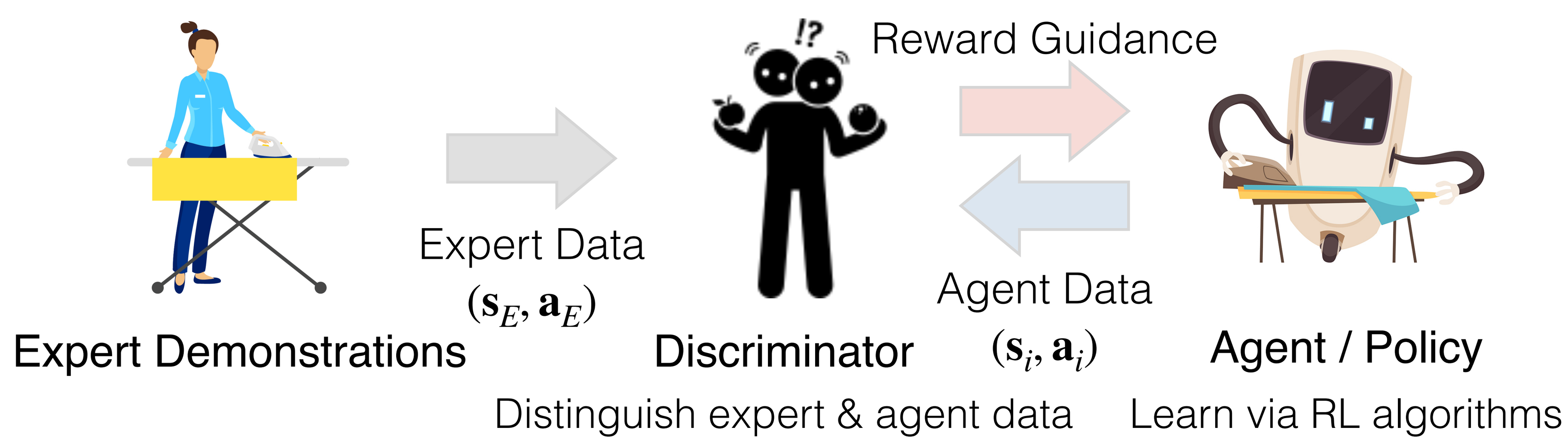
Diffusion Rewards Guided Adversarial Imitation Learning

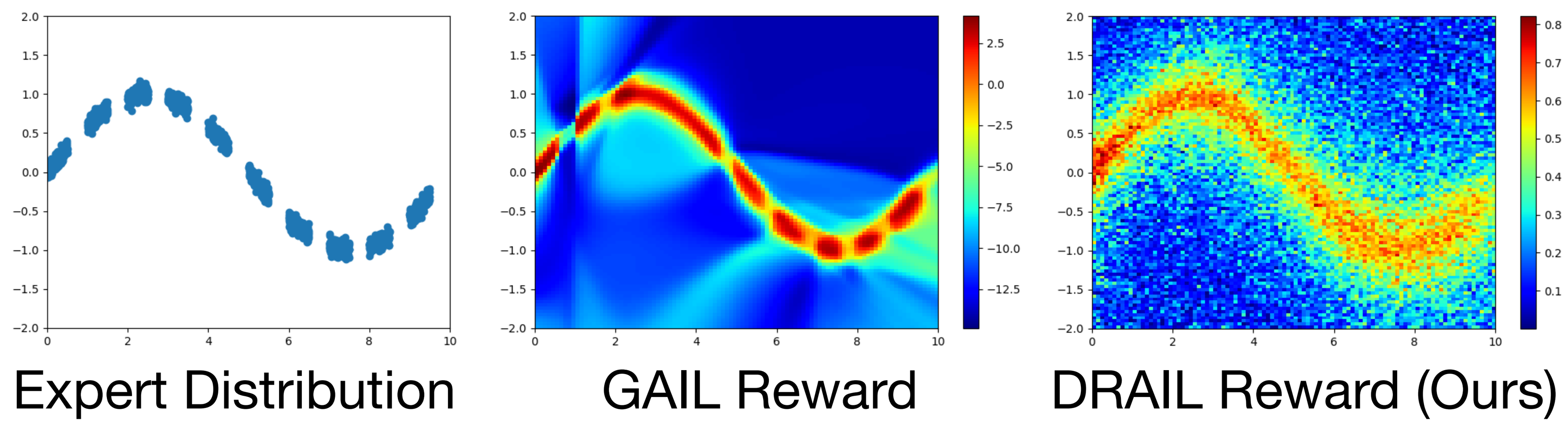
Chun-Mao Lai* Hsiang-Chun Wang* Ping-Chun Hsieh Yu-Chiang Frank Wang Min-Hung Chen Shao-Hua Sun

National Taiwan University

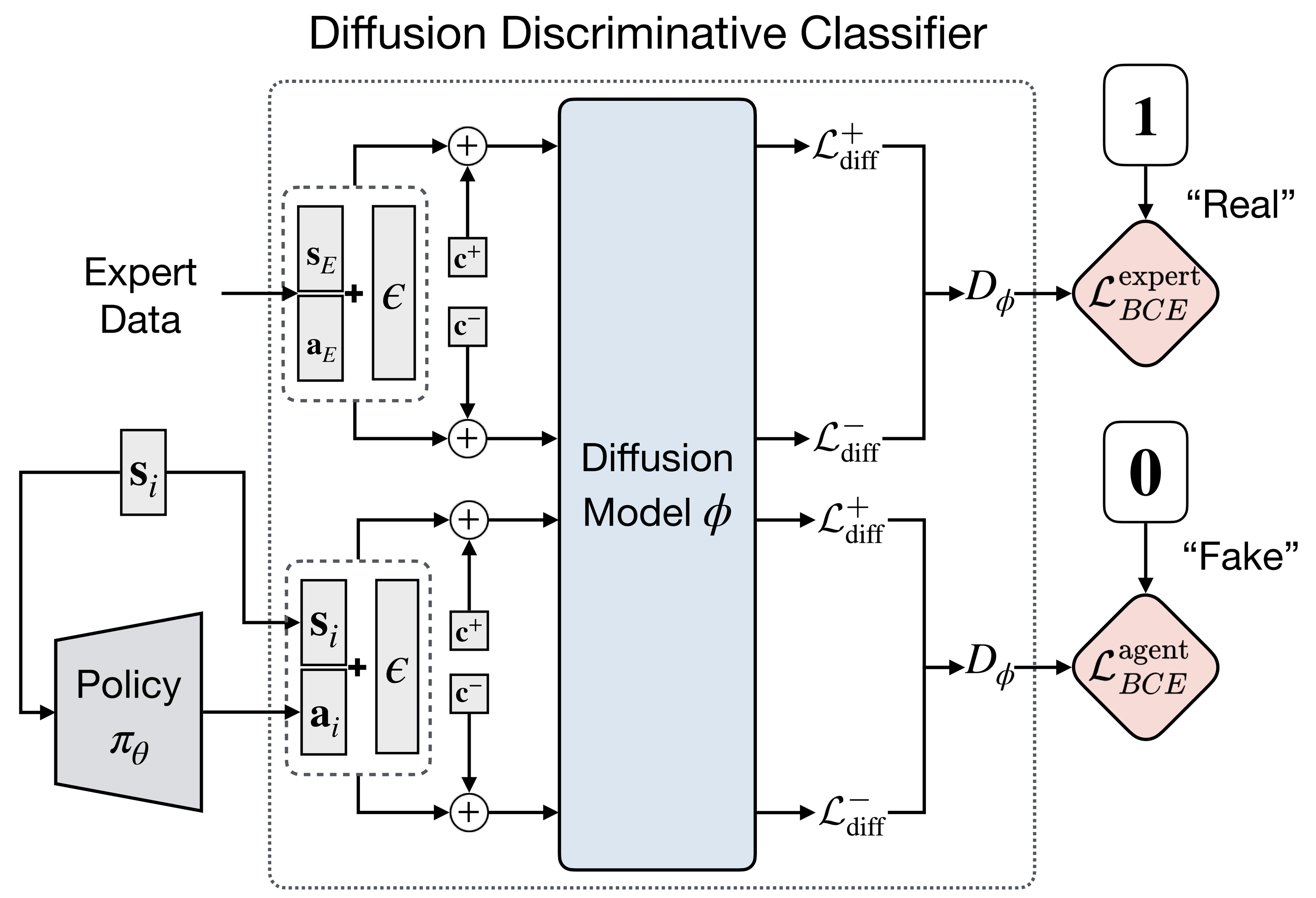
Adversarial Imitation Learning



Reward Function Visualization



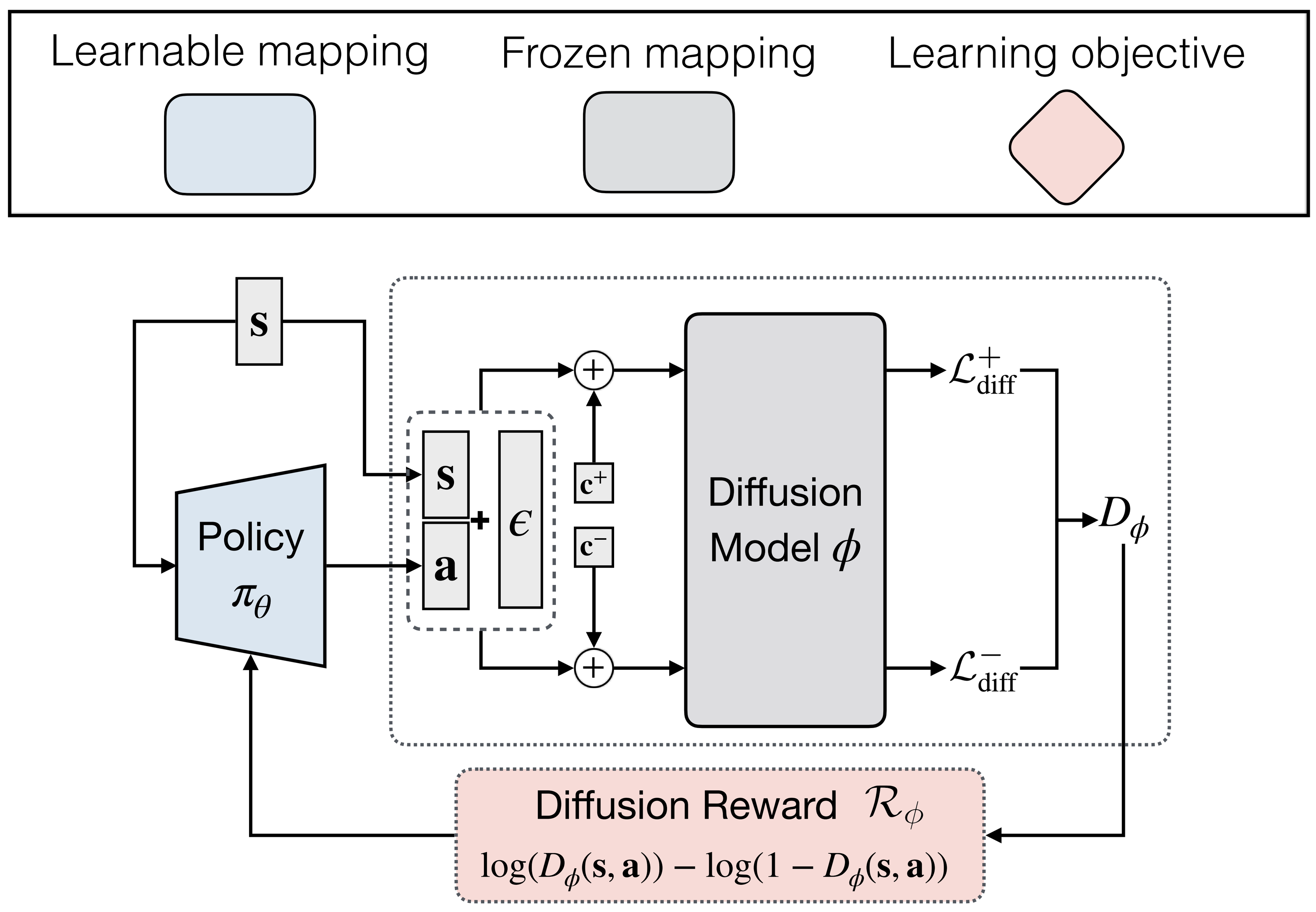
Diffusion Rewards Guided Adversarial Imitation Learning (DRAIL)



(a) Learning Diffusion Discriminative Classifier

Diffusion discriminative classifier learns to distinguish expert data $(\mathbf{s}_E, \mathbf{a}_E)$ from agent data $(\mathbf{s}_i, \mathbf{a}_i)$ using a diffusion model ϕ by denoising expert and agent state-action pairs concatenated with a real/expert label $\mathbf{c}^+$ or a fake/agent label $\mathbf{c}^-$.

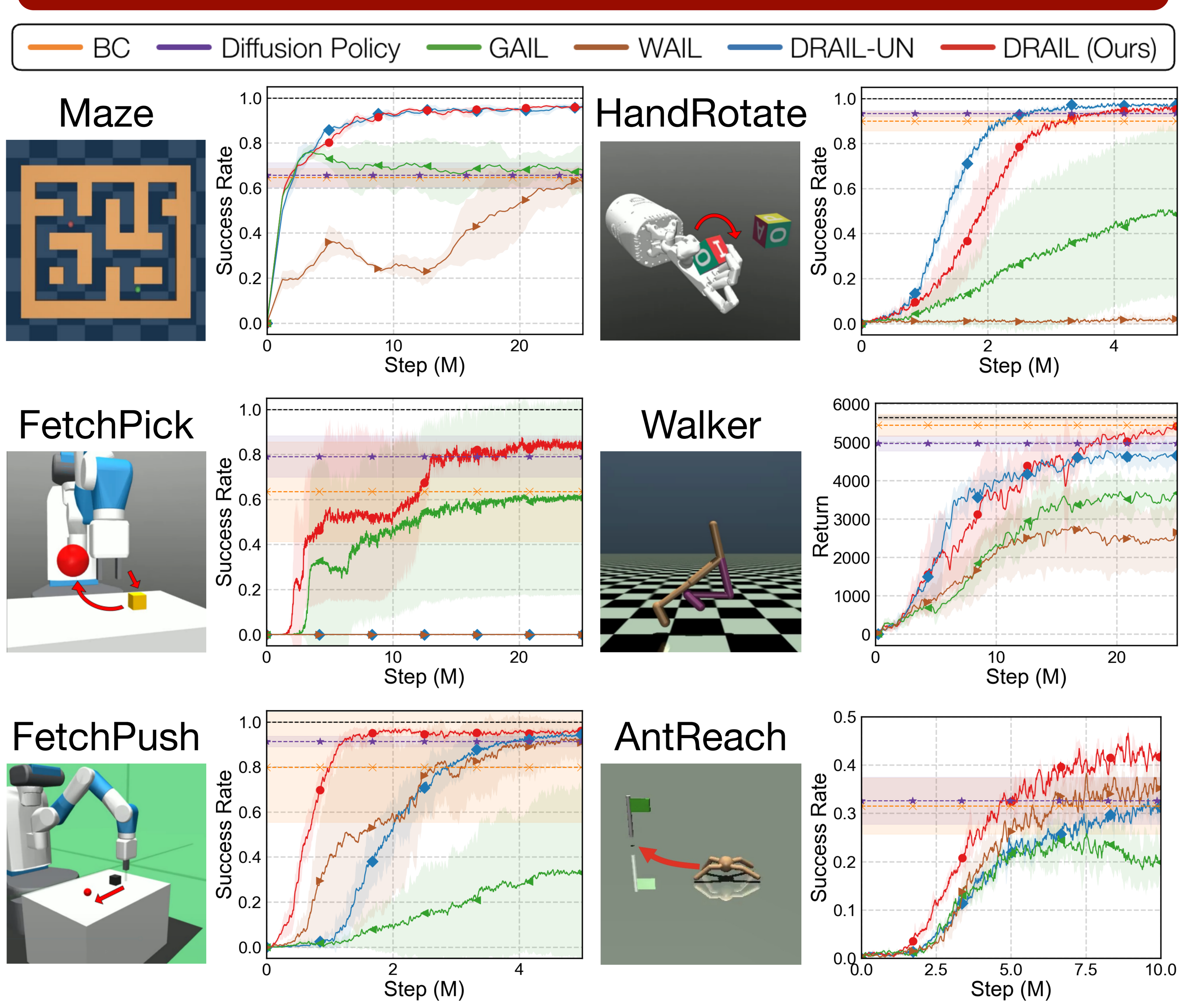
- **Diffusion Loss** $\mathcal{L}_{\text{diff}}(\mathbf{s},\mathbf{a},\mathbf{c}) = \mathbb{E}_{t\sim T}\left[\|\hat{\epsilon}_\phi(\mathbf{s},\mathbf{a},\boldsymbol{\epsilon},t|\mathbf{c}) - \boldsymbol{\epsilon}\|^2\right]$
- **"Realness" of $(\mathbf{s},\mathbf{a})$** $D_\phi(\mathbf{s},\mathbf{a}) = \dfrac{e^{-\mathcal{L}_{\text{diff}}(\mathbf{s},\mathbf{a},\mathbf{c}^+)}}{e^{-\mathcal{L}_{\text{diff}}(\mathbf{s},\mathbf{a},\mathbf{c}^+)} + e^{-\mathcal{L}_{\text{diff}}(\mathbf{s},\mathbf{a},\mathbf{c}^-)}}$
- **Diffusion Reward** $\mathcal{R}_\phi(\mathbf{s},\mathbf{a}) = \log(D_\phi(\mathbf{s},\mathbf{a})) - \log(1 - D_\phi(\mathbf{s},\mathbf{a}))$



(b) Learning Policy with Diffusion Rewards

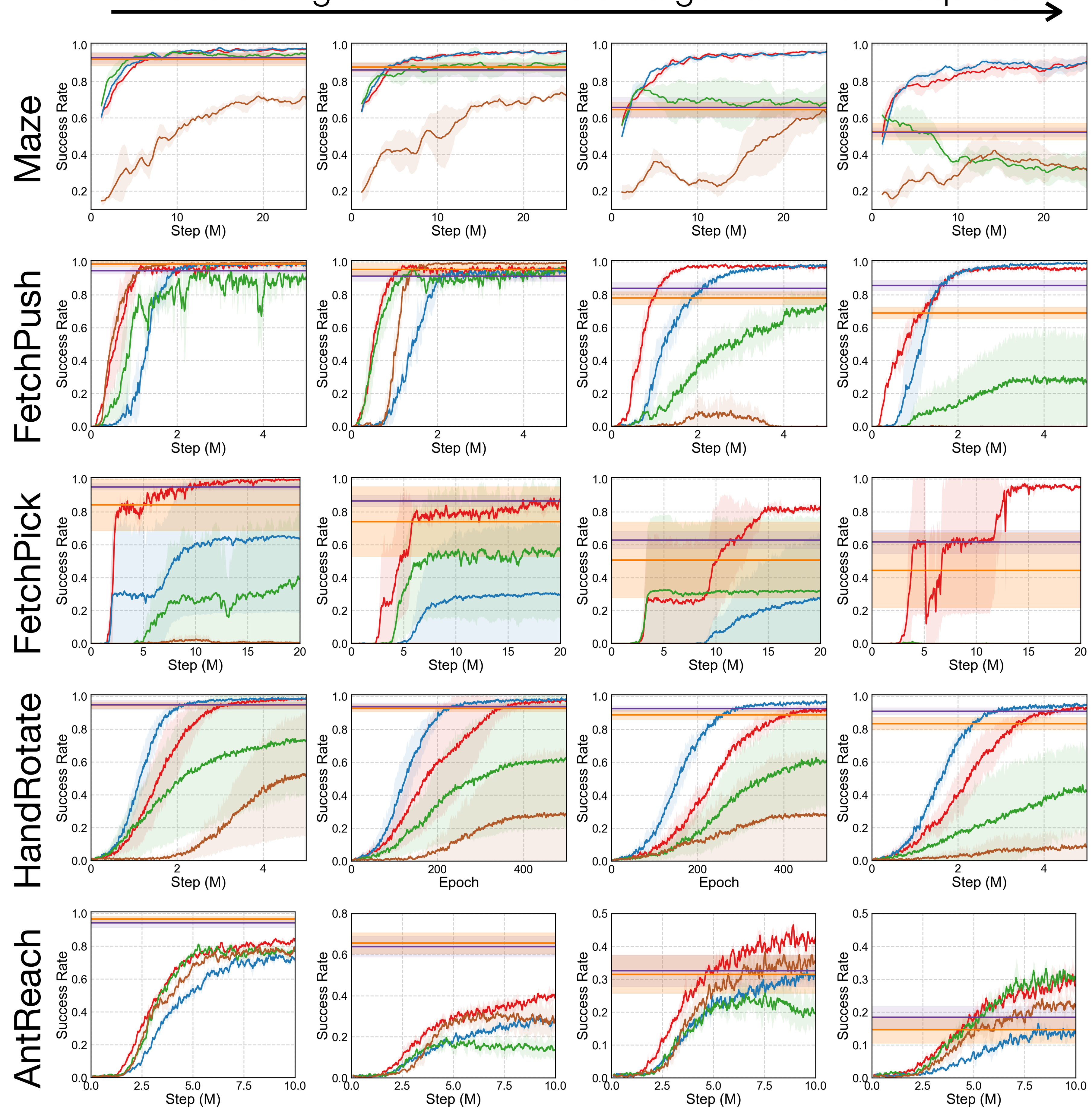
Policy π_θ learns to maximize the diffusion reward computed based on the output of the diffusion discriminative classifier D_ϕ that takes the state-action pairs from the policy as input.

Sample Efficiency



Generalizability

Generalizing to more novel states/goals/tasks is required



Data Efficiency