

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

Ruizhe Shi*¹, Yuyao Liu*¹, Yanjie Ze², Simon Shaolei Du³, Huazhe Xu¹²⁴

¹Tsinghua University, IIS ²Shanghai Qi Zhi Institute

³University of Washington ⁴Shanghai AI Lab

*Equal contribution. Order is decided by coin flip.



Language model



Motion control model



Motivation



Internet-Scale VQA + Robot Action Data

Q: What is happening in the image?
A: A grey donkey walks down the street.

Q: Que puis-je faire avec ces objets?
A: Faire cuire un gâteau.

Q: What should the robot do to <task>?
A Translation = [0.1, -0.2, 0]
A Rotation = [10°, 25°, -7°]

Co-Fine-Tune

Vision-Language-Action Models for Robot Control
RT-2

Deploy



VILA

Closed-Loop Robot Control

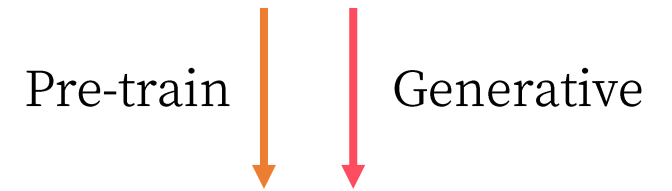
Put the strawberry into the correct bowl

Pick the nearly falling bag

Pick object that is different



Transformer architecture



LLM era

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

LLM + Robotics control



Introduction

TL

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

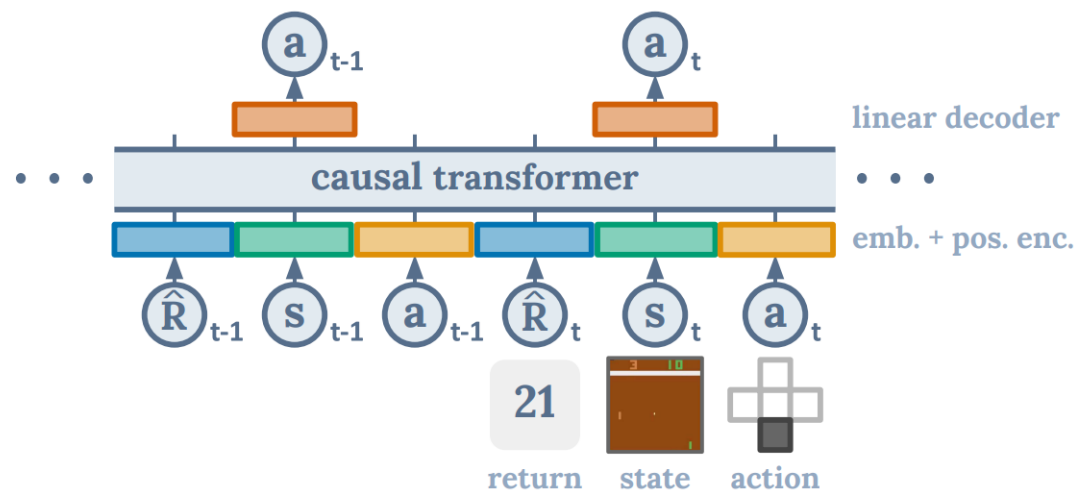


- pre-collecting data is still expensive \Rightarrow **few-shot learning**



Motivation

Offline RL Baseline — Decision Transformer (DT)



LM predict token:

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

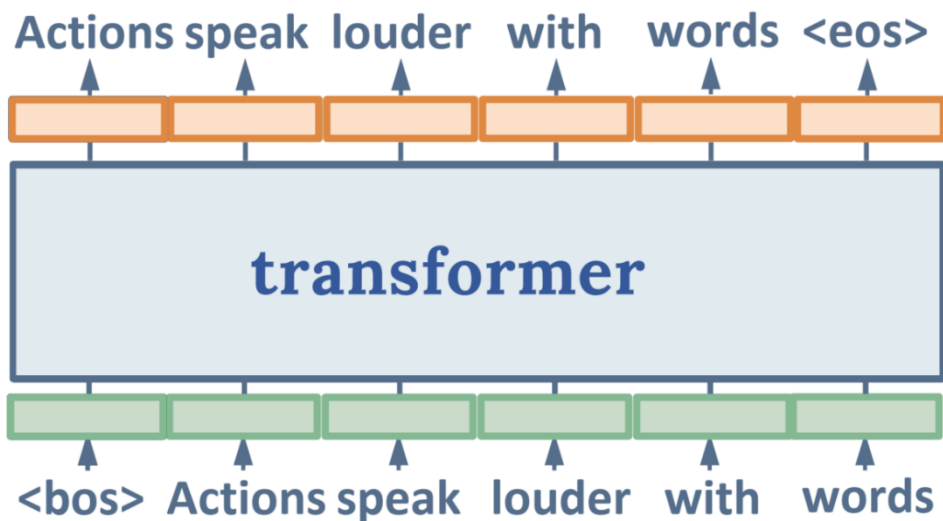
Motion model predict action:

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

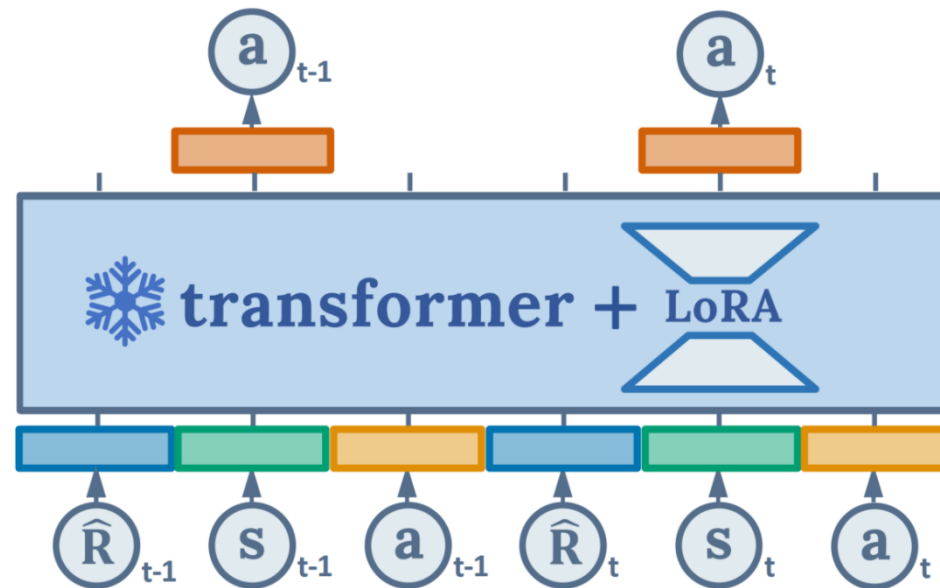


LaMo: Language Models for low level Motion control

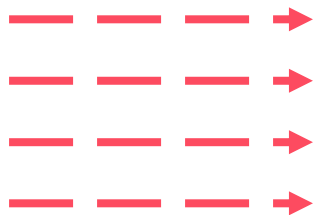
large language model pre-train



downstream offline RL



- knowledge from pre-training
- retain the knowledge
- enhancing representation
- retain the language ability



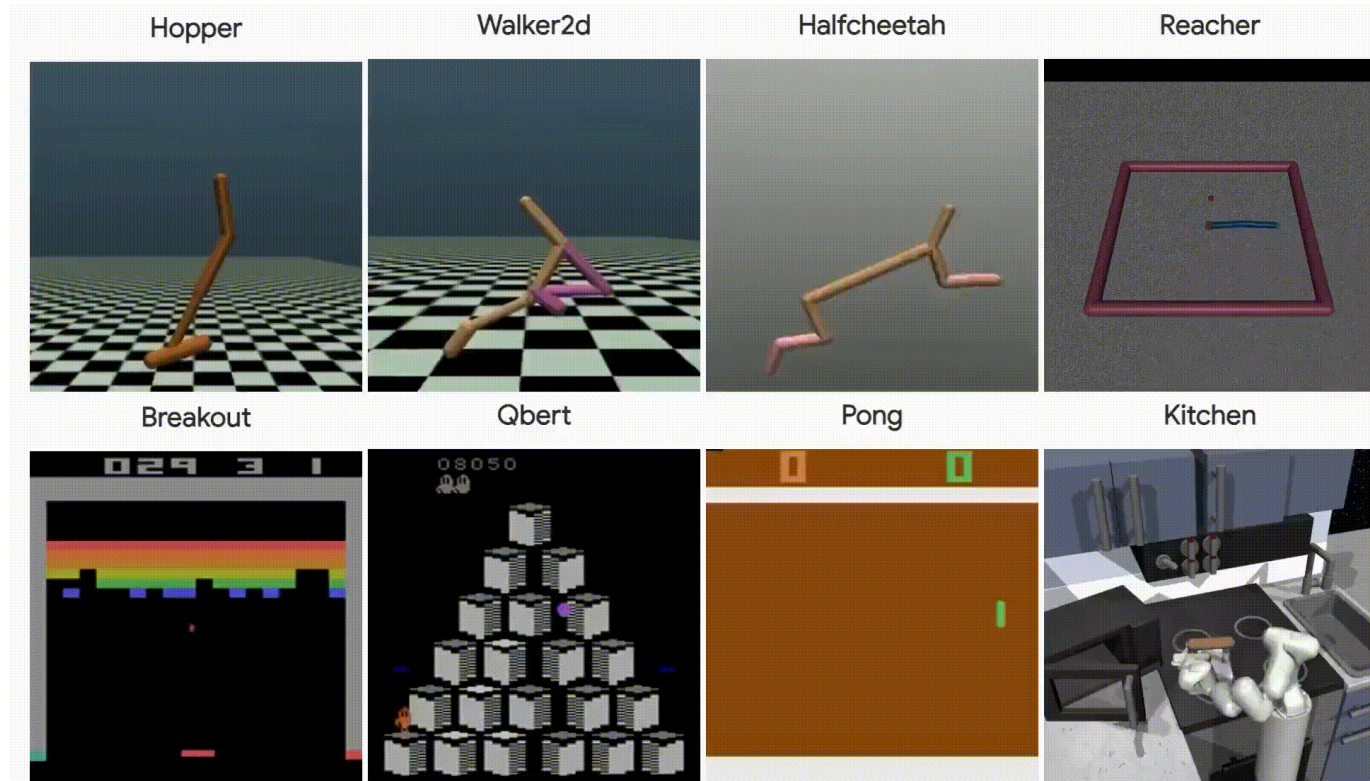
- Initialize with Pretrained LM
- Low Rank Adaptation (LoRA)
- MLP as Embeddings
- Auxiliary Language Object



Experiment: Overview

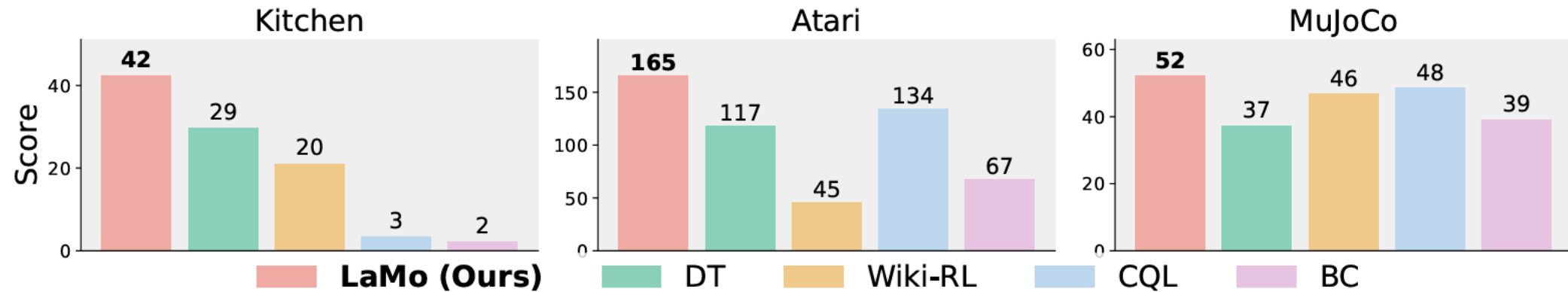
Task selection

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





Experiment: Overview

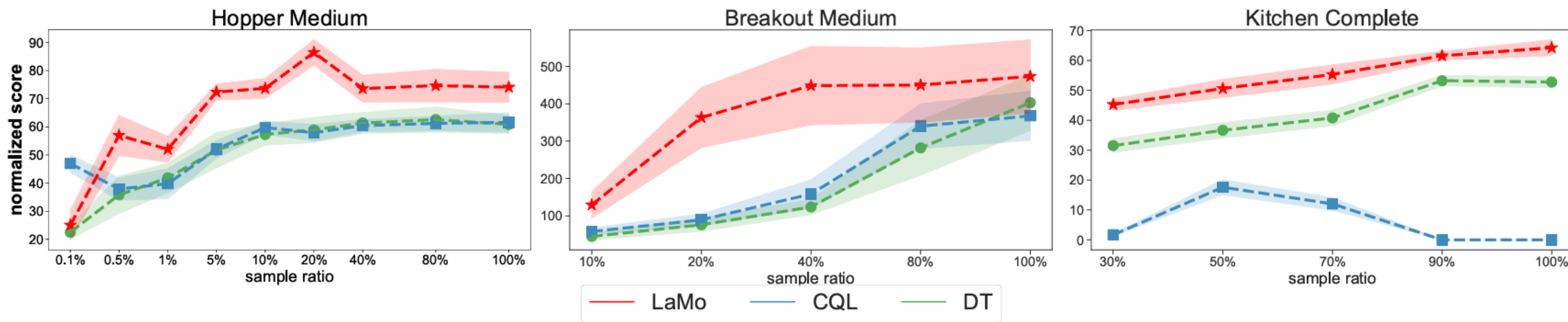


(Average over task and sample ratio)

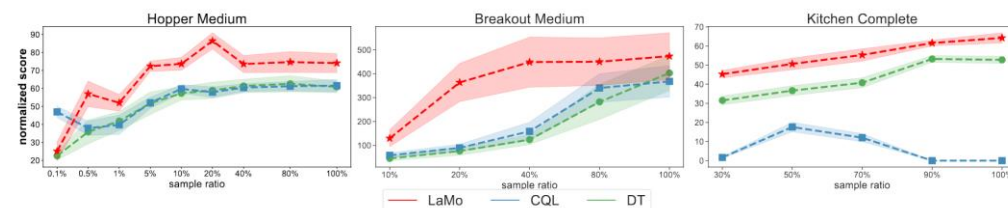
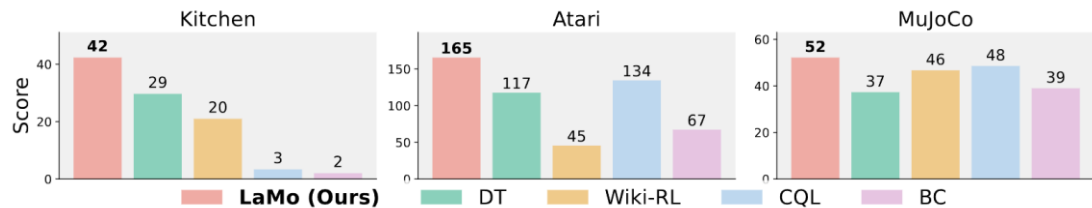
- In sparse-reward tasks (Kitchen, Reacher), **outperform** baselines prominently
- In dense-reward tasks (Locomotion, Atari), **close** the gap between Transformer-based and value-based algorithms



Experiment: Low-Data Regime



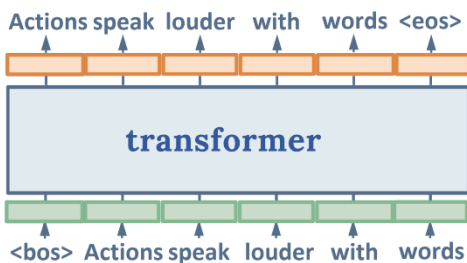
Show strong **few-shot learning ability**



Thank you for your Attention!



large language model pre-train



downstream offline RL

