

Article

Multiple-Object Grasping Using a Multiple-Suction-Cup Vacuum Gripper in Cluttered Scenes

Ping Jiang ^{1,*}, Junji Oaki ^{1,†}, Yoshiyuki Ishihara ² and Junichiro Ooga ¹¹ Corporate Manufacturing Engineering Center, Toshiba Corporation, Yokohama 235-0017, Japan² Corporate Research & Development Center, Toshiba Corporation, Kawasaki 212-8582, Japan

* Correspondence: ping2.jiang@toshiba.co.jp

† These authors contributed equally to this work.

Abstract: Multiple-suction-cup grasping can improve the efficiency of bin picking in cluttered scenes. In this paper, we propose a grasp planner for a vacuum gripper to use multiple suction cups to simultaneously grasp multiple objects or an object with a large surface. To take on the challenge of determining where to grasp and which cups to activate when grasping, we used 3D convolution to convolve the affordable areas inferred by a neural network with the gripper kernel in order to find graspable positions of sampled gripper orientations. The kernel used for 3D convolution in this work was encoded, including cup ID information, which helps to directly determine which cups to activate by decoding the convolution results. Furthermore, a sorting algorithm is proposed to determine the optimal grasp among the candidates. Our planner exhibited good generality and successfully found multiple-cup grasps in previous affordance map datasets. Our planner also exhibited improved picking efficiency using multiple suction cups in physical robot-picking experiments. Compared with single-object (single-cup) grasping, multiple-cup grasping contributed to 1.45×, 1.65×, and 1.16× increases in efficiency for picking boxes, fruits, and daily necessities, respectively.

Keywords: bin picking; grasp planning; suction grasp; graspability; deep learning



Citation: Jiang, P.; Oaki, J.; Ishihara, Y.; Ooga, J. Multiple-Object Grasping Using a Multiple-Suction-Cup Vacuum Gripper in Cluttered Scenes. *Robotics* **2024**, *13*, 85. <https://doi.org/10.3390/robotics13060085>

Academic Editors: Roman Mykhailishyn, Ann Majewicz Fey, Charalampos P. Bechlioulis and Kensuke Harada

Received: 23 April 2024

Revised: 22 May 2024

Accepted: 23 May 2024

Published: 27 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the growth of e-commerce, demand for the automation of bin picking by robots in warehouses has become high [1], particularly in Japan, since the country is faced with a labor shortage due to its aging society. COVID-19 has made the situation worse since picking tasks in warehouses is not amenable to telework. Most state-of-the-art robotic picking systems focus on single-object grasping. To further improve the efficiency of these systems, simultaneous grasping of multiple objects might reduce the number of pick attempts to improve the picking speed, as shown in Figure 1A. In addition, a robot can more stably grasp and hold objects that have a large surface using multiple suction cups to grasp the object, as demonstrated in Figure 1B.

Multiple-object grasp planning for jaw or multi-finger grippers has previously been proposed under various conditions, such as in well-organized scenes [2,3], rearranged scenes [4], and cluttered scenes [5–8]. These studies demonstrated that multiple-object grasping could improve picking speed. However, few studies have examined multiple-object grasping using a vacuum gripper with multiple suction cups. Most studies infer the single-object grasp point for a gripper with only a single suction cup using direct or indirect methods. Direct methods [9–11] use deep convolutional neural networks to directly infer the grasp point, while indirect methods [12–14] first infer the affordance map, which is a pixel-wise map indicating the graspability score for a single-cup vacuum gripper at each pixel, and then find the optimal grasp point in the map. Given that the affordance map contains all possible grasp points for a single suction cup, if all cups in a vacuum gripper have the same geometry (e.g., cup radius) and dynamics (e.g., suction force limit

and friction coefficient), then we can search for a gripper pose where the center positions of at least two of the cups are located at non-zero pixels in the affordance map and satisfy the conditions described in Section 4 for grasping multiple objects or an object with a large surface area.

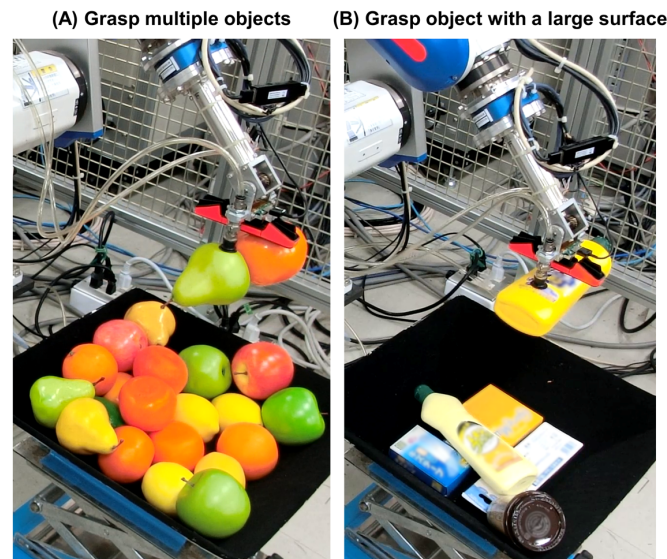


Figure 1. Using multiple suction cups to grasp (A) multiple objects or (B) an object with a large surface area.

In this study, we propose an affordance-map-based grasp planner for a multiple-suction-cup vacuum gripper to grasp multiple objects or grasp an object with a large surface area. We propose a 3D convolution-based method, which takes advantage of the suction affordance map inferred by our prior work, suction graspability U-Net++ (SG-U-Net++) [14], to search for a gripper pose capable of grasping multiple objects or an object with a large surface area. Furthermore, unlike the control of a jaw gripper in which all fingers of the gripper are usually controlled to open or close simultaneously, the suction cups need to be controlled separately. Therefore, we design a kernel that includes an encoded cup ID to determine which suction cup to activate. Furthermore, as there might be many candidates for multiple-suction-cup grasping, we propose an evaluation metric for determining the optimal grasp among candidates. The proposed grasp planner is validated using previous affordance datasets and physical robot experiments.

In short, the contributions of this work include the following:

- A 3D-convolution-based grasp planner for a vacuum gripper with multiple cups to grasp multiple objects or an object with a large surface area.
- Control of suction cup activation by incorporating a convolution kernel, including the encoded cup ID.
- A robotic picking system with a hybrid planner that performs multiple-suction-cup grasp planning preferentially and switches to single-object grasp planning when there are no solutions.
- A sorting algorithm for determining the optimal grasp for multiple-cup grasping.
- Validation of the grasp planner on previous affordance datasets, including Suction FCN [12], SuctionNet-1Billion [13], and SG-U-Net++ [14].
- Experiments for picking boxes, fruits, and daily necessities with a vacuum gripper with two cups and a comparison of multiple- and single-cup grasping results.

2. Related Works

2.1. Single-Object Grasping Based on an Affordance Map

A pixel-wise affordance map contains grasp quality at each pixel when the robot grasps the object in the corresponding pose. Unlike end-to-end deep learning, which is used to

directly predict grasp configurations such as a rotated bounding box [15–17] for a jaw gripper or a suction point for a vacuum gripper with only a single cup [9–11], affordance learning has the advantages that the neural network model can be anchor-free, and there is no need to sample candidate grasps as in [18]. Zeng et al. [12] were one of the first research groups to apply pixel-wise affordance learning to bin picking for four multiple-motion primitives of a hybrid robotic hand with both a jaw and a single suction cup. They used a manually annotated affordance dataset to train fully convolutional networks (FCNs). The precision and generalizability of the FCNs were then further improved by [19,20]. Another representative work is by Morrison et al. [21], who generated affordance and pose maps from the rotated bounding box and designed a generative grasp CNN (GG-CNN) to directly infer the pixel-wise grasp pose and quality. To learn the grasp of a jaw gripper, many researchers [17,22–25] later used similar methods to generate affordance map datasets from grasp configuration annotations represented by a rotated rectangle (e.g., the Cornell Grasp Dataset [15] and Jacquard Dataset [26]).

However, these studies required real images and an expert to perform pixel-wise grasp affordance annotation. To reduce dataset generation costs, datasets are generated in a physical simulator where affordance is evaluated using a designed contact model (e.g., the quasi-static spring model used in Dex-net [27]) on a previously synthesized image. Recently, a similar contact model was used by Cao et al. [13] to generate a larger suction pixel-wise affordance (seal score) dataset. However, these studies required that contact model parameters were determined properly and treated with a vacuum gripper with only a single cup.

In the present study, we used our previously proposed SG-U-Net++ [14] to infer the pixel-wise suction affordance map. SG-U-Net++ was trained on a synthesized dataset annotated by an analysis-based method and was competitive with the method trained on a dataset annotated by a contact model. We propose a grasp planner for multiple-suction-cup grasping that takes advantage of the predicted affordance map.

2.2. Multiple-Object Grasping

Most studies have focused on multiple-object grasping using a multi-finger gripper. Grasp conditions have been analyzed for a multi-finger gripper to stably grasp multiple cylinders [28,29], polyhedral objects [2], planar objects [3], and shaped spatial objects [30]. Recent studies have started to use data-driven methods to deal with the multiple-object grasping problem. Shenoy et al. [5] used a deep neural network to infer the number of objects for a three-finger gripper to grasp when digging into a pile of objects. They later proposed a Markov decision-based method to optimize pick-transfer routines when grasping multiple objects [6]. Sakamoto et al. [4] used mask-RCNN to detect objects and then searched for a gripper pose to push two boxes together in order to simultaneously grasp them. A similar push–grasp task was studied by Agboh et al. for grasping multiple arbitrary convex polygonal objects under frictional and frictionless contact conditions between objects [7,8]. They proposed MOG-Net to infer the maximum number of objects that a two-finger gripper could grasp by a sampled pose. However, the set of objects in these studies was still simple, and simultaneous grasping of objects with more complicated shapes (e.g., daily necessities) is required for more general applications (e.g., picking in warehouses). Mucchiani et al. [31] designed a novel end-effector to sequentially grasp multiple objects with complicated shapes. Similarly, Yao et al. [32] proposed a human-like grasp synthesis algorithm to achieve sequential multiple-object grasping. However, these studies grasped multiple objects sequentially rather than simultaneously.

For multiple-suction-cup grasping, most studies focus on grasping a single object rather than multiple objects using multiple cups. Mantriota [33] analyzed the suction force and friction coefficient to grasp and hold a large object with a four-cup vacuum gripper. Kozák [34] et al. used a deep neural network to estimate the pose of a round part and then used a six-cup vacuum gripper to grasp it. Tanaka et al. [35] designed a two-surface vacuum gripper in which each surface was equipped with multiple cups. They used a

gripper to simultaneously suck two surfaces of a large box to improve the stability of grasping and holding. Leitner et al. [36] used a gripper with two different shaped cups to grasp an object on a shelf. These studies used a multiple-cup vacuum gripper to grasp a single object more stably. Kessens [37] et al. mounted a four-cup vacuum gripper on a drone to achieve sequential multiple-object grasping in the air, but found that simultaneous grasping was challenging. Islam et al. [38] proposed a planner for an unloading task in which the robot used a multiple-suction-cup vacuum gripper to simultaneously grasp and unload multiple cardboard boxes, but it was difficult to apply the planner to objects with complicated shapes in a cluttered scene. To our knowledge, the present study is the first to propose a grasp planner to simultaneously grasp multiple objects using a multiple-suction-cup vacuum gripper. The proposed planner can also determine gripper poses to stably grasp large objects with multiple cups.

3. Problem Statement

This study focuses on a bin-picking task in cluttered scenes. The robot is required to pick multiple objects or an object with a large surface area using multiple suction cups and then place them in a tote.

3.1. Assumption

We assume a suction vacuum gripper with multiple suction cups where all cups have the same specifications (e.g., the right side of Figure 2, in which both cups have the same shape, size, and suction force limits). In addition, the gripper tool center point (TCP) and all cups are in the same plane (e.g., the left side of Figure 2, in which the cup center points and TCP are in the same blue plane).

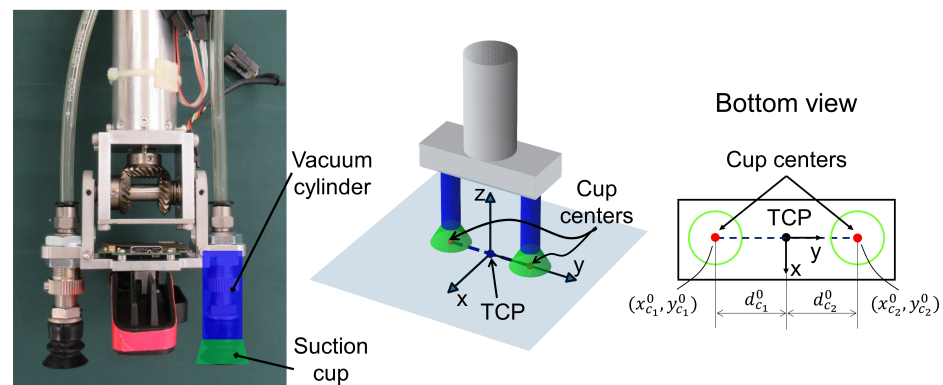


Figure 2. An example of a vacuum gripper with two cups. The TCP is the gripper tool center point. $d_{c_i}^0$ is the distance from the i th cup to the TCP. $(x_{c_i}^0, y_{c_i}^0)$ is the center position of the i th cup in local gripper coordinates. The green trapezium is the suction cup. The blue cylinder is the vacuum cylinder.

3.2. Vacuum Gripper State

The vacuum gripper state G consists of gripper position P , orientation O , suction cup center position C , and cup activation mode A , as shown in Equation (1). P is the position (x_g, y_g, z_g) of the TCP in world coordinates. O is the orientation represented by a ZYZ rotation matrix $(R_z(\theta_g)R_y'(\phi_g)R_z''(\gamma_g))$, where θ_g and ϕ_g are the azimuthal angle and polar angle of the unit vector of gripper axis z , respectively, and γ_g is the rotation angle around gripper axis z . Note that $\sin$ and $\cos$ are abbreviated to c and s in the matrix. C consists of the center position $(x_{c_i}, y_{c_i}, z_{c_i})$ of each suction cup in world coordinates. C^0 is the center position $(x_{c_i}^0, y_{c_i}^0, z_{c_i}^0)$ of each suction cup in local gripper coordinates (see the right side of

Figure 2). A is a one-hot vector representing the activation status (a_{c_i}) of each suction cup, where a_{c_i} is 1 if the i th cup is activated, and 0 if it is disabled.

$$\begin{aligned}
 G &= [P, O, C, A], \\
 P &= [x_g, y_g, z_g], \\
 O &= R_z(\theta_g)R_y'(\phi_g)R_z''(\gamma_g), \\
 &= \begin{bmatrix} c_{\phi_g}c_{\theta_g}c_{\gamma_g} - s_{\theta_g}c_{\gamma_g} & -c_{\phi_g}c_{\theta_g}s_{\gamma_g} - s_{\theta_g}c_{\gamma_g} & s_{\phi_g}c_{\theta_g} \\ c_{\phi_g}s_{\theta_g}c_{\gamma_g} + c_{\theta_g}s_{\gamma_g} & -c_{\phi_g}s_{\theta_g}s_{\gamma_g} + c_{\theta_g}c_{\gamma_g} & s_{\phi_g}s_{\theta_g} \\ -s_{\phi_g}c_{\gamma_g} & s_{\phi_g}s_{\gamma_g} & c_{\phi_g} \end{bmatrix}, \\
 C &= [[x_{c_1}, y_{c_1}, z_{c_1}], [x_{c_2}, y_{c_2}, z_{c_2}], \dots, [x_{c_i}, y_{c_i}, z_{c_i}]], \\
 A &= [a_{c_1}, a_{c_2}, \dots, a_{c_i}].
 \end{aligned} \tag{1}$$

4. Conditions for Grasping Using Multiple Suction Cups

Since all suction cups installed in the gripper are assumed to be the same, the affordance map of each suction cup is the same. Hence, we can determine a gripper pose capable of grasping multiple objects or an object with a large surface area by multiple cups if the following conditions are satisfied (an example is shown in Figure 3).

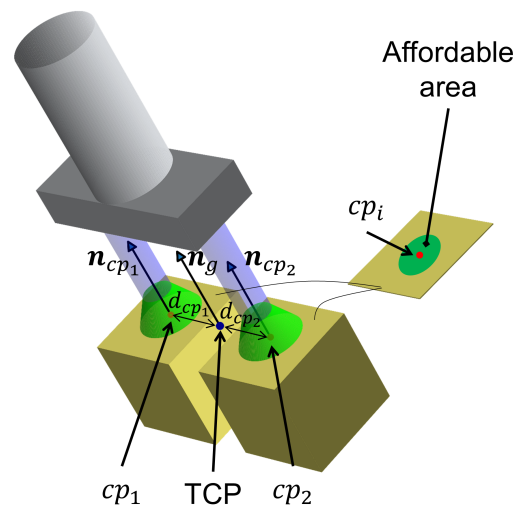


Figure 3. An example of conditions for a vacuum gripper with two cups to grasp two objects. cp is the contact point where the suction cup sucks the surface. Contact points cp_1 and cp_2 need to be located in the affordable areas of objects. cp_1 , cp_2 , and TCP need to be in the same plane perpendicular to $\mathbf{n}_g$. $\mathbf{n}_{cp_1}$ and $\mathbf{n}_{cp_2}$ are the normals of the contact points for the left and right cups, respectively. Both $\mathbf{n}_{cp_1}$ and $\mathbf{n}_{cp_2}$ need to be parallel to $\mathbf{n}_g$. d_{cp_1} and d_{cp_2} are the distances from the contact points to the TCP for the left and right cups in world coordinates, respectively. Both d_{cp_1} and d_{cp_2} need to be equal to the distances from the left cup center ($d_{c_1}^0$) and right cup center ($d_{c_2}^0$) to the TCP in gripper coordinates.

Condition 1. At least two of the contact points are located in affordable areas on the object. If the contact points are located in affordable areas on different objects, the gripper can grasp multiple objects. If the contact points are located in the same affordable area, the gripper can grasp a large surface area using multiple cups.

Condition 2. The gripper TCP and all contact points located in affordable areas are in the same plane, which is perpendicular to the unit vector of the gripper's z-axis ($\mathbf{n}_g$).

Condition 3. Normals of all contact points located in affordable areas ($\mathbf{n}_{cp_i}$) are in the same direction as the unit vector of the gripper's z-axis ($\mathbf{n}_g$), as shown in Equation (2). Note that

$aff_{cp_i} > 0$ indicates that the i th contact point is located in an affordable area where its affordance score is non-zero.

$$\arccos(\mathbf{n}_{cp_i} \cdot \mathbf{n}_g) < \varepsilon_1 \quad \text{where} \quad aff_{cp_i} > 0. \quad (2)$$

Condition 4. The distance from each contact point located in affordable areas to the TCP in world coordinates (d_{cp_i}) needs to be equal to the distance from the corresponding cup center to the TCP in local coordinates ($d_{c_i}^0$ in Figure 2).

$$|d_{cp_i} - d_{c_i}^0| < \varepsilon_2 \quad \text{where} \quad aff_{cp_i} > 0. \quad (3)$$

5. Multiple-Suction-Cup Grasp Planner

5.1. Overview of Architecture

Figure 4 and Algorithm 1 show the overall architecture and workflow of our multiple-suction-cup grasp planner. Given a depth image of I_d , our previous work SG-U-Net++ is used to infer the affordance map I_{aff} for a single cup. The voxel grid generator then extracts the point cloud (I_{pcd}) affiliated with the affordable areas in the map and downsamples them to a voxel grid (V). The orientation generator uses the point normals $\mathbf{n}_{pcd}$ of extracted points to efficiently generate the gripper orientation samples (S_O). The gripper kernel generator generates 3D encoded gripper kernels ($\mathcal{K}$), including cup ID information. The decoder decodes the result ($ConvRes$) of 3D convolution (3D Conv.) of V over $\mathcal{K}$ and generates the gripper pose candidates (G_{cand}). The normal direction checker removes candidates where the $\mathbf{n}_g$ and contact point normals are not in the same direction. If G_{cand} is successfully found, G_{cand} is evaluated and ranked to obtain the optimal grasp (G_{opt}). Otherwise, if no G_{cand} is found, the planner is switched to our previous single-object grasp planner, where the position with the highest affordance score is set as the goal, and the cup that can reach the goal by the shortest trajectory is selected to grasp the object.

Algorithm 1 Multiple-suction-cup grasp planner

Input: I_{aff} : affordance map

I_d : depth image

I_{pcd} : point cloud

l : voxel size

C^0 : local cup center positions (see the right side of Figure 2)

Output: G_{opt} : optimal grasp

- 1: $V \leftarrow \text{GenerateVoxelGrid}(I_{pcd}, I_{aff}, l)$
 - 2: $\mathbf{n}_{pcd} \leftarrow \text{EstimateNormals}(I_{pcd})$
 - 3: $S_O \leftarrow \text{SampleGripperOrientation}(\mathbf{n}_{pcd}, I_{aff})$
 - 4: $\mathcal{K} \leftarrow \text{GenerateEncodedKernels}(S_O, C^0, l)$
 - 5: $ConvRes \leftarrow \text{Conv3D}(V, \mathcal{K})$
 - 6: $G_{cand} \leftarrow \text{Decode}(ConvRes)$
 - 7: $G_{cand} \leftarrow \text{NormalDirectionCheck}(ConvRes)$
 - 8: **if** $\text{len}(G_{cand}) > 0$ **then**
 - 9: $G_{opt} \leftarrow \text{Ranking}(I_{aff}, G_{cand})$
 - 10: **else**
 - 11: # Single-object grasp planning
 - 12: $G_{opt} \leftarrow \text{argmax}(I_{aff})$
 - 13: **end if**
 - 14: **return** G_{opt}
-

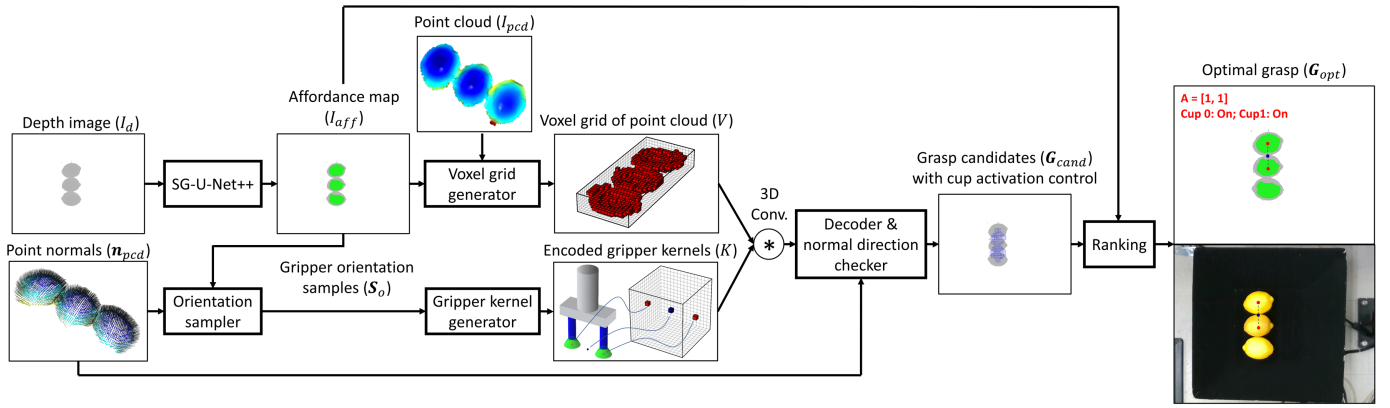


Figure 4. Overall architecture of the multiple-suction-cup grasp planner. $\otimes$ is the convolution operator.

5.2. Affordance Map Inference

We use SG-U-Net++ from our prior work to generate the affordance map. SG-U-Net++ has a nested U-Net structure and infers pixel-wise grasp quality and approachability based on a depth image. Refer to [14] for further details. Pixels with non-zero grasp quality scores are filtered out to generate an affordance map (the green area in the affordance map in Figure 4).

5.3. Voxel Grid Generation

We use voxel downsampling to generate the binary voxel grid (V) of the point cloud. Points located in the affordable areas are extracted and downsampled to a voxel grid with a defined grid size l . The voxel grid is further binarized such that if a grid in the voxel grid contained more than 10 points, the grid value would be 1 or 0. The voxel grid shape is $N_x \times N_y \times N_z$, where $N_x = \lfloor \frac{B_x^{max} - B_x^{min}}{l} \rfloor$, $N_y = \lfloor \frac{B_y^{max} - B_y^{min}}{l} \rfloor$, and $N_z = \lfloor \frac{B_z^{max} - B_z^{min}}{l} \rfloor$. B^{max} and B^{min} are the maximum and minimum bounds of the point cloud.

5.4. Grasp Orientation Candidate Generation

To satisfy Condition 3 in Section 4, Equation (2) must be computed for each point normal to sample the gripper orientations. If the size of the input point cloud is large, online sampling will result in high costs in terms of memory usage and computation time. We propose an efficient sampling method for a vacuum gripper by generating an offline normal to gripper orientation map. Since the Cartesian coordinates of a given vector vec ($[vec_x, vec_y, vec_z]$) can be represented by the azimuthal angle θ and polar angle ϕ as in Equation (4), all possible normals of contact points can be sampled by an angle interval $\Delta\alpha$, as in Equation (5). Meanwhile, as in Equation (1), n_g (the last column of O) depends on only ϕ_g and θ_g and has the same representation as Equation (4), so n_g can be sampled by the same angle interval as in Equation (6).

$$\begin{aligned} vec &= [vec_x, vec_y, vec_z] = [c_\theta s_\phi, s_\theta s_\phi, c_\phi], \\ \theta &= \arctan(vec_y, vec_x), \\ \phi &= \arccos(vec_z), \end{aligned} \quad (4)$$

where θ is the normal polar coordinate in the $x - o - y$ plane, and ϕ is the angle between the vector and the z axis. Assuming the normal is always in the up direction, $\theta \in (-\pi, \pi]$ and $\phi \in [0, \frac{\pi}{2}]$.

$$S_n(ii, jj) = [c_{ii\Delta\alpha - \pi s_{jj\Delta\alpha}}, s_{ii\Delta\alpha - \pi s_{jj\Delta\alpha}}, c_{jj\Delta\alpha}], \quad (5)$$

$$S_{n_g}(ii', jj') = [c_{ii'\Delta\alpha - \pi s_{jj'\Delta\alpha}}, s_{ii'\Delta\alpha - \pi s_{jj'\Delta\alpha}}, c_{jj'\Delta\alpha}], \quad (6)$$

where $ii, ii' = 0, 1, \dots, \frac{\pi}{\Delta\alpha}$ and $jj, jj' = 0, 1, \dots, \frac{\pi}{2\Delta\alpha}$.

For each $S_{n_{cp}}(ii, jj)$, we search for all $S_{n_g}(ii', jj')$ satisfying Equation (2) in order to create a map $\mathcal{M} : (ii, jj) \rightarrow (ii', jj')$, which maps a point normal entry to all n_g in the same direction as the normal vector. This map could be generated offline, and this only needs to be carried out once, thus reducing the computation cost.

Based on $\mathcal{M}$, given the point normals, the feasible candidate n_g could be rapidly obtained so that gripper orientation samples (S_O) could be generated. Given n_{pcd} , normals of points located in affordable areas are extracted and azimuthal and polar angles are computed (lines 1–3 in Algorithm 2). The angles are then used to calculate the entry key ii, jj to query $\mathcal{M}$ to obtain the feasible ii', jj' , based on which samples (S_{θ_g} and S_{ϕ_g}) of θ_g and ϕ_g are obtained (lines 4–7 in Algorithm 2). Note that only unique ii, jj values with top-10% counts are used as entries. This helps to improve the sampling efficiency when the variation in n_{pcd} is small. For example, if the input point cloud is a set of points in a plane, all n_{pcd} and corresponding ii, jj are the same. Hence, using unique values, only one unique ii, jj rather than all ii, jj are used. As n_g depends on only θ_g and ϕ_g , γ_g could be any value if θ_g and ϕ_g are feasible. Hence, γ_g is sampled by the same interval $\Delta\alpha$ in the range $(-\pi, \pi]$ (lines 9–10 in Algorithm 2). The final S_O is obtained by multiplying the rotation matrix of sampled S_{θ_g} , S_{ϕ_g} , and S_{γ_g} .

Algorithm 2 Sample gripper orientation

Input: n_{pcd} : point normals

I_{aff} : affordance map

$\Delta\alpha$: sampling interval

Output: S_O : gripper orientation samples

```

1:  $n \leftarrow n_{pcd}[I_{aff} > 0]$ 
2:  $\theta \leftarrow \arctan(n_y, n_x)$ 
3:  $\phi \leftarrow \arccos(n_z)$ 
4:  $ii \leftarrow \frac{\theta + \pi}{\Delta\alpha}$ 
5:  $jj \leftarrow \frac{\phi}{\Delta\alpha}$ 
6:  $ii, jj \leftarrow \text{Unique}(ii, jj)$ 
7:  $ii', jj' \leftarrow \mathcal{M}(ii, jj)$ 
8:  $S_{\theta_g} \leftarrow ii' \Delta\alpha - \pi$ 
9:  $S_{\phi_g} \leftarrow jj' \Delta\alpha$ 
10:  $kk' \leftarrow 0, 1, \dots, \frac{\pi}{\Delta\alpha}$ 
11:  $S_{\gamma_g} \leftarrow kk' \Delta\alpha - \pi$ 
12:  $S_O \leftarrow R_z(S_{\theta_g})R'_y(S_{\phi_g})R''_z(S_{\gamma_g})$ 
13: return  $S_O$ 

```

5.5. Gripper Orientation Kernel Generation and Suction Cup ID Encoding

The kernel representing each candidate gripper orientation generated in Section 5.4 is created using 3D convolution to determine the graspable position for each S_O , as in Algorithm 3. A binary kernel is used to represent gripper poses in previous studies using 2D convolution [39]. However, the convolution results can only determine the graspable position of the kernel and cannot directly determine which suction cup to activate. For example, as shown in Figure 5, although the convolution results for the four cases are the same, the suction cups chosen to be activated are different and cannot be directly determined from the convolution results. Hence, we design a 3D kernel that includes the suction cup ID information. Algorithm 3 is used to generate kernels S_K of S_O . The shape of one kernel $\mathcal{K}$ is $N_{K_x} \times N_{K_y} \times N_{K_z}$, where $N_{K_x} = N_{K_y} = N_{K_z} = \lfloor \frac{\max(d_{C_i}^0)}{l} \rfloor$. $\max(d_{C_i}^0)$ is the maximum distance of the suction cup center to the TCP in local gripper coordinates, and l is the grid size of the kernel, which is equal to that of the voxel grid. The kernel indices of cup centers are $\lfloor \frac{C}{l} \rfloor + \lfloor \frac{\max(d_{C_i}^0)}{2l} \rfloor$, where C is the cup center position of S_O . The kernel grids at cup center kernel indices are filled with encoded vacuum ID information as in line 9 in

Algorithm 3. Here, the i th suction cup ID information is encoded as 10^{-i} , such that the cup ID is saved in the i th decimal place, and such encoding helps to directly obtain the target suction cups to activate from the decoding convolution results (see Section 5.7).

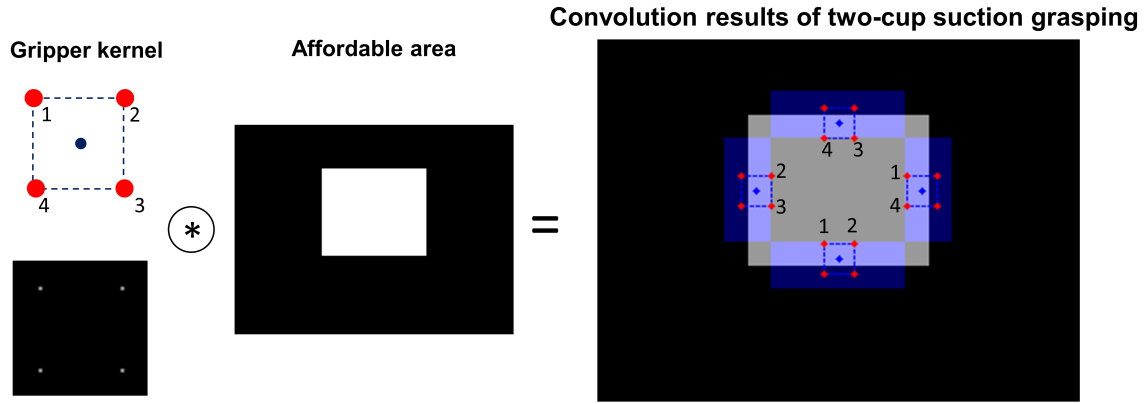


Figure 5. The problem of using a binary kernel. Red dots are cup centers, and blue dots are TCP positions. The transparent blue area indicates graspable positions using two cups. The convolution results for two-cup suction grasping are the same for the four cases in which the convolved values for the four cases are all equal to 2. However, although the suction cup center positions in the affordable area are different so that the cups to activate differ between the four cases, the activation pattern cannot be directly determined from the convolution result.

Algorithm 3 Generated encoded kernels

Input: S_O : gripper orientation samples

C^0 : local cup center positions

(see the right side of Figure 2)

l : voxel size

Output: S_K : kernels of S_O

```

1:  $N_O \leftarrow \text{len}(S_O)$ 
2:  $S_K \leftarrow \text{Zeros}(N_O, N_{K_x}, N_{K_y}, N_{K_z})$ 
3: for  $n \leftarrow 0$  to  $N_O$  do
4:    $K \leftarrow \text{Zeros}(N_{K_x}, N_{K_y}, N_{K_z})$ 
5:   for  $i \leftarrow 0$  to  $N_c$  do
6:      $[C[i], 1] \leftarrow \begin{bmatrix} S_O[n] & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} C^0[i]^T \\ 1 \end{bmatrix}$ 
7:      $C[i] \leftarrow C[i]^T$ 
8:     # encoding
9:      $K[\lfloor \frac{C[i]}{l} \rfloor + \lfloor \frac{\max(d_{C_i}^0)}{2l} \rfloor] \leftarrow 10^{-i}$ 
10:   end for
11:    $S_K[n] \leftarrow K$ 
12: end for
13: return  $S_K$ 

```

5.6. Three-Dimensional Convolution

We perform 3D convolution to determine the indices in V , where the gripper can grasp using multiple suction cups. Because the kernel is generated from an oriented C^0 that is located in the same plane as the TCP, the corresponding kernel indices of cup centers and the TCP are in the same plane, which satisfies Condition 2 in Section 4. Furthermore, as the distances from cup centers to the TCP are represented in a kernel scale that is the same as the voxel grid scale, we can slide the kernel over the voxel grid to determine the voxel grid index where the TCP satisfies Conditions 2 and 4. Specifically, as in Equation (7), the kernel

is set to each grid cell of the voxel grid to calculate the convolution sum. Note that N_K is the number of kernels, which is equal to N_O .

$$\text{ConvRes} = \sum_{n=0}^{N_K} \sum_{m=0}^{N_{V_x}} \sum_{t=0}^{N_{V_y}} \sum_{p=0}^{N_{V_z}} \sum_{i=-\frac{N_{K_x}}{2}}^{\frac{N_{K_x}}{2}} \sum_{j=-\frac{N_{K_y}}{2}}^{\frac{N_{K_y}}{2}} \sum_{k=-\frac{N_{K_z}}{2}}^{\frac{N_{K_z}}{2}} \mathcal{K}[i, j, k] \cdot V[m + i, t + j, p + k]. \quad (7)$$

5.7. Convolution Results Decoding and Normal Direction Check

Algorithm 4 shows the decode function that decodes the 3D convolution results (ConvRes) to generate grasp candidates. As the 3D convolution of the kernel center is set to each grid cell of V and accumulates the kernel values where the corresponding voxel grid value is non-zero (Equation (7)), the cup to be activated can be determined by obtaining each digit of ConvRes . As in line 7 in Algorithm 4, ConvRes is decoded to target i th suction cup activation a_i in Equation (1) by scaling up ConvRes by 10^i times and then calculating the value mod 10. If a_i is 1, a contact point exists for the i th vacuum cup that should be activated. Otherwise, there is no contact point and the cup should be disabled. For example, for the gripper with two suction cups in Figure 2, there are four (2^2) possible values of convolution results: 0.00, 0.10, 0.01, and 0.11, and the decoding results are [0, 0], [1, 0], [0, 1], and [1, 1], indicating non-graspable, graspable for only the first cup, graspable for only the second cup, and graspable for both cups, respectively.

Algorithm 4 Decode

Input: ConvRes : 3D convolution results
Output: G_{cand} : grasp candidates for multiple-cup suction

- 1: $N_K, N_{V_x}, N_{V_y}, N_{V_z} \leftarrow \text{convRes.shape}$
- 2: $A \leftarrow \text{Zeros}(N_K, N_{V_x}, N_{V_y}, N_{V_z}, N_c)$
- 3: **for** Res in ConvRes **do**
- 4: **for** $i \leftarrow 0$ to N_c **do**
- 5: # decoding cup ID
- 6: $A[... , i] \leftarrow \lfloor \frac{10^{N_{cup}}}{10^{N_{cup}-i}} Res \rfloor \bmod 10$
- 7: **end for**
- 8: **end for**
- 9: $validInd \leftarrow \text{sum}(A, \text{dim} = -1) \geq 2$
- 10: $S_P \leftarrow validInd \cdot l + B^{min}$
- 11: $S_O \leftarrow S_O[validInd]$
- 12: $S_C \leftarrow \begin{bmatrix} S_O & S_P \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} C^0 T \\ 1 \end{bmatrix}$
- 13: $S_A \leftarrow A[validInd]$
- 14: $G_{cand} \leftarrow [S_P, S_O, S_C, S_A]$
- 15: **return** G_{cand}

As A is a one-hot vector, the sum of A is the number of suction cups to be used. Therefore, we determine the indices ($validInd$) of V where the sum of A is greater than or equal to two ($\text{sum}(A, \text{dim} = -1) \geq 2$) in order to find the voxel grid indices where multiple suction cups can be used to grasp multiple objects or an object with a large surface area. $validInd$ is further converted to TCP positions in world coordinates (S_P in Algorithm 4), and the corresponding orientation S_P , cup center positions S_C , and target activation status S_A can be obtained to generate the grasp candidates (G_{cand}), as in lines 12–15 in Algorithm 4.

The normal directions of all activated cups ($a_i=1$) of G_{cand} are checked to satisfy Condition 3. Specifically, the closest point to the contact point of each activated cup is

searched for in I_{pcd} , and then the normal of that point is checked for whether it is in the same direction as the gripper's z-axis direction using Equation (2).

5.8. Ranking

Each G_{cand} is evaluated and ranked to determine the optimal grasp G_{opt} . We first perform point clustering on the points with non-zero affordance values, which are extracted from I_{pcd} to generate a label map M_{label} , distance map M_{dist} , and orientation map M_{orient} , as shown in Figure 6. M_{label} contains the ID label of each cluster and is later used to calculate how many objects can be grasped. M_{dist} contains the 3D distance from each point in the cluster to the cluster center. M_{orient} contains the 3D orientation of each cluster. M_{dist} and M_{orient} are generated for later evaluation of the score (J) of G_{opt} . The height and width of the three maps are the same as those of I_d .

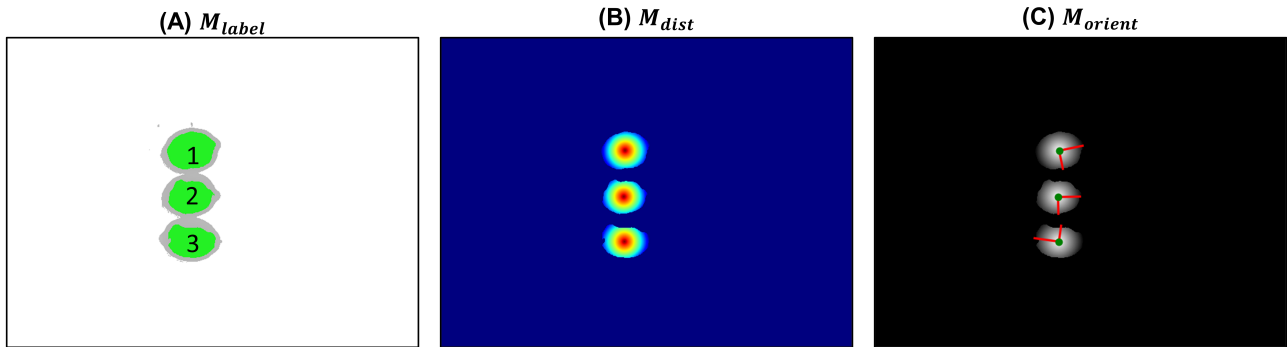


Figure 6. Clustering results. (A) Label map M_{label} . (B) Distance map M_{dist} . (C) Orientation map M_{orient} . Numbers in (A) is the cluster Id. (B) is the heatmap of the distance to the center of the clustered area. Green points and red lines in (C) are center positions and axes of clustered areas.

Lines 3–14 in Algorithm 5 evaluate the maximum number of objects that can be grasped $maxObj$ and the score J of each G in G_{cand} , saving the evaluation results to a dictionary ($rankingRes$). The image coordinates (u_c, v_c) of cup center positions are calculated to obtain the contact point label for each cup in the M_{label} . Note that the contact labels might not be unique. If several cups have the same contact point label, it means that these cups can be used to grasp the same object, which has a large surface area. If all cups have different contact point labels, it means that each cup can grasp a unique object. Therefore, the number of unique contact labels is the maximum number of objects that can be grasped by G . J is the sum of J_{center} , J_{var} , and J_{orient} . J_{center} evaluates the distance from the cup center or the average of cup centers to the cluster center because it is assumed to be more stable to hold the object at its center. As in Equation (8) and Figure 7A, J_{dist} is evaluated as the average sum of distances from cups to the cluster center by obtaining the value of M_{dist} at the position of the average cup center position ($avg(v_c, u_c, contactLabel_i)$) in each cluster.

$$J_{dist} = \frac{\sum_{i=0}^{N_{contactLabel}} M_{dist}[avg(v_c, u_c, contactLabel_i)]}{maxObj}. \quad (8)$$

J_{var} has been incorporated because there are cases where one cup is near, but another cup is far from the cluster center, and J_{dist} cannot evaluate these cases. J_{var} is used to balance the distances of cups to cluster center positions. Specifically, as in Equation (9) and Figure 7B, J_{var} evaluates the variance of the M_{dist} value at the position of average cup center positions.

$$J_{var} = \frac{\sum_{i=0}^{N_{contactLabel}} (M_{dist}[avg(v_c, u_c, contactLabel_i)] - J_{dist})^2}{maxObj}. \quad (9)$$

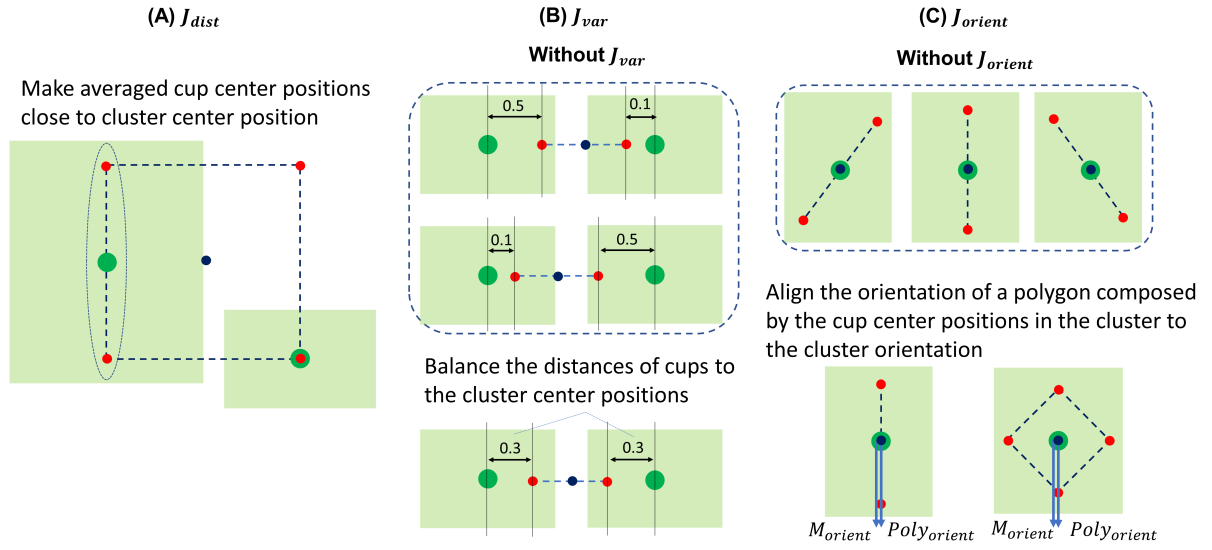


Figure 7. Metrics to evaluate G_{cand} . Red dots are cup centers and blue dots are TCP positions. Green areas are clusters of affordable areas. Green dots are cluster centers. (A) Distance score J_{dist} . (B) Distance variation score J_{var} . (C) Orientation score J_{orient} .

Algorithm 5 Ranking

Input: I_{aff} : affordance map
 I_{pcd} : point cloud
 G_{cand} : grasp candidates
Output: G_{opt} : optimal grasp

```

1:  $M_{label}, M_{dist}, M_{orient} \leftarrow \text{clustering}(I_{aff}, I_{pcd})$ 
2:  $rankingRes \leftarrow \text{Dict}()$ 
3: for  $G$  in  $G_{cand}$  do
4:    $P, O, C, A \leftarrow G$ 
5:    $u_{tcp}, v_{tcp} \leftarrow \text{getImgCoord}(P)$ 
6:    $u_c, v_c \leftarrow \text{getImgCoord}(C)$ 
7:    $contactLabel \leftarrow \text{unique}(M_{label}[v_c, u_c])$ 
8:    $maxObj \leftarrow \text{len}(contactLabel)$ 
9:    $J_{center} \leftarrow \text{calcCenterScore}(M_{dist}, u_c, v_c, contactLabel)$ 
10:   $J_{var} \leftarrow \text{calcVarScore}(M_{var}, u_c, v_c, contactLabel)$ 
11:   $J_{orient} \leftarrow \text{calcOrientScore}(M_{orient}, u_c, v_c, contactLabel)$ 
12:   $J \leftarrow J_{center} + J_{var} + J_{orient}$ 
13:  add  $[maxObj, J]$  to  $rankingRes[contactLabel]$ 
14: end for
15:  $rankingRes \leftarrow \text{sort}(rankingRes)$ 
16: for  $Res$  in  $rankingRes$  do
17:   if  $maxObj$  in  $Res > maxObj$  in  $G_{opt}$  then
18:      $G_{opt} \leftarrow Res$ 
19:   else if  $maxObj$  in  $Res = maxObj$  in  $G_{opt}$  then
20:     if  $J$  in  $Res > J$  in  $G_{opt}$  then
21:        $G_{opt} \leftarrow Res$ 
22:     end if
23:   end if
24: end for
25: return  $G_{opt}, rankingRes$ 

```

J_{orient} has been incorporated to align the orientation of a polygon composed of the cup center positions in the cluster to the cluster orientation. Specifically, we calculate the dot product between the cluster orientation (longer or short axis unit vector) and the polygon orientation, as shown in Equation (10) and Figure 7C.

$$J_{orient} = \frac{\sum_{i=0}^{N_{contactLabel}} M_{orient}[\mathbf{v}_c, \mathbf{u}_c] \cdot Poly_{orient}(\mathbf{v}_c, \mathbf{u}_c, contactLabel_i)}{maxObj}. \quad (10)$$

The G and corresponding $maxObj$ and J are added to the dictionary using *contactLabel* as a key. Key level (local level) sorting is first performed to sort stored J (line 15 in the algorithm). Next, dictionary-level (global level) sorting is performed to determine G_{opt} with the highest $maxObj$ and J (lines 16–24 in the algorithm). Note that both G_{opt} and sorted *rankingRes* are returned because, if the motion planner fails to find a trajectory to G_{opt} , it would search for the trajectory to other goals with high $maxObj$ and J in *rankingRes*.

6. Experiments

The multiple-suction-cup planner was validated using previous affordance map datasets as well as real picking experiments. For both validations, thresholds ε_1 in Equation (2) and ε_2 in Equation (3) were set to 11.5° and 0.01 m, respectively. The voxel grid size l was set to 0.005 m. The angle sampling interval $\Delta\alpha$ was set to 5° . Validations were performed on an Ubuntu 20.04 PC with an 11th Gen Intel Core™ i7-11700K @ 3.60 GHz $\times$ 16 CPU and NVIDIA GeForce RTX 3060 GPU.

6.1. Validation Using a Previous Affordance Map Dataset

We used three datasets to validate the generality of the multiple-suction-cup grasp planner: Suction FCN [12], SuctionNet-1Billion [13], and SG-U-Net++ [14]. These datasets provide real RGB-D or synthesized depth images and the corresponding affordance maps. Point clouds converted from depth images and affordance maps in the dataset were used to determine the optimal multiple-cup graspable poses and the cups to activate two-cup and four-cup vacuum grippers. The accuracy of the position and orientation were evaluated by the average error of Equation (2) and Equation (3), respectively.

6.2. Validation Using the Picking Experiment

To evaluate the robot picking system and efficiency improvement using the multiple-suction-cup grasp planner, we conducted picking experiments and compared the results of single-cup (single-object) grasping and multiple-cup (multiple-object) grasping. The robot with a two-cup vacuum gripper was used to pick items from a bin and then place them into a tote (Figure 8A). Figure 8C shows the gripper architecture. The gripper was equipped with PISCO (PISCO CO., LTD., Nagano, Japan) standard suction cups. We used a cup with a 20 mm diameter because it can generate sufficient force to grasp and hold objects (weight less than 500 g) used in experiments under our vacuum pump setup. The cups were connected to vacuum ejector ports by tubes. Each ejector port controlled the on/off of the corresponding cup by supplying (on) or blocking (off) compressed air. A camera was installed in the center of the gripper, which captured the depth image at the robot's home position. The affordance map was then inferred by SG-U-Net++ based on the depth image. For single-object grasping, the planner in our previous work [14] was used to find the position of maximum affordance value and selected the suction cup that can be used to reach the target grasp point by the shortest trajectory. For multiple-object grasping, a multiple-suction-cup grasp planner was first used to determine grasp poses capable of grasping multiple objects or an object with a large surface area using multiple cups. If there was no solution, the planner was switched to the planner for single-object grasping. Trajectories from the home position to grasp poses were generated by MoveIt. As shown in Figure 8B, the target object set included boxes, fruits, and daily necessities. The robot was required to pick 50 boxes, 50 fruits, and 51 daily necessities in a cluttered scene. The robot

continued grasp attempts until the scene was cleared. A grasp attempt was considered to have failed if the robot could not pick up the item or the item was dropped during the movement of the manipulator. The results of single-object grasping and multiple-object grasping were evaluated and compared by success rate, picks per hour (PPH), and number of pick attempts. The success rate was defined as the number of successful attempts divided by the number of pick attempts. PPH was defined as the number of objects the robot could pick up in 1 h. The number of pick attempts was defined as a number of attempts for the robot to clear the cluttered scene.

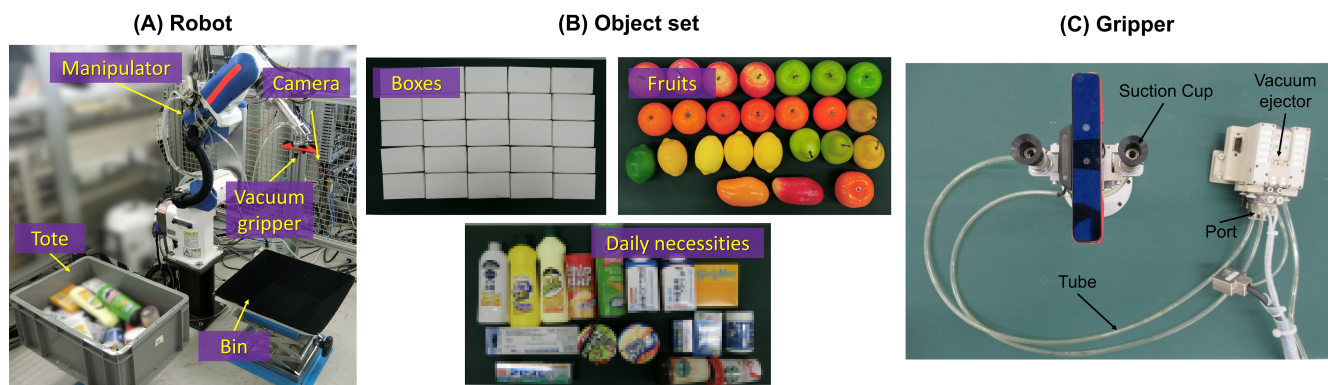


Figure 8. Experiment setup. (A) Robot. (B) Object set. (C) Gripper.

7. Results and Discussion

To our knowledge, this study is the first to propose a grasp planner for multiple-suction-cup grippers to grasp multiple objects or an object with a large surface area. Most of the previous studies used a deep neural network to infer the affordance map for finding the optimal grasp for single-cup grasping. Our planner took advantage of the affordance map to determine the optimal grasp for multiple-cup grasping. The planner was validated on three previous affordance map datasets, and the results are shown in Table 1. Our planner successfully found multiple-suction-cup grasps from the affordance map from Suction FCN, SuctionNet-1Billion, and SG-U-Net++, indicating the high generality of the planner. There were no significant differences in position orientation error between the two-cup and four-cup gripper planning results. The error was the smallest when grasping was planned based on the affordance map from SG-U-Net++ because SG-U-Net++ used synthesized data (e.g., depth image and point cloud normals) without noise values. Figures 9 and 10 show examples of grasp planning results for the two-cup and four-cup vacuum grippers. The planner successfully determined which of the cups to activate when grasping.

Table 1. Position and orientation error of grasp pose.

Dataset	Two-Cup Gripper		Four-Cup Gripper	
	Position Error (Mean (SD) [m])	Orientation Error (Mean (SD) [deg.])	Position Error (Mean (SD) [m])	Orientation Error (Mean (SD) [deg.])
Suction FCN	6.28×10^{-3} (0.80×10^{-4})	4.50 (6.94)	5.94×10^{-3} (0.23×10^{-4})	5.04 (11.00)
suctionNet-1Billion	7.67×10^{-3} (2.43×10^{-4})	4.66 (6.30)	7.64×10^{-3} (1.41×10^{-4})	4.59 (6.98)
SG-U-Net++	2.88×10^{-3} (0.18×10^{-4})	2.85 (10.2)	2.30×10^{-3} (0.07×10^{-4})	2.68 (8.07)

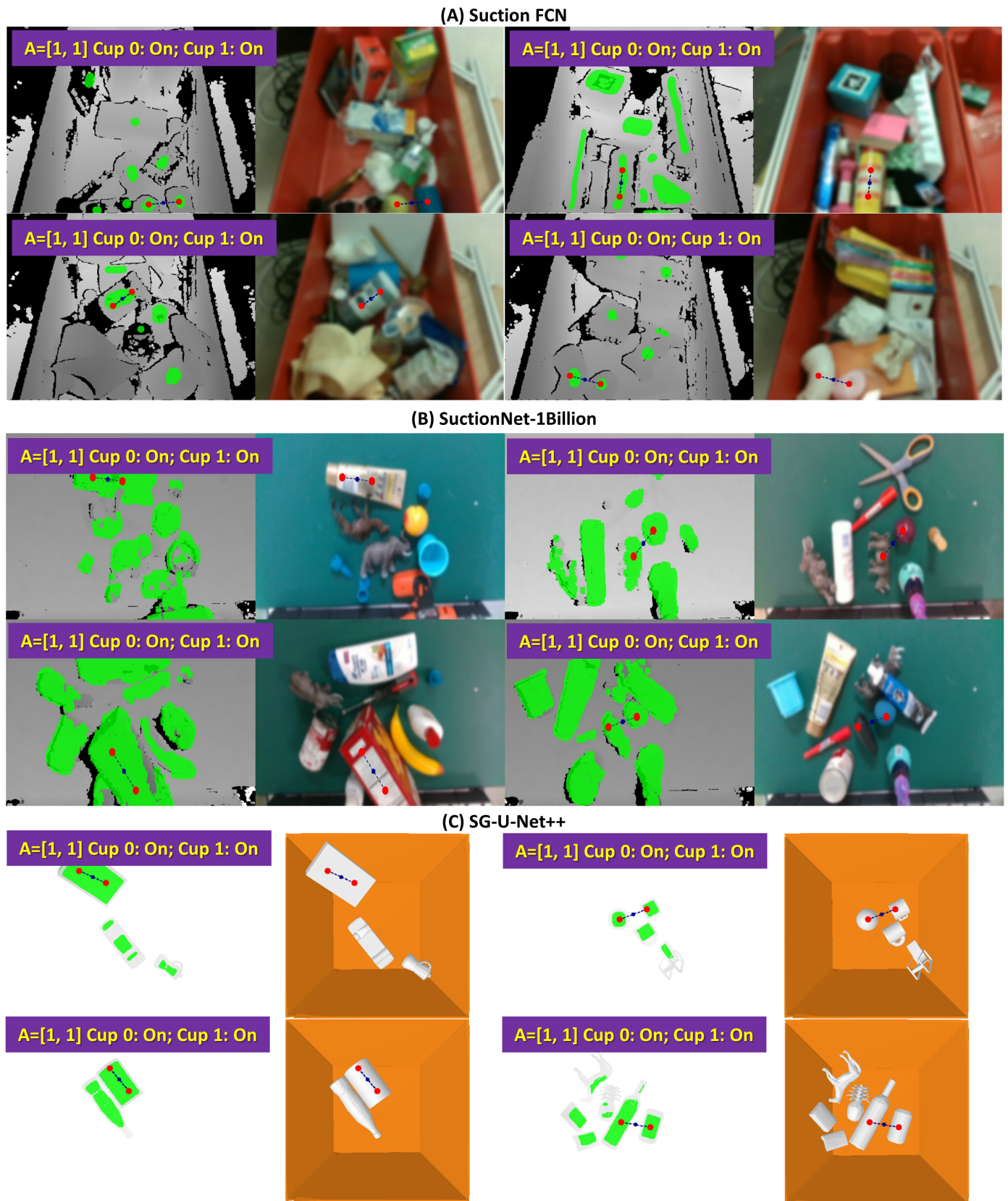


Figure 9. Examples of grasp planning results for two-cup and four-cup vacuum grippers. (A) Suction FCN. (B) SuctionNet-1Billion. (C) SG-U-Net++. Red dots are the center positions of the cups. Dark blue dots are the center positions of the gripper.

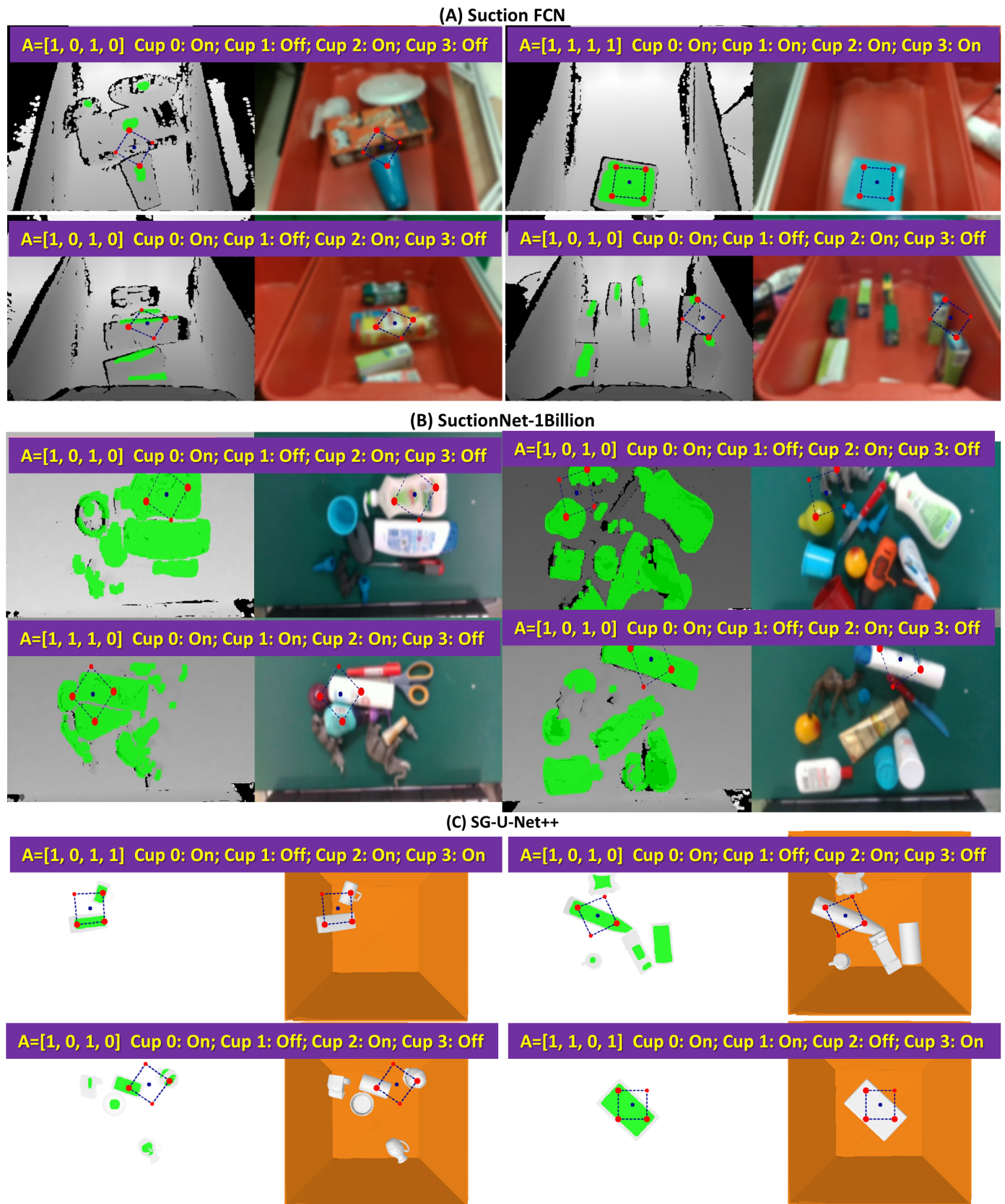


Figure 10. Examples of grasp planning results for two-cup and four-cup vacuum grippers. (A) Suction FCN. (B) SuctionNet-1Billion. (C) SG-U-Net++. Large red spots are centers of activated cups. Small red spots are centers of disabled cups. Dark blue dots are the center positions of the the gripper.

The physical experiment results show that multiple-cup suction grasping could improve the efficiency of picking tasks. Table 2 shows a comparison of experimental results between single-cup (single-object) and multiple-cup (multiple-object) grasping. For single-object grasping, all three object sets could be cleared by the robot. Daily necessities had the highest success rate (91%) and highest PPH (502) among the three object sets. The success rate of picking fruits was the lowest because the objects had a ball-like shape and rolled and slipped when the gripper pushed them along the normal direction during grasping despite having the correct grasp pose. The success rate of picking boxes was lower than that of daily necessities because when two boxes were very close together, the planner treated them as a single box and grasped the center, which was actually the edge between two boxes. This problem did not occur in the case of multiple-suction-cup grasping because even when two boxes were treated as a single big box, the planner set the averaged cup center positions to the center of the affordable area, as shown in Figure 7, so that the cups did not suck the edge between boxes. For multiple-object grasping, all three object sets could also be cleared by the robot. The success rate for grasping boxes (100%) was the highest among the object sets. The robot picked fruits with the highest speed (PPH = 779). Multiple-object grasping improved the picking speed by $1.45\times$ for boxes (PPH: 467 vs. 677), $1.65\times$ for fruits (PPH: 472 vs. 779), and $1.16\times$ for daily necessities (PPH: 502 vs. 583). These results indicate that multiple-suction-cup grasping can improve picking speed. The improvement in picking daily necessities was minor because it was difficult to find multiple-cup graspable poses due to the complicated shapes of the items. Figure 11 shows one picking trial for multiple-suction-cup grasping of boxes, fruits, and daily necessities. More trials are shown in the Supplementary Materials Video S1 file.

Table 2. Experimental results.

Object Set	Method	Total Attempts	Successful Attempts	Success Rate	PPH
Boxes	Single-object grasping	59	50	85%	468
	Multiple-object grasping	36	36	100%	677
Fruits	Single-object grasping	64	50	78%	472
	Multiple-object grasping	33	31	94%	779
Daily necessities	Single-object grasping	56	51	91%	502
	Multiple-object grasping	53	40	75%	583

In this study, the object sets are lightweight items. For the fruit object set, we used lightweight plastic fruit samples instead of real fruit to avoid crushing them. We mainly focused on the correct suction poses and assumed that the suction cup can generate enough force to hold the object if the MSC grasp pose is correct. Although we believe that the maximum load capacity of the selected suction cup is 500 g and that the cup can hold most of the common daily necessities, the influence of object weight on the planner result will be validated using objects with various weights (e.g., using real fruit with different weight) in the future. In addition, the objects in the experiments had rather large and flat surfaces, such that the formulated seals were circular shapes and all contact points between the cup and the surface were in the same plane. In this case, cups are assumed to be capable of fully generating the suction force. However, for a complicated object shape, the seal of a cup might be a polygon (see Figure 1 in [18]). In this case, a more precise contact force model should be used to evaluate the grasp quality (grasp success probability) of the cups, but this will also lead to an increase in planning time. Finding an efficient method to evaluate the contact force for MSC grasp candidates is one of our future works to improve the planner so it is applicable for grasping more complex shape objects in dense scenes.

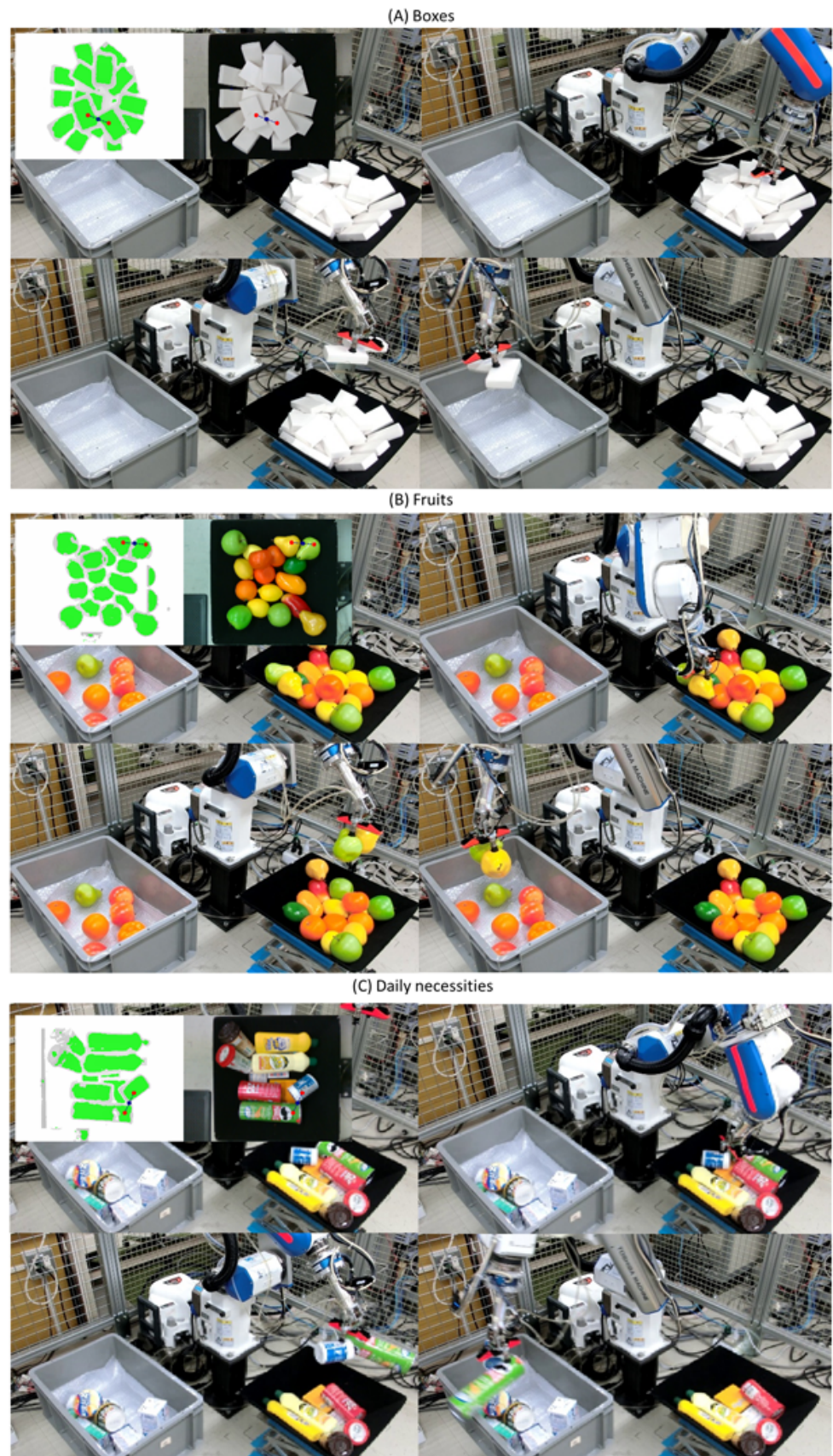


Figure 11. One picking trial for the multiple-suction-cup grasp of (A) boxes, (B) fruits, and (C) daily necessities. Red dots are the center positions of the cups. Dark blue dots are the center positions of the gripper.

The gripper dimensions will affect the planner in two ways. Firstly, the increase in dimensions will lead to an increase in the volume of the gripper kernels in Figure 4, which will further result in an increase in planning time. Therefore, proper kernel interval size should be determined based on gripper dimension. In this study, the gripper kernel interval size was determined empirically. Automatic determination of the size based on gripper dimension will be considered to make the planner applicable to grippers with different dimensions. Secondly, if the gripper is too large, it will easily collide with the neighboring objects or bins so it will be difficult for the planner to find solutions for MSC grasping.

The picking system is expected to be improved in future work aimed at further increasing the picking speed. Our current layout of cups is fixed, and the positions and orientations of cups are fixed in the gripper. However, the real-world scenario is complex in that the items may have complex shapes, and the poses of items are disorganized. Hence, a fixed cup configuration may lead to a low probability of finding MSC grasp poses in a very complex environment (e.g., a dense scene) because there are few candidates who satisfy the MSC grasp condition. Making the positions and poses of cups adjustable by a proper gripper mechanical design may increase the probability of finding MSC grasp poses, which could be one of the directions in our future work to further improve the picking efficiency. In addition, as described above, one common failure is that objects can move (e.g., roll) after being grasped. We intend to analyze the dynamics (e.g., object shape, friction, and contact force between items) after grasping to find a grasp that moves the object and neighboring objects such that grasp success is improved. Another area for improvement is depth filling because incomplete depth results in low accuracy in estimating the affordance map and normals and thus leads to low grasp success. Furthermore, we will consider the picking sequence to improve the possibility of picking multiple objects. Picking experiments using a gripper with more suction cups will be conducted to further validate the performance and application of the planner for MSC grasp in the future. Our system cannot detect whether or not the gripper has successfully grasped the object. There are no sensors to detect whether or not the seal has been successfully formulated. Using a force sensor to detect the grasp success will be conducted in the future. If the force load of the end-effector increases after grasping, the object will be considered to be successfully grasped by the gripper. Based on the success detection result, the robot will be required to re-grasp a certain object if grasp failure occurs without returning to its home position to improve the picking efficiency.

8. Conclusions

In this study, we proposed a grasp planner for a multiple-suction-cup vacuum gripper. The planner took advantage of an affordance map to determine grasp poses for multiple-cup grasping using a 3D convolution-based method. Thanks to the encoded cup ID kernel, the planner could directly determine which cups to activate by decoding the convolution results. The planner exhibited good generality on previous affordance map datasets. The planner also showed the ability to improve picking speed compared with single-cup grasping in physical experiments with a real robot. We will work on improving the planner in future work from several directions, including object state analysis after grasping, point cloud or depth image completion, and picking sequence planning.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/robotics13060085/s1>, Video S1: video.mp4.

Author Contributions: Conceptualization, P.J., J.O. (Junji Oaki), Y.I. and J.O. (Junichiro Ooga); methodology, P.J.; software, P.J.; validation, P.J. and J.O. (Junichiro Ooga); formal analysis, P.J.; investigation, P.J.; resources, P.J.; data curation, P.J.; writing—original draft preparation, P.J.; writing—review and editing, P.J., J.O. (Junji Oaki), Y.I. and J.O. (Junichiro Ooga); visualization, P.J.; supervision, Y.I. and J.O. (Junichiro Ooga); project administration, Y.I. and J.O. (Junichiro Ooga). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets and source code presented in this article are not readily available due to the policy of the authors' institution. Requests to access the datasets should be directed to the corresponding author.

Conflicts of Interest: Ping Jiang, Junji Oaki, Yoshiyuki Ishihara, and Junichiro Ooga are employees of the Toshiba Corporation. Ping Jiang and Yoshiyuki Ishihara are inventors of a patent related to this work. US Patent No.: 11691275. B2; Patent Date: 4 July 2023; Name: HANDLING DEVICE AND COMPUTER PROGRAM PRODUCT.

References

1. Bogue, R. Growth in e-commerce boosts innovation in the warehouse robot market. *Ind. Robot. Int. J.* **2016**, *43*, 583–587. [\[CrossRef\]](#)
2. Yu, Y.; Fukuda, K. Analysis of multifinger grasp internal forces for stably grasping multiple polyhedral objects. *Int. J. Mechatronics Autom.* **2013**, *3*, 203–216. [\[CrossRef\]](#)
3. Yamada, T.; Yamada, M.; Yamamoto, H. Stability analysis of multiple objects grasped by multifingered hands with revolute joints in 2D. In Proceedings of the 2012 IEEE International Conference on Mechatronics and Automation, Chengdu, China, 5–8 August 2012; pp. 1785–1792.
4. Sakamoto, T.; Wan, W.; Nishi, T.; Harada, K. Efficient picking by considering simultaneous two-object grasping. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; pp. 8295–8300.
5. Chen, T.; Shenoy, A.; Kolinko, A.; Shah, S.; Sun, Y. Multi-object grasping—estimating the number of objects in a robotic grasp. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; pp. 4995–5001.
6. Shenoy, A.; Chen, T.; Sun, Y. Multi-Object Grasping—Generating Efficient Robotic Picking and Transferring Policy. *arXiv* **2021**, arXiv:2112.09829.
7. Agboh, W.C.; Ichnowski, J.; Goldberg, K.; Dogar, M.R. Multi-object Grasping in the Plane. In *Robotics Research*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 222–238.
8. Agboh, W.C.; Sharma, S.; Srinivas, K.; Parulekar, M.; Datta, G.; Qiu, T.; Ichnowski, J.; Solowjow, E.; Dogar, M.; Goldberg, K. Learning to Efficiently Plan Robust Frictional Multi-Object Grasps. *arXiv* **2022**, arXiv:2210.07420.
9. Jiang, P.; Ishihara, Y.; Sugiyama, N.; Oaki, J.; Tokura, S.; Sugahara, A.; Ogawa, A. Depth image-based deep learning of grasp planning for textureless planar-faced objects in vision-guided robotic bin-picking. *Sensors* **2020**, *20*, 706. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Pattar, S.P.; Hirakawa, T.; Yamashita, T.; Sawanobori, T.; Fujiyoshi, H. Single Suction Grasp Detection for Symmetric Objects Using Shallow Networks Trained with Synthetic Data. *IEICE Trans. Inf. Syst.* **2022**, *105*, 1600–1609. [\[CrossRef\]](#)
11. Araki, R.; Hirakawa, T.; Yamashita, T.; Fujiyoshi, H. MT-DSSD: Multi-task deconvolutional single shot detector for object detection, segmentation, and grasping detection. *Adv. Robot.* **2022**, *36*, 373–387. [\[CrossRef\]](#)
12. Zeng, A.; Song, S.; Yu, K.T.; Donlon, E.; Hogan, F.R.; Bauza, M.; Ma, D.; Taylor, O.; Liu, M.; Romo, E.; et al. Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching. *Int. J. Robot. Res.* **2022**, *41*, 690–705. [\[CrossRef\]](#)
13. Cao, H.; Fang, H.S.; Liu, W.; Lu, C. Suctionnet-1billion: A large-scale benchmark for suction grasping. *IEEE Robot. Autom. Lett.* **2021**, *6*, 8718–8725. [\[CrossRef\]](#)
14. Jiang, P.; Oaki, J.; Ishihara, Y.; Ooga, J.; Han, H.; Sugahara, A.; Tokura, S.; Eto, H.; Komoda, K.; Ogawa, A. Learning suction graspability considering grasp quality and robot reachability for bin-picking. *Front. Neurorobotics* **2022**, *16*, 806898. [\[CrossRef\]](#)
15. Lenz, I.; Lee, H.; Saxena, A. Deep learning for detecting robotic grasps. *Int. J. Robot. Res.* **2015**, *34*, 705–724. [\[CrossRef\]](#)
16. Xu, R.; Chu, F.J.; Vela, P.A. Gknet: Grasp keypoint network for grasp candidates detection. *Int. J. Robot. Res.* **2022**, *41*, 361–389. [\[CrossRef\]](#)
17. Yu, S.; Zhai, D.H.; Xia, Y.; Wu, H.; Liao, J. SE-ResUNet: A novel robotic grasp detection method. *IEEE Robot. Autom. Lett.* **2022**, *7*, 5238–5245. [\[CrossRef\]](#)
18. Mahler, J.; Matl, M.; Liu, X.; Li, A.; Gealy, D.; Goldberg, K. Dex-net 3.0: Computing robust vacuum suction grasp targets in point clouds using a new analytic model and deep learning. In Proceedings of the 2018 IEEE International Conference on robotics and automation (ICRA), Brisbane, QLD, Australia, 21–25 May 2018; pp. 5620–5627.
19. Utomo, T.W.; Cahyadi, A.I.; Ardiyanto, I. Suction-based Grasp Point Estimation in Cluttered Environment for Robotic Manipulator Using Deep Learning-based Affordance Map. *Int. J. Autom. Comput.* **2021**, *18*, 277–287. [\[CrossRef\]](#)
20. Hasegawa, S.; Wada, K.; Kitagawa, S.; Uchimi, Y.; Okada, K.; Inaba, M. Graspfusion: Realizing complex motion by learning and fusing grasp modalities with instance segmentation. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; pp. 7235–7241.
21. Morrison, D.; Corke, P.; Leitner, J. Learning robust, real-time, reactive robotic grasping. *Int. J. Robot. Res.* **2020**, *39*, 183–201. [\[CrossRef\]](#)

22. Kumra, S.; Joshi, S.; Sahin, F. Antipodal robotic grasping using generative residual convolutional neural network. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 24 October 2020–24 January 2021; pp. 9626–9633.
23. Le, T.N.; Lundell, J.; Abu-Dakka, F.J.; Kyrki, V. Deformation-Aware Data-Driven Grasp Synthesis. *arXiv* **2021**, arXiv:2109.05320.
24. Cao, H.; Chen, G.; Li, Z.; Lin, J.; Knoll, A. Lightweight convolutional neural network with Gaussian-based grasping representation for robotic grasping detection. *arXiv* **2021**, arXiv:2101.10226.
25. Kumra, S.; Joshi, S.; Sahin, F. GR-ConvNet v2: A Real-Time Multi-Grasp Detection Network for Robotic Grasping. *Sensors* **2022**, *22*, 6208. [[CrossRef](#)] [[PubMed](#)]
26. Depierre, A.; Dellandréa, E.; Chen, L. Jacquard: A large scale dataset for robotic grasp detection. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 3511–3516.
27. Mahler, J.; Matl, M.; Satish, V.; Danielczuk, M.; DeRose, B.; McKinley, S.; Goldberg, K. Learning ambidextrous robot grasping policies. *Sci. Robot.* **2019**, *4*, eaau4984. [[CrossRef](#)]
28. Kensuke, H.; Makoto, K. Enveloping grasp for multiple objects. In Proceedings of the 1998 IEEE International Conference on Robotics and Automation (ICRA), Leuven, Belgium, 20 May 1998; pp. 2409–2415.
29. Takayoshi, Y.; Hidehiko, Y.; Tsuji, T. Rolling-Based Manipulation for Multiple Objects. *IEEE Trans. Robot. Autom.* **2000**, *16*, 457–468.
30. Takayoshi, Y.; Hidehiko, Y. Static Grasp Stability Analysis of Multiple Spatial Objects. *J. Control Sci. Eng.* **2015**, *3*, 118–139. [[CrossRef](#)]
31. Mucchiani, C.; Yim, M. A novel underactuated end-effector for planar sequential grasping of multiple objects. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 8935–8941.
32. Yao, K.; Billard, A. Exploiting Kinematic Redundancy for Robotic Grasping of Multiple Objects. *IEEE Trans. Robot.* **2023**, *39*, 1982–2002. [[CrossRef](#)]
33. Mantriota, G. Optimal grasp of vacuum grippers with multiple suction cups. *Mech. Mach. Theory* **2007**, *42*, 18–33. [[CrossRef](#)]
34. Kozák, V.; Sushkov, R.; Kulich, M.; Přeučil, L. Data-driven object pose estimation in a practical bin-picking application. *Sensors* **2021**, *21*, 6093. [[CrossRef](#)]
35. Tanaka, J.; Ogawa, A. Cardboard box depalletizing robot using two-surface suction and elastic joint mechanisms: Mechanism proposal and verification. *J. Robot. Mechatronics* **2019**, *31*, 474–492. [[CrossRef](#)]
36. Leitner, J.; Tow, A.W.; Sünderhauf, N.; Dean, J.E.; Durham, J.W.; Cooper, M.; Eich, M.; Lehnert, C.; Mangels, R.; McCool, C.; et al. The ACRV picking benchmark: A robotic shelf picking benchmark to foster reproducible research. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 4705–4712.
37. Kessens, C.C.; Thomas, J.; Desai, J.P.; Kumar, V. Versatile aerial grasping using self-sealing suction. In Proceedings of the 2016 IEEE international conference on robotics and automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 3249–3254.
38. Islam, F.; Vemula, A.; Kim, S.K.; Dornbush, A.; Salzman, O.; Likhachev, M. Planning, learning and reasoning framework for robot truck unloading. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 5011–5017.
39. Domae, Y.; Okuda, H.; Taguchi, Y.; Sumi, K.; Hirai, T. Fast graspability evaluation on single depth maps for bin picking with general grippers. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 1997–2004.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.