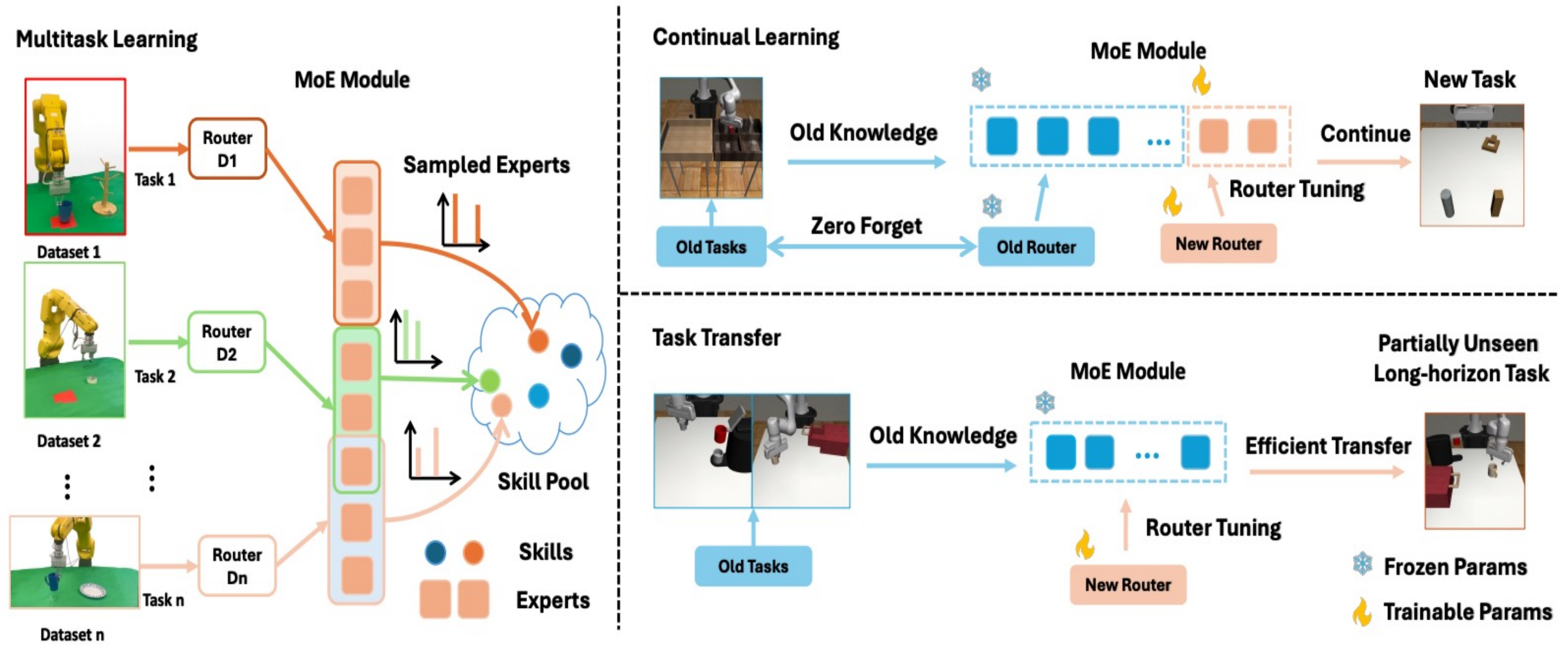
Sparse Diffusion Policy: A Sparse, Reusable, and Flexible Policy for Robot Learning

Yixiao Wang*, Yifei Zhang*, Mingxiao Huo*, Thomas Tian, Xiang Zhang, Yichen Xie, Chenfeng Xu, Pengliang Ji, Wei Zhan, Mingyu Ding†, Masayoshi Tomizuka

UC Berkeley CMU

Motivation

- Traditional dense models rely on a universal policy for all tasks, activating all model parameters, which leads to challenges such as high computational costs, catastrophic forgetting, and limited adaptability to various new tasks.
- We propose a Sparse Diffusion Policy (SDP) that integrates a Mixture of Experts module specifically designed for multitask learning, continual learning, and rapid adaptation to new tasks.

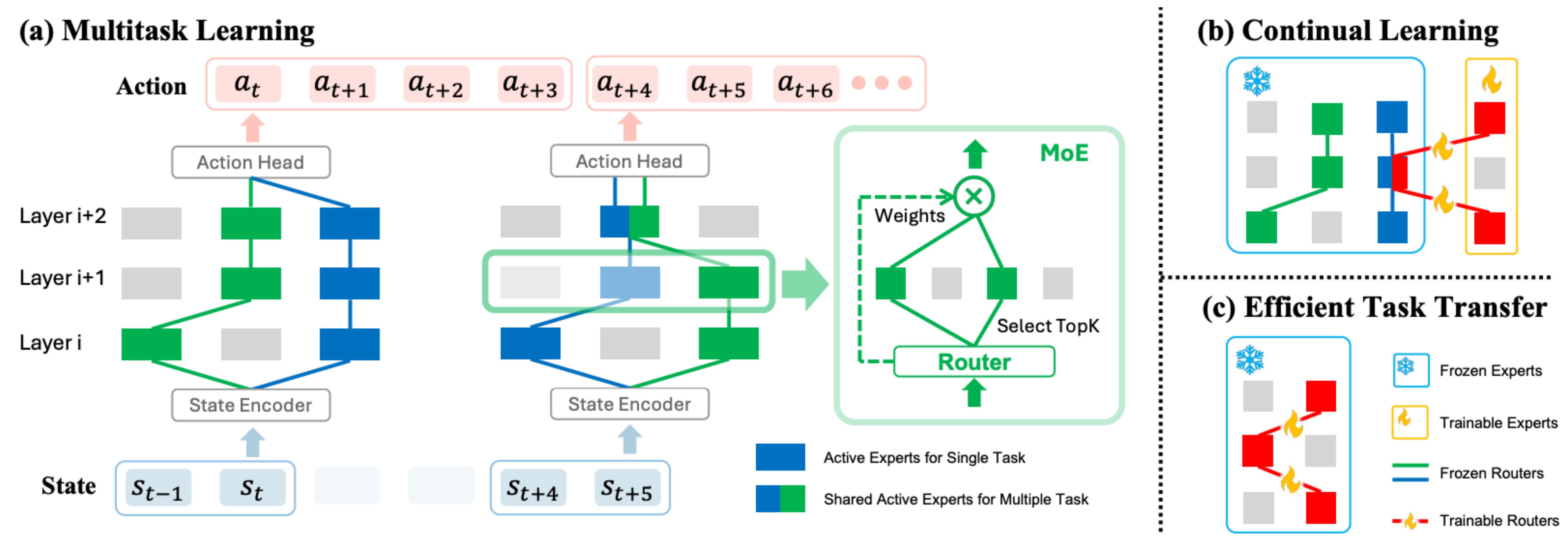


Contribution

- For multi-task learning, SDP achieves superior performance with a more than 20% reduction in active parameters compared to dense models..
- For continual learning, SDP maintains a higher success rate on new tasks without forgetting previously learned tasks.
- By training a highly lightweight router (less than 0.4% of the total parameters), SDP outperforms models trained from scratch for long-horizon task.

Method

- SDP activates sub-experts for multitask learning, which leads to computation efficiency.
- SDP learns new tasks and prevents forgetting of previous ones by freezing existing experts in the MoE, tuning the router, adding more experts.
- SDP freeze all old experts and only tune the router, can transfer to new task with old task knowledge.

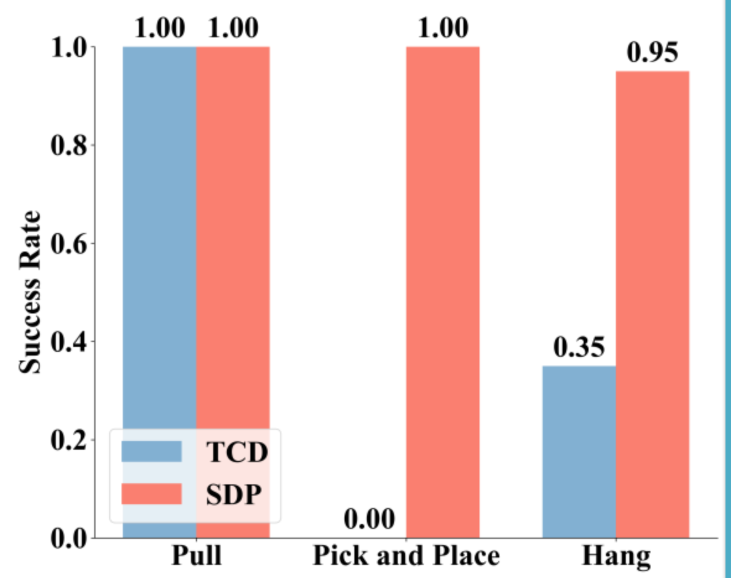


Results

Multi-task learning:

- For 2d simulation results, SDP outperformsbaseline methods in the same level of total parameters (TP), activate parameters (AP).
- For real world experiment, as shown in the right figure, SDP outperforms the previous methods in three tasks.

Method	TP(M)	AP(M)	Square	Stack	Coffee	Hammer	Mug	Nut	Stack Three	Thread	Avg.
TH	52.6	52.6	**0.76**	0.98	0.72	0.97	0.63	**0.52**	0.73	0.55	0.73
TT w/ 3Layer	144.7	52.6	0.73	0.95	0.76	**0.99**	0.65	0.49	0.68	0.59	0.73
TCD [76, 19]	52.7	52.7	0.63	0.95	0.77	0.92	0.53	0.44	0.62	0.56	0.68
Octo [77]	48.4	48.4	0.68	0.96	0.72	0.97	0.48	0.32	0.72	0.64	0.69
SDP	126.9	53.3	0.74	**0.99**	**0.83**	0.98	**0.70**	0.42	**0.76**	0.65	**0.76**
Light SDP	53.3	**38.7**	0.75	0.96	**0.83**	0.97	0.55	0.50	0.74	**0.73**	0.75



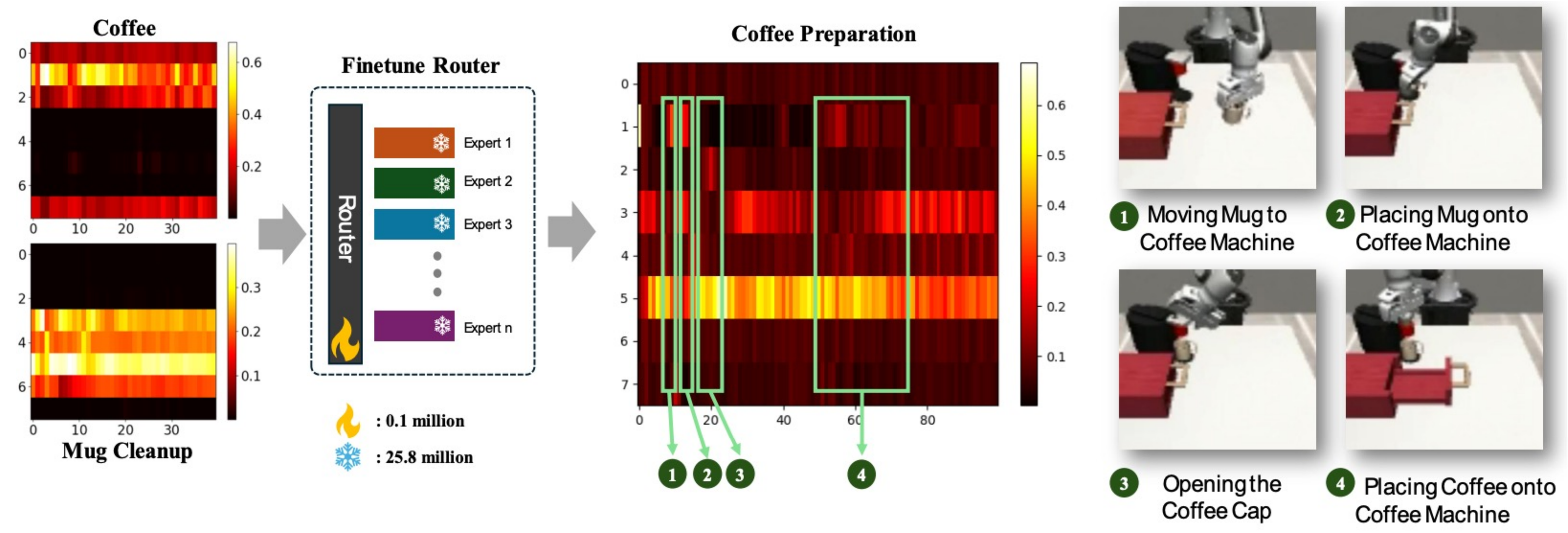
Continue learning:

- Our model outperforms baseline methods in success rate on both previous and new tasks while maintaining computational costs (same # of active parameters)..

Method	Stage 1 Can	Stage 1 AP	Stage 2 Can	Stage 2 Lift	Stage 2 AP	Stage 3 Can	Stage 3 Lift	Stage 3 Square	Stage 3 AP
FFT	0.97	9.0 M	0.00	1.00	9.0 M	0.00	0.00	0.89	9.0 M
LoRA[10]	0.94	9.0 M	0.94	1.00	12.0 M	0.94	1.00	0.73	14.9 M
SDP (Ours)	0.96	9.2 M	0.94	1.00	9.2 M	0.94	1.00	0.75	9.2 M

Task transfer:

- SDP can transfer to long-horizon tasks with only 0.4% of the router's parameters tuned, and also demonstrate a selective mechanism across experts, tasks, and skills.



Task Visualization