Manifold-Aware Point Cloud Completion via Geodesic-Attentive Hierarchical Feature Learning

Jianan Sun[1] , **Dongzhihan Wang**[2] , **Zhangqi Huang**[1] and **Mingyu Fan**[1*]
[1]Donghua University
[2]Shanghai University
2232213@mail.dhu.edu.cn, wdzhihan@shu.edu.cn, ZhangqiHuang1996@outlook.com, fanmingyu@dhu.edu.cn

Abstract

Point cloud completion seeks to recover geometrically consistent shapes from partial or sparse 3D observations. Although recent methods have achieved reasonable global shape reconstruction, they often rely on Euclidean proximity and overlook the intrinsic nonlinear geometric structure of point clouds, resulting in suboptimal geometric consistency and semantic ambiguity. In this paper, we present a manifold-aware point cloud completion framework that explicitly incorporates nonlinear geometry information throughout the feature learning pipeline. Our approach introduces two key modules: a Geodesic Distance Approximator (GDA), which estimates geodesic distances between points to capture the latent manifold topology, and a Manifold-Aware Feature Extractor (MAFE), which utilizes geodesic-based k-NN groupings and a geodesic-relational attention mechanism to guide the hierarchical feature extraction process. By integrating geodesic-aware relational attention, our method promotes semantic coherence and structural fidelity in the reconstructed point clouds. Extensive experiments on benchmark datasets demonstrate that our approach consistently outperforms state-of-the-art methods in reconstruction quality. The code has been available online https://anonymous.4open.science/r/Manifold-Aware-Point-Cloud-Completion--F380/README.md.

1 Introduction

In real-world scenarios, 3D scans are often sparse, incomplete, and noisy due to occlusions, sensor limitations, or environmental disturbances, which pose significant challenges to downstream processing and reconstruction [Chang *et al.*, 2015; Tchapmi *et al.*, 2019; Yu *et al.*, 2021; Xie *et al.*, 2020]. Therefore, point cloud completion has become a fundamental task in 3D computer vision, supporting high-level perception and scene understanding in applications such as autonomous driving, robotics, augmented reality, and cultural heritage preservation [Yuan *et al.*, 2018; Geiger *et al.*, 2012; Newcombe *et al.*, 2011].

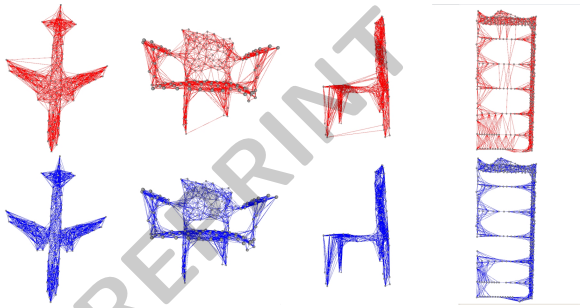

Figure 1: Visualization of point connectivity using Euclidean versus manifold-based neighbors. Graphs in the top row are constructed using Euclidean neighbors, while the graphs in the bottom row are built by manifold-aware neighbors. The manifold-based graphs better preserve the intrinsic topology and continuity of the underlying shape.

Recent learning-based approaches have made notable progress in tackling the shape completion problem [Chen *et al.*, 2023; Rong *et al.*, 2024; Cai *et al.*, 2024; Zhou *et al.*, 2025]. Despite their success, most of these methods treat point clouds as unordered sets in Euclidean space, overlooking the fact that real-world surfaces often lie on low-dimensional manifolds embedded in high-dimensional space [Yu *et al.*, 2023; Zhong *et al.*, 2025]. This oversimplification hinders accurate modeling of intrinsic geometric structures, limiting the capacity to reason about shape continuity beyond local neighborhoods.

Some recent methods incorporate geometry-aware modules for 3D shape modeling. For instance, following the study of DGCNN [Wang *et al.*, 2019], AdaPoinTr [Yu *et al.*, 2023] introduces a geometric block that aggregates features from k-nearest neighbors (k-NN). PointCFormer [Zhong *et al.*, 2025] extends local affinity estimation to high-dimensional feature space by introducing residual vectors. However, both approaches still rely on fixed k-NN groupings in Euclidean space [Qi *et al.*, 2017], making them vulnerable to point clouds with nonlinear geometry. In contrast, PointAttN [Wang *et al.*, 2024] replaces explicit k-NN grouping with global attention weights across all points.

Euclidean k-NN graphs usually contain shortcut con-

*Corresponding author.

nections, linking spatially close but geodesically distant points—leading to features from semantically unrelated regions, especially under sparse sampling (as shown in the top row of Figure 1). To overcome this challenge, we propose a manifold-aware feature extraction framework. The core of the framework is an anchor-based geodesic distance metric, which efficiently approximates true manifold connectivity from partial observations and robustly redefines local neighborhoods (as illustrated in the bottom row of Figure 1). The geodesic k-NN, by comparison, accurately identifies true adjacency points on surfaces, enabling semantic consistency and structural integrity. This geodesic metric serves as the foundation of our framework: it supports geodesic-aware neighborhood grouping, informs attention-based feature aggregation, and facilitates manifold positional embedding.

In summary, our contributions are threefold:

- We propose an efficient anchor-based geodesic distance approximation method that captures intrinsic surface topology and enables robust neighborhood construction in sparse and geometrically complex point clouds.
- We develop a unified manifold-aware feature extraction framework that integrates geodesic guidance into neighborhood grouping, attention-based aggregation, and manifold positional embedding, leading to improved geometry encoding and reasoning.
- We conduct experiments on standard and challenging benchmarks, demonstrating the effectiveness and generalizability of our method, which consistently outperforms state-of-the-art approaches.

2 Related Work

2.1 3D Shape Completion

Early works such as 3D-EPN [Dai *et al.*, 2017] adopted volumetric representations but were limited by resolution and scalability. Recently, point-based methods have become the mainstream due to their flexibility and efficiency, including PCN [Yuan *et al.*, 2018], TopNet [Tchapmi *et al.*, 2019], and AtlasNet [Groueix *et al.*, 2018]. Later works such as FoldingNet [Yang *et al.*, 2019] and GRNet [Xie *et al.*, 2020] further improved reconstruction fidelity. Despite these advances, most methods still assume linear locality and overlook nonlinear topology, limiting performance on complex or incomplete shapes.

2.2 Transformer-based Completion Methods

Transformer-based methods have gained attention for global context modeling. PoinTr [Yu *et al.*, 2021] formulated completion as set-to-set translation with geometry-aware modules, and AdaPoinTr [Yu *et al.*, 2023] improved robustness via adaptive queries and denoising. SnowflakeNet [Xiang *et al.*, 2021] and SeedFormer [Zhou *et al.*, 2022] adopt coarse-to-fine refinement with transformer-based upsampling, while PointCFormer [Zhong *et al.*, 2025] uses relation-weighted progressive extraction. PointAttN [Wang *et al.*, 2024] replaces explicit k-NN with global attention to reduce shortcut edges. Although geodesic relations have been studied for point cloud analysis (e.g., GeoNet [Qi *et al.*, 2018]), these completion Transformers still typically rely on Euclidean neighborhood aggregation and do not explicitly leverage intrinsic topology under sparsity.

2.3 Attention Mechanisms

Some recent approaches have explored geometry-aware attention mechanisms to improve local feature extraction. AdaPoinTr [Yu *et al.*, 2023] incorporates an adaptive geometric block to encode local spatial relationships, while PointCFormer [Zhong *et al.*, 2025] integrates local affinity estimation through residual vector attention. PointAttN [Wang *et al.*, 2024] eliminates fixed k-NN by applying global attention across all points, thereby alleviating the effect of neighborhood errors. Nonetheless, these attention modules are still rooted in Euclidean or feature-space proximity, which can lead to semantic ambiguity and poor locality preservation—particularly on sparsely sampled surfaces. In contrast, we address completion with partial inputs and use an anchor-based geodesic distance as a local reranking and attention cue, rather than estimating exact surface geodesics.

3 Method

3.1 Overview

Point cloud completion aims to recover a complete 3D shape from an incomplete observation. A common limitation of prior methods is that local neighborhoods and relational modeling are largely based on Euclidean proximity, which may introduce shortcut edges on nonlinearly embedded surfaces and under sparsity. To address this, we inject manifold-aware cues into feature learning through two key modules: (1) **Geodesic Distance Approximator (GDA)**, which precomputes an anchor-to-anchor shortest-path matrix and supports efficient geodesic distance queries; (2) **Manifold-Aware Feature Extractor (MAFE)**, which integrates geodesic cues into hierarchical grouping (GNG), geodesic-relational attention (GRA-T), and manifold positional embedding (MPE). The subsequent coarse completion and upsampling follow a standard coarse-to-fine decoder design and are not the focus of our novelty. By combining local geodesic neighborhood modeling and global manifold position encoding, our method better preserves intrinsic geometry and improves completion quality.

3.2 Geodesic Distance Approximator (GDA)

Real-world point clouds often lie on low-dimensional manifolds, where Euclidean distance may poorly reflect intrinsic geometry. As shown in Fig. 3, Euclidean neighborhoods can introduce shortcut edges on nonlinearly embedded surfaces, motivating our geodesic-aware neighborhood construction.

The ideal manifold geodesic distance between two points x_i and x_j on a manifold $\mathcal{M}$ is:

$$d_{\mathcal{M}}(x_i, x_j) = \min_{\gamma \in \Gamma_{ij}} \int_0^1 \|\gamma'(t)\|\, dt, \tag{1}$$

where Γ_{ij} denotes piecewise-smooth paths connecting x_i and x_j.

Direct geodesic computation is intractable on discrete point sets. Instead of computing an all-pairs distance matrix on the

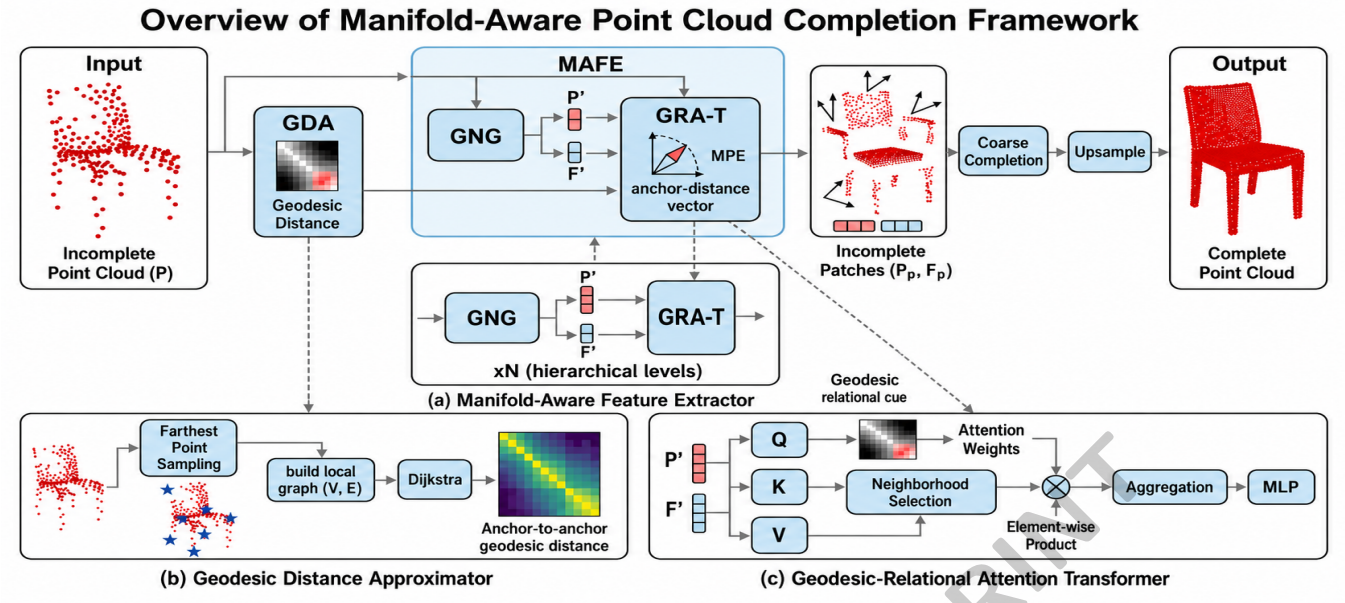

Figure 2: Overview of our framework. Given an incomplete point cloud $\mathcal{P}$, the Geodesic Distance Approximator (GDA) samples anchors and constructs an *anchor* proximity graph, where Dijkstra's algorithm is used to *precompute* the anchor-to-anchor shortest-path matrix $\mathbf{D}_{\mathcal{A}}$ (Fig. 2(b)). Based on $\mathbf{D}_{\mathcal{A}}$, we query approximate point-to-point geodesic cues on demand and inject them into the Manifold-Aware Feature Extractor (MAFE) via geodesic-based grouping (GNG) and the Geodesic-Relational Attention Transformer (GRA-T), enabling manifold-aware feature aggregation (Fig. 2(a,c)). Finally, we employ a *standard* coarse-to-fine completion decoder (coarse completion + upsampling) to generate a dense and complete point cloud.

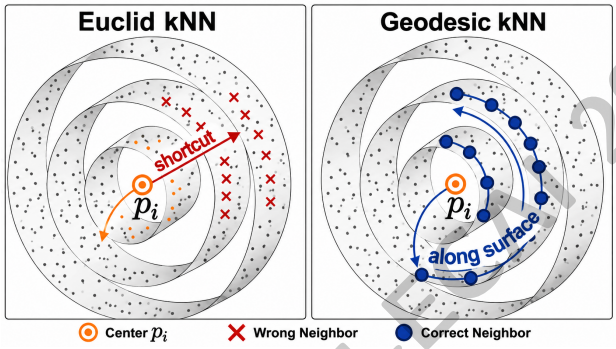

Figure 3: Euclidean vs. geodesic neighborhoods on a rolled manifold. Euclidean kNN may connect across nearby but topologically distant sheets (left, shortcut), whereas geodesic kNN follows the surface and selects consistent neighbors (right).

full point cloud, we adopt an efficient *anchor-based* approximation (Fig. 2(b)).

Anchor graph and precomputed anchor geodesics

We first sample an anchor set $\mathcal{A} = \{a_m\}_{m=1}^{M}$ from $\mathcal{P}$ using farthest point sampling (FPS). We then build a sparse *anchor* k-NN graph $\mathcal{G}_{\mathcal{A}} = (\mathcal{V}_{\mathcal{A}}, \mathcal{E}_{\mathcal{A}})$ with $\mathcal{V}_{\mathcal{A}} = \mathcal{A}$ and edges connecting each anchor to its Euclidean $k_{\mathcal{A}}$ nearest anchors, weighted by Euclidean distances. Running Dijkstra's algorithm [Dijkstra, 1959] from each anchor on $\mathcal{G}_{\mathcal{A}}$ yields an anchor-to-anchor shortest-path matrix:

$$\mathbf{D}_{\mathcal{A}} \in \mathbb{R}^{M\times M}, \qquad \mathbf{D}_{\mathcal{A}}(u,v) = \mathrm{dist}_{\mathcal{G}_{\mathcal{A}}}(a_u, a_v). \tag{2}$$

This $\mathbf{D}_{\mathcal{A}}$ is the *output* of GDA and is reused by all subsequent modules.

Nearest-anchor geodesic approximation

For each point x_i, we assign its nearest anchor:

$$a(i) = \arg\min_{a\in\mathcal{A}} \|x_i - a\|, \qquad \delta_i = \|x_i - a(i)\|. \tag{3}$$

We approximate the geodesic distance between two points x_i and x_j by the nearest-anchor mediated form:

$$\hat{d}_g(x_i, x_j) = \delta_i + \mathbf{D}_{\mathcal{A}}(a(i), a(j)) + \delta_j. \tag{4}$$

In practice, $\mathbf{D}_{\mathcal{A}}$ is precomputed once per input for efficiency, while $\hat{d}_g(\cdot,\cdot)$ is evaluated on demand when forming neighborhoods and computing relational cues (Fig. 2(c)). **Remark.** Although $\hat{d}_g$ in Eq. (4) uses Euclidean offsets δ_i and Euclidean edge weights to build the sparse anchor graph, the *topology-aware* term $D_{\mathcal{A}}(a(i), a(j))$ (anchor shortest paths) is what suppresses cross-surface shortcuts and drives our neighborhood ranking and relational modeling.

3.3 Manifold-Aware Feature Extractor (MAFE)

MAFE injects manifold-aware cues into hierarchical feature learning (Fig. 2(a)) via three components: (1) Geodesic Neighborhood Grouper (GNG) for geodesic neighborhood construction, (2) Geodesic-Relational Attention Transformer (GRA-T) for geodesic-aware relational aggregation, and (3) Manifold Positional Embedding (MPE) for global manifold-aware position encoding.

Geodesic Neighborhood Grouper (GNG)
GNG hierarchically downsamples the input point cloud using FPS to obtain multi-level subsets. For each center point, instead of using Euclidean k-NN, we form its neighborhood using the anchor-based geodesic cue $\hat{d}_g$ in Eq. (4). To reduce cost and improve stability under partial inputs, we first preselect a local candidate set $\mathcal{C}(i)$ by Euclidean L-NN ($L = \rho k, \rho > 1$), and then rerank candidates using $\hat{d}_g$ to obtain geodesic neighbors $\mathcal{N}_g(i)$.

$$\mathcal{N}_g(i) = \mathrm{TopK}_{j\in\mathcal{C}(i)}\big(-\hat{d}_g(p_i, p_j)\big). \tag{5}$$

The features of neighbors in $\mathcal{N}_g(i)$ are aggregated to form local descriptors at each level, yielding geodesic-consistent features for subsequent attention blocks.

Geodesic-Relational Attention Transformer (GRA-T)
We propose GRA-T (Fig. 2(c)) to integrate geodesic cues into local attention. Unlike conventional attention that relies on Euclidean neighborhoods, GRA-T uses the geodesic neighborhood $\mathcal{N}_g(i)$ and an explicit geodesic-relational embedding to enhance semantic coherence and structural fidelity.

Given features $F \in \mathbb{R}^{C\times n}$ and coordinates $P \in \mathbb{R}^{3\times n}$ at a pyramid level, we compute linear projections $Q = W_Q F$, $K = W_K F$, $V = W_V F$. For each center point p_i and neighbor $p_j \in \mathcal{N}_g(i)$, we build a geodesic-relational embedding:

$$r_{ij} = \phi\Big(\big[q_i - k_j,\ \psi(p_i - p_j),\ \eta(\hat{d}_g(p_i, p_j))\big]\Big) \in \mathbb{R}^C, \tag{6}$$

where ϕ is an MLP, and $\psi(\cdot)$ and $\eta(\cdot)$ embed relative coordinates and geodesic cues, respectively. We then apply *channel-wise* softmax along the neighbor dimension:

$$\alpha_{ij}^{(c)} = \frac{\exp\left(r_{ij}^{(c)}\right)}{\sum_{p\in\mathcal{N}_g(i)} \exp\left(r_{ip}^{(c)}\right)}, \quad c = 1, \ldots, C. \tag{7}$$

Finally, we aggregate neighbor values with element-wise (channel-wise) weighting:

$$f_i' = \sum_{p_j\in\mathcal{N}_g(i)} \alpha_{ij} \odot (v_j + \psi(p_i - p_j)). \tag{8}$$

Here $\alpha_{ij} \in \mathbb{R}^C$ and $\odot$ denotes channel-wise multiplication, matching the "Element-wise Product" block in Fig. 2(c).

Manifold Positional Embedding (MPE)
To strengthen global manifold awareness, we introduce MPE that encodes global geometric context into each point's representation. For a point p_i, we first identify its closest anchor $a(i) \in \mathcal{A}$. We then construct an anchor-distance vector by retrieving the geodesic distances from p_i to all anchors, and concatenate it with the point feature:

$$f_i^{\mathrm{aug}} = [f_i,\ \hat{d}_g(p_i, \mathcal{A})], \tag{9}$$

where $\hat{d}_g(p_i, \mathcal{A}) \in \mathbb{R}^{|\mathcal{A}|}$ denotes the geodesic distance vector from p_i to all anchors in $\mathcal{A}$. This embedding provides a compact encoding of the point's global location on the manifold and facilitates better semantic discrimination and geometric consistency.

In summary, MAFE injects intrinsic geometric priors into the feature extraction pipeline by explicitly modeling both local (geodesic grouping + geodesic-relational attention) and global (anchor-based manifold position) structures.

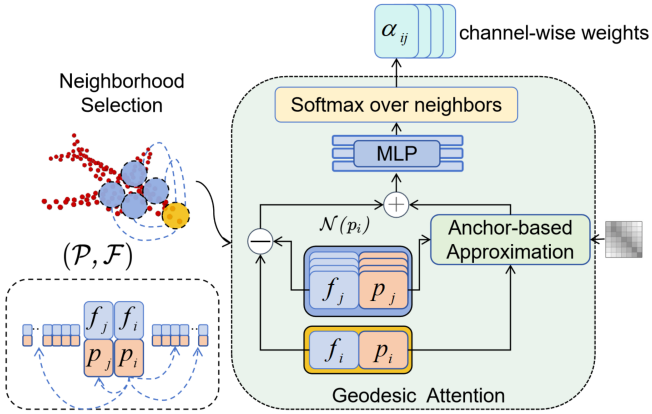

Figure 4: Illustration of the geodesic-relational attention in GRA-T. For each center point, neighborhoods are obtained by Euclidean candidate preselection followed by reranking with the anchor-based geodesic cue $\hat{d}_g$. The geodesic cue is embedded and fused into attention to produce channel-wise weights for feature aggregation.

3.4 Coarse Completion and Upsample Model

Following manifold-aware feature extraction, we adopt a *standard* coarse-to-fine completion decoder. We first generate a coarse completed point cloud and then progressively refine and upsample it with a transformer-based upsampling module. This decoder follows common practice in prior completion architectures (e.g., coarse-to-fine refinement[Zhou *et al.*, 2022]) and serves to convert the enhanced manifold-aware features into dense point sets.

3.5 Loss Function

We adopt the Chamfer Distance (CD) [Fan *et al.*, 2017] as the primary loss for supervising all stages:

$$\mathcal{L} = \mathcal{L}_{CD}(P_c, Q) + \sum_{i=1}^{n} \mathcal{L}_{CD}(P_i, Q), \tag{10}$$

where P_c denotes the initial coarse output, P_i are intermediate/refined outputs, Q is the ground-truth point cloud, and n is the number of refinement stages. This multi-stage supervision encourages progressive alignment and stable training.

4 Experiments

In this section, we conduct a comprehensive evaluation of the proposed method on several widely-used benchmark datasets. We begin by introducing the datasets and evaluation protocols, followed by quantitative comparisons with state-of-the-art methods. We also provide qualitative visualizations to further demonstrate the effectiveness and generalization capability of our approach.

4.1 Datasets and Evaluation Metrics

We evaluate our method on a range of synthetic and real-world datasets, including PCN [Yuan *et al.*, 2018], ShapeNet-55 [Yu *et al.*, 2021], ShapeNet-34, ShapeNet-Unseen21, MVP [Pan *et al.*, 2021], and KITTI [Geiger *et al.*, 2012].

Methods	Average	Plane	Cabinet	Car	Chair	Lamp	Couch	Table	Watercraft
FoldingNet [Yang *et al.*, 2019]	14.31	9.49	15.80	12.61	15.55	16.41	15.97	13.65	14.99
TopNet [Tchapmi *et al.*, 2019]	12.15	7.61	13.31	10.90	13.82	14.44	14.78	11.22	11.12
AtlasNet [Groueix *et al.*, 2018]	10.85	6.37	11.94	10.10	12.06	12.37	12.99	10.33	10.61
PCN [Yuan *et al.*, 2018]	9.64	5.50	22.70	10.63	8.70	11.00	11.34	11.68	8.59
GR-Net [Xie *et al.*, 2020]	8.83	6.45	10.37	9.45	9.41	7.96	10.51	8.44	8.04
SnowFlake [Xiang *et al.*, 2021]	7.19	4.24	9.27	8.20	7.75	5.96	9.25	6.45	6.37
AdaPoinTr [Yu *et al.*, 2023]	6.53	3.68	8.82	7.47	6.85	5.47	8.35	5.80	5.76
SeedFormer [Zhou *et al.*, 2022]	6.74	3.85	9.05	8.06	7.06	5.21	8.85	6.05	5.85
PointSea [Zhu *et al.*, 2025]	6.35	3.61	**8.54**	7.33	6.58	5.21	**8.24**	5.75	5.62
CRA-PCN [Rong *et al.*, 2024]	6.39	3.59	8.70	7.50	6.70	5.06	**8.24**	**5.72**	5.64
ODGNet [Cai *et al.*, 2024]	6.50	3.77	8.77	7.56	6.84	5.09	8.47	5.84	5.66
PointAttN [Wang *et al.*, 2024]	6.86	3.87	9.00	7.63	7.43	5.90	8.68	6.32	6.09
PointCFormer [Zhong *et al.*, 2025]	6.41	3.53	8.73	**7.32**	6.68	5.12	8.34	5.86	5.74
Ours	**6.32**	3.58	8.59	7.45	**6.56**	**4.92**	8.27	5.76	**5.43**

Table 1: Category-wise CD values ($\times 10^{-3}$) on the PCN dataset. Lower is better.

Evaluation Metrics. We employ the following standard metrics to comprehensively assess completion quality:

- **Chamfer Distance (CD):** Measures the average closest-point distance between the predicted and ground-truth point sets. We report both L_1 and L_2 variants as commonly adopted in the literature.
- **F-score:** The harmonic mean of precision and recall, computed with a predefined threshold. It reflects the fidelity and completeness of the reconstructed shape.

4.2 Comparison to the State-of-the-Art

Table 1 reports the category-wise CD values (lower is better) on the PCN dataset. Our method achieves the lowest average CD of 6.32, outperforming all prior state-of-the-art approaches. Notably, our approach obtains the best results in three out of eight categories, including 'Chair' (6.56), 'Lamp' (4.92), and 'Watercraft' (5.43), while remaining competitive on the other categories, indicating strong generalization to both simple and complex geometries. Qualitative results further highlight these advantages. For instance, in the first example of Figure 6, our method reconstructs a chair with multiple thin crossbars, restoring each slat with clear boundaries, whereas prior methods often produce blurred or missing connections and fail to preserve fine structural details. Our method achieves state-of-the-art results on ShapeNet-55, consistently outperforming all baseline methods across simple, medium, and hard settings (as shown in Table 2). Remarkably, despite adopting a more challenging whole-shape completion protocol—where the network must predict the entire object rather than only the missing regions—our model achieves the lowest CD value at every difficulty level [Zhu *et al.*, 2025]. For example, on the hardest split (CD-H), our method attains a CD of 1.09, a substantial improvement over the best competing methods. As can be seen, our approach exhibits excellent generalization ability, maintaining top performance on both seen and unseen categories, which demonstrates its robustness and applicability to new object classes. Moreover, our method demonstrates strong generalization: it achieves leading results not only on the 34 seen categories but also on the 21 unseen categories, with a CD-Avg of 0.64 and 1.05, respectively. These results confirm the effectiveness of our manifold-aware feature extraction in handling both familiar and novel object classes, and emphasize the broad applicability of our framework to diverse and complex shape completion scenarios. On the MVP dataset, where both the input and output point clouds consist of 2048 points, our method achieves state-of-the-art results, as summarized in Table 3. Specifically, our approach obtains the lowest CD (CD-ℓ_2) of 4.58×10^{-4} and the highest F-Score@1% of 0.573. These results represent a relative improvement of 2.8% in CD and 2.6% in F-score compared to the previous best-performing method, AdaPoinTr (CD-ℓ_2: 4.71, F-Score@1%: 0.545). Following the evaluation protocol of Zhou *et al.*, we further assess our model—trained solely on the PCN dataset—under realistic conditions characterized by significant noise and sparsity. As illustrated qualitatively in Figure 5 and quantitatively in Table 4, our method delivers competitive performance; note that the reported FD is the one-sided CD from the partial input to the prediction, thus methods that explicitly include the partial input in their outputs (e.g., PoinTr) can obtain FD$= 0$ by design. Overall, our method achieves state-of-the-art quantitative results and high-fidelity qualitative reconstructions across all four benchmarks, demonstrating both its effectiveness and versatility.

5 Ablation Study

5.1 Effect of Different Module Combinations

To assess the individual and joint contributions of the Geodesic-based Neighborhood Grouper (GNG), Geodesic-Relational Attention Transformer (GRA-T), and Manifold Positional Embedding (MPE), we perform comprehensive ablation experiments, as reported in Table 5. Introducing either GNG or GRA-T alone leads to a substantial reduction in CD, lowering it from 6.74 to 6.42 and 6.39, respectively, relative to the baseline. This indicates that both modules independently enhance local geometric feature extraction. However,

Method	ShapeNet-55					Seen ShapeNet-34					Unseen ShapeNet-21				
	CD-S	CD-M	CD-H	CD-Avg	F@1%	CD-S	CD-M	CD-H	CD-Avg	F@1%	CD-S	CD-M	CD-H	CD-Avg	F@1%
FoldingNet	2.67	2.66	4.05	3.12	0.082	1.86	1.81	3.38	2.35	0.139	2.76	2.74	5.36	3.62	0.095
PCN	1.94	1.96	4.08	2.66	0.133	1.87	1.81	2.97	2.22	0.154	3.17	3.08	5.29	3.85	0.101
TopNet	2.26	2.16	4.30	2.91	0.126	1.77	1.61	3.54	2.31	0.171	2.62	2.43	5.44	3.50	0.121
GRNet	1.35	1.71	2.85	1.97	0.238	1.26	1.39	2.57	1.74	0.251	1.85	2.25	4.87	2.99	0.216
SnowFlake	0.70	1.06	1.96	1.24	0.398	0.60	0.86	1.50	0.99	0.422	0.88	1.46	2.92	1.75	0.388
AdaPoinTr	0.49	0.69	1.24	0.81	0.503	0.48	0.63	1.07	0.73	0.469	0.61	0.96	2.11	1.23	0.416
SeedFormer	0.50	0.77	1.49	0.92	0.472	0.48	0.70	1.30	0.83	0.452	0.61	1.07	2.35	1.34	0.402
CRA-PCN	0.48	0.71	1.37	0.85	0.455	0.45	0.65	1.18	0.76	0.451	0.55	0.97	2.19	1.24	0.412
PointSea	0.43	0.64	1.19	0.75	0.485	0.40	0.57	1.00	0.66	0.492	0.50	0.88	1.92	1.10	**0.541**
AnchorFormer	0.41	0.61	1.26	0.76	**0.558**	0.41	0.57	1.12	0.70	**0.564**	0.52	0.90	2.16	1.19	0.535
PointCFormer	0.42	0.64	1.15	0.73	0.499	0.41	0.55	1.03	0.66	0.459	0.53	0.88	1.97	1.12	0.420
ODGNet	0.47	0.70	1.32	0.83	0.437	0.44	0.64	1.14	0.75	0.451	0.59	1.01	2.26	1.29	0.415
Ours	**0.41**	**0.57**	**1.09**	**0.69**	0.469	**0.40**	**0.53**	**0.99**	**0.64**	0.489	**0.45**	**0.80**	**1.91**	**1.05**	0.415

Table 2: Quantitative results on the ShapeNet55/34 [Chang *et al.*, 2015] dataset for three difficulty levels. CD-L2 $\times 10^{-3}$.

Model	#Points	CD-ℓ_2	F-Score@1%
PCN	2048	9.77	0.321
TopNet	2048	10.11	0.308
PoinTr	2048	6.15	0.456
AdaPoinTr	2048	4.71	0.545
CRA-PCN	2048	5.33	0.529
SymmCompletion	2048	4.89	0.534
Ours	2048	**4.58**	**0.573**

Table 3: Results on the MVP validation set. Both inputs and outputs contain 2048 points. We report CD-ℓ_2 (multiplied by 10,000) and F-Score@1% for each method (including SymmCompletion[Yan *et al.*, 2025])

Methods	PoinTr	PointAttN	SymmCompletion	Ours
FD (↓)	**0.0**	0.67	2.54	1.35
MMD (↓)	8.21	0.50	0.93	**0.31**

Table 4: Quantitative comparison on the KITTI dataset in terms of FD and MMD.

Figure 5: Completion results on the KITTI dataset. Our method recovers plausible shapes under real-world noise and sparsity.

GNG	GRA-T	MPE	CD↓
–	–	–	6.74
✓	–	–	6.42
–	✓	–	6.39
–	–	✓	6.64
✓	✓	–	6.41
✓	–	✓	6.38
–	✓	✓	6.35
✓	✓	✓	**6.32**

Table 5: Ablation study of different MAFE module combinations on PCN.

combining GNG and GRA-T without MPE results in only a marginal further improvement (CD: 6.41), suggesting partial redundancy in their modeling of local context. In contrast, adding the MPE module yields more pronounced gains when paired with either GNG (CD: 6.38) or GRA-T (CD: 6.35), highlighting the importance of global manifold-aware positional information in resolving local ambiguities. The full configuration, integrating GNG, GRA-T, and MPE, achieves the best performance with a CD of 6.32, demonstrating that these components are complementary and jointly critical for optimal feature representation and shape completion accuracy.

5.2 Replacing Different Baseline Backbones

To further demonstrate the generality and adaptability of the proposed MAFE module, we substitute it as the feature extraction component into several representative backbone frameworks, as summarized in Table 6.

For both SnowflakeNet and SeedFormer, replacing the original PointNet++ (PNet++)-based feature extractor with MAFE leads to consistent and substantial improvements in reconstruction accuracy, reducing the CD loss from 7.19 to 6.46 in SnowflakeNet and from 6.74 to 6.32 in SeedFormer. For SVDFormer, the improvement is more modest (CD decreases from 6.54 to 6.41). We attribute this to the unique design of SVDFormer, whose feature extractor leverages additional auxiliary information derived from the input point cloud, such as generated multi-view images, to enhance completion.

Overall, the results demonstrate that MAFE offers consis-

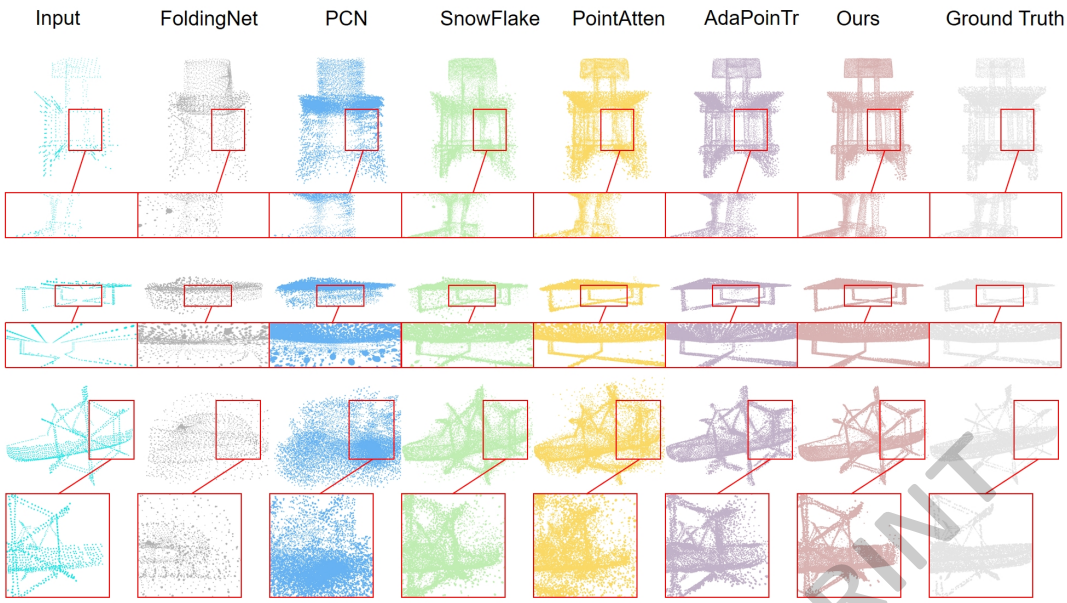

Figure 6: Qualitative results on PCN: our method better restores fine structures and complex topologies compared to prior approaches.

Backbone	Feature Extraction	CD↓
SnowflakeNet	PNet++	7.19
SnowflakeNet	MAFE	**6.46**
SVDFormer	SVFNet	6.54
SVDFormer	MAFE	**6.41**
SeedFormer	PNet++	6.74
SeedFormer	MAFE	**6.32**

Table 6: Ablation study: Performance improvement when replacing different backbones' feature extraction modules with MAFE on PCN.

Anchor Points	CD↓	Time (ms)
64	6.40	**23.11**
128	**6.32**	26.58
256	**6.32**	35.02
2048	–	853.12

Table 7: Effect of anchor point number on completion performance and inference speed on PCN.

tent benefits across diverse backbone architectures.

5.3 Impact of Anchor Point Number

We investigate the effect of the anchor set scale on both model accuracy and computational efficiency through controlled ablation studies, with results summarized in Table 7. As shown, increasing the number of anchor points from 64 to 128 leads to a noticeable improvement in completion accuracy, with the CD decreasing from 6.40 to 6.32, while maintaining a modest increase in inference time (from 23.11 ms to 26.58 ms). Further increasing the anchor count to 256 yields no additional gains in accuracy (CD remains at 6.32) but results in higher inference latency (35.02 ms), highlighting diminishing returns beyond 128 anchors.

Setting the anchor count equal to the number of input points (2048) effectively removes the abstraction layer, necessitating geodesic distance computations across all points. This configuration is computationally prohibitive, as evidenced by a dramatic increase in inference time to 853.12 ms and infeasible memory usage, making it unsuitable for practical deployment.

6 Conclusion

In this work, we proposed a manifold-aware point cloud completion framework that explicitly integrates geodesic distance estimation and manifold-structured attention into the feature extraction process. By capturing the intrinsic nonlinear geometry of point clouds, our method significantly enhances both geometric fidelity and semantic consistency in reconstructed shapes. Comprehensive evaluations across multiple standard benchmarks confirm that our approach consistently surpasses prior state-of-the-art methods, especially in challenging scenarios with sparse or incomplete observations. We plan to extend our framework to handle dynamic scenes and to explore its integration with multi-modal sensory inputs, aiming to further improve robustness and applicability in real-world environments.

Acknowledgements

This work is supported by the Science and Technology Commission of Shanghai under grant No. 24ZR1400400.

References

[Cai *et al.*, 2024] Pingping Cai, Deja Scott, Xiaoguang Li, and Song Wang. Orthogonal dictionary guided shape completion network for point cloud. In *AAAI*, 2024.

[Chang *et al.*, 2015] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.

[Chen *et al.*, 2023] Zhikai Chen, Fuchen Long, Zhaofan Qiu, Ting Yao, Wengang Zhou, Jiebo Luo, and Tao Mei. Anchorformer: Point cloud completion from discriminative nodes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13581–13590, 2023.

[Dai *et al.*, 2017] Angela Dai, Charles R Qi, and Matthias Nießner. Shape completion using 3d-encoder-predictor cnns and shape synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5868–5877, 2017.

[Dijkstra, 1959] Edsger W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, 1959.

[Fan *et al.*, 2017] Haoqiang Fan, Hao Su, and Leonidas Guibas. A point set generation network for 3d object reconstruction from a single image. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul 2017.

[Geiger *et al.*, 2012] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3354–3361, 2012.

[Groueix *et al.*, 2018] Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan Russell, and Mathieu Aubry. Atlasnet: A papier-mâché approach to learning 3d surface generation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 216–224, 2018.

[Newcombe *et al.*, 2011] Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohli, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pages 127–136, 2011.

[Pan *et al.*, 2021] Liang Pan, Xinyi Chen, Zhongang Cai, Junzhe Zhang, Haiyu Zhao, Shuai Yi, and Ziwei Liu. Variational relational point completion network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8524–8533, 2021.

[Qi *et al.*, 2017] Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 5099–5108, 2017.

[Qi *et al.*, 2018] Xiaojuan Qi, Renjie Liao, Zhengzhe Liu, Raquel Urtasun, and Jiaya Jia. Geonet: Geometric neural network for joint depth and surface normal estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 283–291, 2018.

[Rong *et al.*, 2024] Yi Rong, Haoran Zhou, Lixin Yuan, Cheng Mei, Jiahao Wang, and Tong Lu. Cra-pcn: Point cloud completion with intra- and inter-level cross-resolution transformers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 4676–4685, 2024.

[Tchapmi *et al.*, 2019] Lyne P. Tchapmi, Christopher B. Choy, Iro Armeni, JunYoung Gwak, and Silvio Savarese. Topnet: Structural point cloud decoder. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 383–392, 2019.

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

[Wang *et al.*, 2024] Jun Wang, Ying Cui, Dongyan Guo, Junxia Li, Qingshan Liu, and Chunhua Shen. Pointattn: You only need attention for point cloud completion. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(6):5472–5480, Mar. 2024.

[Xiang *et al.*, 2021] Peng Xiang, Xin Wen, Yu-Shen Liu, Yan-Pei Cao, Pengfei Wan, Wen Zheng, and Zhizhong Han. Snowflakenet: Point cloud completion by snowflake point deconvolution with skip-transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5479–5489, 2021.

[Xie *et al.*, 2020] Hanyu Xie, Hongting Yao, Xiaoguang Sun, Shangchen Zhou, Shengping Zhang, Zhengjun Zhang, and Hujun Bao. Grnet: Gridding residual network for dense point cloud completion. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 365–381, 2020.

[Yan *et al.*, 2025] Hongyu Yan, Zijun Li, Kunming Luo, Li Lu, and Ping Tan. Symmcompletion: High-fidelity and high-consistency point cloud completion with symmetry guidance. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 9094–9102, 2025.

[Yang *et al.*, 2019] Yaoqing Yang, Chen Feng, Yiru Shen, and Dong Tian. Foldingnet: Point cloud auto-encoder via deep grid deformation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 206–215, 2019.

[Yu *et al.*, 2021] Qiangeng Yu, Lingxiao Xie, Wenhai Chen, Ying Wei, Chunhua Shen, and Tongliang Liu. Pointr: Diverse point cloud completion with geometry-aware transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19980–19990, 2021.

[Yu *et al.*, 2023] Xumin Yu, Yongming Rao, Ziyi Wang, Jiwen Lu, and Jie Zhou. Adapointr: Diverse point cloud completion with adaptive geometry-aware transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19646–19656, 2023.

[Yuan *et al.*, 2018] Wentao Yuan, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. Pcn: Point completion network. In *3D Vision (3DV), 2018 International Conference on*, 2018.

[Zhong *et al.*, 2025] Yi Zhong, Weize Quan, Dong-Ming Yan, Jie Jiang, and Yingmei Wei. Pointcformer: A relation-based progressive feature extraction network for point cloud completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 2310–2318, 2025.

[Zhou *et al.*, 2022] Haoran Zhou, Yun Cao, Wenqing Chu, Junwei Zhu, Tong Lu, Ying Tai, and Chengjie Wang. Seedformer: Patch seeds based point cloud completion with upsample transformer. In *Computer Vision – ECCV 2022*, volume 13663 of *Lecture Notes in Computer Science*, pages 416–432. Springer, 2022.

[Zhou *et al.*, 2025] Feng Zhou, Qi Zhang, Ju Dai, Lei Li, Qing Fan, and Junliang Xing. Position-aware guided point cloud completion with clip model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 10734–10742, 2025.

[Zhu *et al.*, 2025] Zhe Zhu, Honghua Chen, Xing He, and Mingqiang Wei. Pointsea: Point cloud completion via self-structure augmentation. *International Journal of Computer Vision*, pages 1–25, 2025.