frontiers in Neurorobotics

ORIGINAL RESEARCH
published: 24 March 2022
doi: 10.3389/fnbot.2022.806898

Learning Suction Graspability Considering Grasp Quality and Robot Reachability for Bin-Picking

Ping Jiang, Junji Oaki, Yoshiyuki Ishihara, Junichiro Ooga, Haifeng Han, Atsushi Sugahara, Seiji Tokura, Haruna Eto, Kazuma Komoda and Akihito Ogawa*

Corporate Research & Development Center, Toshiba Corporation, Kawasaki, Japan

Deep learning has been widely used for inferring robust grasps. Although human-labeled RGB-D datasets were initially used to learn grasp configurations, preparation of this kind of large dataset is expensive. To address this problem, images were generated by a physical simulator, and a physically inspired model (e.g., a contact model between a suction vacuum cup and object) was used as a grasp quality evaluation metric to annotate the synthesized images. However, this kind of contact model is complicated and requires parameter identification by experiments to ensure real world performance. In addition, previous studies have not considered manipulator reachability such as when a grasp configuration with high grasp quality is unable to reach the target due to collisions or the physical limitations of the robot. In this study, we propose an intuitive geometric analytic-based grasp quality evaluation metric. We further incorporate a reachability evaluation metric. We annotate the pixel-wise grasp quality and reachability by the proposed evaluation metric on synthesized images in a simulator to train an auto-encoder–decoder called suction graspability U-Net++ (SG-U-Net++). Experiment results show that our intuitive grasp quality evaluation metric is competitive with a physically-inspired metric. Learning the reachability helps to reduce motion planning computation time by removing obviously unreachable candidates. The system achieves an overall picking speed of 560 PPH (pieces per hour).

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

OPEN ACCESS

Edited by:
Yimin Zhou,
Shenzhen Institutes of Advanced Technology (CAS), China

Reviewed by:
Zhifeng Huang,
Guangdong University of Technology, China
Miao Li,
Swiss Federal Institute of Technology Lausanne, Switzerland

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

Received: *01 November 2021*
Accepted: *23 February 2022*
Published: *24 March 2022*

Citation:
Jiang P, Oaki J, Ishihara Y, Ooga J, Han H, Sugahara A, Tokura S, Eto H, Komoda K and Ogawa A (2022) Learning Suction Graspability Considering Grasp Quality and Robot Reachability for Bin-Picking. Front. Neurorobot. 16:806898. doi: 10.3389/fnbot.2022.806898

1. INTRODUCTION

In recent years, growth in retail e-commerce (electronic-commerce) business has led to high demand for warehouse automation by robots (Bogue, 2016). Although the Amazon picking challenge (Fujita et al., 2020) has advanced the automation of the pick-and-place task, which is a common task in warehouses, picking objects from a cluttered scene remains a challenge.

The key to the automation of pick-and-place is to find the grasp point where the robot can approach *via* a collision free path and then stably grasp the target object. Grasp point detection methods can be broadly divided into analytical and data-driven methods. Analytical methods (Miller and Allen, 2004; Pharswan et al., 2019) require modeling the interaction between the object and the hand and have a high computation cost (Roa and Suárez, 2015). For those reasons, data-driven methods are preferred for bin picking.

Many previous studies have used supervised deep learning, which is one of the most widely used data-driven methods, to predict only grasp point configuration (e.g., location, orientation, and open width) without considering the grasp quality. Given an RGB-D image, the grasp configuration for a jaw gripper (Kumra and Kanan, 2017; Chu et al., 2018; Zhang et al., 2019) or a vacuum gripper (Araki et al., 2020; Jiang et al., 2020) can be directly predicted using a deep convolutional neural network (DCNN). Learning was extended from points to regions by Domae et al. (2014) and Mano et al. (2019), who proposed a convolution-based method in which the hand shape mask is convolved with the depth mask to obtain the region of the grasp points. Matsumura et al. (2019) later learned the peak among all regions for different hand orientations to detect a grasp point capable of avoiding multiple objects.

However, in addition to the grasp configuration, the grasp quality is also important for a robot to select the optimal grasp point for bin picking. The grasp quality indicates the graspable probability by considering factors such as surface properties. For example, for suction grasping, although an object with a complicated shape may have multiple grasp points, the grasp points located on flat surfaces need to be given a higher selection priority because they have higher grasp quality (easier for suction by vacuum cup) than do curved surfaces. Zeng et al. (2018b) empirically labeled the grasp quality in the RGB-D images of the Amazon picking challenge object set. They proposed a multi-modal DCNN for learning grasp quality maps (pixel-wise grasp quality corresponding to an RGB-D image) for jaw and vacuum grippers. However, preparing a dataset by manual labeling is time consuming and so the dataset was synthesized in a simulator to reduce the time cost. Dex-Net (Mahler et al., 2018, 2019) evaluated the grasp quality by a physical model and generated a large dataset by simulation. They used the synthesized dataset to train a grasp quality conventional neural network (GQ-CNN) to estimate the success probability of the grasp point. However, defining a precise physical model for the contact between gripper and object is difficult. Furthermore, the parameters of the model needed to be identified experimentally to reproduce the salient kinematics and dynamics features of a real robot hand (e.g., the deformation and suction force of a vacuum cup).

Unlike Dex-Net, this study proposes an intuitive suction grasp quality analytic metric based on point clouds without the need for modeling complicated contact dynamics. Furthermore, we incorporate a robot reachability metric to evaluate the suction graspability from the viewpoint of the manipulator. Previous studies have evaluated grasp quality only in terms of grasp quality for the hand. However, it is possible that although a grasp point has high grasp quality, the manipulator is not able to move to that point. It is also possible for an object to have multiple grasp points with same the level of graspability but varying amounts of time needed for the manipulator to approach due to differences in the goal pose and surrounding collision objects. Bin picking efficiency can therefore be improved by incorporating a reachability evaluation metric. We label suction graspability by the proposed grasp quality and reachability metric and generate a dataset by the physical simulator. An auto-encoder is trained to predict the suction graspability given the depth image input and a graspability clustering and the ranking algorithm is designed to propose the optimal grasp point.

Our primary contributions include (1) Proposal of an intuitive grasp quality evaluation metric without complicated physical modeling. (2) Proposal of a reachability evaluation metric for labeling suction grapability in addition to grasp quality. (3) Performance of a comparison experiment between the proposed intuitive grasp quality evaluation metric and a physically-inspired one (Dex-Net). (4) Performance of an experiment to investigate the effect of learning reachability.

2. RELATED WORKS

2.1. Pixel-Wise Graspability Learning

In early studies, deep neural networks were used to directly predict the candidate grasp configurations without considering the grasp quality (Asif et al., 2018; Zhou X. et al., 2018; Xu et al., 2021). However, since there can be multiple grasp candidates for an object that has a complicated shape or multiple objects in a cluttered scene, learning graspablity is required for the planner to find the optimal grasp among the candidates.

Pixel-wise graspablity learning uses RGB-D or depth-only images to infer the grasp success probability at each pixel. Zeng et al. (2018b) used a manually labeled dataset to train fully convolutional networks (FCNs) for predicting pixel-wise grasp quality (affordance) maps of four pre-defined grasping primitives. Liu et al. (2020) performed active exploration by pushing objects to find good grasp affordable maps predicted by Zeng's FCNs. Recently, Utomo et al. (2021) modified the architecture of Zeng's FCNs to improve the inference precision and speed. Based on Zeng's concept, Hasegawa et al. (2019) incorporated a primitive template matching module, making the system adaptive to changes in grasping primitives. Zeng et al. also applied the concept of pixel-wise affordance learning to other manipulation tasks such as picking by synergistic coordination of push and grasp motions (Zeng et al., 2018a), and picking and throwing (Zeng et al., 2020). However, preparing huge amounts of RGB-D images and manually labeling the grasp quality requires a large amount of effort.

Faced with the dataset generation cost of RGB-D based graspability learning, researchers started to use depth-image-only based learning. The merits of using depth images are that they are easier to synthesize and annotate in a physical simulator compared with RGB images. Morrison et al. (2020) proposed a generative grasping convolutional neural network (GG-CNN) to rapidly predict pixel-wise grasp quality. Based on a similar concept of grasp quality learning, the U-Grasping fully convolutional neural network (UGNet) (Song et al., 2019), Generative Residual Convolutional Neural Network (GRConvNet) (Kumra et al., 2020), and Generative Inception Neural Network (GI-NNet) (Shukla et al., 2021) were later proposed and were reported to achieve higher accuracy than GG-CNN. Le et al. (2021) extended GG-CNN to be capable of predicting the grasp quality of deformable objects by incorporating stiffness information. Morrison et al. (2019) also applied GG-CNN to a multi-view picking controller to avoid bad grasp poses caused by occlusion and collision. However,

the grasp quality dataset of GG-CNN was generated by creating masks of the center third of each grasping rectangle of the Cornell Grasping dataset (Lenz et al., 2015) and Jacquard dataset (Depierre et al., 2018). This annotation method did not deeply analyze the interaction between hand and object, which is expected to lead to insufficient representation of grasp robustness.

To improve the robustness of grasp quality annotation, a physically-inspired contact force model was designed to label pixel-wise grasp quality. Mahler et al. (2018, 2019) designed a quasi-static spring model for the contact force between the vacuum cup and the object. Based on the designed compliant contact model, they assessed the grasp quality in terms of grasp robustness in a physical simulator. They further proposed GQ-CNN to learn the grasp quality and used a sampling-based method to propose an optimal grasp in the inference phase, and also extended their study by proposing a fully convolutional GQ-CNN (Satish et al., 2019) to infer pixel-wise grasp quality, which achieved faster grasping. Recently, (Cao et al., 2021) used an auto-encoder–decoder to infer the grasp quality, which was labeled by a similar contact model to that used in GQ-CNN, to generate the suction pose. However, the accuracy of the contact model depends on the model complexity and parameter tuning. High complexity may lead to a long computation cost of annotation. Parameter identification by real world experiment (Bernardin et al., 2019) might also be necessary to ensure the validity of the contact model.

Our approach also labeled the grasp quality in synthesized depth images. Unlike GQ-CNN, we proposed a more intuitive evaluation metric based on a geometrical analytic method rather than a complicated contact analytic model. Our results showed that the intuitive evaluation metric was competitive with GQ-CNN. A reachability heatmap was further incorporated to help filter pixels that had high grasp quality value but were unreachable.

2.2. Reachability Assessment

Reachability was previously assessed by sampling a large number of grasp poses and then using forward kinematics calculation, inverse kinematics calculation, or manipulability ellipsoid evaluation to investigate whether the sampled poses were reachable (Zacharias et al., 2007; Porges et al., 2014, 2015; Vahrenkamp and Asfour, 2015; Makhal and Goins, 2018). The reachability map was generated off-line, and the feasibility of candidate grasp poses was queried during grasp planning for picking static (Akinola et al., 2018; Sundaram et al., 2020) or moving (Akinola et al., 2021) objects. However, creating an off-line map with high accuracy for a large space is computationally expensive. In addition, although the off-line map considered only collisions between the manipulator and a constrained environment (e.g., fixed bin or wall) since the environment for picking in a cluttered scene is dynamic, collision checking between the manipulator and surrounding objects is still needed and this can be time consuming. Hence, recent studies have started to learn reachability with collision awareness of grasp poses. Kim and Perez (2021) designed a density net to learn the reachability density of a given pose but considered only self-collision. Murali et al. (2020) used a learned grasp sampler to sample 6D grasp poses and proposed a CollisionNet to assess the collision score of sampled poses. Lou et al. (2020) proposed a 3D CNN and reachability predictor to predict the pose stability and reachability of sampled grasp poses. They later extended the work by incorporating collision awareness for learning approachable grasp poses (Lou et al., 2021). These sampling-based methods have required designing or training a good grasp sampler for inferring the reachability. Our approach is one-shot, which directly infers the pixel-wise reachability from the depth image without sampling.

3. PROBLEM STATEMENT

3.1. Objective

Based on depth image and point cloud input, the goal is to find a grasp pose with high graspability for a suction robotic hand to pick items in a cluttered scene and then place them on a conveyor. The depth image and point cloud point are directly obtained from an Intel RealSense SR300 camera.

3.2. Picking Robot

As shown in **Figure 1A**, the picking robot is composed of a 6 degree-of-freedom (DoF) manipulator (TVL500, Shibaura Machine Co., Ltd.) and a 1 DoF robotic hand with two vacuum suction cups (**Figure 1B**). The camera is mounted in the center of the hand and is activated only when the robot is at its home position (initial pose) and, hence, can be regarded as a fixed camera installed above the bin. This setup has the merit that the camera can capture the scene of the entire bin from the view above the bin center without occlusion by the manipulator.

3.3. Grasp Pose

As shown in **Figure 1C**, the 6D grasp pose $\boldsymbol{G}$ is defined as $(\boldsymbol{p}, \boldsymbol{n}, \theta)$, where $\boldsymbol{p}$ is the target point position of the vacuum suction cup center, $\boldsymbol{n}$ is the suction direction, and θ is the rotation angle around $\boldsymbol{n}$. Given the point cloud of the target item and $\boldsymbol{p}$ position, the normal of $\boldsymbol{p}$ can be calculated simply by principal component analysis of a covariance matrix generated from neighbors of $\boldsymbol{p}$ using a point cloud library. $\boldsymbol{n}$ is the direction of the calculated normal of $\boldsymbol{p}$. As $\boldsymbol{n}$ determines only the direction of the center axis of the vacuum suction cup, a further rotation degree of freedom (θ) is required to determine the 6D pose of the hand. Note that the two vacuum suction cups are symmetric with respect to the hand center.

4. METHODS

The overall picking system diagram is shown in **Figure 2**. Given a depth image captured at the robot home position, the auto-encoder SG-U-Net++ predicts the suction graspability maps, including a pixel-wise grasp quality map and a robot reachability map. The auto-encoder SG-U-Net++ is trained using a synthesized dataset generated by a physical simulator without any human-labeled data. Cluster analysis is performed on two maps to find areas with graspbility higher than the

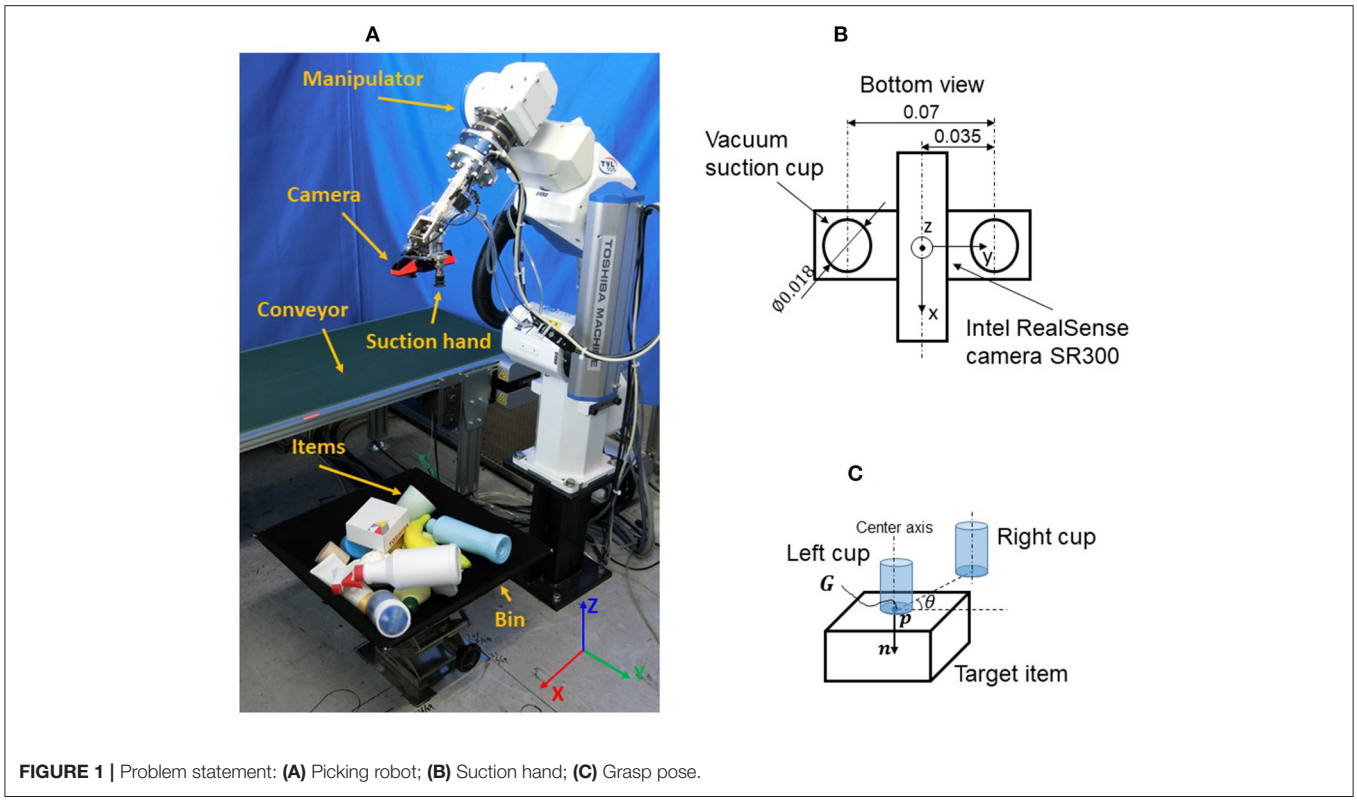

FIGURE 1 | Problem statement: **(A)** Picking robot; **(B)** Suction hand; **(C)** Grasp pose.

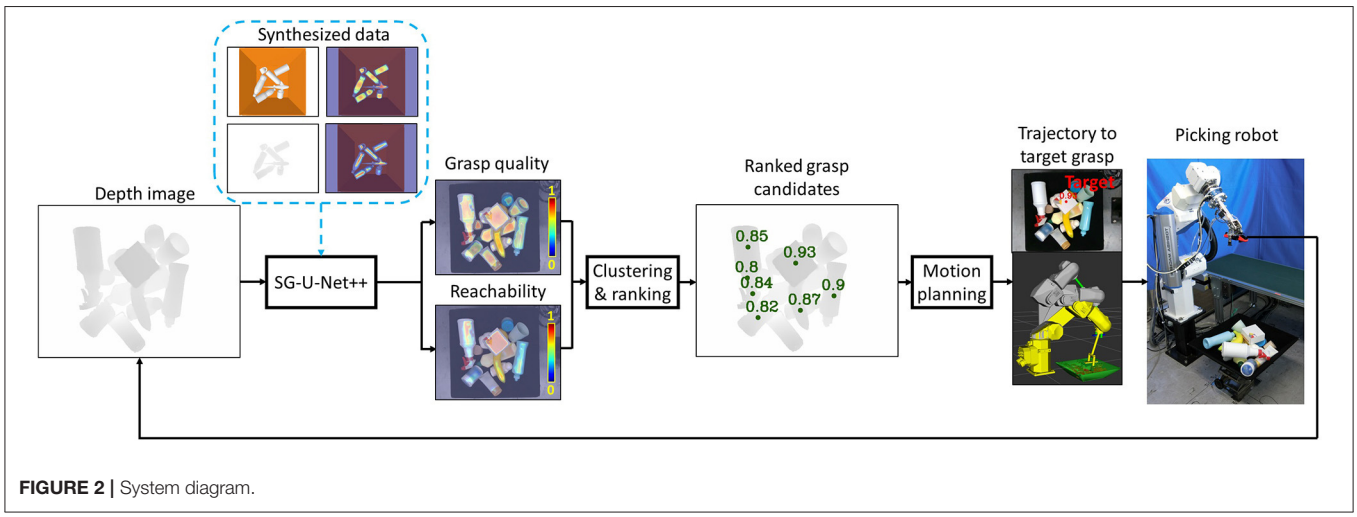

FIGURE 2 | System diagram.

thresholds. Local sorting is performed to extract the points with the highest graspbility values in each cluster as grasp candidates. Global sorting is further performed to sort the candidates of all clusters in descending order of graspbility value, and this is sent to the motion planner. The motion planner plans the trajectory for reaching the sorted grasp candidates in descending order of graspability value. The path search continues until the first successful solution of the candidate is found.

4.1. Learning the Suction Graspability

SG-U-Net++ was trained on a synthesized dataset to learn suction graspability by supervised deep learning. **Figure 3A** shows the overall dataset generation flow. A synthesized cluttered scene is first generated using pybullet to obtain a systematized depth image and object segmentation mask. Region growing is then performed on the point cloud to detect the graspable surfaces. A convolution-based method is further used to find the

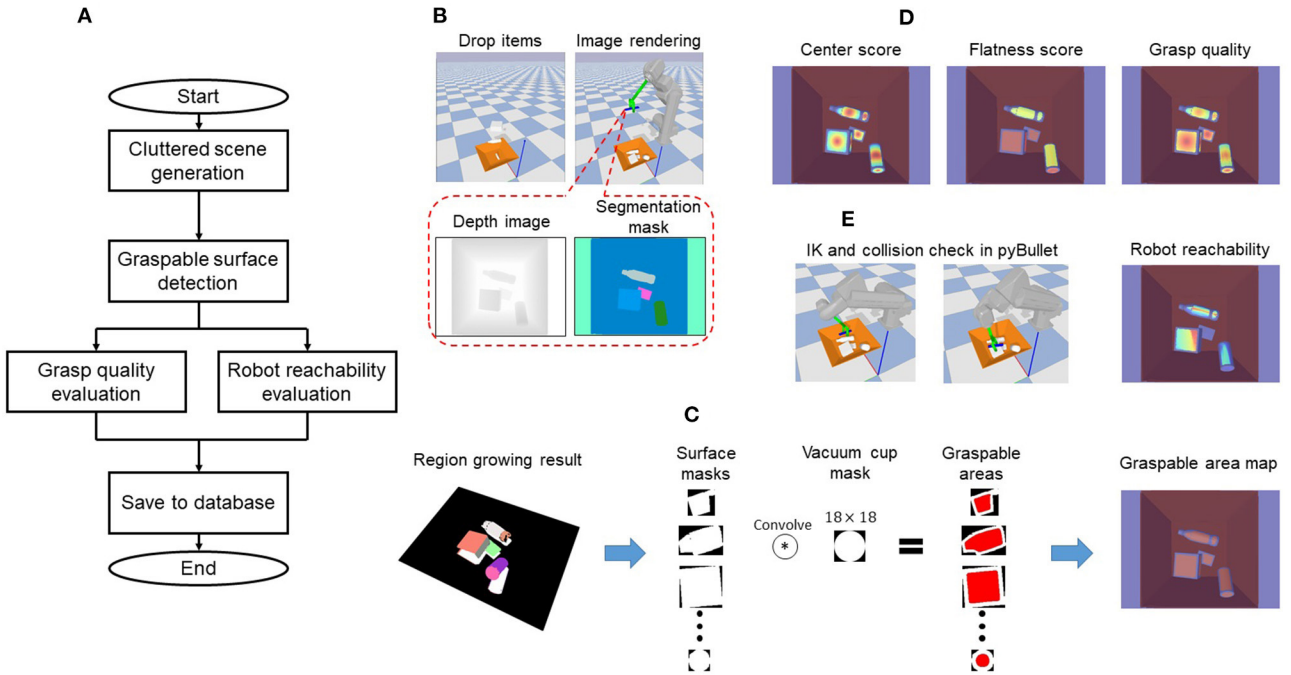

FIGURE 3 | Data generation pipeline: **(A)** Dataset generation flow; **(B)** Cluttered scene generation; **(C)** Graspable surface detection; **(D)** Grasp quality evaluation; **(E)** Robot reachability evaluation.

graspable areas of vacuum cup centers where the vacuum cup can make full contact with the surfaces. The grasp quality and robot reachability are then pixel-wise evaluated by the proposed metrics in the graspable area.

4.1.1. Cluttered Scene Generation

The object set used to synthesize the scene contains 3D CAD models from the 3DNet (Wohlkinger et al., 2012) and KIT Object database (Kasper et al., 2012). These models were used because they had previously been used to generate a dataset for which a trained CNN successfully predicted the grasp quality (Mahler et al., 2017). We empirically removed objects that are obviously difficult for suction to finally obtain 708 models. To generate cluttered scenes, a random number of objects were selected from the object set randomly and were dropped from above the bin in random poses. Once the state of all dropped objects was stable, a depth image and segmentation mask for the cluttered scene was generated, as in **Figure 3B**.

4.1.2. Graspable Surface Detection

As shown in **Figure 3C**, in order to find the graspable area of each object, graspable surface detection was performed. Given the camera intrinsic matrix, the point cloud of each object can be easily created from the depth image and segmentation mask. To detect surfaces that are roughly flat and large enough for suction by the vacuum cup, a region growing algorithm (Rusu and Cousins, 2011) was used to segment the point cloud. To stably suck an object, the vacuum cup needs to be in full contact with the surface. Hence, inspired by Domae et al. (2014), a convolution based method was used to calculate the graspable area (set of vacuum cup center positions where the cup could make full contact with the surface). Specifically, as shown in the middle of **Figure 3C**, each segmented point cloud was projected onto its local coordinates to create a binary surface mask. Each pixel of the mask represents 1 mm. The surface mask was then convolved with a vacuum cup mask (of size 18×18, where 18 is the cup diameter) to obtain the graspable area. At a given pixel, the convolution result is the area of the cup ($\pi * 0.009^2$ for our hand configuration) if the vacuum cup can make full contact with the surface. Refer to Domae et al. (2014) for more details. The calculated areas were finally remapped to a depth image to generate a graspable area map (right side of **Figure 3C**).

4.1.3. Grasp Quality Evaluation

Although the grasp areas of the surfaces were obtained, each pixel in the area may have a different grasp probability, i.e., grasp quality, owing to surface features. Therefore, an intuitive metric J_q (Equation 1) was proposed to assess the grasp quality for each pixel in the graspable area. The metric J_q is made up of J_c which evaluates the normalized distance to the center of the graspable area and J_s which evaluates the flatness and smoothness of the contact area between the vacuum cup and surface.

$$J_q = 0.5J_c + 0.5J_s \tag{1}$$

J_c (Equations 2, 3) was derived based on the assumption that the closer the grasp points are to the center of the graspable area, the closer they are to the center of mass of the object. Hence, grasp

points close to the area center (higher J_c values) are considered to be more stable for the robot to suck and hold the object.

$$J_c = 1 - \text{maxmin}(\|\boldsymbol{p} - \boldsymbol{p}_c\|_2) \tag{2}$$

$$\text{maxmin}(\boldsymbol{x}) = \frac{\boldsymbol{x} - \min(\boldsymbol{x})}{\max(\boldsymbol{x}) - \min(\boldsymbol{x})} \tag{3}$$

where $\boldsymbol{p}$ is a point in a graspable area of a surface, $\boldsymbol{p}_c$ is the center of the graspable area, and maxmin($\boldsymbol{x}$) is a max-min normalization function.

J_s (Equations 4–6) was derived based on the assumption that a vacuum cup generates a higher suction force when in contact with a flat and smooth surface than a curved one. We defined $\boldsymbol{p}_s$ as the point set of the contact area between the vacuum cup and the surface when the vacuum cup is sucked at a certain point in the graspable area. As reported in Nishina and Hasegawa (2020), the surface flatness can be evaluated by the variance of the normals, the first term of J_s assesses the surface flatness by evaluating the variance of the normals of $\boldsymbol{p}_s$ as in Equation (5). However, it is not sufficient to consider only the flatness. For example, although a vicinal surface has a small normal variance, the vacuum cup cannot achieve suction to this kind of step-like surface. Hence, the second term (Equation 6) was incorporated to assess the surface smoothness by evaluating the residual error to fit $\boldsymbol{p}_s$ to a plane Plane($\boldsymbol{p}_s$) where the sum of the distance of each point in $\boldsymbol{p}_s$ to the fitted plane is calculated. Note that the weights in the equations were tuned manually by human observations. We adjusted the weights and parameters until we observed that the Jq map was physically plausible for grasping. We finally empirically set weights of J_c and J_s to 0.5, scaled res($\boldsymbol{p}_s$) by 5.0, and added weights 0.9 and 0.1 to two terms in Equation 4 to obtain plausible grasp quality values.

$$J_s = 0.9\text{var}(\boldsymbol{n}_s) + 0.1e^{-5\text{res}(\boldsymbol{p}_s)} \tag{4}$$

$$\text{var}(\boldsymbol{n}_s) = \frac{\sum_{i=1}^{N} \boldsymbol{n}_{s,i} - \bar{\boldsymbol{n}}_s}{N-1} \tag{5}$$

$$\text{res}(\boldsymbol{p}_s) = \sum_{i=1}^{N} \|\boldsymbol{p}_{s,i} - \text{Plane}(\boldsymbol{p}_s)\|_2 \tag{6}$$

where $\boldsymbol{p}_s$ are the points in the contact surface when the vacuum cup sucks at a point in the graspable area, N is the number of points in $\boldsymbol{p}_s$, $\boldsymbol{n}_s$ are the point normals of $\boldsymbol{p}_s$, var($\boldsymbol{n}_s$) is the function to calculate the variance of $\boldsymbol{n}_s$, Plane($\boldsymbol{p}_s$) is a plane equation fitted by $\boldsymbol{p}_s$ using the least squares method, and res($\boldsymbol{p}_s$) is the function to calculate the residual error of the plane fitting by calculating the sum of the distance from each point in $\boldsymbol{p}_s$ to the fitted plane.

Figure 3D shows an example of the annotated grasp quality. Points closer to the surface center had higher grasp quality values, and points located on flat surfaces had higher grasp quality (e.g., surfaces of boxes had higher grasp quality values than cylinder lateral surfaces).

4.1.4. Robot Reachability Evaluation

The grasp quality considers only the interaction between the object and the vacuum cup without considering the manipulator. As a collision check and inverse kinematics (IK) solution search for the manipulator are needed, online checking and searching for all grasp candidates is costly. Learning robot reachability helped to rapidly avoid the grasp points where the hand and manipulator may collide with the surroundings. It also assessed the ease of finding IK solutions for the manipulator.

As described in Section 3.3, $\boldsymbol{p}$ and $\boldsymbol{n}$ of a grasp pose $\boldsymbol{G}$ can be calculated from the point cloud. θ is the only undetermined variable for defining a $\boldsymbol{G}$. We sampled the θ from 0° to 355° in step intervals of 5°. IKfast (Diankov, 2010) and Flexible Collision Library (FCL) (Pan et al., 2012) were used to calculate the inverse kinematics solution and detect the collision check for each sampled θ. The reachability evaluation metric (Equations 7–8) assessed the ratio of the number of IK valid θ (had collision free IK solution) to the sampled size N_θ.

$$J_a = \frac{\sum_{i=1}^{N_\theta} \text{Solver}(\boldsymbol{p}, \boldsymbol{n}, \theta_i)}{N_\theta} \tag{7}$$

$$\text{Solver}(\boldsymbol{p}, \boldsymbol{n}, \theta_i) = \begin{cases} 1 & \text{if collision free and IK solution exists} \\ 0 & \text{else} \end{cases} \tag{8}$$

where N_θ is the size of sampled θ and Solver is the IK solver and collision checker for the robot.

Note that because the two vacuum cups are symmetric with respect to the hand center, we evaluated the reachability score of only one cup. **Figure 3E** shows an example of the robot reachability evaluation.

4.1.5. SG-U-Net++

As shown in **Figure 4**, a nest structured auto-encoder–decoder called suction graspability U-Net++ (SG-U-Net++) was used to learn the suction graspability. We used the nested architecture because it was previously reported to have high performances for semantic segmentation. Given a 256 × 256 depth image, SG-U-Net++ outputs 256 × 256 shape grasp quality and robot reachability maps. SG-U-Net++ resembles the structure of U-Net++ proposed by Zhou Z. et al. (2018). SG-U-Net++ consists of several sub encoder–decoders connected by skip connections. For example, $X^{0,0} \rightarrow X^{1,0} \rightarrow X^{0,1}$ is one of the smallest sub encoder–decoders, and $X^{0,0} \rightarrow X^{1,0} \rightarrow X^{2,3} \rightarrow X^{3,0} \rightarrow X^{4,0} \rightarrow X^{3,1} \rightarrow X^{2,2} \rightarrow X^{1,3} \rightarrow X^{0,4}$ is the largest encoder–decoder. The dense block for $X^{i,j}$ consists of two $3 \times 3 \times 32 * 2^i$ convolution (conv) layers, each of which is followed by batch normalization and rectified linear unit (ReLU) activation. The output layer connected to $X^{0,4}$ is a $1 \times 1 \times 2$ conv layer. MSELoss (Equation 9) was used for supervised pixel-wise heatmap learning.

$$Loss = \frac{1}{H}\frac{1}{W}\sum_{i=0}^{H}\sum_{j=0}^{W} 0.5*(J_q(i,j) - \widehat{J}_q(i,j))^2 + 0.5*(J_a(i,j) - \widehat{J}_a(i,j))^2 \tag{9}$$

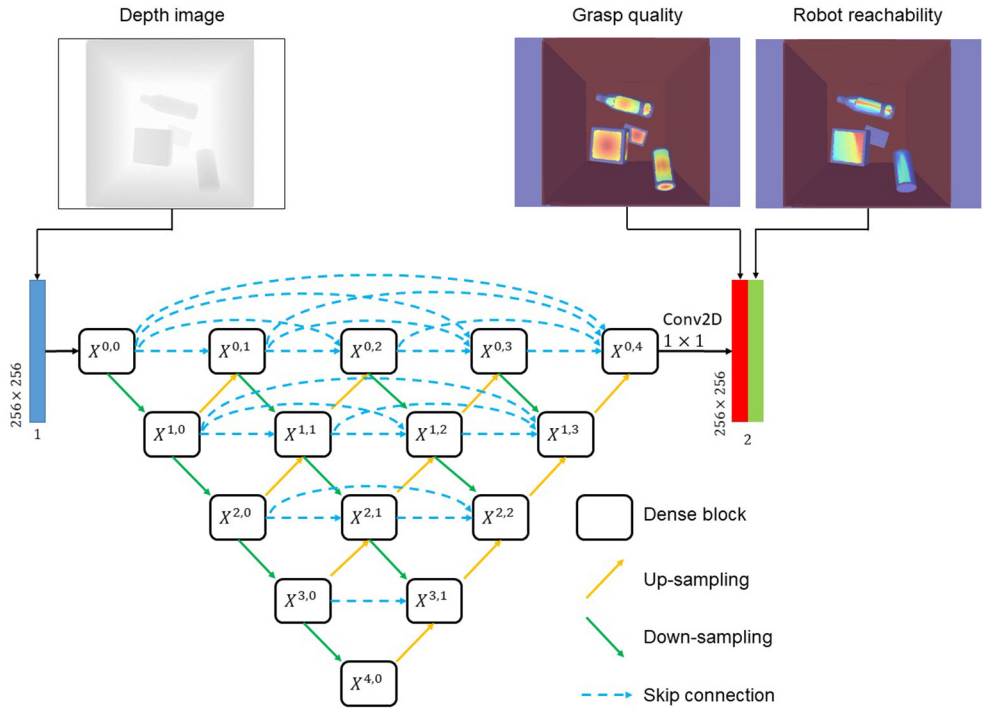

FIGURE 4 | The architecture of SG-U-Net++.

where H and W are the image height and width. $\widehat{J}_q$ and $\widehat{J}_a$ indicate the ground truth.

4.2. Clustering and Ranking

The clustering and ranking block in **Figure 2** outputs the ranked grasp proposals. To validate the role of learning reachability, we proposed two policies (Policy 1: use only grasp quality; Policy 2: use both grasp quality and reachability) to propose the grasp candidates. Policy 1 extracted the area of grasp quality values larger than threshold th_g. Policy 2 extracted the area of grasp quality score values larger than threshold th_g and the corresponding reachability score values larger than th_r. Filtering by reachability score value was assumed to help to remove pixels with high grasp quality values that are not reachable by the robot due to collision or IK error. The values of th_g and th_r were empirically set to 0.5 and 0.3, respectively. The extracted areas were clustered by scipy.ndimage.label (Virtanen et al., 2020). Points in each cluster were ranked (local cluster level) by the grasp quality values, and the point with the highest grasp quality was used as the grasp candidate for its owner clusters (refer to Ranked grasp candidates in **Figure 2**). Finally, the grasp candidates were further ranked (global level) and sent to the motion planner.

4.3. Motion Planning

Given the grasp candidates, goal poses were created for move. It (Chitta et al., 2012) to plan a trajectory. As described in 3.3, the values of $\boldsymbol{p}$ and $\boldsymbol{n}$ of a goal pose could be obtained from the corresponding point cloud information of the grasp candidates so that only θ was undetermined. As a cartesian movement path is required for the hand to suck the object, $\boldsymbol{p}$ was set to a 1 cm offset away from the object along the $\boldsymbol{n}$ direction. θ was sampled from 0° to 180° at step intervals of 5°. For each sampled goal pose, the trajectory was planned for left and right vacuum cups, respectively, and the shorter trajectory was selected as the final solution. The planned trajectory was further time parametrized by Time-Optimal Path Parameterization (toppra) (Pham and Pham, 2018) to realize position control for the robot to approach the goal pose. After reaching the goal pose, the robot hand moved down along $\boldsymbol{n}$ to suck the object. Once the contact force between the vacuum cup and object, which was measured by a force sensor, exceeded the threshold, the object was assumed to be sucked by the vacuum cup and was then lifted and placed on the conveyor.

5. EXPERIMENTS

5.1. Data Collection, Training, and Precision Evaluation

We used the proposed suction graspability annotation method in pyBullet to generate 15,000 data items, which were split into 10,000 for training and 5,000 for testing. The synthesized data was then used to train SG-U-Net++, which was implemented by pyTorch. The adam optimizer (learning rate = 1.0e−4) was used to update the parameters of the neural network during the training. The batch size was set to 16. Both data collection and training were conducted on an Intel Core i7-8700K 3.70 GHz PC with 64G RAM and 4 Nvidia Geforce GTX 1080 GPUs.

To evaluate the learning results, we used a similar evaluation method to that reported in Zeng et al. (2018b) on the testing set. For practical utilization, it is important for SG-U-Net++ to find at least one point in ground truth suction graspable area or manipulator reachable area. We defined suction graspable area as the pixels whose ground truth grasp quality scores are larger than 0.5 and approachable area as the pixels whose ground truth reachability scores are larger than 0.5. The inferred grasp quality and reachability scores were divided by thresholds into Top 1%, Top 10%, Top 25%, and Top 50%. If pixels larger than the threshold were located in the ground truth area, the pixels were considered true positive, otherwise, the pixels were considered false positive. We report the inference precision for the four thresholds above for SG-U-Net++ and compare them with Dex-Net.

5.2. Real World Picking Experiments

To evaluate and compare the performance of different policies for the picking system, a pick-and-placement task experiment was conducted. In order to investigate whether SG-U-Net++ could predict the graspability of objects with different shape complexities, we used primitive solids (a simple shape with large surfaces), commodities (general shape), and 3D-printed objects (a complex shape with small surfaces) as experimental object set (refer to **Figure 5**). All of the objects are novel objects that were not used during training. During each trial, the robot was required to pick 13 randomly posed objects (except for the cup) from a bin and then place them on the conveyor. Note that the cup was placed in the lying pose because it could not be grasped if it was in a standing pose. A grasp attempt was treated as a failure if the robot could not grasp the object in three attempts.

We conducted 10 trials for Policy 1, Policy 2, and Dex-Net 4.0 (suction grasp proposal by fully convolutional grasping policy), respectively. Note that because Dex-Net had its own grasp planning method, we directly sorted the inferred grasp quality values without clustering. To compare our proposed intuitive grasp quality evaluation metric (Equation 1) with the one used in Dex-Net, we evaluated and compared the grasp planning computation time cost and success rate of Policy 1 and Dex-Net. To evaluate the effect of incorporating the reachability score, we evaluated and compared the grasp planning computation time cost, motion planning computation time cost, and success rate of Policy 1 and Policy 2.

FIGURE 5 | Experiment object set.

6. RESULTS AND DISCUSSION

6.1. Inference Precision Evaluation

Table 1 shows the inference precision of grasp quality and reachability. Both SQ-U-Net++ and Dex-Net achieved high precisions for Top 1% and Top 10% but the precision of Dex-Net decreased to lower than 0.9 for Top 25% and Top 50%. This result indicates that the performance of our proposed intuitive grasp quality evaluation metric (Equation 1) was as good as a physically inspired evaluation metric. Learning the suction graspability annotation by point cloud analytic methods might not be so bad compared to dynamics analytic methods for the suction grasp task. However, the inference precision of the reachability for SQ-U-Net++ also achieved larger than 0.9 for Top 1% and Top 10%, but decreased sharply for Top 25% and Top 50%. The overall performance of reachability inference was poorer than grasp quality, indicating that reachability is more difficult to learn than grasp quality. This is probably because grasp quality can be learned from the surface features, but reachability learning requires more features such as the features of surrounding objects in addition to the surface features, leading to more difficult learning.

6.2. Picking Experiments

6.2.1. Overall Performance

Table 2 shows the experimental results of Dex-Net and our proposed method. Although all three methods achieved a high grasp success rate (>90%), our method took a shorter time for grasp planning. Moreover, the motion planning computation time was reduced by incorporating the learning of reachability. The SQ-U-Net++ Policy 2 achieved a high speed picking of approximately 560 PPH (piece per hour) (refer to **Supplementary Video**).

6.2.2. Comparison With Physically-Inspired Grasp Quality Evaluation Metric

As shown in **Table 2**, although our method was competitive with Dex-Net, it was faster for grasp planning. This result indicates that our geometric analytic based grasp quality evaluation is good enough for the picking task compared with a physically-inspired one. The evaluation of contact dynamics between a vacuum cup

TABLE 1 | Inference precision.

Score	Method	Top 1%	Top 10%	Top 25%	Top 50%
Grasp quality	Dex-Net	91.9	91.0	88.7	84.2
	SQ-U-Net++	99.8	99.6	99.2	97.5
Reachability	SQ-U-Net++	95.8	91.1	80.7	61.2

and the object surface might be simplified to just analyze the geometric features of the vacuum cup (e.g., the shape of the cup) and surfaces (e.g., surface curvature, surface smoothness, and distance from the cup center to the surface center). In addition, similar to the report in Zeng et al. (2018b), the grasp proposal of Dex-Net was farther from the center of mass. **Figure 6** shows an example of our method and Dex-Net. Our predicted grasps were closer to the center of mass of the object than the ones inferred by Dex-Net. This is because we incorporated J_c (Equation 2) to evaluate the distance from the vacuum cup center to the surface center, helping the SQ-U-Net++ to predict grasp positions much closer to the center of mass.

TABLE 2 | Experiment results.

Method	Success rate (%)	Grasp planning cost (s)	Motion planning cost (s)
Dex-Net 4.0 Suction (FC-GQCNN-4.0-SUCTION)	91.5	0.60	2.91
SQ-U-Net++ Policy1 (grasp quality only)	94.6	**0.15**	1.71
SQ-U-Net++ Policy2 (grasp quality + reachability)	**95.4**	0.17	**0.90**

Bold values indicates the best performance among three methods in the Table. For the success rate, the higher the better. For the cost (computation time) of grasp planning and motion planning, the shorter the better.

6.2.3. Role of Learning Reachability

Despite that the grasp success rate might be dominant by the grasp quality score, it is possible that although a grasp point has high grasp quality, the manipulator is not able to move to that point, leading to a longer time for motion planning. The success rate and overall system efficiency are both important for the task of bin picking. Hence, reachability learning was incorporated to assess the grasp success probability from the view point of the manipulator. The reachability heatmap helped to filter out the candidates which were with high grasp quality but the manipulator could not reach to improve the efficiency. As shown in **Table 2**, although learning reachability increased the grasp planning cost a little bit by 0.02 s due to the processes such as clustering and ranking of the reachability heatmap, it helped to reduce the motion planning cost (Policy 2: 0.90 s vs. Policy 1: 1.71s) to improve the overall system efficiency, indicating that learning reachability is worthy.

Figure 7 shows an example of the role of learning reachability. Policy 2 predicted grasps with lower collision risks with

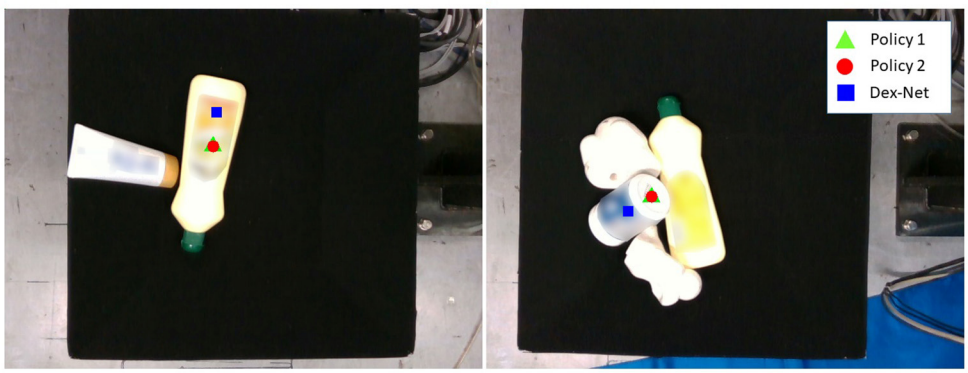

FIGURE 6 | Example of Dex-Net grasp prediction that is farther from the center of mass of the object.

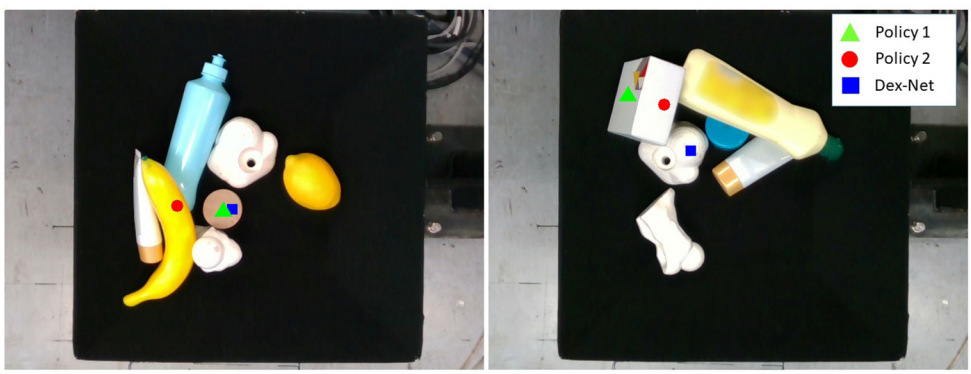

FIGURE 7 | Example of grasps predicted by Dex-Net and Policy 1 that are unreachable or difficult to reach.

neighboring objects than did Policy 1 and Dex-Net (e.g., **Figure 7** Left: Policy 1 and Dex-Net predicted grasps on a wooden cylinder that had high collision risks between the hand and 3D printed objects). Furthermore, an object might have surfaces with the same grasp quality (e.g., **Figure 7** Right: box with two flat surfaces). Whereas, Policy 2 selected the surface that was easier to reach, Policy 1 might select the one that is difficult to reach (**Figure 7** Right), since it does not consider the reachability. Therefore, Policy 2 was superior to Policy 1 and Dex-Net because it removed the grasp candidates that were obviously unable or difficult to approach. However, for Policy 1 and Dex-Net, as they considered only the grasp quality, the motion planner might first search the solutions for the candidates with high grasp quality, but those candidates might be unreachable for the manipulator and, thus, increase the motion planning effort.

6.2.4. Limitations and Future Work

Our study was not devoid of limitations. Several grasp failures occurred when picking 3D printed objects. Since the synthesized depth images differ from real ones because real images are noisy and incomplete, the neural network prediction error increased for real input depth images. This error was tolerable for objects with larger surfaces like cylinders and boxes but intolerable for 3D printed objects that have complicated shapes where the graspable areas are quite small. In the future, we intend to conduct sim-to-real (Peng et al., 2018) or depth missing value prediction (Sajjan et al., 2020) to improve the performance of our neural network. Another failure was that although not very often, the objects fell down during holding and placement because the speed of the manipulator was too high to hold the object stably. We addressed this problem by slowing down the manipulator movement during the placement action but this sacrificed the overall system picking efficiency. In the future, we want to consider a more suitable method for object holding and placement trajectory such as model based control.

Our study determined the grasping sequence by finding the grasp pose with the highest predicted grasp quality score among the filtered grasp pose candidates. The effect of other strategies such as the one that selects the target object which will not contact with the adjacent objects during the whole pick-and-place actions, or the reinforcement learning based policy (Mahler and Goldberg, 2017) will be investigated in the future.

Experiment results showed that our intuitive grasp quality evaluation metric was competitive with a physically-inspired metric, indicating that our method was plausible for bin picking of common rigid objects (e.g., primitive solids and commodities) in an electronic commerce warehouse. However, to apply our method to general industrial bin picking, object dynamics might need to be considered because the mass and materials of objects may vary in an industrial warehouse. We will investigate the effect of grasp quality metric incorporating object deformability (Xu et al., 2020; Huang et al., 2021), friction and mass distribution (Price et al., 2018; Zhao et al., 2018; Veres et al., 2020), and instability caused by robot acceleration (Khin et al., 2021) in the future.

Moreover, there is a trade-off between learning grasp quality and reachability. Increasing the weight of grasp quality loss in Equation (9) might improve the accuracy of grasp quality prediction and, thus, improve the success rate. However, it might also lead to an increased error of reachability, resulting in a long time for the motion planner to find the trajectory. Currently, we empirically set both weights to 0.5 in Equation 9, and the experimental result indicated that such a setup of weights was fine. In the future, we will investigate the influence of different weight values on the experimental result so as to find the optimal setup of weights to ensure both success rate and overall system efficiency.

Furthermore, the reachability heatmap considered the collision status of the hand goal pose for sucking the target object. The motion planner further checked whether the trajectory from the initial pose to the goal pose was collision free. This ensured that the robot could avoid colliding with other objects when grasping the target object. However, the grasped object might contact its neighboring objects when the robot lifted it after grasping. One way to avoid that is to learn the occlusion of the target object (Yu et al., 2020). If the target object was not occluded by any other objects, there would be a lower risk to make the movement of its neighboring objects when it was lifted. Another way is to predict the locations of objects by object segmentation (Araki et al., 2020; Hopfgarten et al., 2020) or object pose estimation (Tremblay et al., 2018) to make sure that there is a safe distance between the target object and its neighboring objects.

We will also extend the proposed framework for grasping by a gripper in the future. Previous studies reported that the grasp quality evaluation metric for a gripper could be designed based on geometric features (Domae et al., 2014), force closure (Miller and Allen, 2004; Roa and Suárez, 2015), or simulated gripper-object interaction (Eppner et al., 2019). For the reachability evaluation metric, the open width of a gripper should also be considered in addition to the grasp poses during evaluation.

7. CONCLUSION

We proposed an auto-encoder–decoder to infer the pixel-wise grasp quality and reachability. Our method is intuitive but competitive with CNN trained by data annotated using physically-inspired models. The reachability learning improved the efficiency of the picking system by reducing the motion planning effort. However, the performance of the auto-encoder–decoder deteriorated because of differences between synthesized and real data. In the future, sim-to-real technology will be adopted to improve performance under various environments.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

PJ made substantial contributions to conceiving the original ideas, designing the experiments, analyzing the results, and writing the original draft. JOa, YI, JOo, HH, AS, ST, HE, KK, and AO helped to conceptualize the final idea. PJ, YI, and JOa conducted the experiments and revised the manuscript. YI and AO supervised the project. All the authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The preprint of this manuscript is available from https://arxiv.org/abs/2111.02571.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnbot.2022.806898/full#supplementary-material

REFERENCES

Akinola, I., Varley, J., Chen, B., and Allen, P. K. (2018). "Workspace aware online grasp planning," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Madrid: IEEE), 2917–2924.

Akinola, I., Xu, J., Song, S., and Allen, P. K. (2021). Dynamic grasping with reachability and motion awareness. *arXiv preprint* arXiv:2103.10562. doi: 10.1109/IROS51168.2021.9636057

Araki, R., Onishi, T., Hirakawa, T., Yamashita, T., and Fujiyoshi, H. (2020). "Mt-dssd: deconvolutional single shot detector using multi task learning for object detection, segmentation, and grasping detection," in *2020 IEEE International Conference on Robotics and Automation (ICRA)* (Paris: IEEE), 10487–10493.

Asif, U., Tang, J., and Harrer, S. (2018). "Graspnet: an efficient convolutional neural network for real-time grasp detection for low-powered devices," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)* (Stockholm), *Vol.* 7, 4875–4882.

Bernardin, A., Duriez, C., and Marchal, M. (2019). "An interactive physically-based model for active suction phenomenon simulation," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Macau: IEEE), 1466–1471.

Bogue, R. (2016). Growth in e-commerce boosts innovation in the warehouse robot market. *Ind. Robot*. 43, 583–587. doi: 10.1108/IR-07-2016-0194

Cao, H., Fang, H. -S., Liu, W., and Lu, C. (2021). Suctionnet-1billion: A large-scale benchmark for suction grasping. *IEEE Robot Autom. Lett*. 8, 8718–8725. Available online at: https://arxiv.org/pdf/2103.12311.pdf (accessed October 29, 2021).

Chitta, S., Sucan, I., and Cousins, S. (2012). Moveit![ros topics]. *IEEE Robot. Automat. Mag*. 19, 18–19. doi: 10.1109/MRA.2011.2181749

Chu, F.-J., Xu, R., and Vela, P. A. (2018). Real-world multiobject, multigrasp detection. *IEEE Robot. Automat. Lett*. 3, 3355–3362. doi: 10.1109/LRA.2018.2852777

Depierre, A., Dellandréa, E., and Chen, L. (2018). "Jacquard: a large scale dataset for robotic grasp detection," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Madrid: IEEE), 3511–3516.

Diankov, R. (2010). *Automated Construction of Robotic Manipulation Programs* (Ph.D. thesis). Carnegie Mellon University, Robotics Institute.

Domae, Y., Okuda, H., Taguchi, Y., Sumi, K., and Hirai, T. (2014). "Fast graspability evaluation on single depth maps for bin picking with general grippers," in *2014 IEEE International Conference on Robotics and Automation (ICRA)* (Hong Kong: IEEE), 1997–2004.

Eppner, C., Mousavian, A., and Fox, D. (2019). A billion ways to grasp: an evaluation of grasp sampling schemes on a dense, physics-based grasp data set. *arXiv preprint* arXiv:1912.05604.

Fujita, M., Domae, Y., Noda, A., Garcia Ricardez, G., Nagatani, T., Zeng, A., et al. (2020). What are the important technologies for bin picking? technology analysis of robots in competitions based on a set of performance metrics. *Adv. Robot*. 34, 560–574. doi: 10.1080/01691864.2019.1698463

Hasegawa, S., Wada, K., Kitagawa, S., Uchimi, Y., Okada, K., and Inaba, M. (2019). "Graspfusion: realizing complex motion by learning and fusing grasp modalities with instance segmentation," in *2019 International Conference on Robotics and Automation (ICRA)* (Montreal, QC: IEEE), 7235–7241.

Hopfgarten, P., Auberle, J., and Hein, B. (2020). "Grasp area detection of unknown objects based on deep semantic segmentation," in *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)* (Hong Kong,: IEEE), 804–809.

Huang, I., Narang, Y., Eppner, C., Sundaralingam, B., Macklin, M., Hermans, T., et al. (2021). Defgraspsim: Simulation-based grasping of 3d deformable objects. *arXiv preprint* arXiv:2107.05778.

Jiang, P., Ishihara, Y., Sugiyama, N., Oaki, J., Tokura, S., Sugahara, A., et al. (2020). Depth image-based deep learning of grasp planning for textureless planar-faced objects in vision-guided robotic bin-picking. *Sensors* 20, 706. doi: 10.3390/s20030706

Kasper, A., Xue, Z., and Dillmann, R. (2012). The kit object models database: An object model database for object recognition, localization and manipulation in service robotics. *Int. J. Rob. Res*. 31, 927–934. doi: 10.1177/0278364912445831

Khin, P. M., Low, J. H., Ang Jr, M. H., and Yeow, C. H. (2021). Development and grasp stability estimation of sensorized soft robotic hand. *Front. Rob. AI* 8, 619390. doi: 10.3389/frobt.2021.619390

Kim, S., and Perez, J. (2021). "Learning reachable manifold and inverse mapping for a redundant robot manipulator," in *2021 IEEE International Conference on Robotics and Automation (ICRA)* (Xi'an: IEEE).

Kumra, S., Joshi, S., and Sahin, F. (2020). "Antipodal robotic grasping using generative residual convolutional neural network," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Las Vegas, NV: IEEE), 9626–9633.

Kumra, S., and Kanan, C. (2017). "Robotic grasp detection using deep convolutional neural networks," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Vancouver, BC: IEEE), 769–776.

Le, T. N., Lundell, J., Abu-Dakka, F. J., and Kyrki, V. (2021). Deformation-aware data-driven grasp synthesis. *arXiv preprint* arXiv:2109.05320. doi: 10.1109/LRA.2022.3146551

Lenz, I., Lee, H., and Saxena, A. (2015). Deep learning for detecting robotic grasps. *Int. J. Rob. Res*. 34, 705–724. doi: 10.1177/0278364914549607

Liu, H., Deng, Y., Guo, D., Fang, B., Sun, F., and Yang, W. (2020). An interactive perception method for warehouse automation in smart cities. *IEEE Trans. Ind. Inf*. 17, 830–838. doi: 10.1109/TII.2020.2969680

Lou, X., Yang, Y., and Choi, C. (2020). "Learning to generate 6-dof grasp poses with reachability awareness," in *2020 IEEE International Conference on Robotics and Automation (ICRA)* (Paris: IEEE), 1532–1538.

Lou, X., Yang, Y., and Choi, C. (2021). Collision-aware target-driven object grasping in constrained environments. *arXiv preprint* arXiv:2104.00776. doi: 10.1109/ICRA48506.2021.9561473

Mahler, J., and Goldberg, K. (2017). "Learning deep policies for robot bin picking by simulating robust grasping sequences," in *Conference on Robot Learning* (Mountain View, CA: PMLR), 515–524.

Mahler, J., Matl, M., Liu, X., Li, A., Gealy, D., and Goldberg, K. (2017). Dex-net 3.0: computing robust robot vacuum suction grasp targets in point clouds using a new analytic model and deep learning. *arXiv preprint* arXiv:1709.06670. doi: 10.1109/ICRA.2018.8460887

Mahler, J., Matl, M., Liu, X., Li, A., Gealy, D., and Goldberg, K. (2018). "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)* (Brisbane, QLD: IEEE), 5620–5627.

Mahler, J., Matl, M., Satish, V., Danielczuk, M., DeRose, B., McKinley, S., et al. (2019). Learning ambidextrous robot grasping policies. *Sci. Rob*. 4, 4984. doi: 10.1126/scirobotics.aau4984

Makhal, A., and Goins, A. K. (2018). "Reuleaux: Robot base placement by reachability analysis," in *2018 Second IEEE International Conference on Robotic Computing (IRC)* (Laguna Hills, CA: IEEE), 137–142.

Mano, K., Hasegawa, T., Yamashita, T., Fujiyoshi, H., and Domae, Y. (2019). "Fast and precise detection of object grasping positions with eigenvalue templates," in *2019 International Conference on Robotics and Automation (ICRA)* (Montreal, QC: IEEE), 4403–4409.

Matsumura, R., Domae, Y., Wan, W., and Harada, K. (2019). "Learning based robotic bin-picking for potentially tangled objects," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Macau: IEEE), 7990–7997.

Miller, A. T., and Allen, P. K. (2004). Graspit! a versatile simulator for robotic grasping. *IEEE Rob. Automat. Mag.* 11, 110–122. doi: 10.1109/MRA.2004.1371616

Morrison, D., Corke, P., and Leitner, J. (2019). "Multi-view picking: Next-best-view reaching for improved grasping in clutter," in *2019 International Conference on Robotics and Automation (ICRA)* (Montreal, QC: IEEE), 8762–8768.

Morrison, D., Corke, P., and Leitner, J. (2020). Learning robust, real-time, reactive robotic grasping. *Int. J. Rob. Res.* 39, 183–201. doi: 10.1177/0278364919859066

Murali, A., Mousavian, A., Eppner, C., Paxton, C., and Fox, D. (2020). "6-dof grasping for target-driven object manipulation in clutter," in *2020 IEEE International Conference on Robotics and Automation (ICRA)* (Paris: IEEE), 6232–6238.

Nishina, Y., and Hasegawa, T. (2020). Model-less grasping points estimation for bin-picking of non-rigid objects and irregular-shaped objects. *Omron Tech.* 52, 1–8. Available online at: https://www.omron.com/global/en/technology/omrontechnics/vol52/012.html (accessed March 9, 2022).

Pan, J., Chitta, S., and Manocha, D. (2012). "Fcl: a general purpose library for collision and proximity queries," in *2012 IEEE International Conference on Robotics and Automation* (Saint Paul, MN: IEEE), 3859–3866.

Peng, X. B., Andrychowicz, M., Zaremba, W., and Abbeel, P. (2018). "Sim-to-real transfer of robotic control with dynamics randomization," in *2018 IEEE International Conference on Robotics and Automation (ICRA)* (Brisbane, QLD: IEEE), 3803–3810.

Pham, H., and Pham, Q.-C. (2018). A new approach to time-optimal path parameterization based on reachability analysis. *IEEE Trans. Rob.* 34, 645–659. doi: 10.1109/TRO.2018.2819195

Pharswan, S. V., Vohra, M., Kumar, A., and Behera, L. (2019). "Domain-independent unsupervised detection of grasp regions to grasp novel objects," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Macau: IEEE), 640–645.

Porges, O., Lampariello, R., Artigas, J., Wedler, A., Borst, C., and Roa, M. A. (2015). "Reachability and dexterity: Analysis and applications for space robotics," in *Workshop on Advanced Space Technologies for Robotics and Automation-ASTRA* (Noordwijk).

Porges, O., Stouraitis, T., Borst, C., and Roa, M. A. (2014). "Reachability and capability analysis for manipulation tasks," in *ROBOT2013: First IBERIAN robOtics* Conference (Madrid: Springer), 703–718.

Price, A., Balakirsky, S., and Christensen, H. (2018). Robust grasp preimages under unknown mass and friction distributions. *Integr. Comput. Aided Eng.* 25, 99–110. doi: 10.3233/ICA-180568

Roa, M. A., and Suárez, R. (2015). Grasp quality measures: review and performance. *Auton. Rob.* 38, 65–88. doi: 10.1007/s10514-014-9402-3

Rusu, R. B., and Cousins, S. (2011). "3d is here: Point cloud library (pcl)," in *2011 IEEE International Conference on Robotics and Automation* (Shanghai: IEEE), 1–4.

Sajjan, S., Moore, M., Pan, M., Nagaraja, G., Lee, J., Zeng, A., et al. (2020). "Clear grasp: 3d shape estimation of transparent objects for manipulation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)* (Paris: IEEE), 3634–3642.

Satish, V., Mahler, J., and Goldberg, K. (2019). On-policy dataset synthesis for learning robot grasping policies using fully convolutional deep networks. *IEEE Rob. Automat. Lett.* 4, 1357–1364. doi: 10.1109/LRA.2019.2895878

Shukla, P., Pramanik, N., Mehta, D., and Nandi, G. (2021). Gi-nnet\rgi-nnet: Development of robotic grasp pose models, trainable with large as well as limited labelled training datasets, under supervised and semi supervised paradigms. *arXiv preprint* arXiv:2107.07452.

Song, Y., Fei, Y., Cheng, C., Li, X., and Yu, C. (2019). "Ug-net for robotic grasping using only depth image," in *2019 IEEE International Conference on Real-time Computing and Robotics (RCAR)* (Irkutsk: IEEE), 913–918.

Sundaram, A. M., Friedl, W., and Roa, M. A. (2020). "Environment-aware grasp strategy planning in clutter for a variable stiffness hand," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Las Vegas, NV,: IEEE), 9377–9384.

Tremblay, J., To, T., Sundaralingam, B., Xiang, Y., Fox, D., and Birchfield, S. (2018). Deep object pose estimation for semantic robotic grasping of household objects. *arXiv preprint* arXiv:1809.10790.

Utomo, T. W., Cahyadi, A. I., and Ardiyanto, I. (2021). Suction-based grasp point estimation in cluttered environment for robotic manipulator using deep learning-based affordance map. *Int. J. Automat. Comput.* 18, 277–287. doi: 10.1007/s11633-020-1260-1

Vahrenkamp, N., and Asfour, T. (2015). Representing the robot's workspace through constrained manipulability analysis. *Auton. Rob.* 38, 17–30. doi: 10.1007/s10514-014-9394-z

Veres, M., Cabral, I., and Moussa, M. (2020). Incorporating object intrinsic features within deep grasp affordance prediction. *IEEE Rob. Automat. Lett.* 5, 6009–6016. doi: 10.1109/LRA.2020.3010444

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat. Methods* 17, 261–272. doi: 10.1038/s41592-020-0772-5

Wohlkinger, W., Aldoma, A., Rusu, R. B., and Vincze, M. (2012). "3dnet: Large-scale object class recognition from cad models," in *2012 IEEE International Conference on Robotics and Automation* (Saint Paul, MN: IEEE), 5384–5391.

Xu, J., Danielczuk, M., Ichnowski, J., Mahler, J., Steinbach, E., and Goldberg, K. (2020). "Minimal work: a grasp quality metric for deformable hollow objects," in *2020 IEEE International Conference on Robotics and Automation (ICRA)* (Paris: IEEE), 1546–1552.

Xu, R., Chu, F.-J., and Vela, P. A. (2021). Gknet: grasp keypoint network for grasp candidates detection. *arXiv preprint* arXiv:2106.08497. doi: 10.1177/02783649211069569

Yu, Y., Cao, Z., Liang, S., Geng, W., and Yu, J. (2020). A novel vision-based grasping method under occlusion for manipulating robotic system. *IEEE Sens. J.* 20, 10996–11006. doi: 10.1109/JSEN.2020.2995395

Zacharias, F., Borst, C., and Hirzinger, G. (2007). "Capturing robot workspace structure: representing robot capabilities," in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems* (San Diego, CA: IEEE), 3229–3236.

Zeng, A., Song, S., Lee, J., Rodriguez, A., and Funkhouser, T. (2020). Tossingbot: Learning to throw arbitrary objects with residual physics. *IEEE Trans. Rob.* 36, 1307–1319. doi: 10.1109/TRO.2020.2988642

Zeng, A., Song, S., Welker, S., Lee, J., Rodriguez, A., and Funkhouser, T. (2018a). "Learning synergies between pushing and grasping with self-supervised deep reinforcement learning," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Madrid: IEEE), 4238–4245.

Zeng, A., Song, S., Yu, K.-T., Donlon, E., Hogan, F. R., Bauza, M., et al. (2018b). "Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching," in *2018 IEEE International Conference on Robotics and Automation (ICRA)* (Brisbane, QLD: IEEE), 3750–3757.

Zhang, H., Lan, X., Bai, S., Zhou, X., Tian, Z., and Zheng, N. (2019). "Roi-based robotic grasp detection for object overlapping scenes," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Macau: IEEE), 4768–4775.

Zhao, Z., Li, X., Lu, C., and Wang, Y. (2018). "Center of mass and friction coefficient exploration of unknown object for a robotic grasping manipulation," in *2018 IEEE International Conference on Mechatronics and Automation (ICMA)* (Changchun: IEEE), 2352–2357.

Zhou, X., Lan, X., Zhang, H., Tian, Z., Zhang, Y., and Zheng, N. (2018). "Fully convolutional grasp detection network with oriented anchor box," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Madrid,: IEEE), 7223–7230.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. (2018). "Unet++: a nested u-net architecture for medical image segmentation," in *Deep Learning in*

Medical Image Analysis and Multimodal Learning for Clinical Decision Support (Springer), 3–11.

Conflict of Interest: PJ, JOa, YI, JOo, HH, AS, ST, HE, KK, and AO were employed by Toshiba Corporation.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.