DYNAMO-GRASP: DYNAMics-aware Optimization for GRASP Point Detection in Suction Grippers

Anonymous Author(s)
Affiliation
Address
email

Abstract: In this research, we introduce a novel approach to the challenge of suction grasp point detection. Our method, exploiting the strengths of physics-based simulation and data-driven modeling, accounts for object dynamics during the grasping process, markedly enhancing the robot's capability to handle previously unseen objects in real-world scenarios. We benchmark DYNAMO-GRASP against established approaches via comprehensive evaluations in both simulated and real-world environments. Our method outperformed the alternatives, achieving a success rate of 98% in diverse simulated picking tasks and 80% in real-world, adversarially-designed scenarios. Demonstrating a strong ability to adapt to complex and unexpected object dynamics, our method offers robust generalization to real-world challenges. The results of this research set the stage for more reliable and resilient robotic manipulation in intricate real-world situations. Experiment videos and code will be made available at: https://sites.google.com/view/dynamo-grasp

Keywords: Suction Grasping, Manipulation, Deep Learning, Vision

1 Introduction

Grasp point detection is essential for successful robotic manipulation, as it requires identifying the optimal location on an object for a robot to securely grasp and manipulate. Rapid and reliable grasping capabilities for a wide range of objects can benefit various applications, such as warehouse and service robots. Suction grasping is a popular grasping modality in real-world settings due to its simplicity and reliability when handling objects with nonporous, flat surfaces compared to parallel-jaw or multi-finger grasping. Existing methods for finding suitable grasping areas for suction grippers typically focus on maximizing suction seal quality and robustness against wrenches, taking into account the object's shape, size, and surface properties [1, 2].

Most existing methods for suction grasping assume a top-down manipulation setting, where objects are initially placed on a stable, flat surface before being grasped, and the robot attempts to grasp the object from above. This is due to the suction cup gripper requiring the robot to apply a specific amount of force to press the suction cup against the object's surface, which causes the cup to deform and create an air seal, resulting in a secure suction grasp. Consequently, an object being grasped needs sufficient and stable support in the direction opposite the robot's pushing. Without such support, the object may move in an unfavorable direction, leading to the suction cup's failure to form the air seal. However, numerous real-world scenarios require a robot to grasp objects without stable support, such as grasping from a container with a side opening or from an unstable pile of objects. In these situations, the objects may exhibit significantly more complex dynamics during the manipulation process, due to the displacement caused by the robot's motion and the objects' interactions with one another. State-of-the-art grasp point detection methods for suction grasping could suffer from these complex object-picking scenarios because they do not consider the objects' movement during the manipulation process. This limitation greatly restricts the range of scenarios

Submitted to the 7th Conference on Robot Learning (CoRL 2023). Do not distribute.

in which suction grippers can be applied, preventing them from reaching their full potential in real-world manipulation tasks. Fig.1.a. illustrates a real-world manipulation task.

In this work, our goal is to fully exploit the potential of suction grippers by developing a grasp point detection model that not only examines quantitative metrics such as mechanical wrench resistance and suction quality but, more importantly, considers the object dynamics during the picking process. This paper makes the following contributions: **1. Suction grasping simulation:** We developed a high-performance suction grasping simulation environment using Isaac Gym[3], a physics simulation platform that supports parallelization on GPUs. In this environment, we simulated robot-picking tasks in a warehouse setting. By implementing a suction grasping model tuned with real-world suction grasping data, this simulation environment provides a physically realistic experience in terms of the robot's kinematic and form factor constraints, suction grasping characteristics, and object dynamics during manipulation. **2. Robot learning framework:** Utilizing the simulation environment, we generated a dataset that contains about 100,000 simulated grasps and trained a grasp point detection model that takes into account how the movement of objects and the robot's kinematics impact the success of grasping. **3. Real robot evaluation in a real-world warehouse setting:** We evaluated two grasp point detection approaches and the model we created using the proposed learning framework. Our method outperformed the alternatives, achieving a success rate of 98% in diverse simulated picking tasks and 80% in real-world, adversarially-designed scenarios. We plan to make our dataset publicly available and open-source our simulation environment.

2 Related Work

Suction-based robot manipulators have gained widespread popularity in both real-world applications. For instance, suction grasping methods are used in manufacturing [4, 5, 6], warehousing [7, 8], underwater manipulation [9, 10], food and fruit manipulation [11, 12, 13, 14], etc. Another major direction where suction grasping has been applied is in the exploration of end-effector modalities [15, 16, 17].

Analytic Models. In the realm of conventional suction cup grippers, the effective analysis of grasp quality necessitates the modeling of various cup properties. Given that these suction cups are typically fashioned from elastic materials, such as rubber or silicone, researchers frequently employ spring-mass systems to represent their deformations [1, 2, 18]. Upon establishing a secure grasp on an object using a suction gripper, the suction cup is typically modeled as a rigid entity. The subsequent analysis involves assessing the forces imposed on the object, encompassing those along the surface normal, friction-induced tangential forces, and suction-generated pulling forces [19]. Mahler et al. [1] introduced a combined model in DexNet3.0, incorporating both torsional friction and contact moment within a compliant model of the contact ring between the cup and the object. This amalgamated model has demonstrated its efficacy and is em-

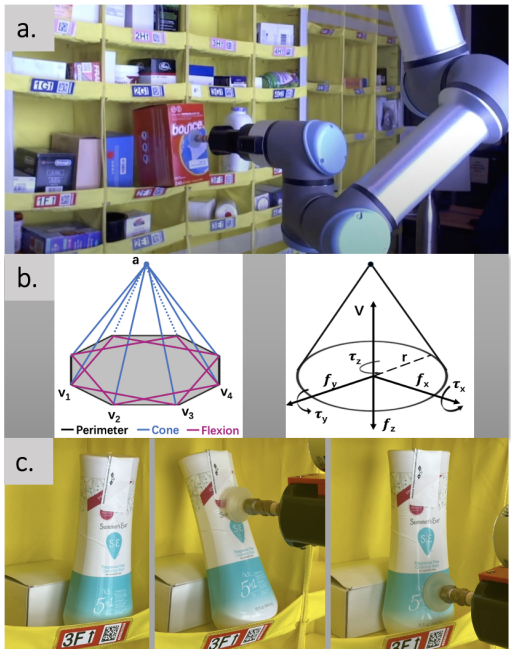

Figure 1: a. Suction grasping for real-world scenarios remains challenging due to limited analysis of object movements. b. SOTA methods only reason for object's surface properties. *Left*: The quasi-static spring model. *Right*: Wrench basis for the suction cup. [1] c. *Left*: A warehouse picking scenario. *Middle*: DexNet failing the grasp due to object toppling. *Right*: An effective grasp point that prevents unfavorable object movements. See the project website for experiment videos.

ployed in subsequent works [2, 20]. Additionally, this work adapts the analytic models from DexNet3.0 for the purpose of data annotation.

Learning Suction Grasps. Machine learning research in robotics has been actively exploring the selection of optimal grasp points to enhance suction grasping for intricate manipulation tasks. These tasks include novel object picking, object stewing, picking from containers, etc. Existing approaches generate training data through either human expertise [21] or simulations [1, 22, 2, 23]. DexNet3.0 [1], for instance, synthesizes training data and proposes suction grasp points that aid in forming an effective suction seal and ensuring wrench resistance. Several other studies center around clustered scenarios by creating models that take RGB-D input and predict graspable points [2, 22, 21]. Jiang et al. [23] proposed a methodology that simultaneously considers grasping quality and robot reachability for bin-picking tasks. Despite these studies primarily focusing on analyzing surface properties or robot configuration, they largely overlook how the displacement of the object during the picking process might impact the success of the task. Addressing this particular aspect is the main focus of our work.

Visual Pushing. This project also shares relevance with the active research area of object displacement modeling during manipulation [24, 25]. Effective non-prehensile manipulation strategies have been successfully applied to enhance grasping operations [26, 27, 28]. Recently, reasoning object translation via visual input has gained huge advances. Transporter and its variants [29, 30] have introduced a data-efficient learning paradigm that links visual inputs to desired robotic actions. Nonetheless, these methods are underpinned by a strong assumption of translational equivariance in visual representation, a condition that is often not met in non-table-top settings. Visual foresight methods [31, 32] have offered a model-based framework that predicts future observations based on a state-action pair. However, these approaches necessitate searching through the action space given a specific task, which can be time-consuming for intricate real-world problems. Other existing studies [33, 34, 35] have examined robotic manipulation from a sideways perspective. However, none of them have explicitly modeled the complex dynamics caused by the interaction between the robot and the objects.

3 Problem Statement

Our objective is to identify grasp points on a target object within a container filled with multiple items, using a single-view depth image observation. The identified grasp points should enable a robot to successfully establish a suction grasp, even when the object lacks stable support in the direction opposite to the robot's push. Consistent with previous suction grasp point detection studies [1, 2, 20], a grasp point is defined by a target point $[\mathbf{p}, \mathbf{v}]$. Here, $\mathbf{p} \in \mathbb{R}^3$ represents the center of the contact ring between the suction cup and the object, while $\mathbf{v} \in \mathbb{S}^2$ denotes the gripper's approach direction. The grasp labeling function is defined as 1 if the grasp successfully forms a suction grasp on the target object, and 0 if it does not. Our final system evaluates the quality of a suction grasp based on the following crucial criteria:

3.1 Seal Quality and Wrench Resistance

A suction cup is capable of lifting objects owing to a differential in air pressure. This differential is created across the cup's membrane by a vacuum generator, which pulls the object toward the cup. Ensuring a tight seal between the suction cup and the target object is crucial for successful operation. For sealing evaluation, we follow the highly effective quasi-static spring-based model proposed in Dexnet 3.0 [1]. As shown in Fig.1.b., this model uses a combination of three spring systems to represent suction cup deformation. A perimeter springs system is used to assess the deformation between adjacent vertices, namely v_i and v_{i+1}. The cone springs system signifies the deformation of the suction cup's physical structure, as determined by the distance between v_i and a. Lastly, the flexion springs, which connect vertex v_i to v_{i+2}, are employed to resist bending along the surface of the cup.

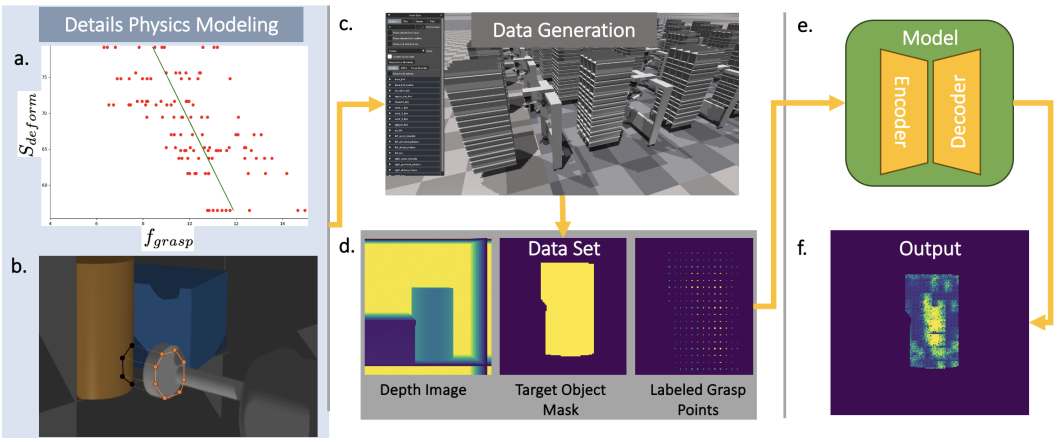

Figure 2: An overview of the proposed pipeline: **a.** We conducted system identification using 19 everyday objects of diverse shapes, weights, volumes, and materials to ascertain the function F discussed in Section 4. **b.** Calculation of deformation score at each simulation time step. **c.&d.** Generating dataset with our simulation environment. **e.&f.** Trained DYNAMO-GRASP model, outputs an affordance map highlighting optimal grasp areas.

When the suction cup forms an air seal with the object, the suction gripper should be able to resist wrenches that are caused by gravity or other disturbances. The suction ring contact model proposed in Dexnet 3.0 [1] efficiently encapsulates the forces experienced by a suction cup during a grasp. As depicted in Fig.1.b, this model takes into account five forces. The actuated normal force (f_z) and vacuum force V represent the gripper pressing into the object along the contact axis and air suction pulling the object, respectively. The friction forces (f_x, f_y) and torsional friction (τ_z) result from the normal force exerted between the suction cup and the object, acting as resistive forces. Lastly, the elastic restoring torques (τ_x, τ_y) result from the elastic restoring forces within the suction cup, which apply torque on the object along the boundary of the contact ring.

3.2 Object Movement

Most established suction grasping techniques assume little to no movement of the object during the process, which facilitates the deformation of the suction cup, thereby enabling the formation of an air seal for a secure grasp. However, in various practical manipulation scenarios, the target object might not have ample and steady support opposite the robot's push. This lack of support can lead to undesirable shifts in the object's position, preventing the successful creation of the air seal. The situation becomes even more complex when other objects are located near the target, due to the interactions among them. This work addresses these complexities by modeling the movement of objects during the picking process, which enhances the applicability and efficiency of suction-based grippers in real-world manipulation tasks. Assuming that an object's state is denoted by its Cartesian pose and velocity in a workspace, represented as $s = (p, \delta p)$, the states of i objects in a container at time t can be represented as $\mathbf{s_t} = \{s_{t_0}, s_{t_1}, ..., s_{t_i}\}$. At each time step, a robot equipped with a suction gripper performs a pushing action $a_t = (f_t, \mathbf{p}, \mathbf{v})$, applying a force f_t to a specific location, $\mathbf{p}$, on the object's surface in the direction of $\mathbf{v}$. The state transition model $p(\mathbf{s_{t+1}}) = T(\mathbf{s_t}, a_t)$ provides a distribution over the potential movements of the objects during the picking process.

4 DYNAMO-GRASP

This section proposes a robot learning pipeline designed to create a grasp point detection model. This model suggests suction grasp points by analyzing combined information regarding object surface properties and object movement during the picking process. We first implemented a new suction grasping simulation environment that accurately simulate suction cup properties and objects' displacement caused by the robot's motion and the objects' interactions with one another. A

transformer-based model is trained to take a depth image as input to analyze the implicit object displacements during the picking process. The model generates an affordance map over the target object's surface, indicating the likelihood of a successful suction grasp if a robot executes a pushing action along the surface normal across various areas of the object. Please note that our method primarily focuses on analyzing the impact of physical interaction between robots and objects on the quality of a suction grasp. During execution, we filter out grasp points that offer inadequate air seals and wrench resistance, based on Dexnet's output. Fig.2 shows the system architecture.

4.1 Simulation

In order to eliminate the need for expensive real robot data collection, we carefully designed a simulation environment that accurately replicates the physical properties of the suction cup, the motion of objects caused by robot grasping, as well as the robot's kinematics during the picking process. We chose to implement our grasping simulation environment based on Isaac Gym, allowing all computations to be accelerated via GPUs. While Isaac Gym lacks important features that emulate detailed suction grasping properties, our environment integrates several custom-implemented functional modules. It provides a pipeline that accurately simulates a suction-picking process by taking into account factors such as suction cup properties, robot kinematic constraints, collisions, control noise, and object dynamics.

Modeling Suction Properties. The majority of popular physics simulations for robotics merely simulate suction grasp through simplistic mechanisms. These mechanisms typically involve directly attaching the object to the robot's end-effector or creating an attracting force between the object and the effector. However, these approaches neglect critical physical details. Specifically, to successfully register a suction grasp, the suction cup must be pushed and deformed to a sufficient extent that the rim of the suction cup attaches to the surface of the object, thereby forming an air seal. Modeling the amount of force required to form an air seal is crucial for this problem. This is because when the target object lacks rigid support, exerting sufficient force directly causes the object's displacement. Understanding the magnitude of the force that the robot exerts on the target object is instrumental in recreating accurate object dynamics. To model the deformation properties of the suction cup, we first adopted the Perimeter Springs in the quasi-static spring system, as discussed in Sec 3.1. Given a grasp point $\mathbf{p}$ on the object's surface and the angle of incident $\mathbf{v}$, this model calculates a suction deformation score $S_{deform} = 1 - \max(r_1, r_2, ..., r_n)$, where $r_i = \min(1, |(l_i^{'} - l_i)/l_i|)$. Here, l_i represents the original length of the perimeter spring linking vertex v_i and v_{i+1}, and $l_i^{'}$ is the length after projecting the vertices onto the object's surface. Using real-world data, we then conduct a system identification process to ascertain the function F. This function signifies how forcefully the robot needs to press the suction cup to achieve a successful grasp, given a deformation score of a specific grasp point: $F(S_{deform}) \rightarrow f_{grasp}$.

Simulating Grasping Physics. *(1) Kinematics:* Our simulation accepts a robot's model as input and controls the robot using an end-effector controller to attempt various suction grasps. This approach enables the simulation to demonstrate how the robot's form factor and kinematic properties impact its grasp. For instance, some grasp points might be physically unattainable for the robot due to its manipulability and reachability constraints, or collisions. *(2) Generating Scenarios:* Our experiment primarily focuses on a warehouse lateral picking scenario. During our data generation process, the simulation randomly selects one to three objects from our object set and spawns them into the same container with random positions and orientations. We also implement domain randomization for observation noise, objects' weights, and controller parameters, ensuring the dataset reflects a range of diverse physical properties and robot behaviors. One of the objects in the container is randomly assigned as the target object to be picked. *(3) Sampling Grasp Points:* Given a picking scenario, we sample two sets of candidate grasp points from the visible surface of the target object. The first set is derived from uniform sampling across the entire surface, ensuring that the robot explores diverse picking strategies. The second set contains the grasp points with the highest score returned by DexNet via the Cross-Entropy Method (CEM) sampling strategy, ensuring the robot explores areas that DexNet deems preferable.

Labeling Data. After sampling the candidate grasp points, our simulated robot 'physically' executes each candidate $\mathbf{p}$ by performing a sequence of pushing actions $\mathbf{A} = \{a_t\}_{t=0}^{T}$, where each action $a_t = (f_t, \mathbf{p}, \mathbf{v})$. Here, $\mathbf{v}$ is determined by the surface normal at $\mathbf{p}$. The robot exerts a constant force $f_t = f_c$ if the target object is unstable and moves in response to the gripper's push. Once the object finds a position with adequate support against the push, f_t gradually increases until the suction cup deforms enough to form an air seal, or until the object starts moving again. The simulation of the suction cup's deformation and the precise estimation of suction grasp registration involve continuous calculation of S_{deform} and f_{grasp} at each timestep. This process considers the suction cup's current position relative to the target object, as shown in Fig.2.b. A force sensor on the end-effector continuously monitors f_t, and a grasp is deemed successful if $f_t \geq f_{grasp}$. Any failure to meet this condition, such as collisions or inaccuracies in end-effector positioning due to manipulability or reachability issues, results in the grasp point being marked as unsuccessful.

4.2 Robot Learning

We employ the suction grasping simulation, as described above, to generate a dataset. This dataset represents a warehouse scenario where a robot, equipped with a suction gripper, is tasked with extracting a target object from a small container filled with multiple unorganized items. The experimental setting is detailed in Sec.5. This dataset was used to train a model for grasp point selection. The model takes a single-view point cloud of the container's interior, a segmentation mask identifying each object within the container and its boundaries as inputs. It then outputs an affordance map representing the estimated probability of successful grasps at all potential grasp points on the target object. The largest value in the affordance map indicates the optimal grasp point, $(\mathbf{p}^*, \mathbf{v}^*)$, for the given scenario. As shown in Fig.2.e, our model employs an auto-encoder architecture, integrating a transformer encoder and a deconvolutional decoder. As previously mentioned, our data generation process is designed to capture the inherent variabilities of complex real-world robotic suction grasping tasks. These include variations stemming from the physical properties of different objects, robot constraints, and stochasticity in the controller, among others. Empirically, we discovered that the following loss function can effectively mitigate the adverse effects of the high aleatoric uncertainty in our dataset during training: $\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \forall y_i \in Y_{max}$. Here, Y_{max} represents a subset of samples containing the n highest-scored grasp points on an object.

5 Experiment

Our experiment focuses on a warehouse robot suction grasping scenario, as shown in Fig.1.a. The opening of the container is located on the side, introducing a side-way manipulation nature to the problem. This makes suction grasping significantly more challenging compared to top-down manipulation scenarios, as the robot's movements can trigger a series of object displacements, leading to objects being shifted or even toppled over. Consequently, this scenario serves as an excellent evaluation environment for our work. In our experiment, we evaluated four methods: DYNAMO-GRASP, DexNet3.0, a centroid picking method, and DYNAMO+DexNet. DexNet3.0 is a SOTA suction-picking technique, serving as a strong baseline. Meanwhile, the centroid picking method, a straightforward approach involving suctioning on the object's centroid, has proven effective in similar tasks at the Amazon Robotics Challenge [35, 36].

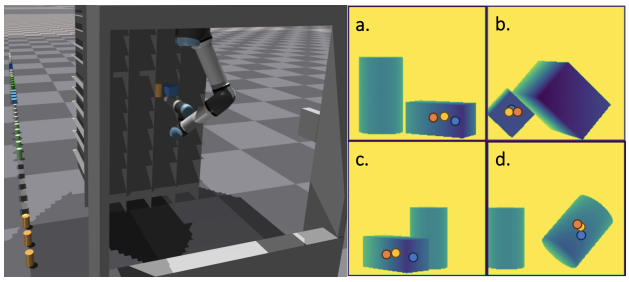

Figure 3: **Left:** The simulation environment for data generation and experiments. The simulated objects with different weights, sizes, and shapes are displayed on the left side of the robot. **Right:** In Section 5.1, challenging test cases are presented where only DYNAMO-GRASP was successful in grasping the target object. The orange, blue, and yellow points indicate the grasp points proposed by DYNAMO-GRASP, DexNet, and the centroid method, respectively.

Our method uses a relatively simple suction cup model compared to DexNet3.0's comprehensive quasi-static spring and wrench basis model. Our focus lies on simulating and modeling robot-object dynamics during the manipulation process. Including a detailed suction cup model in every simulation step could significantly increase data generation time. Consequently, our approach may have a lower performance in assessing air seal quality and wrench resistance of a grasp point. To make up for this shortcoming, we implemented DYNAMO+DexNet. This approach densely samples grasp points on the object's surface and filters out 90% deemed suboptimal by DexNet. The robot will then use the grasp point that earns the highest score from our model.

5.1 Large-scale, Diverse Scenario Assessment

To thoroughly evaluate the performance of various methods on the suction grasping problem at hand, we randomly generated 50 diverse picking scenarios. These 50 scenarios were generated by sampling from a distribution that incorporates even greater randomness in object orientation than the dataset used for model training. These scenarios incorporate a wide range of object configurations, leading to potentially complex object movements during picking. As it's overly labor-intensive and time-consuming to repeatedly set up the 50 scenarios in the real world for each method, we've chosen to conduct these experiments in our simulation environment. We didn't include the DYNAMO+DexNet method in this experiment because the objects in our simulation environment lack complex surface geometries. Consequently, we anticipated its performance to align closely with that of DYNAMO-GRASP.

As presented in Table.1, even though the test cases were sourced from a more complex distribution than that used for training, our method demonstrated robust performance with a high success rate of 98%. This outcome implies our model's effective generalization to unforeseen challenges. Our method was able to adapt its grasping strategy to accommodate potential object movements in challenging scenarios. In contrast, the DexNet and Centroid methods, though capable of handling most common scenarios, struggle with more challenging cases, leading to lower success rates of 84% and 76%, respectively. We illustrate some challenging cases in Fig.3 where only our method successfully grasped the target object. Interestingly, in scenarios involving complex object dynamics, slight differences in grasp point location can lead to different outcomes. For instance, in cases b and d (Fig.3), successful grasp points were located just one to two centimeters away from the unsuccessful ones.

	Successful Grasp	Success Rate
DYNAMO	49	98%
DexNet	42	84%
Centroid	38	76%

Table 1: Experimental results for 50 configurations experiment.

5.2 Real-world Evaluation

To assess real-world effectiveness, we established the warehouse side-way picking setting in our lab and used a UR16 robot to evaluate the proposed grasping strategies of the four different methods. Before observing the robot in action, a human operator well-versed in suction gripper mechanisms created five adversarial picking scenarios that could pose potential challenges. In each picking scenario, the UR16 robot executes the optimal grasp point as recommended by each of the methods. As shown in Fig.4, our method (e.g. DYNAMO-GRASP) succeeded in four out of five adversarial scenarios. With an 80% success rate, it outperformed all other methods tested. Surprisingly, the centroid method, with a 40% success rate, outperformed DexNet, which only achieved a 20% success rate in this experiment. This result, however, is understandable upon reflection. DexNet was trained exclusively on a dataset created using top-down picking scenarios. As a result, the model does not anticipate potential object movements, such as toppling or sliding, that could occur prior to the successful establishment of a suction grasp. In contrast, the centroid method, which seeks to attach the suction cup to the object's geometric center, often results in more stable movement and reduces unexpected displacements. While the DYNAMO+Dex method outperformed DexNet in this experiment, its performance fell short of our expectations. We suspect that the grasp point selection strategy used for this method may not have been optimal for this experimental setup.

	1.	2.	3.	4.	5.
DYN	✓	✓	✗ Dynamics	✓	✓
Dex	✗ Dynamics	✓	✗ Kinematics	✗ Kinematics	✗ Dynamics
Cen	✓	✗ Dynamics	✗ Dynamics	✓	✗ Dynamics
D+D	✓	✗ Dynamics	✗ Dynamics	✗ Dynamics	✓

Figure 4: Real-world adversarial evaluation: DYNAMO GRASP (our method), DexNet, centroid-based, and a DYNAMO+DexNet hybrid. The color-coded points represent the suggested grasp points from various algorithms. The successfully identified grasp points are marked by a green tick and failed ones are in red. Failures are classified into 'Kinematics Constraint Failures' (mechanical/geometric limitations) and 'Dynamics Failures' (object displacement during grasp).

Qualitative Analysis. In the first scenario, shown in Fig.4.1, a tall bottle is partially propped up by a box in the back. This setting examines the grasp point selection method's awareness that the target object may topple during the pick. Both the DYNAMO-GRASP and DYNAMO+DexNet methods opted to pick from the lower portion of the bottle, ensuring that the pick was firmly supported by the box. On the other hand, the other two methods chose grasp points higher up on the bottle, which both led to toppling movements. Despite the bottle's toppling motion, the centroid method fortuitously formed an air seal. Another intriguing scenario was the third case, as seen in Fig.4.3. The slender bottle is less tolerant of end-effector positioning errors, and the diagonal orientation of the box induces lateral movement in the bottle, further compounding the challenge. Consequently, all four methods failed to successfully pick up the object. The final case, illustrated in Fig.4, involves both a small target object and disorganized object configurations. As consistent with our observation in Sec.5.1, slight variations in grasp point locations in this scenario yielded different outcomes. Notably, the grasp points suggested by both DYNAMO-GRASP and DYNAMO+Dex led to successful grasps and were nearly identical. This demonstrates a certain level of robustness in this area, as opposed to mere coincidence.

Consistency Test. We further evaluate the robustness of our method by conducting five additional experimental trials on scenario four, using DYNAMO-GRASP and DexNet. Out of a total of six attempts, DYNAMO-GRASP achieved successful grasping five times, while DexNet only succeeded three times. Additionally, our method consistently suggested grasp points within the same local region across different trials, while DexNet suggested points across a much broader distribution.

6 Conclusion and Limitation

We've presented DYNAMO-GRASP, a dynamic-aware grasp point detection method, which takes the object dynamics into account during the grasping process. Our quantitative evaluation both in simulation as well as in real-world scenarios yields an increase in the pick success rate to 98% (simulation) or to 80% (real-world). **Limitations and future work:** Firstly, the dataset used in our simulation environment primarily includes objects with relatively simple geometric shapes. This aspect could limit the efficacy of our method when dealing with objects of uncommon or complex shapes. Similarly, our real-world experiments primarily involved simple geometric objects, such as boxes and bottles. Lastly, due to time constraints, we were unable to explore more effective heuristics for selecting grasp points using the DYNAMO-GRASP + DexNet combination. This exploration might have allowed us to leverage the strengths of both methods and potentially achieve even greater performance.

References

[1] J. Mahler, M. Matl, X. Liu, A. Li, D. Gealy, and K. Goldberg. Dex-net 3.0: Computing robust vacuum suction grasp targets in point clouds using a new analytic model and deep learning. In *2018 IEEE International Conference on robotics and automation (ICRA)*, pages 5620–5627. IEEE, 2018.

[2] H. Cao, H.-S. Fang, W. Liu, and C. Lu. Suctionnet-1billion: A large-scale benchmark for suction grasping. *IEEE Robotics and Automation Letters*, 6(4):8718–8725, 2021.

[3] V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa, and G. State. Isaac gym: High performance gpu-based physics simulation for robot learning, 2021.

[4] T. Zhang, C. Zhang, and T. Hu. A robotic grasp detection method based on auto-annotated dataset in disordered manufacturing scenarios. *Robotics and Computer-Integrated Manufacturing*, 76:102329, 2022.

[5] M. Yang, L. Yu, C. Wong, C. Mineo, E. Yang, I. Bomphray, and R. Huang. A cooperative mobile robot and manipulator system (co-mrms) for transport and lay-up of fibre plies in modern composite material manufacture. *The International Journal of Advanced Manufacturing Technology*, pages 1–17, 2021.

[6] A. S. Olesen, B. B. Gergaly, E. A. Ryberg, M. R. Thomsen, and D. Chrysostomou. A collaborative robot cell for random bin-picking based on deep learning policies and a multi-gripper switching strategy. *Procedia Manufacturing*, 51:3–10, 2020.

[7] S. Hasegawa, K. Wada, K. Okada, and M. Inaba. A three-fingered hand with a suction gripping system for warehouse automation. *Journal of Robotics and Mechatronics*, 31(2):289–304, 2019.

[8] M. Schwarz, A. Milan, C. Lenz, A. Munoz, A. S. Periyasamy, M. Schreiber, S. Schüller, and S. Behnke. Nimbro picking: Versatile part handling for warehouse automation. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3032–3039. IEEE, 2017.

[9] H. S. Stuart, M. Bagheri, S. Wang, H. Barnard, A. L. Sheng, M. Jenkins, and M. R. Cutkosky. Suction helps in a pinch: Improving underwater manipulation with gentle suction flow. In *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 2279–2284. IEEE, 2015.

[10] H. Kumamoto, N. Shirakura, J. Takamatsu, and T. Ogasawara. Underwater suction gripper for object manipulation with an underwater robot. In *2021 IEEE International Conference on Mechatronics (ICM)*, pages 1–7. IEEE, 2021.

[11] P. Y. Chua, T. Ilschner, and D. G. Caldwell. Robotic manipulation of food products–a review. *Industrial Robot: An International Journal*, 30(4):345–354, 2003.

[12] R. Morales, F. Badesa, N. Garcia-Aracil, J. Sabater, and L. Zollo. Soft robotic manipulation of onions and artichokes in the food industry. *Advances in Mechanical Engineering*, 6:345291, 2014.

[13] K. Gilday, J. Lilley, and F. Iida. Suction cup based on particle jamming and its performance comparison in various fruit handling tasks. In *2020 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pages 607–612. IEEE, 2020.

[14] C. Blanes, M. Mellado, C. Ortiz, and A. Valera. Technologies for robot grippers in pick and place operations for fresh fruits and vegetables. *Spanish Journal of Agricultural Research*, 9 (4):1130–1141, 2011.

[15] S. Jeong, P. Tran, and J. P. Desai. Integration of self-sealing suction cups on the flexotendon glove-ii robotic exoskeleton system. *IEEE Robotics and Automation Letters*, 5(2):867–874, 2020.

[16] T. M. Huh, K. Sanders, M. Danielczuk, M. Li, Y. Chen, K. Goldberg, and H. S. Stuart. A multi-chamber smart suction cup for adaptive gripping and haptic exploration. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1786–1793. IEEE, 2021.

[17] B. Mazzolai, A. Mondini, F. Tramacere, G. Riccomi, A. Sadeghi, G. Giordano, E. Del Dottore, M. Scaccia, M. Zampato, and S. Carminati. Octopus-inspired soft arm with suction cups for enhanced grasping tasks in confined environments. *Advanced Intelligent Systems*, 1(6): 1900041, 2019.

[18] X. Provot et al. Deformation constraints in a mass-spring model to describe rigid cloth behaviour. In *Graphics interface*, pages 147–147. Canadian Information Processing Society, 1995.

[19] R. Kolluru, K. P. Valavanis, and T. M. Hebert. Modeling, analysis, and performance evaluation of a robotic gripper system for limp material handling. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 28(3):480–486, 1998.

[20] J. Mahler, M. Matl, V. Satish, M. Danielczuk, B. DeRose, S. McKinley, and K. Goldberg. Learning ambidextrous robot grasping policies. *Science Robotics*, 4(26):eaau4984, 2019.

[21] A. Zeng, S. Song, K.-T. Yu, E. Donlon, F. R. Hogan, M. Bauza, D. Ma, O. Taylor, M. Liu, E. Romo, et al. Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching. *The International Journal of Robotics Research*, 41(7): 690–705, 2022.

[22] Q. Shao, J. Hu, W. Wang, Y. Fang, W. Liu, J. Qi, and J. Ma. Suction grasp region prediction using self-supervised learning for object picking in dense clutter. In *2019 IEEE 5th International Conference on Mechatronics System and Robots (ICMSR)*, pages 7–12. IEEE, 2019.

[23] P. Jiang, J. Oaki, Y. Ishihara, J. Ooga, H. Han, A. Sugahara, S. Tokura, H. Eto, K. Komoda, and A. Ogawa. Learning suction graspability considering grasp quality and robot reachability for bin-picking. *Frontiers in Neurorobotics*, 16, 2022.

[24] M. T. Mason. Mechanics and planning of manipulator pushing operations. *The International Journal of Robotics Research*, 5(3):53–71, 1986.

[25] K. M. Lynch and M. T. Mason. Stable pushing: Mechanics, controllability, and planning. *The international journal of robotics research*, 15(6):533–556, 1996.

[26] M. Dogar and S. Srinivasa. A framework for push-grasping in clutter. *Robotics: Science and systems VII*, 1, 2011.

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

[28] S. Levine, C. Finn, T. Darrell, and P. Abbeel. End-to-end training of deep visuomotor policies. *The Journal of Machine Learning Research*, 17(1):1334–1373, 2016.

[29] A. Zeng, P. Florence, J. Tompson, S. Welker, J. Chien, M. Attarian, T. Armstrong, I. Krasin, D. Duong, V. Sindhwani, et al. Transporter networks: Rearranging the visual world for robotic manipulation. In *Conference on Robot Learning*, pages 726–747. PMLR, 2021.

[30] H. Wu, J. Ye, X. Meng, C. Paxton, and G. S. Chirikjian. Transporters with visual foresight for solving unseen rearrangement tasks. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10756–10763. IEEE, 2022.

[31] C. Finn and S. Levine. Deep visual foresight for planning robot motion. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2786–2793. IEEE, 2017.

[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*, 7(1):231–238, 2022. doi:10.1109/LRA.2021.3123373.

[33] R. Wang, Y. Miao, and K. E. Bekris. Efficient and high-quality prehensile rearrangement in cluttered and confined spaces. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 1968–1975. IEEE, 2022.

[34] C. Eppner, S. Höfer, R. Jonschkowski, R. Martín-Martín, A. Sieverling, V. Wall, and O. Brock. Lessons from the amazon picking challenge: Four aspects of building robotic systems. In *Robotics: science and systems*, pages 4831–4835, 2016.

[35] C. Hernandez, M. Bharatheesha, W. Ko, H. Gaiser, J. Tan, K. van Deurzen, M. de Vries, B. Van Mil, J. van Egmond, R. Burger, et al. Team delft's robot winner of the amazon picking challenge 2016. In *RoboCup 2016: Robot World Cup XX 20*, pages 613–624. Springer, 2017.

[36] K.-T. Yu, N. Fazeli, N. Chavan-Dafle, O. Taylor, E. Donlon, G. D. Lankenau, and A. Rodriguez. A summary of team mit's approach to the amazon picking challenge 2015. *arXiv preprint arXiv:1604.03639*, 2016.