This paper presents an efficient deep reinforcement learning (DRL) framework for online 3D bin packing (3D-BPP). The 3D-BPP is an NP-hard problem significant in logistics, warehousing, and transportation, involving the optimal arrangement of objects inside a bin. Traditional heuristic algorithms often fail to address dynamic and physical constraints in real-time scenarios.

We introduce a novel DRL framework that integrates a reliable physics heuristic algorithm and object rearrangement and stable placement. Our experiment show that the proposed framework achieves higher space utilization rates effectively minimizing the amount of wasted space with fewer training epochs.

The physics heuristic algorithm aims to ensure the stability of packed objects. Three variants of this algorithm are introduced: convexHull-1, convexHull-k, and convexHull--$¥alpha$. The convexHull-1 algorithm checks stability by evaluating whether the center of a sliding window, matching the dimensions of the incoming object, lies within the convex hull formed by the highest points within that window. However, it was found to be insufficient in complex multi-stack scenarios. The convexHull-k algorithm extends this concept by considering the intersection of convex hulls across multiple layers, but it still underestimates stability in certain cases. To address these limitations, the convexHull-$¥alpha$ algorithm is proposed, which ensures that the supporting force originates vertically from the ground, using an empty map to track wasted voxels along the z-axis. This algorithm significantly improves the accuracy of stability checks, as demonstrated through simulations.

To further enhance packing efficiency, the framework incorporates an object rearrangement process. This allows the robot manipulator to change the orientation of incoming objects, an action that directly improves space utilization without requiring additional time costs. The DRL model is formulated as a Markov Decision Process (MDP), with two separate agents, each responsible for predicting the optimal orientation and position for the packing object. The reward function is designed to balance the increase in space utilization and the minimization of wasted space.

The proposed DRL framework is trained and evaluated through both simulation experiments. In the simulations, two vision sensors capture visual information of the bin and incoming objects, enabling the model to predict orientations, stable positions, and placement scores. The results show that the proposed framework achieves higher space utilization rates compared to baseline methods, while requiring fewer training epochs.

The real-robot experiments involve connecting a uFactory850 robotic arm to the PC, performing hand-eye calibration using ArUco QR codes, and recognizing and suctioning target objects. The target objects' dimensions and poses are accurately estimated using a semantic segmentation model and OpenCV's pose estimation functions. The robotic arm then adjusts the object's orientation and places it in the bin, as predicted by the DRL model.

The experimental results shows that: 1.The physics heuristic algorithm, especially convexHull-$alpha$, significantly outperforms previous methods in terms of stability check accuracy.2. the proposed DRL framework achieves higher space utilization rates effectively minimizing the amount of wasted space with fewer training epochs.