

Multi-task Reinforcement Learning with Soft Modularization

Ruihan Yang, Huazhe Xu, Yi Wu, Xiaolong Wang

project page: <https://rchalyang.github.io/SoftModule>

UC San Diego

Berkeley
UNIVERSITY OF CALIFORNIA

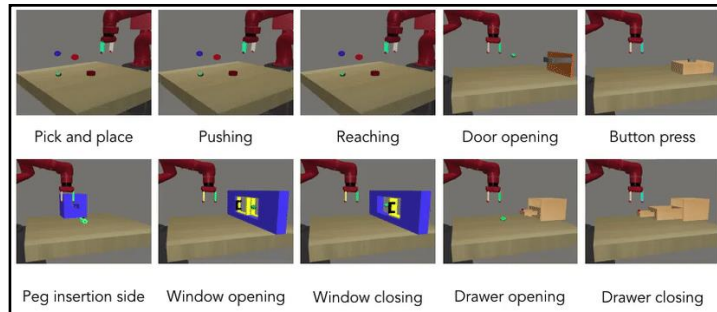

OpenAI

Motivation

Most RL: specialized

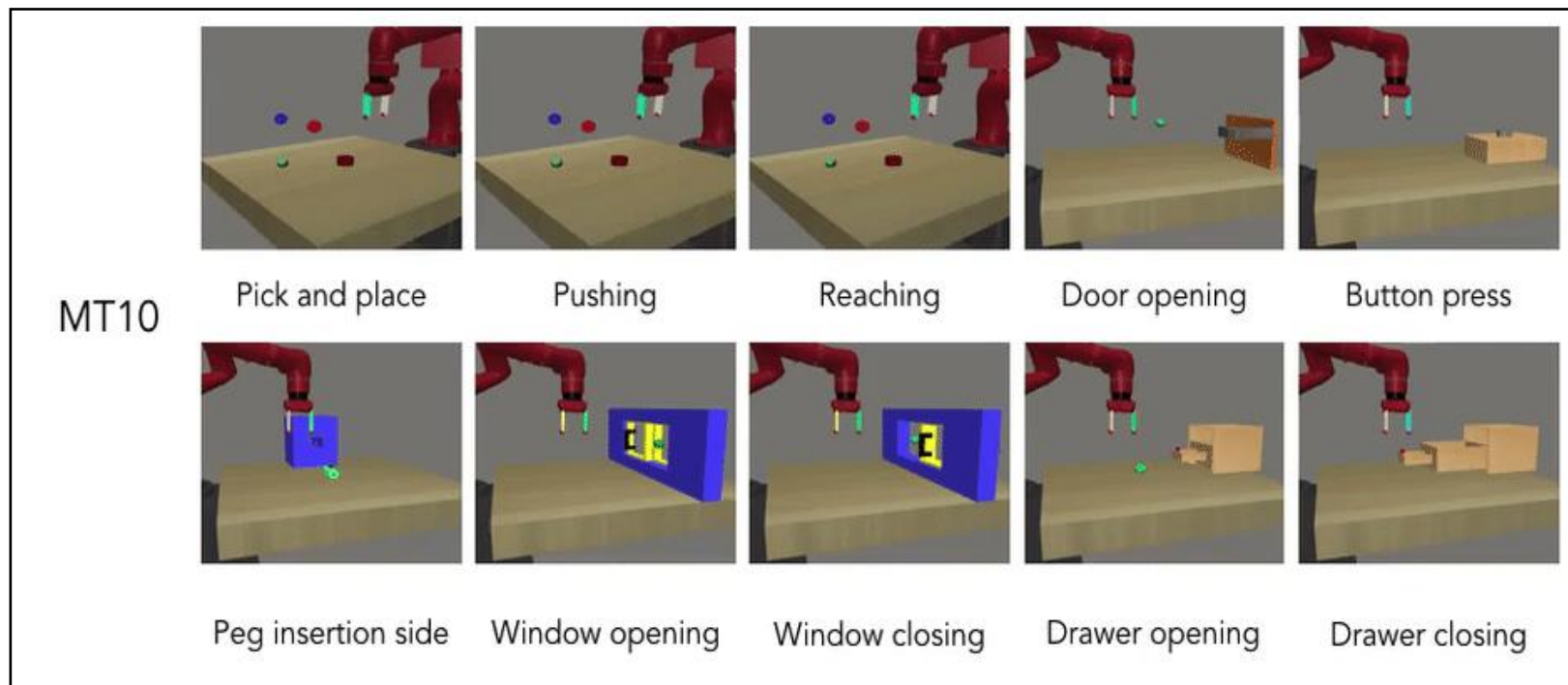
How? Multi-task RL

Goal: Apply to real-world

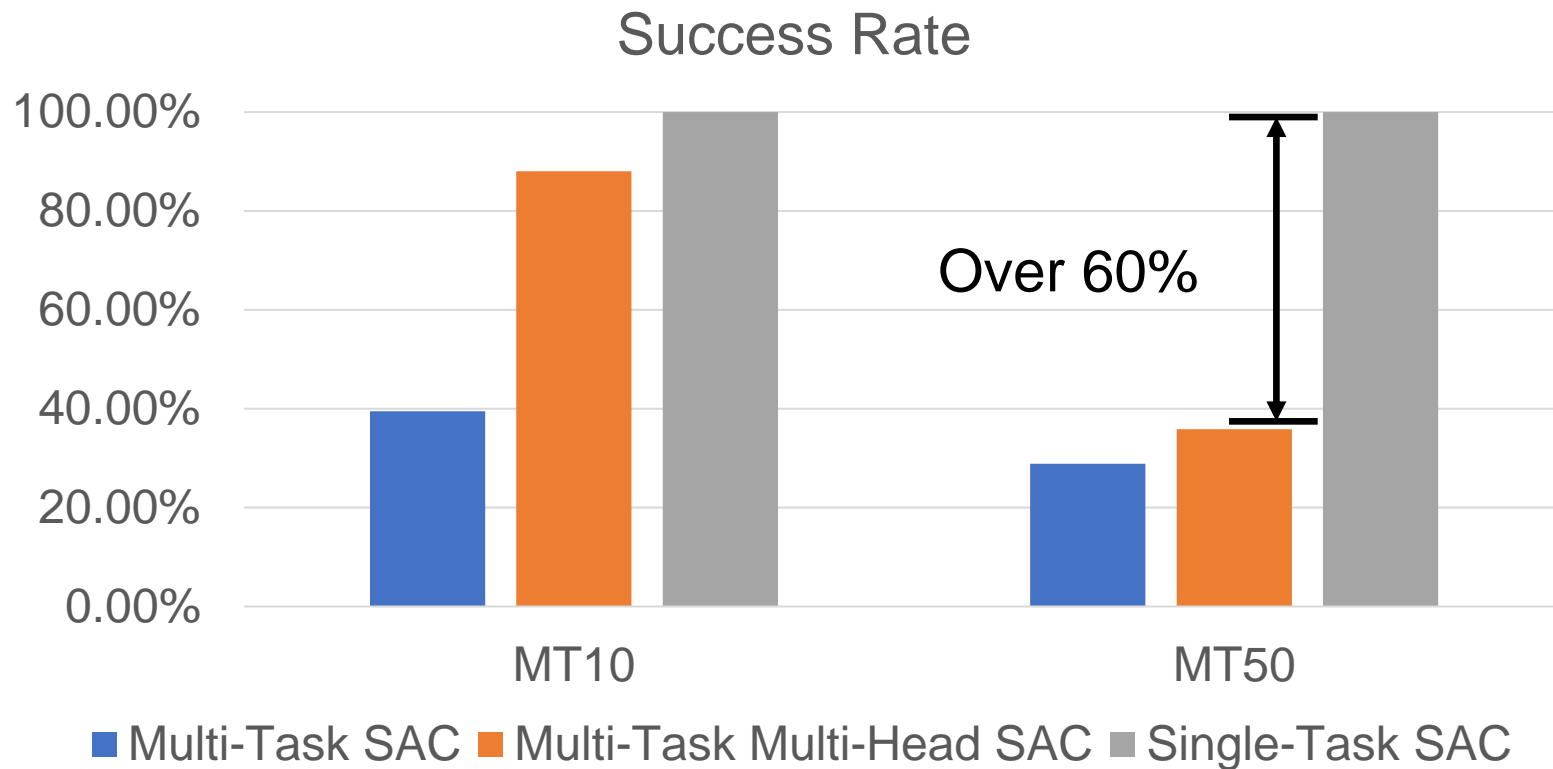


Current Multi-task RL Benchmark: Meta-World

Containing dozens of robotics manipulation tasks.

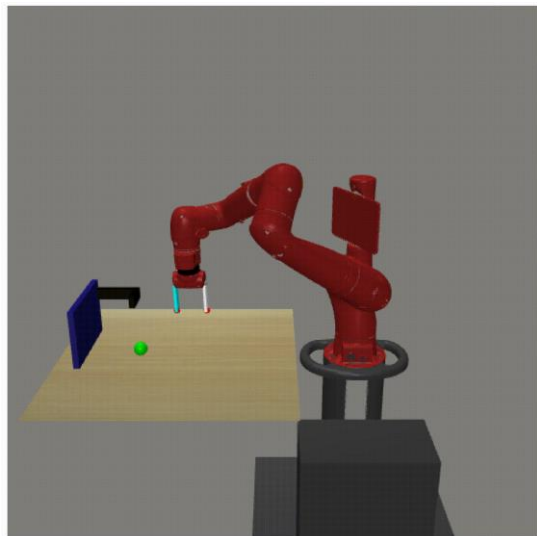


Meta-World

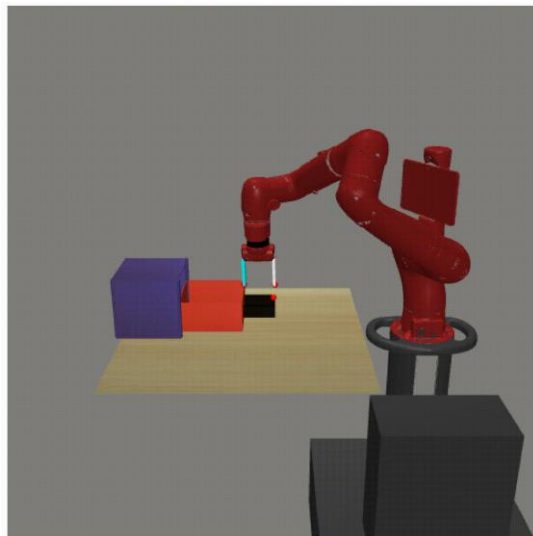


Failed Examples

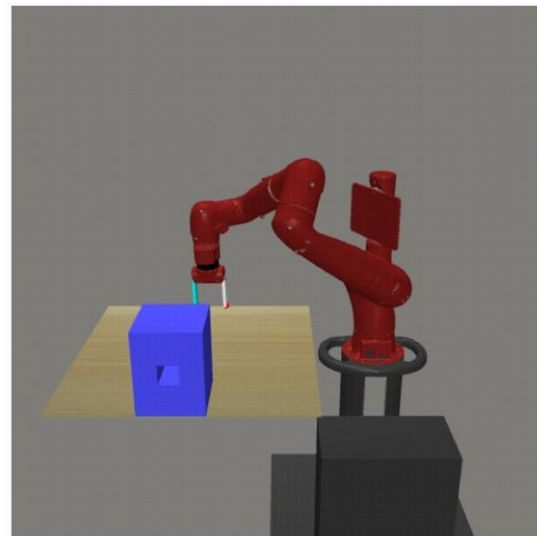
Baseline failed to generalize to different tasks



Open Door



Close Drawer

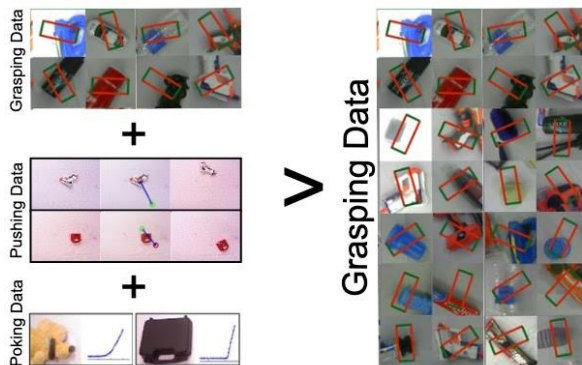


Insert Peg

General Multi-Task Learning



Computer Vision:
Detection + Segmentation



Robot Learning:
Pushing + Grasping + Poking >
Grasping

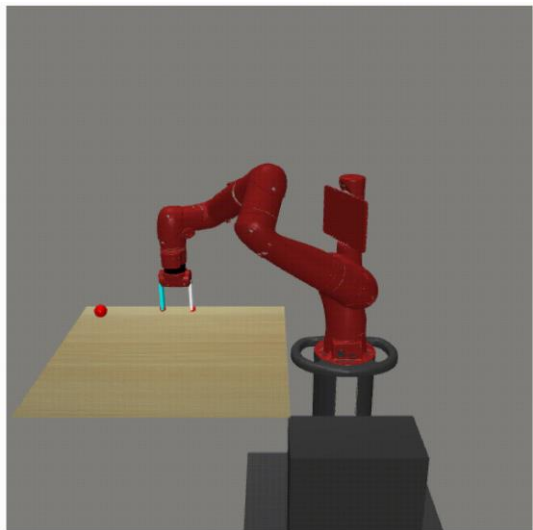
He, K et al. Mask R-CNN, 2017

Pinto, L et al. Learning to Push by Grasping: Using multiple tasks for effective Learning, 2017

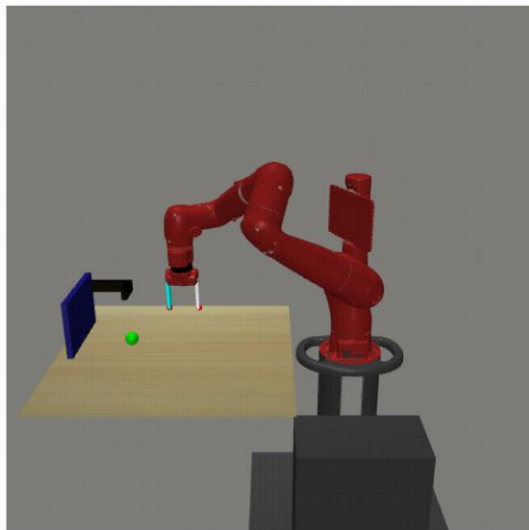
Two Challenges in Multi-Task Reinforcement Learning

Avoid negative interference between irrelevant tasks

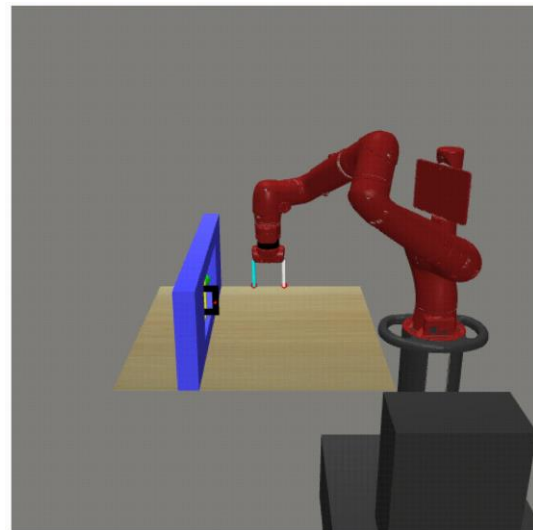
Reuse shared components across **similar** tasks



Reach

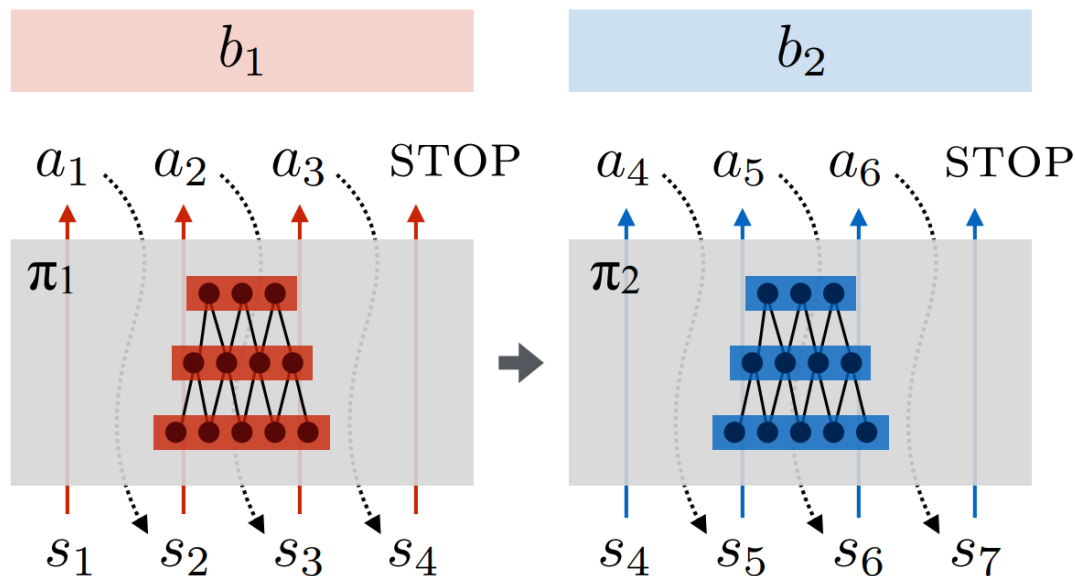


Open Door



Open Window

Modularization



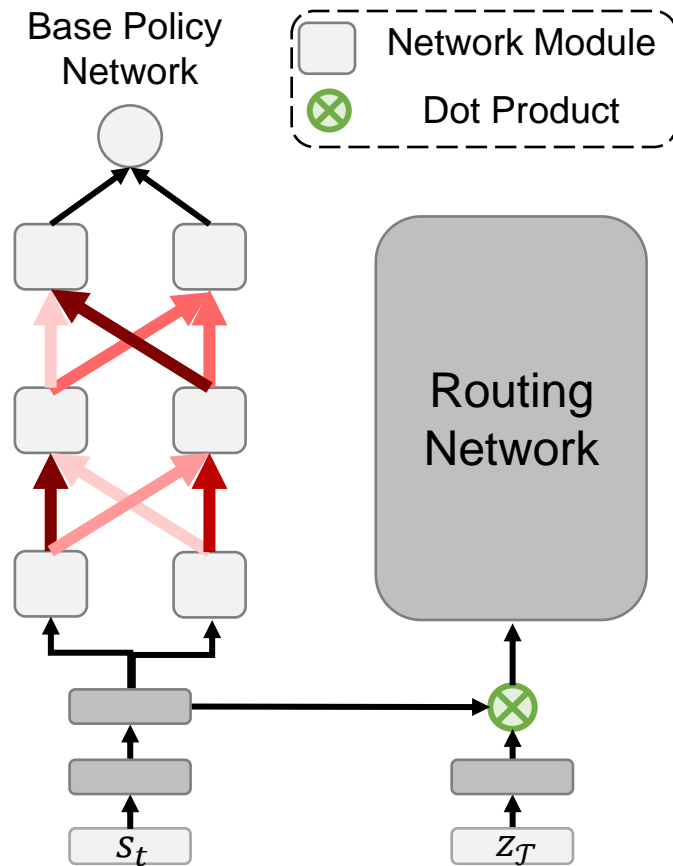
Previous Modular
network for multi-task
RL

In hierarchical manner

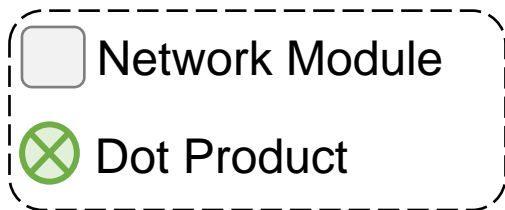
Soft Modularization

Base Policy Network + Routing Network

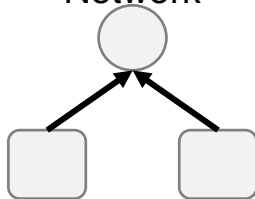
Differentiable



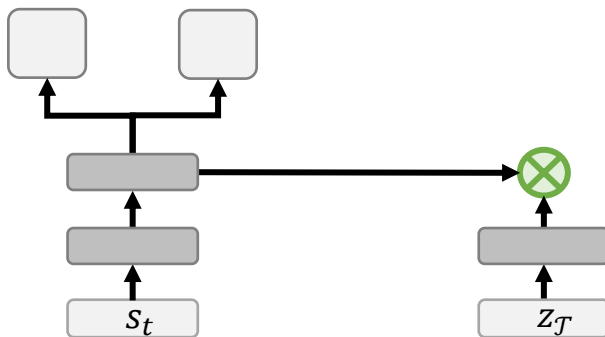
Soft Modularization



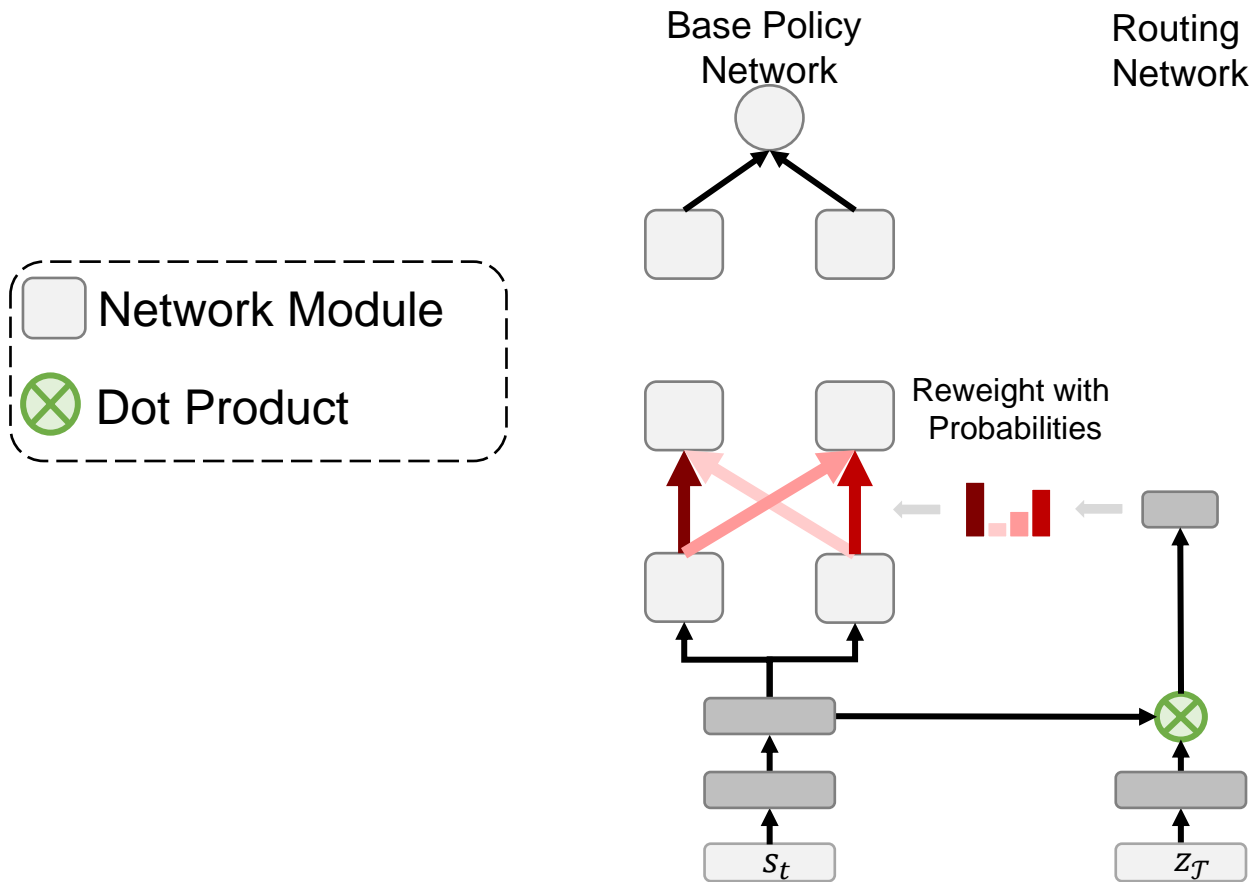
Base Policy
Network



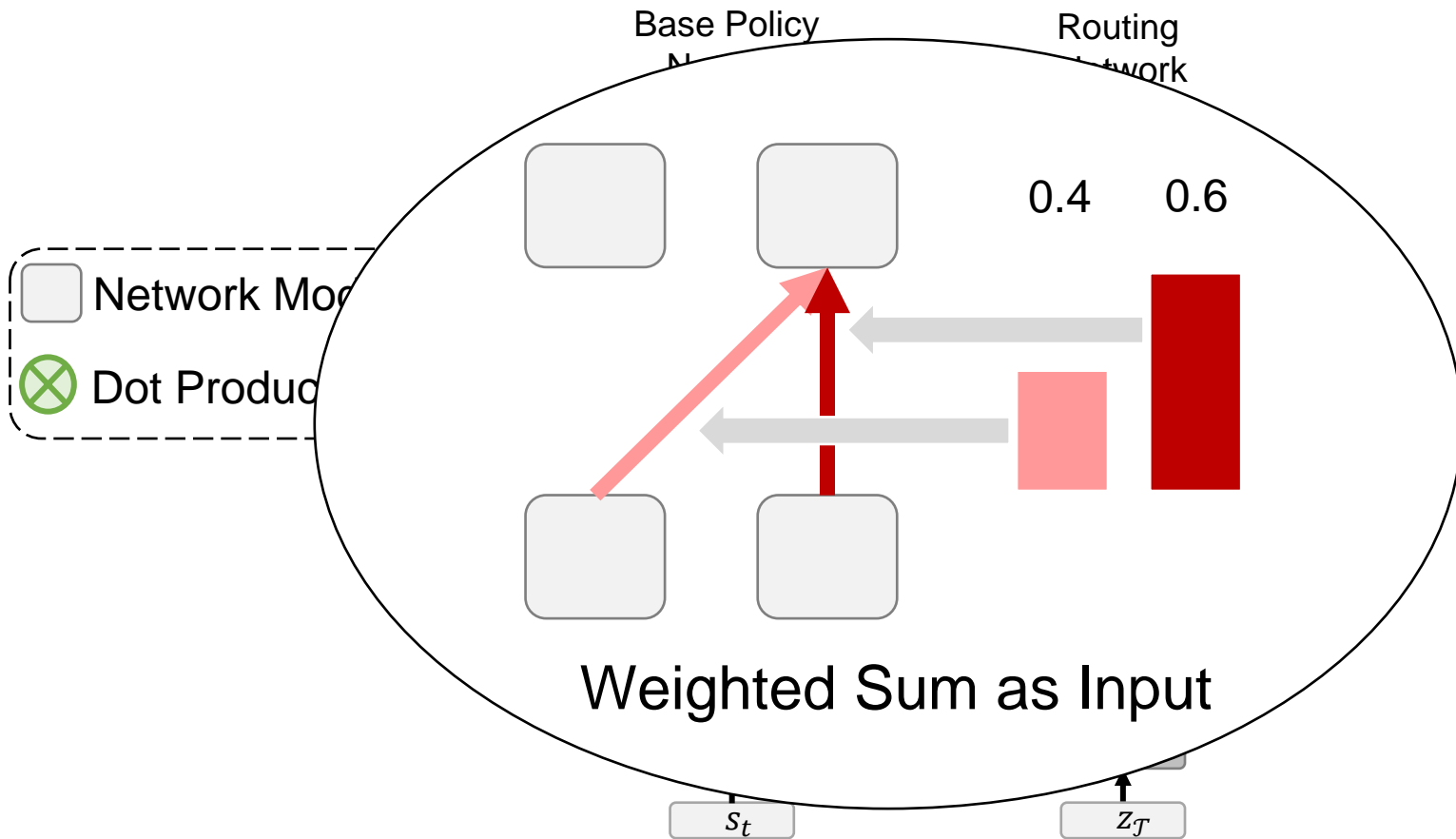
Routing
Network



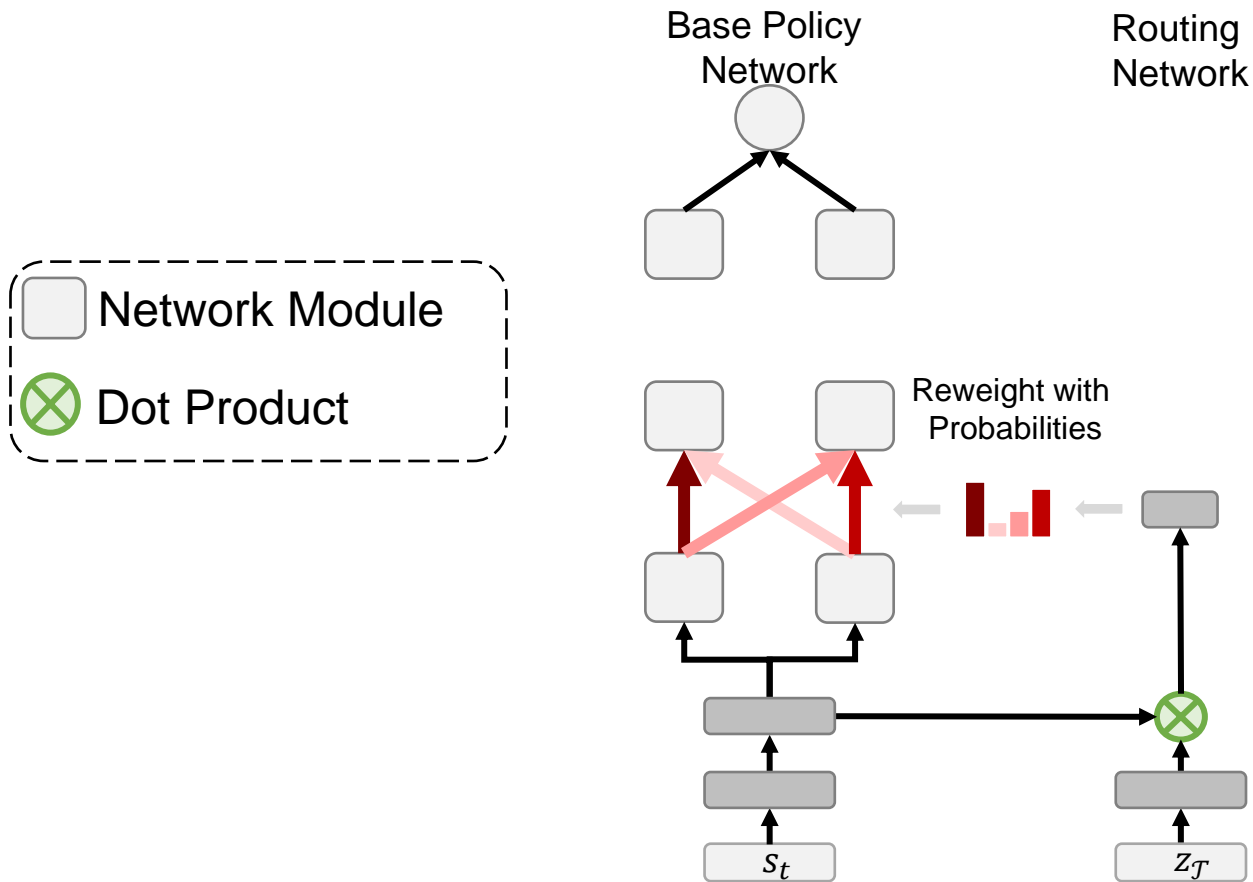
Soft Modularization



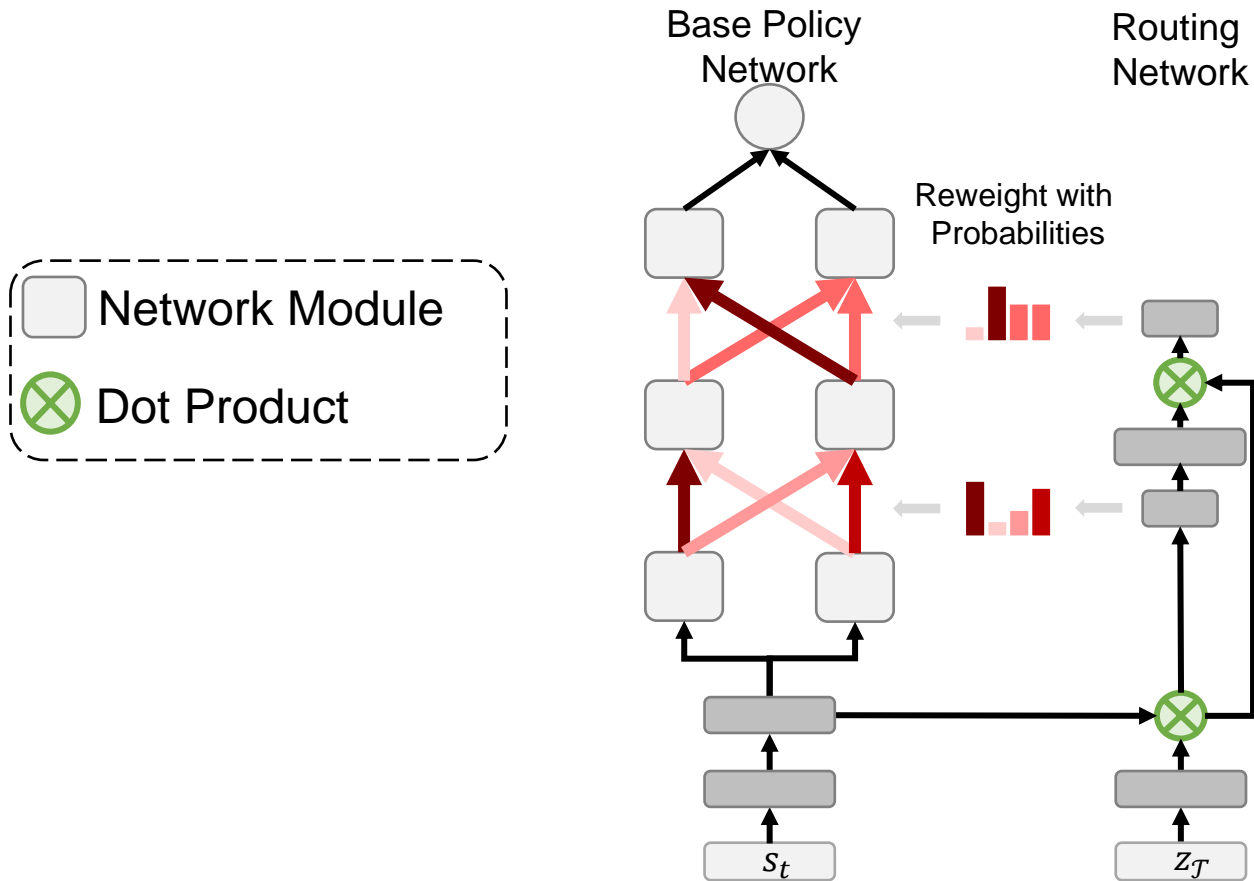
Soft Modularization



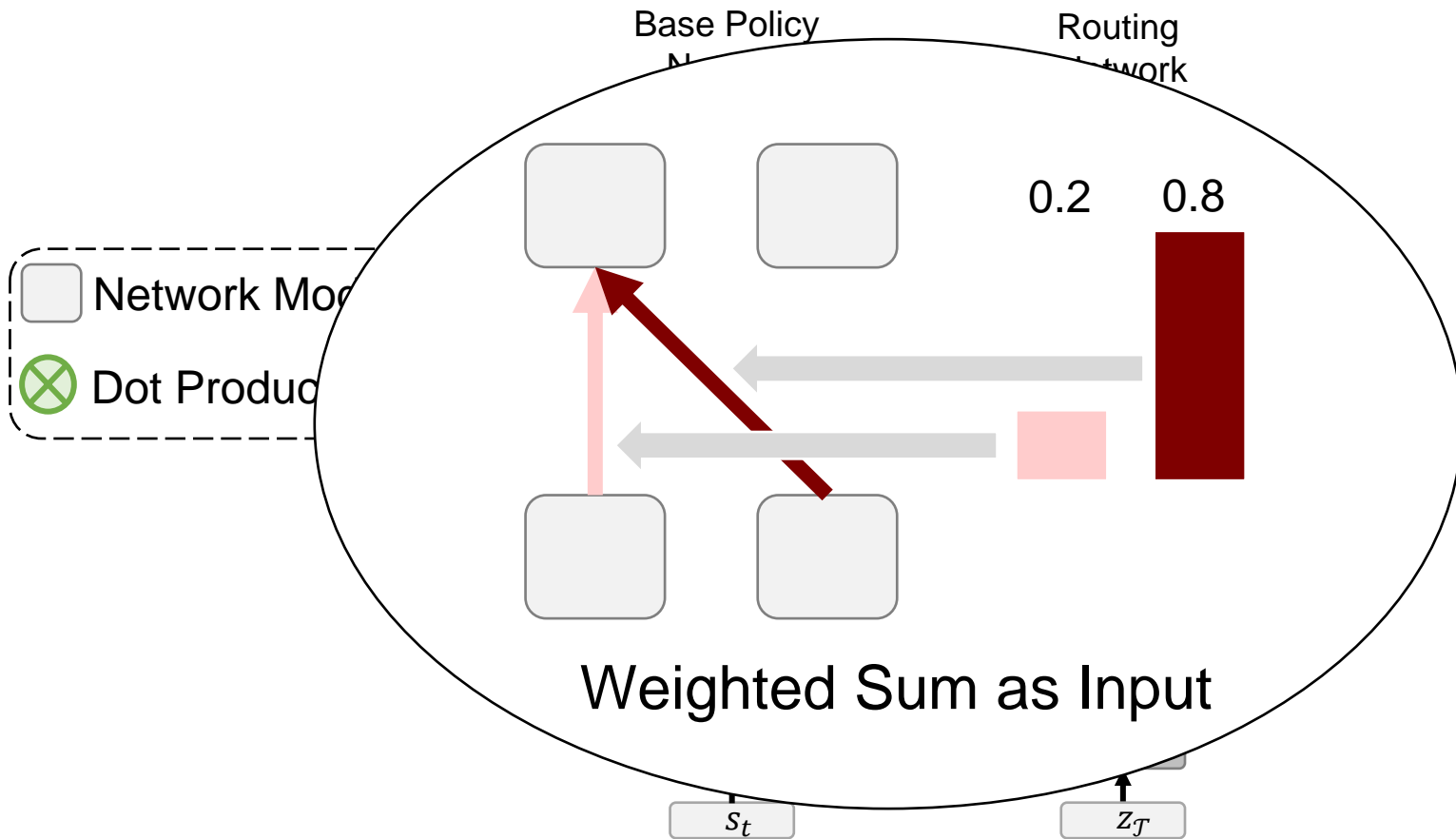
Soft Modularization



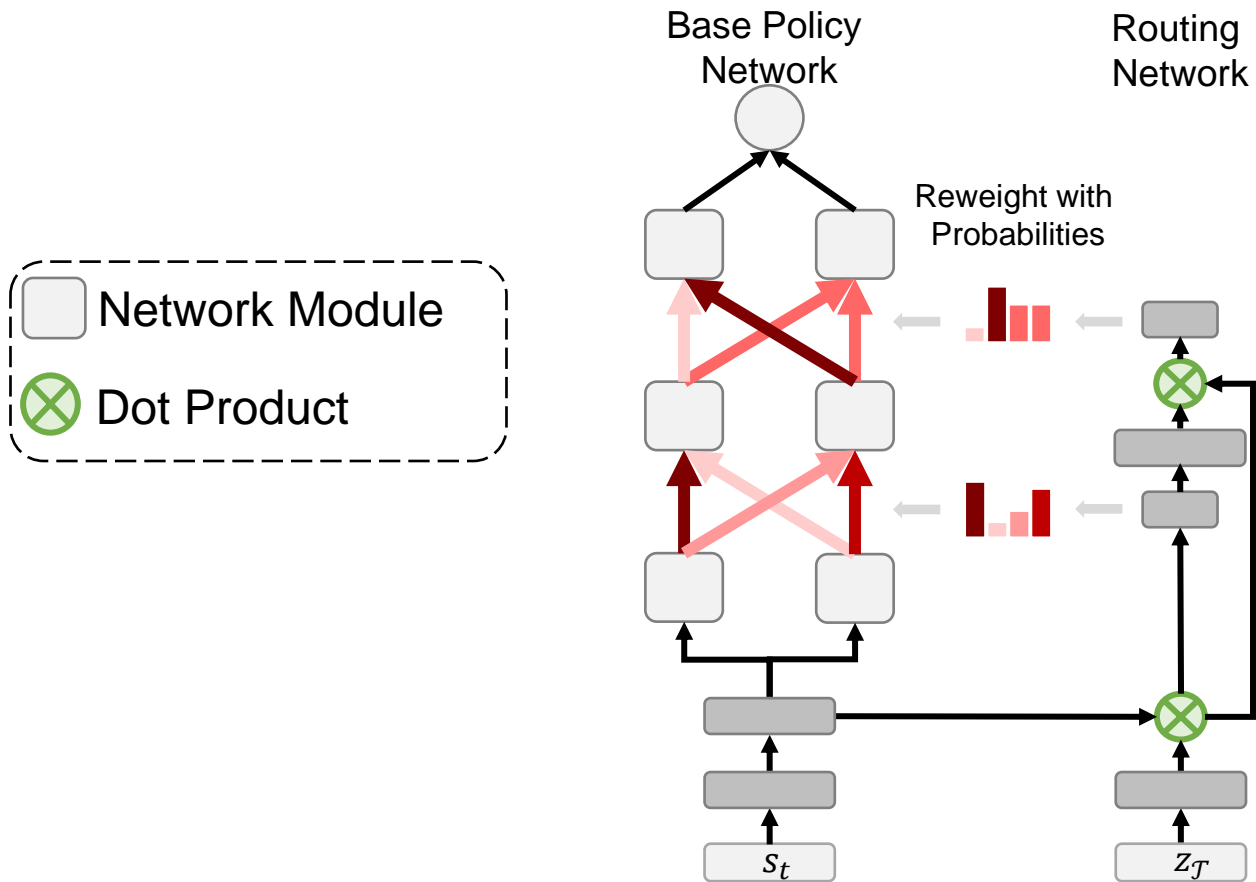
Soft Modularization



Soft Modularization



Soft Modularization



Balance Multi-task learning in SAC

Temperature weight for multi-task RL:

$$J_{\pi}(\phi) = \mathbb{E}_{\mathcal{T} \sim p(\mathcal{T})}[w_{\mathcal{T}} \cdot J_{\pi, \mathcal{T}}(\phi)] \quad (1)$$

Temperature adjustment in SAC:

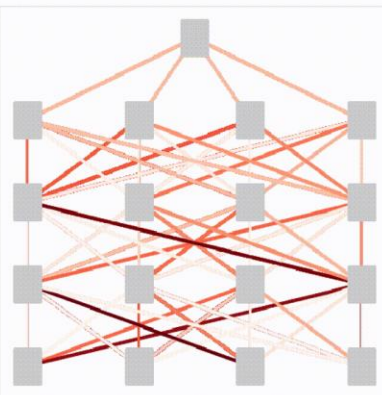
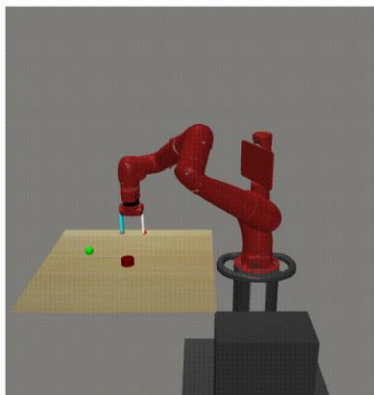
$$J(\alpha) = E_{a_t \sim \pi_{\phi}}[-\alpha \log \pi_{\phi}(a_t | s_t) - \alpha \overline{\mathcal{H}}] \quad (2)$$

Temperature weight for multi-task RL:

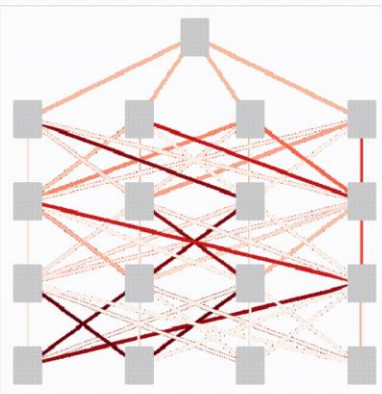
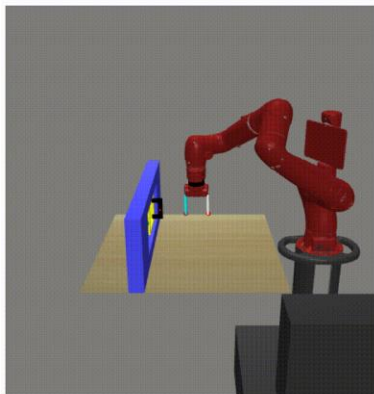
$$w_i = \frac{\exp(-\alpha_i)}{\sum_{j=1}^M \exp(-\alpha_j)} \quad (3)$$

Soft Modularization

Push

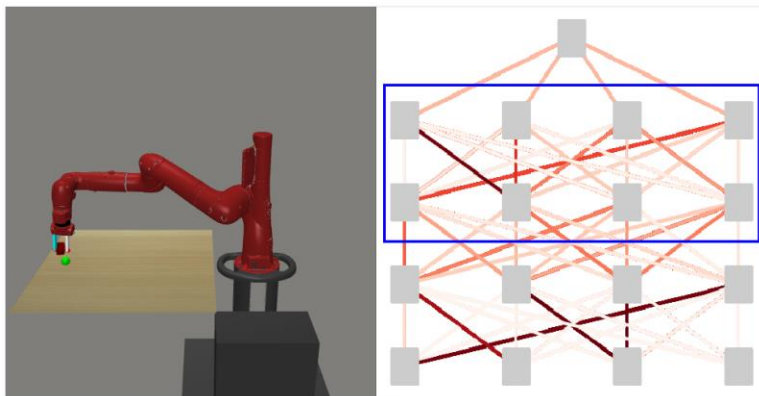


Close Window

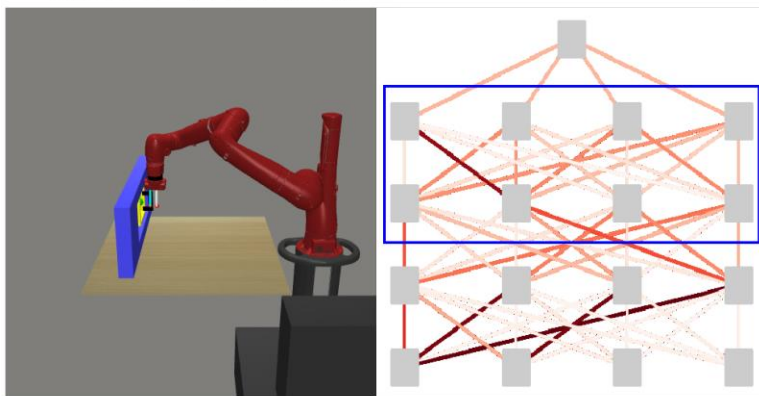


Module Sharing : Knowledge Sharing.

Push

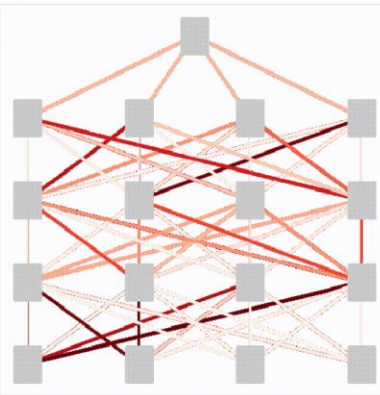
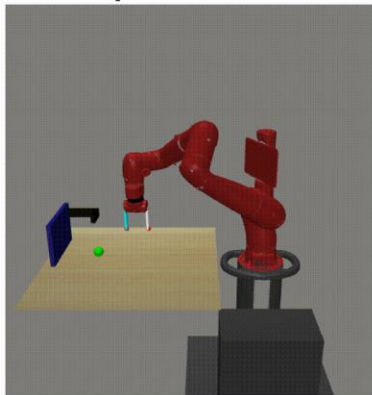


Close Window

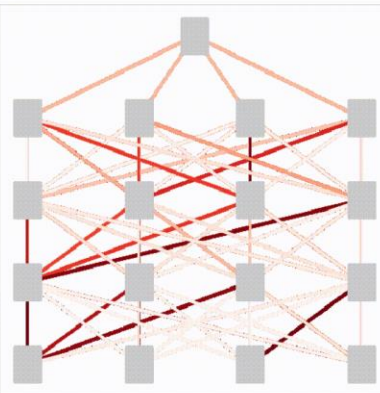
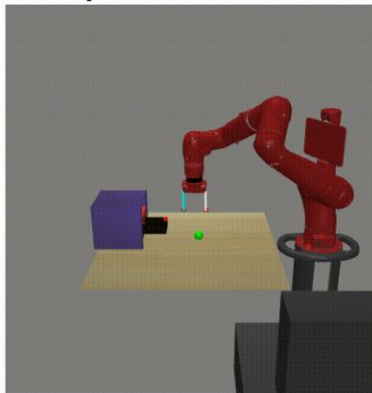


Module Sharing : Knowledge Sharing.

Open Door

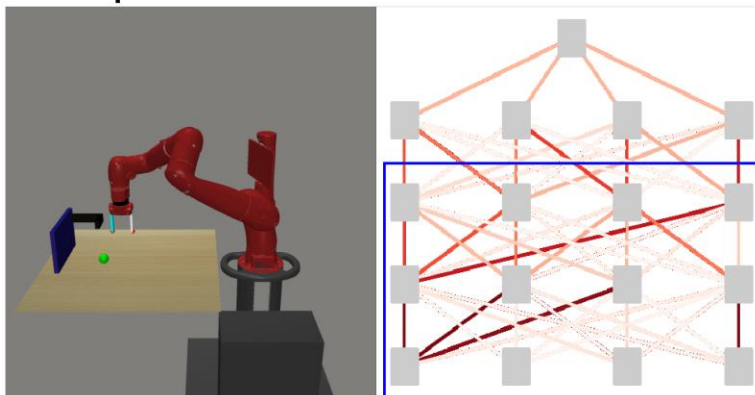


Open Drawer

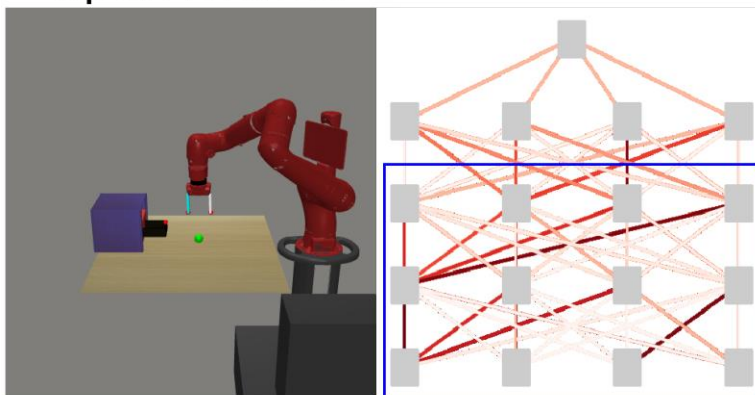


Module Sharing : Knowledge Sharing.

Open Door



Open Drawer



Experiments

Original Meta-World:

Fixed goal.

MT10-Fixed / MT50-Fixed

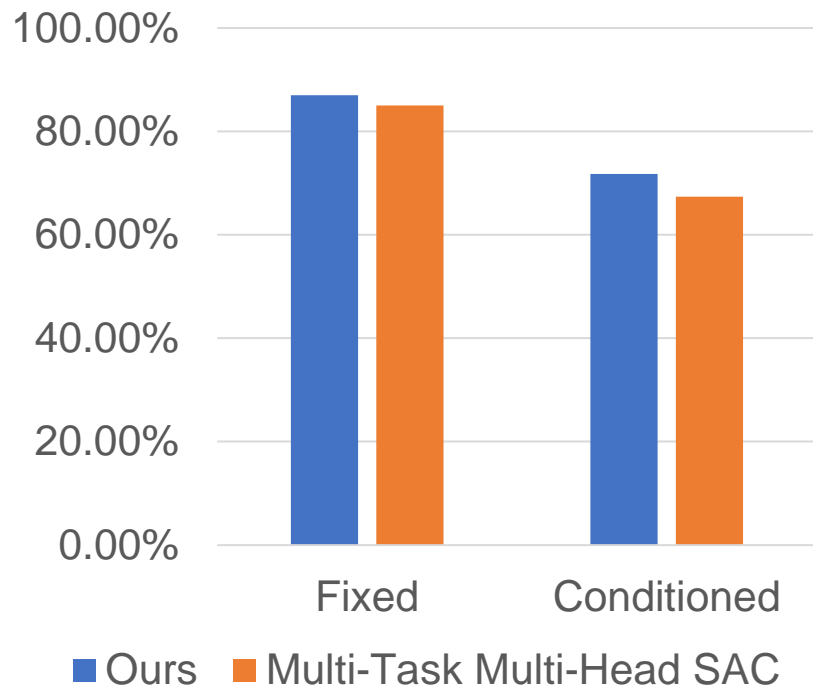
More realistic and more challenging:

Goal conditioned

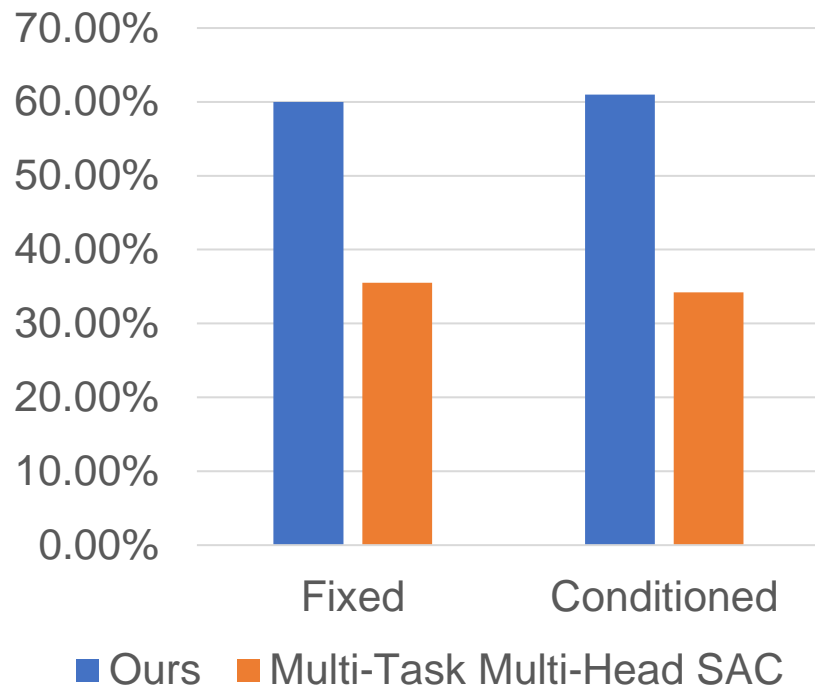
MT10-Conditioned / MT50-Conditioned

Experiments : Over All

MT10

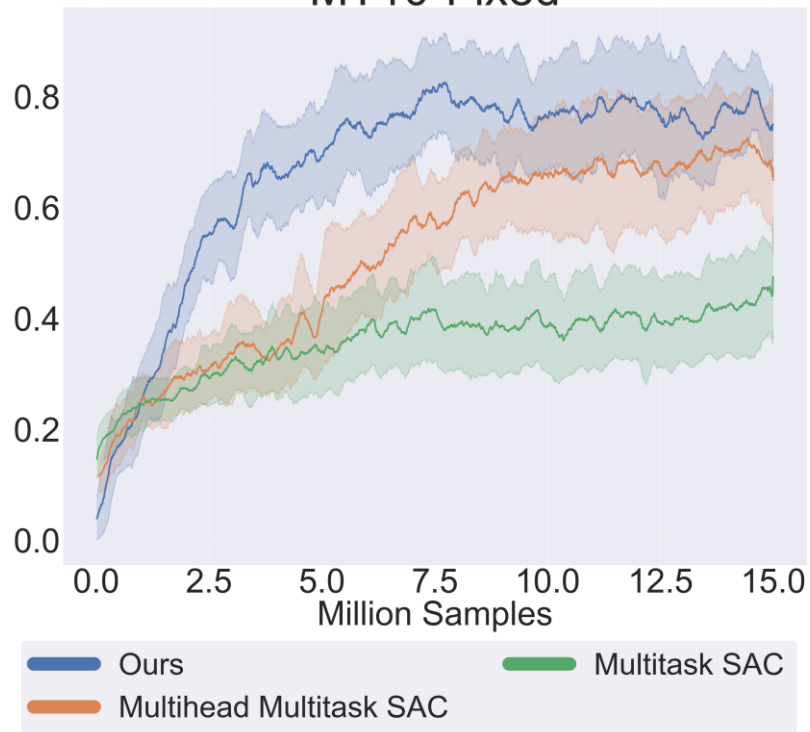


MT50

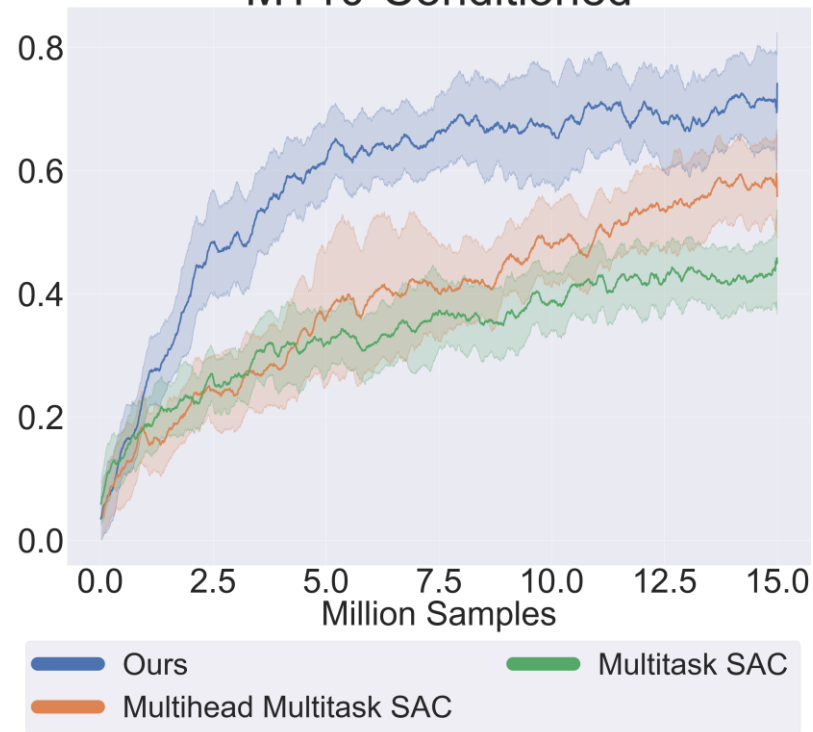


Experiments : MT10

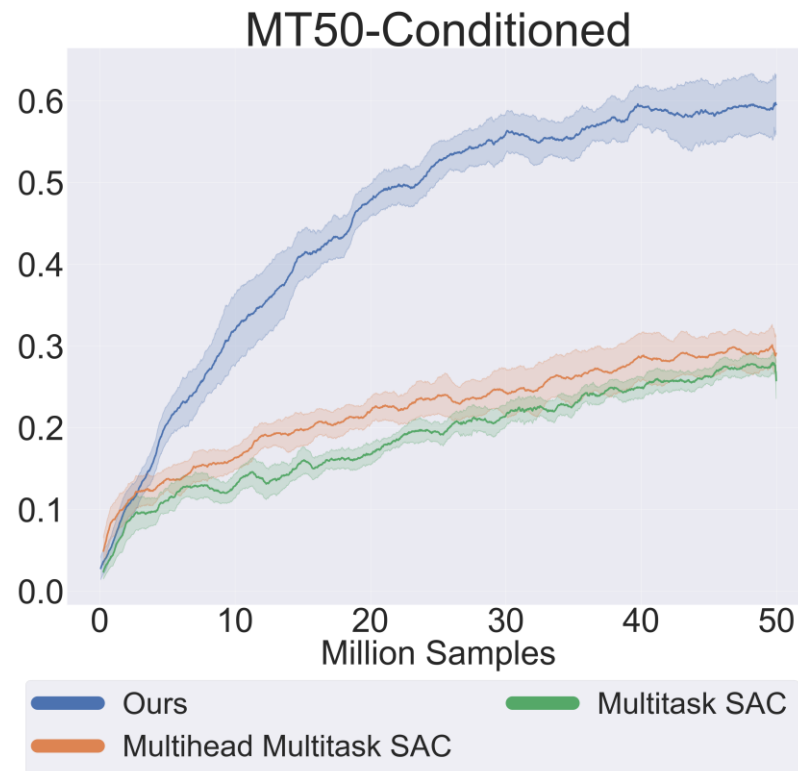
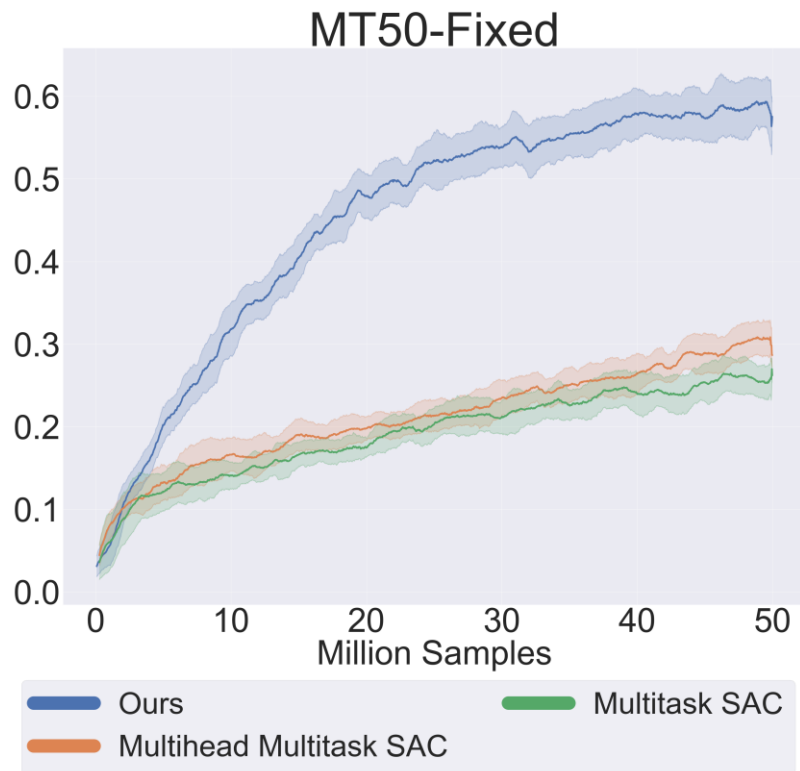
MT10-Fixed



MT10-Conditioned



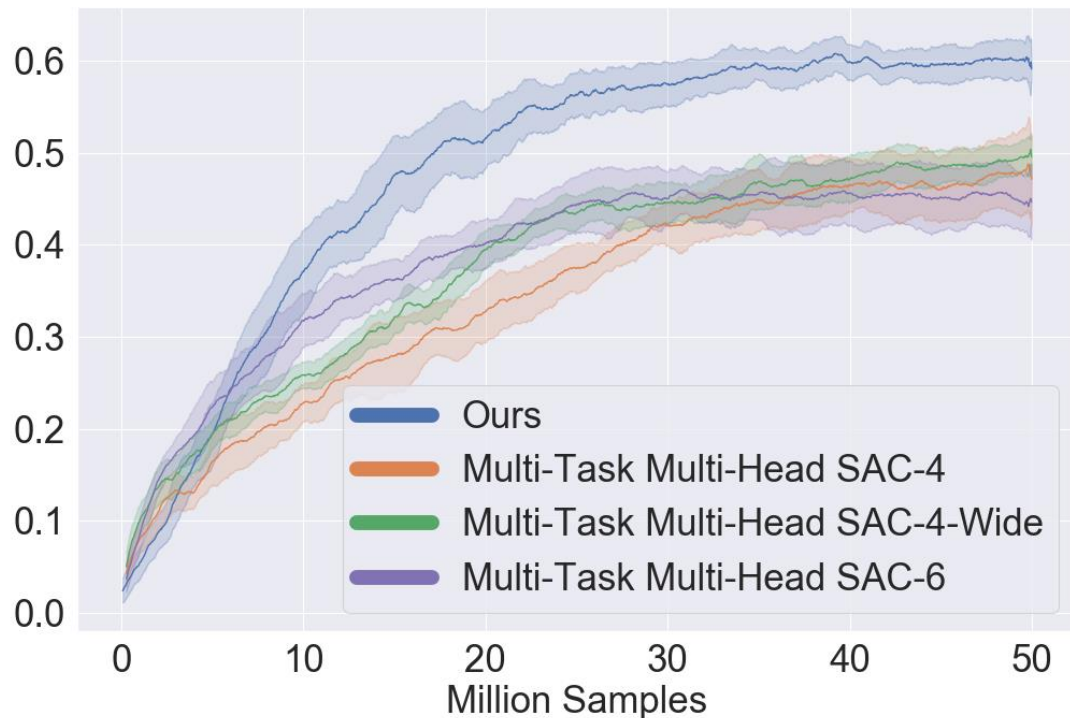
Experiments : MT50



Effects on Network Capacity

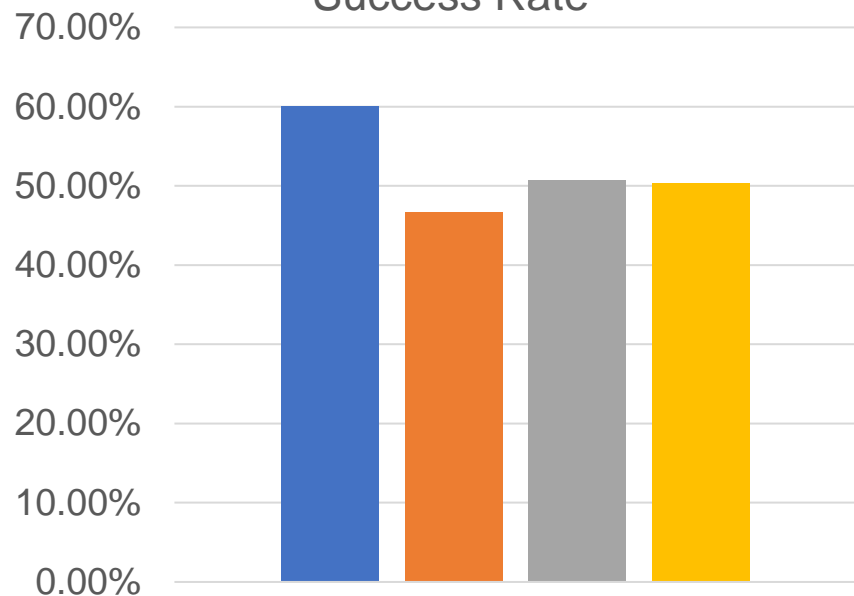
Performed on MT50-Fixed

Larger network can not solve multi-task RL.



Effects on Network Capacity

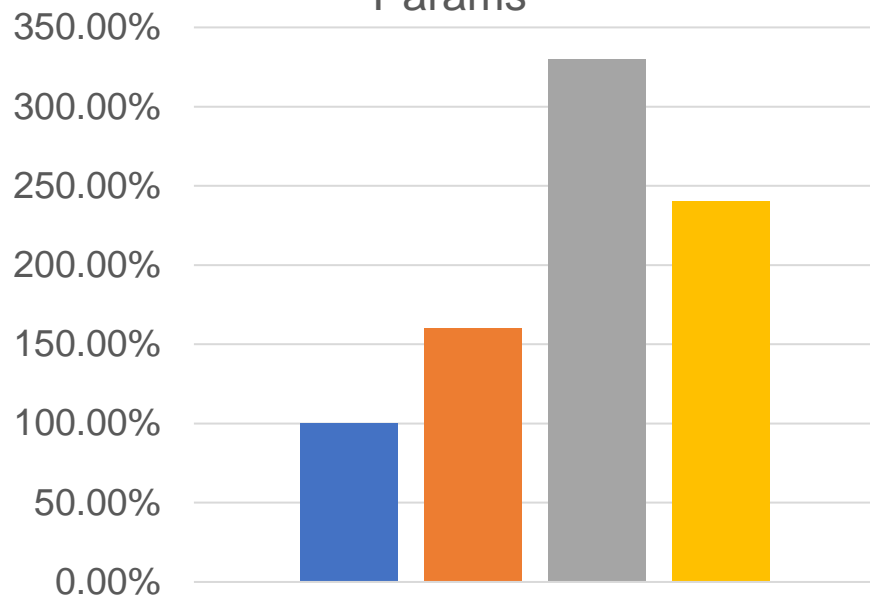
Success Rate



MT-50-Fixed

- Ours
- Multi-Task Multi-Head SAC-4
- Multi-Task Multi-Head SAC-4-Wide
- Multi-Task Multi-Head SAC-6

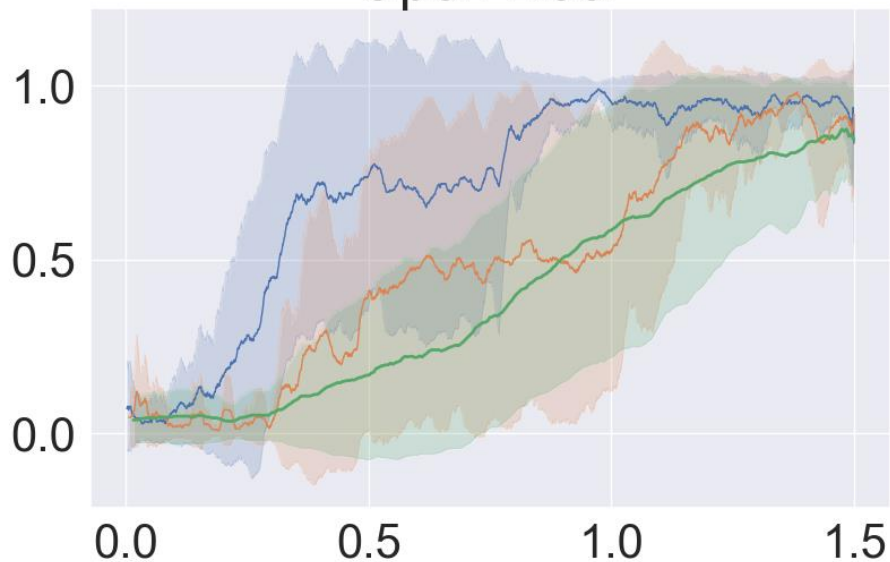
Params



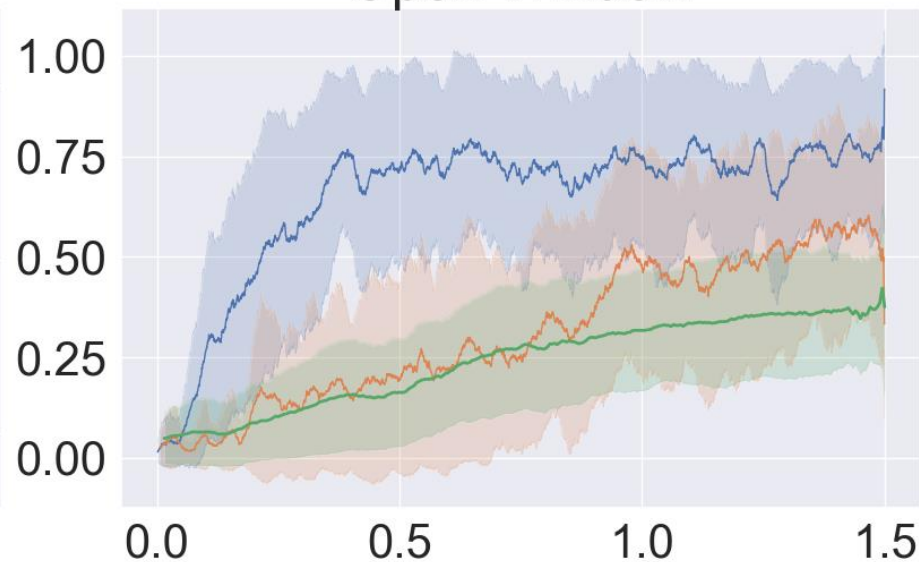
- Ours
- Multi-Task Multi-Head SAC-4
- Multi-Task Multi-Head SAC-4-Wide
- Multi-Task Multi-Head SAC-6

Comparison with Single Task Policy

Open Door

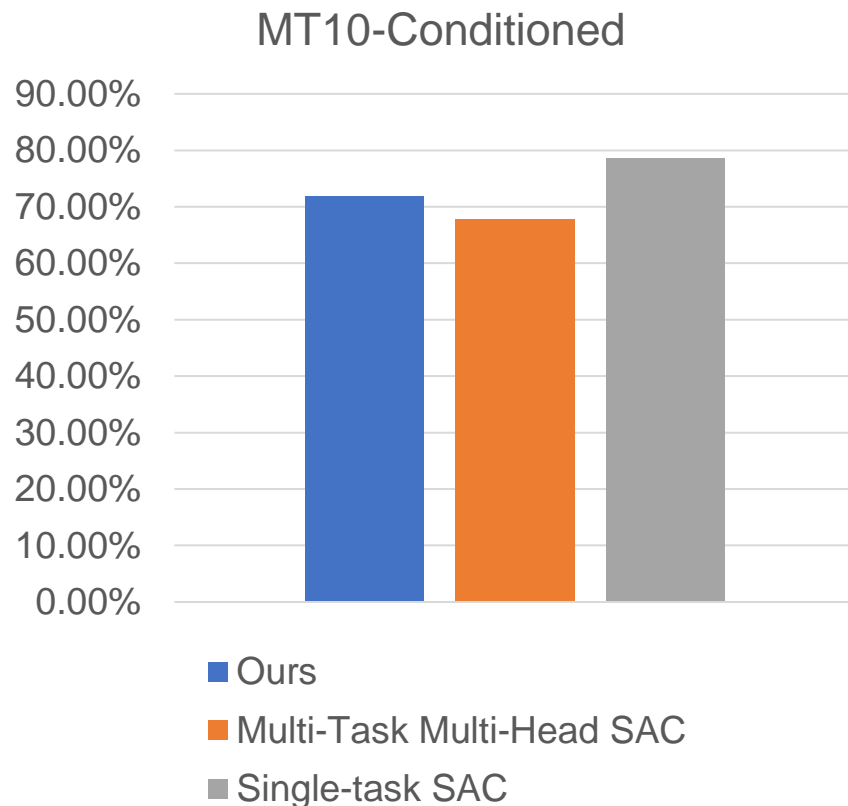


Open Window



— Ours — Multi-task Multi-head SAC — Single Task SAC

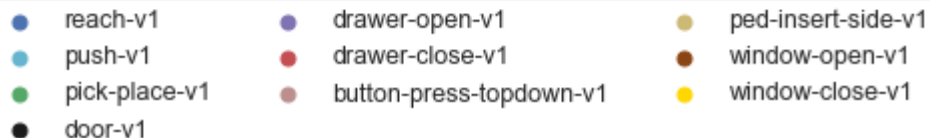
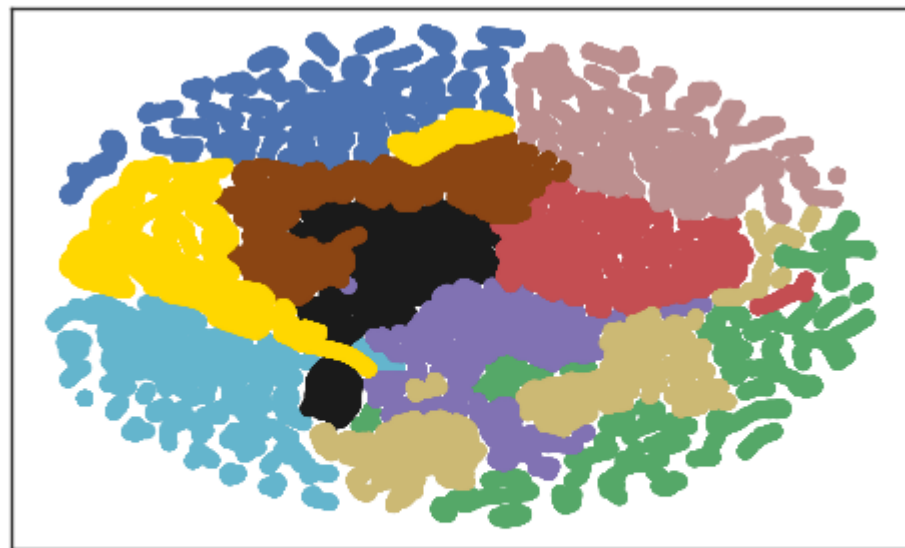
Comparison with Single Task Policy



Routing Visualization

Different Task:
Separated

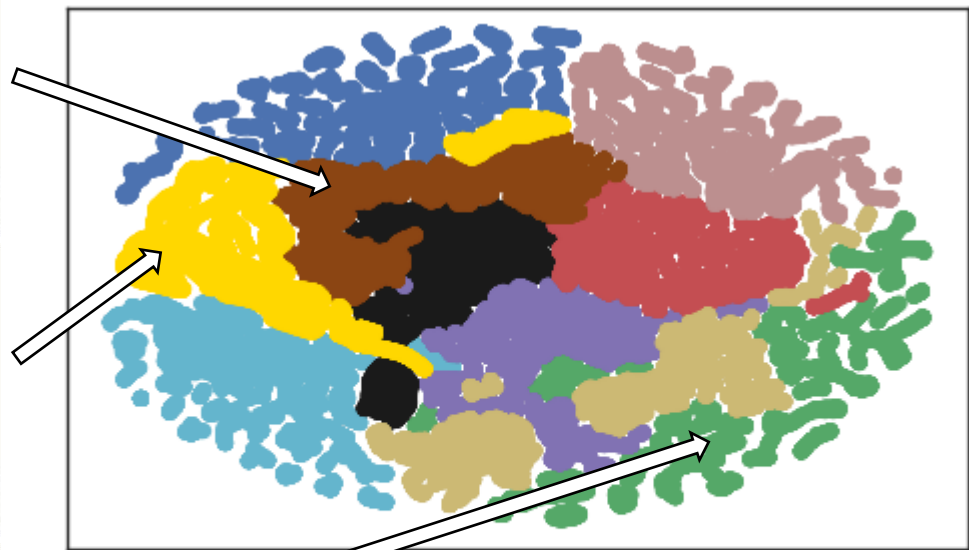
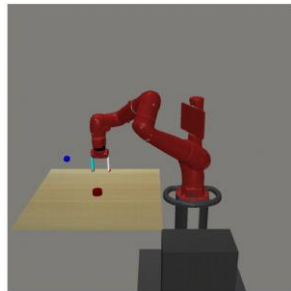
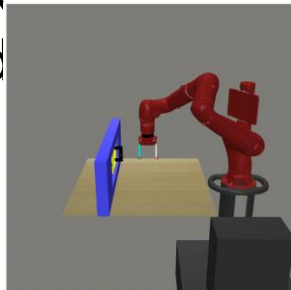
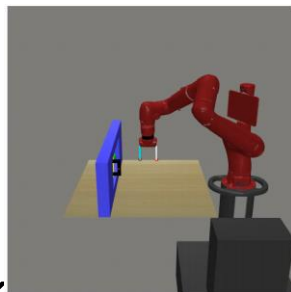
Similar Task:
Closer



Routing Visualization

Different Task
Separated

Similar Task:
Closer



reach-v1	drawer-open-v1	ped-insert-side-v1
push-v1	drawer-close-v1	window-open-v1
pick-place-v1	button-press-topdown-v1	window-close-v1
door-v1		

Thanks!

Our project page: <https://rchalyang.github.io/SoftModule>