# **KT-Net: Knowledge Transfer for Unpaired 3D Shape Completion**

Zhen Cao<sup>1\*</sup>, Wenxiao Zhang<sup>2\*</sup>, Xin Wen<sup>3†</sup>, Zhen Dong<sup>1†</sup> Yu-Shen Liu<sup>4</sup>, Xiongwu Xiao<sup>1</sup>, Bisheng Yang<sup>1</sup>

<sup>1</sup>Wuhan University <sup>2</sup>Singapore University of Technology and Design <sup>3</sup>JD Logistics <sup>4</sup>Tsinghua University {zhen.cao,dongzhenwhu,xwxiao,bshyang}@whu.edu.cn, wenxxiao.zhang@gmail.com wenxin16@jd.com, liuyushen@tsinghua.edu.cn

### Abstract

Unpaired 3D object completion aims to predict a complete 3D shape from an incomplete input without knowing the correspondence between the complete and incomplete shapes. In this paper, we propose the novel KT-Net to solve this task from the new perspective of knowledge transfer. KT-Net elaborates a teacher-assistant-student network to establish multiple knowledge transfer processes. Specifically, the teacher network takes complete shape as input and learns the knowledge of complete shape. The student network takes the incomplete one as input and restores the corresponding complete shape. And the assistant modules not only help to transfer the knowledge of complete shape from the teacher to the student, but also judge the learning effect of the student network. As a result, KT-Net makes use of a more comprehensive understanding to establish the geometric correspondence between complete and incomplete shapes in a perspective of knowledge transfer, which enables more detailed geometric inference for generating high-quality complete shapes. We conduct comprehensive experiments on several datasets, and the results show that our method outperforms previous methods of unpaired point cloud completion by a large margin. Code is available at https://github.com/a4152684/KT-Net.

# Introduction

Point clouds, as a popular representation form of 3D geometries, have drawn a growing research concern in many fields, such as computer vision (Qi et al. 2017a), robotics (Liu 2015) and auto-navigation (Yue et al. 2018). However, the raw point clouds obtained by 3D scanning devices are often sparse, noisy and incomplete, which requires pre-processing (e.g. completion, denoising) before the downstream tasks. In this paper, we focus on the specific task to predict the complete shape for an incomplete point cloud, where the missing shape is usually caused by the limited scanning view and occlusion. Previous deep learning based studies in this field (Yuan et al. 2018; Liu et al. 2020; Huang et al. 2020; Xie et al. 2020) usually rely on the paired supervision of incomplete shape and its corresponding complete ground truth.

<sup>†</sup>Corresponding author.

However, in spite of the great progress achieved by these supervised methods, the completion performance is still limited by the insufficiency of paired training data, which is caused by the difficulty of obtaining the complete ground truth for incomplete shape in real-world scenarios.

A promising solution to this problem is to train the completion network in an unpaired manner. That is, the network is only fed by collections of complete and incomplete shapes without paired correspondence. Such intuition is driven by the easy access to the incomplete scanning in the real-world, and the large amount of complete shapes from many large-scaled 3D shape datasets (Chang et al. 2015; Pan et al. 2021). However, a typical challenge of unpaired shape completion task is the absence of the one-toone strong supervision for each incomplete shape. Previous methods (e.g. Cycle4Completion (Wen et al. 2021a) and Pcl2Pcl (Chen, Chen, and Mitra 2019)) usually consider the adversarial global feature matching framework as the basic solution to this challenge, which can implicitly bridge the distribution gap between the incomplete and the complete shapes. However, the global feature based adversarial learning framework may lose the detailed geometric information encoded in different levels of the network, which may lead to low completion quality. On the other hand, implicitly aligning the feature can not fully explore the corresponding relationship between incomplete and complete shapes. Another representative work ShapeInversion (Zhang et al. 2021) introduces GAN inversion for shape completion with the help of a well-pretrained generator and a proper latent code. But it takes much time to search for the appropriate latent code during the inference stage, and an inappropriate latent code might cause disaster completion results.

Inspired by the great success of knowledge distillation (Hinton et al. 2015), we propose a new perspective to the unpaired shape completion task, which is how to transfer the knowledge from the complete shape domain into the incomplete shape domain. In this paper, the term "knowl-edge" specifically refers to the common representation in the latent space of the complete shape, which should be distilled to guide the completion process for incomplete input. This idea raises several questions: (1) How to implement a knowledge transfer process? (2) Where should the knowledge transfer take place? (3) How to evaluate the effect of knowledge transfer? (4) How to restore high-fidelity shape

<sup>\*</sup>These authors contributed equally.

Copyright © 2023, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.



Figure 1: An illustration of our starting point. The teacher network is responsible for learning the knowledge of complete shape and transferring it to the student network through the assistant. The student network combines the specific incomplete shape and learns the knowledge of complete shape with the help of the assistants.

### after knowledge transfer?

To solve the above problems, we design a novel Teacher-Assistant-Student framework as shown in Fig. 1, which consists of a series of elaborating knowledge transfer processes. A Knowledge Transfer Implementation: The teacher network takes a complete shape as the input and learns the common knowledge of complete shape by reconstructing it. The student network takes an incomplete shape as input and restores its complete shape with the guidance of the complete shape knowledge. To implement a knowledge transfer, the proposed Knowledge Recovery Assistant (KRA) inputs the incomplete shape features learned by the student network and adopts a residual module to explicitly infer and supplement the missing complete shape knowledge, thus enhancing the incomplete shape features by encoding the complete shape knowledge. Multi-Stage Knowledge Transfer: The knowledge in the different layers encodes irreplaceable geometric information of complete shape. Therefore, we propose a coarse-to-fine knowledge transfer strategy to sufficiently learn the multi-level knowledge of complete shape by iteratively employing the KRA module in different layers. Effectiveness Evaluation of Knowledge Transfer: We propose a mini-GAN based Knowledge Discrimination Assistant (KDA) to evaluate the effectiveness of knowledge transfer. Specifically, the KDA accepts complete shape feature from the teacher network as exemplary knowledge and discriminates against the enhanced features after knowledge transfer from the student network. High-fidelity Complete Shape Restoration: We introduce a training strategy to restore high-fidelity complete shapes. For the teacher network, the gradient shielding technology is adapted to ensure the gradient of incomplete shape does not flow into the teacher network during backpropagation. For the student network, we share the restoration module structure and parameters of the teacher network, which makes it possible to restore highfidelity complete shapes based on enhanced features.

We conduct experiments on the common benchmarks of unpaired shape completion task, including synthetic datasets (3D-EPN dataset (Dai, Ruizhongtai Qi, and Nießner 2017) and CRN dataset (Wang, Ang Jr, and Lee 2020)) and realworld scans (MatterPort3D (Chang et al. 2017), ScanNet (Dai et al. 2017) and KITTI (Geiger, Lenz, and Urtasun 2012)). The results demonstrate that our method outperforms the state-of-the-art (SOTA) unpaired shape completion methods. The contribution of our work mainly includes:

- We formulate the unpaired shape completion task as the knowledge transfer task, and design a novel end-to-end completion network composed of the paralleled teacher and student networks called KT-Net. Under the guidance of complete shape knowledge, the student network can restore a reasonable high-quality point cloud.
- We introduce the Knowledge Recovery Assistant and the Knowledge Discrimination Assistant to aid the student network in explicitly inferring and supplementing the missing complete shape knowledge, thus enhancing the incomplete shape feature by encoding the complete shape knowledge.
- We conduct comprehensive experiments on several datasets and show that our method can achieve SOTA performance over the previous unpaired shape completion methods.

# **Related Work**

**Traditional Shape Completion Methods.** Traditional point cloud completion methods usually utilize the geometric attributes of objects (Hu, Fu, and Guo 2019; Mitra, Guibas, and Pauly 2006) or retrieval of the complete structure in the database (Han and Zhu 2008; Li et al. 2015). These methods usually leverage the artificially constructed features for completion, which cannot robustly generalize to the cases of complex 3D surfaces with large missing parts.

**3D Shape Completion with Pair Supervision.** With the development of the deep neural network, researchers tend to leverage learning-based methods for 3D shape completion. Early methods are based on the 3D voxel grid (Yang et al. 2017; Wang et al. 2017), but they are limited by the computational cost, which increases cubically to the shape resolution. On the other hand, point cloud based methods have emerged a lot in recent years, which is benefited from the pioneering work PointNet (Qi et al. 2017a) of point cloud feature extraction and its subsequent studies (Qi et al. 2017b; Wang et al. 2019). PCN (Yuan et al. 2018) is the first learning-based completion network that can directly operate on point clouds, which uses an encoder-decoder structure to predict complete point clouds. More recent works (Zhang, Yan, and Xiao 2020; Wen et al. 2020; Yu et al. 2021; Huang et al. 2021; Gong et al. 2021; Xie et al. 2021; Wen et al. 2021b; Xiang et al. 2021; Zhang, Huang, and Wang 2021; Xia et al. 2021; Wang et al. 2022a; Long et al. 2022; Wang et al. 2022b; Yan et al. 2022) have made efforts to preserve the observed geometric details from the local features.VRC-Net (Pan et al. 2021) proposes a variational framework by leveraging the relationship between structures during the completion process. LAKe-Net (Tang et al. 2022) proposes a novel topology-aware point cloud completion model by localizing aligned keypoints, with a



Figure 2: The overall architecture of KT-Net, which consists of the paralleled teacher and student networks. (a) The above is the student network, which converts an incomplete shape into a complete one. (b) And the below is the teacher network, which reconstructs the complete shape and transfers the knowledge to the student network. (c) The Knowledge Recovery Assistant (KRA) in the student network is adopted to supplement the missing complete shape knowledge to adapt to the task of point cloud completion. (d) The Knowledge Discrimination Assistant (KDA) is responsible for identifying which features come from the teacher network.

**Overview** 

novel Keypoints-Skeleton-Shape prediction manner. Seed-Former (Zhou et al. 2022) introduces a new shape representation, which not only captures general structures from partial inputs but also preserves regional information of local patterns. There are also some networks using a voxelbased completion process. GRNet (Xie et al. 2020) proposes a gridding network for dense point reconstruction. VE-PCN (Wang, Ang, and Lee 2021) develops a voxel-based network for point cloud completion by leveraging edge generation.

3D Shape Completion without Pair Supervision. On the other hand, a few studies on unpaired shape completion has been proposed. AML (Stutz and Geiger 2018) uses the maximum likelihood method to measure the distance between complete and incomplete point clouds in the latent space. Pcl2Pcl (Chen, Chen, and Mitra 2019) pretrains two auto-encoders, and directly learns the mapping from partial shapes to the complete ones in the latent space. Wu et al. (Wu et al. 2020) refer to the method of VAE to achieve multimodal outputs. MM-Flow (Zhao et al. 2021) proposes a flow-based network together with a multi-modal mapping strategy for 3D point cloud completion. Cycle4Completion (Wen et al. 2021a) designs two cycle transformations to establish the geometric correspondence between incomplete and complete shapes in both directions. ShapeInversion (Zhang et al. 2021) incorporates a well-trained GAN as an effective prior for shape completion. Cai et al. (Cai et al. 2022) establish a unified and structured latent space to achieve partial-complete geometry consistency.

Our KT-Net intends to transfer the knowledge from complete shape to incomplete shape based on two paralleled teacher and student networks. Fig. 2 shows the overall framework of our method.  $X_{in}$  and  $Y_{in}$  denote the incomplete and the complete point clouds, respectively. Note that for unpaired completion, there is no correspondence between  $Y_{in}$  and  $X_{in}$ . The teacher network takes  $Y_{in}$  as input and learns the knowledge of complete shape by reconstructing the same complete shape  $Y_c$  as  $Y_{in}$ . The student network takes  $X_{in}$  as input and expects to restore the complete prediction  $X_c$ . Feature extraction module is responsible for extracting the global features  $f_x$  and  $f_y$  from the input point cloud  $X_{in}$  and  $Y_{in}$ . The knowledge transfer module accepts the features extracted from both networks, and then guides the student network to explicitly infer and supplement the missing complete shape knowledge with the help of our proposed assistants, thus enhancing the incomplete shape features by encoding the complete shape knowledge. Finally, a unified MLP is used to convert the enhanced features into complete shapes.

Method

# **Feature Extraction Module**

With a given point set  $P \in \mathbb{R}^{N \times 3}$ , our encoder extracts a k-dimensional global feature  $f \in \mathbb{R}^k$ . Similar to PCN, the encoder consists of MLP layers for point-wise feature extraction and finally performs maxpooling operation to obtain

the global feature. For both incomplete and complete shapes, we leverage the same encoder architecture and parameters to get their global representations, which guarantees that the extracted features from each network lie in the same latent space for subsequent knowledge transfer.

### **Knowledge Transfer Module**

This module takes the k-dimensional latent features  $f_x$  and  $f_y$  from the feature extraction module as input, which can promote the student network to learn the complete shape knowledge from the teacher network.

**KRA for the Knowledge Transfer.** For the teacher network, the feature  $f_y^i$  of the *i*-th layer contains the knowledge of complete shape, which are gradually pushed from high dimensional latent space to Euclidean space. For the student network, the KRA is introduced to enhance the incomplete shape features by explicitly learning and encoding the complete shape knowledge. In our design, the knowledge of the missing region is inferred and supplemented through residual connection. As shown in Fig. 2, we select a simple but effective model to accomplish this work.

$$KRA(x) = FC \circ ReLU \circ FC(x), \tag{1}$$

where FC means a fully connected layer and ReLU is an activation function.

The knowledge in each stage is irreplaceable, and the shallow knowledge is assumed to be more specific than the deep one. To enhance 3D shape understanding ability, we propose a multi-stage transfer strategy by iteratively employing the KRA module in different layers, which contributes to the mastery of more comprehensive complete shape knowledge for the student network via a continuous learning procedure.

KDA for the Effect Evaluation of Knowledge Transfer. An additional assistant KDA that is applied in each layer is adopted to receive the complete shape knowledge from the teacher network, and evaluate the learning effectiveness of the student on this basis. However, in the unpaired task, it is unreasonable to directly transfer the unprocessed complete shape knowledge to the incomplete shape of the same batch due to the missing one-to-one correspondence. Considering this, and in order to avoid over complex knowledge transfer design, the KDA employs the generative adversarial architecture to make the student network automatically obtain the knowledge from the teacher network. Specifically, the KDA regards the feature from the teacher network as the representative of the complete shape knowledge and assigns it to True, while the enhanced feature from the student network is assigned to False. Under the guidance of the KDA, the student network strives to make its enhanced feature mimic the one from the teacher network. The complete shape knowledge has been transferred from the teacher network to the student network when the KDA can hardly distinguish the two types of features. The discrimination loss can be defined as follows:

$$\mathcal{L}_D = \frac{1}{2n} \sum \left\{ E_x [D_i(f_x^i)]^2 + E_y [D_i(f_y^i) - 1]^2 \right\}.$$
 (2)

where  $D_i$  represents the *i*-th KDA, *E* denotes the mathematical expectation,  $f_x^i$  represents the feature of the student

network at the *i*-th layer of the knowledge transfer module,  $f_y^i$  is the feature of the teacher network at the *i*-th layer and *n* is the number of layers of the knowledge transfer module. On the other hand, the generated loss used to evaluate the student network's learning achievements is defined as:

$$\mathcal{L}_G = \frac{1}{n} \sum E_x [D_i(\boldsymbol{f}_x^i) - 1]^2.$$
(3)

# **Point Cloud Restoration**

To restore the complete shapes from the features, an MLP is applied to map the feature f from  $R^k$  to  $R^{3N}$ , where N is the number of generated points. Then the feature is reshaped from  $R^{3N}$  to  $R^{N\times3}$  to get the point cloud P.

The teacher network follows an auto-encoder architecture for self-reconstruction. After obtaining the reconstructed complete shape  $Y_c$ , we leverage a common permutationinvariant metrics Earth Mover's Distance (EMD) to supervise the training, which is shown below.

$$\mathcal{L}_{teacher} = \mathcal{L}_{EMD}(Y_{in}, Y_c). \tag{4}$$

For the student network, the enhanced feature is mapped to the complete shape  $X_c$  by the shared MLP in the teacher network. Nevertheless, there is no corresponding ground truth for the predicted complete shape  $X_c$ , which indicates a direct implementation of the bi-direction EMD may not yield desirable results. Therefore, in order to prevent the mode collapse and preserve the input information, we apply Unidirectional Chamfer Distance(UCD) and the loss of the student network is expressed as:

$$\mathcal{L}_{student} = \mathcal{L}_{UCD}(X_{in}, X_c). \tag{5}$$

# **Training Strategy**

Since the teacher and student networks share the restoration module, the quality of the student network's completion results depends on the teacher network's ability to reconstruct complete shapes. To enable the teacher network to reconstruct a high-quality complete shape, we adopt a special training strategy, which ensures the gradient of incomplete shape does not flow into the teacher network at the knowledge transfer module during backpropagation.

We use FE and KRA to represent the feature extraction module and the knowledge recovery assistants in the knowledge transfer module, respectively. In particular, KTdenotes the remaining structure of the knowledge transfer module.  $\theta_D$  is used to represent the trainable parameters in  $\{D_i|i = 1, ..., k\}$  (i.e. all the knowledge discrimination assistants).  $\theta_{teacher}$  is adopted to stand for the trainable parameters in  $\{FE, KT\}$ , while  $\theta_{student}$  denotes the trainable parameters in  $\{FE, KRA\}$ .

For a given learning rate  $\gamma$ , we first update  $\theta_D$  of all the discriminators and freeze all other parts, which is shown as:

$$\theta_D \leftarrow \theta_D - 2\gamma \frac{\partial \mathcal{L}_D}{\partial \theta_D}.$$
 (6)

|          | Methods          | Average | Plane | Cabinet | Car  | Chair | Lamp | Sofa | Table | Boat |
|----------|------------------|---------|-------|---------|------|-------|------|------|-------|------|
| paired   | PCN              | 7.6     | 2.0   | 8.0     | 5.0  | 9.0   | 13.0 | 8.0  | 10.0  | 6.0  |
| unpaired | Pcl2Pcl          | 17.4    | 4.0   | 19.0    | 10.0 | 20.0  | 23.0 | 26.0 | 26.0  | 11.0 |
|          | Cycle4Completion | 14.3    | 3.7   | 12.6    | 8.1  | 14.6  | 18.2 | 26.2 | 22.5  | 8.7  |
|          | Ours             | 10.0    | 2.6   | 10.9    | 6.3  | 12.4  | 15.1 | 10.5 | 15.8  | 6.5  |

Table 1: Shape completion results on 3D-EPN dataset. The numbers shown are CD (lower is better), which is scaled by  $10^4$ . KT-Net outperforms other unpaired methods by a large margin, and is comparable to the paired method.

|          | Methods        | Average | Plane | Cabinet | Car  | Chair | Lamp | Sofa | Table | Boat |
|----------|----------------|---------|-------|---------|------|-------|------|------|-------|------|
| paired   | PCN            | 9.1     | 3.5   | 11.3    | 6.4  | 11.0  | 11.6 | 11.5 | 10.4  | 7.4  |
| unpaired | Pcl2Pcl        | 22.4    | 9.8   | 27.1    | 15.8 | 26.9  | 25.7 | 34.1 | 23.6  | 15.7 |
|          | ShapeInversion | 14.9    | 5.6   | 16.1    | 13.0 | 15.4  | 18.0 | 24.6 | 16.2  | 10.1 |
|          | Ours           | 11.5    | 3.8   | 12.4    | 9.0  | 13.9  | 14.1 | 18.1 | 11.7  | 8.7  |

Table 2: Shape completion results on CRN dataset. The numbers shown are CD (lower is better), which is scaled by  $10^4$ . KT-Net outperforms other unpaired methods by a large margin, and is comparable to the paired method.

Then we use  $\mathcal{L}_{teacher}$  to regularize the teacher network. The gradient descent step is shown below.

$$\theta_{teacher} \leftarrow \theta_{teacher} - \gamma \frac{\partial \mathcal{L}_{teacher}}{\partial \theta_{teacher}}.$$
 (7)

To prevent the teacher network from generating bad reconstructed complete shapes,  $\mathcal{L}_{student}$  and  $\mathcal{L}_G$  are only adopted to update  $\theta_{student}$ . Such a strategy enables the teacher network only to learn the complete shape knowledge, thus reconstructing a high-quality point cloud.

$$\theta_{student} \leftarrow \theta_{student} - \gamma [\lambda_g \frac{\partial \mathcal{L}_G}{\partial \theta_{student}} + \lambda_p \frac{\partial \mathcal{L}_{student}}{\partial \theta_{student}}], (8)$$

where  $\lambda_g$  and  $\lambda_p$  are hyper-parameters and we set  $\lambda_g = 0.1$ and  $\lambda_p = 1$ .

# Experiments

# **Datasets and Evaluation Metrics**

We evaluate our network on two systhetic datasets 3D-EPN (Dai, Ruizhongtai Qi, and Nießner 2017) and CRN (Wang, Ang Jr, and Lee 2020), and three real-word datasets Matter-Port3D (Chang et al. 2017), ScanNet (Dai et al. 2017) and KITTI (Geiger, Lenz, and Urtasun 2012).

For evaluation on synthetic datasets, we use Chamfer Distance(CD) to compare the difference between the predicted point cloud and the ground truth. For the realworld datasets without ground truth, we follow ShapeInversion(Zhang et al. 2021) to use Unidirectional Hausdorff Distance(UHD) to evaluate the similarity between the input point cloud and the predicted point cloud. More details are shown in Appendix.

## **Evaluation on Synthetic Dataset**

In this section, we follow the train/test split of the previous unpaired shape completion methods (i.e., Pcl2Pcl and Cycle4Completion) for a fair comparision. Note that there is no correspondence between incomplete and complete shapes during training. The quantitative results are shown in Tab. 1. The experimental results show that our KT-Net achieves the best results over the counterpart unpaired completion methods. Especially, compared with the second-best method Cycle4Completion, KT-Net reduces the CD loss with a nontrivial margin by 30%, and achieves the best results in terms of CD loss over all categories. KT-Net's achieves better performance over Cycle4Completion mainly comes from the more detailed geometric recovery ability of the proposed two assistant modules. In Fig. 3, we visually compare the performance of KT-Net with the other unpaired completion methods. We can find that KT-Net can predict more accurate and detailed shapes in the missing regions, and generate a better completion quality. For example, in the fifth column of Fig. 3, KT-Net successfully predicts the correct shape of table legs, while the other two methods (i.e. Pcl2Pcl and Cycle4Completion) failed to reconstruct the same legs as ground truth.

To verify the effectiveness of our network on different datasets and make a fair comparison with ShapeInversion, we also conducted an evaluation on CRN dataset. The quantitative results are shown in Tab. 2. KT-Net outperforms the second-best unpaired method ShapeInversion with a large margin by more than 20%, and achieves the best results in all the classes, which proves the superiority of our method. The visual comparison with other unpaired shape completion methods on CRN dataset is shown in Appendix.

### **Generalization Performances to Real-World Scans**

In this section, we evaluate our network trained on CRN datasets on three real-world dataset, i.e., MatterPort3D, ScanNet and KITTI without fine-tuning. Quantitative and qualitative results shown in Tab. 3 and Fig. 4, respectively. Our network can predict a reasonable complete point cloud even on cases of severe sparseness and incompleteness (such as cars in KITTI). Quantitative results show that our method can well preserve the input information rather than predict an average shape of this category only. For example, in the first line of Fig. 4, KT-Net predicts reasonable armrest and legs of chair, while the results of other methods may be



Figure 3: Visual comparison with the state-of-the-art unpaired shape completion methods on 3D-EPN dataset.

inconsistent. In particular, Pcl2Pcl tends to output a dense complete shape, which doesn't often represent the same object as partial inputs because of the stronger shape prior.



Figure 4: Shape completion on real-world partial scans.

# **Analysis of Different Training Strategies**

In our model, the student loss  $\mathcal{L}_{student}$  and the adversarial loss  $\mathcal{L}_G$  are used to update  $\theta_{student}$  only. However,  $\mathcal{L}_{student}$  and  $\mathcal{L}_G$  involve not only the KRA but also the remaining part of the knowledge transfer module. To evaluate the effectiveness of other potential training strategies, similar to Training Strategy subsection, we use  $\theta_{all}$  to denote the trainable

| Methods        | ScanNet |       | Matter | KITTI |      |
|----------------|---------|-------|--------|-------|------|
| Wiethous       | Chair   | Table | Chair  | Table | Car  |
| Pcl2Pcl        | 10.1    | 11.8  | 10.5   | 11.8  | 14.1 |
| ShapeInversion | 10.1    | 11.9  | 10.0   | 11.8  | 13.8 |
| Ours           | 6.4     | 6.3   | 8.4    | 10.0  | 6.0  |

Table 3: Shape completion results on the real scans. The numbers shown are [UHD](lower is better), where UHD is scaled by  $10^2$ .

parameters in  $\{FE, KT, KRA\}$ . Then we develop the variations of (a)  $\partial \mathcal{L}_{student} / \partial \theta_{all}$  and (b)  $\partial \mathcal{L}_G / \partial \theta_{all}$  to replace  $\partial \mathcal{L}_{student} / \partial \theta_{student}$  or  $\partial \mathcal{L}_G / \partial \theta_{student}$  in Eq. 8.

The results are reported in Tab. 4. We observed dropping performances in the variations of (a) and (b), where the teacher network is affected by the incomplete shape knowledge, which leads to a failure to reconstruct the complete shape input  $Y_{in}$ .

| Methods   | Average | Chair | Lamp | Sofa | Table |  |
|---|---------|-------|------|------|-------|--|
| $\partial \mathcal{L}_{com}/\partial 	heta_{all}$     | 26.2    | 22.3  | 27.3 | 30.1 | 25.0  |  |
| $\partial \mathcal{L}_{match} / \partial 	heta_{all}$ | 24.5    | 22.0  | 26.1 | 32.3 | 17.5  |  |
| Original Model  | 14.5    | 13.9  | 14.1 | 18.1 | 11.7  |  |

Table 4: Results of different training strategies  $(CD \times 10^4)$ .



(a) Visualization of latent space in our model.

(b) Visualization of latent space in Pcl2Pcl.

Figure 5: Visualization of latent space. (a) shows the results of our model, (b) shows the results of Pcl2Pcl. We further visualize the shape that these points represent.

### Visual Analysis of Latent Feature Distribution

In Fig. 5, we use t-SNE (Van der Maaten and Hinton 2008) to visualize the features before and after the knowledge transfer module in the high-dimensional latent space as 2D points. The distance of points can reflect the similarity of the corresponding latent features to a certain extent. The red and blue dots represent the features from complete and incomplete shapes before knowledge transfer module, respectively. The green dots represent the transferred features of the incomplete shape after the knowledge transfer module.

As shown in Fig. 5(a), the features from the teacher and student networks before the knowledge transfer module are not exactly the same, indicating that the correspondences in the incomplete shape and the complete shape have not been fully explored. After our knowledge transfer module, the features could be similar enough to adapt to the task of point cloud completion. We also visualize the features of Pcl2Pcl on the same objects, as shown in Fig. 5(b). Due to the stronger shape prior, Pcl2Pcl tends to output a dense complete shape without consistency, resulting in the fact that its transformed features are still far away from those of the corresponding ground truth, which is sometimes misleading. Furthermore, we visualize the shape represented by these points, including the incomplete inputs, ground truths and predicted complete shapes.

### **More Ablation Studies**

We further analyze each part of our network with results shown in Tab. 5, which is performed on CRN dataset.

**Recovery Assistant Evaluation.** To evaluate the effectiveness of the components in our framework, we give a study about removing the KRA. In this case, we retrained the network for evaluation. The results are shown as *w/o KRA* in Tab. 5 and the network performance decreases without KRA. **Discrimination Assistant Evaluation.** The KDA is used to transfer the complete shape knowledge from the teacher network to the student network and judge the learning effectiveness of the student network. In order to evaluate the necessity of introducing the KDA, we retrained the network without using it and the results are shown as *w/o KDA* in Tab. 5. The student network without the guidance of the KDA cannot learn the knowledge of complete shape, which leads to poor completion results.

**Residual Connection Evaluation.** In KT-Net, we use the residual connection to reduce the learning pressure of the student network. We try to test the performance of the network without residual connection. The result is shown as *w/o Residual*, which shows that residual connection can significantly improve the ability of the network.

**Evaluation of**  $\mathcal{L}_{student}$ .  $\mathcal{L}_{student}$  (shown in Eq. 5) uses incomplete shape input  $X_{in}$  to constrain the the complete shape output  $X_c$ , which makes  $X_{in}$  and  $X_c$  represent the same object. We try to train the network without  $\mathcal{L}_{student}$ and the result is shown as  $w/o \mathcal{L}_{student}$ .

| Methods                     | Average | Chair | Lamp | Sofa | Table |
|-----------------------------|---------|-------|------|------|-------|
| w/o KRA                     | 21.9    | 20.4  | 19.1 | 30.5 | 17.6  |
| w/o KDA                     | 24.9    | 17.3  | 18.0 | 51.0 | 13.3  |
| w/o Residual                | 17.9    | 17.6  | 17.4 | 20.2 | 16.5  |
| w/o $\mathcal{L}_{student}$ | 26.2    | 22.3  | 27.3 | 30.1 | 25.0  |
| Original Model              | 14.5    | 13.9  | 14.1 | 18.1 | 11.7  |

Table 5: More Analysis  $(CD \times 10^4)$ .

# Conclusion

We propose the KT-Net for unpaired point cloud completion task. Compared with the previous methods, our method pays more attention to the process of knowledge transfer, and uses a teacher-student network combined with ingenious design to enhance the student network's understanding of shape. Furthermore, our model is evaluated on several widely used datasets, achieving state-of-the-art performance by a large margin compared with other unpaired completion methods. To further improve the performance, we will develop a refine module that is suitable for unpaired shape completion.

# Acknowledgements

This research is jointly sponsored by the National Natural Science Foundation of China Projects (No. 42171431), Autonomous Driving Division of X Research Department, JD Logistics and Hunan Provincial Key Laboratory of Geo-Information Engineering in Surveying, Mapping and Remote Sensing, Hunan University of Science and Technology.

# References

Cai, Y.; Lin, K.-Y.; Zhang, C.; Wang, Q.; Wang, X.; and Li, H. 2022. Learning a Structured Latent Space for Unsupervised Point Cloud Completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 

Chang, A.; Dai, A.; Funkhouser, T.; Halber, M.; Niessner, M.; Savva, M.; Song, S.; Zeng, A.; and Zhang, Y. 2017. Matterport3d: Learning from rgb-d data in indoor environments. *arXiv preprint arXiv:1709.06158*.

Chang, A. X.; Funkhouser, T.; Guibas, L.; Hanrahan, P.; Huang, Q.; Li, Z.; Savarese, S.; Savva, M.; Song, S.; Su, H.; et al. 2015. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*.

Chen, X.; Chen, B.; and Mitra, N. J. 2019. Unpaired Point Cloud Completion on Real Scans using Adversarial Training. In *International Conference on Learning Representations*.

Dai, A.; Chang, A. X.; Savva, M.; Halber, M.; Funkhouser, T.; and Nießner, M. 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5828–5839.

Dai, A.; Ruizhongtai Qi, C.; and Nießner, M. 2017. Shape completion using 3d-encoder-predictor cnns and shape synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5868–5877.

Geiger, A.; Lenz, P.; and Urtasun, R. 2012. Are we ready for autonomous driving? the kitti vision benchmark suite. In 2012 IEEE conference on computer vision and pattern recognition, 3354–3361. IEEE.

Gong, B.; Nie, Y.; Lin, Y.; Han, X.; and Yu, Y. 2021. ME-PCN: Point Completion Conditioned on Mask Emptiness. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12488–12497.

Han, F.; and Zhu, S.-C. 2008. Bottom-up/top-down image parsing with attribute grammar. *IEEE transactions on pattern analysis and machine intelligence*, 31(1): 59–73.

Hinton, G.; Vinyals, O.; Dean, J.; et al. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2(7).

Hu, W.; Fu, Z.; and Guo, Z. 2019. Local frequency interpretation and non-local self-similarity on graph for point cloud inpainting. *IEEE Transactions on Image Processing*, 28(8): 4087–4100.

Huang, T.; Zou, H.; Cui, J.; Yang, X.; Wang, M.; Zhao, X.; Zhang, J.; Yuan, Y.; Xu, Y.; and Liu, Y. 2021. RFNet: Recurrent Forward Network for Dense Point Cloud Completion. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 12508–12517.

Huang, Z.; Yu, Y.; Xu, J.; Ni, F.; and Le, X. 2020. Pf-net: Point fractal network for 3d point cloud completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7662–7670.

Li, Y.; Dai, A.; Guibas, L.; and Nießner, M. 2015. Databaseassisted object retrieval for real-time 3d reconstruction. In *Computer graphics forum*, volume 34, 435–446. Wiley Online Library.

Liu, M. 2015. Robotic online path planning on point cloud. *IEEE transactions on cybernetics*, 46(5): 1217–1228.

Liu, M.; Sheng, L.; Yang, S.; Shao, J.; and Hu, S.-M. 2020. Morphing and sampling network for dense point cloud completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 11596–11603.

Long, C.; Zhang, W.; Li, R.; Wang, H.; Dong, Z.; and Yang, B. 2022. PC2-PU: Patch Correlation and Point Correlation for Effective Point Cloud Upsampling. In *Proceedings of the 30th ACM International Conference on Multimedia*, 2191–2201.

Mitra, N. J.; Guibas, L. J.; and Pauly, M. 2006. Partial and approximate symmetry detection for 3d geometry. *ACM Transactions on Graphics (TOG)*, 25(3): 560–568.

Pan, L.; Chen, X.; Cai, Z.; Zhang, J.; Zhao, H.; Yi, S.; and Liu, Z. 2021. Variational Relational Point Completion Network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8524–8533.

Qi, C. R.; Su, H.; Mo, K.; and Guibas, L. J. 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 652–660.

Qi, C. R.; Yi, L.; Su, H.; and Guibas, L. J. 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30.

Stutz, D.; and Geiger, A. 2018. Learning 3d shape completion from laser scan data with weak supervision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1955–1964.

Tang, J.; Gong, Z.; Yi, R.; Xie, Y.; and Ma, L. 2022. LAKe-Net: Topology-Aware Point Cloud Completion by Localizing Aligned Keypoints. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), 1726–1735.

Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).

Wang, H.; Liu, Y.; Dong, Z.; and Wang, W. 2022a. You only hypothesize once: Point cloud registration with rotation-equivariant descriptors. In *Proceedings of the 30th ACM International Conference on Multimedia*, 1630–1641.

Wang, W.; Huang, Q.; You, S.; Yang, C.; and Neumann, U. 2017. Shape inpainting using 3d generative adversarial network and recurrent convolutional networks. In *Proceedings* of the IEEE international conference on computer vision, 2298–2306.

Wang, X.; Ang, M. H.; and Lee, G. H. 2021. Voxel-based Network for Shape Completion by Leveraging Edge Generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 13189–13198.

Wang, X.; Ang Jr, M. H.; and Lee, G. H. 2020. Cascaded refinement network for point cloud completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 790–799.

Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S. E.; Bronstein, M. M.; and Solomon, J. M. 2019. Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*, 38(5): 1–12.

Wang, Y.; Tan, D. J.; Navab, N.; and Tombari, F. 2022b. Learning Local Displacements for Point Cloud Completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1568–1577.

Wen, X.; Han, Z.; Cao, Y.-P.; Wan, P.; Zheng, W.; and Liu, Y.-S. 2021a. 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*, 13080–13089.

Wen, X.; Li, T.; Han, Z.; and Liu, Y.-S. 2020. Point cloud completion by skip-attention network with hierarchical folding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1939–1948.

Wen, X.; Xiang, P.; Han, Z.; Cao, Y.-P.; Wan, P.; Zheng, W.; and Liu, Y.-S. 2021b. Pmp-net: Point cloud completion by learning multi-step point moving paths. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7443–7452.

Wu, R.; Chen, X.; Zhuang, Y.; and Chen, B. 2020. 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,* 281–296. Springer.

Xia, Y.; Xia, Y.; Li, W.; Song, R.; Cao, K.; and Stilla, U. 2021. ASFM-Net: Asymmetrical Siamese Feature Matching Network for Point Completion. *arXiv preprint arXiv:2104.09587*.

Xiang, P.; Wen, X.; Liu, Y.-S.; Cao, Y.-P.; Wan, P.; Zheng, W.; and Han, Z. 2021. Snowflakenet: Point cloud completion by snowflake point deconvolution with skip-transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5499–5509.

Xie, C.; Wang, C.; Zhang, B.; Yang, H.; Chen, D.; and Wen, F. 2021. Style-based Point Generator with Adversarial Rendering for Point Cloud Completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4619–4628.

Xie, H.; Yao, H.; Zhou, S.; Mao, J.; Zhang, S.; and Sun, W. 2020. Grnet: Gridding residual network for dense point cloud completion. In *European Conference on Computer Vision*, 365–381. Springer.

Yan, X.; Lin, L.; Mitra, N. J.; Lischinski, D.; Cohen-Or, D.; and Huang, H. 2022. ShapeFormer: Transformer-Based

Shape Completion via Sparse Representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6239–6249.

Yang, B.; Wen, H.; Wang, S.; Clark, R.; Markham, A.; and Trigoni, N. 2017. 3d object reconstruction from a single depth view with adversarial learning. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 679–688.

Yu, X.; Rao, Y.; Wang, Z.; Liu, Z.; Lu, J.; and Zhou, J. 2021. Pointr: Diverse point cloud completion with geometryaware transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12498–12507.

Yuan, W.; Khot, T.; Held, D.; Mertz, C.; and Hebert, M. 2018. Pcn: Point completion network. In *2018 International Conference on 3D Vision (3DV)*, 728–737. IEEE.

Yue, X.; Wu, B.; Seshia, S. A.; Keutzer, K.; and Sangiovanni-Vincentelli, A. L. 2018. A lidar point cloud generator: from a virtual world to autonomous driving. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, 458–464.

Zhang, J.; Chen, X.; Cai, Z.; Pan, L.; Zhao, H.; Yi, S.; Yeo, C. K.; Dai, B.; and Loy, C. C. 2021. Unsupervised 3D Shape Completion through GAN Inversion. In *Proceedings* of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1768–1777.

Zhang, W.; Yan, Q.; and Xiao, C. 2020. 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, 512–528. Springer.

Zhang, Y.; Huang, D.; and Wang, Y. 2021. PC-RGNN: Point Cloud Completion and Graph Neural Network for 3D Object Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 3430–3437.

Zhao, Y.; Zhou, Y.; Chen, R.; Hu, B.; and Ai, X. 2021. MM-Flow: Multi-modal Flow Network for Point Cloud Completion. In *Proceedings of the 29th ACM International Conference on Multimedia*, 3266–3274.

Zhou, H.; Cao, Y.; Chu, W.; Zhu, J.; Lu, T.; Tai, Y.; and Wang, C. 2022. SeedFormer: Patch Seeds based Point Cloud Completion with Upsample Transformer. *arXiv preprint arXiv:2207.10315*.