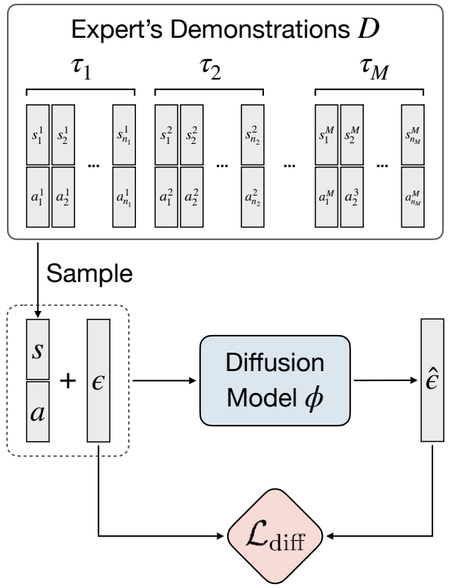
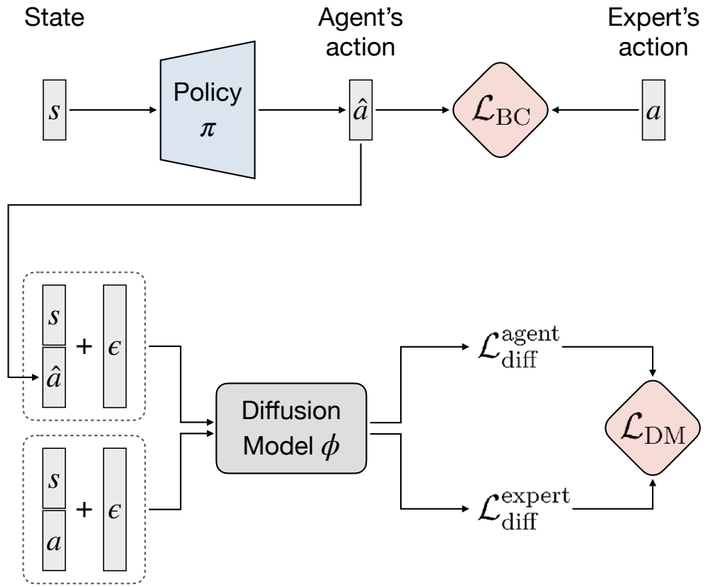
Diffusion Model-Augmented Behavioral Cloning

Submitted to NeurIPS 2023



(a) Learning a Diffusion Model



(b) Learning a Policy with the Learned Diffusion Model


Hsiang-Chun Wang*


Shang-Fu Chen*

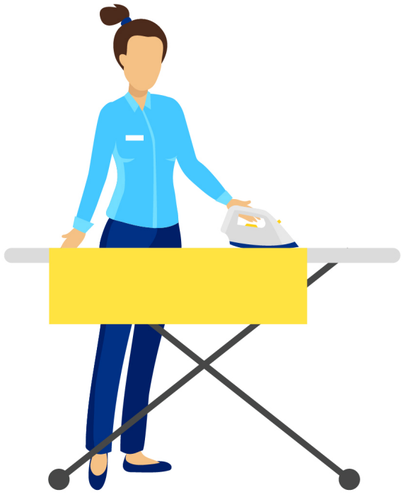

Ming-Hao Hsu


Chun-Mao Lai

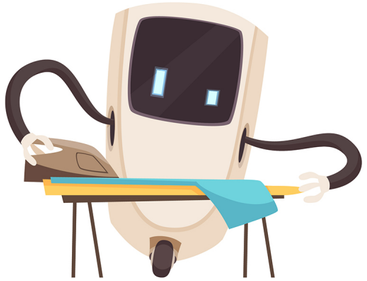

Shao-Hua Sun

Imitation Learning

Expert Demonstrations

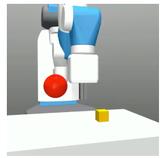
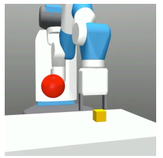
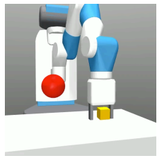
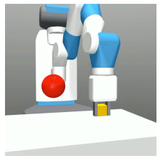
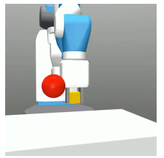


Imitate the Expert



Learning from **Demonstration**

$s^1_{1:n_1}$

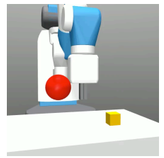
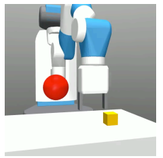
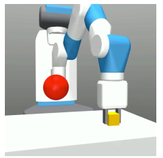
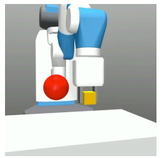
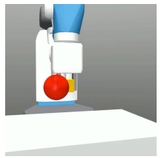






$a^1_{1:n_1}$

7.34	-2.00	-1.47	4.98	4.68
-3.17	8.51	1.07	-9.56	-1.41
2.06	-8.63	-1.88	6.92	9.85
3.18	3.61	3.04	-1.80	-1.73

⋮

$s^1_{1:n_M}$

$a^1_{1:n_M}$

5.32	3.91	1.64	2.56	5.31
2.59	6.38	-1.17	-2.50	-2.12
-2.40	-5.21	1.43	2.13	1.66
6.86	1.27	-1.61	-1.26	-1.24

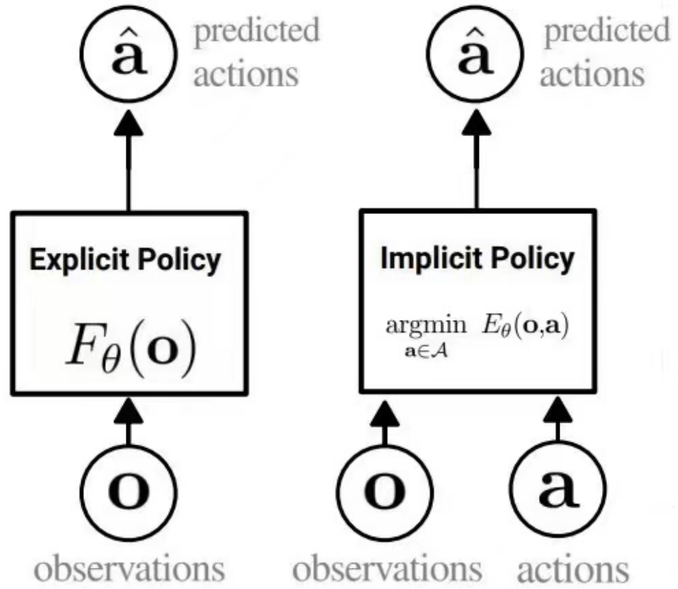
Dataset: Expert **states** and **actions**

Modeling Conditional Probability $p(a \mid s)$ vs Joint probability $p(s, a)$

Behavioral Cloning (BC): Directly learns a state-to-action mapping as a policy

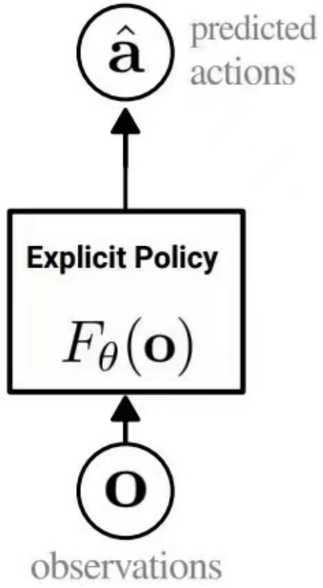
Implicit BC: Learns an energy-based model E(o, a) and samples actions given an observation

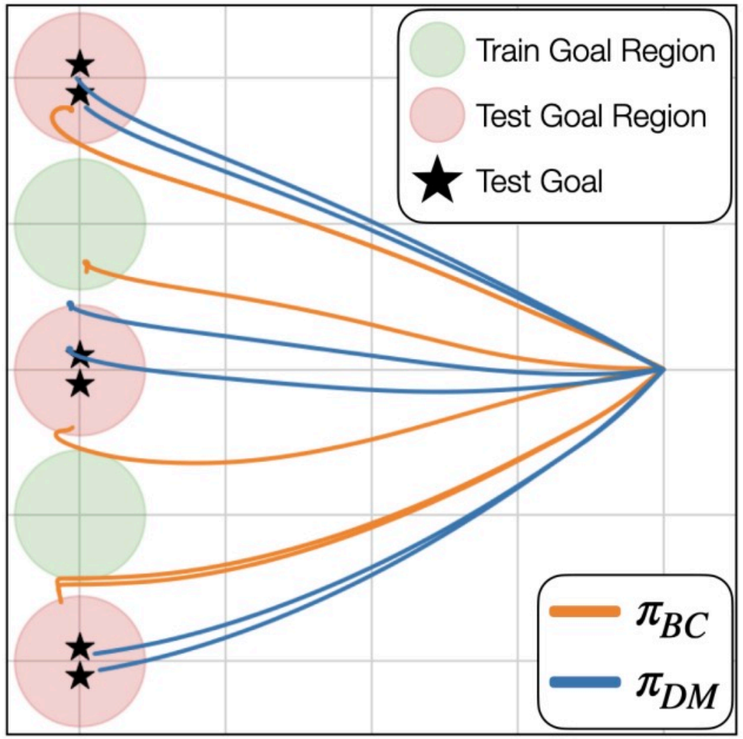
	Advantage	Disadvantage
BC (Conditional)	Simplicity & Training Stability	Generalize poorly
Implicit BC (Joint)	Generalize well	Time-consuming & Manifold Overfitting



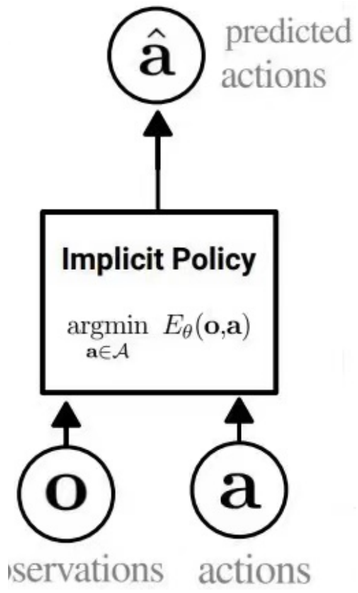
Reference: https://arxiv.org/pdf/2109.00137.pdf

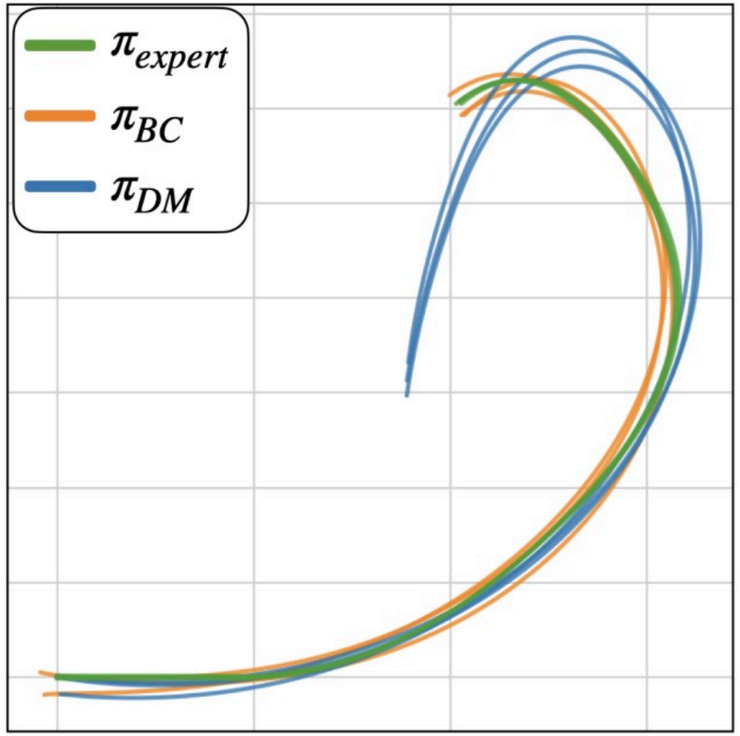
Conditional Probability $p(a \mid s)$ → Struggles at Generalizing





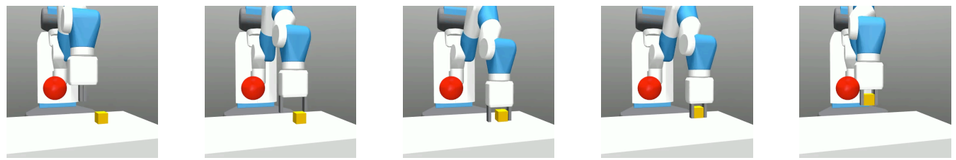
Joint probability $p(s, a)$ → Suffers from Manifold Overfitting





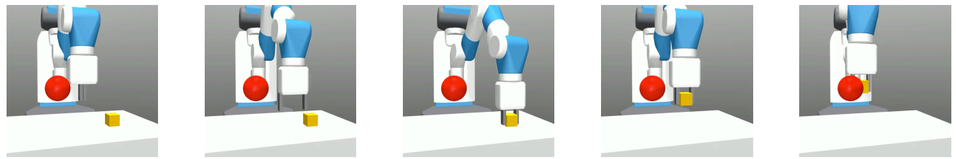
Modeling Expert State-Action Pair Distribution with Diffusion Model

Learning from **Demonstration**



$a^1_{1:n_1}$

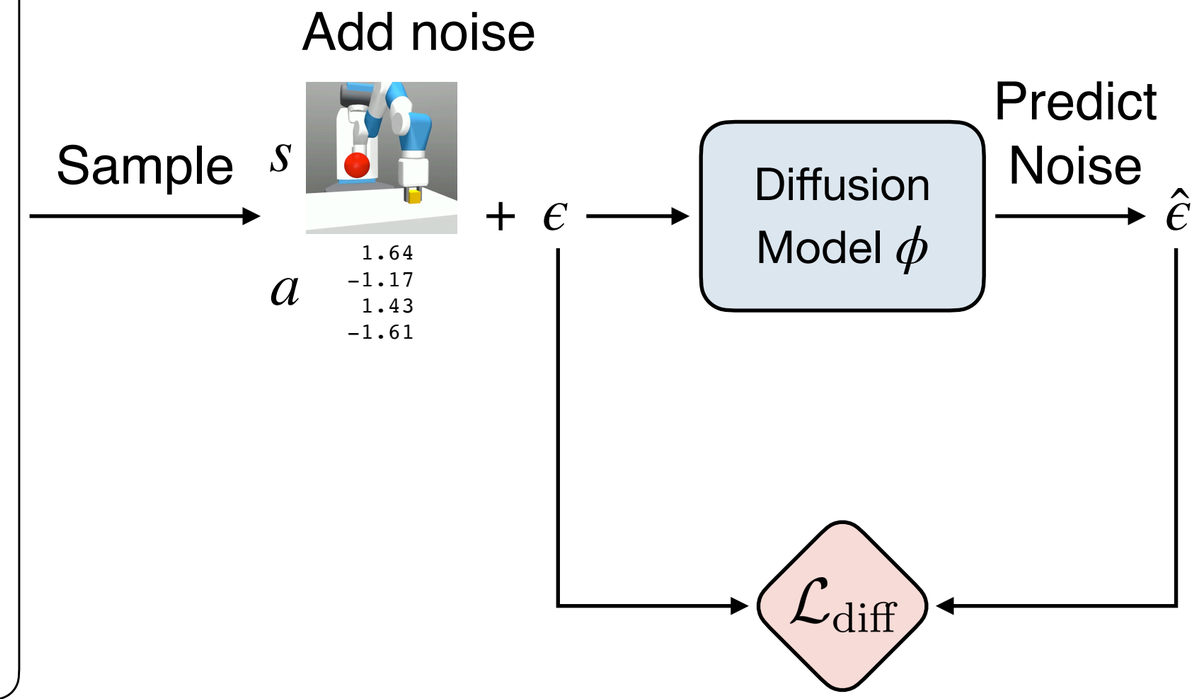
```
 7.34   -2.00   -1.47    4.98    4.68
-3.17    8.51    1.07   -9.56   -1.41
 2.06   -8.63   -1.88    6.92    9.85
 3.18    3.61    3.04   -1.80   -1.73
```



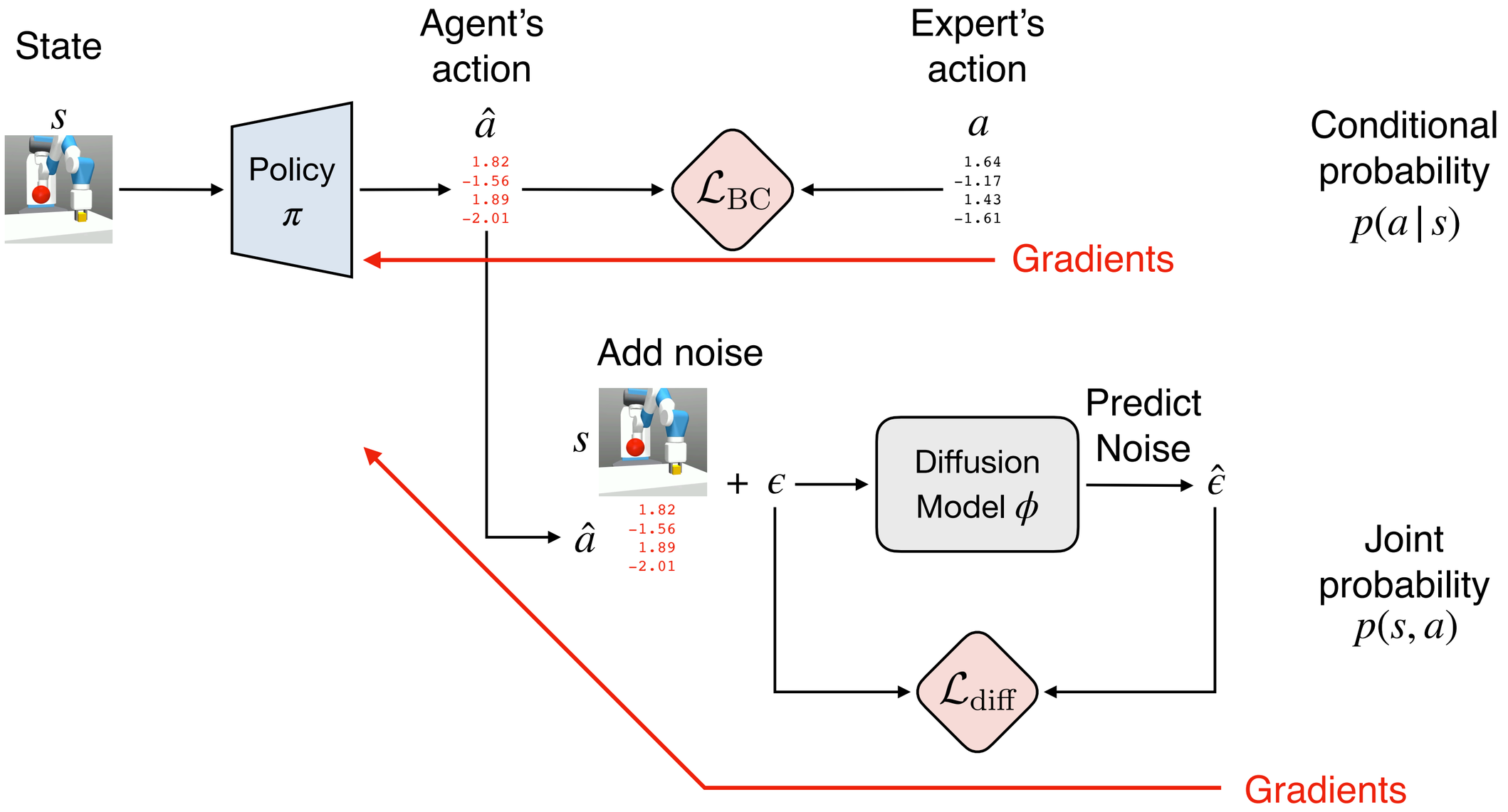
$a^1_{1:n_M}$

```
 5.32    3.91    1.64    2.56    5.31
 2.59    6.38   -1.17   -2.50   -2.12
-2.40   -5.21    1.43    2.13    1.66
 6.86    1.27   -1.61   -1.26   -1.24
```

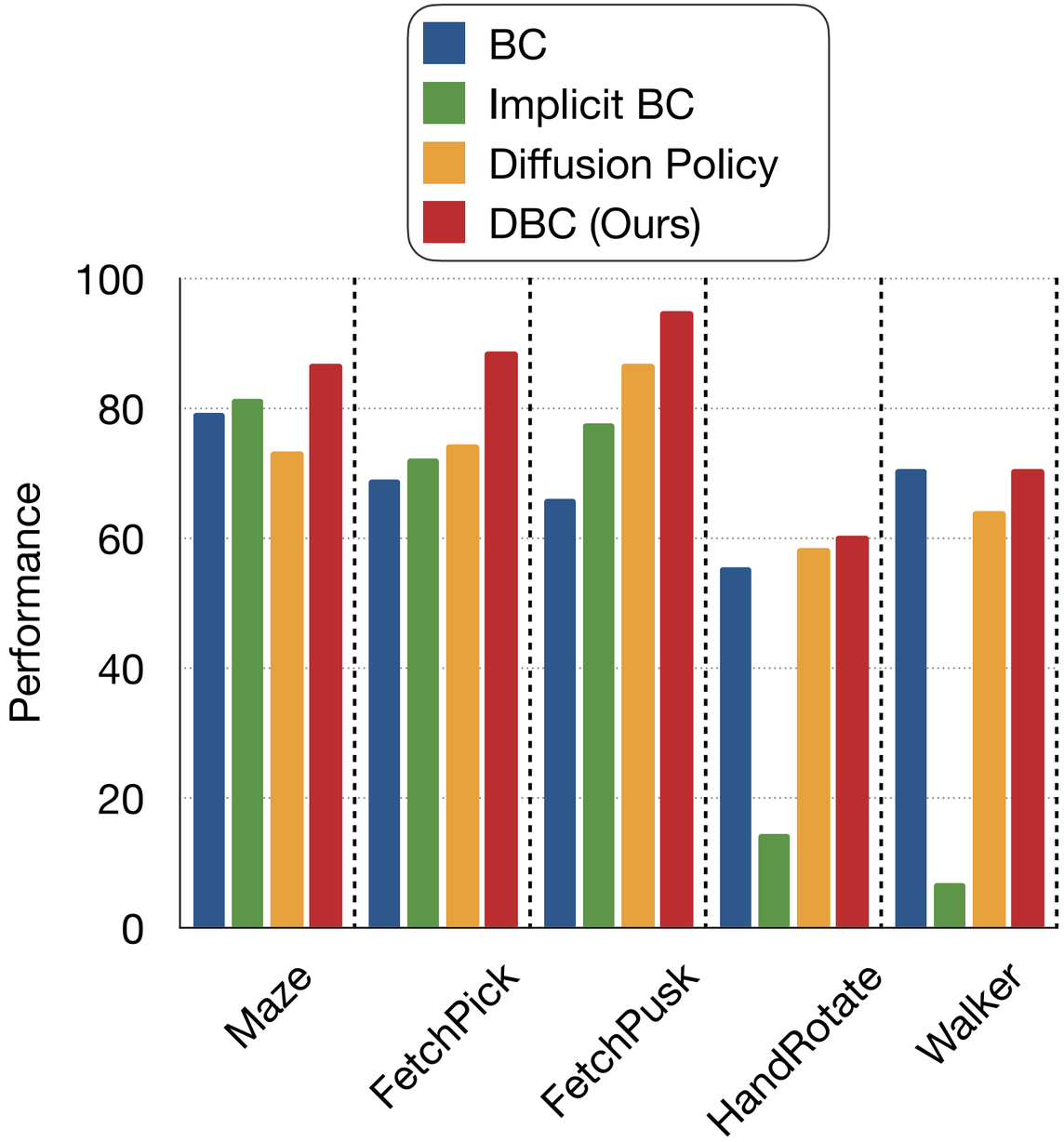
Dataset: Expert **states** and **actions**

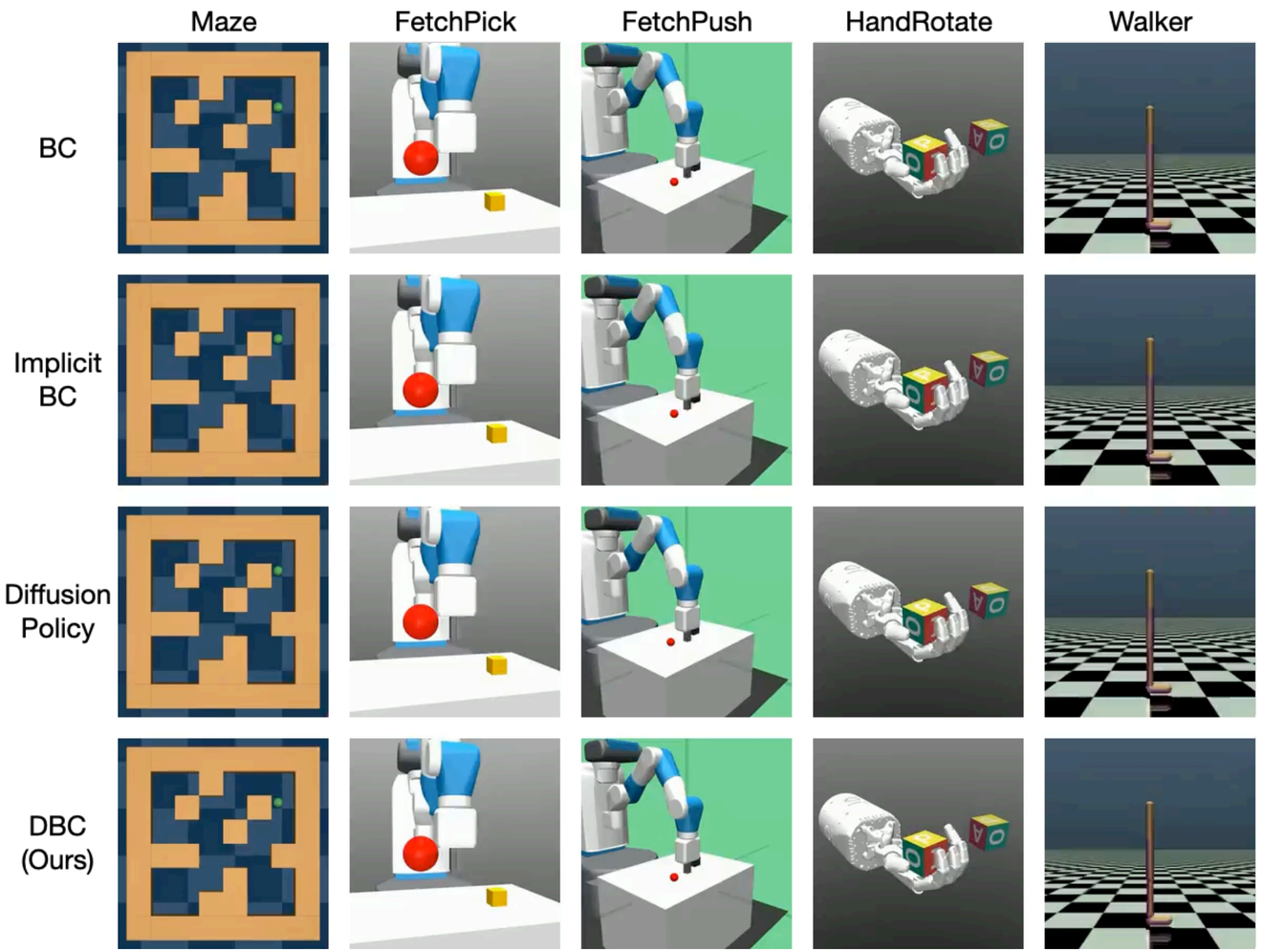


Learning a Policy with the Learned Diffusion Model



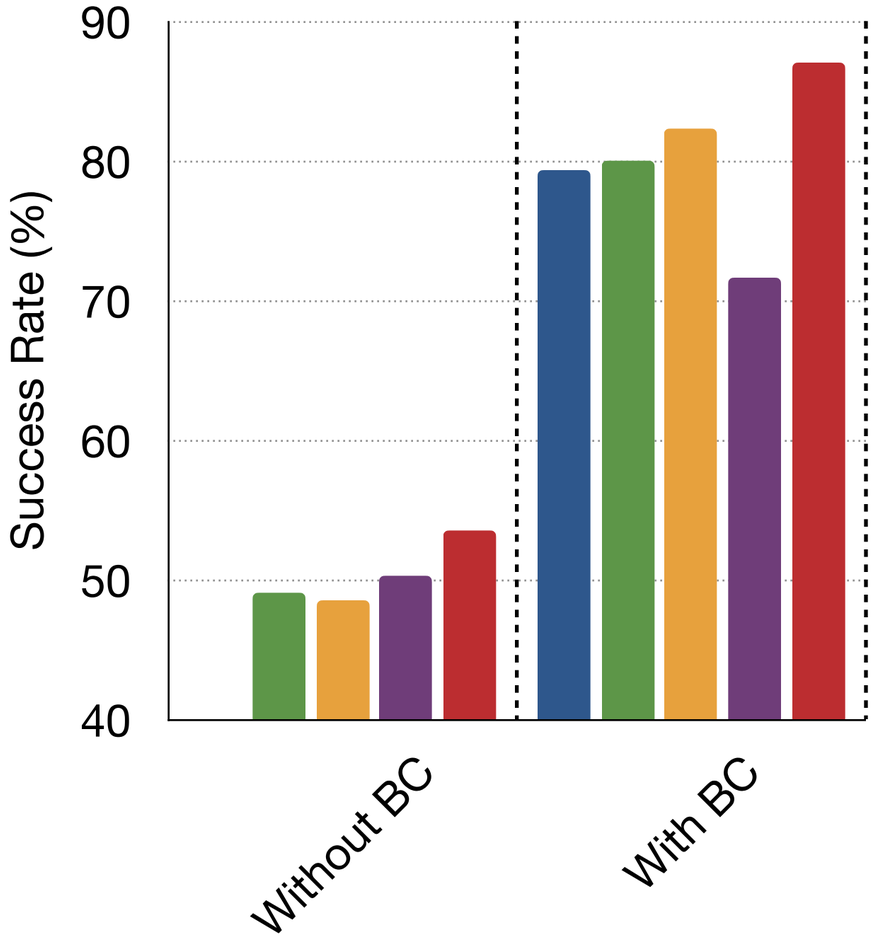
Experiments





Experiments - Generative Models

- Behavioral Cloning (BC)
- Energy-Based Model (EBM)
- Variational Autoencoder (VAE)
- Generative Adversarial Netwok (GAN)
- Diffusion Model (Ours)

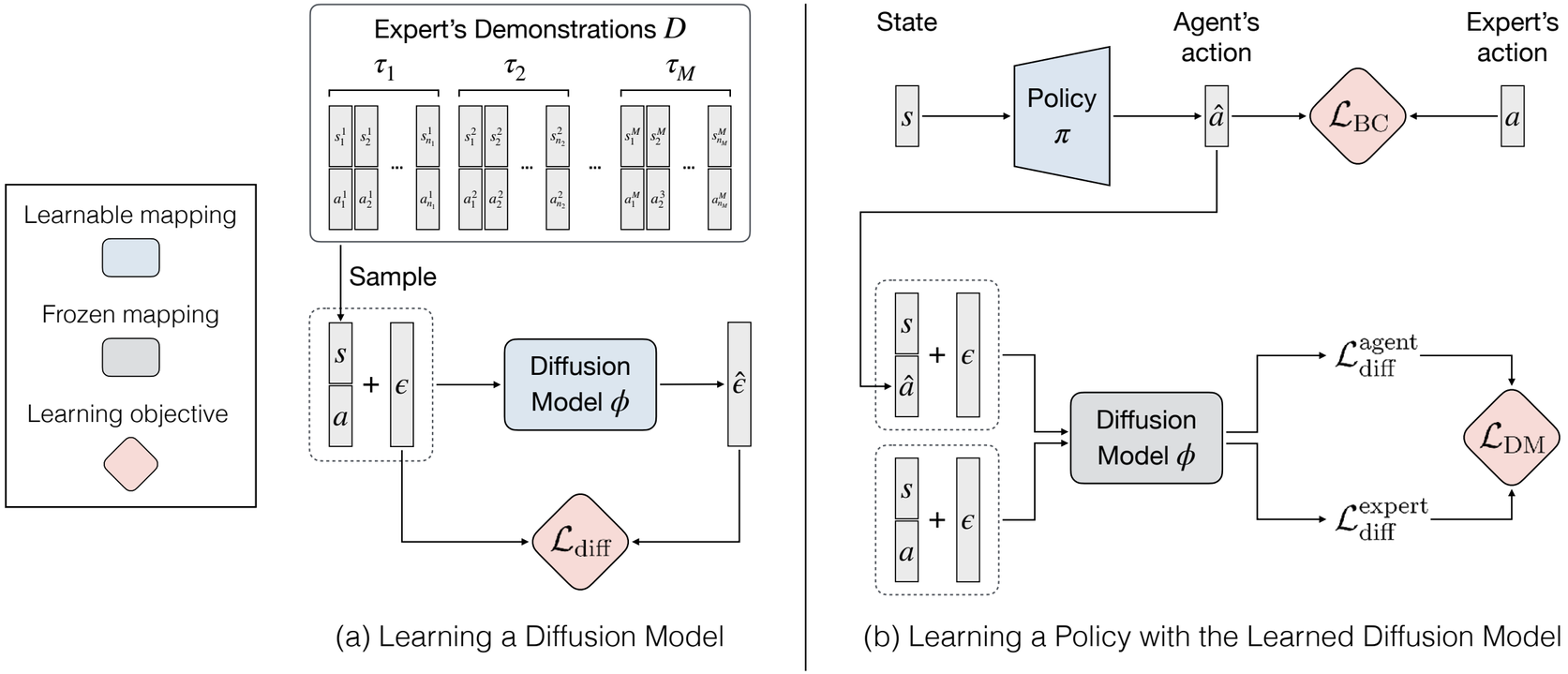


Findings

- Combining any GM with BC improves its performance
- Diffusion model (our choice) achieves the best performance

Takeaway

- Combining both the advantages of modeling the conditional probability p(a|s) and the joint probability p(s, a) of expert state-action pair distributions
- Model the expert state-action pair distribution with a diffusion model
- Leverage the learned diffusion model to learn a policy that mimics the expert



(a) Learning a Diffusion Model

(b) Learning a Policy with the Learned Diffusion Model

Rewards From This Project (Hsiang-Chun)

- Learn how to adjust the project progress dynamically
  - There are always problems when doing research
  - It is important to organize and summarize the progress
- Paper Writing Skill is Important
- Evaluate your own ability properly is essential
- Next time, I want to push myself and earn leeway on the project deadline
- Gratitude to our respected professor and my teammate

Rewards From This Project (Ming-Hao)

- Code Synchronization
  - During the research process, code synchronization can pose a significant challenge. It's essential to establish a robust system that allows all team members to work on the same codebase and maintain consistency.
- Experiment Settings
  - Deciding on and setting up experiments is another area that requires effective teamwork. Developing a shared understanding and agreement on experimental conditions can significantly enhance the team's overall efficiency.

Rewards From This Project (Ming-Hao)

- Explaining Results
  - A crucial part of team collaboration is the ability to interpret and explain experiment results logically and concisely. This includes identifying key findings, drawing valid conclusions, and linking results to the research objectives.
- Experiment Design for Hypothesis Checking
  - Following the interpretation of results, new experiments may be needed to validate or refute hypotheses. The team must work together to design follow-up experiments, ensuring they are effectively targeted towards hypothesis testing.

Rewards From This Project (Ming-Hao)

- Resource Management
  - Efficient team communication also extends to the allocation and management of resources. This involves sharing, requesting, and efficiently utilizing resources, including equipment, software, and information.
- Reporting Results
  - Another key aspect of team communication is the reporting of results. This involves sharing progress updates, preliminary findings, and final results in a manner that is clear, concise, and accessible to all team members.

Conclusion

- Reinforce the importance of these skills in fostering effective team collaboration. The ability to collaborate efficiently is essential in the modern workplace, helping to solve complex problems, improve productivity, and ultimately contribute to the successful completion of projects.

Acknowledgement

- Gratitude
  - I would like to express my heartfelt appreciation to our esteemed professor for providing me with the opportunity to delve into the world of collaborative paper writing. This invaluable learning experience has equipped me with vital skills for our academic and professional development.
- Guidance
  - I am also grateful for the continuous guidance and support I have received throughout this process. The professor and my teammates' expertise, insightful feedback, and constructive critiques have played a pivotal role in enriching my understanding and enhancing my proficiency in collaborative