Supplementary Material *for* ACL-SPC: Adaptive Closed-Loop system *for* Self-Supervised Point Cloud Completion

Sangmin Hong[1*] Mohsen Yavartanoo[2*] Reyhaneh Neshatavar[2] Kyoung Mu Lee[1,2]
[1]IPAI, [2]Dept. of ECE & ASRI, Seoul National University, Seoul, Korea
{mchiash2,myavartanoo,reyhanehneshat,kyoungmu}@snu.ac.kr

S1. Ablation on network architecture

We use PolyNet [47] as the encoder is that it is an advanced GraphCNN that can learn local features efficiently. To justify the design choice of our proposed architecture, we have tested our framework by changing both the encoder and decoder modules. We compare the chamfer distance, per epoch training time, and per sample testing time when converting the encoder, decoder, or both to PCN [49] or TopNet [37]. The results are shown in Table S1. The results demonstrate that our method can leverage PolyNet to improve performance. We further show that using our shallower decoder $\mathcal{D}$ can save the test time while obtaining slightly better performance. We note that even with a much simpler encoder (PCN) our method is still able to achieve relatively satisfactory performance.

Network Architecture	CD↓	Train time (s)↓	Test time (ms)↓
TopNet [37]	3.21	229.6	61.5
PCN [49]	2.99	245.1	40.2
PCN encoder [49] + $\mathcal{D}$	3.07	**118.1**	39.9
$\mathcal{E}$(PolyNet[47]) + PCN decoder [49]	2.93	314.3	42.3
Ours ($\mathcal{E}$(PolyNet[47]) + $\mathcal{D}$)	**2.88**	179.8	**38.5**

Table S1. **Ablation studies for the choice of the network architecture.** The values of CD are multiplied by 100. The train time is measured per epoch and the test time is measured per sample.

S2. Analysis of losses

We use a global average pooling after the last PolyConv layer to eliminate the spatial dependency and obtain permutation-invariant representative features of an object. Therefore, for different partial observations of the same object, the encoder must generate the same representative features. Accordingly, the decoder can generate the complete point clouds C_v with the same point orders for different observations. Moreover, in our $\mathcal{L}^{\text{cons}}$ loss, instead of using the original or GT point clouds which have different orders from the generated complete point clouds C_v, we reduce the MSE between the points of the generated complete point clouds C_0 and C_v which have the same orders. We further justify our claim by replacing our loss function $\mathcal{L}^{\text{cons}}$ with the Chamfer loss $\mathcal{L}^{\text{chamf}}$ between C_0 and C_v and showing the effectiveness of $\mathcal{L}^{\text{cons}}$ as shown in Table S2.

Loss function	P↓	C↓	CD↓
$\mathcal{L}^{\text{chamf}}$	**1.69**	2.53	4.22
$\mathcal{L}^{\text{cons}}$	1.70	**1.18**	**2.88**

Table S2. **Evaluation on loss function.** P, C, and CD refers to precision, coverage, and chamfer distance respectively. All the avalues are multiplied by 100.

S3. Qualitative results on Asymmetric Data

We extract different partial views from asymmetric objects and used them to train and evaluate our method. According to Figure S1, the results show that our method reconstructs the asymmetric shapes reliably as well as symmetric shapes.

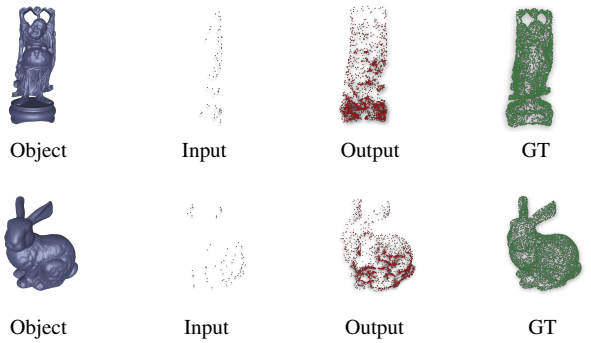

Figure S1. **Qualitative results on Asymmetric Data.**

*equal contribution

Supervision	Method	Plane		Cabinet		Car		Chair		Lamp		Sofa		Table		Vessel		Average	
		P↓	C↓	P↓	C↓	P↓	C↓	P↓	C↓	P↓	C↓	P↓	C↓	P↓	C↓	P↓	C↓	P↓	C↓
Supervised	PCN [49]	0.58	0.52	1.01	1.12	0.90	**0.83**	1.17	1.03	1.33	0.94	1.23	1.10	0.88	0.84	1.07	0.86	1.02	0.91
	GRNet[43]	0.68	0.61	1.08	**1.00**	0.98	0.91	1.01	0.87	0.89	0.71	1.12	0.99	0.89	0.80	0.88	0.73	0.94	0.83
	SFNet [42]	**0.46**	**0.39**	**0.85**	1.01	**0.80**	0.84	**0.72**	**0.82**	**0.55**	**0.64**	**0.90**	**0.96**	**0.61**	**0.68**	**0.60**	**0.67**	**0.69**	**0.75**
Self-supervised	**Ours**	1.81	0.58	2.80	1.28	1.96	0.98	3.43	1.10	5.31	0.98	3.08	1.26	3.17	1.05	3.40	0.82	3.12	1.01

Table S3. **Evaluation and comparison on eight categories of PCN [49] dataset.** P and C refers to precision and coverage, respectively. All the values are multiplied by 100. Compared to supervised methods, our result shows competitive results on coverage.

S4. Evaluation on PCN

PCN [49] dataset originates from ShapeNet [5] dataset where a set of $16,384$ points are uniformly sampled from the surface of each shape. Partial point clouds are obtained by projecting the 3D point cloud to 2.5D data from eight random views and back projecting to 3D for each object. Note that the points are resampled to a fixed number of $3,096$ points for each partial point cloud.

As shown in Table S3, we train and evaluate our method on the PCN [49] dataset and compare it with the supervised methods. Even though our self-supervised method ACL-SPC does not achieve the best performance among the supervised methods, it shows decent performance in the coverage, which demonstrates the ability of our ACL-SPC method to cover the missing parts without using the ground-truth. However, since there is no such strict supervision as a complete point cloud, our method has higher precision, implying that it is noisier than the supervised methods.

We also qualitatively compare our results to the supervised methods PCN [49], GRNet [43], and SFNet [42] on the PCN [49] dataset in Figure S2. The results demonstrate that our method can fill the missing parts of the input almost par with the supervised methods. Furthermore, we quantitatively evaluate our model on unseen categories of PCN [49] dataset. The results in Table S4 illustrate that our method can achieve superior performance on unseen categories compared with the supervised methods in terms of coverage in average. We can conclude from the results that our method is applicable in real-world situations where there is no information about the object's category.

S5. Learning progress

We visualize the learning progress of our ACL-SPC framework by the output results from different epochs in Figure S3. We visualize the results at the epoch of 1, 10, 100, 300, and that when it has the best chamfer distance result. The results demonstrate how our method can gradually learn to fill (epoch best) a given partial point cloud starting from a point cloud with naive shape (epoch 1).

S6. Robustness to Viewpoint

We also illustrate the effect of viewpoint in Figure S4. The results indicate that the quality of the generated complete point clouds is almost invariant to the point of view. This robustness can be due to the attribute of our designed ACL-SPC framework that receives synthesized partial point clouds from random views during training.

S7. Synthesized partial point clouds

We visualize the generated synthetic partial point clouds from the predicted complete output in Figure S5. It shows that our partial point cloud generator can synthesize realistic partial point clouds from a predicted complete output.

S8. Training dataset

In Figure S6, we compare our method trained individually on each category to one trained on multiple categories. Our qualitative comparisons show that our method trained on a training dataset with multiple categories can fill in the missing region similarly to the single case. As a result, our method is practical in real-world scenarios involving a variety of samples from various categories.

S9. Failure cases

We also show our failure results in Figure S7. We show that our method generates redundant points when similar objects are rare in the training dataset. However, it can still restore the missing parts in the given partial point cloud. As mentioned in the manuscript, it is due to no restrictions to not generate redundant points which leads to high precision values. We will develop our self-supervised framework in denoising tasks to improve the results for the failure cases.

S10. Ablation on α and β

We quantitatively evaluate our method under different settings of α and β parameters in equation 5 of the main manuscript. The results in Table S5 demonstrate that our method can achieve the best performance with $\alpha = 0.1$ and $\beta = 0.9$ in average.

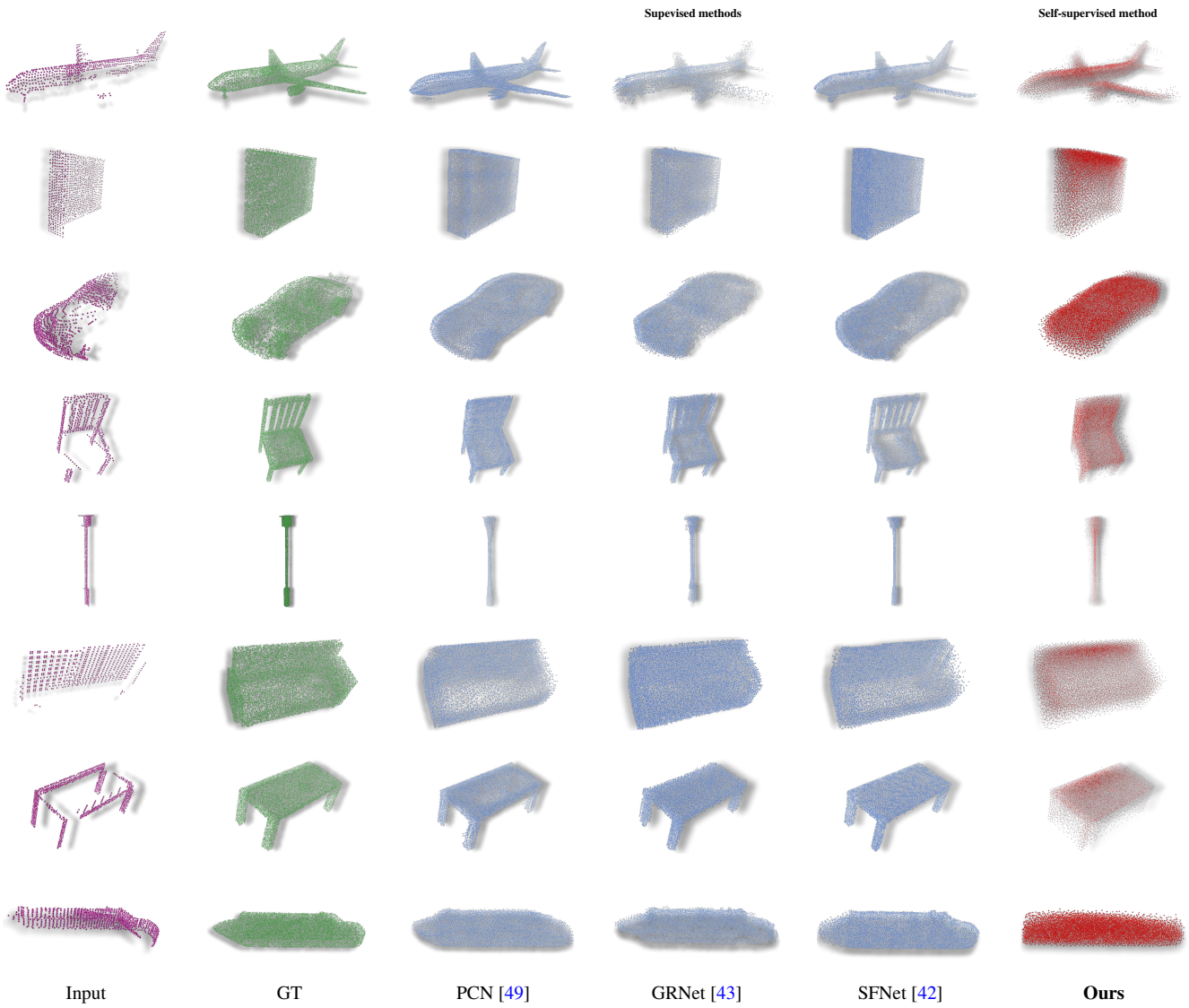

Figure S2. **Qualitative comparison on the PCN [49] dataset.** PCN [49], GRNet [43], and SFNet [42] are supervised methods, whereas our method is self-supervised and is trained without the use of ground-truth.

Supervision	Method	Bus		Bed		Bookshelf		Bench		Guitar		Motorbike		Skateboard		Pistol		Average	
		P	C	P	C	P	C	P	C	P	C	P	C	P	C	P	C	P	C
Supervised	PCN [49]	**0.96**	**0.93**	2.68	1.65	1.75	1.21	**1.32**	**0.88**	1.20	0.88	1.74	1.21	1.57	0.83	1.50	1.34	1.59	1.12
	GRNet [43]	1.13	1.18	**2.38**	2.10	**1.63**	1.34	1.38	0.92	1.03	0.75	1.14	1.23	1.43	0.83	1.12	1.53	**1.41**	1.24
	SFNet [42]	1.13	1.25	2.90	2.80	2.39	2.01	1.60	1.21	**0.49**	**0.62**	**1.13**	1.12	**0.74**	0.82	**1.10**	1.36	1.44	1.40
Self-supervised	**Ours**	2.25	0.96	4.90	**1.43**	2.98	**1.17**	3.32	0.99	4.57	0.81	3.14	**0.87**	2.90	**0.70**	5.47	**0.92**	3.69	**0.98**

Table S4. **Evaluation and comparison on eight unseen categories of pcn [49] dataset.** P and C refers to precision and coverage, respectively. All the values are multiplied by 100. Compared to supervised methods, our result shows competitive results on coverage.

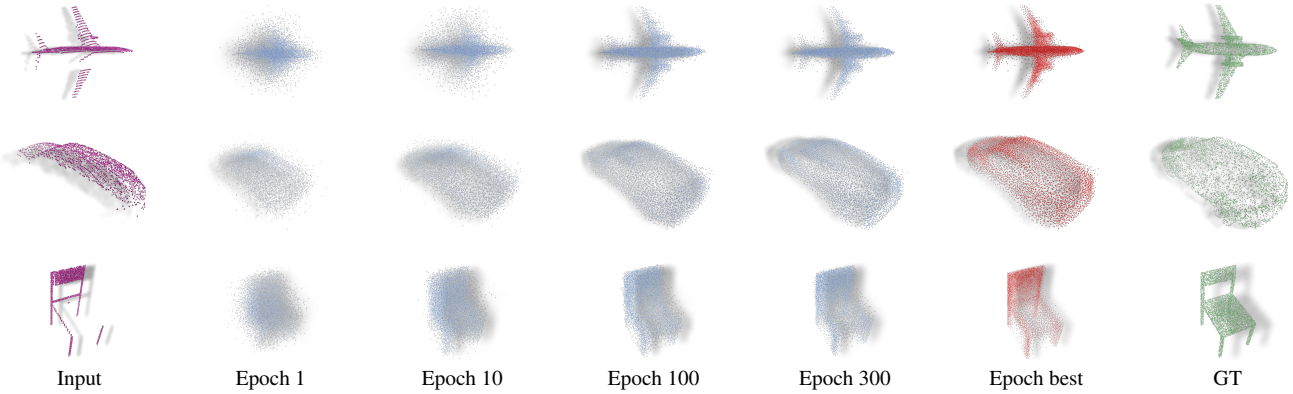

Figure S3. **Qualitative improvement by iteration.**

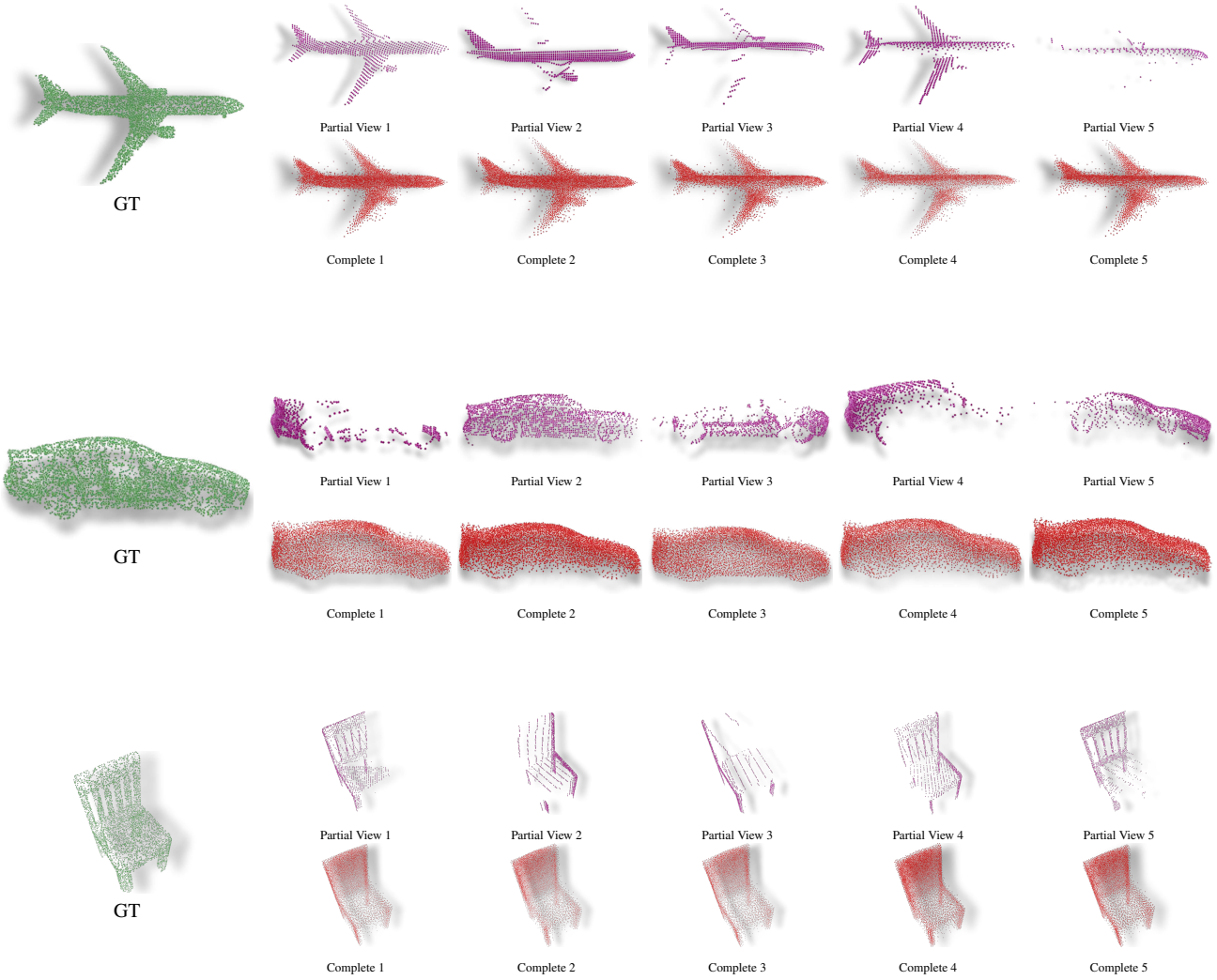

Figure S4. **The qualitative effect of viewpoint.**

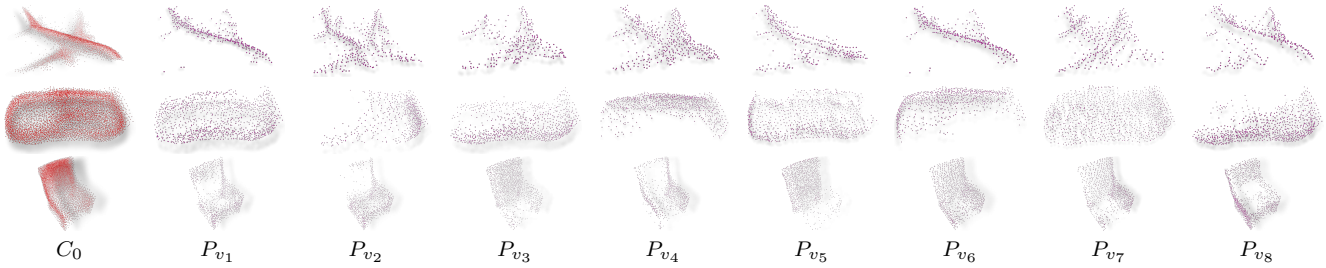

Figure S5. **Visualization of the synthesized partial point clouds.**

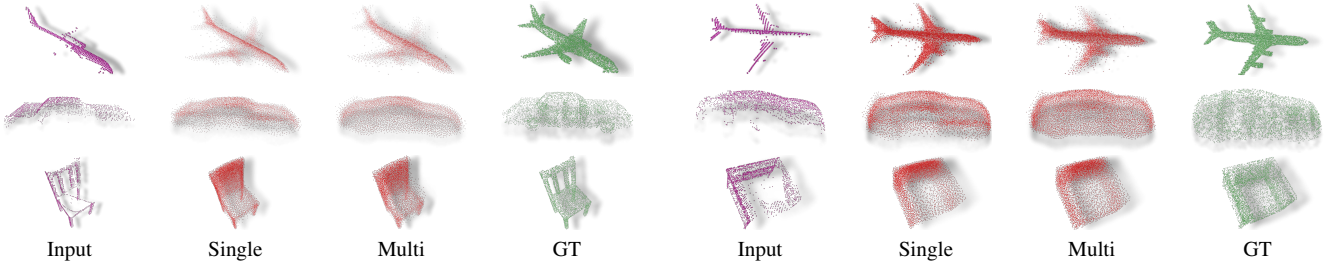

Figure S6. **Qualitative comparisons of our method trained on multi-class vs. single-class.**

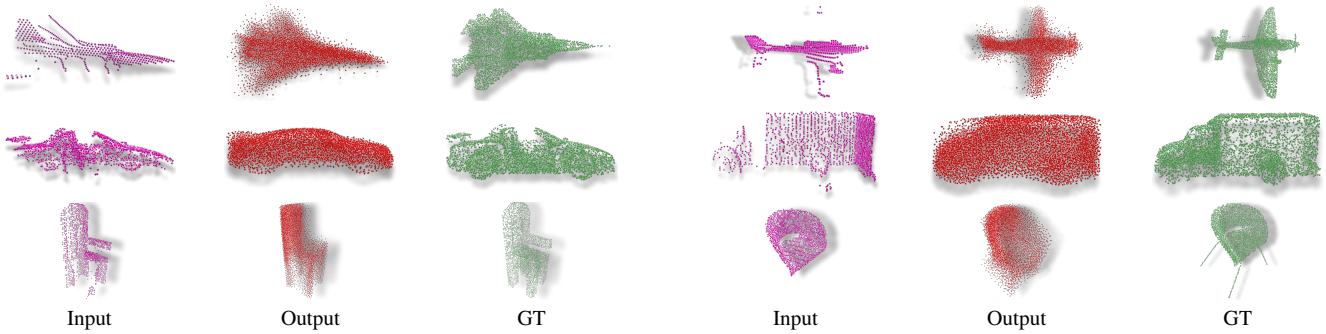

Figure S7. **Visualization of failure results.**

α	β	Airplane			Car			Chair		
		P	C	CD	P	C	CD	P	C	CD
1	0	0.72	23.42	24.14	1.25	27.99	29.23	1.22	37.61	38.84
0.9	0.1	0.67	2.57	3.24	1.26	4.94	6.2	1.05	4.21	5.26
0.5	0.5	0.74	1.71	2.45	1.25	3.17	4.42	1.21	2.6	3.81
0.1	**0.9**	1.20	0.80	**2.01**	1.65	1.28	2.93	2.25	1.46	**3.71**
0	1	3.33	0.75	4.08	1.79	1.09	**2.89**	4.49	1.23	5.72

Table S5. **Ablation Studies for α and β.** We also calculate average values among the categories. P, C, and CD refers to precision, coverage, and Chamfer distance, respectively. All the values are multiplied by 100.