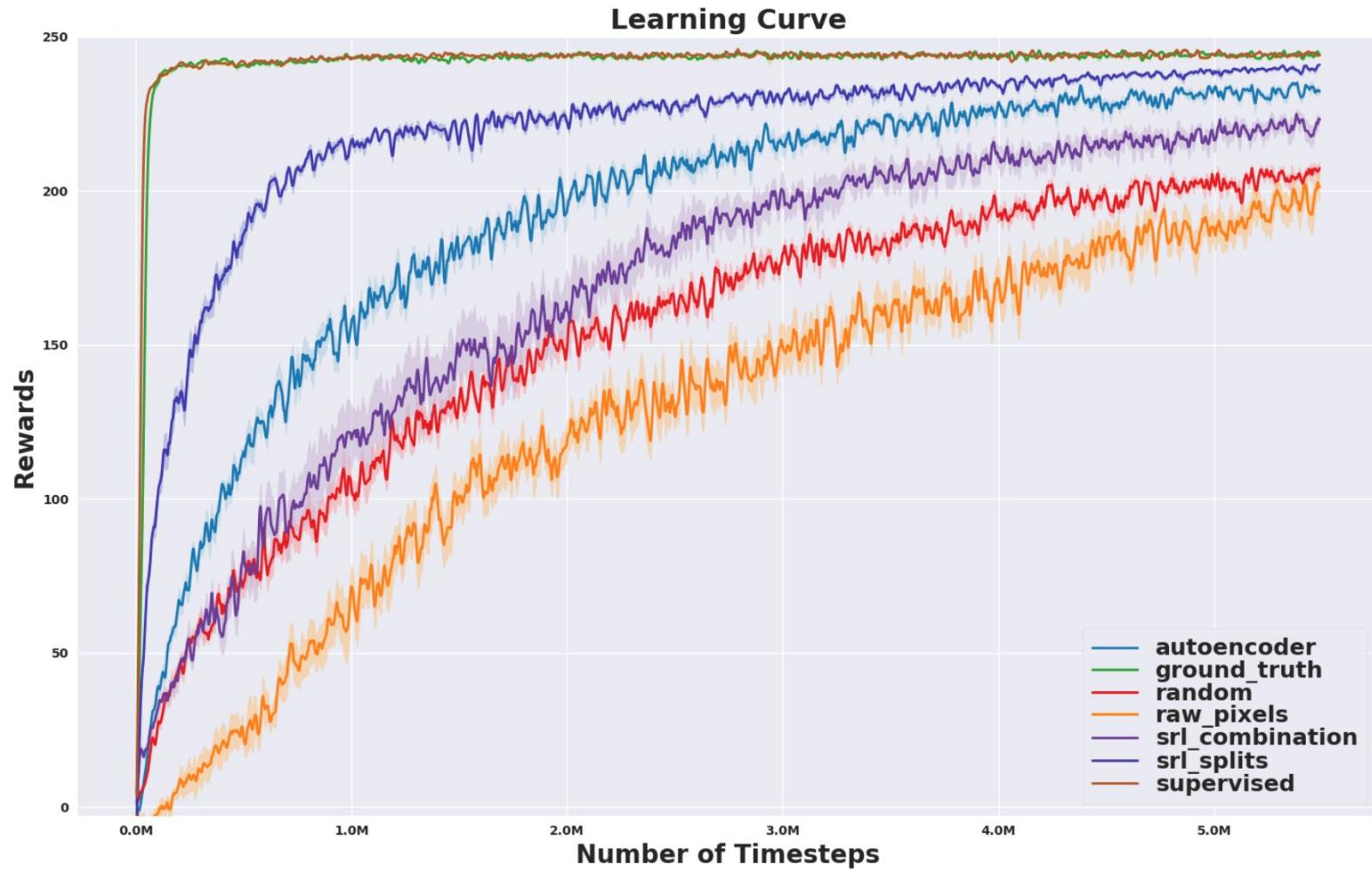
Decoupling feature extraction from policy learning



Baselines and Results

Environments	<i>Nav. 1D Target</i>	<i>Nav. 2D Target</i>	<i>Arm Random Target</i>	<i>Pseudo-Real Omnidot</i>
Ground Truth	211.6 ± 14.0	234.4 ± 1.3	4.2 ± 0.5	243.7 ± 1.2
Supervised	189.7 ± 14.8	213.5 ± 6.0	3.1 ± 0.3	243.9 ± 1.8
Raw Pixels	215.7 ± 9.6	231.5 ± 3.1	2.6 ± 0.3	185.2 ± 7.83
Random Features	211.9 ± 10.0	208 ± 6.1	4.1 ± 0.3	201.5 ± 5.7
Auto-Encoder	188.8 ± 13.5	192.6 ± 8.9	3.4 ± 0.3	230.27 ± 3.2
SRL Combination	216.3 ± 10.0	183.6 ± 9.6	2.9 ± 0.3	216.8 ± 5.6
SRL Splits	205.1 ± 11.7	232.1 ± 2.2	3.7 ± 0.3	237.8 ± 2.1

Table 1: End-to-end vs State Representation Learning: Mean reward performance and standard error in RL (using PPO) per episode (average on 100 episodes) at the end of training for all the environments tested.



$$GTC_{(i)} = \max_j |\rho_{s,\tilde{s}}(i, j)| \in [0, 1] \quad (1)$$

where $\rho_{s,\tilde{s}}$ is the Pearson correlation coefficient for the pair (s, \tilde{s}) , where \tilde{s} is the GT state, s the learned state, $i \in \llbracket 0, |\tilde{s}| \rrbracket$, $j \in \llbracket 0, |s| \rrbracket$, $\tilde{s} = [\tilde{s}_1; \dots; \tilde{s}_n]$, and \tilde{s}_k being the k^{th} dimension of the GT state vector.

$$GTC_{(i)} = \max_j |\rho_{s,\tilde{s}}(i, j)| \in [0, 1] \quad (1)$$

where $\rho_{s,\tilde{s}}$ is the Pearson correlation coefficient for the pair (s, \tilde{s}) , where \tilde{s} is the GT state, s the learned state, $i \in \llbracket 0, |\tilde{s}| \rrbracket$, $j \in \llbracket 0, |s| \rrbracket$, $\tilde{s} = [\tilde{s}_1; \dots; \tilde{s}_n]$, and \tilde{s}_k being the k^{th} dimension of the GT state vector.

Ground Truth Correlation	x_{robot}	y_{robot}	x_{target}	y_{target}	Mean	Mean Reward
Ground Truth	1	1	1	1	1	243.7 ± 1.2
Supervised	0.69	0.73	0.6	0.61	0.66	243.9 ± 1.8
Random Features	0.59	0.54	0.50	0.42	0.51	201.5 ± 5.7
Robotic Priors	0.1	0.1	0.45	0.54	0.30	-1.1 ± 2.4
Auto-Encoder	0.50	0.54	0.20	0.25	0.37	230.27 ± 3.2
SRL Combination	0.95	0.96	0.22	0.20	0.58	216.8 ± 5.6
SRL Splits	0.98	0.98	0.61	0.73	0.83	237.8 ± 2.1

Table 2: GTC , GTC_{mean} , and mean reward performance in RL (using PPO) per episode after 5 millions steps, with standard error (SE) for each SRL method in 2D Simulated omnibot with a random target environment.

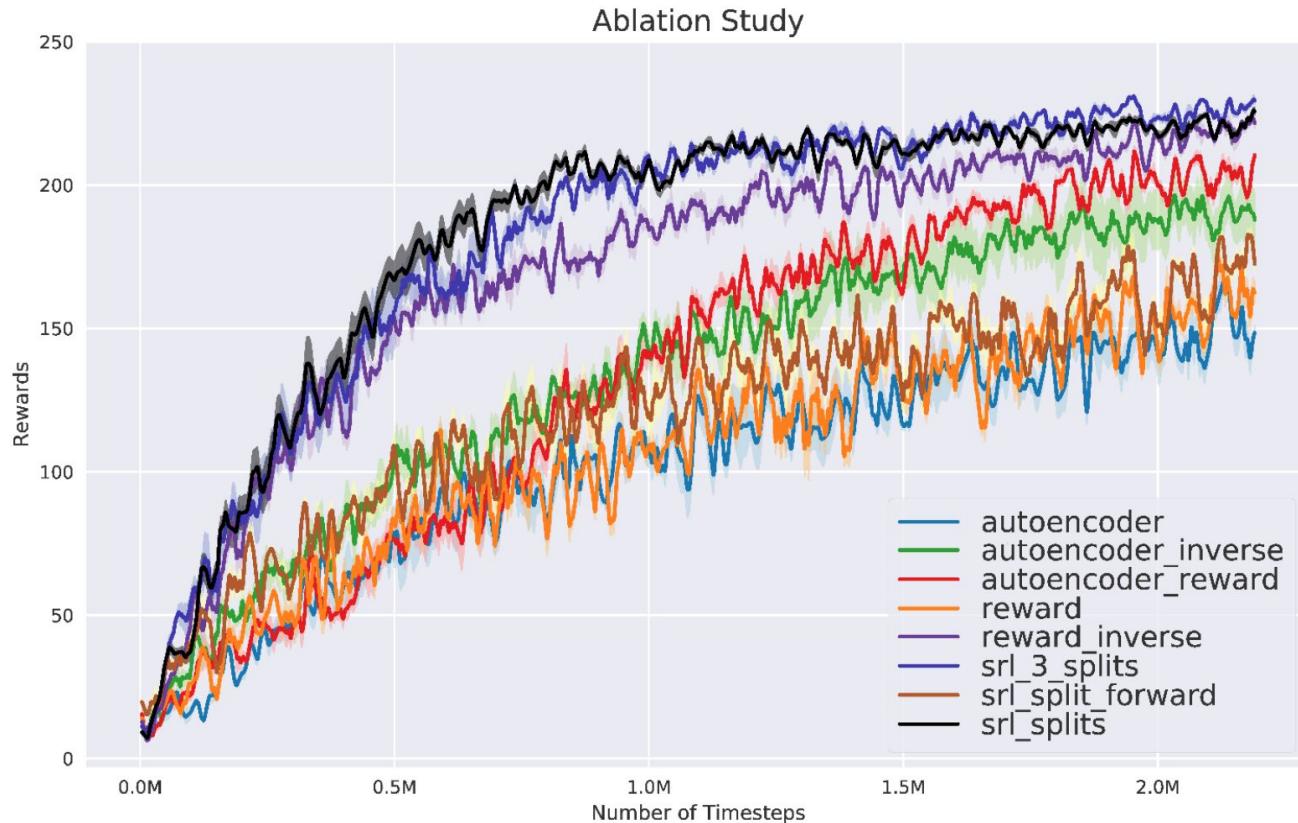


Figure 8: Ablation study of *SRL Splits* (mean and standard error for 10 runs) for PPO algorithm in Navigation 2D random target environment. Models details are explained in Table 10, e.g., *SRL_3_splits* model allocates separate parts of the state representation to each loss (reconstruction/reward/inverse).

Random Features are a great baseline....

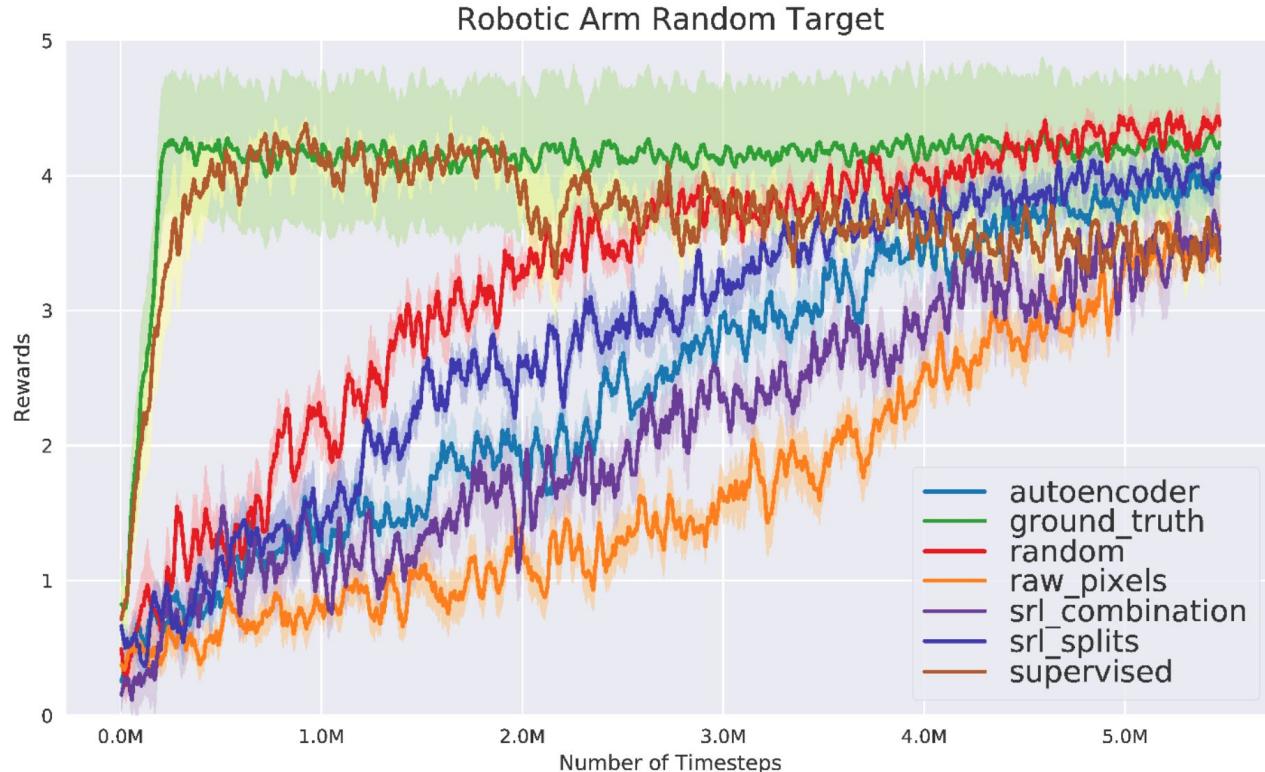


Figure 7: Performance (mean and standard error for 10 runs) for PPO algorithm for different state representations learned in robotic arm with random target environment.

But does not transfer as well...

Sim2Real Transfer:

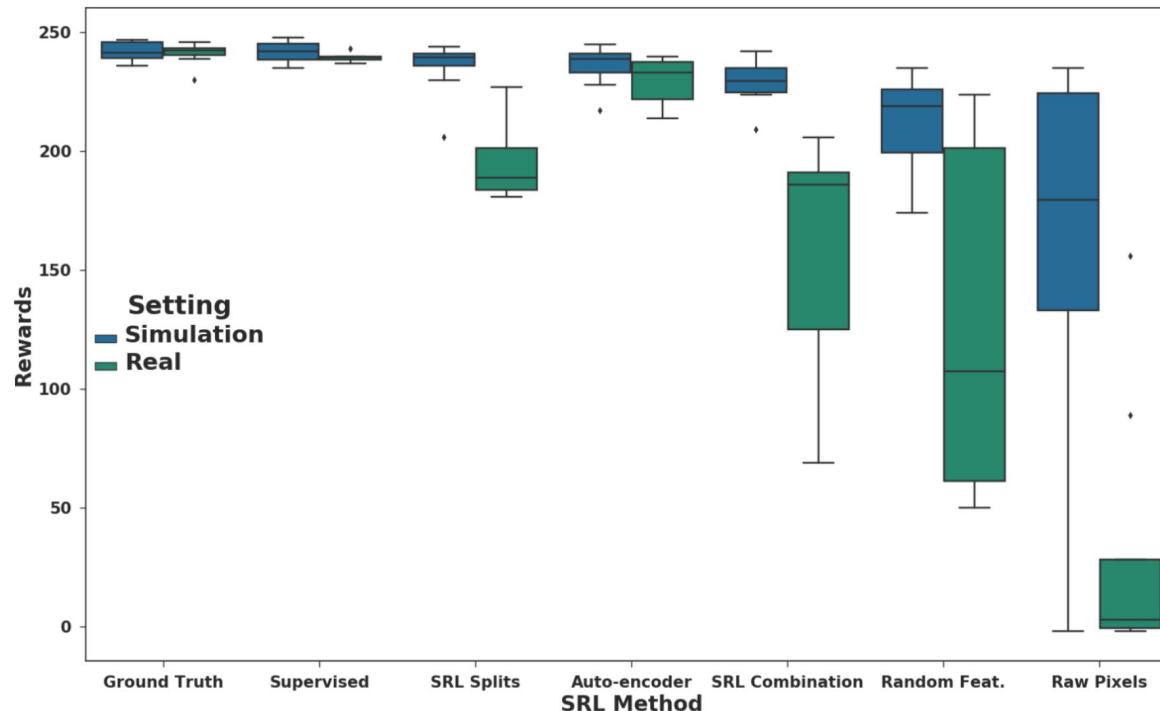
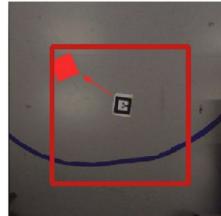


Figure 12: From simulation to real robot: Mean reward and standard deviation for policies trained in simulation (5M steps budget) and replayed in Simulated and Real Omnidot (250 steps, 8 runs).

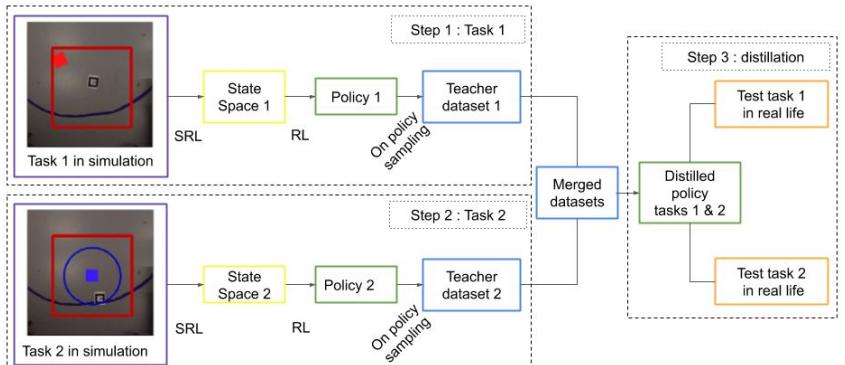
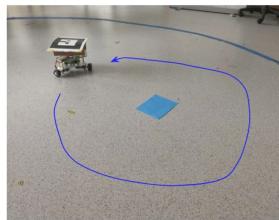
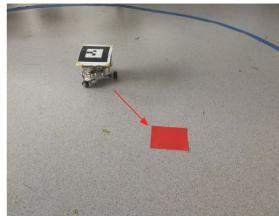
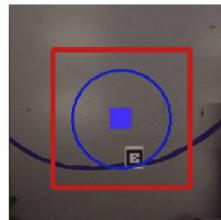
Future Work: Continual RL in Multi-Task, Lifelong, Real Life Settings



Task 1



Task 2



No test time task label

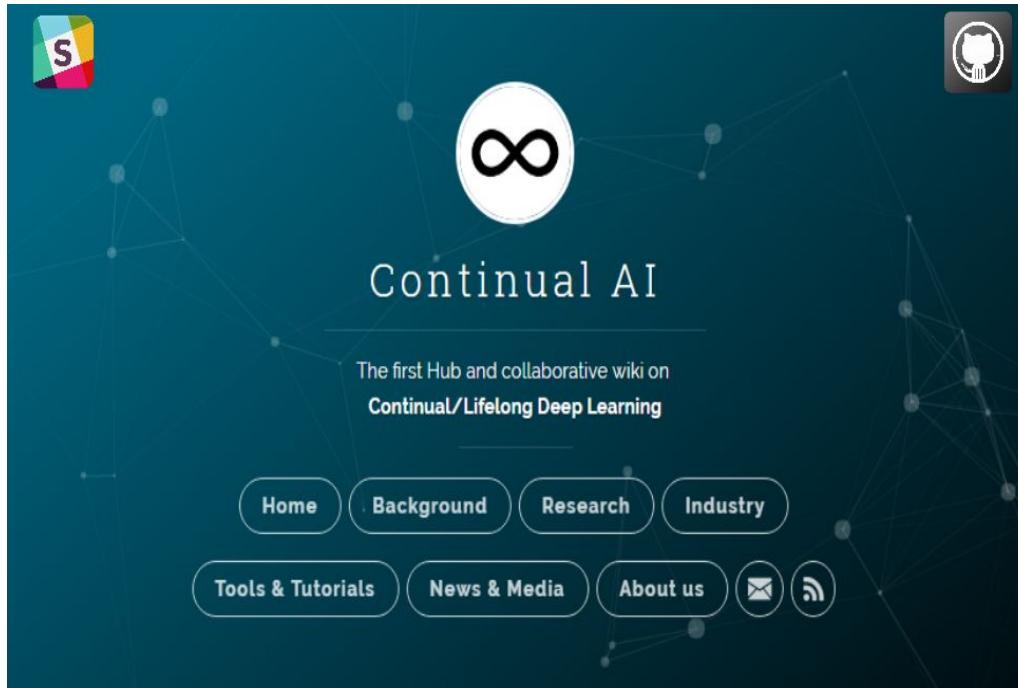
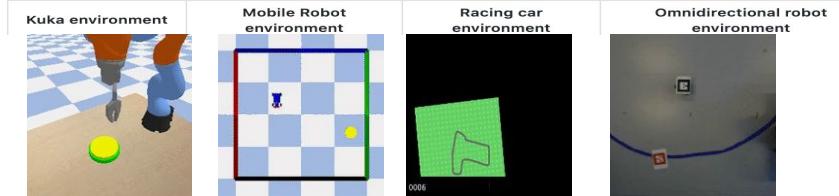
Thank you!

CODE: https://github.com/kalifou/robotics-rl-srl/tree/circular_movement_omnibot

S-RL Toolbox: RL & SRL for Robotics <https://github.com/araffin/robotics-rl-srl>

natalia.diaz@ensta-paristech.fr <https://nataliadiaz.github.io/>

@NataliaDiazRodr



The Continual AI website features a dark teal background with a network graph overlay. In the center is a large white circle containing a black infinity symbol. Below it, the text "Continual AI" is displayed in a white serif font. To the left, there is a logo consisting of a stylized letter 'S' inside a square divided into four colored quadrants (blue, pink, yellow, and light blue). On the right side, there is a GitHub icon. At the bottom, there are several navigation links: "Home", "Background", "Research", "Industry", "Tools & Tutorials", "News & Media", "About us", an envelope icon, and a Wi-Fi icon.

The first Hub and collaborative wiki on
Continual/Lifelong Deep Learning

Join! <https://www.continualai.org/>
Slack channel: <https://continualai.herokuapp.com/>