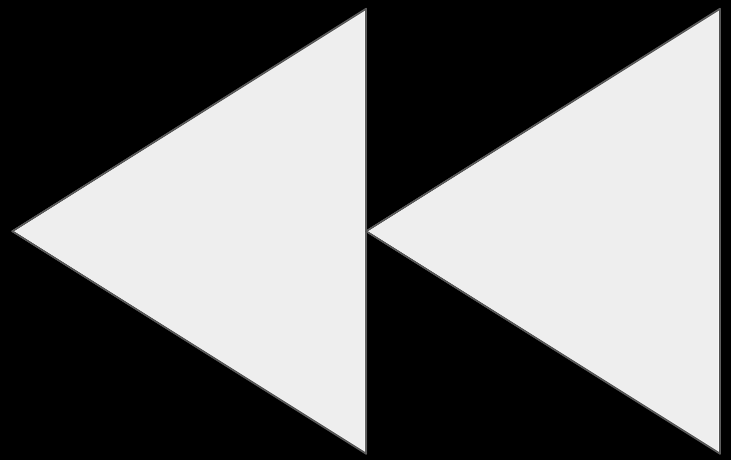
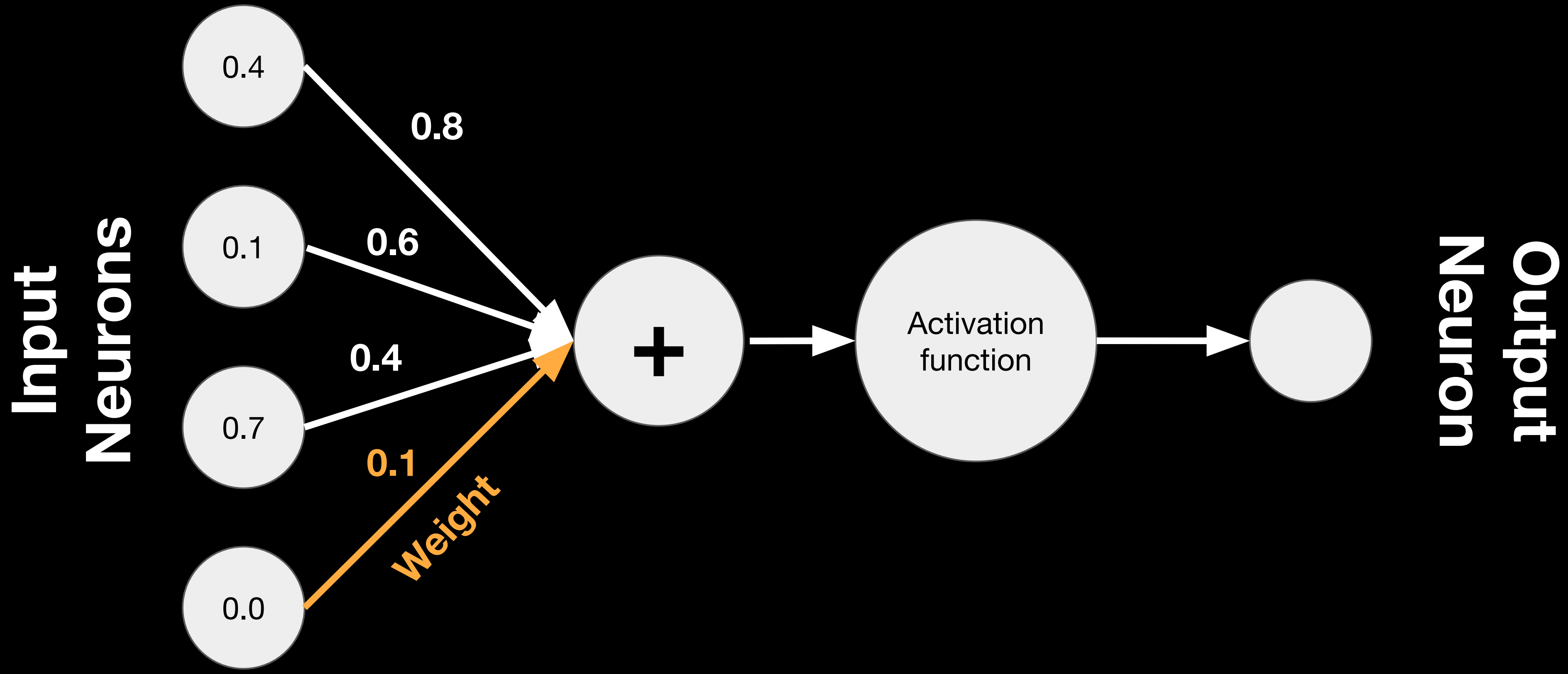
Reinforcement and Imitation Learning

Robert Katzschmann

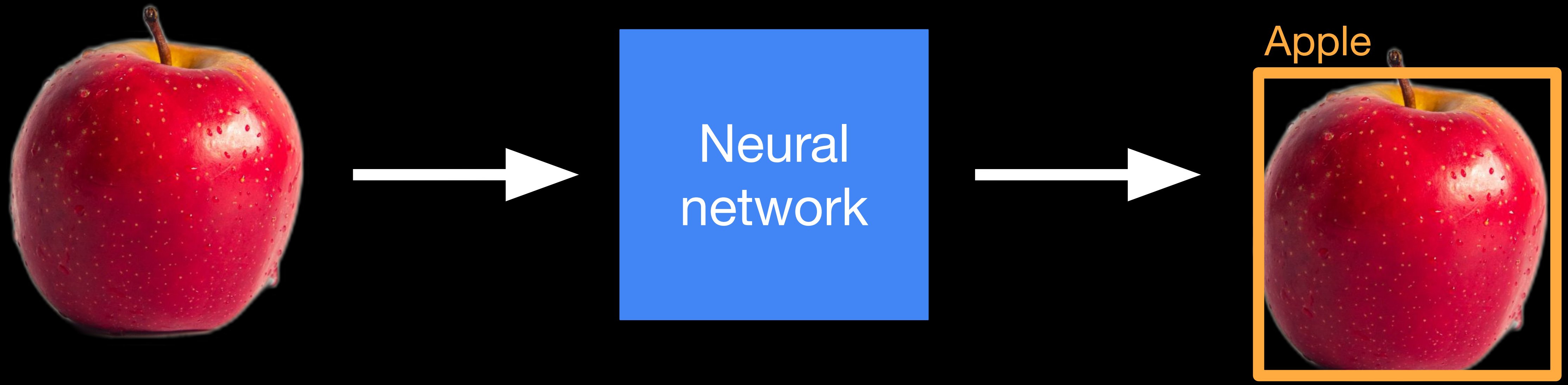
Assistant Professor of Robotics, Soft Robotics Lab



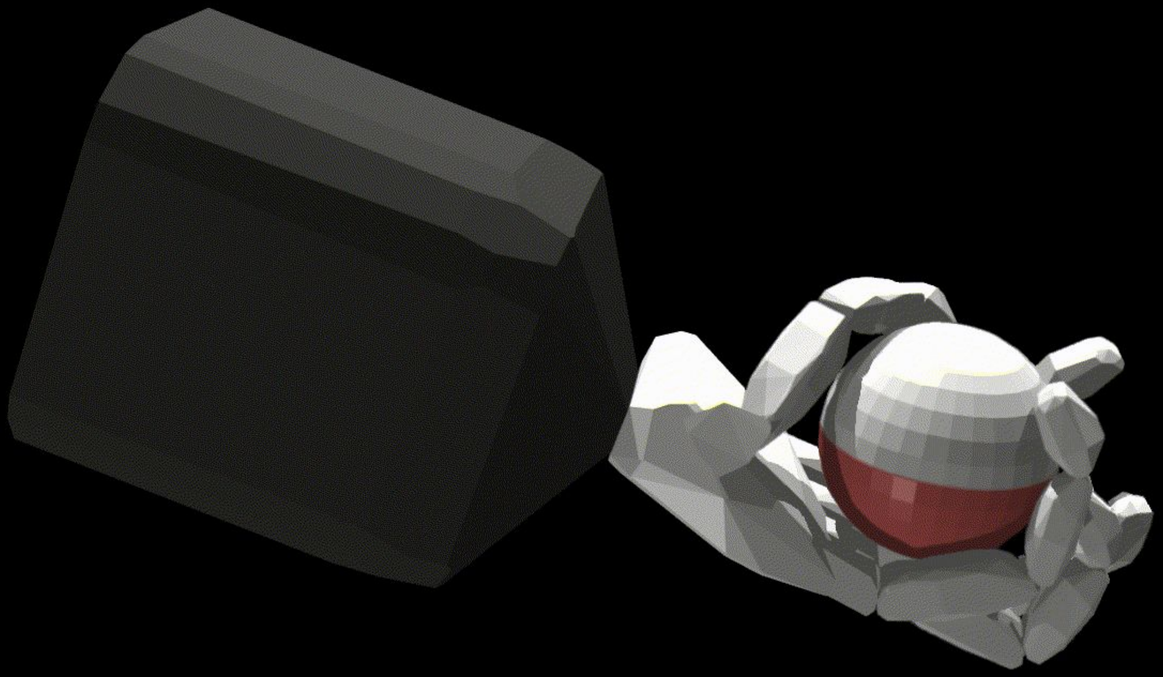
Recap: Neural networks



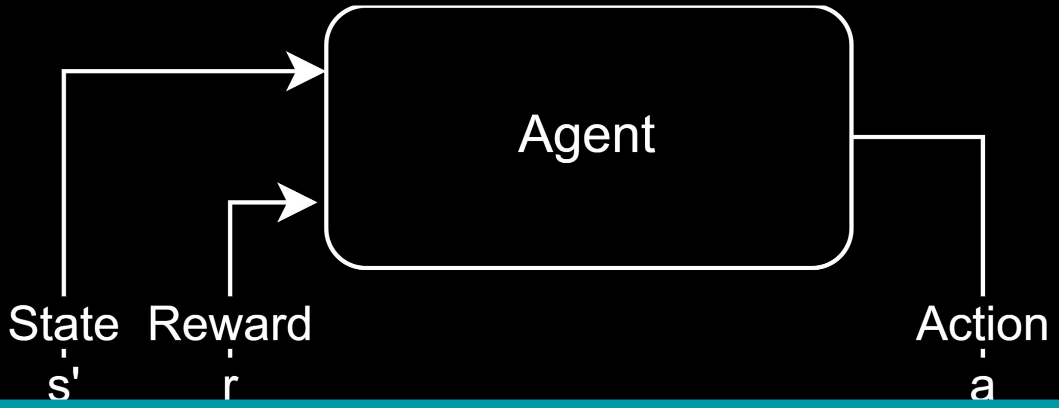
Recap: Neural networks



ETH zürich



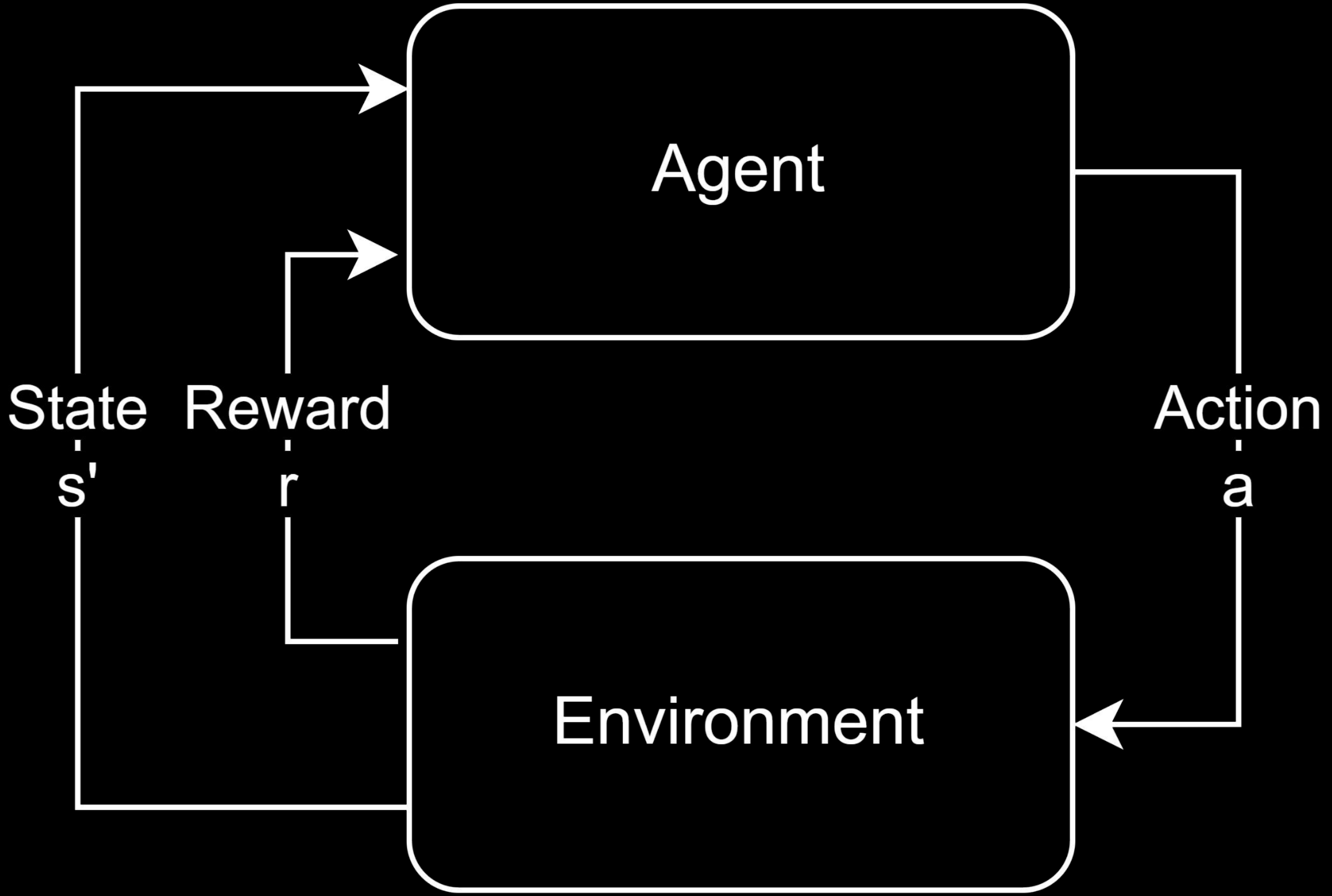
SRL



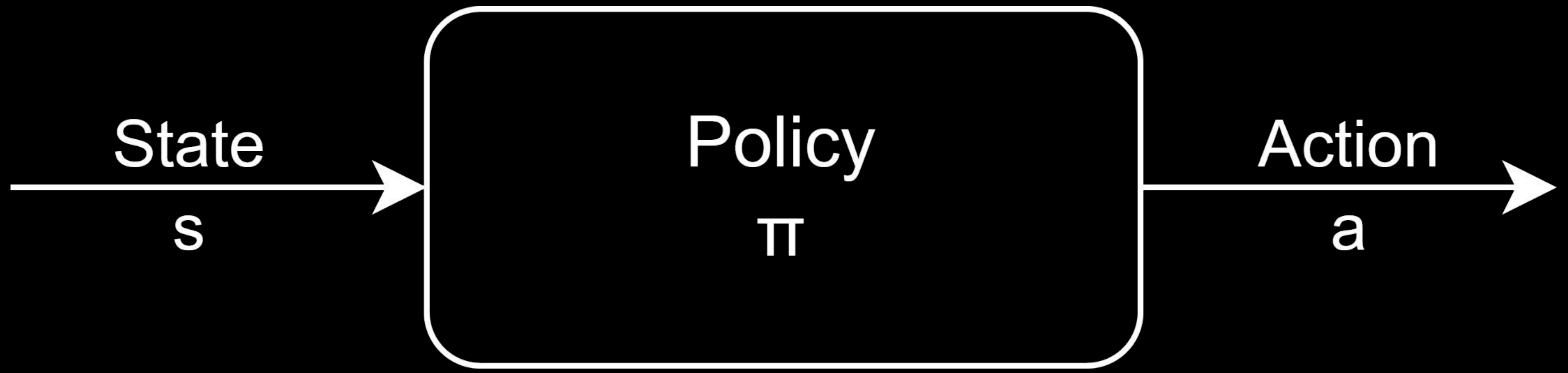
Part 1:
Reinforcement Learning

SoftRobotics Laboratory

Markov Process



Policy



Reward and Discount Factor

Cumulative reward

Reward at timestep k

Action at timestep k

$$R_t = \sum_{k=t}^{T} \gamma^{(k-t)} r_k\left(s_k, a_k\right)$$

Discount Factor

State at timestep k

ETH zürich SoftRobotics Laboratory

Q and Value functions

Value function in state s given policy π

Expected cumulative reward

$$V^{\pi}(s) = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1}\right] \qquad \forall s \in \mathbb{S}$$

Set of all possible states

Q and Value functions

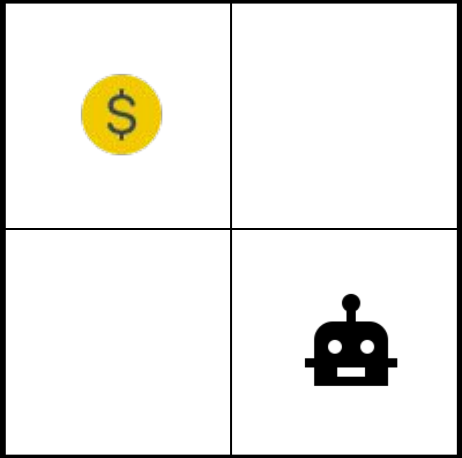
Q function in state s and action a given policy π

Expected cumulative reward

$$Q^{\pi}(s, a) = \mathbb{E}_{\pi}\left[\sum_{k=0}^{\infty} \gamma^{k} r_{t+k+1} \mid s_t = s, a_t = a\right]$$

Given that in state s action a is applied

Q and Value functions



Original map

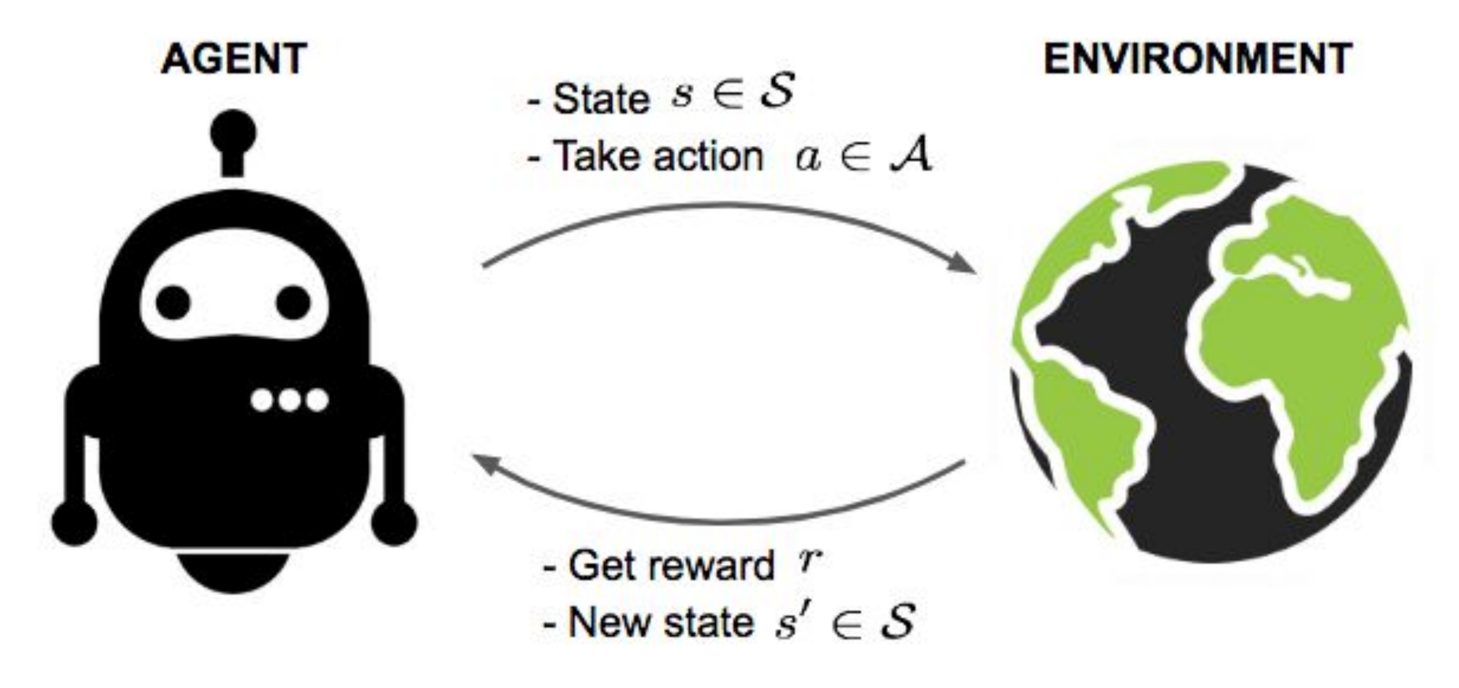
1.0	0.75
0.75	0.5

Value function for each cell

↑ 1.0	↓ 1.0	↑ 0.0	↓ 0.25
← 1.0	→ 1.0	← 0.75	→ 0.0
↑ 0.75	↓ 0.0	↑ 0.5	↓ 0.0
← 0.0	→ 0.25	← 0.5	→ 0.0

Q function for each cell and action

High level intuition



Altamimi, S., 2018. *QoE-Fair Video Streaming over DASH* (Doctoral dissertation, Université d'Ottawa/University of Ottawa).

Q Learning

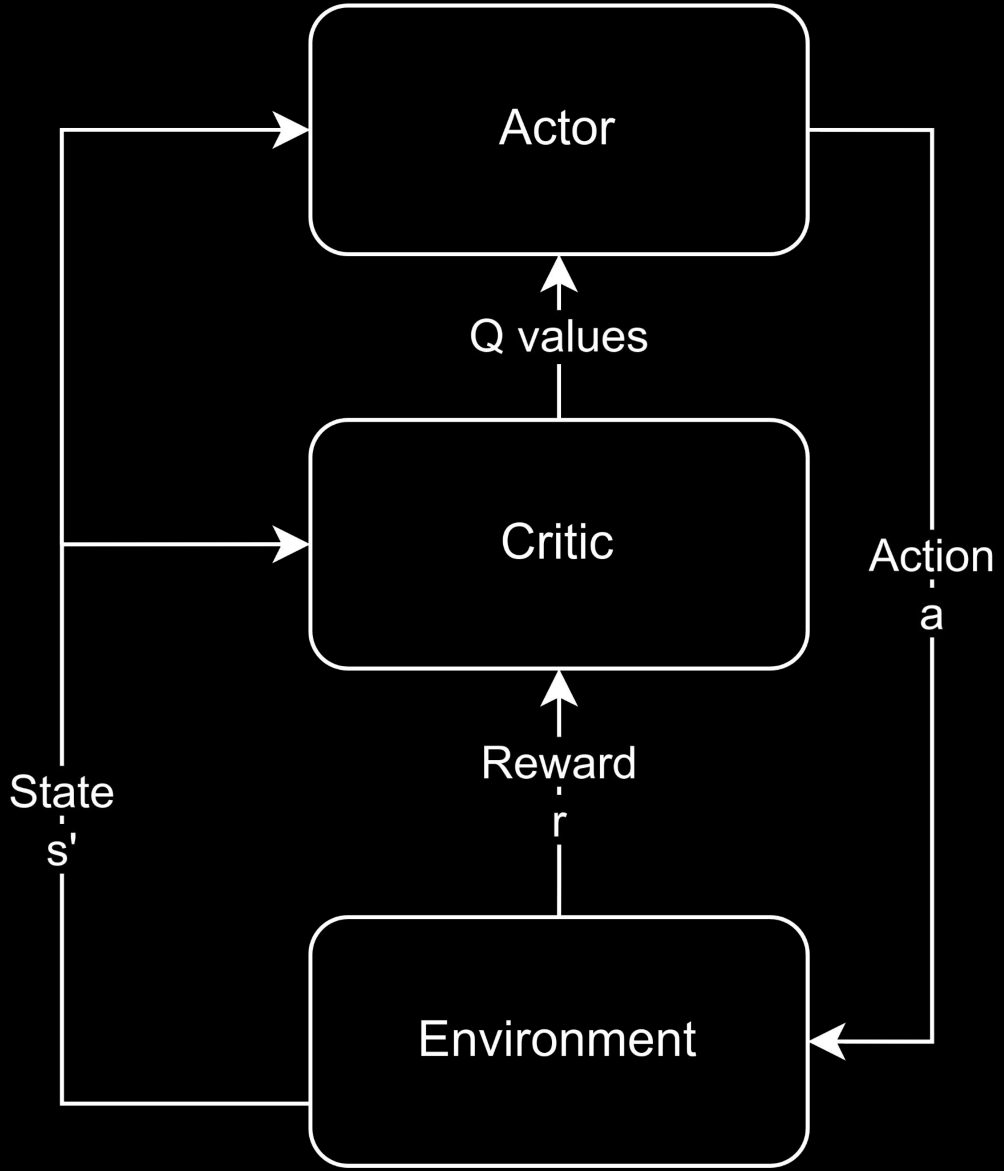
for each step t **do**

Observe (s_t, a_t, r_t, s_{t+1})

Update $Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha_t[\overbrace{r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)}^{\text{Temporal difference}}]$

end for

Actor-Critic structure



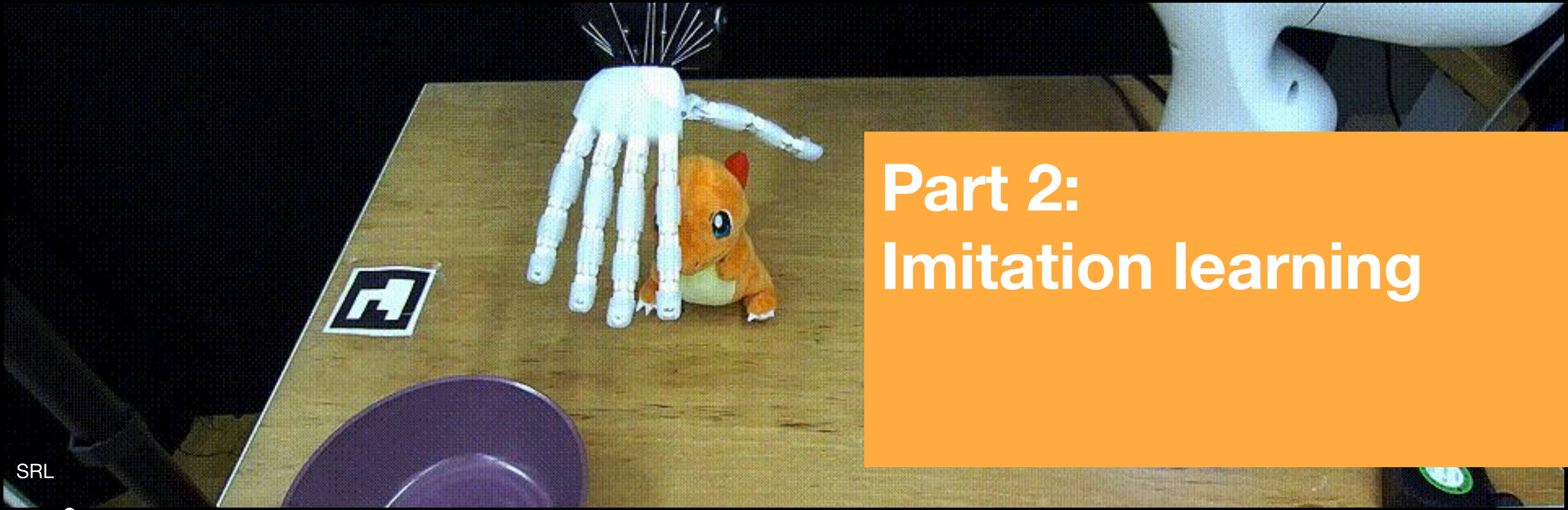
ETH zürich SoftRobotics Laboratory

State-of-the-art algorithms

DDPG

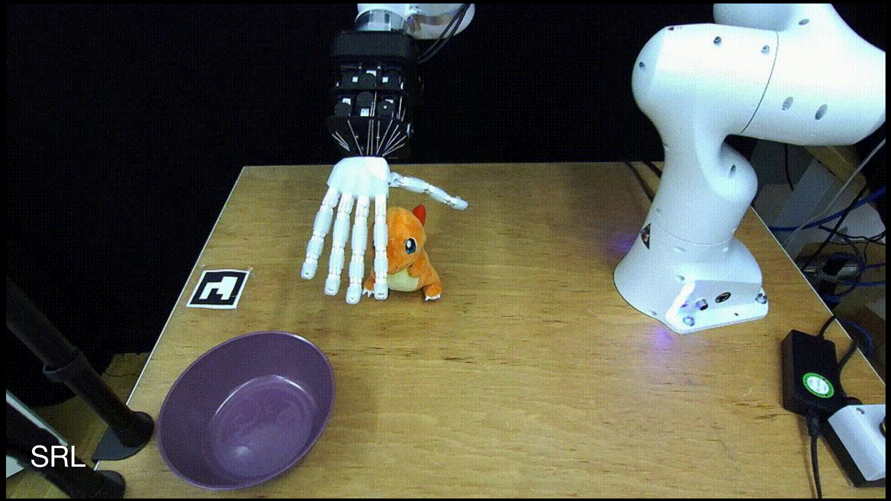
PPO

SAC



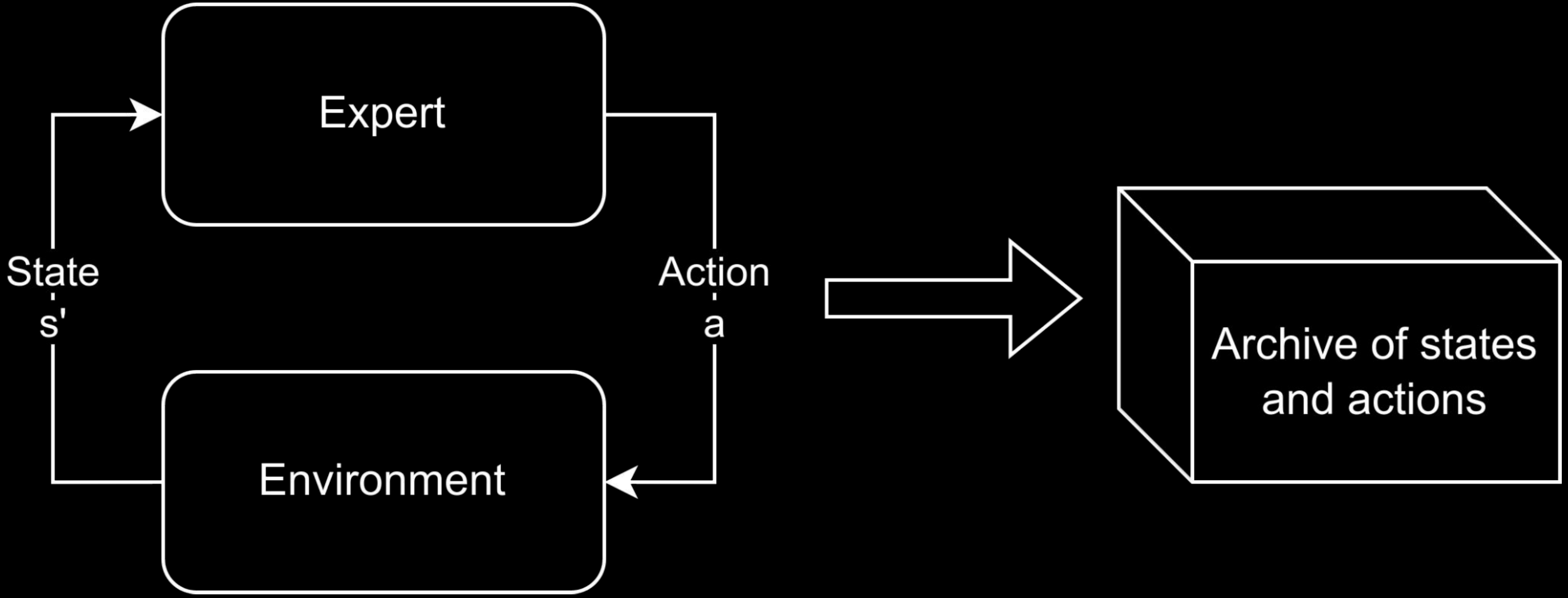
Part 2:
Imitation learning

Differences with Reinforcement Learning

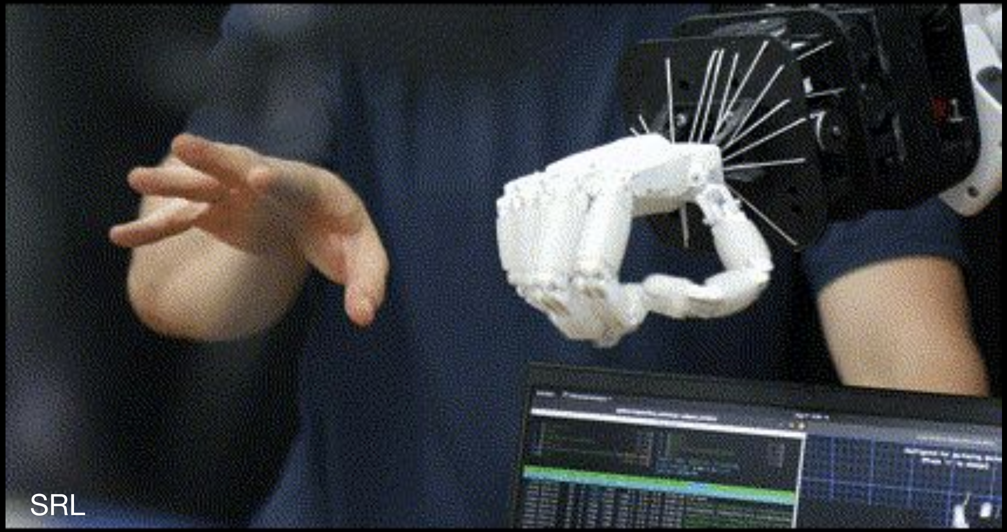


Reward function?

Differences with Reinforcement Learning

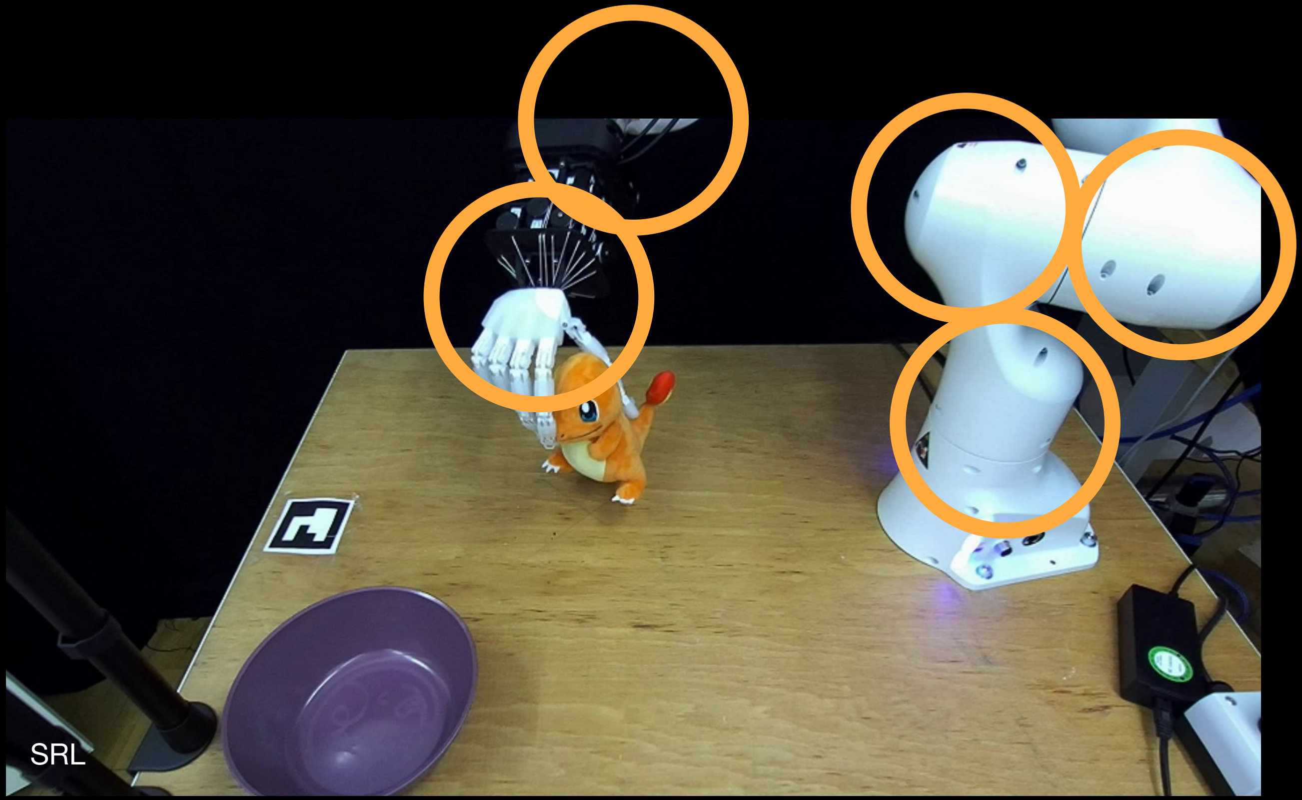


Expert choice



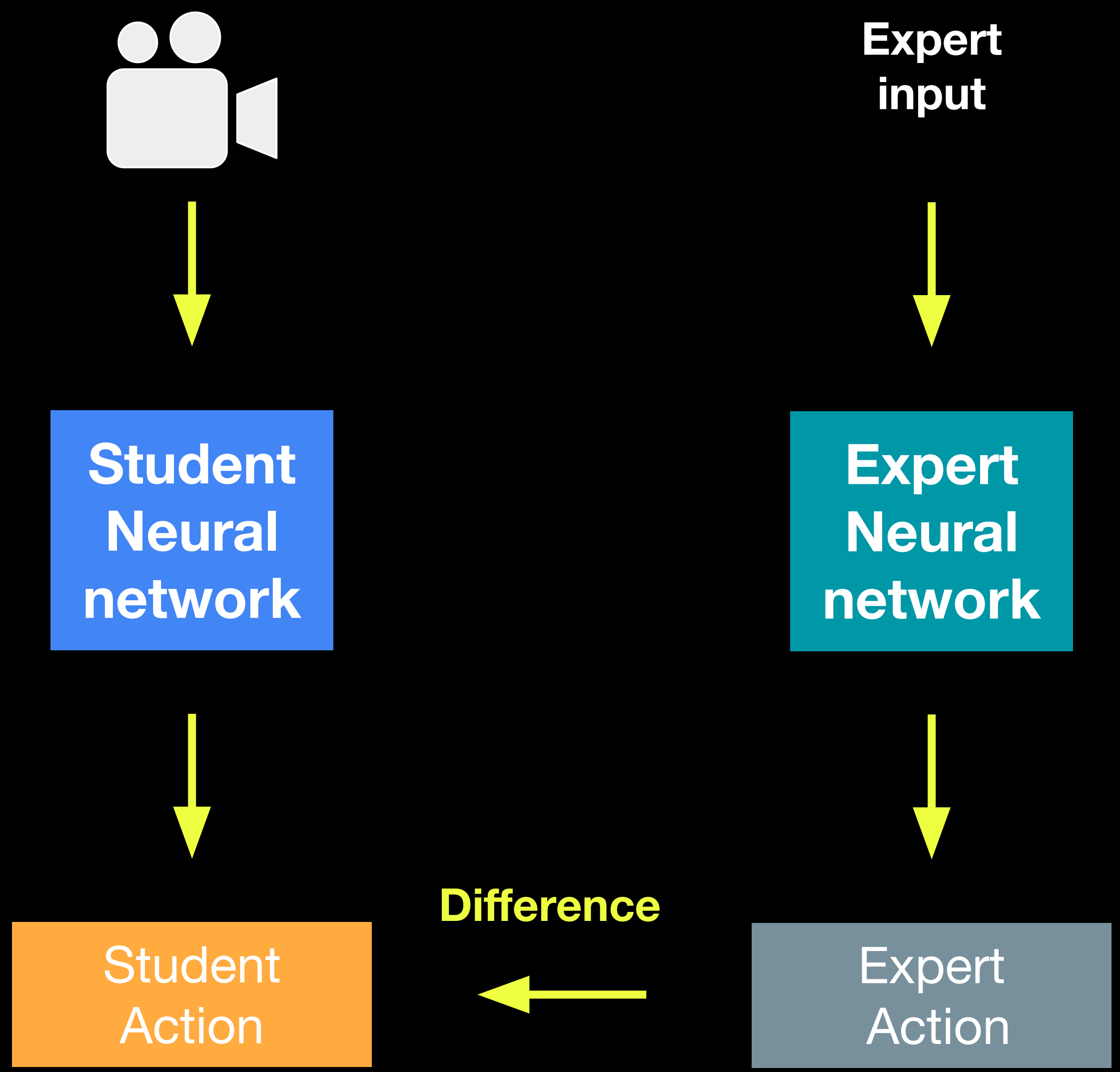
Expert choice

Student Neural network

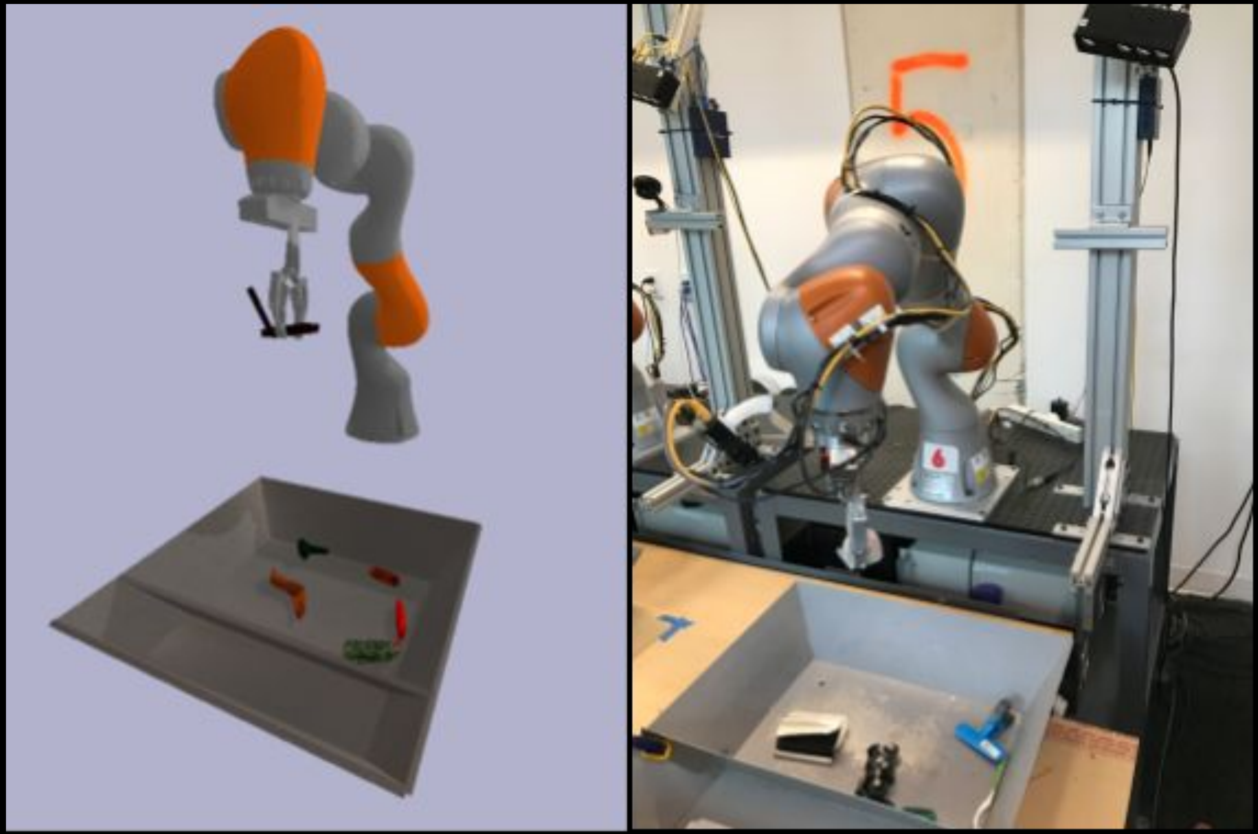


Expert Neural network

Behavioral cloning

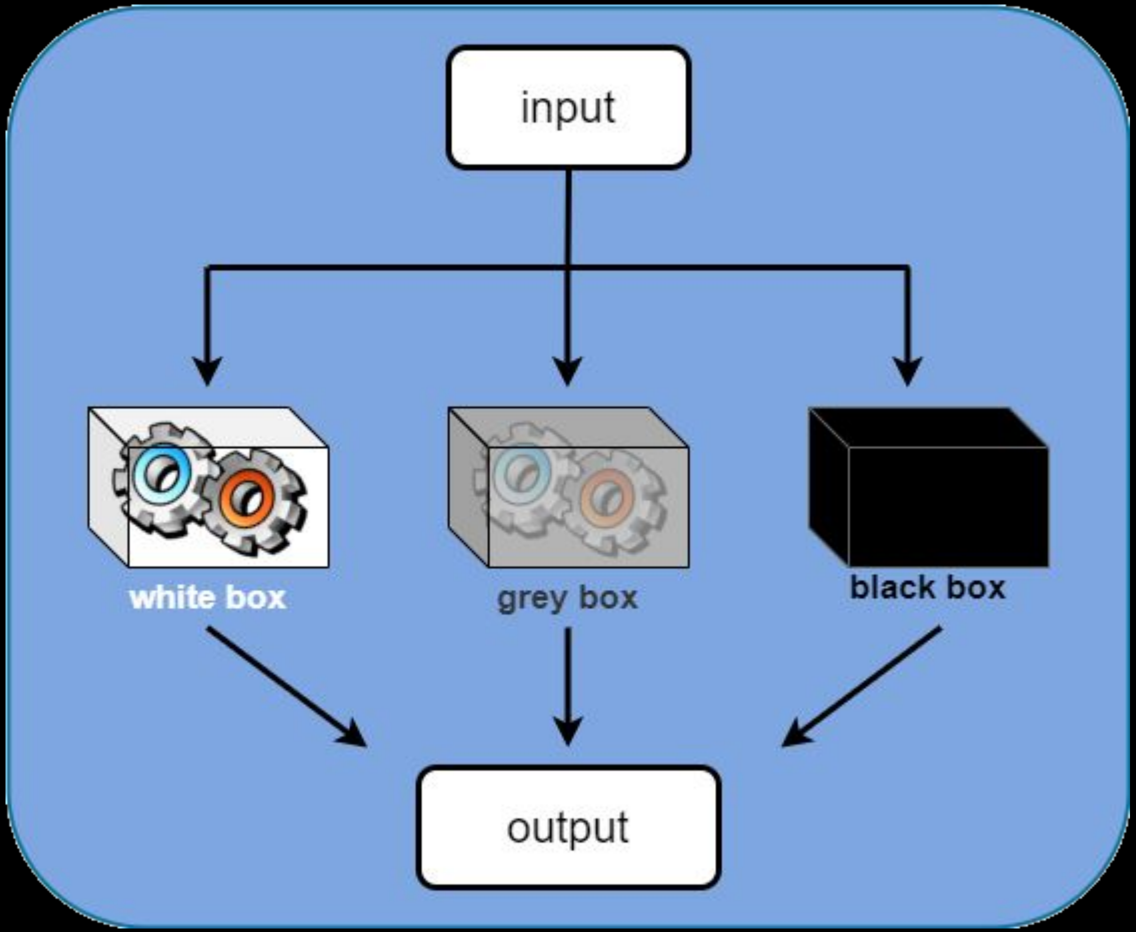


Training to Overcome the sim2real Gap



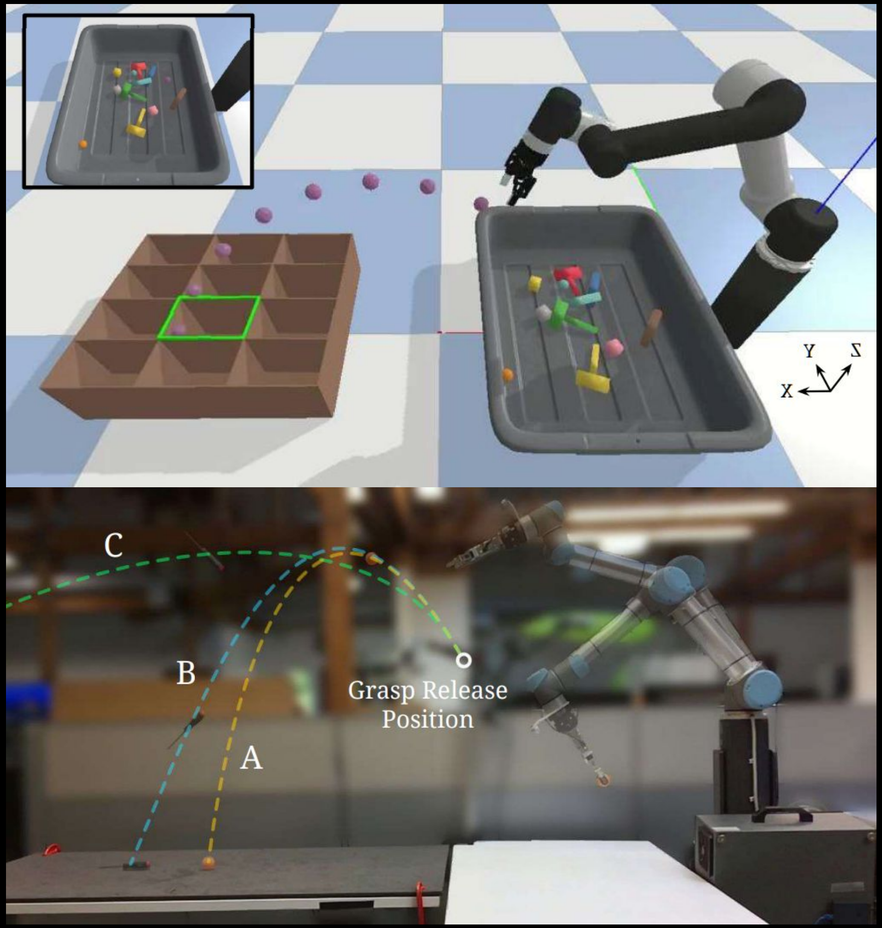
https://blog.research.google/2017/10/closing-simulation-to-reality-gap-for.html

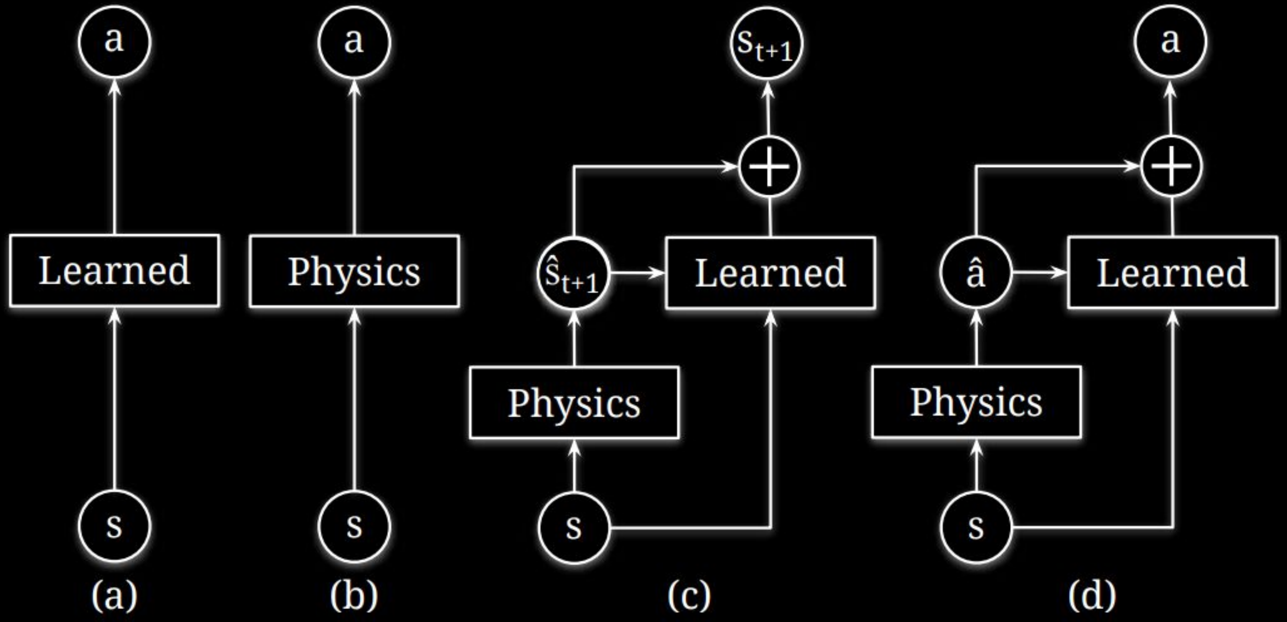
Training to Overcome the sim2real Gap



Wikipedia

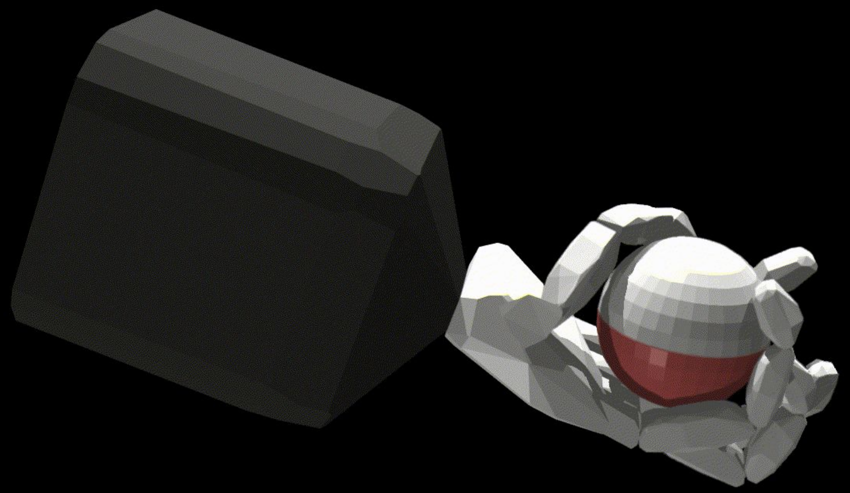
Training to Overcome the sim2real Gap



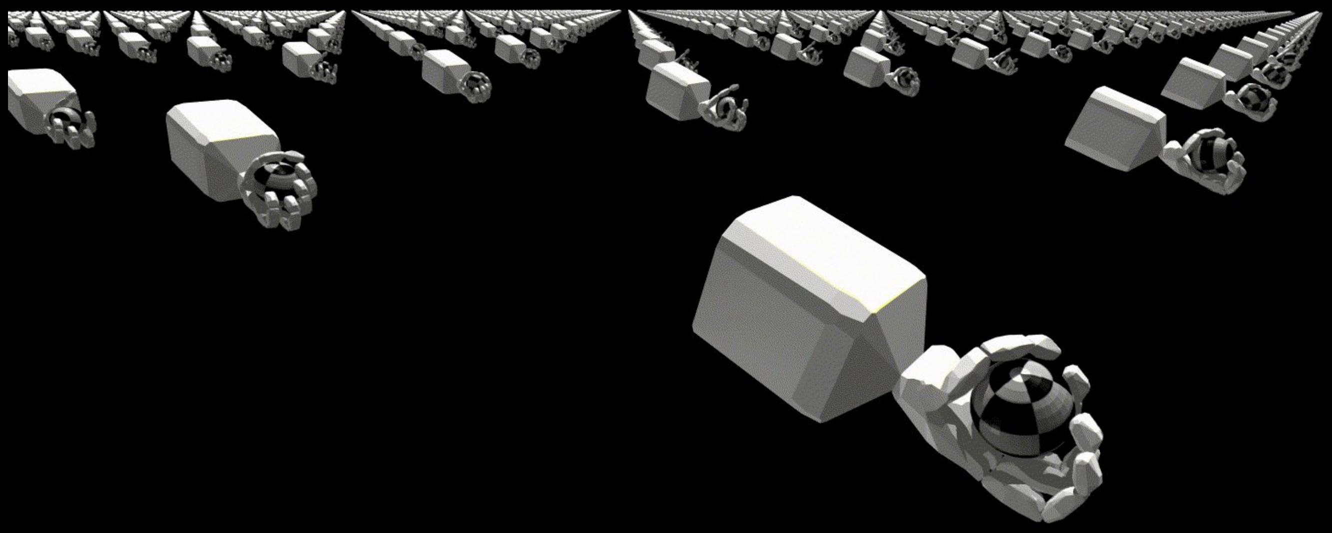


Zeng et al., RSS (2020)

Training to Overcome the sim2real Gap



SRL



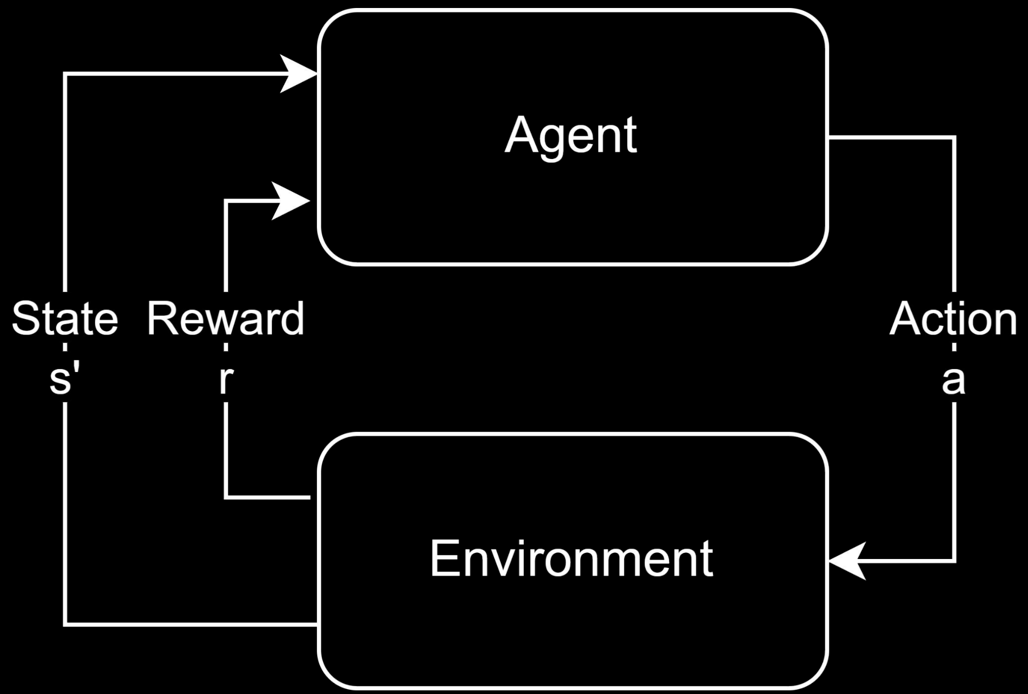
SRL

Wikipedia

Faive Robotics

Outro
Recap

Recap: Markov process



$$R_t = \sum_{k=t}^{T} \gamma^{(k-t)} r_k \left(s_k, a_k\right)$$

Recap: From Q and Value Functions to State-of-the-art algorithms

$$V^\pi(s) = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1}\right] \quad \forall s \in \mathbb{S}$$

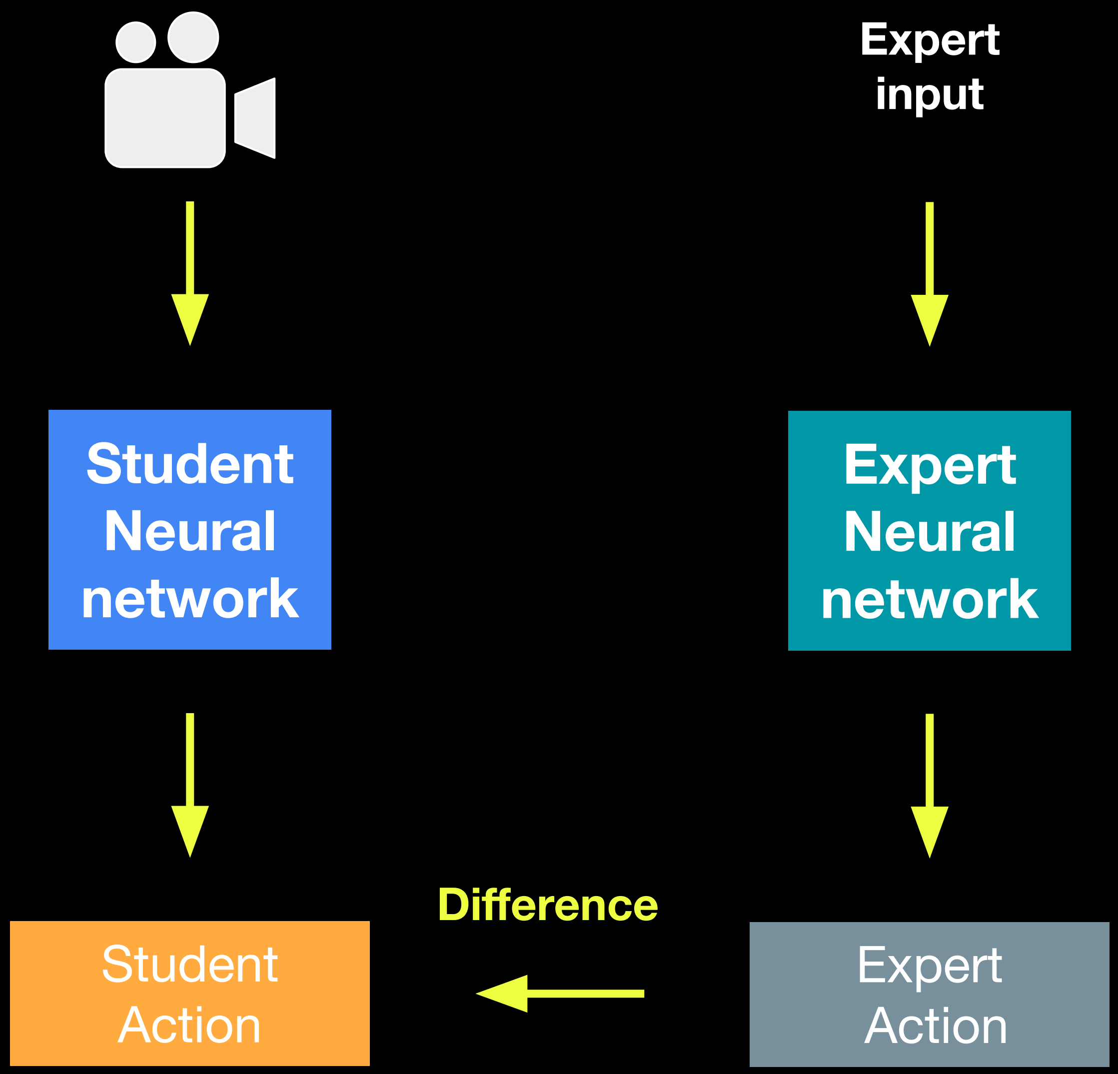
$$Q^\pi(s,a) = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s, a_t = a\right]$$

DDPG

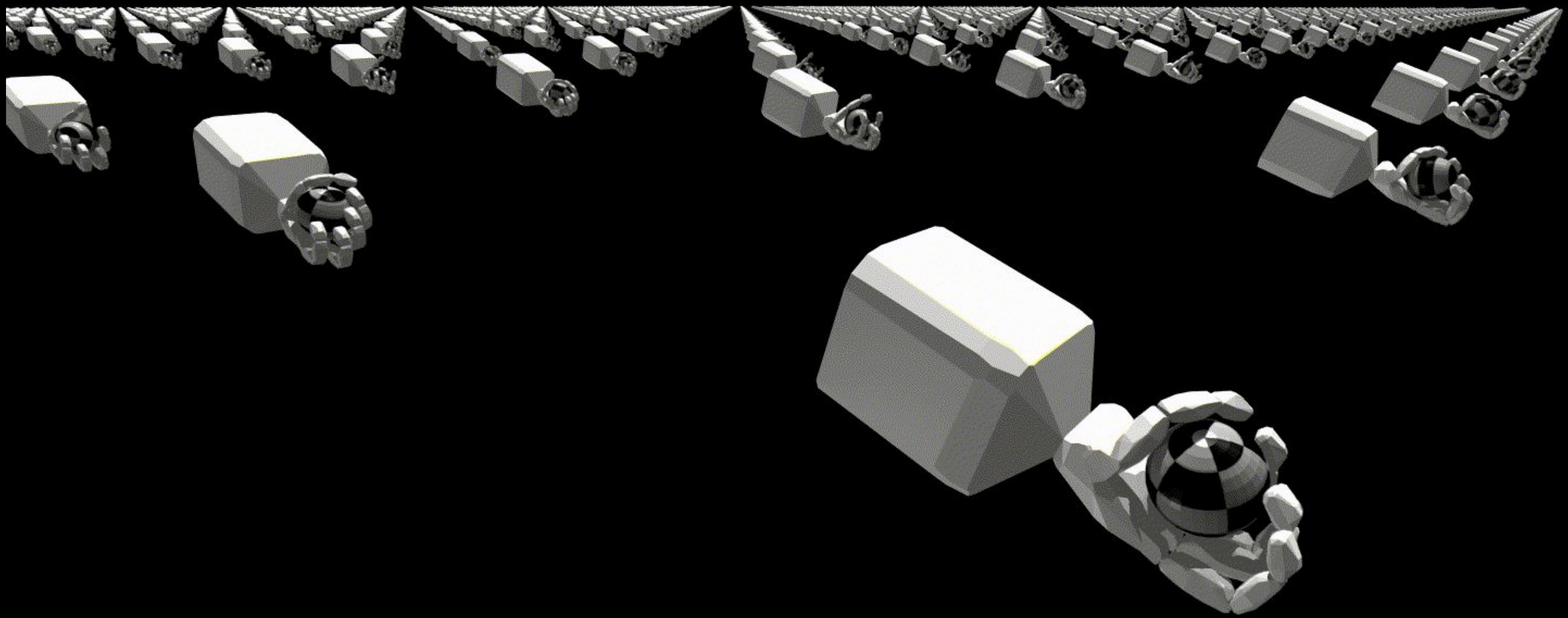
PPO

SAC

Recap: Imitation learning



Recap: Training to Reality Gap



SRL