2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Self-supervised Learning for Joint Pushing and Grasping Policies in Highly Cluttered Environments

Yongliang Wang[1*], Kamal Mokhtar[1*], Cock Heemskerk[2], Hamidreza Kasaei[1]

***Abstract*— Robotic systems often face challenges when attempting to grasp a target object due to interference from surrounding items. We propose a Deep Reinforcement Learning (DRL) method that develops joint policies for grasping and pushing, enabling effective manipulation of target objects within untrained, densely cluttered environments. In particular, a dual RL model is introduced, which presents high resilience in handling complicated scenes, reaching an average of 98% task completion in simulation and real-world scenes. To evaluate the proposed method, we conduct comprehensive simulation experiments in three distinct environments: densely packed building blocks, randomly positioned building blocks, and common household objects. Further, real-world tests are conducted using actual robots to confirm the robustness of our approach in various untrained and highly cluttered environments. The results from experiments underscore the superior efficacy of our method in both simulated and real-world scenarios, outperforming recent state-of-the-art methods. To ensure reproducibility and further the academic discourse, we make available a demonstration video, the trained models, and the source code for public access. https://sites.google.com/view/pushandgrasp/home.**

I. Introduction

Grasping is fundamental to a myriad of robotic applications, facilitating tasks that are untenable without adept object manipulation [1]. In many situations, objects are not isolated, making grasping challenging due to environmental occlusions and clutter [2]–[4]. Drawing inspiration from human dexterity, robots should employ motion primitives, like pushing, to isolate target objects from clutter and facilitate grasping [5], [6]. For effective grasping, robots need to integrate visual observations of the goal object(s) and understand spatial relationships, rather than just executing a basic antipodal grasp. This necessitates a harmonious system where all components collaborate [7]. Given the potentially flawed sensory information a robot obtains from its environment, it needs to effectively address these inconsistencies. Additionally, dynamic scene changes after each action necessitate the adaptability of the robot before task completion [8]. Moreover, planning the manipulator's motion to achieve the desired pose is inherently challenging, requiring continuous environment observation and adjustments [4], [9]–[13].

This paper primarily addresses self-supervised learning for pushing and grasping in cluttered settings through DRL [14]–[16]. As depicted in Fig. 1, when a direct grasp of the target object is unfeasible, our method employs a combination of pushing and grasping to guarantee successful manipulation. A multitude of studies have delved into the conjunction of pushing and grasping, probing their combined impact on object manipulation [17]–[19]. While most existing studies focus on bin-picking tasks, which entail transferring objects between bins, our research investigates situations where surrounding clutter renders the target object initially ungraspable. The interplay between pushing and grasping cultivates a unified behavior, elevating the proficiency of object manipulation [20], [21]. Instead of considering pushing and grasping as separate actions, which can result in unforeseen scene alterations and problem-solving hurdles, our method utilizes DRL to harness the power of synergy and address its limitations. The complete system architecture is depicted in Fig. 2 and detailed in Section III. Broadly, our paper provides the following contributions:

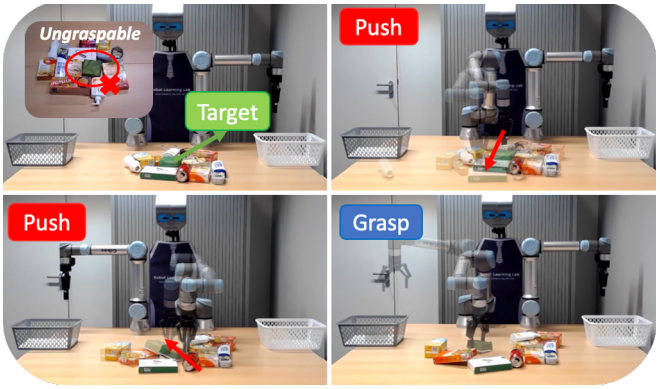

Fig. 1. Achieving Grasping Amidst Highly Cluttered Objects: The green target object, initially ungraspable due to surrounding blocks, is repositioned using pre-manipulation pushes by the robot. This strategic maneuver ensures a feasible grasp, showcasing the efficacy of our method.

- We propose a self-supervised learning method enabling a service robot to concurrently learn pushing and grasping policies, aiming to efficiently clear occlusions around target objects. Our approach addresses intricate tasks like obstacle removal while probing the constraints of the model and system.
- We conduct extensive simulations and real-world experiments to validate our method. Simulations encompass scenarios with building blocks in both packed and random arrangements. Empirical studies using the real robot further affirm the robustness of our approach. The results convincingly outperform current state-of-the-art strategies in task completion and grasp success rates, detailed in Section IV.

* These authors contributed equally to this work.
[1]Department of Artificial Intelligence, Bernoulli Institute, University of Groningen, 9747 AG, The Netherlands.
[2] Heemskerk Innovative Technology, Delft, The Netherlands
Yongliang Wang is funded by the China Scholarship Council.
Email: yongliang.wang@rug.nl, mokhtar.kamal@icloud.com, hamidreza.kasaei@rug.nl

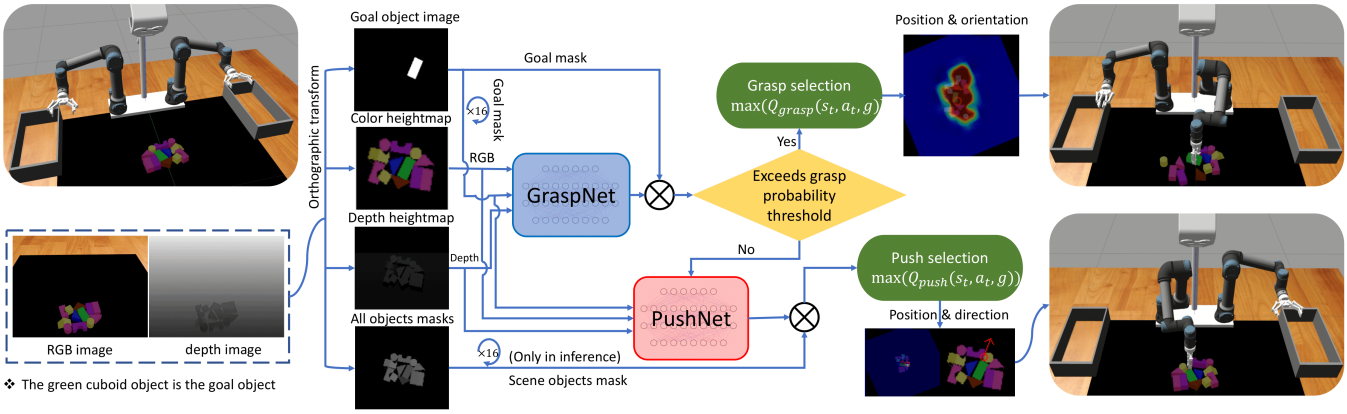

Fig. 2. **System Overview**: Within the Gazebo environment, an RGB-D camera and dual-arm UR5e robot are integrated. The camera transforms sensory data into orthographic projections, generating goal and scene masks. Using 360° rotation, the grasp and push nets assess heightmaps and the goal mask, outputting Q values to dictate push or grasp actions.

- For the broader research community's benefit, we make our source code and trained models publicly accessible, facilitating result replication.[1]

II. Related Work

In recent years, advancements in robotic grasping have been noteworthy. This section explores object grasping and pre-grasp manipulation [11], [22].

A. Grasping

Autonomous grasping has advanced rapidly, becoming foundational for robotic tasks in complex environments [23]. While traditional methods depended on precise object modeling for grasp metrics, they often faced impractical demands [24]. Ensuring stable interactions between the robot's hand and various objects necessitated accurate mathematical representations [25]. With recent computational advancements, the paradigm has shifted to deep-learning and data-driven methods. Leveraging image and depth data, these methods enable robots to simulate and perform grasps without exhaustive object knowledge, overcoming previous constraints [23].

Data-driven grasping methods prioritize the analysis of successful grasp instances over object mechanics. By utilizing grasp simulators, these approaches assess grasp stability based on manually gathered or simulated 3D environment data [26]–[28]. The rise of geometry-based antipodal grasp evaluation has invigorated the use of deep learning in this field [29], [30], with innovations like the GR-ConvNet [31]. The emergence of multi-view 3D grasping methods, such as those representing objects through three perspective depth images, provides grasp heatmaps [32]. While grasp simulators are adept at evaluating known objects, addressing unknown ones involves shape completion prior to grasping [33], [34] or image-based grasp pose inference [35]–[38]. Current methods typically concentrate on isolated objects in uncluttered environments. For cluttered scenarios, most solutions employ parallel grippers. Mousavian et al. proposed a grasp-generating method through a variational autoencoder, circumventing random sampling [39]. However, antipodal-centric approaches might falter in cluttered areas necessitating precise grasping [40], [41]. To address these challenges, Kiatos et al. proposed a system that combines scene point clouds with robot hand geometry. However, this system encounters difficulties in achieving collision-free grasps in restricted spaces [42].

B. Pre-grasp manipulation

In cluttered environments, pre-grasp manipulations are crucial for effective grasps. While RL excels in isolating objects from obstructions [43], a research gap exists in extracting specific objects from such environments, essential for service robots. Successful grasping often requires manipulating surrounding objects. Within RL, integrating a Markov Decision Process (MDP) is fundamental. However, the intrinsic data-heavy nature of RL, as well as the issue of scarce rewards in time-bound tasks, presents difficulties. An adjustment in the MDP, tying state-to-action and refining skill allocation in grasping, offers a solution [16]. Despite the prevalence of top-down grasping approaches, the diverse orientations of objects necessitate a range of grasp evaluations. One method involves an initial learning phase from isolated items, followed by a discriminative model that considers possible collisions. Following that, a Variational Autoencoder (VAE) is used to evaluate grasps and validate collision outcomes [17].

Visual Pushing for Grasping (VPG) proposed by Zeng et al. emphasized experiential learning, enabling robots to adapt their grasp based on object orientation from 16 potential push directions [44]. Bao et al. employed deep learning for task segmentation, though their reliance on color was restrictive [45]. Ewerton et al. introduced an algorithm optimized for cylindrical objects, aiming to reduce push actions [46]. Baris et al. leveraged deep Q-learning in cluttered settings [47], while Guo et al.'s graph-based DRL targeted occluded items but had a limited application range [48].

[1] https://github.com/Kamalnl92/Self-Supervised-Learning-for-pushing-and-grasping

While much of the existing research focuses on identifying objects in cluttered environments, our study extends to both densely packed and sporadically distributed object contexts. In highly cluttered environments, the efficient interplay of pushing and grasping remains challenging when success-based rewards are the only incentives. To enhance efficiency, we integrate Hindsight Experience Replay (HER) for goal relabeling, enriching the replay buffer, and accelerating learning [49]. Building upon Xu et al. [15], our paper further probes the push-grasp interrelation. By framing it as a MDP, a nuanced balance between deterministic and stochastic actions is maintained in our approach.

III. Method

A. Overall system

As depicted in Fig. 2, our system chooses between two actions: push and grasp, depending on environmental conditions. If grasp likelihood surpasses a threshold, it opts to grasp; otherwise, it pushes the target or adjacent objects approximately $1.3\,cm$ to enhance subsequent grasp opportunities. Grasp/push coordinates, initially in the image domain, are later converted to the robot's frame for action.

In our RL framework, we address pushing and grasping in highly cluttered environments. The policy, reward, and Q-value function are denoted by $\pi(s_t|g)$, $R(s_t, a_t, g)$, and $Q(s_t, a_t, g)$, respectively. Capturing the scene, a camera produces orthographic images which are rotated 16 times, each at 22.5 degrees, cumulatively spanning 360 degrees. This rotation facilitates the learning of grasp and push orientations. The action chosen aligns with the highest Q-value rotation. If this value exceeds 1.8, an object grasp is pursued. Otherwise, a push action is employed, as elaborated by Xu et al. (2021) [15]. Mirroring Generative Adversarial Networks (GAN) principles, the grasp network ϕ_g operates as a discriminator, while the push network ϕ_p acts akin to a generator. ϕ_p modifies the scene to optimize the grasp's Q-value, centering on Q_{grasp} (refer Fig. 2). Our model selectively processes the target object's discrete mask, isolating relevant scene features.

Our method diverges from Xu et al.(2021) [15], who assign specific masks to each object. Rather than restricting our model to Q-values of relevant objects during training, we expand its exploration domain. The model assesses all Q-values, acting on the peak value. During inference, we overlay the object mask to guide actions. The discrete masking method by [15] can lead to erroneous outcomes. As illustrated at the left of Fig.3, high pixel-wise Q-values in vacant areas might misdirect robotic tasks. This inconsistency, highlighted by the red circle, causes unintended maneuvers. Their strategy also necessitates different masking for training and testing, posing scalability challenges. Advanced segmentation methods, like those by Xiang et al. (2020) [50], provide solutions. However, with our primary focus away from perception, we employ RGB-color segmentation, recognizing the target's green color. The effectiveness of our masking is evident at the right of Fig. 3, following its application to pixel Q-values.

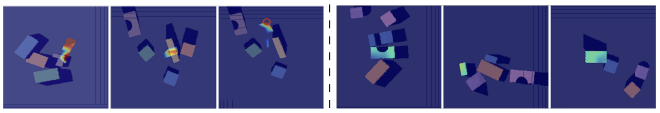

Fig. 3. Post-mask Pixel Q-values: (*left*) Xu et al. (2021) flaws [15]; (*right*) our refined method.

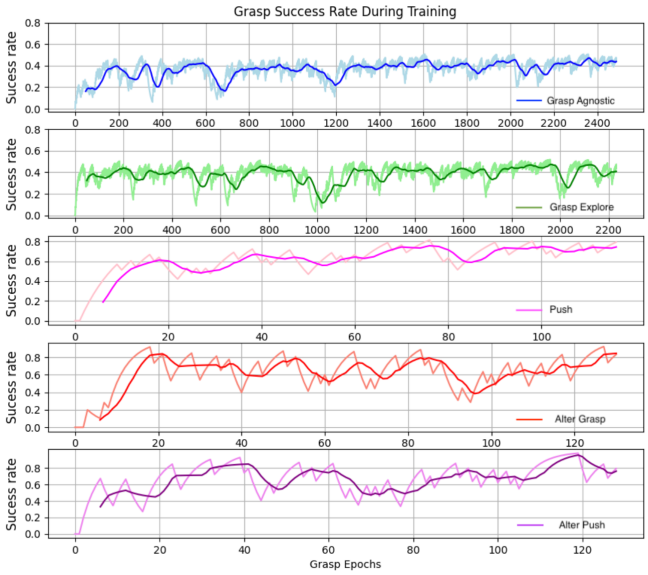

Fig. 4. Grasp success rate versus the number of grasp epochs. The top two are related to the grasping network ϕ_g training, and the middle one is for the push network ϕ_p training. The last two are when the training is alternated between the grasp net and push net.

B. Models hierarchical framework

Our architecture features two distinct networks: the grasp (ϕ_g) and push (ϕ_p) networks, setting it apart from Xu et al. (2021) [15]. We employ dual 121-layer DenseNets [20], previously trained on ImageNet [51], reducing our model's scale to one-third of that presented by Xu et al. These DenseNets independently process RGB inputs and normalized depth images. Their resulting outputs merge into a cohesive feature vector, channeled into a fully convolutional layer [52] equipped with 1×1 kernels, ReLU activation functions, and batch normalization [53]. Finally, bilinear upsampling is applied to ensure the output's dimensions align with the input.

C. Models training

Our training process unfolds in three stages: starting with goal-conditioned grasping, transitioning to goal-conditioned pushing, and concluding with alternating training between both. The rewards for grasping R_g, and pushing R_p, are defined as follows:

$$R_g = \begin{cases} 1, & \text{if grasp success} \\ 0, & \text{if not} \end{cases} \tag{1}$$

$$R_p = \begin{cases} 0.5, & Q_g^{improved} > 0.1 \text{ \& scene change} \\ -0.5, & \text{no scene change} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where $Q_g^{improved}$ is calculated as

$$Q_g^{improved} = Q_g^{post-push} - Q_g^{pre-push} \tag{3}$$

Furthermore, changes in the scene are gauged from the depth map around the target object. Drawing inspiration from the Bellman equation in RL, we define the state-action function relative to the goal as $\pi(s|a, g)$. Employing the epsilon-greedy action selection, denoted as $\epsilon(\pi(s|a, g))$, allows the agent to balance exploration and exploitation [54].

$$\nu(S_t, g) = E[R_{[t+1]} + \rho\nu(S_{[t+1]})|S_t = S, g] \tag{4}$$

Where ρ represents a discount function, an overview of the training process and the grasp success rate for the different stages of training are depicted in Fig. 4. In the following sub-sections, more details about the training are discussed.

1) Goal-conditioned grasping: During training, we set five objects in a sparse workspace, as depicted in Fig. 3. Training is episodic, with each grasp marking an episode through two phases. In the *grasp agnostic* phase, we use target relabeling to optimize sampling. Due to the better gripping ability of the model, we transition to the *grasp explores* phase, eschewing goal relabeling, assuming the model efficiently gauges orientation. Both stages utilize an ϵ-greedy strategy to balance exploration and exploitation [55]. Unlike Xu et al. (2021) [15], who terminate training based on Q-values, we rely on grasp success rates. As evident in Fig. 4, training plateaus after 1400 epochs. It should be noted that, during the *grasp explores* phase, despite a consistent success rate, the task's complexity heightens, given only specific grasps are deemed successful, more so with an unpolished push model.

2) Goal-conditioned pushing: In this training phase, we fix the grasp model and exclusively refine the push model through adversarial training. Each episode encompasses up to five pushes, culminating in a grasp. If the target object's grasp Q-value exceeds a certain threshold, the episode terminates with an immediate grasp by the robot. Within a mere 120 epochs, the push model's scene adjustments elevate the goal-directed grasp success rate significantly (see Fig. 4, third row). Rewards are allotted to the push model solely when it heightens the future grasp probability for the goal object.

3) Goal-conditioned alternating: Initially, the grasp model trains in sparse environments with 5-12 objects to minimize occlusion, leading to a distribution disparity. To reconcile this, after push model training, we fine-tune the grasp model. Both models then train collaboratively in 10-object scenarios for optimal coordination. Recognizing the erratic behavior of certain objects, like bottles, when pushed, we devise an alternative: the model prioritizes grasping non-target items to clear a path to the primary object, sequentially engaging with these obstacles until the main target becomes accessible.

Our objective is to grasp a designated target, highlighted in green in Fig. 5. Instead of conventional multi-predictive methods, we determine the target using aggregated pixel-wise Q-values, which inform both its orientation and position. After training, challenges persisted near target areas, with success rates plateauing at 40%. This approach does not work for such cases. The consistent collision of the robot arm with objects in the scene and the sequence decoding of removing collision objects seem impossible for the model to learn. Approaches such as curriculum learning could mitigate the issue [56]. It addresses a complex problem where a sequence of tasks is presented with an increase in difficulty.

IV. Experiment

To assess our push-grasping strategy, we conducted tests in both simulated and real-world scenarios. The objective was to compare its efficiency with established policies and verify its stability in real robotic systems, with minimal reduction in performance.

A. Simulation Experiments

1) Experimental Setups and Metrics: Extensive experiments were finished utilizing the proposed method on the Coppeliasim and Gazebo simulators. We trained on Coppeliasim [57] and conducted preliminary block-building tests to guarantee a fair comparison for both strategies. Subsequent Gazebo experiments delved into scenarios mimicking real-world conditions with common colored objects. The model trained on Coppeliasim was fine-tuned and deployed in Gazebo, allowing for the handling of real-world objects. The whole system, depicted in Fig. 2, features dual UR5e robot arms and an RGB-D Intel RealSense SR300. For motion mapping, we employed an inverse kinematics (IK) solver [58], and for efficient training, the networks used the Adam optimizer [59] on an NVIDIA V100 GPU. For the valid comparison, we compared our method with Xu et al. [15]. As far as we know, their methodology is currently state-of-the-art in the sphere of goal-object-oriented scenarios. We trained and evaluated both approaches on identical hardware and the same set of evaluation scenes to ensure an equitable comparison. The evaluation metrics we utilized are the same as those previously employed by [14], [15]:

- **Completion (C)**: The mean percentage completion over n test runs. Completions are successful and equal to 1 in a test run; if the system does not exceed in failing to grasp the goal object $n = 5$ times, it is 0. The metric measures the system's ability to complete the task.
- **Grasp success (GS)**: The mean percentage of successful grasp over all grasp attempts. This metric represents the accuracy of the model and its ability to estimate the goal object successfully grasped.
- **Motion number (MN)**: The mean number of push actions per completion. It reflects action efficiency.

2) Sim-to-Sim: CoppeliaSim is well-known for robust robotic testing due to its versatile models and interface [60], [61]. In contrast, Gazebo closely mirrors real-world conditions with its advanced physics [62], [63]. While CoppeliaSim excels in grasping simulations, Gazebo, relying on the Grasp-fix-plugin [64], sometimes inaccurately executes grasping [65]. Our Sim-to-Sim method trains in CoppeliaSim and fine-tuned in Gazebo (Fig. 6), merging the strengths of both for optimal accuracy.

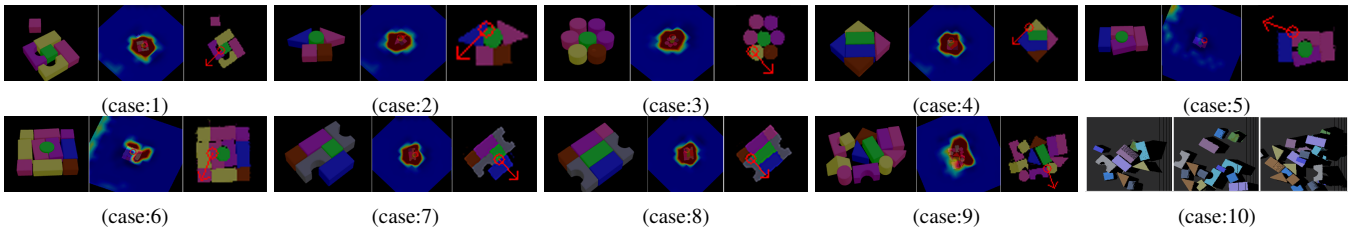

Fig. 5. The first column shows the original image; the next depicts the output angle and position; the last highlights the initial action. Case $1-9$: Simulation experiments with 9 distinct packed scenes featuring dense adversarial clutter; each scene's target is the green object. Case 10: Typical scenes showcasing random 10, 15, and 20 objects from left to right.

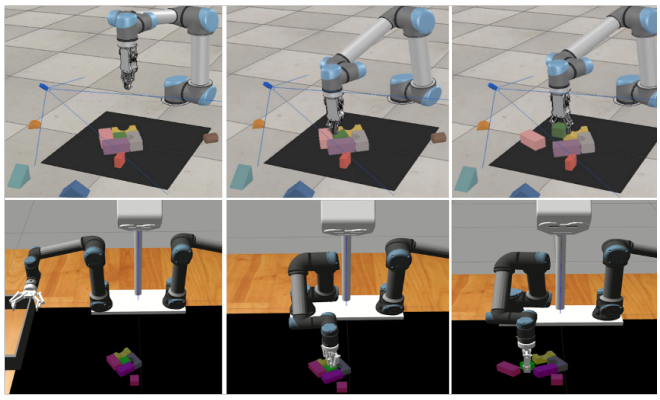

Fig. 6. Snapshot Sequence: Robot declutters for target grasp: The sequence includes (*left*) the initial scene; (*middle*) the process of pushing surrounding objects; (*right*) the final act of grasping the goal object. Top: Coppeliasim; Bottom: Gazebo. Model trained in Coppeliasim, fine-tuned in Gazebo.

B. Real Robot Experiments

We verified our methodology through real-world testing experiments using robots and objects that mirror our simulations. In untrained and highly cluttered environments and amid household items, our method, as depicted in Fig. 8, showcased performance akin to human-like cognition, emphasizing its viability for kitchen service robots.

C. Results Analysis

We conducted four experiments to assess our approach. In Coppeliasim, we evaluated densely arranged and randomized object scenes. Subsequently, we used Gazebo to examine household objects. The final test gauged a physical robot's performance with a target object in multifaceted sequences, echoing real-world challenges.

1) Packed Scenes: In the initial experiment, we assessed 9 intricate scenarios, each iterated 100 times. Results and their heat maps are found in Fig. 5 (case:1-9) and Table I. The minimal standard error showcases the method's reliable action sequence. The model appears to value push rewards (see Eq. 2), indicating efficient surrounding object clearance by relocating the target. Elevated Q-values adjacent to the target in Fig. 6 emphasize that creating space facilitates grasping. From the illustrative examples in Fig. 5, we posit that a single push is often sufficient to effectively grasp the target object.

In packed scenes, we choose variously shaped target objects, consistently arranged from scenes case:1 to case:9, with 5-12 objects each. Table I shows that our method reaches an impressive completion rate of over 95%, outdoing Xu et al. [15] by 16%. Moreover, we exceed their Grasp Success (GS) by approximately 60%. Xu et al. struggle due to inaccurate space estimation during training and inference, primarily from ineffective masking, causing early experiment conclusions. In contrast, our model accurately discerns optimal grasps, leading to greater success. In scenes (case:2, case:3, and case:5), the cylindrical object's tendency to roll poses challenges for both methods, as detailed in Table I. However, our method, with its subtle push strategy, ensures stability and outpaces Xu et al. [15], showcasing its real-world applicability.

2) Random Scenes: In our second experiment, we assess scenes with 10, 15, and 20 objects, evaluating our method on 100 scenes for each count. As shown in Fig. 5 case:10, object layouts are randomly generated, sometimes hiding the goal object. The results in Table II raise questions about the goal mask's necessity. Though models with and without the mask show comparable completion rates, the latter needs an extra action. However, its reduced size makes it an appealing alternative. Although our approach takes more actions in 10-object scenes, it ensures higher success rates. While adding objects initially does not impact performance, complexities in 20-object scenarios reduce efficiency. However, our method consistently outperforms Xu et al. [15], especially in 20-object scenes with a 20% higher success rate. Despite having a higher mean number of actions (MN) than Xu et al. [15], the robustness of our method is evident across varying scene complexities.

3) Household Scenes: In our household object experiments, we utilize the Gazebo simulator to evaluate our approach on 4 untrained simulated household items, always targeting to grasp a green milk box (see Fig. 7). The design of scenes mimics real-world challenges in complex environments. Based on the data in Table III, after conducting 100 repetitions for each scenario, our approach consistently yields not only a superior GS but also maintains a completion rate of 100%.

4) Real Robot Scenes: As shown in Fig. 8, our strategy enables the robot to adeptly grasp a target in real-world cluttered environments using our devised method. The figure underscores that, even in densely populated settings such as

(case:11) (case:12) (case:13) (case:14)

Fig. 7. Four simulated scenes with novel, untrained objects in dense household clutter. The target in each is the green box. The initial column presents the original image; the middle one is the model's angle and position; the final is the action.

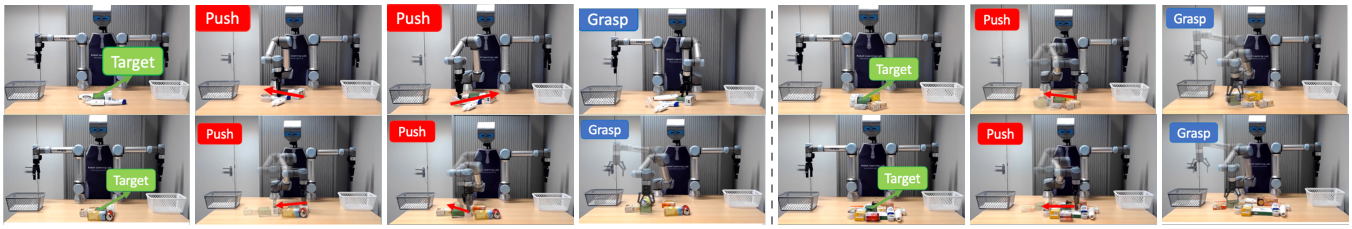

Fig. 8. Real-world Experiments: Four untrained objects in scenes where clutter obstructs direct goal grasping.

TABLE I

SIMULATION RESULTS FOR 9 COMPLEX PACKED SCENES WITH 100 TRIALS PER SCENE

Approach	#scene	C%	GS%	MN
Xu et al. [15]	case:1	83.0 ± 3.75	26.48 ± 2.16	3.43 ± 0.31
	case:2	97.98 ± 1.42	51.82 ± 3.63	$\mathbf{2.78 \pm 0.29}$
	case:3	100.0 ± 0.0	58.96 ± 3.75	$\mathbf{5.64 \pm 0.39}$
	case:4	99.0 ± 0.99	67.11 ± 3.88	$\mathbf{3.53 \pm 0.47}$
	case:5	$\mathbf{100.0 \pm 0.0}$	60.89 ± 3.66	$\mathbf{5.33 \pm 0.52}$
	case:6	98.0 ± 1.41	57.47 ± 3.77	$\mathbf{11.36 \pm 1.54}$
	case:7	96.0 ± 1.97	44.84 ± 3.32	$\mathbf{2.44 \pm 0.32}$
	case:8	95.0 ± 2.19	46.95 ± 3.41	11.1 ± 1.26
	case:9	100.0 ± 0.0	68.49 ± 3.86	4.49 ± 0.32
Ours	case:1	$\mathbf{98.98 \pm 1.01}$ ↑	$\mathbf{86.08 \pm 3.32}$ ↑	$\mathbf{1.12 \pm 0.03}$ ↑
	case:2	$\mathbf{99.0 \pm 0.99}$ ↑	$\mathbf{66.67 \pm 3.88}$ ↑	4.99 ± 0.43
	case:3	$\mathbf{100.0 \pm 0.0}$ ↑	$\mathbf{77.52 \pm 3.69}$ ↑	5.44 ± 0.31
	case:4	$\mathbf{100.0 \pm 0.0}$ ↑	$\mathbf{83.33 \pm 5.44}$ ↑	5.88 ± 0.46
	case:5	98.98 ± 1.02	$\mathbf{70.5 \pm 3.91}$ ↑	8.97 ± 0.72
	case:6	$\mathbf{99.0 \pm 0.99}$ ↑	$\mathbf{86.21 \pm 3.3}$ ↑	17.67 ± 1.50
	case:7	$\mathbf{100 \pm 0.0}$ ↑	$\mathbf{61.73 \pm 3.83}$ ↑	3.04 ± 0.14
	case:8	$\mathbf{99.0 \pm 1.0}$ ↑	$\mathbf{62.89 \pm 3.86}$ ↑	$\mathbf{2.41 \pm 0.089}$ ↑
	case:9	$\mathbf{100.0 \pm 0.0}$ ↑	$\mathbf{91.74 \pm 2.65}$ ↑	14.34 ± 1.60

TABLE II

SIMULATION RESULTS FOR RANDOM OBJECTS SCENES WITH 100 TRIALS PER SCENE

Approach	#objects	C%	GS%	MN
Xu et al. [15]	10	71.56 ± 4.48	20.35 ± 1.57	$\mathbf{0.64 \pm 0.25}$
	15	71.0 ± 4.56	22.72 ± 1.75	$\mathbf{1.1 \pm 0.54}$
	20	77.89 ± 4.27	21.59 ± 1.78	$\mathbf{1.13 \pm 0.45}$
Ours (mask)	10	$\mathbf{98.97 \pm 1.02}$ ↑	66.21 ± 3.91	1.02 ± 0.14
	15	$\mathbf{100 \pm 0.0}$ ↑	$\mathbf{74.60 \pm 3.89}$ ↑	3.67 ± 0.98
	20	97.22 ± 1.95	$\mathbf{69.23 \pm 4.62}$ ↑	6.09 ± 1.82
Ours (no mask)	10	$\mathbf{98.78 \pm 1.22}$ ↑	$\mathbf{68.33 \pm 4.29}$ ↑	3.19 ± 0.56
	15	$\mathbf{100 \pm 0.0}$ ↑	71.42 ± 4.28	3.27 ± 0.47
	20	$\mathbf{99.22 \pm 0.1}$ ↑	67.36 ± 4.83	7.12 ± 2.04

the fourth environment, our method often accomplishes the task with a mere 1-2 pushes. For a more comprehensive view, please refer to the accompanying video.

V. CONCLUSION AND FUTURE WORK

In this paper, we propose an innovative self-supervised DRL approach that enables robots to grasp target objects in highly cluttered and untrained environments. We conduct a comprehensive evaluation of our method across a spectrum of challenging scenarios, including densely packed arrangements of building blocks, random environments with varying object counts (10, 15, and 20 objects), and real-world object manipulation tasks. Notably, our approach achieves superior performance while maintaining a significantly smaller model size.

TABLE III

SIMULATION RESULTS FOR 4 HOUSEHOLD SCENES WITH 100 TRIALS PER SCENE

Approach	#scene	C%	GS%	MN
Xu et al. [15]	case:11	95.96 ± 1.99	46.48 ± 3.41	3.75 ± 0.88
	case:12	94.0 ± 2.39	41.67 ± 3.16	2.58 ± 0.46
	case:13	97.0 ± 1.71	57.80 ± 3.78	$\mathbf{1.73 \pm 0.12}$
	case:14	99.0 ± 0.99	58.82 ± 3.79	$\mathbf{1.47 \pm 0.28}$
Ours	case:11	$\mathbf{100.0 \pm 0.0}$ ↑	$\mathbf{56.18 \pm 3.73}$ ↑	$\mathbf{3.33 \pm 0.29}$ ↑
	case:12	$\mathbf{100.0 \pm 0.0}$ ↑	$\mathbf{68.97 \pm 3.86}$ ↑	$\mathbf{1.26 \pm 0.05}$ ↑
	case:13	$\mathbf{100.0 \pm 0.0}$ ↑	$\mathbf{75.76 \pm 3.74}$ ↑	2.49 ± 0.10
	case:14	$\mathbf{100.0 \pm 0.0}$ ↑	$\mathbf{76.92 \pm 3.71}$ ↑	3.69 ± 0.21

We design two distinct scenarios using building blocks, enhancing complexity by altering the shape of the target object. The results consistently demonstrate our method's proficiency in interacting with its surroundings and successfully grasping the target object. Overall, our performance in terms of task completion and grasp success surpasses Xu et al.'s, with minor variations in the Mean Number of Actions (MN) metric, primarily attributable to occasional misjudgments in their model. In the third experimental phase, we transition to real-world object manipulation tasks, where our approach consistently outperforms others, especially for some untrained objects in environments. To ensure a reliable evaluation, we employ a Sim-to-Sim strategy to mitigate inaccuracies introduced by Gazebo, ensuring a more trustworthy assessment. Impressively, our approach maintains its robustness and effectiveness in real-world testing, mirroring its strong performance in simulation. Our future research explores curriculum learning to further enhance our system's object manipulation proficiency.

ACKNOWLEDGMENT

We thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Hábrók high-performance computing cluster.

REFERENCES

[1] R. Newbury, M. Gu, L. Chumbley, A. Mousavian, C. Eppner, J. Leitner, J. Bohg, A. Morales, T. Asfour, D. Kragic *et al.*, "Deep learning approaches to grasp synthesis: A review," *IEEE Transactions on Robotics*, 2023.

[2] 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.

[3] 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.

[4] N. Lu, Y. Cai, T. Lu, X. Cao, W. Guo, and S. Wang, "Picking out the impurities: Attention-based push-grasping in dense clutter," *Robotica*, vol. 41, no. 2, pp. 470–485, 2023.

[5] R. Huang, F. Mu, W. Li, H. Liu, and H. Cheng, "Estimating 6d object poses with temporal motion reasoning for robot grasping in cluttered scenes," *IEEE Robotics and Automation Letters*, 2022.

[6] M. Kang, H. Kee, J. Kim, and S. Oh, "Grasp planning for occluded objects in a confined space with lateral view using monte carlo tree search," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 10 921–10 926.

[7] B. Burgess-Limerick, C. Lehnert, J. Leitner, and P. Corke, "Dgbench: An open-source, reproducible benchmark for dynamic grasping," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 3218–3224.

[8] S. Duan, G. Tian, Z. Wang, S. Liu, and C. Feng, "A semantic robotic grasping framework based on multi-task learning in stacking scenes," *Engineering Applications of Artificial Intelligence*, vol. 121, p. 106059, 2023.

[9] T. Kunz, U. Reiser, M. Stilman, and A. Verl, "Real-time path planning for a robot arm in changing environments," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 5906–5911.

[10] 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.

[11] L. Wu, J. Wu, Y. Chen, Z. Li, and Z. Liu, "Learning pre-grasp manipulation of multiple flat target objects in clutter," in *2023 9th International Conference on Control, Automation and Robotics (ICCAR)*. IEEE, 2023, pp. 371–376.

[12] W. Zhou and D. Held, "Learning to grasp the ungraspable with emergent extrinsic dexterity," in *Conference on Robot Learning*. PMLR, 2023, pp. 150–160.

[13] H. Zhang, H. Liang, L. Cong, J. Lyu, L. Zeng, P. Feng, and J. Zhang, "Reinforcement learning based pushing and grasping objects from ungraspable poses," *arXiv preprint arXiv:2302.13328*, 2023.

[14] M. Fujita, Y. Domae, A. Noda, G. Garcia Ricardez, T. Nagatani, A. Zeng, S. Song, A. Rodriguez, A. Causo, I.-M. Chen *et al.*, "What are the important technologies for bin picking? technology analysis of robots in competitions based on a set of performance metrics," *Advanced Robotics*, vol. 34, no. 7-8, pp. 560–574, 2020.

[15] 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.

[16] L. Berscheid, P. Meißner, and T. Kröger, "Robot learning of shifting objects for grasping in cluttered environments," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 612–618.

[17] A. Murali, A. Mousavian, C. Eppner, C. Paxton, and D. Fox, "6-dof grasping for target-driven object manipulation in clutter," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 6232–6238.

[18] B. Tang, M. Corsaro, G. Konidaris, S. Nikolaidis, and S. Tellex, "Learning collaborative pushing and grasping policies in dense clutter," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 6177–6184.

[19] 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.

[20] 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.

[21] 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.

[22] S. Ghorbani, Z. Samadikhoshkho, and F. Janabi-Sharifi, "Dual-arm aerial continuum manipulation systems: modeling, pre-grasp planning, and control," *Nonlinear Dynamics*, vol. 111, no. 8, pp. 7339–7355, 2023.

[23] H. Duan, P. Wang, Y. Huang, G. Xu, W. Wei, and X. Shen, "Robotics dexterous grasping: The methods based on point cloud and deep learning," *Frontiers in Neurorobotics*, vol. 15, p. 73, 2021.

[24] 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.

[25] D. Prattichizzo and J. C. Trinkle, "Grasping," in *Springer handbook of robotics*. Springer, 2016, pp. 955–988.

[26] A. T. Miller, S. Knoop, H. I. Christensen, and P. K. Allen, "Automatic grasp planning using shape primitives," in *2003 IEEE International Conference on Robotics and Automation (Cat. No. 03CH37422)*, vol. 2. IEEE, 2003, pp. 1824–1829.

[27] C. Goldfeder, P. K. Allen, C. Lackner, and R. Pelossof, "Grasp planning via decomposition trees," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007, pp. 4679–4684.

[28] M. Ciocarlie, C. Goldfeder, and P. Allen, "Dexterous grasping via eigengrasps: A low-dimensional approach to a high-complexity problem," in *Robotics: Science and systems manipulation workshop-sensing and adapting to the real world*, 2007.

[29] A. t. Pas and R. Platt, "Using geometry to detect grasps in 3d point clouds," *arXiv preprint arXiv:1501.03100*, 2015.

[30] A. ten Pas, M. Gualtieri, K. Saenko, and R. Platt, "Grasp pose detection in point clouds," *The International Journal of Robotics Research*, vol. 36, no. 13-14, pp. 1455–1473, 2017.

[31] S. Kumra, S. Joshi, and F. Sahin, "Antipodal robotic grasping using generative residual convolutional neural network," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 9626–9633.

[32] H. Kasaei and M. Kasaei, "Mvgrasp: Real-time multi-view 3d object grasping in highly cluttered environments," *arXiv preprint arXiv:2103.10997*, 2021.

[33] J. Varley, C. DeChant, A. Richardson, J. Ruales, and P. Allen, "Shape completion enabled robotic grasping," in *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2017, pp. 2442–2447.

[34] M. Van der Merwe, Q. Lu, B. Sundaralingam, M. Matak, and T. Hermans, "Learning continuous 3d reconstructions for geometrically aware grasping," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 11 516–11 522.

[35] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *The International Journal of Robotics Research*, vol. 34, no. 4-5, pp. 705–724, 2015.

[36] 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.

[37] P. Schmidt, N. Vahrenkamp, M. Wächter, and T. Asfour, "Grasping of unknown objects using deep convolutional neural networks based on depth images," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 6831–6838.

[38] Q. Lu, M. Van der Merwe, and T. Hermans, "Multi-fingered active grasp learning," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 8415–8422.

[39] A. Mousavian, C. Eppner, and D. Fox, "6-dof graspnet: Variational grasp generation for object manipulation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 2901–2910.

[40] Z. Xu, B. Qi, S. Agrawal, and S. Song, "Adagrasp: Learning an adaptive gripper-aware grasping policy," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 4620–4626.

[41] B. Wu, I. Akinola, A. Gupta, F. Xu, J. Varley, D. Watkins-Valls, and P. K. Allen, "Generative attention learning: A "general" framework for high-performance multi-fingered grasping in clutter," *Autonomous Robots*, vol. 44, no. 6, pp. 971–990, 2020.

[42] M. Kiatos, S. Malassiotis, and I. Sarantopoulos, "A geometric ap-

proach for grasping unknown objects with multifingered hands," *IEEE Transactions on Robotics*, vol. 37, no. 3, pp. 735–746, 2020.

[43] M. Q. Mohammed, K. L. Chung, and C. S. Chyi, "Review of deep reinforcement learning-based object grasping: Techniques, open challenges, and recommendations," *IEEE Access*, vol. 8, pp. 178 450–178 481, 2020.

[44] 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.

[45] J. Bao, G. Zhang, Y. Peng, Z. Shao, and A. Song, "Learn multi-step object sorting tasks through deep reinforcement learning," *Robotica*, vol. 40, no. 11, pp. 3878–3894, 2022.

[46] E. R. Vieira, D. Nakhimovich, K. Gao, R. Wang, J. Yu, and K. E. Bekris, "Persistent homology for effective non-prehensile manipulation," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 1918–1924.

[47] 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.

[48] G. Zuo, J. Tong, Z. Wang, and D. Gong, "A graph-based deep reinforcement learning approach to grasping fully occluded objects," *Cognitive Computation*, pp. 1–14, 2022.

[49] M. Andrychowicz, F. Wolski, A. Ray, J. Schneider, R. Fong, P. Welinder, B. McGrew, J. Tobin, O. Pieter Abbeel, and W. Zaremba, "Hindsight experience replay," *Advances in neural information processing systems*, vol. 30, 2017.

[50] Y. Xiang, C. Xie, A. Mousavian, and D. Fox, "Learning RGB-D feature embeddings for unseen object instance segmentation," in *Conference on Robot Learning (CoRL)*, 2020.

[51] 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.

[52] 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.

[53] 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.

[54] M. Gimelfarb, S. Sanner, and C.-G. Lee, "ϵ-bmc: A bayesian ensemble approach to epsilon-greedy exploration in model-free reinforcement learning," *arXiv preprint arXiv:2007.00869*, 2020.

[55] C. D'Eramo, A. Cini, and M. Restelli, "Exploiting action-value uncertainty to drive exploration in reinforcement learning," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.

[56] S. Narvekar, B. Peng, M. Leonetti, J. Sinapov, M. E. Taylor, and P. Stone, "Curriculum learning for reinforcement learning domains: A framework and survey," *arXiv preprint arXiv:2003.04960*, 2020.

[57] E. Rohmer, S. P. N. Singh, and M. Freese, "V-rep: A versatile and scalable robot simulation framework," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 1321–1326.

[58] R. Diankov, "Automated construction of robotic manipulation programs," 2010.

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

[60] J. Chen, J. Cao, Z. Cheng, and Y. Sahni, "Maniware: An easy-to-use middleware for cooperative manipulator teams," *IEEE Internet of Things Journal*, 2023.

[61] W. Zhao, J. P. Queralta, and T. Westerlund, "Sim-to-real transfer in deep reinforcement learning for robotics: a survey," in *2020 IEEE symposium series on computational intelligence (SSCI)*. IEEE, 2020, pp. 737–744.

[62] K. Kleeberger, R. Bormann, W. Kraus, and M. F. Huber, "A survey on learning-based robotic grasping," *Current Robotics Reports*, vol. 1, pp. 239–249, 2020.

[63] G. Bellegarda, Y. Chen, Z. Liu, and Q. Nguyen, "Robust high-speed running for quadruped robots via deep reinforcement learning," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 10 364–10 370.

[64] D. Wang, B. Lutz, P. J. Cobb, and P. Dames, "Rascal: Robotic arm for sherds and ceramics automated locomotion," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 6378–6384.

[65] J. P. Hanna, S. Desai, H. Karnan, G. Warnell, and P. Stone, "Grounded action transformation for sim-to-real reinforcement learning," *Machine Learning*, vol. 110, no. 9, pp. 2469–2499, 2021.