Cross-Domain Alignment with Fine Geometric Perception for Detail-Preserving Point Cloud Completion

Chen Huang
School of Computer Science, Hubei University
Wuhan, China

Haobo Ma
School of Computer Science, Hubei University
Wuhan, China

Yan Zhang
School of Computer Science, Hubei University
Wuhan, China
zhangyan@hubu.edu.cn

Chao Yang
School of Computer Science, Hubei University
Wuhan, China

Jianhua Song
School of Cyber Science and Technology, Hubei University
Wuhan, China

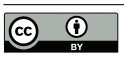
ABSTRACT

Point cloud completion involves inferring and reconstructing the full structure of an object or scene from incomplete 3D point cloud data. Deep learning-based methods typically use encoder-decoder architectures to learn geometric priors from partial inputs for reconstruction. However, these methods often prioritize global features over local geometric details, leading to coarse completions lacking high-frequency information. Sequential application of such models can also cause error accumulation and increased computational costs. To address these issues, we propose CAM-FGP, a **C**ross-domain **A**lignment **M**ethod with **F**ine **G**eometric **P**erception, designed to enhance structural integrity and restore details, especially in regions with missing geometry. CAM-FGP first employs a Fine Geometry Detail Extraction Network (FGDE) to gather high-resolution local details from visible point clouds while integrating low-resolution global information to reinforce the missing areas' structure. Then, a Hierarchical Optimal Transport Network (HOTN) aligns multi-source point cloud distributions, improving the transferability of local geometric features. Lastly, CAM-FGP utilizes a multi-stage hidden state completion and fusion strategy to merge local and global features. This approach preserves continuity, reduces memory-induced information loss, and lowers computational costs. CAM-FGP achieves state-of-the-art performance on several benchmark datasets, demonstrating its superiority in point cloud completion.

KEYWORDS

Point Cloud Completion;Cross-domain Alignment;Fine Geometric Perception;Hierarchical Optimal Transport;Multi-stage Feature Fusion

ACM Reference Format:
Chen Huang, Haobo Ma, Yan Zhang, Chao Yang, and Jianhua Song. 2026. Cross-Domain Alignment with Fine Geometric Perception for Detail-Preserving Point Cloud Completion. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 9 pages. https://doi.org/10.65109/MQXV5147



Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/MQXV5147

1 INTRODUCTION

Point cloud completion plays a crucial role in the development of 3D vision and serves as a core component in a variety of applications such as autonomous driving [2, 13], robotics [6], and augmented reality [9]. Due to occlusions during the scanning process or the limited sensing range of hardware, the collected point cloud data often contain missing regions, posing significant challenges for downstream tasks.

Although numerous studies have proposed promising approaches, challenges remain in fully reconstructing object structures and restoring fine geometric details in generated point clouds. Recent advances in 3D deep set architectures [5, 8, 18, 30, 32] have spurred the development of deep learning-based point cloud completion methods, which typically reconstruct complete shapes from partial inputs. These methods generally fall into two categories: supervised and unsupervised learning, depending on whether paired partial-complete samples are used during training. Deep models [10, 28, 29] usually adopt multi-layer architectures to map inputs into geometric representations for completion. Compared to traditional techniques, they require fewer prior assumptions and often achieve superior performance [33].

However, most point cloud completion networks employ feature extraction techniques from classical architectures such as PointNet++ [17] and PCN [27], which mainly rely on point coordinates for feature encoding, often neglecting important local geometric structures. To address this, some studies have explored voxelization-based approaches to extract local features. However, as the resolution increases, voxel-based methods incur exponential computational costs, significantly reducing efficiency. Therefore, efficiently capturing fine-grained local geometry remains a major challenge in large-scale and complex environments. Moreover, the distribution of point cloud features often varies significantly across different domains, further complicating the completion process. This domain discrepancy limits the transferability and generalization ability of learned features. Therefore, achieving effective cross-domain feature alignment while enhancing local geometric representation is key to improving the quality of point cloud completion.

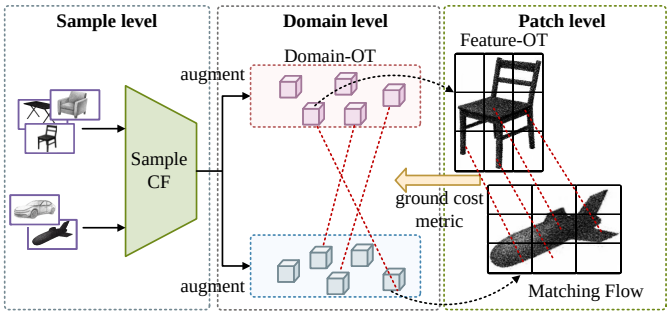


Figure 1: The workflow of the proposed HOTN. The airplane's nose, wings, and tail receive higher matching flows, while background regions are ignored. At the domain level, Domain-OT captures shared structures across categories for cross-domain alignment. At the sample level, Sample CF aligns point cloud samples to enhance consistency and support geometric structure recovery in point cloud completion.

To address these challenges, we propose CAM-FGP, a **C**ross-domain **A**adaptation **M**ethod with **F**ine **G**eometric **P**erception, aimed at resolving domain discrepancies and missing local structures in point cloud completion. CAM-FGP adopts a collaborative optimization strategy to enhance structural integrity and detail restoration through domain alignment and hierarchical geometric modeling. We first introduce the Fine Geometry Detail Extraction Network (FGDE) to extract high-resolution local geometric features from partial point clouds while incorporating auxiliary low-resolution global context to enhance the structure of missing regions. As shown in Figure 1, to mitigate distribution shifts across sources, we employ the HOTN framework, which aligns multi-source point cloud distributions using hierarchical optimal transport and improves feature transferability through local geometric alignment. Finally, CAM-FGP employs a multi-stage hidden state completion fusion strategy to improve the expressiveness of hidden states while reducing memory bottlenecks. This strategy effectively combines coarse structures with fine geometric details, achieving high-fidelity point cloud reconstruction. To the best of our knowledge, CAM-FGP is the first model to align multi-source point cloud distributions using hierarchical optimal transport from the perspective of domain alignment.

The contributions can be summarized as follows:

- We propose a 3D point cloud completion method for cross-domain fine geometric perception, performing detailed geometric modeling and cross-domain alignment on local point clouds to improve detail restoration accuracy.
- We introduce the HOTN framework for aligning multi-source point cloud distributions and enhance the discriminative and transfer capabilities of embedded features through local fragment geometric alignment.
- Our CAM-FGP achieves state-of-the-art performance on three common datasets for point cloud completion.

2 RELATED WORK

2.1 Optimal Transport Theory

Optimal Transport (OT) theory [12] provides a powerful tool for comparing probability distributions by minimizing transport costs, which is crucial for tasks like point cloud completion. Point cloud completion involves filling in sparse data and restoring the original shape, and OT theory can help compare sparse and dense point clouds to enable effective completion. While OT theory is well-developed, applying it to large-scale point cloud problems remains challenging. The Sinkhorn algorithm [7] has accelerated OT distance computation, making it more efficient for point cloud completion. This paper uses the Sliced Wasserstein Distance (SWD) [3] to calculate Feature-OT, reducing the computational complexity of double-layer OT operations for large-scale point cloud processing.

We also utilize the "geometry" concept from OT theory. In point cloud completion, when the cost function is based on the geometric structure of the point cloud, optimal coupling captures this geometry [21]. This geometry depends on key properties of the implicit "ground" space that defines the point cloud distribution. Meier et al. [16] showed that geometric properties remain consistent across cost functions. Studies suggest that these geometric structures can be effectively applied to point cloud completion. Liu et al. [15] demonstrated that OT theory generates meaningful distances by exploiting the geometry of the underlying space, unlike traditional Euclidean distance, which ignores this geometry.

2.2 Point Cloud Completion

In previous studies, various decoders have been developed for point cloud completion. L-GAN [1] proposed a fully connected layer decoder to directly predict point cloud coordinates. PCN introduced the concept of sparse-to-dense point cloud completion, where sparse point clouds are first generated to depict the general shape, and then refined into a smooth, dense point cloud using a folding-based method. CRN [22] also adopted this approach and introduced a refinement process that progressively generates point clouds through upsampling and folding. PF-Net proposed a pyramid decoder inspired by FPN [14], which improves accuracy by supervising key points at different resolutions. VEPCN developed a multiscale voxel-based network to generate detailed features and integrated object structure information into shape completion through edge generation. LAKe-Net [19] introduced a topology-aware model by localizing aligned key points using a Keypoints-Skeleton-Shape method. LSLS [4] achieves shape encoding by learning a structured latent space. These networks first generate sparse point clouds to represent the shape and then produce dense point clouds based on this sparse representation. However, the accuracy of the sparse point clouds may limit the effectiveness of the final output. To address this issue, this paper proposes collaborative optimization of cross-domain alignment and fine geometric modeling for point cloud completion.

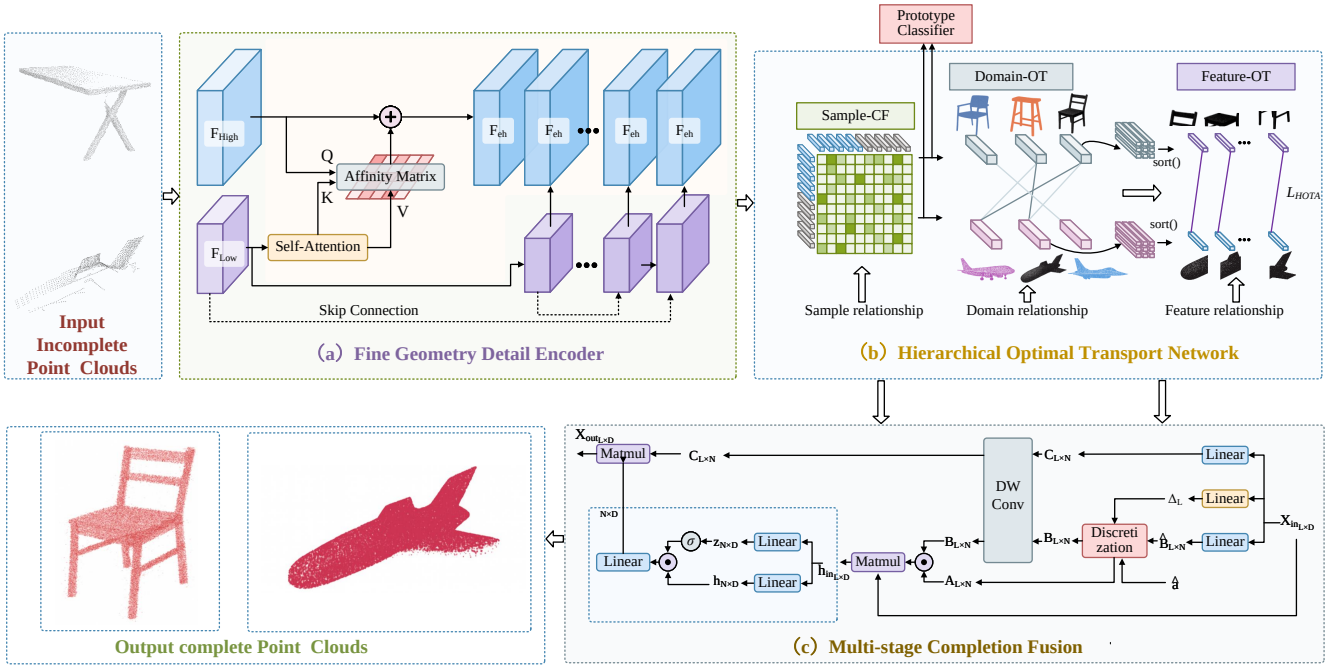


Figure 2: Overview of the proposed model. The input incomplete point clouds are first processed by a fine geometry detail extraction network with self-attention and affinity matrix to capture detailed structures while maintaining multi-scale consistency. A hierarchical optimal transport network then learns sample, domain, and feature relationships to enhance feature alignment. Finally, a multi-stage completion fusion module refines the point cloud progressively, enabling accurate and structure-preserving generation of the output complete point clouds.

3 METHODOLOGY

The proposed CAM-FGP follows a coarse-to-fine approach for point cloud completion. It utilizes a Fine Geometry Detail Extraction Network (FGDE) to capture high-resolution local features, a Hierarchical Optimal Transport Network for cross-domain alignment, and a multi-stage fusion strategy to enhance structural integrity and geometric restoration.

As shown in Figure 2, given a partial point cloud, local features are first extracted using the Fine Geometry Detail Extraction module to enhance the structural representation of missing regions. Then, the Hierarchical Optimal Transport Network aligns the extracted features across domains, ensuring consistency among multi-source data. Finally, a multi-stage fusion module integrates both local and global features to generate a high-fidelity completed point cloud.

In the following, we provide a detailed description of the Fine Geometry Detail Extraction module, the Hierarchical Optimal Transport Network, and the fusion module.

3.1 Fine Geometry Detail Extraction Network

Sparse or incomplete point clouds often display a uniformly distributed spatial pattern, particularly in regions with subtle geometric variations. This uniformity poses significant challenges for point cloud completion models, as it hampers the accurate reconstruction of fine structural details. The lack of such information frequently results in overly smoothed or blurred outputs, thereby diminishing the fidelity of geometric features and reducing overall completion accuracy.

To address this, we propose the Fine Geometry Detail Extraction Network, a module designed to preserve the geometric cues in subtle structural changes of dense point cloud regions. FGDE focuses on capturing high-resolution local features, such as curvature, boundary sharpness, and surface continuity, which are essential for completing missing regions with complex structures.

However, relying solely on local geometric information is insufficient for fully restoring object structures, particularly when global semantic understanding is necessary. Thus, FGDE incorporates Deep Contextual Point Matching Module(DCPM) to provide richer semantic guidance.

The process can be formally expressed as:

$$F_{\text{el}} = \text{SA}(F_{\text{low}}), \tag{1}$$

Where SA(·) represents the self-attention transformation, F_{low} represents the low-resolution contextual point cloud features, and F_{el} represents the enhanced contextual geometric features.

Through this mechanism, DCPM enhances the ability to model local geometric details, which helps to improve the quality of point cloud completion in regions where fine details are missing.

Next, we introduce the Cross-Attention Module to preserve the details in high-resolution local point cloud features to the greatest extent, while integrating supplementary information from low-resolution contextual features. In this module, high-resolution local

point cloud features serve as the query, and the enhanced contextual geometric features serve as the key and value, enabling fine fusion of cross-scale features. This process can be formally expressed as:

$$F_{\text{eh}} = \text{CA}(F_{\text{high}}, F_{\text{el}}, F_{\text{el}}). \tag{2}$$

Where CA represents the cross-attention operation, F_{high} represents the high-resolution local point cloud features, F_{el} represents the enhanced contextual geometric features, and F_{eh} represents the fused output detail-aware features.

It is important to note that in FGDE, we use the DCPM multiple times at different scales to aggregate high-resolution local geometric features with low-resolution contextual information, fully extracting and preserving the detailed geometric structures during the point cloud completion process.

3.2 Hierarchical Optimal Transport Network

Although FGDE captures details well, it struggles with point cloud alignment across shapes and distributions. To address this, we introduce the Hierarchical Optimal Transport Network, utilizing Domain-OT and Feature-OT to learn transferable, discriminative representations by minimizing hierarchical optimal transport distances. This alignment helps handle structural differences between source and target domains, facilitating effective geometric completion and transfer.

Let μ_s and μ_t represent the discrete probability distributions of the source and target domains, respectively. These are defined as:

$$\mu_s = \sum_{i=1}^{I} \frac{1}{I}\delta(x_i^s), \quad \mu_t = \sum_{j=1}^{J} \frac{1}{J}\delta(x_j^t), \tag{3}$$

Where $\delta(x_i^s)$ and $\delta(x_j^t)$ are Dirac delta functions supported on the source domain sample $x_i^s \in x^s$ and the target domain sample $x_j^t \in x^t$.

The domain-level optimal transport distance is given by:

$$\mathcal{D}_{\text{domain}}(x_s, x_t) = \min_{\gamma \in \Pi(\mu_s, \mu_t)} \langle \gamma, C_d \rangle_F \tag{4}$$

$$= \min_{\gamma \in \Pi(\mu_s, \mu_t)} \sum_{i=1}^{I}\sum_{j=1}^{J} \gamma(x_i^s, x_j^t)\, \mathcal{D}_{\text{geo}}(x_i^s, x_j^t), \tag{5}$$

where $\langle \cdot, \cdot \rangle_F$ denotes the Frobenius inner product, and $C_d \in (\mathbb{R}^+)^{I \times J}$ is a non-negative cost matrix. The element $C_d(x_i^s, x_j^t)$ represents the geometric structural distance between the source sample x_i^s and the target sample x_j^t. $\Pi(\mu_s, \mu_t)$ represents the joint probability distribution space (coupling set), which consists of all joint distributions constructed from the empirical distributions μ_s and μ_t, defined as:

$$\Pi(\mu_s, \mu_t) = \left\{ \gamma \in (\mathbb{R}^+)^{\text{I} \times \text{J}} \mid \gamma \mathbf{1}_{\text{J}} = \mu_s, \gamma^{\text{T}} \mathbf{1}_{\text{I}} = \mu_t \right\}, \tag{6}$$

where $\mathbf{1}_d$ denotes the d-dimensional vector of ones. Using Eq. 4, we can align the source and target point cloud fragments at the domain level, facilitating the transfer of fine details and geometric completion across shape variations in the scene.

Given two point cloud sub-block samples $x_i^s \in x^s$ and $x_i^t \in x^t$ from the source and target domains, we input them into a shared feature extractor to obtain their local feature representations $h_i^s, h_i^t \in \mathbb{R}^{H \times W \times C}$. Here, $H \times W$ denotes the spatial dimensions, and C is the number of channels. Thus, h_i^s and h_i^t can be viewed as collections of local geometric sub-feature patches, each represented as a C-dimensional feature vector:

$$\begin{cases} h_i^s = \{u_i^1, \ldots, u_i^n, \ldots, u_i^{HW}\}, & u_i^n \in \mathbb{R}^C \\ h_i^t = \{v_j^1, \ldots, v_j^m, \ldots, v_j^{HW}\}, & v_j^m \in \mathbb{R}^C \end{cases} \tag{7}$$

In the subsequent steps, we will introduce an optimal transport mechanism at the patch level to measure the fine-grained correspondences between the two sets of local sub-structures, thereby enabling more structure-aware point cloud alignment and completion.

Similarly, for the feature maps h_i^s and h_j^t extracted from the source and target domain samples, we define their corresponding discrete probability distributions μ_s^i and μ_t^j as:

$$\mu_s^i = \sum_{n=1}^{HW} \frac{1}{HW}\delta(u_i^n), \quad \mu_t^j = \sum_{m=1}^{HW} \frac{1}{HW}\delta(v_j^m), \tag{8}$$

where $\delta(u_i^n)$ and $\delta(v_j^m)$ are Dirac distributions supported on the source domain feature patch u_i^n and the target domain patch v_j^m, respectively.

Based on this, we define the feature-level optimal transport distance between the sample pair x_i^s and x_j^t as:

$$D_{\text{geo}}(x_i^s, x_j^t) = \min_{\gamma \in \Pi(\mu_s^i, \mu_t^j)} \langle \gamma, C_f \rangle_F \tag{9}$$

$$= \min_{\gamma \in \Pi(\mu_s^i, \mu_t^j)} \sum_{n=1}^{HW}\sum_{m=1}^{HW} \gamma(u_i^n, v_j^m) \|u_i^n - v_j^m\|. \tag{10}$$

where $C_f(u_i^n, v_j^m)$ represents the Euclidean distance between two local point cloud geometric fragments, and $\Pi(\mu_s^i, \mu_t^j)$ denotes the set of joint probability distributions defined in the local feature space.

Feature-level Optimal Transport (Feature-OT) enhances local structure alignment by focusing transport mass on correlated regions, improving local detail fidelity. To address the computational cost of Hierarchical OT, we use Sliced Wasserstein Distance (SWD), which approximates D_{geo} efficiently. SWD slices high-dimensional features with random projections, calculates Euclidean distances between sorted projections, and averages across directions for a robust approximation of Feature-OT.

$$D_{\text{geo}}^{\text{swd}}(x_i^s, x_j^t) = \frac{1}{M}\sum_{m=1}^{M} \left\| \text{sort}(\theta_m^T h_i^s) - \text{sort}(\theta_m^T h_j^t) \right\|_2, \tag{11}$$

where $\{\theta_m\}_{m=1}^M$ represents M random projection vectors that map the high-dimensional embeddings to 1D, and $\text{sort}(\cdot)$ indicates sorting the one-dimensional vectors in ascending (or descending) order. Finally, the Euclidean distance between the sorted vectors is computed, and the results across all projection directions are averaged, providing an approximation of the feature OT distance.

3.3 Multi-stage Completion Fusion

Considering that the features extracted by FGDE at different stages of the completion process have varying abstraction levels, we introduce a Multi-Stage Hidden-State Fusion (MSF) strategy to aggregate geometric representations from multiple stages for unified modeling. Let $\{\mathbf{h}^{(s)}\}_{s=1}^{S}$ denote the hidden features extracted at the end of each stage s.

For each hidden state $\mathbf{h}^{(s)}$, we calculate its corresponding global geometric representation $\hat{\mathbf{h}}^{(s)}$ by applying average pooling over all point-level features:

$$\hat{\mathbf{h}}^{(s)} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{h}_i^{(s)}, \tag{12}$$

Subsequently, each global representation $\hat{\mathbf{h}}^{(s)} \in \mathbb{R}^D$ is normalized and projected into the predicted shape representation $\mathbf{z}^{(s)} \in \mathbb{R}^c$. The final output representation $\mathbf{z}$ is obtained by weighted fusion of the predictions from all stages, including the raw prediction $\mathbf{z}^{(0)}$ from the final stage, and is specifically defined as:

$$\mathbf{z} = \sum_{s=0}^{S}\hat{\beta}^{(s)}\mathbf{z}^{(s)}, \tag{13}$$

$$\hat{\beta}^{(s)} = \frac{\exp(\beta^{(s)})}{\sum_{i=0}^{S}\exp(\beta^{(i)})}. \tag{14}$$

Here, $\beta^{(s)}$ is a learnable scalar that represents the relative importance of the s-th stage. This fusion strategy explicitly guides the hidden states from different stages to jointly participate in the point cloud reconstruction process, effectively integrating both coarse-grained global structural information and fine-grained local geometric details, thereby improving the fidelity and generalization ability of the completion results.

In our preliminary experiments, we observed that traditional multi-head attention structures create efficiency bottlenecks in point cloud completion, particularly due to memory access during tensor reshaping in high-resolution scenes.

To address this, we eliminate redundant multi-head configurations in the Fine Geometry Detail Extraction Network and introduce a lightweight state-level importance modeling mechanism. This reduces computational cost while maintaining expressiveness. We construct a state weight matrix $\Delta \in \mathbb{R}^{L\times N}$ and a saliency vector $\hat{a} \in \mathbb{R}^N$, generating an attention matrix $\mathbf{A} \in \mathbb{R}^{L\times N}$ to estimate the importance of each local region (token) across different geometric states.

The final form of the fused input representation is as follows:

$$\mathbf{h}_{\text{in}} = (\mathbf{A}\odot\mathbf{B})^{T}_{\mathbf{X}_{\text{in}}}. \tag{15}$$

where $\odot$ denotes element-wise multiplication, $\mathbf{X}_{\text{in}}$ represents the original local features of the point cloud, and $\mathbf{B}$ is the shared transformation matrix. This design not only effectively simulates the multi-head mechanism's ability to model diverse geometric relationships, but also significantly reduces the computational overhead caused by memory operations. In practical point cloud completion tasks, this mechanism achieves higher inference throughput while maintaining completion accuracy and substantially improving the geometric recovery quality in detailed regions.

4 EXPERIMENTS

4.1 Datasets and Metrics

The comparison is performed on three datasets: ShapeNet-55/34 (Yuan et al., 2018) and PCN (Yuan et al., 2018).

ShapeNet-55 comprises synthetic 3D objects from 55 distinct categories, each represented by a uniformly sampled point cloud containing 8,192 points. Following the protocol in [26], we decompose each complete point cloud into two subsets: a 2,048-point partial input and a 6,144-point ground-truth completion target. To introduce diversity in partial observations, we randomly select a camera viewpoint and remove the N farthest points from that view. The remaining points are then downsampled to 2,048 to serve as partial inputs. This strategy ensures that the missing regions vary across training epochs, thereby improving the model's robustness to diverse occlusion patterns. During evaluation, we employ eight fixed viewpoints and set N to 2,048, 4,096, and 6,144 points, corresponding to *easy*, *medium*, and *hard* difficulty levels, respectively. These settings simulate partial observations with 25%, 50%, and 75% missing data, providing a graded assessment of model performance under varying degrees of incompleteness.

ShapeNet-34 serves as an unseen-category benchmark for testing generalization. It consists of 3D objects from 34 categories, each containing 8,192 points. The testing process mirrors the ShapeNet-55 protocol, ensuring consistency in viewpoint selection, occlusion generation, and difficulty settings. This cross-category evaluation enables us to rigorously assess the model's ability to generalize to novel object types that were not present during training.

PCN dataset [27] contains 30,714 high-resolution 3D objects spanning 8 categories, with each complete shape uniformly sampled to 16,384 points. Partial point clouds are obtained by back-projecting 2.5D depth maps from synthetic viewpoints into 3D space, resulting in partial observations containing exactly 2,048 points. This benchmark poses significant challenges due to its higher resolution and larger scale, which demand more precise geometric reasoning from the model. For fair comparison, we adopt the same experimental configuration, preprocessing steps, and train/test splits as in prior work.

We evaluate our method using four complementary metrics to capture both overall reconstruction quality and local geometric fidelity. Specifically, the L2 Chamfer Distance (CD) measures the average squared bidirectional distance between predicted and ground-truth point sets, reflecting coverage and alignment accuracy. The F-Score computes the harmonic mean of precision and recall under a distance threshold, jointly assessing accuracy and completeness. The Earth Mover's Distance (EMD) estimates the minimal cost required to transform the predicted point distribution into the ground truth, providing a fine-grained measure of geometric similarity. Finally, the Unidirectional Hausdorff Distance (UHD) calculates the maximum distance from any predicted point to the nearest ground-truth point, serving as a stringent indicator of completion fidelity. Together, these metrics offer a comprehensive evaluation of the model's performance from both global and local perspectives.

Method	Table	Chair	Plane	Car	Sofa	CD-S	CD-M	CD-H	CD-Avg	F1
FoldingNet	2.59	2.74	1.48	1.95	2.47	2.70	2.66	4.12	3.15	0.087
PCN	2.12	2.25	1.07	1.83	2.07	1.95	1.94	4.04	2.64	0.137
TopNet	2.22	2.53	1.14	2.18	2.36	2.26	2.16	4.31	2.92	0.129
PFNet	3.94	4.17	1.81	2.55	3.34	3.83	3.86	7.93	5.20	0.335
GRNet	1.68	1.88	1.06	1.64	1.76	1.38	1.71	2.85	1.98	0.243
PointAttN	0.75	0.83	0.43	0.89	0.71	0.52	0.77	1.52	0.93	0.484
AnchorFormer	0.85	0.97	0.45	0.91	0.81	0.60	0.88	1.81	1.10	0.475
CAM-FGP	**0.70**	**0.79**	**0.39**	**0.81**	**0.65**	**0.44**	**0.71**	**1.41**	**0.89**	**0.491**

Table 1: Quantitative results on ShapeNet-55 dataset in terms of L2 Chamfer Distance ×1000 (lower is better) and F-Score@1%.

	34 seen categories					21 unseen categories				
	CD-S	CD-M	CD-H	CD-Avg	F1	CD-S	CD-M	CD-H	CD-Avg	F1
FoldingNet	1.88	1.72	3.35	2.32	0.150	2.62	2.59	5.12	3.48	0.102
PCN	1.85	1.79	2.93	2.25	0.165	3.12	3.06	5.19	3.83	0.110
TopNet	1.75	1.59	3.48	2.30	0.180	2.55	2.43	5.30	3.47	0.130
PFNet	3.10	3.20	7.75	4.68	0.360	5.23	5.85	13.40	8.15	0.325
GRNet	1.30	1.45	2.55	1.77	0.270	1.90	2.30	4.95	3.05	0.225
PointAtt	0.53	0.75	1.40	0.88	0.465	0.64	1.10	2.45	1.38	0.420
AnchorFormer	0.81	1.10	1.90	1.27	0.440	1.12	1.75	3.50	2.10	0.405
CAM-FGP	**0.49**	**0.69**	**1.30**	**0.81**	**0.560**	**0.60**	**1.03**	**2.33**	**1.32**	**0.540**

Table 2: Quantitative comparison of CAM-FGP and state-of-the-art methods on ShapeNet-34 in terms of L2 Chamfer Distance ×1000 and F-Score@1% (higher is better).

Methods	CD-Avg	Plane	Chair	Table
FoldingNet	14.31	9.49	15.65	13.75
TopNet	12.15	7.61	13.44	11.22
AtlasNet	10.85	6.37	12.09	11.01
PCN	9.64	5.20	8.85	11.54
GRNet	8.83	6.45	9.47	8.55
PointAttN	6.74	3.91	6.74	5.85
AnchorFormer	8.38	4.85	9.50	7.90
CAM-FGP	**6.70**	**3.80**	**3.69**	**5.71**

Table 3: Quantitative results on PCN benchmark in terms of per-point L1 Chamfer Distance ×1000 (lower is better).

Methods	Avg.	Plane	Chair	Table
W/o FGDE	7.82	4.93	6.41	7.13
W/o DCPM	7.58	4.89	6.29	7.32
W/o HOTN	7.45	4.77	6.15	7.28
W/o Multi-Stage Fusion	7.60	4.84	6.38	7.10
CAM-FGP	**6.70**	**3.80**	**3.69**	**5.71**

Table 4: Comparison of methods with and without specific components

4.2 Baselines

We compare our method with several advanced methods, including FoldingNet [25], PCN [27], TopNet [20], PFNet [11], GRNet [24], SeedFormer [31], AnchorFormer [5], PointAttN [23].

4.3 Comparisons with State-of-the-Art Methods

4.3.1 Results on ShapeNet-55. The ShapeNet-55 dataset, which contains 55 categories, has become a standard benchmark for evaluating point cloud completion methods. As shown in Table 1, CAM-FGP outperforms all other methods across various metrics. Specifically, it achieves the lowest average Chamfer Distance (CD-Avg) of 0.89, improving by 0.04 compared to SeedFormer and 0.21 compared to PoinTr. CAM-FGP also excels in the CD-S, CD-M, and CD-H categories, showing improvements across all shape categories. Additionally, CAM-FGP achieves an F-Score of 0.491, the highest among the methods listed, surpassing PointAttN's 0.484 and AnchorFormer's 0.475. These results demonstrate the superior performance of CAM-FGP in both geometric accuracy and shape completion.

Figure 3 shows visual comparisons on ShapeNet-55. The leftmost column shows partial inputs. While methods like PointAttN and AnchorFormer recover general structures, our CAM-FGP better preserves fine details (e.g., table legs, chair backrests) and geometric consistency. Notably, on challenging cases like sofa and car, CAM-FGP recovers clearer contours and more complete geometry, producing results closer to the ground truth in fidelity and accuracy.

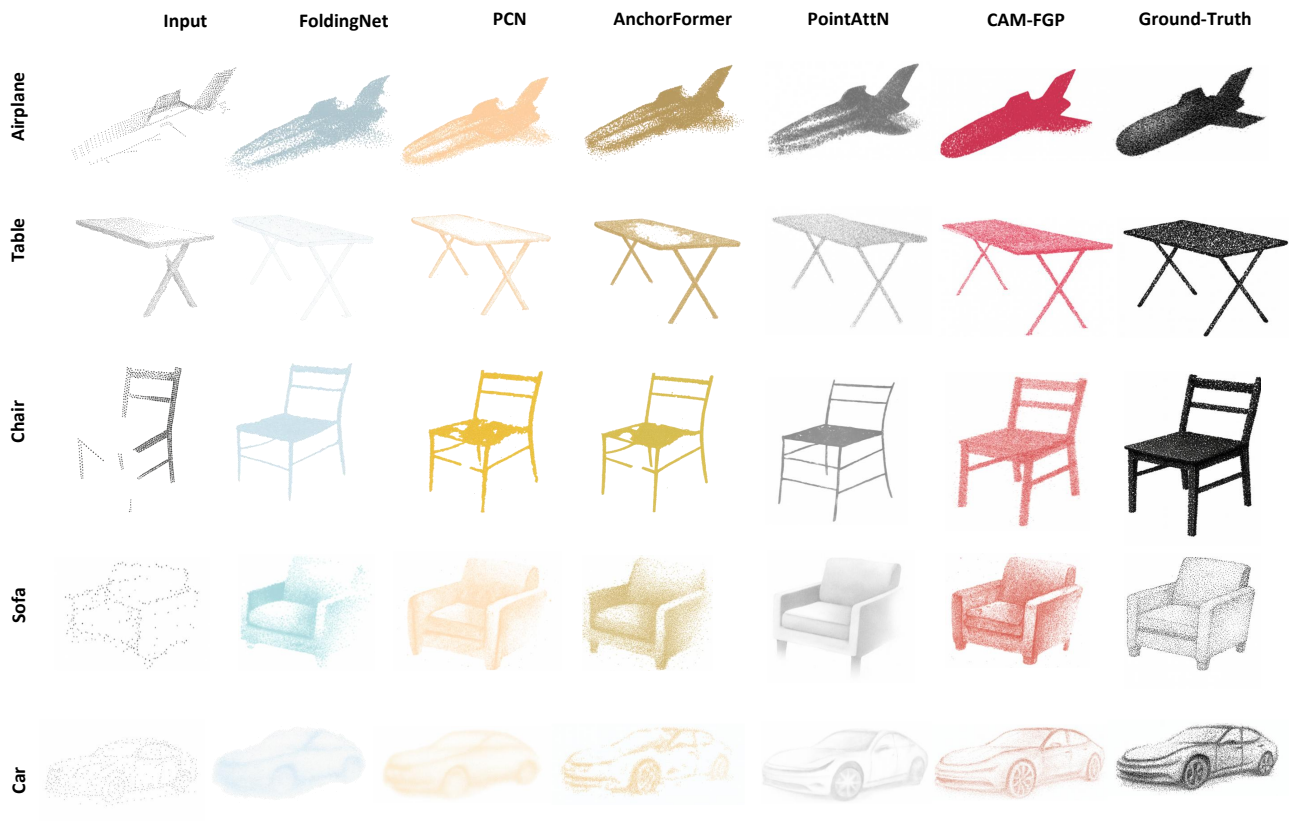


Figure 3: Visual examples of point cloud completion results on the ShapeNet-55 dataset using different methods.

4.3.2 Results on ShapeNet-34. We conduct extensive experiments on the ShapeNet-34 dataset, evaluating both seen and unseen categories to assess the generalization capability of CAM-FGP. As shown in Table 2, CAM-FGP achieves the best performance on both the CD and F1 metrics across the 34 seen categories and 21 unseen categories. Despite its close F1 score compared to other methods, CAM-FGP outperforms in CD metrics, especially with the smallest CD-S and CD-M values. This highlights CAM-FGP's superior generalization ability and robustness in point cloud completion tasks, demonstrating its effectiveness in handling diverse data scenarios.

4.3.3 Results on PCN. The PCN benchmark dataset contains a total of 28,974 shapes for training and 1,200 shapes for testing, distributed across multiple categories. It is widely regarded as one of the most frequently used datasets for evaluating the performance of point cloud completion methods. As shown in Table 3, CAM-FGP outperforms all other methods across various metrics. Specifically, CAM-FGP achieves an average CD-Avg of 6.70, improving by 0.04 compared to PointAttN, and 1.68 compared to AnchorFormer. The method also excels in the individual categories, including Plane, Chair, and Table, showing improvements across all these shapes. CAM-FGP demonstrates superior performance in both the overall per-point Chamfer Distance and category-specific results, solidifying its position as a top-performing method for point cloud completion.

Figure 4: Visualized comparison of removing each module in the proposed completion pipeline. From left to right: removing FGDE, removing DCPM, removing HOTN, and removing Multi-Stage Fusion.

4.4 Ablation Study

CAM-FGP conducts four ablation experiments on a representative subset of the PCN dataset, as summarized in Table 4. These experiments systematically evaluate the contribution of each core module. The removal of the Fine Geometry Detail Extraction Network results in the loss of critical fine-scale geometric features, leading to noticeable inaccuracies in reconstructing intricate structural details. Excluding the Deep Contextual Point Matching Module significantly impairs the model's ability to establish contextual correspondences between points, reducing completion accuracy in regions where contextual understanding is essential. The absence of the Higher-Order Temporal Attention mechanism weakens the model's ability to capture temporal dependencies across frames

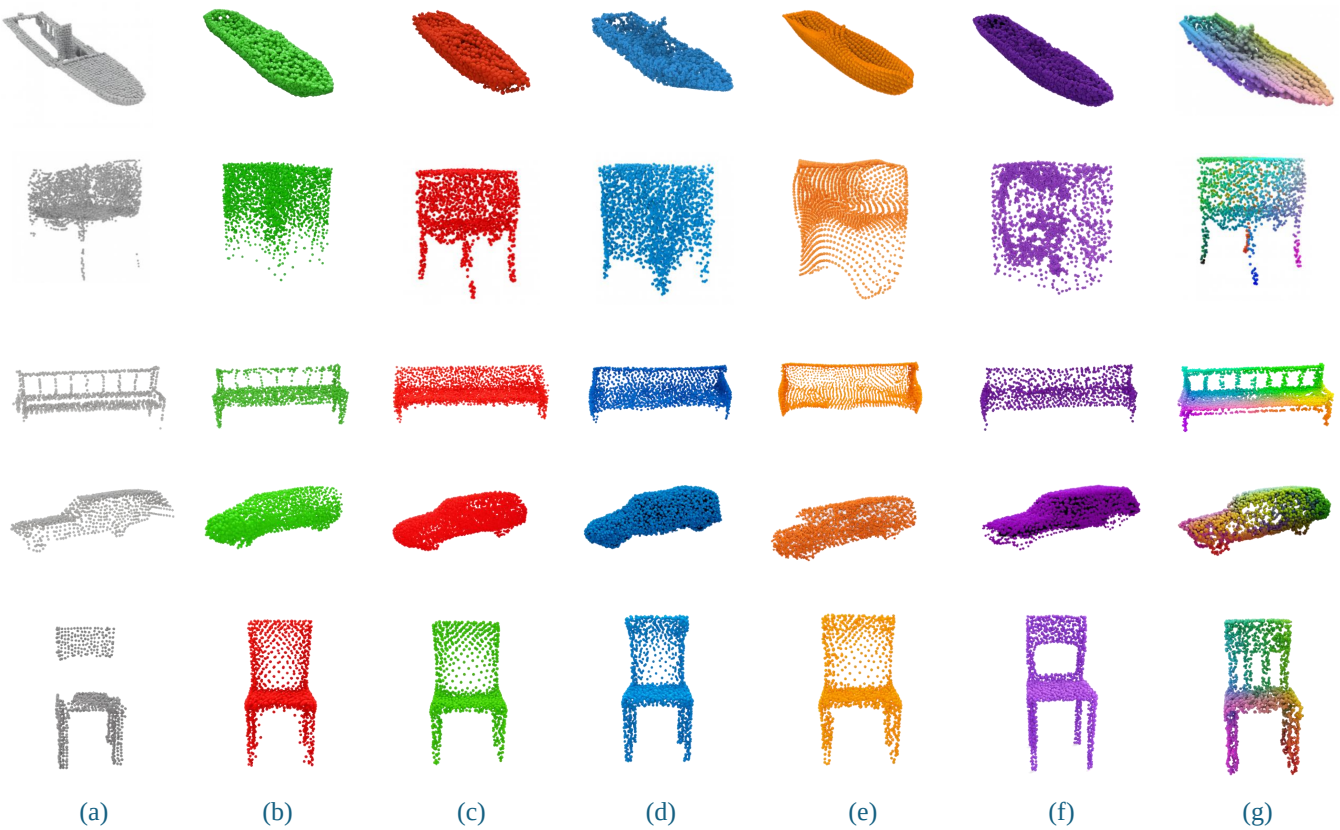


Figure 5: Comparison of the visual results by different methods on the synthetic dataset. Note that we change the background color of the view images in our dataset from black to white for better visualization. (a) Partial points. (b) FoldingNet. (c) PCN. (d) TopNet. (e) PFNet. (f) PointAttN. (g) Ours.

or partial observations, which adversely affects both structural consistency and reconstruction fidelity. Furthermore, eliminating the multi-stage fusion mechanism compromises the integration of multi-level features throughout the pipeline, leading to a marked decline in point cloud completion quality. When all modules are retained, CAM-FGP achieves optimal performance, highlighting the indispensable role each component plays in enhancing the model's robustness, precision, and generalization. Corresponding visualizations are provided in Fig. 4, further demonstrating the impacts of omitting each module and reinforcing their collective importance.

4.4.1 Results on Synthetic Data. Figure 5 compares our method with state-of-the-art approaches on the synthetic dataset. Our method effectively completes missing geometry while preserving the original shape, unlike the baselines, which distort the input, leading to significant deviations from both the partial observation and the ground truth.

A key strength of our method is its ability to retain fine details, such as thin legs, rails, and holes. For example, in the chair and bench categories, existing methods fill hollow backrest regions, while our method accurately reconstructs them. Similarly, for boats and cars, our method maintains structural continuity, producing complete and consistent shapes.

Our method also demonstrates strong generalization across different object categories, generating plausible and detailed completions, unlike existing methods that often produce distorted or oversmoothed outputs, especially in complex or occluded regions. These results highlight our method's superior shape awareness and structural fidelity, achieved through local geometry preservation and global shape priors.

5 CONCLUSION

CAM-FGP performs alignment and geometric modeling to enhance structural integrity and recover fine details. It uses a Fine Geometry Detail Extraction Network to capture local features from visible point clouds and preserve details in incomplete areas. A Hierarchical Optimal Transport Network aligns multi-source point clouds, enhancing feature transfer. To integrate local and global features, CAM-FGP adopts a multi-stage hidden state fusion strategy, maintaining continuity and reducing overhead. Experiments and ablation studies on benchmark datasets confirm the method's superior performance in structural accuracy and detail recovery.

REFERENCES

[1] P. Achlioptas, O. Diamanti, I. Mitliagkas, and L. Guibas. 2018. Learning representations and generative models for 3D point clouds. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, Vol. 80. 40–49.

[2] Pravallika Ambati, Mohammad Farukh Hashmi, and Aditya Gupta. 2024. Deep Learning Frontiers in 3D Object Detection: A Comprehensive Review for Autonomous Driving. *IEEE Access* (2024).

[3] N. Bonneel, J. Rabin, G. Peyré, and H. Pfister. 2015. Sliced and Radon Wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision* 51, 1 (2015), 22–45.

[4] Y. Cai, K.-Y. Lin, C. Zhang, Q. Wang, X. Wang, and H. Li. 2022. Learning a structured latent space for unsupervised point cloud completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 5533–5543.

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

[6] A. Chitta. 2024. Replacing Objects in Point Cloud Stream with Real-time Meshes using Semantic Segmentation. *J* (2024).

[7] M. Cuturi. 2013. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 26.

[8] Manoj Kumar Goshisht. 2024. Machine learning and deep learning in synthetic biology: Key architectures, applications, and challenges. *ACS Omega* 9, 9 (2024), 9921–9945.

[9] Yun-Chih Guo, Tzu-Hsuan Weng, Robin Fischer, and LiChen Fu. 2022. 3D semantic segmentation based on spatial-aware convolution and shape completion for augmented reality applications. *Computer Vision and Image Understanding* 224 (2022), 103550.

[10] A. Heidari, N. Jafari Navimipour, H. Dag, and M. Unal. 2024. Deepfake detection using deep learning methods: A systematic and comprehensive review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 14, 2 (2024), e1520.

[11] Z. Huang, Y. Yu, J. Xu, F. Ni, and X. Le. 2020. PF-Net: Point Fractal Network for 3D Point Cloud Completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Computer Vision Foundation / IEEE, 7659–7667.

[12] L. V. Kantorovich. 2006. On the translocation of masses. *Journal of Mathematical Sciences* 133, 4 (2006), 1381–1382.

[13] S. Kato, S. Tokunaga, Y. Maruyama, S. Maeda, M. Hirabayashi, Y. Kitsukawa, A. Monrroy, T. Ando, Y. Fujii, and T. Azumi. 2018. Autoware on board: Enabling autonomous vehicles with embedded systems. In *2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS)*. IEEE, 287–296.

[14] T. Y. Lin, P. Dollàr, R. Girshick, K. He, B. Hariharan, and S. Belongie. 2017. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2117–2125.

[15] F. Liu, M. Kim, Z. Ren, et al. 2024. Distilling CLIP with dual guidance for learning discriminative human body shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 256–266.

[16] Timon Meier et al. 2024. Obtaining auxetic and isotropic metamaterials in counterintuitive design spaces: an automated optimization approach and experimental characterization. *npj Computational Materials* 10, 1 (2024), 3.

[17] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. 2017. PointNet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems* 30 (2017).

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

[19] J. Tang, Z. Gong, R. Yi, Y. Xie, and L. Ma. 2022. LAKe-Net: Topology-aware point cloud completion by localizing aligned keypoints. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 1716–1725.

[20] L. P. Tchapmi, V. Kosaraju, H. Rezatofighi, I. D. Reid, and S. Savarese. 2019. TopNet: Structural point cloud decoder. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Computer Vision Foundation / IEEE, 383–392.

[21] C. Villani et al. 2009. *Optimal Transport: Old and New*. Vol. 338. Springer.

[22] X. Wang, M. H. Ang, and G. H. Lee. 2020. Cascaded refinement network for point cloud completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 787–796.

[23] Y. Wang, Y. Cui, D. Guo, J. Li, Q. Liu, and C. Shen. 2024. POINTATTN: You Only Need Attention for Point Cloud Completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 1–12.

[24] H. Xie, H. Yao, S. Zhou, J. Mao, S. Zhang, and W. Sun. 2020. GRNet: Gridding residual network for dense point cloud completion. In *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, 365–381.

[25] Y. Yang, C. Feng, Y. Shen, and D. Tian. 2018. FoldingNet: Point cloud auto-encoder via deep grid deformation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 206–215.

[26] X. Yu, Y. Rao, Z. Wang, Z. Liu, J. Lu, and J. Zhou. 2021. PoinTr: Diverse point cloud completion with geometry-aware transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 12478–12487.

[27] W. Yuan, T. Khot, D. Held, C. Mertz, and M. Hebert. 2018. PCN: Point completion network. In *Proceedings of the 2018 International Conference on 3D Vision (3DV)*. IEEE, 728–737.

[28] T. Yun, J. Li, L. Ma, J. Zhou, R. Wang, M. P. Eichhorn, and H. Zhang. 2024. Status, advancements and prospects of deep learning methods applied in forest studies. *International Journal of Applied Earth Observation and Geoinformation* 131 (2024), 103938.

[29] M. Zhang, N. Tsoulakos, P. Kujala, and S. Hirdaris. 2024. A deep learning method for the prediction of ship fuel consumption in real operational conditions. *Engineering Applications of Artificial Intelligence* 130 (2024), 107425.

[30] X. Zhang, D. Chu, X. Zhao, et al. 2024. Machine learning-driven 3D printing: a review. *Applied Materials Today* 39 (2024), 102306.

[31] H. Zhou, Y. Cao, W. Chu, J. Zhu, T. Lu, Y. Tai, and C. Wang. 2022. SeedFormer: Patch seeds based point cloud completion with upsample transformer. In *Computer Vision - ECCV (Lecture Notes in Computer Science, Vol. 13663)*. Springer, 416–432.

[32] Zhe Zhu, Honghua Chen, Xing He, Weiming Wang, Jing Qin, and Mingqiang Wei. 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 (ICCV)*. 14508–14518.

[33] Z. Zhuang, Z. Zhi, T. Han, Y. Chen, J. Chen, C. Wang, and L. Ma. 2024. A survey of point cloud completion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17 (2024), 5691–5711.