

Supplementary file for “GraspNet-1Billion: A Large-Scale Benchmark for General Object Grasping”

Hao-Shu Fang, Chenxi Wang, Minghao Gou, Cewu Lu¹
Shanghai Jiao Tong University

fhaoshu@gmail.com, {wcx1997, gmh2015, lucewu}@sjtu.edu.cn

1. Data Collection

We provide the detailed settings of our data collection process in Figure 1. (a) Both the RealSense and Kinect camera are fixed on the end link of a robot arm. (b) 256 data collection points are sampled from a quarter sphere.

2. Grasp Pose NMS

For two grasps \mathbf{G}_1 and \mathbf{G}_2 , we define grasp pose distance $D(\mathbf{G}_1, \mathbf{G}_2)$ as a tuple:

$$D(\mathbf{G}_1, \mathbf{G}_2) = (d_t(\mathbf{G}_1, \mathbf{G}_2), d_\alpha(\mathbf{G}_1, \mathbf{G}_2)), \quad (1)$$

where $d_t(\mathbf{G}_1, \mathbf{G}_2)$ and $d_\alpha(\mathbf{G}_1, \mathbf{G}_2)$ denote translation distance and rotation distance of two grasps respectively. Let a grasp pose \mathbf{G} be denoted by a translation vector \mathbf{t} and a rotation matrix \mathbf{R} , then $d_t(\cdot)$ and $d_\alpha(\cdot)$ is defined as:

$$\begin{aligned} d_t(\mathbf{G}_1, \mathbf{G}_2) &= \|\mathbf{t}_1 - \mathbf{t}_2\|, \\ d_\alpha(\mathbf{G}_1, \mathbf{G}_2) &= \arccos \frac{1}{2}(\text{tr}(\mathbf{R}_1 \cdot \mathbf{R}_2^T) - 1), \end{aligned} \quad (2)$$

where $\text{tr}(\mathbf{M})$ denotes the trace of matrix \mathbf{M} .

Since translation and rotation are not in the same metric space, we define the NMS threshold as a tuple too. Let $TH = (th_d, th_\alpha)$, we say $D(\mathbf{G}_1, \mathbf{G}_2) < TH$ if and only if

$$d_t(\mathbf{G}_1, \mathbf{G}_2) < th_d, \quad d_\alpha(\mathbf{G}_1, \mathbf{G}_2) < th_\alpha. \quad (3)$$

Based on the tuple metric, two grasps are merged when their distance is lower than TH . Meanwhile, only the top K grasps from each object are considered according to confidence scores and other grasps are omitted. In evaluation, we set $th_d = 3$ cm, $th_\alpha = 15$ degree and $K = 10$.

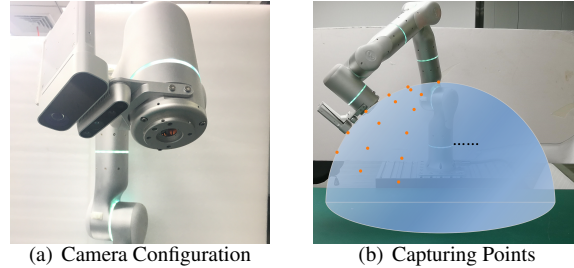


Figure 1. The setting of data collection.