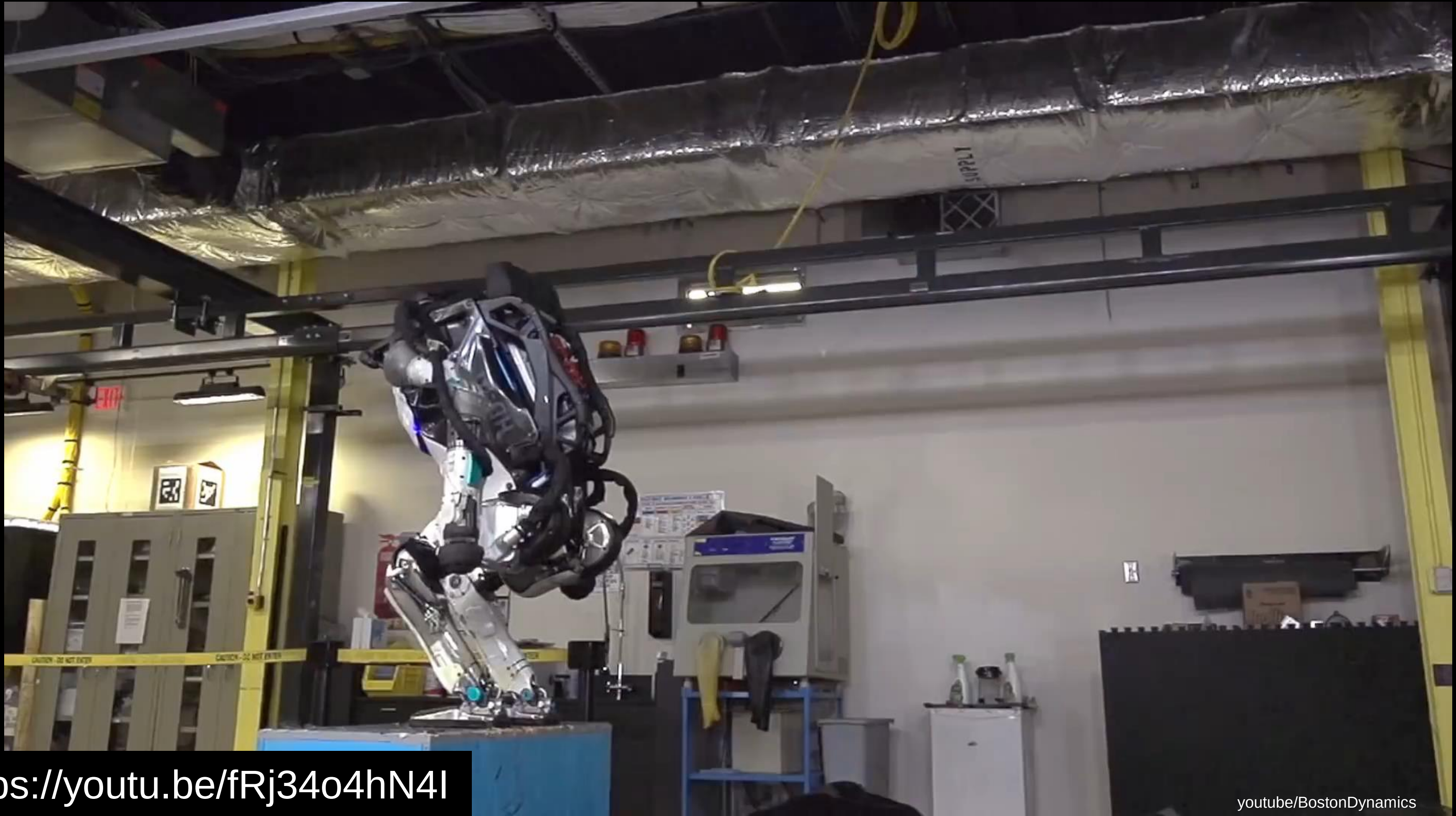
Robots Learning (Through) Interactions

May 7th 2020

Jens Kober

TU Delft

Advanced Robot Hardware + Manual Programming



https://youtu.be/fRj34o4hN4I

What Makes Tasks Hard?

- Complex dynamics
- Uncertainties and variations
  - Objects
  - Environment
  - Tasks
  - Human behaviour
- Occurring in all robot domains!

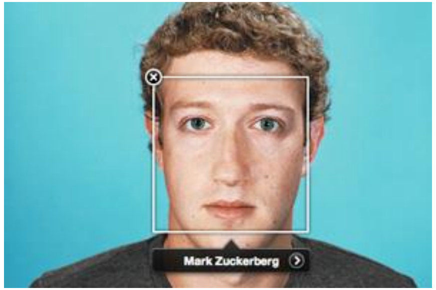
Learning to Interact



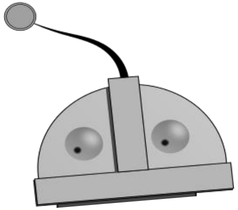
TUDelft

Why is Robotics Different?

IBM Research



sociable.co


openclipart.org


europe1.fr

DeepMind

(Reinforcement) Learning in Robotics

Mahadevan & Connell, AI 1992

Schaal, NIPS 1996

Rottmann et al., IROS 2007

Kormushev & Calinon, IROS 2010

- Safe with real robots
- Fast learning
  - Sample efficient – “small” data
  - Incorporate prior knowledge
  - Few open parameters
- Real time computations

Domain appropriate methods needed!

Google

Kober, Bagnell & Peters, IJRR 2013

Naïve Deep RL

Tim de Bruin

https://youtu.be/vv85S4Z-ZG0

de Bruin, Kober, Tuyls, Babuška, NIPS-WS 2015.

Imitation Learning

- Teacher demonstrates skill, student tries to mimic

testmyprep.com

Reinforcement Learning

- Practice, practice, practice…

cnn.com

incaavalanche.com

How to Learn?

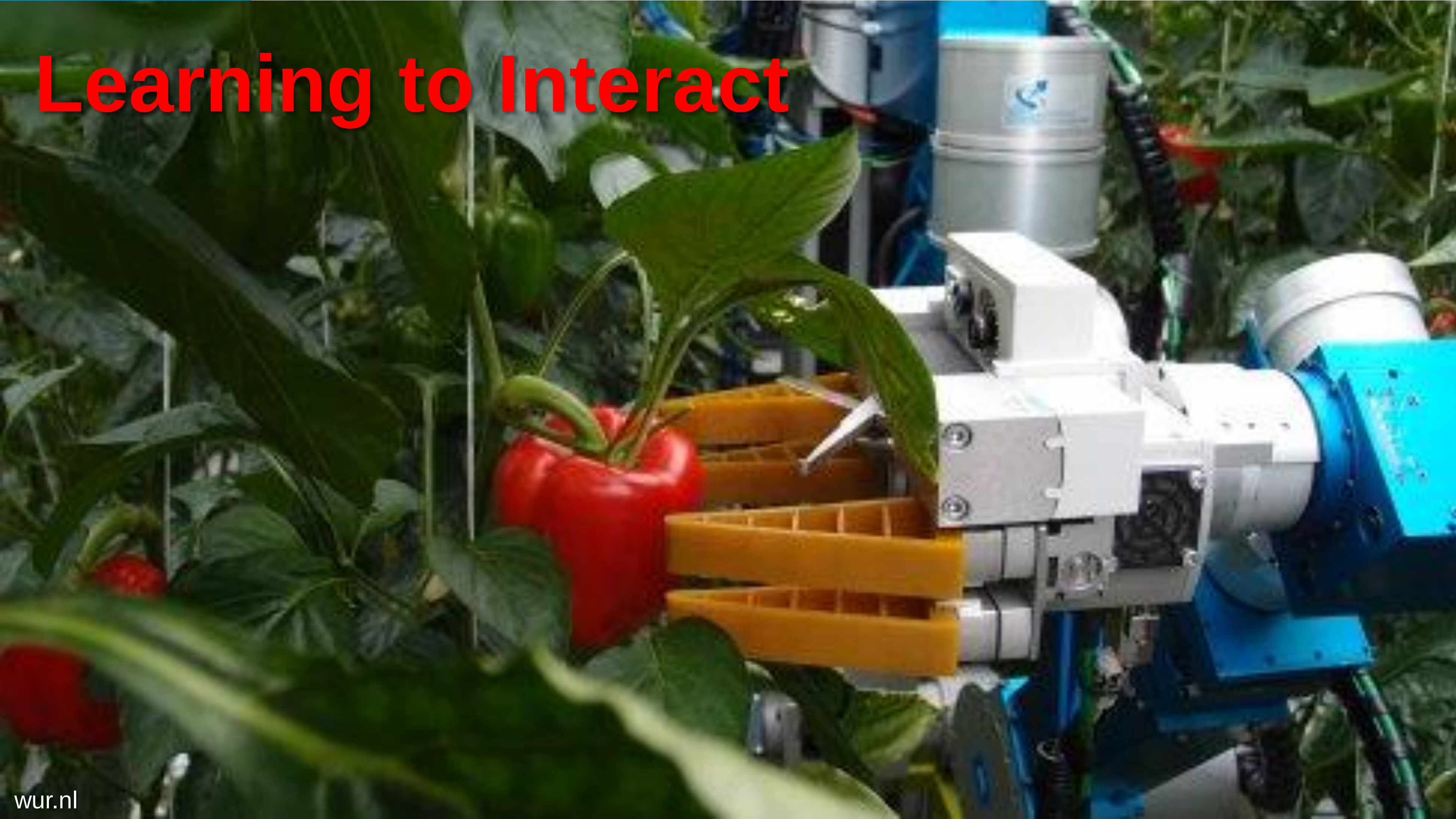
- Continued student-teacher interaction
  - Additional demonstrations
  - Intermittent feedback
- Largely missing in robot learning!
- Benefits
  - Speed-up
  - Complex tasks
  - Intuitive

Interacting to Learn

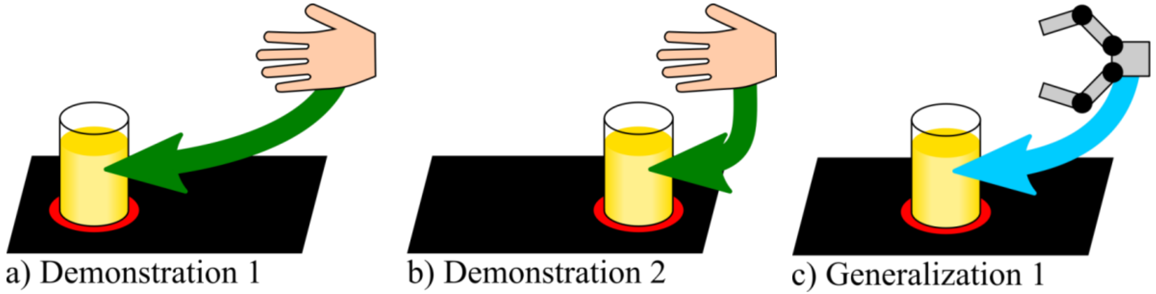

karinprinsloo.com


brainbalancecenters.com

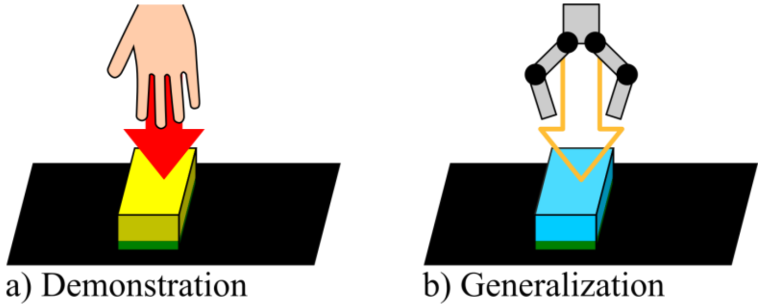
Learning to Interact



Imitation Learning



- Combinatorial explosion
- Force → Position? Position → Force?



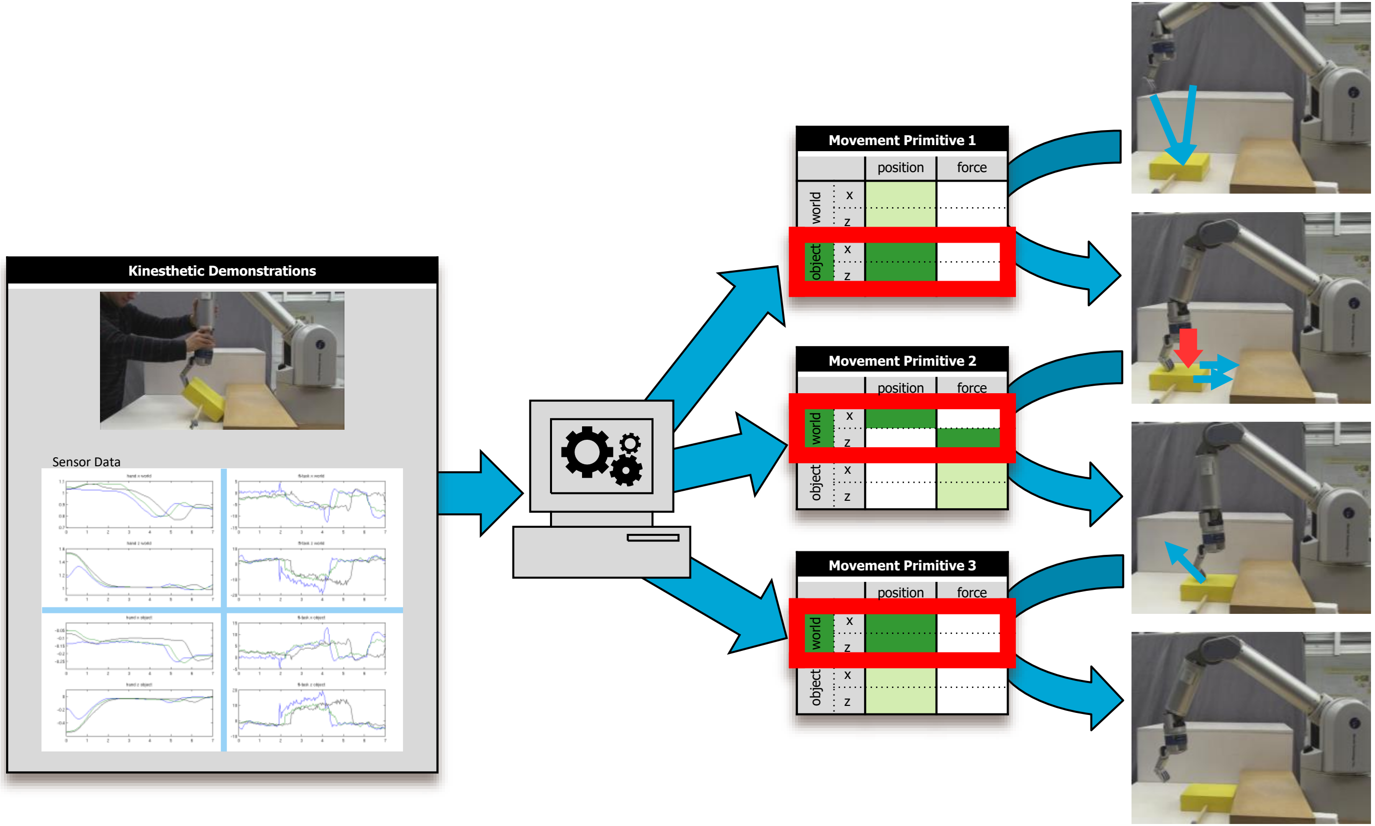
https://youtu.be/t_ZoiKcEM0M

Kinesthetic teaching

Skill reproduction

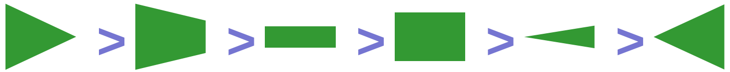
Kober, Gienger, & Steil, ICRA 2015

Primitives within a Sequence

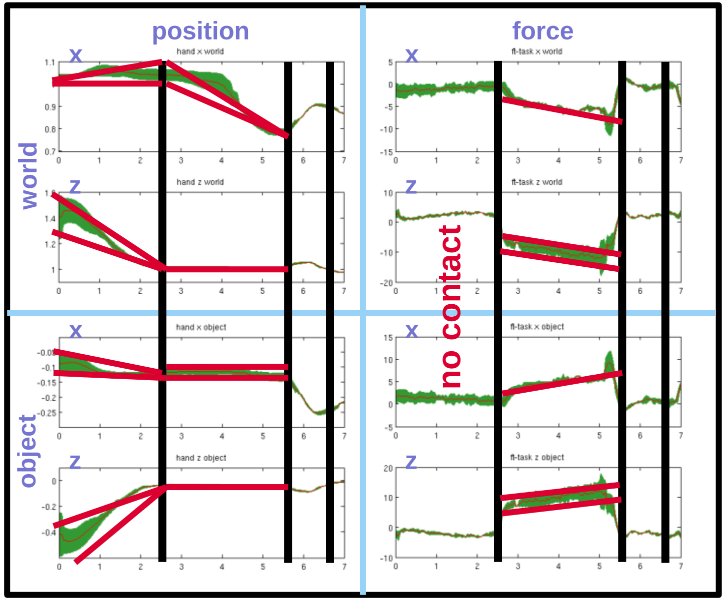


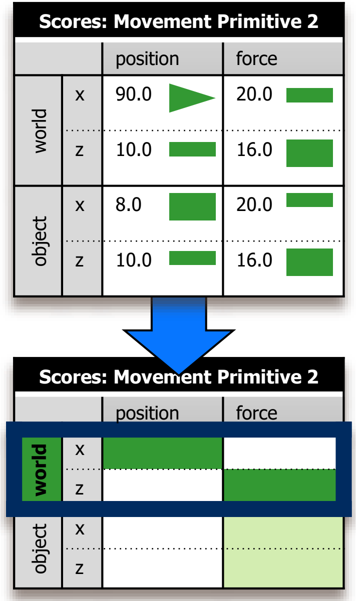
Primitives within a Sequence

Score per MP/frame/component/modality based on statistics & contact



converging/attractor constant diverging





Kober, Gienger, & Steil, ICRA 2015

TU Delft

https://youtu.be/t_ZoiKcEM0M

Kober, Gienger, & Steil, ICRA 2015

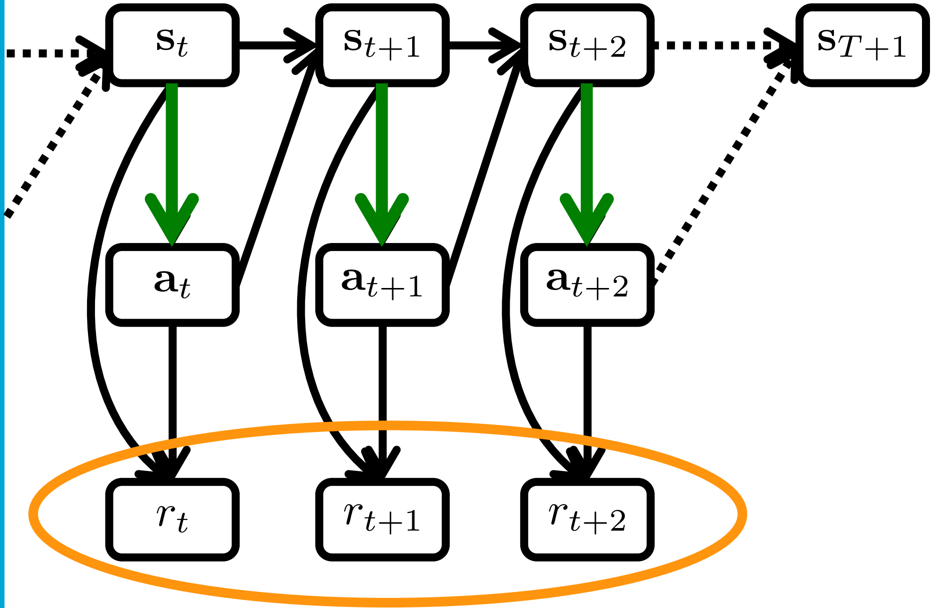
https://youtu.be/WCayQ8xIiU8

Simon Manschitz

Manschitz, Kober, Gienger, & Peters, RAS 2015

Reinforcement Learning

Sutton & Barto 1998



states
$\mathbf{s}_{1:T+1}$

actions
$\mathbf{a}_{1:T}$

rewards
$\mathbf{r}_{1:T}$

policy
$\pi(\mathbf{a}_t|\mathbf{s}_t, t, \boldsymbol{\theta})$

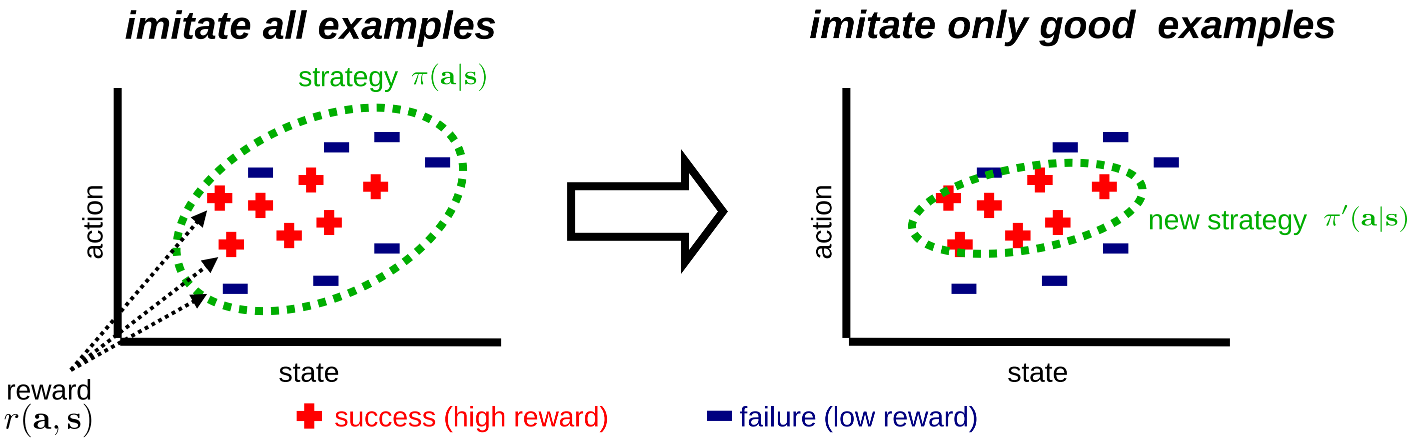
policy parameters

Objective: maximize expected return $J(\boldsymbol{\theta})$

TU Delft

Reward-Weighted Imitation

- Maximize reward = optimize strategy
- One possibility: reward-weighted imitation



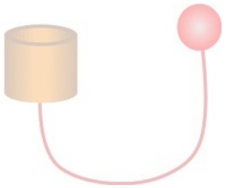
$$\boldsymbol{\theta}' = \boldsymbol{\theta} + E\left\{\sum_{t=1}^{T} Q_t^{\mathbf{sa}} \mathbf{W}_t^{\mathbf{s}}\right\}^{-1} E\left\{\sum_{t=1}^{T} Q_t^{\mathbf{sa}} \mathbf{W}_t^{\mathbf{s}} \boldsymbol{\varepsilon}_t\right\}$$

Kober & Peters, NIPS 2008

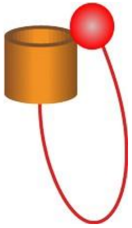
Reward-weighted Imitation

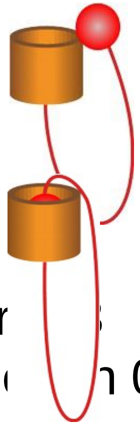
Trial 1
Return 0,1

Trial 2
Return 0,3



Trial 3
Return 0,8





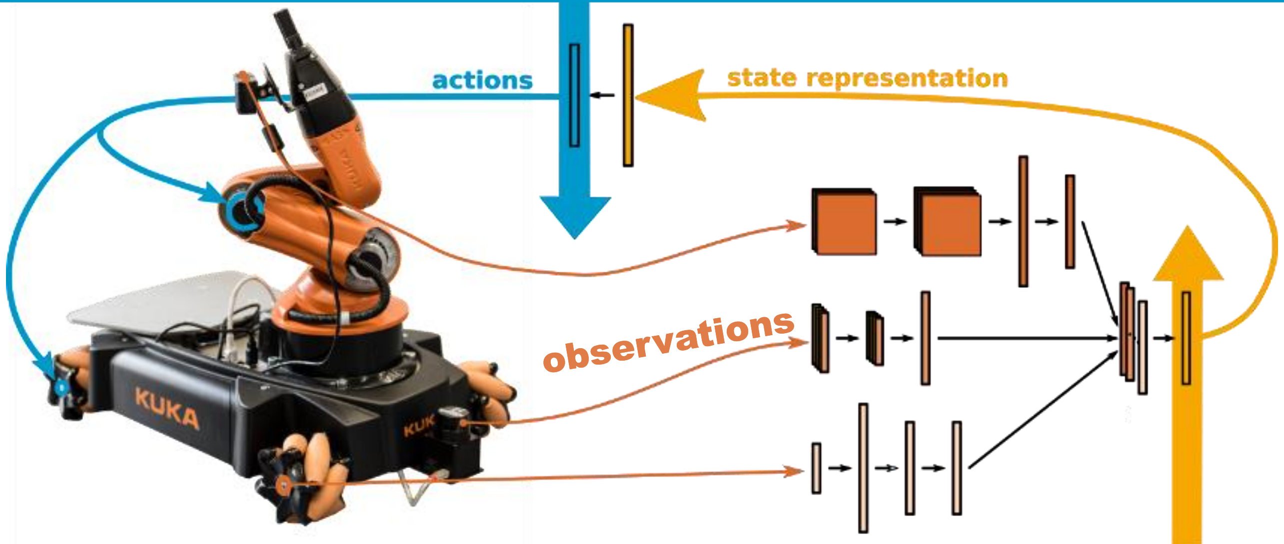
Tr
R n 0,8

Trial 4
Return 1,0

https://youtu.be/cNyoMVZQdYM

Kober & Peters, NIPS 2008

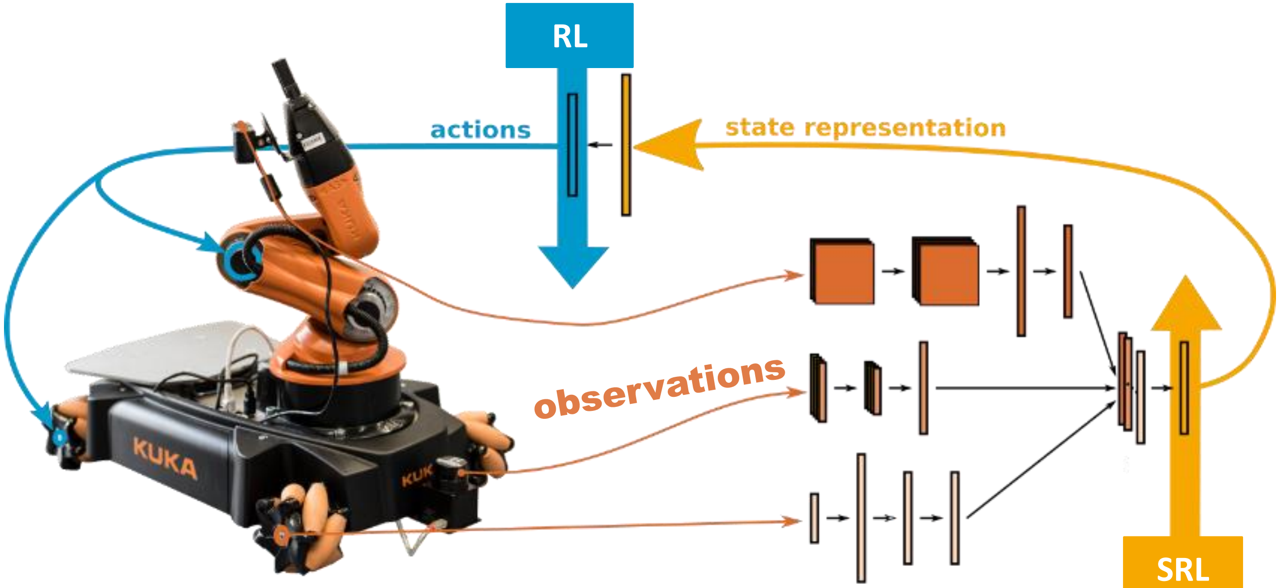
Reinforcement Learning:
- Learn behaviors using trial and error and a scalar reward function



State Representation Learning:
- **Cost functions acting on the state representation**
- **Prior knowledge about the world**
- **Force learning a more general representation**

TU Delft

Tim de Bruin

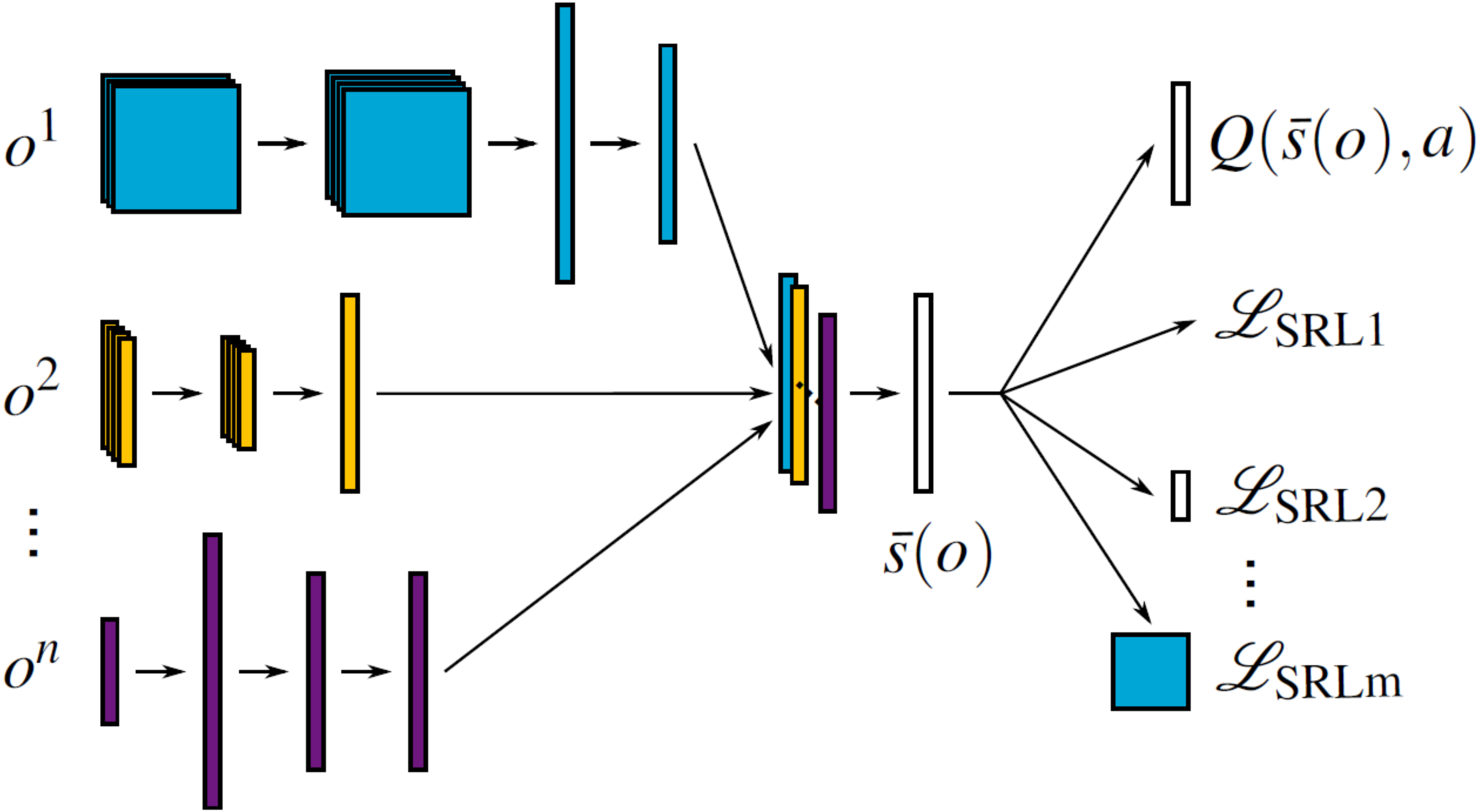


- Auto encoding
- Instantaneous reward prediction
- (Inverse) state dynamics
- Slowness and diversity
- Reinforcement Learning

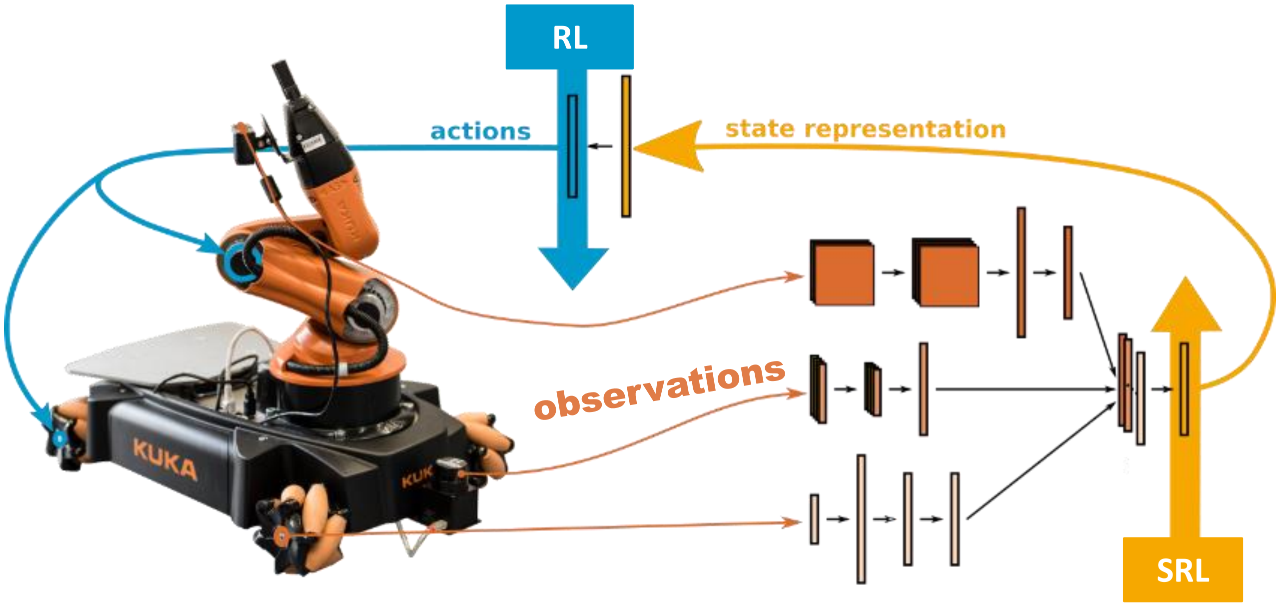
TUDelft

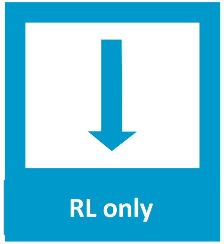
de Bruin, Kober, Tuyls, Babuška, RAL 2018

Tim de Bruin



de Bruin, Kober, Tuyls, Babuška, RAL 2018

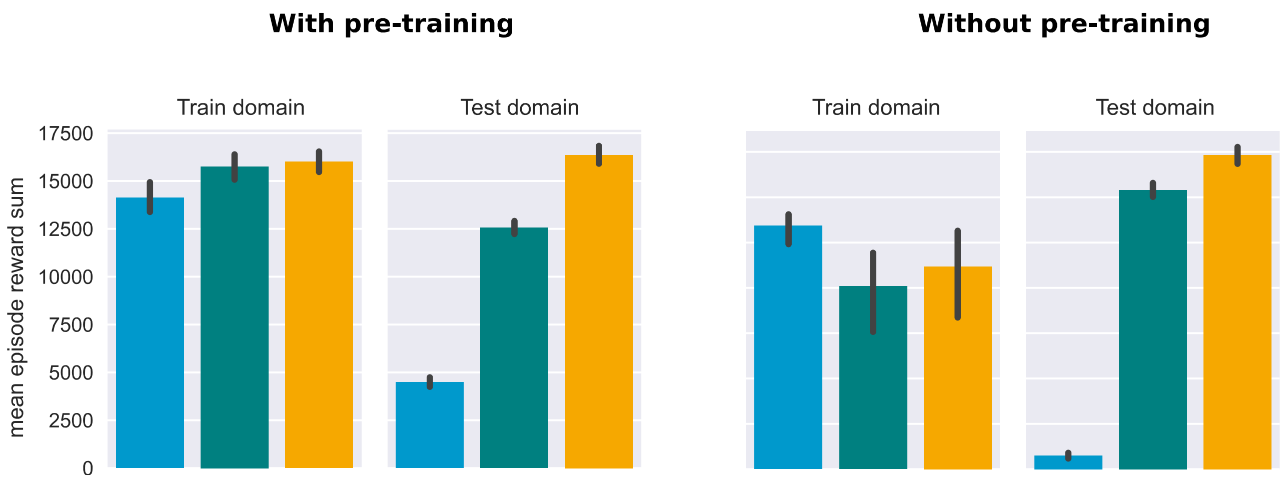




de Bruin, Kober, Tuyls, Babuška, RAL 2018

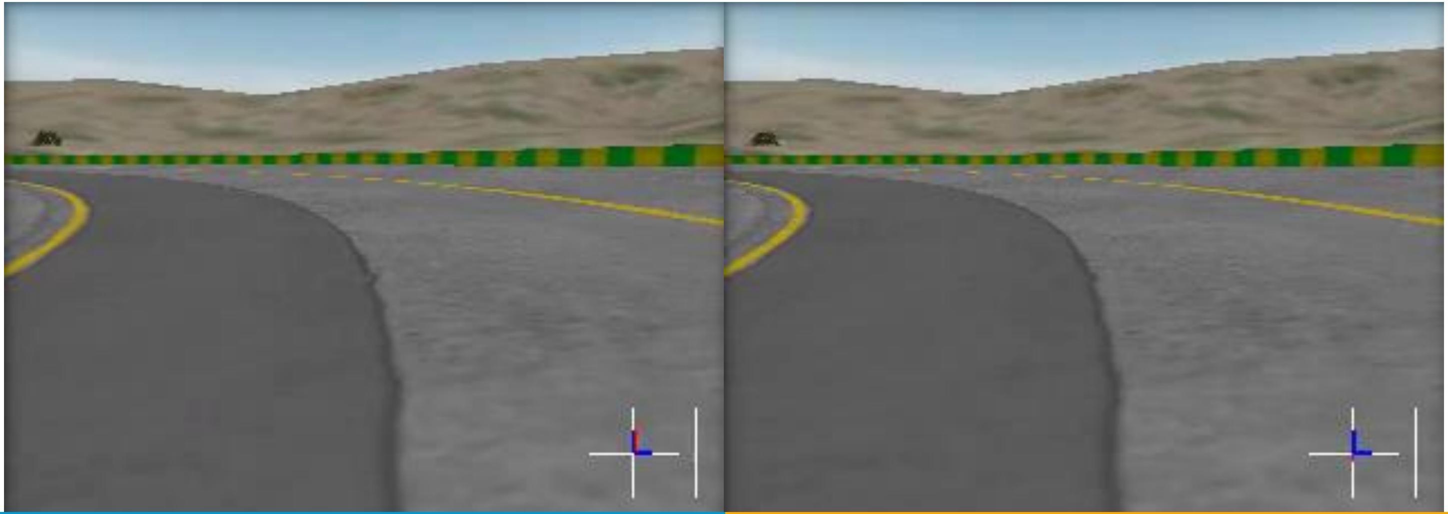
Experiment

de Bruin, Kober, Tuyls, Babuška, RAL 2018



de Bruin, Kober, Tuyls, Babuška, RAL 2018

Train Track



RL only

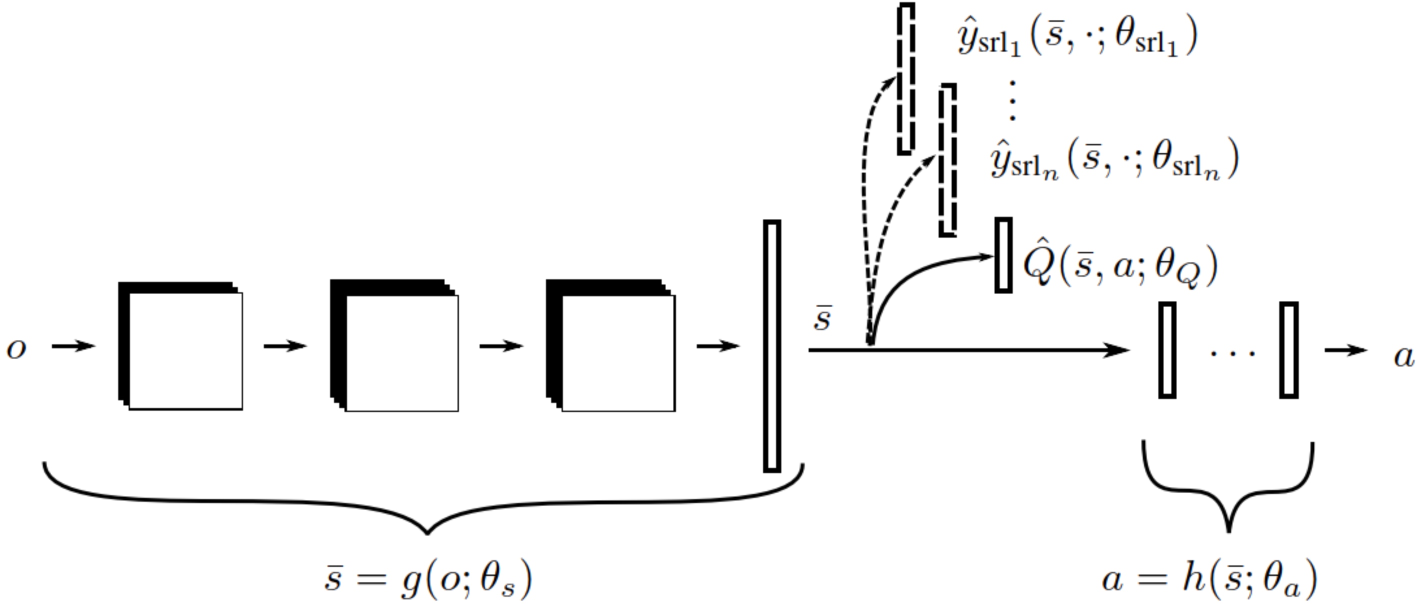
RL + SRL

de Bruin, Kober, Tuyls, Babuška, RAL 2018

Test Track

RL only | RL + SRL

Splitting SRL and RL

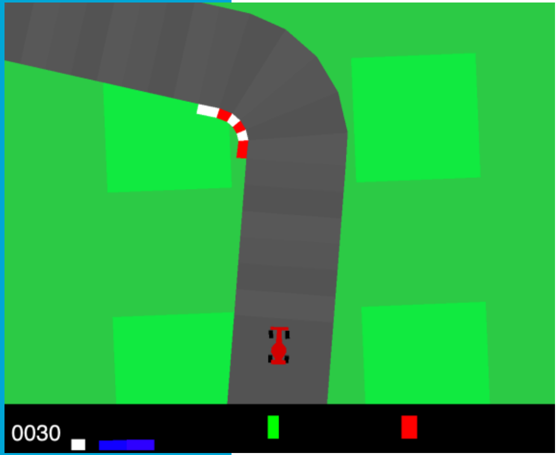


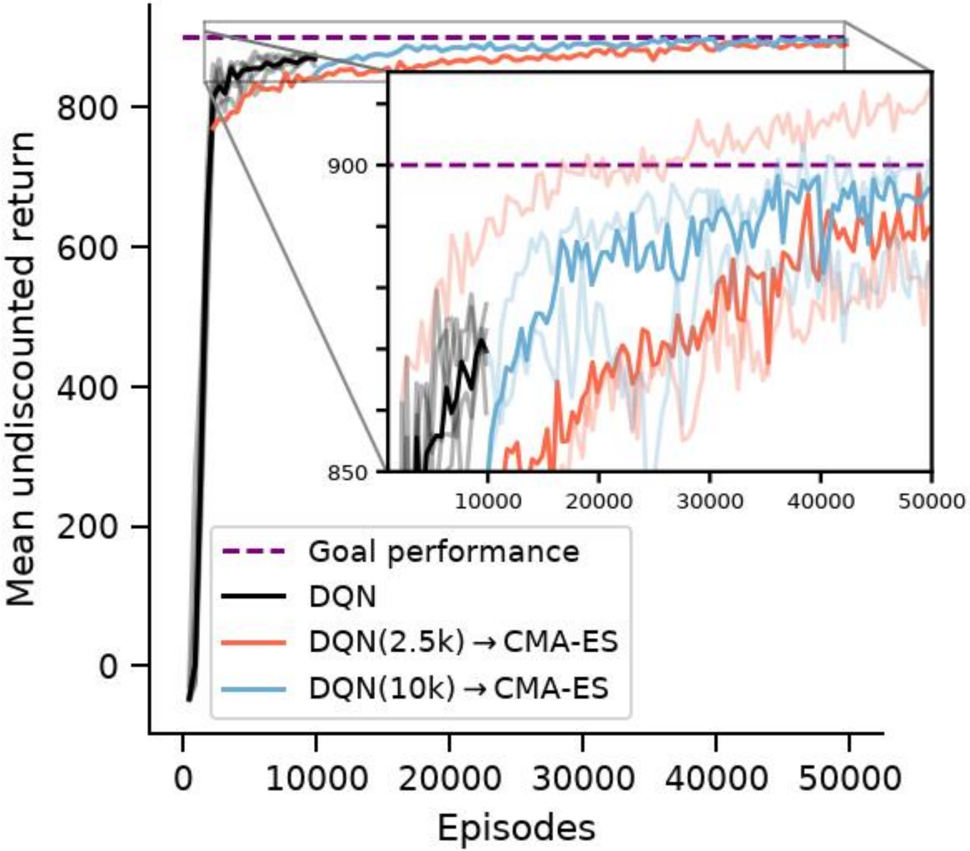
Phase 1
Phase 2

DQN/DDPG
freeze

DQN/DDPG
CMA-ES

de Bruin, Kober, Tuyls, Babuška, IFAC 2020



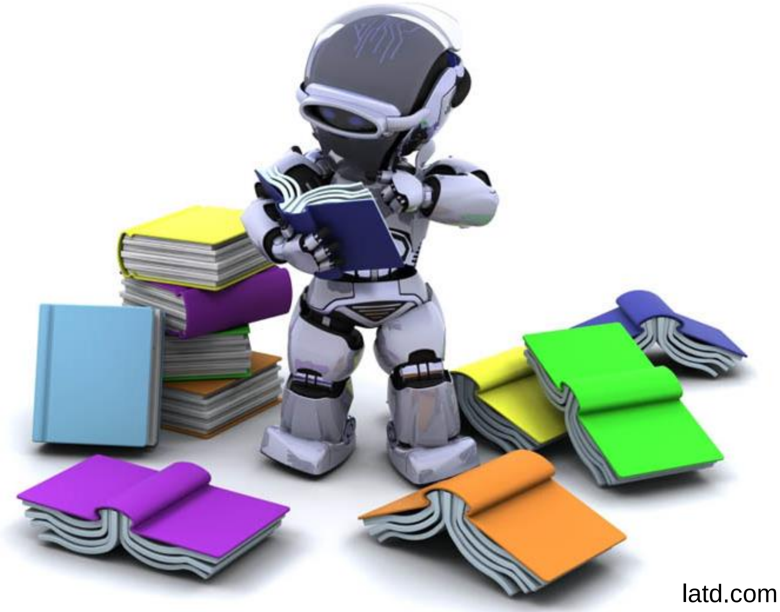


de Bruin, Kober, Tuyls, Babuška, IFAC 2020 https://youtu.be/D7zqglDkEq4

rotate box 180 degrees counterclockwise

Linda van der Spaa

Van der Spaa, Gienger, Bates, Kober, ICRA 2020


latd.com

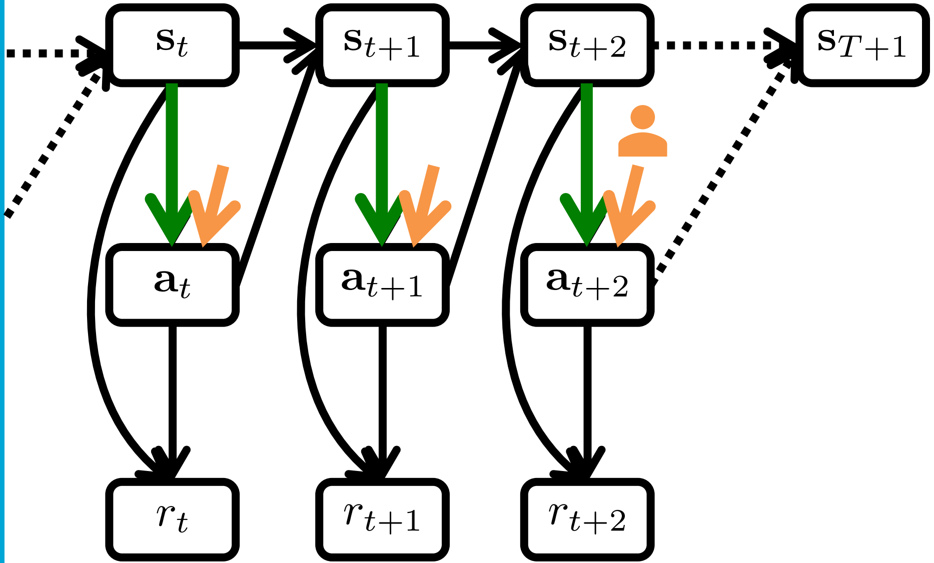
Questions so far?

TU Delft

Interacting to Learn

Reinforcement Learning: Exploration & Human Advice

Sutton & Barto 1998



states
$\mathbf{s}_{1:T+1}$

actions
$\mathbf{a}_{1:T}$

rewards
$\mathbf{r}_{1:T}$

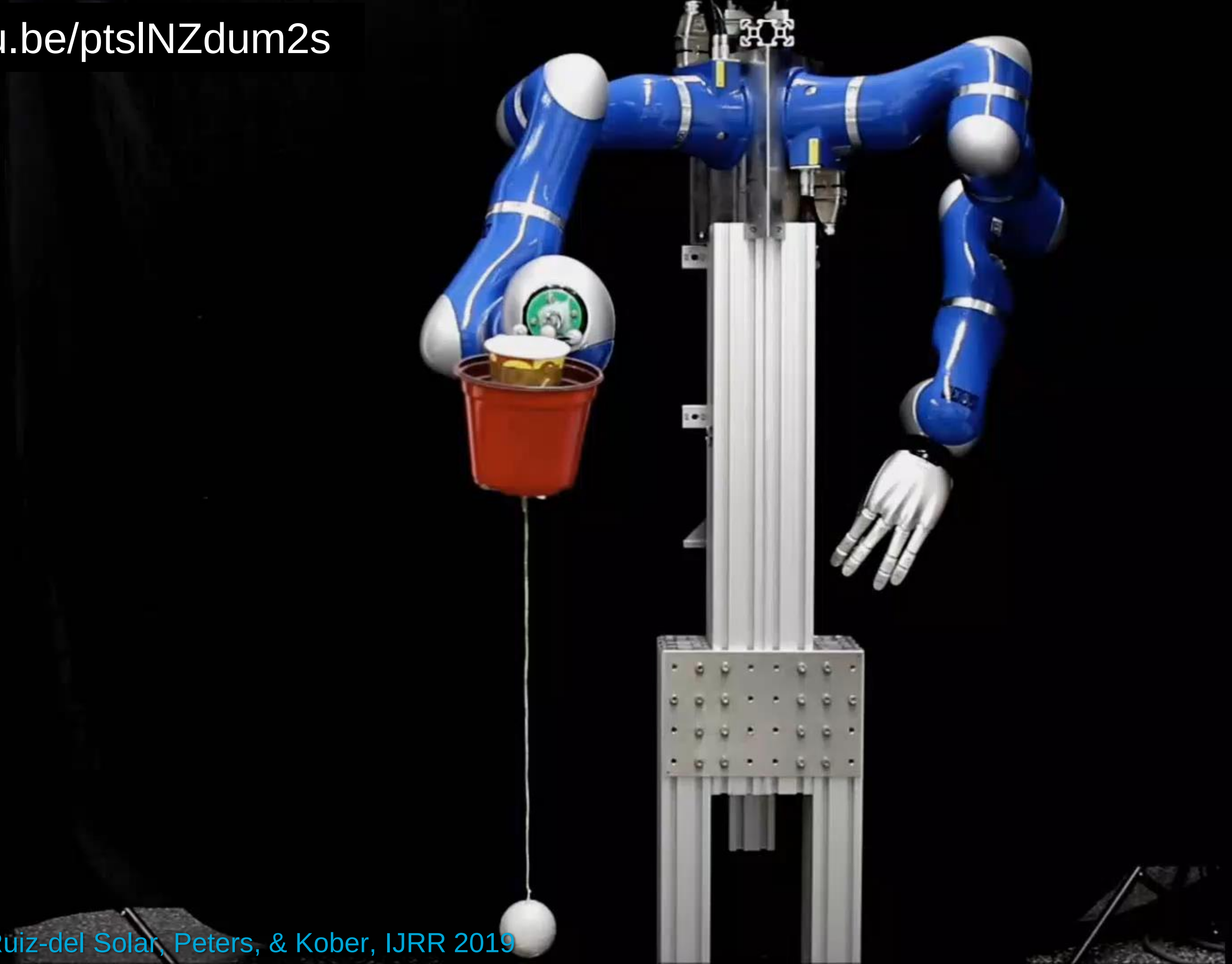
policy
$\pi(\mathbf{a}_t|\mathbf{s}_t, t, \boldsymbol{\theta})$

exploration
ε

Model human advice as exploration

TU Delft

https://youtu.be/ptslNZdum2s



Carlos Celemin

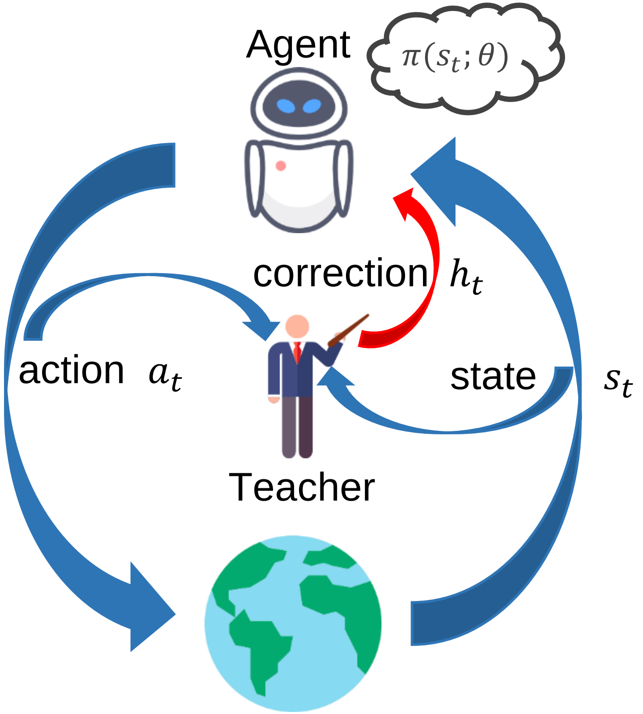
Celemin, Maeda, Ruiz-del Solar, Peters, & Kober, IJRR 2019

Back to Imitation Learning…

Interactive Learning: COACH

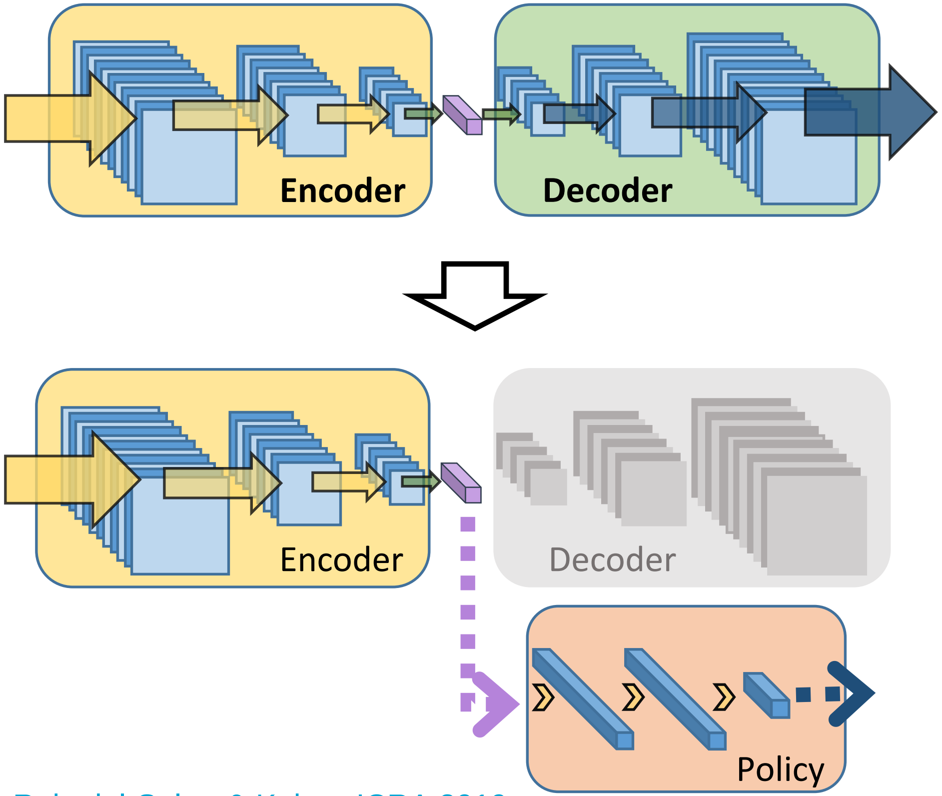
Rodrigo Pérez

Carlos Celemin



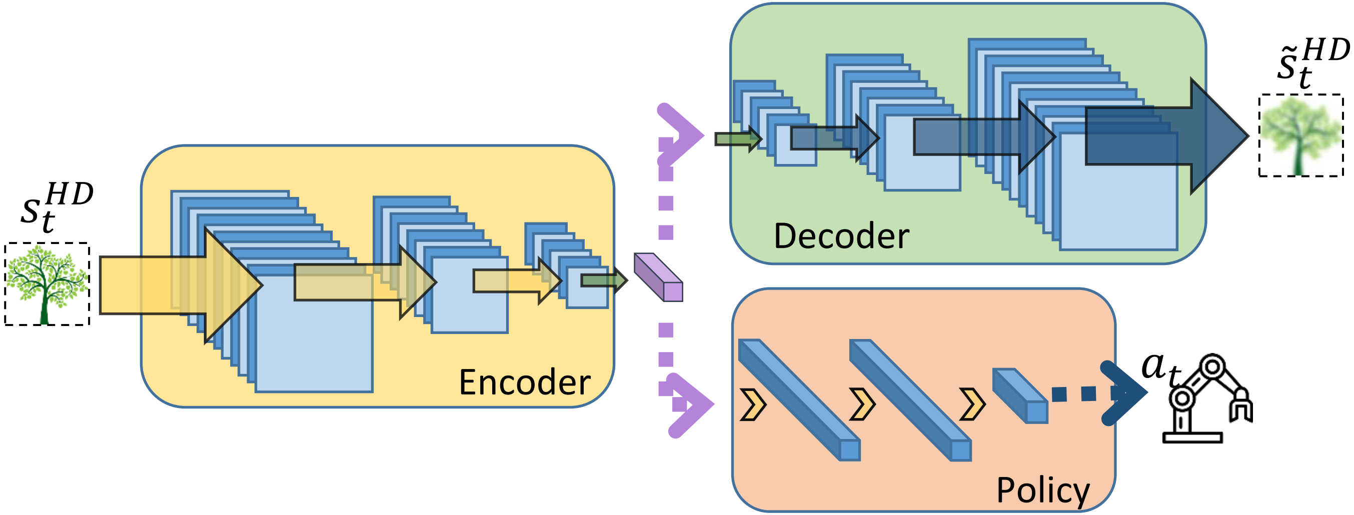
Pérez, Celemin Paez, Ruiz-del-Solar, & Kober, ICRA 2019

Deep COACH - Offline



TUDelft

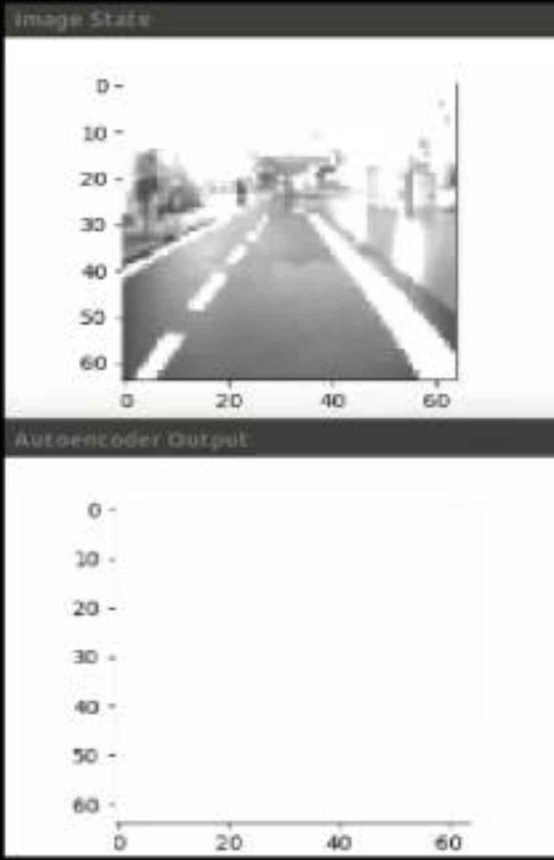
Deep COACH - Online



Pérez, Celemin Paez, Ruiz-del-Solar, & Kober, ICRA 2019

https://youtu.be/i4f1D4CH26E

Pérez, Celemin Paez, Ruiz-del-Solar, & Kober, ICRA 2019



https://youtu.be/i4f1D4CH26E

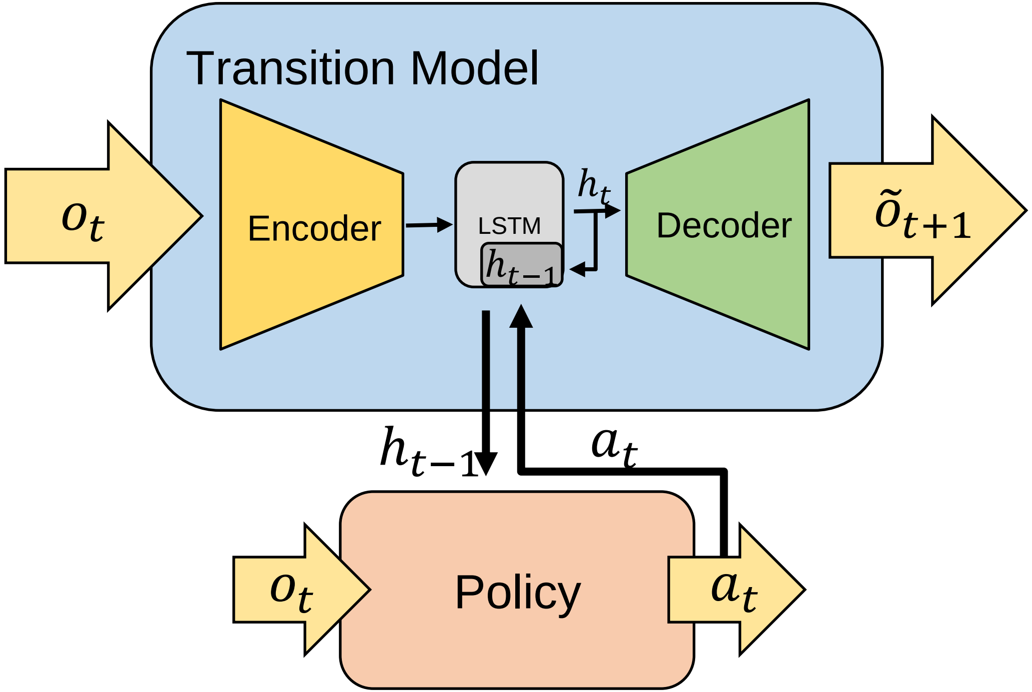

Rodrigo Pérez

Giovanni Franzese

Carlos Celemin

TU Delft

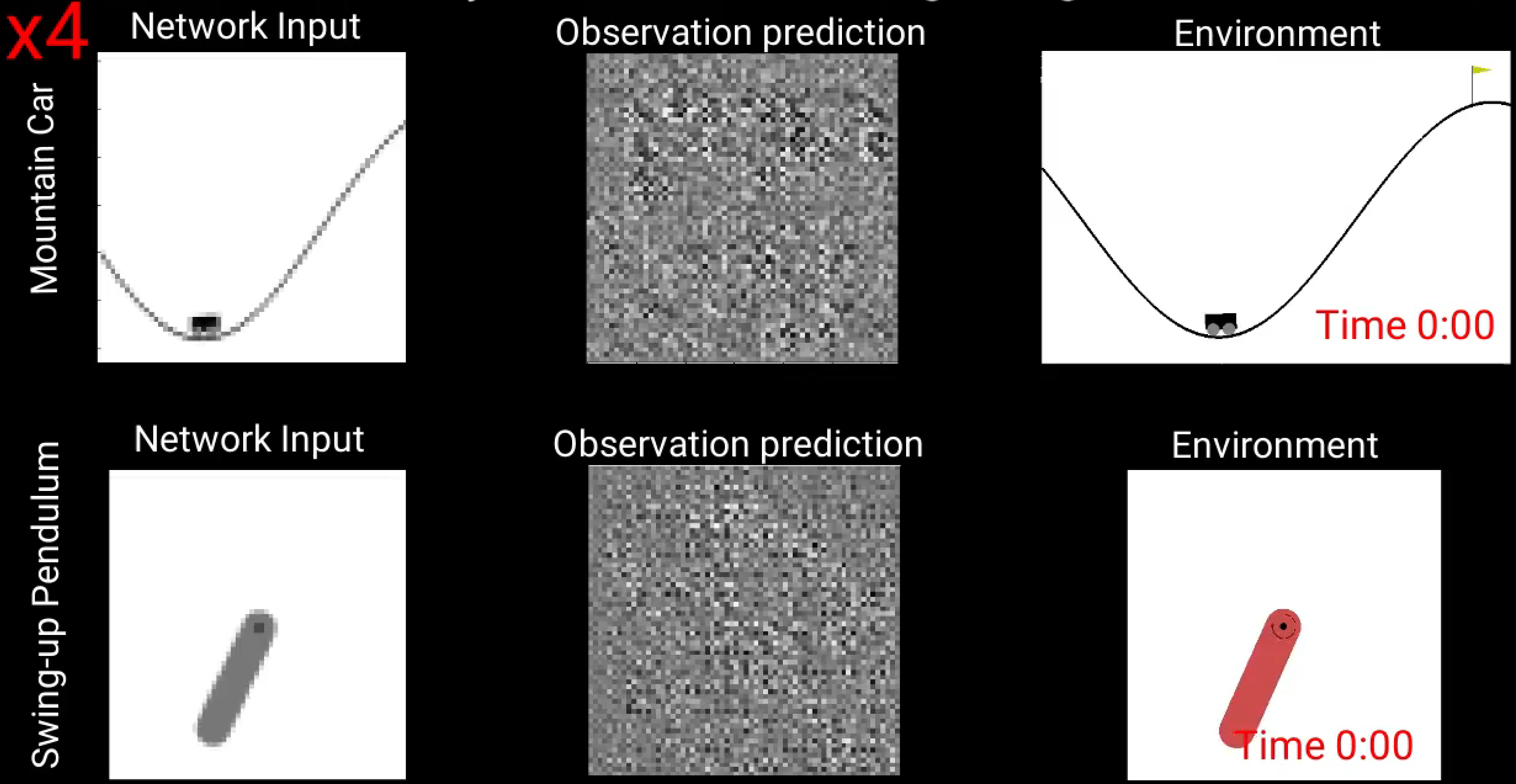
Deep COACH - Memory



Pérez, Celemin Paez, Franzese, Ruiz-del-Solar, & Kober, RAM 2020

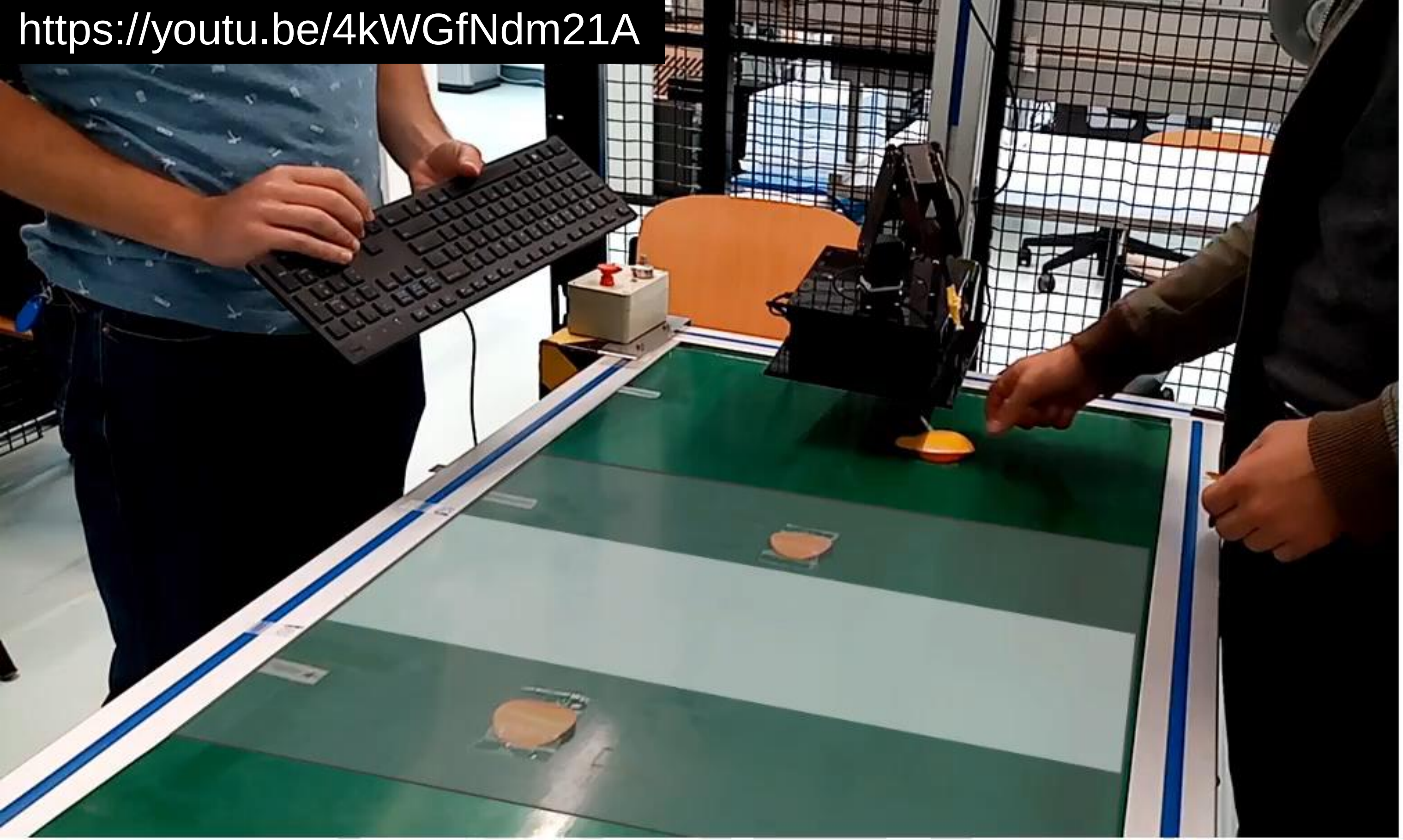
https://youtu.be/4kWGfNdm21A

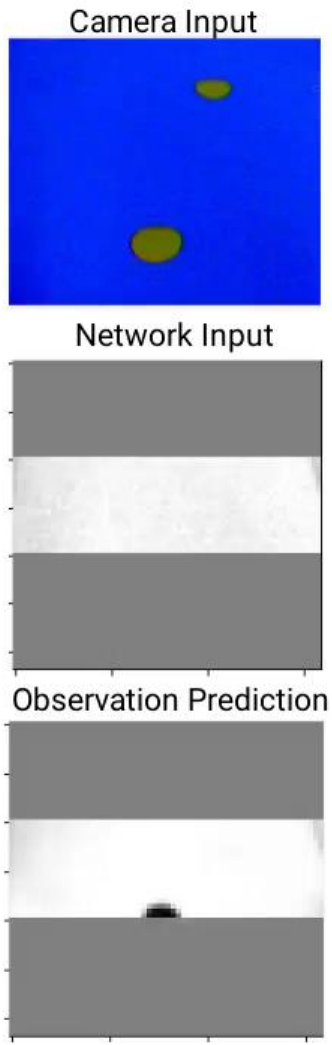
... Model Learning using D- COACH and SRL



Pérez, Celemin Paez, Franzese, Ruiz-del-Solar, & Kober, RAM 2020

https://youtu.be/4kWGfNdm21A





x4

Start Training

Pérez, Celemin Paez, Franzese, Ruiz-del-Solar, & Kober, RAM 2020

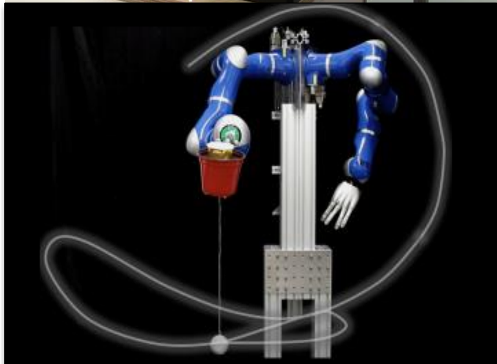
https://youtu.be/apxeeW-hm6I

Walt Disney Studios – Real Steel

Conclusion

Summary

- Robot learning constraints
  - “Small” data
  - Safety
- Interactive Learning
  - Efficient & intuitive
- Challenges
  - Uncertainty
  - Variations



Questions?

j.kober@tudelft.nl

www.jenskober.de

latd.com