ENSTA Autonomous Systems and Robotics
Computer Science and System Engineering Laboratory

INSTITUT POLYTECHNIQUE DE PARIS

ContinualAI

DREAM

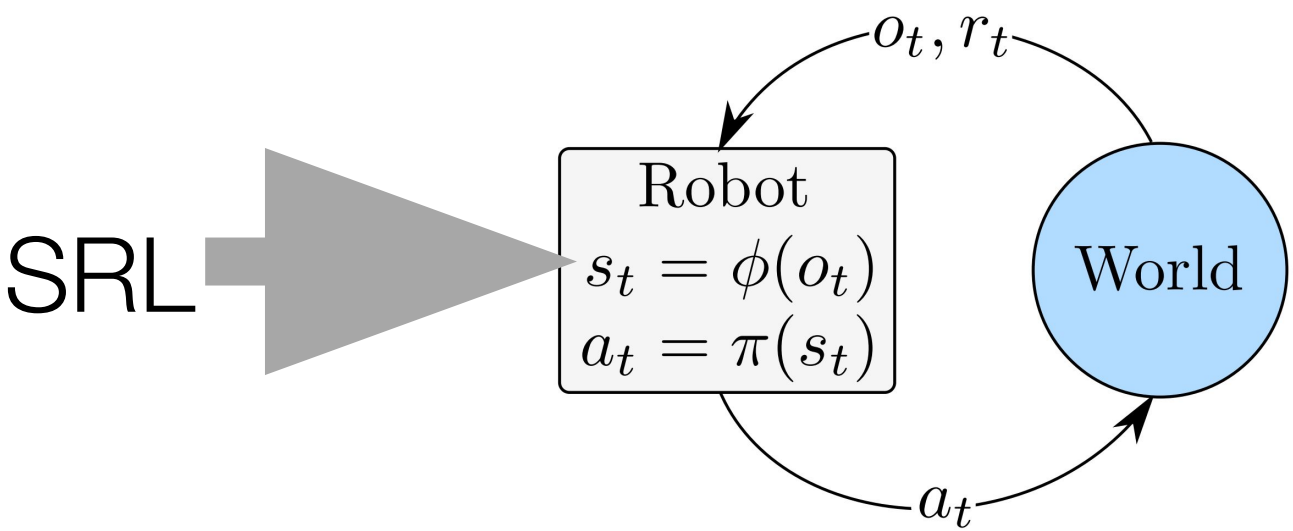
www.robotsthatdream.eu

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

Natalia Díaz Rodríguez, PhD

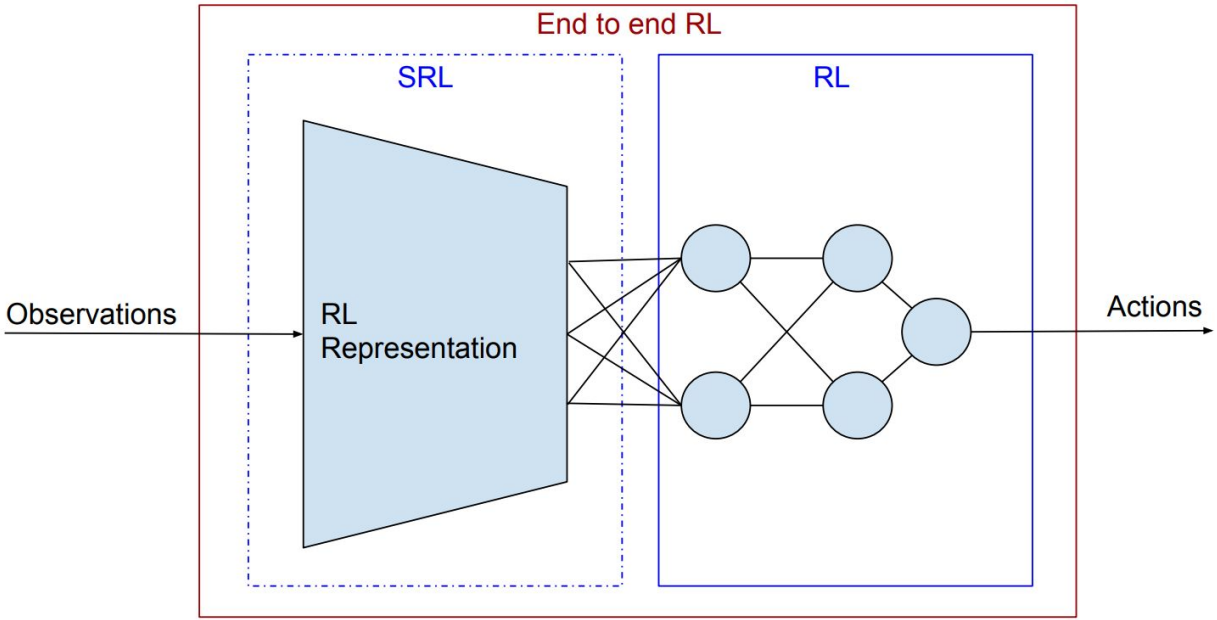
Antonin Raffin, Ashley Hill, René Traoré, Timothée Lesort, Natalia Díaz-Rodríguez, David Filliat
U2IS & INRIA FLOWERS Team, ENSTA ParisTech, Palaiseau, France

State Representation Learning (SRL) in RL context



[Lesort'18, State representation learning for control: An overview. Neural Networks]

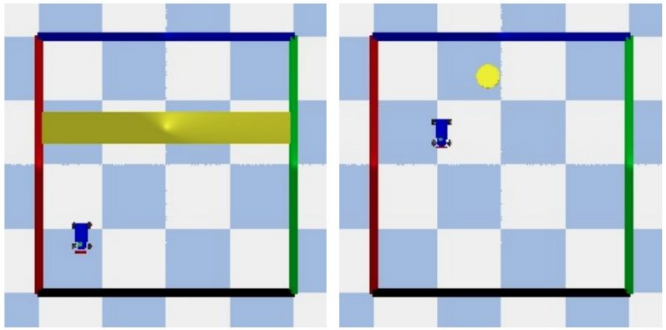
State Representation Learning (SRL) in RL context

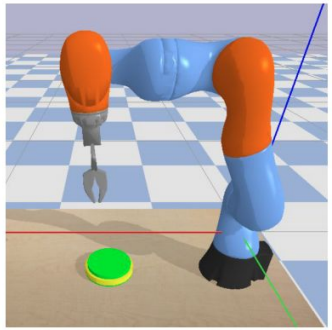


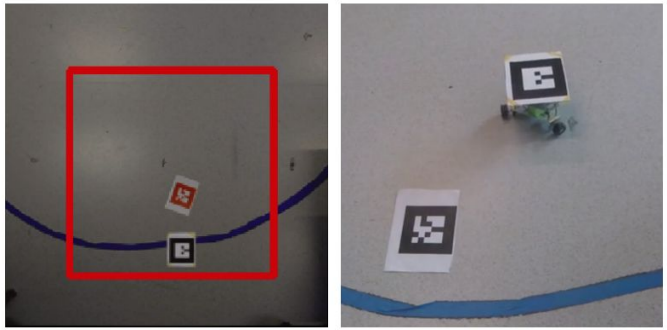
[Lesort'18, State representation learning for control: An overview. Neural Networks]

Goal based robotic tasks

Mobile Navigation	Robotic Arm	Simulated & Real Omnibot

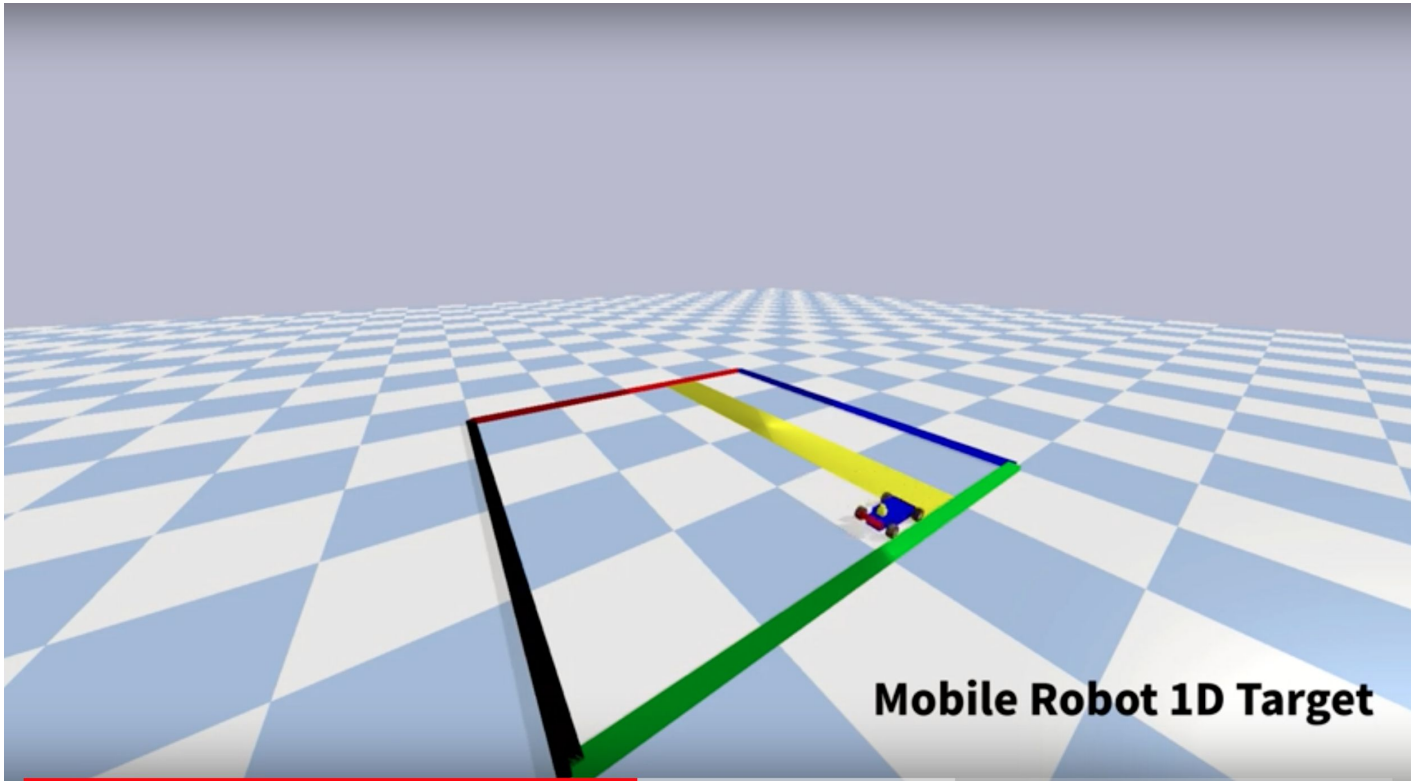




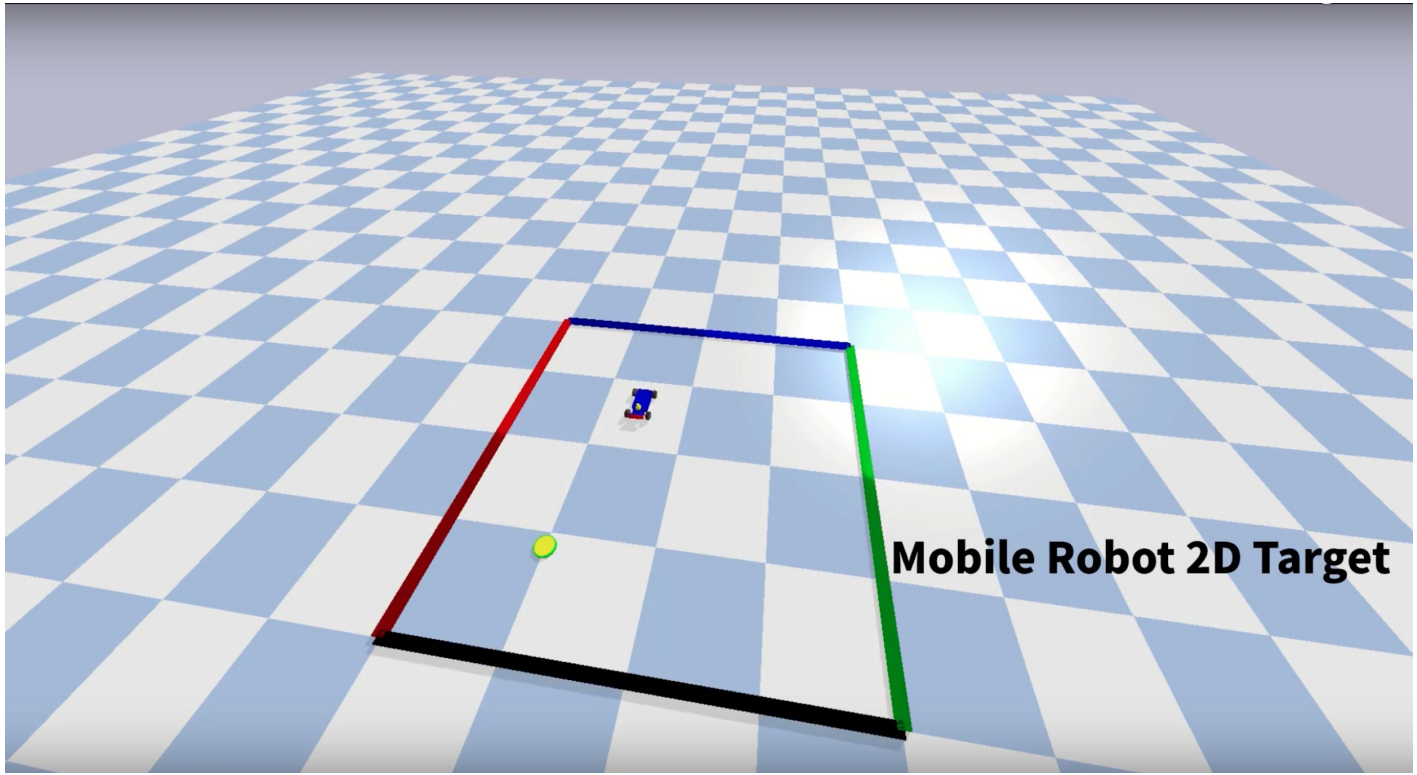


Environments, code and data are available at `https://github.com/araffin/robotics-rl-srl`.

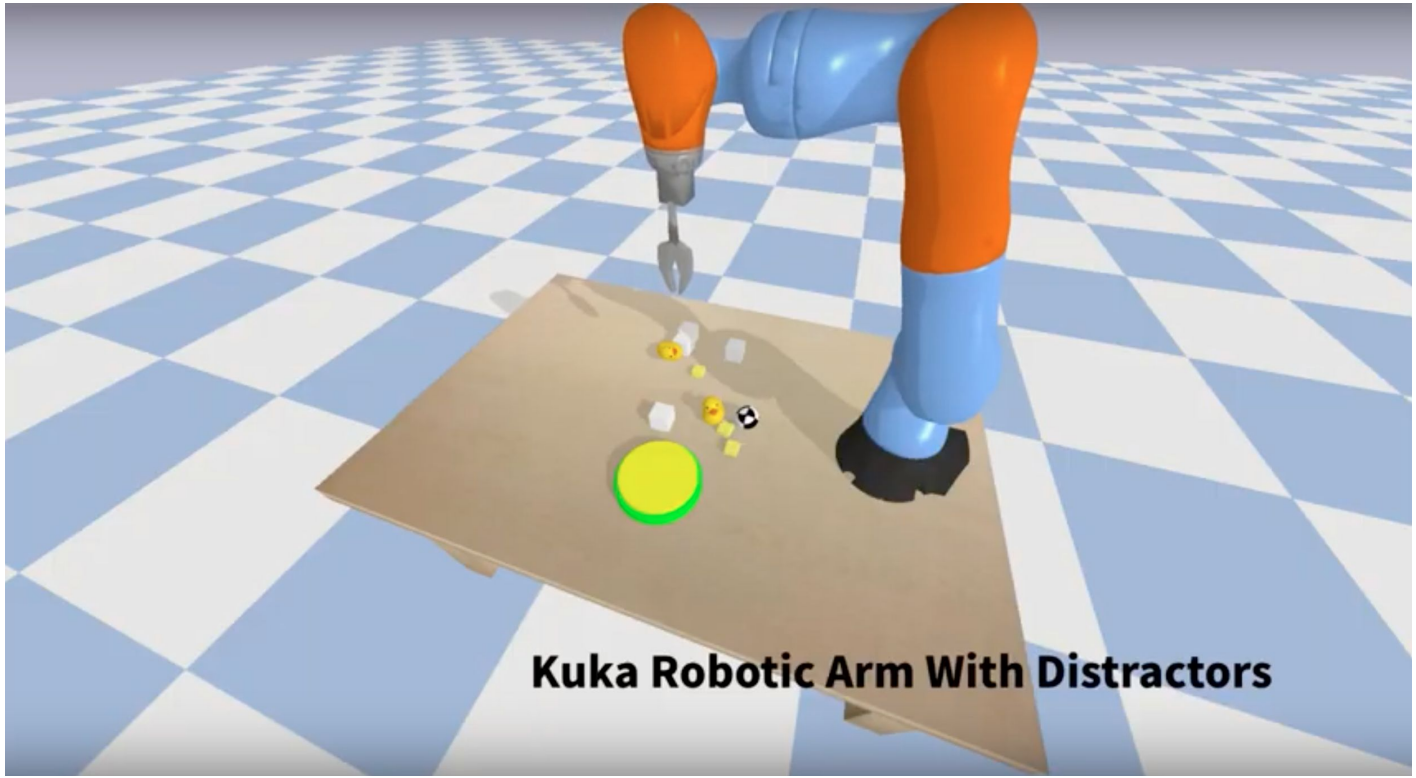
S-RL Toolbox



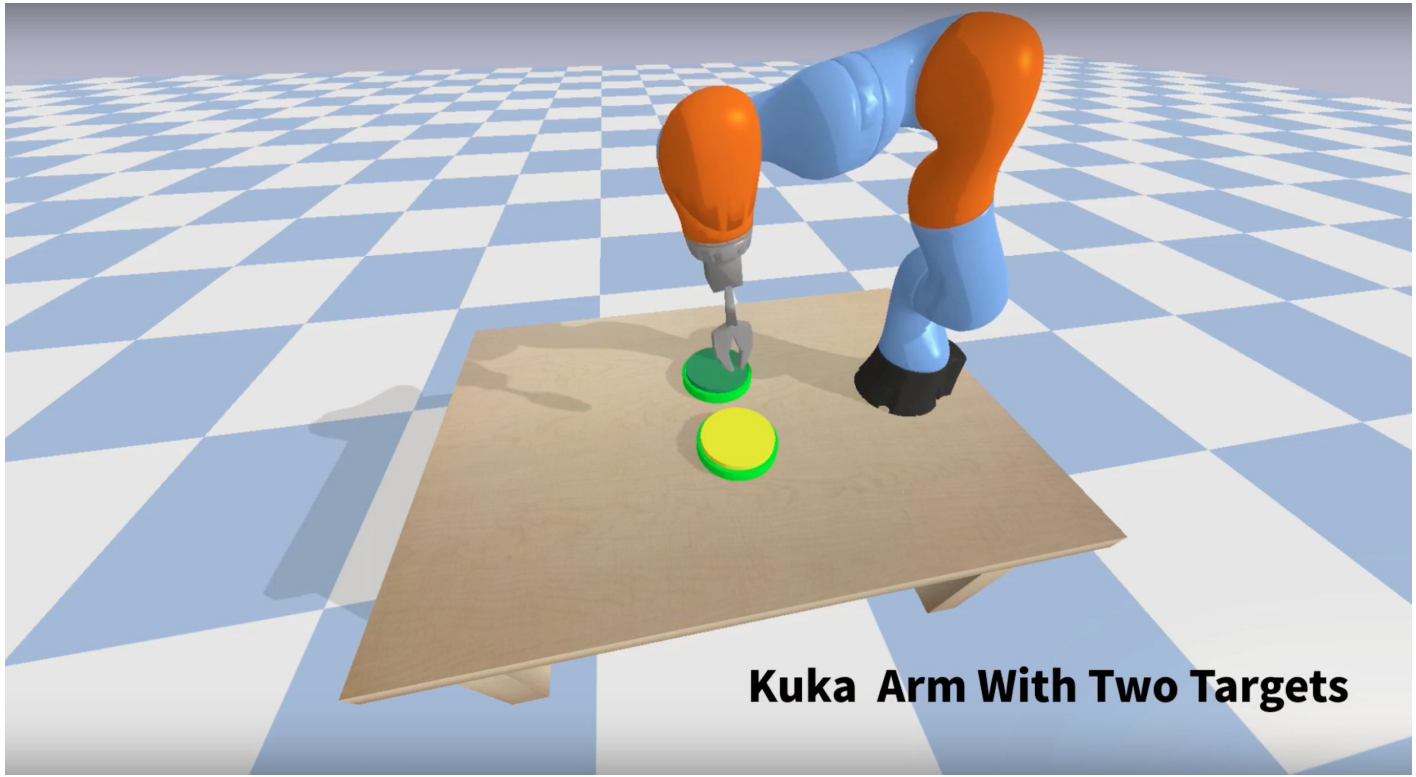
S-RL Toolbox



S-RL Toolbox

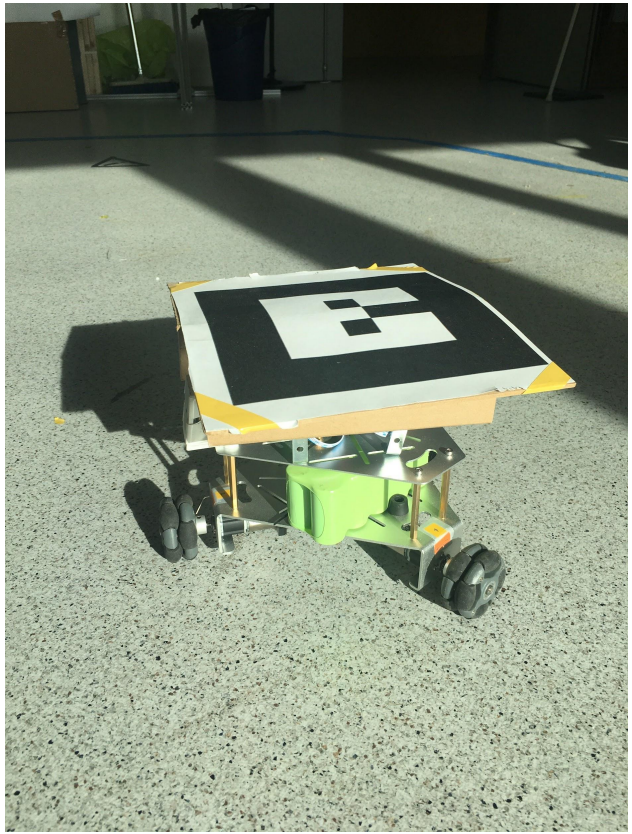


S-RL Toolbox

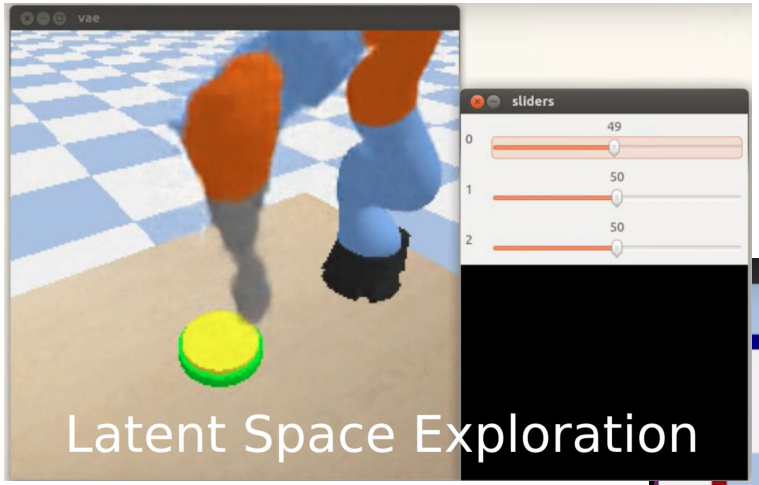


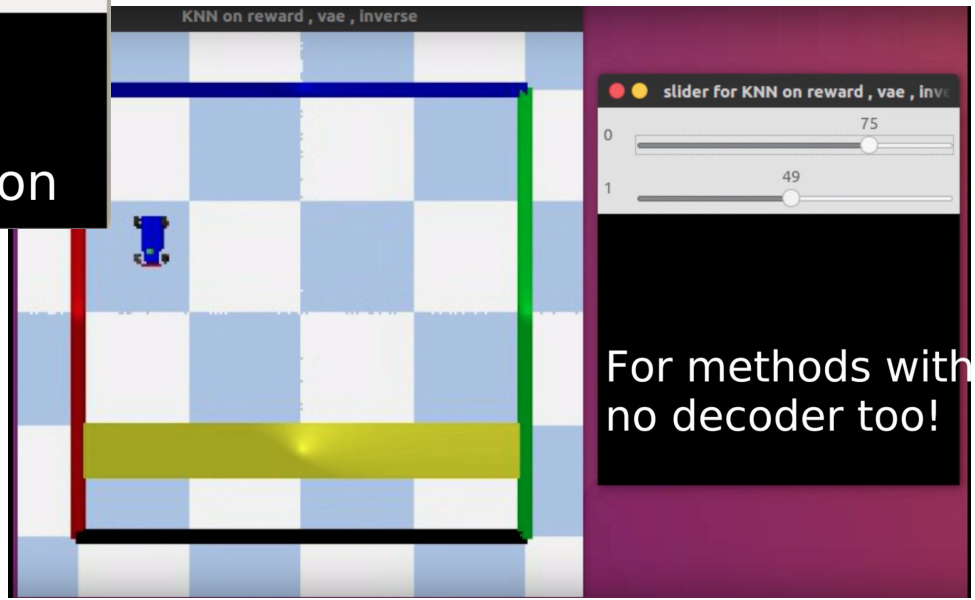
S-RL Toolbox

S-RL Toolbox

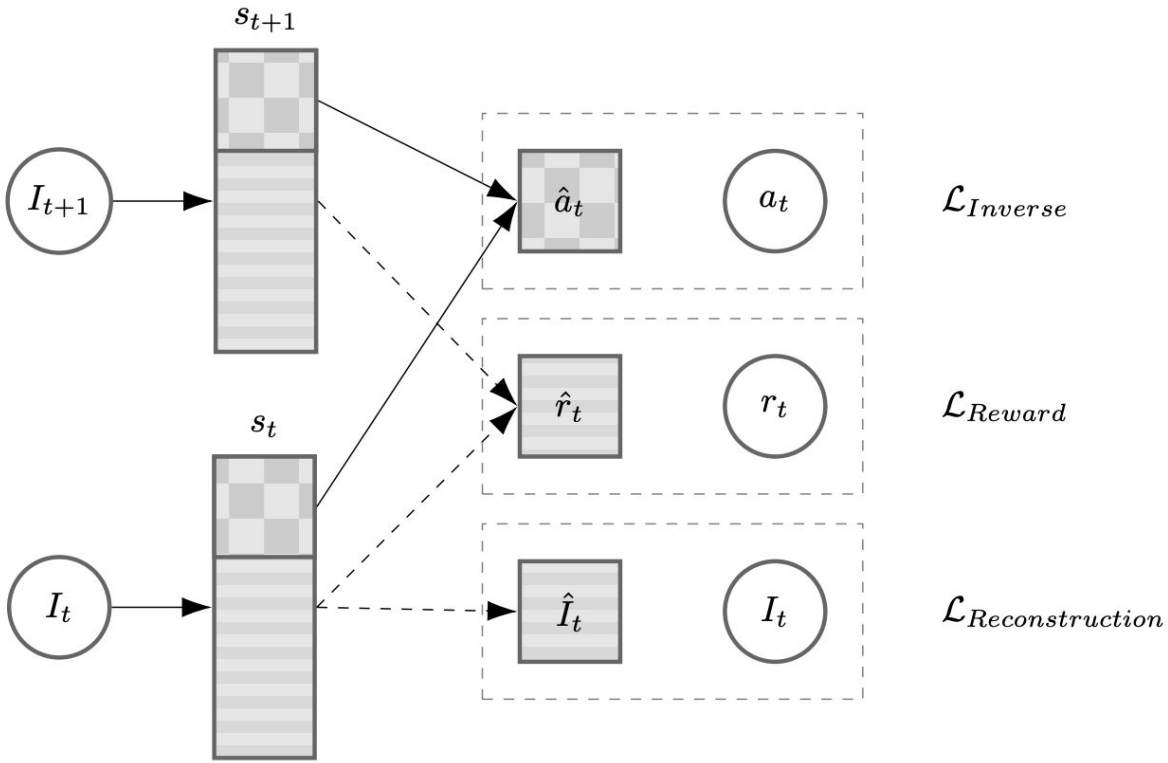


S-RL Toolbox


vae
sliders
0
49
1
50
2
50
Latent Space Exploration


KNN on reward , vae , inverse
slider for KNN on reward , vae , inve
0
75
1
49
For methods with
no decoder too!

Decoupling feature extraction from policy learning

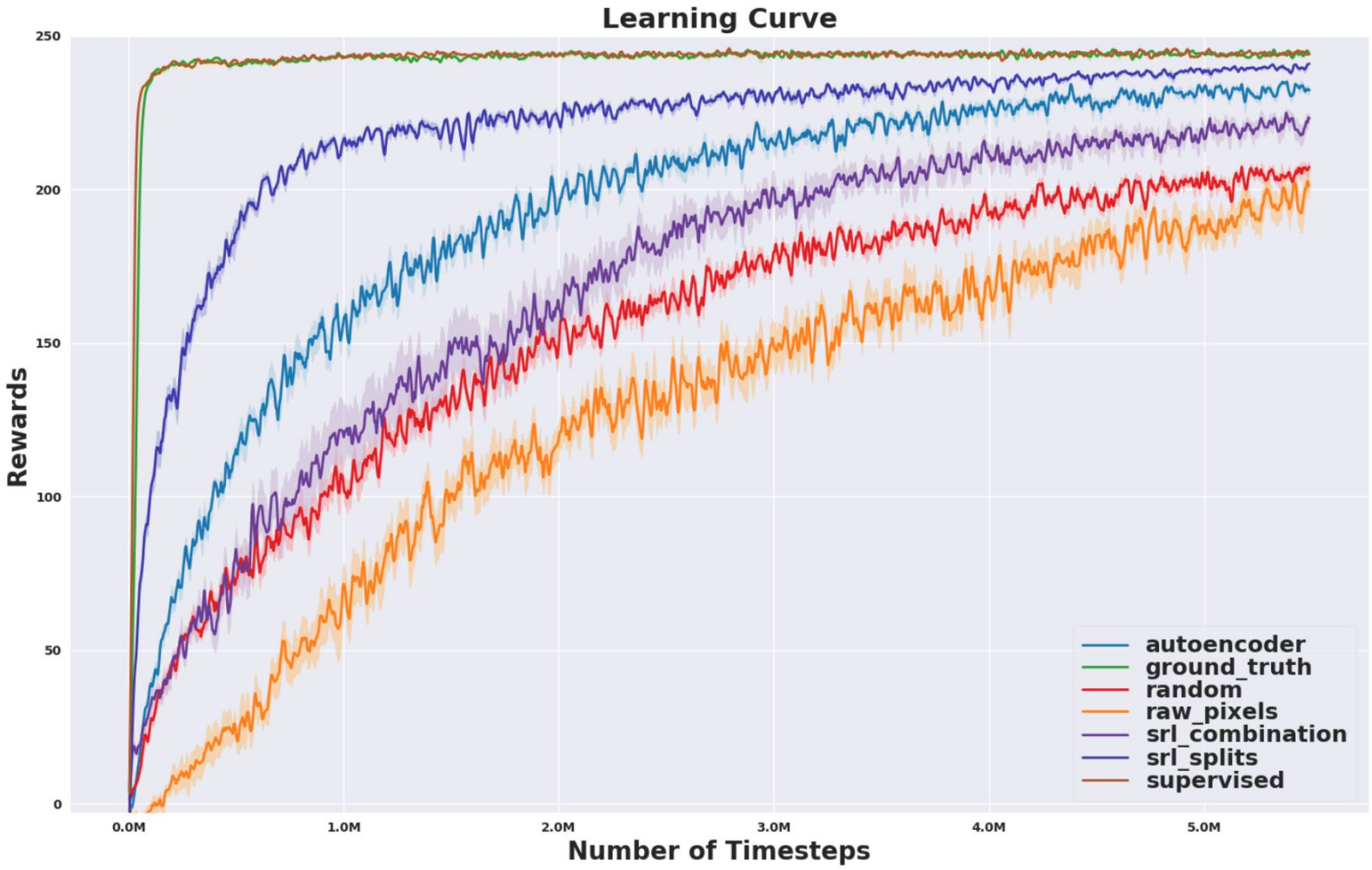


[Raffin et al.18]

Baselines and Results

Environments	*Nav. 1D Target*	*Nav. 2D Target*	*Arm Random Target*	*Pseudo-Real Omnibot*
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.


Learning Curve
Rewards
Number of Timesteps
250
200
150
100
50
0
0.0M
1.0M
2.0M
3.0M
4.0M
5.0M
autoencoder
ground_truth
random
raw_pixels
srl_combination
srl_splits
supervised

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

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, $i \in [\![0, |\tilde{s}|]\!]$, $j \in [\![0, |s|]\!]$, $\tilde{s} = [\tilde{s}_1; ...; \tilde{s}_n]$, 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, $i \in [\![0, |\tilde{s}|]\!]$, $j \in [\![0, |s|]\!]$, $\tilde{s} = [\tilde{s}_1; ...; \tilde{s}_n]$, 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.

$w_{reconstruction}$	w_{reward}	$w_{inverse}$	*Mean Reward*
1	1	1	225.2 ± 6.3
1	1	10	223.5 ± 8.0
1	10	10	215.1 ± 7.1
1	10	5	217.8 ± 12.2
1	5	1	217.8 ± 6.7
1	5	10	228.8 ± 4.2
5	1	1	221.0 ± 7.4
5	1	10	209.1 ± 19.5
5	10	10	226.3 ± 5.2
5	5	1	194.6 ± 14.6
5	5	10	224.5 ± 5.5
10	1	1	176.5 ± 16.2
10	1	10	218.9 ± 8.0
10	1	5	182.4 ± 15.8
10	5	10	225.5 ± 5.7
10	5	5	210.2 ± 8.3

Table 9: Influence of the weights on the SRL Splits model performance, Navigation 2D random target environment.

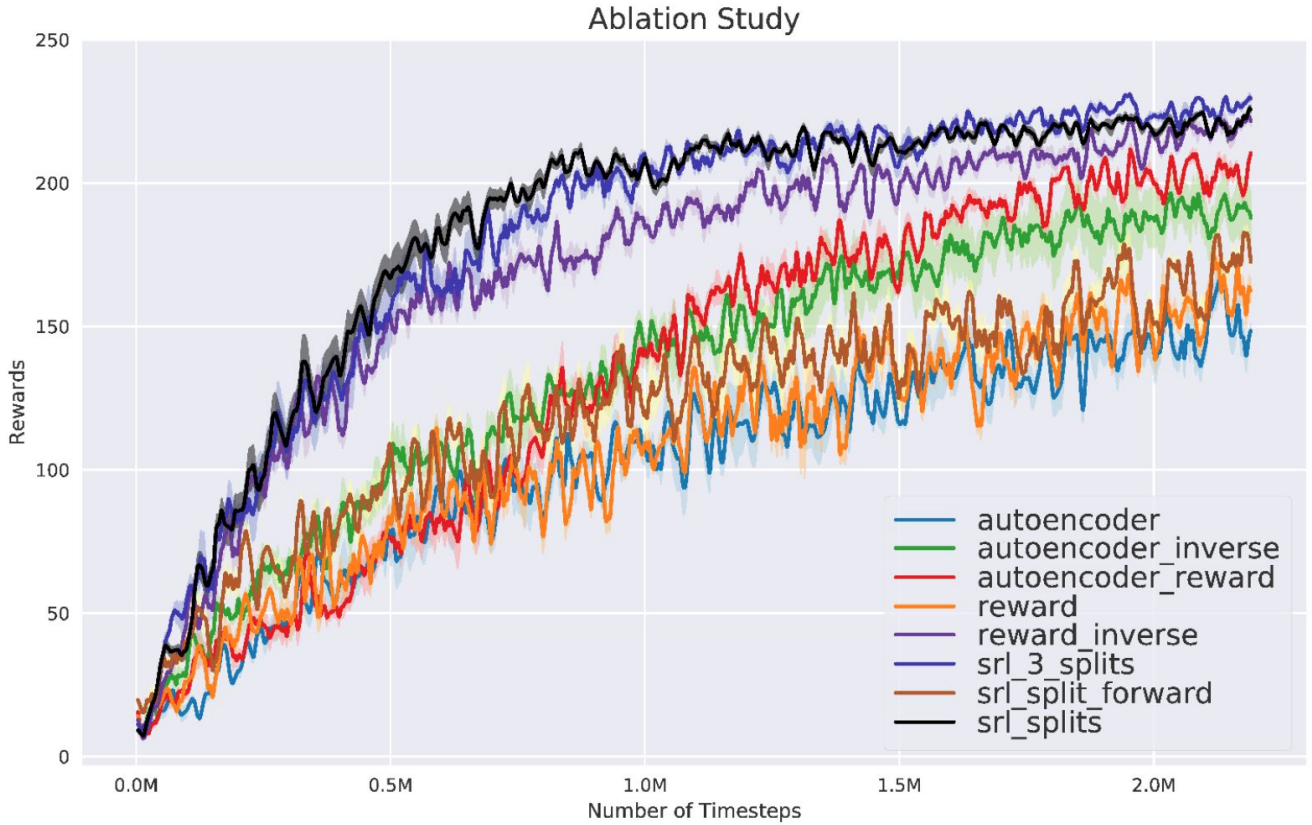


Figure 8: Ablation study of *SRL Splits* (mean and standard error for 10 runs) for PPO algorithm in Navigation 2D random target environment. Models details are explained in Table 10, e.g., *SRL_3_splits* model allocates separate parts of the state representation to each loss (reconstruction/reward/inverse).

Random Features are a great baseline….

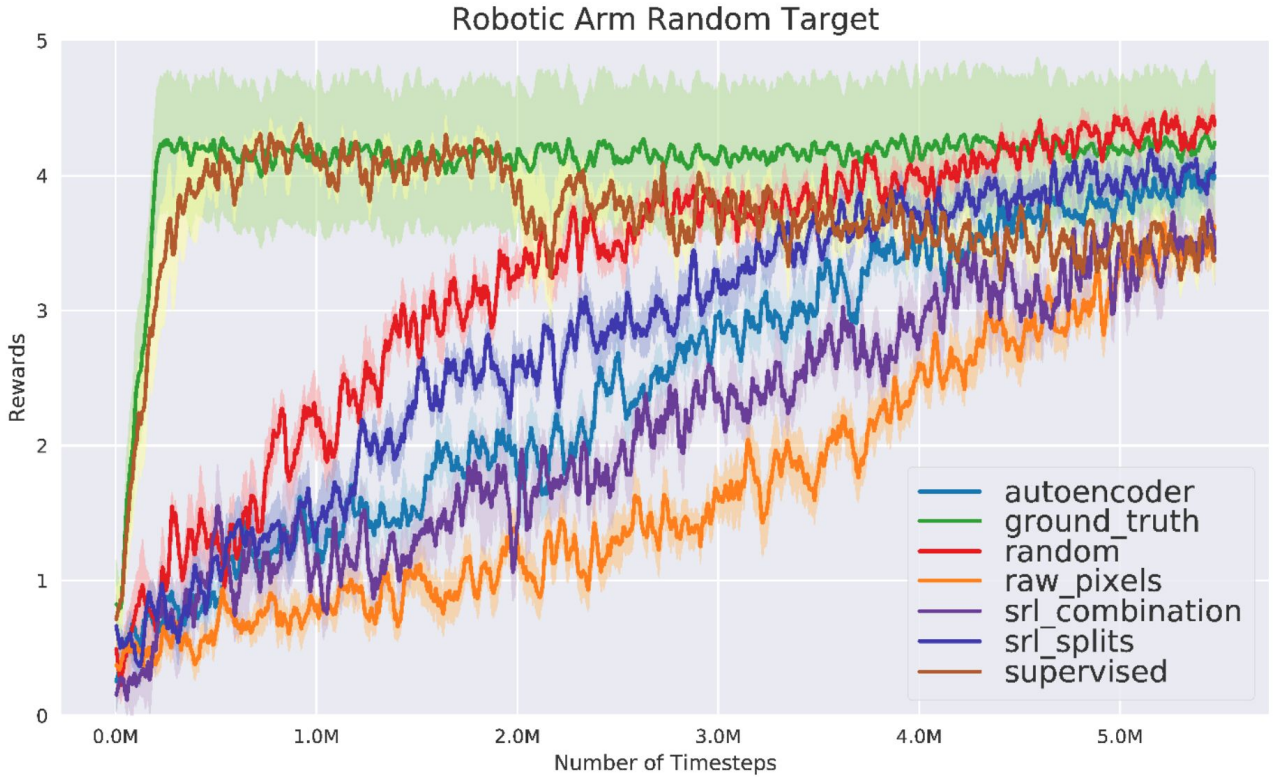


Figure 7: Performance (mean and standard error for 10 runs) for PPO algorithm for different state representations learned in robotic arm with random target environment.

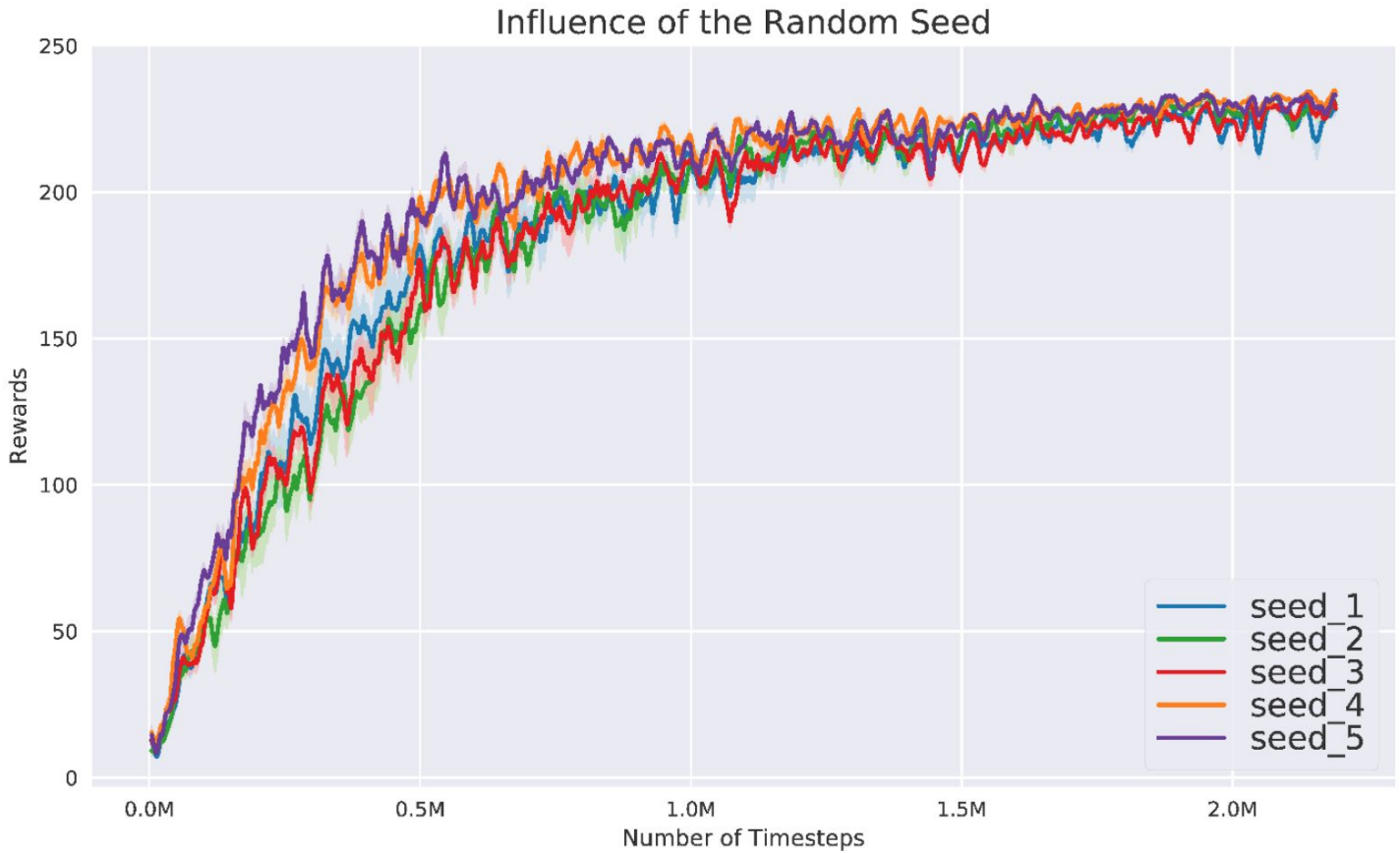


Figure 9: Influence of random seed (mean and standard error for 10 runs) for PPO algorithm for SRL Splits in Navigation 2D random target environment

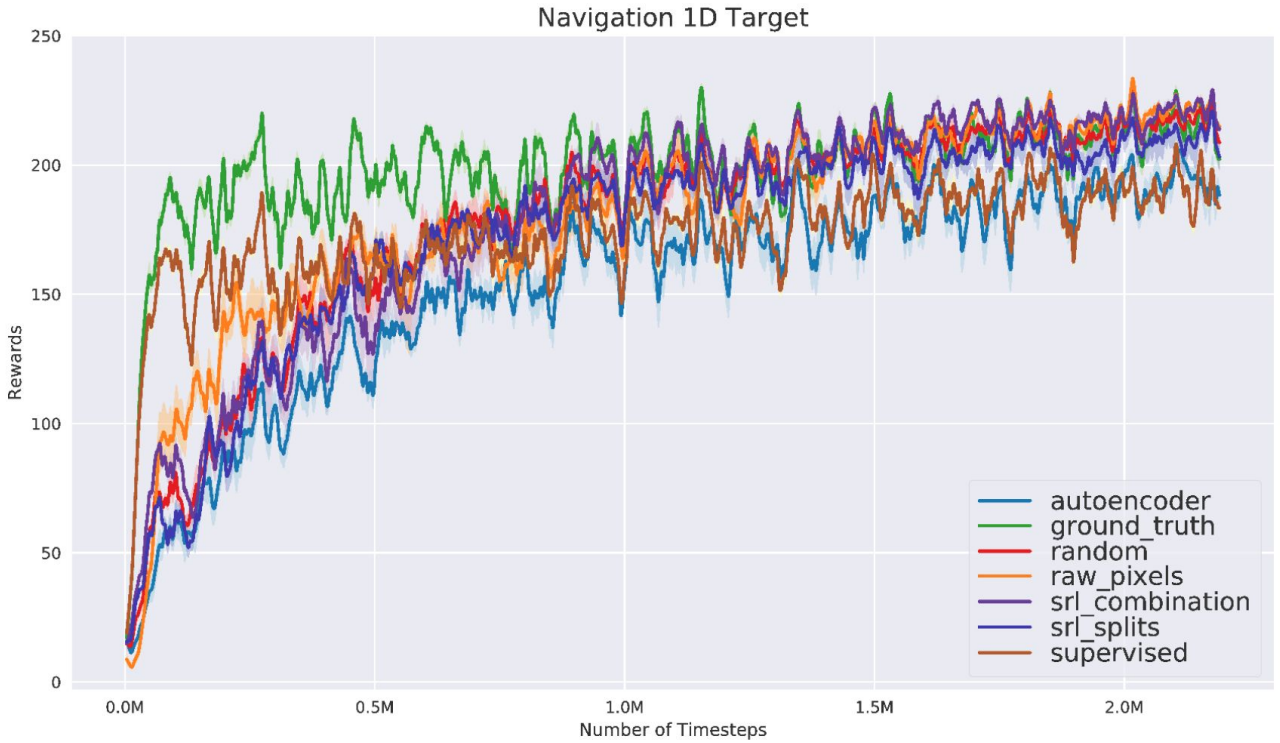


Figure 4: Performance (mean and standard error for 10 runs) for PPO algorithm for different state representations learned in Navigation 1D target environment.

But does not transfer as well…
Sim2Real Transfer:

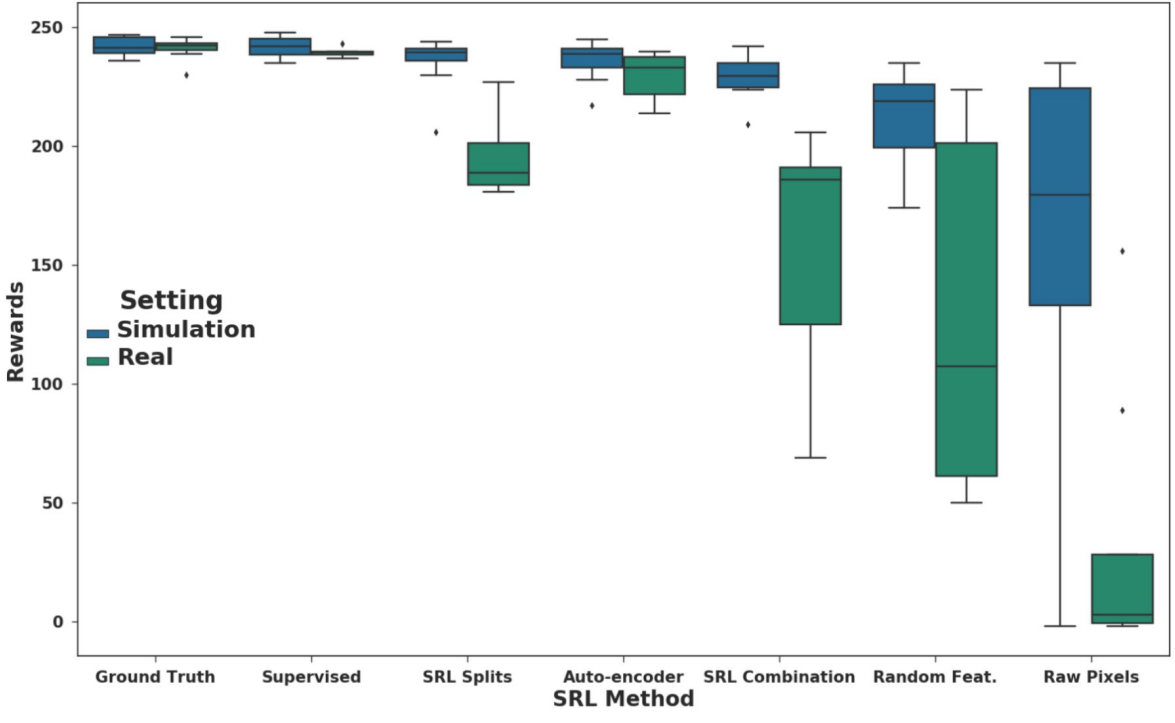
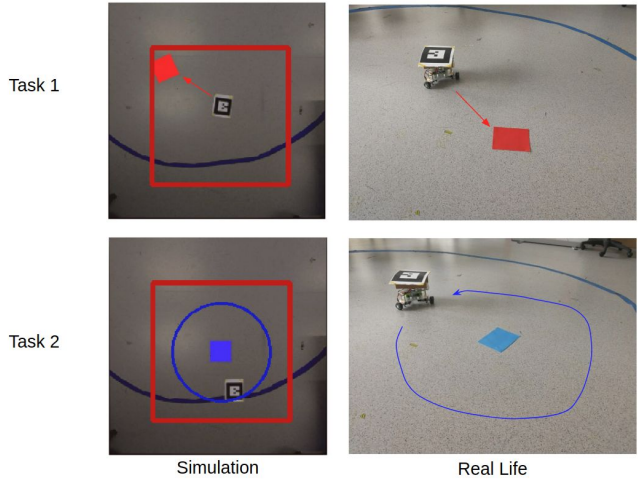
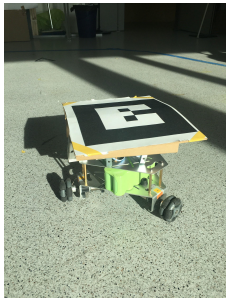
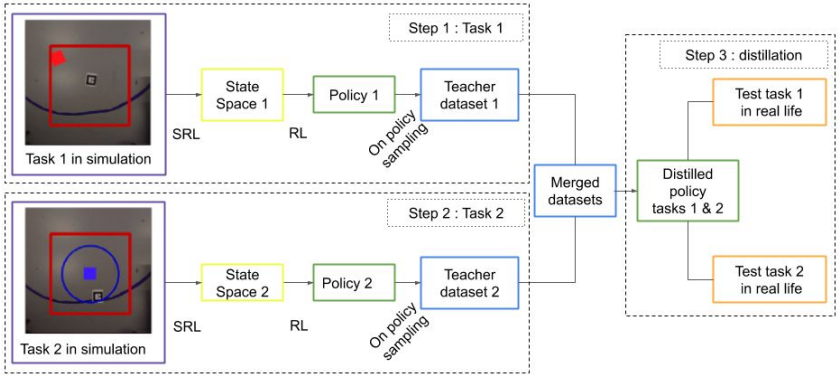


Figure 12: From simulation to real robot: Mean reward and standard deviation for policies trained in simulation (5M steps budget) and replayed in Simulated and Real Omnibot (250 steps, 8 runs).

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







No test time task label

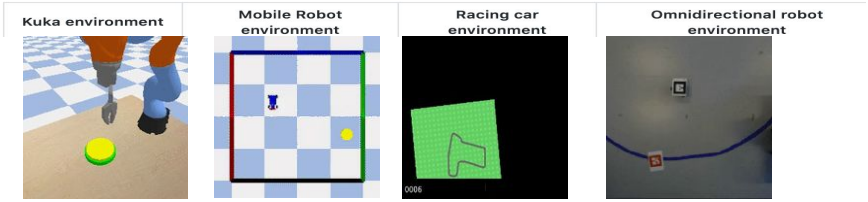
Continual Reinforcement Learning deployed in Real-life using Policy Distillation and Sim2Real Transfer. [Traoré'19]
State representation learning for control: An overview, Neural Networks, [Lesort'18]

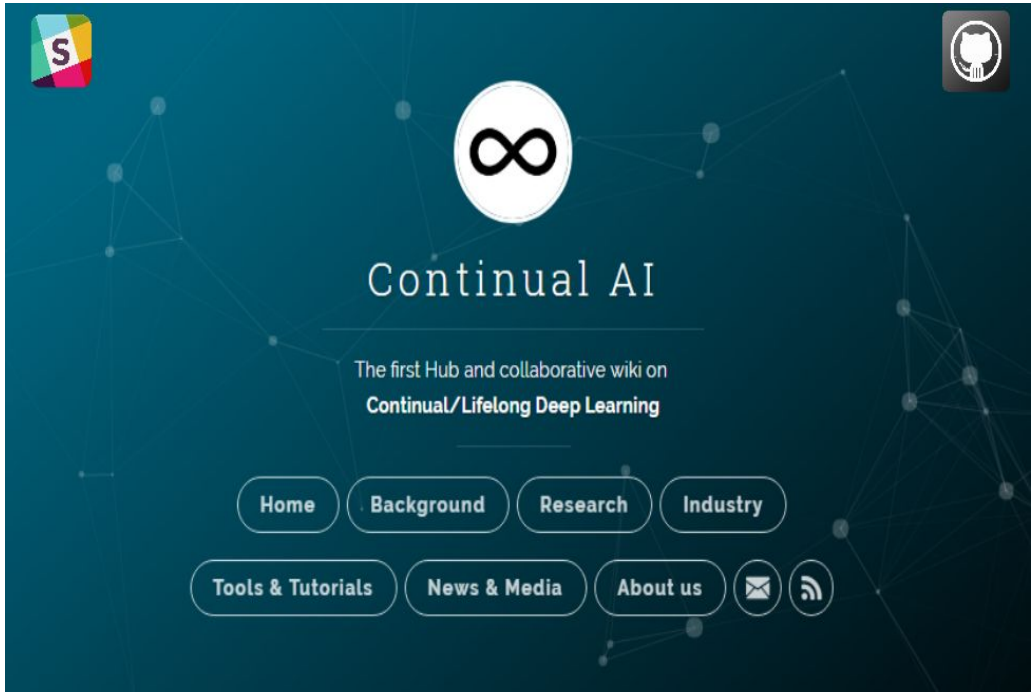
Thank you!

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





Join! https://www.continualai.org/

Slack channel: https://continualai.herokuapp.com/