AEDNet: Adaptive Embedding and Multiview-Aware Disentanglement for Point Cloud Completion

Zhiheng Fu[1], Longguang Wang[4], Lian Xu[1], Zhiyong Wang[2], Hamid Laga[3], Yulan Guo*[4], Farid Boussaid[1], and Mohammed Bennamoun[1]

[1] The University of Western Australia
[2] The University of Sydney [3] Murdoch University
[4] The Shenzhen Campus of Sun Yat-Sen University, Sun Yat-sen University

Abstract. Point cloud completion involves inferring missing parts of 3D objects from incomplete point cloud data. It requires a model that understands the global structure of the object and reconstructs local details. To this end, we propose a global perception and local attention network, termed AEDNet, for point cloud completion. The proposed AEDNet utilizes designed adaptive point cloud embedding and disentanglement (AED) module in both the encoder and decoder to globally embed and locally disentangle the given point cloud. In the AED module, we introduce a global embedding operator that employs the devised slot attention to compose point clouds into different embeddings, each focusing on specific parts of 3D objects. Then, we proposed a multiview-aware disentanglement operator to disentangle geometric information from those embeddings in the 3D viewpoints generated on a unit sphere. These 3D viewpoints enable us to observe point clouds from the outside rather than from within, resulting in a comprehensive understanding of their geometry. Additionally, the arbitrary number of points and point-wise features can be disentangled by changing the number of viewpoints, reaching high flexibility. Experiments show that our proposed method achieves state-of-the-art results on both MVP and PCN datasets.

Keywords: Point Cloud · Point Cloud Completion · Slot Attention

1 Introduction

The field of 3D computer vision has seen considerable growth, driven by the development of technologies such as Light Detection And Ranging (LiDAR) and depth cameras. Yet, the 3D models generated by these sensors are often incomplete and sparse due to inherent limitations in sensor resolution and obstacles with objects only partially visible from the sensor's perspective. These shortcomings pose significant challenges for downstream applications, including robotic manipulation and autonomous vehicle navigation, highlighting the importance of accurately reconstructing full, detailed 3D structures from incomplete scans [11, 50].

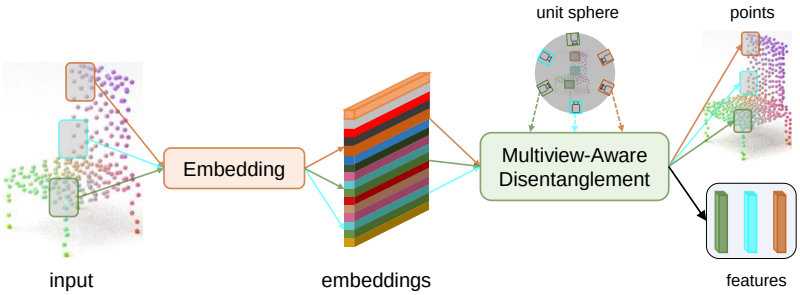


Fig. 1: Illustration of AED Module in AEDNet. The geometric patterns of the input point clouds are initially organized into separate embeddings and each embedding focuses on a specific part of a 3D object. Next, 3D viewpoints ("'cameras') are evenly distributed around a unit sphere using the Fibonacci point sequence generation algorithm, and they are adapted to the model's surface based on these embeddings. An arbitrary number of samples can be obtained by varying the count of 3D viewpoints.

Point cloud object completion involves creating complete 3D models from partial 3D scans. The encoder-decoder framework is commonly employed in deep learning approaches for this task, as noted in several studies [1, 14, 23, 24, 29, 36, 38, 41, 47, 48, 52, 54]. During the encoding phase, the challenge lies in capturing local geometries, particularly in defining an ambiguous local neighborhood. Typically, anchor points are first selected using farthest point sampling (FPS) followed by a k-nearest neighbor (KNN) search to establish local neighborhoods for feature extraction [23, 38, 41, 52]. However, FPS's sensitivity to noise has been highlighted in [18, 35, 55]. This leads to the development of FPS-inspired methods [35, 44, 44, 55] that refine point selection through attention mechanisms. However, these methods typically rely solely on local geometric information instead of considering the global structure, leading to sub-optimal sampling for point cloud completion. In the decoding phase, a common technique involves encoding the observed points into a global feature vector to predict the complete points via a generative process based on this vector [48]. Yet, this approach often struggles to generate high-quality shapes since a global feature vector cannot fully represent the varied patterns within an object. Recently, there has been a growing use of local pattern query techniques [1, 47, 52], which employ anchor points as queries to establish correlations with complete 3D shapes through the transformer [30] model. This method enhances the accuracy of predicting an initial coarse object shape, which is then further refined using foldingNet [45] or deconvolution operations [41, 52]. The concatenation approach employed by vanilla folding methods restricts their ability to generate intricate and faithful shapes [42], while deconvolutions lack the ability to capture necessary global context for comprehensive shape perception. Consequently, there is a need for a framework that integrates both global perception and local attention, specifically tailored for point cloud completion.

In this paper, we address this need by proposing a global perception and local attention network. We propose an Adaptive point cloud Embedding and Disentanglement (AED) module, including a global embedding operator and a

multiview-aware disentanglement operator (refer to Fig. 1). It first compresses input point clouds into embeddings, allowing each embedding to focus on a specific part of a 3D object. These embeddings act as global geometric representations of the input point cloud. Then, we propose to employ multi-view information to further enhance local geometric disentanglement. Particularly, we approach embedding learning as a form of set learning and propose a devised slot attention mechanism to automatically group input point clouds into distinct embeddings. Specifically, we replace Gaussian noise initialization, used in [19], with dictionary embedding and apply self-attention to further globally compose embeddings, thereby enhancing the ability to capture the complex structures within point clouds. For the multiview-aware disentanglement operator, we first use the Fibonacci point sequence generation algorithm [31] to evenly distribute 3D viewpoints ("cameras" in Fig. 1) on a unit sphere, and then encode the relationships between these viewpoints and the input point cloud into viewpoint encodings. Then, geometric details from these embeddings are locally disentangled to points on a 3D object's surface as well as point-wise features within these viewpoints. Different from previous methods [1, 41, 48, 52], the generated 3D viewpoints enable us to observe objects from the outside rather than from within, resulting in a more comprehensive understanding of their geometry.

In the encoder, we use the proposed AED module in conjunction with the point transformer [51] to hierarchically extract point-wise features and sample anchor points from input partial point clouds. Rather than relying solely on FPS for k-nearest neighbor aggregation [26], our AED module can establish long-distance dependencies among input points, resulting in higher-quality representation. Moreover, the anchor points are sampled with a global understanding of input point clouds in a learning manner instead of using FPS, making less sensitive to noisy input. In the decoder, we follow a coarse-to-fine completion process by hierarchically employing the proposed module. After obtaining the coarse prediction, the proposed AED module first aggregates its geometric patterns into embeddings and then disentangles more points from it by adjusting the number of 3D viewpoints sampled from a unit sphere. Thus, the proposed AED module can be used as a down-sampler in the encoder and an up-sampler in the decoder.

Our experimental results confirm that our proposed AEDNet attains leading performance on both the MVP [24] and PCN datasets [48]. Our contributions are as follows:

- We introduce AEDNet, a novel network specifically designed for point cloud completion, aimed at significantly enhancing global geometry embedding and local disentanglement.
- We propose a devised slot attention mechanism to significantly enhance the embedding of complex structures within point clouds.
- We propose a multiview-aware disentanglement to improve local geometry reconstruction for point cloud completion.
- Our approach achieves unparalleled completion results on the MVP and PCN benchmark datasets, achieving state-of-the-art performance in the field.

2 Related Work

In this section, we examine existing research on point cloud processing, focusing on point cloud completion and slot attention.

Point Cloud Completion. Historically, significant progress in 3D reconstruction and shape completion has been made using structured volumetric methods and robust 3D convolutions [3, 4, 7, 12, 40]. These approaches, however, come with high computational and memory requirements. Sparse representation techniques [28, 33] attempt to address these issues, but they often result in the loss of detailed information due to the quantization involved.

Recent shifts toward unstructured point clouds as representations for 3D objects have helped to reduce memory usage and better preserve fine details. This transition brings new challenges, as standard convolution operations do not translate well to the unordered nature of point clouds. Innovations like PointNet and its variants [25, 26] have allowed for the direct handling of 3D points across multiple downstream tasks. The PCN network [48] adopts a global feature extraction method inspired by PointNet [25] and introduces a folding technique [45] for point generation. Efforts to capture local structures in point clouds have led to multi-scale feature extraction methods [49]. Lyu et al. [21] approached point cloud completion as a conditional generation task using denoising diffusion probabilistic models (DDPM) [13, 20, 53]. While their method achieves fine-grained completion using a simple mean squared error loss function [27], it is computationally demanding, limiting its use to preliminary stages of coarse point cloud generation. More recently, Chen et al. [2] leverages the conditional DDPM in the 3D latent space for shape reconstruction. Examining the role of viewpoint information, Fu et al. [6] found it enhances completion quality and performance but requires an additional trained model for viewpoint representation learning. More advanced architectures like SnowflakeNet [41] and PointTr [47] emphasize decoder structures with Transformer-like designs [30], and PointAttN [32] and CompleteDT [17] introduce a Transformer-based model tailored to point cloud completion. Recently, Seedformer [52] and Acchorformer [1] explored the impact of local pattern propagation both in seed generation and point up-sampling, highlighting the importance of local geometry propagation for detailed point cloud completion. Differently, in this paper, we propose a multiview-aware disentanglement to recover the details of 3D shapes.

Slot Attention. Slot Attention Networks (SANs) infer a set of latent variables, each representing an object within the image. Methods focused on "object-centric learning" [5, 8, 9, 16, 19] strive to identify generative factors that correlate with parts or objects in the scene. SANs employ a feed-forward pass that uses the attention mechanism [30] to assign latent variables to permutation-invariant slots, making this feature ideal for deep learning approaches to point cloud processing. In this paper, slot attention serves as a dynamic clustering tool, automatically partitioning the input point cloud into distinct local areas and capturing the geometric details of these areas into slots. This approach differs from the grouping layer in PointNet++ [26] as our aggregation operates in a high-dimensional space without relying on Farthest Point Sampling (FPS),

thereby minimizing the adverse effects of noisy input data. However, current slot attention-based segmentation techniques [8, 16, 19] exhibit a notable drawback: they predominantly perform well on synthetic images featuring uniform colors and layouts or on objects distinctly separated by color and shape, or simple textures. Their performance deteriorates in more complex, real-world environments due to several factors, such as the challenge of randomly initializing slots to represent meaningful context in intricate scenes and the lack of established relationships among different slots, resulting in a weak semantic linkage between them.

To address these issues, we propose altering the initial slot attention approach by applying dictionary mapping for initialization and by constructing inter-slot relationships. Furthermore, we introduce an adaptive embedding and disentanglement method that leverages the proposed slot attention together with a new local geometry disentanglement module, targeting enhanced accuracy and structural coherence.

3 Methodology

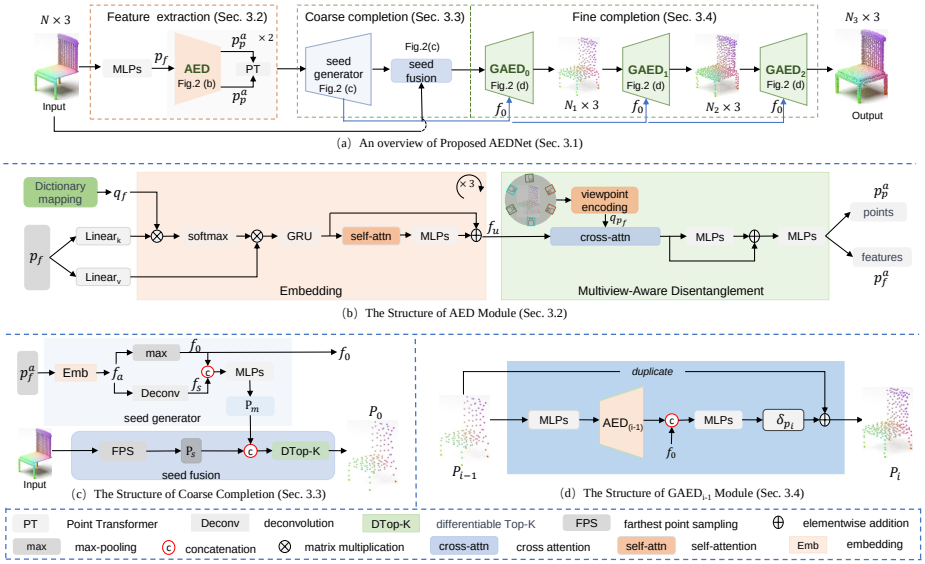
3.1 Overview

Our point cloud completion framework, as shown in Fig. 2(a), operates through two principal phases: encoder and decoder. The encoder stage is responsible for extracting features, while the decoder stage employs a coarse-to-fine strategy for completion. Initially, a partial scan is fed to the encoder, where it is processed by our proposed AED module in conjunction with a point transformer (PT), as depicted in Fig. 2(b). This setup allows for the hierarchical extraction of point-wise features. Subsequently, these features are forwarded to the coarse completion stage, as shown in Fig. 2(c), where the preliminary shape of the complete 3D object and its global feature is predicted. The final phase of this process is the refinement phase, where the initial shape is enhanced through systematic up-sampling of the point cloud, facilitated by our GAED module, a procedure illustrated in Fig. 2(d).

3.2 Feature Extraction

Many current point cloud completion methods [1, 38, 41, 52] rely on Farthest Point Sampling (FPS) and K-Nearest Neighbors (KNN) to identify the local neighborhoods of points for local feature extraction, making them vulnerable to noise. In this paper, we use the proposed AED and PT techniques to perform feature extraction without the need of FPS. For an incomplete point cloud $P_{in} \in \mathbb{R}^{N\times 3}$, we apply our AED module (illustrated in Fig. 2(b)) to execute both embedding and disentanglement operations. This process results in M points (with $M < N$) and their associated features, which are then refined using PT.

AED: Embedding (Fig. 2(b)). "Object-centric learning" methods [5, 8, 9, 16, 19] leverage slot attention to uncover generative factors related to components or



(a) An overview of Proposed AEDNet (Sec. 3.1)

(b) The Structure of AED Module (Sec. 3.2)

(c) The Structure of Coarse Completion (Sec. 3.3)

(d) The Structure of GAED$_{i-1}$ Module (Sec. 3.4)

Fig. 2: An Overview of AEDNet. (a) The complete AEDNet framework encompasses the feature extraction, coarse and fine completion processes. (b) The detailed structure of the proposed AED module, includes an embedding operator and a multiview-aware disentanglement operator. (c) The coarse completion phase, includes a seed generator and a seed fusion technique to predict the initial shape of the 3D object. (d) Finally, the fine completion phase involves the use of GAED modules for the hierarchical upsampling of the point cloud to achieve the final detailed output.

objects within a scene. However, traditional slot attention techniques primarily excel with synthesized images characterized by uniform colors and layouts, or objects easily distinguishable by color, shape, or simple textures, but struggle to handle the complexities of real-world scenes. The diverse shapes of different objects and recurring similar structures within a single object present a substantial challenge for conventional slot attention in point cloud completion tasks. To address this, we adapt slot attention by refining how slots are initialized and enhancing the semantic interactions between slots.

Specifically, we initially use multi-layer perceptrons to generate point-wise features p_f, which are then processed through two linear layers to produce key and value features. We improve on the standard slot attention mechanism [19] by employing a learnable dictionary for mapping, as opposed to using Gaussian-initialized slots, yielding a set of slot features $q_{f_0} \in \mathbb{R}^{M \times C}$, with $M < N$. These slot features, along with the key and value features, are introduced to the attention mechanism and a Gated Recurrent Unit (GRU) for updating the slot features. Furthermore, we apply self-attention in conjunction with MLPs to forge semantic links among slot features, resulting in the final slot feature q_{f_i}. This embedding process is iteratively conducted three times (i.e., i ranging from 1 to 3), resulting in the ultimate embedding feature $f_u = q_{f_3}$.

AED: Multiview-aware Disentanglement (Fig. 2(b)). Following the embedding of point features, we propose a viewpoint-aware method to disentangle

local geometry from these embeddings. We first use the Fibonacci point sequence generation algorithm [31] to evenly distribute n 3D viewpoints on a unit sphere. The produced 3D viewpoints allow us to view objects externally rather than internally, leading to a deeper grasp of their geometry and structure. Then, relative position encoding between viewpoints and sparse input point clouds is calculated using MLPs. These viewpoint encodings q_{p_f}, together with f_u, are then applied to a cross-attention module, effectively segregating geometric details and linking them to the viewpoint encodings q_{p_f}. Subsequent use of MLPs enables the regression of the coordinates and features of the sampled points p_p^a and p_f^a, respectively. A point transformer module [51] further refines the sampled features by analyzing the positional relations among the sampled points. Throughout this feature extraction phase, our AED module plays a crucial role in both reducing the size of the point cloud and accurately learning the features.

3.3 Coarse Completion

Seed Generator (Fig. 2(c)). As highlighted in previous studies [1, 41, 47, 52], the diffusion of local geometry is crucial for the successful completion of point clouds. In our approach, we employ an embedding and disentanglement strategy to predict the foundational structure of complete objects. After acquiring the extracted features f_e from the feature extraction phase, we proceed with our coarse completion technique (illustrated in Fig. 2(c)) to predict the initial complete point cloud, named as $P_0 \in \mathbb{R}^{N_0 \times 3}$. Specifically, we leverage our newly designed embedding module to compile f_e into f_a. Subsequently, f_a undergoes processing via two separate pathways: the first employs max-pooling to extract a global feature f_0, and the second uses deconvolution to expand f_a into f_s, diverging from SnowflakeNet's [41] method of directly generate point-wise features by enhancing the use of geometric information through the up-sampling of f_a. Following this, we merge f_0 with the up-sampled f_s and apply MLPs to generate N_0 points, represented by P_m.

Seed Fusion (Fig. 2(c)). Recognizing the presence of detailed structures within the incomplete inputs, we aim to integrate the predicted $P_m \in \mathbb{R}^{N_0 \times 3}$ with the original incomplete point cloud. This integration begins by combining P_m with P_s, which is derived from the incomplete point cloud P_{in} through Farthest Point Sampling (FPS). Following this, we employ a differentiable Top-K selection method to isolate N_0 points from this combined set, establishing the initial shape of the complete point cloud as $P_0 \in \mathbb{R}^{N_0 \times 3}$. To refine the accuracy of this initial shape selection, we have formulated a specialized loss function, which is detailed further in Section 3.5.

3.4 Fine Completion

To reconstruct objects with detailed completeness, we employ the proposed GAED method for the hierarchical up-sampling of the initial complete point cloud shape P_0, as shown in Fig. 2(d). Each iteration of GAED takes the previous point cloud prediction P_{i-1} and the global feature f_0 to derive point-wise

features p_{w_i} of a higher resolution. This approach differs from the feature extraction phase, where GAED samples a greater number of points from a unit sphere than the input points N_{i-1} in the up-sampling stage. Moreover, at this phase, we focus solely on predicting point-wise features, as opposed to predicting both points and point-wise features as done in the encoder stage. Subsequently, the point-wise features are concatenated with the global feature f_0 to estimate the residuals of the duplicated P_{i-1}.

3.5 Loss Functions

To quantify the disparity between two point clouds, we adopted the Chamfer Distance (CD) for its efficiency compared to Earth Mover's Distance (EMD).

$$\mathcal{L}_{CD}(P,G) = \frac{1}{\|P\|_1}\sum_{x\in P}\min_{y\in G}\|x-y\|_2 + \frac{1}{\|G\|_1}\sum_{y\in G}\min_{x\in P}\|y-x\|_2, \tag{1}$$

where P and G represent the predicted complete point clouds and the actual ground truth, respectively.

To explicitly constrain point clouds sampled from the input point cloud, we treat the input point cloud as ground truth and use a variant CD loss.

$$\mathcal{L}_{sampling}(P_{in},P_s) = \lambda\frac{1}{\|P_s\|_1}\sum_{x\in P_s}\min_{y\in P_{in}}\|x-y\|_2 + \frac{1}{\|P_{in}\|_1}\sum_{y\in P_{in}}\min_{x\in P_s}\|y-x\|_2, \tag{2}$$

In this case, $\lambda = \|P_{in}\|_1/\|P_s\|_1$ where P_{in} and P_s are the input points and sampled points.

To impose specific constraints on point clouds created during coarse and fine completion, we down-sampled the ground truth point clouds to match the sampling density of P_0, P_1, P_2, P_3. During coarse completion, we apply the Chamfer Distance (CD) loss in combination with the repulsion loss [46] ($\mathcal{L}_{rep}$) to form the seed loss, labeled $\mathcal{L}_{seed}$. In the fine completion stage, the aggregate of the three CD losses is termed the completion loss, represented by $\mathcal{L}_{completion}$.

$$\mathcal{L}_{seed} = \mathcal{L}_{CD}(P_0, gt_0) + \lambda_s\mathcal{L}_{rep}(P_0), \tag{3}$$

$$\mathcal{L}_{generation} = \mathcal{L}_{CD}(P_1, gt_1) + \mathcal{L}_{CD}(P_2, gt_2) + \mathcal{L}_{CD}(P_3, gt_3), \tag{4}$$

where gt_0, gt_1, gt_2, gt_3 are down-sampled ground truth point clouds corresponding to P_0, P_1, P_2, P_3. λ_s here is set to 0.05.

We also exploit the partial matching loss from [37] to preserve the structural shape integrity of the input point cloud. This is a unidirectional constraint designed to align one shape to another. The partial matching loss ensures that the output point cloud partially matches the input to a certain extent, which we refer to as the preservation loss, $\mathcal{L}_{preservation}$. The overall training loss is:

$$\mathcal{L} = \mathcal{L}_{sampling} + \mathcal{L}_{seed} + \mathcal{L}_{generation} + \mathcal{L}_{preservation}. \tag{5}$$

4 Experiments

In this section, we will first introduce the datasets and discuss the implementation details (Sec. 4.1), then move on to present our completion results on both the MVP (Sec. 4.2) and PCN datasets (Sec. 4.3).

4.1 Datasets and Implementation Details

MVP Dataset [24]: The MVP dataset includes 16 categories with 4000 CAD models. Each model is virtually scanned from 26 camera positions to create partial scans.
PCN Dataset [48]: Originating from a subset of the ShapeNet dataset [4], the PCN dataset includes complete point clouds of 16384 points against incomplete point clouds with 2048 points. The training set has 28,974 models across 8 categories.
Implementation Details: For feature extraction, we used two down-sampling operations in our AEDNet to generate 512 and then 64 points. For coarse completion, we predicted an initial complete shape with 512 points. During fine completion, we performed three up-sampling operations with our AED module, resulting in 512, 1024, and 2048 points, respectively. Specifically, we performed uniform sampling of 512, 1024, and 2048 points from the unit sphere during the three consecutive AED module operations. We opted for the Adam optimization method [15] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ for training, running for 50 epochs on the MVP dataset and 400 on the PCN dataset. The learning rate was set to 10^{-4}, and was reduced by 30% every 20 epochs. All tests were conducted on an NVIDIA 3090Ti GPU.

4.2 Completion on the MVP Dataset

We assessed the MVP dataset against a variety of leading-edge baseline methods using the $\mathcal{L}_2$ Chamfer Distance and F-Score@1% as performance metrics. The results for the baseline methods [1, 21, 24, 35, 41] were produced from the codes and pre-trained models available in their respective official Github projects. The results for the remaining methods were taken directly from [21, 24] and the original publication [6].

Table 1: Shape completion results (CD loss multiplied by 10^4) on the multi-view partial (MVP) point cloud dataset (16,384 points). The lower, the better.

Method	airplane	cabinet	car	chair	lamp	sofa	table	watercraft	bed	bench	bookshelf	bus	guitar	motorbike	pistol	skateboard	Avg.
PCN [48]	2.95	4.13	3.04	7.07	14.93	5.56	7.06	6.08	12.72	5.73	6.91	2.46	1.02	3.53	3.28	2.99	6.02
TopNet [29]	2.72	4.25	3.40	7.95	17.01	6.04	7.42	6.04	11.56	5.62	8.22	2.37	1.37	3.90	3.97	2.09	6.36
MSN [18]	2.07	3.82	2.76	6.21	12.72	4.74	5.32	4.80	9.93	3.89	5.85	2.12	0.69	2.48	2.91	1.58	4.90
Wang et al. [34]	1.59	3.64	2.60	5.24	9.02	4.42	5.45	4.26	9.56	3.67	5.34	2.23	0.79	2.23	2.86	2.13	4.30
ECG [23]	1.41	3.44	2.36	4.58	6.95	3.81	4.27	3.38	7.46	3.10	4.82	1.99	0.59	2.05	2.31	1.66	3.58
GRNet [43]	1.61	4.66	3.10	4.72	5.66	4.61	4.85	3.53	7.82	2.96	4.58	2.97	1.28	2.24	2.11	1.61	3.87
NSFA [49]	1.51	4.24	2.75	4.68	6.04	4.29	4.84	3.02	7.93	3.87	5.99	2.21	0.78	1.73	2.04	2.14	3.77
VRCNet [24]	1.15	3.20	2.14	3.58	5.57	3.58	4.17	2.47	6.90	2.76	3.45	1.78	0.59	1.52	1.83	1.57	3.12
Our AEDNet	**0.74**	**2.99**	**2.27**	**2.75**	**2.79**	**2.96**	**2.64**	**1.98**	**4.73**	**1.96**	**3.01**	**1.71**	**0.38**	**1.50**	**1.55**	**0.81**	**2.24**

Table 2: Shape completion results (F-Score@1%) on the multi-view partial (MVP) point cloud dataset (16,384 points). The higher, the better.

Method	airplane	cabinet	car	chair	lamp	sofa	table	watercraft	bed	bench	bookshelf	bus	guitar	motorbike	pistol	skateboard	Avg.
PCN [48]	0.861	0.641	0.686	0.517	0.455	0.552	0.646	0.628	0.452	0.694	0.546	0.779	0.906	0.665	0.774	0.861	0.638
TopNet [29]	0.798	0.621	0.612	0.443	0.387	0.506	0.639	0.609	0.405	0.680	0.524	0.766	0.868	0.619	0.726	0.837	0.601
MSN [18]	0.879	0.692	0.693	0.599	0.604	0.627	0.730	0.696	0.569	0.797	0.637	0.806	0.935	0.728	0.809	0.885	0.710
Wang et al. [34]	0.898	0.688	0.725	0.670	0.681	0.641	0.748	0.742	0.600	0.797	0.659	0.802	0.931	0.772	0.843	0.902	0.740
ECG [23]	0.906	0.680	0.716	0.683	0.734	0.651	0.766	0.753	0.640	0.822	0.706	0.804	0.945	0.780	0.835	0.897	0.753
GRNet [43]	0.861	0.641	0.686	0.517	0.455	0.552	0.646	0.628	0.452	0.694	0.546	0.779	0.906	0.665	0.774	0.861	0.638
NSFA [49]	0.903	0.694	0.721	0.737	0.783	0.705	0.817	0.799	0.687	0.845	0.747	0.815	0.932	0.815	0.858	0.894	0.783
VRCNet [24]	0.928	0.721	0.756	0.743	0.789	0.696	0.813	0.800	0.674	0.863	0.755	0.832	0.960	0.834	0.887	0.930	0.796
Our AEDNet	**0.947**	**0.766**	**0.757**	**0.794**	**0.850**	**0.754**	**0.856**	**0.828**	**0.735**	**0.893**	**0.807**	**0.848**	**0.974**	**0.843**	**0.899**	**0.956**	**0.832**

Table 3: Shape completion results (CD loss multiplied by 10^4) on multi-view partial point cloud (MVP) dataset with various point cloud resolutions.

#Points	2048		4096		8192		16384	
	CD↓	F1↑	CD↓	F1↑	CD↓	F1↑	CD↓	F1↑
PCN [48]	9.77	0.320	7.96	0.458	6.99	0.563	6.02	0.638
TopNet [29]	10.11	0.308	8.20	0.440	7.00	0.533	6.36	0.601
MSN [18]	7.90	0.432	6.17	0.585	5.42	0.659	4.90	0.710
Wang et al. [34]	7.25	0.434	5.83	0.569	4.90	0.680	4.30	0.740
ECG [23]	6.64	0.476	5.41	0.585	4.18	0.690	3.58	0.753
GRNet [43]	7.61	0.353	5.73	0.493	4.51	0.616	3.54	0.700
VRCNet [24]	5.96	0.499	4.70	0.636	3.64	0.727	3.12	0.791
PoinTr [47]	5.79	0.499	4.29	0.638	3.52	0.725	2.95	0.783
PMP-Net++ [39]	-	-	-	-	-	-	3.38	0.687
Wang et al. [35]	-	-	-	-	-	-	-	0.816
SnowflakeNet [41]	5.71	0.503	4.45	0.648	3.48	0.743	2.69	0.796
PDR [21]	5.66	0.499	4.26	0.649	3.35	0.754	2.61	0.817
VAPCNet [6]	5.40	0.521	3.96	0.658	3.02	0.763	2.40	0.829
AnchorFormer [1]	5.89	0.482	4.35	0.655	3.21	0.763	2.60	0.819
Our AEDNet	**5.12**	**0.522**	**3.75**	**0.675**	**2.90**	**0.770**	**2.24**	**0.832**

Quantitative comparison. The performance of all methods, measured by CD loss and F-score@1%, is reported in Tables 1 and 2. Our proposed AEDNet outperforms all other competitors in terms of CD and F-score@1%. In particular, our approach shows a substantial improvement, achieving nearly 50% better performance in the lamp category when compared to VRCNet. We also evaluated our method against others that support multi-resolution completion, as shown in Table 3. This comparison is important since our AED module is capable of performing various levels of down-sampling and up-sampling on point clouds. In this comparison, AEDNet demonstrated superior performance over all the compared methods.

Qualitative comparison. Visual comparison, as displayed in Fig. 3, shows that AEDNet can produce accurate complete shapes than competing methods. The effectiveness of our AED module is particularly evident in our completed shapes. For instance, the chair's missing legs (first row of Fig. 3) are recoverd by referencing the visible legs, taking into account the perspective from which they were scanned. Moreover, in the second row of Fig. 3, the lamp base is reconstructed more accurately, resulting in a better-shaped lampstand compared to other meth-

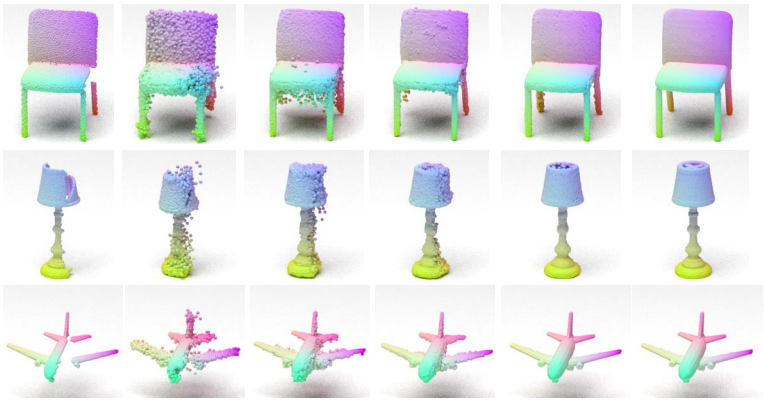


Fig. 3: Visual comparisons on MVP dataset. Note that, the partial point clouds (2048 points) are sparse and self-occluded, as opposed to the reconstructed and ground truth point clouds (16,384 points), which are dense and complete.

ods. This achievement is attributed to the embedding-and-disentangling strategy employed by AEDNet, which efficiently reconstructs complete shapes by learning structural relations at each up-sampling stage through the AED module.

4.3 Completion on the PCN Dataset

On the PCN dataset, we benchmarked our network against SOTA baseline methods. The $\mathcal{L}_1$ Chamfer Distance served as our metric for evaluation. The baseline methods' results of [23,49] were produced from the codes and pre-trained models provided in their official Github repositories. We gathered results for other methods from [38, 41, 43, 52] along with their respective original publications [35, 47].

Table 4: Quantitative comparison of SOTA methods on the PCN dataset, using $\mathcal{L}_1$ Chamfer Distance $\times 10^3$ as the evaluation metric. Lower $\mathcal{L}_1$ Chamfer Distance values indicate better performance.

Models	Avg.	airplane	cabinet	car	chair	lamp	couch	table	watercraft
AtlasNet [10]	10.58	6.37	11.94	10.10	12.06	12.37	12.99	10.33	10.61
FoldingNet [45]	14.31	9.49	15.80	12.61	15.55	16.41	15.97	13.65	14.99
PCN [48]	9.64	5.50	22.70	10.63	8.70	11.00	11.34	11.68	8.59
TopNet [29]	12.15	7.61	13.31	10.90	13.82	14.44	14.78	11.22	11.12
GRNet [43]	8.83	6.45	10.37	9.45	9.41	7.96	10.51	8.44	8.04
Wang et al. [34]	8.51	4.79	9.97	8.31	9.49	8.94	10.69	7.81	8.05
PMP-Net [38]	8.73	5.65	11.24	9.64	9.51	6.95	10.83	8.72	7.25
ECG [23]	8.63	5.23	10.12	8.36	9.43	8.53	10.94	7.98	8.16
NSFA [49]	8.32	5.03	10.51	9.11	9.16	7.45	10.46	7.56	7.28
SK-PCN [22]	8.49	5.09	9.98	8.22	9.29	8.39	10.80	7.84	8.02
PoinTr [47]	8.38	4.75	10.47	8.68	9.39	7.75	10.93	7.78	7.29
SnowflakeNet [24]	7.21	4.29	9.16	8.08	7.89	6.07	9.23	6.55	6.40
VAPCNet [6]	7.02	4.10	9.28	8.15	7.51	5.55	9.18	6.28	6.10
Anchorformer [1]	6.59	3.70	8.94	7.57	**7.05**	5.21	8.40	6.03	5.81
Our AEDNet	**6.52**	**3.61**	**8.90**	**7.51**	**7.05**	**5.15**	**8.32**	**5.82**	**5.74**

Quantitative comparison. The data, presented in Table 4, demonstrate that our network achieves the lowest average $\mathcal{L}_1$ Chamfer Distance (CD). Specifically, in the chair category, AEDNet displays performance on par with that of Anchorformer [1]. Across other categories, however, our approach surpasses Anchorformer. Unlike Anchorformer, which reconstructs the 3D shape from a set of key points, our proposed method reconstructs the 3D shape by disentangling the embedding. The key points extracted from the incomplete point clouds typically establish relationships only with their immediate local neighbourhoods. In contrast, our embedding strategy establishes connections with all input points, facilitating a comprehensive understanding of the input partial objects.
Qualitative comparison. Fig. 4 presents the qualitative comparison results. Our method stands out by predicting shapes with greater accuracy and finer details. For example, as shown in the second and third rows of Fig. 4, our approach more effectively restores the complex structures on the wings of the airplane and the lampshade of the lamp, while the reconstructions from other methods appear significantly noisier. This highlights our network's ability in refining the shape with localized details.

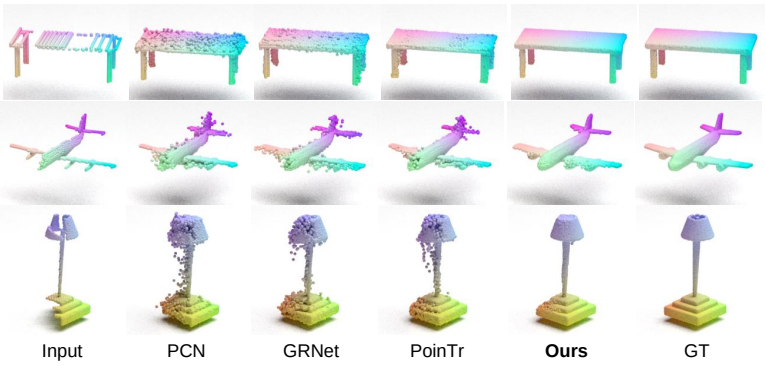


Fig. 4: Visual comparisons on PCN dataset. Note that the partial point clouds (2048 points) are sparse and self-occluded, as opposed to the reconstructed and ground truth point clouds (16,384 points) which are dense and complete.

5 Ablation Study

This section analyzes the performance of the feature extraction, coarse completion, fine completion and multi view-aware disentanglement modules within our model, specifically targeting the MVP dataset and working with point clouds consisting of 2048 points. An additional area of evaluation is the efficacy of the altered slot attention mechanism in enhancing point cloud completion tasks. To quantify the performance and improvements brought by these modifications, we employ CD and F-Score @1% threshold as our evaluation metrics.

Feature Extraction (FE): In the encoding phase, our AED module was used to down-sample the point cloud and extract features. To validate the efficiency of our feature extraction technique, we compared it with the feature extractor from SnowFlakeNet [41], hereafter referred to as $Model1$. According to the results presented in Table 5, AEDNet surpasses $Model1$ in both CD and F-Score @1%, confirming the superiority of our feature extraction process for point cloud completion tasks. This is primarily because our model is capable of automatically identifying and selecting the keypoints that are most crucial for the completion task, unlike the points selected through FPS, as illustrated in Fig. 5.

Table 5: Ablation studies for out AEDNet, examining the contributions of the Feature Extraction (FE) ($Model1$), Coarse Completion (CC) module ($Model2$), Fine Completion (FC) module ($Model3$), and multiview-aware Disentanglement operation ($Model4$) to the overall performance. 'SPD' means snowflake point deconvolution in [41], 'Emb' indicates embedding operator, 'Deconv' represents deconvolution, 'CD' denotes $\mathcal{L}_2$ Chamfer Distance (multiplied by 10^4) and 'F1' represents F Score @1%.

Model	Description	CD ↓	F1 ↑
1	Backbone [41] +CC+ FC	5.38	0.510
2	FE + CC [41] + FC	5.43	0.506
3	FE + CC + SPD [41]	5.54	0.501
4	Emb + Deconv [52]	5.44	0.504
5	**Our AEDNet**	**5.12**	**0.522**

Coarse Completion (CC): Previous methods [41,47,48] typically reconstructed a 3D shape either from a global feature alone or by directly using anchor points for shape query. Our approach diverges by initially reconstructing a coarse 3D object through the combination of the global feature and the learned embedding. Subsequently, we employ a differentiable Top-K technique to automatically choose informative points for the up-sampling phase. When comparing our coarse completion with the one from SnowFlakeNet, denoted as $Model2$, it becomes clear that our method significantly boosts performance.

Fine Completion (FC): Traditional point cloud up-sampling methods [10, 24, 41, 45, 47, 52] concentrate on generating points using a variety of techniques. To evaluate the impact of our GAED module, we substituted it with the snowflake point deconvolution (SPD) [41], referred to as $Model3$. The comparative results in Table 5 show that AEDNet outperforms $Model3$, evidencing the efficiency of our GAED module in enhancing point cloud generation.

Multiview-aware Disentanglement: We test the effectiveness of our mutliview aware disentanglement operator in AED module by replacing it with the

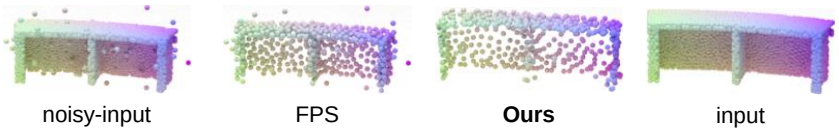


Fig. 5: Visualization of Point Cloud Down-sampling Techniques. The figures are ordered from left to right: noisy input, FPS down-sampled results, our method (labeled "Ours"), and the clean input for comparison.

deconvolution [52], named as embedding and deconvolution module (denoted as $Model4$). The comparative results in Table 5 show that multiview-aware disentanglement operator based network (**ours**) outperforms deconvolution based network ($Model4$), evidencing the efficiency of our multiview-aware disentanglement operator in enhancing detailed point cloud upsampling.

Slot Attention Modifications: We examine the effects of altering slot initialization and fostering slot interactions. To this end, we transitioned from Gaussian Noise initialization to dictionary mapping initialization. Moreover, we incorporated self-attention mechanisms within each iteration to facilitate slot interactions. According to the results presented in Table 6, these adjustments significantly improve model performance. This improvement suggests that the modifications to slot attention are effective in better capturing the intricate structures of 3D objects, highlighting the value of our proposed changes in enhancing the model's ability to understand complex spatial relationships.

6 Conclusion

This paper introduced an Adaptive Embedding and Disentanglement Network (AEDNet) for point cloud completion. The central innovation is the emphasis on global perception and local attention as essential components for successful point cloud completion. We refined the original slot attention mechanism to better achieve adaptive embedding of complex 3D structures, focusing on improving how slots are initialized and interact. We introduced multi-view information to disentanglement to observe objects from the outside rather than from within, resulting in a more comprehensive geometric understanding. The quantity of 3D points sampled can be changed as needed to allow for flexible sampling. Experiments on both the MVP and PCN datasets validate the performance of the proposed AEDNet.

Table 6: Ablation studies examining the effect of slot initialization and slot interaction. 'DMI' represents dictionary mapping initialization, 'CD' denotes $\mathcal{L}_2$ Chamfer Distance (multiplied by 10^4), and 'F1' represents F Score @1%.

Model	Modifications	CD ↓	F1 ↑
i	original slot attention	5.45	0.488
ii	only DMI	5.34	0.502
iii	only slot interaction	5.38	0.510
iv	**Ours**	**5.12**	**0.522**

Acknowledgements

This research was financially supported by the Australian Research Council (ARC DP210101682, DP210102674, DP220102197), UWA Research Collaboration Award (2023/GR001286) and received additional partial funding from the National Natural Science Foundation of China (No. U20A20185, 62372491), the Guangdong Basic and Applied Basic Research Foundation (2022B1515020103, 2023B1515120087), the Shenzhen Science and Technology Program (No. RCYX20200714114641140).

References

1. Chen, Z., Long, F., Qiu, Z., Yao, T., Zhou, W., Luo, J., Mei, T.: Anchorformer: Point cloud completion from discriminative nodes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13581–13590 (2023)
2. Chen, Z., Long, F., Qiu, Z., Yao, T., Zhou, W., Luo, J., Mei, T.: Learning 3d shape latent for point cloud completion. IEEE Transactions on Multimedia (2024)
3. Choy, C.B., Xu, D., Gwak, J., Chen, K., Savarese, S.: 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction. In: European Conference on Computer Vision. pp. 628–644. Springer (2016)
4. Dai, A., Ruizhongtai Qi, C., Nießner, M.: Shape completion using 3d-encoder-predictor cnns and shape synthesis. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 5868–5877 (2017)
5. Eslami, S., Heess, N., Weber, T., Tassa, Y., Szepesvari, D., Hinton, G.E., et al.: Attend, infer, repeat: Fast scene understanding with generative models. Advances in Neural Information Processing Systems **29** (2016)
6. Fu, Z., Wang, L., Xu, L., Wang, Z., Laga, H., Guo, Y., Boussaid, F., Bennamoun, M.: Vapcnet: Viewpoint-aware 3d point cloud completion. In: IEEE International Conference on Computer Vision. pp. 12108–12118 (2023)
7. Girdhar, R., Fouhey, D.F., Rodriguez, M., Gupta, A.: Learning a predictable and generative vector representation for objects. In: European Conference on Computer Vision. pp. 484–499. Springer (2016)
8. Greff, K., Kaufman, R.L., Kabra, R., Watters, N., Burgess, C., Zoran, D., Matthey, L., Botvinick, M., Lerchner, A.: Multi-object representation learning with iterative variational inference. In: International Conference on Machine Learning. pp. 2424–2433. PMLR (2019)
9. Greff, K., Van Steenkiste, S., Schmidhuber, J.: Neural expectation maximization. Advances in Neural Information Processing Systems **30** (2017)
10. Groueix, T., Fisher, M., Kim, V.G., Russell, B.C., Aubry, M.: A papier-mâché approach to learning 3d surface generation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 216–224 (2018)
11. Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M.: Deep learning for 3d point clouds: A survey. IEEE transactions on pattern analysis and machine intelligence **43**(12), 4338–4364 (2020)
12. Han, X., Li, Z., Huang, H., Kalogerakis, E., Yu, Y.: High-resolution shape completion using deep neural networks for global structure and local geometry inference. In: IEEE International Conference on Computer Vision. pp. 85–93 (2017)
13. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020)
14. Huang, Z., Yu, Y., Xu, J., Ni, F., Le, X.: Pf-net: Point fractal network for 3d point cloud completion. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7662–7670 (2020)
15. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
16. Kosiorek, A., Kim, H., Teh, Y.W., Posner, I.: Sequential attend, infer, repeat: Generative modelling of moving objects. Advances in Neural Information Processing Systems **31** (2018)
17. Li, J., Guo, S., Wang, L., Han, S.: Completedt: Point cloud completion with information-perception transformers. Neurocomputing **592**, 127790 (2024)

18. Liu, M., Sheng, L., Yang, S., Shao, J., Hu, S.M.: Morphing and sampling network for dense point cloud completion. In: AAAI conference on Artificial Intelligence. vol. 34, pp. 11596–11603 (2020)
19. Locatello, F., Weissenborn, D., Unterthiner, T., Mahendran, A., Heigold, G., Uszkoreit, J., Dosovitskiy, A., Kipf, T.: Object-centric learning with slot attention. Advances in Neural Information Processing Systems **33**, 11525–11538 (2020)
20. Luo, S., Hu, W.: Diffusion probabilistic models for 3d point cloud generation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2837–2845 (2021)
21. Lyu, Z., Kong, Z., Xu, X., Pan, L., Lin, D.: A conditional point diffusion-refinement paradigm for 3d point cloud completion. arXiv preprint arXiv:2112.03530 (2021)
22. Nie, Y., Lin, Y., Han, X., Guo, S., Chang, J., Cui, S., Zhang, J., et al.: Skeleton-bridged point completion: From global inference to local adjustment. Advances in Neural Information Processing Systems **33**, 16119–16130 (2020)
23. Pan, L.: Ecg: Edge-aware point cloud completion with graph convolution. IEEE Robotics and Automation Letters **5**(3), 4392–4398 (2020)
24. Pan, L., Chen, X., Cai, Z., Zhang, J., Zhao, H., Yi, S., Liu, Z.: Variational relational point completion network. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 8524–8533 (2021)
25. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 652–660 (2017)
26. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in Neural Information Processing Systems **30** (2017)
27. Shao, J.: Mathematical statistics. Springer Science & Business Media (2003)
28. Su, H., Jampani, V., Sun, D., Maji, S., Kalogerakis, E., Yang, M.H., Kautz, J.: Splatnet: Sparse lattice networks for point cloud processing. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2530–2539 (2018)
29. Tchapmi, L.P., Kosaraju, V., Rezatofighi, H., Reid, I., Savarese, S.: Topnet: Structural point cloud decoder. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 383–392 (2019)
30. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in Neural Information Processing Systems **30** (2017)
31. Viswanath, D.: Random fibonacci sequences and the number 1.13198824.... Mathematics of Computation **69**(231), 1131–1155 (2000)
32. Wang, J., Cui, Y., Guo, D., Li, J., Liu, Q., Shen, C.: Pointattn: You only need attention for point cloud completion. arXiv preprint arXiv:2203.08485 (2022)
33. Wang, P.S., Liu, Y., Guo, Y.X., Sun, C.Y., Tong, X.: O-cnn: Octree-based convolutional neural networks for 3d shape analysis. ACM Transactions On Graphics (TOG) **36**(4), 1–11 (2017)
34. Wang, X., Ang Jr, M.H., Lee, G.H.: Cascaded refinement network for point cloud completion. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 790–799 (2020)
35. Wang, Y., Tan, D.J., Navab, N., Tombari, F.: Learning local displacements for point cloud completion. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1568–1577 (2022)
36. Wang, Y., Wang, L., Hu, Q., Liu, Y., Zhang, Y., Guo, Y.: Panoptic segmentation of 3d point clouds with gaussian mixture model in outdoor scenes. Visual Intelligence **2**(1), 10 (2024)

37. Wen, X., Han, Z., Cao, Y.P., Wan, P., Zheng, W., Liu, Y.S.: Cycle4completion: Unpaired point cloud completion using cycle transformation with missing region coding. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 13080–13089 (2021)
38. Wen, X., Xiang, P., Han, Z., Cao, Y.P., Wan, P., Zheng, W., Liu, Y.S.: Pmp-net: Point cloud completion by learning multi-step point moving paths. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 7443–7452 (2021)
39. Wen, X., Xiang, P., Han, Z., Cao, Y.P., Wan, P., Zheng, W., Liu, Y.S.: Pmp-net++: Point cloud completion by transformer-enhanced multi-step point moving paths. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(1), 852–867 (2022)
40. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3d shapenets: A deep representation for volumetric shapes. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1912–1920 (2015)
41. Xiang, P., Wen, X., Liu, Y.S., Cao, Y.P., Wan, P., Zheng, W., Han, Z.: Snowflakenet: Point cloud completion by snowflake point deconvolution with skip-transformer. In: IEEE International Conference on Computer Vision. pp. 5499–5509 (2021)
42. Xie, C., Wang, C., Zhang, B., Yang, H., Chen, D., Wen, F.: Style-based point generator with adversarial rendering for point cloud completion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4619–4628 (2021)
43. Xie, H., Yao, H., Zhou, S., Mao, J., Zhang, S., Sun, W.: Grnet: Gridding residual network for dense point cloud completion. In: European Conference on Computer Vision. pp. 365–381. Springer (2020)
44. Yan, X., Zheng, C., Li, Z., Wang, S., Cui, S.: Pointasnl: Robust point clouds processing using nonlocal neural networks with adaptive sampling. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 5589–5598 (2020)
45. Yang, Y., Feng, C., Shen, Y., Tian, D.: Foldingnet: Point cloud auto-encoder via deep grid deformation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 206–215 (2018)
46. Yu, L., Li, X., Fu, C.W., Cohen-Or, D., Heng, P.A.: Pu-net: Point cloud upsampling network. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2790–2799 (2018)
47. Yu, X., Rao, Y., Wang, Z., Liu, Z., Lu, J., Zhou, J.: Pointr: Diverse point cloud completion with geometry-aware transformers. In: IEEE International Conference on Computer Vision. pp. 12498–12507 (2021)
48. Yuan, W., Khot, T., Held, D., Mertz, C., Hebert, M.: Pcn: Point completion network. In: International Conference on 3D Vision (3DV). pp. 728–737. IEEE (2018)
49. Zhang, W., Yan, Q., Xiao, C.: Detail preserved point cloud completion via separated feature aggregation. In: European Conference on Computer Vision. pp. 512–528. Springer (2020)
50. Zhang, Y., Wang, L., Li, K., Fu, Z., Guo, Y.: Slfnet: A stereo and lidar fusion network for depth completion. IEEE Robotics and Automation Letters **7**(4), 10605–10612 (2022)
51. Zhao, H., Jiang, L., Jia, J., Torr, P.H., Koltun, V.: Point transformer. In: IEEE International Conference on Computer Vision. pp. 16259–16268 (2021)
52. Zhou, H., Cao, Y., Chu, W., Zhu, J., Lu, T., Tai, Y., Wang, C.: Seedformer: Patch seeds based point cloud completion with upsample transformer. In: European Conference on Computer Vision. pp. 416–432. Springer (2022)

53. Zhou, L., Du, Y., Wu, J.: 3d shape generation and completion through point-voxel diffusion. In: IEEE International Conference on Computer Vision. pp. 5826–5835 (2021)
54. Zhu, Z., Chen, H., He, X., Wang, W., Qin, J., Wei, M.: Svdformer: Complementing point cloud via self-view augmentation and self-structure dual-generator. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14508–14518 (2023)
55. Zong, D., Sun, S., Zhao, J.: Ashf-net: Adaptive sampling and hierarchical folding network for robust point cloud completion. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 3625–3632 (2021)