OCTOBER 1 - 5, 2023
IEEE/RSJ International Conference on Intelligent Robots and Systems

Robotics and Autonomous Driving Lab

GOATS: Goal Sampling Adaption for Scooping with Curriculum Reinforcement Learning

Yaru Niu[1], Shiyu Jin[2], Zeqing Zhang*[2,3], Jiacheng Zhu[1], Ding Zhao[1], Liangjun Zhang[2]*

[1]Carnegie Mellon University, [2]Baidu Research, [3]The University of Hong Kong

Full Paper

Project Page

Motivation

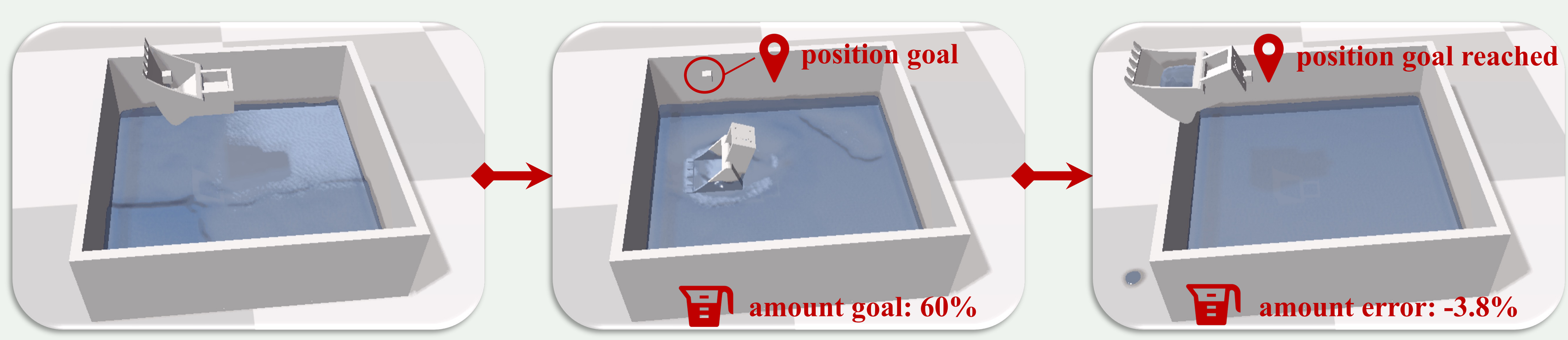
Robotic Water Scooping

- Scooping is an essential skill for human beings
- Robotic scooping has mainly focused on scooping solid materials
- Robotic liquid scooping can be helpful to many downstream tasks

Prior Works on Goal-Conditioned Deformable Object Manipulation

- Relatively simple goal state spaces
- Many rely on heuristics, primitives, demonstrations

Objectives & Challenges



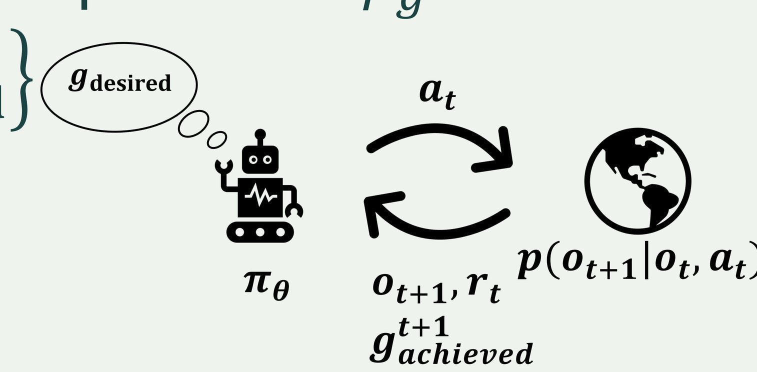
Challenges

- A long-horizon task for RL with a multi-modal goal state space
  - Position goal
  - Water amount goal
- Randomly initialized over a large combined space of water states and goal states
- Complex dynamics of water

Problem Formulation

A multi-goal reinforcement learning problem

- Goal-conditioned Markov Decision Process (MDP): $(\mathcal{S}, \mathcal{G}, \mathcal{A}, p, r, \rho_0, \rho_g)$
  - $\mathcal{G}$: a set of goals
  - ρ_g: goal distribution
- $g_{\text{desired}} = \{g^p_{\text{desired}}, g^a_{\text{desired}}\}$ sampled from ρ_g
- $g^t_{\text{achieved}} = \left\{g^{p(t)}_{\text{achieved}}, g^{a(t)}_{\text{achieved}}\right\}$
- $a_t \sim \pi_\theta(o_t, g_{\text{desired}}, g^t_{\text{achieved}})$
- $o_{t+1}, g^{t+1}_{\text{achieved}}$
- $r_t = r(g^{t+1}_{\text{achieved}}, g_{\text{desired}})$



Methodology

Goal-Factorized Reward Formulation

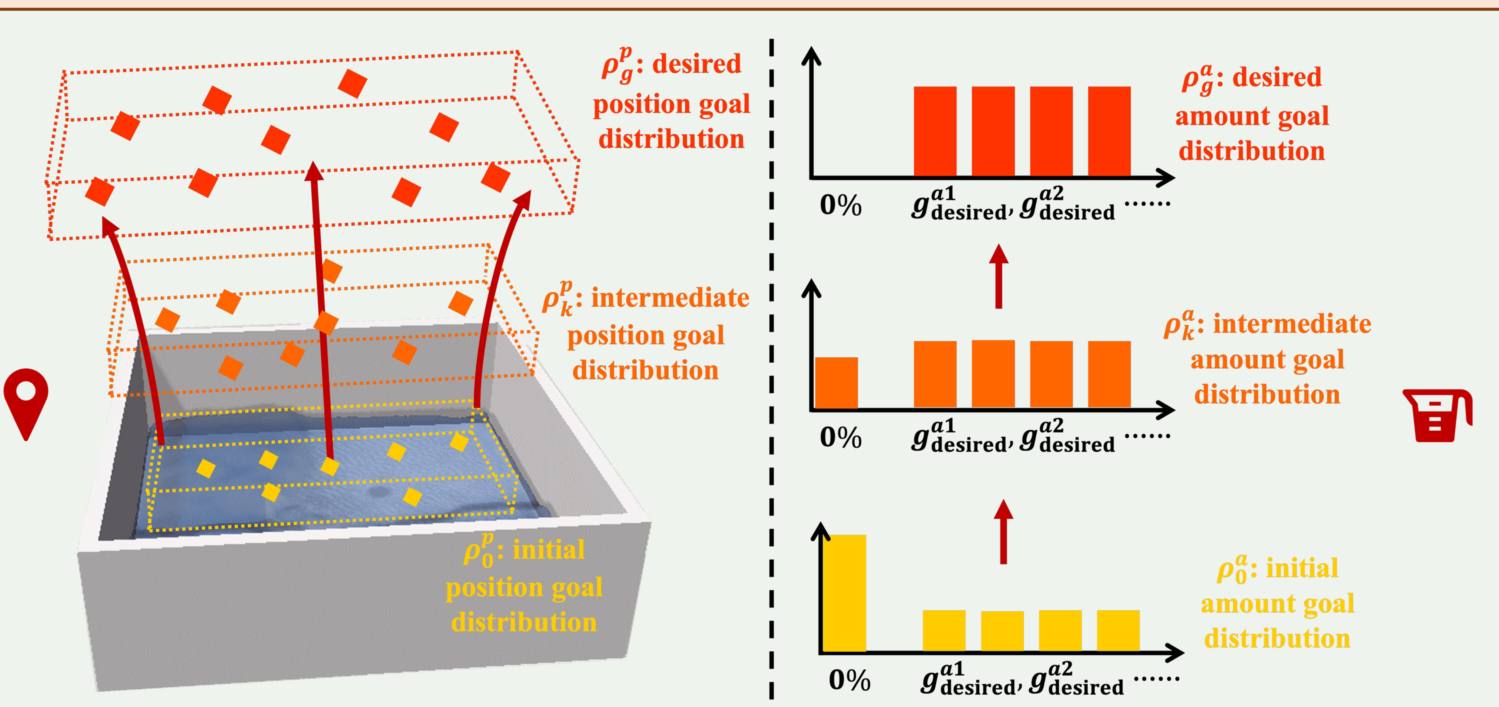
$$r(g^{t+1}_{\text{achieved}}, g_{\text{desired}})$$

$$\mathbb{1}\left(\left\|g^{p(t+1)}_{\text{achieved}} - g^p_{\text{desired}}\right\| \le \epsilon\right)\left(1 - \left\|g^{a(t+1)}_{\text{achieved}} - g^a_{\text{desired}}\right\|\right) - 1$$

Sparse Rewards
+ Reward shaping is hard for the position-reaching motions of scooping
+ Encourages exploration

Dense Rewards
+ Reward shaping is simple
+ Dense signals for training

Factorized Goal Sampling Adaptation

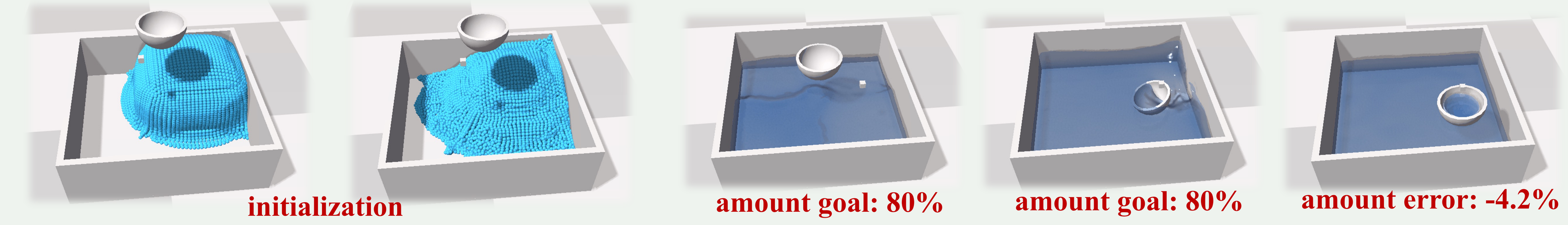


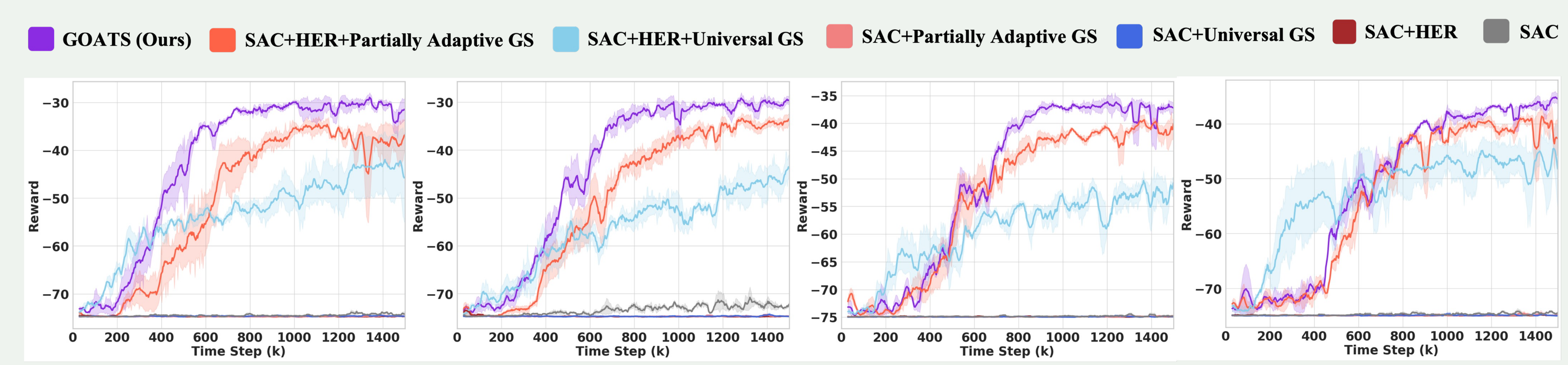
Position Goal Sampling Adaptation   Amount Goal Sampling Adaptation

Construct curriculum through interpolations between the desired and the initial goal distributions

Experiments & Results

Simulation



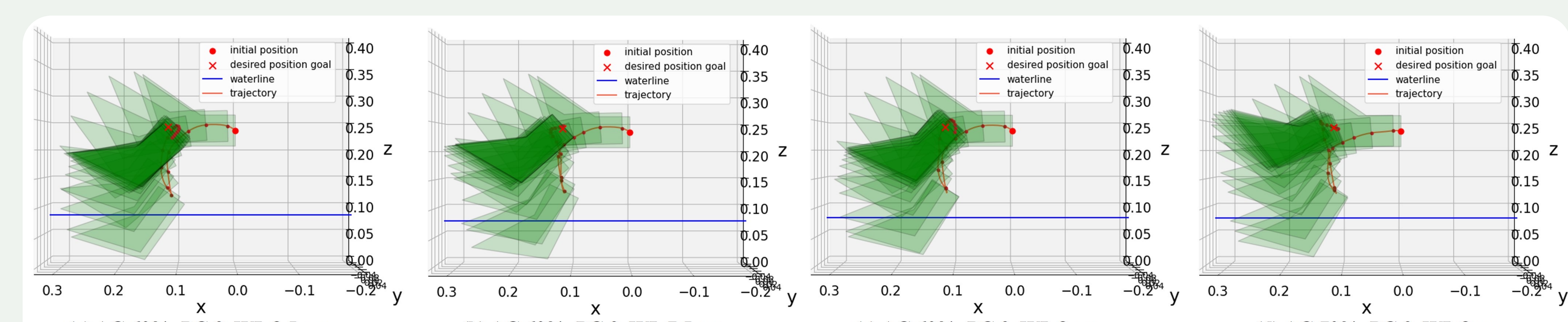


(a) Bowl Scooping w/ Single Amount Goal (b) Bowl Scooping w/ Multi. Amount Goals (c) Bucket Scooping w/ Single Amount Goal (d) Bucket Scooping w/ Multi. Amount Goals

Method	Bowl Scooping				Bucket Scooping			
	Single Amount Goal		Multi. Amount Goals		Single Amount Goal		Multi. Amount Goals	
	Reward ↑	Amount Error↓	Reward↑	Amount Error↓	Reward↑	Amount Error↓	Reward↑	Amount Error↓
SAC	-69.41 ± 0.78	$69.60\% \pm 0.33\%$	-61.21 ± 2.00	$71.02\% \pm 0.34\%$	-71.20 ± 1.12	$69.99\% \pm 0.01\%$	-69.47 ± 0.77	$71.00\% \pm 0.35\%$
SAC+HER	-72.72 ± 0.32	$67.28\% \pm 1.66\%$	-69.59 ± 2.32	$63.36\% \pm 5.91\%$	-73.40 ± 0.47	$52.35\% \pm 13.79\%$	-72.15 ± 1.05	$55.76\% \pm 0.35\%$
SAC+Universal GS	-71.7 ± 0.69	$69.51\% \pm 0.40\%$	-72.05 ± 0.41	$71.02\% \pm 0.34\%$	-72.96 ± 0.65	$70.00\% \pm 0.00\%$	-71.48 ± 1.28	$70.83\% \pm 0.43\%$
SAC+Partially Adaptive GS	-72.89 ± 0.59	$70.00\% \pm 0.00\%$	-71.87 ± 0.18	$67.51\% \pm 2.23\%$	-73.73 ± 0.24	$69.81\% \pm 0.15\%$	-73.14 ± 1.01	$70.98\% \pm 0.35\%$
SAC+HER+Universal GS	-36.45 ± 4.41	$26.18\% \pm 14.33\%$	-37.88 ± 2.48	$11.24\% \pm 2.51\%$	-42.48 ± 1.04	$12.76\% \pm 2.60\%$	-37.32 ± 1.24	$13.39\% \pm 0.69\%$
SAC+HER+Partially Adaptive GS	-28.80 ± 0.41	$8.54\% \pm 1.11\%$	-28.98 ± 0.43	$7.43\% \pm 1.41\%$	-35.22 ± 0.35	$\mathbf{9.61\% \pm 2.68\%}$	-33.12 ± 0.60	$14.16\% \pm 3.16\%$
GOATS (Ours)	$\mathbf{-25.67 \pm 0.32}$	$\mathbf{5.93\% \pm 1.20\%}$	$\mathbf{-25.77 \pm 0.60}$	$\mathbf{4.99\% \pm 0.37\%}$	$\mathbf{-33.36 \pm 0.69}$	$9.97\% \pm 2.09\%$	$\mathbf{-32.51 \pm 0.61}$	$\mathbf{7.45\% \pm 1.65\%}$

- Our method achieves 5.46% and 8.71% amount errors on bowl and bucket scooping in simulation, respectively, outperforming baselines across fours tasks

Real-Robot Scooping



(a) AG 60%, PG 2, WL 8.5 cm (b) AG 60%, PG 2, WL 7.5 cm (c) AG 60%, PG 2, WL 8 cm (d) AG 70%, PG 2, WL 8 cm

- Our method can adapt to diverse configurations (position goals, amount goals, initial water states), and generalize to unseen settings, e.g., initial bucket heights