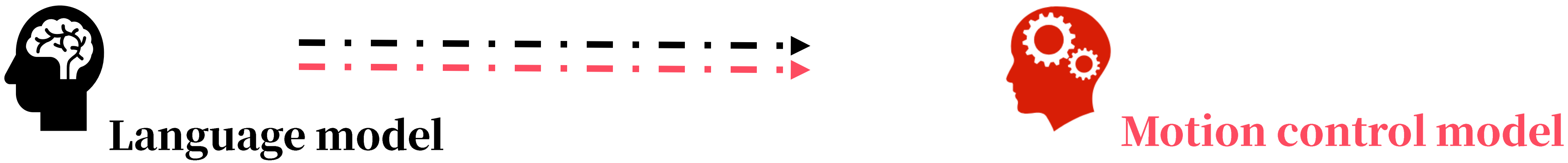
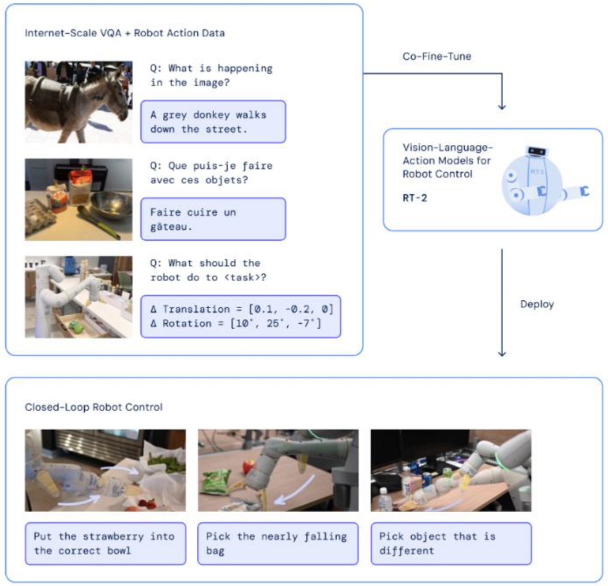
Unleashing the Power of Pre-trained Language Models for Offline Reinforcement Learning

Ruizhe Shi*[1], Yuyao Liu*[1], Yanjie Ze[2], Simon Shaolei Du[3], Huazhe Xu[124]

[1]*Tsinghua University, IIIS* [2]*Shanghai Qi Zhi Institute*
[3]*University of Washington* [4]*Shanghai AI Lab*
*Equal contribution. Order is decided by coin flip.



Motivation



Transformer architecture

Pre-train ↓ ↓ Generative

LLM era

QA, text translations, coding writing, image (or even video) generation…
Can LMs do more?

LLM + Robotics control

Introduction

IL

Fragile

Time-consuming

Security concern

Massive high-quality trajectories
Hard to collect/manual design

RL

Learn optimal policy from sub-optimal data by learning reward functions

Introduction

RL

Learn optimal policy from sub-optimal data by learning reward functions

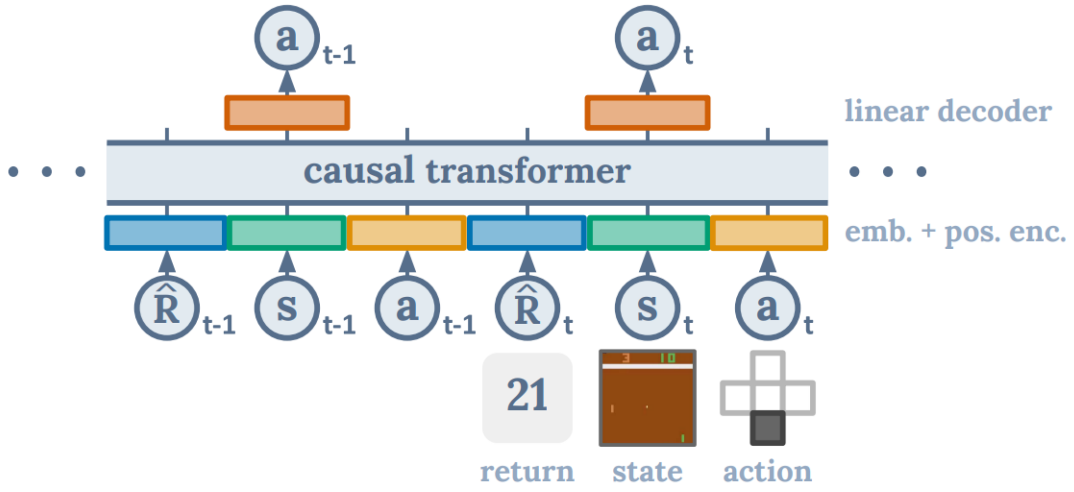
Online: collect data through interactions ✗

Offline: learn on pre-collected datasets ✓

- pre-collecting data is still expensive ⇒ few-shot learning

Motivation

Offline RL Baseline ——**Decision Transformer (DT)**



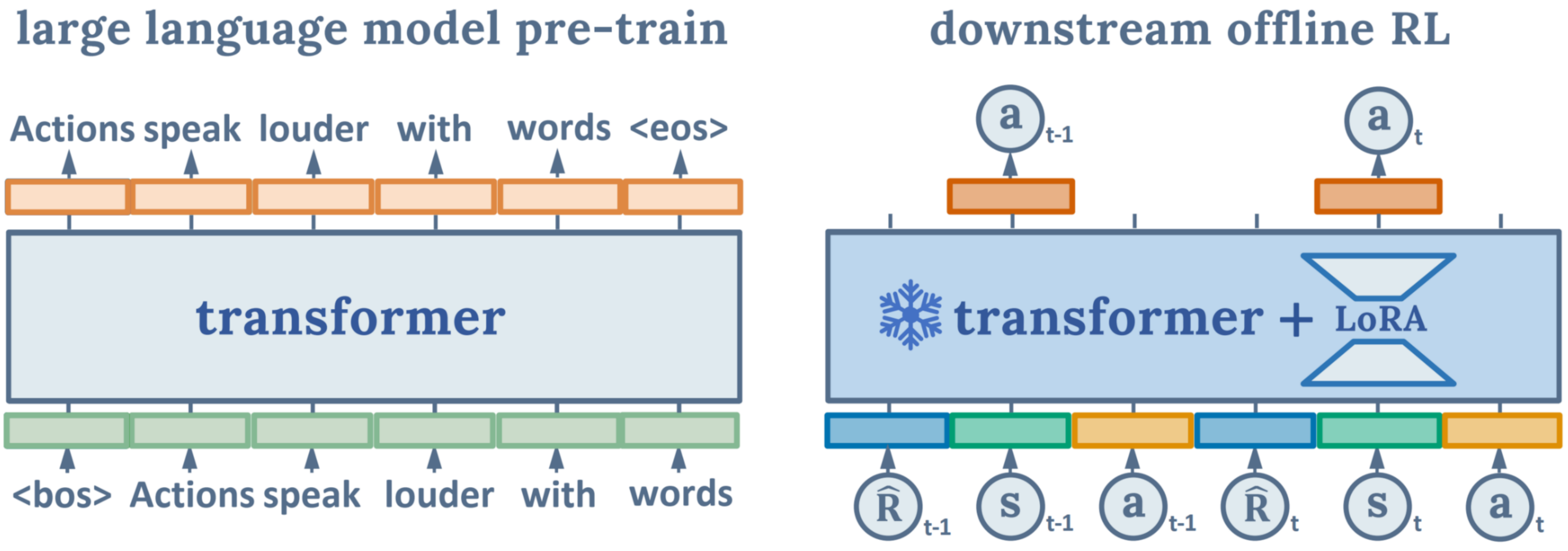
LM predict token:

$$P(\text{“you”}|[\text{“How”}, \text{“ ”}, \text{“are”}, \text{“ ”}])$$

Motion model predict action:

$$\pi(a_t|s_1, a_1, r_1, \dots, s_t)$$

LaMo: Language Models for low level Motion control

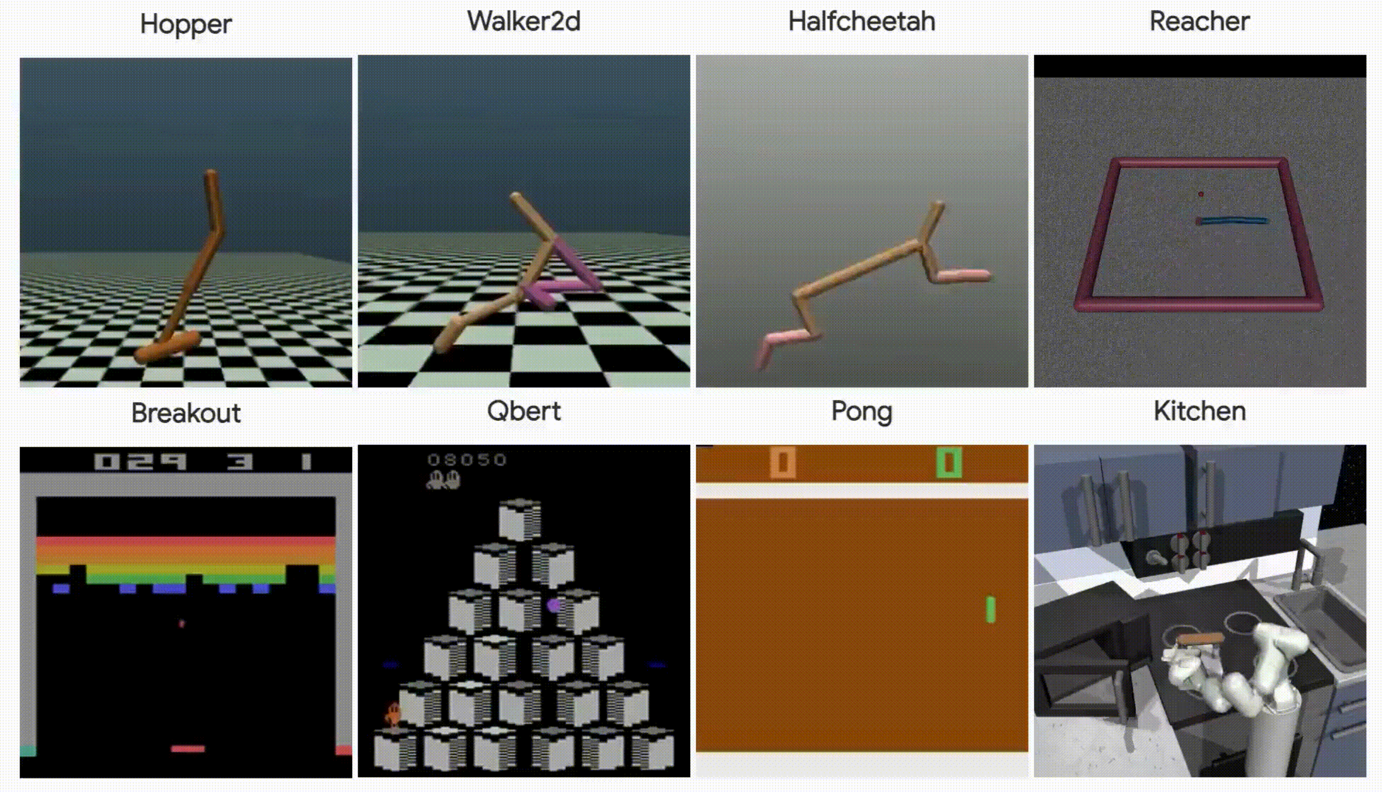


- knowledge from pre-training → Initialize with Pretrained LM
- retain the knowledge → Low Rank Adaptation (LoRA)
- enhancing representation → MLP as Embeddings
- retain the language ability → Auxiliary Language Object

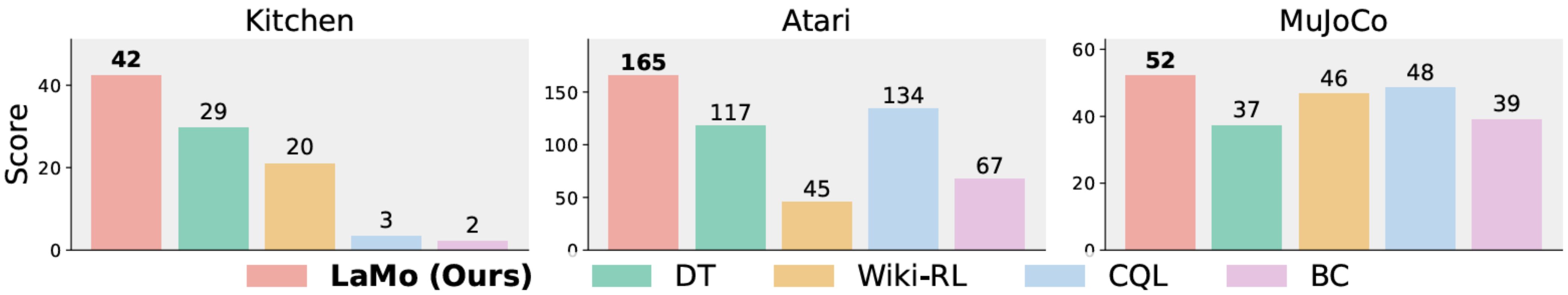
Experiment: Overview

Task selection

- **Action space**（continuous、discrete）
- **Reward distribution**（sparse、dense）
- **Data size**（0.1%-100% sampling ratio）



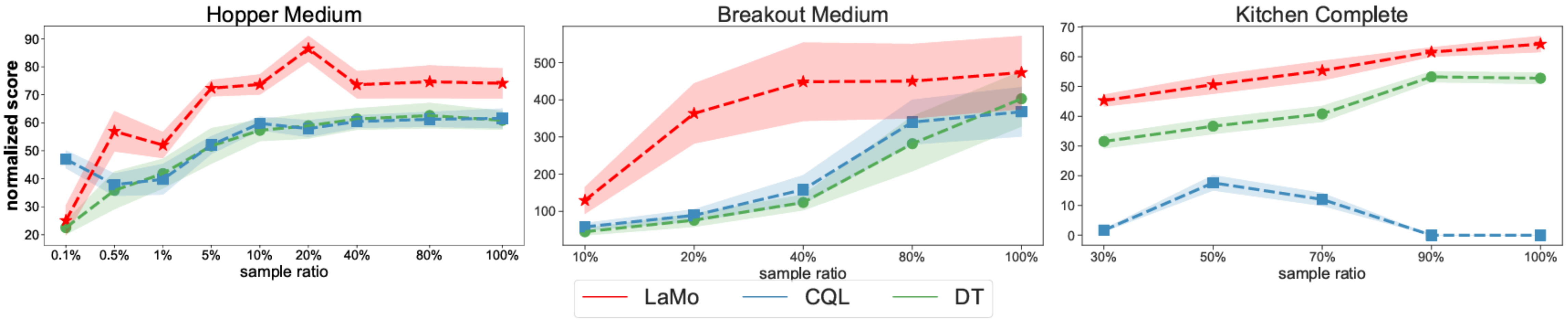
Experiment: Overview



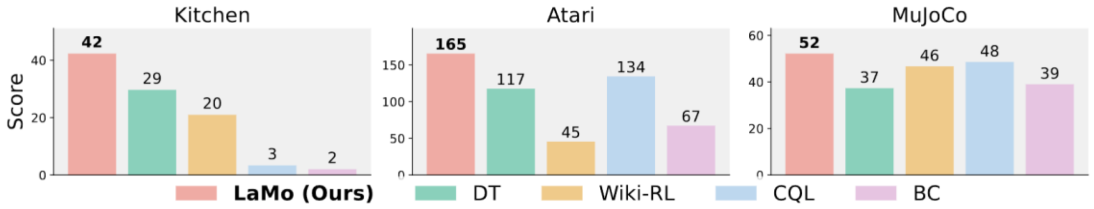
(Average over task and sample ratio)

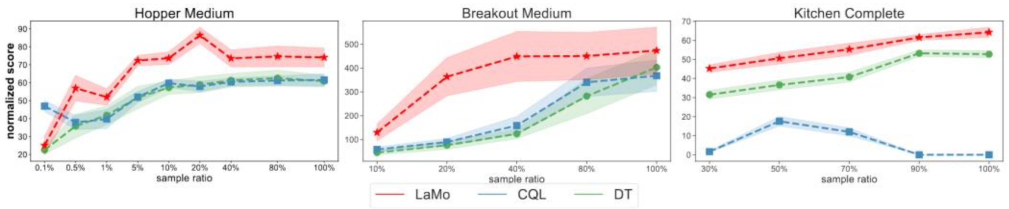
- In sparse-reward tasks (Kitchen, Reacher), **outperform** baselines prominently
- In dense-reward tasks (Locomotion, Atari), **close** the gap between Transformer-based and value-based algorithms

Experiment: Low-Date Regime



Show strong few-shot learning ability





Thank you for your Attention!