

Project overview: model-free and model-based agents

SAC

TD3

TD-MPC2

Actor $\pi_\theta(s) \rightarrow a$

Critic $Q_\phi(s, a)$ (x2)

Entropy α, \mathcal{H}

Replay buffer

Model-free, max entropy

Actor $\pi_\theta(s) \rightarrow a$

Critic $Q_\phi(s, a)$ + target

Delayed policy, clipped Q

Replay buffer

Model-free, deterministic

Encoder $s \rightarrow \vec{z}$

Dynamics $d(\vec{z}, a)$

Reward, Q , policy heads

Opponent model

Model-based, latent planning

Hockey Environments s_t, a_t, r_t, s_{t+1}