Training P-H-D (σ=2)

REINFORCE
DQN

training returns
episode