Towards More Sample Efficiency in Reinforcement Learning with Data Augmentation

Yijiong Lin[†], Jiancong Huang[†], Matthieu Zimmer*, Yisheng Guan[†], Juan Rojas[†], Paul Weng*

[†] Guangdong University of Technology, *Shanghai Jiao Tong University

Corresponding author: paul.weng@sjtu.edu.cn

Abstract

Deep reinforcement learning (DRL) is a promising approach for adaptive robot control, but its current application to robotics is currently hindered by high sample requirements. We propose two novel data augmentation techniques for DRL in order to reuse more efficiently observed data. The first one called Kaleidoscope Experience Replay exploits reflectional symmetries, while the second called Goal-augmented Experience Replay takes advantage of lax goal definitions. Our preliminary experimental results show a large increase in learning speed.

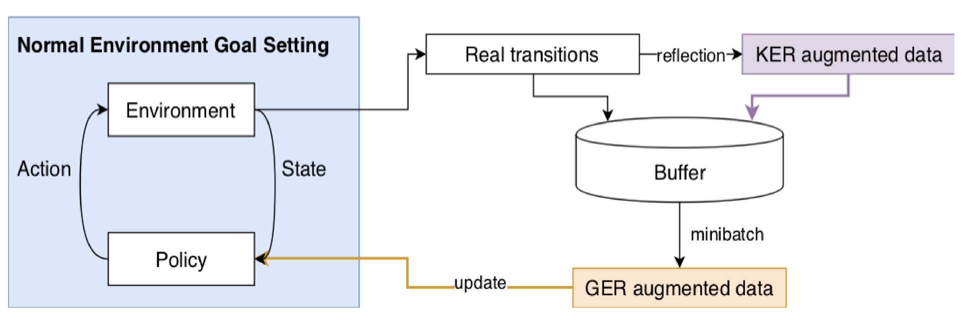

Fig. 1. ITER framework overview: real and symmetrically transformed transitions are stored in the replay buffer. Sampled minibatches are then augmented with GER before updating the policy.

Methodology

As a general approach to increase data efficiency in DRL, one can leverage the symmetries of valid trajectories for data augmentation. As an illustration, we propose **ITER (Invariant Transform Experience Replay)**, an architecture that combines our two proposed techniques (explained in details below):

- **Kaleidoscope experience replay (KER)** is based on decomposable symmetries of valid transitions set.
- **Goal-Augmented Experience Replay (GER)** is based on reward-preserving decomposable symmetries, but can be applied to all feasible trajectories in the same fashion as HER.

Kaleidoscope Experience Replay

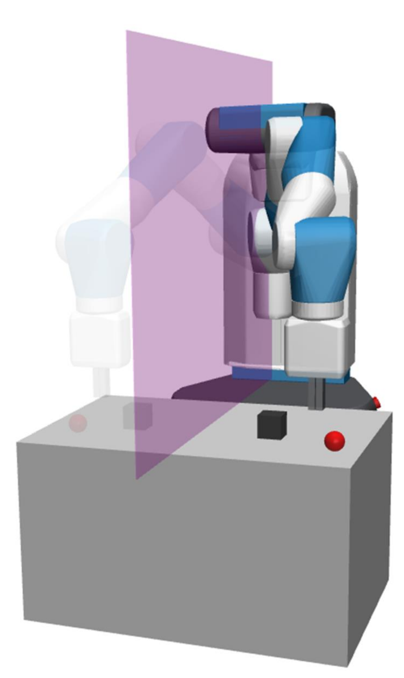

Fig. 2. Symmetric Plane of KER

KER uses reflectional symmetry (Though more general invariant transformations, e.g., rotation, translation, could also be used in place of reflectional symmetry.). Consider a 3D workspace with a bisecting plane *xoz* as shown in Fig. 2. If a feasible trajectory is generated in the workspace (red in Fig.3), natural symmetry would then yield a new feasible trajectory reflected on this plane. More generally, the *xoz* plane may be rotated by some angle θ along axis z and still define an invariant symmetry for the robotic task.

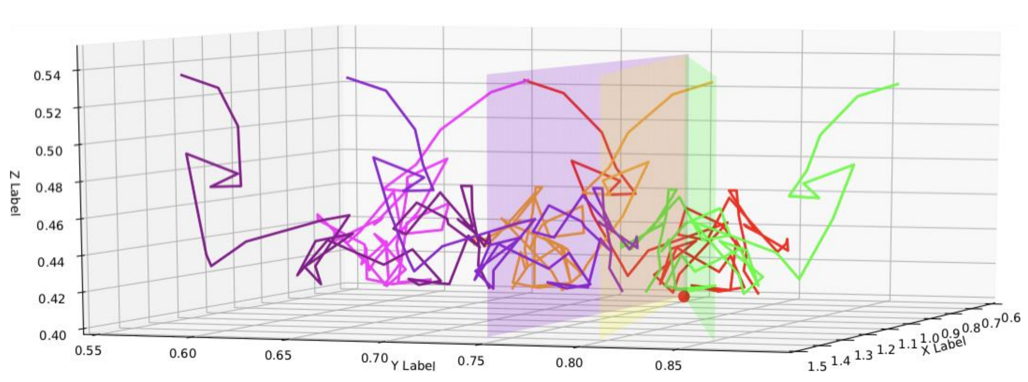

Fig. 3. The bottom figure illustrates this with xoz rotated twice to obtain the orange and green planes. The three symmetries are then applied to the red trajectory to obtain five new ones.

Goal-Augmented Experience Replay

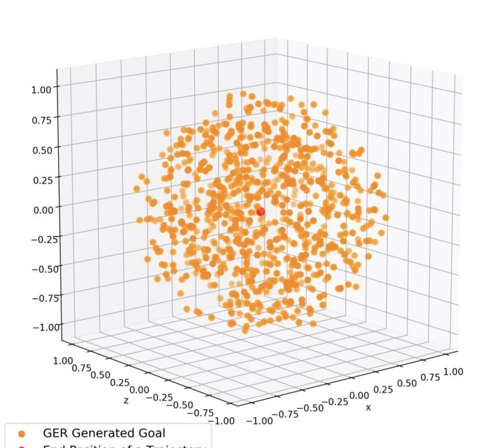

Fig. 4. Generated Goals in Orange Color

GER exploits the formulation of any reward function that defines a successful trajectory as one whose end position is within a small radial threshold (a ball) centered around the goal. When the robot obtains a valid trajectory, we therefore know that it can in fact be considered successful for any goal within a ball centered around its each position. Based on this observation, GER augments successful trajectories by replacing the original goal with a random goal sampled within that ball.

Results

We use a simulated 7-DOF Fetch Robotics arm trained with DDPG on the pushing, sliding, and pick-and-place tasks from OpenAI Gym, to perform our experimental evaluation and demonstrate the effectiveness of our propositions and answer those questions:

- How does ITER perform compared to HER on single and multi-goal tasks ?
- How much KER contributes to the performance of ITER?
- What is the contribution of GER to the performance of ITER?

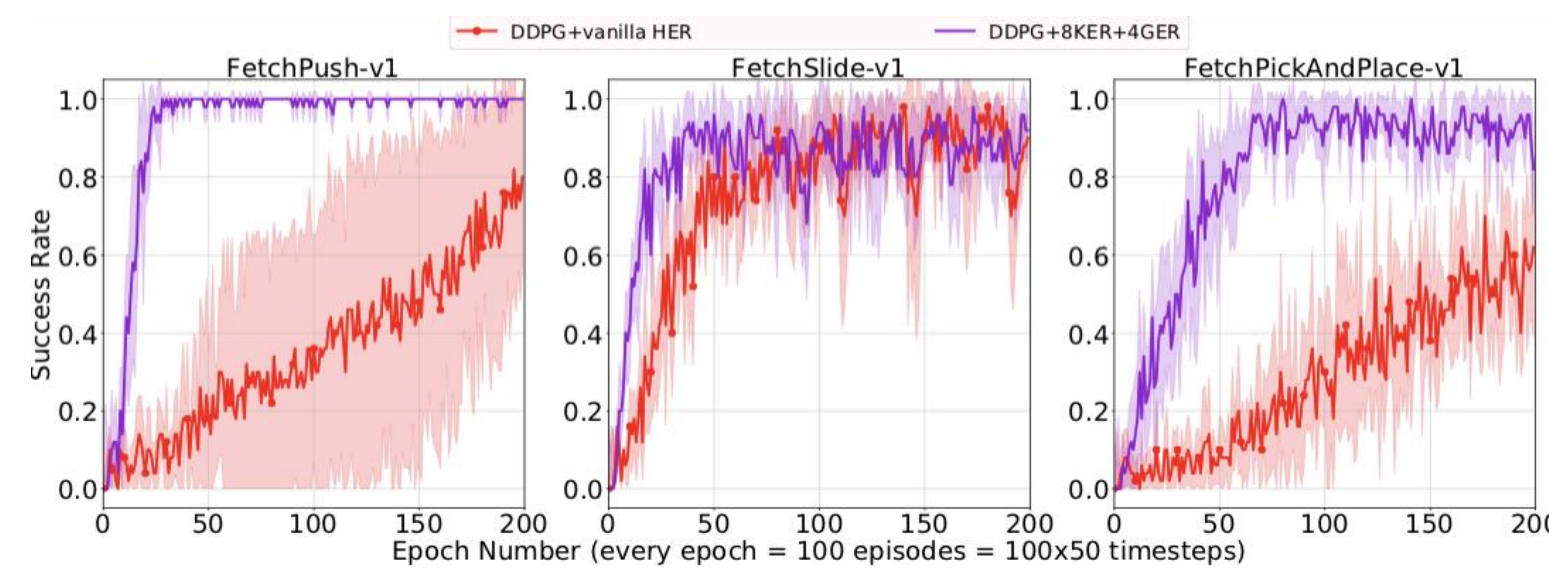

Fig. 4. Comparison of vanilla HER and ITER with *8KER* symmetries and *4GER* applications on multi-goal tasks

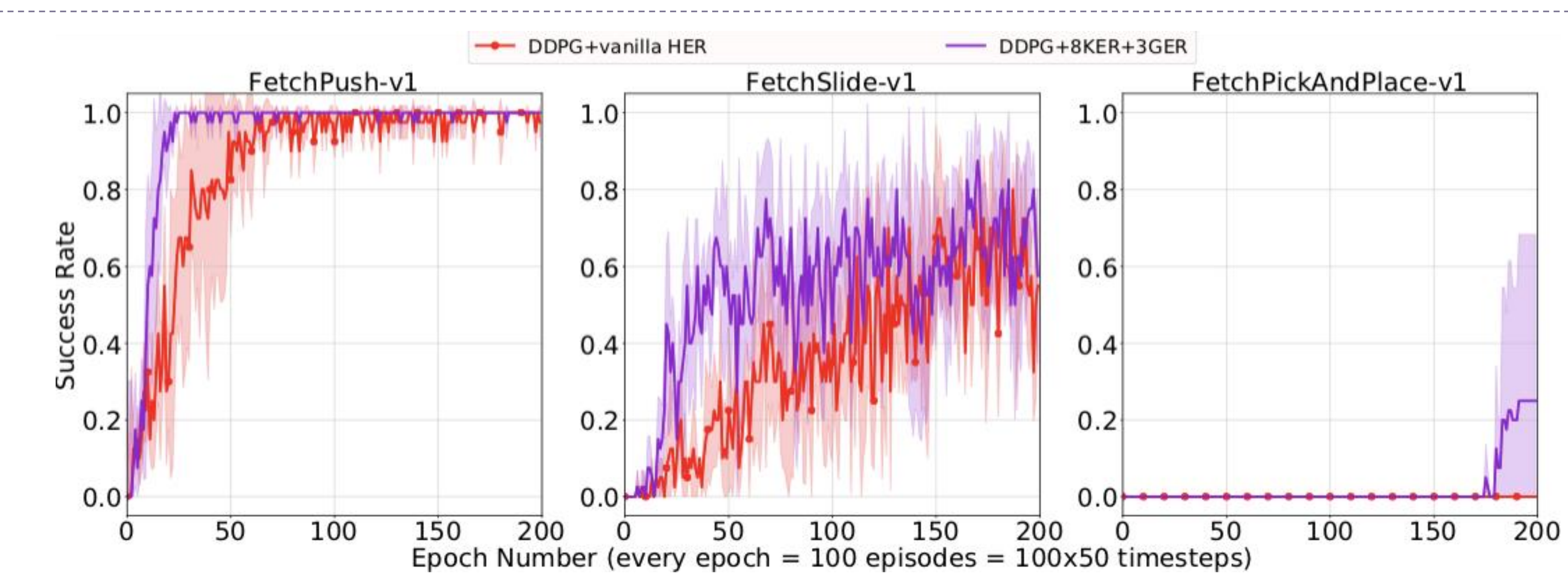

Fig. 5. Comparison of vanilla HER and ITER with *8KER* symmetries and *4GER* applications on single-goal tasks.

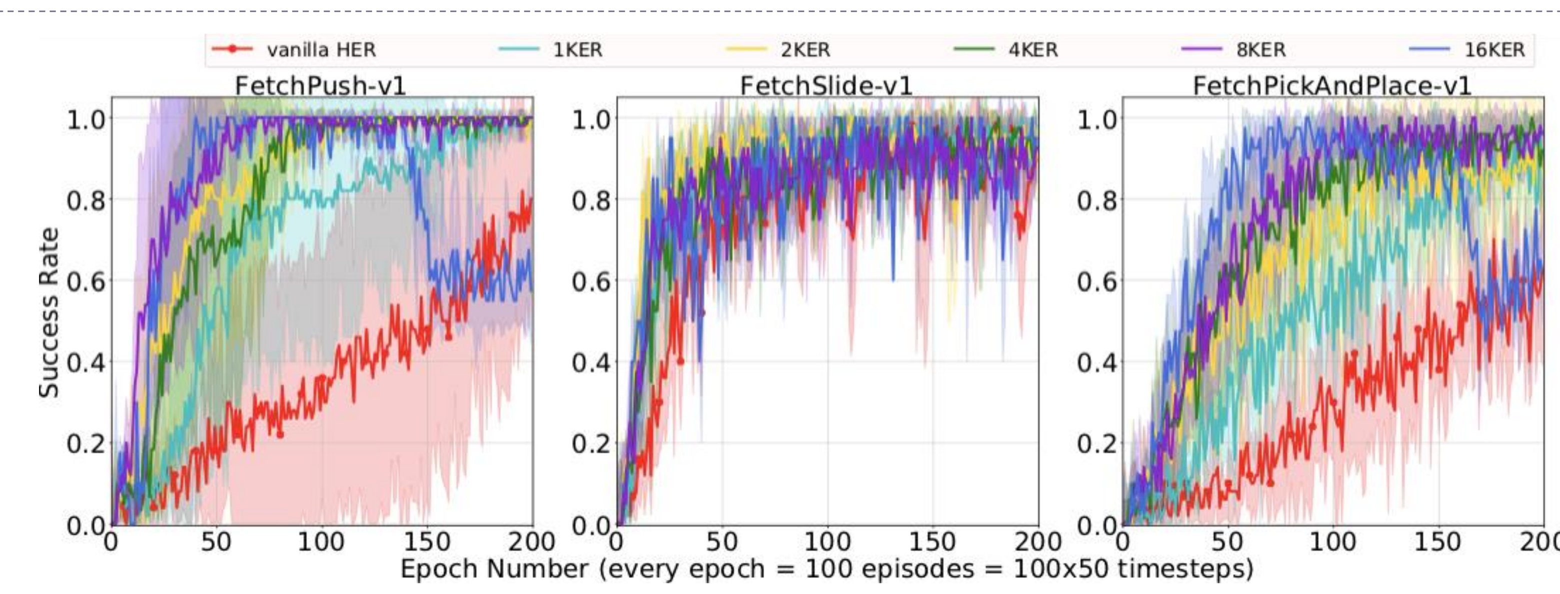

Fig. 6. Comparison of different KER parameters with a single GER on multi-goal tasks.

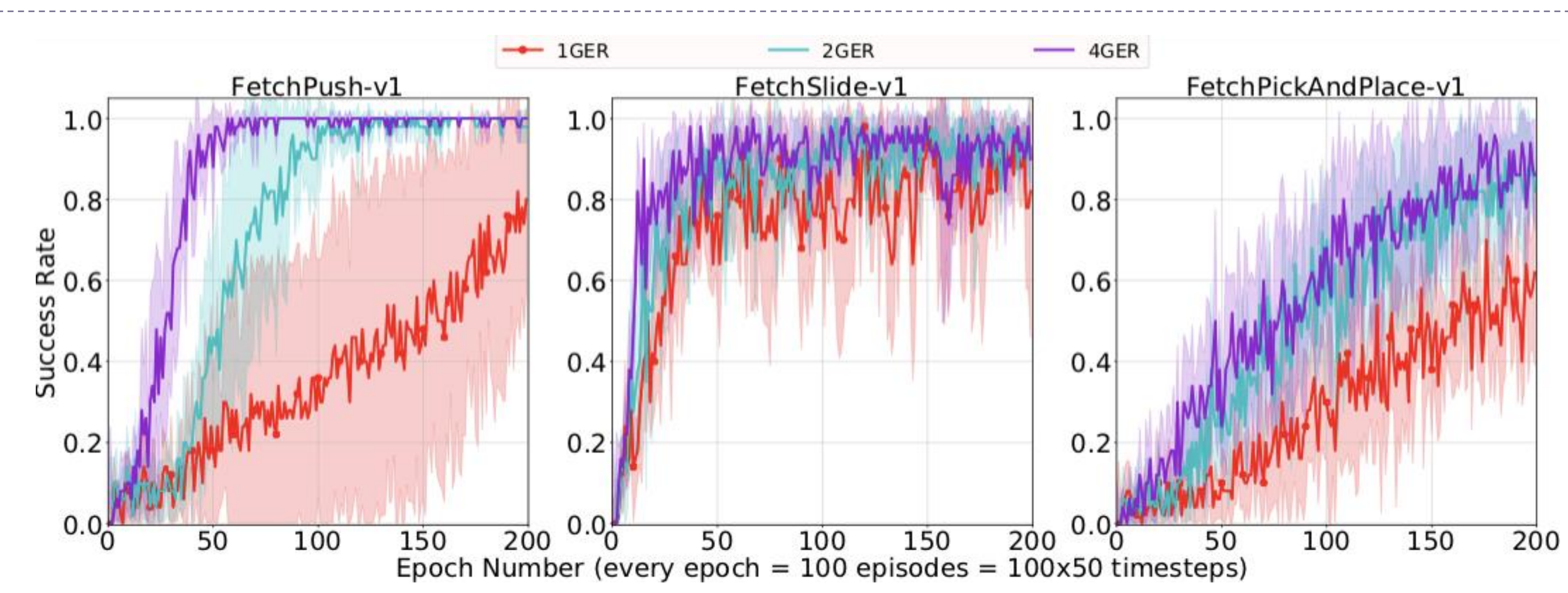

Fig. 7. Comparison of different GER parameters without KER on multi-goal tasks.

Conclusion

We proposed two novel data augmentation techniques KER and GER to amplify the efficiency of observed samples in a memory replay mechanism. KER exploited reflectional symmetry in the valid workspace (though in general it could be employed with other types of symmetries). GER, as an extension of HER, is specific to goal-oriented tasks where success is defined in terms of a thresholded distance. The combination of these techniques greatly accelerated learning as demonstrated in our experiments.