



香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

基于Isaac gym的机械臂抓取强化学习项目复现

刘宇飞

数据科学学院
香港中文大学 (深圳)

2025. 12. 30

目录

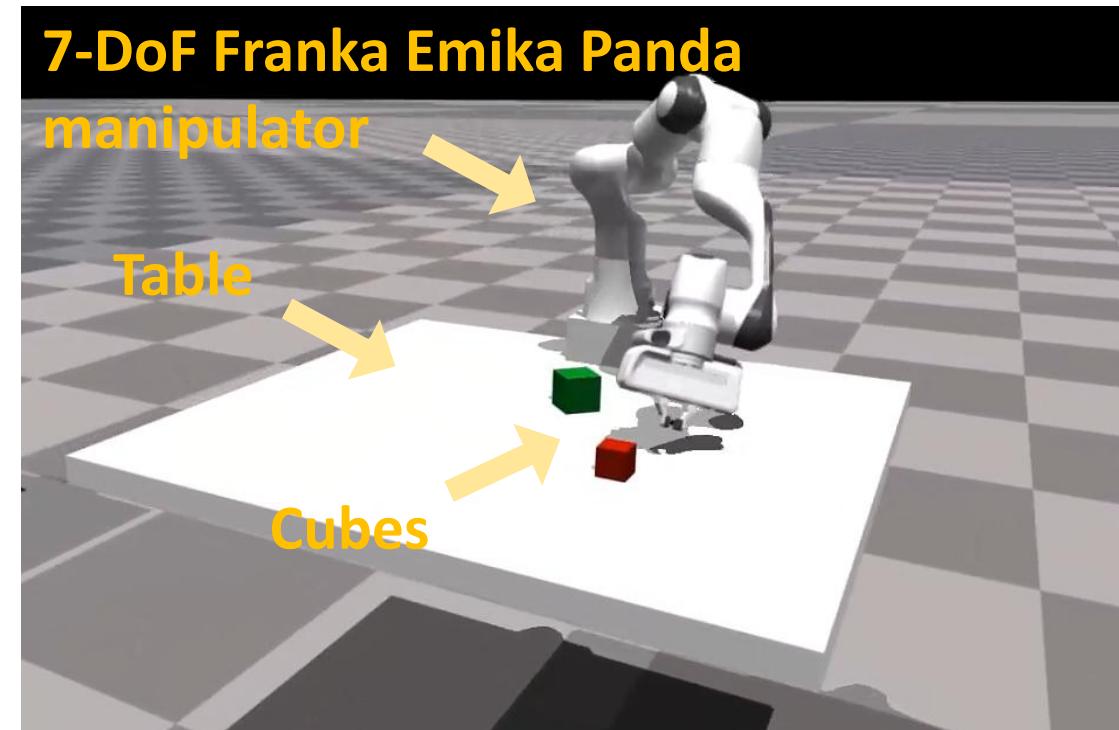
1. 项目介绍

2. 项目方法

3. 总结

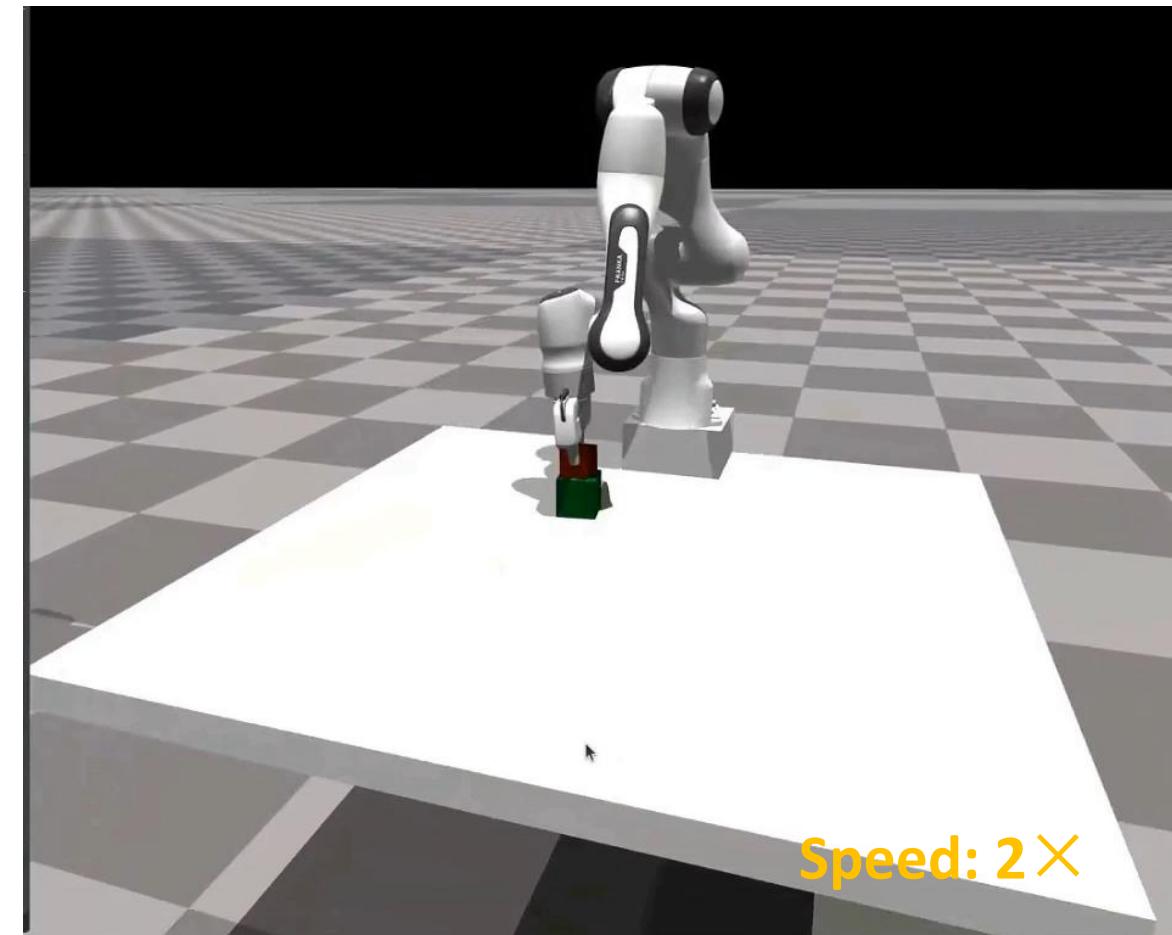
1. 项目介绍

- 训练了一个策略使得机械臂可以抓取随机生成在桌面的 **cube A** 并将其堆叠到 **cube B** 之上.
- 在 **Isaac Gym** 平台上进行仿真训练, 采(**Proximal Policy Optimization**) **PPO** 算法进行训练.
- 绘制 **loss and reward** 曲线并且可视化训练结果



1. Introduction

- 训练了一个策略使得机械臂可以抓取随机生成在桌面的 **cube A** 并将其堆叠到 **cube B** 之上.
- 在 **Isaac Gym** 平台上进行仿真训练, 采(Proximal Policy Optimization) **PPO** 算法进行训练.
- 绘制 **loss and reward** 曲线并且可视化训练结果



Content

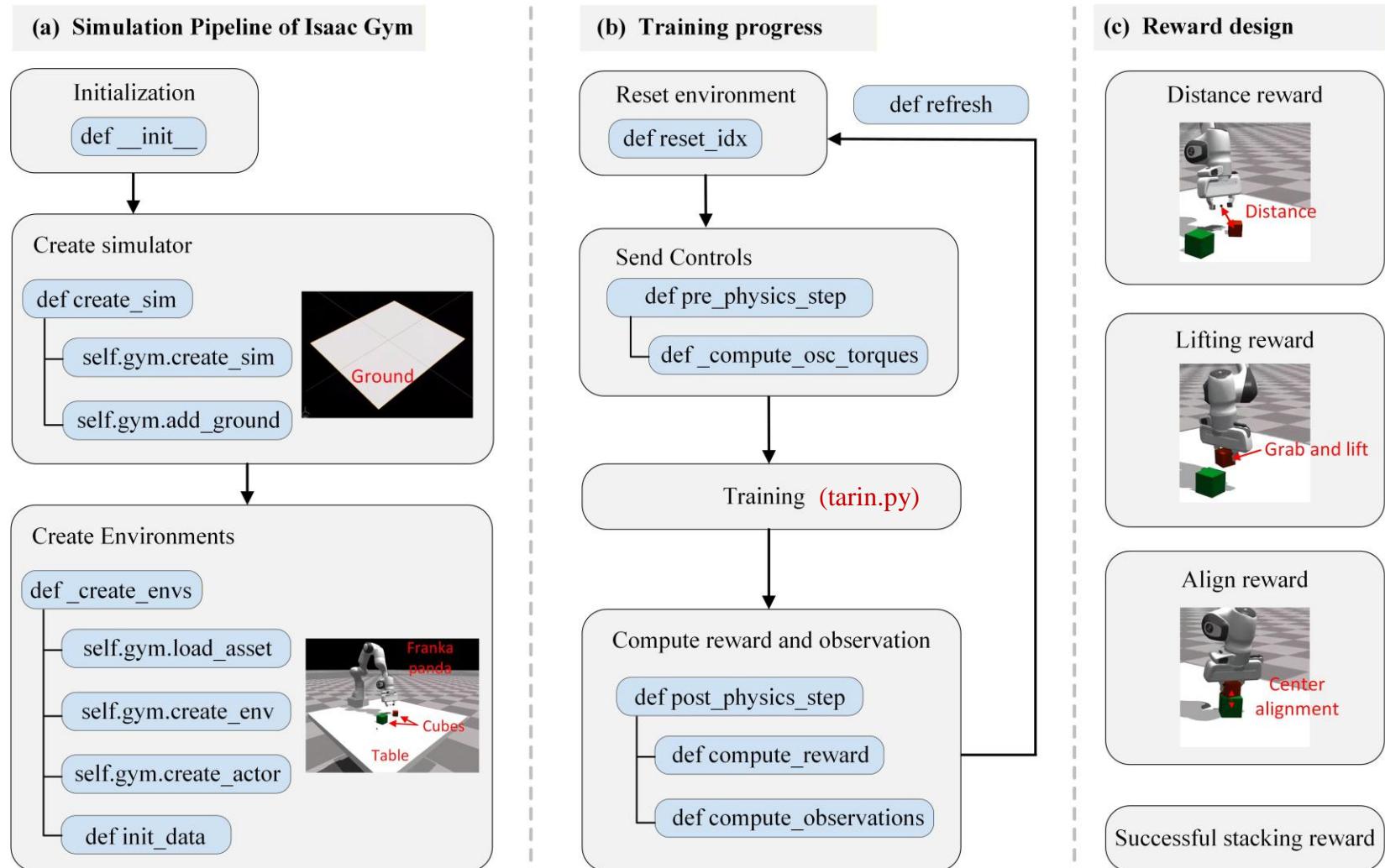
1. 项目介绍

2. 项目方法

3. 总结

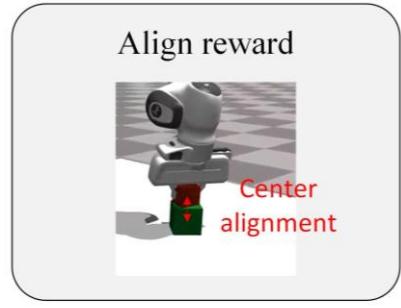
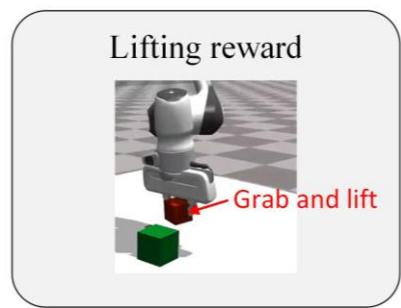
2. 1 Isaac gym仿真框架

- 仿真环境搭建、资产加载.
- 设计仿真框架pipeline



2. 2 奖励设计

(c) Reward design



□ 通过设计多阶段奖励函数，引导策略逐步学习并掌握复杂操作行为

Table 2: Reward setting

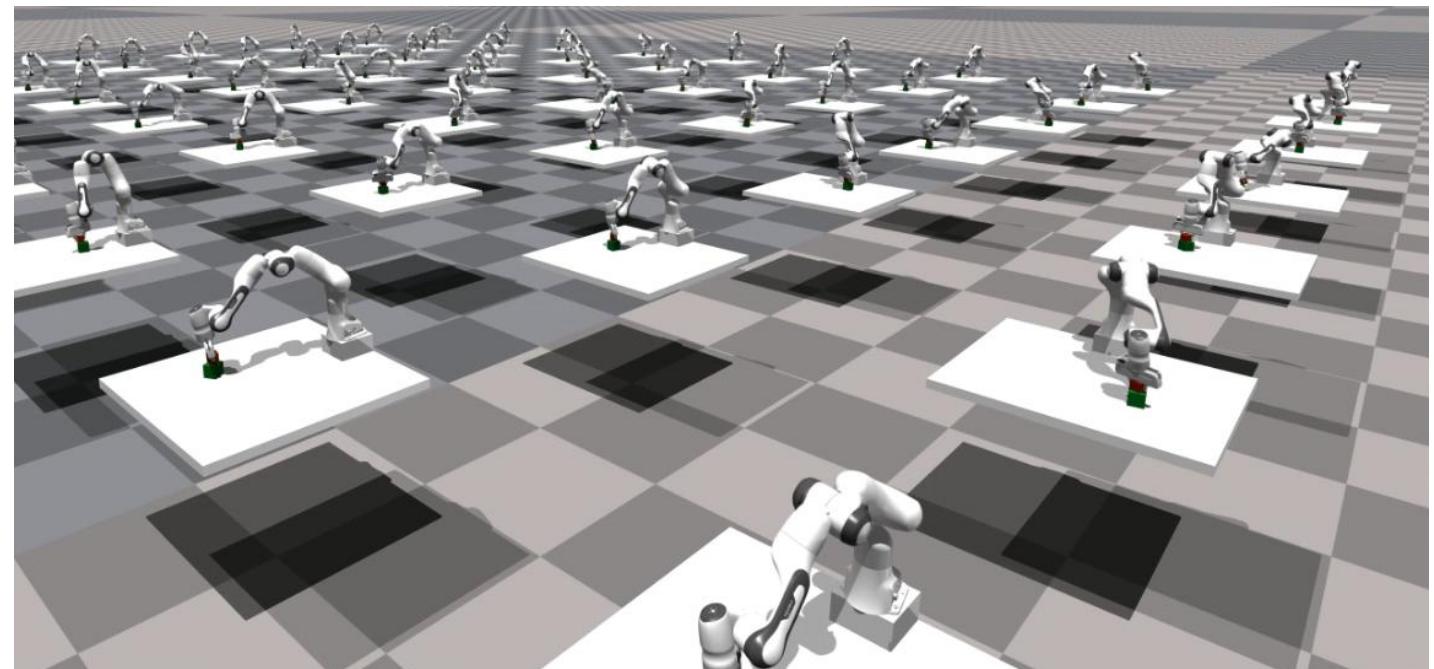
Reward type	Compute value	Max value	Reward condition (mm)
Distance reward	$1 - \tanh\left(10 \cdot \frac{d_G + d_L + d_R}{3}\right)$	0.1	Always
Lifting reward	$h_A - w_A$	1.5	$h_A - w_A > 40$
Align reward	$1 - \tanh(10 \cdot d_{A-B})$	2.0	$h_A - w_A > 40$
Shaped rewards			$\ p_A^{xy} - p_B^{xy}\ _2 < 20$
Successful stacking reward	16.0	16.0	$h_A - w_A > 40$
Sparse rewards			$d_{A-hand} > 40$

2.3 GPU并行环境训练

利用 GPU 并行环境显著加速训练过程

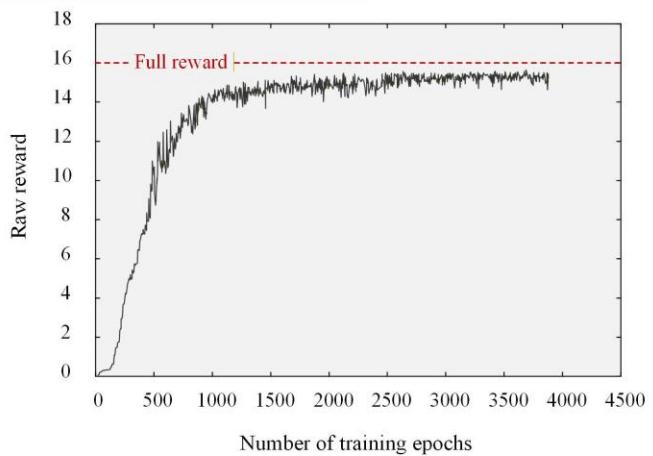
Table 1: Training parameters of PPO

Parameter	Value
Episode length of each iteration	500
Maximize epochs	10000
Horizon length	32
Learning rate	$5 \cdot 10^{-4}$
Number of environments	2048
Value loss coefficient	4
Target KL	0.008
PPO clip	0.2

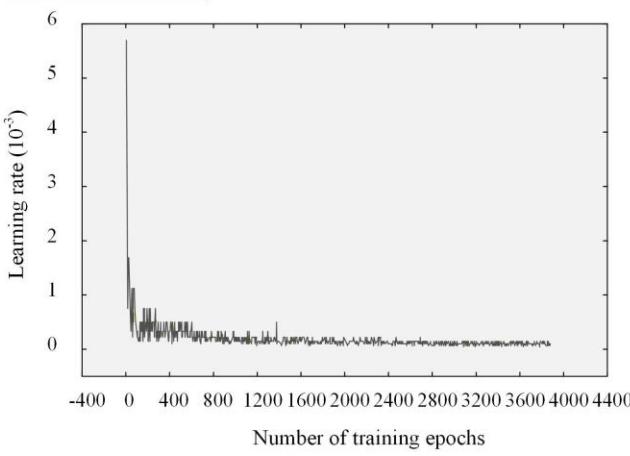


2.4 训练曲线绘制

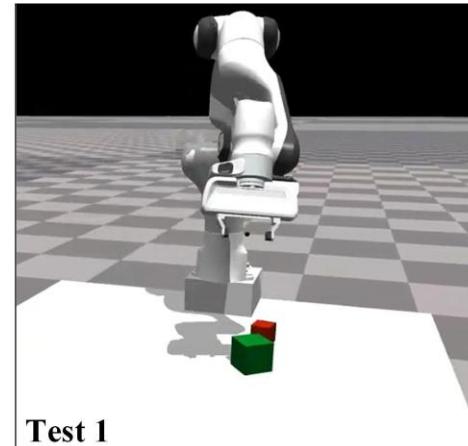
(a) Reward of each training epoch



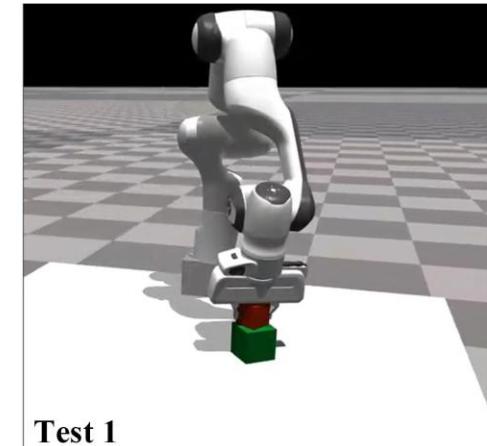
(b) Learning rate



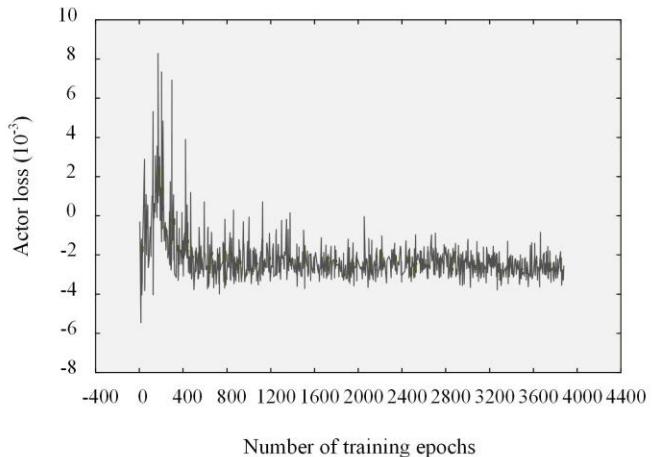
Test 1



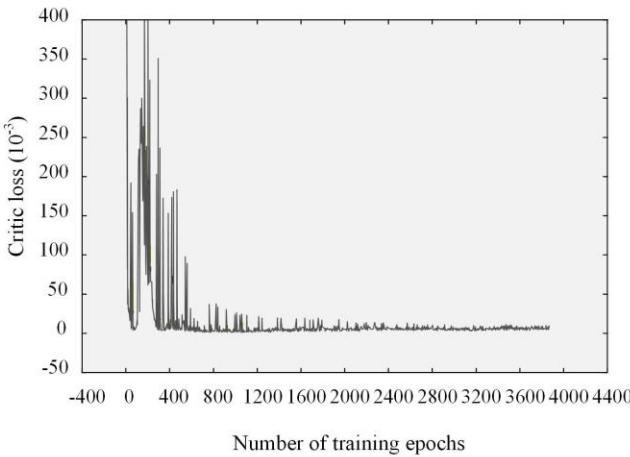
Test 1



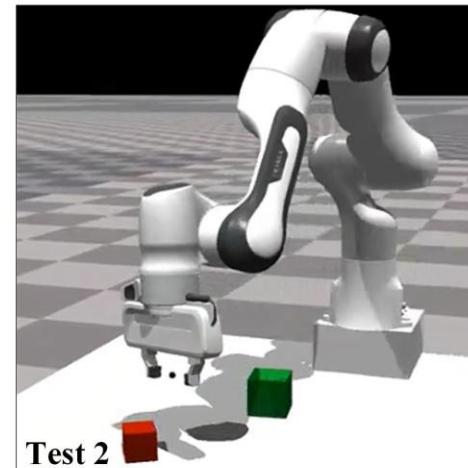
(c) Actor loss



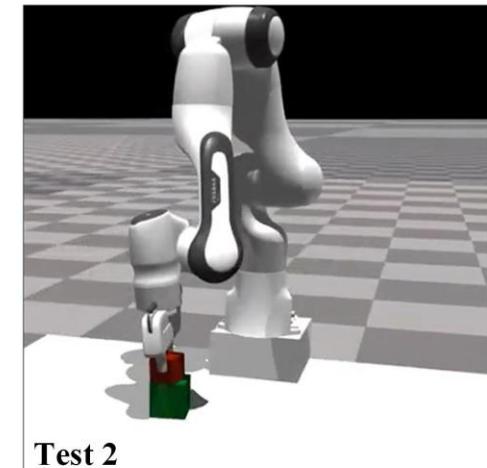
(d) Critic loss



Test 2



Test 2



Content

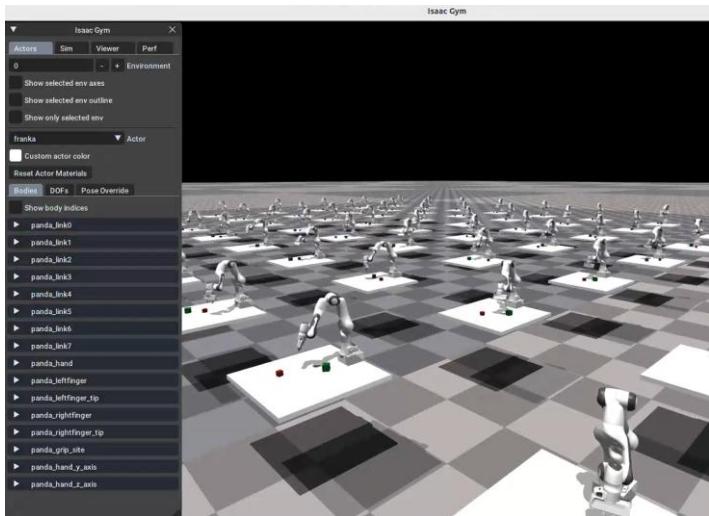
1. 项目介绍
2. 项目方法
3. 总结

3. 总结

基于Isaac gym的机械臂抓取强化学习项目复现

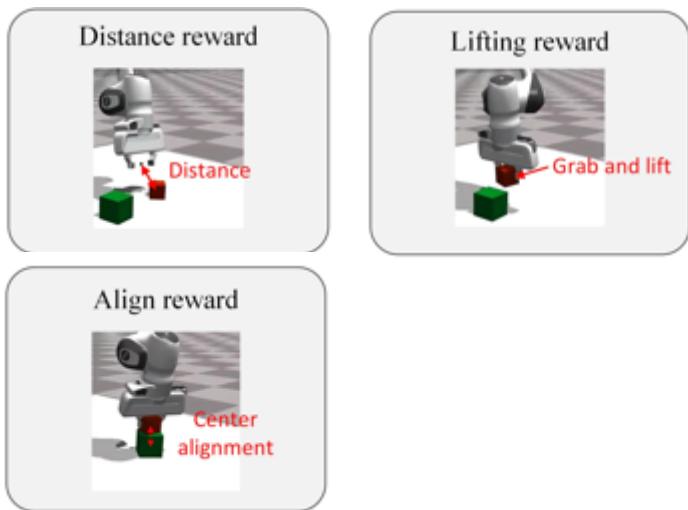
1 利用 Isaac Gym 进行仿真

- 在 Isaac Gym 中搭建仿真环境，进行资产加载.
- 利用GPU并行环境加速训练.



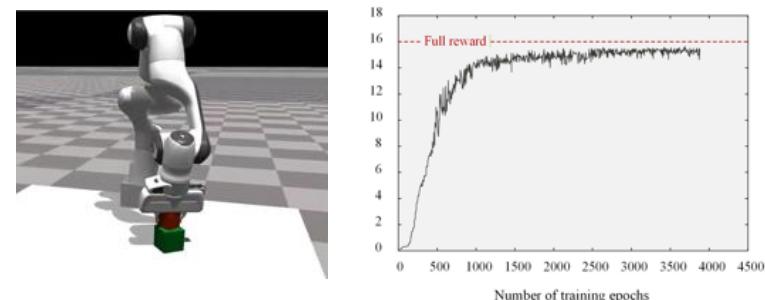
2 设计 PPO 算法奖励函数

- 采用 PPO 算法进行学习，设计了四个奖励函数，通过设计多阶段奖励函数，引导策略逐步学习并掌握复杂操作行为.



3 可视化训练结果

- 绘制训练过程中的loss和reward曲线，对训练结果进行可视化.





谢谢！

刘宇飞

数据科学学院
香港中文大学（深圳）

2025. 12. 30