Learning to Backflip: An RL Approach

Nishanth Rao, Kunpeng Liu, Nithasree Somanathan

INTRODUCTION

- In Reinforcement Learning (RL), an agent is trained to maximize the cumulative reward accrued based on the observations from environment.
- Model-free RL learns a controller directly from the reward and observations from the environment. Here, a DQN agent is used.
- RL is useful in scenarios where precise models might not be available or external environmental disturbances and uncertainties come into play.
- The motive of the project is to compare model-based control with RL-based control by executing a backflip on a Bipedal Spring-Loaded Inverted Pendulum (SLIP).
- The performance of RL is evaluated against Model Predictive Control (MPC) with trajectory optimization using iLQR.
- The MuJoCo simulator is integrated with Gym, along with sensor noise and random external disturbances.

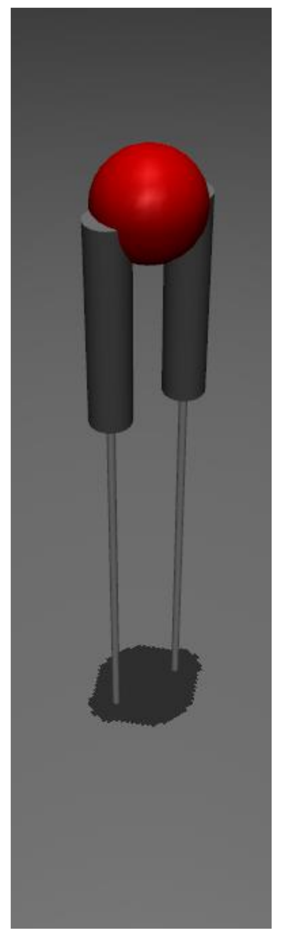


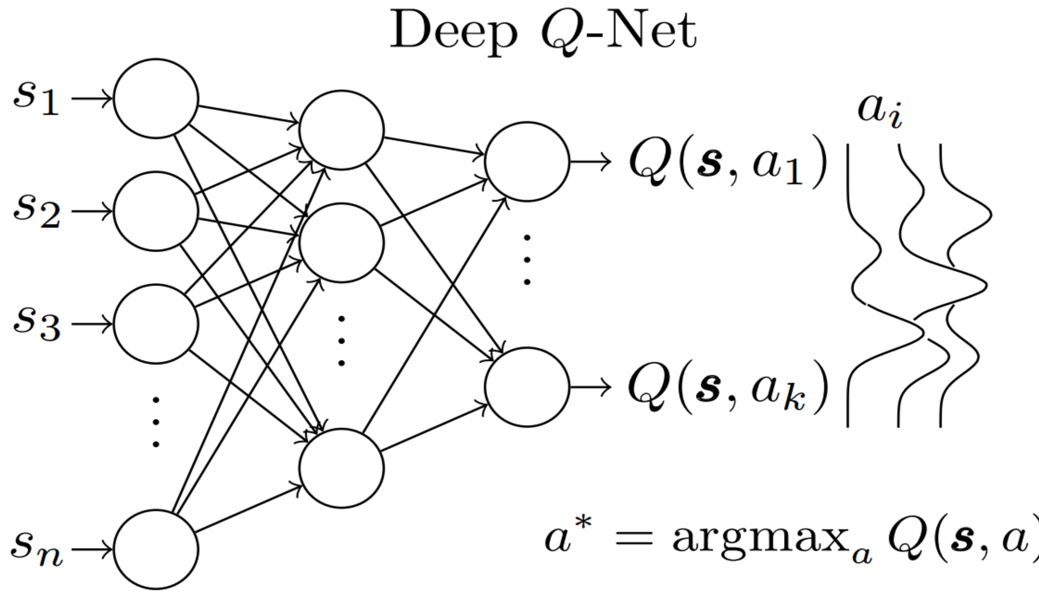
Fig 1: SLIP model (MuJoCo)

BACKGROUND

Q-learning:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_{t+1} + \gamma \max_a Q(s_t, a_t) - Q(s_t, a_t) \right]$$

Deep Q-Net



$$a^* = \text{argmax}_a\, Q(\boldsymbol{s}, a)$$

MPC:

$$\min_{x,u} \sum_{k=0}^{T-1} \left(\|x_k - \bar{x}_k\|_Q^2 + \|u_k - \bar{u}_k\|_R^2 \right) + \|x_k\|_{Q_f}^2$$

$$s.t \quad x_{k+1} = f(x_k, u_k)$$

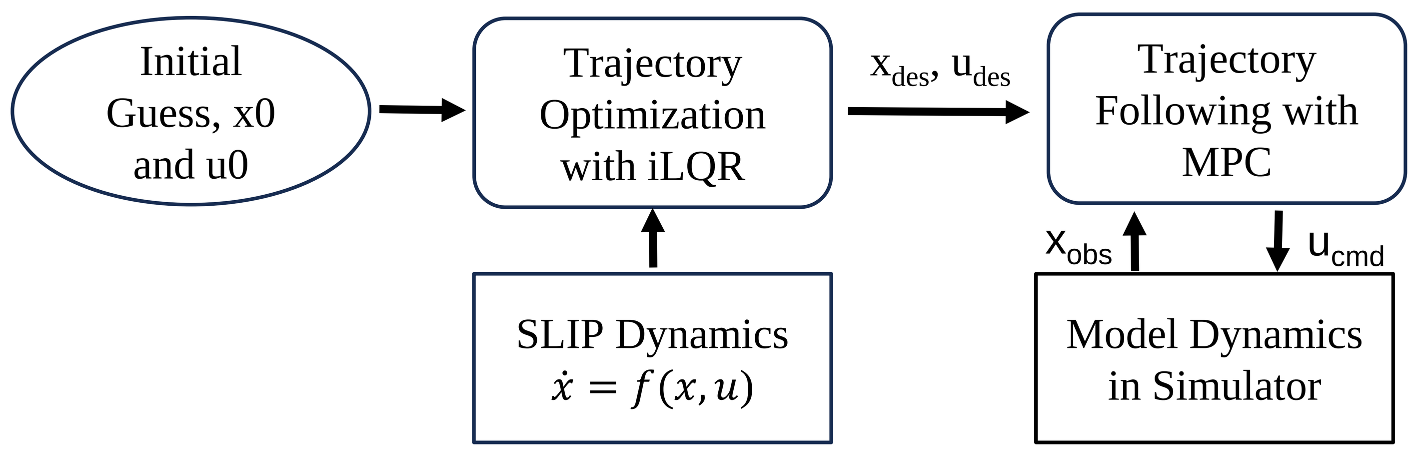
$$A \begin{bmatrix} x_k \\ u_k \end{bmatrix} \leq b$$

METHODS

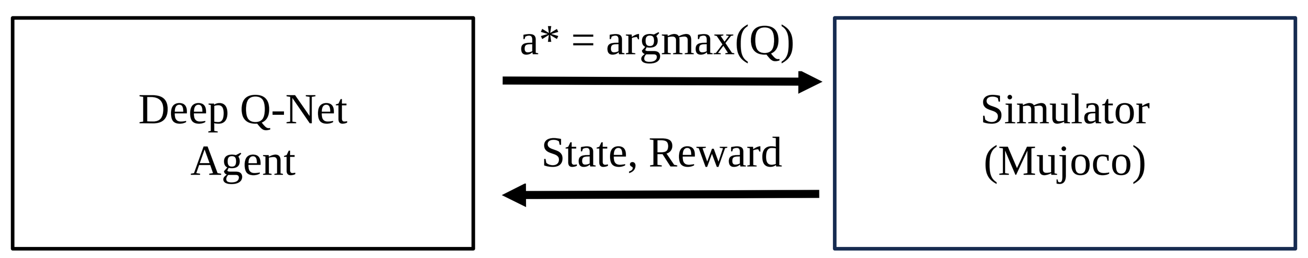
State and Control Variables:

$$x = [y, z, \theta, \dot{y}, \dot{z}, \dot{\theta}]$$
$$u = [\tau_1, \tau_2]$$

Model Based Control:



Model Free Control – RL :

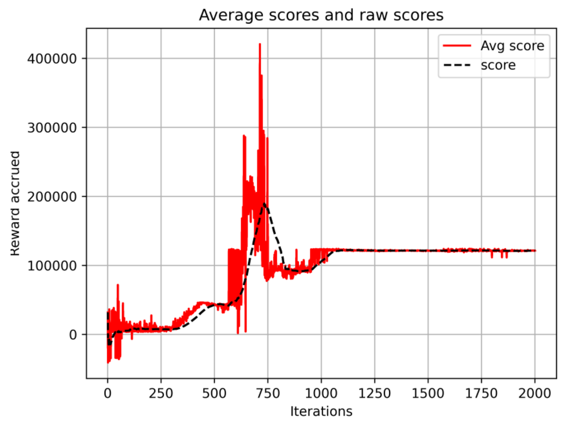


Reward Function:

If alive:
$r_t \propto \dot{\theta}_t$
$r_t \mathrel{+}= k_{alive}$
if contact and $|\theta| \leq 20°$:
$r_t \mathrel{+}= k_{upright}$
else:
$r_t \mathrel{-}= k_{fail}$

Training Objective:

$$\min \|Q_d - \hat{Q}\|^2$$
$$s.t \quad Q_d = r_t + \gamma \max_a Q$$



CONCLUSION

- Our comparative study highlights the efficiency of Reinforcement Learning in executing a backflip maneuver on the bipedal SLIP in presence of sensor noise and external disturbances against model-based control approach.

- Future prospects include exploring model-free RL algorithms to more complex tasks of robotic systems like Humanoids.

RESULTS

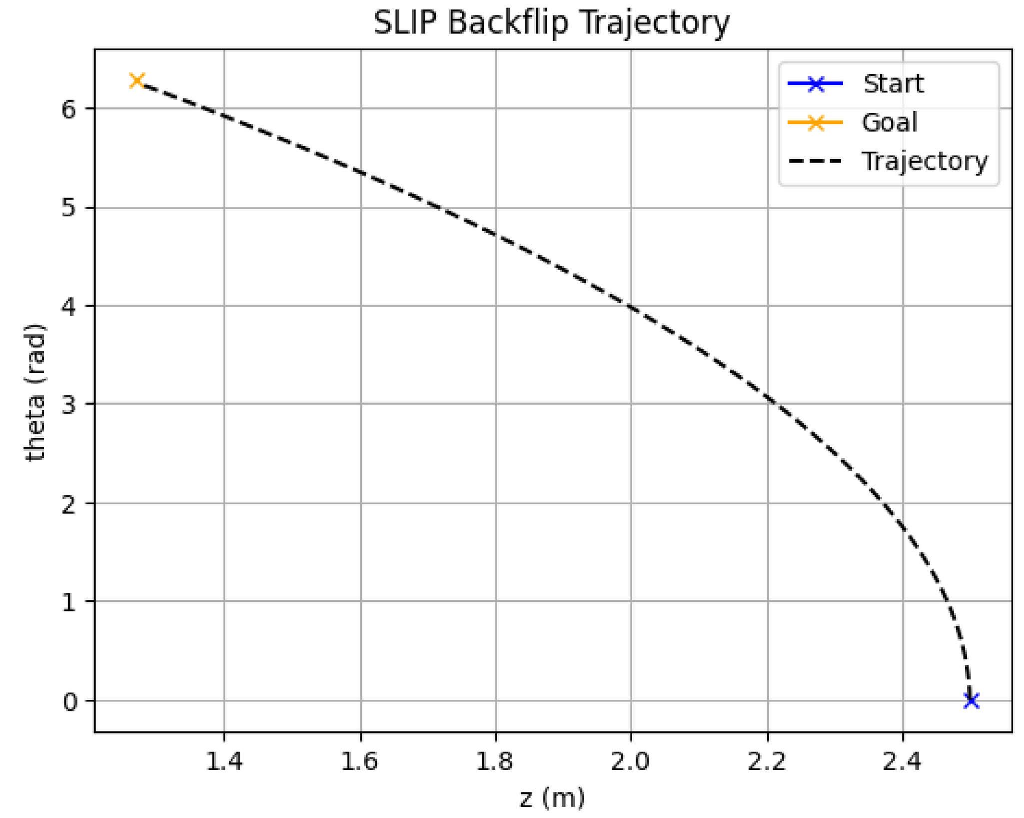


Fig 2: Numerical Trajectory tracking

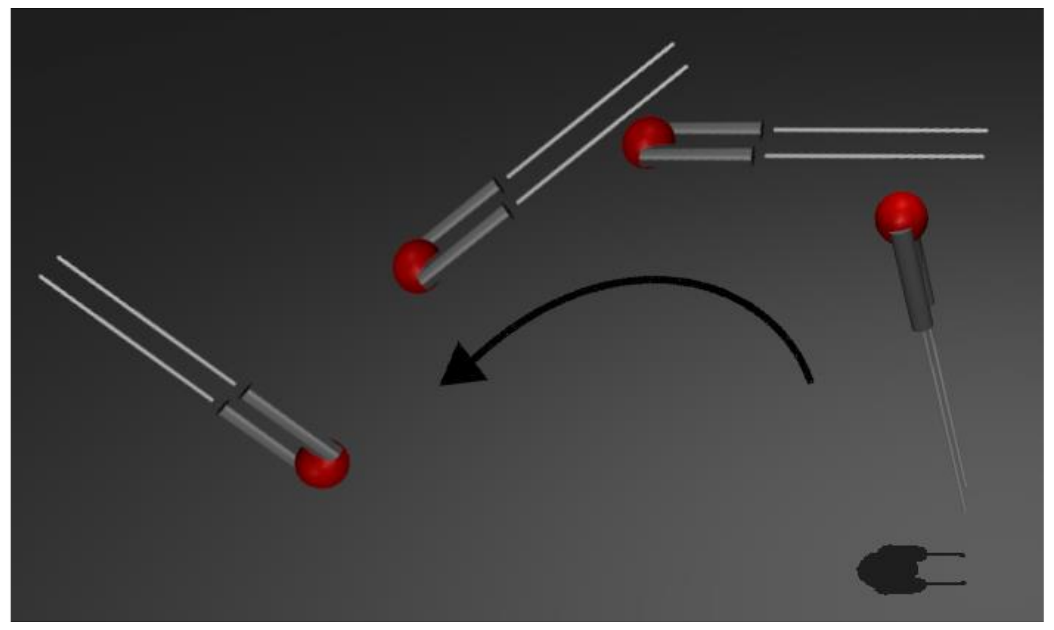


Fig 3: Model Based Control in MuJoCo with disturbances

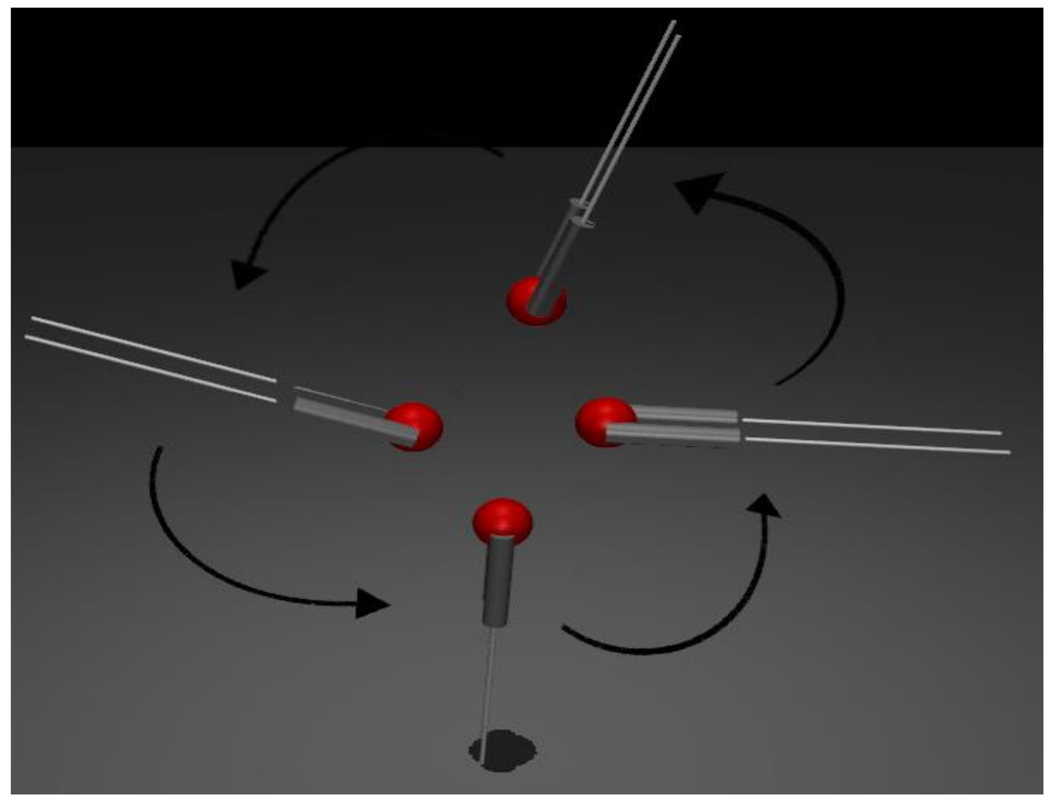


Fig 4: RL Control in MuJoCo with disturbances