上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

SJTU

CS3316强化学习课程项目

2025年2月28日

任务介绍

下列任务任选其一

- **IsaacGym-based Robot Learning Projects**
  - **RMA on quadrupeds**
  - **Humanoid Locomotion**
- **王者荣耀开悟(1v1)**
- **RFT on LLMs**
- **自选问题或环境（RL相关）**
  - 问题设定、预期方案
  - 请在3月4日前联系助教

单人项目要求

评价标准：

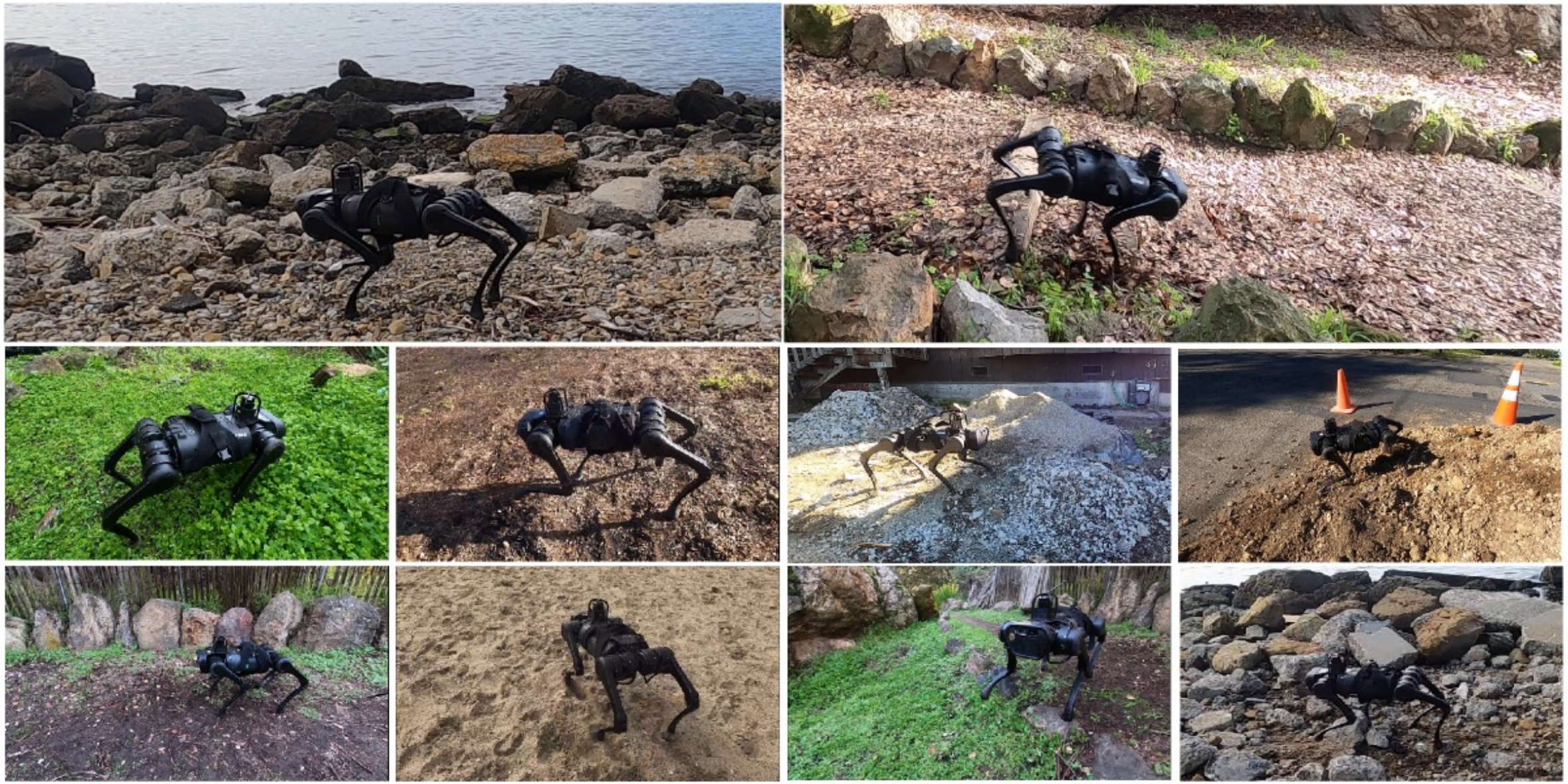
* 任务环境/完成度

提交方式：

提交内容：

- 正文长度4-6页
- Methods/Eval/Contrib…
- 代码压缩包

RMA for quadrupeds



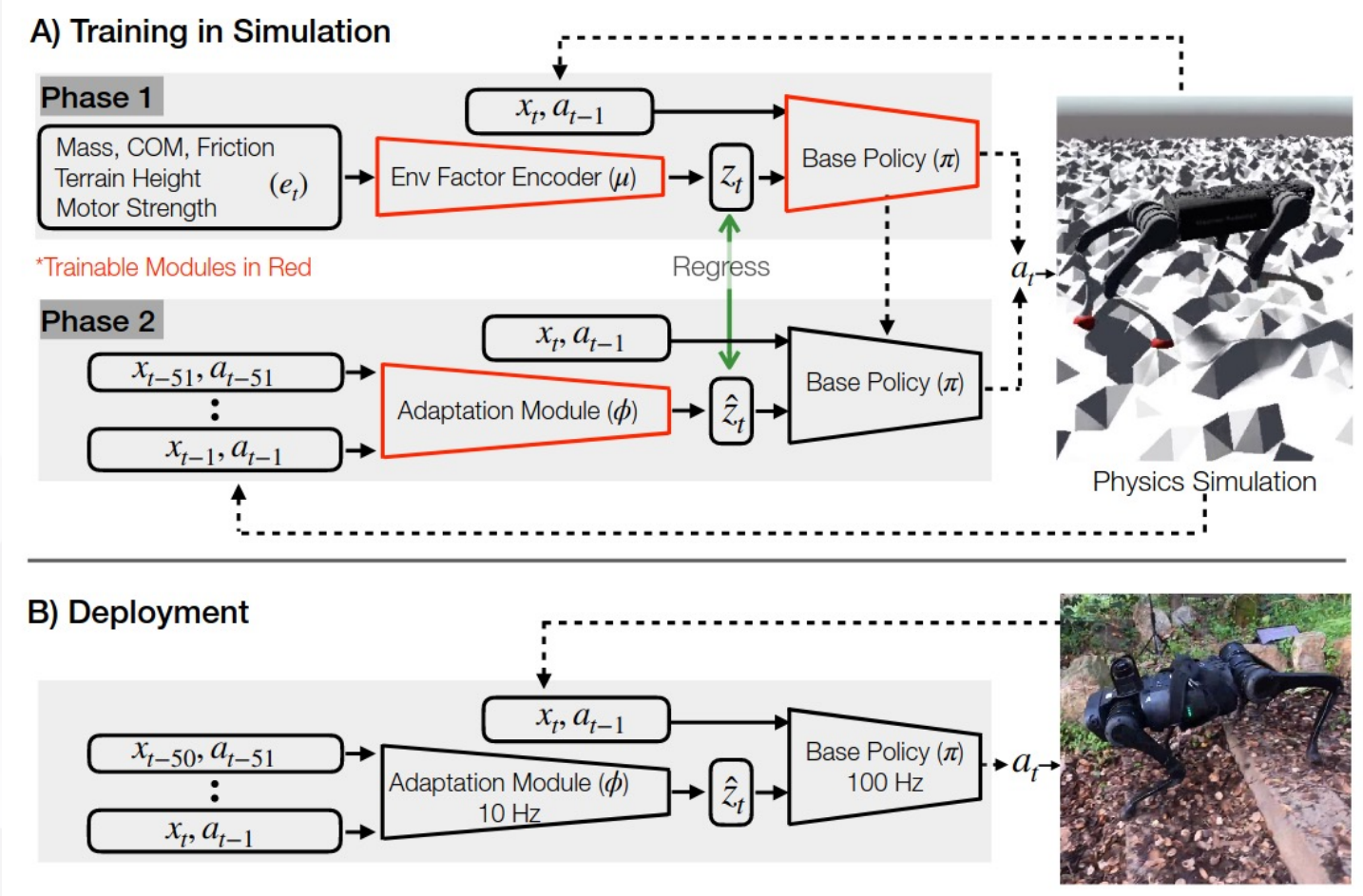
RMA for quadrupeds

- Rapid Motor Adaptation (RMA) is the baseline

  Based on legged_gym

  https://github.com/leggedrobotics/legged_gym
- You need to beat RMA in any of the aspects:
  - Overall Perf in varied terrains
  - Overall Perf in varied envs (mass, fric)
  - Learning efficiency in RL
- Some hints:
  - Improve RMA's z_t
  - Improve learning pipeline
  - Learning from real world (real2sim)
  - Tune rewards (not recommended)
- Experiments in simulation is enough
- Real-world onboarding is BONUS!

(We have Unitree A1 quadrupedal robots)

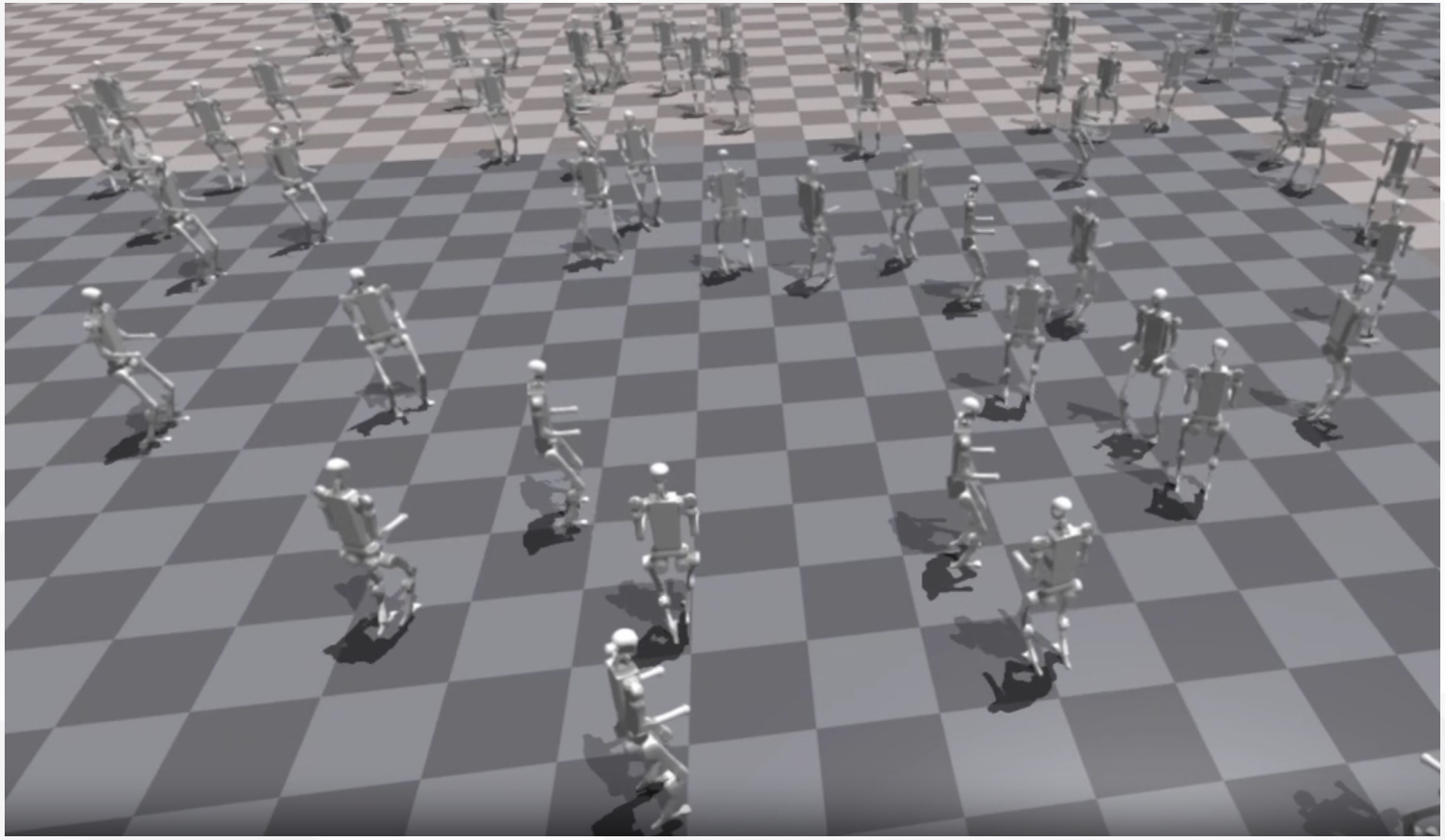


Humanoid Locomotion

Humanoid Locomotion

- Based on HumanoidVerse
  https://github.com/LeCAR-Lab/HumanoidVerse
- Baseline: Upper body fixed
- Goal: Upper body unfixed
- Little Bonus: Humanoid locomotion with upper body motions

- Experiments in simulation is enough
- Real-world onboarding is BONUS
- We have H1, G1 for onboarding



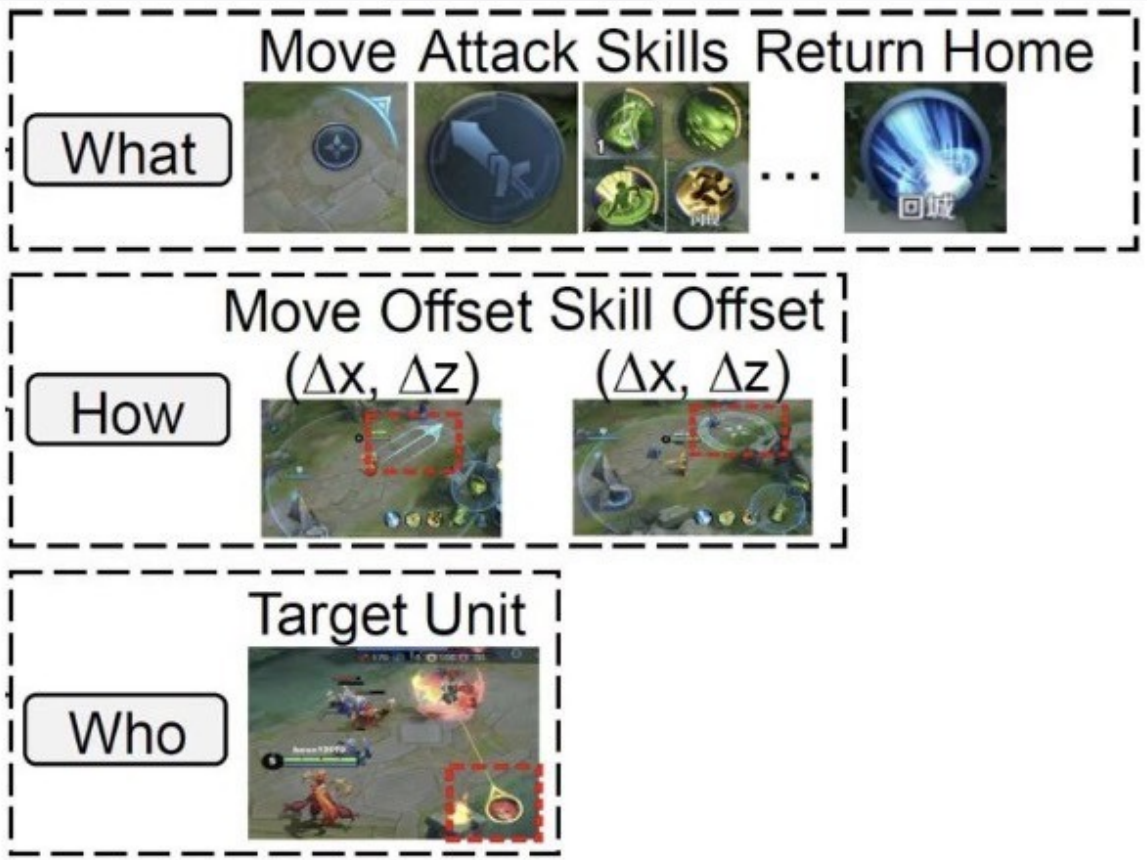
王者荣耀开悟

- 2-Player Competitive Game 1v1
- 双方控制一个游戏角色摧毁敌方基地即可获得胜利
- 2-3人一组

王者荣耀开悟

- 提供了简单的PPO算法实现，每个agent通过self play方式进行训练，请基于已有代码，对模型训练效果进行改进。

Evaluation:对baseline胜率或天梯赛



Reinforcement Fine-Tuning LLMs

- 推荐框架&Baseline：https://github.com/hkust-nlp/simpleRL-reason
- 通过改进RFT算法/架构，使得LLM在Math Reasoning上提升表现
- 需要有RL层面上的novelty，如算法、训练架构等
- Hint：
  - Baseline = Qwen2.5-Math+PPO RFT
  - Implement GRPO over PPO, do ablations
- 本课程可提供阿里云平台算力

自选项目要求

前期要求：

- 提交说明文档
  - 问题设定预期方案
- **RL相关！**

提交、评价方式：

- 和可选方案一致

额外说明

- 除自选项目外每个题目会提供简单baseline，只要各组实验结果超过baseline、mini paper和代码完整提交、过程符合学术诚信标准就可以获得大量基础分
- 机器人/狗、王者选题最好能提供视频（海报展示时）
  - Simply High Reward != High Performance
- 鼓励大家锻炼学术技能和创新性探索，不注重模型绝对性能

注意要点

- 课后确定问题后于**3月7日**前**填写选题表格**
- 自选课题组**3月7日**前**提交前期说明文档**
- **个人项目-第7周末**（4月6日24:00）前**提交paper和代码附件**
- 预计**第8周**举行模型和结果的答辩