



● : 环境样本      ●/● : 模拟样本       $P$  : 环境状态转移分布       $M$  : 模型状态转移分布       $D_{env}/D_{model}$  : 经验回放缓存