Supplementary Material for
SCoDA: Domain Adaptive Shape Completion for Real Scans

Table S1. Statistics of the proposed dataset ScanSalon.

	Chair	Desk	Sofa	Bed	Lamp	Car	Total
Synthetic Scans	6,579	8,071	3,091	233	2,318	3,514	23,806
Real Scans	4,651	1,630	428	365	133	437	7,644
Paired Models	497	161	43	36	20	43	800

1. More ScanSalon Details

We provide a comparison between the number of synthetic and real scans in Tab. S1. The synthetics scans are from the ShapeNet dataset [1]. Besides the class "Bed", there are more samples of synthetic scans than real scans.

2. More Implementation Details

Synthetic Scan Generation To get point clouds from the ShapeNet models, we adopt a popular simulation toolbox, BlenSor [3], which supports scanning simulation with different sensors (*e.g.* Velodyne, Kinect, and Time of Flight camera) and parameters. To simulate the sparsity in real scans from the ScanNet dataset, we conduct a random down-sampling with ratio 13%, which is computed according to the average point number of scans in ScanNet and simulated point clouds from ShapeNet. We also add Gaussian noise with a max scale 0.01 to simulate the noise in scanning. The incompleteness is also introduced in the simulated scanning by self-occlusion. Besides, we adopt the unsupervised clustering-based way introduced in the main paper to partition the point clouds and randomly drop 1∼4 clusters in the training process. Some simulation results (before dropping some clusters in training) are presented in the second row of Fig. S1. As shown in Fig. S1, in spite of careful simulation, the generated synthetic scans are still different from real scans (the first row of Fig. S1) in (i) sparsity: dependent of the object materials and object-scanner distance, the sparsity of real scans has a larger variance; (ii) noise: both the scanning intrinsic error the segmentation from background can introduce complex noise; and (iii) incompleteness: the incompleteness of real scans results from the occlusion and surroundings, which is more complex. Thus, there still exists a domain gap between two kinds of point clouds, which can hardly be handled by simulation.

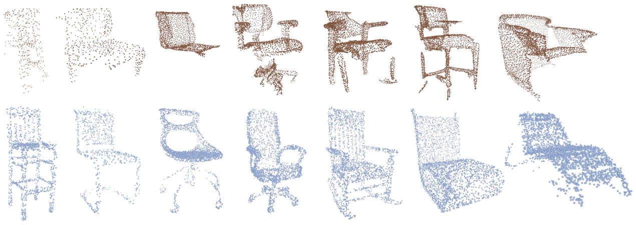

Figure S1. Comparison between real scans (the first row, from the ScanNet dataset) and synthetic scans (the second row, simulated from models in the ShapeNet dataset).

Implementation of Baselines The implemented baselines include: (i) IF-Net: it consists of a 6-layer 3DCNN with a 4-layer MLP, and the network structure is used in all baselines; (ii) SelfSup: a mean-square-error (MSE) loss is used to minimize the Euclidean distance between the normalized top-layer feature vectors generated from the two views, which are created in the same way as in our method; (iii) PtComp: two 3DCNN-based UNets (5 layers in the encoder/decoder) are used for the encoding-decoding of voxelized point clouds from the real and synthetic domains for point completion, respectively. An adversarial loss is used to encourage the domain invariance of the codes output by the encoders following [2], and for supervised samples, an additional MSE loss is used to minimize the distance between the generations and the ground truths. The completion results share the same resolution with the input, so are then fed into a standard IF-Net for reconstruction; (iv) Adversarial: based on a standard IF-Net-based reconstruction framework, an adversarial loss is adopted to minimize the domain discriminativeness of the top-layer feature vectors generated by the 3DCNN, where a three-layer MLP (the latent dimensions are 256 and 512) is used as the discriminator, and the adversarial loss is rescaled by a ratio of 0.01 to be integrated.

3. More Qualitative Results

We present the qualitative comparison between our method and all baselines on the 3% label setting, and here we give the visualized results on the 5% label setting in Fig. S2. It can be seen that the proposed method is superior in reconstruction also on the 5% setting.

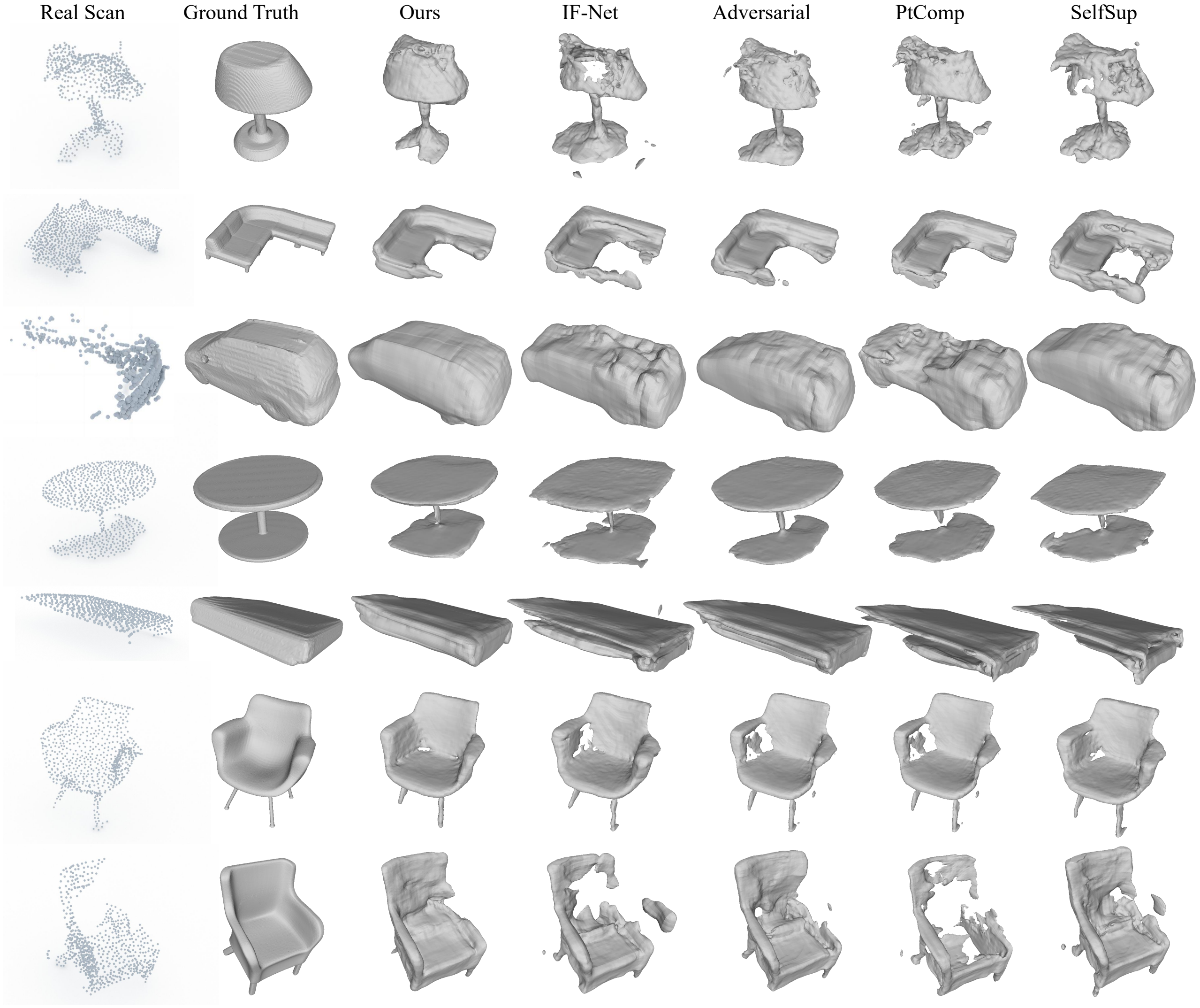

Figure S2. Qualitative comparison between different methods on shape completion with only 5% labels for training.

4. More Ablative Results

Single Module of CDFF and VCST In the main paper, we list the results of using only CDFF or VCST module in 3 classes for the space limit. We provide the results on all 6 classes on the 3% label setting in Tab. S2.

CDFF variants For qualitative comparison of the ablative experiments, we provide the visualization of 3 chair samples generated by different variants of the CDFF module in Fig. S3. Among all variants of CDFF, the completion results are relatively worse without feature fusion (columns (4) and (5)). When only using knowledge from the source domain (column (4)), the reconstruction results tend to rely more on the input scan but create less completion. A potential reason is the training data in the source domain are cleaner and with supervision, which introduces the domain bias. Differently, when only using knowledge from the target domain (column (5)), the reconstructions tend to produce more completion but have a worse global shape (*e.g.* the second row of column (5)).

VCST variants The reconstruction results by different variants of the VCST module are visualized in Fig. S4. We can observe that (i) when using random down-sampling only, the network tends to give worse completion for the missing object components, which proves the necessity of employing surface-aware augmentation. (ii) when integrating the volume-aware augmentation (columns (6), (8), and (9)), the reconstructions tend to over-complete the shapes (the first two rows of columns (6), (8), and (9)), of which a potential reason is that our clustering-based surface-aware augmentation implies some object-specific information, which creates incompleteness that is more close to the ones in real scans. These results further validate the superiority of surface-aware augmentation.

Table S2. Experiment results on the 3% setting of the SCoDA task (the complete results on 6 classes of Tab. 3 in the main paper). The units of CD and mIoU value are 1×10^{-3} and %, respectively. **Red** text indicates the best result.

	Chair		Desk		Sofa		Bed		Lamp		Car		Average	
Method	CD↓	mIoU↑	CD↓	mIoU↑	CD↓	mIoU↑	CD↓	mIoU↑	CD↓	mIoU↑	CD↓	mIoU↑	CD↓	mIoU↑
CDFF+VCST	1.58	**60.77**	**2.36**	**48.62**	**0.42**	**82.00**	**1.57**	**73.05**	**1.62**	**58.57**	**0.41**	**80.96**	**1.33**	**67.33**
CDFF only	**1.49**	58.55	2.84	48.20	0.53	79.42	1.91	71.19	1.73	53.18	0.57	79.17	1.51	64.95
VCST only	2.08	59.42	2.89	46.86	0.43	81.60	1.63	72.62	1.85	50.18	0.62	78.62	1.58	64.88

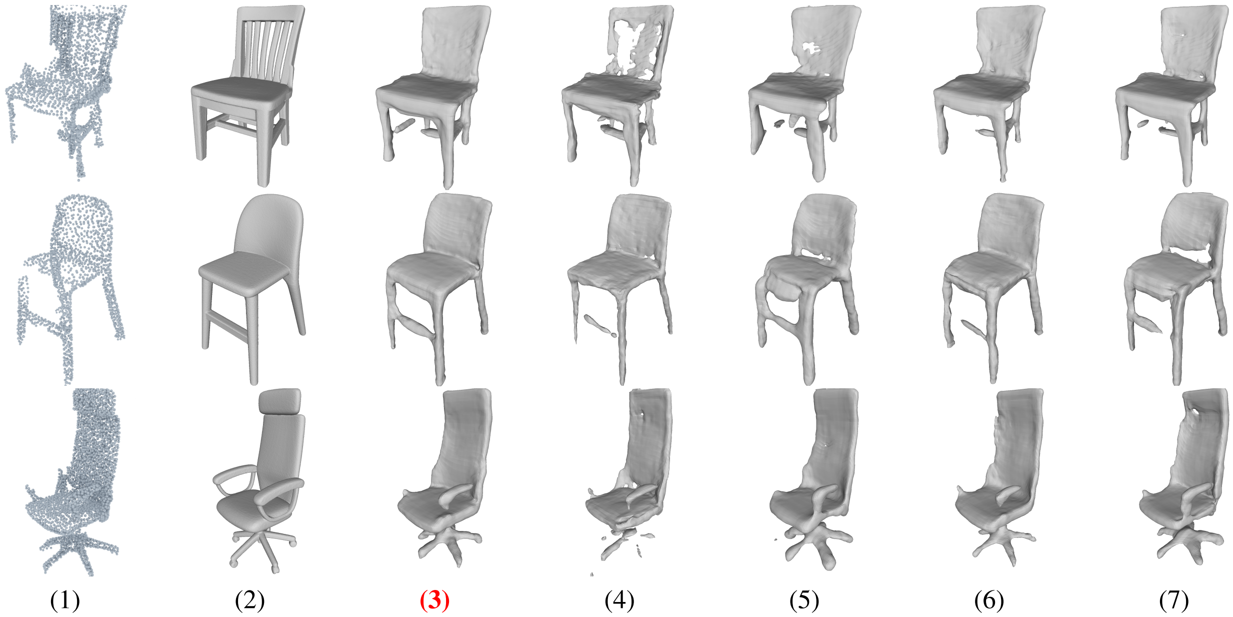

Figure S3. Visualization of the ablation experiments in Tab. 4 of the main paper. (1) Real Scan. (2) Ground Truth. **(3) Ours**. (4) $\mathbf{F} = \mathbf{F}_s$. (5) $\mathbf{F} = \mathbf{F}_t$. (6) Non-Adap. (7) Contrad.

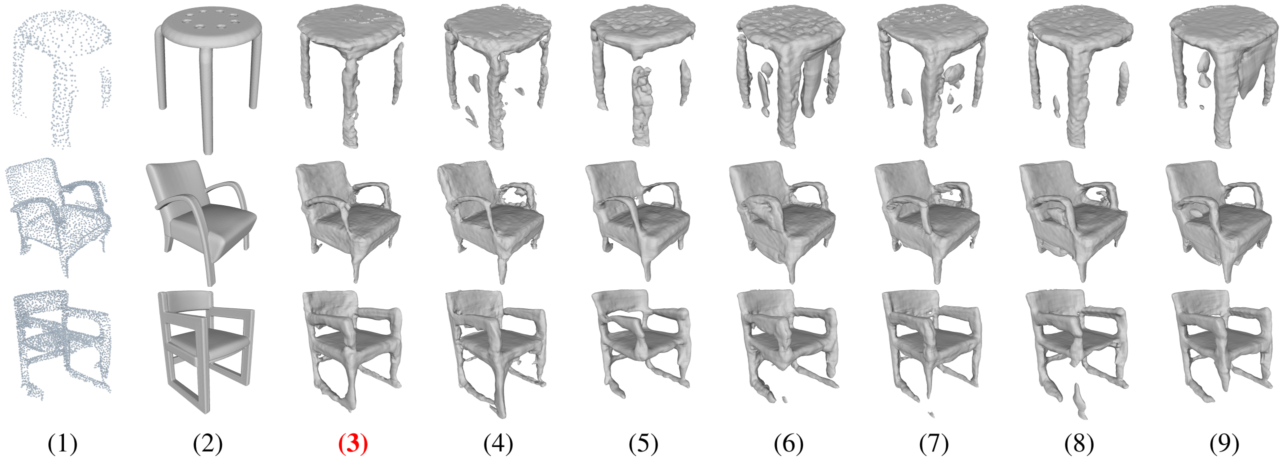

Figure S4. Visualization of the ablation experiments in Tab. 5 of the main paper. (1) Real Scan. (2) Ground Truth. **(3) Ours (Random & Surface)**. (4) w/o consistency training. (5) Random Only. (6) Volume Only. (7) Surface Only. (8) Random & Volume. (9) Random , Volume & Surface.

5. Failure Case Analysis

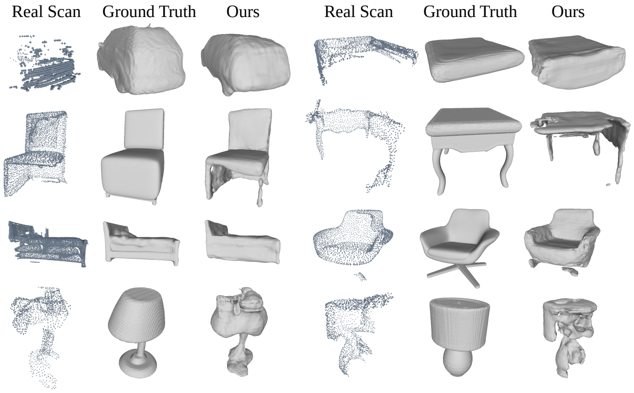

Figure S5. Three kinds of failure cases of reconstruction. The first two rows give 4 examples for case (i), and the third and the forth row gives 2 examples for case (ii) and (iii), respectively.

From abundant reconstruction results generated by our method, we conclude 3 kinds of cases that easily lead to failure, which are also shared by other baselines. The 3 kinds of cases are: (i) much incompleteness makes completion harder, which is also the most common reason that leads to poor reconstruction, *e.g.* the first two rows in Fig. S5; (ii) the distribution bias makes completion fail in some parts, *e.g.* the recovery failure of bed and chair legs in the third row of Fig. S5; (iii) the strong noise in the input scans misleads the reconstruction, *e.g.* the poor reconstruction quality of lamps in the last row of Fig. S5.

References

[1] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 1

[2] Xuelin Chen, Baoquan Chen, and Niloy J Mitra. Unpaired point cloud completion on real scans using adversarial training. In *ICLR*, 2019. 1

[3] Michael Gschwandtner, Roland Kwitt, Andreas Uhl, and Wolfgang Pree. Blensor: Blender sensor simulation toolbox. In *International Symposium on Visual Computing*, 2011. 1