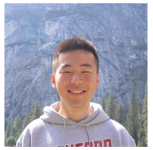
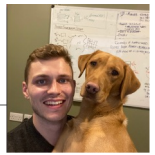


Synthetic Experience Replay

CONG LU*, PHILIP J. BALL*,

YEE WHYE TEH, JACK PARKER-HOLDER

NEURIPS 2023

Motivation

- RL agents are typically bottlenecked by useful data which they need to gather themselves
- Recent advances in diffusion generative modelling have shown that **generated synthetic data** is a powerful method to boost downstream performance, e.g. in image classification or robotics [1, 2]
- Proposed solution: **upsample** agent replay data using a diffusion model!

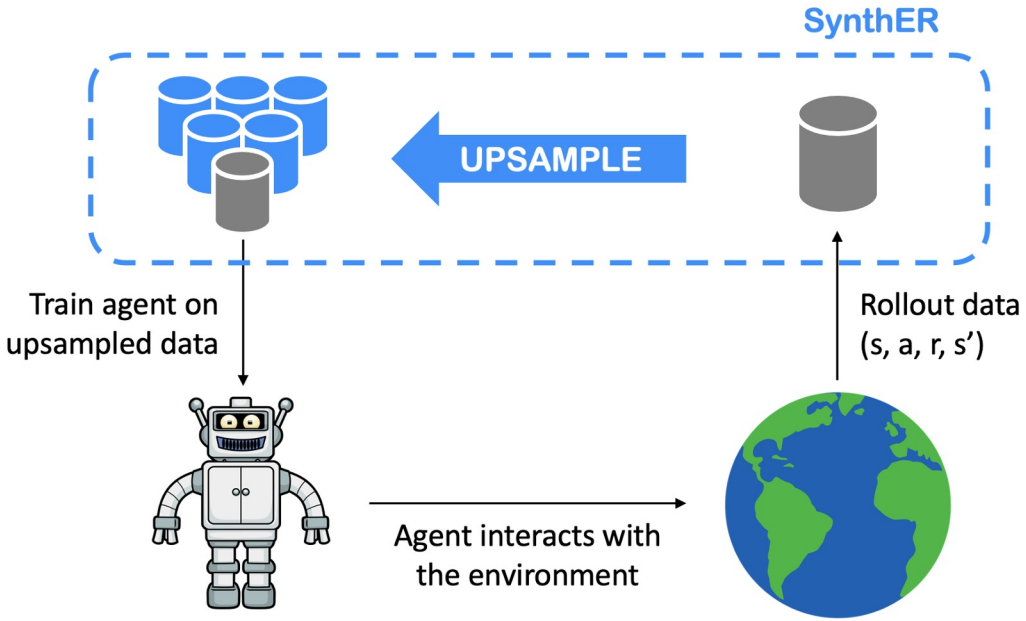
[1] Synthetic Data from Diffusion Models Improves ImageNet Classification. Shekoofeh Azizi, Simon Kornblith, Chitwan Saharia, Mohammad Norouzi, David J. Fleet. 2023
[2] GenAug: Retargeting behaviors to unseen situations via Generative Augmentation. Zoey Chen, Sho Kiami, Abhishek Gupta, Vikash Kumar. 2023

Background and Notation

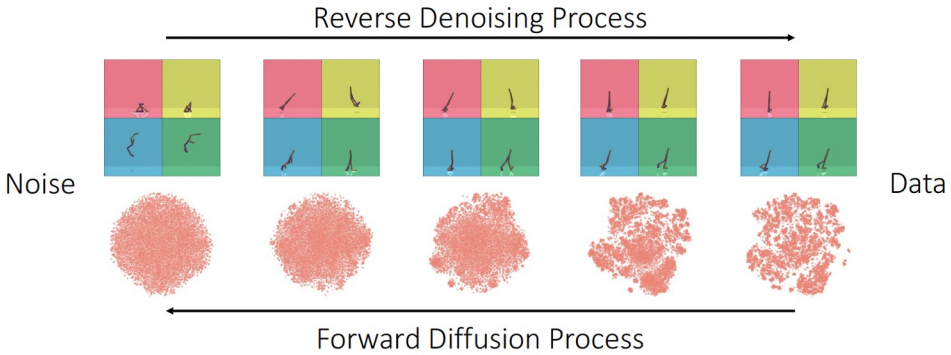
- Reinforcement Learning
  - We model the environment as a MDP $M = (S, A, P, R, \gamma)$
  - Agents train on $\mathcal{D}=\{(s_i, a_i, r_i, s_i')\}$ in order to learn a policy $\pi(a|s)$ to maximize expected return in the environment M

- Diffusion Generative Models
  - A class of models that learn to model a data distribution ***p(x)***
  - Learns to iteratively reverse a forward noising process and generate samples starting from pure noise

Synthetic Experience Replay

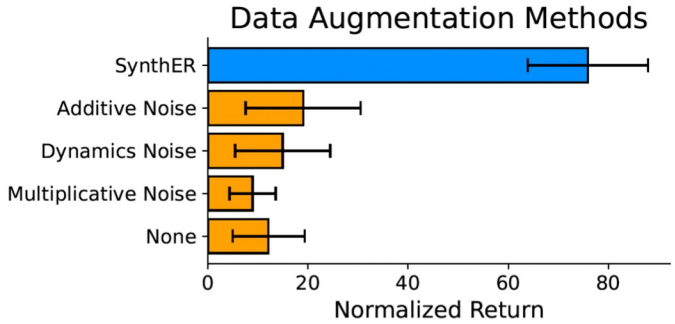


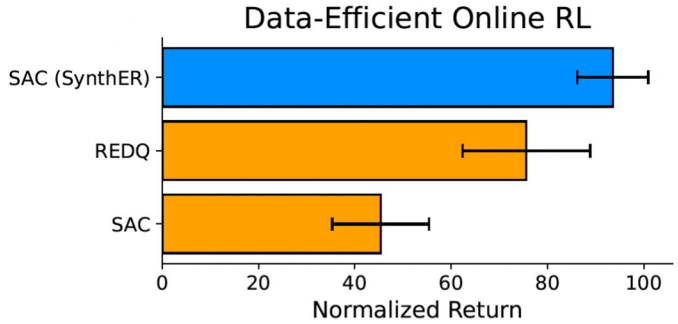
Visualization of the Data Generation Process



Summary of Proprioceptive Results

Upsampling data using SynthER greatly outperforms explicit data augmentation schemes for small offline datasets and data-efficient algorithms in online RL **without any algorithmic changes.**





(a) IQL (Kostrikov et al., 2022) on a reduced 15% subset of walker2d medium-replay (Fu et al., 2020).

(b) SAC (Haarnoja et al., 2018) on 6 DeepMind Control Suite and OpenAI Gym environments.

Comparison To Traditional Data Augmentation

SynthER generates samples that both **more faithful** to the true dynamics and **more diverse** than traditional data augmentations

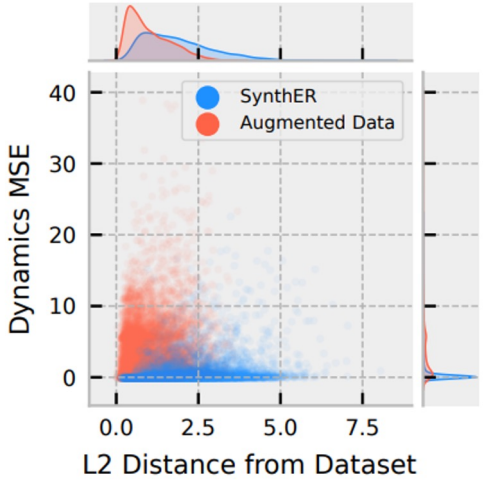


Figure 5: Comparing L2 distance from training data and dynamics accuracy under SYNTHER and augmentations.

Future Work

- Guided diffusion for targeted data generation
  - For example, high TD-error, low-data tasks for multitask settings
- Fine-tuning pretrained diffusion models
- Extensions to different formulations of experience replay
  - For example, n-step methods

Please do get in touch with any questions!

PAPER: HTTPS://OPENREVIEW.NET/FORUM?ID=6JNQ1AY1UF

ALL CODE AVAILABLE AT: HTTPS://GITHUB.COM/CONGLU1997/SYNTHER