Advances in Intelligent Systems Research, volume 146
2nd International Conference on Artificial Intelligence: Technologies and Applications (ICAITA 2018)

Research on Object Grasping Point Selection Based on Deep Learning

Bo Yuan[1], Shukai Qin[1,*], Hualiang Zhang[2], Tao Zhang[2] and Xiaolong Yu[2]
[1]Northeastern University, Shenyang, Liaoning Province, China
[2]Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, Liaoning Province, China
[*]Corresponding author

***Abstract*—The research on robot grasping involves mechanical, control, computer, artificial intelligence and so on. Robot Grasping is also a good emplementation of minimal research to support other related research. Efforts on flexibility and interactivity of robot grasping can promote many related studies. In this paper, convolutional neural network is used to study the grasp candidates selection of two-finger robot grasp. The experient explained the processes of select candidate points in detail, the results of the convolutional neural network in the grasp candidates selection is verified through experiments.**

Keywords—deep learning; depth image; grasping points

I. INTRODUCTION

Intelligent autonomous grasping is a major challenge for robotics research. At present, the level of robotic intelligence is weak, and its execution capability is insufficient. Traditional robots focus on improving their perception, calculation, and execution capabilities to deal with grasping tasks in complex environments. This often results in robots that are bulky and expensive. In view of the fact that there are many types of items and the environment is complex in the real environment, it is difficult to accurately identify the items. Deep learning is a new research direction in the field of machine learning. Its purpose is to build a multilayer neural network in order to be able to imitate the human brain mechanism to analyze and interpret data such as images, audio and text. It forms more abstract high-level features by combining shallow features to discover distributed representations of data. This paper uses deep learning to study the method of intelligent grasping point selection, We use a deep learning network to evaluate the stability of a pair of grasping points which makes the grasping not completely rely on the traditional object recognition and pose estimation methods and more easily and effectively to find out the appropriate grasp points to new objects.

II. OVERVIEWS OF THE GRASP

The basic problem of grasping is how to find an appropriate position for an object to grasp the object relatively robustness. The traditional grasping methods are mostly based on object recognition[1,2], pose estimation[3], and point cloud matching[4] to find the grasping points. This is prone to errors, may not generalize well to new objects, and can be slow to match point clouds to known models during execution. Last few years, some research in this area has largely focused on associating human labels with graspable regions in images[8,10],and part of them use the dataset to train a Convolutional Neural Network (CNN)[4] to predict human labels. Google researchers also use machine learning[6] training robots to pick up random objects and predicted the results based on camera input and motor commands. However, this required over 3,000 hours of training (800,000 scraping attempts) across 14 robots.

In this paper, we use a deep learning network -- Grasp Quality Convolutional Neural Network (GQ-CNN)[5] model to evaluate the stability of a pair of grasp points. Through experimental simulation, this method can effectively and accurately select the grasping position that meets human experience.

III. ROBOT GRASP PROBLEM DESCRIPTION

The key to solving the graping problem lies in to find a suitable location can grasp the object relatively stable ,the robot firstly perceives the object for object recognition (object segmentation, extraction the effective features, then infer the optimal pose (grasping position and direction) for the grasping. For the above research objectives, the robot grasping problem can be described as follows:

By acquiring the depth image of the object through the sensor, we extract the n-dimensional feature sequence $X(t) = (x_1(t); x_2(t); ...; x_n(t))$,supposed there are m possible grasp positions for this goal $G(t) = (g_1(t); g_2(t); ...; g_m(t))$,let the grasp point feature set $\varphi_i = (g_i(t);X(t)),g_i(t)(i = 1; 2; ...; x)$ represents the specific grasp pose, the float variable y(t) belonging to {0, 1} represents the stability of different grasping points.

Given X(t) denote object feature , $g_i(t)$ denote the optimal grasp position, the robot's optimal picking discrimination problem translates into maximizing the following probability model:

$$g^*(t) = \arg \max_{g_i(t) \in G(t)} \left\{ P(\hat{y}_i(t) | \phi(X(t)), W \right\} \quad (1)$$

where $\phi(X(t)) \in R^n$ is the abstract expression of initial feature set X(t),W is the weight vector of the model. In this paper, L-layer deep learning network is used to construct an optimal robot robot grasp model, in which the input layer output is:

$$h_j^{[1]}(t) = \delta\left(\sum_{i=1}^{n} x_i(t) W_{i,j}^{[1]}\right) \tag{2}$$

where is ReLU function as(3),The L-1 layer of the network's output is input to the next layer.Eq. (4) is the relationship of Hidden layers between input and output as.

$$\delta(\mathrm{a}) = \max\{0, \mathrm{a}\} \tag{3}$$

$$h_j^{[l]}(t) = \delta\left(\sum_{i=1}^{k^{[l-1]}} h_i^{[l-1]}(t) W_{i,j}^{[l]}\right), \quad l = 2,3\cdots, L-1 \tag{4}$$

where the superscript indicates the number of network layers and the subscript indicates the network node. The last layer (Layer L) of the network is the logical output layer as (5):

$$P\{\hat{y}(t) \mid X(t), \Theta\} = \delta\left(\sum_{i=1}^{k^{[l-1]}} h_i^{[L-1]}(t) w_i^{[L]}\right) \tag{5}$$

We can acquire weight variable W=(w1;w2;...wL) through deep learning training data .

IV. Learning A Grasp Robustness Function

The Neural Network architecture contains four convolutional layers in pairs of two separated by ReLU nonlinearities followed by 3 fully connected layers and a separate input layer for the gripper height reflects the height of the grasping object in the grasp position. The Neural Network is trained offline to predict robustness of candidate grasps. The network outputs an estimate of grasp robustness R_θ, which can be used to rank grasp candidates.

Training database is Synthetic training datapoints for the eight training objects (contains189k datapoints).Dataset include five different attributes, but in our experimental we only need three types of files:

1) depth_ims_tf_table: depth images transformed to align the grasp center with the image center and the grasp axis with the middle row of pixels

2) hand_poses: gripper center from the camera that took the corresponding depth image

3) grasp_metrics: value of the robust epsilon metric computed according to the Dex-Net 2.0 graphical model

We optimize the parameters of our model using Adam Optimizer. Meanwhile, Gaussian noise is added to the images during training. To using the network to estimate grasp robustnes. We first generate grasp candidates from a depth image. Then, each image inputed to the network is rotated, translated, cropped, and scaled to align the grasp pixel location with the image center and the grasp axis with the middle row of the image, creating a 32*32 grasp image from the original depth image. The image alignment with grasp center removes the need to learn rotational invariances that can be modeled by known, computationally-efficient image transformations and allows us to evaluate grasp robustness at any orientation in the image.

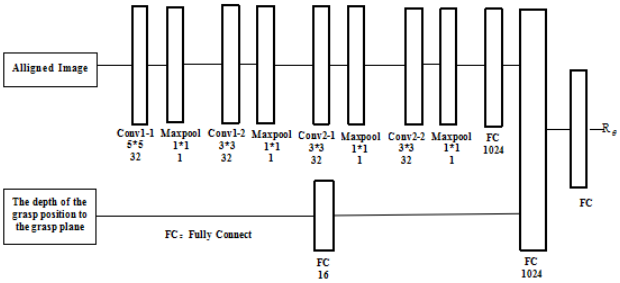

FIGURE I. THE ARCHITECTURE OF THE CNN

V. Grasp Candidates Generation and Process

Before the network evaluates the stability of the grasping point, we need to create a certain number of grasping points. The method for generating grasping points is as follows:

A. *Apply Gaussian Filtering on Depth Images*

Gaussian filtering is a linear smoothing filter, which is suitable for eliminating Gaussian noise and is widely used in image processing to reduce noise.

B. *Find Boundary Points Based on Depth Image Gradients*

Compared to using opencv to find the boundary of an object on a color image, the method of finding the boundary point according to the gradient of the depth image has greater advantages, such as being less susceptible to the influence of light and the continuity of the boundary is better. Boundary of the object can be detected clearly as Figure II, using which we are able to sample antipodal grasps.

FIGURE II. EDGE MAP

C. *Crasp Sampling*

FIGURE III. .EDGE PIXELS AND NORMALS

We firstly combine boundary points and their normal vectors to make them snappable so that they can satisfy the

force-closure as much as possible. To reduce the number of points of the boundary, we downsample original edge map and compute normal vectors to verify the force-closure[9] requirement. Then remove the similar grasp points depend on the similarity between the two groups of grasp points reference formula(6). Meanwhile, the grasping points should not be selected in the boundary area of the image for this area is easy to misjudgment. Then we get a series of grasp candidates as FIGURE IV.

$$\varsigma = \mathrm{E}\left(\varphi_{\mathrm{m}},\ \varphi_{\mathrm{n}}\right) + \partial * \theta \tag{6}$$

where ς is the similarity between the two groups of grasp points, $\mathrm{E}(\varphi_m, \varphi_n)$ is the Euclidean distance between grasps axis center, θ is the grasp angle difference between the two groups of grasp points.

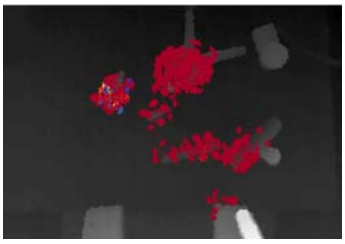

FIGURE IV. GRASP SAMPLES

D. *Align Image*

After we get We get a series of grasp candidates that satisfy basic grasp conditions, then we align the image by the grasp center and grasp axis, aligned results as shown in Figure III. Most of them have high possibility to catch up by the two graspers, which make the convolution network have a small number of input.

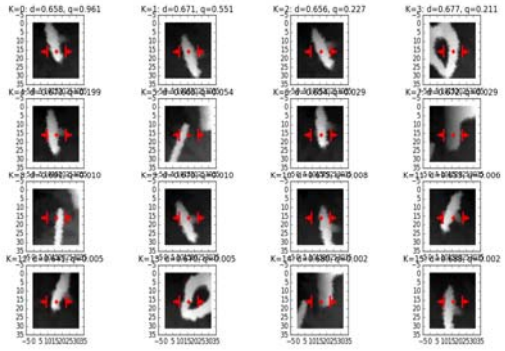

FIGURE V. ALIGNED IMAGES ON GRASP CANDIDATES

E. *GMM Clustering*

The Gaussian Mixture Model (GMM) has applications in image segmentation, object recognition, video analysis, etc. The Gaussian model uses the Gaussian probability density function (normal distribution curve) to accurately quantify things and decomposes one thing into several Gaussian probability density functions (normal distribution curves).We believe that the location of a reliable grasping point on an object should also be approximately meet Gaussian distribution, and so apply it to GMM clustering to remove some redundant grasp points. The Gaussian Mixture Model can limits the number of candidates without missing any possible grasp modes.

FIGURE VI. GMM CLUSTER ON GRASP CANDIDATES

F. *Select Best Grasps*

The grasp points obtained after GMM clustering are input into the depth network GQCNN and are sorted according to network output results and then GMM clustering of results. We iterate this process several times, we input the last clustered sampling point into GQCNN and find the maximum value from the output as the optimal grasping point. Figure VII shows the whole process of select best grasps. T is the threshold for the similarity between two grasp candidates.

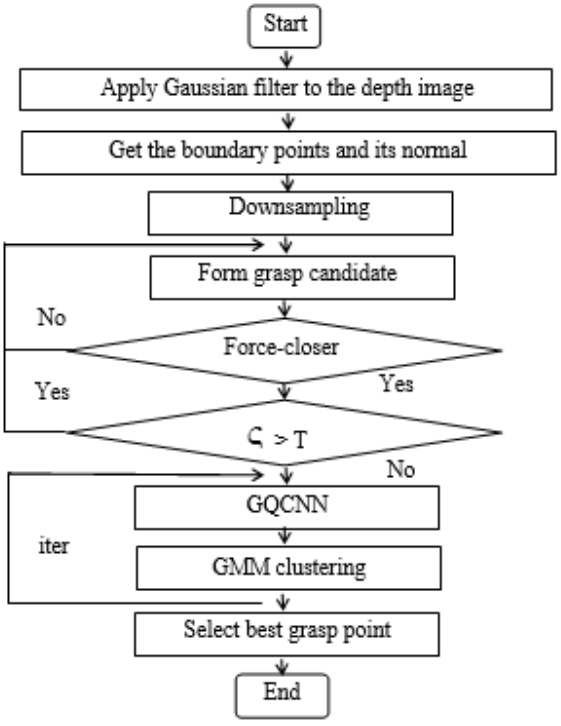

FIGURE VII. THE FLOW DIAGRAM OF SELECT BEST SAMPLE

VI. RESULTS

In order to test the validity of this paper's deep learning method in select grasp point, we apply our method to several common objects. Using Microsoft Kinect cameras to capture the object's depth image, such as cup, hammer, Clamp, pen, the final test results are shown in Figure VIII. Can be seen from the

results, the GQCNN model satisfy the basic grasp condition. Results show deep learning model GQCNN has the ability to identify the robust grasp point. Part of the test object in the experiment are different from the objects in the training data set, this illustrates the GQCNN model has strong generalization capability to a certain degree.

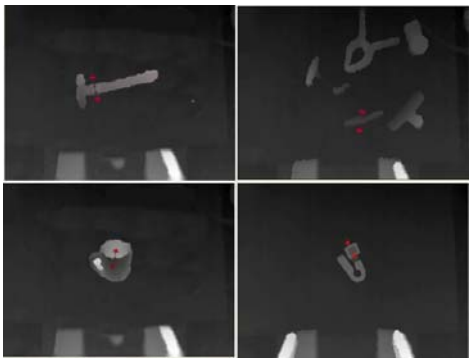

FIGURE VIII. BEST GRASP CANDIDATES

VII. CONCLUSIONS

In this paper we apply Grasp Quality Convolutional Neural Network to evaluate and rank each matched pair of grasping points. Through experimental verification, our method can meet the requirements, and it can quickly and stably find the grasp point of the test object. Although our training object has only 8 classes of objects, the input of the neural network is the local shape information of the object, we also extracted the local shape information of the object when we sample candidate candidates from the depth image of object , so our method for object grasping point selection has good generalization ability.

ACKNOWLEDGMENT

This work was partially supported by National Natural Science Foundation of China under Grant(NO.91648204); National Science and Technology Major Project under Grant (NO.2017ZX02101007-004).I would like to thank all the people for helping finish the paper.

REFERENCES

[1] Alexander Herzog, Peter Pastor, Mrinal Kalakrishnan, Ludovic Righetti, Jeannette Bohg, Tamim Asfour, and Stefan Schaal.Learning of grasp selection based on shape-templates. AutonomousRobots, 36(1-2):51–65, 2014.

[2] Stefan Hinterstoisser, Stefan Holzer, Cedric Cagniart, SlobodanIlic, Kurt Konolige, Nassir Navab, and Vincent Lepetit. Multimodaltemplates for real-time detection of texture-less objects inheavily cluttered scenes. In Computer Vision (ICCV), 2011 IEEE.International Conference on, pages 858–865. IEEE, 2011.

[3] Ken Goldberg, Brian V Mirtich, Yan Zhuang, John Craig, Brian RCarlisle, and John Canny. Part pose statistics: Estimators and experiments. IEEE Transactions on Robotics and Automation, 15(5):849–857, 1999.

[4] Andreas ten Pas and Robert Platt. Using geometry to detect grasp poses in 3d point clouds. In Intl Symp. on Robotics Research, 2015

[5] Jeffrey Mahler, Matthew Matl, Xinyu Liu, Albert Li, David Gealy, and Ken Goldberg. "Dex-Net 2.0: Deep Learning to Plan Robust Grasps with Synthetic Point Clouds and Analytic Grasp Metrics" IEEE International Conference on Robotics and Automation, 2017.

[6] Pelossof R, Miller A, Allen P, Jebara T. An SVM learning approach to robotic grasping. In:Proceedings of the 2004 IEEE International Conference on Robotics and Automation. New Orleans, USA:IEEE, 2004. 3512-3518

[7] Varley, C. DeChant, A. Richardson, A. Nair, J. Ruales, and P. Allen. Shape completion enabled robotic grasping. arxiv preprint: 1609.08546, 2016.

[8] Renaud Detry, Carl Henrik Ek, Marianna Madry, and DanicaKragic. Learning a dictionary of prototypical grasp-predicting parts from grasping experience. In Proc. IEEE Int. Conf. Robotics and Automation (ICRA), pages 601–608. IEEE, 2013.

[9] I-Ming Chen and Joel W Burdick. Finding antipodal point grasps on irregularly shaped objects. IEEE transactions on Robotics and Automation, 9(4):507–512, 1993.

[10] Alexander Herzog, Peter Pastor, Mrinal Kalakrishnan, LudovicRighetti, Jeannette Bohg, Tamim Asfour, and Stefan Schaal.Learning of grasp selection based on shape-templates. Autonomous Robots, 36(1-2):51–65, 2014.