

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

# DexSinGrasp: Learning a Unified Policy for Dexterous Object Singulation and Grasping in Densely Cluttered Environments

Lixin Xu\*, Zixuan Liu\*, Zhewei Gui, Jingxiang Guo,  
Zeyu Jiang, Tongzhou Zhang, Zhixuan Xu, Chongkai Gao, Lin Shao†

**Abstract**—Grasping objects in cluttered environments remains a fundamental yet challenging problem in robotic manipulation. While prior works have explored learning-based synergies between pushing and grasping for two-fingered grippers, few have leveraged the high degrees of freedom (DoF) in dexterous hands to perform efficient singulation for grasping in cluttered settings. In this work, we introduce *DexSinGrasp*, a unified policy for dexterous object singulation and grasping. *DexSinGrasp* enables high-dexterity object singulation to facilitate grasping, significantly improving efficiency and effectiveness in cluttered environments. We incorporate clutter arrangement curriculum learning to enhance success rates and generalization across diverse clutter conditions, while policy distillation enables a deployable vision-based grasping strategy. To evaluate our approach, we introduce a set of cluttered grasping tasks with varying object arrangements and occlusion levels. Experimental results show that our method outperforms baselines in both efficiency and grasping success rate, particularly in dense clutter. Codes, appendix, and videos are available on our website <https://nus-lins-lab.github.io/dexsingweb/>.

**Index Terms**—Grasping, dexterous manipulation, deep learning in grasping and manipulation.

## I. INTRODUCTION

**D**EXTEROUS grasping of target objects in cluttered environments is crucial for various applications, from production lines [1] to assembly processes [2], [3] and beyond. While dexterous hands offer high degrees of freedom (DoF) and substantial potential for complex manipulation tasks [4]–[9], effectively leveraging their capabilities for grasping in cluttered settings remains a challenging problem. The importance of dexterous grasping in cluttered environments lies in its ability to singulate surroundings efficiently, where the flexibility of fingers can be leveraged to create sufficient space for a successful grasp. Recent dexterous grasping approaches [10], [11] focus primarily on grasping target objects in scenarios without the need to rearrange surrounding objects. These methods are typically designed for loosely cluttered environments, where they extract scene information from segmented point clouds [10] or segmented images [11] to identify suitable grasping positions. However, due to the lack

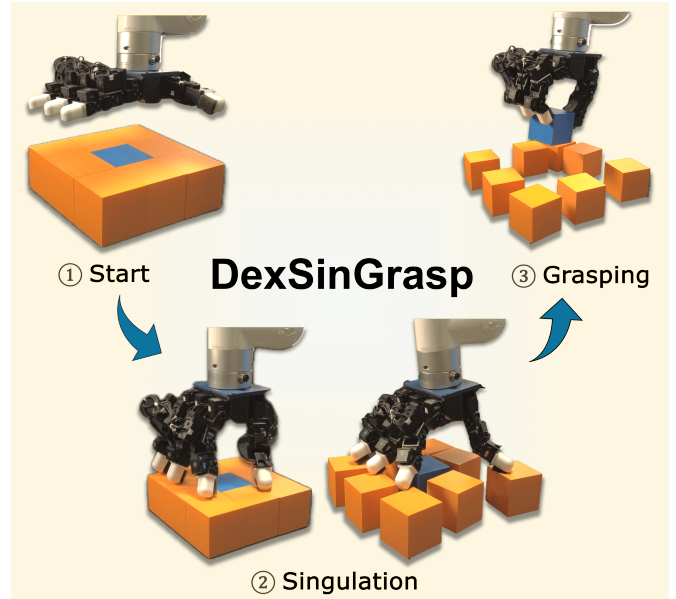


Fig. 1. We propose DexSinGrasp to learn a unified policy for dexterous object singulation and grasping in densely cluttered environments

of explicit singulation training, these approaches struggle in denser clutter, where avoiding interaction with surrounding objects is insufficient to ensure grasp success.

One approach to handling densely cluttered environments is to *singulate*—spatially separate—the target object from surrounding objects. Researchers have explored frameworks to learn the synergies between pushing and grasping [12]–[14] for two-fingered grippers. These methods emphasize that in densely cluttered scenes where the target object is initially ungraspable, singulation must first be performed before grasping [15]–[23]. Naturally, this leads to a pushing-and-grasping process, where Zeng et al. [12] proposed a method using two networks to coordinate pushing and grasping. However, their approach relies solely on a grasp success reward to guide coordination, leading to low sample efficiency during training. Other methods introduced adversarial training to enhance synergy learning efficiency [13], [14]. Despite these advancements, the mechanical constraints of grippers require the target object to be fully isolated from surrounding clutter, making singulation inefficient. In contrast, dexterous hands perform singulation using only their fingers, minimizing movement of the end-effector (i.e., the palm) and providing a more flexible and efficient execution in cluttered settings.

Learning singulation for grasping with dexterous hands offers a promising approach to handling objects in cluttered environments. However, the high degrees of freedom (DoF) of

\* denotes equal contribution

† denotes the corresponding author

Manuscript received: August, 14, 2025; Revised October, 21, 2025; Accepted November, 10, 2025.

This paper was recommended for publication by Editor Julia Borrás Sol upon evaluation of the Associate Editor and Reviewers’ comments. This research is supported by the Ministry of Education, Singapore, under the Academic Research Fund Tier 1 (FY2024).

The authors are with the School of Computing, National University of Singapore. davidxulixin@gmail.com, zixuanliu@u.nus.edu, linshao@nus.edu.sg

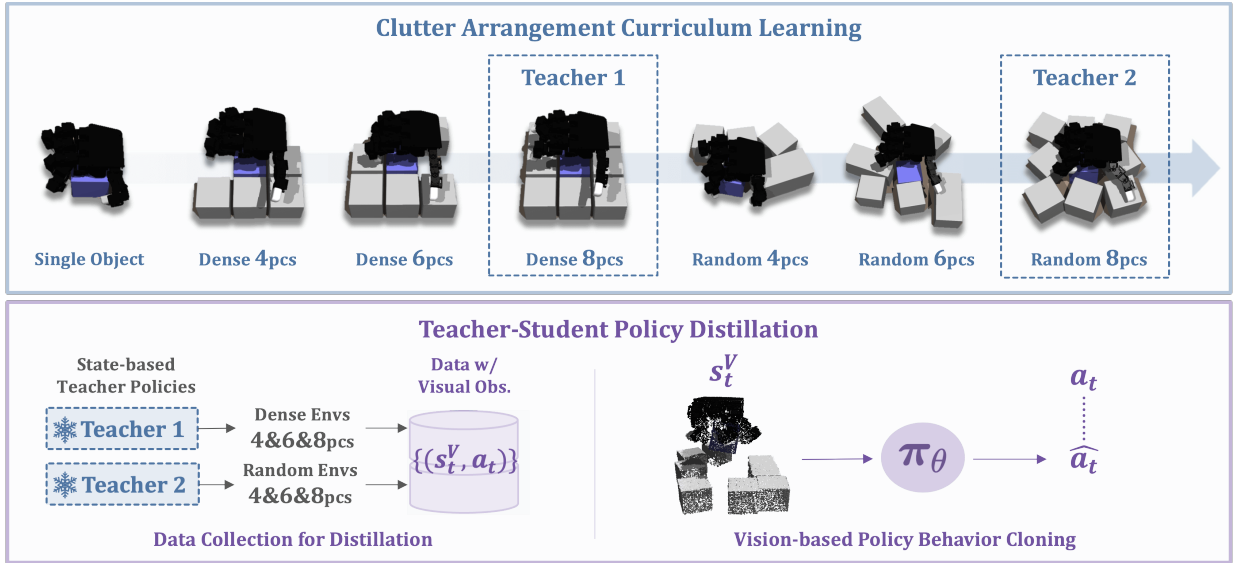


Fig. 2. Framework of DexSinGrasp. Firstly, we adopt clutter arrangement curriculum learning to progressively improve the performance of our teacher policy to address the challenge of training from scratch in dense or random clutter arrangements, and acquire two teacher policies for dense and random arrangement tasks, respectively. We then collect data with visual observation from these two teachers and finally train a vision-based student policy via behavior cloning, which better facilitates real-world deployment.

dexterous hands and the complexity of cluttered scenes make this synergy challenging to learn. One approach to addressing this challenge is task decomposition. Jiang et al. [24] propose a three-stage framework where separate policies are trained for isolating, grasping, and lifting. While this decomposition simplifies learning by breaking the problem into manageable sub-tasks, it also limits the synergy between singulation and grasping. In particular, minimal singulation may often suffice for dexterous grasping, as illustrated in Fig. 1. Alternatively, curriculum learning [25] has proven effective in tackling complex tasks and has already been successfully applied to dexterous grasping policies [8], [26]. More recently, Bai et al. [27] utilize a dexterous hand to expose the target object from clutter for grasping, but their reward design primarily encourages palm movement during singulation. In contrast, we argue that finger dexterity should be leveraged throughout the singulation to improve efficiency. For instance, in a top-down grasp, the palm can remain above the target while the fingers actively separate surrounding objects until a feasible grasp is possible. Therefore, our work builds on this success by exploring curriculum learning to develop a unified policy that integrates both singulation and grasping with finger dexterity.

In this work, the framework enables a dexterous hand to efficiently retrieve the target object from tightly cluttered environments, as illustrated in Fig. 1. We develop a reinforcement learning (RL) framework to seamlessly integrate object singulation and grasping through a unified reward design. Due to the challenges of directly solving grasping tasks in densely cluttered environments, our method leverages clutter arrangement curriculum learning to progressively enhance the performance of the teacher policy in generated cluttered environments with increasing complexity in object quantity, types, and arrangements. Furthermore, through teacher-student policy distillation, we obtain a vision-based student policy that generalizes across diverse cluttered environments and can be deployed on a real-world robot.

- We develop a unified RL policy for dexterous object singulation and grasping, enabling dexterous hands to effectively and efficiently retrieve objects in tightly cluttered environments.
- We incorporate clutter arrangement curriculum learning to improve policy performance across various cluttered scenes and employ policy distillation to obtain a vision-based policy suitable for real-world deployment.
- We design a set of cluttered grasping tasks and experiments with varying difficulty levels and conduct extensive experiments to demonstrate the effectiveness and efficiency of our proposed DexSinGrasp, and to verify the significance of finger dexterity.

## II. RELATED WORK

**Dexterous Grasping Policy.** Dexterous grasping with high degrees of freedom (DoF) presents significant learning challenges. To address these challenges, some approaches decompose grasping into hierarchical levels and use large grasping dataset to train high-level grasp proposal module [4]–[7] in a supervised setting, followed by motion planning or low-level policies. Modern approaches on dexterous grasping benefit from RL and massive parallel simulations for dynamic hand–object interactions. Building on this foundation, others focus on obtaining deployable vision-based policies for real-world applications through policy distillation [8], [9]. Researchers have also explored population-based training for reward-scale tuning, which can significantly reduce the cost of manual hyperparameter tuning [28]. Recently, dexterous grasping in cluttered environments has gained increasing attention. Through encouraging visual exposure of the target object [10], and extracting scene information from segmented point clouds [10] or segmented images [11], grasping in relatively loose clutter conditions is successfully achieved. However, rather than solely focusing on grasping, our approach learns to perform object singulation for grasping, leading to a sig-

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

nificantly higher success rate, especially in tightly cluttered environments.

**Grasping in Object Clutters.** Prior studies on robotic grasping in cluttered environments with grippers have explored various approaches to address the challenges such as occlusions [15]–[17] and limited grasping space [18]–[22], [29]. A common requirement in these scenarios is decluttering or singulation [12]–[14], [23], [30]–[37], which isolates the target object to facilitate successful grasping. Zeng *et al.* [12] utilized two separate networks to learn the synergy between pushing and grasping with a parallel gripper, inspiring subsequent studies on the adversarial training of these two networks [13], [14]. Chen *et al.* [38] optimize contact and object trajectories and solve hand motion via inverse kinematics in reverse order, forming an open-loop control strategy, while our learning-based approach achieves adaptive closed-loop control, making it more effective in dynamic and contact-rich tasks. Due to large space needed for palm-driven pushing and limited efficiency of parallel grippers, SOPE [24] uses fingers of a dexterous hand to isolate the target object as one of the pre-defined stages. However, our method uses clutter arrangement curriculum learning to obtain a synergistic dexterous singulation that greatly improves grasping efficiency. Recent work [27] has explored using dexterous hands to expose target objects from clutter for grasping, but primarily relies on palm-driven motion during singulation. In contrast, we highlight the importance of using finger dexterity throughout the singulation to improve efficiency through our unified reward design.

**Curriculum Learning.** Curriculum learning has been proven to improve both task success rates and generalization [25]. In challenging dexterous grasping tasks, it has been used to design training curricula based on object geometry [8], [26], significantly enhancing grasping generalization. In contrast, we use clutter arrangement curriculum learning to start from single object grasping to structured clutter distributions, thus reducing the difficulty of direct learning in fully compact or random clutter and leading to improved success rates and better generalization of the teacher policy.

### III. METHOD

**Problem Formulation** We first define a *Singulation-Necessary Clutter* (SNC) as a clutter configuration in which the target object is initially ungraspable and singulation is required before any feasible grasp can be executed.

To construct SNC scenes in a controlled and reproducible way, we impose the following three sufficient conditions: (1) The target object is fully enclosed by non-target objects and located at the geometric center of the clutter; (2) The horizontal distance between the target and each surrounding object is smaller than the fingertip size; (3) The target and surrounding objects have identical height.

These constraints ensure that direct grasping is physically infeasible, and that singulation is required to create graspable space. Based on the conditions, we consider a 2D cluttered scene on a horizontal tabletop, where the target object lies in the same plane as surrounding distractor objects, with no vertical occlusion. The densely cluttered setting mentioned in

the Sec. I serves as a concrete example of SNC. We further study two representative types of SNC:

**Dense Arrangement.** Objects are tightly packed with minimal spatial variability due to strong geometric constraints.

**Random Arrangement.** Objects are more loosely placed with slightly increased inter-object distances, resulting in more diverse spatial configurations.

Throughout training, we use cuboid-shaped objects as the standard shape for training. This choice facilitates the design of our curriculum learning scheme by providing consistent and controllable geometric structure. Moreover, we empirically find that policies trained on cuboids exhibit robust generalization to unseen shapes during later evaluations.

We then formulate dexterous grasping and object singulation as a RL task. Specifically, we consider a tabletop scene with the target object  $b^{\text{target}}$  and  $n$  surrounding objects  $\{b_i\}_{i=1}^n$ . To overcome this challenge, we train a policy to singulate the target object from its surroundings, thereby creating sufficient space for inserting fingers during dexterous grasping. Once singulation and grasping are complete, the target object is lifted to a goal position above the table. As a solution, we propose DexSinGrasp, a unified policy that jointly optimizes dexterous object singulation and grasping using RL.

a) *Observation Space:* The observation space of this singulation and grasping is defined as

$$s_t \triangleq [s_t^R, a_{t-1}, s_t^O, d_t^{HO}, T_t, d_t^S] \in \mathbb{R}^{168}, \quad (1)$$

where the proprioceptive robot state  $s_t^R \in \mathbb{R}^{72}$  includes the wrist pose as well as joint positions, velocities, and forces for each finger and wrist dummy joints (All joint-related dimensions hereafter are based on the Leap Hand [39] and omitted for brevity); the action  $a_{t-1}$  at the previous time step will be discussed later; the object state  $s_t^O \in \mathbb{R}^{16}$  consists of the object’s position and quaternion, linear and angular velocity, and object-hand position difference; the hand-object distances  $d_t^{HO} \in \mathbb{R}^{21}$  present the minimum distances between each hand links and points on the object; the time encoding  $T_t \in \mathbb{R}^{29}$  encodes the current time along with a sine-cosine time embedding. The singulation distance  $d_t^S \in \mathbb{R}^8$  presents the distances between the target object and surrounding objects, indicating the level of enclosure within the clutter. If the number of surrounding objects satisfies  $n < 8$ , the corresponding dimensions are padded with 0.

b) *Action Space:* The action space  $a_t \triangleq [a_t^P, a_t^F] \in \mathbb{R}^{22}$  includes palm delta pose  $a_t^P \in \mathbb{R}^6$  and linearly smoothed finger joint positions  $a_t^F := \lambda a_t^F + (1 - \lambda)a_{t-1}^F \in \mathbb{R}^{16}$  for each finger.

The reward function, as a crucial component of our unified policy, will be discussed in detail in Sec. III-A.

**Overview.** As illustrated in Fig. 2, we train a unified policy for dexterous object singulation and grasping through a structured learning framework, and adopt teacher-student policy distillation for real-world deployment. First, we introduce a unified reward design (Sec. III-A) that seamlessly integrates singulation and grasping into a single objective, enabling more efficient policy learning. Then, to improve learning efficiency in SNCs, we adopt Clutter Arrangement Curriculum Learning (Sec. III-B) to progressively train state-based teacher policies.

**IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.**

Finally, we employ Teacher-Student Policy Distillation (Sec. III-C) to transfer knowledge from the teacher policies to a vision-based student policy, allowing deployment on a real robot by mapping high-dimensional visual observations to effective actions.

### A. Unifying Dexterous Object Singulation and Grasping

Training a policy for dexterous grasping from scratch in SNCs poses significant challenges due to the high-dimensional action space and complex contact interactions. A naive approach trains a policy with rewards only for successful grasps in SNCs, but this suffers from low sample efficiency due to the need to learn singulation and grasping simultaneously. Alternatively, a two-stage method trains separate policies for singulation and grasping, but it leads to longer execution times and suboptimal transitions, as it fails to leverage their synergy.

To address these issues, we propose a unified reward design that seamlessly integrates dexterous object singulation and grasping into a single learning objective. The piece-wise reward function is defined as

$$r_t = \begin{cases} r_t^P + r_t^J + r_t^S, & \text{if } d_t^P \geq 0.06 \text{ or } d_t^J \geq 0.2, \\ r_t^P + r_t^J + r_t^F + r_t^L \\ \quad + r_t^G + r_t^S + r_t^B, & \text{if } d_t^P < 0.06 \text{ and } d_t^J < 0.2, \end{cases} \quad (2)$$

Detailed definitions of these rewards can be found in Tab. I, where the subscript  $t$  is omitted for simplicity. The rewards are defined as follows:  $r_t^P$  encourages the palm to approach the target object;  $r_t^J$  and  $r_t^F$  promote grasp formation through joint and fingertip contacts;  $r_t^L$  encourages lifting once contact is established;  $r_t^S$  drives separation of the target from obstacles;  $r_t^G$  guides motion toward the goal position; and  $r_t^B$  provides a bonus for successful singulation and grasping.  $d_t^P$  and  $d_t^J$  denote the minimum distances from the palm and hand links to the target, respectively.  $\{p_i^{\text{target}}\}_{i=1}^{N_p}$  are the target surface points;  $p^{\text{palm}}$ ,  $\{p_j^{\text{link}}\}_{j=1}^m$ , and  $\{p_j^{\text{fingertip}}\}_{j=1}^h$  denote the palm, link, and fingertip positions;  $p^{\text{goal}}$  and  $p^{\text{target}}$  are the goal and current positions of the target object;  $\{p_i\}_{i=1}^n$  are the obstacle positions; and  $a^{Pz}$  represents palm translation along  $+z$ , corresponding to the lifting motion.

The reward function consists of two main components: an approach reward that encourages the hand to move toward the target object and a lifting reward that promotes object elevation after contact is established. To further facilitate singulation, the singulation reward  $r_t^S$  is incorporated into both components, incentivizing the hand to separate the target object from surrounding obstacles. The transition between these two reward stages is achieved by a contact criterion specified by  $d_t^P < 0.06$  and  $d_t^J < 0.2$ . These values are selected such that the palm and fingers are close enough for grasping. The reward transitions from approaching to grasping, ensuring a smooth progression toward a successful grasp.

Despite this unified learning framework, training remains highly challenging due to the increasing complexity of clutters. To further improve learning efficiency and policy generalization, we introduce Clutter Arrangement Curriculum Learning, which progressively increases clutter complexity.

### B. Clutter Arrangement Curriculum Learning

To ensure successful and efficient object singulation and grasping in SNCs, we begin by training the teacher policy with privileged information from the simulation and adopt clutter arrangement curriculum learning, allowing our teacher policy to progressively improve as object diversity and spatial complexity increase. Using Proximal Policy Optimization (PPO) [40], we optimize the policy to maximize the cumulative discounted reward  $E[\sum_{t=1}^T \gamma^{t-1} r_t]$ , enabling effective RL.

1) *Clutter Generation*: To facilitate the training and testing of policies across various object configurations, we introduce a clutter generation module designed to create diverse object-based tasks. Our clutters consist of cuboid-shaped objects with varying quantities (from 4 to 8) and shapes (1×1, 1×2, and 1×3 cuboids). The generation for two types of clutters are introduced below, as shown in Fig. 3.

**Dense Arrangements Generation.** This type of task arranges different quantities of the surrounding 1×1 cuboids densely near the target object to create an extreme scenario that challenges the singulation and grasping policies under dense and narrow conditions.

**Random Arrangements Generation.** This type of task arranges objects of different quantities and shapes randomly around the target object for grasping, mainly to test the generalization of the singulation and grasping policies.

For simplicity, we use D/R- $n$  to denote task setting with  $n$  objects for dense (D) or random (R) arrangements, such as D-8. In our curriculum, only D-8 and R-8 are SNC tasks, while D-4/6 and R-4/6 are intermediate stages to aid learning and are not necessarily SNCs.

2) *Clutter Arrangement Curriculum Learning*: Since it is challenging to learn object singulation and grasping with dexterous hands in compact or diverse environments, such as the D-8 or R-8 tasks, we design clutter arrangement curriculum to gradually increase the object diversity and spatial complexity. We begin by training a grasping policy designed exclusively for Single-Object (SO) scenarios, where the objective is to grasp a single cuboid. Based on this initial policy, we continuously follow the curriculum and train on increasingly complex singulation and grasping tasks. Specifically, we first train on dense arrangements with D-4, then D-6, and finally D-8 tasks. We then use the expert obtained from the D-8 training as the starting point for training on random arrangements with R-4, R-6, and R-8 tasks. At the end of the training process, we extract the final policies from the D-8 training and the R-8 training as two teacher policies.

### C. Teacher-Student Policy Distillation

Since privileged observations—such as object states and singulation distances—are difficult to obtain in the real world, and some proprioceptive data, like finger-joint forces, are limited by hardware constraints, we learn a vision-based student policy to ensure feasible real-world deployment.

1) *Data Collection for Distillation*: The data collection phase involves preparing training data using two distinct teacher policies: the dense-clutter teacher policy for the D-4/6/8 tasks and the random-clutter teacher policy for the R-4/6/8 tasks, respectively. In total, 1000 episodes of observation

TABLE I  
 REWARD-RELATED TERMS

Term	Equation	Explanation
$d^P$	$\min_{i=1}^p \ p^{\text{palm}} - p_i^{\text{target}}\ _2$	Minimized distance of palm to target object.
$r^P$	$-2.0 \times d^P$	Palm reward to encourage hand approaching the target object.
$d^J$	$\sum_{j=1}^m \min_{i=1}^p \ p_j^{\text{link}} - p_i^{\text{target}}\ _2$	Minimized summed distance of hand links to target object.
$r^J$	$-d^J$	Joint reward to encourage the finger to grasp the target object.
$r^F$	$-\sum_{j=1}^h \min_{i=1}^p \ p_j^{\text{fingertip}} - p_i^{\text{target}}\ _2$	Fingertip reward to encourage fingertip to grasp the target object.
$r^L$	$0.2 + 0.6 \times a^{Pz}$	Lifting reward to encourage lifting actions.
$r^G$	$0.9 - 2.5 \times \ p^{\text{goal}} - p^{\text{target}}\ _2$	Goal reward to encourage target object to approach the goal.
$r^S$	$50 \times \min_{i=1}^n \ p^{\text{target}} - p_i\ _2$	Singulation reward for synergistic dexterous singulation.
$r^B$	$(1 + 10 \times \ p^{\text{goal}} - p^{\text{target}}\ _2)^{-1}$	Bonus term to target object to approach the goal.

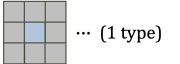
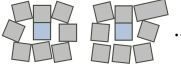
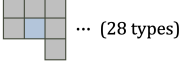
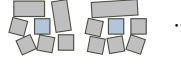
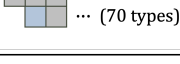

Obstacle Num.	Dense Arrangements	Random Arrangements ( $\infty$ types)
8	 ... (1 type)	 ... various pose
6	 ... (28 types)	 ... various shapes
4	 ... (70 types)	 ...

Fig. 3. Dense and random arrangement settings. We introduce a cluttered environment generation module to create diverse object settings.

and action data, along with scene point cloud, are prepared as  $\{(s_t^V, a_t)\}$ , where  $s_t^V$  will be discussed later. The dataset is structured so that the D-4 and R-4 tasks each account for 10% of the total, while the D-6, D-8, R-6, and R-8 tasks each contribute 20%. This balanced distribution better captures SNC tasks and ensures a comprehensive representation of varying task complexities, which is critical for effective policy distillation and better generalizability.

2) *Vision-Based Policy Behavior Cloning*: Our vision-based student policy uses scene point cloud instead of object poses, which cannot be accurately obtained in such heavily occluded SNCs. We specifically use behavior cloning to train the student policy, with data collected from two teachers. The different visual observation  $s_t^V$  for the vision-based student policy is defined as

$$s_t^V \triangleq [s_t^R, a_{t-1}, s_t^{O'}, d_t^{HO}, T_t, v_t] \in \mathbb{R}^{275} \quad (3)$$

where singulation distance  $d_t^S$  is removed from  $s_t$  in the state-based teacher policy, and object state  $s_t^O \in \mathbb{R}^{16}$  is substituted with center position of scene point cloud  $s_t^{O'} \in \mathbb{R}^3$ . Moreover, the vision-based policy includes the visual features  $v_t \in \mathbb{R}^{128}$  encoded from the scene point cloud using a pre-trained point cloud encoder from UniGraspTransformer [9], with the encoder weights frozen during the policy distillation.

#### IV. EXPERIMENT

In this section, we conduct comprehensive experiments to evaluate our proposed method, DexSinGrasp, in both simulation and real-world tasks. Through these experiments, we aim to address the following key questions: (1) How effective and efficient is our method for grasping in SNCs? (2) How does our method generalize to different spatial arrangements?

(3) How effective is our clutter arrangement curriculum learning? (4) How does our method generalize to different object geometries? (5) How does finger dexterity affect singulation and grasping performance in SNCs? (6) How does our method perform on real-world tasks?

##### A. Baselines

To evaluate our approach, we compare our proposed approach with two baseline methods:

**GraspReward-Only Method.** In this baseline, pure dexterous grasping is conducted without singulation. This baseline is trained from scratch in a single target object environment with the singulation reward set to zero [8], [9].

**Multi-Stage Method.** This baseline is a two-stage framework where separately trained singulation and grasping policies operate in sequence as adopted by SOPE [24]. We train the separate singulation policy without the grasping reward stage as mentioned in Sec. III-A. The singulation stage is switched to the grasping stage when  $\sum_{i=1}^n \|p^{\text{target}} - p_i\|_2/n > 0.16$ , where  $n$  is the number of surrounding objects.

##### B. Evaluation Metrics

To evaluate the performance of our approach, we introduce two key metrics that assess both effectiveness and efficiency in dexterous object singulation and grasping.

**Success Rate.** The proportion of trials where the target object reaches the predefined goal position above the table, defined as  $\|p^{\text{goal}} - p^{\text{target}}\|_2 < 0.05$ . Denoted as SR.

**Average Steps.** The average number of steps to singulate and grasp the target object, denoted as AS. Failed episodes are counted as 300 steps (the maximum per episode).

The best results in each table are highlighted in bold. Parenthesized values show the difference from our method.

##### C. Implementation Details

We use Isaac Gym for clutter arrangement curriculum learning. For each D/R- $n$  task, we used 1000 simulated environments and trained the PPO policy network over 10K iterations with a learning rate of  $3 \times 10^{-4}$ . We then evaluated and selected the best-performing iteration as the policy for the next-stage clutter arrangement curriculum learning. The student policy is trained over 200 epochs with a batch size of 12 trajectories, each composed of 300 steps of recorded

TABLE II  
EVALUATION ON DENSE ARRANGEMENTS.

Method		SR(%) $\uparrow$				AS $\downarrow$			
		D-4	D-6	D-8	Avg.	D-4	D-6	D-8	Avg.
State-Based	GraspReward-Only	66	40	10	39	202	252	292	249
	Multi-Stage	77	76	64	72	199	210	235	215
	Ours (Teacher)	<b>98</b>	<b>99</b>	<b>97</b>	<b>98</b>	<b>100</b>	<b>115</b>	<b>138</b>	<b>118</b>
Vision-Based	Ours (Student)	90	92	84	89	122	123	159	135

TABLE III  
EVALUATION ON RANDOM ARRANGEMENTS.

Method		SR(%) $\uparrow$				AS $\downarrow$			
		R-4	R-6	R-8	Avg.	R-4	R-6	R-8	Avg.
State-Based	GraspReward-Only	73	61	33	56	179	207	261	216
	Multi-Stage	88	72	78	79	156	170	178	168
	Ours (Teacher)	<b>97</b>	<b>96</b>	<b>94</b>	<b>96</b>	<b>94</b>	<b>95</b>	<b>103</b>	<b>97</b>
Vision-Based	Ours (Student)	91	86	88	88	101	113	114	109

simulation data and a learning rate of  $1 \times 10^{-4}$ . In both simulation and real-world experiments, hand rotation is disabled, restricting the end-effector to translation and finger motion. This encourages dexterous finger use for singulation and grasping, prevents unrealistic floating-hand movements that may cause arm–table collisions on the xArm, and eases sim-to-real transfer. We use Isaac Gym’s default friction (1.0) for experiments without domain randomization, and maintain small inter-object gaps (1–1.5 cm) to enable horizontal contact forces and avoid self-locking.

#### D. Main Results and Analysis

We tested the dense-clutter teacher policy and the student policy on D-4/6/8 tasks and compared their performance with all the baselines. All methods were evaluated for 100 times.

Based on the results presented in Tab. II, the multi-stage singulation policy demonstrates a higher success rate than the GraspReward-Only baseline, suggesting that the singulation stage plays a positive role in task performance. Our dense-clutter teacher policy achieves a significantly higher average success rate of 98%, with a substantially lower AS compared to the Multi-Stage policy. While the student policy has a lower SR than the teacher, it still outperforms the baselines and maintains a comparable AS. Our results show that combining separate singulation and grasping policies increases action steps and reduces efficiency, whereas our unified policy balances effectiveness and efficiency for grasping in cluttered environments including SNCs, thus addressing Q1.

In response to Q2, we evaluate the random-clutter teacher policy and the student policy on R-4/6/8 tasks with different object arrangements to test the spatial generalization ability of our policy. In Tab. III, we first observe that baselines perform better in random than in dense arrangements, where looser gaps allow direct grasps. Our teacher policy can achieve an average SR of 96% across all tasks. While the student policy achieves a lower SR than the teacher policy, it still outperforms all the baselines in both effectiveness and efficiency.

During training and testing, we found the policy learned several singulation patterns, including finger flickering and

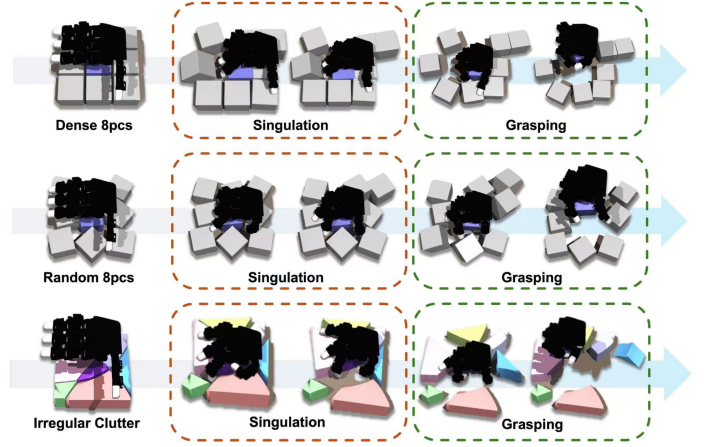


Fig. 4. Qualitative results on object singulation and grasping in simulation.

TABLE IV  
EVALUATION ON DIFFERENT CURRICULUMS.

Curriculum	SR(%) $\uparrow$ on D- $n$ tasks				SR(%) $\uparrow$ on R- $n$ tasks			
	D-4	D-6	D-8	Avg.	R-4	R-6	R-8	Avg.
Training from scratch	90 (8 $\downarrow$ )	75 (17 $\downarrow$ )	97 (0)	87 (9 $\downarrow$ )	74 (12 $\downarrow$ )	26 (70 $\downarrow$ )	0 (94 $\downarrow$ )	33 (62 $\downarrow$ )
Random-to-dense	96 (2 $\downarrow$ )	87 (5 $\downarrow$ )	81 (16 $\downarrow$ )	88 (8 $\downarrow$ )	<b>97 (1<math>\uparrow</math>)</b>	92 (4 $\downarrow$ )	<b>97 (3<math>\uparrow</math>)</b>	<b>95 (0)</b>
Dense-to-random (Ours)	<b>98</b>	<b>92</b>	<b>97</b>	<b>96</b>	96	<b>96</b>	94	<b>95</b>

finger-palm vibration to displace and nudge surrounding objects, effectively singulating targets from clutters, as shown in Fig. 4. These patterns rely heavily on the finger dexterity, and a more detailed discussion is provided in Sec. IV-G.

#### E. Clutter Arrangement Curriculum Learning Analysis

The clutter arrangement curriculum learning process is designed to enhance success rates in increasingly complex scenes. We evaluate each policy trained on D/R- $n$  tasks under various curriculum directions—dense to random (SO, D-4, D-6, D-8, R-4, R-6, R-8), random to dense (SO, R-4, R-6, R-8, D-4, D-6, D-8), and no curriculum (training each D/R- $n$  task from scratch)—as shown in Tab. IV. For the two curricula, we use the best-performing checkpoints from trained 10k iterations for each setting. The curriculum with dense-to-random direction consistently yields the best performance across tasks. The results indicate that with the progression of the curriculum, the teacher policy demonstrate greater accuracy and efficiency, addressing Q3.

#### F. Generalization to irregular clutters

To test whether our policy generalizes to geometries beyond the cuboid clutter, we evaluate its performance on randomly generated irregular clutters that are tightly packed—a particularly challenging case for SNCs. Each clutter is constructed by dividing a 24 $\times$ 24 $\times$ 8 cm cuboid using three random curves, ensuring three intersections to form seven distinct pieces, shown in Fig. 5. We fine-tuned our policy from D-8 expert on 200 such irregular clutters and tested it on 50 unseen instances together with previous baselines. As shown in Table V, our method achieves good performance, and therefore exhibits geometric generalization, in response to Q4.

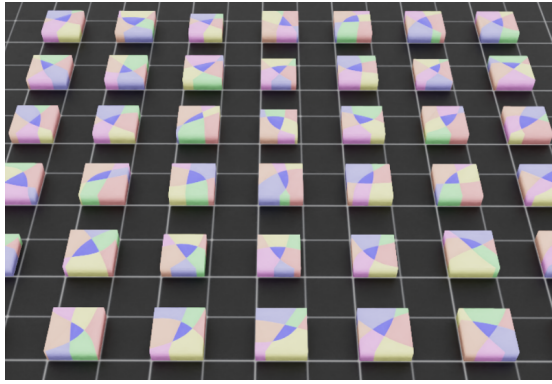


Fig. 5. Irregular Clutter. The target object is marked as purple in the center.

TABLE V  
EVALUATION ON IRREGULAR CLUTTERS

Method		SR(%) $\uparrow$	AS $\downarrow$
State-Based	GraspReward-Only	26	271
	Multi-Stage	53	245
	Ours (Teacher, D-8 Finetuned)	<b>81</b>	<b>211</b>
Vision-Based	Ours (Student, D-8 Finetuned)	60	231

TABLE VI  
EVALUATION OF TEACHER POLICY WITH VARYING HAND DEXTERITY.

SR(%) $\uparrow$	D-4	D-6	D-8	Irregular
Low-DoF LEAP Hand	82 (16 $\downarrow$ )	70 (19 $\downarrow$ )	71 (16 $\downarrow$ )	24 (57 $\downarrow$ )
Mid-DoF LEAP Hand	91 (7 $\downarrow$ )	90 (9 $\downarrow$ )	87 (10 $\downarrow$ )	47 (34 $\downarrow$ )
Full-DoF LEAP Hand (Ours)	<b>98</b>	<b>99</b>	<b>97</b>	<b>81</b>
AS $\downarrow$	D-4	D-6	D-8	Irregular
Low-DoF LEAP Hand	119 (19 $\uparrow$ )	153 (38 $\uparrow$ )	166 (28 $\uparrow$ )	272 (61 $\uparrow$ )
Mid-DoF LEAP Hand	113 (13 $\uparrow$ )	135 (20 $\uparrow$ )	165 (27 $\uparrow$ )	245 (34 $\uparrow$ )
Full-DoF LEAP Hand (Ours)	<b>100</b>	<b>115</b>	<b>138</b>	<b>211</b>

### G. Discussion of Dexterity

In singulation and grasping tasks for SNCs—where limited space restricts finger articulation—the use of high-degree-of-freedom (DoF) hands becomes essential. We evaluate three variants of the LEAP Hand with varying degrees of dexterity:

- **Low-DoF:** The index, middle, and ring fingers are underactuated with fixed abduction–adduction, where each finger’s 3 joints are coupled into 1 DoF. The thumb retains 3 DoFs. (Total: 6 DoF)
- **Mid-DoF:** Index, middle, and ring fingers have independently actuated 3 joints with fixed abduction–adduction. The thumb is fully actuated with 4 DoFs. (total: 13 DoF)
- **Full-DoF:** A fully actuated LEAP Hand with control over all lateral and longitudinal joints of every digit. The thumb is fully actuated with 4 DoFs. (total: 16 DoF)

To quantify the impact of dexterity, we benchmarked these 3 embodiments across D-4/6/8 tasks and irregular clutters. The results, summarized in Table VI, show significant improvement with increased hand dexterity, in response to Q5. High finger dexterity enables complex contact interactions, which are critical in SNCs where efficient singulation requires delicate finger skills instead of coarse palm movement.

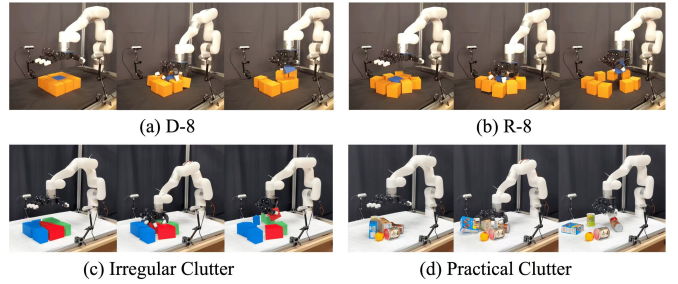


Fig. 6. Real-world experiments for D-8, R-8, irregular and practical clutters.

TABLE VII  
EVALUATION OF VISION-BASED STUDENT POLICY IN REAL-WORLD.

Method	D-8	R-8	Irregular Clutter	Practical Clutter
GraspReward-Only	3/20	2/20	0/20	17/100
Multi-Stage	7/20	4/20	3/20	22/100
Ours	<b>15/20</b>	<b>12/20</b>	<b>12/20</b>	<b>61/100</b>

### H. Real-World Experiments

We conduct real-world experiments using a uFactory xArm6 robot equipped with a LEAP Hand and two side-view RealSense D435 RGB-D cameras, as shown in Fig. 6. The synchronized RGB-D data from both cameras are converted into point clouds, transformed to the world frame using calibrated extrinsics, and fused into a single scene point cloud, which is further refined using the Iterative Closest Point (ICP) algorithm. A predefined crop around the clutter center removes background noise, and the fused cloud is downsampled to 1,024 points before being fed into the vision-based policy.

We accomplish singulation and grasping for D-8, R-8, irregular clutter and practical clutter consisting of previously unseen practical objects randomly arranged in the real-world environment and conduct 20 trials for each setting (5 settings for practical clutter). To further improve sim-to-real transfer, we additionally retrained the teacher policy with domain randomization on hand joint damping, stiffness, and friction (hand friction varies from 0.7 to 1.8 and table friction varies from 0.4 to 1.3 in Isaac Gym). As shown in Tab. VII, the GraspReward-Only baseline struggles in cluttered environments due to the lack of singulation. In contrast, the Multi-Stage method suffers from the sensitivity to sim-to-real discrepancies, often causing failures when singulation generates unseen configurations. Notably, in practical clutter, our policy achieves a 61% success rate in a zero-shot setting, demonstrating strong generalization across both shape and spatial configurations, even though it was trained solely on a cuboid-shaped curriculum.

### V. CONCLUSION

Our proposed approach demonstrates that a unified RL framework can effectively integrate object singulation and grasping in SNCs using dexterous hands. By unifying singulation and grasping, our method not only achieves higher grasp success rates and improved efficiency compared to traditional multi-stage approaches but also enables subtle maneuvers directly with finger dexterity. The combination of clutter arrangement curriculum learning and policy distillation further enhances the generalization of the vision-based policy,

**IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.**

ensuring successful skill transfer from simulation to the real world, and from seen blocks to unseen objects. Additionally, the introduction of various cluttered grasping environments provides a comprehensive testbed for evaluating performance across various clutter configurations. Interestingly, a curriculum trained solely on cuboid objects exhibits strong zero-shot generalization to diverse practical clutters, suggesting that spatial relationships—rather than geometric variations alone—are critical to grasping in densely cluttered environments and validating our curriculum design.

Future work could extend to more complex clutters, broader object shapes, and improved real-world robustness, while further exploring how spatial diversity in clutter arrangement curricula enhance generalization to unseen clutters.

## REFERENCES

- [1] M. Q. Mohammed, L. C. Kwek, S. C. Chua, A. Al-Dhaqm, S. Nahavandi, T. A. E. Eisa, M. F. Miskon, M. N. Al-Mhiqani, A. Ali, M. Abaker *et al.*, “Review of learning-based robotic manipulation in cluttered environments,” *Sensors*, vol. 22, no. 20, p. 7938, 2022.
- [2] B. Wen, W. Lian, K. Bekris, and S. Schaal, “Catgrasp: Learning category-level task-relevant grasping in clutter from simulation,” in *2022 International Conference on Robotics and Automation (ICRA)*, 2022.
- [3] M. Laskey, J. Lee, C. Chuck, D. Gealy, W. Hsieh, F. T. Pokorny, A. D. Dragan, and K. Goldberg, “Robot grasping in clutter: Using a hierarchy of supervisors for learning from demonstrations,” in *2016 IEEE International Conference on Automation Science and Engineering (CASE)*, 2016, pp. 827–834.
- [4] S. Christen, M. Kocabas, E. Aksan, J. Hwangbo, J. Song, and O. Hilliges, “D-Grasp: Physically Plausible Dynamic Grasp Synthesis for Hand-Object Interactions,” Mar. 2022.
- [5] H. Zhang, S. Christen, Z. Fan, O. Hilliges, and J. Song, “GraspXL: Generating Grasping Motions for Diverse Objects at Scale,” Jul. 2024.
- [6] Z. Xu, C. Gao, Z. Liu, G. Yang, C. Tie, H. Zheng, H. Zhou, W. Peng, D. Wang, T. Hu *et al.*, “Manifoundation model for general-purpose robotic manipulation of contact synthesis with arbitrary objects and robots,” in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 10905–10912.
- [7] Z. Wei, Z. Xu, J. Guo, Y. Hou, C. Gao, Z. Cai, J. Luo, and L. Shao, “D (r, o) grasp: A unified representation of robot and object interaction for cross-embodiment dexterous grasping,” *arXiv preprint arXiv:2410.01702*, 2024.
- [8] W. Wan, H. Geng, Y. Liu, Z. Shan, Y. Yang, L. Yi, and H. Wang, “UniDexGrasp++: Improving Dexterous Grasping Policy Learning via Geometry-aware Curriculum and Iterative Generalist-Specialist Learning,” Apr. 2023.
- [9] W. Wang, F. Wei, L. Zhou, X. Chen, L. Luo, X. Yi, Y. Zhang, Y. Liang, C. Xu, Y. Lu, J. Yang, and B. Guo, “UniGraspTransformer: Simplified Policy Distillation for Scalable Dexterous Robotic Grasping,” Dec. 2024.
- [10] M. Mosbach and S. Behnke, “Grasp anything: Combining teacher-augmented policy gradient learning with instance segmentation to grasp arbitrary objects,” in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 7515–7521.
- [11] J. Lundell, F. Verdoja, and V. Kyrki, “Ddgc: Generative deep dexterous grasping in clutter,” *IEEE Robotics and Automation Letters*, 2021.
- [12] 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,” Sep. 2018.
- [13] 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, Oct. 2021.
- [14] Y. Wang, K. Mokhtar, C. Heemskerck, 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)*, 2024, pp. 13840–13847.
- [15] A. Kurenkov, J. Taglic, R. Kulkarni, M. Dominguez-Kuhne, A. Garg, R. Martín-Martín, and S. Savarese, “Visuomotor mechanical search: Learning to retrieve target objects in clutter,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.
- [16] 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)*. Kyoto, Japan: IEEE, Oct. 2022, pp. 1388–1395.
- [17] H. Zhang, Y. Lu, C. Yu, D. Hsu, X. Lan, and N. Zheng, “INVIGORATE: Interactive Visual Grounding and Grasping in Clutter,” Jan. 2024.
- [18] 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)*, 2021, pp. 4620–4626.
- [19] B. Wu, I. Akinola, and P. K. Allen, “Pixel-attentive policy gradient for multi-fingered grasping in cluttered scenes,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 1789–1796.
- [20] X. Wang and Q. Xu, “Transferring grasping across grippers: Learning-optimization hybrid framework for generalized planar grasp generation,” *IEEE Transactions on Robotics*, vol. 40, pp. 3388–3405, 2024.
- [21] S. Song, A. Zeng, J. Lee, and T. Funkhouser, “Grasping in the wild: Learning 6dof closed-loop grasping from low-cost demonstrations,” *IEEE Robotics and Automation Letters*, 2020.
- [22] M. Breyer, J. J. Chung, L. Ott, R. Siegwart, and J. Nieto, “Volumetric grasping network: Real-time 6 dof grasp detection in clutter,” in *Proceedings of the 2020 Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, J. Kober, F. Ramos, and C. Tomlin, Eds., vol. 155. PMLR, 16–18 Nov 2021, pp. 1602–1611. [Online]. Available: <https://proceedings.mlr.press/v155/breyer21a.html>
- [23] I. Sarantopoulos, M. Kiatos, Z. Doulgeri, and S. Malassiotis, “Split deep q-learning for robust object singulation,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 6225–6231.
- [24] H. Jiang, Y. Wang, H. Zhou, and D. Seita, “Learning to singulate objects in packed environments using a dexterous hand,” *arXiv preprint arXiv:2409.00643*, 2024.
- [25] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, “Curriculum learning,” in *International Conference on Machine Learning*, 2009. [Online]. Available: <https://api.semanticscholar.org/CorpusID:873046>
- [26] Y. Xu, W. Wan, J. Zhang, H. Liu, Z. Shan, H. Shen, R. Wang, H. Geng, Y. Weng, J. Chen, T. Liu, L. Yi, and H. Wang, “UniDexGrasp: Universal Robotic Dexterous Grasping via Learning Diverse Proposal Generation and Goal-Conditioned Policy,” Mar. 2023.
- [27] F. Bai, Y. Li, J. Chu, T. Chou, R. Zhu, Y. Wen, Y. Yang, and Y. Chen, “Retrieval dexterity: Efficient object retrieval in clutters with dexterous hand,” *arXiv preprint arXiv:2502.18423*, 2025.
- [28] A. Petrenko, A. Allshire, G. State, A. Handa, and V. Makoviychuk, “Dexpbt: Scaling up dexterous manipulation for hand-arm systems with population based training,” 05 2023.
- [29] S. Chen, A. Wu, and K. Liu, “Synthesizing dexterous nonprehensile pregrasp for ungraspable objects,” 07 2023, pp. 1–10.
- [30] B. Huang, S. D. Han, A. Boularias, and J. Yu, “DIPN: Deep Interaction Prediction Network with Application to Clutter Removal,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021.
- [31] M. Kiatos, L. Koutras, I. Sarantopoulos, and Z. Doulgeri, “Learning a pre-grasp manipulation policy to effectively retrieve a target in dense clutter,” in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024, pp. 7543–7549.
- [32] B. Huang, S. D. Han, J. Yu, and A. Boularias, “Visual foresight trees for object retrieval from clutter with nonprehensile rearrangement,” *IEEE Robotics and Automation Letters*, vol. 7, no. 1, pp. 231–238, 2022.
- [33] C. Gao, Z. Li, H. Gao, and F. Chen, “Iterative interactive modeling for knotting plastic bags,” in *Conference on Robot Learning*. PMLR, 2023.
- [34] M. Kiatos, I. Sarantopoulos, L. Koutras, S. Malassiotis, and Z. Doulgeri, “Learning Push-Grasping in Dense Clutter,” *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 8783–8790, Oct. 2022.
- [35] Y. Chen, Z. Ju, and C. Yang, “Combining reinforcement learning and rule-based method to manipulate objects in clutter,” in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–6.
- [36] A. Zeng, S. Song, K.-T. Yu, E. Donlon, F. R. Hogan, M. Bauza, D. Ma, O. Taylor, M. Liu, E. Romo, N. Fazeli, F. Alet, N. C. Daffe, R. Holladay, I. Morona, P. Q. Nair, D. Green, I. Taylor, W. Liu, T. Funkhouser, and A. Rodriguez, “Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching,” *The International Journal of Robotics Research*, vol. 41, no. 7, pp. 690–705, 2022. [Online]. Available: <https://doi.org/10.1177/0278364919868017>
- [37] P. K. Murali, A. Dutta, M. Gentner, E. Burdet, R. Dahiya, and M. Kaboli, “Active visuo-tactile interactive robotic perception for accurate object pose estimation in dense clutter,” *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4686–4693, 2022.
- [38] S. Chen, A. Wu, and C. K. Liu, “Synthesizing dexterous nonprehensile pregrasp for ungraspable objects,” in *ACM SIGGRAPH 2023 Conference Proceedings*, 2023, pp. 1–10.
- [39] K. Shaw, A. Agarwal, and D. Pathak, “Leap hand: Low-cost, efficient, and anthropomorphic hand for robot learning,” *arXiv preprint arXiv:2309.06440*, 2023.
- [40] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” *arXiv preprint arXiv:1707.06347*, 2017.