**During training the random picking policy**

**During testing maximum value under picking policy**

**Terminology**

Model-free visual Reinforcement learning

Sample inefficiency

Action space

Structural encoding

Policy

Maximum Value under Placing

Sim-to-real transfer

Domain Randomization\

nsteadof a two step planner and local controller, we propose todirectly use model-free visual learning, which should alleviatethe state-estimation problem along with working with the truecomplex dynamics of the manipulated objects.

vIn model-based RL, we learn the dynamics of the system. In the ACframework, we learn both the policy (actor) and value function(critic). Finally, in Q-learning we often learn only the valuefunction, and choose actions that maximize it