# Enhancing Robotic Manipulation: Harnessing the Power of Multi-Task Reinforcement Learning and Single Life Reinforcement Learning in Meta-World


Ghadi Nehme [1] Ishan Sabane [2] Tejas Y. Deo [1]

[1]Department of Mechanical Engineering, Stanford University [2]Department of Electrical Engineering, Stanford University


## Introduction

Developing a model that can efficiently handle multiple tasks simultaneously, in a single shot, presents a significant challenge in the field of robotics. In this project, we leverage multi-task reinforcement learning (Multi-Task Soft Actor Critic) and single life reinforcement learning (Q-weighted adversarial learning) to enhance the performance of a robotic arm. Using the Meta World environment [1], we train the arm to master seven tasks and generate prior data for QWALE [2], and then, evaluate the performance of our agent on the seven tasks. We will also explore different task embeddings and identify the optimal approach to improve the performance of the Q-weighted adversarial learning (QWALE) algorithm. Our research aims to improve the arm's generalization capabilities and pave the way for more versatile and efficient robotic applications.

## Background

### Single Life Reinforcement Learning

Single life reinforcement learning enables autonomous task completion in a single trial, in the presence of a novel distribution shift, by leveraging prior data.

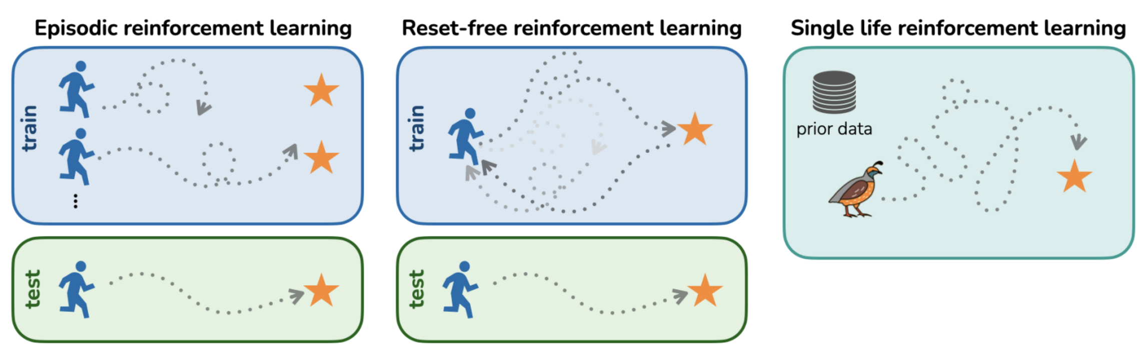



Figure 1. Single Life Reinforcement learning

### Meta-World

Meta-World is an open-source simulated benchmark for meta-reinforcement learning and multi-task learning consisting of 50 distinct robotic manipulation tasks.

We will be training our models on the following 7 tasks:

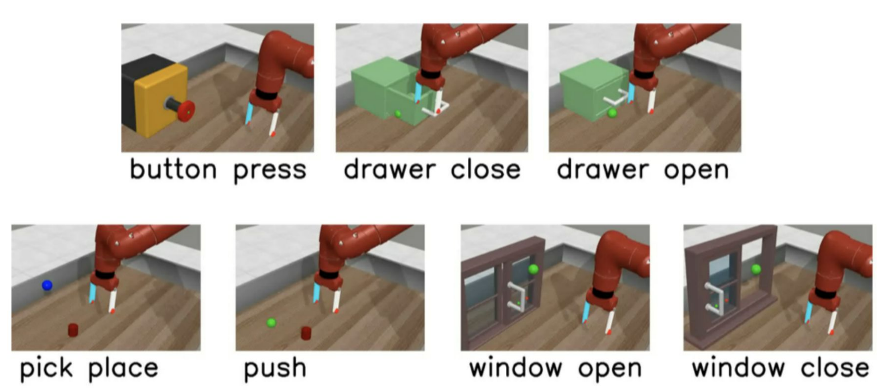



Figure 2. The 7 tasks on which our models are trained

### Observation

The observation is the input of our model.

$$\mathbf{o} = \begin{bmatrix} \mathbf{x}_E^{(i)} : \text{3D Cartesian coordinates of End-Effector} \\ \delta_G^{(i)} : \text{Measurement of how open the gripper is} \\ \mathbf{x}_1^{(i)} : \text{3D position of the first object} \\ \lambda_1^{(i)} : \text{Quaternion of the first object} \\ \mathbf{x}_2^{(i)} : \text{3D position of the second object} \\ \lambda_2^{(i)} : \text{Quaternion of the second object} \\ \mathbf{x}_{\text{Goal}} : \text{3D position of the goal} \end{bmatrix} \in \mathbb{R}^{39}, i \in \{t, t-1\}$$

### Action

The action is the output of our model.

$$\mathbf{a} = \begin{bmatrix} \delta\mathbf{x} : \text{3D space of the end-effector} \\ \tau : \text{normalized torque that the gripper fingers should apply} \end{bmatrix} \in \mathbb{R}^4$$

## Technical Methods

### Multi-Task Soft Actor Critic (MTSAC)

Multi-Task Soft Actor-Critic (SAC) is a RL algorithm that enables a single agent to effectively handle multiple tasks simultaneously.

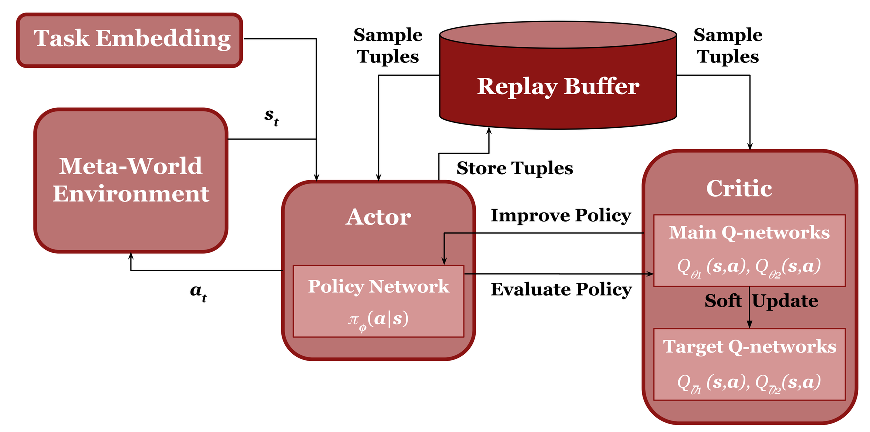



Figure 3. Multi-Task Soft Actor Critic (MTSAC)

### Task Embedding

We concatenate the observation of Meta-World with three different types of task embedding:

| Task Embedding | Example |
|---|---|
| One Hot Encoding | $[1, 0, 0, 0, 0, 0, 0]$ |
| Sine Encoding | $[\sin(k), \sin(2k), \sin(3k), \sin(4k), \sin(5k), \sin(6k), \sin(7k)]$ |
| Learned Encoding | $M[1, 0, 0, 0, 0, 0, 0]$, where $M$ is a learned matrix |

Table 1. Task Embeddings

### Multi-Task Q-weighted adversarial learning

- Weight states in the prior data based on their estimated Q-values.
- Incentivizes the agent to move towards states closer to task completion.

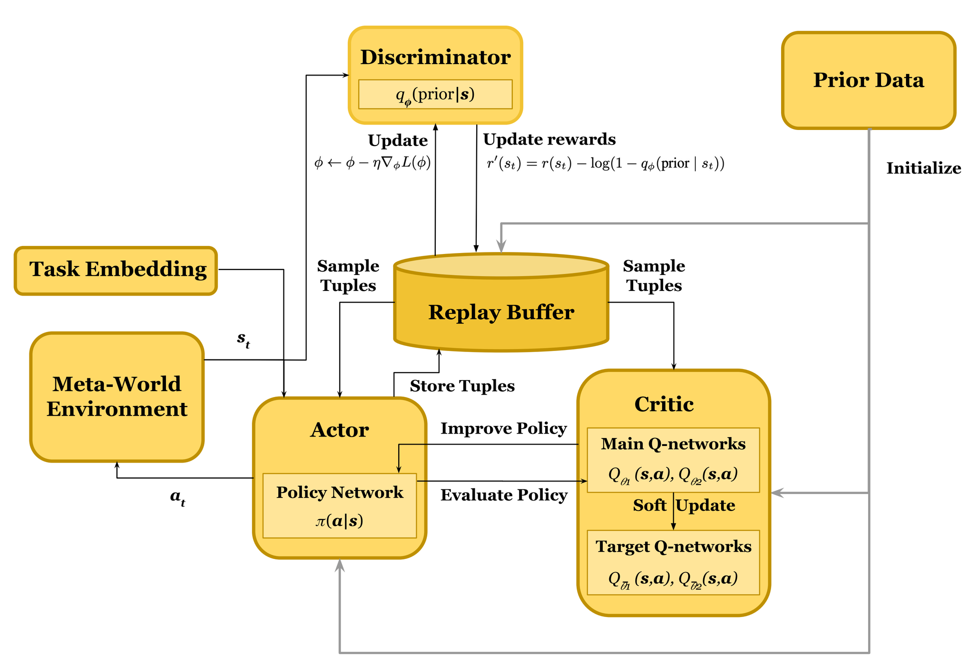



Figure 4. Multi-Task Q-weighted Adversarial Learning

## Experiments & Results

We are developing a model capable of completing **multiple tasks** in a **single trial**, in the presence of **novel distribution shifts**. In our case, the tasks are the ones in Fig. 2. The novelties are to randomly vary the positions of the objects in the environments. To achieve this:

- We first train a MT-SAC on the 7 tasks to collect the prior data needed for MT-QWALE for 10000 episodes per task.
- We then train a MT-QWALE on the 7 tasks and observe its performance.
- We also repeat this experiment for different tasks encoding, to optimize the performance of Multi-Task QWALE.

### Multi-Task SAC

- We observe that using the sine encoding leads to the best performance.
- The algorithm is less accurate when executing tasks combining grasping and moving.

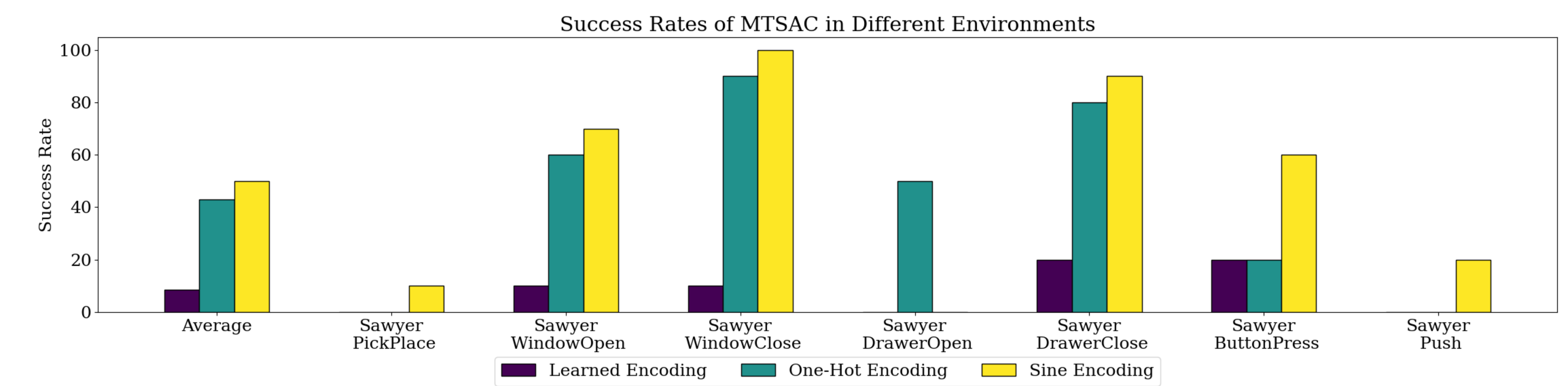



Figure 5. Success Rates of MTSAC in different environments

### Multi-Task QWALE

We observe that MT-QWALE provides a big improvement in the performance of MT-SAC in the environments with novelty.

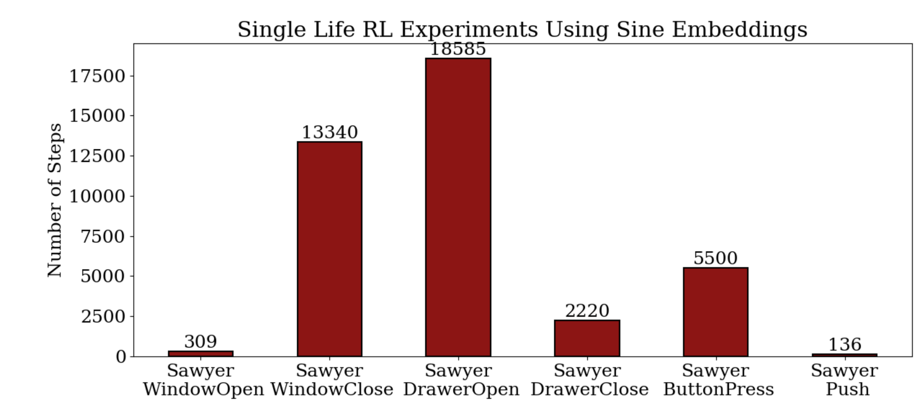
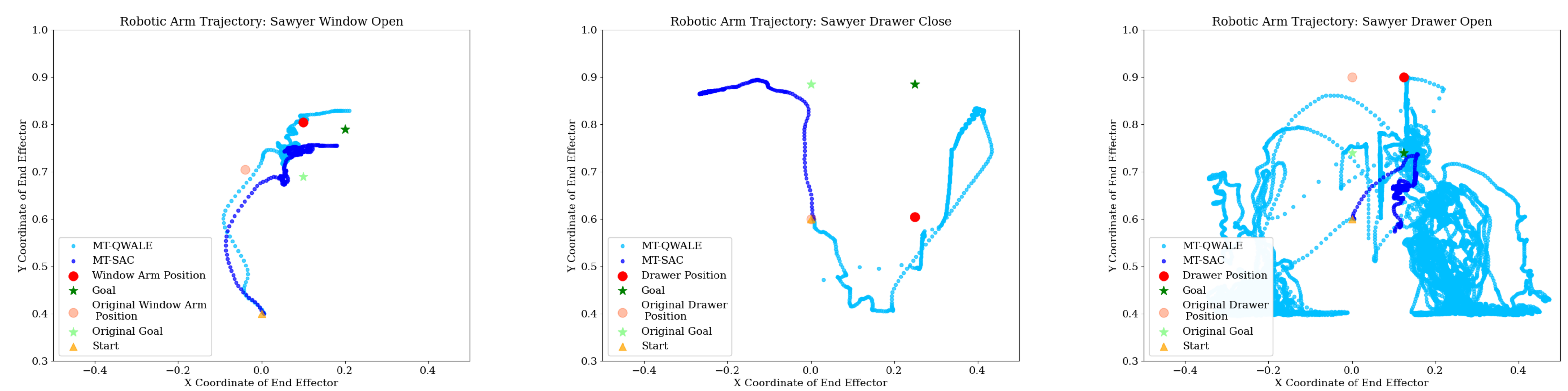



Figure 6. Number of steps of single life RL experiments in different environments

#### Comparaison of MT-SAC and MT-QWALE

1. In Fig 7.(a), we observe that both algorithms achieve the tasks in similar number of steps.
2. In Fig 7.(b), we observe that MT-QWALE achieves the task while MT-SAC fails.
3. In Fig 7.(c), we observe that MT-QWALE achieves the task with many steps while MT-SAC fails.




(a) Window Open (b) Drawer Close (c) Drawer Open

Figure 7. Trajectories of end-effector comparing the performance of MT-SAC and MT-QWALE in different environments with novelty

## References


[1] Tianhe Yu, Deirdre Quillen, Zhanpeng He, Ryan Julian, Karol Hausman, Chelsea Finn, and Sergey Levine. Meta-world: A benchmark and evaluation for multi-task and meta reinforcement learning. In *Conference on robot learning*, pages 1094–1100. PMLR, 2020.

[2] Annie Chen, Archit Sharma, Sergey Levine, and Chelsea Finn. You only live once: Single-life reinforcement learning. *Advances in Neural Information Processing Systems*, 35:14784–14797, 2022.



ghadi@stanford.edu ishancs@stanford.edu tejasdeo@stanford.edu