Additional Experiments

Due to page limitations, some redundant experiments were removed from the original manuscript.

1 Experiments of training policies on different gravitational conditions

In this section, we explore the policy performance trained under conditions that include different gravity effects. We compared three cases: our original policy trained only on the horizontal plane (denoted as OnlyZ) and a policy trained at workspace tilt angles including 0 and 30 degrees (denoted as Zand30) and 0 and 60 degrees (denoted as Zand60). The training objects for all cases are the same, including printer books and winter suit fabrics. Unseen objects for testing are those cases in Fig. 3D, excluding the training objects. During training, the two robotic arms are each responsible for a tilt angle. After every five attempts, the researcher exchanges objects.

Fig. 1A and 1B show the difference in performance when training includes different gravity effects, compared to OnlyZ. The results show that Zand60 significantly improves the average performance in the test under the 30 and 60-degree conditions (Fig. 1B). This proves that including different gravity effects in the training can enhance the overall performance. Meanwhile, Zand30 also achieved performance improvement under corresponding conditions, although the improvement was smaller (Fig. 1A). This reflects that objects on the physical characteristics of a workspace with a 30-degree tilt angle are closer to those of a 0-degree workspace.

Fig. 1C shows the training curves of these three cases. Including different workspace tilt angles increases the diversity of situations encountered during the training phase. This variation demands more exploration from the RL agent, thus extending the training time to reach convergence.

These results also demonstrate the potential that, when hardware conditions permit, our approach can be extended to more parallel environments. This can not only significantly accelerate the learning process but also cover a wider range of training objects and gravitational effects, further improving the robot's performance.

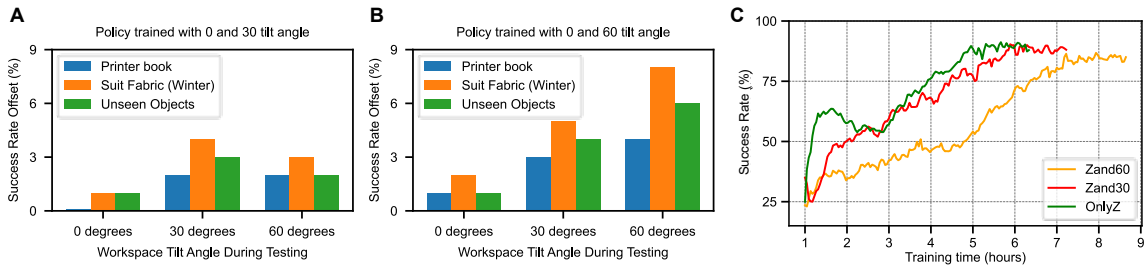

Figure 1: **Experiments of including different gravitational effects in policy training.** (**A** and **B**) Success rate offsets, calculated as the difference in performance of test objects between policies trained on various workspace tilt angles and the success rate of our original policy, which was trained exclusively on a horizontal workspace. (**C**) Training curves for policy learned with different tilt angles of the workspace. The time when humans exchange objects is not included. Training stops when performance no longer increases.