

Point Set Self-Embedding

Ruihui Li, Xianzhi Li, Tien-Tsin Wong, and Chi-Wing Fu

Abstract—This work presents an innovative method for point set self-embedding, that encodes the structural information of a dense point set into its sparser version in a visual but imperceptible form. The self-embedded point set can function as the ordinary downsampled one and be visualized efficiently on mobile devices. Particularly, we can leverage the self-embedded information to fully restore the original point set for detailed analysis on remote servers. This task is challenging, since both the self-embedded point set and the restored point set should resemble the original one. To achieve a learnable self-embedding scheme, we design a novel framework with two jointly-trained networks: one to encode the input point set into its self-embedded sparse point set and the other to leverage the embedded information for inverting the original point set back. Further, we develop a pair of up-shuffle and down-shuffle units in the two networks, and formulate loss terms to encourage the shape similarity and point distribution in the results. Extensive qualitative and quantitative results demonstrate the effectiveness of our method on both synthetic and real-scanned datasets. The source code and trained models will be publicly available at <https://github.com/liruihui/Self-Embedding>.

Index Terms—Point set self-embedding, jointly-trained networks, shape similarity, point distribution.

1 INTRODUCTION

Point clouds become increasingly accessible in various mobile devices, due to the popularity of 3D scanning sensors. To fit the low-profile devices, e.g., VR headset and mobile phone, the captured point sets are usually downsampled but still maintaining visual recognizability for real-time graphics rendering and user interaction. Later, when the downsampled point set is transferred to the connected hosts or remote servers for further analysis, a post-upsampling operation is followed to restore the original details. Thus, it is desirable to develop an effective downsampling & upsampling pipeline to make such application more practical, e.g., video streaming and analytics [3], [4], [5].

However, existing downsampling techniques [6], [7], [8], [9], [10], [11] typically select a representative subset from the input and drop all the remaining points, so the fine structures represented by the dropped points may unavoidably be lost. Even leveraging the state-of-the-art upsampling methods [1], [2], [12], [13], [14], [15], [16], precisely inferring the dropped points is still very challenging, particularly for sparse areas. Figures 1(e)&(f) show the obvious deviations in fine structures of upsampled points from two state-of-the-art methods, compared to the original one (Figure 1(a)).

We then raise a thought - Can we embed the original structural information of a point set into its sparse version, so that the self-embedded information can be leveraged for a better restoration in the later upsampling process? We call this brand new task as *point set self-embedding*. To achieve a learnable self-embedding scheme, in this paper, we design a new framework, consisting of (i) a self-embedding network to encode the input point set into its self-embedded sparse version, and (ii) a restoration network to leverage the

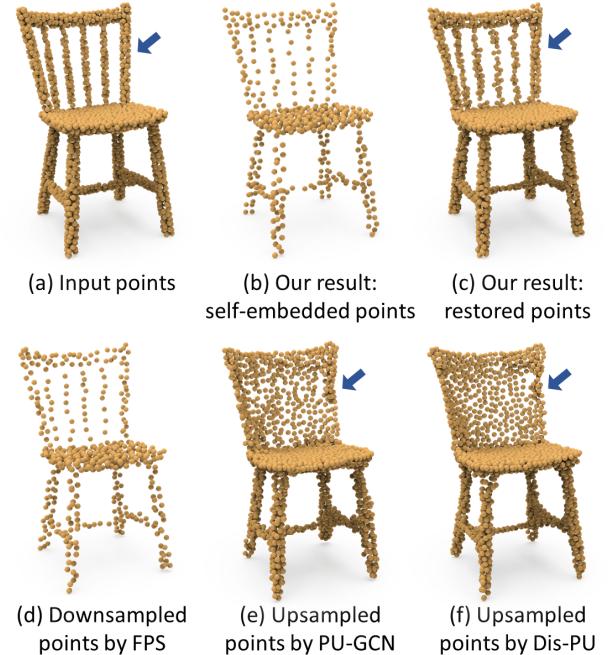


Fig. 1: Given the original input point set (a), our method creates downsampled points (b) with self-embedded information that can later be consumed to restore (c) that looks similar to the original one. As a contrast, given the downsampled points (d) via farthest-point sampling (FPS) from (a), the upsampled points by PU-GCN [1] (e) and Dis-PU [2] (f) retain excessive noise, especially on the Chair back. Though our learned self-embedded point set (b) looks not much different from (d) visually, the embedded information in (b) is quite helpful for a high-quality restoration (c) compared with the SOTA upsampling methods (e&f).

• R.Li is with Hunan University. E-mail: liruihui@hnu.edu.cn.
X.Li is with Huazhong University of Science and Technology. E-mail: xzli@hust.edu.cn.
T.-T. Wong and C.-W. Fu are with the Chinese University of Hong Kong. E-mail: {ttwong, cwf} @cse.cuhk.edu.hk

embedded information to invert the original point set back. This formulation enables a self-supervised learning without the need of preparing labeled training data. Having said that, point set self-embedding goes beyond the conventional downsampling process and aims to create self-embedded point sets that not just look like the original ones but are also restorable to produce dense point sets that are similar to the originals. This also means that we change point cloud upsampling from an inference nature to a restoration nature.

Achieving such restorable self-embedding is more challenging than the conventional point cloud downsampling task. First, to self-embed a point set, we cannot simply select a subset of points. According to the Nyquist-Shannon sampling theorem [17], it is inevitable that geometric information is likely lost after the downsampling. Our goal is to reduce the “lost” information as much as possible for an accurate restoration. Second, a well-restored point set should be consistent to the original input in terms of both the global shape structure and local point distribution.

To meet these goals, we formulate a residual-learning-like approach to first create an initially-downsampled point set that looks like the input. Then, in the self-embedding network, we design the down-shuffle unit to learn to generate small offset vectors that represent the missing structural information. These offsets are added to the pre-downsampled point set to form the final self-embedded point set. On the other hand, we design the up-shuffle unit in the restoration network to learn to recover the original information. Using this approach, keeping small offset vectors in the self-embeddings ensures the similarity between our self-embedded point set and the ordinary downsampled one for a better visualization. So, the network training can focus on the information embedding by optimizing a restoration objective, in which we design losses to encourage the restored point set close to the original one in terms of shape similarity and point distribution as the original one.

To sum up, our point set self-embedding aims for both good visualization and shape restoration simultaneously. Figure 1(b) shows the self-embedded sparse points from (a), which is uniformly distributed and visually recognizable as the original one. Also, the restored dense point set in Figure 1(c) from (b) better conforms to the original one, when compared with the dense point sets (e) and (f) from the downsampled points (d), validating the advantage of our self-embedded point set. The main objective of this paper is to *self-embed structural information into the downsampled point set*, such that we effectively turn the information-losing downsampling process into an information-embedding process. Also, the self-embedded information can be helpful for recovering the original input. More extensive experimental results on both synthetic and real-scanned inputs demonstrate the effectiveness of our self-embedding method.

2 RELATED WORK

As far as we know, there seems no other research shares the same spirit as ours. Hence, we mainly discuss the related works on point cloud downsampling and upsampling. We also discuss recent steganography-related methods.

Point cloud downsampling. Traditional methods downsample point sets mostly rely on handcrafted rules that

are geometry- or random-based. Geometry-based methods [9], [10], [11], [18], [19], [20], [21], [22], [23] explore shape characteristics to drop or create points. However, these methods typically require expensive computation for shapes with complex structures. Random-based methods [6], [7], [8], *e.g.*, farthest point sampling and Poisson disk sampling, iteratively subsample a point set with certain randomness, while avoiding points that are too close to aim for a more uniform coverage. However, these methods generally focus on preserving the overall shape but not on considering the local structures and the downstream tasks.

Recently, learning-based approaches [24], [25], [26] were proposed to select a representative subset guided by a pre-trained task network. They aim to reduce the performance drop when applying the selected subset for a few specific downstream tasks, *e.g.*, classification and reconstruction. However, they do not aim at preserving the dropped points, which may be useful for other previously unconsidered downstream tasks. In contrast, we aim at restoring the original shapes and local details, by self-embedding the input structure into its sparse counterpart, so that future unforeseeable downstream tasks can still be performed. In addition, their selected subsets may not act as a visually-pleasing preview of the original geometry (see the visual results in Figure 7), we aim at producing a downsampled version for both better visualizations and shape restoration simultaneously. Lastly, instead of using a pre-trained task network in these methods, we jointly optimize the downsampling and upsampling networks, and carefully design the framework modules and loss functions to achieve effective information embedding.

Point cloud upsampling. Rather than using shape priors to constrain the point generation [27], [28], [29], [30], [31], recent deep-learning-based methods synthesize points directly in the feature space. Yu *et al.* [12] propose PU-Net to upsample points by expanding features via a multi-branch convolution. Edge-aware upsampling is later proposed [14]. Wang *et al.* [13] develop MPU, a progressive network to learn the multi-level features for upsampling. Li *et al.* [15] design PU-GAN by exploring the power of the generative adversarial network, while Qian *et al.* [16] propose PU-GeoNet to generate samples in a 2D domain, then lift them to 3D via a linear transformation. Recently, Qian *et al.* [1] propose PU-GCN to better represent locality and aggregate the point neighborhood information via Graph Convolutional Networks. Li *et al.* [2] introduce Dis-PU to disentangle the task into two cascaded networks via a divide-and-conquer strategy. As these methods mainly operate on patch level, patch cropping and stitching may introduce significant information loss. Generally, upsampling is an ill-posed task, meaning that there could exist multiple possible outputs given a sparse input. In contrast, the restoration module in our framework is to leverage the self-embedded point set for accurately restoring the original input.

Concurrently, PointLIE [32] adopts an invertible neural network [33], [34] for point cloud sampling and recovery. Similar to its image counterpart [35], PointLIE learns to sample points from a dense input and encodes the remaining points into a case-agnostic latent variable that follows by a Gaussian distribution; the recovered points are obtained

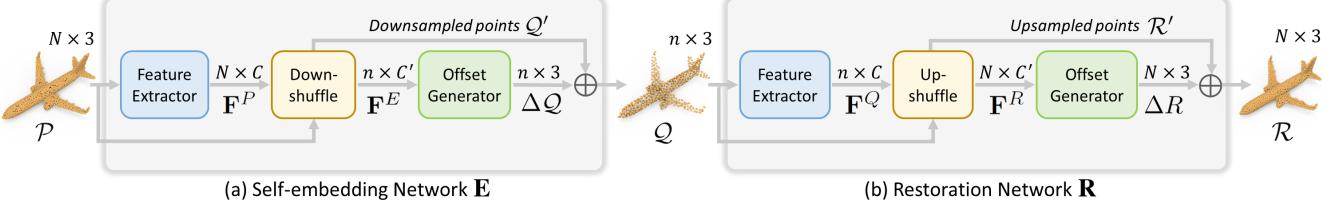


Fig. 2: Overview of our approach. Given input point set \mathcal{P} of N points, self-embedding network \mathbf{E} (a) produces sparse point set \mathcal{Q} of $n = N/r$ points (sampling rate $r > 1$), whereas restoration network \mathbf{R} (b) consumes \mathcal{Q} to produce restored point set \mathcal{R} . Both \mathcal{Q} and \mathcal{R} should look like \mathcal{P} , and \mathcal{R} should also be consistent to \mathcal{P} with similar point distribution.

by combining the sampled points and a randomly-drawn embedding via an invertible operation. Yet, PointLIE utilizes a case-agnostic embedding, so the distribution of the upsampled points may not well follow that of the original input; Figure 8 of the supplemental material. In contrast, we achieve a case-specific self-embedding, targeting not only to restore the original shape but also to conform to the original distribution; see the various visual results in Section 4 and the supplement. Essentially, PointLIE shares a similar objective as upsampling methods, instead of trying to restore the original inputs like ours.

Steganography on various representations. Recent steganography methods [36], [37], [38], [39], [40] conceal confidential information in images, videos, or audios into a reversible container, from which the secret information is recoverable. Among them, the most widely-adopted media is 2D digital image. Zhu *et al.* [38] propose to hide secret messages in images through noise interference, while Xia *et al.* [37] formulate a neural network to generate a reversible grayscale from a color image, where the colors can be restored from the grayscale image. Two recent works [41], [42] explores invertible conversion for halftoning and binocular videos. In this work, our attempt of exploring 3D point clouds with self-embeddings also belongs to the stream of works.

Deep learning on point clouds. Inspired by the success of PointNet [43], a wide range of deep-learning methods have been developed for assorted point cloud processing tasks, including classification [44], [45], [46], segmentation [47], [48], detection [49], [50], [51], generation [52], [53], [54], completion [55], [56], registration [57], [58], [59], and other applications [60], [61]. In this paper, we aim to learn a self-embedded point set that is restorable after downsampling.

3 METHOD

3.1 Overview

Figure 2 shows the overall framework for producing self-embedded point set: (i) the *self-embedding network* \mathbf{E} encodes input point cloud $\mathcal{P} \in \mathbb{R}^{N \times 3}$ of N points into self-embedded sparse point set $\mathcal{Q} \in \mathbb{R}^{n \times 3}$ of $n = N/r$ points with a sampling rate r ; and (ii) the *restoration network* \mathbf{R} recovers point set $\mathcal{R} \in \mathbb{R}^{N \times 3}$ from \mathcal{Q} . To achieve an effective self embedding, we should meet the following goals:

- G1: \mathcal{Q} should look like input \mathcal{P} but with fewer points;
- G2: \mathcal{Q} should self-embed the potentially missing geometric information of \mathcal{P} for better restoring \mathcal{R} later; and
- G3: \mathcal{R} should also look like \mathcal{P} , but its point distribution and density should conform to \mathcal{P} , in terms of both the global structure and the local point distribution.

For \mathbf{E} to learn to produce \mathcal{Q} (G1 & G2) and for \mathbf{R} to learn to consume the embedded information in \mathcal{Q} to recover \mathcal{R} (G3), we jointly train the two networks in an end-to-end manner. After that, we can employ \mathbf{E} for self-embedding and a \mathbf{R} for recovering details in various devices separately. Section 3.2 details the architecture of \mathbf{E} and \mathbf{R} , Section 3.3 presents the down-shuffle unit in \mathbf{E} , whereas Section 3.4 presents the up-shuffle unit in \mathbf{R} . Lastly, Section 3.5 presents our losses designed specifically to encourage \mathcal{R} to look like \mathcal{P} , both globally and locally.

3.2 Network Architecture

Self-embedding network \mathbf{E} . To start, we use a feature extractor (Figure 2(a)) to extract point features $\mathbf{F}^P \in \mathbb{R}^{N \times C}$ from \mathcal{P} , where C is the number of channels. In this work, we adopt the feature extractor used in [13], [15], where EdgeConv [62] is taken as the basic convolution layer with dense connections between layers to enhance the features.

We then feed \mathbf{F}^P into our down-shuffle unit (to be presented in Section 3.3) to obtain the self-embedded point features $\mathbf{F}^E \in \mathbb{R}^{n \times C'}$, where $C' > C$ is the number of channels. Now, to produce the self-embedded point set \mathcal{Q} , a straightforward approach is to directly regress \mathcal{Q} from \mathbf{F}^E via multi-layer perceptrons (MLPs). However, to meet the goals of self-embedding, we should try to embed more structural information of \mathcal{P} into \mathcal{Q} for a better recovery. At the same time, \mathcal{Q} has to look like \mathcal{P} . Therefore, we first pre-downsample \mathcal{P} into an initial downsampled point set $\mathcal{Q}' \in \mathbb{R}^{n \times 3}$, then regress offset vectors $\Delta \mathcal{Q} \in \mathbb{R}^{n \times 3}$ from \mathbf{F}^E via an offset generator (Figure 2(a)) implemented as MLPs. Lastly, we produce \mathcal{Q} as $\mathcal{Q}' + \Delta \mathcal{Q}$.

The above approach has two advantages. First, thanks to the guidance point set \mathcal{Q}' , which is already very similar to \mathcal{P} , we only need to ensure a small $\Delta \mathcal{Q}$ to keep the geometric similarity between \mathcal{Q} and \mathcal{P} . Second, since geometric similarity has been achieved with least effort, the self-embedding network can focus on preserving the valuable geometric information of \mathcal{P} using the regressed offsets. As shown in an experiment later, though $\Delta \mathcal{Q}$ is very small, it embeds important geometric information for restoring a higher-quality \mathcal{R} that is more consistent to \mathcal{P} .

Restoration network \mathbf{R} . Figure 2(b) shows the architecture of \mathbf{R} . First, we use a feature extractor of same architecture as that in \mathbf{E} to extract point features $\mathbf{F}^Q \in \mathbb{R}^{n \times C}$ from \mathcal{Q} . We then feed \mathbf{F}^Q into our up-shuffle unit (to be presented in Section 3.4) to generate the restored point features $\mathbf{F}^R \in \mathbb{R}^{N \times C'}$. Next, we create r copies of \mathcal{Q} to form the initial restored point set $\mathcal{R}' \in \mathbb{R}^{N \times 3}$, regress offset vectors

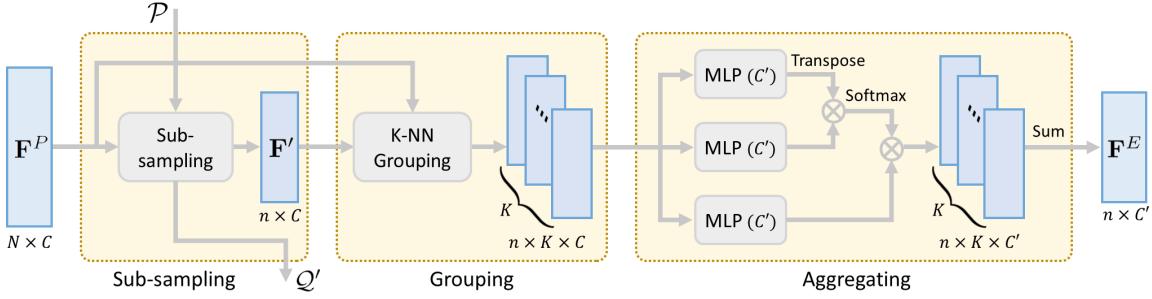


Fig. 3: The down-shuffle unit aims to group and aggregate input features $\mathbf{F}^P \in \mathbb{R}^{N \times C}$ into self-embedded point features $\mathbf{F}^E \in \mathbb{R}^{n \times C'}$ for the pre-downsampled point set Q' . Here, we aim to maximize the preservation of point features of $\mathcal{P} - Q'$ in the self-embeddings. Note that C and C' are the number of feature channels and K is the number of nearest neighbors.

$\Delta\mathcal{R} \in \mathbb{R}^{N \times 3}$ from \mathbf{F}^R via another offset generator (MLPs), and then add the offset vectors to \mathcal{R}' to produce the final restored point set \mathcal{R} , which is $\mathcal{R}' + \Delta\mathcal{R}$.

3.3 Down-shuffle Unit

Given input points \mathcal{P} with associated features \mathbf{F}^P , the down-shuffle unit aims to generate self-embedded point features $\mathbf{F}^E \in \mathbb{R}^{n \times C'}$ for producing the self-embedded context. To reduce the information loss, we group and aggregate neighboring point features into each sampled point feature (associated with the points in Q') and maximize the amount of original information in the self-embeddings.

Figure 3 shows the architecture of the down-shuffle unit, which has the following three steps.

- (i) *Sampling*. We obtain Q' from \mathcal{P} using farthest point sampling (FPS), then use FPS's sampling indices to obtain associated point features $\mathbf{F}' \in \mathbb{R}^{n \times C}$ from \mathbf{F}^P ;
- (ii) *Grouping*. To embed and retain the features of the points to be dropped (*i.e.*, those in $\mathcal{P} - Q'$), for each point $q' \in Q'$, we propose to locate the K nearest neighbors of q' in original \mathcal{P} and group their point features into an $n \times K \times C$ feature volume, where we set $K > N/n$ for a better coverage of points in \mathcal{P} ; and
- (iii) *Aggregation*. Instead of directly using a pooling operation [45] in the K dimension, we use a self-attention mechanism [63] to learn to better embed the local neighbor features around each downsampled point via a weighted aggregation, which ensures the embedded features \mathbf{F}^E to be as informative as possible.

Formally, given feature vector $f_i \in \mathbf{F}'$ associated with each sampled point $q_i \in Q'$, step (iii) can be written as

$$f_i^E = \mathcal{A}(\mathcal{W}(q_i^j, q_i^k)\gamma(f_i^j)), \forall q_i^j, q_i^k \in \mathcal{N}^{\mathcal{P}}(q_i),$$

where $\mathcal{N}^{\mathcal{P}}(q_i)$ is the set of K -nearest neighbors of q_i in \mathcal{P} ; q_i^j and q_i^k is the j -th and k -th neighbor of q_i , respectively; f_i^j denotes the feature vector associated with q_i^j ; and \mathcal{A} is a weighted sum over all the K neighbors, such that each updated point feature (*i.e.*, f_i^E) captures the information from all points in $\mathcal{N}^{\mathcal{P}}(q_i)$. Inside \mathcal{A} , the unary function γ is a linear transformation, which lifts the channel number in f_i^j from C to C' ; and the pairwise relation weight \mathcal{W} computes high-level relationships between the two neighbors q_i^k and q_i^j , which is a dot product similarity between f_i^k and f_i^j :

$$\mathcal{W}(q_i^j, q_i^k) = \text{Softmax}(\phi(f_i^k)^T \theta(f_i^j) / \sqrt{C'}),$$

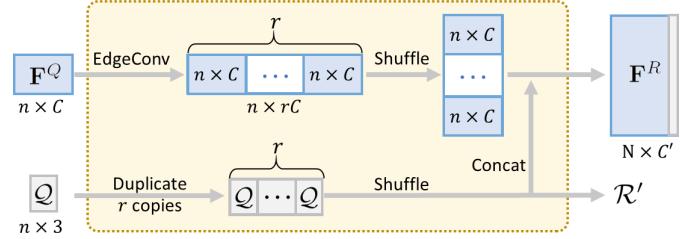


Fig. 4: The up-shuffle unit expands input features \mathbf{F}^Q to generate restored point features \mathbf{F}^R . Here, we expand \mathbf{F}^Q via a feature transformation (*i.e.*, EdgeConv [62]), reshape the result to $N \times C$, and concatenate it with r copies of Q to generate \mathbf{F}^R . So, $C' = C + 3$ with last three channels from Q .

where ϕ and θ are linear transformations that are implemented as independent MLPs (Figure 3). In this way, we can produce the self-embedded feature $f_i^E \in \mathbf{F}^E$ that encodes the local geometry of \mathcal{P} centered around q_i .

3.4 Up-shuffle Unit

From self-embedded points Q with features \mathbf{F}^Q , the up-shuffle unit aims to obtain restored point features $\mathbf{F}^R \in \mathbb{R}^{N \times C'}$. In existing upsampling methods, *e.g.* [13], [15], point feature expansion is achieved by feature duplication, then concatenating the results with a random 2D grid. However, such operation may introduce redundant information or even noise. Unlike the general upsampling, we restore point clouds by consuming the embedded information. Particularly, we aim to restore the original point features in the latent space to reduce the artifacts in the 3D data space.

Inspired by pixel periodic shuffle [64] for image super-resolution, we propose the up-shuffle unit shown in Figure 4. First, we use a graph convolutional layer (*i.e.*, EdgeConv [62]) to expand \mathbf{F}^Q from C to rC channels, where $r = N/n$. Note that, EdgeConv [62] is effective in capturing non-local neighboring point features, thus enabling feature expansion with long-range dependencies. Next, we shuffle the expanded features from $n \times rC$ to $N \times C$; see Figure 4. Also, we duplicate r copies of Q and shuffle it into the initial restored point set $\mathcal{R}' \in \mathbb{R}^{N \times 3}$. Last, we concatenate the expanded $N \times C$ features with \mathcal{R}' to produce the final restored point features $\mathbf{F}^R \in \mathbb{R}^{N \times C'}$, where $C' = C + 3$.

3.5 Loss Functions

To train \mathbf{E} and \mathbf{R} to produce Q and \mathcal{R} subject to the goals enlisted in Section 3.1, we formulate (i) shape similarity loss,

(ii) point distribution loss, and (iii) geometry-conformity loss. The first two losses are collectively referred to as the restoration loss, which encourages \mathcal{R} to be similar to \mathcal{P} both *globally* and *locally*, whereas the last one is for keeping $\Delta\mathcal{Q}$ to be small, such that \mathcal{Q} can look similar to \mathcal{P} .

Shape similarity loss. To ensure \mathcal{R} to be similar to \mathcal{P} , we may simply use an averaged per-point mean square error:

$$\mathcal{L}_{\text{shape}} = \frac{1}{N} \sum_{i=1}^N \|p_i - q_i\|_2, p_i \in \mathcal{P}, q_i \in \mathcal{R}. \quad (1)$$

However, Eq. (1) requires a fixed point-to-point correspondence (*i.e.*, $p_i \leftrightarrow q_i$) between \mathcal{P} and \mathcal{R} . Constraining the network output to follow a fixed order will greatly complicate the training, due to the unordered nature of points. So, we employ the Chamfer Distance (CD) [65] to encourage the global geometric consistency between \mathcal{P} and \mathcal{R} :

$$\mathcal{L}_{\text{shape}} = \sum_{r_i \in \mathcal{R}} \min_{p_j \in \mathcal{P}} \|r_i - p_j\|_2 + \sum_{p_i \in \mathcal{P}} \min_{r_j \in \mathcal{R}} \|p_i - r_j\|_2$$

CD, in fact, finds a flexible point-to-point correspondence by searching the closest point between \mathcal{R} and \mathcal{P} .

Point distribution loss. Though $\mathcal{L}_{\text{shape}}$ helps encourage a global shape similarity between \mathcal{R} and \mathcal{P} , it may not be sufficient to encourage a consistent local point distribution. We thus further formulate the point distribution loss $\mathcal{L}_{\text{dist}}$.

The key idea is to encourage local neighborhoods of the same point in \mathcal{R} and \mathcal{P} to be similar. Specifically, for each point p_i in \mathcal{P} , we search its m nearest neighbors independently in \mathcal{P} and \mathcal{R} ; the two nearest-neighbor sets are denoted as $\mathcal{N}^{\mathcal{P}}(p_i)$ and $\mathcal{N}^{\mathcal{R}}(p_i)$, respectively. We then construct their respective local distribution vectors

$$\begin{aligned} \mathcal{D}(p_i, \mathcal{P}) &= \{\overrightarrow{p_ip_{i1}}, \overrightarrow{p_ip_{i2}}, \dots, \overrightarrow{p_ip_{im}}\}, p_{ij} \in \mathcal{N}^{\mathcal{P}}(p_i) \\ \text{and } \mathcal{D}(p_i, \mathcal{R}) &= \{\overrightarrow{p_iq_{i1}}, \overrightarrow{p_iq_{i2}}, \dots, \overrightarrow{p_iq_{im}}\}, q_{ij} \in \mathcal{N}^{\mathcal{R}}(p_i), \end{aligned}$$

where $\overrightarrow{p_ip_{ij}}$ denotes the vector from p_i to p_{ij} . Also, we sort the m point-wise vectors in $\mathcal{D}(p_i, \mathcal{P})$ (and also in $\mathcal{D}(p_i, \mathcal{R})$) in ascending order of the vector magnitude. Hence, when $\mathcal{D}(p_i, \mathcal{P})$ and $\mathcal{D}(p_i, \mathcal{R})$ are similar, $\mathcal{N}^{\mathcal{P}}(p_i)$ and $\mathcal{N}^{\mathcal{R}}(p_i)$ should have similar point distributions.

Therefore, we formulate $\mathcal{L}_{\text{dist}}$ by minimizing the L_2 distance and the cosine angle difference between any two corresponding distribution vectors in $\mathcal{D}(p_i, \mathcal{P})$ and $\mathcal{D}(p_i, \mathcal{R})$:

$$\begin{aligned} \mathcal{L}_{\text{dist}} &= \frac{1}{N} \sum_{i=1}^N [\mathcal{L}_{\text{norm}}(i) + \beta \mathcal{L}_{\text{angle}}(i)], \quad (2) \\ \mathcal{L}_{\text{norm}}(i) &= \frac{1}{m} \sum_{j=1}^m \|\overrightarrow{p_ip_{ij}} - \overrightarrow{p_iq_{ij}}\|_2, \\ \text{and } \mathcal{L}_{\text{angle}}(i) &= \frac{1}{m} \sum_{j=1}^m \frac{\overrightarrow{p_ip_{ij}} \cdot \overrightarrow{p_iq_{ij}}}{\|\overrightarrow{p_ip_{ij}}\|_2 \cdot \|\overrightarrow{p_iq_{ij}}\|_2}, \end{aligned}$$

where β is a weight.

Geometry-conformity loss. To encourage \mathcal{Q} to be similar to \mathcal{P} , all we need to do is to keep $\Delta\mathcal{Q}$ small as explained earlier in Section 3.2. Hence, we formulate the geometry-conformity loss as an averaged truncated L_2 norm of $\Delta\mathcal{Q}$:

$$\mathcal{L}_Q(\mathcal{Q}, \mathcal{P}) = \frac{1}{n} \sum max\{0, \|\Delta\mathcal{Q}\|_2 - \tau\}, \quad (3)$$

where τ is a small threshold.

Overall loss. In summary, we jointly train the whole framework by minimizing the following objective function:

$$\mathcal{L} = \mathcal{L}_R(\mathcal{R}, \mathcal{P}) + \lambda \mathcal{L}_Q(\mathcal{Q}, \mathcal{P}), \quad (4)$$

where $\mathcal{L}_R(\mathcal{R}, \mathcal{P}) = \mathcal{L}_{\text{shape}} + \alpha \mathcal{L}_{\text{dist}}$ is the restoration loss; and α and λ are hyperparameters.

4 EXPERIMENTS

This section presents various experiments we conducted for method evaluation. Since we work on exploring self-embedding the structural information of a point set into its sparse version, we focus mainly on analyzing the design efficiency and intuition of the self-embedding. First, Section 4.2 shows a gallery of results. Section 4.3 shows comparisons with assorted related methods, including upsampling, downsampling, and cascade of them. Section 4.4 presents experimental results on point sets of large scale, varying input densities, and sampling rates. Last, Section 4.5 shows the ablation studies and Section 4.6 shows visualizations of the embedded information and discusses the limitations. More experimental results on both synthetic and real scans are provided in the supplemental material.

4.1 Experimental Setting

Datasets. We employ both synthetic and real-scanned data in experiments. For synthetic data, we use ModelNet40 [66] of 9,843 training and 2,468 testing shapes from 40 categories, and follow the train-test split in [43]. In detail, we uniformly sample 10,000 points on the surface of each shape using Poisson disk sampling. For real-scanned data, we use ScanObjectNN [67], which contains 2,902 point cloud objects (each 2,048 points) in 15 categories. Compared with ModelNet40, ScanObjectNN poses more practical challenges, including noise, object partiality, non-uniform point distribution, and deformation variants.

In our experiments, input \mathcal{P} has $N=2,048$ points, corresponding to the point cloud size in ScanObjectNN, whereas downsampled point set \mathcal{Q} has $n=512$ points with sampling rate $r=4$. To train our network, we employ the ModelNet40 training split and randomly sample 2,048 points from the 10,000 points in each training object. Also, we normalize each input \mathcal{P} to fit a unit sphere centered at the origin.

For a comprehensive generalization, we employ the following three kinds of data in testing: (i) *uniform set*: use FPS to sample 2,048 points from the 10,000 points in each test object in ModelNet40; (ii) *random set*: like (i) but randomly sample the points; and (iii) *partial set*: directly use the real-scanned 2,048 points in ScanObjectNN. Also, we compare with related methods on the PU-147 dataset [15], which contains 147 objects, and on two large-scale point cloud datasets, *i.e.*, Waymo open [68] and ScanNet [69].

Metrics. To study the effect of information preserving in the self-embedded point sets, we employ three commonly-used metrics to compare the restored points with the original points: (i) Earth Mover’s distance (EMD), (ii) Hausdorff distance (HD), and (iii) Chamfer distance (CD). EMD [65] measures the point-to-point distance using a bijection mapping between \mathcal{P} and \mathcal{R} . HD and CD, respectively, measure the maximum and average closest point distance between

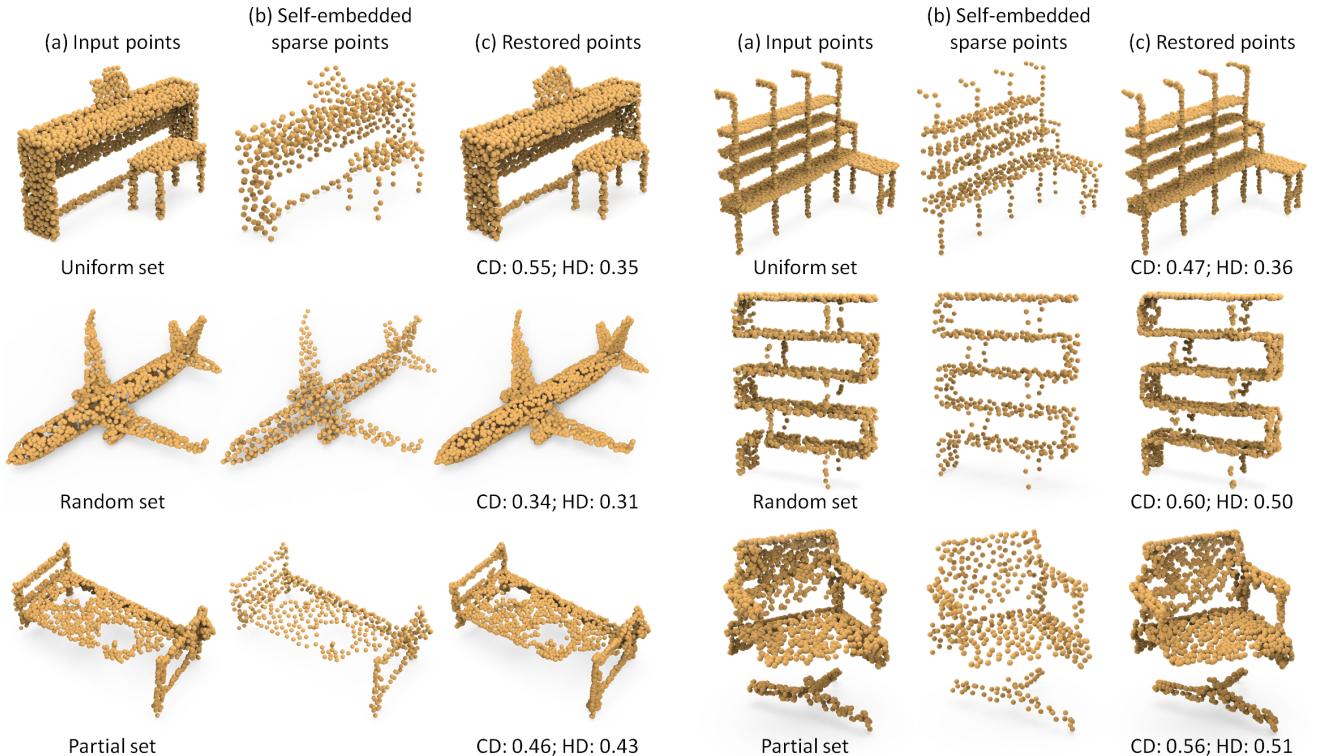


Fig. 5: Gallery of our self-embedded sparse points (b) and the restored points (c) given the original inputs (a) from three kinds of testing data: uniform set (top row), random set (middle row), and partial set (bottom row). The CD and HD values are calculated between (c) and (a), where the units are 10^{-3} and 10^{-2} , respectively.

TABLE 1: Quantitative comparisons between our method and two state-of-the-art upsampling methods. The units of EMD, HD, and CD are 10^{-2} , 10^{-2} , and 10^{-3} , respectively.

Methods	Uniform set			Random set			Partial set		
	EMD	HD	CD	EMD	HD	CD	EMD	HD	CD
PU-GCN [1]	6.57	1.11	1.13	7.48	1.19	1.30	6.82	0.83	0.72
Dis-PU [2]	6.24	1.29	0.99	7.32	1.41	1.14	6.41	0.99	0.63
Our	4.51	0.74	0.76	6.45	0.85	0.90	4.63	0.49	0.47

\mathcal{P} and \mathcal{R} . For these metrics, a small value indicates a large shape similarity between \mathcal{P} and \mathcal{R} .

Implementation details. We empirically set α , β , λ , and τ as 5.0, 2.0, 100.0, and 10^{-6} , respectively. We train our framework with a mini-batch size of 16 for 100 epochs on the TensorFlow platform, and adopt common augmentation strategies, including random scaling, rotation, and point perturbation. We use the Adam optimization with the learning rate of 0.001, which is linearly decreased by a decay rate of 0.5 per 20 epochs until 10^{-6} . The inference takes only 4.04ms for point set self-embedding and 6.85ms for restoration on a single 1080Ti GPU.

4.2 Restoration Visualization

We first demonstrate the self-embedded ability of our framework on point clouds of various geometric structures and point distributions. Figure 5 shows examples from the uniform (top), random (middle), and partial (bottom) test sets. Clearly, our generated self-embedded sparse point sets (Figure 5 (b)) look similar to the original inputs (a). Benefited by the self-embedded information in (b), our framework

restores high-quality dense points (c) that are very similar to the originals, regardless of the point distribution of the inputs. This is also evidenced by the small CD and HD values for all three test sets. Particularly, as shown on the right-hand side of Figure 5, even the input objects are complex with fine structures, our method can still yield high-quality restored point sets with small CD and HD values.

4.3 Comparison with Related Works

Comparing with upsampling methods. To study how our self-embedded point sets promote high-quality restorations, we compare our method with two state-of-the-art inference-based upsampling methods, PU-GCN [1] and Dis-PU [2]. We followed the setting in [1], [2] to re-train their networks using our training data. Table 1 shows the quantitative evaluation and Figure 6 shows the visual comparisons. Note that, this comparison may not be very appropriate, since [1], [2] are designed for upsampling instead of restoration; yet, the comparison can reveal the ability of our method in consuming the self-embedded information for better restorations. In detail, we feed the point set downsampled by FPS (b) to these upsampling methods and feed the self-embedded point set (e) to our method for restoration. Our restored point sets are more similar to the originals with the smallest CD and HD values. In contrast, these upsampling-based methods cannot infer fine structures, such as the sharp narrow edges shown on the bottom row example.

Further, we study whether the results, that are restored from the self-embedded point sets, are still effective for downstream tasks like the original ones. To do so, we apply the restored point sets to shape classification and compare

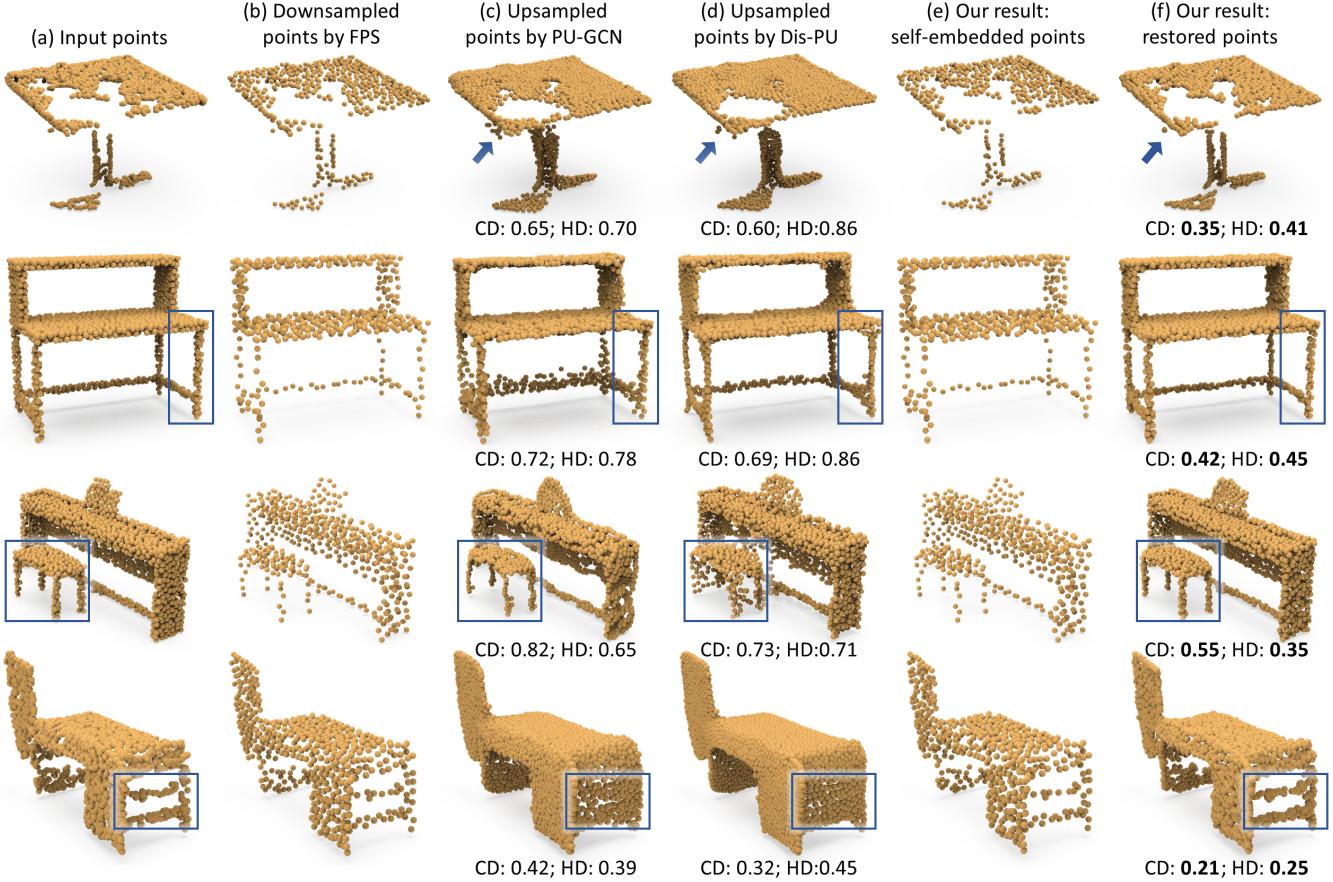


Fig. 6: Given the input points (a), our framework can well leverage the self-embedded sparse points (e) to produce high-quality restorations (f) that are more faithful to the input point clouds; however, the upsampled results by PU-GCN [1] (c) and Dis-PU [2] (d) from the downsampled points (b) via farthest-point sampling (FPS) are far from the originals.

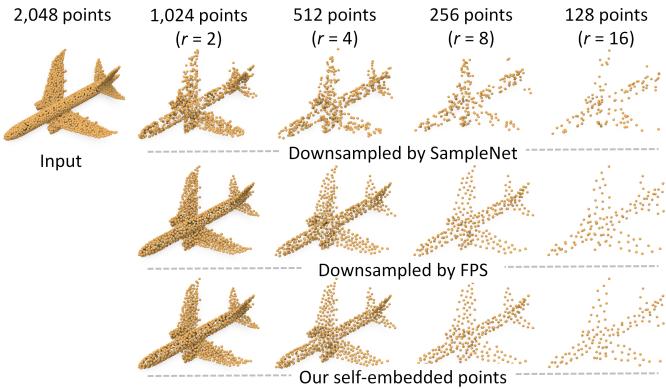


Fig. 7: Comparing the downsampling results produced by SampleNet [25] (top), FPS (middle), and our method (bottom) for decreasing sampling rates. Our self-embedded point sets are more similar to those produced by the ordinary FPS, presenting visually-recognizable shapes as the originals. In contrast, it is harder to recognize the results of SampleNet, particularly for large downsampling rates.

the classification performance with (i) the original point sets (Ori) and (ii) using Dis-PU [2] to upsample the FPS-downsampled point sets (Ups). Specifically, we employ both the synthetic ModelNet40 [70] and real-scanned ScanObjectNN [67] to train different classifiers [43], [45], [59] via the same settings in their papers. The pre-trained models

TABLE 2: Comparing the overall shape classification accuracy (%) on the original input (Ori.), upsampled results (Ups.), and our restored output (Res.). Clearly, the performance of our restored outputs are closer to the originals.

Methods	ModelNet40 [70]			ScanObjectNN [67]		
	Ori.	Ups.	Res.	Ori.	Ups.	Res.
PointNet [43]	89.2	81.0	88.9	79.2	68.4	78.6
PointNet++ [45]	91.9	85.0	91.2	84.3	77.2	83.4
DGCNN [59]	92.2	86.1	92.0	86.2	79.0	85.9

are directly applied to different test sets. Table 2 enlists the quantitative results for comparison, showing that our restored point sets achieve similar classification accuracies as the originals. In contrast, there is a large performance drop when directly testing on the upsampled point sets, due to permanent information loss after the downsampling.

Comparing with downsampling methods. Regularly point sampling [6], [7], [8] generally leads to better visualizations. To show that our self-embedded point sets can function as ordinary downsampled ones, we compare them with those produced by a learning-based method (*i.e.*, SampleNet [25]) and regular sampling (*i.e.*, FPS [7]). Figure 7 shows the results for increasing downsampling rates. Compared with SampleNet, our results exhibit similar distributions as those of FPS, capable of serving as better visually-pleasing previews of the original geometry.

TABLE 3: User study results. Statistically, our self-embedded point sets and FPS point sets better conform to the original geometry; they are more preferred by the participants.

	Random	SampleNet	Self-Embed	FPS
Preference (%)	7.5 ± 2.5	10.0 ± 5.5	42.5 ± 5.5	40.0 ± 5.0
Conformity (0-5)	1.5 ± 1.3	2.5 ± 1.1	4.5 ± 1.0	4.5 ± 0.9

TABLE 4: Quantitative comparison with methods combining downsampling and upsampling on PU-147 [15].

	PU-GAN +SampleNet	PU-GCN +SampleNet	Dis-PU +SampleNet	PointLIE	Ours
EMD (10^{-2})	3.04	3.11	2.70	-	2.13
HD (10^{-3})	2.70	2.96	1.88	1.71	1.58
CD (10^{-3})	0.26	0.24	0.16	0.21	0.14

Further, we evaluate how well our self-embedded point sets visually conform to the original geometry through a user study with 20 participants (12 males and 8 females, aged 22 to 30). We randomly selected 15 example shapes and side-by-side show to each participant: the original input, the randomly-downsampled points, the FPS-downsampled points, the point samples produced by SampleNet [25], and our self-embedded point set. To avoid bias, we randomized the location of the four down-sampling point sets. First, for each example shape, we asked the participant to choose the most representative one relative to the original (*preference*). Next, we asked the participant to rate the geometric *conformity* of each down-sampled point set relative to the original: from 0 (completely different) to 5 (completely the same). Table 3 summarizes the results in terms of mean and standard deviation. Overall, our self-embedded point sets and the FPS-downsampled ones are more preferred, since they exhibit better visual recognizability as the originals.

Comparing with downsampling-aware upsampling methods. Next, we compared the restoration results on the PU-147 dataset [15] with (i) approaches that cascade a learning-based downsampling method with a recent upsampling method and (ii) the concurrent work PointLIE [32]¹. Specifically, we trained SampleNet [25] with pre-trained upsampling networks, *i.e.*, PU-GAN [15], PU-GCN [1], and Dis-PU [2]. Table 4 and Figure 8 show the quantitative and qualitative results, respectively, demonstrating that the cascaded methods and the recent PointLIE cannot effectively improve the reconstruction performance as ours.

4.4 Robustness Test

Restoration results on large-scale real scans. Figure 9 shows the restoration results on large-scale inputs (left column), comparing the results of our method (right column) with SampleNet+Dis-PU (middle column). Note that, the input in row(a) is from the outdoor Waymo open dataset [68], while the input in row(b) is from the indoor ScanNet [69] dataset, and row(c) shows the associated reconstructed meshes. We directly tested all the networks on these large-scale inputs without any re-training. Like upsampling

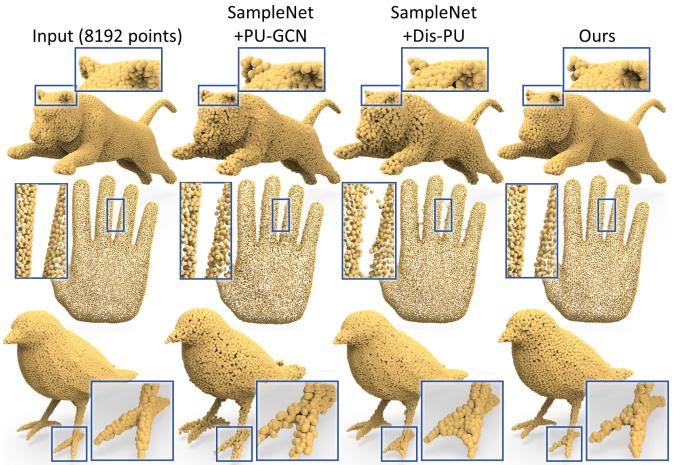


Fig. 8: Restoration results on PU-147 [15]. Compared with methods by cascading SampleNet with PU-GCN or Dis-PU, our results exhibit more details and are closer to the inputs.

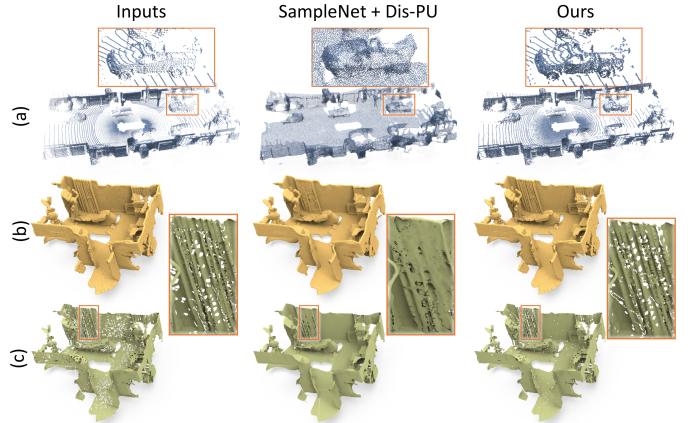


Fig. 9: Restoration results on large-scale real scans. (a) results on the outdoor Waymo open dataset [68]; (b)&(c) results on the indoor ScanNet [69] and associated 3D reconstructed meshes. Compared with SampleNet+Dis-PU, our results exhibit details that are significantly closer to the inputs.

methods [1], [15], we split the input points into patches, feed each patch with 2048 points into our framework, and merge the output patches as the restoration results. Our restored results are significantly much closer to the inputs, in terms of preserving the original scanlines in (a) and well recovering the details in the inputs in (b), thus promoting an accurate and similar surface reconstruction in (c) as the original; see more visual comparison results in the supplement.

Restoration results on varying sampling rates. We further show restored results for increasing downsampling rate r in Figure 10. From the results, we can see that our method can be applied to different sampling rates and produce stable restored geometry for decreasing point numbers, even for extremely sparse self-embedded point set, which has only 128 points with a large sampling rate of $r = 16$. Table 5 shows the corresponding quantitative evaluation results.

Restoration results on varying input sizes. Figure 11 and Table 6 show another set of comparison results on restoring input point sets of different sizes. The results show that our method performs stably for varying input sizes; more visual results are shown in the supplemental material.

1. Since there is no publicly released code so far, we directly take its quantitative results from its original paper [32] in the comparison.

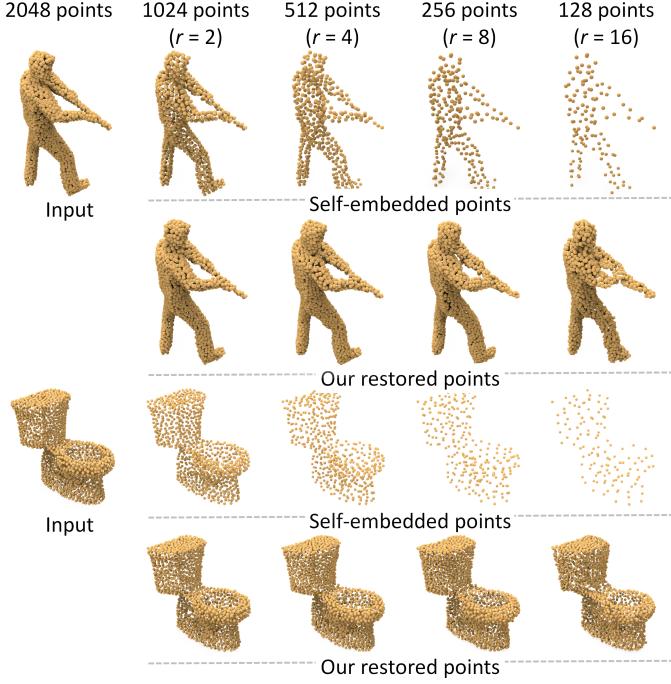


Fig. 10: Restoration results of increasing sampling rates.

TABLE 5: Quantitative results on inputs with various downsample rates. The input point number is 2048.

r	2	4	8	16
EMD (10^{-2})	3.65	4.51	5.95	7.02
HD (10^{-2})	0.26	0.74	0.85	1.01
CD (10^{-3})	0.25	0.76	0.92	1.18

4.5 Ablation Studies

We conducted a series of ablation studies to analyze the major components in our framework, including the self-embedding network E , restoration network R , and point distribution loss $\mathcal{L}_{\text{dist}}$. Table 7 summarizes the evaluation results by comparing the restored points with the original inputs, in which we use the uniform set for testing.

Self-embedding network E . First, we replace the attention module in the down-shuffle unit (see Section 3.3) with the direct max-pooling operation in [45] to aggregate the neighbor features; see the first row of Table 7 for the results. By comparing with our full pipeline in the bottom row, we can see that the weighted aggregation in the attention module leads to a better performance by maximizing the information in the embedded features. Second, instead of regressing the offsets, we modify it to directly generate the embedded sparse points; see the second row. Yet, such a one-step regression leads to a worse performance.

Restoration network R . Instead of using the up-shuffle operation (see Section 3.4) to expand the point features, we replace it with the commonly-used duplication operation in the existing upsampling methods [13], [15]. The first row in the R section of Table 7 shows the results, indicating the superiority of our up-shuffle operation over simple duplication. Next, we remove the offset regression and directly produce the restored points. Similarly, as shown in the

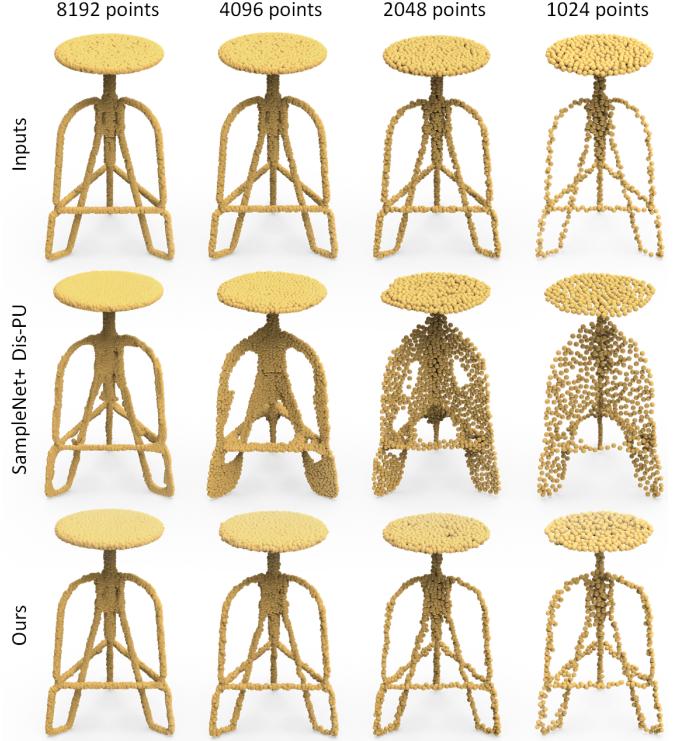


Fig. 11: Restoration results of decreasing input density. Our method performs stably, even restoring with a sparse input.

TABLE 6: Quantitative results (EMD (10^{-2})) on inputs with various point numbers (default downsampling rate is 4).

N	1024	2048	4096	8192
SampleNet+Dis-PU	6.67	6.26	5.93	4.17
Ours	5.78	4.51	3.22	2.71

second row in the R section, regressing the offset is much better than directly regressing the restored points.

Point distribution loss $\mathcal{L}_{\text{dist}}$. As detailed in Section 3.5, besides the shape similarity loss $\mathcal{L}_{\text{shape}}$ (Eq. (2)), we further propose the point distribution loss $\mathcal{L}_{\text{dist}}$ (Eq. (2)) to promote the similarity in local point distribution by $\mathcal{L}_{\text{norm}}$ and $\mathcal{L}_{\text{angle}}$. To analyze the contribution of $\mathcal{L}_{\text{dist}}$, we either remove each term in $\mathcal{L}_{\text{dist}}$ or directly remove $\mathcal{L}_{\text{dist}}$ by keeping only $\mathcal{L}_{\text{shape}}$. The results presented on the $\mathcal{L}_{\text{dist}}$ section of Table 7 show that each term contributes to a better performance.

4.6 Discussion on Self-embedding

How does self-embedding work? First, we investigate whether our framework can restore the original point set without the tiny offset ΔQ . So, we omit ΔQ and directly feed sparse point set Q' to the trained restoration network, which is essentially degraded into an upsampling network. As Figure 12(a) shows, the network is unable to recover high-quality dense points without ΔQ , comparing to restoring with ΔQ in Figure 12(b). Conversely, upsampling our self-embedded sparse points Q using the existing inference-based method, *i.e.*, Dis-PU, cannot yield high-quality restoration (Figure 12(e) vs. (b)), as it does not know how to utilize the structural information encoded

TABLE 7: Ablation analysis of our framework: the self-embedding network E , restoration network R , and point distribution loss $\mathcal{L}_{\text{dist}}$. We consider various adaptations in each component and show the result of our full pipeline on the last row for comparison. The units of EMD, HD, and CD are 10^{-2} , 10^{-2} , and 10^{-3} , respectively.

	pooling	attention	offset	EMD	HD	CD
E	✓		✓	5.02	0.95	0.98
		✓		5.16	1.02	1.07
R	duplicate	up-shuffle	offset	EMD	HD	CD
	✓		✓	4.91	1.05	1.01
$\mathcal{L}_{\text{dist}}$	$\mathcal{L}_{\text{shape}}$	$\mathcal{L}_{\text{norm}}$	$\mathcal{L}_{\text{angle}}$	EMD	HD	CD
	✓			4.74	0.87	0.89
	✓	✓		4.64	0.79	0.81
Full			✓	4.68	0.81	0.85
				4.51	0.74	0.76

in ΔQ . Also, there is not much difference between the Dis-PU-upsampling results from the sparse points via FPS (Figure 12(d)) and from our self-embedded points (e).

How is self-embedding represented? Figure 13 shows the magnitude and direction of the offset vectors in ΔQ as point size and color, respectively, of Q' . Since offset vectors in ΔQ are tiny, we enlarge the magnitude non-linearly for viewing. At first glance, the embedded information seems to reveal some geometric meanings, e.g., symmetry on airplane wings. As we have no control on how the structural information is encoded in ΔQ , the network can freely encode in its own way, provided the resultant offsets are small.

Limitation & Discussion. Manipulating the self-embedded point sets can hurt the restorability. For example, if we perturb the self-embedded points by randomly permuting the offset ΔQ , we can observe how the restored dense point sets are affected in Figure 12(c). Its quality is significantly lower than that in (e). It is because the manipulations ruin the visually-embedded information, thus interfering the restoration network. Similarly, upsampling the perturbed point set by Dis-PU (Figure 12(f)) also performs badly, as Dis-PU cannot utilize the embedded information anyway. Also, like most methods on point cloud downsampling [24], [25] and upsampling [1], [2], [13], [15], our self-embedding framework requires users to specify the downsampling rate. In the future, we plan to further formulate a perception metric to quantify the difference between the restored point set and the input. By this means, we may automatically determine the downsampling rate.

Self-embeddings in other forms. Though our framework is for self-embedding the original structure information, the general idea of point cloud self-embedding can be extended further for embedding other information such as colors, normals, or labels in indoor/outdoor 3D scenes.

5 CONCLUSION

We present an innovative method, capable of imperceptibly self-embedding the shape context of a point set into its sparse version. The self-embedded point set not only

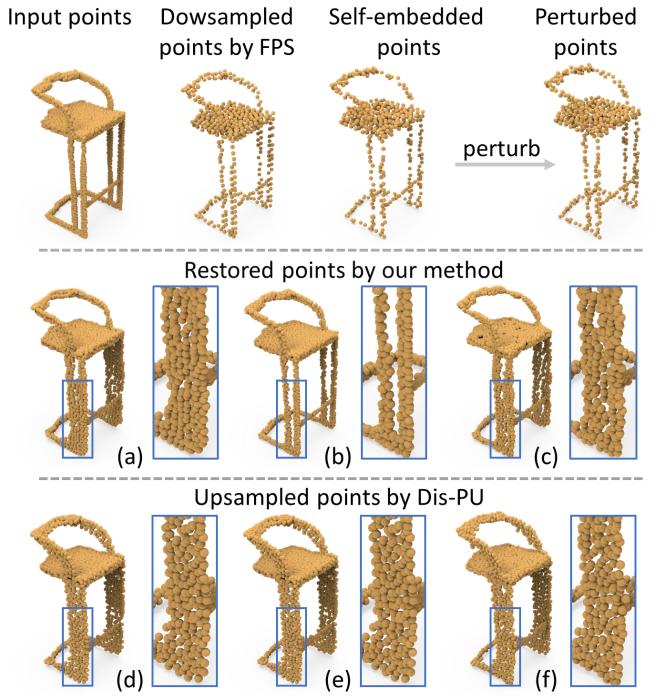


Fig. 12: Given FPS-downsampled points, our self-embedded points, and perturbed self-embedded points (on top), (a)-(c) show the corresponding restoration results produced by our method, whereas (d)-(f) show the corresponding upsampled points produced by Dis-PU [2]. (b) shows that our method can consume the self-embedded information to produce a higher quality restoration that is closer to the input.



Fig. 13: Visualizing the self-embedded offsets ΔQ by mapping $||\Delta Q||$ to point size and $\Delta Q/||\Delta Q||$ to color in Q' .

functions as an ordinary downsampled point set for visualizations but also allows us to restore the original density for viewing the details and for further analysis. To achieve a learnable self-embedding scheme, we design a novel framework, consisting of two jointly-optimized networks: a self-embedding network to encode the input point set into a self-embedded sparse version and a restoration network to leverage the embedded information to reconstruct the original point set. Both qualitative and quantitative experimental results show the effectiveness of our approach.

ACKNOWLEDGMENTS

We thank reviewers for their valuable comments. The work is supported by the Research Grants Council of the Hong Kong Special Administrative Region, China [Project No.: CUHK 14201921] and CUHK Direct Grant [Project No. 4055152].

REFERENCES

- [1] G. Qian, A. Abualshour, G. Li, A. Thabet, and B. Ghanem, "PU-GCN: Point cloud upsampling using graph convolutional networks," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 11 683–11 692.
- [2] R. Li, X. Li, P.-A. Heng, and C.-W. Fu, "Point cloud upsampling via disentangled refinement," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 344–353.
- [3] B. Han, Y. Liu, and F. Qian, "ViVo: Visibility-aware mobile volumetric video streaming," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1–13.
- [4] G. Ananthanarayanan, P. Bahl, P. Bodík, K. Chintalapudi, M. Philipose, L. Ravindranath, and S. Sinha, "Real-time video analytics: The killer app for edge computing," *computer*, vol. 50, no. 10, pp. 58–67, 2017.
- [5] H. Zhang, B. Han, C. Y. Ip, and P. Mohapatra, "Slimmer: Accelerating 3D semantic segmentation for mobile augmented reality," in *IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, 2020, pp. 603–612.
- [6] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.
- [7] Y. Eldar, M. Lindenbaum, M. Porat, and Y. Y. Zeevi, "The farthest point strategy for progressive image sampling," *IEEE Trans. Image Proc. (TIP)*, vol. 6, no. 9, pp. 1305–1315, 1997.
- [8] X. Ying, S.-Q. Xin, Q. Sun, and Y. He, "An intrinsic algorithm for parallel poisson disk sampling on arbitrary surfaces," *IEEE Trans. Vis. & Comp. Graphics (TVCG)*, vol. 19, no. 9, pp. 1425–1437, 2013.
- [9] M. Pauly, M. Gross, and L. P. Kobbelt, "Efficient simplification of point-sampled surfaces," in *Proceedings of the conference on Visualization'02*, 2002, pp. 163–170.
- [10] Y. Miao, R. Pajarola, and J. Feng, "Curvature-aware adaptive re-sampling for point-sampled geometry," *Computer-Aided Design (CAD)*, vol. 41, no. 6, pp. 395–403, 2009.
- [11] Z. Chen, T. Zhang, J. Cao, Y. J. Zhang, and C. Wang, "Point cloud resampling using centroidal Voronoi tessellation methods," *Computer-Aided Design (CAD)*, vol. 102, pp. 12–21, 2018.
- [12] L. Yu, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng, "PU-Net: Point cloud upsampling network," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 2790–2799.
- [13] W. Yifan, S. Wu, H. Huang, D. Cohen-Or, and O. Sorkine-Hornung, "Patch-based progressive 3D point set upsampling," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 5958–5967.
- [14] L. Yu, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng, "EC-Net: An edge-aware point set consolidation network," in *European Conf. on Computer Vision (ECCV)*, 2018, pp. 386–402.
- [15] R. Li, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng, "PU-GAN: A point cloud upsampling adversarial network," in *IEEE Intl. Conf. on Computer Vision (ICCV)*, 2019, pp. 7203–7212.
- [16] Y. Qian, J. Hou, S. Kwong, and Y. He, "PUGeo-Net: A geometry-centric network for 3D point cloud upsampling," in *European Conf. on Computer Vision (ECCV)*, 2020.
- [17] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.
- [18] Z. Yu, H.-S. Wong, H. Peng, and Q. Ma, "ASM: An adaptive simplification method for 3D point-based models," *Computer-Aided Design (CAD)*, vol. 42, no. 7, pp. 598–612, 2010.
- [19] B.-Q. Shi, J. Liang, and Q. Liu, "Adaptive simplification of point cloud using k-means clustering," *Computer-Aided Design (CAD)*, vol. 43, no. 8, pp. 910–922, 2011.
- [20] A. P. Witkin and P. S. Heckbert, "Using particles to sample and control implicit surfaces," in *Proceedings of ACM SIGGRAPH*, 1994, pp. 269–277.
- [21] J. Proença, J. A. Jorge, and M. C. Sousa, "Sampling point-set implicits," in *Proceedings of Eurographics Symposium on Point-based Graphics*, 2007, pp. 11–18.
- [22] L. Linsen, *Point cloud representation*. Technical Report, Germany: Faculty of Computer Science, University of Karlsruhe, 2001.
- [23] H. Song and H.-Y. Feng, "A progressive point cloud simplification algorithm with preserved sharp edge data," *The International Journal of Advanced Manufacturing Technology*, vol. 45, pp. 583–592, 2009.
- [24] O. Dovrat, I. Lang, and S. Avidan, "Learning to sample," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2760–2769.
- [25] I. Lang, A. Manor, and S. Avidan, "SampleNet: Differentiable point cloud sampling," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 7578–7588.
- [26] Y. Qian, J. Hou, Q. Zhang, Y. Zeng, S. Kwong, and Y. He, "MOPS-Net: A matrix optimization-driven network for task-oriented 3D point cloud downsampling," *arXiv preprint:2005.00383*, 2020.
- [27] M. Alexa, J. Behr, D. Cohen-Or, S. Fleishman, D. Levin, and C. T. Silva, "Computing and rendering point set surfaces," *IEEE Trans. Vis. & Comp. Graphics (TVCG)*, vol. 9, no. 1, pp. 3–15, 2003.
- [28] Y. Lipman, D. Cohen-Or, D. Levin, and H. Tal-Ezer, "Parameterization-free projection for geometry reconstruction," *ACM Trans. on Graphics (SIGGRAPH)*, vol. 26, no. 3, pp. 22:1–5, 2007.
- [29] H. Huang, D. Li, H. Zhang, U. Ascher, and D. Cohen-Or, "Consolidation of unorganized point clouds for surface reconstruction," *ACM Trans. on Graphics (SIGGRAPH Asia)*, vol. 28, no. 5, pp. 176:1–7, 2009.
- [30] H. Huang, S. Wu, M. Gong, D. Cohen-Or, U. Ascher, and H. Zhang, "Edge-aware point set resampling," *ACM Trans. on Graphics (TOG)*, vol. 32, no. 1, pp. 9:1–12, 2013.
- [31] S. Wu, H. Huang, M. Gong, M. Zwicker, and D. Cohen-Or, "Deep points consolidation," *ACM Trans. on Graphics (SIGGRAPH Asia)*, vol. 34, no. 6, pp. 176:1–13, 2015.
- [32] W. Zhao, X. Yan, J. Gao, R. Zhang, J. Zhang, Z. Li, S. Wu, and S. Cui, "PointLIE: Locally invertible embedding for point cloud sampling and recovery," in *Intl. Joint Conf. on Artificial Intell. (IJCAI)*, 2021, pp. 1345–1351.
- [33] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using real NVP," in *Intl. Conf. on Learning Representations (ICLR)*, 2017.
- [34] D. P. Kingma and P. Dhariwal, "GLOW: Generative flow with invertible 1x1 convolutions," in *Conference and Workshop on Neural Information Processing Systems (NeurIPS)*, 2018, pp. 10 215–10 224.
- [35] M. Xiao, S. Zheng, C. Liu, Y. Wang, D. He, G. Ke, J. Bian, Z. Lin, and T.-Y. Liu, "Invertible image rescaling," in *European Conf. on Computer Vision (ECCV)*, 2020, pp. 126–144.
- [36] S. Baluja, "Hiding images in plain sight: Deep steganography," in *Conference and Workshop on Neural Information Processing Systems (NeurIPS)*, 2017, pp. 2066–2076.
- [37] M. Xia, X. Liu, and T.-T. Wong, "Invertible grayscale," *ACM Trans. on Graphics (TOG)*, vol. 37, no. 6, pp. 1–10, 2018.
- [38] J. Zhu, R. Kaplan, J. Johnson, and L. Fei-Fei, "Hidden: Hiding data with deep networks," in *European Conf. on Computer Vision (ECCV)*, 2018, pp. 657–672.
- [39] Z. Wang, N. Gao, X. Wang, J. Xiang, D. Zha, and L. Li, "Hidinggan: High capacity information hiding with generative adversarial network," in *Computer Graphics Forum (CGF)*, vol. 38, no. 7, 2019, pp. 393–401.
- [40] E. Wengrowski and K. Dana, "Light field messaging with deep photographic steganography," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 1515–1524.
- [41] M. Xia, W. Hu, X. Liu, and T.-T. Wong, "Deep halftoning with reversible binary pattern," in *IEEE Intl. Conf. on Computer Vision (ICCV)*, 2021, pp. 14 000–14 009.
- [42] W. Hu, M. Xia, C.-W. Fu, and T.-T. Wong, "Mononizing binocular videos," *ACM Trans. on Graphics (SIGGRAPH Asia)*, vol. 39, no. 6, pp. 228:1–16, 2020.
- [43] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 652–660.
- [44] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, and B. Chen, "PointCNN: Convolution on \mathcal{X} -transformed points," in *Conference and Workshop on Neural Information Processing Systems (NeurIPS)*, 2018, pp. 828–838.
- [45] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Conference and Workshop on Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5099–5108.
- [46] R. Li, X. Li, P.-A. Heng, and C.-W. Fu, "PointAugment: an auto-augmentation framework for point cloud classification," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 6378–6387.
- [47] K. Mo, S. Zhu, A. X. Chang, L. Yi, S. Tripathi, L. J. Guibas, and H. Su, "PartNet: A large-scale benchmark for fine-grained and hierarchical part-level 3D object understanding," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 909–918.

- [48] L. Wang, Y. Huang, Y. Hou, S. Zhang, and J. Shan, "Graph attention convolution for point cloud semantic segmentation," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 10296–10305.
- [49] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "PointPillars: Fast encoders for object detection from point clouds," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 12697–12705.
- [50] S. Shi, X. Wang, and H. Li, "PointRCNN: 3D object proposal generation and detection from point cloud," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 770–779.
- [51] C. R. Qi, O. Litany, K. He, and L. J. Guibas, "Deep Hough voting for 3D object detection in point clouds," in *IEEE Intl. Conf. on Computer Vision (ICCV)*, 2019, pp. 9277–9286.
- [52] C.-L. Li, M. Zaheer, Y. Zhang, B. Poczos, and R. Salakhutdinov, "Point cloud GAN," *arXiv preprint arXiv:1810.05795*, 2018.
- [53] G. Yang, X. Huang, Z. Hao, M.-Y. Liu, S. Belongie, and B. Hariharan, "PointFlow: 3D point cloud generation with continuous normalizing flows," in *IEEE Intl. Conf. on Computer Vision (ICCV)*, 2019, pp. 4541–4550.
- [54] R. Li, X. Li, K.-H. Hui, and C.-W. Fu, "SP-GAN: Sphere-guided 3D shape generation and manipulation," *ACM Trans. on Graphics (SIGGRAPH)*, vol. 40, no. 4, pp. 151:1–12, 2021.
- [55] W. Yuan, T. Khot, D. Held, C. Mertz, and M. Hebert, "PCN: Point completion network," in *Intl. Conf. on 3D Vision (3DV)*, 2018, pp. 728–737.
- [56] X. Chen, B. Chen, and N. J. Mitra, "Unpaired point cloud completion on real scans using adversarial training," *Intl. Conf. on Learning Representations (ICLR)*, 2020.
- [57] Y. Aoki, H. Goforth, R. A. Srivatsan, and S. Lucey, "PointNetLK: Robust & efficient point cloud registration using PointNet," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 7163–7172.
- [58] W. Lu, G. Wan, Y. Zhou, X. Fu, P. Yuan, and S. Song, "DeepVCP: An end-to-end deep neural network for 3D point cloud registration," in *IEEE Intl. Conf. on Computer Vision (ICCV)*, 2019, pp. 12–21.
- [59] Y. Wang and J. M. Solomon, "Deep closest point: Learning representations for point cloud registration," in *IEEE Intl. Conf. on Computer Vision (ICCV)*, 2019, pp. 3523–3532.
- [60] P. Hermosilla, T. Ritschel, and T. Ropinski, "Total Denoising: Unsupervised learning of 3D point cloud cleaning," in *IEEE Intl. Conf. on Computer Vision (ICCV)*, 2019, pp. 52–60.
- [61] Z. Chen, W. Zeng, Z. Yang, L. Yu, C.-W. Fu, and H. Qu, "LassoNet: Deep lasso-selection of 3D point clouds," *IEEE Trans. Vis. & Comp. Graphics (TVCG)*, vol. 26, no. 1, pp. 195–204, 2019.
- [62] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph CNN for learning on point clouds," *ACM Trans. on Graphics (TOG)*, vol. 38, no. 5, pp. 146:1–12, 2019.
- [63] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Conference and Workshop on Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008.
- [64] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1874–1883.
- [65] H. Fan, H. Su, and L. J. Guibas, "A point set generation network for 3D object reconstruction from a single image," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 605–613.
- [66] A. X. Chang, T. Funkhouser, L. J. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su *et al.*, "ShapeNet: An information-rich 3D model repository," *arXiv preprint arXiv:1512.03012*, 2015.
- [67] M. A. Uy, Q.-H. Pham, B.-S. Hua, T. Nguyen, and S.-K. Yeung, "Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data," in *IEEE Intl. Conf. on Computer Vision (ICCV)*, 2019, pp. 1588–1597.
- [68] P. Sun, H. Kretschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine *et al.*, "Scalability in perception for autonomous driving: Waymo open dataset," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 2446–2454.
- [69] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, "ScanNet: Richly-annotated 3D reconstructions of indoor scenes," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5828–5839.
- [70] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3D ShapeNets: A deep representation for volumetric shapes," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1912–1920.



Ruihui Li is currently an associate professor at Hunan University. Before that, he was a post-doctoral fellow at the Chinese University of Hong Kong. He received his Ph.D. degree in the Department of Computer Science and Engineering from the Chinese University of Hong Kong. He serves as the reviewer of several conferences and journals, including TPAMI, IJCV, TVCG, CVPR, ICCV, etc. His research interests include deep geometry learning, generative modeling, 3D vision, and computer graphics.



Xianzhi Li is currently an associated professor at Huazhong University of Science and Technology. Prior to that, she was a post-doctoral fellow at the Chinese University of Hong Kong. She received her Ph.D. degree in the Department of Computer Science and Engineering from the Chinese University of Hong Kong. She serves as the reviewer of several conferences and journals, including TVCG, CVPR, ICCV, etc. Her research interests focus on 3D vision, computer graphics, and deep learning.



Tien-Tsin Wong received the BSc degree in computer science from the Chinese University of Hong Kong, in 1992, and the MPhil and PhD degrees in computer science from the same university, in 1994 and 1998 respectively. In August 1999, he joined the Computer Science & Engineering Department, Chinese University of Hong Kong. He is currently a professor. He is a core member of Virtual Reality, Visualization and Imaging Research Centre in the Chinese University of Hong Kong. His main research interests include computer graphics, computational manga, computer vision, machine learning, image-based rendering, and medical visualization. He is a senior member of the ACM.



Chi-Wing Fu Chi-Wing Fu is currently a full professor in the Chinese University of Hong Kong. He served as the co-chair of SIGGRAPH ASIA Technical Brief and Poster program, associate editor of IEEE Computer Graphics & Applications, and Computer Graphics Forum, panel member in SIGGRAPH 2019 Doctoral Consortium, and program committee members in various research conferences, including SIGGRAPH Technical papers, SIGGRAPH Asia Technical Brief, SIGGRAPH Asia Emerging tech., IEEE visualization, CVPR, IEEE VR, VRST, Pacific Graphics, GMP, etc. His recent research interests include point cloud processing, 3D computer vision, computation fabrication, user interaction, and data visualization.

Supplementary material

Point Set Self-Embedding

IEEE Transactions on Visualization and Computer Graphics

OVERVIEW

This supplemental material consists of the following sections:

- In Section A, we provide more visual restoration results on point sets of various structures.
- In Section B, we provide one more visual example to demonstrate the importance of the self-embedded information.
- In Section C, we provide more visual comparison results with different downsampling methods.
- In Section D, we provide more visual comparison results on inputs of varying sizes.
- In Section E, we provide more visual comparison results on large-scale real-scanned inputs.
- In Section F, we provide a visual comparison with PointLIE [1].

APPENDIX A MORE RESTORATION VISUALIZATION RESULTS

Figures 1-3 demonstrate the restoration ability of our framework on point clouds of various geometric structures and point distributions from the uniform, random, and partial test sets, respectively. Benefited by the self-embedded information in (b), our framework restores high-quality dense point sets (c) that are very similar to the originals, regardless of the point distribution of the inputs (a).

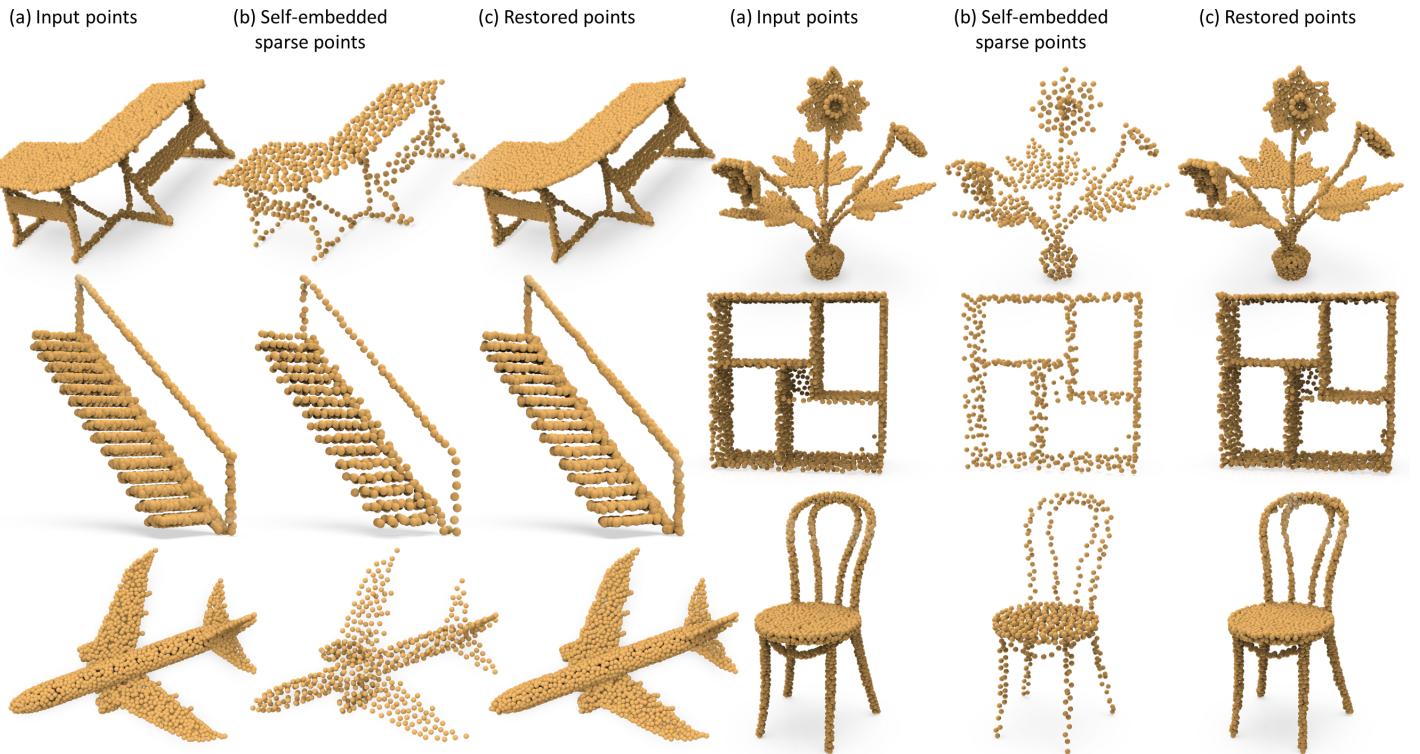


Fig. 1. Gallery of our restoration results on the *uniform set*. Given the original inputs (a), our method produces the self-embedded sparse points (b) and leverages the embedded information to achieve high-quality restorations (c).

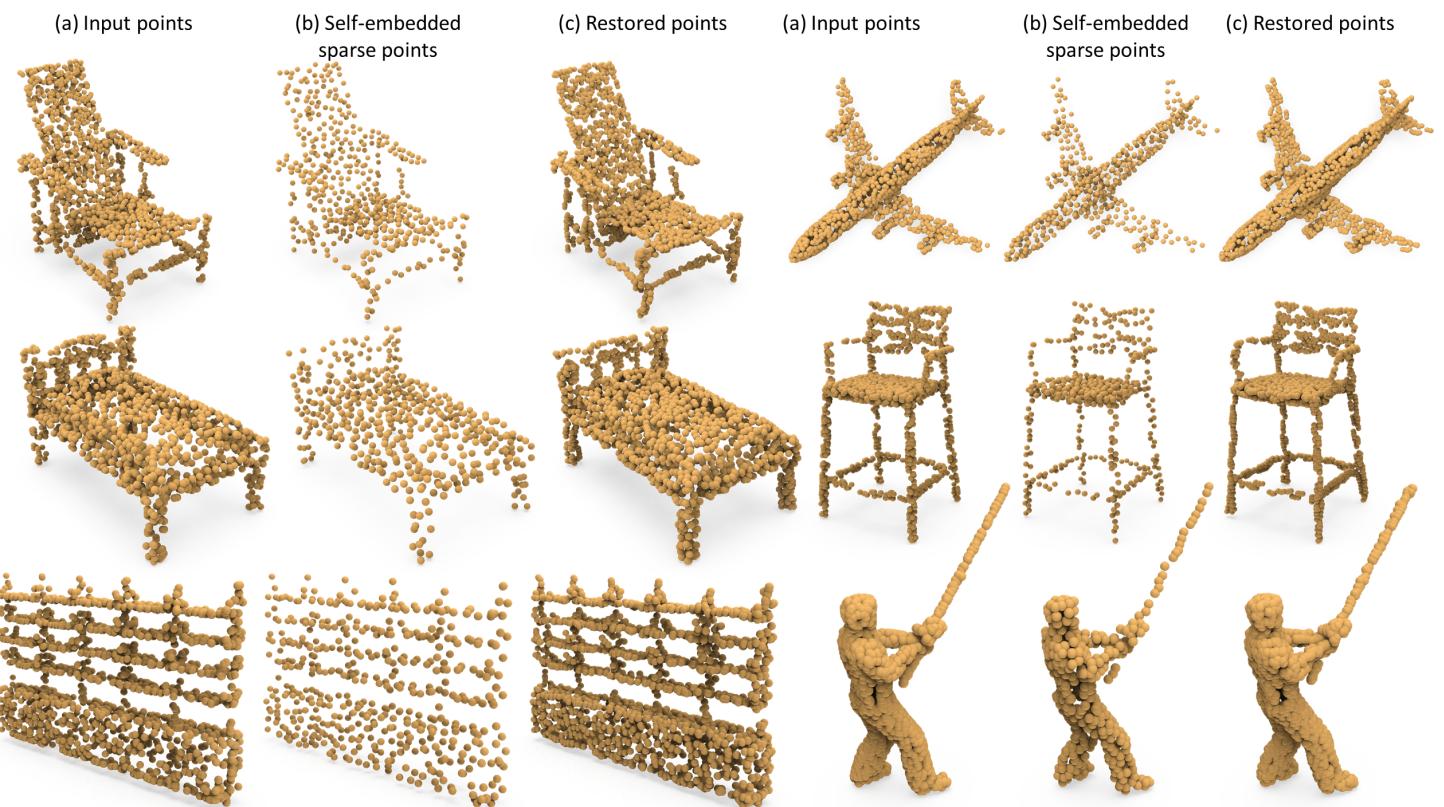


Fig. 2. Gallery of our restoration results on the *random set*. Given the original inputs (a), our method produces the self-embedded sparse points (b) and leverages the embedded information to achieve high-quality restorations (c).

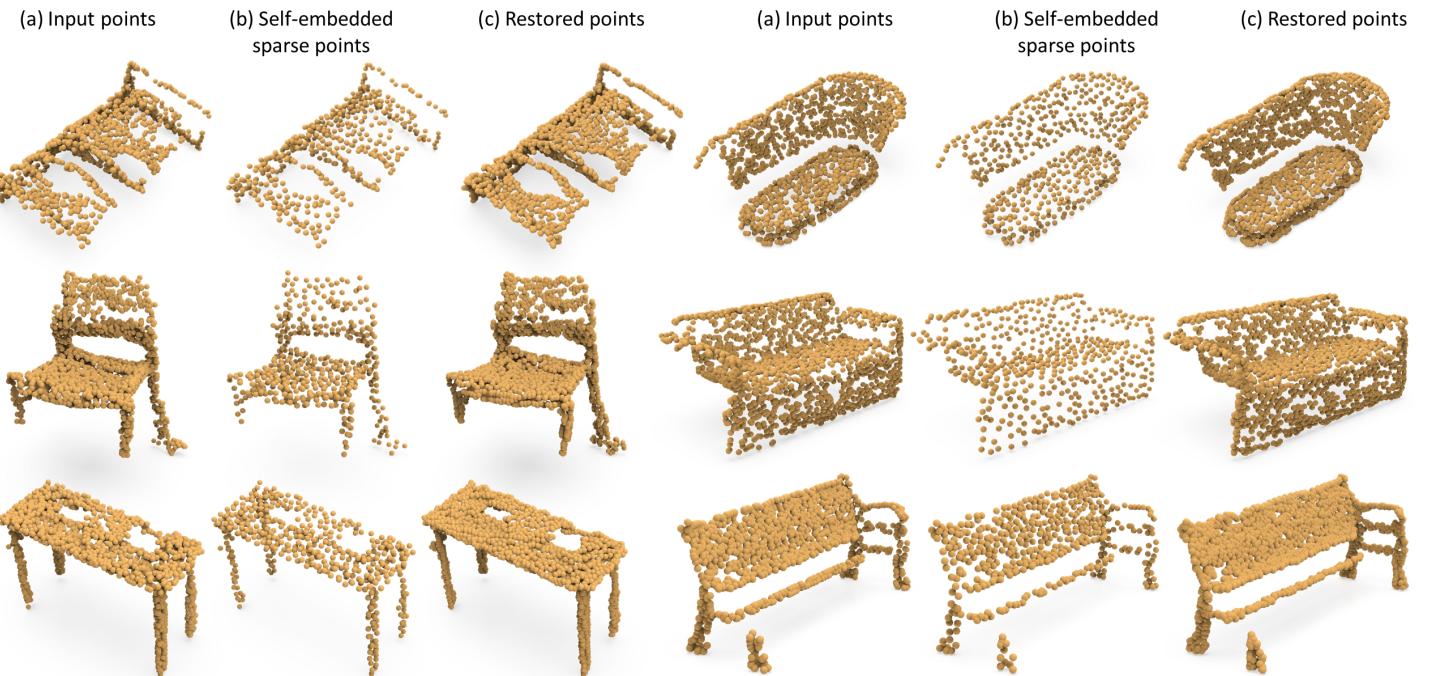


Fig. 3. Gallery of our restoration results on the *partial* set. Given the original inputs (a), our method produces the self-embedded sparse points (b) and leverages the embedded information to achieve a high-quality restoration (c).

APPENDIX B

MORE VISUAL ANALYSIS ON SELF-EMBEDDED INFORMATION

Figure 4 shows one more visual result to demonstrate the importance of the self-embedded information in the restoration. Please refer to Section 4.6 in the main paper for a detailed description.

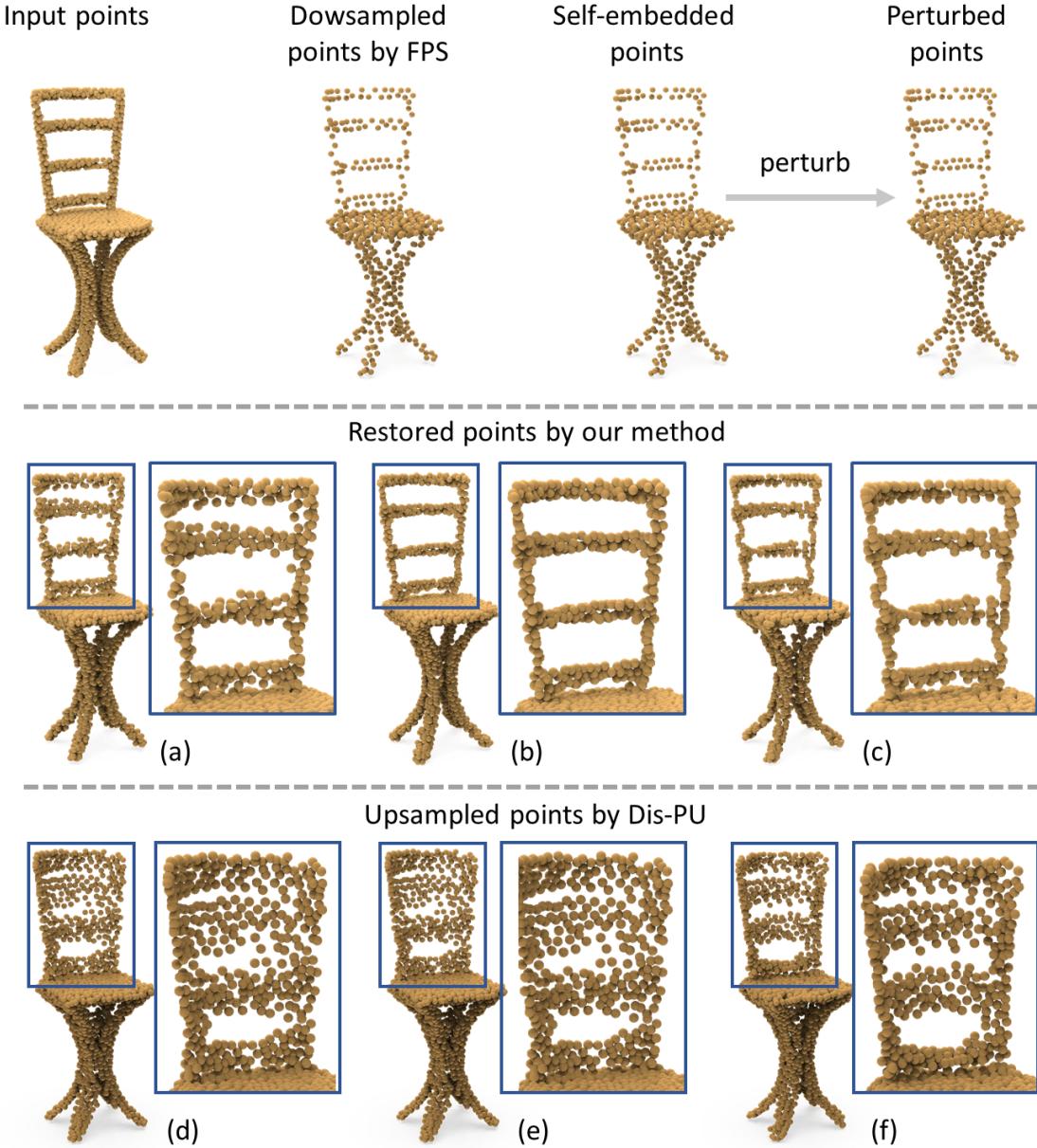


Fig. 4. Top row, from left to right: input points, downsampled points via FPS, our self-embedded points, and perturbed self-embedded points, (a)-(c) show restoration results produced by our methods, correspondingly from the three sparse inputs on the top row, whereas (d)-(f) show the corresponding upsampled points by Dis-PU [2]. We can see that the point set restored from our self-embedded points (b) achieve the best restoration vs. other results, showing the importance of the self-embedded information and that our method is able to consume the self-embedded information to produce the high-quality restoration.

APPENDIX C

COMPARISON RESULTS WITH DOWNSAMPLING METHODS

Figure 5 shows one more comparison result with FPS and SampleNet [3] (corresponding also to Figure 7 in the main paper).

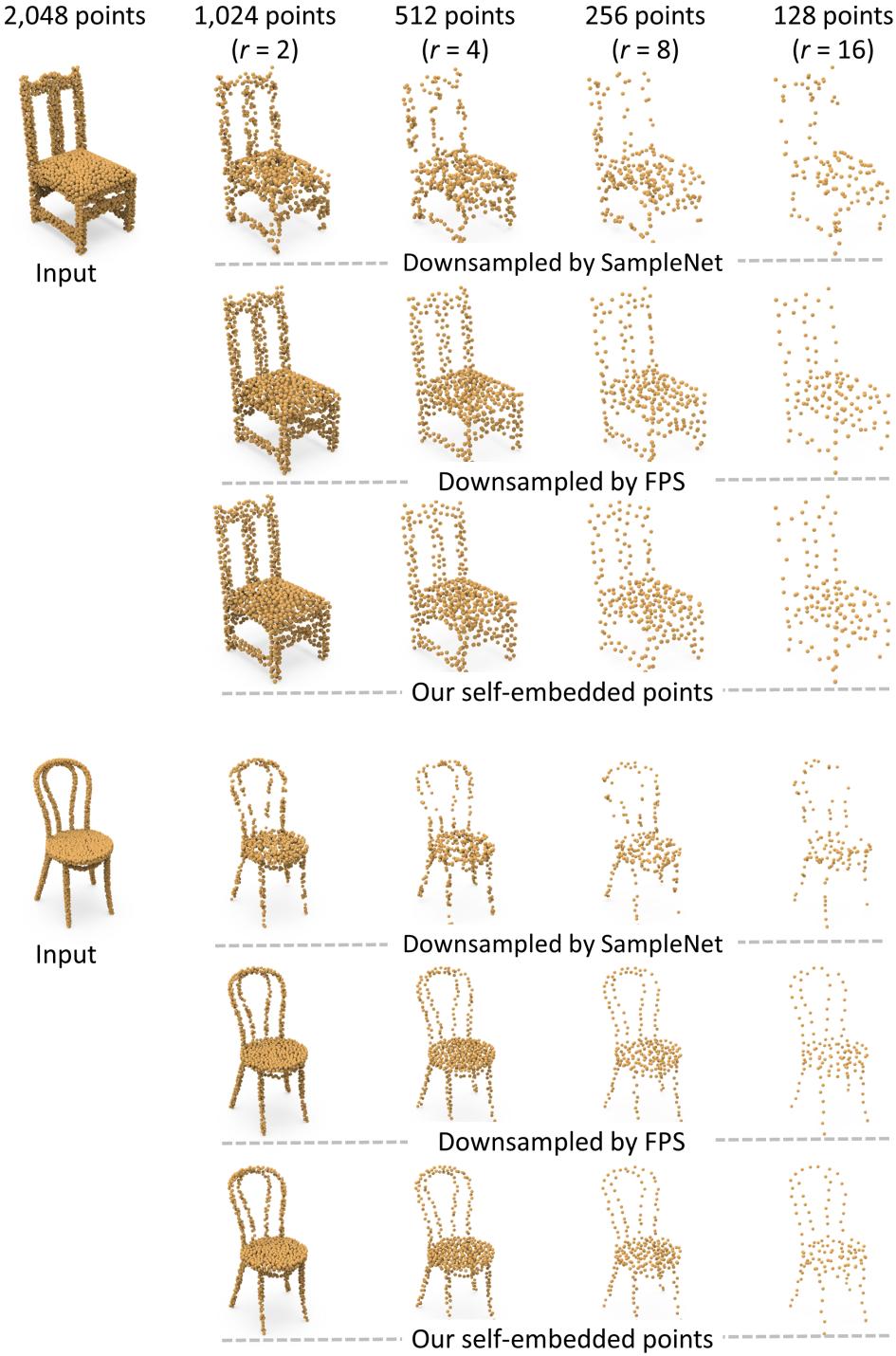


Fig. 5. Comparing the downsampling results produced by SampleNet [3] (top), FPS (middle), and our method (bottom) for decreasing sampling rates. Our self-embedded point sets are more similar to those produced by the ordinary FPS, presenting visually-recognizable shapes as the originals. In contrast, it is harder to recognize the results of SampleNet, particularly for large downsampling rates.

APPENDIX D

MORE COMPARISON RESULTS IN INPUTS OF VARYING SIZES

Figure 6 shows two more restoration results on restoring point sets of varying sizes (additional to Figure 11 in main paper). The results show that our method performs stably for varying input sizes.

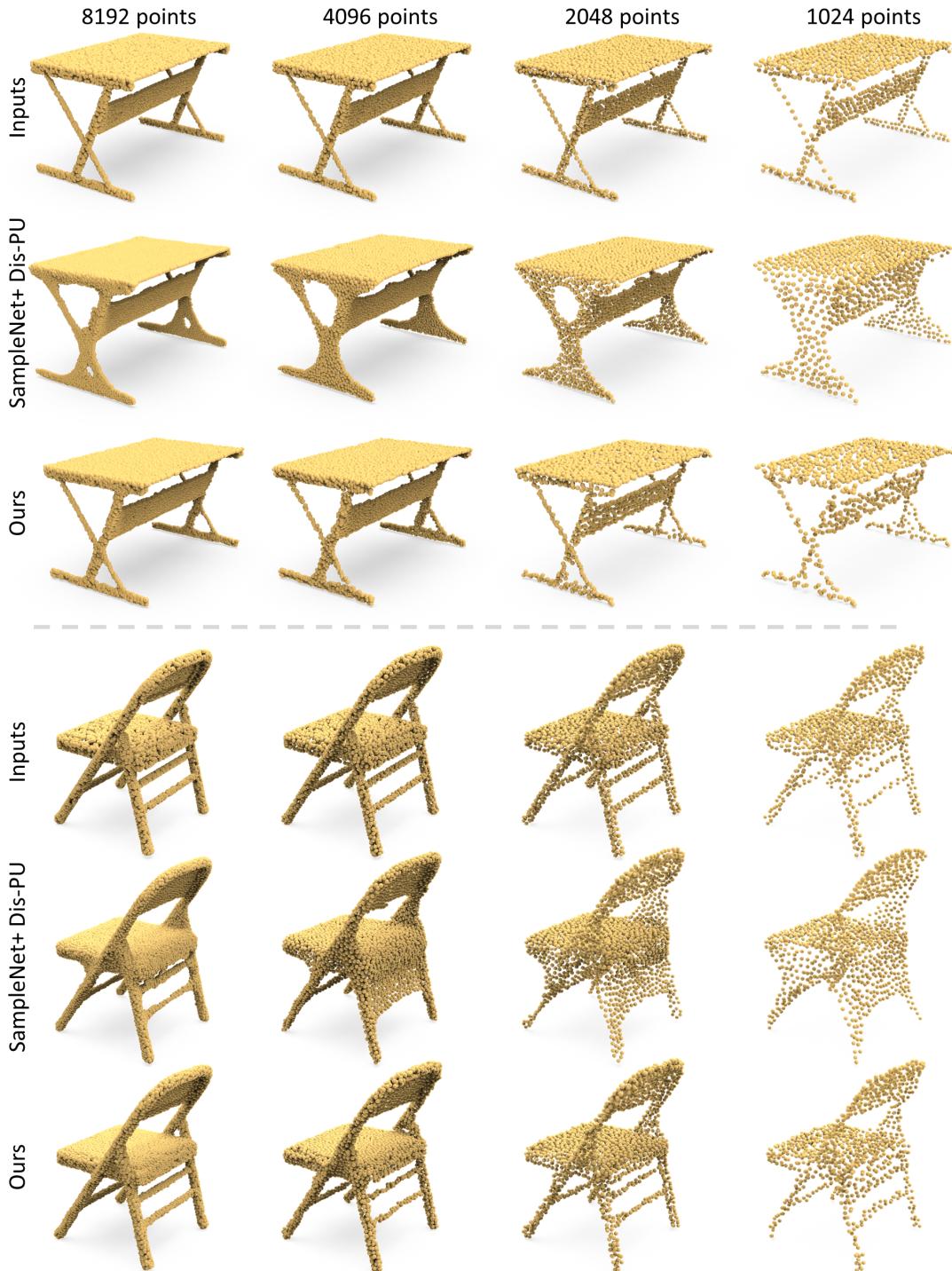


Fig. 6. Restoration results on input point sets of varying sizes (left to right).

APPENDIX E

MORE COMPARISON RESULTS ON LARGE-SCALE REAL-SCANNED INPUTS

Figure 7 shows more restoration results on large-scale real-scanned inputs (left column) additional to Figure 9 in the main paper, comparing the results of our method (right column) with SampleNet+Dis-PU (middle column). Note that, the input in row (a) is from the outdoor Waymo open dataset [4], while the input in row(b) is from the indoor ScanNet [5] dataset, and row (c) shows the associated reconstructed meshes. Our restored results are significantly much closer to the inputs, *e.g.*, in terms of preserving the original scanlines in (a) and well recovering the details in the inputs in (b), thus promoting an accurate and similar surface reconstruction in (c) as the original.

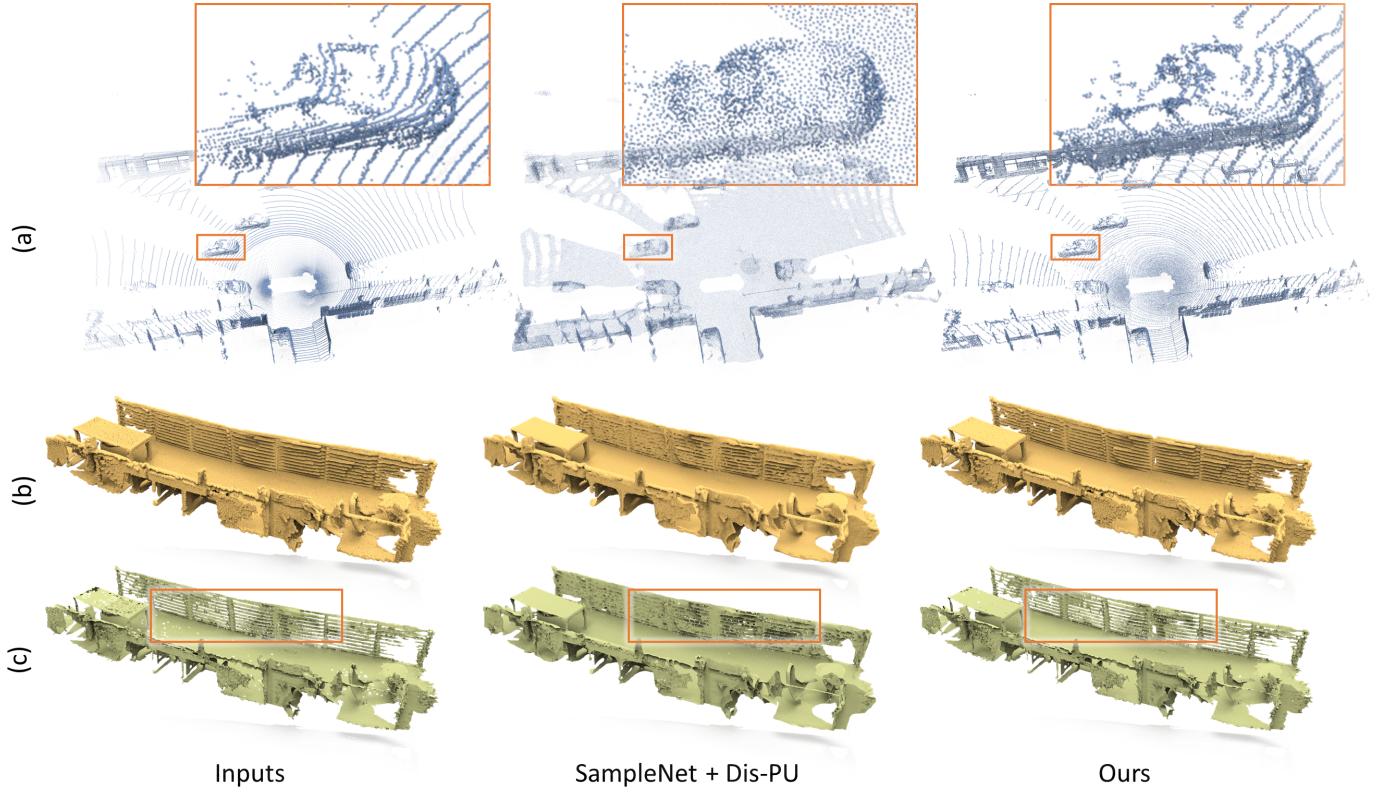


Fig. 7. Restoration results on large-scale real scans. (a) Results on the outdoor Waymo open dataset [4] (b)&(c) Results on the indoor ScanNet [5] and associated 3D reconstructed meshes. Compared with SampleNet+Dis-PU, our results exhibit details that are significantly closer to the inputs.

APPENDIX F

VISUAL COMPARISON WITH POINTLIE

Last, we present a visual comparison between our method and PointLIE [1]. First, note again that for PointLIE, there is no publicly released code so far, so we directly show the visual result from Figure 7 of its original paper [1] on Figure 8(a). Comparing with our result in Figure 8(b), we can see that our result is significantly closer to the inputs, better preserving the original scanlines, whereas PointLIE will smooth out the original details by filling the gaps, similar to the conventional upsampling methods.

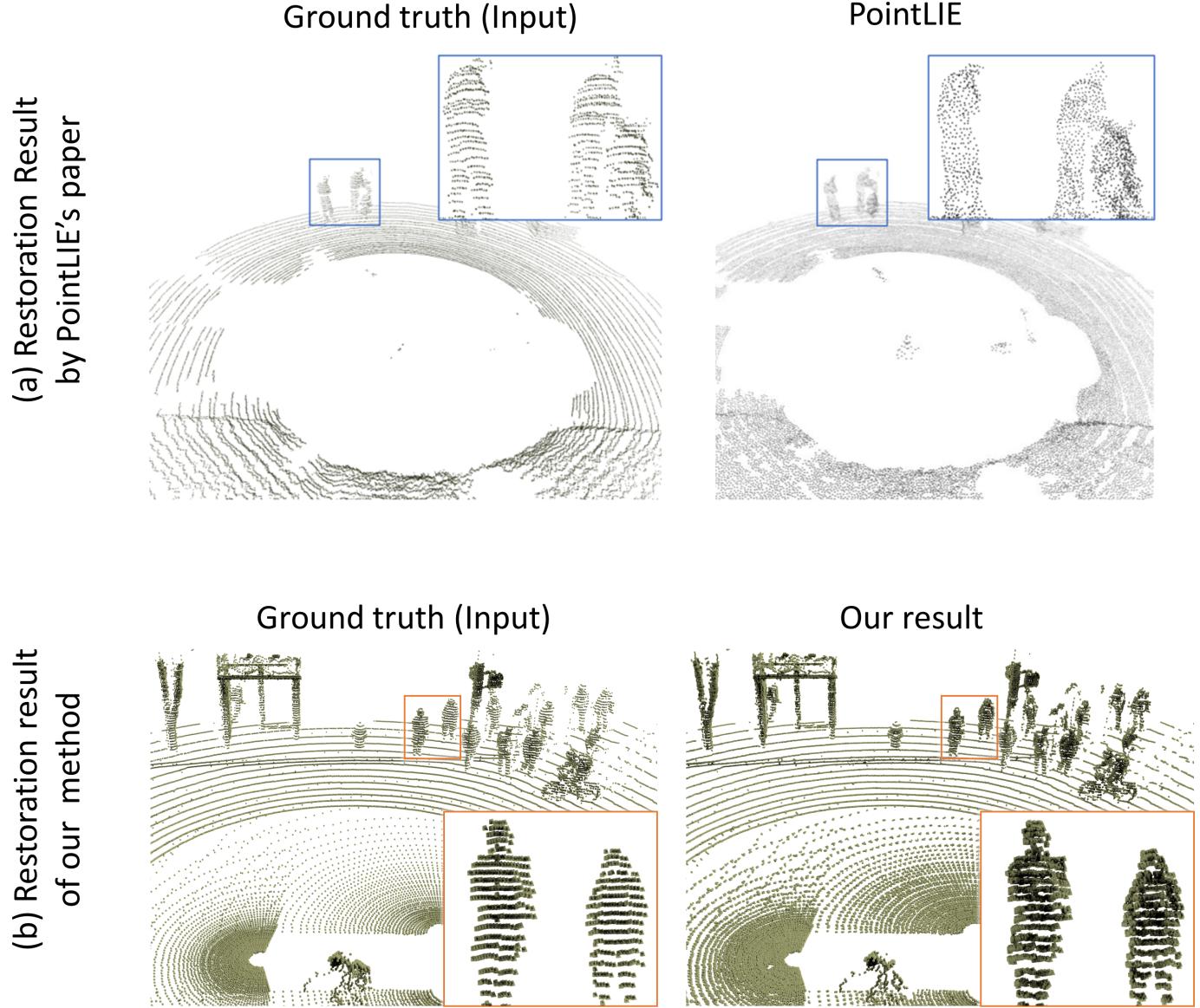


Fig. 8. Restoration on LiDAR inputs. (a): the top row shows a result from PointLIE's paper (the top row is from Figure 7 in PointLIE's appendix); and (b): the bottom row shows a result of our method.

REFERENCES

- [1] W. Zhao, X. Yan, J. Gao, R. Zhang, J. Zhang, Z. Li, S. Wu, and S. Cui, “PointLIE: Locally invertible embedding for point cloud sampling and recovery,” in *Intl. Joint Conf. on Artificial Intell. (IJCAI)*, 2021, pp. 1345–1351. [1](#), [9](#)
- [2] R. Li, X. Li, P.-A. Heng, and C.-W. Fu, “Point cloud upsampling via disentangled refinement,” in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 344–353. [5](#)
- [3] I. Lang, A. Manor, and S. Avidan, “SampleNet: Differentiable point cloud sampling,” in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 7578–7588. [6](#)
- [4] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine *et al.*, “Scalability in perception for autonomous driving: Waymo open dataset,” in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 2446–2454. [8](#)
- [5] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, “ScanNet: Richly-annotated 3D reconstructions of indoor scenes,” in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5828–5839. [8](#)

The End