2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Learning Push-Grasp Synergy for Occluded Objects in Cluttered Environments

Ziang Li[1], Haorui Wu[1], Zhiqi Chen[1], Haozhe Zhang[1], Yuzhe Huang[2], and Changshui Zhang[1†]

***Abstract*— Successfully executing grasping tasks within highly cluttered spaces is still a significant hurdle in robotics, especially in scenarios involving severe target occlusion. To tackle this, we present a novel self-supervised framework driven by deep reinforcement learning that enables robots to acquire push–grasp synergy for reliable manipulation under occlusions. The core contribution of this research is the target switching mechanism that dynamically selects alternative targets when the goal object is severely occluded. Furthermore, we utilize a strategy for selecting actions based on object masks to reduce the action space, thereby improving efficiency and minimizing ineffective operations. Comprehensive evaluations across both simulated and physical environments confirm that our method achieves robust grasping performance under severe or complete occlusions. Notably, the learned policy is readily transferable to physical environments and generalizes effectively to previously unseen objects. To guarantee experimental reproducibility and encourage further studies, our source code is available at https://github.com/lzalza/Learning-Push-Grasp-Synergy-for-Occluded-Objects-in-Cluttered-Environments and demonstration videos can be viewed at https://youtu.be/hDm-vIlaymw.**

I. INTRODUCTION

As a cornerstone of robotic manipulation, object grasping provides the essential foundation for executing diverse and complex tasks [1], including the utilization of robotic tools [2], [3]. In many real-world applications, robots are often required to operate in highly cluttered environments, where target objects may be densely surrounded or severely occluded [4]–[6]. Inspired by human strategies, robots have increasingly adopted push–grasp synergy as a solution in such scenarios [7]–[10]. By pushing away surrounding clutter, the robot can isolate the target object and thereby facilitate grasp execution. To accomplish this effectively, robots must infer spatial relationships from visual observations and formulate grasping strategies that are adaptive to different cluttered and occluded scenarios, ultimately aiming to maximize grasp success with minimal pushing actions.

Research efforts have focused on developing learning strategies to enable coordinated pushing and grasping [11]–[13]. To concurrently acquire pushing and grasping strategies, Zeng et al. [11] proposed an architecture utilizing two parallel neural networks. A multi-level reinforcement learning structure was formulated by Xu et al. [12] with the specific aim of improving the acquisition of push–grasp synergies and to facilitate performance in goal-oriented tasks. More recently, Wang et al. [13] enhanced push–grasp policies with object masks, improving the efficiency of both pushing and grasping. Nevertheless, most existing approaches remain limited in handling occlusion-dominated scenarios, where pushing actions are often inefficient and may even impair grasp performance.

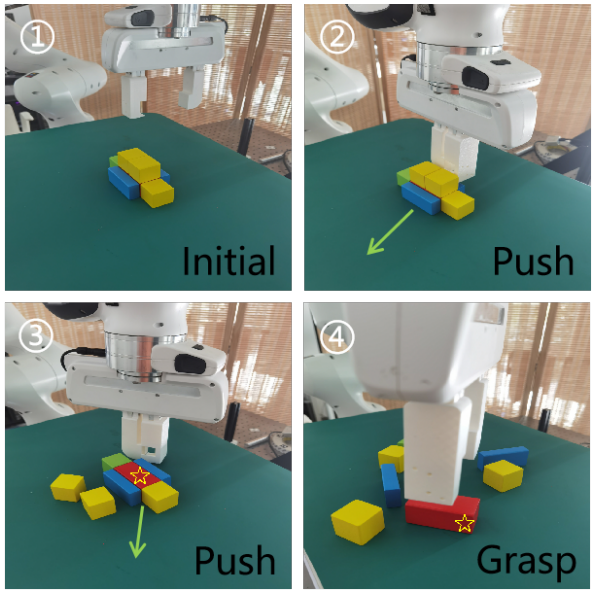


Fig. 1: Grasping an occluded target in clutter. The red target object is initially blocked by surrounding items. Our method adaptively selects the yellow object above as an alternative for pushing to resolve the occlusion. Once the target is exposed, it can be successfully retrieved by leveraging the coordinated execution of both pushing and grasping primitives.

Our current research focuses on configurations in which the desired item begins in a state of occlusion caused by surrounding clutter. We propose a self-supervised method based on deep reinforcement learning to jointly learn pushing and grasping actions towards effective collaboration. We optimize the existing pipeline by incorporating a human intuition–inspired target switching mechanism to better handle occlusions. As shown in Figure 1, when the target object is blocked, our method adaptively selects alternative objects for coordinated pushing and grasping, enabling the robot to progressively clear the scene until the target becomes visible. Moreover, we introduce an action selection strategy guided by object masks to heavily constrain the searchable state-action space and enhance action efficiency. In summary, our main contributions are as follows:

[1] Beijing National Research Center for Information Science and Technology (BNRist) Department of Automation, Tsinghua University Beijing, P.R.China
[2] Beihang University, Beijing, P.R.China
† Corresponding author.

- We introduce a self-supervised learning framework based on DQN networks that enables robots to simultaneously acquire pushing and grasping policies, thereby facilitating effective push–grasp synergy under occlusion scenarios.
- We incorporate a human intuition–guided target switching mechanism together with a mask-based action selection strategy to address occlusions and reduce the exploration space, ultimately enhancing grasping efficiency.
- We conduct extensive simulation experiments, including cases with severe occlusions, to rigorously evaluate the performance and reliability of our proposed approach. Moreover, we implement the proposed framework on a physical robotic system, demonstrating its practicality and direct transferability to real-world environments without additional retraining while maintaining consistent performance.

II. RELATED WORK

A. Goal-Oriented Grasping

Recent advances in robotic grasping have drawn significant attention, leading to substantial progress in both methodology and practical applications. Current methodologies generally fall into two primary groups: analytical model-driven techniques [14], [15] and data-driven learning paradigms [16], [17]. Furthermore, grasping tasks are typically distinguished by their objectives as either goal-agnostic grasping [18], [19] or goal-oriented grasping [20], [21].

Compared with goal-agnostic grasping, goal-oriented grasping has been less extensively studied but constitutes the central focus of this work. Chen et al. [20] proposed a network that directly predicts command-satisfying grasps from natural language instructions and images. Lu et al. [21] developed an interactive grasping policy that integrates visual grounding with grasp pose detection, enabling robots to manipulate objects specified through natural language directives. Li et al. [22] introduced a binocular stereo vision–based approach that explicitly infers occlusion relationships to segment and localize targets and estimate multi-target grasp poses.

Despite these advances, existing methods remain limited in severely cluttered or highly occluded environments, particularly when no feasible one-step grasping solution exists. This limitation highlights the necessity of exploring push–grasp synergy as a means to effectively uncover and access target objects under such challenging conditions.

B. Grasping under Occlusion

Robotic grasping under occlusion continues to pose a major challenge for real-world manipulation, as objects in cluttered scenes are frequently partially or completely obscured from direct observation. To address this issue, recent studies have investigated diverse strategies. Cui et al. [23] presented a Gaussian process–based probabilistic active learning framework that models uncertainties in system dynamics and observation functions, enabling effective search for occluded objects. Hoang et al. [24] designed a voting-based approach that leverages contextual information to achieve collision-free grasping in partially occluded settings. Yu et al. [25] introduced an image inpainting–based method to reconstruct occluded objects and subsequently generate grasp poses for robotic manipulation.

More recently, increasing attention has been directed toward fully occluded environments, in which target objects are entirely hidden from perception. In such scenarios, robots must actively interact with the environment to uncover the target, rather than relying on a single grasping attempt. Danielczuk et al. [26] proposed a framework where the robot iteratively performs pushing, suction, and grasping actions, guided by RGB-D sensing, to retrieve targets from cluttered bins. Yang et al. [7] developed a unified framework that integrates two complementary strategies: an exploration policy for fully occluded targets and a push–grasp policy for partially visible ones. Zuo et al. [4] incorporated graph-based reasoning to capture feature relationships across regions, thereby enabling more efficient exploration and target retrieval.

C. Push-Grasp Synergy

An increasing number of studies have investigated the synergy between pushing and grasping as a means to achieve efficient manipulation in cluttered environments. The foundation for this direction was laid by Zeng et al. [11] through their Visual Pushing for Grasping system. This model-free deep reinforcement learning setup simultaneously acquires strategies for both action types. Furthermore, Xu et al. [12] formulated a structured reinforcement learning method that incorporates both goal relabeling and sequential alternating training phases, which enhances the learning of push–grasp coordination by improving sample efficiency. Liu et al. [27] introduced GE-Grasp, which integrates multiple action primitives within a generator–evaluator architecture to achieve collision-free grasping. Ren et al. [28] combined a bi-functional network with hierarchical reinforcement learning to simultaneously address goal-agnostic and goal-oriented grasping tasks. Li et al. [6] developed MPGNet, which incorporates a moving action to strengthen push–grasp synergy, thereby improving target localization and manipulation. Most recently, Wang et al. [10] proposed a dual-reinforcement learning framework for joint pushing and grasping in dual-arm robotic systems.

Whereas densely packed workspaces have been the primary focus of most prior investigations, our study extends to occlusion scenarios. We find that it remains challenging to learn push–grasp synergy under severe occlusions [12], [13], since pushing and grasping actions are often ineffective when the target is heavily blocked. Inspired by human intuition, we introduce an object selection mechanism to address occlusions and to collaborate with existing push–grasp synergy.

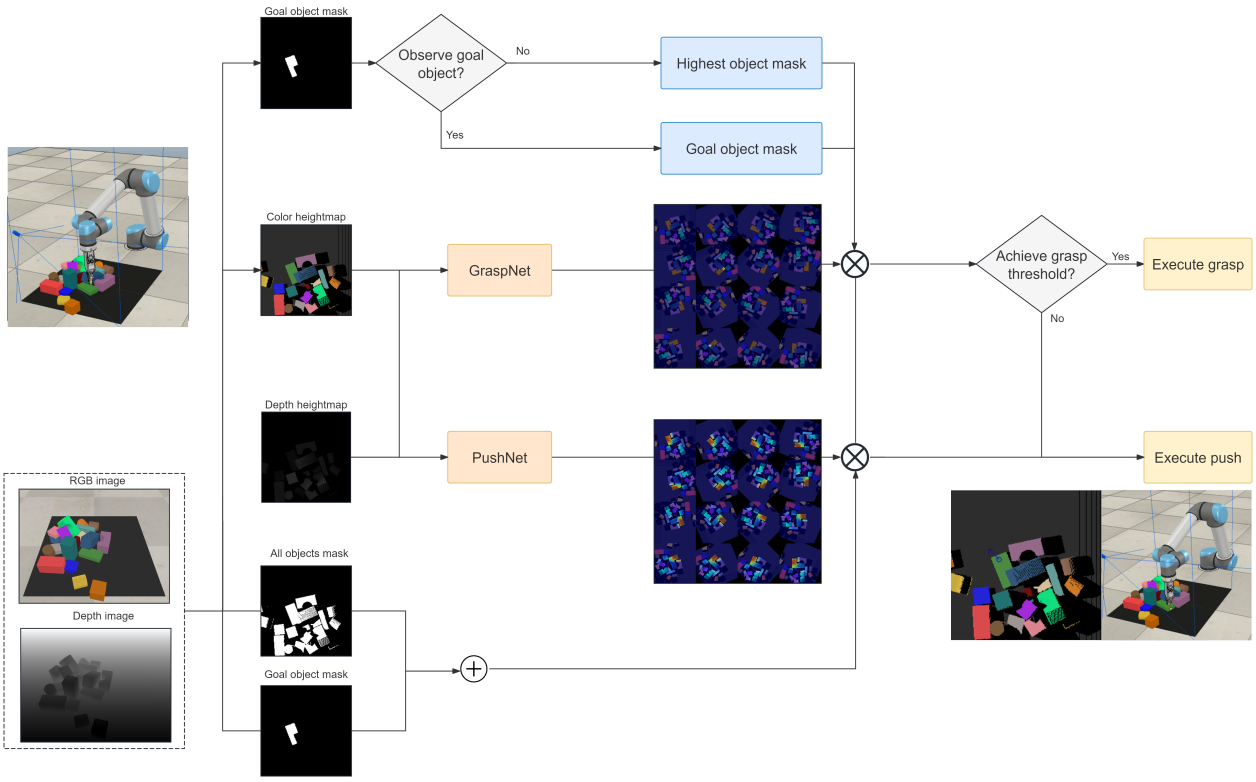


Fig. 2: A high-level illustration depicting the overall system. The RGB-D image of the workspace is orthographically projected from a top-down view to generate color and depth heightmaps, which are rotated 16 times and processed by the GraspNet ϕ_g and the PushNet ϕ_p. The current target object is determined through the target switching mechanism based on the occlusion of the predefined goal. The network outputs are then weighted by the object masks, and the action with the highest Q-value is selected for execution as either a grasp or a push.

III. METHOD

A. Problem Formulation

Similar to previous work [11], [12], the coordinated pushing and grasping problem is modeled as a goal-conditioned Markov decision process (MDP). During each discrete time step t, the robot receives a state observation s_t alongside the target representation g_t, subsequently executing an action a_t derived from the policy distribution $\pi(s_t \mid g_t)$. Following this execution, the system evolves into the subsequent state s_{t+1} and yields an instantaneous reward signal $R_{a_t}(s_t, s_{t+1}, a_t, g_t)$.

The primary aim within this reinforcement learning paradigm is to acquire an optimal control policy π^* capable of maximizing the anticipated long-term return, which is formulated as the γ-discounted cumulative reward.

$$R_t = \sum_{i=t}^{\infty} \gamma^{i-t} R_{a_i}(s_i, s_{i+1}, a_i, g_i). \tag{1}$$

In this work, we employ a Deep Q-Network (DQN) to learn a greedy deterministic policy $\pi(s_t \mid g_t)$, where actions are selected by maximizing the action-value function $Q_\pi(s_t, a_t, g_t)$.

B. System Architecture

Figure 2 depicts the complete architecture of our proposed system. Visual data (RGB-D) from the manipulation area is acquired via a statically mounted camera. This visual input is then orthographically projected downward aligned with gravity, yielding both a color heightmap c_t and a spatial depth heightmap d_t. The combination of these two maps constitutes the state representation $s_t = (c_t, d_t)$ at any given moment. The target object is represented by a mask heightmap m_t. We specifically choose this top-down orthographic representation over raw perspective RGB-D inputs to decouple the visual features from the specific camera viewpoint. This design choice significantly enhances the transferability of our framework. Although more advanced segmentation methods exist [29], we adopt a simple RGB-color segmentation approach, since the focus here is on robotic grasping rather than perception.

Primitive actions a_t are represented by the tuple (x, y, z, w, ϕ), where (x, y, z) denote the gripper's center position, w specifies the action orientation, and $\phi \in \{\text{grasp}, \text{push}\}$ indicates the action type. For pushing, the robot evaluates 16 orientations evenly spaced at 22.5° to cover the full 360°. Each push begins 5 cm behind the target position and proceeds forward for 10 cm. For grasping, the

robot similarly considers 16 orientations and executes top-down parallel-jaw grasps at the designated position, lowering the gripper 2 cm below the position to ensure stability.

Our method is built on a DQN framework with two networks, ϕ_g (GraspNet) and ϕ_p (PushNet), that estimate Q-values. Each fully convolutional network (FCN) takes the state heightmaps s_t as input and outputs a Q-value map of the same resolution, where each pixel corresponds to the expected reward of executing the associated action. To accommodate different orientations, the input heightmaps are rotated 16 times, with only rightward horizontal actions considered in each rotation. This rotation strategy explicitly hardcodes rotation equivariance into the network architecture, which enhances the sample efficiency of the reinforcement learning process. Each FCN thus outputs 16 Q-value maps corresponding to the orientations. If the maximum Q-value from the GraspNet ϕ_g exceeds a predefined threshold and the surrounding clutter is below a certain proportion, the robot executes a grasp; otherwise, it executes a push corresponding to the maximum Q-value predicted by the PushNet ϕ_p. Conceptually, this mechanism is analogous to a GAN [30], with the GraspNet acting as a discriminator and the PushNet as a generator: the generator modifies the scene to increase the discriminator's Q-value for grasping.

The GraspNet ϕ_g and PushNet ϕ_p share an identical architecture. Each network utilizes a dual-stream architecture comprising 121-layer DenseNets [31], initialized with ImageNet weights [32] for feature extraction from the respective depth and color inputs. These resulting feature representations are fused via concatenation before being fed into a fully convolutional network (FCN) [33]. This FCN block is constructed using a pair of 1×1 convolutions, followed by ReLU non-linearities [34] and standard batch normalization techniques [35]. Bilinear upsampling is applied to restore the resolution to match the input heightmaps.

Due to the large action space, the robot frequently explores invalid actions in the early stages of training, significantly reducing learning efficiency. Even after training, it may still attempt actions at inappropriate locations, leading to failed grasps and potential safety concerns [12]. Following Wang et al. [13], we employ object masks to restrict exploration to meaningful regions, thereby improving sample efficiency by reducing wasted interactions.

Specifically, we introduce a target switching mechanism to address occlusions by dynamically selecting the target object in the current state. The predefined goal object is identified by matching its prior visual features (e.g., color) within the workspace. If the predefined goal object is severely or even completely occluded, the robot naturally fails to detect these features and generates an empty or near-empty mask. Once the detected mask area falls below a specific threshold, the system automatically switches the target to the highest object in the workspace to initiate obstacle removal. This reactive strategy allows the robot to progressively clear the scene without requiring complex pose estimation of invisible objects, until the predefined goal is sufficiently exposed.

For grasping, the current target object mask is used to constrain the action space during action selection, while the networks still take the global workspace as input. For pushing, we compute a weighted combination of the current target object mask and all object masks. Compared to Wang et al. [13], which only uses the global mask of all objects to constrain the action space, our framework explicitly increases the likelihood of pushing the current target object rather than arbitrary surrounding objects. This targeted approach improves sample efficiency and facilitates directed occlusion removal.

C. Training Strategy

Following prior works [6], [12], [13], we adopt a hierarchical reinforcement learning framework to learn coordinated push–grasp synergy. The training procedure is divided into three stages: goal-conditioned grasping training, goal-conditioned pushing training, and goal-conditioned alternating training.

1) Goal-conditioned Grasping Training: In the first stage, the goal-conditioned GraspNet ϕ_g is optimized to accurately evaluate states based on its Q-values Q_g. To improve sampling efficiency, we employ a two-phase training scheme. In the goal-agnostic phase, each successful grasp is relabeled with the grasped object as the goal. Once the model develops basic grasping capability, training transitions to the goal-conditioned phase.

The grasp reward function is defined as:

$$R_g = \begin{cases} 1, & \text{if grasp success,} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Here, "grasp success" refers to grasping any object during the goal-agnostic phase, but only the designated goal object in the goal-conditioned phase.

2) Goal-conditioned Pushing Training: In this stage, the GraspNet ϕ_g is fixed while the PushNet ϕ_p is trained. A push is considered effective if it facilitates subsequent grasping. Specifically, the PushNet receives a positive reward when the grasp success probability (Q_g) predicted by ϕ_g increases. To penalize ineffective actions, we compute the occupancy ratio around the target object to detect environmental changes, and assign a negative reward if the push fails to alter the surroundings.

The push reward is defined as:

$$R_p = \begin{cases} 0.5, & \text{if } Q_g^{\text{in}} > 0.1, \\ -0.5, & \text{if } Q_g^{\text{in}} < 0.1 \text{ and } \text{Occ}^{\text{de}} < 0.1, \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where Q_g^{in} denotes the increase in Q_g after pushing, and Occ^{de} represents the decrease in occupancy ratio. During training, each episode allows up to five push actions before executing a grasp. If Q_g exceeds the predefined threshold Q_g^*, the grasp is performed immediately.

3) Goal-conditioned Alternating Training: In the first stage, the GraspNet is trained to handle occluded objects, whereas in the second stage, the PushNet is trained within

a push–grasp synergy setting. This sequential training introduces a distribution mismatch: the grasping network may fail to recognize graspable configurations in cluttered environments encountered during push–grasp interactions. Since the reward of PushNet depends on grasp evaluation, errors in the GraspNet can negatively affect its learning.

To mitigate this issue, we employ alternating training for the GraspNet and PushNet. The two networks are updated in turns, with one network's parameters fixed while the other is optimized, thereby promoting effective coordination in cluttered environments. Throughout every training episode, pushing operations are iteratively performed by the agent until the grasping network's peak Q-value, evaluated specifically within the target object's mask, surpasses a predefined threshold Q_g^*.

In this stage, we also revise the reward function of the GraspNet to discourage ineffective grasp attempts. Unlike the strict penalty schemes in prior works, we introduce a more forgiving reward structure for exploratory interactions. Specifically, if a grasp attempt fails but inadvertently alters the environment, we assign a neutral reward of 0 rather than a penalty. The modified grasp reward R'_g is defined as:

$$R'_g = \begin{cases} 1, & \text{if grasp success,} \\ -1, & \text{if grasp fail and no change detected,} \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

4) Implementation Details: In the first stage, $k = 10$ objects are randomly placed in the workspace, while in the second and third stages, the number of objects is increased to $k = 20$. The first object placed in the workspace is designated as the target object. To generate random occlusions during training, object positions are sampled from a 2D Gaussian distribution centered on the workspace, which is different from the 2D uniform distribution widely adopted in prior works. Following Xu et al. [12], we determine the grasping threshold Q_g^*. At the end of the first stage, the maximum Q-value was approximately 1.78, and therefore Q_g^* was set to 1.78. To optimize the model weights, we minimize a Huber loss function utilizing the Adam algorithm [36], employing default parameters. For exploration, we adopt an ϵ-greedy policy [37]. The exploration rate ϵ starts at an initial value of 0.5 and is gradually decayed to a minimum of 0.1. Additionally, we utilize a discount factor of $\gamma = 0.5$ for future rewards.

IV. EXPERIMENT

To rigorously assess our coordinated push-grasp framework, this section details a comprehensive series of trials executed across both simulated setups and physical deployments. We compare our method with that of Wang et al. [13], which represents a recent and highly competitive baseline.

A. Evaluation Metrics

Our approach is evaluated across a series of test cases, each repeated for N trials. In each trial, the robot must remove occlusions and grasp the designated target object. Performance is assessed using the same metrics as in prior studies [11]–[13].

1) Task Completion Rate (TCR): The proportion of successful target grasps over N trials. A trial is considered incomplete if the robot fails in five consecutive grasp attempts or executes five consecutive actions without affecting the environment. This metric reflects the model's ability to complete the overall task and is regarded as the most critical indicator.

2) Grasp Success Rate (GSR): The proportion of successful grasps among all grasp attempts in completed trials. Grasp attempts from unfinished trials are excluded. This metric evaluates the grasping capability of the model.

3) Motion Number (MN): The average number of actions per trial for completed tasks. Similarly, actions from unfinished trials are not included. This metric reflects the efficiency of the model's actions, particularly in pushing.

B. Simulation Experiments

To improve training efficiency, our model is trained in a simulated environment, as illustrated in Fig. 2. The simulation is implemented in CoppeliaSim [38], where we configure a UR5e robotic arm with an RG2 gripper and employ an RGB-D camera with known extrinsic parameters for image acquisition.

We begin by evaluating our method in random occlusion scenarios (Fig. 3). Experiments are conducted with 10, 20, and 30 objects, respectively, generating 500 scenes for each setting. The first or second object placed in the workspace is designated as the target. Random occlusions are generated similarly to training, by placing objects according to a 2D Gaussian distribution over the workspace.

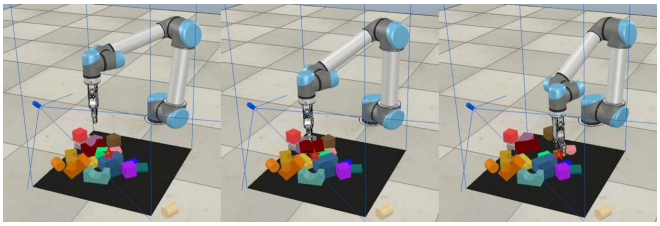
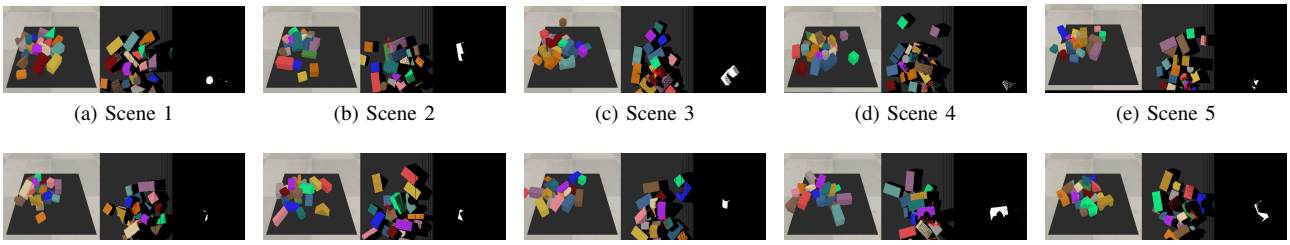


Fig. 3: Grasping in random occlusion scenarios in simulation. Left: the initial scene, in which the desired green target is occluded with a red star indicating its actual spatial location. Middle: surrounding objects are pushed to remove the occlusion. Right: the target object is successfully grasped.

As reported in Table I, our method achieves significantly higher task completion rates, along with improved grasp success rates and reduced motion numbers, particularly in cases with larger numbers of objects, which naturally result in more severe occlusions.

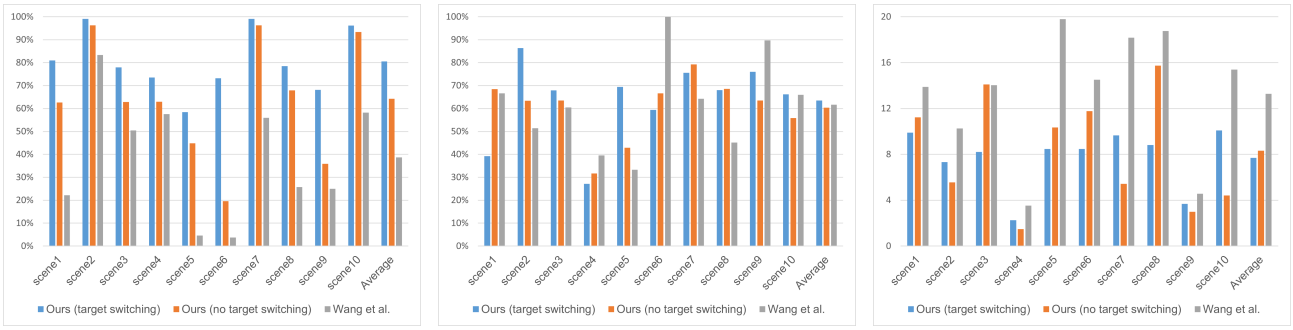
To further assess performance under extreme occlusion, we construct 10 challenging scenarios (Fig. 4), each evaluated across 100 trials. In Scene 1, 3, 8, and 9, the goal entity remains fully obscured from the camera's view during the starting configuration. The results, shown in Fig. 5, indicate that our method maintains higher task completion rates even



(a) Scene 1 (b) Scene 2 (c) Scene 3 (d) Scene 4 (e) Scene 5

(f) Scene 6 (g) Scene 7 (h) Scene 8 (i) Scene 9 (j) Scene 10

Fig. 4: Ten challenging scenarios with severe occlusions. Left: the original image of each scene. Middle: the initial action position and orientation. Right: the initial target masks. Note that in (a), (c), (h), and (i), the predefined goal object is completely occluded, and the target is switched according to our target switching mechanism.



(a) Task Completion Rate (b) Grasp Success Rate (c) Motion Number

Fig. 5: Simulation results for ten challenging scenarios with severe occlusions.

TABLE I: Simulation results for random occlusion scenarios.

Approach	Objects	TCR	GSR	MN
Wang et al. [13]	10	88.99%	46.84%	3.63
	20	61.63%	46.74%	6.33
	30	58.35%	47.37%	9.61
Ours (target switching)	10	**97.58%**	**47.36%**	**2.80**
	20	**83.11%**	**57.19%**	**5.04**
	30	**73.27%**	**53.36%**	**5.8**
Ours (no target switching)	10	95.69%	43.00%	3.14
	20	80.57%	45.73%	5.60
	30	70.27%	47.92%	7.57

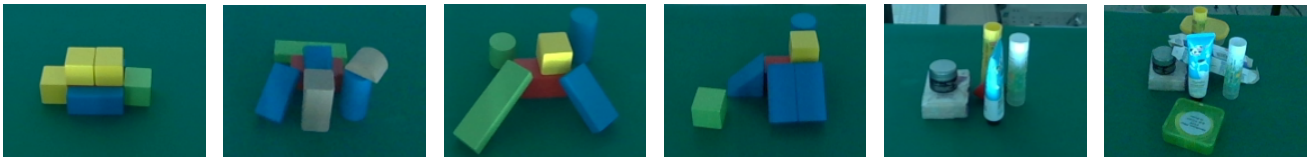
under severe occlusions. It is worth noting that in certain scenarios, grasp success rates and motion numbers are not always superior. This is because these metrics are calculated only over successfully completed tasks. By achieving higher task completion rates, our method is able to solve more complex tasks, which naturally demand additional grasp attempts and motions.

Notably, across all these evaluations, the baseline method (Wang et al. [13]) exhibits lower overall performance in our experiments compared to the results reported in their original work. This discrepancy primarily arises because our evaluation scenarios are significantly more complex. Specifically, our experiments feature much more densely cluttered environments with up to 30 objects and highly challenging configurations with severe or even complete occlusions. These extreme conditions inherently increase the task difficulty and expose the limitations of the baseline. Ultimately, these results strongly validate the effectiveness and robustness of our approach in heavily cluttered and occluded environments.

C. Ablation Study

To further evaluate the effectiveness of our approach, we conduct an ablation study. Specifically, we implement a variant that strictly relies on the mask of the predefined goal object, without applying the target switching mechanism under occlusion. As shown in Table I, the performance difference between the versions with and without target switching is relatively small in random occlusion environments. However, when evaluated in the highly challenging scenarios with severe or complete occlusions (Fig. 5), the target switching mechanism demonstrates a substantial advantage, significantly improving the task completion rate. This contrast clearly indicates that the target switching mechanism is particularly crucial and effective for resolving severe occlusions.

Notably, the results indicate that even without target switching, our method consistently outperforms the baseline of Wang et al. [13]. This improvement is primarily attributed to our mask-based action selection strategy for pushing. Our PushNet utilizes a weighted combination of the target object mask and the global mask, which fundamentally biases the robot to prioritize pushing the target object away from the



(a) Scene 1 (b) Scene 2 (c) Scene 3 (d) Scene 4 (e) Scene 5 (f) Scene 6

Fig. 6: Six real-world scenarios with challenging occlusions, captured from the observation camera. The red object indicates the target. In (a)–(d), the block objects are those encountered during simulation training, with the target in (a) completely occluded behind a blue block. In (e)–(f), novel objects with more complex shapes and varying heights are introduced, which were unseen during training. Notably, in (f) the target is also completely occluded behind a blue object.

surrounding clutter. In addition, our refined reward scheme and training strategy further contribute to the performance gain of our base model.

Finally, it is observed that target switching does not produce a notable increase in grasp success rate or a reduction in motion numbers. However, because the task completion rate is significantly improved in difficult scenarios, our method is able to solve more complex tasks that naturally demand additional grasp attempts and motions. This is consistent with our earlier analysis.

D. Real-World Experiments

We further validate our method in real-world environments by directly transferring the model trained in simulation without any retraining. To highlight its generalization ability, experiments are conducted using a different robotic platform consisting of a Franka Emika Panda arm equipped with the Franka Hand gripper. RGB-D images with a resolution of 640×480 are captured using an Intel RealSense D405c camera.

Experiments are performed in four real-world scenarios (Fig. 6 (a)-(d)), where the red target object is heavily occluded. In particular, in the first scenario the target object is completely invisible in the initial state. Each scenario is tested with 10 valid trials, and results are compared against the baseline of Wang et al. [13], as summarized in Table II. Our method outperforms the baseline across all metrics, with especially pronounced improvements in the most severely occluded Scene 1. These findings demonstrate that our approach can effectively generalize from simulation to real-world environments and achieve reliable grasping under occluded conditions.

Furthermore, the generalization capacity of our approach was assessed using novel physical items with varying heights and more complex shapes. Additional experiments are conducted in two challenging scenarios (Fig. 6 (e)-(f)). As reported in Table II, our method successfully generalizes to these previously unseen objects, further validating its robustness.

TABLE II: Real-word Results for six challenging scenarios.

Approach	Scene	TCR	GSR	MN
Wang et al. [13]	Scene 1	20%	50%	11
	Scene 2	40%	62.5%	5.5
	Scene 3	30%	75%	7.67
	Scene 4	30%	**60%**	13.33
	Scene 5	60%	**52.94%**	8.33
	Scene 6	30%	57.14%	11.33
	Average	35%	59.60%	9.53
Ours	Scene 1	**80%**	**62.5%**	**6.625**
	Scene 2	**100%**	**77.27%**	**3.8**
	Scene 3	**100%**	**90.91%**	**5.7**
	Scene 4	**100%**	**60%**	**7.7**
	Scene 5	**100%**	50%	**7.1**
	Scene 6	**100%**	**63.16%**	**6.2**
	Average	**96.67%**	**67.31%**	**6.19**

V. CONCLUSION

In this paper, we presented a self-supervised learning framework based on deep reinforcement learning that enables robots to acquire push–grasp synergy for grasping target objects under occlusions in cluttered environments. To address occlusions and improve efficiency, we introduce a target switching mechanism to handle severe occlusions and reduce ineffective actions, followed by a mask-based action selection strategy that enables more precise action choices and further improves efficiency. Comprehensive validations across virtual and physical platforms confirmed that the introduced strategy surpasses a competitive baseline, particularly by delivering superior task success ratios. Continued development will focus on further improving pushing efficiency and enhancing the stability and effectiveness of the approach in practical robotic applications.

VI. ACKNOWLEDGEMENT

This work was supported by National Science and Technology Major Project (No. 2022ZD0114903) and Beijing Natural Science Foundation (funding Number L258013)

REFERENCES

[1] Y. Zheng, L. Yao, Y. Su, Y. Zhang, Y. Wang, S. Zhao, Y. Zhang, and L.-P. Chau, "A survey of embodied learning for object-centric robotic manipulation," *Machine Intelligence Research*, pp. 1–39, 2025.

[2] M. Qin, J. Brawer, and B. Scassellati, "Robot tool use: A survey," *Frontiers in Robotics and AI*, vol. 9, p. 1009488, 2023.

[3] Z. Lu, N. Wang, and C. Yang, "A dynamic movement primitives-based tool use skill learning and transfer framework for robot manipulation," *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 1748–1763, 2024.

[4] G. Zuo, J. Tong, Z. Wang, and D. Gong, "A graph-based deep reinforcement learning approach to grasping fully occluded objects," *Cognitive Computation*, vol. 15, no. 1, pp. 36–49, 2023.

[5] X.-M. Wu, J.-F. Cai, J.-J. Jiang, D. Zheng, Y.-L. Wei, and W.-S. Zheng, "An economic framework for 6-dof grasp detection," in *European Conference on Computer Vision*. Springer, 2024, pp. 357–375.

[6] D. Li, C. Zhao, S. Yang, R. Song, X. Li, and W. Zhang, "Mpgnet: Learning move-push-grasping synergy for target-oriented grasping in occluded scenes," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 5064–5071.

[7] Y. Yang, H. Liang, and C. Choi, "A deep learning approach to grasping the invisible," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2232–2239, 2020.

[8] L. Wu, Y. Chen, Z. Li, and Z. Liu, "Efficient push-grasping for multiple target objects in clutter environments," *Frontiers in Neurorobotics*, vol. 17, p. 1188468, 2023.

[9] X. Cao, T. Lu, L. Zheng, Y. Cai, and S. Wang, "Plot: Human-like push-grasping synergy learning in clutter with one-shot target recognition," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 16, no. 4, pp. 1391–1404, 2024.

[10] Y. Wang and H. Kasaei, "Learning dual-arm push and grasp synergy in dense clutter," *IEEE Robotics and Automation Letters*, 2025.

[11] A. Zeng, S. Song, S. Welker, J. Lee, A. Rodriguez, and T. Funkhouser, "Learning synergies between pushing and grasping with self-supervised deep reinforcement learning," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 4238–4245.

[12] K. Xu, H. Yu, Q. Lai, Y. Wang, and R. Xiong, "Efficient learning of goal-oriented push-grasping synergy in clutter," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 6337–6344, 2021.

[13] Y. Wang, K. Mokhtar, C. Heemskerk, and H. Kasaei, "Self-supervised learning for joint pushing and grasping policies in highly cluttered environments," in *2024 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2024, pp. 13 840–13 847.

[14] A. Sahbani, S. El-Khoury, and P. Bidaud, "An overview of 3d object grasp synthesis algorithms," *Robotics and autonomous systems*, vol. 60, no. 3, pp. 326–336, 2012.

[15] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesis—a survey," *IEEE Transactions on robotics*, vol. 30, no. 2, pp. 289–309, 2013.

[16] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojea, and K. Goldberg, "Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics," *arXiv preprint arXiv:1703.09312*, 2017.

[17] C. Chen, S. Yan, M. Yuan, C. Tay, D. Choi, and Q. D. Le, "A minimal collision strategy of synergy between pushing and grasping for large clusters of objects," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 6817–6822.

[18] H.-S. Fang, C. Wang, H. Fang, M. Gou, J. Liu, H. Yan, W. Liu, Y. Xie, and C. Lu, "Anygrasp: Robust and efficient grasp perception in spatial and temporal domains," *IEEE Transactions on Robotics*, vol. 39, no. 5, pp. 3929–3945, 2023.

[19] S. Chen, W. Tang, P. Xie, W. Yang, and G. Wang, "Efficient heatmap-guided 6-dof grasp detection in cluttered scenes," *arXiv preprint arXiv:2403.18546*, 2024.

[20] Y. Chen, R. Xu, Y. Lin, and P. A. Vela, "A joint network for grasp detection conditioned on natural language commands," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 4576–4582.

[21] Y. Lu, Y. Fan, B. Deng, F. Liu, Y. Li, and S. Wang, "Vl-grasp: a 6-dof interactive grasp policy for language-oriented objects in cluttered indoor scenes," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 976–983.

[22] L. Li, A. Cherouat, H. Snoussi, and T. Wang, "Grasping with occlusion-aware ally method in complex scenes," *IEEE Transactions on Automation Science and Engineering*, 2024.

[23] Y. Cui, J. Ooga, A. Ogawa, and T. Matsubara, "Probabilistic active filtering with gaussian processes for occluded object search in clutter," *Applied Intelligence*, vol. 50, no. 12, pp. 4310–4324, 2020.

[24] D.-C. Hoang, J. A. Stork, and T. Stoyanov, "Context-aware grasp generation in cluttered scenes," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 1492–1498.

[25] Y. Yu, Z. Cao, S. Liang, W. Geng, and J. Yu, "A novel vision-based grasping method under occlusion for manipulating robotic system," *IEEE Sensors Journal*, vol. 20, no. 18, pp. 10 996–11 006, 2020.

[26] M. Danielczuk, A. Kurenkov, A. Balakrishna, M. Matl, D. Wang, R. Martín-Martín, A. Garg, S. Savarese, and K. Goldberg, "Mechanical search: Multi-step retrieval of a target object occluded by clutter," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 1614–1621.

[27] Z. Liu, Z. Wang, S. Huang, J. Zhou, and J. Lu, "Ge-grasp: Efficient target-oriented grasping in dense clutter," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 1388–1395.

[28] D. Ren, S. Wu, X. Wang, Y. Peng, and X. Ren, "Learning bifunctional push-grasping synergistic strategy for goal-agnostic and goal-oriented tasks," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 2909–2916.

[29] B. Serhan, H. Pandya, A. Kucukyilmaz, and G. Neumann, "Push-to-see: learning non-prehensile manipulation to enhance instance segmentation via deep q-learning," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 1513–1519.

[30] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[31] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[32] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[33] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[34] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[35] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. pmlr, 2015, pp. 448–456.

[36] D. P. Kingma, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[37] R. S. Sutton, A. G. Barto, *et al.*, *Reinforcement learning: An introduction*. MIT press Cambridge, 1998, vol. 1, no. 1.

[38] E. Rohmer, S. P. N. Singh, and M. Freese, "Coppeliasim (formerly v-rep): a versatile and scalable robot simulation framework," in *Proc. of The International Conference on Intelligent Robots and Systems (IROS)*, 2013, www.coppeliarobotics.com.