论文汇报

北京航空航天大學 BEIHANG UNIVERSITY

题目：

Precise and Dexterous Robotic Manipulation via Human-in-the-Loop Reinforcement Learning

Related Work： **RLPD：加入先验数据的强化学习**

SERL：加入人类演示的强化学习 **Intervention**

HG-DAgger：人类修正，但训练有监督网络 **RL**

HIL-SERL: Precise and Dexterous Robotic Manipulation via Human-in-the-Loop Reinforcement Learning

机器人强化学习问题建模

北京航空航天大學 BEIHANG UNIVERSITY

数学模型 Markov Decision Process

$$\text{MDP } \mathcal{M} = \{\mathcal{S}, \mathcal{A}, \rho, \mathcal{P}, r, \gamma\}$$

- $\mathcal{S}$ 状态空间
- $\mathcal{A}$ 动作集合
- ρ 初始状态分布
- $\mathcal{P}$ 转移概率
- r 奖励函数
- γ 折扣因子（调整长期与近期）

Policy：$\pi \longrightarrow E[\sum_{t=0}^{H} \gamma^t r(\mathrm{s}_t, \mathrm{a}_t)]\uparrow$

训练过程 $\longrightarrow \pi$

Target Policy
Behavior Policy
$\nearrow$ On - Policy
$\searrow$ Off - Policy

常见的 Policy：$\pi(a|s)$ 是一个高斯分布，去训练分布参数

Sparse reward function：何为**稀疏**

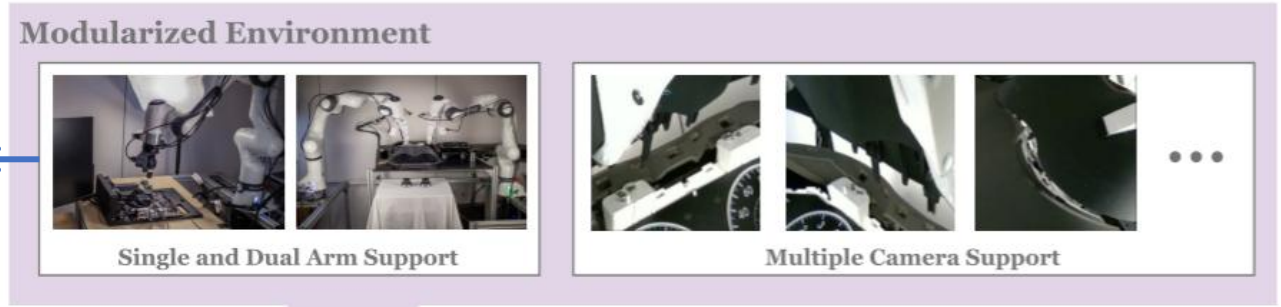
RLPD:

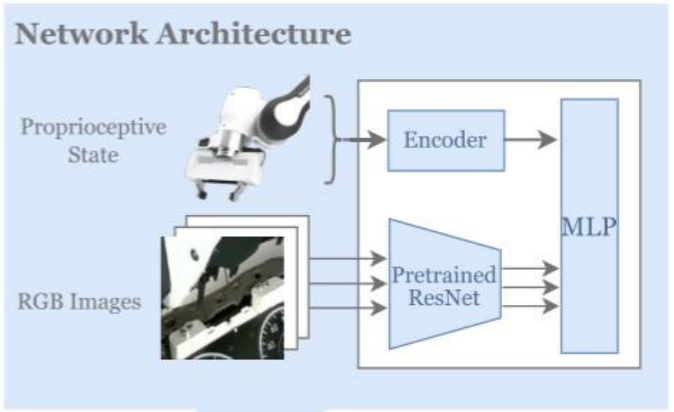
$$\mathcal{L}_Q(\phi) = E_{\mathrm{s,a,s'}}\left[\left(Q_\phi(\mathrm{s},\mathrm{a}) - \left(r(\mathrm{s},\mathrm{a}) + \gamma E_{\mathrm{a'}\sim\pi_\theta}[Q_{\bar{\phi}}(\mathrm{s'},\mathrm{a'})]\right)\right)^2\right]$$

$$\mathcal{L}_\pi(\theta) = -E_{\mathrm{s}}\left[E_{\mathrm{a}\sim\pi_\theta(\mathrm{a})}[Q_\phi(\mathrm{s},\mathrm{a})] + \alpha\mathcal{H}(\pi_\theta(\cdot|\mathrm{s})\right],$$

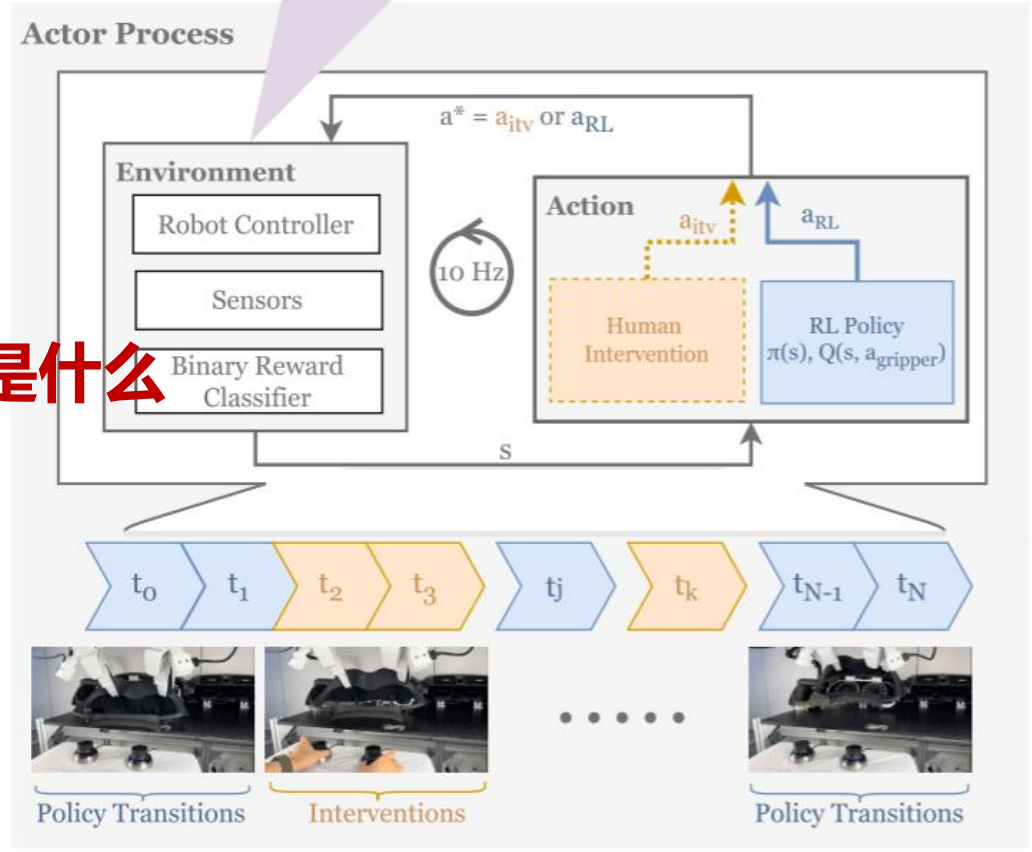
System Overview

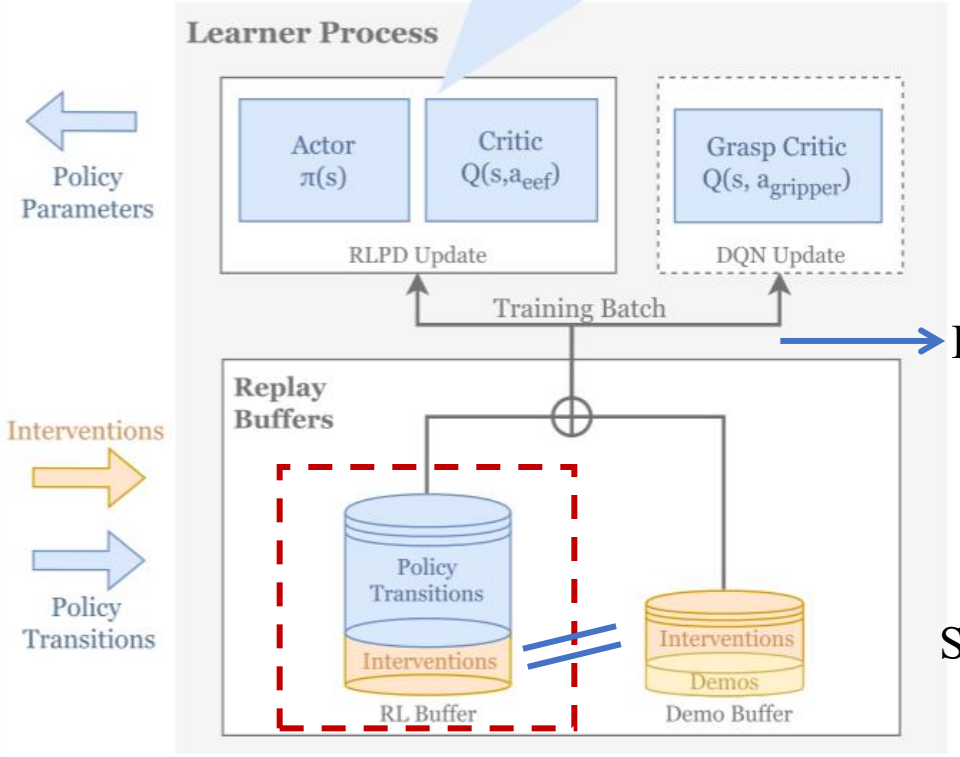
底层控制器
Careful
数据捕捉





Encoder 是什么





Policy Parameters

Interventions

Policy Transitions

RLPD

Interventions
>
RL Policy

SERL的区别

How to do in real world

Sample-efficient　Contact-rich tasks

1. Pretrained Vision Backbones

ImageNet → ResNet-10

与proprioceptive数据融合

2. Sparse Reward Function

3. Downstream Robot

相对末端坐标系描述感知信息 → Wrist Camera

阻抗控制器力控

4. Gripper Control → 设计小负惩罚，防止不必要的动作

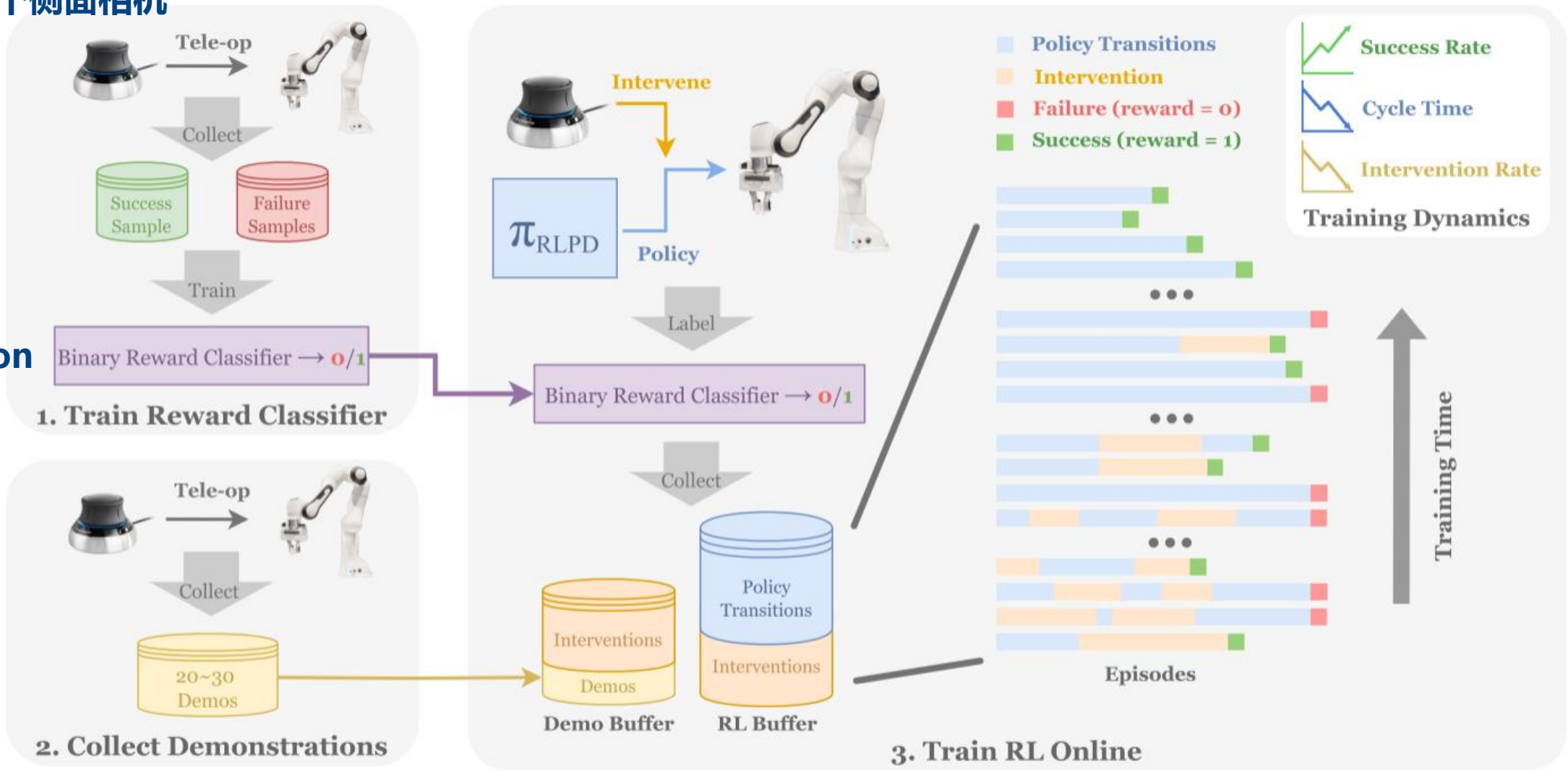
离散动作难训练：

弄一个离散动作MDP使用DQN训练

Training Process

腕部相机➕几个侧面相机

eward Function



避免早期的长期稀疏干预

Shortcoming

并非所有任务都能HIL

本身需要一个reward function

并没有进行长任务和广泛泛化性的实验

Shortcoming

什么是BC

什么是BC