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

Lixin Xu[1], Zixuan Liu[1], Zhewei Gui[1], Jingxiang Guo[1],
Zeyu Jiang[1], Zhixuan Xu[1], Chongkai Gao[1], Lin Shao[1]

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 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. 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 project website https://nus-lins-lab.github.io/dexsingweb/.

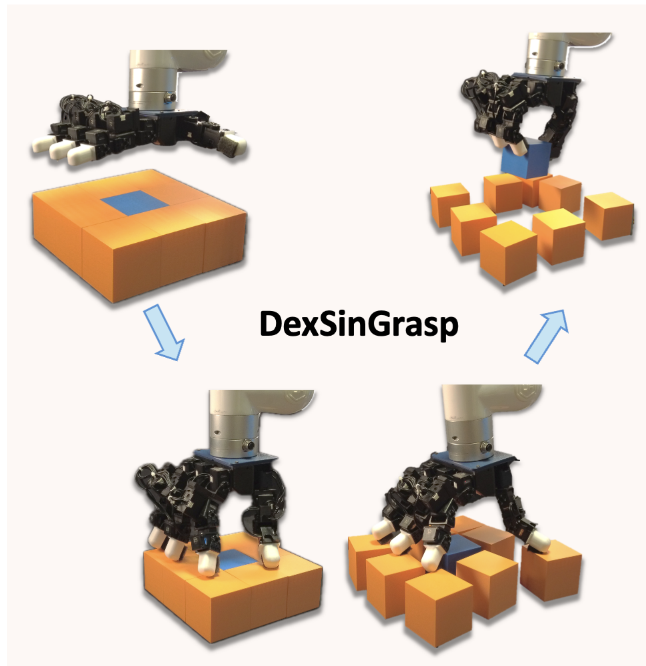

Fig. 1. We propose *DexSinGrasp* to learn a unified policy for dexterous object singulation and grasping in cluttered environments

I. Introduction

Grasping target objects in cluttered environments remains a significant challenge in robotics. Enabling robots to perform this task 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. Recent dexterous grasping approaches [10] 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 object-specific key points [11] to identify suitable grasping positions. However, due to the lack 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* 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. We 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 approach to object rearrangement 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 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.

Lixin Xu, Zixuan Liu, Zhewei Gui, Jingxiang Guo, Zeyu Jiang, Zhixuan Xu, Chongkai Gao, Lin Shao are with the School of Computing, National University of Singapore

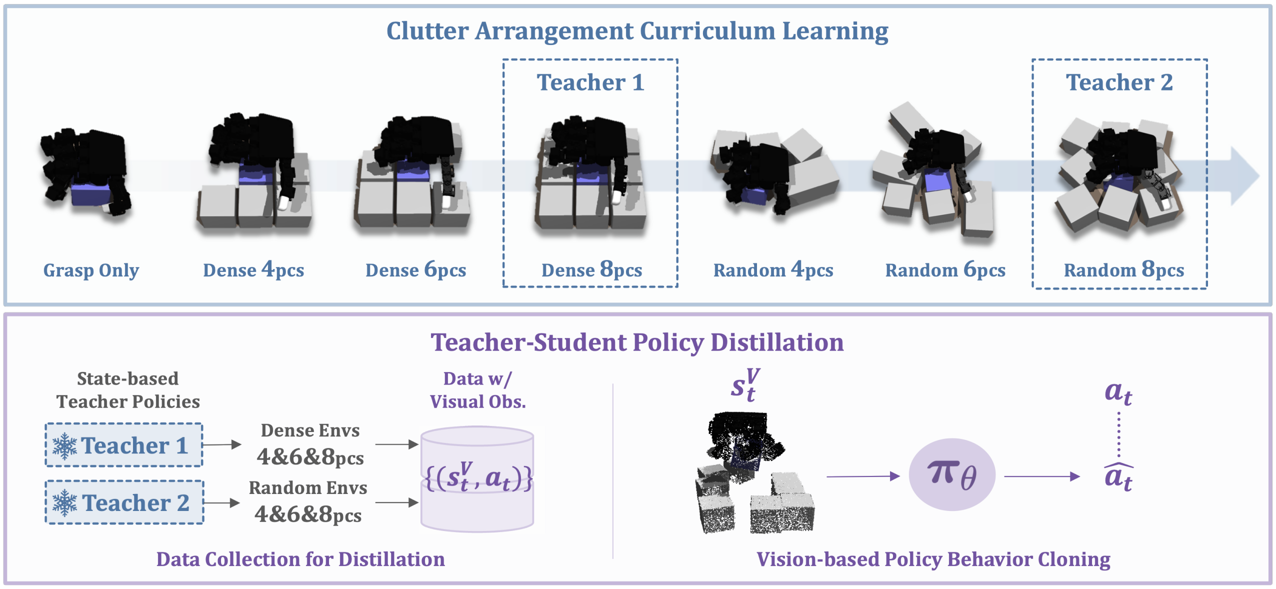

Fig. 2. Framework of *DexSinGrasp*. First, cluttered environments are generated for the following training. Curriculum learning is then leveraged to reduce the difficulty of training state-based teacher policies. Finally, the vision-based student policy is distilled from teacher policies for deployment on a real robot.

[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]. Our work builds on this success by exploring curriculum learning to develop a unified policy that integrates both singulation and grasping for dexterous manipulation.

In this work, we develop a reinforcement learning framework to train a unified policy that seamlessly integrates object singulation and grasping. This framework enables a dexterous hand to efficiently grasp target objects from tightly cluttered environments, as illustrated in Fig. 1. Due to the challenges of directly solving grasping tasks in general cluttered environments, our method leverages 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 singulation policy that generalizes across diverse cluttered environments and can be deployed on a real-world robot.

- We develop a unified reinforcement learning policy for dexterous object singulation and grasping, enabling dexterous hands to effectively and efficiently grasp objects in tightly cluttered environments.
- We incorporate curriculum learning to improve policy performance across various cluttered scenes and employ policy distillation to obtain a vision-based grasping 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*.

Our method leverages a unified reward design to transform target objects—initially ungraspable due to tight occlusions—into graspable states, ultimately completing the grasp.

II. Related Work

Dexterous grasping policy. Recent advancements in dexterous grasping policies leverage reinforcement learning (RL) and physics simulations to model dynamic hand-object interactions. Due to the learning challenges introduced by high DoF, some approaches decompose grasping into hierarchical levels or unify multiple grasping objectives [4]–[7] to improve scalability. In contrast, others employ curriculum learning or policy distillation [8], [9] to obtain deployable vision-based policies for real-world applications. Recently, dexterous grasping in cluttered environments has gained increasing attention. Through encouraging visual exposure of the target object [10] or penalizing collisions with obstacles [11], grasping in relatively loose clutter conditions is successfully achieved. Compared to these grasp-centric methods, our approach learns to perform synergistic singulation for grasping, leading to a significantly higher success rate due to the sufficient grasping space, 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 posed by occlusions [15]–[17] and limited grasping space [18]–[22]. A common requirement in these scenarios is decluttering or singulation [12]–[14], [23], [27]–[34], which isolates the

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	$0.2 + 0.6 \times a^{P_{t_z}}$	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 to 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.

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]. To overcome the dexterity constraints of grippers, SOPE [24] uses fingers to isolate the target object as one of the pre-defined stages. However, our method uses curriculum learning to obtain a synergistic dexterous singulation that efficiently contributes to grasping.

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, our approach designs to starting curriculum learning with structured clutter distributions, thus reducing the difficulty of direct learning in fully random clutter and leading to improved success rates and better generalization of teacher policies.

III. METHOD

Problem Formulation We formulate dexterous grasping and object singulation as a reinforcement learning task. Specifically, we consider a tabletop scene with the target object b^{target} and n surrounded objects $\{b_i\}_{i=1}^{n}$. This setup represents common cluttered scenarios and introduces significant grasping challenges due to the tight arrangement of surrounding objects. The target object is positioned at the center of the clutter, making direct grasping difficult. To overcome this challenge, we train robots to *singulate* the target object from its surroundings, thereby creating sufficient space for inserting fingers during dexterous grasping. In this work, we propose *DexSinGrasp*, a unified policy that jointly optimizes object singulation and dexterous grasping using reinforcement learning.

Overview. As depicted in Fig. 2, we train our unified policy for dexterous grasping and object singulation through (1) first leveraging curriculum learning to reduce the difficulty of training state-based teacher policies, as described in Sec. III-B; (2) then for deployment on a real robot, distilling knowledge from teacher policies to a vision-based student policy taking in high-dimensional visual observation to extract features and output action, as described in Sec. III-C.

A. Unifying Dexterous Object Singulation and Grasping

Training a policy for dexterous grasping from scratch in cluttered environments poses significant challenges due to the high-dimensional action space and complex contact interactions. The naive approach is to directly train a policy that only rewards successful grasps in cluttered environments. However, this often struggles with sample efficiency, as the policy must simultaneously learn to resolve contact occlusions and grasp effectively. Another alternative is a two-stage method, where a singulation policy is trained separately to first singulate the object before a grasping policy is trained. However, this approach is inefficient in execution, as it fails to fully exploit the synergy between singulation and grasping, leading to longer task completion time and suboptimal transitions between the two stages.

To overcome these limitations, we propose Clutter Arrangement Curriculum Learning, which progressively increases clutter complexity and allows our teacher policy to develop robust strategies in a structured manner. Our approach avoids the execution inefficiency of separate stage-wise training while improving sample efficiency compared to direct training. We compare our approach with the other two methods in II, demonstrating that Clutter Arrangement Curriculum Learning achieves higher efficiency in diverse environments.

B. Clutter Arrangement Curriculum Learning

Due to the challenge of training the unified policy dexterous grasping and object singulation from scratch in compact or diverse environments, we employ clutter arrangement curriculum learning to progressively improve our teacher policy's performance as the object diversity and spatial complexity increase.

1) Cluttered Environment Generation: To facilitate the training and testing of policies across various object configurations, we introduce a cluttered environment generation module designed to create diverse object-based tasks. Our cluttered environment primarily consists of block-shaped objects with varying quantities (from 4 to 8) and shapes (1×1, 1×2, and 1×3 blocks). Based on the degree of enclosure, the tasks are generally divided into two categories, as shown in Fig. 3.

Dense Arrangements. This type of task arranges different quantities of the surrounding 1x1 blocks densely near the target object to create an extreme scenario that challenges the singulation and grasping policies under dense and narrow conditions.

Random Arrangements. 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.

2) State-based Teacher Policy Learning: In order to train a teacher policy capable of achieving successful and efficient object singulation and grasping performances for a given D/R-n task setting, we start by training with some privileged information from the simulation.

Observation space. The observation space of our state-based teacher policy is defined as

$$s_t \triangleq \left[s_t^P, a_{t-1}, s_t^O, d_t^{HO}, T_t, d_t^S\right] \in \mathbb{R}^{168}, \tag{1}$$

where the proprioceptive state $s_t^P \in \mathbb{R}^{72}$ includes the wrist pose as well as joint positions, velocities, and forces for each finger and wrist dummy joints; 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 $n < 8$, the corresponding dimensions are padded with 0.

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

Rewards. The piece-wise reward function is defined as

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

Detailed definitions of these rewards can be found in Tab. I, where t subscript is omitted for simplicity. In the rewards, r_t^P encourages the hand palm to stay close to the target object; r_t^J and r_t^F both encourage the hand to grasp the target object; r_t^L encourage the hand to lift the target object once contact is made, r_t^S encourages separation of the target object from the obstacles, r_t^G encourages the hand to move to the goal position; r_t^B is a bonus term for a success singulation and grasping process; d_t^P is the minimum distance between hand palm and target object; d_t^J is the minimum distance between links and target object. $\{p_i^{\text{target}}\}_{i=1}^{p}$ are the positions of p points on the target object; p^{palm} is the palm position of the hand; $\{p_j^{\text{link}}\}_{j=1}^{m}$ are the positions of m links of the hand; $\left\{p_j^{\text{fingertip}}\right\}_{j=1}^{h}$ are the positions of h fingertips of the hand; p^{goal} and p^{target} are the goal position and current position for the target object; $\{p_i\}_{i=1}^{n}$ are the positions of n obstacles; $a^{P_{t_z}}$ refers to the palm translation in the $+z$ direction, which corresponds to the lifting motion of the target object.

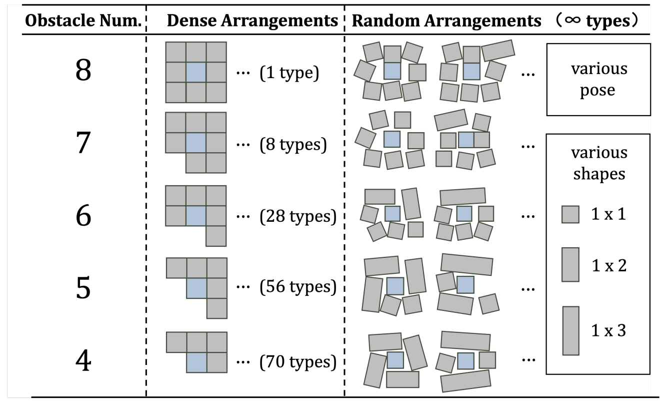

Fig. 3. Dense and random arrangement settings. We introduce a cluttered environment generation module to create diverse object settings, including different obstacle quantities from 0 to 8, dense and random arrangements, various poses, and block shapes. For simplicity, we only show obstacles with numbers 4 to 8.

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 $d_t^J < 0.2$. The reward transitions from approaching to grasping, ensuring a smooth progression toward a successful grasp. We use PPO [35] for reinforcement learning to maximize the cumulative discounted reward $E[\sum_{t=1}^{T} \gamma^{t-1} r_t]$.

3) Clutter Arrangement Curriculum Learning: Due to the challenge of training synergistic dexterous singulation from scratch in compact or diverse environments, such as the D-8 or R-8 tasks, we employ curriculum learning to progressively improve our teacher policy's performance as the object diversity and spatial complexity increase. We begin by training a grasping-only policy designed exclusively for single-object scenarios, where the objective is to grasp a single block. Based on this initial policy, we directly transfer its weights—treating it as our first "expert"—and iteratively train on increasingly complex tasks. Specifically, we first train on dense arrangements with D-4, then D-6, and finally

D-8 objects. 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 objects. At the end of the training process, we extract the policy trained on D-8 as the dense-clutter teacher policy and the one trained on R-8 as the random-clutter teacher policy.

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, D-6, and D-8 tasks and the random-clutter teacher policy for the R-4, R-6, and R-8 tasks. In total, 1000 episodes of observation and action data, along with point cloud data, 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 ensures a comprehensive representation of varying task complexities, which is critical for effective policy distillation.

2) Vision Policy Behavior Cloning: Our vision-based student policy uses point clouds instead of object poses, which cannot be accurately obtained in such heavily occluded, cluttered environments. We specifically use behavior cloning to train the student policy, using data collected from two teachers.

Observation space. The observation space for the vision-based student policy is defined as

$$s_t^V \triangleq \left[s_t^P, a_{t-1}, s_t^{O'}, d_t^{HO}, T_t, v^t\right] \in \mathbb{R}^{275} \tag{3}$$

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.

Action space. The action space of the vision-based student policy is identical to that of the state-based teacher policy. It consists of 16 dimensions for the linearly smoothed finger joint actions and 6 dimensions for the delta palm pose actions.

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 clutter environments? (2) How does our method generalize to different objects and tasks? (3) How effective is our curriculum learning? (4) How does our method perform on real-world tasks?

A. Baselines

To evaluate our approach, we design three experimental configurations with two baseline methods and our proposed approach:

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 singulation 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-B.2. We also include a singulation bonus to encourage singulation. 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

We introduce two key performance metrics:

Success rate. The proportion of trials in which the target object successfully reaches the predefined target position above the table surface is defined by $\|p^{\text{goal}} - p^{\text{target}}\|_2 < 0.05$. We denote the success rate as SR to evaluate the performance of each method.

Average steps. The average number of simulation steps required to singulate and grasp the target object to goal positions. The unsuccessful trials are excluded from the calculation. We denote average steps as AS to evaluate the efficiency of each method.

C. Implementation Details

We use Isaac Gym for teacher policy training. 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 3e-4. We then evaluated and selected the best-performing iteration as the policy for the next-stage curriculum learning. The student policy is trained over 200 epochs with a batch size of 12 trajectories, each composed of 300 steps of recorded simulation data and a learning rate of 1e-4.

D. Main Results and Analysis

We tested the dense-clutter teacher policy and the distilled vision student policy on D-4, D-6, and D-8 tasks and compared their performance with the grasping-only and multi-stage singulation methods. The grasping-only and multi-stage singulation baselines and our dense-clutter teacher policy were tested in 100 environments for one iteration, and the distilled vision student policy was tested in 10 environments for 10 iterations. Based on the results presented in Tab. II, the multi-stage singulation policy demonstrates a higher success rate than the grasping-only baseline. Nevertheless, although both efficiency and effectiveness decline

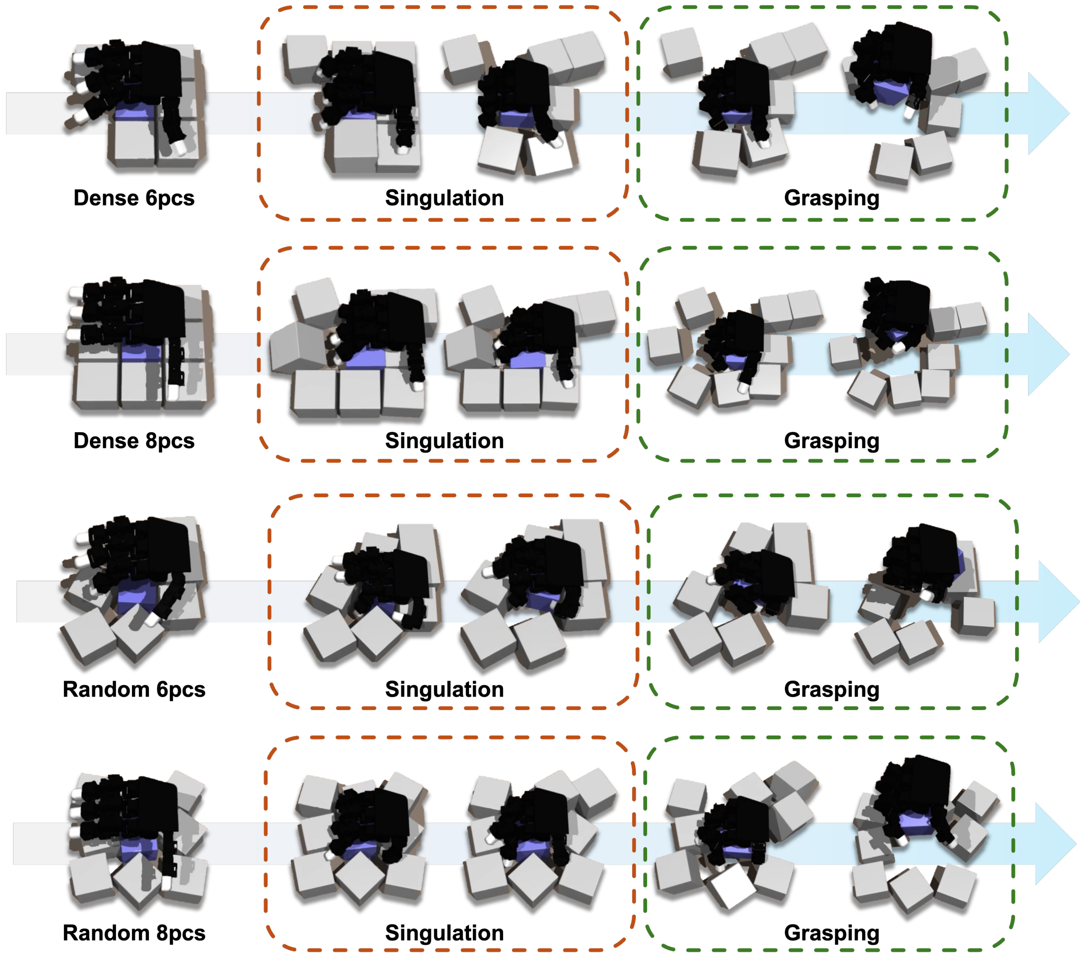

Fig. 4. Simulation Quantitive Result.

as AS increases, these results suggest that the singulation stage plays a positive role in task performance. Our dense-clutter teacher policy achieves a significantly higher success rate of 93%, with a substantially lower AS compared to the multi-stage singulation policy. While the distilled vision student policy exhibits a lower SR than the teacher policy, it still outperforms the baseline policies. It maintains an AS comparable to that of the teacher policy. Our results show that while combining separate optimal singulation and grasping policies can achieve a higher success rate, it increases action steps and reduces efficiency, whereas our single-stage policy balances effectiveness and efficiency for grasping in clutter environments, addressing Q1.

TABLE II

EVALUATION ON DENSE ARRANGEMENTS.

Method	SR(%)↑				AS↓			
	D-4	D-6	D-8	Avg.	D-4	D-6	D-8	Avg.
GraspReward-only	66%	40%	10%	39%	152	180	223	185
Multi-stage singulation	77%	76%	64%	72%	169	181	199	183
Ours (Teacher)	92%	96%	92%	93%	111	123	134	123
Ours (Student)	89%	87%	64%	80%	106	129	151	129

During training and testing, we found the policy learned several singulation patterns, including finger flickering, palm rubbing, and finger-palm vibration, to displace, nudge, or destabilize surrounding objects, effectively singulating targets in cluttered environments.

In response to Q2, we evaluate the random-clutter teacher policy and the distilled vision student policy on R-4, R-6, and R-8 tasks with different object arrangements to test the generalization ability of our policy. In Tab. III, we first observe the baseline policies perform better in random arrangements than dense arrangements, as the presence of random gaps provides more opportunities to grasp the target object directly. Despite that the diverse random arrangements pose challenges for our policies, our teacher policy can still achieve an average SR of 91% across all tasks. While the distilled vision student policy exhibits an SR drop compared to the teacher policy, it still outperforms the baseline policies with a lower AS for better task efficiency.

E. Curriculum Learning Analysis

The expert curriculum learning process is designed to enhance success rates in increasingly complex scenes. We assess the success rate and average steps for each expert within the curriculum learning framework in the R-8 clutter

TABLE III
EVALUATION ON RANDOM ARRANGEMENTS.

Method	SR(%)↑				AS↓			
	R-4	R-6	R-8	Avg.	R-4	R-6	R-8	Avg.
GraspReward-only	73%	61%	33%	56%	134	148	182	155
Multi-stage singulation	88%	72%	78%	79%	136	120	143	133
Ours (Teacher)	93%	89%	91%	91%	100	106	128	111
Ours (Student)	81%	64%	71%	72%	92	121	128	114

environment, as depicted in Fig. 5. The results indicate that as the curriculum advances, the expert policies demonstrate greater accuracy and efficiency, addressing Q3.

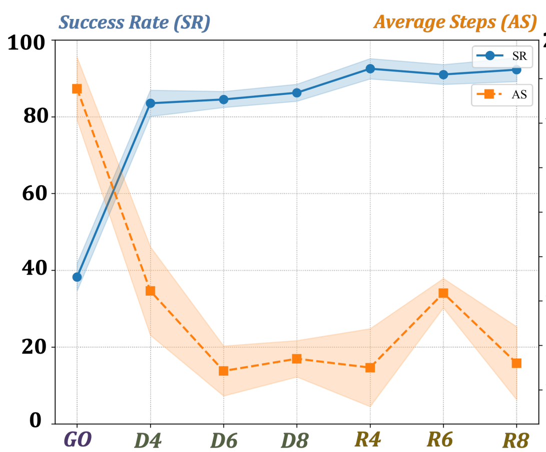

Fig. 5. Improving success rates during curriculum learning. GO stands for Grasping-only.

F. Real-World Experiments

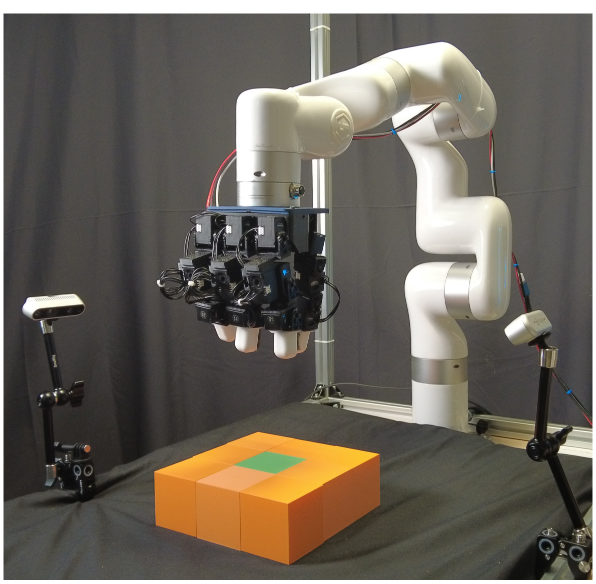

Fig. 6. Real-world experiment setting.

We conduct real-world experiments using a uFactory xArm6 robot equipped with the LEAP Hand [36] and two side view Realsense D435 RGB-D cameras, as illustrated in Fig. 6. We mount our LEAP hand vertically to the end-effector of the xArm6. In the experimental setup, we calibrate the camera's internal reference with 1280x720 RGB and depth pixels and fuse two RGB-D real-time point cloud outputs in the world coordinate system using the Iterative Closest Point (ICP) algorithm. We go through a spatial position filtering and downsampling step in acquiring 1024 neat points at 20 Hz as point cloud input. We accomplish the task of singulation and grasping in the real-world environment. For experiment videos, please visit our website at https://nus-lins-lab.github.io/dexsingweb/.

V. CONCLUSION

Our proposed approach demonstrates that a unified reinforcement learning framework can effectively integrate object singulation and grasping in densely cluttered environments using dexterous robotic 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 the dexterous hand to perform subtle non-prehensile maneuvers—such as nudging or sliding obstacles—directly during the grasping process. The integration of curriculum learning and policy distillation further enhances the generalization of the vision-based policy, ensuring successful skill transfer from simulation to real-world applications. Additionally, the introduction of cluttered grasping benchmarks provides a comprehensive testbed for evaluating performance across various clutter configurations, reinforcing the superiority of our approach over conventional methods.

Future work can extend these promising results by addressing more complex object arrangements and incorporating a broader range of object shapes and dynamic clutter scenarios to push the limits of the current framework. Improving robustness in real-world settings by mitigating challenges such as sensor noise and sim-to-real discrepancies, as well as exploring alternative learning paradigms—including imitation learning and hierarchical control strategies—will be crucial for advancing dexterous manipulation. These directions offer valuable pathways toward developing a robust, general-purpose robotic system capable of efficient and reliable object retrieval in increasingly unstructured environments.

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, pp. 6401–6408.

[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. 10 905–10 912.

[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] W. Liu, Y. Wan, J. Wang, Y. Kuang, X. Shi, H. Li, D. Zhao, Z. Zhang, and H. Wang, “Fetchbot: Object fetching in cluttered shelves via zero-shot sim2real,” *arXiv preprint arXiv:2502.17894*, 2025.

[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. Heemskerk, and H. Kasaei, “Self-supervised learning for joint pushing and grasping policies in highly cluttered environments,” in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 13 840–13 847.

[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, pp. 8408–8414.

[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*, vol. 5, no. 3, pp. 4978–4985, 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] 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)*. Xi’an, China: IEEE, May 2021, pp. 4694–4701.

[28] 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.

[29] 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.

[30] 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)*, 2021, pp. 6177–6184.

[31] 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.

[32] 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.

[33] 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. Dafle, 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

[34] 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.

[35] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” *arXiv preprint arXiv:1707.06347*, 2017.

[36] K. Shaw, A. Agarwal, and D. Pathak, “Leap hand: Low-cost, efficient, and anthropomorphic hand for robot learning,” *arXiv preprint arXiv:2309.06440*, 2023.