澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU

Sparse Diffusion Policy: A Sparse, Reusable, and Flexible Policy for Robot Learning

GONG Ying
gong.ying@um.edu.mo

2024.8.14

Sparse Diffusion Policy: A Sparse, Reusable, and Flexible Policy for Robot Learning

Yixiao Wang[1*†], Yifei Zhang[1*], Mingxiao Huo[2*†], Ran Tian[1], Xiang Zhang[1], Yichen Xie[1], Chenfeng Xu[1], Pengliang Ji[2], Wei Zhan[1], Mingyu Ding[1‡], Masayoshi Tomizuka[1]
[1]UC Berkeley [2]Carnegie Mellon University

arXiv:2407.01531v1
1 Jul 2024

Abstract: The increasing complexity of tasks in robotics demands efficient strategies for multitask and continual learning. Traditional models typically rely on a universal policy for all tasks, facing challenges such as high computational costs and catastrophic forgetting when learning new tasks. To address these issues, we introduce a *sparse*, *reusable*, and *flexible* policy, Sparse Diffusion Policy (SDP). By adopting Mixture of Experts (MoE) within a transformer-based diffusion policy, SDP selectively activates experts and skills, enabling efficient and task-specific learning without retraining the entire model. SDP not only reduces the burden of active parameters but also facilitates the seamless integration and reuse of experts across various tasks. Extensive experiments on diverse tasks in both simulations and real world show that SDP 1) excels in multitask scenarios with negligible increases in active parameters, 2) prevents forgetting in continual learning of new tasks, and 3) enables efficient task transfer, offering a promising solution for advanced robotic applications. Demos and codes can be found in https://forrest-110.github.io/sparse_diffusion_policy/.

Keywords: Robot learning, Multitask and continual learning, Mixture of experts

Contents

Introduction

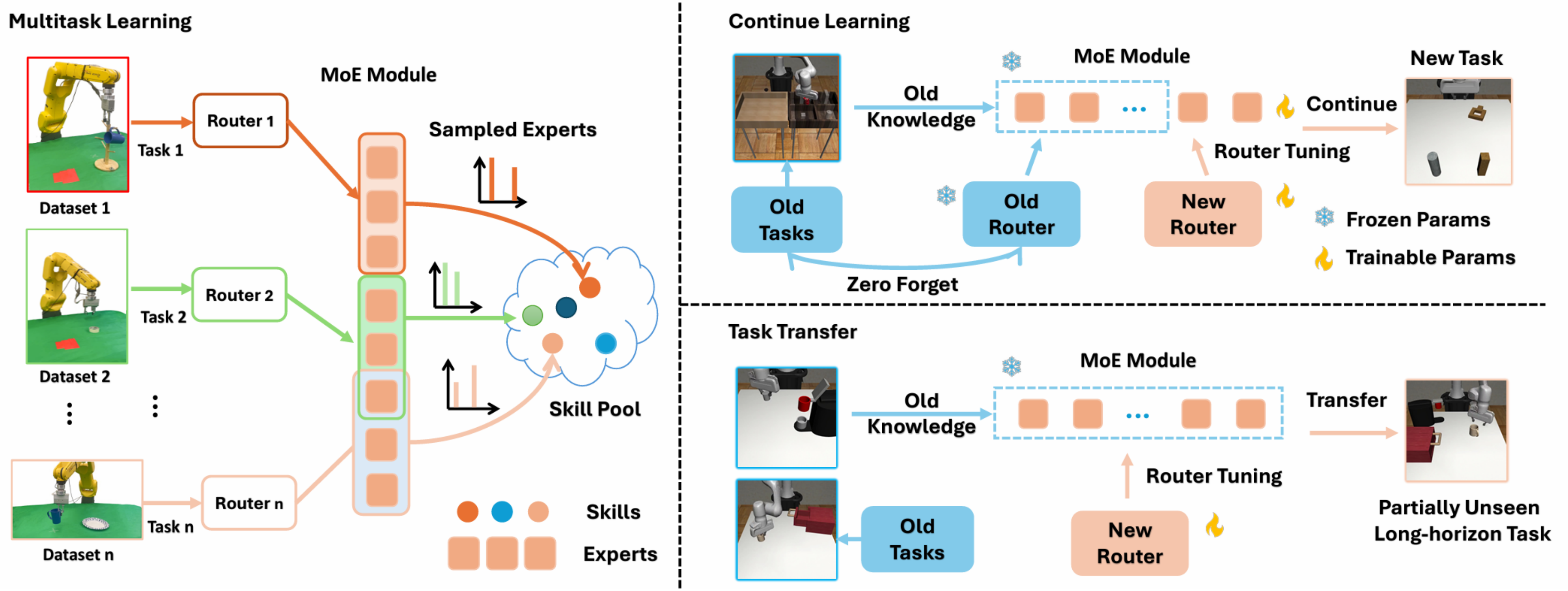
Motivation

- Traditional approaches often rely on a universal and monolithic policy for all tasks, activating all the parameters in the large network for even simple tasks.
- When encountering a new task, these approaches require costly fine-tuning.

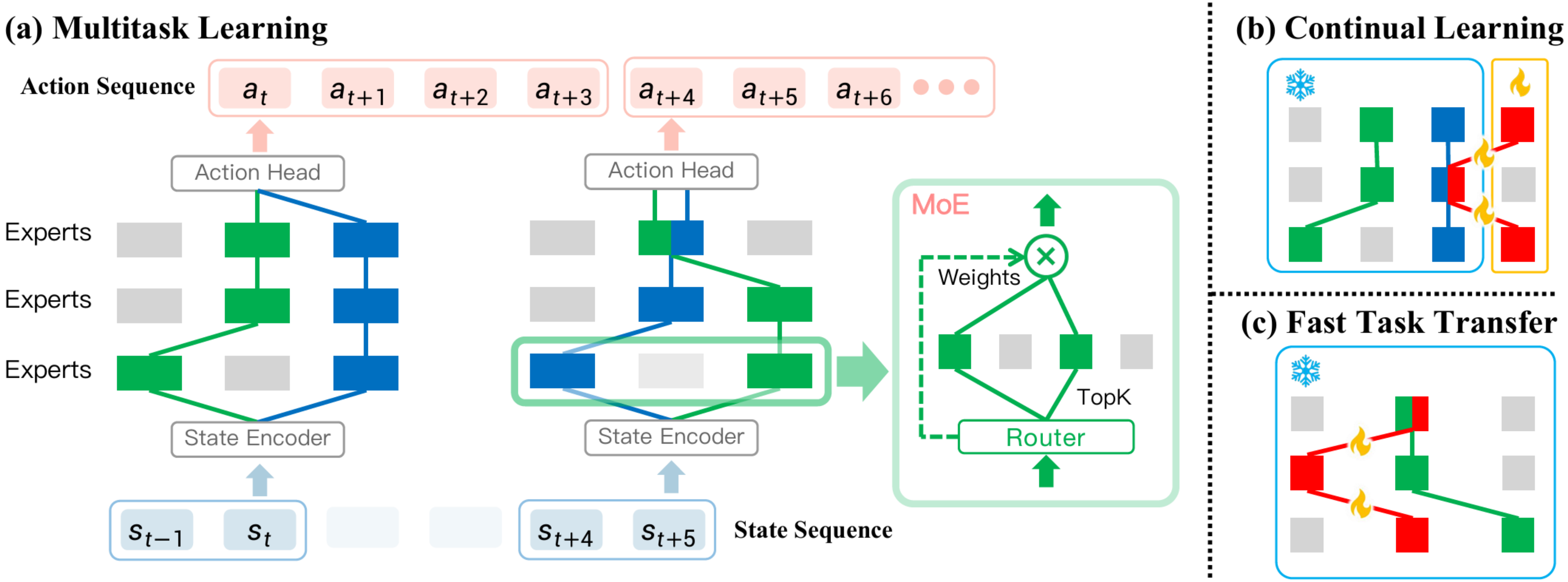
Introduction

Contributions

- **Multitask Learning**: Due to its sparsity, SDP can activate different experts for different tasks. Additionally, with its reusability, SDP can activate the same expert to share knowledge among tasks.
- **Continual Learning**: SDP can transfer to new tasks by adding only a few new experts to learn the new tasks.
- **Task Transfer**: SDP can transfer to new tasks by tuning the old experts and routers for expert selection.

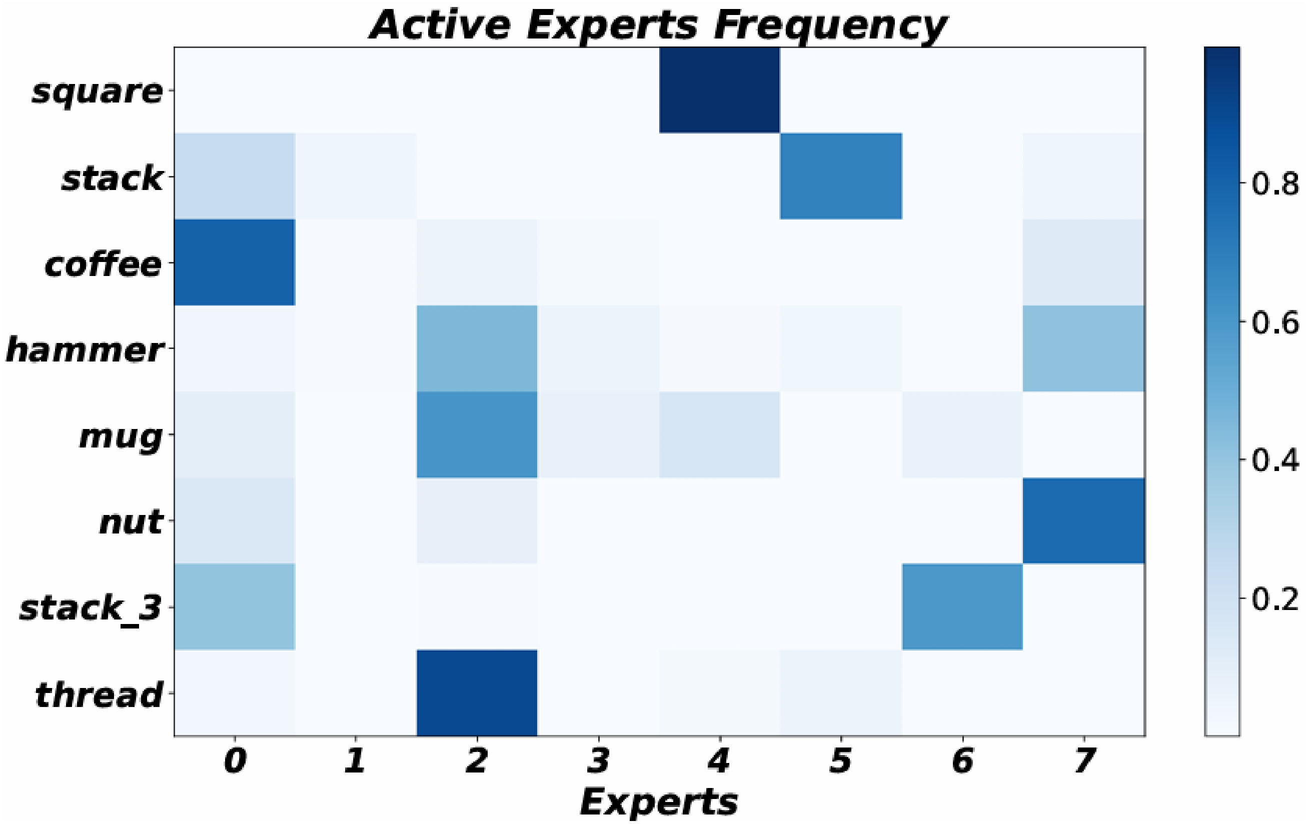


Method


(a) Multitask Learning
Action Sequence
a_t a_{t+1} a_{t+2} a_{t+3} a_{t+4} a_{t+5} a_{t+6}
Action Head
Action Head
Experts
Experts
Experts
MoE
Weights
TopK
Router
State Encoder
State Encoder
s_{t-1} s_t s_{t+4} s_{t+5}
State Sequence
(b) Continual Learning
(c) Fast Task Transfer

Experiments

Expert Selection Frequency in 2D Simulation



Experiments

2D Simulation

Method	Active Params	Square	Stack	Coffee	Hammer	Mug	Nut	Stack three	Thread	Avg.
TH	52.6 M	**0.76**	0.98	0.72	0.97	0.63	**0.52**	0.73	0.55	0.73
TT w/ 3Layer	52.6 M	0.73	0.95	0.76	**0.99**	0.65	0.49	0.68	0.59	0.73
TCD [76, 19]	52.7 M	0.63	0.95	0.77	0.92	0.53	0.44	0.62	0.56	0.68
SDP(Ours)	53.3 M	0.74	**0.99**	**0.83**	0.98	**0.70**	0.42	**0.76**	**0.65**	**0.76**

3D Simulation

Method	Toilet	Faucet	Laptop	Avg.
TT w/ 1Layer	0.73	0.35	**0.85**	0.64
TCD [76, 19]	0.72	0.33	0.80	0.62
SDP(Ours)	**0.75**	**0.43**	0.82	**0.67**

Real-world

Method	Pull	Pick Place	Hang	Avg.
TCD [76, 19]	**1.0**	0.0	0.35	0.45
SDP(Ours)	**1.0**	**1.0**	**0.95**	**0.98**

Experiments

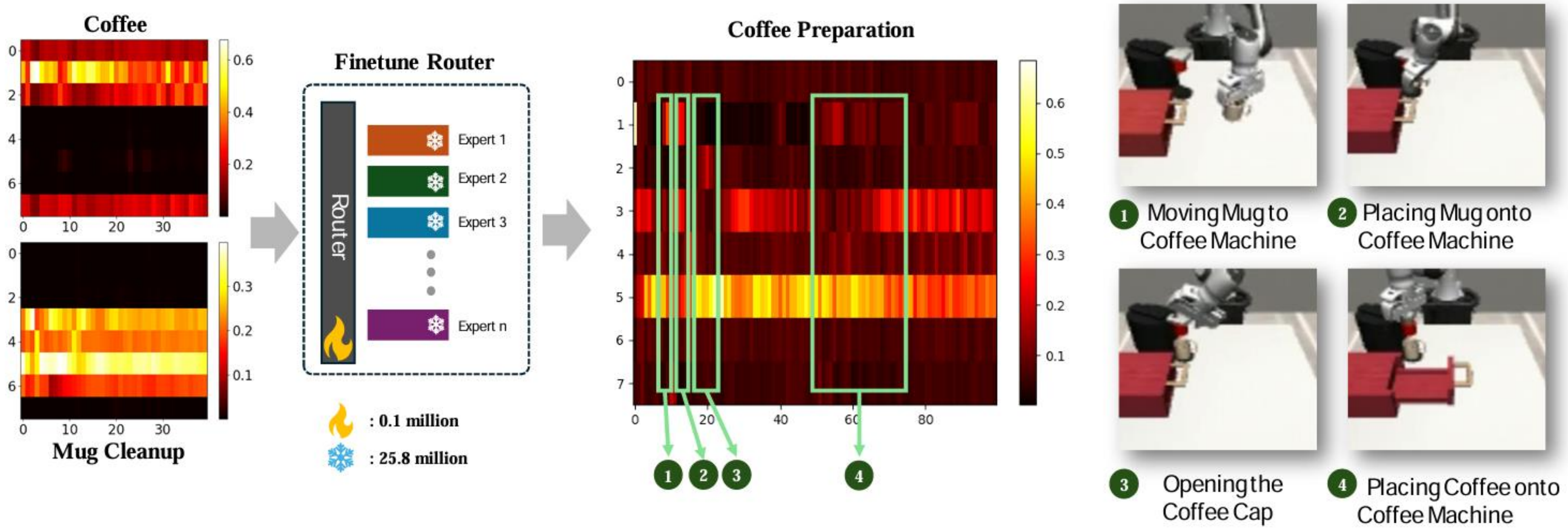
Continual Learning

Policy Decoder	Stage 1		Stage 2			Stage 3			
	Can	AP	Can	Lift	AP	Can	Lift	Square	AP
FFT	0.97	9.0 M	0.00	1.00	9.0 M	0.00	0.00	0.89	9.0 M
LoRA	0.94	9.0 M	0.94	1.00	12.0 M	0.94	1.00	0.73	14.9 M
MOE (Ours)	0.96	9.2 M	0.94	1.00	9.2 M	0.94	1.00	0.75	9.2 M

AP denotes Active Parameters of the policy network. Grey blocks indicate performance on new tasks; light-blue blocks indicate performance on previous tasks.

Experiments

Task Transfer



Method	Trainable Params	Coffee Preparation
Scratch	25.9M	0.70
Rou. only	**0.1M**	**0.80**

Thank you.

Task	Lift			Square			Can		
	Epoch	Steps	Accuracy	Epoch	Steps	Accuracy	Epoch	Steps	Accuracy
DP		2			2			2	
		10		100	10	0.94	50	5	0.96
	5	20	0.94		20			20	
Ours		2			2			2	
		10		100	10	**0.98**	50	5	**1**
	5	20	**1**		20			20	