

DANCE: Density-Agnostic and Class-Aware Network for Point Cloud Completion

Da-Yeong Kim¹, Yeong-Jun Cho¹

¹Department of Artificial Intelligence Convergence
Chonnam National University, Gwangju 61186, South Korea
qetuo0909@jnu.ac.kr, yj.cho@jnu.ac.kr

Abstract

Point cloud completion aims to recover missing geometric structures from incomplete 3D scans, which often suffer from occlusions or limited sensor viewpoints. Existing methods typically assume fixed input/output densities or rely on image-based representations, making them less suitable for real-world scenarios with variable sparsity and limited supervision. In this paper, we introduce Density-agnostic and Class-aware Network (DANCE), a novel framework that completes only the missing regions while preserving the observed geometry. DANCE generates candidate points via ray-based sampling from multiple viewpoints. A transformer decoder then refines their positions and predicts opacity scores, which determine the validity of each point for inclusion in the final surface. To incorporate semantic guidance, DANCE includes a classification head and fusion network trained directly on geometric features, enabling category-consistent completion without relying on external image supervision. Extensive experiments on the PCN and MVP benchmarks show that DANCE outperforms state-of-the-art methods in accuracy and structural consistency, while remaining robust to varying input densities and noise levels.

Code — <https://github.com/ayeong0909/DANCE>

1 Introduction

Point cloud completion aims to recover missing geometric structures from incomplete 3D point clouds, typically caused by occlusions, limited sensor coverage, or constrained viewpoints. This task is essential for downstream applications like autonomous driving, robotics, and 3D reconstruction, where complete and accurate 3D representations are critical. The objective is to generate a plausible and structurally consistent point cloud that fills in missing regions while preserving the geometry of the observed input.

Recent advances in deep learning have led to point cloud completion methods (Yuan et al. 2018) that generate complete shapes from partial inputs by learning global geometric priors. Building on this, recent generative approaches (Li, Zhu, and Wei 2025; Wei et al. 2025) leverage powerful image-to-3D models by converting partial point clouds into 2D images and performing shape completion using

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

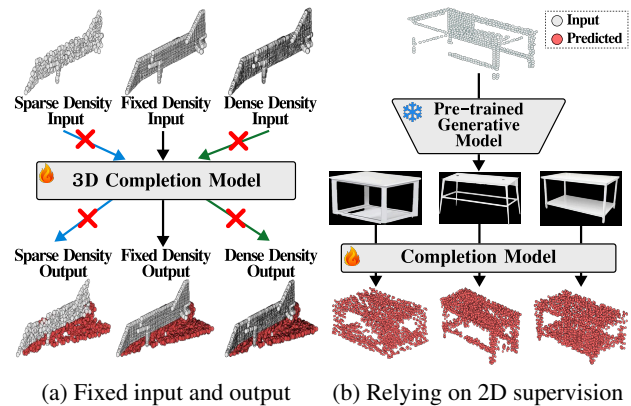


Figure 1: Limitations of previous methods. (a) Fixed input/output densities. Prior methods assume fixed-size inputs and outputs, which limits their applicability in real-world scenarios. (b) Image-based generative frameworks. Methods relying on 2D image representations often produce shapes that do not align with the original 3D input geometry.

2D generative priors. Meanwhile, partial completion methods (Huang et al. 2020; Yu et al. 2021) focus on reconstructing only the missing regions, aiming to preserve the observed geometry and avoid regenerating known areas.

Although these methods have achieved significant progress, they still suffer from two key limitations. First, most approaches assume fixed input and output densities, which is unsuitable for real-world scenarios (see Fig.1(a)). In practice, input sparsity varies depending on object distance and sensor resolution, yet many methods generate a fixed number of points regardless of the actual level of incompleteness or the desired output density. Second, many methods incorporate semantic priors by relying on generative frameworks that use image-based representations. This reliance on external image representations prevents direct learning from 3D geometry. As shown in Fig. 1(b), it often leads to structurally inconsistent outputs that deviate from the input object due to the strong bias of well-trained 2D generative models.

To overcome these limitations, we propose Density-agnostic and Class-aware Network (DANCE) for point

cloud completion. DANCE incorporates a lightweight classification head and fusion network trained directly on geometric features to address the lack of semantic guidance. Unlike prior methods that depend on image-based representations through generative frameworks, this module learns semantic priors purely from 3D point cloud data. In addition, DANCE adopts a density-agnostic design inspired by NeRF-style (Mildenhall et al. 2021) sampling. Given a partial point cloud, it generates candidate points via ray-based sampling over the observed surface and refines them using a transformer decoder. An opacity score is also predicted for each candidate to determine its inclusion in the output. This enables DANCE to process inputs of varying density and produce outputs without relying on fixed output density.

We conduct extensive experiments on point cloud completion benchmarks (PCN, MVP) and demonstrate that DANCE outperforms state-of-the-art methods in both accuracy and structural consistency. Thanks to its density-agnostic design, our model maintains stable performance across varying input densities and generates outputs with flexible resolution. In addition, DANCE achieves superior class-aware completion quality, producing plausible and category-consistent shapes without relying on image supervision. Overall, DANCE provides a density-agnostic and class-aware completion framework that achieves flexible, semantically guided outputs without relying on image supervision. Our main contributions are as follows:

- We propose DANCE, a novel point cloud completion framework that is density-agnostic and class-aware, designed to overcome key limitations of existing methods.
- DANCE achieves state-of-the-art performance in accuracy and structural consistency on standard benchmarks such as PCN and MVP.
- Our method is scalable and practical for real-world scenarios, requiring no image supervision and robustly handling varying input sparsity.

To the best of our knowledge, this is the first attempt at density-agnostic 3D point cloud completion.

2 Related Works

Thanks to recent progress in deep learning, many studies have focused on learning-based point cloud completion. Point cloud completion methods aim to generate complete shapes from partial inputs. PCN (Yuan et al. 2018) introduced a pioneering encoder-decoder framework that extracts global features from partial point clouds and reconstructs complete shapes using a FoldingNet (Yang et al. 2017)-based decoder. Recent studies have begun leveraging well-designed generative models, particularly 2D image-to-3D architectures, to improve point cloud completion. For example, GenPC (Li, Zhu, and Wei 2025) converts partial point clouds into intermediate 2D image representations and applies pre-trained image-to-3D generative models to complete shapes. PCDreamer (Wei et al. 2025) adopts a diffusion-based framework that gradually refines 3D shapes by learning latent distributions. It combines geometric cues and learned shape priors to generate high-quality completions, even in the presence of large missing regions.

Although generative model-based completion methods demonstrate impressive completion performance, they still have several limitations. First, regenerating the entire point cloud, including visible regions, can lead to geometric distortions and the loss of fine details originally present in the input. Second, methods that require paired image-point cloud datasets are impractical in real-world settings due to the high cost and difficulty of data collection.

While many recent methods adopt generative models that regenerate the entire point cloud, some earlier approaches have focused on completing the missing regions to preserve the observed geometry. For instance, PF-Net (Huang et al. 2020) introduces a pyramid decoder that generates missing points at multiple scales while retaining the original input points throughout the process. PoinTr (Yu et al. 2021) applies transformers to point cloud completion by dividing the input into small groups called “point proxies” that represent local regions. Recently, although not limited to missing region completion, transformer-based methods such as SeedFormer (Zhou et al. 2022) and SVDFormer (Zhu et al. 2023) combined partial completion with global shape modeling.

Despite their strengths, these completion methods still have important limitations. One key issue is their fixed input and output density, which is not suitable for real-world scenarios. In practice, input point clouds can be either sparse or dense, depending on factors such as the distance to the object or the resolution of the 3D sensor. Moreover, the appropriate output density is typically undefined. For example, they generate a fixed number of points regardless of how much of the shape is missing, resulting in unrealistic and uneven point distributions. Another limitation is that these methods generally lack semantic awareness and treat all object categories uniformly, without leveraging category-specific shape priors that could guide more accurate completion.

3 Proposed Methods

The goal of point cloud completion is to reconstruct the complete 3D point cloud of an object by predicting the missing geometric structures from an incomplete point cloud $P^I \in \mathbb{R}^{N \times 3}$, where N is the number of points. The proposed DANCE first generates candidate points to predict missing regions in 3D space (Sec. 3.1). The candidate points and given incomplete point cloud are processed through an encoder-decoder architecture. The encoder (Sec. 3.2) extracts 3D features from both candidate points and given incomplete point cloud, while the decoder (Sec. 3.3) reconstructs the geometry of the missing regions P^{out} . The overall pipeline of the proposed DANCE is illustrated in Fig. 2.

Most existing point cloud completion methods (Li, Zhu, and Wei 2025; Wei et al. 2025) generate outputs with a fixed density, regardless of the input’s sparsity or density. In contrast, the proposed DANCE is density-agnostic: it works with inputs of varying densities and does not require a fixed output size. Additionally, DANCE leverages the classification results of the incomplete point cloud P^I to guide the completion process, allowing semantic information to improve completion accuracy. Thus, DANCE performs density-agnostic and class-aware point cloud completion, enabling significantly more natural and accurate recon-

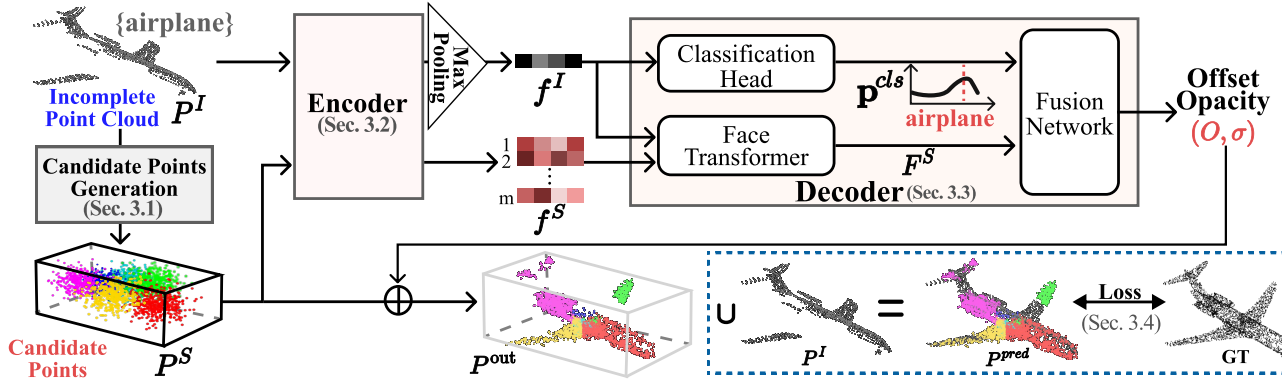


Figure 2: The overall pipeline of DANCE. It generates candidate points using ray-based sampling strategy to reconstruct missing regions of incomplete point clouds. The encoder extracts features from both the incomplete input and generated points, and the decoder processes them to estimate the offset and opacity of each point. Final completion results are obtained by refining candidate points using the predicted offset and selecting valid ones based on the predicted opacity. Best viewed in color.

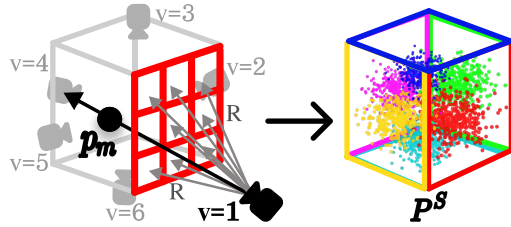


Figure 3: Candidate point generation in DANCE. Candidate points are generated using a ray-based sampling strategy. Colors represent the corresponding faces and their associated candidate points. Best viewed in color.

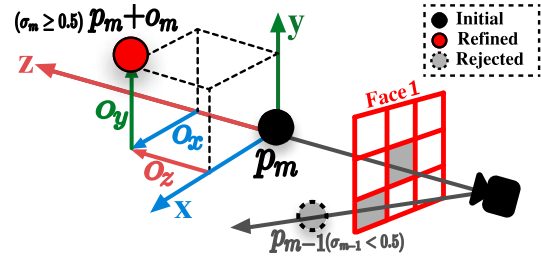


Figure 4: Offset and opacity estimation in DANCE. The estimated offset and opacity refine the position of each candidate point and control its inclusion in the final completion.

struction results, and making it well-suited for real-world applications with varying input conditions.

Candidate Points Generation

Motivated by NeRF (Mildenhall et al. 2021), which samples 3D points along multi-view camera rays for novel view synthesis, we adopt a similar ray-based sampling strategy for point cloud completion. NeRF predicts color and opacity at sampled 3D points to synthesize novel views, as its goal is image rendering rather than 3D reconstruction. In contrast, our task focuses on completing missing 3D geometry from incomplete point clouds. Therefore, instead of predicting color as in NeRF, we focus on estimating the 3D positions and opacity of the candidate points.

To reconstruct the missing regions of an incomplete given point cloud P^I , we generate a set of candidate points P^S . To this end, we define V viewpoints surrounding P^I , each associated with a face oriented toward the object, as illustrated in Fig. 3. For example, setting $V = 6$ allows the six viewpoints and faces to form a hexahedral structure that surrounds the object. An $R \times R$ grid is placed on each face, and rays are cast from the corresponding viewpoint through each grid location. A single 3D point is then sampled along each ray based on a Gaussian distribution centered between

the face and the object surface. As object point clouds are defined only on the surface, sampling one point per ray is sufficient for representing the geometry. This results in the initial candidate point set $P^S = \{p_m\}_{m=1}^M \in \mathbb{R}^{M \times 3}$, where $M = VR^2$ denotes the total number of the points. Here, $P_v^S \in \mathbb{R}^{R^2 \times 3}$ denotes the set of candidate points corresponding to the v -th viewpoint. Each P_v^S is visualized with a different color in Fig. 3. Note that P^S is not perfect yet, but serves as the starting point for reconstruction. It is subsequently refined into accurate 3D positions through the encoder-decoder network.

Encoder for 3D Feature Extraction

A set of candidate points P^S and the incomplete point cloud P^I are fed into the 3D encoder E to extract candidate and global 3D features as follows:

$$f^S = E(P^S) \in \mathbb{R}^{M \times d_{\text{en}}}, \quad (1)$$

$$f^I = \text{maxpool}(E(P^I)) \in \mathbb{R}^{1 \times d_{\text{en}}}, \quad (2)$$

where d_{en} denotes the dimension of the feature vector. Note that P^I additionally passes through a max-pooling layer, as defined in Eq. 2, to extract a global feature. Although the global feature f^I is derived from the incomplete point cloud P^I , it serves as a compact latent representation that encodes

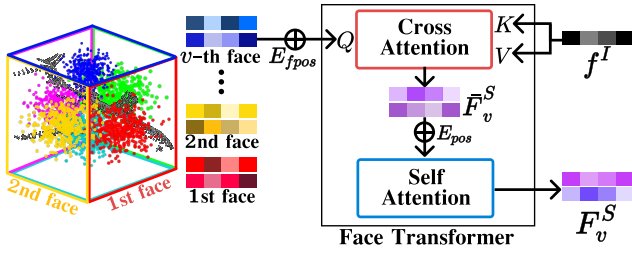


Figure 5: Architecture of the transformer in DANCE. Cross-attention between f_v^S and f^I followed by self-attention within viewpoint groups v .

global shape priors of the visible regions. In parallel, the candidate feature f^S , extracted from the candidate points P^S , contributes to the prediction of the missing structures.

As the backbone for the encoder, any architecture such as the PointNet series (Qi et al. 2017a,b) or DGCNN (Wang et al. 2019) can be used. P^S and P^I share the same encoder E , which not only improves efficiency but also ensures that both the candidate points and the incomplete point cloud are represented in a consistent feature space.

Decoder for Point Cloud Completion

The decoder consists of three main components: 1) a face transformer; 2) a classification head; and 3) a fusion network. It performs point cloud completion by predicting the 3D coordinate offset and opacity value for each candidate point in P^S . Based on these predictions, the missing 3D point cloud P^{out} is estimated. Importantly, the proposed decoder focuses solely on reconstructing the missing regions, rather than regenerating the entire object point cloud. Moreover, thanks to its transformer-based design, the decoder can handle input point clouds with varying densities.

We define the offset $o_m = \{o_x, o_y, o_z\}$ to represent the (x, y, z) displacements of the m -th candidate point p_m . For each candidate point p_m , a local coordinate system is defined with p_m as the origin, where the sampling ray is aligned with the z -axis, and the two axes of the corresponding face define the x - and y -axes, as shown in Fig. 4. The opacity value σ represents the influence of each point, similar to its role in NeRF (Mildenhall et al. 2021). Only points with σ above a certain value (≥ 0.5) are considered meaningful, while the rest are filtered out. This decoder design allows the model to adjust the output point cloud density. Then, the missing 3D point cloud P^{out} is defined as follows:

$$P^{\text{out}} = \{p_m + o_m \mid \sigma_m \geq 0.5, m = 1, \dots, M\}. \quad (3)$$

The final completion result is obtained by combining the predicted points P^{out} with the incomplete input P^I , yielding $P^{\text{pred}} = P^I \cup P^{\text{out}}$.

Face Transformer. The face transformer enhances the interaction between the candidate feature f^S and the global feature f^I through cross- and self-attention layers as shown in Fig. 5. We assume that points generated from the same viewpoint share similar geometric and visibility characteristics. To reflect this, we process the candidate feature set f_v^S

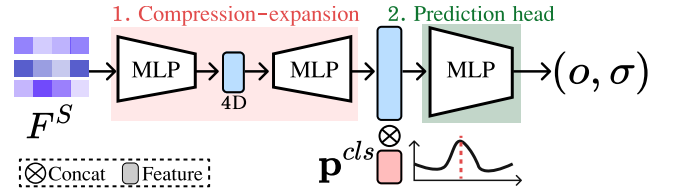


Figure 6: The architecture of the fusion network for offset and opacity prediction. The geometric feature F^S is refined by incorporating classification results \mathbf{p}^{cls} to predict o, σ .

from each viewpoint v independently. Each candidate feature f_v^S undergoes cross-attention with the global feature f^I as follows:

$$\bar{F}_v^S = \text{crossAttn}(Q = f_v^S + E_v^{f_{pos}}, K = f^I, V = f^I), \quad (4)$$

where $E_v^{f_{pos}}$ is the positional embedding of viewpoint, maintaining the spatial relationships between different viewpoints. Since f^I is a compact latent representation encoding global shape priors from the visible regions, this cross-attention allows each candidate point to incorporate global structural context. As a result, the point features are refined to better guide the completion of missing regions.

To further enhance local geometric consistency, we apply self-attention within each \bar{F}_v^S as follows:

$$F_v^S = \text{selfAttn}(\bar{F}_v^S + E_{pos}), \quad (5)$$

where Q, K and V denote the query, key, and value in the transformer. E_{pos} is a positional embedding for each candidate point. Points generated from the same viewpoint exhibit strong spatial correlation; therefore, this intra-group self-attention helps enforce consistency among neighboring points. After applying self-attention to each set of features \bar{F}_v^S , the whole feature representation $F^S = \{F_1^S, F_2^S, \dots, F_V^S\}$ is obtained. The proposed two-stage attention simultaneously achieves both global consistency and local precision for reconstructing missing regions.

Classification Head. Most existing studies rely solely on spatial structure of the object for point cloud completion. However, even from incomplete point clouds P^I , it is possible to infer the object’s class distribution. In the proposed DANCE, this semantic information is utilized as a strong prior to guide and improve the reconstruction of missing regions. The classification head is implemented using multi-layer perceptrons (MLP) followed by a softmax function:

$$\mathbf{p}^{cls} = \text{softmax}(\text{MLP}(f^I)) \in \mathbb{R}^c, \quad (6)$$

where c is the number of categories, and \mathbf{p}^{cls} represents the class probability distribution.

Fusion Network Finally, the fusion network integrates point feature representation F^S with the predicted class distribution \mathbf{p}^{cls} to estimate the final offsets and opacities of the candidate points P^S . The fusion network consists of a compression-expansion MLP and a prediction head as shown in Fig. 6. First, the compression-expansion MLP compresses the input feature representation $F^S \in \mathbb{R}^{M \times d_{en}}$

to a 4-dimensional vector and then reconstructs it back to the original dimension. The 4D bottleneck is designed to match the dimensionality of the prediction head’s output (offset and opacity), encouraging the model to focus on essential task-specific cues.

Next, the output vector of the MLP is concatenated with the class probability distribution \mathbf{p}^{cls} . The combined representation is then passed through the prediction head to infer the offset and opacity for each candidate point. Finally, we obtain the offset $o_m = \{o_x, o_y, o_z\}$ and the opacity value σ_m for each candidate point p_m . Based on the predicted offsets and opacity values, the missing 3D point cloud P^{out} is computed as described in Eq. 3.

Loss Functions

To train the proposed method, we jointly optimize the completion loss and the classification loss. The completion loss measures the geometric difference between the predicted point cloud P^{pred} and the ground truth point cloud P^{GT} . To this end, we employ Chamfer Distance (CD), defined as:

$$\mathcal{L}_{CD} = CD(P^{pred}, P^{GT}). \quad (7)$$

It measures the similarity between two point clouds by summing the distances from each point in one set to its nearest neighbor in the other set.

The classification loss encourages the model to correctly predict the object category from the incomplete point cloud feature f^I . We compute the cross-entropy loss between the predicted class probability distribution \mathbf{p}^{cls} and the ground truth one-hot label vector $\mathbf{y} = \{y_1, \dots, y_c\}$ as follows:

$$\mathcal{L}_{cls} = - \sum_{i=1}^c y_i \log(\mathbf{p}_i^{cls}). \quad (8)$$

The classification loss provides shape priors that guide the network to complete partially missing regions more reliably. The total loss function is formulated as a weighted sum of the completion and classification losses:

$$\mathcal{L}_{total} = \lambda \mathcal{L}_{CD} + (1 - \lambda) \mathcal{L}_{cls}. \quad (9)$$

where λ is a hyperparameter that balances the influence of the classification loss.

4 Experimental Results

Datasets and Settings

Datasets. We evaluate our method on two widely used point cloud completion benchmarks: PCN (Yuan et al. 2018) and MVP (Pan et al. 2021), both constructed from subsets of ShapeNet (Chang et al. 2015). PCN includes 8 object categories (e.g., Plane, Chair, Table) and provides approximately 30,000 complete point clouds and 240,000 incomplete ones. The incomplete point clouds are generated by removing parts from 8 different viewpoints, with 2,048 points per sample, while complete shapes consist of 4,096 points. MVP, on the other hand, covers 16 categories and introduces more diverse partiality by simulating 26 viewpoints. Although the input size remains fixed at 2,048 points, it offers complete ground truths at multiple resolutions (4,096, 8,192, and 16,384 points).

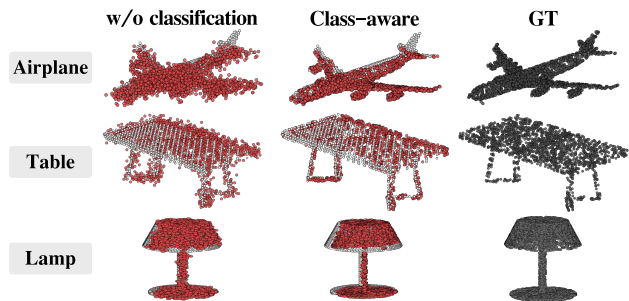


Figure 7: Effects of the proposed class-aware approach. Completion results with and without classification head. Gray and red points represent the input and predicted completions, respectively.

Methods	CD-Avg↓	DCD-Avg↓	F1↑
w/o classification	6.46	0.528	0.856
w/o face attention	6.52	0.534	0.849
Ours	6.42	0.524	0.859

Table 1: Ablation study of proposed methods. Effectiveness of each component measured on PCN dataset. The full model achieves the best performance across all evaluation metrics.

Settings. All experiments were conducted with ground truth point clouds containing 4,096 points. We set $V=6$ and $R=21$ for ray-based candidate points generation on each face. The opacity threshold was fixed at 0.5 to filter candidate points based on their predicted opacity values. We employ Chamfer Distance (CD), Density-aware CD (DCD), and F1-Score@1% as evaluation metrics. CD-avg measures the average nearest-neighbor distance between predicted and ground truth point clouds, while DCD-avg improves upon CD by taking local point density into account. F1-Score@1% is defined as the harmonic mean of precision and recall at a 1% distance threshold. Following prior works, we use L1-based CD for the PCN and L2-based CD.

Effectiveness of the Proposed Methods

To evaluate the individual contributions of our proposed methods, we conducted comprehensive ablation studies on the PCN dataset (Yuan et al. 2018). We first evaluated the effectiveness of class-aware completion using the proposed decoder structure introduced in Sec.3.3. As shown in Tab.1, the model with the classification head outperforms the baseline without the head across all evaluation metrics. Notably, we observed a reduction in CD-Avg from 6.46 to 6.42, indicating improved reconstruction accuracy. The results in Fig. 7 also demonstrate the effectiveness of the class-aware approach. It produces more geometrically coherent completions and better preserves category-specific structural characteristics. This result supports that semantic priors extracted from incomplete inputs help the network resolve ambiguities in missing regions.

We evaluated the effectiveness of the face transformer in the decoder, which performs self-attention within viewpoint-

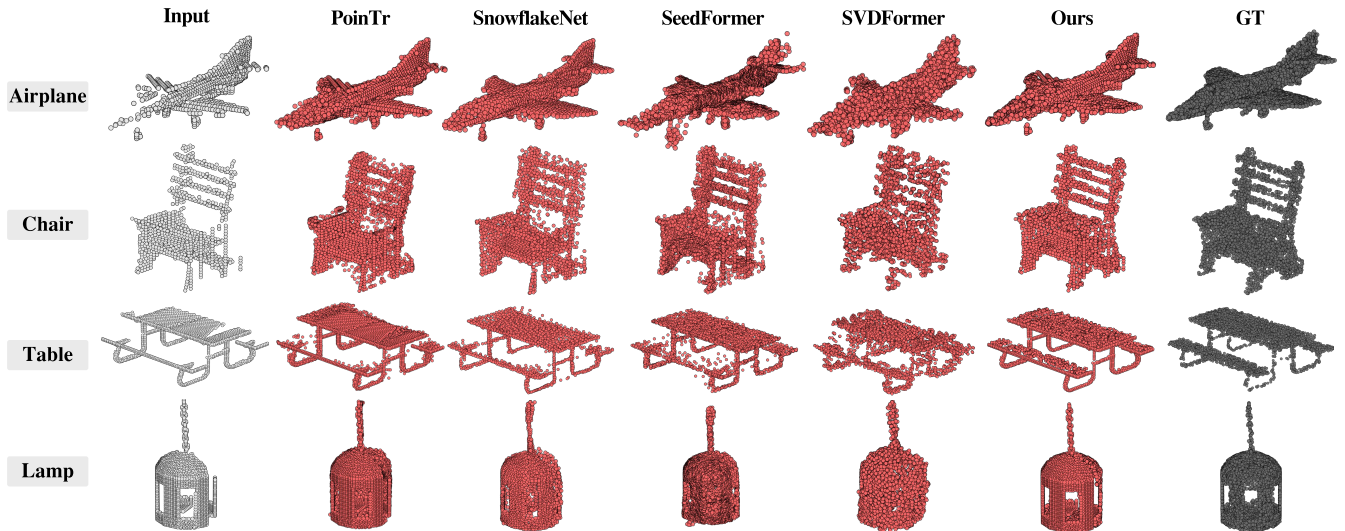


Figure 8: Qualitative Comparison with state-of-the-art methods on the PCN dataset. DANCE generates more complete and structurally consistent shapes compared to prior methods. The completions better preserve the observed geometry while producing plausible details, demonstrating the effectiveness of our density-agnostic and class-aware design.

Methods	Plane	Cabinet	Car	Chair	Lamp	Couch	Table	Boat	CD-Avg↓	DCD-Avg↓	F1↑
PCN (2018)	5.82	10.91	9.00	11.09	11.91	12.06	9.56	9.12	9.93	–	0.657
GRNet (2020)	8.51	12.84	10.17	12.13	11.83	12.64	11.16	9.84	11.14	0.765	0.563
PointTr (2021)	4.31	9.23	7.60	8.35	8.27	9.38	7.97	7.01	7.76	0.562	0.810
PMP-Net++ (2022)	4.80	11.70	8.96	8.93	7.04	12.25	8.57	6.88	8.64	0.714	0.683
SnowflakeNet (2021)	3.95	8.82	7.52	7.48	6.34	8.88	6.61	6.10	6.96	0.567	0.828
SeedFormer (2022)	3.87	9.05	7.53	7.63	6.12	9.26	6.61	6.03	7.01	0.573	0.824
AnchorFormer (2023)	3.62	8.79	7.20	7.12	6.18	8.71	6.50	6.01	6.77	0.539	0.841
HyperCD (2023)	3.87	9.05	7.40	7.50	5.95	9.25	6.50	5.89	6.93	0.568	0.822
SVDFormer (2023)	3.68	8.73	7.10	6.95	5.64	8.58	6.26	5.91	6.61	0.534	0.848
CRA-PCN (2024)	3.62	8.77	7.00	6.92	5.46	8.59	6.27	5.86	6.56	0.537	0.846
PCDreamer (2025)	3.51	8.62	6.92	6.91	5.66	8.31	6.27	5.90	6.52	0.531	0.856
Ours	3.53	8.6	6.88	6.89	5.44	8.28	6.24	5.82	6.46	0.528	0.859

Table 2: Performance comparison with state-of-the-art methods on the PCN dataset. DANCE achieves the best overall performance across all metrics, consistently outperforming existing methods. The best results are highlighted in bold.

specific groups, by comparing models with and without this design. As shown in Tab. 1, removing this component consistently degraded performance across all metrics. The results highlight that performing self-attention within viewpoint-specific groups enhances completion accuracy by exploiting local geometric coherence.

Comparison with State-of-the-Art Methods

We compared DANCE against state-of-the-art point cloud completion methods on the PCN and MVP benchmarks. As shown in Tab. 2, our method achieves superior performance across most object categories, with an overall CD-Avg of 6.46 on the PCN dataset. Figure 8 presents qualitative comparisons with other point cloud completion methods. Our approach achieves more accurate reconstruction of fine geometric details while maintaining global shape consistency. For objects with complex structures, such as lamps and chairs, DANCE effectively recovers thin parts and symmetric components that existing methods often fail to recon-

struct.

This improvement is mainly due to our design, which infers only the missing regions while preserving the input geometry, and leverages semantic guidance from the classification head to provide category-specific shape priors for ambiguous areas. To further support this observation, we conducted a qualitative comparison with PoinTr (Yu et al. 2021), a method that also focuses on completing only the missing regions while preserving the input point cloud. As shown in Fig. 9, DANCE achieves more accurate reconstruction of the missing parts, especially in recovering fine-grained geometric structures that PoinTr often fails to capture.

To assess the robustness of our method under varying densities and challenging conditions, we conducted experiments on the MVP dataset (Pan et al. 2021), which provides diverse partial inputs from 26 viewpoints. Table 3 presents comparisons at resolutions of 4,096 and 8,192 points. DANCE outperforms existing methods across both density settings, demonstrating robustness to variations in point cloud den-

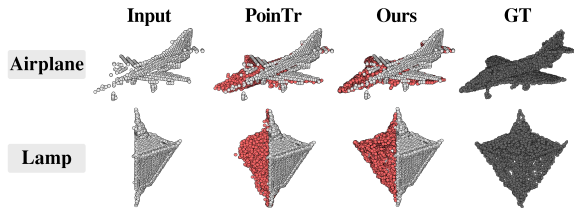


Figure 9: Qualitative comparison with PoinTr and DANCE. Both methods complete only the missing regions, but DANCE achieves more accurate and consistent shapes by leveraging stronger structural and semantic cues.

Methods	4096		8192	
	CD-Avg ↓	F1 ↑	CD-Avg ↓	F1 ↑
PCN (2018)	7.96	0.463	6.86	0.565
TopNet (2019)	7.69	0.434	6.64	0.526
ECG (2020)	7.31	0.506	4.23	0.695
PoinTr (2021)	4.69	0.598	3.52	0.712
VRCNet (2021)	4.82	0.629	3.68	0.720
SeedFormer (2022)	4.61	0.605	3.59	0.710
PDR (2021)	4.42	0.638	3.45	0.739
DualGenerator (2023)	4.29	0.643	3.38	0.747
Ours	4.19	0.662	3.37	0.754

Table 3: Performance comparison on MVP dataset. DANCE achieves superior results under both 4096 and 8192 input densities, highlighting its robustness to input sparsity.

sity. We set $R = 21$ for the 4,096-point output and $R = 33$ for the 8,192-point output to approximately match the target output density. These are experimental settings to ensure that the number of candidate points exceeds the desired output size. However, the choice of R is not critical to performance, as discussed in Sec.4.4. Since DANCE filters unnecessary candidates based on predicted opacity, the final output naturally approximates the target density. Moreover, in practical scenarios, the output density can be flexibly controlled by adjusting R as needed.

Real-World Robustness of DANCE

We evaluate the robustness of DANCE in two real-world scenarios: (1) noisy inputs from sensor artifacts, and (2) varying input and output point densities. Real-world point clouds often contain noise due to sensor limitations or environmental interference. To evaluate the robustness of DANCE to such noise, we added Gaussian noise at varying levels to PCN dataset. Figure 10 shows the change in CD-Avg as the noise level increases. DANCE outperforms state-of-the-art methods SVDFormer and SeedFormer, and exhibits greater robustness to various noise levels.

Previous methods typically assume fixed input and output densities, limiting their flexibility in handling varying point cloud sparsity, which does not reflect real-world scenarios. In contrast, DANCE achieves density-agnostic completion by leveraging a transformer-based decoder that flexibly operates regardless of the input point cloud density, and adjusts the output density through its opacity prediction process. Figure 11 shows that DANCE allows flexible control over

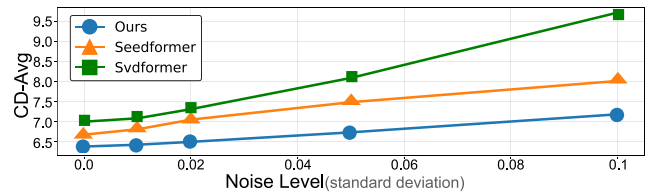


Figure 10: Performance comparison under varying noise levels. DANCE shows strong robustness to noise, exhibiting smaller performance degradation compared to other methods across all noise levels.

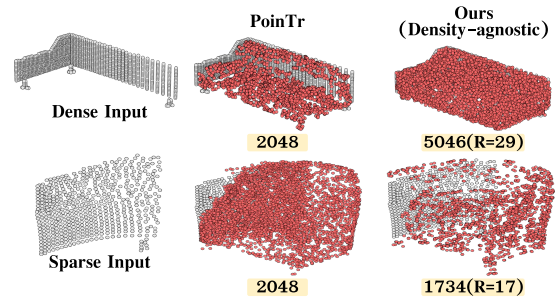


Figure 11: Density-agnostic completion of DANCE. It operates robustly across varying input densities, and the output density is determined according to the sampling parameter R and the predicted opacity values.

the output density. This is achieved by adjusting the face sampling parameter R and filtering out low-opacity candidate points. Although DANCE is trained with a fixed setting of $R = 21$, the output density can be modified at inference time by changing R to values like 17 or 29, without requiring retraining. This density-agnostic capability makes DANCE more adaptable to real-world scenarios.

5 Conclusions and Future Works

In this paper, we proposed DANCE, a density-agnostic and class-aware network for point cloud completion. DANCE generates candidate points using a ray-based sampling strategy and refines them with a transformer decoder that predicts their position and opacity. To provide semantic guidance, we include a simple classification module that learns shape categories directly from incomplete point clouds—without using image-based supervision. Thanks to this design, DANCE works well with inputs of different densities and produces flexible, high-quality completions without the need for retraining in different settings.

However, DANCE currently adopts a fixed sampling configuration, using predetermined viewpoint positions and a constant number of views regardless of the input characteristics. While this setting works well in general, it may be suboptimal for objects with highly complex or asymmetric structures. As future work, adaptive view sampling strategies can be explored to dynamically adjust the sampling positions and view counts based on the geometric patterns of the input, further improving completion quality in challenging cases.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2025-24683045), and by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grants funded by the Korea government (MSIT): the Artificial Intelligence Convergence Innovation Human Resources Development (IITP-2023-RS-2023-00256629) and the University ICT Research Center (ITRC) support program (IITP-2025-RS-2024-00437718).

References

- Chang, A. X.; Funkhouser, T.; Guibas, L.; Hanrahan, P.; Huang, Q.; Li, Z.; Savarese, S.; Savva, M.; Song, S.; Su, H.; et al. 2015. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*.
- Chen, Z.; Long, F.; Qiu, Z.; Yao, T.; Zhou, W.; Luo, J.; and Mei, T. 2023. Anchorformer: Point cloud completion from discriminative nodes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 13581–13590.
- Huang, Z.; Yu, Y.; Xu, J.; Ni, F.; and Le, X. 2020. Pf-net: Point fractal network for 3d point cloud completion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7662–7670.
- Li, A.; Zhu, Z.; and Wei, M. 2025. GenPC: Zero-shot Point Cloud Completion via 3D Generative Priors. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 1308–1318.
- Lin, F.; Yue, Y.; Hou, S.; Yu, X.; Xu, Y.; Yamada, K. D.; and Zhang, Z. 2023. Hyperbolic chamfer distance for point cloud completion. In *Proceedings of the IEEE/CVF international conference on computer vision*, 14595–14606.
- Lyu, Z.; Kong, Z.; Pan, L.; Lin, D.; et al. 2021. A Conditional Point Diffusion-Refinement Paradigm for 3D Point Cloud Completion. In *International Conference on Learning Representations*.
- Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106.
- Pan, L. 2020. ECG: Edge-aware point cloud completion with graph convolution. *IEEE Robotics and Automation Letters*, 5(3): 4392–4398.
- Pan, L.; Chen, X.; Cai, Z.; Zhang, J.; Zhao, H.; Yi, S.; and Liu, Z. 2021. Variational relational point completion network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8524–8533.
- Qi, C. R.; Su, H.; Mo, K.; and Guibas, L. J. 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 652–660.
- Qi, C. R.; Yi, L.; Su, H.; and Guibas, L. J. 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30.
- Rong, Y.; Zhou, H.; Yuan, L.; Mei, C.; Wang, J.; and Lu, T. 2024. Cra-pcn: Point cloud completion with intra-and inter-level cross-resolution transformers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 4676–4685.
- Shi, P.; Cheng, H.; Han, X.; Zhou, Y.; and Zhu, J. 2023. DualGenerator: Information interaction-based generative network for point cloud completion. *IEEE Robotics and Automation Letters*, 8(10): 6627–6634.
- Tchapmi, L. P.; Kosaraju, V.; Rezatofighi, H.; Reid, I.; and Savarese, S. 2019. Topnet: Structural point cloud decoder. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 383–392.
- Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S. E.; Bronstein, M. M.; and Solomon, J. M. 2019. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)*, 38(5): 1–12.
- Wei, G.; Feng, Y.; Ma, L.; Wang, C.; Zhou, Y.; and Li, C. 2025. Pcdreamer: Point cloud completion through multi-view diffusion priors. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 27243–27253.
- Wen, X.; Xiang, P.; Han, Z.; Cao, Y.-P.; Wan, P.; Zheng, W.; and Liu, Y.-S. 2022. 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.
- Xiang, P.; Wen, X.; Liu, Y.-S.; Cao, Y.-P.; Wan, P.; Zheng, W.; and Han, Z. 2021. Snowflakenet: Point cloud completion by snowflake point deconvolution with skip-transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, 5499–5509.
- Xie, H.; Yao, H.; Zhou, S.; Mao, J.; Zhang, S.; and Sun, W. 2020. Grnet: Gridding residual network for dense point cloud completion. In *European conference on computer vision*, 365–381. Springer.
- Yang, Y.; Feng, C.; Shen, Y.; and Tian, D. 2017. Foldingnet: Interpretable unsupervised learning on 3d point clouds. *arXiv preprint arXiv:1712.07262*, 2(3): 5.
- Yu, X.; Rao, Y.; Wang, Z.; Liu, Z.; Lu, J.; and Zhou, J. 2021. PointR: Diverse point cloud completion with geometry-aware transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, 12498–12507.
- Yuan, W.; Khot, T.; Held, D.; Mertz, C.; and Hebert, M. 2018. Pcn: Point completion network. In *2018 international conference on 3D vision (3DV)*, 728–737. IEEE.
- Zhou, H.; Cao, Y.; Chu, W.; Zhu, J.; Lu, T.; Tai, Y.; and Wang, C. 2022. Seedformer: Patch seeds based point cloud completion with upsampler transformer. In *European conference on computer vision*, 416–432. Springer.
- Zhu, Z.; Chen, H.; He, X.; Wang, W.; Qin, J.; and Wei, M. 2023. Svdformer: Complementing point cloud via self-view augmentation and self-structure dual-generator. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 14508–14518.