

# Temporal Difference Learning for Model Predictive Control

Nicklas Hansen, Xiaolong Wang\*, Hao Su\*

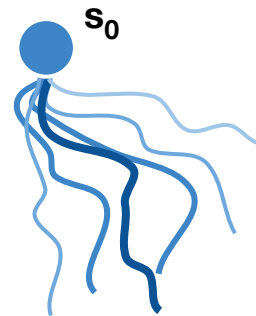
ICML 2022

UC San Diego

# Data-Driven Model Predictive Control (MPC)

- Plan using a **learned** model of the environment

- Objective  $\mathbb{E}_{\Gamma} \sim \Pi_{\theta} \left[ \sum_{t=0}^{\infty} \gamma^t r(s_t, \mathbf{a}_t) \right]$  intractable



(repeat for  $\infty$  steps)

# Data-Driven Model Predictive Control (MPC)

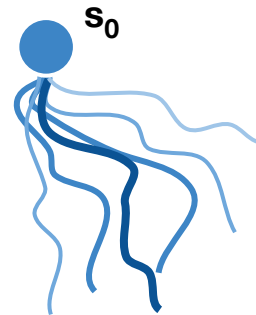
- Plan using a **learned** model of the environment

- Objective  $\mathbb{E}_{\Gamma} \sim \Pi_{\theta} \left[ \sum_{t=0}^{\infty} \gamma^t r(s_t, \mathbf{a}_t) \right]$  intractable

- Instead find **locally optimal** trajectory  $\mathbb{E}_{\Gamma} \sim \Pi_{\theta} \left[ \sum_{t=0}^H \gamma^t r(s_t, \mathbf{a}_t) \right]$

- **Two major challenges:**

- Compounding model errors
- Cost of long-horizon planning



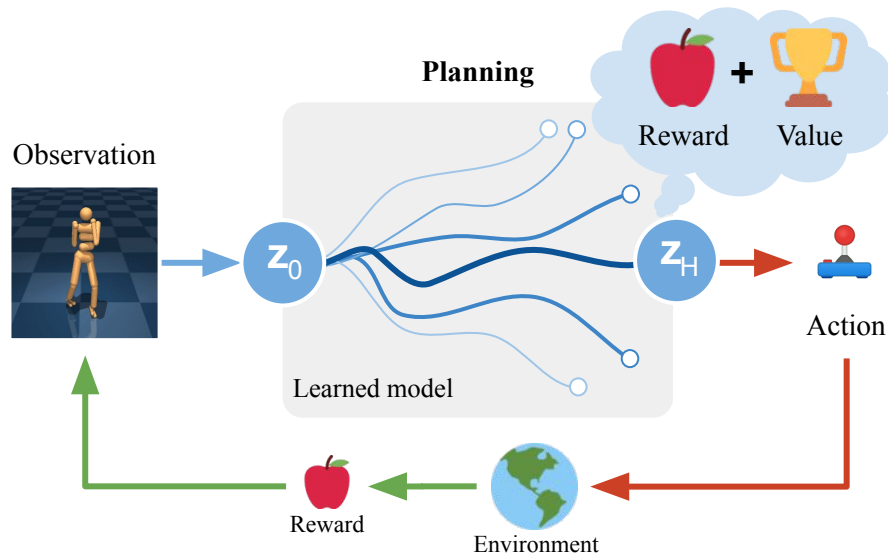
(repeat for  $H$  steps)

# How can TD-learning help MPC?

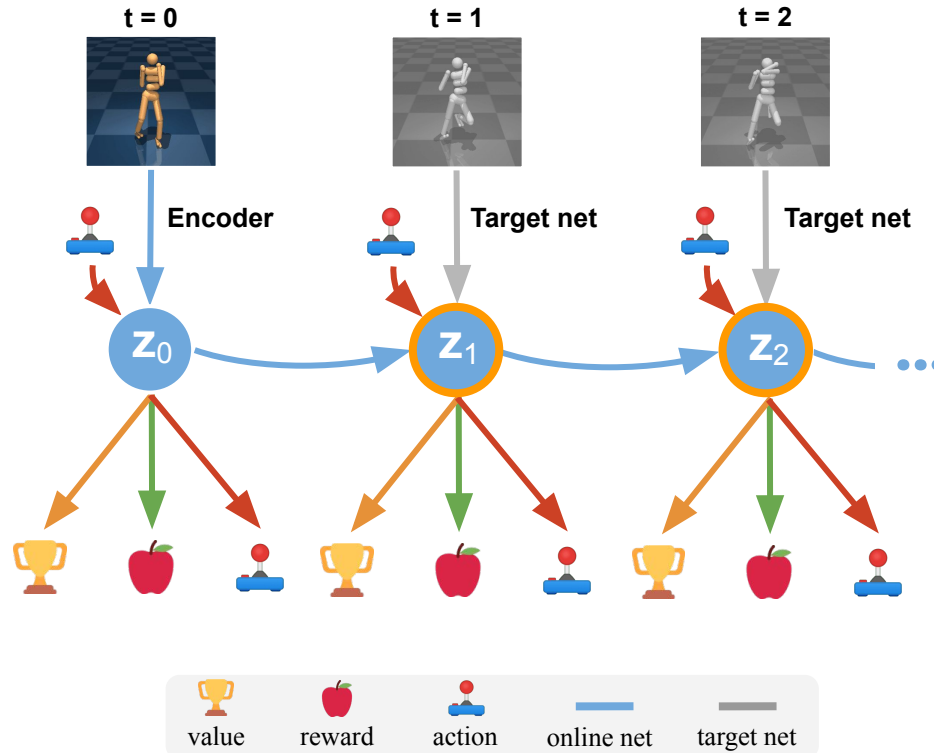
## Inference (planning)

- Planning in latent space
- Return estimate:

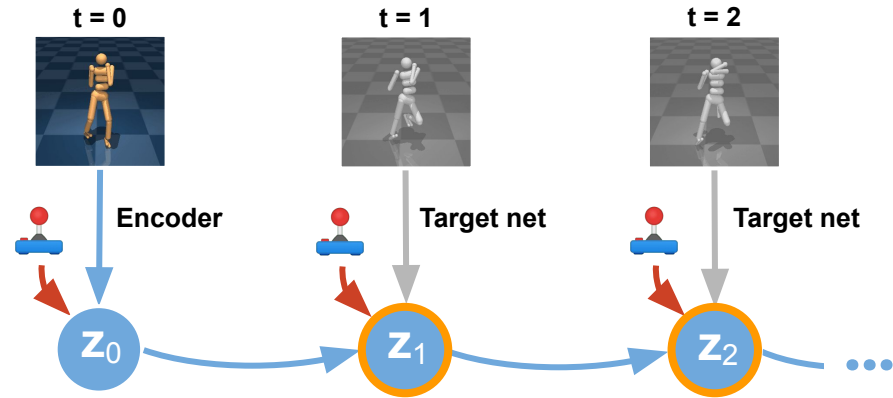
$$\mathbb{E}_{\Gamma} \left[ \underbrace{\gamma^H Q_{\theta}(\mathbf{z}_H, \mathbf{a}_H)}_{\text{Value}} + \underbrace{\sum_{t=0}^{H-1} \gamma^t R_{\theta}(\mathbf{z}_t, \mathbf{a}_t)}_{\text{Rewards}} \right]$$



# TD-MPC



# TD-MPC



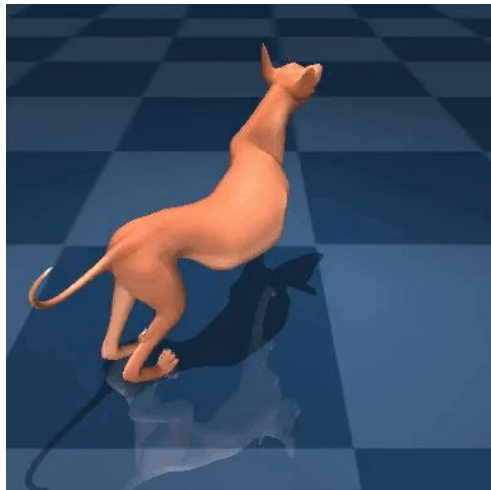
Minimize diff. between **recurrent prediction** and **target encoding**



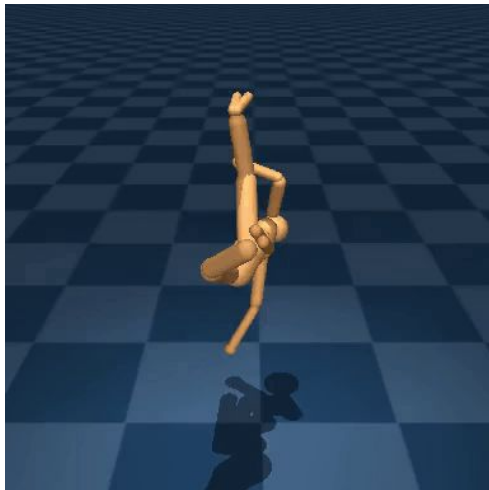
# Results

TD-MPC solves *challenging* continuous control problems

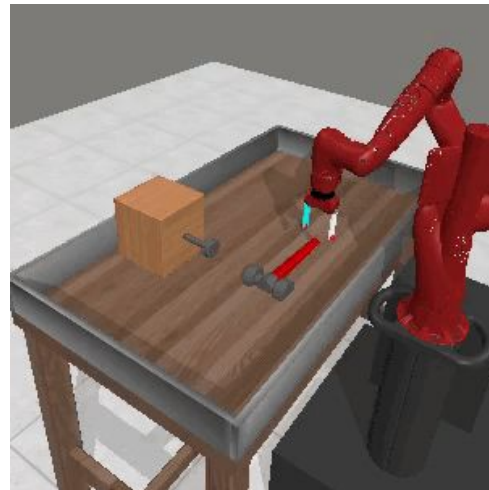
Dog Run



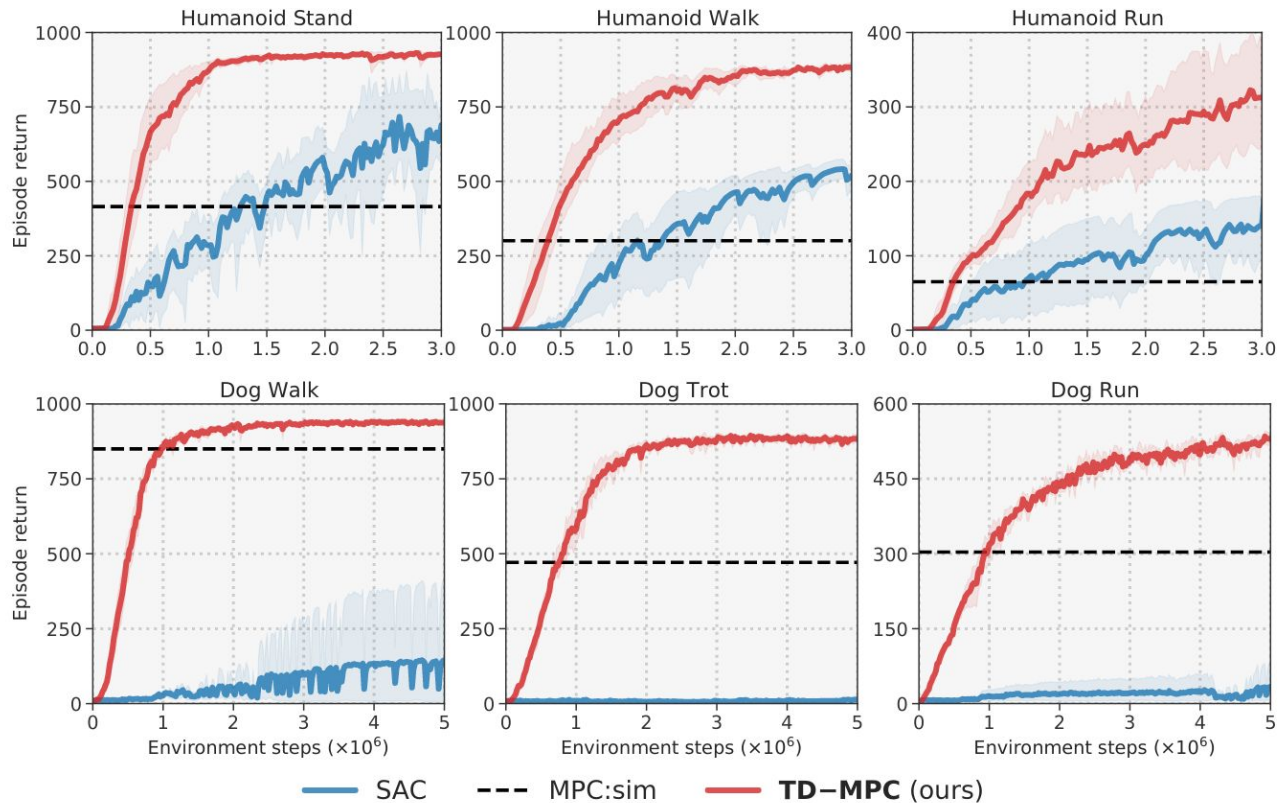
Humanoid Run



Hammer



# Results







**[nicklashansen.github.io/td-mpc](https://github.com/nicklashansen/td-mpc)**