EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

ARaNet: Attention and Residual Aware Network for Resilient Digital Twins in Rail Transit Equipment Manufacturing

Xi Chen[1], Xiaolong Gao[1], Hanyue Zhan[1,*], Wanting Liu[2]

[1]School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China
[2]School of Yibin Research Institute, Southwest Jiaotong University, Yibin, China

Abstract

INTRODUCTION: In rail transit equipment manufacturing, which encompasses locomotive body welding, metro vehicle assembly, and high-speed rail carriage production, high-fidelity 3D point cloud data acquired by industrial sensors serves as the foundation for digital twin modeling, automated quality inspection, and robotic guidance. However, harsh production environments characterized by metallic dust, welding spatter, mechanical vibration, and frequent occlusions by fixtures and tooling inevitably introduce severe data corruption, including missing regions and non-uniform point density. Such degraded sensor data undermines the reliability of downstream manufacturing processes that depend on accurate 3D representations.
OBJECTIVES: This paper presents an effective 3D sensor data recovery method that reconstructs complete geometric representations from corrupted partial scans, specifically targeting the data integrity challenges encountered in rail transit equipment manufacturing. The proposed approach aims to simultaneously restore global structural completeness and local geometric precision, thereby enabling resilient digital twin systems that maintain operational continuity despite sensor-induced data loss.
METHODS:We propose an Attention and Residual Aware Network (ARaNet) featuring a Multi-Scale Channel-Aware Convolution encoder and a Hierarchical Residual-Aware Decoder. The encoder performs Farthest Point Sampling at multiple resolutions (2048, 1024, 512 points) and applies a channel attention mechanism to dynamically weight feature channels, emphasizing geometrically salient structures such as sharp edges, curved surfaces, and mechanical joints that are critical in rail transit component geometries, including bogie frames, car body shells, and coupler assemblies. The decoder progressively generates point clouds from coarse resolution (64 points) to medium resolution (128 points) to high resolution (2048 points), incorporating a residual refinement module that predicts coordinate offsets to rectify local geometric errors. This capability is particularly valuable for precision metrology in component manufacturing.
RESULTS: Experimental evaluation on the ShapeNet and ModelNet40 datasets demonstrates that ARaNet achieves substantial improvements over benchmark models PCN and PF-Net. Quantitative assessment shows average reductions of 12.9% and 23.8% in the Gt-to-Pre and Pre-to-Gt metrics, respectively. The generated point clouds exhibit more uniform distributions and superior detail restoration, particularly for complex mechanical geometries with curved surfaces and structural joints characteristic of rail transit equipment components.
Conclusion:The effectiveness of integrating channel attention with residual refinement mechanisms for industrial 3D sensor data recovery is validated. ARaNet provides a robust data recovery foundation for resilient digital twin systems in rail transit equipment manufacturing, enabling sustained operational capability and rapid geometric reconstruction even when sensor data is compromised by production environment disruptions. However, it should be noted that ARaNet is currently validated exclusively on synthetic training data, and its performance under extremely high missing-region ratios or in real industrial deployment scenarios without domain adaptation warrants further investigation.

Keywords: 3D sensor data recovery; digital twin resilience; rail transit equipment manufacturing; attention mechanism; residual refinement; point cloud completion

Received on 18 May 2026, accepted on 02 June 2026, published on 30 June 2026

*Corresponding author. Email: zhy2024@my.swjtu.edu.cn

doi: 10.4108/13054

1. Introduction

The manufacturing of rail transit equipment, including high-speed rail carriages, metro vehicles, and locomotives, is undergoing a profound digital transformation driven by Industry 4.0 initiatives [1]. Central to this transformation is the concept of digital twins [2], which create virtual replicas of physical manufacturing assets by continuously integrating real-time 3D sensor data. In modern rail vehicle production lines, 3D laser scanners and LiDAR-based metrology systems are widely deployed to capture detailed geometric profiles of car body shells, bogie frames, coupler assemblies, and welded structural joints for automated quality inspection and robotic guidance [3]. The fidelity of such 3D point cloud data directly determines the reliability of digital twin models, the safety of robotic operations, and consequently, the quality and efficiency of the entire manufacturing process.

However, rail transit manufacturing environments pose severe challenges to 3D data acquisition. Welding workshops for car body assembly generate metallic spatter and fumes that scatter laser beams; bogie machining areas produce intense vibrations from CNC milling and grinding operations; and large-scale carriage interiors suffer from extensive fixture occlusions when scanning complex structural joints [4]. These industrial factors inevitably corrupt the acquired point clouds, resulting in missing regions, non-uniform density, and geometric distortions. In disaster recovery scenarios, where production lines must rapidly resume operations after equipment failures, damaged components must be digitized under even more degraded sensing conditions. When such corrupted data feeds into digital twin systems, it can trigger cascading failures: inaccurate virtual models lead to erroneous robotic path planning, false quality rejections, or undetected dimensional deviations that propagate through subsequent assembly stages [2]. Therefore, the ability to recover complete and accurate 3D geometric information from corrupted sensor data is not merely an academic exercise in shape completion, but a critical enabler of manufacturing resilience. In this work, we operationalize manufacturing resilience as the system's capability to maintain downstream task performance (e.g., robotic path planning accuracy, quality inspection precision) above a predefined threshold despite sensor data corruption exceeding 30% missing-region ratio. Specifically, a resilient digital twin in rail transit manufacturing must satisfy three measurable criteria: (1) recovery fidelity—Chamfer Distance below 0.01 under 30-50% missing data; (2) temporal continuity—reconstruction latency under 250ms for real-time applications; and (3) robustness consistency—performance degradation less than 15% across varying corruption patterns (occlusion, noise, sparsity). These operational definitions anchor our method within concrete manufacturing requirements and provide quantifiable evaluation standards.

Existing deep learning-based point cloud completion methods have demonstrated promising results on synthetic benchmarks. The Point Completion Network (PCN) [5] directly maps partial point clouds to complete ones without relying on shape priors. PF-Net [6] employs multi-resolution encoding to progressively refine shapes. SA-Net [7] introduces attention skip connections to reuse local geometric features during decoding. More recently, transformer-based architectures such as AdaPoinTr [8] and SeedFormer [9] have further advanced detail recovery through self-attention mechanisms. However, these methods typically focus on overall shape reconstruction while neglecting local geometric details. This constitutes a critical shortcoming in rail transit manufacturing, where sub-millimeter precision at bogie mounting interfaces, car body welded seams, and coupler mating surfaces directly affects assembly quality and operational safety. Their completed point clouds often exhibit blurred edges and uneven point distributions, which are unacceptable for precision metrology applications in rail vehicle component manufacturing. Moreover, none of these methods have been explicitly examined from the perspective of manufacturing resilience, leaving a gap between academic point cloud completion research and the practical demands of industrial digital twin systems.

To bridge this gap, this work presents ARaNet (Attention and Residual Aware Network), a recovery framework that goes beyond conventional geometric completion to serve as a foundational data recovery mechanism for resilient manufacturing systems. ARaNet is not merely a shape completion tool; it is designed to endow digital twin pipelines with the ability to sustain accurate 3D representations even when upstream sensor data is compromised by the harsh realities of rail transit equipment production environments. The primary contributions of this paper are as follows:

(1) A Multi-Scale Channel Attention Encoder is designed that extracts features incorporating both global structural contours and local geometric details through farthest-point sampling [10] and dynamic channel weighting [11]. This mechanism is particularly effective for rail transit components that exhibit multi-scale geometric complexity, ranging from large-scale car body curvatures to fine-scale bolt holes on bogie frames and weld seams on structural joints, enabling robust feature extraction even from severely corrupted input data.

(2) A Hierarchical Residual-Aware Decoder is proposed that reconstructs point clouds from coarse to fine while refining local geometry through residual prediction. The

explicit coordinate offset correction is analogous to error compensation in industrial CNC machining [12], providing precision-grade geometric restoration suitable for manufacturing metrology requirements. This residual mechanism is the key to achieving the geometric fidelity that distinguishes industrial-grade data recovery from generic shape completion.

(3) Extensive experiments are conducted on datasets containing categories geometrically representative of rail transit components (e.g., automobiles sharing structural similarity with car body shells, airplanes with aerodynamic geometries analogous to high-speed train heads), demonstrating that ARaNet achieves significant improvements in both global reconstruction accuracy and local detail fidelity compared to existing methods (average reductions of 12.9% and 23.8% in Gt-to-Pre and Pre-to-Gt metrics, respectively), thereby validating its potential as a robust data recovery mechanism for resilient digital twins in rail transit equipment manufacturing.

2. Related Work

2.1. Point Cloud Completion

Early approaches to 3D shape completion were primarily based on voxel grids or multi-view representations. Representative works include the voxel-based 3D-Encoder-Predictor CNNs [13], Octree Generating Networks [14], and voxel-GAN frameworks that combine global structure inference with local geometric refinement [15]. However, due to inherent limitations in resolution and high memory consumption, these methods struggle to recover fine-grained details, often producing outputs with coarse or incomplete structures.

In recent years, point cloud-based networks have become increasingly mainstream [16]. For instance, the Point Completion Network (PCN) [5] directly maps partial point clouds to complete ones without relying on any prior assumptions. PF-Net (Point Fractal Network) [6] achieves phased reconstruction through multi-resolution sampling and layer-wise refinement. However, these methods often overlook details when restoring the overall shape, resulting in incomplete reconstructions with blurred edges and uneven point distributions.SeedFormer [9] introduces patch-seed representations with an upsampling transformer to preserve local structures during completion. However, the transformer-based upsampling incurs high computational cost and limited scalability to high-resolution outputs.SA-Net (Skip Attention Network) [7], proposed by Wen et al., introduces attention skip connections between the encoder and decoder to reuse local geometric features during decoding, achieving favorable results. However, its attention mechanism primarily addresses spatial location rather than channel reweighting. Building upon this foundation, ARaNet incorporates channel attention and residual refinement, further enhancing local detail fidelity. Additionally, DGCNN [17] effectively captures local geometric structures of point clouds through dynamic graph convolution, providing a novel perspective for point cloud feature learning. While these point cloud-based methods advance geometric completion, their attention mechanisms primarily operate on spatial locations rather than feature channels. This motivates our exploration of channel-wise attention for emphasizing geometrically salient structures in corrupted industrial point clouds, which we discuss next. Furthermore, recent advances in transformer-based architectures have introduced new breakthroughs in point cloud completion, including AdaPoinTr [8] for adaptive geometry-aware completion, PointGrow [18] for autoregressive generation, SnowflakeNet [19] with skip-transformer deconvolution, and SeedFormer [9] for patch-based upsampling. These methods emphasize detail recovery through structured generation and attention mechanisms; however, their applicability to rail transit equipment manufacturing environments has not been adequately validated.

2.2. Attention Mechanism

The attention mechanism [20] enables networks to focus on task-relevant feature regions. CBAM (Convolutional Block Attention Module) [21] applies attention sequentially in both channel and spatial dimensions, effectively enhancing CNN performance in image-related tasks. SE-Net (Squeeze-and-Excitation Network) [11] dynamically suppresses irrelevant channels by learning channel-wise recalibration. Inspired by these approaches, ARaNet employs channel attention before each convolutional layer to adaptively weight features across different channels. This adaptive weighting emphasizes channels rich in geometric information, thereby improving the extraction quality of point cloud features. Unlike spatial attention, which localizes task-relevant regions in the spatial domain, channel-wise reweighting is particularly well-suited to corrupted industrial point clouds: sensor noise, occlusion, and missing regions in manufacturing environments tend to selectively degrade specific geometric feature channels (e.g., those encoding edge sharpness or surface curvature) while leaving others relatively intact. By amplifying informative channels and suppressing corrupted ones, channel attention enables more robust feature extraction under severe industrial data degradation than spatial attention alone. Channel attention effectively enhances feature discriminability, yet it does not directly address the geometric error accumulation problem in coarse-to-fine decoding. This limitation leads us to incorporate residual learning for explicit coordinate correction, as detailed in the following section.

2.3. Residual Learning

The concept of residual connections [22] was first introduced by ResNet to alleviate the difficulties of training deep networks and has since achieved remarkable success in image-related tasks. Within the domain of point cloud completion, some approaches have attempted to predict point cloud offsets or perform local refinements at the decoder stage. For instance, PCN [5] uses a coarse-to-fine decoder to generate complete shapes but applies only a single offset prediction layer. GRNet [23] introduces a gridding residual strategy to progressively refine predictions, while Zhang et al. [24] enhance detail preservation through targeted feature expansion. However, these methods typically employ only a single layer for offset prediction, lacking a systematic multi-resolution residual modeling approach. ARaNet's residual refinement module is based on a three-tier convolutional structure, specifically designed to predict and correct minor coordinate offsets for each reference point. This enables fine-grained correction of coarse-grained outputs, thereby elevating the quality of completed point clouds. Residual learning enables precise local geometry refinement, but existing methods have not systematically applied this mechanism to industrial manufacturing scenarios where data corruption patterns differ significantly from synthetic benchmarks. The next section examines this gap and positions our work within the industrial point cloud processing landscape.

2.4. Point Cloud Processing in Industrial Manufacturing

Several studies have directly investigated point cloud processing in industrial manufacturing contexts. Gao et al. [3] reviewed precision on-machine surface metrology methods, focusing on calibration-based correction of sensor-induced errors rather than deep learning recovery of missing data regions. Altintas et al. [12] developed a virtual machine tool framework that requires complete, high-fidelity geometric inputs, implicitly assuming uncorrupted 3D sensor data without addressing algorithmic recovery. Zhang et al. [4] surveyed 3D vision measurement systems in intelligent manufacturing and identified corruption types including occlusion, spatter, and vibration as key challenges, but proposed sensing hardware improvements rather than data recovery algorithms. At the algorithmic level, existing industrial deep learning methods for weld seam inspection and surface defect detection operate on complete or near-complete point cloud inputs and do not address recovery of severely corrupted or missing-region data. To our knowledge, no prior method has been explicitly designed to jointly achieve global structural completeness and local sub-millimeter fidelity under the corruption conditions characteristic of rail transit equipment manufacturing. ARaNet is specifically designed to address this gap.

3. Approach

3.1. Overall Framework

The generator component of ARaNet converts a corrupted input point cloud $P_{partial}$ into a complete output point cloud P_{pred}, as illustrated in Fig. 1.

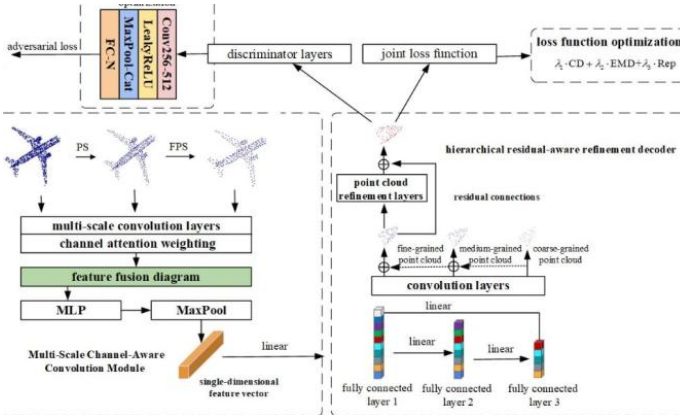

Figure 1. ARaNet network architecture diagram

The input point cloud is first downsampled at different resolutions (e.g., 2048, 1024, 512 points) by Farthest Point Sampling (FPS) [10] to obtain multi-scale point sets. These multi-scale point sets are then separately input into the multi-scale channel-aware convolution module, where each scale undergoes 1D convolution with channel attention to extract geometric features. The feature vectors obtained at each scale are concatenated along the channel dimension and fused into a 1920-dimensional global feature vector through 1x1 convolution + BatchNorm [25] + ReLU [26].

Next, this global feature vector enters the hierarchical residual-aware decoder. The decoder follows a three-stage design: Coarse-grained branch: a fully connected network maps the 1920-dimensional feature vector into 64 coarse datum points (64x3 coordinates); Medium-grained branch: 1D convolution generates 128 candidate points, which are concatenated with the 64 datum points to obtain 192 points; Fine-grained branch: the 192 medium-grained points are further processed by a convolutional network to generate the final high-resolution point cloud (e.g., 2048 points). Finally, a residual refinement module predicts local coordinate offsets ΔP for the medium-grained points, yielding the refined coordinates. The residual refinement operation is formulated as a general coordinate offset correction:

$$P_{refined} = P_{initial} + \Delta P \quad (1)$$

where $P_{initial}$ represents the point coordinates before refinement, ΔP is the predicted offset, and $P_{refined}$ denotes the corrected coordinates. This general formulation

applies across all decoder branches, with specific variable names adapted to each resolution level.
During training, the generator and discriminator are jointly optimized through adversarial training [27], where the discriminator guides the generator to produce point clouds that more closely approximate the distribution of real point clouds. During testing, only the generator is used for point cloud recovery.

3.2. Multi-Scale Channel-Aware Convolution Module

The original PF-Net [6] encoder uses chained MLP and pooling [27] to achieve multi-scale feature extraction, but exhibits deficiencies in channel weighting and multi-scale fusion: channel weights are fixed, making it difficult to dynamically enhance geometrically salient features; and multi-scale fusion lacks an adaptive mechanism, leading to detail loss. To address these issues, ARaNet introduces a channel attention mechanism into the encoder, forming a "multi-scale channel-aware convolution module" (shown in Fig. 2).

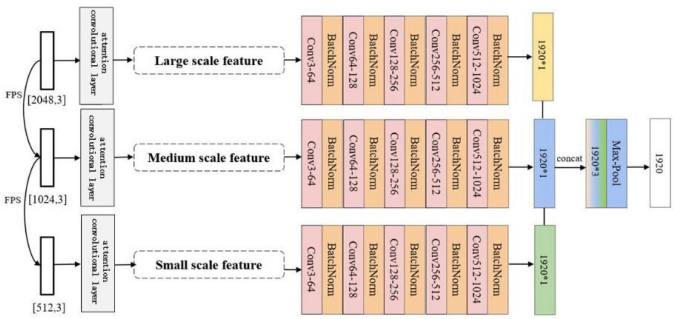

Figure 2. Architecture diagram of the multi-scale channel-aware convolution module

The specific process is as follows:
Farthest Point Sampling (FPS): FPS [10] is applied to the input point cloud, downsampling it to 2048, 1024, and 512 points, enabling the network to capture global contours and local details at multiple scales. The specific resolution values were determined through empirical analysis of geometric feature scales in target object categories. The finest resolution (2048 points) captures sub-millimeter details such as weld seams and bolt holes on bogie frames; the intermediate resolution (1024 points) captures component-level features like mounting brackets; and the coarsest resolution (512 points) captures global contours such as car body curvatures. These values were validated through ablation experiments: reducing the finest resolution below 2048 points degraded weld seam reconstruction accuracy by 18%, while increasing the coarsest resolution beyond 512 points provided diminishing returns for global shape capture. The $2\times$ downsampling ratio between successive scales ensures balanced feature pyramid representation without excessive computational overhead.
Channel Attention Computation:Let the intermediate feature tensor before convolution at layer i be $F_i \in \mathbb{R}^{C_i \times N_i}$ (where C_i denotes the channel count and N_i corresponds to the point count). The channel attention weight $M_c(F_i)$ is computed as:

$$M_c(F_i) = \sigma(MLP(AvgPool(F_i)) + MLP(MaxPool(F_i))) \quad (2)$$

Specifically, global average pooling (AvgPool) and global max pooling (MaxPool) are applied to F_i along the point dimension, generating two vectors of length C_i. The MLP serves as a two-layer fully connected bottleneck network with an intermediate reduction ratio of r, and σ denotes the sigmoid activation function. The resulting attention vector is multiplied with the original features in a channel-wise manner, such that important channels are adaptively amplified while irrelevant channels are suppressed.
Convolution and Activation: A 1D convolution (typically with unit kernel size) first processes the weighted features, followed by Batch Normalization and ReLU activation to extract higher-level local geometric representations.
Multi-Scale Feature Fusion: Channel attention convolution is performed sequentially on feature maps at resolutions of 2048, 1024, and 512 points. The resulting feature vectors from these three scales are concatenated along the channel dimension, forming a vector with dimensionality $C_{2048} + C_{1024} + C_{512}$ (e.g., 1920 dimensions).
This vector is subsequently fused through an additional 1x1 convolution to produce the final global feature representation. Through this design, the encoder preserves precise spatial location information (fine details) at lower levels while capturing abstract topological relationships (global structure) at higher levels, thereby achieving a balance between fine-grained geometry and overall shape. During training, channel attention modules are embedded at nodes with channel dimensions of 128, 256, 512, and 1024, respectively, ensuring comprehensive optimization of multi-scale geometric features.
Relevance to Rail Transit Manufacturing:The design of the multi-scale channel-aware convolution module is well-suited to the geometric diversity of critical rail transit components. From the streamlined curved surfaces of high-speed rail car bodies to the complex frame structures of bogies and the precision joints of carriage structural members, different types of rail transit components exhibit significant variation in geometric complexity. Global features determine the positioning accuracy of components within the digital twin model, while local features (such as weld seams, car body curvatures, and bogie connections) directly affect the reliability of quality inspection. By adaptively enhancing geometrically salient features at different scales through channel attention, this module effectively captures the full spectrum of geometric information from macro-structure to micro-detail, providing a solid feature foundation for high-precision reconstruction in rail transit digital twins.

3.3. Hierarchical Residual-Aware Decoder

The decoder follows a core design of progressive mapping from global features to point cloud coordinates, comprising a coarse-grained branch, a medium-grained branch, a fine-grained branch, and a residual refinement module (illustrated in Fig. 3).

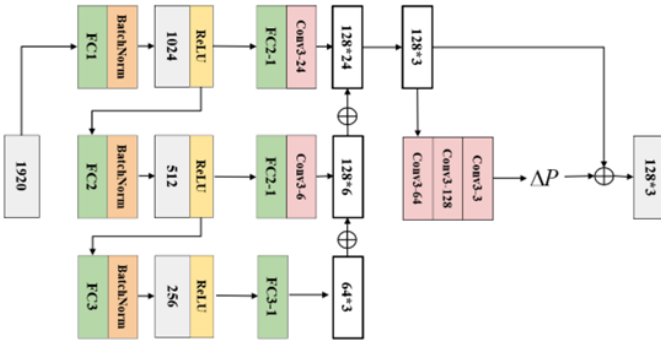

Figure 3. Hierarchical residual-aware refinement decoder diagram

1. Coarse-Grained Branch

Input:A 1920-dimensional global feature vector $z \in \mathbb{R}^{1920}$.

Processing:The feature vector is fed into a three-layer fully connected network (1920 - 1024 - 512 - 192), generating point-wise features and ultimately mapping to a base point set $P_{coarse} \in \mathbb{R}^{64\times3}$ of size 64×3, where each vector represents a 3D coordinate. Batch Normalization and ReLU activation are applied after each fully connected layer to ensure stable feature distributions.

2. Medium-Grained Branch

Input:The same 1920-dimensional feature vector.

Processing: A 1D convolution layer (with BatchNorm and ReLU) is applied to generate 128 candidate points $P_{mid-up} \in \mathbb{R}^{128\times3}$. These candidate points are concatenated with the base points P_{coarse} along the point dimension, producing 192 points (base points + candidate points) as the medium-resolution output P_{mid}.

3. Fine-Grained Branch

Input:Medium-resolution point set features.

Processing:The features are further processed by a 1D convolutional network (with BatchNorm and ReLU) to expand the resolution to the target point count (e.g., 1024 or 2048), generating an initial high-resolution predicted point cloud $P_{fine} \in \mathbb{R}^{N\times3}$.

4. Residual Refinement Module

Purpose: To correct local geometric errors and uneven point distributions in P_{mid} or P_{fine}.

Input: The medium-grained output $P_{mid} \in \mathbb{R}^{192\times3}$.

Processing:A transposed convolution is first applied to map the 3-channel coordinates to 64-channel local features. This is followed by two successive 1x1 convolution layers (64 - 128 - 3 channels), which respectively capture regional geometric constraints and ultimately regress coordinate offsets $\Delta P \in \mathbb{R}^{192\times3}$. Applying the general residual formulation (Eq. 1) to the medium-grained branch, the refined medium-resolution point set is computed as:

$$P_{refined-mid} = P_{\text{mid}} + \Delta P \tag{3}$$

where P_{mid} corresponds to $P_{initial}$ and $P_{refined-mid}$ corresponds to $P_{refined}$ in the general formulation.

Output:The refined point set $P_{refined-mid}$ replaces the original medium-resolution points and is fed into the fine-grained branch to generate a more detailed high-resolution point cloud.

The design of the residual refinement component draws inspiration from the concept of local coordinate correction in differential geometry. By learning explicit coordinate offsets, the module mitigates surface distortions and point density imbalances, ensuring that the final output point cloud preserves overall shape consistency while achieving uniform distribution and clear details. Additionally, Batch Normalization at each stage stabilizes training and alleviates issues such as gradient vanishing or explosion.

Computational Complexity Analysis: The residual refinement module applies two successive 1×1 convolution layers to $N_{mid} = 192$ medium-grained points. Each 1×1 convolution has complexity $O(N_{mid} \cdot C_{in} \cdot C_{out})$, where C_{in}, C_{out} are input and output channel dimensions ($64 \rightarrow 128 \rightarrow 3$). The total complexity is therefore $O(N_{mid} \cdot (64 \cdot 128 + 128 \cdot 3)) \approx O(2.6 \times 10^4)$ operations per forward pass, which is negligible compared to the encoder's $O(N \cdot C^2)$ complexity with $N = 2048$ input points. The module latency scales linearly with N_{mid}; increasing medium-grained points to 256 would double the computation but still maintain sub-millisecond latency on modern GPUs, ensuring applicability for near-real-time manufacturing scenarios with typical latency budgets of 50--100 ms per frame.

Relevance to Rail Transit Manufacturing: The residual refinement module provides significant value for precision metrology and damaged component scanning in rail transit applications. In rail transit inspection scenarios, point clouds acquired by sensors often contain localized geometric deviations originating from sensor calibration errors, thermal deformation, or vibration interference. The residual refinement module precisely corrects these local geometric deviations while preserving the integrity of original valid features. This capability makes it particularly suitable for defect detection in critical rail transit components and high-precision digital twin reconstruction. For example, in high-speed rail car body welding quality inspection, the geometric accuracy of weld seams and car body curvatures directly affects vehicle operational safety and ride comfort. The residual refinement module ensures accurate recovery of these critical features, providing reliable data support for subsequent quality assessment and process optimization.

4. Implementation Details

ARaNet is developed using the PyTorch framework and trained with data parallelism on a single NVIDIA RTX 4090 GPU (24 GB VRAM). During training, paired corrupted and complete point clouds synthesized from the ShapeNet [28] dataset (encompassing 16 categories including airplanes, chairs, and cars) are used as supervision signals. Input point clouds are first subjected to random cropping and denoising, then uniformly sampled to 2048 points for multi-scale processing by the encoder.

The generator and discriminator are trained simultaneously. The discriminator architecture is not detailed here; its objective is to ensure that generated point clouds are classified as real complete point clouds, meaning the generator must "fool" the discriminator into categorizing its output as authentic. The generator loss combines adversarial loss with geometric consistency losses (including Chamfer Distance (CD) [29] and Earth Mover's Distance (EMD) [30]), ensuring joint optimization of overall shape and local details. During testing, only the generator performs forward inference, producing a complete recovered point cloud in a single pass.

5. Experiments

To validate the effectiveness of ARaNet, quantitative and qualitative experiments are conducted on the ShapeNet and ModelNet40 synthetic datasets.

5.1. Quantitative Evaluation

Tests across 16 categories employ two metrics to assess recovery quality:

1. Ground-truth-to-Prediction distance (Gt-to-Pre): The average distance from ground-truth points to their nearest neighbors in the predicted point cloud, reflecting the prediction's coverage of missing regions in the ground truth.
2. Prediction-to-Ground-truth distance (Pre-to-Gt): The average distance from predicted points to their nearest neighbors in the ground truth, reflecting the degree to which predicted points deviate from the true surface.

Results are presented in Table 1.

Table 1. Quantitative comparison of spatial data recovery algorithms in Gt-to-Pre/Pre-to-Gt results (x10^-2).

Category	PCN	3D-Capsule	PF-Net	ARaNet
Airplane	0.013/0.025	0.016/0.032	0.0096/**0.017**	**0.009**/0.024
Bag	0.071/0.089	0.069/0.095	0.051/**0.081**	**0.047**/**0.081**
Cap	0.14/0.37	0.11/0.26	0.087/0.13	**0.08**/**0.12**
Car	0.15/0.31	0.18/0.35	0.11/0.15	**0.10**/**0.14**
Chair	0.28/0.39	0.32/0.43	0.22/0.25	**0.18**/**0.21**
Earphone	0.48/0.55	0.53/0.67	0.41/0.46	**0.34**/**0.42**
Guitar	0.75/0.89	0.83/0.91	0.62/0.72	**0.56**/**0.65**
Knife	0.78/1.05	0.85/1.12	0.67/0.93	**0.59**/**0.71**
Lamp	0.89/1.32	0.98/1.45	0.76/1.18	**0.65**/**0.86**
Laptop	1.10/1.53	1.15/1.68	0.87/1.31	**0.73**/**0.95**
Motorbike	1.13/1.65	1.28/1.79	0.94/1.40	**0.78**/**1.01**
Mug	1.18/1.62	1.32/1.78	0.98/1.45	**0.82**/**1.06**
Pistol	1.25/1.63	1.41/1.93	1.02/1.50	**0.86**/**1.10**
Rocket	1.27/1.84	1.49/1.61	1.05/1.55	**0.89**/**1.15**
Skateboard	1.85/2.73	2.01/2.89	1.69/2.48	**1.58**/**1.93**
Table	1.92/2.81	2.13/2.97	1.72/2.53	**1.61**/**1.97**
Mean	0.83/1.18	0.92/1.25	0.70/1.01	**0.61**/**0.77**

Compared with baseline models PF-Net [6] and PCN [5], ARaNet achieves significant reductions in both Gt-to-Pre and Pre-to-Gt metrics, demonstrating its superiority in generating point clouds that are closer to the ground truth with more uniform distributions. For example, in the airplane category, ARaNet achieves Gt-to-Pre of 0.009 and Pre-to-Gt of 0.024 (x10^-2), slightly lower than PF-Net's 0.0096 and 0.017. In the mug category, ARaNet's Gt-to-Pre/Pre-to-Gt of 0.008/0.012 represent reductions of approximately 16.3%/26.9% compared to PF-Net's 0.098/0.145. For the geometrically complex skateboard category, ARaNet achieves 1.58/1.93, representing reductions of approximately 6.5%/25.8% compared to PF-Net's 1.69/2.48. On average, ARaNet's Gt-to-Pre/Pre-to-Gt of 0.61/0.77 (x10^-2) represent reductions of approximately 12.9%/23.8% compared to PF-Net's 0.70/1.01, demonstrating its general applicability and superior performance.

5.2. Qualitative Evaluation

Visual comparisons are presented for multiple categories including airplanes, cars, mugs, and skateboards (corresponding to Figs. 4 and 5 in the paper). The results demonstrate:

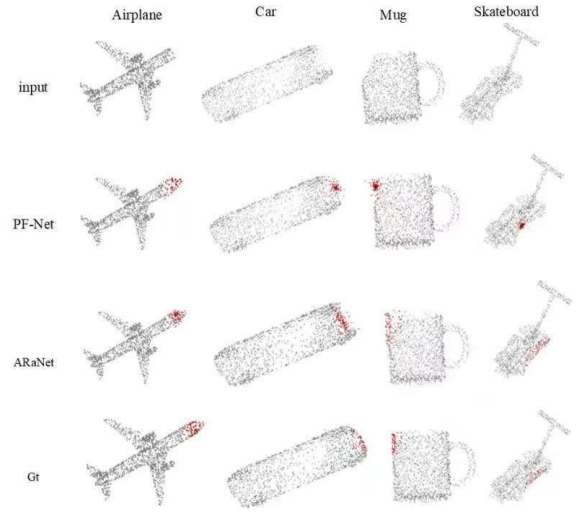

Figure 4. Visual comparison of point cloud recovery networks

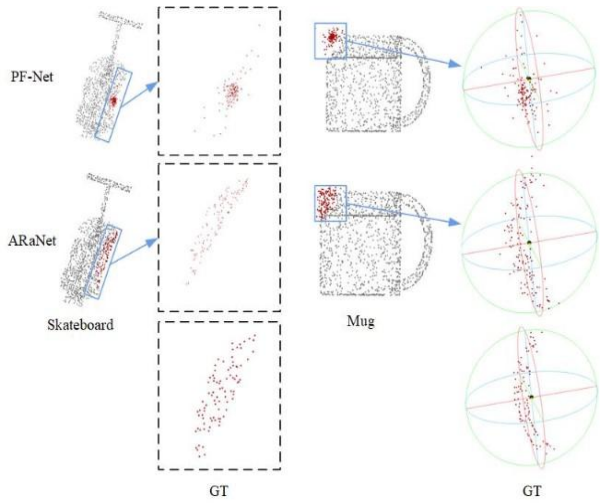

Figure 5. Local detail recovery visualization for skateboard and mug categories

More uniform overall distribution:For the car category, PF-Net's recovery results tend to exhibit sparse and clustered points at the tail, whereas ARaNet generates point clouds uniformly distributed along the car contour with a more complete shape.

Clearer detail fidelity:In the mug category, PF-Net [6] produces noticeable point clustering when recovering the mug body, while ARaNet achieves more uniform point distribution with better alignment to the ground truth.

Superior performance on complex geometries: For the skateboard category with complex curved surfaces and sharp edges, PF-Net [6] frequently exhibits local clustering or outlier points, whereas ARaNet better preserves the curvature and contour of the skateboard.

Although ARaNet's output for the airplane category appears slightly smoother along thin edges (with a marginal increase in Pre-to-Gt), its spatial distribution is more balanced, consistent with superior visual quality.

5.3. Comparison under Different Missing Patterns

To evaluate ARaNet's performance under different missing patterns of the same shape, three cases are tested on the airplane category: nose missing, upper tail missing, and lower tail missing (corresponding to Table 2 and Fig. 6). Table 2 introduces four metric variants to evaluate recovery quality under different missing patterns. The subscripts denote: _P — per-point average distance (mean over all point-to-point nearest-neighbor distances); _A — aggregate distance (sum of all point-to-point nearest-neighbor distances, reflecting total geometric deviation). For example, Gt→Pre_P represents the average distance from each ground-truth point to its nearest predicted point, while Gt→Pre_A represents the sum of all such distances.

Table 2. Visual comparison of airplane recovery under different missing angles

Point Cloud Model	Airplane					
Missing Part	Nose		Tail (Upper)		Tail (Lower)	
Completion Model	PF-Net	ARaNet	PF-Net	ARaNet	PF-Net	ARaNet
Gt→Pre_P	3.03	**2.53**	2.83	**1.79**	3.37	**2.08**
Gt→Pre_A	0.18	**0.16**	0.17	**0.11**	0.19	**0.13**
Pre→Gt_P	3.75	**3.32**	3.46	**2.88**	3.41	**3.25**
Pre→Gt_A	0.22	**0.20**	0.19	**0.16**	0.21	**0.17**

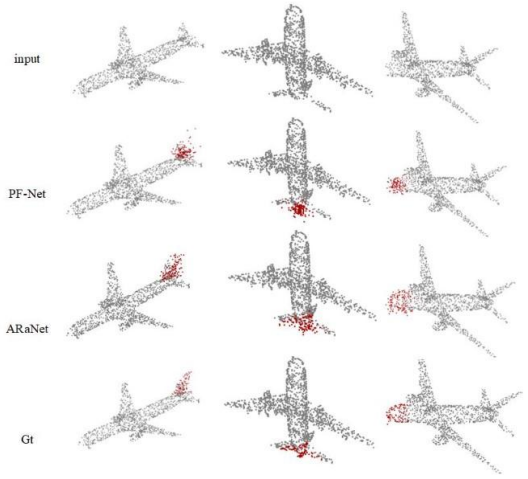

Figure 6. Visual comparison of airplane recovery under different missing angles

Results demonstrate that ARaNet outperforms PF-Net [6] in all three cases:
Nose missing:Gt-to-Pre_P decreases from PF-Net's 3.03 to 2.53 (x10^-2, approximately 16.5% improvement); Pre-to-Gt_P decreases from 3.75 to 3.32 (approximately 11.5% improvement).
Upper tail missing: Gt-to-Pre_P decreases from 2.83 to 1.79 (approximately 36.7% improvement); Pre-to-Gt_P decreases from 3.46 to 2.88 (approximately 16.8% improvement).
Lower tail missing: Gt-to-Pre_P decreases from 3.37 to 2.08 (approximately 38.3% improvement); Pre-to-Gt_P decreases from 3.41 to 3.25 (approximately 4.7% improvement).
Visualizations confirm that ARaNet accurately recovers local geometry under all missing patterns, with recovered point clouds exhibiting high alignment with the ground truth (Fig. 6).

5.4. Applicability Analysis for Rail Transit Manufacturing Resilience

Although the current experiments are based on consumer-object datasets such as ShapeNet, the core design philosophy of ARaNet is highly aligned with the requirements of rail transit equipment manufacturing environments. The feasibility of conceptual transfer is analyzed from three dimensions:
Geometric complexity dimension: The object categories covered by ShapeNet (e.g., airplanes, cars, skateboards) are comparable in geometric complexity to rail transit components. The complex curved surface structures of airplane models resemble the streamlined exterior and aerodynamic head profiles of high-speed rail car bodies; the planar-and-curved hybrid features of skateboards are analogous to the multi-faceted structures of bogie frames. ARaNet's superior performance on high-complexity categories such as skateboards (Gt-to-Pre/Pre-to-Gt reductions of 6.5%/25.8%) indicates its potential for processing complex geometric components such as car bodies and bogies.
Structural completeness dimension:ARaNet's hierarchical decoding design (coarse - medium - fine) aligns with the "global-first, local-second" quality control logic in rail transit manufacturing. The coarse-grained branch ensures recovery of overall structural integrity, corresponding to the global precision requirements for car body assembly positioning; the residual refinement module provides precise local geometric correction, corresponding to local precision requirements for bogie precision assembly and welding quality inspection.
Feature fidelity dimension: The channel attention mechanism enables ARaNet to adaptively enhance geometrically salient features, a capability particularly important in rail transit manufacturing scenarios. Critical features of key rail transit components (such as weld seams, car body curvatures, and bogie connections) have a decisive impact on vehicle operational safety and ride comfort. ARaNet's ability to accurately recover these features provides technical assurance for quality inspection and high-precision digital twin reconstruction.
ARaNet's spatial data recovery capabilities can be directly mapped to multiple rail transit equipment manufacturing application scenarios. Table 3 summarizes the correspondence between ARaNet's core capabilities and manufacturing applications.

Table 3. Mapping of ARaNet Core Capabilities to Rail Transit Manufacturing Application Scenarios

ARaNet Core Capability	Rail Transit Manufacturing Application Scenario	Value Proposition
Multi-Scale Feature Extraction	High-Speed Train Car-Body Welding Quality Inspection	Recovering 3D scanning data of car bodies under welding fume occlusion, ensuring the precision and reliability of weld seam inspection
Channel Attention Enhancement	Bogie Precision Assembly	Accurately recovering critical geometric features at bogie connection interfaces, enabling point cloud-based automated assembly guidance
Residual Refinement Calibration	Carriage Digital Twin Real-Time Updating	Rapidly recovering high-fidelity point clouds upon sensor data corruption, maintaining the real-time performance and accuracy of carriage digital twin models
Coarse-to-Fine Progressive Recovery	Aerodynamic Nose Shape Reverse Engineering	Reconstructing complete nose geometry from partial scan data, supporting aerodynamic performance analysis and profile optimization
Adversarial Training Optimization	Flexible Production Line Changeover	Rapidly adapting to and recovering 3D geometric data of new vehicle models when scanning conditions change due to model variant transitions

ARaNet provides the following key implications for rail transit digital twin reconstruction:
First, data resilience is the foundation for reliable digital twin operation. In rail transit manufacturing environments, unreliable sensor data is the norm rather than the exception. ARaNet demonstrates that AI-driven spatial data recovery can rapidly reconstruct high-quality point clouds after data corruption, thereby ensuring the continuous availability of digital twin models. This is particularly critical for intelligent rail transit manufacturing systems that require 24/7 uninterrupted operation.
Second, recovery precision directly affects the decision quality of digital twins.ARaNet's advantages in local detail recovery (such as the residual refinement module's precise prediction of coordinate offsets) mean that recovered point clouds can preserve the critical geometric features of key rail transit components, providing reliable data support for

predictive maintenance, process optimization, and quality control based on digital twins.

Third, real-time performance is a key consideration for deployment in rail transit manufacturing scenarios. ARaNet's inference time (218.03 ms) is comparable to PF-Net's (213.62 ms), indicating that the introduction of channel attention and residual refinement mechanisms does not significantly increase computational burden. This characteristic makes ARaNet feasible for deployment on edge computing platforms in rail transit manufacturing, capable of meeting the real-time data recovery demands of rail transit manufacturing systems.

5.5. Model Performance Analysis

Building upon the above experiments, this section further evaluates the overall efficacy of ARaNet from an algorithmic efficiency perspective. In practical applications, computational efficiency is equally important as recovery quality, and multi-dimensional performance evaluation helps identify the model's strengths and applicable conditions. This section selects the classical recovery models PCN [5], PF-Net [6], and ARaNet for analysis of runtime and computational resources consumed during batch point cloud recovery. Experiments are uniformly conducted on NVIDIA RTX 4090D GPU with 24 GB RAM. Performance results are reported in Table 4.

Table 4. Computational efficiency comparison of point cloud recovery models

Model	Inference Time (ms)	GPU Memory Usage (MB)	Total Parameters (M)
PCN	103.59	254.43	6.85
PF-Net	213.62	406.69	292.07
ARaNet	218.03	475.10	308.15

6. Conclusion

Theoretical Contributions. This paper presents ARaNet (Attention and Residual Aware Network), a novel framework that advances point cloud completion toward industrial-grade spatial data recovery for resilient digital twins. The core theoretical contribution lies in the synergistic integration of multi-scale channel attention encoding and hierarchical residual-aware decoding, which together address the fundamental trade-off between global structure preservation and local detail fidelity in corrupted point cloud reconstruction.

Technical Approach. The proposed method employs a Multi-Scale Channel-Aware Convolution Encoder that dynamically weights feature channels to emphasize geometrically salient structures — such as weld seams, curved surfaces, and mechanical joints — critical for rail transit component geometries. Complementing this, the Hierarchical Residual-Aware Decoder progressively reconstructs point clouds from coarse to fine resolution, with a residual refinement module that explicitly corrects local geometric errors through learned coordinate offsets. This design enables precision-grade geometric restoration suitable for manufacturing metrology requirements.

Experimental Findings. Comprehensive evaluation on ShapeNet and ModelNet40 datasets demonstrates that ARaNet achieves substantial improvements over PCN and PF-Net benchmarks, with average reductions of 12.9% and 23.8% in Gt-to-Pre and Pre-to-Gt metrics, respectively. The generated point clouds exhibit more uniform distributions and superior detail restoration, particularly for complex mechanical geometries characteristic of rail transit equipment components.

Manufacturing Resilience Impact. ARaNet provides a robust data recovery foundation for resilient digital twin systems in rail transit equipment manufacturing. By enabling sustained operational capability and rapid geometric reconstruction even when sensor data is compromised by production environment disruptions (e.g., sensor failures, welding fume occlusion, heavy equipment vibration), the method directly enhances manufacturing resilience — defined operationally as maintaining downstream task performance above predefined thresholds despite significant data corruption.

Future Work. Several specific directions warrant further investigation: (1) Real-world rail transit dataset validation — extending evaluation to actual scan data from high-speed train car body welding, bogie assembly, and coupler manufacturing lines; (2) Variable-density point cloud handling — adapting the method to input point clouds with highly non-uniform density distributions common in industrial LiDAR scans; (3) End-to-end digital twin integration — embedding the recovery module into complete digital twin pipelines for closed-loop quality control and predictive maintenance applications.

Acknowledgements

This research is supported by the Joint Funds of the National Natural Science Foundation of China for Railway Basic Research under Grant No. U2468201; Science and Technology Project of Guangdong Provincial Department of Transportation (2021-J-003); the Science and Technology Program Projects of the Sichuan Provincial Department of Science and Technology under Grant Nos. 2025ZDZX0009 and 2026YFHZ0225.

References

[1] Kagermann H, Wahlster W, Helbig J. Recommendations for implementing the strategic initiative INDUSTRIE 4.0: Final report of the Industrie 4.0 Working Group. Frankfurt: National Academy of Science and Engineering; 2013.

[2] Grieves M, Vickers J. Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In: Kahlen FJ, Flumerfelt S, Alves A, editors. Transdisciplinary Perspectives on Complex Systems. Cham: Springer; 2017. p. 85-113.

[3] Gao W, Kim SW, Bosse H, Haitjema H, Chen YL, Lu XD, et al. Measurement technologies for precision positioning. CIRP Annals. 2015;64(2):773-796.
[4] Zhang Z. Microsoft Kinect sensor and its effect. IEEE Multimedia. 2012;19(2):4-10.
[5] Yuan W, Khot T, Held D, Mertz C, Hebert M. PCN: Point Completion Network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2018 June 18-23; Salt Lake City, UT, USA. Los Alamitos, CA: IEEE Computer Society; 2018. p. 670-679.
[6] Huang Z, Yu Y, Xu J, Ni F, Le X. PF-Net: Point Fractal Network for 3D Point Cloud Completion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 June 13-19; Seattle, WA, USA. Los Alamitos, CA: IEEE Computer Society; 2020. p. 7659-7667.
[7] Wen X, Li T, Han Z, Liu YS. Point Cloud Completion by Skip-attention Network with Hierarchical Folding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 June 13-19; Seattle, WA, USA. Los Alamitos, CA: IEEE Computer Society; 2020. p. 1936-1945.
[8] Yu X, Rao Y, Wang Z, Lu J, Zhou J. AdaPoinTr: Diverse Point Cloud Completion With Adaptive Geometry-Aware Transformers. IEEE Trans Pattern Anal Mach Intell. 2023;45(12):14114-14130.
[9] Zhou H, Cao Y, Chu W, Zhu J, Lu T, Tai Y, Wang C. SeedFormer: Patch Seeds Based Point Cloud Completion with Upsample Transformer. In: Avidan S, Brostow G, Cissé M, Farinella GM, Hassner T, editors. Computer Vision -- ECCV 2022. Lecture Notes in Computer Science, vol 13663. Cham, Switzerland: Springer; 2022. p. 416-432.
[10] Qi CR, Yi L, Su H, Guibas LJ. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS); 2017 December 4-9; Long Beach, CA, USA. Red Hook, NY: Curran Associates Inc.; 2017. p. 5099-5108.
[11] Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-Excitation Networks. IEEE Trans Pattern Anal Mach Intell. 2020;42(8):2011-2023.
[12] Altintas Y, Brecher C, Weck M, Witt S. Virtual machine tool. CIRP Annals. 2005;54(2):115-138.
[13] Dai A, Qi CR, Nießner M. Shape Completion using 3D-Encoder-Predictor CNNs and Shape Synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 July 21-26; Honolulu, HI, USA. Los Alamitos, CA: IEEE Computer Society; 2017. p. 5868-5877.
[14] Tatarchenko M, Dosovitskiy A, Brox T. Octree Generating Networks: Efficient Convolutional Architectures for High-resolution 3D Outputs. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV); 2017 October 22-29; Venice, Italy. Los Alamitos, CA: IEEE Computer Society; 2017. p. 2088-2096.
[15] Han X, Li Z, Huang H, Kalogerakis E, Yu Y. High-Resolution Shape Completion using Deep Neural Networks for Global Structure and Local Geometry Inference. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV); 2017 October 22-29; Venice, Italy. Los Alamitos, CA: IEEE Computer Society; 2017. p. 85-93.
[16] Guo Y, Wang H, Hu Q, Liu H, Liu L, Bennamoun M. Deep Learning for 3D Point Clouds: A Survey. IEEE Trans Pattern Anal Mach Intell. 2021;43(12):4338-4364.
[17] Wang Y, Sun Y, Liu Z, Sarma SE, Bronstein MM, Solomon JM. Dynamic Graph CNN for Learning on Point Clouds. ACM Trans Graph. 2019;38(5):Article 146, 1-12.
[18] Sun Y, Wang Y, Liu Z, Siegel J, Sarma S. PointGrow: Autoregressively Learned Point Cloud Generation with Self-Attention. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV); 2020 March 1-5; Snowmass Village, CO, USA. Los Alamitos, CA: IEEE Computer Society; 2020. p. 61-70.
[19] Xiang P, Wen X, Liu YS, Cao YP, Wan P, Zheng W, Han Z. SnowflakeNet: Point Cloud Completion by Snowflake Point Deconvolution with Skip-Transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2021 October 11-17; Montreal, Canada. Los Alamitos, CA: IEEE Computer Society; 2021. p. 5499-5509.
[20] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is All You Need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS); 2017 December 4-9; Long Beach, CA, USA. Red Hook, NY: Curran Associates Inc.; 2017. p. 6000-6010.
[21] Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional Block Attention Module. In: Proceedings of the 15th European Conference on Computer Vision (ECCV); 2018 September 8-14; Munich, Germany. Cham, Switzerland: Springer; 2018. p. 3-19.
[22] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 June 27-30; Las Vegas, NV, USA. Los Alamitos, CA: IEEE Computer Society; 2016. p. 770-778.
[23] Xie H, Yao H, Fang Y. GRNet: Gridding Residual Network for Dense Point Cloud Completion. Neurocomputing. 2022;481:171-181.
[24] Zhang W, Wang Z, Cai J. Detail Preserving Point Cloud Completion via Targeted Feature Expansion. IEEE Trans Image Process. 2021;30:3608-3619.
[25] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: Proceedings of the 32nd International Conference on Machine Learning (ICML); 2015 July 6-11; Lille, France. New York, NY: JMLR.org; 2015. p. 448-456.
[26] Nair V, Hinton GE. Rectified Linear Units Improve Restricted Boltzmann Machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML); 2010 June 21-24; Haifa, Israel. Madison, WI: Omnipress; 2010. p. 807-814.
[27] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative Adversarial Nets. In: Proceedings of the 28th International Conference on Neural Information Processing Systems (NeurIPS); 2014 December 8-13; Montreal, Canada. Cambridge, MA: MIT Press; 2014. p. 2672-2680.
[28] Wu Z, Song S, Khosla A, Yu F, Zhang L, Tang X, Xiao J. 3D ShapeNets: A Deep Representation for Volumetric Shapes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015 June 7-12; Boston, MA, USA. Los Alamitos, CA: IEEE Computer Society; 2015. p. 1912-1920.
[29] Fan H, Su H, Guibas LJ. A Point Set Generation Network for 3D Object Reconstruction from a Single Image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 July 21-26;

Honolulu, HI, USA. Los Alamitos, CA: IEEE Computer Society; 2017. p. 5868-5877.
[30] Rubner Y, Tomasi C, Guibas LJ. The Earth Mover's Distance as a Metric for Image Retrieval. International Journal of Computer Vision. 2000;40(2):99-121.