

# Learning a Structured Latent Space for Unsupervised Point Cloud Completion

Yingjie Cai<sup>1</sup>, Kwan-Yee Lin<sup>1\*</sup>, Chao Zhang<sup>2</sup>, Qiang Wang<sup>2</sup>, Xiaogang Wang<sup>1</sup>, Hongsheng Li<sup>1\*</sup>

<sup>1</sup>CUHK-SenseTime Joint Laboratory, The Chinese University of Hong Kong

<sup>2</sup>Samsung Research Institute China - Beijing (SRC-B)

caiyijinge@link.cuhk.edu.hk

## Abstract

Unsupervised point cloud completion aims at estimating the corresponding complete point cloud of a partial point cloud in an unpaired manner. It is a crucial but challenging problem since there is no paired partial-complete supervision that can be exploited directly. In this work, we propose a novel framework, which learns a unified and structured latent space that encoding both partial and complete point clouds. Specifically, we map a series of related partial point clouds into multiple complete shape and occlusion code pairs and fuse the codes to obtain their representations in the unified latent space. To enforce the learning of such a structured latent space, the proposed method adopts a series of constraints including structured ranking regularization, latent code swapping constraint, and distribution supervision on the related partial point clouds. By establishing such a unified and structured latent space, better partial-complete geometry consistency and shape completion accuracy can be achieved. Extensive experiments show that our proposed method consistently outperforms state-of-the-art unsupervised methods on both synthetic ShapeNet and real-world KITTI, ScanNet, and Matterport3D datasets.

## 1. Introduction

Point cloud completion aims at estimating the corresponding complete point cloud of a partial point cloud, which is an important task and can assist downstream applications such as shape classification [17, 26–28, 34], robotics navigation [12, 31] and scene understanding [1, 2, 10, 19], as raw point clouds are often noisy, sparse and partial. Although fully supervised point cloud completion methods [21, 22, 25, 43, 45, 48–51] have achieved impressive performance, they heavily rely on large-scale paired partial-complete training data. It is, however, difficult to collect paired data from real-world scans. Additionally, such com-

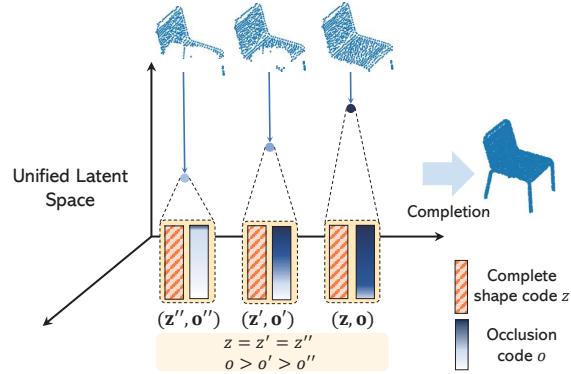


Figure 1. Illustration of the unified and structured latent space where any point cloud can be represented as a complete shape code  $\mathbf{z}$  and a corresponding occlusion code  $\mathbf{o}$ . We complete the partial point cloud via properly regularizing such codes in an unsupervised manner. Best viewed in color.

pletion networks trained on paired real data or paired synthetic data cannot sufficiently generalize to actual scans, as their data distributions might not well match those of training samples.

A promising alternate solution is to learn a completion network in an unpaired manner following the setup of [6, 41, 44, 52]. However, it is a more challenging setup since there is no paired and accurate point-wise supervision that can be adopted directly. To tackle the problem, different methods are proposed to adopt different types of supervision from the unpaired data. A representative work [52] adopts GAN inversion for 3D shape completion. It trains a complete point cloud generator with adversarial losses. During inference, an optimal shape code can be recovered via hundreds of gradient descent iterations by minimizing a partial-complete consistency loss. The consistency<sup>1</sup> between the predicted point cloud and the input can be maintained. However, the inverse optimization is generally unstable and easy to stuck at local minima if using unsuitable

\*H. Li and K. Lin are the co-corresponding authors.

<sup>1</sup>Consistency describes whether the predicted point cloud represents the same object as the partial input.

initial code, unsuitable learning rate or too many iterations, etc. The inversion process is much more time-consuming than direct methods ( $\sim 3500x$ ). Another representative unsupervised work [41] exploits cycle supervisions to enhance consistency indirectly by learning bidirectional transformations between the latent spaces of complete and incomplete shapes (point clouds). However, the bidirectional transformations need to be separately modeled and are difficult to learn, especially for the complete-to-partial mapping. If one direction is not learned well, the other direction would be influenced correspondingly. In summary, without direct and accurate paired supervision, designing proper supervision and applying to unsupervised point cloud completion is of great importance to this task.

To this end, we propose to create a unified and structured latent space for encoding both partial and complete shapes. To apply strong supervisions for unsupervised point cloud completion, we make an assumption that each partial shape is created by occluding a complete one. If a complete shape is occluded to become partial in the 3D space, its code in the latent space should also be “occluded” from a complete shape code accordingly. We model the “occlusion” of a complete code in the latent space as weighting each of its dimension with a weight in  $[0, 1]$ . However, instead of manually determining the occlusion weights, we make them learned from the training data. In this way, the complete and partial shapes are strongly bounded in a unified latent space. In addition, to better regularize the relation between partial point clouds from the same complete shape, the occlusion code of a more occluded shape is required to have smaller weights than that of a less occluded shape.

Specifically, to learn the unified latent space, we represent any partial or complete point cloud by two codes: a complete shape code and an occlusion code. The complete shape code can be fed into a completion decoder to reconstruct the corresponding complete point cloud. The “occluded” shape code via multiplying the above two codes can be fed into a partial decoder to reconstruct the partial shape. Furthermore, we create a series of related partial point clouds by gradually removing more points from a partial shape and apply ranking constraints to their occlusion codes by N-pair loss [36] according to their relative occlusion degrees. Their complete shape codes are required to be equal since they represent the same object. By adopting such properly designed strong supervisions, more accurate complete point clouds with better geometric consistency and shape details can be reconstructed.

We experiment on popular point cloud completion benchmarks, including a synthetic dataset (ShapeNet [23]) and real datasets (KITTI [13], ScanNet [9] and Matterport3D [3]). The proposed method outperforms state-of-the-art unsupervised methods [6, 41, 44, 52]. Our main contributions are summarized as follows:

- We propose to learn a unified and structured latent space for unsupervised point cloud completion, which encodes both partial and complete point clouds to improve partial-complete geometry consistency and lead to better shape completion accuracy.
- We propose to constrain the complete and occlusion codes of a series of related partial point clouds to enhance the learning of the structured latent space.
- Experimental results demonstrate the superiority of the proposed method over state-of-the-art unsupervised point cloud completion methods on both synthetic and real datasets.

## 2. Related Work

**Point Cloud Completion.** Point cloud completion has played an important role for many downstream applications such as robotics [12, 31] and perception [1, 2, 10, 20], which has seen significant development since the pioneering work PCN [51] proposed. Most existing approaches like [6, 11, 22, 22, 25, 29, 37, 39, 42, 48, 50, 53] are trained in a fully-supervised manner. Although supervised point cloud completion methods have achieved impressive results, they are difficult to generalize to real-world scans, since the paired data is difficult to collect for actual scans and their data distributions might not match well. Pcl2pcl [6] first proposes to complete the partial shapes in an unsupervised manner without the need for paired data, which trains two separate auto-encoders, for reconstructing complete shapes and partial ones respectively and learns a latent code transformation from the latent space of partial shapes to that of the complete ones. Its subsequent work [44] outputs multiple plausible complete shapes from a partial input. Base on Pcl2pcl, Cycle4completion [41] exploits an extra complete-to-partial latent spaces transformation in addition of partial-to-complete direction to capture the bidirectional geometric correspondence between incomplete and complete shapes. Another unsupervised work ShapeInversion [52] proposes to apply GAN inversion which utilizes the shape prior learned from a pre-trained generator to complete the partial shape in an unsupervised manner. However, the inverse optimization process is time-consuming compared with forward-based methods and the results are easy to stuck at local minima, which greatly limits the practical application of inversion-based methods. Different from existing methods, we propose to learn a unified latent space supervised by tailored structured latent constraints to reconstruct better complete shapes.

**Structured Ranking Losses.** Deep metric learning plays an important role in various applications of computer vision, such as image retrieval [30, 36], clustering [18], and transfer learning [33]. The loss function is one of the essential components in successful deep metric learning frame-

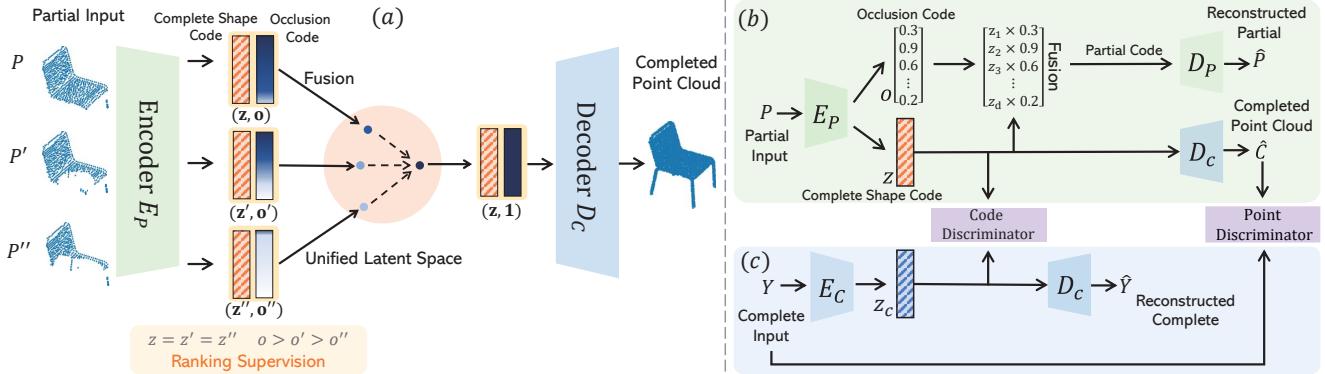


Figure 2. **Overview.** (a) A series of related partial point clouds encoded to multiple complete shape and occlusion code pairs. Their element-wise multiplication are the representation in the unified latent space. (b) Reconstructing the partial input  $\hat{P}$  and predicting the completed point cloud  $\hat{C}$  simultaneously with a shape latent code discriminator and a complete point cloud discriminator. (c) The real complete shape codes and point clouds are provided by the complete point cloud auto-encoder. Best viewed in color.

works and a large variety of loss functions have been proposed. Contrastive loss [7, 16] captures the relationship between pairwise data points, i.e., similarity or dissimilarity. Triplet-based loss is widely studied [8, 35, 38] and composed of an anchor point, a positive data point, and a negative point and aims to pull the anchor point closer to the positive point than to the negative point by a fixed margin  $\delta$ . Inspired by this, recent ranking-motivated approaches [24, 30, 32, 33, 35, 36] have proposed taking into consideration richer structured information across multiple data points and achieve impressive performance. Different from triplet loss who considers one negative point, N-pair loss [36] aims to identify one positive example from multi negative examples.

### 3. Method

The goal of our work is to reconstruct the complete point cloud from an input partial point cloud with only unpaired data. Designing proper and strong supervisions is of great importance to tackle this challenging problem. We propose to learn a unified latent space for encoding both complete and partial point clouds (shapes). We first introduce the unified latent space in Section 3.1, which encodes complete and partial point clouds in a joint space. Then structured latent supervisions of a series of related partial point clouds are adopted to further regularize the learning of the space (Section 3.2). The overall architecture is depicted in Figure 2.

#### 3.1. A Unified Latent Space for Point Cloud Encoding

We introduce the unified latent space to establish the relations between partial and complete point clouds in an unpaired manner. Partial point clouds can be considered as being created by occluding the complete shapes. A partial point cloud represents the same object as its correspond-

ing complete point cloud and the difference between the complete and partial point is just their occlusion degrees, as shown in Figure 2 (a). Therefore, we embed the incomplete and the complete point clouds into a unified latent space equipping with different occlusion degrees.

Specifically, as illustrated in Figure 2 (b), we map any partial point cloud  $P$  into a complete shape code  $z \in \mathbb{R}^d$  and a corresponding occlusion code  $o \in \mathbb{R}^d$  via a point cloud encoder  $E_p$  [46] consisting of EdgeConv [40] layers. Each entry  $z_i, i \in [1, \dots, d]$  of the occlusion code is bounded in  $[0, 1]$  by a sigmoid function and has the same length as the complete shape code. Occlusion of the complete shape in the latent space is modeled as softly “gating” each dimension of the complete shape code. A smaller occlusion value denotes more occlusion to a complete shape. The embedding of the partial shape in the unified latent space can then be obtained by element-wise multiplication of the two codes. The complete and partial codes are then fed into two separate decoders,  $D_c$  and  $D_p$ , to generate completed point cloud  $\hat{C}$  and reconstruct the input partial point cloud  $\hat{P}$ , respectively. The two separate decoders adopt the same architecture made up of a multiple layers perception (MLP) following [41]. Both  $\hat{C}$  and  $\hat{P}$  are supervised by a point-wise Chamfer Distance (CD) loss with respect to the partial input. The point-wise reconstruction loss is expressed as:

$$\mathcal{L}_{rec} = \mathcal{L}_{CD}(P, \hat{P}) + \mathcal{L}_{CD}(P, \text{Deg}(\hat{C})). \quad (1)$$

For  $(P, \hat{C})$ , the bi-direction Chamfer Distance cannot be utilized directly, but only the Unidirectional Chamfer Distance (UCD) cannot provide enough supervision for the inference of the missing parts, so we degrade (*i.e.*  $\text{Deg}$ ) the  $\hat{C}$  into a partial point cloud following [52]’s degradation module, where only top- $k$  nearest points with respect to partial point cloud are kept. In order to further encourage the predicted complete point clouds to represent reasonable

shapes, a point cloud discriminator is adopted. We formulate the point cloud discriminator with WGAN-GP [15] loss as

$$\mathcal{L}_D^p = \mathbb{E}_{\hat{C}} D(\hat{C}) - \mathbb{E}_Y D(Y) + \lambda_{gp} \mathcal{T}_D, \quad (2)$$

where  $\lambda_{gp}$  is a pre-defined weight factor and  $\mathcal{T}_D$  is gradient penalty term, denoted as

$$\mathcal{T}_D = \mathbb{E}_{\hat{C}} \left[ \left( \left\| \nabla_{\hat{C}} D(\hat{C}) \right\|_2 - 1 \right)^2 \right]. \quad (3)$$

The code adversarial training loss for the encoder  $E_p$  and decoder  $D_c$  is

$$\mathcal{L}_G^p = -\mathbb{E}_{\hat{C}} D(\hat{C}). \quad (4)$$

Note that, during inference, only the encoder  $E_p$  and decoder  $D_c$  are needed.

### 3.2. Structural Regularization of the Unified Space

To further regularize the learning of the structured latent space, we create a series of related partial point clouds and propose several properly designed latent code supervisions, including structured ranking regularization, latent code swapping constraint, and latent code distribution supervision to enhance the learning of the structured latent space. Specifically, given a partial input  $P$ , we can create a series of related partial point clouds (see  $\{P, P', P''\}$  in Figure 2 (a)) by gradually removing more points. For  $P'$  and  $P''$ , there are  $K$  and  $2K$  points removed from the initial partial shape  $P$ . Therefore, for a triplet of such related partial point clouds  $S=\{P, P', P''\}$ , their occlusion degrees gradually increase.

**Structured Ranking Regularization.** For their complete shape codes, since the point clouds represent the same object, their complete shape latent codes,  $\mathbf{z}, \mathbf{z}', \mathbf{z}''$  are required to be equal. And we adopt *Smooth L1* loss to constrain them

$$L_z = L_1(\mathbf{z}, \mathbf{z}') + L_1(\mathbf{z}, \mathbf{z}''). \quad (5)$$

Furthermore, since their occlusion codes represent the increasing degree of occlusion, their corresponding occlusion codes' weights shall be smaller as their occlusion degrees increase. Such relations between their occlusion codes can be expressed as

$$o''_i \leq o'_i \leq o_i \leq 1 \quad \text{for } i = 1, \dots, d, \quad (6)$$

where  $o''_i, o'_i, o_i$  are the  $i$ th entry of the occlusion codes for  $P'', P', P$ , respectively. To implement such a constraint, we adopt the N-pair loss [36] to constrain the proposed relative relations in Eq. (6),

$$L \left( \mathbf{a}, \mathbf{p}, \{\mathbf{n}_j\}_{j=1}^N \right) = \log \left( 1 + \sum_{j=1}^N \exp \left( \mathbf{a}^\top \mathbf{n}_j - \mathbf{a}^\top \mathbf{p} \right) \right). \quad (7)$$

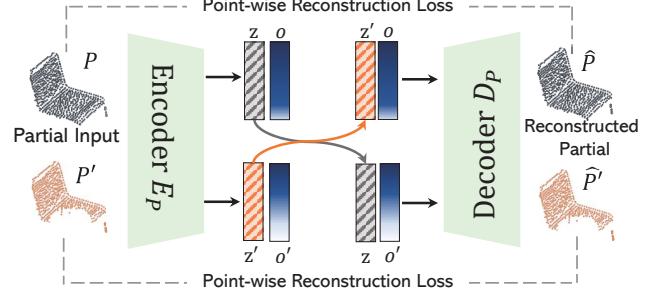


Figure 3. **Illustration of the latent code swapping.** To better decouple the complete shape and occlusion codes, we swap the complete shape codes between related partial point clouds and apply point-wise reconstruction loss to supervise the reconstructed partial point clouds. Best viewed in color.

For the N-pair loss, there are one anchor sample  $\mathbf{a} \in \mathbb{R}^d$ , one positive sample  $\mathbf{p} \in \mathbb{R}^d$  and  $N$  negative samples  $\mathbf{n}_j \in \mathbb{R}^d$ , as shown in Eq. (7). By minimizing the loss function, the positive sample would be pulled closer to the anchor, and negative samples are pushed farther away from the anchor. Here, we adopt N-pair loss by choosing different occlusion codes as anchors, which can be written as the following sets, where  $\mathbf{1} \in \mathbb{R}^d$  is an all-one vector:

$$\begin{cases} \mathbf{a} = \mathbf{1}, & \mathbf{p} = \mathbf{o}, & \{\mathbf{n}_j\}_{j=1}^N = \{\mathbf{o}', \mathbf{o}''\} \end{cases}, \quad (8)$$

$$\begin{cases} \mathbf{a} = \mathbf{o}, & \mathbf{p} = \mathbf{o}', & \{\mathbf{n}_j\}_{j=1}^N = \{\mathbf{o}''\} \end{cases},$$

$$\begin{cases} \mathbf{a} = \mathbf{o}'', & \mathbf{p} = \mathbf{o}', & \{\mathbf{n}_j\}_{j=1}^N = \{\mathbf{o}\} \end{cases},$$

Therefore, the proposed relative ranking relation Eq. (6) between occlusion codes are constrained via applying the N-pair loss on each of the sets in Eq. (8). Through adopting such strong ranking regularization, the unified latent space is trained to be more structured.

**Latent Code Swapping.** To further regularize the complete shape codes and occlusion codes, we employ a latent code swapping constraint. Specifically, as illustrated in Figure 3, we swap the complete and occlusion codes extracted from a partial point cloud  $P$  and a more occluded version of  $P'$  to reconstruct the corresponding partial point clouds. Based on our assumption on the unified space,  $\mathbf{z}$  and  $\mathbf{z}'$  represent the same complete object and the partial degrees are decided by their occlusion codes. Therefore, no matter  $\mathbf{o}'$  is combined with  $\mathbf{z}$  or  $\mathbf{z}'$ , the same partial point cloud should be reconstructed. Therefore, we also feed the fused code from  $\mathbf{z}$  and  $\mathbf{o}'$  to the decoder  $D_p$  and apply point-wise reconstruction loss  $\mathcal{L}_{rec}$  to constrain. And the fused shape code from  $\mathbf{z}'$  and  $\mathbf{o}$  is similarly processed. Through such a latent code swapping constraint, the disentanglement of the complete shape codes and the occlusion codes are greatly improved, which leads to better shape completion.

**Latent Code Distribution.** In order to further constrain the reality of the complete shape codes, a shape latent code dis-

criminator is applied to directly supervise whether the complete shape codes learned from partial point clouds match well with the real complete shape codes extracted from a complete shape auto-encoder. As illustrated in Figure 2 (c), the real shape latent code  $z_c \in \mathbb{R}^d$  can be obtained from a complete point cloud auto-encoder following [41]. The input of the auto-encoder  $Y$  is a shape randomly sampled from a complete point clouds set which is not paired with  $P$ . The objective functions for updating latent code discriminator  $\mathcal{L}_D^c$  and latent code generator  $\mathcal{L}_G^c$  are similar as Eqs. (2) and (4) respectively.

In summary, through the constraints on the unified latent space, the complete shape and occlusion codes can be well learned to enhance the relation between the complete and partial point clouds.

**Overall Loss.** The overall training objective for the two discriminators is

$$\mathcal{L}_D = \mathcal{L}_D^p + \mathcal{L}_D^c. \quad (9)$$

And the overall training loss for the encoders and decoders including  $E_p$ ,  $D_p$ ,  $E_c$  and  $D_c$  is the weighted sum of pointwise reconstruction loss, structured latent supervision and adversarial losses:

$$\mathcal{L} = \gamma \mathcal{L}_{rec} + \beta \mathcal{L}_z + \mathcal{L}_{npair} + \mathcal{L}_G^p + \mathcal{L}_G^c, \quad (10)$$

where  $\gamma$  and  $\beta$  are pre-defined weight factors.

## 4. Experiments

We evaluate the proposed method through extensive experiments. Besides shape completion on the virtual scan benchmarks, we also demonstrate its effectiveness compared with other methods on the widely used real-world partial scans.

**Datasets.** For a comprehensive comparison, we conduct experiments on both synthetic and real-world partial shapes following state-of-the-art unsupervised point cloud completion methods [6, 41, 44, 52]. We evaluated our method on three synthetic datasets CRN [39], 3D-EPN [11] and PartNet [29], which are all derived from ShapeNet [5]. For real-world scans, we evaluate on objects extracted from three datasets covering indoor and outdoor scenes, KITTI (cars) [14], ScanNet (chairs and tables) [9], and MatterPort3D (chairs and tables) [4].

**Evaluation Metrics.** For datasets equipped with ground truth, we evaluate the shape completion performance using CD and F1-score following previous unsupervised point cloud completion methods [6, 41, 52], where F1-score is the harmonic average of the accuracy and the completeness. The Chamfer Distance is defined as:

$$\begin{aligned} \mathcal{L}_{CD}(\mathbf{x}_{out}, \mathbf{x}_{in}) &= \frac{1}{|\mathbf{x}_{out}|} \sum_{p \in \mathbf{x}_{out}} \min_{q \in \mathbf{x}_{in}} \|p - q\|_2^2 \\ &+ \frac{1}{|\mathbf{x}_{in}|} \sum_{q \in \mathbf{x}_{in}} \min_{p \in \mathbf{x}_{out}} \|p - q\|_2^2, \end{aligned} \quad (11)$$

where  $\mathbf{x}_{out}$  and  $\mathbf{x}_{in}$  are two point clouds. The smaller the distance value is, the more accurate the reconstructed point cloud is. For synthetic dataset PartNet utilized by [44], we follow the method and also use Minimum Matching Distance (MMD) metric to evaluate the accuracy of the completion. The MMD measures the quality of the completed shape and we calculate the MMD between the set of completion shapes and the set of test shapes. For real-world scans where no ground truth is provided, we follow [44, 47] to evaluate the generated shapes in terms of UCD and MMD, respectively. The UCD evaluates the consistency and computes the Chamfer distance from the partial input to the predicted complete point cloud.

**Implementation Details.** The proposed method follows previous unsupervised point cloud methods [6, 41, 44, 52] to train single-class models separately for better fidelity. The number of points of the predicted complete shapes is 2048 for all datasets. We use 8 TITAN GPUs to implement our experiments. Specifically, we adopt an Adam optimizer with a learning rate  $10^{-4}$  and a batch size of 16 per GPU to train the framework for 500 epochs. The top 5 points are kept for complete point cloud degradation. The dimension  $d$  of complete shape code and occlusion code are both 96 and a number of 500 points (i.e.,  $K=500$ ) are gradually removed to generate more partial point clouds. The  $\gamma=100$ ,  $\beta=10$  and  $\lambda_{gp}=1$  are set for the combination of losses.

### 4.1. Completion Results on ShapeNet Benchmark

We conduct experiments on CRN, 3D-EPN, and PartNet synthetic datasets generated from ShapeNet to demonstrate the superiority of our method over state-of-the-art unsupervised methods.

**Comparison on CRN and 3D-EPN datasets.** For synthetic datasets CRN and 3D-EPN equipped with ground truth, we evaluate the shape completion performance using CD and F1-score following [6, 41, 52]. Tables 1 and 2 show the experiment results on the two datasets across eight categories where “Cycle.” and “Inversion.” represent Cycle4Completion and ShapeInversion respectively. As illustrated in Table 1, the proposed method outperforms state-of-the-art unsupervised method ShapeInversion [52] by large margins across most categories and achieves 12.2 CD and 85.6 F1-score surpassing [52] by 2.7 and 1.7 for average CD and F1-Score metrics, respectively. For ShapeNet 3D-EPN dataset, as shown in Table 2, our method consistently achieves the best completion performance on most categories, especially for categories like “chair” and “table”, whose shape diversity and number of training samples are relatively rich compared with other categories. For chair and table, the CD and F1-score metrics show significant improvements (from 14.6/84.2 to 12.1/86.4 for chair and from 22.5/82.7 to 19.8/85.5 for table). There is a 0.9 gap between our method and [41] on the car category. Through evalua-

Table 1. Shape completion performance on CRN benchmark. The numbers shown are  $[CD \downarrow / F1 \uparrow]$ , where CD is scaled by  $10^4$ .

Methods	Plane	Cabinet	Car	Chair	Lamp	Sofa	Table	Boat	Average
Pcl2pcl [6]	9.7/89.1	27.1/68.4	15.8/80.8	26.9/70.4	25.7/70.4	34.1/58.4	23.6/79.0	15.7/77.8	22.4/74.2
Cycle. [41]	5.2/94.0	14.7/82.1	12.4/82.1	18.0/77.5	17.3/77.4	21.0/75.2	18.9/81.2	11.5/84.8	14.9/81.8
Inversion. [52]	5.6/94.3	16.1/77.2	13.0/85.8	15.4/81.2	18.0/ <b>81.7</b>	24.6/78.4	<b>16.2/85.5</b>	10.1/87.0	14.9/83.9
Ours	<b>3.9/95.9</b>	<b>13.5/83.3</b>	<b>8.7/90.4</b>	<b>13.9/82.3</b>	<b>15.8/81.0</b>	<b>14.8/81.6</b>	17.1/82.6	<b>10.0/87.6</b>	<b>12.2/85.6</b>

Table 2. Shape completion performance on 3D-EPN benchmark. The numbers shown are  $[CD \downarrow / F1 \uparrow]$ , where CD is scaled by  $10^4$ .

Methods	Plane	Cabinet	Car	Chair	Lamp	Sofa	Table	Boat	Average
Pcl2pcl [6]	4.0/–	19.0/–	10.0/–	20.0/–	23.0/–	26.0/–	26.0/–	11.0/–	17.4/–
Cycle. [41]	3.7/96.4	12.6/ <b>87.1</b>	<b>8.1/91.8</b>	14.6/84.2	18.2/80.6	26.2/71.7	22.5/82.7	8.7/89.8	14.3/85.5
Inversion. [52]	4.3/96.2	20.7/79.4	11.9/86.0	20.6/81.1	25.9/78.4	54.8/74.7	38.0/80.2	12.8/85.2	23.6/82.7
Ours	<b>3.5/96.8</b>	<b>12.2/86.4</b>	9.0/88.4	<b>12.1/86.4</b>	<b>17.6/81.6</b>	<b>26.0/75.5</b>	<b>19.8/85.5</b>	<b>8.6/89.8</b>	<b>13.6/86.3</b>

Table 3. Shape completion performance on PartNet benchmark. We evaluate the results with  $MMD \downarrow$ , which is scaled by  $10^2$ .

Methods	Chair	Lamp	Table	Average
Pcl2pcl [6]	1.90	2.50	1.90	2.10
MPC [44]	1.52	1.97	1.46	1.65
Cycle. [41]	1.71	3.46	1.56	2.24
Inversion. [52]	1.68	2.54	1.74	1.98
Ours	<b>1.43</b>	<b>1.95</b>	<b>1.37</b>	<b>1.58</b>

tion on the two popular synthetic datasets across eight categories, our method outperforms existing methods consistently, which proves the superiority of the proposed framework that learns a unified latent space with effective and efficient structured regularization.

**Comparison on PartNet dataset.** We also conduct experiments on PartNet dataset utilized by MPC [44]. The PartNet benchmark is generated by removing semantic parts on ShapeNet dataset. We follow [44] to adopt Minimum Matching Distance on three categories to evaluate the quality of the completed shapes. As shown in Table 3, our method outperforms existing state-of-the-art unsupervised methods on the three categories consistently.

**Qualitative Results.** Figure 4 illustrates the qualitative results of the same samples from the ShapeNet dataset. Despite the efforts of previous approaches, they usually fail to deal with severe occlusion cases and can not maintain shape details. As shown by the couch and car in Figure 4 first two rows, in the case of severe occlusion, the completed point clouds of previous methods do not represent the target objects (see the large missing regions which are not recovered correctly). However, our method can accurately recover the complete point cloud of the target object, even when only fairly limited information is available under severe occlusion. What’s more, our method reconstructs more accurate complete point clouds equipped with better fine-grained shape details. As shown by the red dotted boxes, our method can generate more accurate complete shapes at the corner of the lamps, the tail of the planes, and the legs of chairs and tables. We attribute the great results to

the learned unified latent space and properly applied structured latent supervisions, which leads to more reasonable predicted complete point clouds and better consistency between partial and complete point clouds.

## 4.2. Completion Results on Real-World Scans

We investigate the generalization of the proposed method on various real-world datasets including both outdoor and indoor scenes, where the objects tend to be more incomplete and noisier. The trained car, chair and table models on CRN dataset are directly utilized to predict complete point clouds on KITTI, ScanNet and MatterPort3D datasets without any further fine-tuning process. As shown in Table 4 (Cycle. [41] vs. Ours), our method significantly outperforms Cycle4Completion [41] across multiple categories on all the three real-scan datasets. For the comparison with ShapeInversion [52], as the inversion process is to minimize the UCD loss directly between partial-complete pairs, it is unfair to compare our method that does not involve GAN inversion with ShapeInversion [52]. However, our method is also compatible with ShapeInversion [52]. When integrating GAN inversion on the top of our method, it can surpass ShapeInversion on all various real-world scans, as shown in Table 4 (Inversion. [52] vs. Ours + Inversion), which demonstrates that our method can enhance consistency between the predicted complete point cloud and the partial input. On the other hand, the results of our method in Table 5 also surpass other unsupervised methods consistently.

In addition, as shown in Tables 4 and 5, we also compare the generalization of our models with those of state-of-the-art fully-supervised methods [48, 50] on the real scans. Their authors’ official released models are used here. Our unsupervised model can outperform them on multiple categories of real scans, which demonstrates that the proposed unsupervised method has better generalization ability on real-world scans than the supervised methods, which are specifically trained to only fit their original synthetic data. Figure 5 shows the completion results of our method on

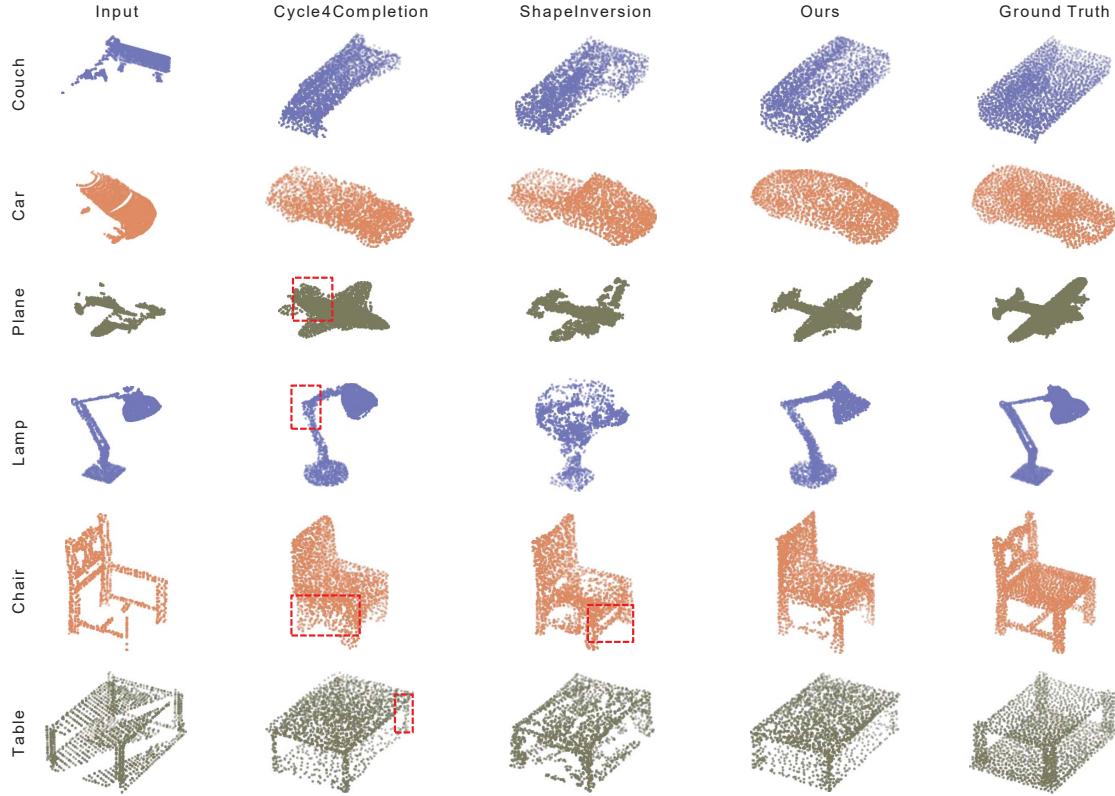


Figure 4. **Point cloud completion results on ShapeNet dataset.** From left to right: partial input, results of Cycle4completion [41], ShapeInversion [52] ours and ground truth. Our results achieve more accurate completion under severe occlusion and recover better fine-grained shape details compared with state-of-the-art methods. Better viewed in color and zoom in.

Table 4. Shape completion performance on the real scans. The results are evaluated by  $UCD \downarrow$ , where  $UCD$  is scaled by  $10^4$ . *sup.*: supervised methods.

Methods	<i>sup.</i>	ScanNet		MatterPort3D		KITTI
		Chair	Table	Chair	Table	
GRNet [48]	yes	1.6	1.6	1.6	1.5	2.2
PoinTr [50]	yes	1.7	1.5	1.8	1.3	1.9
Pcl2pcl [6]	no	17.3	9.1	15.9	6.0	9.2
Cycle. [41]	no	9.4	4.3	4.9	4.9	9.4
Inversion. [52]	no	3.2	3.3	3.6	3.1	2.9
Ours	no	3.2	2.7	3.3	2.7	4.2
Ours + Inversion	no	<b>1.1</b>	<b>0.87</b>	<b>1.1</b>	<b>0.87</b>	<b>0.76</b>

Table 5. Shape completion performance on the real scans. We evaluate the results with  $MMD \downarrow$ , where  $MMD$  is scaled by  $10^2$ . *sup.*: supervised methods.

Methods	<i>sup.</i>	ScanNet		MatterPort3D		KITTI
		Chair	Table	Chair	Table	
GRNet [48]	yes	6.070	6.302	6.147	6.911	2.845
PoinTr [50]	yes	6.001	6.089	6.248	6.648	2.790
Cycle. [41]	no	6.278	5.727	6.022	6.535	3.033
Inversion. [52]	no	6.370	6.222	6.360	7.110	2.850
Ours	no	<b>5.893</b>	<b>5.541</b>	<b>5.770</b>	<b>6.076</b>	<b>2.742</b>

real data, which indicates that even under severe occlusions (such as KITTI’s car), our method can still generate reasonable complete shapes.

### 4.3. Ablation Study

To verify the effectiveness of each component of the proposed method, we conduct a series of experiments on four representative categories on ShapeNet CRN dataset.

**Effect of Unified Latent Space for Point Clouds Encoding.** To evaluate the benefits of introducing the unified latent code space for unsupervised point cloud completion, we compare two alternative strategies that do not encode occlusion codes as soft weighting vectors. Instead of fusing the codes via element-wise multiplication, we test fusing the shape and occlusion codes via concatenation or element-wise addition. Table 6 shows the quantitative results of the compared schemes. We employ our simplified model, which multiples the shape and occlusion codes, and has point and code discriminators and code swapping constraints under our unified space design but without using the ranking constraints as the baseline (denoted as “Uni. Space”) in Table 6. The average CD drops from 18.3 to 19.1 and 18.6 when fusing the two codes via concatenation or addition, which proves that the learned unified latent space is conducive to unsupervised point cloud completion, and also creates a foundation for integrating our stronger ranking supervision.

**Effect of Structured Ranking Supervision.** To evaluate

Table 6. Comparison of different schemes for fusing complete shape and occlusion codes. CD $\downarrow$  scaled by  $10^4$  are reported here.

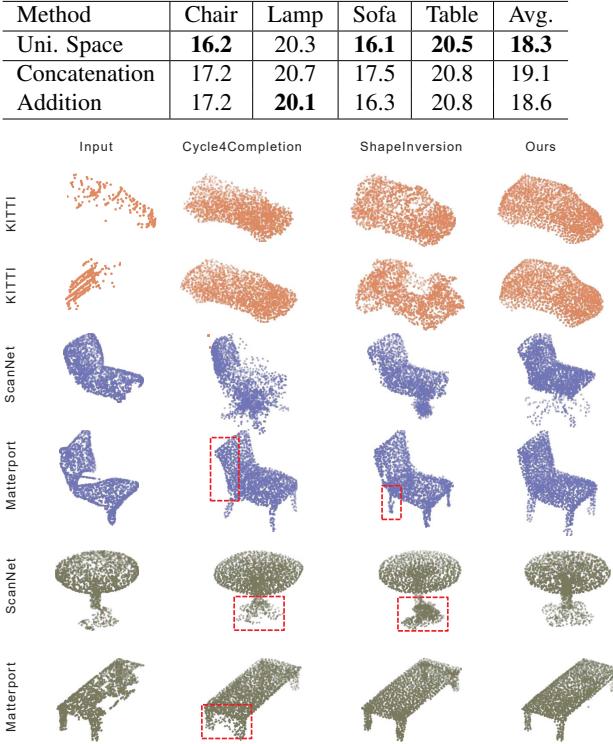


Figure 5. **Point cloud completion results on real-world scans.** From left to right: partial input, results of Cycle4completion [41], ShapeInversion [52] and ours, respectively. Our method predicts more reasonable complete shape results compared with other state-of-the-art methods. Better viewed in color and zoom in.

Table 7. Effect of different ranking supervisions. CD $\downarrow$  scaled by  $10^4$  are reported here.

Method	Chair	Lamp	Sofa	Table	Avg.
Uni. Space	16.2	20.3	16.1	20.5	18.3
+triplet	14.7	19.5	15.0	19.5	17.2
+n-pair	<b>13.9</b>	<b>15.8</b>	<b>14.8</b>	<b>17.1</b>	<b>15.4</b>

the effect of our proposed structured ranking supervision, we adopt different ranking supervisions, N-pair loss and triplet loss with varying hyper-parameters. As shown in Table 7, our model with the triplet loss with different hyper-parameters (denoted as “+triplet”) shows better performance compared with not using the ranking supervisions. Specifically the average Chamfer Distance improves from 18.3 to 17.2. Furthermore, when equipped with N-pair loss, the results are further improved and the average CD has an obvious improvement from 18.3 to 15.4. It demonstrates that the proposed structured ranking supervisions assist the learning of the unified latent space. Therefore, for our full model, we use N-pair loss without any hyper-parameter as our final structured ranking supervision.

**Effect of Discriminators.** We also test the effects of the two discriminators used in our system via deducting the code

Table 8. Effect of discriminators and latent code swapping. CD $\downarrow$  scaled by  $10^4$  are reported here.

Method	Chair	Lamp	Sofa	Table	Avg.
Full Model	<b>13.9</b>	<b>15.8</b>	<b>14.8</b>	<b>17.1</b>	<b>15.4</b>
w/o pointD	14.7	23.6	20.0	19.4	19.4
w/o codeD	18.5	21.7	22.0	20.8	20.8
w/o code swap	14.7	20.0	17.0	19.2	17.7

discriminator (denoted as w/o codeD) or point cloud discriminator (denoted as w/o pointD) from our final model to compare with our full solution (denoted as Full Model). As shown in Table 8, when removing the point cloud discriminator or latent code discriminator, the average Chamfer Distances drops 4 and 5.4 points, respectively. It proves that the integration of the discriminators can assist the complete model to generate better predictions via providing direct guidance about the distributions of complete shape latent codes and point clouds.

**Effect of Latent Code Swapping.** The occlusion codes swapping between a series of related partial point clouds is designed to improve the disentanglement of complete shape and occlusion codes. Here we test removing the design of swapping occlusion code for reconstructing pairs of partial point clouds (denoted as w/o code swap) to compare with our solution. As shown in the last row of Table 8, without latent codes swapping, the system shows much worse performance on all categories, decreasing for 2.3 CD from 15.4 to 17.7. It demonstrates that generating related partial point clouds and swapping their latent codes provide effective supervisions on decoupling the complete shape and occlusion codes more thoroughly.

## 5. Limitations and Conclusion

In this paper, we propose to learn a unified and structured latent space, which encodes both partial and complete point clouds to improve partial-complete geometry consistency under an unsupervised manner. Furthermore, we apply tailored structured latent supervisions between a series of related partial point clouds to enhance the learning of the structured latent space. Extensive experiments prove that the proposed method consistently achieves state-of-the-art performance on both synthetic and real-world benchmarks.

Although our method has made great progress in accuracy, there are still limitations for some fine-grained structure reconstruction of objects, such as the complex texture structure of chairs. These limitations may be solved by designing better decoders or introducing implicit functions.

**Acknowledgments:** This work is supported in part by Centre for Perceptual and Interactive Intelligence Limited, in part by the General Research Fund through the Research Grants Council of Hong Kong under Grants (Nos. 14204021, 14207319, 14203118, 14208619), in part by Research Impact Fund Grant No. R5001-18, in part by CUHK Strategic Fund.

## References

- [1] Yingjie Cai, Xuesong Chen, Chao Zhang, Kwan-Yee Lin, Xiaogang Wang, and Hongsheng Li. Semantic scene completion via integrating instances and scene in-the-loop. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 324–333, 2021. [1](#) [2](#)
- [2] Yingjie Cai, Buyu Li, Zeyu Jiao, Hongsheng Li, Xingyu Zeng, and Xiaogang Wang. Monocular 3d object detection with decoupled structured polygon estimation and height-guided depth estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10478–10485, 2020. [1](#) [2](#)
- [3] Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. Matterport3d: Learning from rgbd data in indoor environments. *International Conference on 3D Vision (3DV)*, 2017. [2](#)
- [4] Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. Matterport3d: Learning from rgbd data in indoor environments. *arXiv preprint arXiv:1709.06158*, 2017. [5](#)
- [5] 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. [5](#)
- [6] Xuelin Chen, Baoquan Chen, and Niloy J Mitra. Unpaired point cloud completion on real scans using adversarial training. *arXiv preprint arXiv:1904.00069*, 2019. [1](#) [2](#) [5](#) [6](#) [7](#)
- [7] Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a similarity metric discriminatively, with application to face verification. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 539–546. IEEE, 2005. [3](#)
- [8] Yin Cui, Feng Zhou, Yuanqing Lin, and Serge Belongie. Fine-grained categorization and dataset bootstrapping using deep metric learning with humans in the loop. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1153–1162, 2016. [3](#)
- [9] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5828–5839, 2017. [2](#) [5](#)
- [10] Angela Dai, Daniel Ritchie, Martin Bokeloh, Scott Reed, Jürgen Sturm, and Matthias Nießner. Scancomplete: Large-scale scene completion and semantic segmentation for 3d scans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4578–4587, 2018. [1](#) [2](#)
- [11] Angela Dai, Charles Ruizhongtai Qi, and Matthias Nießner. Shape completion using 3d-encoder-predictor cnns and shape synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5868–5877, 2017. [2](#) [5](#)
- [12] Jakob Engel, Thomas Schöps, and Daniel Cremers. Lsd-slam: Large-scale direct monocular slam. In *European conference on computer vision*, pages 834–849. Springer, 2014. [1](#) [2](#)
- [13] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. [2](#)
- [14] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361. IEEE, 2012. [5](#)
- [15] Ishaaq Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved training of wasserstein gans. *arXiv preprint arXiv:1704.00028*, 2017. [4](#)
- [16] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1735–1742. IEEE, 2006. [3](#)
- [17] Zhizhong Han, Xiyang Wang, Chi-Man Vong, Yu-Shen Liu, Matthias Zwicker, and CL Chen. 3dviewgraph: Learning global features for 3d shapes from a graph of unordered views with attention. *arXiv preprint arXiv:1905.07503*, 2019. [1](#)
- [18] John R Hershey, Zhuo Chen, Jonathan Le Roux, and Shinji Watanabe. Deep clustering: Discriminative embeddings for segmentation and separation. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 31–35. IEEE, 2016. [2](#)
- [19] Ji Hou, Angela Dai, and Matthias Nießner. 3d-sis: 3d semantic instance segmentation of rgbd scans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4421–4430, 2019. [1](#)
- [20] Ji Hou, Angela Dai, and Matthias Nießner. 3d-sis: 3d semantic instance segmentation of rgbd scans. In *Proc. Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2019. [2](#)
- [21] Tao Hu, Zhizhong Han, and Matthias Zwicker. 3d shape completion with multi-view consistent inference. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10997–11004, 2020. [1](#)
- [22] Zitian Huang, Yikuan Yu, Jiawen Xu, Feng Ni, and Xinyi Le. Pf-net: Point fractal network for 3d point cloud completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7662–7670, 2020. [1](#) [2](#)
- [23] Marcin Kopaczka, Justus Schock, and Dorit Merhof. Superrealtime facial landmark detection and shape fitting by deep regression of shape model parameters. *arXiv preprint*, 2019. [2](#)
- [24] Marc T Law, Raquel Urtasun, and Richard S Zemel. Deep spectral clustering learning. In *International conference on machine learning*, pages 1985–1994. PMLR, 2017. [3](#)
- [25] Minghua Liu, Lu Sheng, Sheng Yang, Jing Shao, and Shi-Min Hu. Morphing and sampling network for dense point cloud completion. In *Proceedings of the AAAI conference on*

*artificial intelligence*, volume 34, pages 11596–11603, 2020. 1, 2

[26] Xinhai Liu, Zhizhong Han, Fangzhou Hong, Yu-Shen Liu, and Matthias Zwicker. Lrc-net: Learning discriminative features on point clouds by encoding local region contexts. *Computer Aided Geometric Design*, 79:101859, 2020. 1

[27] Xinhai Liu, Zhizhong Han, Yu-Shen Liu, and Matthias Zwicker. Fine-grained 3d shape classification with hierarchical part-view attention. *IEEE Transactions on Image Processing*, 30:1744–1758, 2021. 1

[28] Xinhai Liu, Zhizhong Han, Xin Wen, Yu-Shen Liu, and Matthias Zwicker. L2g auto-encoder: Understanding point clouds by local-to-global reconstruction with hierarchical self-attention. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 989–997, 2019. 1

[29] Kaichun Mo, Shilin Zhu, Angel X Chang, Li Yi, Subarna Tripathi, Leonidas J Guibas, and Hao Su. Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 909–918, 2019. 2, 5

[30] Yair Movshovitz-Attias, Alexander Toshev, Thomas K Leung, Sergey Ioffe, and Saurabh Singh. No fuss distance metric learning using proxies. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 360–368, 2017. 2, 3

[31] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. Orb-slam: a versatile and accurate monocular slam system. *IEEE transactions on robotics*, 31(5):1147–1163, 2015. 1, 2

[32] Hyun Oh Song, Stefanie Jegelka, Vivek Rathod, and Kevin Murphy. Deep metric learning via facility location. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5382–5390, 2017. 3

[33] Hyun Oh Song, Yu Xiang, Stefanie Jegelka, and Silvio Savarese. Deep metric learning via lifted structured feature embedding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4004–4012, 2016. 2, 3

[34] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 1

[35] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015. 3

[36] Kihyuk Sohn. Improved deep metric learning with multi-class n-pair loss objective. In *Advances in neural information processing systems*, pages 1857–1865, 2016. 2, 3, 4

[37] Lyne P Tchapmi, Vineet Kosaraju, Hamid Rezatofighi, Ian Reid, and Silvio Savarese. Topnet: Structural point cloud decoder. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 383–392, 2019. 2

[38] Jiang Wang, Yang Song, Thomas Leung, Chuck Rosenberg, Jingbin Wang, James Philbin, Bo Chen, and Ying Wu. Learning fine-grained image similarity with deep ranking. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1386–1393, 2014. 3

[39] Xiaogang Wang, Marcelo H Ang Jr, and Gim Hee Lee. Cascaded refinement network for point cloud completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 790–799, 2020. 2, 5

[40] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*, 38(5):1–12, 2019. 3

[41] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13080–13089, 2021. 1, 2, 3, 5, 6, 7, 8

[42] Xin Wen, Tianyang Li, Zhizhong Han, and Yu-Shen Liu. Point cloud completion by skip-attention network with hierarchical folding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1939–1948, 2020. 2

[43] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 1

[44] Rundi Wu, Xuelin Chen, Yixin Zhuang, and Baoquan Chen. Multimodal shape completion via conditional generative adversarial networks. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16*, pages 281–296. Springer, 2020. 1, 2, 5, 6

[45] 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 *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2021. 1

[46] Chulin Xie, Chuxin Wang, Bo Zhang, Hao Yang, Dong Chen, and Fang Wen. Style-based point generator with adversarial rendering for point cloud completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4619–4628, June 2021. 3

[47] Chulin Xie, Chuxin Wang, Bo Zhang, Hao Yang, Dong Chen, and Fang Wen. Style-based point generator with adversarial rendering for point cloud completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4619–4628, 2021. 5

[48] Haozhe Xie, Hongxun Yao, Shangchen Zhou, Jiageng Mao, Shengping Zhang, and Wenxiu Sun. Grnet: Gridding residual network for dense point cloud completion. In *European Conference on Computer Vision*, pages 365–381. Springer, 2020. 1, 2, 6, 7

[49] Kangxue Yin, Hui Huang, Daniel Cohen-Or, and Hao Zhang. P2p-net: Bidirectional point displacement net for shape

transform. *ACM Transactions on Graphics (TOG)*, 37(4):1–13, 2018. 1

[50] Xumin Yu, Yongming Rao, Ziyi Wang, Zuyan Liu, Jiwen Lu, and Jie Zhou. Pointr: Diverse point cloud completion with geometry-aware transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12498–12507, 2021. 1, 2, 6, 7

[51] Wentao Yuan, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. Pcn: Point completion network. In *3DV*, pages 728–737. IEEE, 2018. 1, 2

[52] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1768–1777, 2021. 1, 2, 3, 5, 6, 7, 8

[53] Wenxiao Zhang, Qingan Yan, and Chunxia Xiao. Detail preserved point cloud completion via separated feature aggregation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 512–528. Springer, 2020. 2