

## 1 Appendix

2 In this appendix, we first describe the implementation details in Section A, followed by additional  
3 experimental results in Section B.

### 4 A Implementation Details

#### 5 A.1 Training Details

6 We train our model on 8 NVIDIA RTX 4090 GPUs with a batch size of 64. We use the AdamW  
7 optimizer with a learning rate of  $6e^{-5}$  and trained the model for 130 epochs. The training procedure  
8 takes about 2 days.

#### 9 A.2 Hyperparameters

10 The hyperparameters employed in our experiments are detailed in Table 1. The size of IBS volume  
11 is configured to  $0.2\text{ m} \times 0.2\text{ m} \times 0.2\text{ m}$ , adequately encompassing the interaction space between  
12 the dexterous hand and the object, while remaining sufficiently compact to focus on the critical  
13 local grasping region. The resolution of the IBS volume is set to  $40 \times 40 \times 40$ , striking a balance  
14 between computational efficiency and the accuracy of the IBS surface representation. Both the IBS  
15 sampling and grasp pose optimization processes are executed concurrently, with the number of IBS  
16 candidates and grasp poses each limited to 5. The weights in the contact energy  $\mathbf{E}_d$  are meticulously  
17 adjusted to balance the contacts between the object and the thumb, as well as the other fingers,  
18 whereas the weights in the overall energy  $\mathbf{E}$  are calibrated to harmonize the various energy terms.  
19 The hyperparameter for denoising timesteps is adopted from [1].

Hyperparameter	Value
IBS Volume Size	$0.2m \times 0.2m \times 0.2m$
IBS Resolution ( $n$ )	$40 \times 40 \times 40$
Number of IBS Candidates ( $m$ )	5
Number of Grasp Poses ( $k$ )	5
Weights in Contact Energy $\mathbf{E}_d$ ( $\alpha_1, \alpha_2, \alpha_3$ )	80, 100, 2
Weights in Overall Energy $\mathbf{E}$ ( $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ )	5, 1, 1000, 1
Denoising Timesteps	50

Table 1: Hyperparameters used in our experiments.

### 20 B More Results

21 In this section, we provide additional results of our method to demonstrate the effectiveness, robust-  
22 ness, scalability, and generalization ability of our method.

#### 23 B.1 Qualitative Results

24 In this section, we present additional qualitative evaluations of our proposed method. Figure 1  
25 showcases the perception results of our method in both simulated and real-world environments. The  
26 results clearly demonstrate that our approach successfully discerns the sparse IBS volume pertinent to  
27 a success grasp pose from a single-view point cloud in cluttered settings. It effectively differentiates  
28 the contact regions of the thumb and other fingers with the object. Furthermore, the second-stage  
29 optimization, guided by Sparse IBS constraints, adaptly refines collision-free and plausible grasp  
30 poses. Notably, the adoption of point cloud representation minimizes the sim-to-real gap, ensuring  
31 that our method can generalize to real-world settings without additional training, thereby achieving  
32 robust dexterous grasping capabilities.

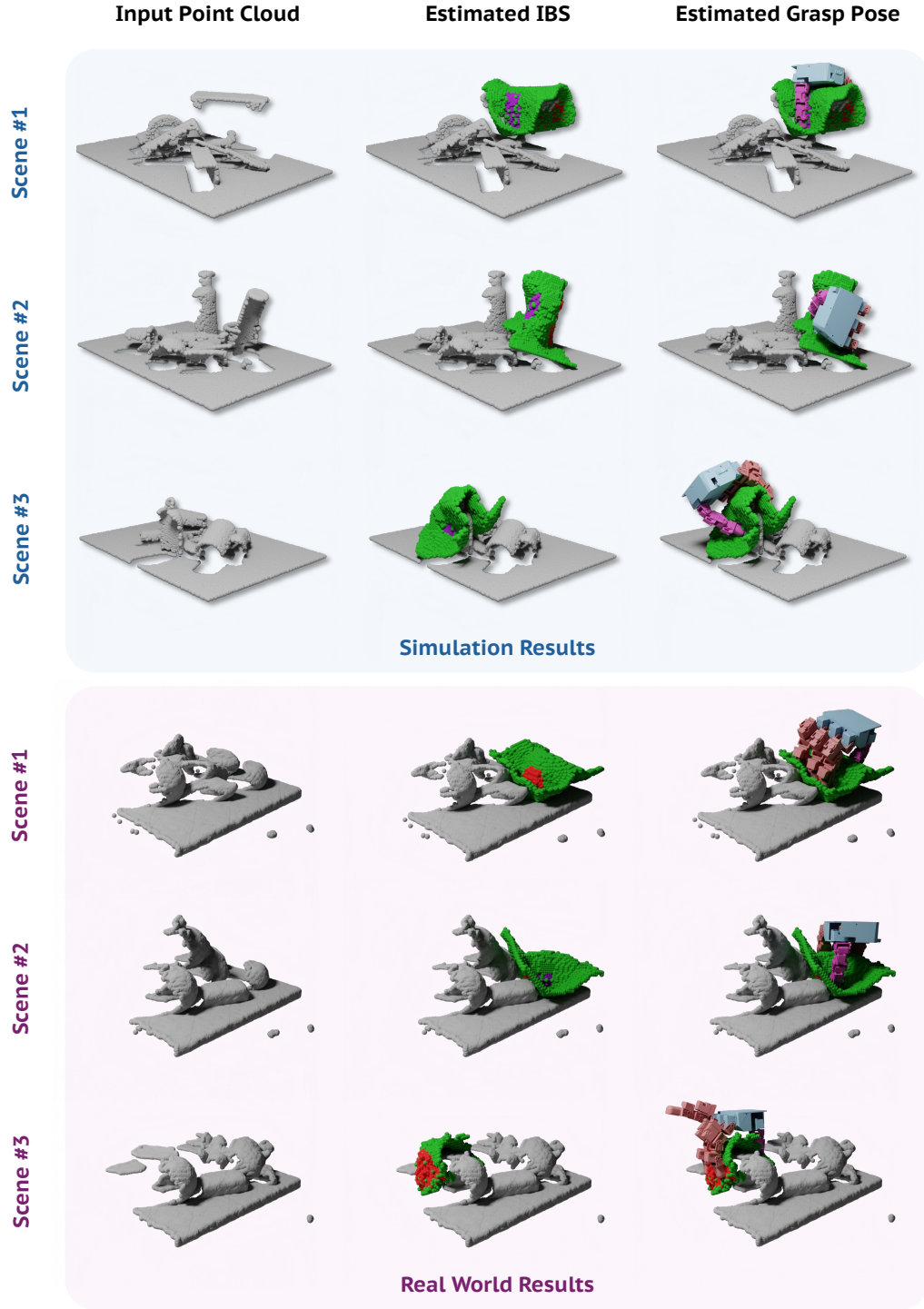


Figure 1: Qualitative results of our method in simulated (upper panel) and real-world environments (lower panel). From left to right: the initial single-view point cloud input, the sparse IBS prediction from the initial stage, and the optimized grasp pose from the subsequent stage. The purple areas indicate thumb contact, red areas denote contact by other fingers, and green areas represent non-contact regions on the IBS surface.

## 33 B.2 Grasp Diversity Analysis

34 We analyze the distribution of joint configurations for all predicted grasp poses by **CADGrasp** and  
 35 DexGraspNet2.0 [2] within the GraspNet-1B loose scenarios [2]. This analysis aimed to compare the  
 36 diversity of grasp poses generated by the two methods. Taking the thumb as an example, Figure 2  
 37 illustrates that our method can generate diverse grasps with significantly higher dexterity compared  
 38 to DexGraspNet2.0. This highlights our approach’s capability to produce a wider range of effective  
 grasp configurations, enhancing its applicability in complex scenarios.

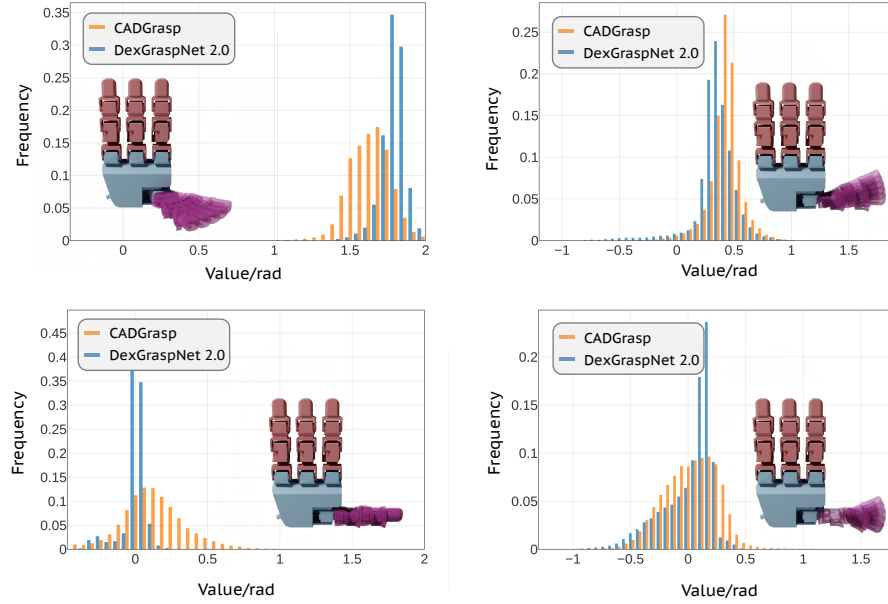


Figure 2: Grasp diversity analysis of **CADGrasp** and DexGraspNet2.0 within the GraspNet-1B loose scenarios. The histogram illustrates the distribution of joint configurations for the thumb.

39

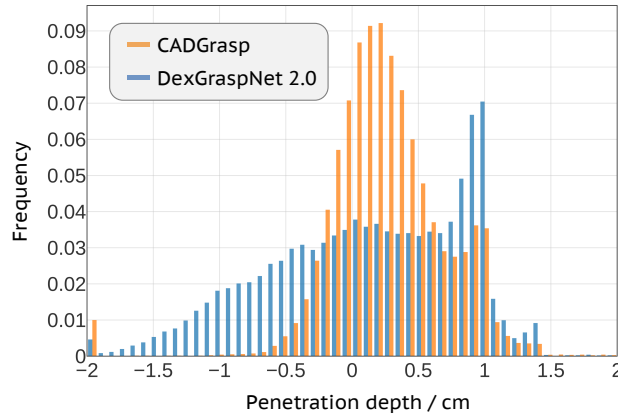


Figure 3: Penetration depth analysis of predicted grasp poses by **CADGrasp** and DexGraspNet2.0 within the GraspNet-1B loose scenarios. The histogram illustrates the distribution of maximal penetration depths.

### 40 **B.3 Grasp Quality Analysis**

41 In evaluating grasp quality, the proximity and penetration between the predicted grasp pose and the  
42 object serve as crucial indicators. We analyzed the maximal penetration depth (in cm) of all predicted  
43 grasp poses by **CADGrasp** and DexGraspNet2.0 within the GraspNet-1B loose scenarios. This  
44 metric is defined as the maximal penetration depth from the object point cloud to the hand meshes.  
45 As illustrated in Figure 3, our method demonstrates a concentration of penetration depths near the  
46 object’s surface (penetration depth = 0), attributed to the constraints imposed by the IBS Surface  
47 and contact points. This further underscores the superiority of our approach in achieving precise and  
48 effective grasping.

49 **References**

- 50 [1] Xin-Yang Zheng, Hao Pan, Peng-Shuai Wang, Xin Tong, Yang Liu, and Heung-Yeung Shum.  
51 Locally attentional sdf diffusion for controllable 3d shape generation. *ACM Transactions on*  
52 *Graphics (ToG)*, 42(4):1–13, 2023.
- 53 [2] Jialiang Zhang, Haoran Liu, Danshi Li, XinQiang Yu, Haoran Geng, Yufei Ding, Jiayi Chen,  
54 and He Wang. Dexgraspnet 2.0: Learning generative dexterous grasping in large-scale synthetic  
55 cluttered scenes. In *8th Annual Conference on Robot Learning*, 2024.