

# SCoDA: Domain Adaptive Shape Completion for Real Scans

Yushuang Wu<sup>1,2,3</sup> Zizheng Yan<sup>1,2,3</sup> Ce Chen<sup>1,2</sup> Lai Wei<sup>4</sup> Xiao Li<sup>5</sup>  
Guanbin Li<sup>6,7</sup> Yihao Li<sup>1,2</sup> Shuguang Cui<sup>2,1</sup> Xiaoguang Han<sup>2,1#</sup>

<sup>1</sup>FNii, CUHKSZ <sup>2</sup>SSE, CUHKSZ <sup>3</sup>SRIBD <sup>4</sup>SDS, CUHKSZ <sup>5</sup>Microsoft Research Asia  
<sup>6</sup>Sun Yat-sen University <sup>7</sup>Research Institute, Sun Yat-sen University, Shenzhen

{yushuangwu, zizhengyan, cechen, laiweil}@link.cuhk.edu.cn luoheliyihao@mail.ustc.edu.cn  
{shuguangcui, hanxiaoguang}@cuhk.edu.cn liguanbin@mail.sysu.edu.cn li.xiao@microsoft.com

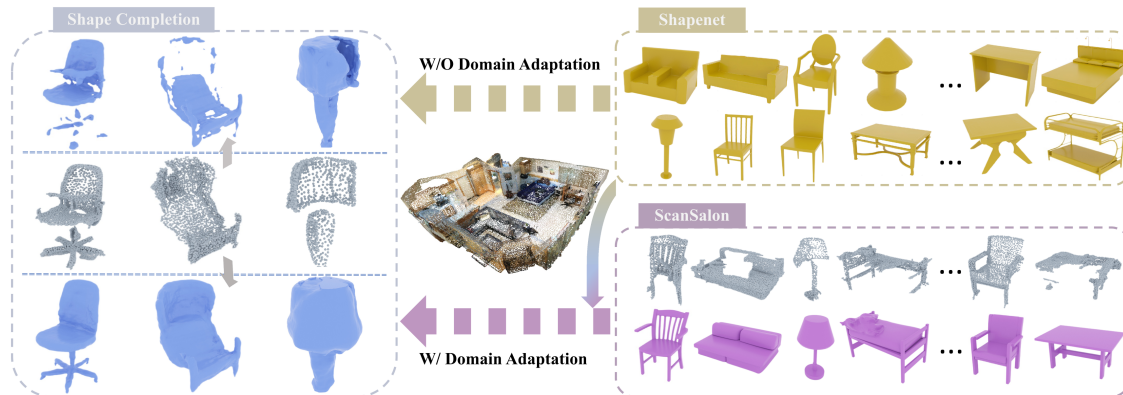


Figure 1. The proposed task **SCoDA** aims to transfer the knowledge in the synthetic domain to the reconstruction of noisy and incomplete real scans. A dataset, **ScanSalon**, with paired real scans and 3D models is contributed. Project page: [yushuang-wu.github.io/SCoDA](https://yushuang-wu.github.io/SCoDA).

## Abstract

3D shape completion from point clouds is a challenging task, especially from scans of real-world objects. Considering the paucity of 3D shape ground truths for real scans, existing works mainly focus on benchmarking this task on synthetic data, e.g. 3D computer-aided design models. However, the domain gap between synthetic and real data limits the generalizability of these methods. Thus, we propose a new task, **SCoDA**, for the domain adaptation of real scan shape completion from synthetic data. A new dataset, **ScanSalon**, is contributed with a bunch of elaborate 3D models created by skillful artists according to scans. To address this new task, we propose a novel cross-domain feature fusion method for knowledge transfer and a novel volume-consistent self-training framework for robust learning from real data. Extensive experiments prove our method is effective to bring an improvement of 6%~7% mIoU.

## 1. Introduction

Shape completion and reconstruction from scans is a practical 3D digitization task that is of great significance

in applications of virtual and augmented reality. It takes a scanned point cloud as input and aims to recover the 3D shape of the target object. The completion of real scans is challenging for the poor quality of point clouds and the deficiency of 3D shape ground truths. Existing methods exploit synthetic data, e.g. 3D computer-aided design (CAD) models to alleviate the demand for real object shapes. For example, authors of [18, 26, 35] simulate the scanning process to obtain point clouds from CAD models with paired ground truth shapes to train learning-based reconstruction models. However, there still exist distinctions between the simulated and real scan because the latter is scanned from a real object with complex scanning noise and occlusion, which limits the generalization quality.

Considering the underexploration in this field, we propose a new task, **SCoDA**, **D**omain **A**daptive **S**hape **C**ompletion, that aims to transfer the knowledge from the synthetic domain with rich clean shapes (source domain) into the shape completion of real scans (target domain), as illustrated in Fig. 1. To this end, we, for the first time, build an object-centric dataset, **ScanSalon**, which consists of real **Scans** with **Shape manual annotations**. Our ScanSalon contains a bunch of 3D models that are paired with real scans of objects in six categories: chair, desk/table, sofa, bed, lamp,

and car. The 3D models are manually created of high quality by skillful artists for around 10% of real scans, which can serve as the evaluation ground truths of shape completion or as the few labels for semi-supervised domain adaptation. See Fig. 1 for some examples, and details of **ScanSalon** are exposed in Sec. 4.

The main challenge of the proposed SCoDA task lies in the domain gap between synthetic and real point clouds. Due to the intrinsic complexity of the scanning process, *e.g.*, the scanner parameters and object materials, it is difficult to simulate the scans in terms of sparsity, noisy extent, *etc.* More importantly, real scans are usually incomplete resulting from the scene layout and object occlusion during scanning, which can hardly be simulated. Thus, we propose a novel domain adaptive shape completion approach to transfer the rich knowledge from the synthetic domain to the real one. At first, the reconstruction module in our approach is based on an implicit function for continuous shape completion [18, 54, 58]. For an effective transfer, we observe that although the local patterns of real scans (*e.g.*, noise, sparsity, and incompleteness) are distinct from the simulated synthetic ones, the global topology or structure in a same category is usually similar between the synthetic and real data, for example, a chair from either the synthetic or real domain usually consists of a seat, a backrest, legs, *etc.* (see Fig. 1). In other words, the global topology is more likely to be domain invariant, while the local patterns are more likely to be domain specific. Accordingly, we propose a cross-domain feature fusion (CDFS) module to combine the global features and local features learned in the synthetic and real domain, respectively, which helps recover both fine details and global structures in the implicit reconstruction stage. Moreover, a novel volume-consistent self-training (VCST) framework is developed to encourage self-supervised learning from the target data. Specifically, we create two views of real scans by dropping different clusters of points to produce incompleteness of different extent, and the model is forced to make consistent implicit predictions at each spatial volume, which encourages the model’s robustness to the incompleteness of real scans.

To construct a benchmark on the proposed new task, we implement some existing solutions to related tasks as baselines, and develop extensive experiments for the baselines and our method on **ScanSalon**. These experiments also demonstrate the effectiveness of the proposed method.

In summary, our key contributions are four-fold:

- We propose a new task, **SCoDA**, namely domain adaptive shape completion for real scans; A dataset **ScanSalon** that contains 800 elaborate 3D models paired with real scans in 6 categories is contributed.
- A novel cross-domain feature fusion module is designed to combine the knowledge of global shapes and local patterns learned in the synthetic and real domain,

respectively. Such a feature fusion manner may also inspire the works in the 2D domain adaption field.

- A volume-consistent self-training framework is proposed to improve the robustness of shape completion to the complex incompleteness of real scans.
- A benchmark with multiple methods evaluated is constructed for the task of **SCoDA** based on **ScanSalon**; Extensive experiments also demonstrate the superiority of the proposed method.

## 2. Related Work

**Shape Completion** Shape completion or reconstruction from point clouds is a branch of 3D reconstruction task [20, 47, 54, 68, 72, 91, 94]. The recovered shape can be represented as dense points [45, 46, 60] (*i.e.*, point cloud completion or consolidation), polygonal meshes [28, 35, 48, 78], manifold atlases [4, 25, 27, 32, 83], voxel grids [20, 34, 85], or implicit representations [18, 26, 40, 61, 70, 82]. Among them, the task of point cloud completion is related to our work, which aims to recover a complete, dense, and clean point cloud for the 3D shape, from a partial, sparse, and noisy input scanned by sensors. The pioneering work, PCN [99], proposes a coarse-to-fine completion framework with an encoder-decoder architecture, which encodes the partial points into global features and recovers the fined results in the decoding stage. The following works explore further to better exploit local or multi-scale features for higher-quality completion [38, 49, 57, 73, 76, 77, 80, 81, 89, 92, 95, 98, 99, 101]. In addition, the implicit representation based approaches now attract increasing attention because of its superior property that enables continuous shape recovery for objects with arbitrary topologies [17, 18, 54, 58]. Although numerous efforts have been contributed, the existing approaches are usually performed on point clouds scanned from synthetic data via simulation, and a straightforward application on real scans results in an unsatisfactory performance due to the domain gap between synthetic and real scans [18, 26, 64]. For the task of point completion, some works also adopt unsupervised or weakly-supervised learning techniques to learn from unlabeled data [8, 14, 33, 79, 86, 87, 100]. Another line of works retrieve suitable models from 3D CAD repositories and deform them as the estimated shape ground truths of object scans [3, 75]. Differently, our work first proposes the task of domain adaptive shape completion for real scans, which aims to sufficiently transfer the knowledge from the label-rich synthetic domain into real scan shape completion.

**2D Domain Adaptation** Domain adaptation has been a popular research topic in various 2D vision tasks, including image recognition [5, 11, 15, 22, 24, 29, 50, 67], semantic segmentation [16, 36, 52, 53], and object detection [12, 42, 55, 65]. Existing works can broadly be categorized as (i)

adversarial-based methods [12, 13, 15, 16, 22, 67] that train a discriminator to discriminate the source and the target domain, (ii) alignment-based methods [11, 24, 29, 65, 78] that use pre-defined metrics to align the source and the target domain, and (iii) self-training-based methods [5, 36, 52, 53, 55] that exploit the pseudo labels of the target domain to train the model. Recently, consistency-based self-training methods have become the most popular approach due to their impressive performance. One key factor in the success of consistency training is the augmentations applied to the inputs [31, 71, 93], *e.g.*, stronger augmentations have better performance in domain adaptive image recognition [5]. To this end, we design a novel consistency training framework specialized for the shape completion task. Furthermore, we propose to conduct knowledge transfer via a cross-domain feature fusion manner, which has not been explored by existing works, to our best knowledge.

**3D Domain Adaptation** In the 3D vision field, existing works on domain adaptation are mainly committed to the task of point cloud understanding, including classification, part segmentation, and semantic segmentation. For object-centric point clouds, PointDAN [63] constructs a domain adaptive classification benchmark by collecting point clouds from ShapeNet [10], ModelNet [88], and ScanNet [23] datasets to compose 3 different domains. It also proposes an adversarial representation learning approach with feature alignment at both local and global levels. The following works improve the classification precision by adopting techniques of consistency training [44], self-training [9, 103], multimodal learning [2], or learning from self-supervised tasks [1, 37, 51, 69]. Another line of works are committed to scene-centric point cloud analysis, *i.e.*, point cloud semantic segmentation [6, 7, 39, 41, 59, 66, 84, 90, 96, 102]. Considering the exhaustive annotation process for 3D point semantics, they propose to transfer the segmentation knowledge in synthetic scenes into real-world point clouds [41, 84, 90, 102]. Besides, there are a few attempts contributed to domain adaptive single-view reconstruction (SVR) [43, 62, 97], which is highly related to our work. Except for the difference in tasks (SVR *v.s.* shape completion), these works usually apply existing domain adaptation techniques to the target task, while we propose a novel domain adaptation method based on cross-domain feature fusion, which may also inspire others working on domain adaption.

### 3. Method

IF-Nets are a promising reconstruction method for their impressive performance [18], which is also successfully used in many applications [19, 56]. Constructed on IF-Nets, our approach first improves the representation learning by proposing a novel cross-domain feature fusion module, which aims to transfer the knowledge on the global-level object shape learned in the label-rich synthetic domain into

the real domain. Secondly, to exhaust the domain-specific information in the real data, a novel volume-consistent self-training method is proposed for the specific shape completion task, which also leads to a robust implicit prediction of IF-Nets. An illustration is presented in Fig. 2.

#### 3.1. Implicit Feature Networks

We first introduce IF-Nets [18] as our reconstruction framework. An IF-Net is composed of a 3D convolution neural network encoder  $g(\cdot)$  for multi-scale feature extraction and a multi-layer perceptron (MLP)  $f(\cdot)$  for implicit shape decoding.

Given a point cloud sample  $P$ , it is first converted into a discrete voxel representation  $\mathbf{X} \in \mathbb{R}^{N \times N \times N}$ , where  $N \in \mathbb{N}$  is the resolution of the input space.  $\mathbf{X}$  is then fed into the  $L$ -layer encoder  $g(\cdot)$  to generate a set of multi-scale features  $\{\mathbf{F}_1, \dots, \mathbf{F}_L\}$ , then they are upsampled to the same spatial dimension and been concatenated along the channel dimension to create the final feature  $\mathbf{F} = \text{concat}(\{\text{upsample}(\mathbf{F}_1), \dots, \text{upsample}(\mathbf{F}_L)\})$ . Formally:

$$g(\mathbf{X}) = \mathbf{F}, \quad \mathbf{X} \in \mathbb{R}^{N \times N \times N}, \quad \mathbf{F} \in \mathbb{R}^{d \times N \times N \times N},$$

where  $d$  is the feature channel number that equals the summation of the channel numbers of  $\mathbf{F}_i$ . Note that  $\mathbf{F}$  has a 3D structure aligned with the input  $\mathbf{X}$ . Given a point query  $\mathbf{p} \in \mathbb{R}^3$ , continuous features  $\mathbf{F}(\mathbf{p})$  at this point can be extracted from  $\mathbf{F}$  using trilinear interpolation.

Then, the encoding at the point  $\mathbf{p}$  is fed into a point-wise decoder  $f(\cdot)$  to give a binary prediction indicating if the point lies inside or outside the shape:

$$f(\mathbf{F}, \mathbf{p}) = f(\mathbf{F}(\mathbf{p})) \mapsto \{0, 1\}.$$

Given the occupancy value  $o(\mathbf{p}) \in \{0, 1\}$  at each position  $\mathbf{p}$  pre-computed according to the ground truth shape mesh, a binary cross-entropy loss is used to train  $g$  and  $f$ :

$$\min L_{IF} = \text{BCE}(f(\mathbf{F}, \mathbf{p}), o(\mathbf{p})).$$

Note that the prediction values can also be continuous signed distance values as in [58].

#### 3.2. Cross-Domain Feature Fusion

The training of a naive IF-Net demands ground truth shapes for generating numerous occupancy labels to supervise the output of  $f(\cdot)$ , yet most of the real scans are unpaired with shape ground truths. So we develop a cross-domain feature fusion module to well transfer the knowledge in the label-rich source domain. Our idea of exploiting the source knowledge stems from an important **observation**: although synthetic data have distinct local patterns from real scans, the global-level knowledge (*e.g.*, the common structure and coarse shape) of a certain category can

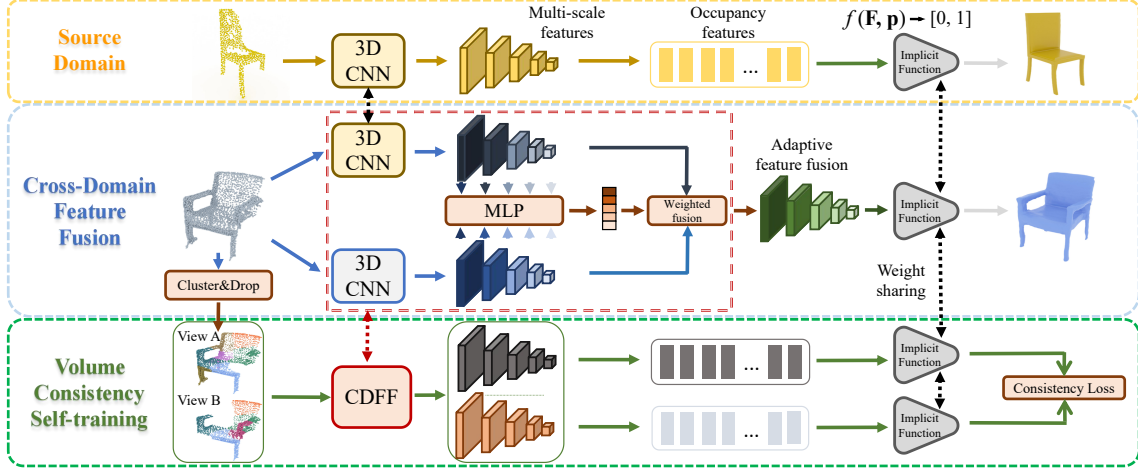


Figure 2. Overview of the proposed method. Two IF-Net encoders are used for the source and the target domain, respectively, and they share an implicit function decoder. The cross-domain feature fusion (CDFF) works by adaptively combining the global-level and local-level knowledge learned from the source and target domain, respectively. The volume-consistency self-training (VCST) works by enforcing the prediction consistency between two different augmented views to learn the local details.

be shared in both domains. Besides, the target data with few labels may be enough for learning rich local-level information. On the other hand, IF-Nets conduct shape completion via exploiting local representations, so it is important to extract high-quality local-level features without introducing bias from the source domain.

To this end, we develop a cross-domain feature fusion module (CDFF) for knowledge transfer. First, two shape encoders,  $g_s(\cdot)$  and  $g_t(\cdot)$ , are used for feature extraction of source and target data, generating  $\mathbf{F}_s$  and  $\mathbf{F}_t$ , respectively. As explained above,  $\mathbf{F}_s$  contains rich global information, and  $\mathbf{F}_t$  is more reliable in providing domain-specific local-level representations. A simple linear combination is adopted to fuse them into  $\mathbf{F}$ :

$$\mathbf{F} = \mathbf{w} \cdot \mathbf{F}_s + (1 - \mathbf{w}) \cdot \mathbf{F}_t,$$

where  $\mathbf{w} \in [0, 1]^d$  is a channel-wise weight vector. To better combine the advantages of  $\mathbf{F}_s$  and  $\mathbf{F}_t$ , the computation of  $\mathbf{w}$  takes two aspects into account: (i) exploiting the global-level features in  $\mathbf{F}_s$  while believing  $\mathbf{F}_t$  for local ones (based on our observation), and (ii) learning  $\mathbf{w}$  adaptively to maintain the flexibility. The adaptive weight vector  $\mathbf{w}$  is computed as:

$$\mathbf{w} = \alpha \cdot h(\mathbf{F}_s \odot \mathbf{F}_t) + \mathbf{w}^0,$$

where  $\alpha \in \mathbb{R}^+$  is the ratio of adaptiveness,  $h(\cdot) \mapsto [0, 1]^d$  is a two-layer MLP with a sigmoid activation at the output layer, and  $\odot$  indicates the operation of element-wise multiplication plus a global pooling on the spatial dimension (returns a vector of dimension  $d$ ).  $\mathbf{w}^0 \in [0, 1]^d$  is a constant weight vector that implies a prior derived from our observation, of which each value  $w_i^0$  is simply defined by a linear mapping as follows:

$$w_i^0 = \frac{l_i}{L+1}, i \in \{1, 2, \dots, d\},$$

where  $l_i \in \{1, 2, \dots, L\}$  indicates which layer the  $i$ -th channel belongs to. In this way, the deeper layer one channel is from, the more the computation of  $\mathbf{F}$  in this channel relies on  $\mathbf{F}_s$ , as the deeper layers of IF-Nets capture more global information. In addition, a clipping operation is applied to  $\mathbf{w}$  to limit all values to  $[0, 1]$ .

### 3.3. Volume-Consistent Self-training

With few shape ground truths in the target domain, it is hard to learn rich information in a supervised manner. A typical way is to adopt self-supervised learning to learn from the data itself. In the task of SCoDA, the **incompleteness** is a significant characteristic of real scans that hampers the shape completion quality. Thus, we create two views of scans with different levels of incompleteness, and encourage the model to make consistent implicit predictions on their volume occupancy. Such a volume-consistent self-training (VCST) encourages the model to keep robust to various incompleteness and “imagine” the missing parts for shape completion.

Specifically, we first adopt an unsupervised clustering algorithm (e.g.,  $k$ -means clustering) to partition the point cloud into  $K$  different parts (see Fig. 2 for examples). The clustering is based only on the spatial positions of points and thus coarsely splits an object scan into multiple components. With the pre-computed partitions, two views  $\mathbf{X}^A$  and  $\mathbf{X}^B$  can be generated by randomly dropping different clusters of points from the original real scan, with  $K_A$  and  $K_B$  clusters left, respectively, where  $K \geq K_A > K_B$ . Besides, two random downsamplings are also conducted on the two views to create different sparsity. Note that the clustering-based augmentation is actually a surface-aware augmentation. Compared with volume-based ones that randomly mask some spatial volumes, our augmentation strat-

egy implies more object-part knowledge to some extent.

Given the two views of input, their features are extracted by  $g(\cdot)$  and then taken by  $f(\cdot)$  to give implicit predictions:

$$g(\mathbf{X}^v) = \mathbf{F}^v, v = A, B,$$

$$f(\mathbf{F}^v, \mathbf{p}) \mapsto \{0, 1\}.$$

Note that  $\mathbf{F}^v$  can also be generated via our CDFD module. Here we simplify the formulation to better present our VCST. The consistency constraint is imposed on the implicit predictions of  $f(\mathbf{F}^A, \mathbf{p})$  and  $f(\mathbf{F}^B, \mathbf{p})$  with the same point query  $\mathbf{p}$ :

$$\min L_{CT} = \text{BCE}(f(\mathbf{F}^B, \mathbf{p}), f(\mathbf{F}^A, \mathbf{p})),$$

where  $\text{BCE}(\cdot, \cdot)$  indicates the binary cross-entropy loss. The consistency objective can be viewed as using the prediction from view  $A$  as the pseudo label of view  $B$ , because  $\mathbf{X}^B$  has a poorer completeness than  $\mathbf{X}^A$  with  $K_B < K_A$ . As there will be noises in the implicit predictions due to the limited shape ground truths of target data, using the noisy prediction as the pseudo label will mislead the training. Therefore, a thresholding operation is performed to mask the predictions with low confidence in  $f(\mathbf{F}^A, \mathbf{p})$ . The overall loss function is a summation of  $L_{IF}$  and  $L_{CT}$ .

## 4. Dataset

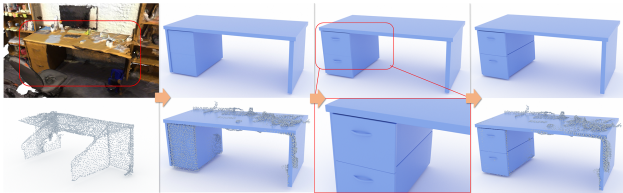


Figure 3. The procedure of 3D model crafting. From left to right: (i) extracting the object scan from a scene; (ii) constructing a coarse mesh frame to fit the scan; (iii) adding fine details; (iv) adjusting the defects detected by inspectors.

**Overview** The proposed ScanSalon dataset collects 7,644 real scans, of which 800 objects are equipped with 3D shape (artificial) ground truths. All models or scans are from 6 categories of objects, with 5 common indoor objects: chair, desk (or table), sofa, bed, and lamp, and 1 outdoor object: car. Detailed statistics are listed in Tab. 1. In Fig. 4, we exhibit some samples of real scans and corresponding 3D shapes in ScanSalon. More visualizations could be seen in the supplementary materials.

**Data Collection** The real scans are collected from two datasets, the ScanNet [23] and KITTI [30] dataset for indoor and outdoor objects, respectively. The object scans are extracted from a scene point cloud according to the point-level instance segmentation annotations provided by the two

Table 1. Statistics of the proposed dataset ScanSalon.

	Chair	Desk	Sofa	Bed	Lamp	Car	Total
Real Scans	4,651	1,630	428	365	133	437	7,644
Paired Models	497	161	43	36	20	43	800

datasets, and then rotated and normalized into the same pose and scale, with the aid of pose annotations provided by the Scan2CAD [3] and KITTI [30] dataset. These data are also aligned with models in the ShapeNet database [10], which provides a bunch of synthetic models as the source domain.

**Shape Annotation for Real Scans** Two skillful artists participate in the annotation procedure. In addition to the real scans, we provide several photos for each scan from different viewpoints (selected from the videos contained in the ScanNet and KITTI datasets), which provide 2D references to improve the model quality. A brief creation pipeline is shown in Fig. 3, where the artists are required to create the shapes according to both scans and photo references (see Fig. 4), overcome the distractions of scan noise, and recover the incompleteness by his(her) rich experience when the photo references are poor or inadequate. They use professional 3D software, Maya<sup>1</sup> to create models. On average, it takes 0.5~1.0 person-hours to create a 3D model and the creation of all 800 models takes 2.5 months in total. For each created model, another 8 inspectors are invited to verify their (i) recovery extent compared with the real objects in photos and (ii) matching extent with the given scan, and the defects will be fed back to artists for further adjustments until no defects are detected by any inspector. Considering the inaccessibility of the objects in scans from ScanNet and KITTI, we have made our greatest efforts to create **artificial ground truth** shapes for these scans.

## 5. Experiments

### 5.1. Settings

**Benchmark setting** We define two settings for the proposed task, which give only 3% and 5% shape ground truths (e.g. 47 and 140 chair models randomly sampled from all 4,651 chair scans, respectively) of real scans in the training set, and the rest of samples with corresponding 3D models compose the test set. Note that the 3% labeled samples are covered in the 5% labels for observing a cross-setting comparison when given more labels. For evaluation, we consider two metrics to measure the shape completion quality, mIoU: mean volumetric intersection over union (higher is better), and CD:  $l_2$  chamfer distance (lower is better). The former measures the matching extent of the defined volumes and the latter measures the surface accuracy.

**Baselines** We implement four baseline methods to construct the benchmark of the new task. (i) IF-Net: a naive IF-

<sup>1</sup><https://www.autodesk.com/products/maya>



Figure 4. ScanSalon data visualization. Each tuple includes: point cloud (left), created mesh (middle), and photo references (right).

Table 2. Experiment results on the 3% and 5% labels setting of the SCoDA task. The units of CD and mIoU value are  $1 \times 10^{-3}$  and %, respectively. **Red** text indicates the best and **blue** text indicates the second best result, respectively (similarly hereinafter).

(a) Results on the 3% labels setting.

Method	Chair		Desk		Sofa		Bed		Lamp		Car		Average	
	CD↓	mIoU↑	CD↓	mIoU↑	CD↓	mIoU↑	CD↓	mIoU↑	CD↓	mIoU↑	CD↓	mIoU↑	CD↓	mIoU↑
IF-Net	<b>1.57</b>	56.10	2.44	43.04	0.65	79.03	1.64	67.30	<b>1.67</b>	39.89	0.74	74.77	1.45	60.02
SelfSup	<b>1.49</b>	<b>58.55</b>	3.49	42.97	0.55	81.16	1.59	68.58	2.41	51.42	0.62	78.75	1.69	63.57
PtComp	1.61	57.33	<b>2.16</b>	44.26	0.51	79.90	<b>1.52</b>	68.23	1.95	46.97	<b>0.59</b>	<b>80.35</b>	<b>1.39</b>	62.84
Adversarial	1.74	58.54	2.99	<b>46.02</b>	<b>0.46</b>	<b>81.42</b>	<b>1.37</b>	<b>71.32</b>	2.43	<b>56.39</b>	0.67	78.91	1.61	<b>65.43</b>
<b>Ours</b>	1.58	<b>60.77</b>	<b>2.36</b>	<b>48.62</b>	<b>0.42</b>	<b>82.00</b>	1.57	<b>73.05</b>	<b>1.62</b>	<b>58.57</b>	<b>0.41</b>	<b>80.96</b>	<b>1.32</b>	<b>67.32</b>

(b) Results on the 5% labels setting.

Method	Chair		Desk		Sofa		Bed		Lamp		Car		Average	
	CD↓	mIoU↑	CD↓	mIoU↑	CD↓	mIoU↑	CD↓	mIoU↑	CD↓	mIoU↑	CD↓	mIoU↑	CD↓	mIoU↑
IF-Net	1.88	56.98	2.14	44.87	0.50	82.04	0.66	76.05	1.72	51.33	0.52	80.13	1.24	65.23
SelfSup	2.08	59.42	2.73	46.39	0.51	82.25	0.61	77.22	1.46	62.02	0.43	81.89	1.09	68.06
PtComp	<b>1.34</b>	57.98	<b>1.83</b>	46.20	<b>0.32</b>	82.66	0.61	79.07	<b>1.44</b>	61.61	<b>0.43</b>	<b>81.89</b>	<b>1.00</b>	68.24
Adversarial	1.71	<b>60.58</b>	2.13	<b>48.46</b>	0.41	<b>83.54</b>	<b>0.51</b>	<b>80.81</b>	<b>1.33</b>	<b>64.22</b>	<b>0.41</b>	81.86	1.08	<b>69.91</b>
<b>Ours</b>	<b>1.37</b>	<b>61.48</b>	<b>2.09</b>	<b>50.93</b>	<b>0.31</b>	<b>82.71</b>	<b>0.41</b>	<b>82.27</b>	1.57	<b>67.80</b>	0.46	<b>83.12</b>	<b>1.04</b>	<b>71.39</b>

Net [18] is adopted without any other designs and is trained on all supervised samples (including synthetic and real labels); (ii) SelfSup: a self-supervised IF-Net that conducts feature-level consistency (by minimizing the distance between two domain feature vectors) on unsupervised samples [50]; (iii) PtComp: a domain adaptive point completion method [14] is re-implemented to align the real scan inputs with the synthetic ones, incorporated with an extra IF-Net for implicit shape reconstruction. Note the encoder and decoder used in [14] are revised into a voxel-based 3D CNN to incorporate with IF-Net; (iv) Adversarial: a domain adaptive IF-Net is exploited with adversarial feature learning [43, 74, 87] to learn domain-invariant features. The de-

tailed implementation details of each baseline are included in the supplementary materials for the space limit.

**Implementation details** We adopt a 6-layer 3DCNN as the encoder with a 4-layer MLP as the decoder to implement the IF-Nets, using the official codes released by [18]. Following [18], the channel dimension of multi-scale features is 2,583, and a resolution  $N=32$  is used for the training of reconstruction. For evaluation, we use a higher resolution with  $N=128$ . In the CDF module,  $\alpha$  is set to 0.2 to better combine the advantages of  $F_s$  and  $F_t$ . For the augmentation strategy in VCST,  $K_A$  is randomly selected from  $\{7, 8\}$ , and  $K_B$  is randomly selected from  $\{5, 6, 7\}$  for each data sample. The downsampling randomly sam-

Table 3. Ablation on using the single module of CDFD or VCST.

	Chair		Bed		Lamp	
	CD↓	mIoU↑	CD↓	mIoU↑	CD↓	mIoU↑
CDFD+VCST	1.58	<b>60.77</b>	<b>1.57</b>	<b>73.05</b>	<b>1.62</b>	<b>58.57</b>
CDFD only	<b>1.49</b>	58.55	1.91	71.19	1.73	53.18
VCST only	2.08	59.42	1.63	72.62	1.85	50.18

Table 4. Analysis of the design of the CDFD module.

	Fusion	Observ.	Adaptive	CD↓	mIoU↑
	<b>Ours</b>	✓	✓	✓	<b>1.49</b>
$\mathbf{F} = \mathbf{F}_s$				1.57 (+0.08)	56.10 (-2.45)
$\mathbf{F} = \mathbf{F}_t$				2.22 (+0.73)	54.98 (-3.57)
Non-Adap.	✓	✓		1.92 (+0.43)	57.77 (-0.78)
Contrad.	✓		✓	1.96 (+0.47)	56.62 (-1.93)

points 1k points from each scan (no sampling when has <1k points). A thresholding operation is adopted to filter out predictions with a confidence in (0.1, 0.9) (viewed as unconfident predictions) from view  $A$  to guarantee the quality of consistency training [71]. All models are trained with an Adam optimizer and a learning rate  $1 \times 10^{-4}$ , with a mini-batch size 4.  $L_{CT}$  only works on unsupervised data. More implementation details of our method are introduced in the supplementary materials.

## 5.2. Results

**Quantitative comparison** The performance of baseline methods and ours are listed in Tab. 2 on the 3% and 5% labels settings. It can be seen that our method is superior to all baselines. On the 3% labels setting, our method achieves  $1.32 \times 10^{-3}$  on CD and 67.32% on mIoU on average, which is higher than the second best “Adversarial” by  $\sim 0.3 \times 10^{-3}$  CD and  $\sim 2\%$  mIoU. Besides, our method achieves good performances consistently over all 6 categories, most of which get a top rank over all 5 methods. The adversarial domain adaption method achieves the top mIoU among all baseline methods. Besides, our method also achieves the best performance on the 5% setting for both average CD and mIoU. In addition, it can be observed that an extra 2% labels bring an improvement of around 5% mIoU to all methods, and the improvements are especially significant for the categories “bed” and “lamp”, which shows the promising practical value to provide a few more real-data labels to boost the shape completion quality from real scans.

**Qualitative comparison** We visualize the shape completion results of different methods in Fig. 5 to qualitatively compare baselines and our method. With the implicit prediction at each grid, a threshold of 0.5 is used to justify whether an occupancy is predicted here. The resulting occupancy grids with resolution 128 are then transformed into a mesh using the marching cubes algorithm [21]. The visualizations present the superiority of the proposed method, which achieves better completion for missing parts and generates fewer defects almost in all samples of Fig. 5.

Table 5. Analysis of the design of the VCST module.

Random	Volume	Surface	CD↓	mIoU↑
w/o consistency training			<b>1.57</b>	56.10
✓			3.65 (+2.08)	47.60 (-8.50)
	✓		2.03 (+0.46)	57.55 (+1.45)
		✓	2.37 (+0.80)	59.03 (+2.93)
✓		✓	2.08 (+0.51)	<b>59.42</b> (+3.32)
✓	✓		2.37 (+0.80)	58.13 (+2.03)
✓	✓	✓	2.30 (+0.73)	58.55 (+2.45)

## 5.3. Ablation

**Single module** We first conduct ablative studies on using the CDFD or VCST module only. The results on 3 categories of the 3% label setting confirm the effectiveness of either of them, see Tab. 3. Besides, integrating both two methods brings an extra performance gain, which shows the complementary functions of these two modules.

**CDFD variants** Next, we implement ablations on different variants of the proposed two modules, and all experiments are conducted on the “chair” category under the 3% label setting. For the CDFD module, we experiment such variants: (i) using the whole features generated by  $g_s$  rather than via feature fusion ( $\mathbf{F} = \mathbf{F}_s$ ); (ii) without fusion but using the whole features generated by  $g_t$  ( $\mathbf{F} = \mathbf{F}_t$ ); (iii) conducting feature fusion but with a non-adaptive manner, specifically,  $w_i$  is set to 0 for the bottom 3 layers and 1 for the top 3 layers (**Non-Adap.**); (iv) conducting adaptive feature fusion but in a manner contradictory to our observation (**Contradict.**), where we compute  $w_0$  via  $w_i^0 = 1 - l_i / (L + 1)$  instead. The results in Tab. 4 show that (i) fully using  $\mathbf{F}_s$  as  $\mathbf{F}$  is equivalent to using a naive IF-Net that has a lower performance; (ii) using  $\mathbf{F}_t$  as  $\mathbf{F}$  results in a significant performance drop because  $g_t$  can hardly learn high-quality global representations with a few labels; (iii) using a non-adaptive feature fusion that complies with our observation also leads to a performance drop but performs better than those without fusion; (iv) a fusion contradictory to our observation brings a  $\sim 2\%$  mIoU drop, which further supports the reasonableness of our observation.

**VCST variants** Related to our VCST design, we try different combinations to achieve the best augmentation strategy. Alternative strategies include random downsampling, surface-aware sampling (clustering-based), volume-aware sampling, and adding noise. Among them, adding noise has no effect on the performance whatever other augmentations are used, which may be caused by the voxelization operation in IF-Nets. Thus we try the following variants with the combination of the other three augmentation strategies. Note that we control the dropping ratio to guarantee the expected number of dropped points (if only using dropping) are the same for all settings. The results in Tab. 5 show that: (1) the consistency training with random downsampling only crashed (the 2nd row), using volume-

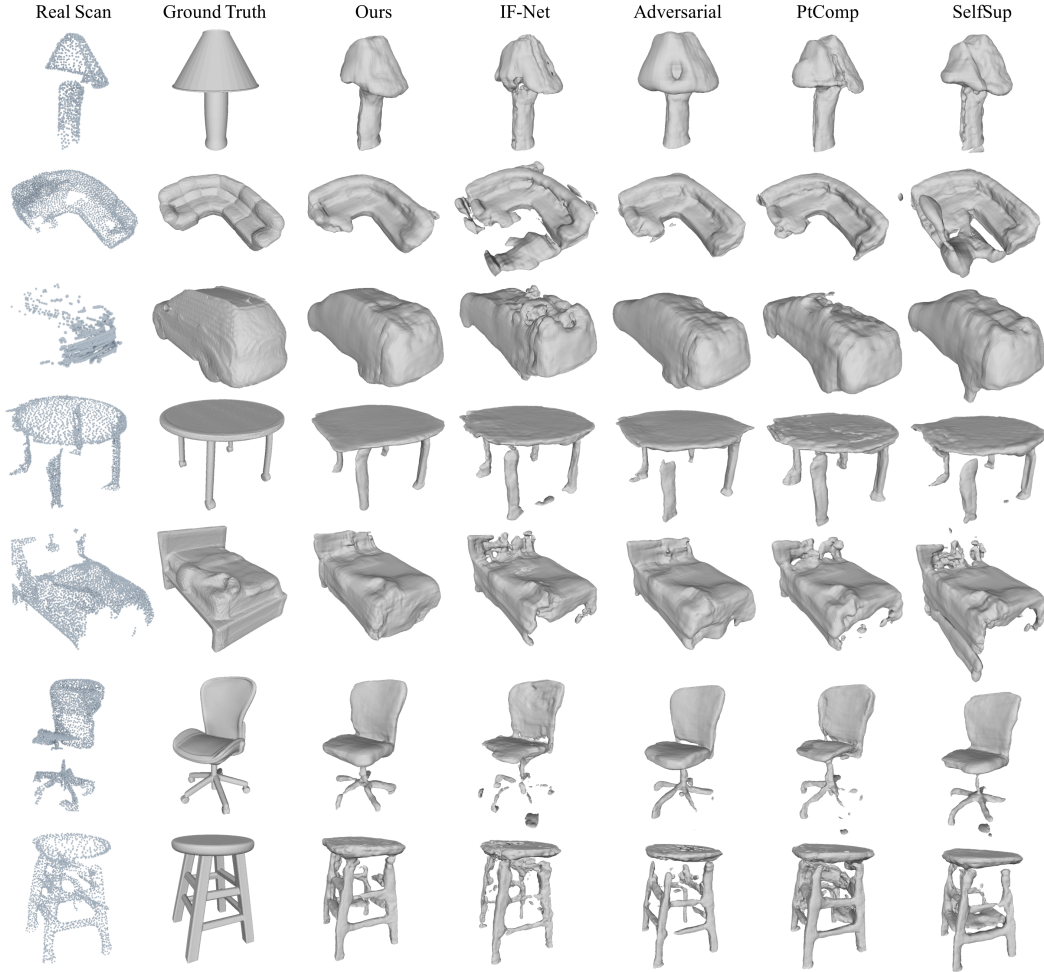


Figure 5. Qualitative comparison between different methods on shape completion with only 3% labels for training.

aware or surface-aware augmentation only can still work, but the surface-aware one performs better (the 3rd and 4th row), and the 4th row is the version used in our standard VCST; (3) the volume-aware augmentation can bring minor improvements on mIoU with or without random downsampling (the 3rd and 5th row), which is worse than our surface-aware one, and using both two of them even makes a negative effect (the last row). The reason for the failure of using random downsampling only is that the consistency training fails when two views are very similar in a global fashion. An intuitive explanation for (3) is that compared with our surface-aware augmentation, the incompleteness created by the volume-aware one has a larger gap with real scans. The qualitative comparisons between different variants are included in the supplementary materials.

## 6. Conclusion

In this paper, we propose a new task, domain adaptive shape completion of real scans, which aims to transfer the knowledge in the label-rich synthetic domain into the more challenging real domain. We construct a new dataset that

contains elaborate 3D models created by professional artists for 10% of real scans in 6 categories, which can serve for the learning of deep reconstruction models and evaluation. Besides, we develop a new domain adaptive shape completion framework with two novel modules, cross-domain feature fusion and volume-consistent self-training, to better exploit both the synthetic and real data. Extensive experiments validate the superiority and effectiveness of our method and a benchmark is constructed on the proposed task.

**Acknowledge** The work was supported in part by NSFC-62172348, the Basic Research Project No. HZQB-KCZYZ-2021067 of Hetao Shenzhen-HK S&T Cooperation Zone. It was also partially supported by NSFC-62172348, Outstanding Young Fund of Guangdong Province with No. 2023B1515020055, the National Key R&D Program of China with grant No. 2018YFB1800800, by Shenzhen Outstanding Talents Training Fund 202002, by Guangdong Research Projects No. 2017ZT07X152 and No. 2019CX01X104, by the Guangdong Provincial Key Laboratory of Future Networks of Intelligence (Grant No. 2022B1212010001), and by Shenzhen Key Laboratory of Big Data and Artificial Intelligence (Grant No. ZDSYS201707251409055).



## References

- [1] Idan Achituve, Haggai Maron, and Gal Chechik. Self-supervised learning for domain adaptation on point clouds. In *WACV*, 2021. 3
- [2] Mohamed Afham, Isuru Dissanayake, Dinithi Dissanayake, Amaya Dharmasiri, Kanchana Thilakarathna, and Ranga Rodrigo. Crosspoint: Self-supervised cross-modal contrastive learning for 3d point cloud understanding. In *CVPR*, 2022. 3
- [3] Armen Avetisyan, Manuel Dahnert, Angela Dai, Manolis Savva, Angel X Chang, and Matthias Nießner. Scan2cad: Learning cad model alignment in rgb-d scans. In *CVPR*, 2019. 2, 5
- [4] Abhishek Badki, Orazio Gallo, Jan Kautz, and Pradeep Sen. Meshlet priors for 3d mesh reconstruction. In *CVPR*, 2020. 2
- [5] David Berthelot, Rebecca Roelofs, Kihyuk Sohn, Nicholas Carlini, and Alex Kurakin. Adamatch: A unified approach to semi-supervised learning and domain adaptation. In *ICLR*, 2022. 2, 3
- [6] Yikai Bian, Le Hui, Jianjun Qian, and Jin Xie. Unsupervised domain adaptation for point cloud semantic segmentation via graph matching. *arXiv preprint arXiv:2208.04510*, 2022. 3
- [7] Yikai Bian, Jin Xie, and Jianjun Qian. Unsupervised domain adaptive point cloud semantic segmentation. In *ACPR*, 2022. 3
- [8] Yingjie Cai, Kwan-Yee Lin, Chao Zhang, Qiang Wang, Xiaogang Wang, and Hongsheng Li. Learning a structured latent space for unsupervised point cloud completion. In *CVPR*, 2022. 2
- [9] Adriano Cardace, Riccardo Spezialetti, Pierluigi Zama Ramirez, Samuele Salti, and Luigi Di Stefano. Refrec: Pseudo-labels refinement via shape reconstruction for unsupervised 3d domain adaptation. In *3DV*, 2021. 3
- [10] 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. 3, 5
- [11] Chao Chen, Zhihang Fu, Zhihong Chen, Sheng Jin, Zhaowei Cheng, Xinyu Jin, and Xian-Sheng Hua. Homm: Higher-order moment matching for unsupervised domain adaptation. In *AAAI*, 2020. 2, 3
- [12] Chaoqi Chen, Zebiao Zheng, Xinghao Ding, Yue Huang, and Qi Dou. Harmonizing transferability and discriminability for adapting object detectors. In *CVPR*, 2020. 2, 3
- [13] Lin Chen, Huaian Chen, Zhixiang Wei, Xin Jin, Xiao Tan, Yi Jin, and Enhong Chen. Reusing the task-specific classifier as a discriminator: Discriminator-free adversarial domain adaptation. In *CVPR*, 2022. 3
- [14] Xuelin Chen, Baoquan Chen, and Niloy J Mitra. Unpaired point cloud completion on real scans using adversarial training. In *ICLR*, 2019. 2, 6
- [15] Xinyang Chen, Sinan Wang, Mingsheng Long, and Jianmin Wang. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. In *ICML*, 2019. 2, 3
- [16] Yun-Chun Chen, Yen-Yu Lin, Ming-Hsuan Yang, and Jia-Bin Huang. Crdoco: Pixel-level domain transfer with cross-domain consistency. In *CVPR*, 2019. 2, 3
- [17] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *CVPR*, 2019. 2
- [18] Julian Chibane, Thiemo Alldieck, and Gerard Pons-Moll. Implicit functions in feature space for 3d shape reconstruction and completion. In *CVPR*, 2020. 1, 2, 3, 6
- [19] Julian Chibane and Gerard Pons-Moll. Implicit feature networks for texture completion from partial 3d data. In *ECCV*, 2020. 3
- [20] Christopher B Choy, Danfei Xu, JunYoung Gwak, Kevin Chen, and Silvio Savarese. 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction. In *ECCV*, 2016. 2
- [21] Marching Cubes. A high resolution 3d surface construction algorithm. In *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 1987. 7
- [22] Shuhao Cui, Shuhui Wang, Junbao Zhuo, Chi Su, Qingming Huang, and Qi Tian. Gradually vanishing bridge for adversarial domain adaptation. In *CVPR*, 2020. 2, 3
- [23] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *CVPR*, 2017. 3, 5
- [24] Bharath Bhushan Damodaran, Benjamin Kellenberger, Rémi Flamary, Devis Tuia, and Nicolas Courty. Deepjdot: Deep joint distribution optimal transport for unsupervised domain adaptation. In *ECCV*, 2018. 2, 3
- [25] Theo Deprelle, Thibault Groueix, Matthew Fisher, Vladimir Kim, Bryan Russell, and Mathieu Aubry. Learning elementary structures for 3d shape generation and matching. *NeurIPS*, 2019. 2
- [26] Philipp Erler, Paul Guerrero, Stefan Ohrhallinger, Niloy J Mitra, and Michael Wimmer. Points2surf learning implicit surfaces from point clouds. In *ECCV*, 2020. 1, 2
- [27] Matheus Gadelha, Rui Wang, and Subhransu Maji. Deep manifold prior. In *ICCV*, 2021. 2
- [28] Jun Gao, Wenzheng Chen, Tommy Xiang, Alec Jacobson, Morgan McGuire, and Sanja Fidler. Learning deformable tetrahedral meshes for 3d reconstruction. *NeurIPS*, 2020. 2
- [29] Zhiqiang Gao, Shufei Zhang, Kaizhu Huang, Qiufeng Wang, and Chaoliang Zhong. Gradient distribution alignment certificates better adversarial domain adaptation. In *ICCV*, 2021. 2, 3
- [30] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *CVPR*, 2012. 5
- [31] Atin Ghosh and Alexandre H Thiery. On data-augmentation and consistency-based semi-supervised learning. In *ICLR*, 2020. 3
- [32] Thibault Groueix, Matthew Fisher, Vladimir G Kim, Bryan C Russell, and Mathieu Aubry. A papier-mâché approach to learning 3d surface generation. In *CVPR*, 2018. 2

- [33] Jiayuan Gu, Wei-Chiu Ma, Sivabalan Manivasagam, Wenyuan Zeng, Zihao Wang, Yuwen Xiong, Hao Su, and Raquel Urtasun. Weakly-supervised 3d shape completion in the wild. In *ECCV*, 2020. 2
- [34] Christian Häne, Shubham Tulsiani, and Jitendra Malik. Hierarchical surface prediction for 3d object reconstruction. In *3DV*, 2017. 2
- [35] Rana Hanocka, Gal Metzer, Raja Giryes, and Daniel Cohen-Or. Point2mesh: A self-prior for deformable meshes. *arXiv preprint arXiv:2005.11084*, 2020. 1, 2
- [36] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *CVPR*, 2022. 2, 3
- [37] Junxuan Huang, Junsong Yuan, and Chunming Qiao. Generation for unsupervised domain adaptation: A gan-based approach for object classification with 3d point cloud data. In *ICASSP*, 2022. 3
- [38] Zitian Huang, Yikuan Yu, Jiawen Xu, Feng Ni, and Xinyi Le. Pf-net: Point fractal network for 3d point cloud completion. In *CVPR*, 2020. 2
- [39] Maximilian Jaritz, Tuan-Hung Vu, Raoul de Charette, Emilie Wirbel, and Patrick Pérez. xmuda: Cross-modal unsupervised domain adaptation for 3d semantic segmentation. In *CVPR*, 2020. 3
- [40] Chiyu Jiang, Avneesh Sud, Ameesh Makadia, Jingwei Huang, Matthias Nießner, Thomas Funkhouser, et al. Local implicit grid representations for 3d scenes. In *CVPR*, 2020. 2
- [41] Peng Jiang and Srikanth Saripalli. Lidarnet: A boundary-aware domain adaptation model for point cloud semantic segmentation. In *ICRA*, 2021. 3
- [42] Mehran Khodabandeh, Arash Vahdat, Mani Ranjbar, and William G Macready. A robust learning approach to domain adaptive object detection. In *ICCV*, 2019. 2
- [43] Brandon Leung, Siddharth Singh, and Arik Horodniceanu. Domain adaptation for real-world single view 3d reconstruction. *arXiv preprint arXiv:2108.10972*, 2021. 3, 6
- [44] Qing Li, Xiaojiang Peng, and Qi Hao. Self-ensembling for 3d point cloud domain adaptation. *arXiv preprint arXiv:2112.05301*, 2021. 3
- [45] Ruihui Li, Xianzhi Li, Ka-Hei Hui, and Chi-Wing Fu. Spgan: Sphere-guided 3d shape generation and manipulation. *TOG*, 2021. 2
- [46] Chen-Hsuan Lin, Chen Kong, and Simon Lucey. Learning efficient point cloud generation for dense 3d object reconstruction. In *AAAI*, 2018. 2
- [47] Chen-Hsuan Lin, Chaoyang Wang, and Simon Lucey. Sdf-sm: Learning signed distance 3d object reconstruction from static images. *NeurIPS*, 2020. 2
- [48] Or Litany, Alex Bronstein, Michael Bronstein, and Ameesh Makadia. Deformable shape completion with graph convolutional autoencoders. In *CVPR*, 2018. 2
- [49] Minghua Liu, Lu Sheng, Sheng Yang, Jing Shao, and Shi-Min Hu. Morphing and sampling network for dense point cloud completion. In *AAAI*, 2020. 2
- [50] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Deep transfer learning with joint adaptation networks. In *ICML*, 2017. 2, 6
- [51] Xiaoyuan Luo, Shaolei Liu, Kexue Fu, Manning Wang, and Zhijian Song. A learnable self-supervised task for unsupervised domain adaptation on point clouds. *arXiv preprint arXiv:2104.05164*, 2021. 3
- [52] Ke Mei, Chuang Zhu, Jiaqi Zou, and Shanghang Zhang. Instance adaptive self-training for unsupervised domain adaptation. In *ECCV*, 2020. 2, 3
- [53] Luke Melas-Kyriazi and Arjun K Manrai. Pixmatch: Unsupervised domain adaptation via pixelwise consistency training. In *CVPR*, 2021. 2, 3
- [54] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *CVPR*, 2019. 2
- [55] Muhammad Akhtar Munir, Muhammad Haris Khan, M Sarfraz, and Mohsen Ali. Ssal: Synergizing between self-training and adversarial learning for domain adaptive object detection. *NeurIPS*, 2021. 2, 3
- [56] Amine Ouasfi and Adnane Boukhayma. Few-zero level set-shot learning of shape signed distance functions in feature space. *arXiv preprint arXiv:2207.04161*, 2022. 3
- [57] Liang Pan, Xinyi Chen, Zhongang Cai, Junzhe Zhang, Haiyu Zhao, Shuai Yi, and Ziwei Liu. Variational relational point completion network. In *CVPR*, 2021. 2
- [58] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: Learning continuous signed distance functions for shape representation. In *CVPR*, 2019. 2, 3
- [59] Duo Peng, Yinjie Lei, Wen Li, Pingping Zhang, and Yulan Guo. Sparse-to-dense feature matching: Intra and inter domain cross-modal learning in domain adaptation for 3d semantic segmentation. In *CVPR*, 2021. 3
- [60] Songyou Peng, Chiyu Jiang, Yiyi Liao, Michael Niemeyer, Marc Pollefeys, and Andreas Geiger. Shape as points: A differentiable poisson solver. *NeurIPS*, 2021. 2
- [61] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *ECCV*, 2020. 2
- [62] Pedro O Pinheiro, Negar Rostamzadeh, and Sungjin Ahn. Domain-adaptive single-view 3d reconstruction. In *ICCV*, 2019. 3
- [63] Can Qin, Haoxuan You, Lichen Wang, C-C Jay Kuo, and Yun Fu. Pointdan: A multi-scale 3d domain adaptation network for point cloud representation. *NeurIPS*, 2019. 3
- [64] Sulzer Raphael, Landrieu Loïc, Boulch Alexandre, Marlet Renaud, and Vallet Bruno. Deep surface reconstruction from point clouds with visibility information. *arXiv preprint arXiv:2203.09440*, 2020. 2
- [65] Farzaneh Rezaeianaran, Rakshith Shetty, Rahaf Aljundi, Daniel Olmeda Reino, Shanshan Zhang, and Bernt Schiele. Seeking similarities over differences: Similarity-based domain alignment for adaptive object detection. In *ICCV*, 2021. 2, 3

- [66] Mrigank Roachan, Shubhra Aich, Eduardo R Corral-Soto, Amir Nabatchian, and Bingbing Liu. Unsupervised domain adaptation in lidar semantic segmentation with self-supervision and gated adapters. In *ICRA*, 2022. 3
- [67] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *CVPR*, 2018. 2, 3
- [68] Shunsuke Saito, Zeng Huang, Ryota Natsume, Shigeo Morishima, Angjoo Kanazawa, and Hao Li. Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In *ICCV*, 2019. 2
- [69] Yuefan Shen, Yanchao Yang, Mi Yan, He Wang, Youyi Zheng, and Leonidas J Guibas. Domain adaptation on point clouds via geometry-aware implicits. In *CVPR*, 2022. 3
- [70] Vincent Sitzmann, Eric Chan, Richard Tucker, Noah Snavely, and Gordon Wetzstein. Metasdf: Meta-learning signed distance functions. *NeurIPS*, 2020. 2
- [71] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *NeurIPS*, 2020. 3, 7
- [72] Maxim Tatarchenko, Stephan R Richter, René Ranftl, Zhuwen Li, Vladlen Koltun, and Thomas Brox. What do single-view 3d reconstruction networks learn? In *CVPR*, 2019. 2
- [73] Lyne P Tchammi, Vineet Kosaraju, Hamid Rezatofighi, Ian Reid, and Silvio Savarese. Topnet: Structural point cloud decoder. In *CVPR*, 2019. 2
- [74] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *CVPR*, 2017. 6
- [75] Mikaela Angelina Uy, Vladimir G Kim, Minhyuk Sung, Noam Aigerman, Siddhartha Chaudhuri, and Leonidas J Guibas. Joint learning of 3d shape retrieval and deformation. In *CVPR*, 2021. 2
- [76] Xiaogang Wang, Marcelo H Ang Jr, and Gim Hee Lee. Cascaded refinement network for point cloud completion. In *CVPR*, 2020. 2
- [77] Yida Wang, David Joseph Tan, Nassir Navab, and Federico Tombari. Softpoolnet: Shape descriptor for point cloud completion and classification. In *ECCV*, 2020. 2
- [78] Xingkui Wei, Zhengqing Chen, Yanwei Fu, Zhaopeng Cui, and Yinda Zhang. Deep hybrid self-prior for full 3d mesh generation. In *ICCV*, 2021. 2, 3
- [79] Xin Wen, Zhizhong Han, Yan-Pei Cao, Pengfei Wan, Wen Zheng, and Yu-Shen Liu. Cycle4completion: Unpaired point cloud completion using cycle transformation with missing region coding. In *CVPR*, 2021. 2
- [80] Xin Wen, Tianyang Li, Zhizhong Han, and Yu-Shen Liu. Point cloud completion by skip-attention network with hierarchical folding. In *CVPR*, 2020. 2
- [81] Xin Wen, Peng Xiang, Zhizhong Han, Yan-Pei Cao, Pengfei Wan, Wen Zheng, and Yu-Shen Liu. Pmp-net: Point cloud completion by learning multi-step point moving paths. In *CVPR*, 2021. 2
- [82] Francis Williams, Zan Gojcic, Sameh Khamis, Denis Zorin, Joan Bruna, Sanja Fidler, and Or Litany. Neural fields as learnable kernels for 3d reconstruction. In *CVPR*, 2022. 2
- [83] Francis Williams, Teseo Schneider, Claudio Silva, Denis Zorin, Joan Bruna, and Daniele Panozzo. Deep geometric prior for surface reconstruction. In *CVPR*, 2019. 2
- [84] Bichen Wu, Xuanyu Zhou, Sicheng Zhao, Xiangyu Yue, and Kurt Keutzer. Squeezesegv2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud. In *ICRA*, 2019. 3
- [85] Jiajun Wu, Yifan Wang, Tianfan Xue, Xingyuan Sun, Bill Freeman, and Josh Tenenbaum. Marnet: 3d shape reconstruction via 2.5 d sketches. *NeurIPS*, 2017. 2
- [86] Mengxi Wu, Hao Huang, and Yi Fang. 3d point cloud completion with geometric-aware adversarial augmentation. *arXiv preprint arXiv:2109.10161*, 2021. 2
- [87] Rundi Wu, Xuelin Chen, Yixin Zhuang, and Baoquan Chen. Multimodal shape completion via conditional generative adversarial networks. In *ECCV*, 2020. 2, 6
- [88] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *CVPR*, 2015. 3
- [89] Peng Xiang, Xin Wen, Yu-Shen Liu, Yan-Pei Cao, Pengfei Wan, Wen Zheng, and Zhizhong Han. Snowflakenet: Point cloud completion by snowflake point deconvolution with skip-transformer. In *ICCV*, 2021. 2
- [90] Aoran Xiao, Jiaxing Huang, Dayan Guan, Fangneng Zhan, and Shijian Lu. Transfer learning from synthetic to real lidar point cloud for semantic segmentation. In *AAAI*, 2022. 3
- [91] Haozhe Xie, Hongxun Yao, Xiaoshuai Sun, Shangchen Zhou, and Shengping Zhang. Pix2vox: Context-aware 3d reconstruction from single and multi-view images. In *ICCV*, 2019. 2
- [92] Haozhe Xie, Hongxun Yao, Shangchen Zhou, Jiageng Mao, Shengping Zhang, and Wenxiu Sun. Grnet: Gridding residual network for dense point cloud completion. In *ECCV*, 2020. 2
- [93] Qizhe Xie, Zihang Dai, Eduard Hovy, Thang Luong, and Quoc Le. Unsupervised data augmentation for consistency training. *NeurIPS*, 2020. 3
- [94] Qiangeng Xu, Weiyue Wang, Duygu Ceylan, Radomir Mech, and Ulrich Neumann. Disn: Deep implicit surface network for high-quality single-view 3d reconstruction. *NeurIPS*, 2019. 2
- [95] Xuejun Yan, Hongyu Yan, Jingjing Wang, Hang Du, Zhihong Wu, Di Xie, Shiliang Pu, and Li Lu. Fbnet: Feedback network for point cloud completion. In *ECCV*, 2022. 2
- [96] Li Yi, Boqing Gong, and Thomas Funkhouser. Complete & label: A domain adaptation approach to semantic segmentation of lidar point clouds. In *CVPR*, 2021. 3
- [97] Kangxue Yin, Jun Gao, Maria Shugrina, Sameh Khamis, and Sanja Fidler. 3dstylenet: Creating 3d shapes with geometric and texture style variations. In *ICCV*, 2021. 3
- [98] Xumin Yu, Yongming Rao, Ziyi Wang, Zuyan Liu, Jiwen Lu, and Jie Zhou. Pointr: Diverse point cloud completion with geometry-aware transformers. In *ICCV*, 2021. 2

- [99] Wentao Yuan, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. Pcn: Point completion network. In *3DV*, 2018. [2](#)
- [100] Junzhe Zhang, Xinyi Chen, Zhongang Cai, Liang Pan, Haiyu Zhao, Shuai Yi, Chai Kiat Yeo, Bo Dai, and Chen Change Loy. Unsupervised 3d shape completion through gan inversion. In *CVPR*, 2021. [2](#)
- [101] Wenxiao Zhang, Zhen Dong, Jun Liu, Qingan Yan, Chunxia Xiao, et al. Point cloud completion via skeleton-detail transformer. *TVCG*, 2022. [2](#)
- [102] Sicheng Zhao, Yezhen Wang, Bo Li, Bichen Wu, Yang Gao, Pengfei Xu, Trevor Darrell, and Kurt Keutzer. epointda: An end-to-end simulation-to-real domain adaptation framework for lidar point cloud segmentation. In *AAAI*, 2021. [3](#)
- [103] Longkun Zou, Hui Tang, Ke Chen, and Kui Jia. Geometry-aware self-training for unsupervised domain adaptation on object point clouds. In *ICCV*, 2021. [3](#)