Supplementary File for "Graspness Discovery in Clutters for Fast and Accurate Grasp Detection"

Chenxi Wang*, Hao-Shu Fang*, Minghao Gou, Hongjie Fang, Jin Gao, Cewu Lu†
Shanghai Jiao Tong University
wcx1997@sjtu.edu.cn, fhaoshu@gmail.com, {gmh2015, galaxies}@sjtu.edu.cn,
gao3944677@126.com, lucewu@sjtu.edu.cn

1. Video Demo

A video demo is attached for real grasping using the results predicted by GSNet which is trained on GraspNet-1Billion. Watch the video "demo.mp4" for more details. Notably, some objects (chain, mesh bag with marbles, slipper, etc.) in the demo are not collected from GraspNet-1Billion [1] and our model shows robustness on novel objects.

2. Grasping Experiment Configuration

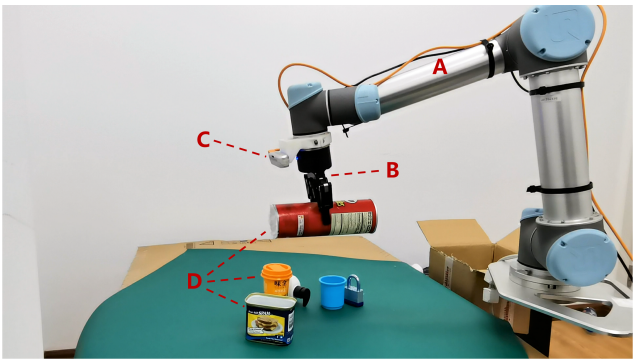


Figure 1. Configuration of real grasping experiments. A: UR-5 robotic arm. B: Robotiq two-finger gripper. C: RealSense D435 camera. D: object models from GraspNet-1Billion dataset.

Fig. 1 shows the configuration of our grasping experiments. A grasp with a high score output by GSNet is chosen and sent to the robotic arm. The program attempts to grab one object each time, and repeat execution until all the objects are cleaned from the table.

* denotes equal contribution.

Cewu Lu is the corresponding author, member of Qing Yuan Research Institute and MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, China and Shanghai Qi Zhi institute.

3. Robotic Experiments on Baselines

Method	Success Rate		
	Set A	Set B	Set C
GPD [3]	71.43%	74.07%	67.80%
PointNetGPD [2]	76.92%	78.43%	70.18%
Fang *et al.* [1]	81.63%	80.00%	74.07%
Ours	**88.89%**	**90.91%**	**85.11%**

Table 1. Success rate on real robot experiments.

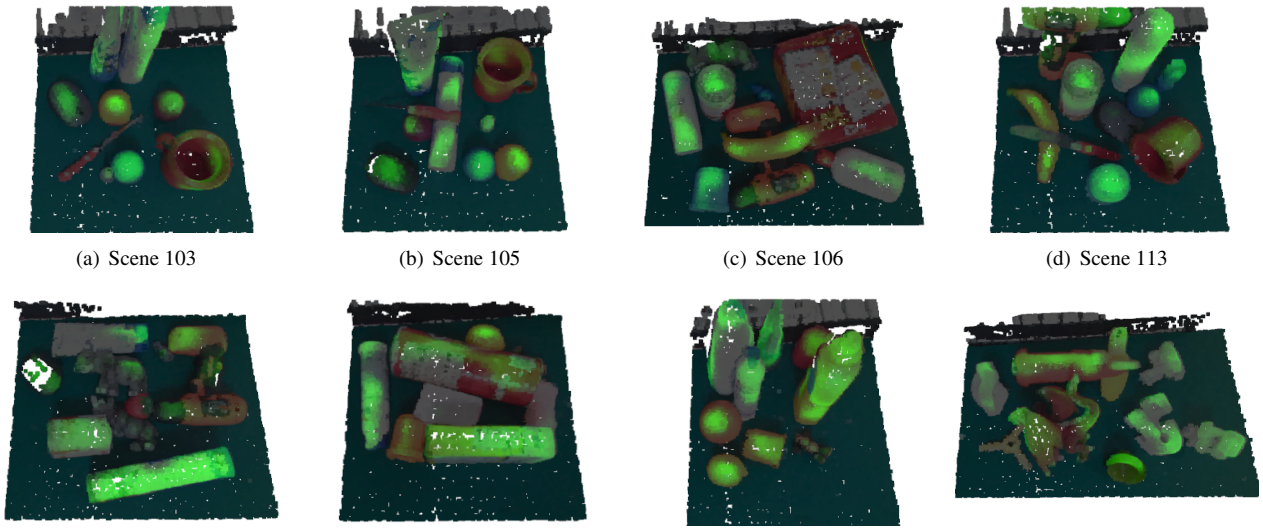
We compare our methods with other baselines [1, 2, 3] in real experiments. Objects are divided into three sets, each containing 10 objects from [1]. These methods are used to remove all the objects in the workspace with single-view point clouds as input. Four repeated experiments are conducted for each object set. Tab. 1 shows the results of different methods, where success rate is defined in Sec. 5.6. GSNet outperforms other methods on all three sets.

4. Details of Grasp Operation Model

Grasp operation model (GOM) in GSNet is designed based on OperationNet in GraspNet baseline model [1] with several improvements. The main differences between the two components are listed as follows.

Simplified Cylinder Representation In [1], points are cropped and transformed into a cylinder region for each depth bin, which leads to multiple groups with repeated points on one grasp proposal. GOM replaces them with a single cylinder region, where the height is determined by the maximum depth (0.04m) used in our experiments. Depth classification is moved to final output accordingly.

Scaled Point Coordinates In [1], points are transformed without scaling. Since all groups shared the same gripper coordinate frame and the transformed coordinates are relative small in abusolute value ($<=$ 0.05m), we scaled the points with the cylinder radius (0.05m). The width prediction is modified accordingly.



(a) Scene 103 (b) Scene 105 (c) Scene 106 (d) Scene 113

(e) Scene 118 (f) Scene 146 (g) Scene 150 (h) Scene 183

Figure 2. Visualization of point-wise graspness predicted by GSNet. Regions with higher graspness are annotated with brighter colors.

Shared Point Features For each depth bin on one grasp point, OperationNet in [1] directly samples 64 points with only xyz coordinates from the original input (about 20k points). GOM samples 16 points from the seeds (about 1k points). The xyz coordinates are then concatenated with point features output by cascaded grasp model. This modification helps reduce computing overhead of point sampling.

Output Parameters [1] output grasp scores, in-plane rotation angles and gripper widths for each depth bin, and choose the parameters with the highest angle classification scores. In GOM, grasp scores and gripper widths are predicted for each (in-plane rotation)-(approach depth) combination and output the parameters of the combination with the highest grasp score. In addition, output grasp scores and gripper widths are replaced with relative value from 0 to 1.

5. Visualization of Point-wise Graspness

Point-wise graspness predicted by GSNet is visualized in Fig. 2. Regions with higher graspness are annotated with brighter colors. We can see that graspness is not only decided by the object itself, but also influenced by its position. Most of the low-graspness areas are caused by collision with tables, which have higher graspness in single object. Graspness of an object is also influenced by its neighbours. For example, the knife on the banana (Fig. 2d) cut off the contiguous graspness of the latter. The object size is also an important factor. In Fig. 2c, the box has no areas with high graspness although the shape is relative simple. That is because there are few areas for a gripper with width up to 0.1m to grasp on when the box are lying on the table.

6. t-SNE Visualization of Point Features

We visualize the point features output by GSNet. Fig. 3 shows the t-SNE visualization of point features in different test setting, where points are obtained from GraspNet-1Billion dataset and feature vectors are output by GSNet. Tab. 2 details the experimental settings. The points with high graspness are labeled as positive samples, and other points are labeled as negative samples. We can see that graspable points are quite distinguishable from others, which demonstrates the generality of graspness model across different settings.

References

[1] Hao-Shu Fang, Chenxi Wang, Minghao Gou, and Cewu Lu. Graspnet-1billion: A large-scale benchmark for general object grasping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11444–11453, 2020. 1, 2

[2] Hongzhuo Liang, Xiaojian Ma, Shuang Li, Michael Görner, Song Tang, Bin Fang, Fuchun Sun, and Jianwei Zhang. Pointnetgpd: Detecting grasp configurations from point sets. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 3629–3635. IEEE, 2019. 1

[3] Andreas ten Pas, Marcus Gualtieri, Kate Saenko, and Robert Platt. Grasp pose detection in point clouds. *The International Journal of Robotics Research (IJRR)*, 36(13-14):1455–1473, 2017. 1

	Object Variation				Viewpoint Variation				Camera Variation		
	Train	Test			Train	Test			Train	Test	
Scene	0-99	100-129	130-159	160-189	0-99	100-129	100-129	100-129	0-99	100-129	100-129
View	0-255	0-255	0-255	0-255	0-127	0-127	128-191	192-255	0-255	0-255	0-255
Camera	Kinect	Kinect	Kinect	Kinect	Kinect	Kinect	Kinect	Kinect	Kinect	Kinect	Realsense
Serial Number	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3

Table 2. Serial number on different test setting. Each number corresponds to a t-SNE visualization result in Fig. 3.

Figure 3. t-SNE visualization of point features on different test setting. The three rows show the results of object/viewpoint/camera variation respecctively. Orange points denote the samples with high graspness, and blue points denote the samples with low graspness. The setting details are listed in Tab. 2.