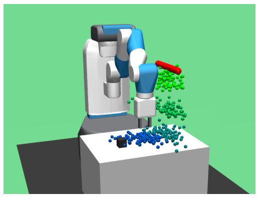
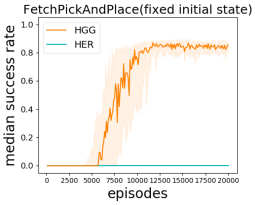
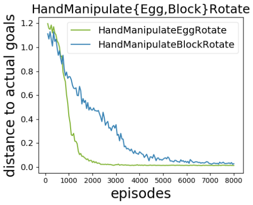
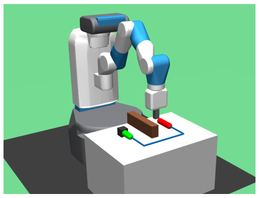
We thank the reviewers for their constructive comments and helpful feedback.

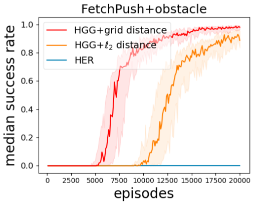






More illustrations on difficult tasks (R1,R3) To give better intuitive illustrations on our motivation, we provide an additional visualization of goal distribution generated by HGG on a complex manipulation task FetchPickAndPlace (left and middle figures). In the left figure, "blue to green" corresponds to the generated goals during training. HGG will guide the agent to understand the location of the object in the early stage, and move it to its nearby region. Then it will learn to move the object towards the easiest direction, i.e. pushing the object to the location underneath the actual goal, and finally pick it up. For those tasks which are hard to visualize, such as the HandManipultation tasks, we plotted the curves of distances between proposed exploratory goals and actually desired goals (right figure), all experiment followed the similar learning dynamics. We will add more intuitive demonstrations in the next version.





Regarding the difficulty of complex tasks (R1,R2) HGG does help to improve sample efficiency in difficult tasks, but its *final performance* is still limited by back-end RL algorithm (i.e. DDPG). For example, HGG provides a better warming-up in the experiment of FetchSlide, despite the final accuracy seems to be similar. For complex tasks which DDPG provides reasonable training, HGG also demonstrates large improvement over HER, as shown in the paper. We will add extra analysis and discussion about it.

About the metric we used in this paper, we agree that a crafted or a learned metric may be more suitable than ℓ_2 for difficult tasks. To show this, we created an environment with an obstacle (left figure). The object and the goal are uniformly generated in the green and the red segments respectively. The brown block is a static wall which cannot be moved. In addition to ℓ_2, we also construct a distance metric based on the graph distance of a mesh grid on the plane, as shown in the left figure, the blue line is a successful trajectory in such hand-craft distance measure. Intuitively, this distance should be better ℓ_2 due to the existence of the obstacle. Experimental results (in the right Figure) also suggest that such a crafted distance metric can provide better for goal generation and training, and significantly improve sample efficiency over ℓ_2 (right figure). It would be a future direction to investigate ways to obtain or learn a good metric.

Regarding the theorem (R3) We are sorry about the insufficiency of motivation and our inattention to the presentation formality of this theorem. Our theorem serves as a tool, as well as a justification, to get the objective function that we use, especially the term of Wasserstein distance between distributions of surrogate goals and underlying goals. We also believe that the lower bound we obtained in this theorem, although we agree that it is quite simple to prove, is not absolutely vacuous. For example, when the value function is linear, the lower bound becomes tight and cannot be further improved. We also agree with R3 that there can be also a clear and intuitive motivation of our method. We'll rewrite this part by improving the presentation, make its usage easier to understand, increase the rigorousness of formal mathematical statements, and provide further intuitive explanation of our method in the next version.

Regarding the Lipschitz condition (R3) Thank you for the suggestion about the validness of this condition. Similar generalizability assumptions were also required by many previous works on goal curriculum (e.g. two related works mentioned at line 118), they assumed the learned policy could generalize to close-by regions. In our method, Lipschitz continuity is assumed to simplify the introduction of algorithmic framework and objective. It is a formalization of the intuition that, the performance of a policy is similar for tasks that are close to each other. In principle, we require the smoothness of validated policy and environment dynamics to derive an approximated version of Lipschitz continuity. These two preconditions are commonly satisfied in many continuous robotics tasks with parameterized policy classes. We will provide empirical ways to verify this condition on several robotics tasks in the OpenAI gym benchmark.

As suggested by R3, we will include more intuitive discussion of this condition and provide further mathematical and practical guidelines of it in the next version.

Response to other helpful suggestions (R3) Thank you for pointing out the related references and suggesting comparison to simpler optimization schemes. We will add the suggested related works and discussions in the next version. We will investigate a simpler implementation which chooses the argmax goal instance in a random subset of trajectories, as suggested. Some preliminary results indicate that it achieves reasonable performance on simple manipulation tasks. We will add comprehensive results on this part in the next version.