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

I. Introduction

Dexterous 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. Recent dexterous grasping approaches [10], [11] focus primarily on grasping target objects in scenarios without the need to rearrange surrounding objects. 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. However, due to the mechanical limitations of parallel grippers, the target object must be fully isolated from surrounding clutter, often requiring multiple inefficient steps of pushing and grasping. 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. 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 [15] which simplifies learning by breaking the problem into manageable sub-tasks, but it limits the synergy between singulation and grasping. Alternatively, curriculum learning [16] has proven effective in tackling complex tasks and has already been successfully applied to dexterous grasping policies [8], [17].

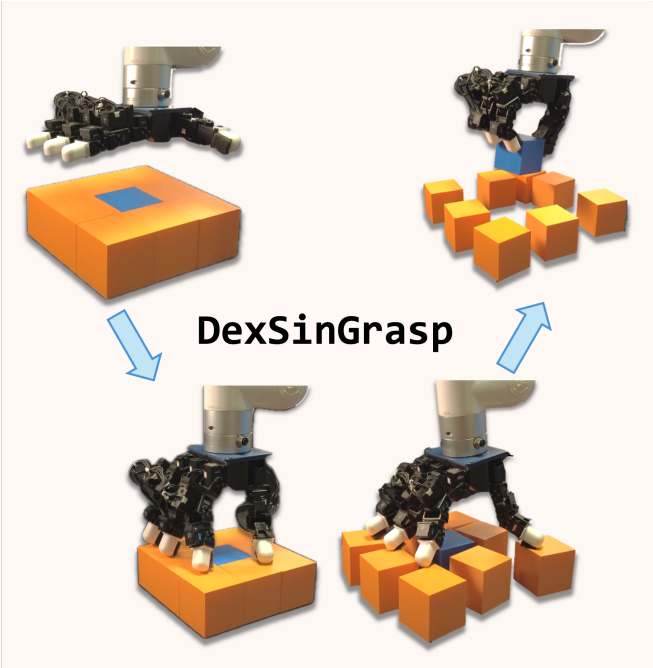

Fig. 1. We propose *DexSinGrasp* to learn a unified policy for dexterous object singulation and grasping in cluttered environments

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 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

* denotes equal contribution
† denotes the corresponding author

[1]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. davidxulixin@gmail.com, zixuanliu@u.nus.edu, linshao@nus.edu.sg

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 clutter arrangement 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*.

II. Method

Overview. We formulate dexterous grasping and object singulation as a reinforcement learning task. 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. II-A) that seamlessly integrates singulation and grasping into a single objective, enabling more efficient policy learning. Then, to improve learning efficiency in cluttered environments, we adopt Clutter Arrangement Curriculum Learning (Sec. II-B) to progressively train state-based teacher policies. Finally, we employ Teacher-Student Policy Distillation (Sec. II-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. Further details of problem formulation, unified reward function, clutter arrangement curriculum learning, and teacher-student policy distillation are provided in Appendix A, B, C, and D.

A. Unifying Dexterous Object Singulation and Grasping

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

$$r_t = \begin{cases} r_t^P + r_t^J + r_t^S, & \text{if } d_t^P \geq 0.06 \text{ or } d_t^J \geq 0.2, \\ r_t^P + r_t^J + r_t^F + r_t^L & \\ \quad + r_t^G + r_t^S + r_t^B, & \text{if } d_t^P < 0.06 \text{ and } d_t^J < 0.2, \end{cases} \tag{1}$$

Refer to Appendix for further explanation. Despite this unified learning framework, training remains highly challenging due to the increasing complexity of cluttered environments. 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 tightly cluttered environments, 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) [18], we optimize the policy to maximize the cumulative discounted reward $E[\sum_{t=1}^{T} \gamma^{t-1} r_t]$, enabling effective reinforcement learning in cluttered environments.

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. The teacher policy leverages simulator-provided privileged information, including object pose, singulation distances, and relative positions, to facilitate training. We then collect demonstration data using point cloud-based approximations of object location and hand-object distances, replacing privileged inputs during data collection, and train the vision-based student policy through behavior cloning.

III. 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 clutter arrangement 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 [15]. We train the separate singulation policy without the grasping reward stage as mentioned in Sec. II-A. 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

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.

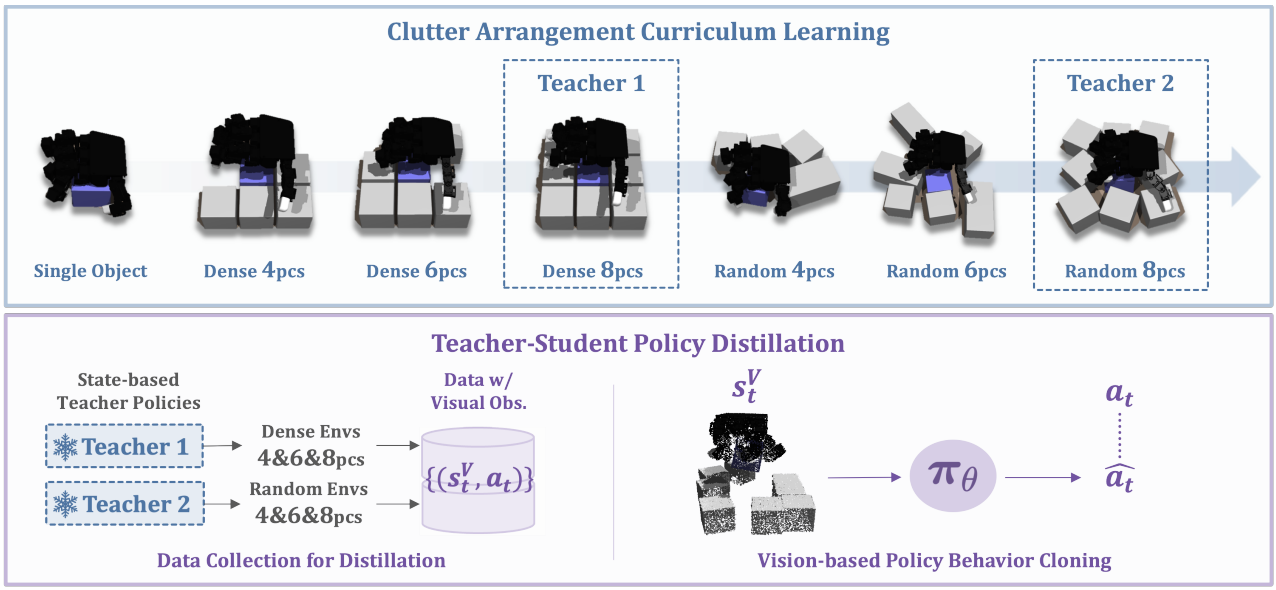

Fig. 2. Framework of *DexSinGrasp*. Firstly, we adopt clutter arrangement curriculum learning to progressively improve the performance of our teacher policy to address the challenge of training from scratch in dense or random clutter arrangements, and acquire two teacher policies for dense and random 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.

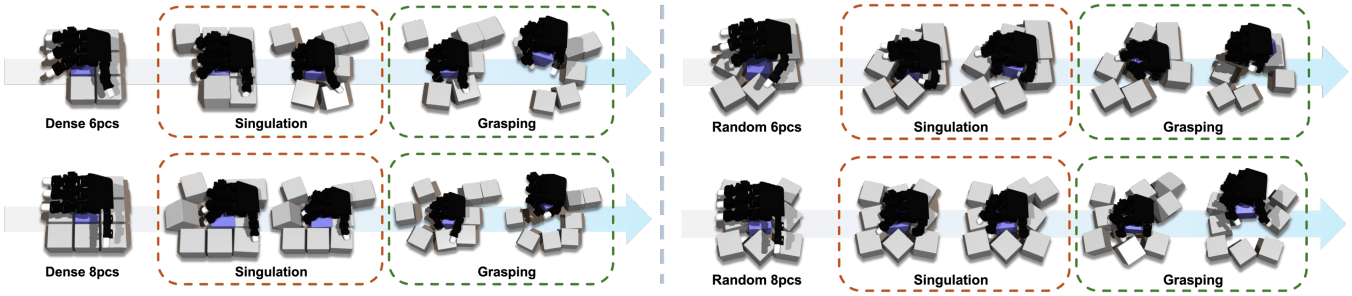

Fig. 3. Qualitative results on object singulation and grasping in simulation environments.

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 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 3e-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 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 GraspReward-only and multi-stage singulation methods. All methods were evaluated in 10 environments over 10 episodes each, except the multi-stage singulation policy, which was tested in a single environment for 100 episodes due to its non-parallelizable stage-switching mechanism.

Based on the results presented in Tab. I, 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 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 singulation and grasping policies can achieve a higher success rate, it increases action steps and reduces efficiency, whereas our unified policy balances effectiveness and efficiency for grasping in clutter environments, addressing Q1.

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. II, we

TABLE I

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)	98%	99%	97%	98%	96	113	133	114
Ours (Student)	90%	92%	84%	89%	102	108	132	114

first observe the baseline policies perform better in random arrangements than dense arrangements, as the presence of relatively loose gaps provides more opportunities to grasp the target object directly. Our teacher policy can achieve an average SR of 96% 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 higher SR and lower AS for better task efficiency.

TABLE II

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)	97%	96%	94%	96%	88	86	90	88
Ours (Student)	91%	86%	88%	88%	81	82	89	84

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, as shown in Fig. 3.

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. III. For the dense-to-random curriculum, we use the best-performing checkpoints at iterations 2k, 7.8k, 9.2k, 4.6k, 2.9k, 2.7k, and 2.5k respectively from trained 10k iterations; for the random-to-dense curriculum, at iterations 2k, 6.5k, 7.5k, 3.3k, 5.2k, 7.4k, and 1.6k respectively from trained 10k iterations. 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. Real-World Experiments

We conduct real-world experiments using a uFactory xArm6 robot equipped with the LEAP Hand [19] and two

TABLE III

EVALUATION ON DIFFERENT CURRICULUMS.

Curriculum	SR(%)↑ on D-n tasks				SR(%)↑ on R-n tasks			
	D-4	D-6	D-8	Avg.	R-4	R-6	R-8	Avg.
Training from scratch	90%	75%	97%	87%	74%	26%	0%	33%
Random-to-dense	96%	87%	81%	88%	97%	92%	97%	95%
Dense-to-random	98%	92%	97%	96%	96%	96%	94%	95%

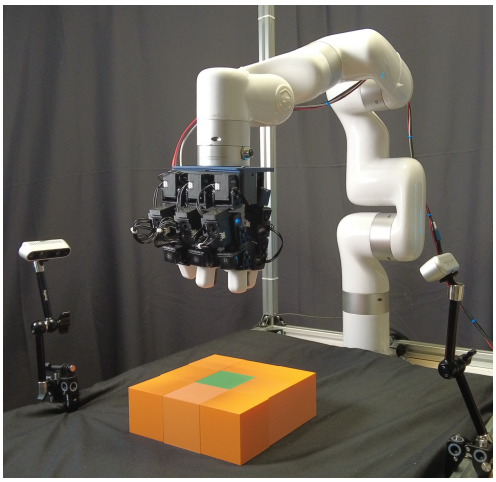

Fig. 4. Real-world experiment setting.

side view Realsense D435 RGB-D cameras, as illustrated in Fig. 4. We mount our LEAP hand vertically to the end-effector of the xArm6. In the experimental setup, we calibrate the camera intrinsics 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 to obtain 1024 clean points at 20 Hz.We accomplish singulation and grasping on the D-4, D-6, D-8, R-4, R-6, and R-8 tasks in the real-world environment. For experiment videos, please visit our website at https://nus-lins-lab.github.io/dexsingweb/.

IV. CONCLUSION

Our proposed approach demonstrates that a unified reinforcement learning framework can effectively integrate object singulation and grasping in densely or randomly cluttered environments using dexterous robotic hands. The integration of clutter arrangement 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 various cluttered grasping environments 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.

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] J. Lundell, F. Verdoja, and V. Kyrki, "Ddgc: Generative deep dexterous grasping in clutter," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 6899–6906, 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. 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] 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.

[16] 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

[17] 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.

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

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

APPENDIX

A. Problem Formulation

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.

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

$$s_t \triangleq \left[s_t^R, a_{t-1}, s_t^O, d_t^{HO}, T_t, d_t^S\right] \in \mathbb{R}^{168}, \tag{2}$$

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; 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.

2) 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 finger joint positions $a_t^F := \lambda a_t^F + (1-\lambda) a_{t-1}^F \in \mathbb{R}^{16}$ for each finger.

B. Unifying Dexterous Object Singulation and Grasping

TABLE IV
REWARD-RELATED TERMS

Term	Equation
d^P	$\min_{i=1}^p \lVert p^{\text{palm}} - p_i^{\text{target}} \rVert_2$
r^P	$-2.0 \times d^P$
d^J	$\sum_{j=1}^m \min_{i=1}^p \lVert p_j^{\text{link}} - p_i^{\text{target}} \rVert_2$
r^J	$-d^J$
r^F	$-\sum_{j=1}^h \min_{i=1}^p \lVert p_j^{\text{fingertip}} - p_i^{\text{target}} \rVert_2$
r^L	$0.2 + 0.6 \times a^{P_{t_z}}$
r^G	$0.9 - 2.5 \times \lVert p^{\text{goal}} - p^{\text{target}} \rVert_2$
r^S	$50 \times \min_{i=1}^n \lVert p^{\text{target}} - p_i \rVert_2$
r^B	$(1 + 10 \times \lVert p^{\text{goal}} - p^{\text{target}} \rVert_2)^{-1}$

The terms in the piece-wise reward function are summarized in Tab. IV, 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 successful 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 hand 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.

The reward function consists of 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. 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 a successful full-hand grasping, ensuring grasp stability and robustness in cluttered environment.

C. Clutter Arrangement Curriculum Learning

1) Cluttered Environment Generation: 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. 5.

Dense Arrangements. This type of task arranges different quantities of the surrounding 1×1 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) 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 scenarios, where the objective is to grasp a single block. Based on this initial policy, we continuously follow the curriculum and train on

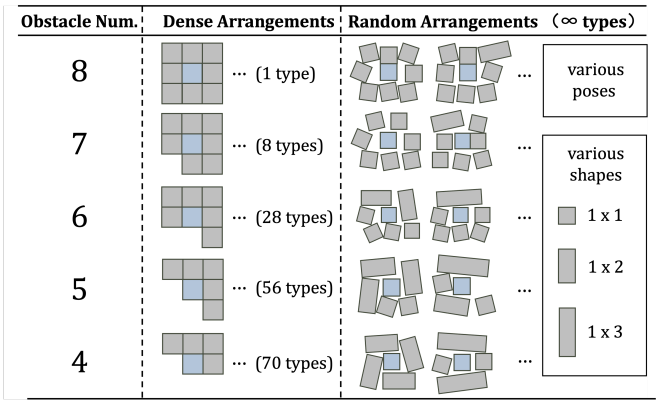

Fig. 5. We introduce a cluttered environment generation module to create diverse object settings, including different obstacle quantities from 0 to 8, dense and random arrangements, various poses, and block shapes. For simplicity, we only show obstacles with numbers 4 to 8.

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 policy from the D-8 training as the dense-clutter teacher policy (optimal on D-n tasks) and the final policy from the R-8 training as the random-clutter teacher policy (optimal on R-n tasks).

D. Teacher-Student Policy Distillation

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 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 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, cluttered environments. 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

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

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 $s_t^{O'} \in \mathbb{R}^3$ as the object state is hard to acquire. 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.