

RESEARCH

Open Access



Ganet: graph attention based Terracotta Warriors point cloud completion network

Jian Gao¹, Yuhe Zhang¹, Gaoxue Shiqin¹, Pengbo Zhou², Yue Wen³ and Guohua Geng^{1*}

Abstract

Point cloud completion technology is used to address incomplete three-dimensional point cloud data, predicting and reconstructing the original shape and details to achieve virtual restoration. While existing learning-based methods have made significant progress in point cloud completion, they still face challenges when dealing with noise and invisible data. To address these issues, this paper proposes a multi-layer upsampling network based on a graph attention mechanism, called GANet. GANet consists of three main components: (1) feature extraction; (2) seed point generation; (3) State Space Model-based Point Cloud Upsampling Layer. GANet demonstrates exceptional robustness in handling noise and invisible data. To validate the effectiveness of GANet, we applied it to Terracotta Warrior data. The Terracotta Warriors, as important cultural heritage, present a challenging test case due to damage and missing parts caused by prolonged burial and environmental factors. We trained and tested GANet on both the PCN dataset and Terracotta Warrior data, comparing it with several recent learning-based methods. Experimental results show that GANet can effectively reconstruct missing or damaged parts of 3D point clouds, providing more detailed and structurally accurate completion results. These completion models not only validate GANet's effectiveness but also offer valuable references for cultural heritage restoration work.

Keywords Terracotta Warriors, Virtual repair, Point Cloud Completion, Deep learning

Introduction

The Terracotta Warriors and Horses are an important legacy of ancient Chinese sculpture, discovered in the 1970s at the mausoleum of Qin Shi Huang. High-quality clay is used to make the Terracotta Warriors and Horses, which are fired at high temperatures through mold molding and hand-carving processes. About 8000 terracotta warriors and horses have been discovered so far, mainly in Pits 1, 2 and 3. Pit 1 is the largest and contains infantry and chariots; Pit 2 has cavalry, chariots and infantry; and Pit 3 is believed to be the command headquarters. As

a result of prolonged burial, the terracotta warriors have suffered weathering, cracking, and human damage that has affected their integrity. A three-dimensional model of eight terracotta warriors is shown in Fig. 1. Each Terracotta Warrior has a red box labeling its partially damaged or missing parts, such as the head, arms, or body. This type of damage poses a challenge for archaeologists and restoration specialists. How to efficiently and accurately carry out restoration and improvement is an urgent problem in the field of digitization and cultural heritage preservation.

Point cloud data, as an important form of three-dimensional data, is widely used in the three-dimensional scanning and reconstruction of cultural relics because it can capture and reconstruct the geometric information of three-dimensional object surfaces with high precision. Point cloud completion, a research hotspot in the field of computer vision, is dedicated to recovering the complete geometric structure of objects from incomplete point

*Correspondence:

Guohua Geng
ghgeng@nwnu.edu.cn

¹ School of Information Science and Technology, Northwest University, Xi'an 710127, China

² School of Arts and Communication, Beijing Normal University, Beijing 100875, China

³ School of law, Wuhan University, Wuhan, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

cloud data. Completing point clouds can generate complete data of the Terracotta Warriors from broken ones, which has significant practical application value for heritage conservation.

However, this task usually faces the following challenges: Severe data loss: In practical applications, due to the extensive damage of the Terracotta Warriors themselves or limited scanning view of the device, the collected point cloud data often contains only parts of the objects. Completing these severely incomplete point clouds requires algorithms to infer the structure of unseen parts, which is inherently a highly uncertain problem. Structural diversity and complexity: Different objects and scenes possess a variety of geometric and topological structures and density distributions, which makes it complex to establish a universal and effective completion model. The model needs to be able to handle a variety of structures from simple planes to complex surfaces.

Due to the damaged and unstructured nature of point cloud data, especially in high-precision restoration tasks, point cloud completion technology poses significant challenges. Traditional point cloud completion methods primarily rely on the geometric symmetry, regular arrangement, and surface smoothness of objects to achieve completion. However, these methods often fail to effectively capture the relational structure information of the missing parts in point clouds, especially when the missing areas are large. In recent years, some research has attempted to use voxelized point clouds and 3D convolutional neural networks (3DCNN) for point cloud completion, drawing on techniques from the image restoration field. However, as the resolution of point clouds increases, the computational resources required by these methods also surge dramatically. With

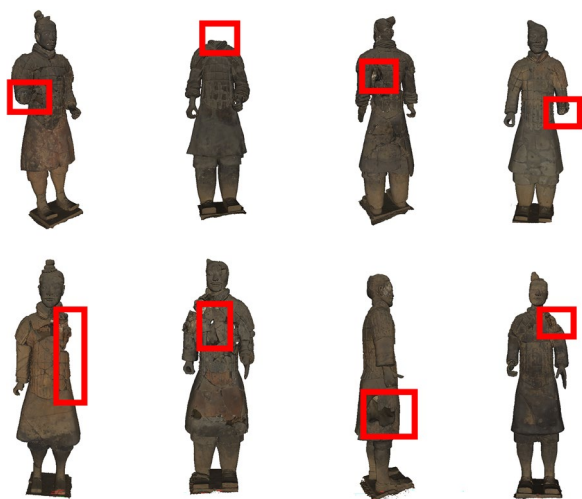


Fig. 1 Schematic diagram of partial damage to the Terracotta Warriors

the introduction of PointNet [1] and PointNet++ [2], directly processing three-dimensional coordinates has become the mainstream method for point cloud-based three-dimensional analysis. Recently, many researchers have proposed a series of new point cloud completion methods, which can mainly be divided into three categories: point-based, attention-based, and graph-based methods. These emerging technologies have shown their unique advantages in handling point cloud completion tasks, yet they lack the capability to represent the details of missing parts, especially in different objects and scenes where their generalization effects are weak.

To address these issues, a novel point cloud completion network based on GANet has been proposed for the virtual restoration of the Terracotta Warriors. The proposed GANet consists of three parts: Firstly, the feature extraction module, which uses PointNet++ [2] to extract features of the damaged point clouds and uses the graph attention encoding module to enhance the integration of local and global features through adaptive feature weighting, and finally uses max pooling to obtain global features. Secondly, the seed point generation module aims to predict the seed points of the complete point cloud (coarse point cloud) through a multi-layer perceptron and external attention mechanism, providing a basis for subsequent fine point clouds. However, directly predicting the complete point cloud (fine point cloud) from seed points (coarse point cloud) would result in the loss of point cloud details, hence the third part of this study adopts a multi-layer State Space Model-based Point Cloud Upsampling Layer. For each layer of the point cloud upsampling module, it takes the point cloud and features generated by the previous layer as input, gradually outputting more detailed point clouds. This module uses the graph attention decoding module to decode the features provided by the previous layer, and uses transposed convolution to generate more detailed point clouds. In this study, we use a three-layer upsampling network. GANet has been tested on the PCN dataset and the Terracotta Warriors dataset. The qualitative and quantitative results obtained show that GANet outperforms current deep learning-based methods.

Overall, our contributions are as follows:

- This study aims to propose a new point cloud completion algorithm, GANet, which can effectively perform virtual restoration of cultural relics such as the Terracotta Warriors.
- Designed a simple and effective graph attention mechanism that integrates local and global features effectively through fusion of positional encoding and adaptive weighted features, making the completion of the Terracotta Warriors more complete and accurate.

- A State Space Model-based Point Cloud Upsampling mechanism is designed to generate point cloud data that more closely matches the dynamic characteristics of real objects by building a state space model and thus generating point cloud data during the complementary process.
- By comparing with various existing deep-learning based methods, it has shown the advantages of GANet in terms of restoration efficiency and accuracy.

Related work

Point-based methods: With the introduction of PointNet [1] and PointNet++ [2], a series of point-based methods have been widely developed. PointNet extracts features of individual points using multilayer perceptrons and integrates global features using a Max-Pooling function, but it does not consider the local neighborhood structure in the point cloud. To address this issue, PointNet++ uses set abstraction layers to capture local information of points and restores the point cloud through an upsampling network. TopNet [3] introduces an innovative decoder that processes geometric information hierarchically. PUNet [4] learns multi-scale features of the point cloud through sub-pixel convolution layers, although it is mainly used for dense reconstruction of point clouds, it is not suitable for completing missing parts. AtlasNet [5] adds a unit square as input, uses it to generate point cloud surfaces, and completes the 3D point cloud by repeating the process from multiple surface elements. MSN [6] uses a deformed decoder to transform a unit plane into a set of surface elements integrated into a coarse point cloud. Hebert and others proposed PCN [7], a coarse-to-fine completion method, which uses dual PointNet layers to extract features and combines the advantages of fully connected and folding-based decoders, surpassing the performance of a single decoder. FinerPCN [8] integrates pointwise convolution into PCN to enhance the local feature information of the point cloud. ASFM-Net [9] uses an asymmetric twin autoencoder to end-to-end generate a coarse point cloud, and outputs the final point cloud through a refinement unit. SDME-NET [10] uses PointNet and PointNet++ in two steps to generate preliminary complete and detailed point clouds from damaged point clouds. LAKe-Net [11] completes point clouds through key point localization alignment and topology awareness. CS-Net [12] combines segmentation and completion models to deal with noisy or anomalous point clouds, obtaining clearer

point cloud inputs for the completion module through segmentation.

Attention-based methods: The successful application of Transformers [13] in the text processing field has attracted the attention of many researchers, and as a result, it has been introduced into computer vision. The core of this technology is the multi-head attention mechanism, which can adaptively learn and weight important information. In the field of computer vision, PUI-Net [14] combines attention convolution units and non-local feature expansion units to extract key features in the point cloud through a channel attention mechanism and uses local feature expansion units to generate high-resolution dense feature maps. Additionally, to address the gaps in incomplete point clouds, it introduces an innovative inner drawing loss mechanism. N-DPC [15] achieves dense point cloud completion by merging self-attention units and local and global features. PointGrow [16] uses a recurrent neural network integrated with an attention mechanism to sample points by analyzing the conditional distribution output by the previous network, enhancing the correlation between points. SnowflakeNet [17] proposes an innovative multi-layer snowflake deconvolution (SPD) module, combined with Skip-Transformer to capture local shape characteristics, optimizing the point splitting process, and enabling neighboring SPD modules to work effectively together. PoinTr [18] designs a geometry-sensitive point cloud completion Transformer by representing the point cloud as a set of unordered point proxies and using the Transformer's Encoder-Decoder structure to generate missing point clouds. PointAttn [19] utilizes cross-attention and self-attention mechanisms to optimize the point cloud completion task from coarse to fine.

Graph-based methods: Point cloud data can be viewed as vertices of a graph, making graph convolutional networks suitable for handling point clouds to explore relationships between points or local areas. DGCNN [20] innovatively introduced graph convolution into point cloud processing, using EdgeConv to compute neighborhood information for each point, combined with multilayer perceptrons and max pooling functions to extract local features of the point cloud, dynamically updating the feature map in space. Inspired by DGCNN, LDGCNN [21] adopts DGCNN's feature extraction method and connects multi-level features from different layers to improve performance and optimize model scale. FoldingNet [22] uses graph convolution to extract features from fixed adjacency information of sampled center points. ECG [23] is divided into generating seed points and refining point cloud details in two stages, each conditioned on the input incomplete point cloud, using a deep hierarchical encoder with graph

convolution capabilities to propagate multi-scale edge features to refine local geometric details. Shi [24] and others proposed a graph-guided deformation network, viewing input data and intermediate products as control points and support points, simulating the minimum square Laplacian deformation process through mesh deformation, providing adaptability for optimizing geometric details in point cloud completion tasks. CRA-Net [12] introduces a cross-regional attention module based on graph attention networks, which evaluates potential connections among all local features and generates a coarse point cloud. Subsequently, it recovers clearer object shape details with lower instability through a folding-based layer. PC-RGNN [25] proposes a local-global attention mechanism and a multi-scale graph-based context aggregation module, focusing on capturing topological relationships between points, significantly enhancing feature learning capability.

Method

Since the point cloud model of damaged Terracotta Warriors typically exhibits complex structures and unique detailed features, completing the missing areas in the point clouds of damaged Terracotta Warriors is a highly

challenging task. Since there is no corresponding complete point cloud for already damaged Terracotta Warriors, we simulate the point clouds of damaged Terracotta Warriors by sampling from the complete point clouds. We use a supervised deep learning network to learn the mapping from damaged point clouds to complete point clouds, and eventually apply this mapping to the damaged Terracotta Warriors to achieve virtual restoration.

Assume T is a set of 3D points on the observed surface of a complete Terracotta Warrior model, and P is the point cloud of a damaged Terracotta Warrior obtained from a single observation or a series of observations from a 3D sensor. Since P and T are sampled independently, there is no explicit correspondence between the points of the two models. Our task is to learn the mapping from P to T .

The Fig. 2 shows the overall framework of our proposed GANet. GANet adopts a classic encoder-decoder structure, mainly comprising three modules: a feature encoding module for predicting the shape of missing parts, a seed point generation module (decoder module) for generating a rough complete point cloud, and an State Space Model-based Point Cloud Upsampling Layer for generating a refined point cloud. The upsampling module

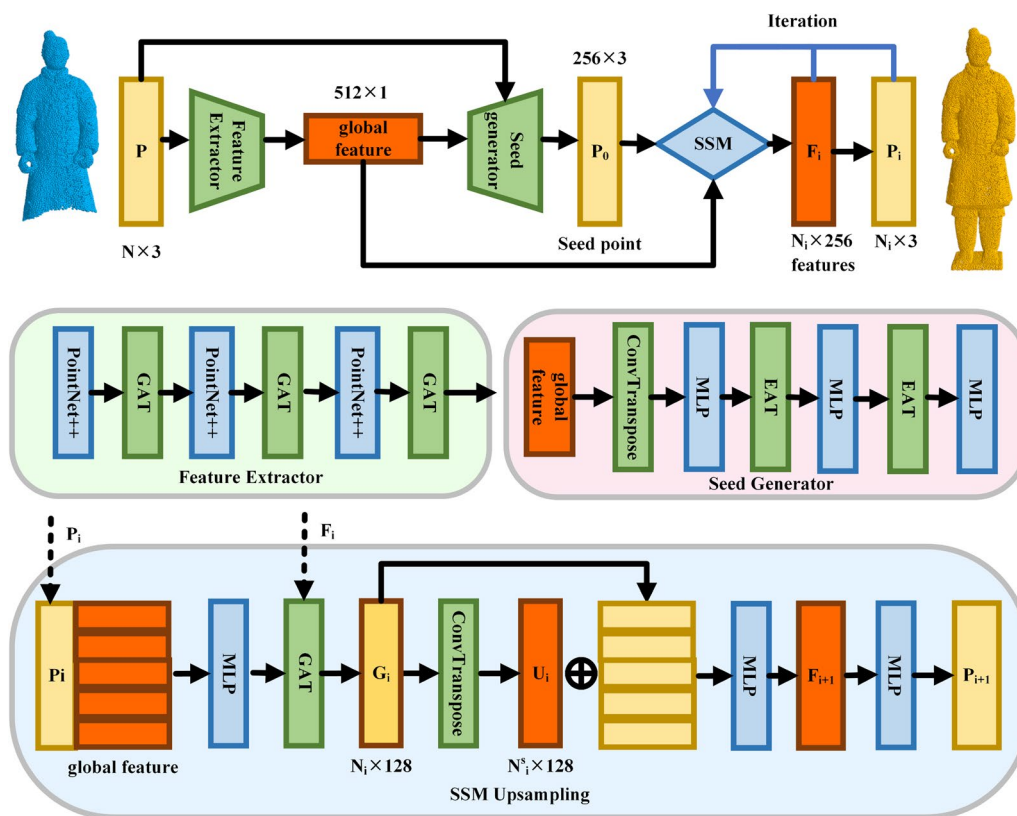


Fig. 2 Framework of the proposed GANet

is iterative and can be iterated multiple times as needed to achieve fine-grained completion.

To construct these three modules, we were inspired by the attention mechanism and PointNet++, and designed two basic units: the Graph Attention Mechanism and State Space Model-based Point Cloud Upsampling Layer. GANet performs excellently in handling multiple object categories and does not require objects to have symmetry or planarity. Through these designs, GANet can effectively complete the point cloud models of damaged Terracotta Warriors, restoring their original complex structures and detailed features. This method not only improves the accuracy of restoration but also can be used to provide an effective shape and detail reference to the physical restoration staff. Below, we will introduce our proposed method in detail.

Feature extraction layer

The feature extraction layer is responsible for extracting information from the input point cloud of damaged Terracotta Warriors and predicting the global feature vector $f_{global} \in \mathbb{R}^l$, where $l = 512$. Inspired by the outstanding performance of attention mechanisms in text processing, our feature encoding module primarily consists of set abstraction, and a graph attention mechanism (GAT) as shown Fig. 3. The set abstraction is used to extract information from the damaged point cloud, while the graph attention mechanism is used to predict the global features of the missing parts from the extracted information.

The feature encoding module achieves hierarchical downsampling through multiple set abstractions obtaining point-wise features at different scales. The output of the final graph attention mechanism can be considered the global feature. The set abstraction consists of three modules: sampling, grouping, and MLP:

$$X_i = \text{FPS}(P) \tag{1}$$

$$X_f = \text{concat}(\text{knn}(X_i, P)) \tag{2}$$

$$X_F = \text{ReLU}(\text{bn}((W_i \cdot X_f + b_i))) \tag{3}$$

where FPS stands for Farthest Point Sampling [26], KNN is the k-nearest neighbors algorithm, ReLU is the Rectified Linear Unit, bn is Batch Normalization, and MLP is the Multi-Layer Perceptron. The sampling process is illustrated in Fig. 4.

After obtaining the point cloud features for the current layer, We have designed a novel Graph Attention Mechanism (GAT) to improve feature extraction and representation in data. This mechanism combines positional encoding and feature encoding in multiple steps to complete feature aggregation. First, the inputs

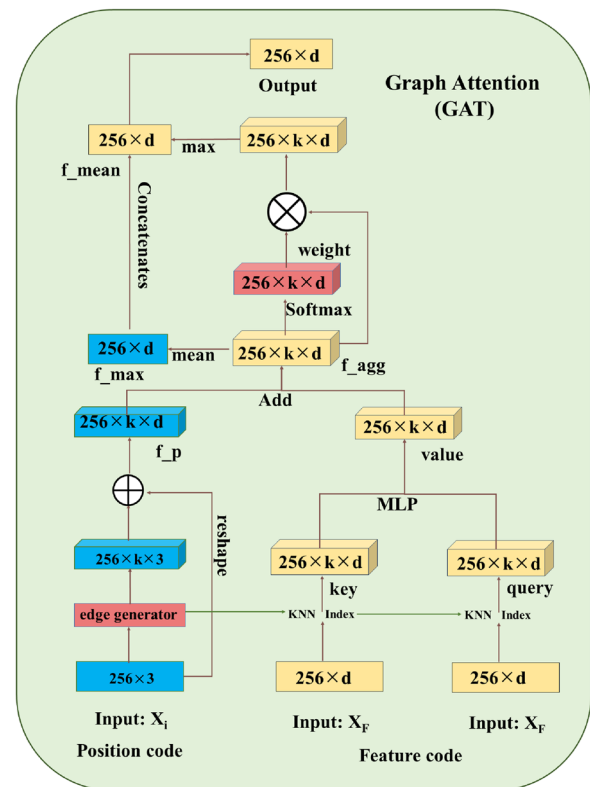


Fig. 3 Schematic of point cloud feature extraction based on graph attention mechanism (GAT)

to the GAT include the point cloud positions $X_i \in \mathbb{R}^{N_i \times 3}$ and the corresponding features $X_F \in \mathbb{R}^{N_i \times d}$, N_i is the number of points in the current feature extraction layer, replacing N_i with 256 in the Fig. 3. Positional encoding constructs a connectivity graph for the point cloud using an edge generator. The edge generator determines the relationships between points based on their spatial position, utilizing the k-nearest neighbors (KNN) algorithm to generate the graph structure, resulting in a set of edge index. Point cloud X_i generates the position graph structure of the point cloud based on this index.

In the positional encoding stage, the point cloud position graph is fed into a Multi Layer Perceptron (MLP), composed of several fully connected layers, to encode the positional information and obtain the positional feature representation $f_p \in \mathbb{R}^{256 \times k \times d}$. Specifically, the positional encoding process can be represented as:

$$f_p = \sigma(W_2 \cdot \sigma(W_1 \cdot \text{KNN}(X_i) + b_1) + b_2) \tag{4}$$

where W_1 and W_2 are weight matrices, b_1 and b_2 are bias vectors, and σ is a non-linear activation function such as ReLU.

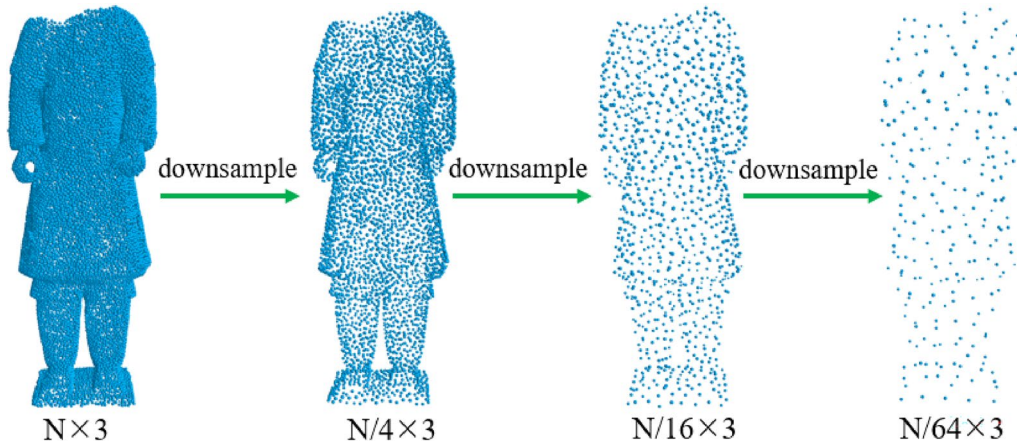


Fig. 4 Schematic of the furthest point sampling

Next is the feature encoding stage, where the process generates the key and query components required by the attention mechanism using the input features. The input features X_F are fed into two independent MLPs to compute the key $K \in \mathbb{R}^{256 \times d}$ and the query $Q \in \mathbb{R}^{256 \times d}$, where d is the feature dimension of each point.

The generated keys and queries summarize the neighborhood information based on the edge indexes generated by the positional encoding to generate $256 \times k \times d$ dimensional features, where k is the number of neighborhoods for each point. The aggregated key and query representations are then fed into another MLP to integrate the information, resulting in the aggregated features $f_{agg} \in \mathbb{R}^{256 \times k \times d}$. This process can be represented as:

$$f_{agg} = \sigma(W_{agg,2} \cdot \sigma(W_{agg,1} \cdot \text{concat}(K, Q) + b_{agg,1}) + b_{agg,2}) \quad (5)$$

Subsequently, the aggregated features f_{agg} are added to the positional features f_p to form the integrated feature representation:

$$f_{int} = f_{agg} + f_p \quad (6)$$

To compute the attention weights, a softmax operation is applied to the integrated features f_{int} . Specifically, for each point i and its neighbor j , the attention weight α_{ij} is calculated as:

$$\alpha_{ij} = \frac{\exp(f_{int,ij})}{\sum_{j=1}^k \exp(f_{int,ij})} \quad (7)$$

The final feature representation is obtained by performing element-wise multiplication of the attention weights α with the integrated features f_{int} :

$$f_{final} = \alpha \cdot f_{int} \quad (8)$$

In the final step, pooling operations are applied to obtain a comprehensive feature representation. The final feature representation undergoes max pooling and mean pooling operations, and these pooled features are concatenated to form the output feature vector:

$$f_{global} = \text{concat}(\text{Max}(f_{int}), \text{Mean}(f_{final})) \quad (9)$$

This method significantly enhances the feature representation of point cloud data by integrating positional encoding and the graph attention mechanism. Positional encoding ensures effective utilization of spatial information, the edge generator extracts important edge features, the graph attention mechanism captures complex relationships between points and their neighbors, and multi-scale pooling further enriches the feature representation. Overall, this mechanism allows the model to better capture intrinsic structural information when handling complex 3D point cloud data, thereby improving model performance.

Seed generator

As shown in Fig. 2, the proposed Seed Generator architecture starts with a 512-dimensional global feature vector input. This global feature undergoes a transposed convolution operation to transform it into a 128x256-dimensional feature map. The Seed Generator then uses a series of three stacked Multi-Layer Perceptrons (MLPs) and External Attention Mechanisms (EAT) for feature extraction and enhancement, EAT as detailed in Fig. 5.

Specifically, the input global feature is denoted as $f_{global} \in \mathbb{R}^{512}$, which is transformed into a high-dimensional feature representation $F_{trans} \in \mathbb{R}^{128 \times 256}$ through

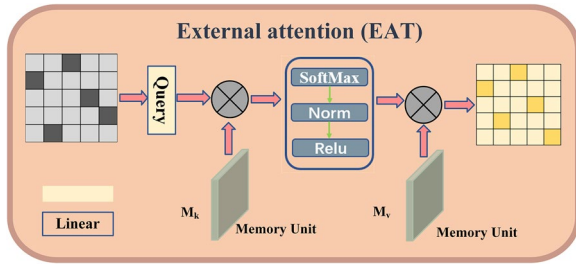


Fig. 5 Flowchart of feature extraction based on external attention mechanism (EAT)

a transposed convolution layer. The External Attention Mechanism (EAT) combines a query-based attention mechanism with external memory units to enhance feature representation. The input feature map to the EAT module is first transformed into a query vector through a linear transformation. Let the input feature map be $F_{trans} \in \mathbb{R}^{N \times d}$, where N is the number of features and d is the feature dimension. The feature map is transformed into a query vector as follows:

$$Q = F_{trans} W_q \tag{10}$$

where $W_q \in \mathbb{R}^{d \times d_k}$ is the query weight matrix, and $Q \in \mathbb{R}^{N \times d_k}$ is the generated query vector. The query vector Q interacts with two external memory units M_k and M_v , which store key and value pairs, respectively. The attention mechanism computes the weighted sum of these values, modulated by the query:

$$\xi = \text{SoftMax} \left(\frac{QM_k^T}{\sqrt{d_k}} \right) \tag{11}$$

where $\xi \in \mathbb{R}^{N \times N_k}$ is the attention matrix, $M_k \in \mathbb{R}^{N_k \times d_k}$ is the key memory unit. This output is then processed through normalization and an activation function:

$$E = \text{ReLU}(\text{Norm}(\xi)) \tag{12}$$

$$F_{EAT} = EM_v \tag{13}$$

$M_v \in \mathbb{R}^{N_k \times d_v}$ is the value memory unit, and $E \in \mathbb{R}^{N \times d_v}$ is the external attention output. The output of the EAT module is combined with the input features to generate the final enhanced feature map:

$$F_{enhanced} = F_{trans} + F_{EAT} \tag{14}$$

Finally, the output is mapped to a 3x256 dimensional point cloud using a MLP. We reformulate the task of point cloud completion into the prediction of missing parts. This reconstructed 3x256 dimensional point cloud is concatenated with the initial incomplete point cloud P and subjected to Farthest Point Sampling (FPS) to generate a

3x512 seed point cloud P_0 . This seed point cloud serves as the input for subsequent processing stages, ensuring robust and accurate point cloud reconstruction.

State space model-based point cloud upsampling layer (SSM)

This paper presents a State Space Model (SSM)[27]-based point cloud upsampling layer designed to improve the resolution and detail capture of point cloud data. Similar to the existing Snowflake Point Deconvolution (SPD) method, SSM increases the number of points by splitting each parent point into multiple child points.

In the SSM upsampling layer, we start with the input features F_i and point cloud P_i passed from the previous layer, the global features f_{global} are the final output of the feature encoding layer. The point cloud P_i and the global features f_{global} are then encoded by another MLP to generate intermediate feature representations. In the first SSM Upsampling layer, copy the intermediate feature as F_i , Seed points as P_i . Next, the GAT is applied to generate shape context features G_i . Convolution transpose operations are then used to upsample these features, producing higher resolution features U_i . These upsampled features are concatenated with the shape context features G_i to form new feature representations F_{i+1} .

Simultaneously, another set of MLP layers processes these concatenated features to generate the upsampled point cloud P_{i+1} . In each upsampling layer, the output point cloud P_{i+1} and the state space features F_{i+1} are used as the input for the next SSM upsampling layer, ensuring the transmission and preservation of state space features throughout the network. Importantly, our model upsamples the point cloud to twice the number of points in each iteration.

The state space model (SSM) updating process proposed in this paper can be simplified and described as follows:

$$F_{i+1} = AF_i + BP_i \tag{15}$$

$$P_{i+1} = CF_{i+1} + DP_i \tag{16}$$

The SSM upsampling layer workflow is illustrated in Fig. 6, refer to the SSM upsampling section of Fig. 2 for specific details. In our model, F_i represents the current layer's features, P_i represents the input point cloud, F_{i+1} represents the updated features, and P_{i+1} represents the upsampled point cloud. This approach effectively updates and transmits state information across layers, enhancing the feature preservation and detail capture during the point cloud upsampling process.

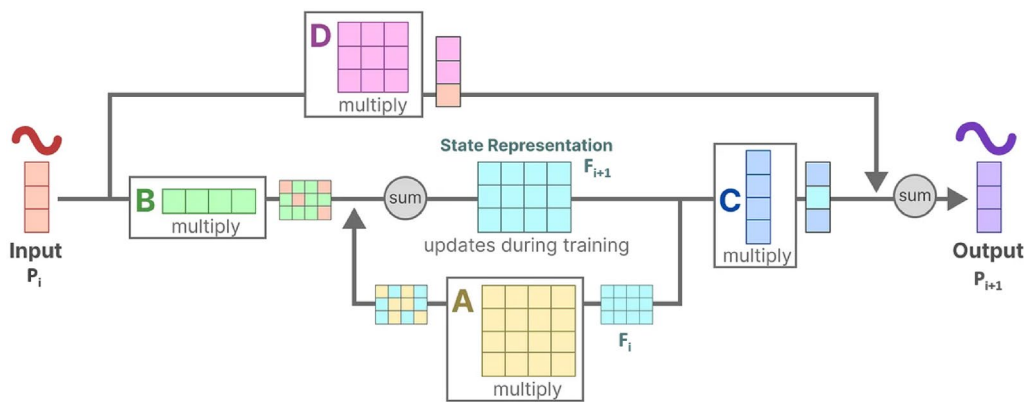


Fig. 6 Framework of the proposed SSM

Loss function

In our experiments, we introduced two different metric functions: Chamfer Distance (CD)[28] and Earth Mover’s Distance (EMD)[29].

Chamfer Distance (CD)

$$CD(P_i, T) = \frac{1}{|P_i|} \sum_{p \in P_i} \min_{t \in T} \|p - t\|_2 + \frac{1}{|T|} \sum_{t \in T} \min_{p \in P_i} \|t - p\|_2 \tag{17}$$

CD calculates the average minimum distance from points in the point cloud P_i to the points in the ground truth point cloud T . We use a symmetric version of CD, where the first term forces the output points to be close to the ground truth points, and the second term ensures that the ground truth points are covered by the output points.

Earth Mover’s Distance (EMD)

$$EMD(P_i, T) = \min_{\phi: P_i \rightarrow T} \frac{1}{|P_i|} \sum_{p \in P_i} \|p - \phi(p)\|_2 \tag{18}$$

EMD finds an optimal bijection that minimizes the average distance between the points in the point cloud P_0 and the corresponding points in the ground truth point cloud T .

Chamfer Distance, as a loss function for point cloud completion, has the advantages of high computational efficiency, simple implementation, and wide applicability. However, its local matching property and sensitivity to noise limit its ability to capture global structures and complex details. EMD considers the global matching between point clouds, rather than just local distances. By finding an optimal bijection, EMD can more comprehensively measure the differences between two point clouds. This is crucial for capturing the global structure and shape of point clouds. However, EMD has high

computational complexity and resource consumption. Therefore, in our experiments, we use both CD and EMD loss functions for P_0 . For the subsequent P_i , we use only the CD distance as the loss function. We train using the Adam[30] optimizer with a batch size of 16, an initial learning rate of 0.001, and a learning rate decay of 0.99 per iteration, for a total of 400 epochs.

Experiments

In this section, we first describe the dataset used to train our model, as well as the comparison methods and evaluation metrics. Next, we conduct qualitative and quantitative comparisons between our method and the comparison methods. Finally, we perform ablation studies on the key modules we proposed.

Dataset and experimental setup

PCN Dataset [31]: To train our model, we use a 3D dataset derived from the PCN dataset, which includes pairs of partial and complete point clouds. The PCN dataset comprises eight categories: airplane, table, boat, chair, cabinet, car, lamp, and sofa, with a total of 30,974 mesh models.

Terracotta Warriors Dataset: The dataset from the National Joint Local Engineering Research Center for Digitization of Cultural Heritage, Northwest University, China. We collected five types of complete Terracotta Warriors models using the Artec EVA scanner, totaling over 500 models. Each model consists of approximately 900,000 points, including XYZ coordinates and normal vectors.

In our experiments, the complete point clouds were generated by uniformly sampling 16,384 points on the mesh surface or within the point cloud. The partial point clouds were generated by back-projecting 2.5D depth images into 3D, with each point cloud consisting of 2,048 points. For the PCN dataset, to ensure a fair comparison,

we followed PCN's training and testing split settings. For the Terracotta Warriors dataset, we used 80% of the data for training and 20% for testing.

Baseline Methods and Evaluation Metrics: To conduct a comprehensive comparison, we selected eleven typical methods as baselines: FoldingNet [22], AtlasNet [5], PCN [7], TopNet [3], GRNet [32], PMP-Net [33], CRN [34], PointTr [18], NSEFA [35], SA-Net [36], SPD [17], and PointAttn [19]. These methods are recent and representative deep learning-based approaches, with implementations based on the authors' code provided on GitHub. They were also trained on our Terracotta Warriors dataset. We used the Chamfer Distance (CD) as the evaluation metric for point cloud completion.

Evaluation on PCN dataset

Quantitative Analysis: The quantitative results of our proposed method and 12 previous baseline methods on the PCN dataset are shown in Table 1. The results of the baseline methods are referenced from SPD and PointAttn. Our proposed method achieved optimal performance in 6 out of 8 categories (Cabinet, Chair, Lamp, Couch, Table, and Boat). Compared to the previously best method, PointAttn, our method reduces the average Chamfer Distance (CD) by 5.8% (6.46 versus 6.86), as can be seen from the last two rows of Table 1. Notably, our method outperforms all categories compared to state-of-the-art attention-based methods

such as PointAttn, SPD, and PointTr. This comparison demonstrates the superiority and effectiveness of our proposed graph attention mechanism framework.

Qualitative Analysis: To further evaluate the performance of our method, we conducted a qualitative analysis of the experimental results, as visualized in Fig. 7. From the figure, we can see that our method also achieves better visual effects on the PCN dataset compared to state-of-the-art methods, including fewer noisy points and more precise geometric structures. Compared to recent attention-based methods like PMPNet and SPD, our method particularly achieves more precise geometric structures and fewer noisy points, as shown in categories such as Chair (noisy points below the chair generated by PMPNet), Couch (the geometric structure of the trunk), and Table (cluttered points at the bottom of the table generated by SPD). This is because our proposed graph attention mechanism dynamically increases the receptive field of the local neighborhood, taking into account the specific graph structure features of the object, thereby generating better local details and smoother surfaces.

Evaluation in the Terracotta Warriors dataset

We present the quantitative results of our proposed method and five recent deep learning-based methods on the Terracotta Warriors dataset, as shown in Table . From the table, it can be seen that our proposed method ranks first in the CD evaluation metric across all categories,

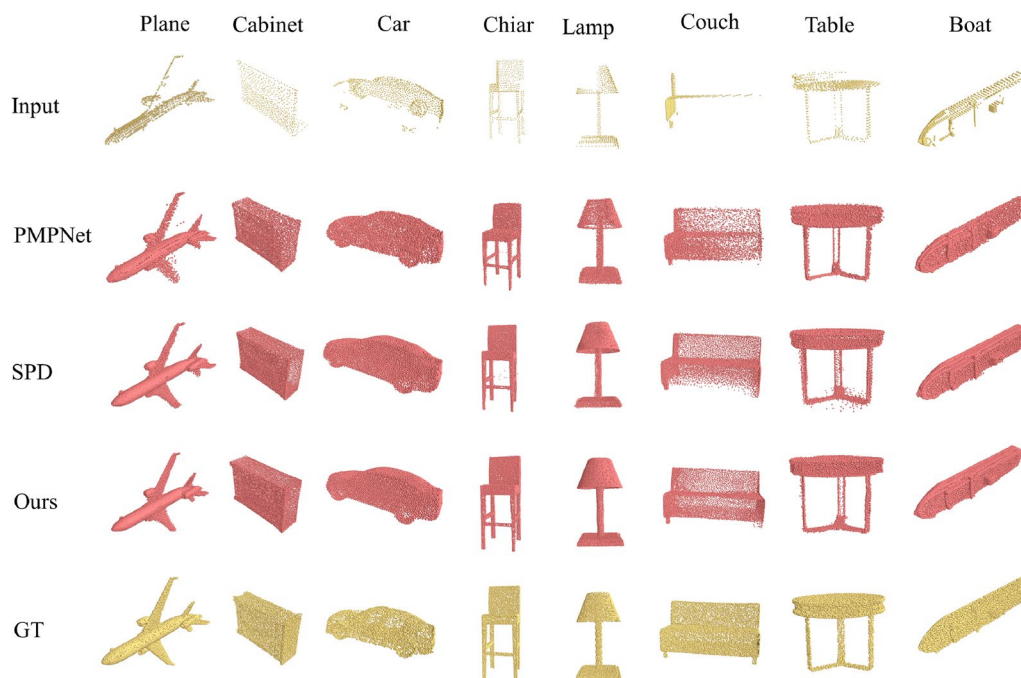


Fig. 7 Visual comparison on PCN dataset. Our method achieves less noisy and smoother completions

Table 1 The point cloud completion of our method and the baseline method on the PCN dataset is expressed in terms of the L1 chamfer distance $\times 10^3$ (lower is better) per point

Methods	Average	Plane	Cabinet	Car	Chair	Lamp	Couch	Table	Boat
FoldingNet	14.31	9.49	15.80	12.61	15.55	16.41	15.97	13.65	14.99
AtlasNet	10.85	6.37	11.94	10.10	12.06	12.37	12.99	10.33	10.61
PCN	9.64	5.50	22.70	10.63	8.70	11.00	11.34	11.68	8.59
TopNet	12.15	7.61	13.31	11.30	13.82	14.44	14.78	11.22	11.12
GRNet	8.83	6.45	10.37	9.45	9.41	7.96	10.51	8.44	8.04
PMP-Net	8.73	5.65	11.24	9.64	9.51	6.95	10.83	8.72	7.25
CRN	8.51	4.79	9.97	8.31	9.49	8.94	10.69	7.81	8.05
PointTr	8.38	4.75	10.47	8.68	9.39	7.75	9.83	7.78	7.29
NSFA	8.06	4.76	10.78	8.10	8.63	8.53	10.53	7.35	7.48
SA-Net	7.74	2.18	9.11	5.56	8.94	9.98	7.83	9.94	7.24
SnowflakeNet	7.21	4.29	9.16	8.08	7.89	6.07	9.23	6.55	6.40
PointAttn	6.86	3.87	9.00	7.63	7.43	5.90	8.68	6.32	6.09
ours	6.46	4.08	8.59	7.80	6.69	5.18	7.78	5.86	5.67

Bold indicates that the minimum loss is achieved in all methods

demonstrating the effectiveness and robustness of our model. Compared to the previous best method, SPD, our proposed method reduces the average CD error distance by 17% (8.23 vs. 9.93), as shown in the last two rows of Table 2. Notably, compared to attention mechanism-based methods such as PMPNet, PCN, and PointAttn, our method achieves significant improvements, primarily due to the graph attention module's ability to capture the global structure of shapes.

To further evaluate the performance of our method, we conducted a qualitative analysis of the experimental results, as visualized in Fig. 8. From the figure, it can be seen that our method achieves higher quality results on the Terracotta Warriors dataset. Compared to recent attention mechanism-based methods such as PMPNet and SPD, our method particularly achieves more accurate geometric structures and fewer noise points, as demonstrated by the legs of the Warrior (more noise points in PMPNet and SPD results), the shoulder of the Kneeling Archer, and the foreleg of War Horse 2 (smoother local details generated by our method). This is because the graph attention mechanism we proposed effectively considers the local graph structural features of objects and dynamically increases the receptive field of the local neighborhood during this process, resulting in better local details and smoother surfaces. Some additional visualization results are shown in Fig. 9.

Ablation experiments

In this section, we validate the effectiveness of each module in our model through ablation experiments on the Terracotta Warriors dataset. We analyzed the following four key components.

To demonstrate the effectiveness of the farthest point sampling (FPS) module in the feature extraction layer, we replaced it with the graph attention mechanism for direct feature extraction. The experiment revealed that the time required increased significantly, but the accuracy was similar to that achieved with the FPS module. To validate the effectiveness of the graph attention mechanism in the feature extraction layer, we used a multi-layer perceptron (MLP) for feature extraction after applying FPS, denoted as GANet1. The experimental results are shown in the first row of Table 3. We found that the error of GANet1 increased significantly compared to using the graph attention mechanism. This is because the feature extraction capability of the MLP is limited and does not adequately capture local information for each point.

To verify the effectiveness of the external attention mechanism in the seed point generation module, we replaced our seed point generation module with the one from SnowflakeNet, denoted as GANet2. The experimental results are shown in the second row of Table 3. We found that the external attention mechanism enhanced the model's generation accuracy, as it can more accurately understand the contextual relationships in the point cloud, leading to the generation of points that better fit the structure of real objects. This helps reduce errors and noise in the completed point cloud.

To demonstrate the effectiveness of the State Space Model-based Point Cloud Upsampling Layer (SSM), we used transposed convolution and upsampling for fine point cloud generation, denoted as GANet3. The results are shown in the third row of Table 3. It is evident that GANet3's generation accuracy is very poor, indicating that the proposed SSM module significantly improves

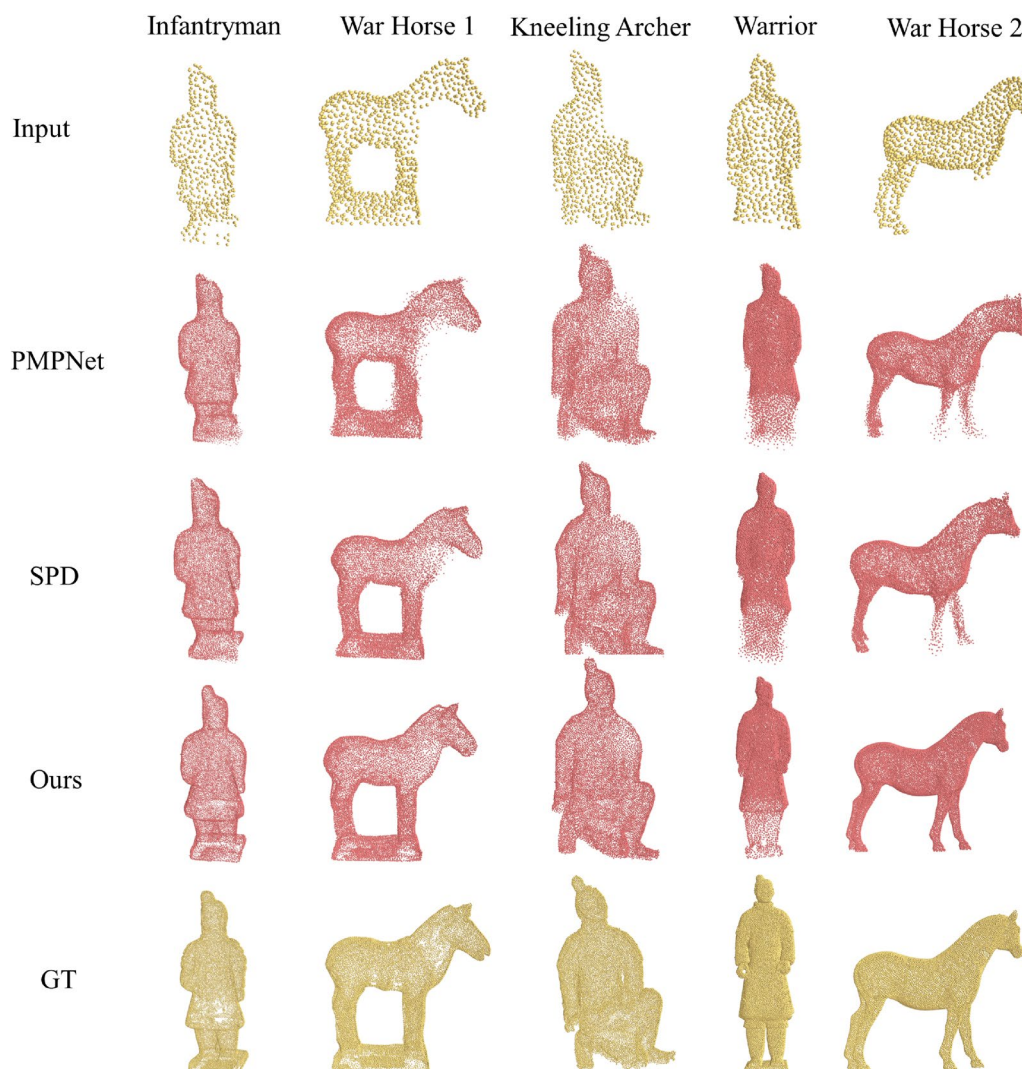


Fig. 8 Visual comparison on the Terracotta Warriors dataset. Our method allows for more accurate and uniform completions

generation performance. This is because the SSM effectively captures the dynamic features in point cloud data. By establishing a state space model, it can better describe the trends in point cloud changes, thus generating point cloud data that conforms to the dynamic characteristics of real objects during the completion process.

To validate the effectiveness of the loss function used, we created a variant that uses only the Chamfer Distance (CD) without the Earth Mover's Distance (EMD), denoted as GANet4. The results are shown in the fourth row of Table 3. We found that the error with only CD was slightly higher than GANet's results, indicating the effectiveness and superiority of our loss function. This is because EMD better reflects the geometric differences in the point cloud compared to other distance metrics like CD L1 distance, leading to the generation of point cloud

data that more accurately conforms to actual shapes during the completion process.

Challenges and future application directions

Challenges and shortfalls

In the field of point cloud complementation and artifact recovery, GANet mainly faces the following challenges and shortcomings: firstly, it can only generate smooth surface structures and details, and cannot generate information such as internal structures and textures. Second, although GANet performs well in dealing with noise, it still needs to be further optimized when recovering tiny cracks or artifact surface details. In addition, although GANet performs well on the Terracotta Warriors dataset, its generalization ability

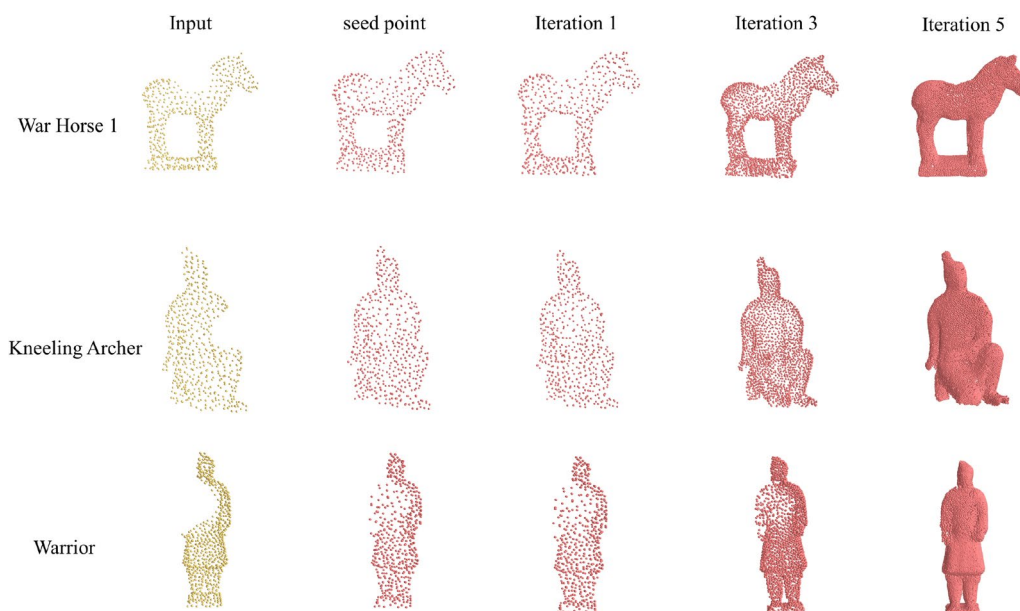


Fig. 9 Visualization results with different number of iterations on Terracotta Warriors dataset

Table 2 The point cloud completion of our method and the baseline method on the Terracotta Warriors dataset is expressed in terms of the L1 chamfer distance $\times 10^3$ (lower is better) per point

Methods	Average	Infantryman 1	War House 1	Kneeling Archer	Warrior	War House 2
PCN	13.15	11.45	16.87	15.75	11.14	11.17
PMP-Net	12.46	10.86	15.67	14.45	11.30	10.59
SA-Net	10.14	8.71	12.74	11.71	8.89	9.05
SPD	9.93	8.72	12.26	11.12	9.06	8.86
PointAttn	10.43	9.08	12.85	11.86	9.23	9.49
ours	8.23	7.09	10.09	9.26	7.41	7.58

Bold indicates that the minimum loss is achieved in all methods

to different artifact types is limited and needs to be adjusted according to the characteristics of different artifacts. Finally, high-resolution point clouds bring high computational complexity, and existing loss functions (e.g., CD and EMD) have limitations in capturing the global structure and computational efficiency.

Future application directions

Future research on GANet can focus on enhancing its generalization ability across different types of cultural heritage, allowing it to handle a variety of materials and structures in artifact restoration. Additionally, integrating multimodal data such as images, textures, and historical records will provide the model with richer

Table 3 Experimental results of the ablation of our proposed method on the Terracotta Warriors dataset. Denoted by the L1 chamfer distance $\times 10^3$ for each point (lower is better)

Methods	Average	Infantryman 1	War House 1	Kneeling Archer	Warrior	War House 2
GANet1	14.72	13.12	18.07	17.49	12.23	12.71
GANet2	9.51	8.28	11.47	10.29	8.68	8.85
GANet3	13.55	12.24	16.91	14.93	11.64	12.03
GANet4	8.45	7.37	10.18	9.51	7.46	7.74
ours	8.23	7.09	10.09	9.26	7.41	7.58

Bold indicates that the minimum loss is achieved in all methods

contextual information, improving its understanding of artifact structures and the restoration of fine details. Moreover, GANet-based technology can be used to develop automated tools for artifact restoration, helping conservation experts perform virtual repairs and analysis more efficiently and accurately, thereby advancing the practical application of digital heritage preservation.

Conclusion

This study aims to explore a new point cloud completion algorithm that can effectively perform virtual restoration of artifacts such as the Terracotta Warriors. We propose a novel graph attention mechanism that models point cloud data using graph structures, enabling it to extract features at different scales. This allows the algorithm to handle both global structures and local details during the completion process. The multi-scale feature extraction helps generate point clouds that are rich in detail and consistent overall. Additionally, we introduce an external attention mechanism during the seed point feature extraction phase, which allows for a more precise understanding of the contextual relationships within the point cloud, resulting in the generation of points that better match the structure of real objects during completion. This helps reduce errors and noise in the completed point cloud. Moreover, we propose a State Space Model-based Point Cloud Upsampling Layer, which, by establishing a state space model, can generate point cloud data that better conforms to the dynamic features of real objects during the completion process. By comparing with recent learning-based methods, this paper will demonstrate the advantages of our algorithm in terms of restoration effects and accuracy. Furthermore, the potential and challenges of applying this technology in the field of artifact conservation and restoration are discussed. Through this study, we hope to provide a new perspective and technical support for the digital preservation and restoration of cultural heritage, further advancing the development of the cultural heritage digitization field.

Author Contributions

Conceptualization: G.J., Z.Y. Methodology: G.J., Z.Y. Formal analysis: G.J., Z.Y. Investigation: G.J., G.G. Resources: G.G. Data curation: G.G. Writing-original draft preparation: G.J., Z.Y., Z.P. Writing-review and editing: G.J., Z.Y., S.Q. Supervision: G.J., Z.Y. All authors read and approved the final manuscript.

Funding

This research was funded by the National Natural Science Foundation of China (62271393), Xi'an Science and Technology Programme Demonstration Project on Science and Technology Innovation for Social Development (24SF5F0002), Northwest University 2024 Graduate Research and Innovation Program (CX2024202).

Availability of data and materials

The PCN dataset used in this paper are public datasets and can be shared; the Terracotta Warriors dataset is not publicly available at this time due to laboratory requirements. The code can be made public if needed.

Declarations

Competing interests

The authors declare that they have no Conflict of interest.

Received: 13 June 2024 Accepted: 9 October 2024

Published online: 19 November 2024

References

- Charles RQ, Su H, Kaichun M, Guibas LJ. Pointnet: deep learning on point sets for 3d classification and segmentation. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), pp. 2017:77–85
- Qi CR, Yi L, Su H, Guibas LJ. Pointnet++: deep hierarchical feature learning on point sets in a metric space. *Adv Neural Inf Process Syst* 2017;30
- Tchapmi LP, Kosaraju V, Rezaatofighi H, Reid I, Savarese S. Topnet: structural point cloud decoder. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR) 2019
- Yu L, Li X, Fu C-W, Cohen-Or D, Heng P-A. Pu-net: point cloud upsampling network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2018:2790–2799
- Groueix T, Fisher M, Kim VG, Russell BC, Aubry M. A papier-mâché approach to learning 3d surface generation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2018:216–224
- Liu M, Sheng L, Yang S, Shao J, Hu S-M. Morphing and sampling network for dense point cloud completion. *Proc AAAI Conf Artif Intell.* 2020;34:11596–603.
- Yuan W, Khot T, Held D, Mertz C, Hebert M. Pcn: point completion network. In: 2018 International conference on 3D vision (3DV), pp. 2018:728–737. IEEE
- Chang Y, Jung C, Xu Y. Finerpcn: high fidelity point cloud completion network using pointwise convolution. *Neurocomputing.* 2021;460:266–76.
- Xia Y, Xia Y, Li W, Song R, Cao K, Stilla U. Asfm-net: asymmetrical siamese feature matching network for point completion. In: Proceedings of the 29th ACM international conference on multimedia, pp. 2021:1938–1947
- Peng Y, Chang M, Wang Q, Qian Y, Zhang Y, Wei M, Liao X. Sparse-to-dense multi-encoder shape completion of unstructured point cloud. *IEEE Access.* 2020;8:30969–78.
- Tang J, Gong Z, Yi R, Xie Y, Ma L. Lake-net: topology-aware point cloud completion by localizing aligned keypoints. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 2022:1726–1735
- Wu H, Miao Y, Fu R. Point cloud completion using multiscale feature fusion and cross-regional attention. *Image Vis Comput.* 2021;111: 104193.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. *Adv Neural Inf Process Syst* 2017;30
- Zhao Y, Xie J, Qian J, Yang J. Pui-net: a point cloud upsampling and inpainting network. In: pattern recognition and computer vision: third Chinese conference, PRCV 2020, Nanjing, China, October 16–18, 2020, Proceedings, Part I 3, pp. 2020:328–340. Springer
- Li G, Chen Y, Cheng M, Wang C, Li J. N-dpc: dense 3D point cloud completion based on improved multi-stage network. In: Proceedings of the 2020 9th international conference on computing and pattern recognition, pp. 2020:274–279
- 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, pp. 2020:61–70
- Xiang P, Wen X, Liu Y-S, Cao Y-P, Wan P, Zheng W, Han Z. Snowflakenet: point cloud completion by snowflake point deconvolution with

- skip-transformer. In: Proceedings of the IEEE/CVF international conference on computer vision, pp. 2021:5499–5509
18. Yu X, Rao Y, Wang Z, Liu Z, Lu J, Zhou J. Pointtr: diverse point cloud completion with geometry-aware transformers. In: Proceedings of the IEEE/CVF international conference on computer vision, pp. 2021:12498–12507
 19. Wang J, Cui Y, Guo D, Li J, Liu Q, Shen C. Pointattn: you only need attention for point cloud completion. *Proc AAAI Conf Artif Intell.* 2024;38:5472–80.
 20. Phan AV, Le Nguyen M, Nguyen YLH, Bui LT. Dgcnn: a convolutional neural network over large-scale labeled graphs. *Neural Netw.* 2018;108:533–43.
 21. Zhang K, Hao M, Wang J, Silva CW, Fu C. Linked dynamic graph cnn: learning on point cloud via linking hierarchical features. *arXiv preprint arXiv:1904.10014* 2019
 22. Yang Y, Feng C, Shen Y, Tian D. Foldingnet: point cloud auto-encoder via deep grid deformation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2018:206–215
 23. Chen C, Liu D, Xu C, Truong T-K. Genecgan: a conditional generative adversarial network based on genetic tree for point cloud reconstruction. *Neurocomputing.* 2021;462:46–58.
 24. Shi J, Xu L, Heng L, Shen S. Graph-guided deformation for point cloud completion. *IEEE Robot Autom Lett.* 2021;6(4):7081–8.
 25. Zhang Y, Huang D, Wang Y. Pc-rgnn: Point cloud completion and graph neural network for 3d object detection. *Proc AAAI Conf Artif Intell.* 2021;35:3430–7.
 26. Moenning C, Dodgson NA. Fast marching farthest point sampling. Technical report, University of Cambridge, Computer laboratory 2003
 27. Hamilton JD. State-space models. *Handb Econom.* 1994;4:3039–80.
 28. Butt MA, Maragos P. Optimum design of chamfer distance transforms. *IEEE Trans Image Process.* 1998;7(10):1477–84.
 29. Pele O, Werman M. Fast and robust earth mover's distances. In: 2009 IEEE 12th international conference on computer vision, pp. 2009:460–467. IEEE
 30. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* 2014
 31. Chang AX, Funkhouser T, Guibas L, Hanrahan P, Huang Q, Li Z, Savarese S, Savva M, Song S, Su H, et al. Shapenet: an information-rich 3d model repository. *arXiv preprint arXiv:1512.03012* 2015
 32. Xie H, Yao H, Zhou S, Mao J, Zhang S, Sun W. Grnet: Gridding residual network for dense point cloud completion. In: European conference on computer vision, pp. 2020:365–381. Springer
 33. Wen X, Xiang P, Han Z, Cao Y-P, Wan P, Zheng W, Liu Y-S. Pmp-net: Point cloud completion by learning multi-step point moving paths. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 2021:7443–7452
 34. Wang X, Ang Jr MH, Lee GH. Cascaded refinement network for point cloud completion. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 2020:790–799
 35. Zhang W, Yan Q, Xiao C. Detail preserved point cloud completion via separated feature aggregation. In: Computer Vision—ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16, pp. 2020:512–528. Springer
 36. Wen X, Li T, Han Z, Liu Y-S. Point cloud completion by skip-attention network with hierarchical folding. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 2020:1939–1948

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.