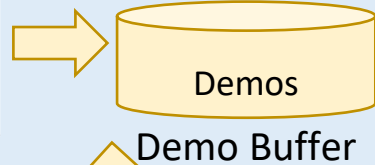


Pre-training



Demo Buffer

Actor
 $\pi_{rl}(s)$

Critic
 $Q(s, a_{eef})$

SAC Update

BC
 $\pi_{bc}(s)$

DBC
 $\pi_{dbc}(s)$

BC Update

MoE
 $G(s)$

MoE Update

Online Fine-tuning

Action

RL Policy

a_{rl}

BC Policy

a_{bc}

DBC Policy

$a_{gripper}$

MoE

a_{eef}

$\oplus$

a_{moe}

Human
Intervention

a_{jtv}

s

Environment

Robot Controller

Sensors

Reward Function

10hz

a

Policy
Transitions

Interventions

Replay Buffer

Interventions

Demos

Demo Buffer

Success
Trajectories

Demos

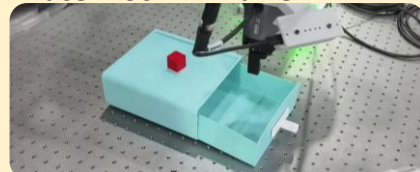
Success Buffer

SAC
Buffer

BC & MoE
Buffer

Final MoE Policy

Place Block in Drawer



Put Towel in Lidded Box



Insert Two Sockets



Double-Fold the Towel

