

RESEARCH

Open Access



# DGPCD: a benchmark for typical official-style Dougong in ancient Chinese wooden architecture

Caochenyu Zhou<sup>1,2</sup>, Youqiang Dong<sup>1,2</sup> and Miaole Hou<sup>1,2\*</sup>

## Abstract

Dougong, a distinctive component of ancient Chinese wooden architecture, holds significant importance for the preservation and restoration of such structures. In particular, the northern official-style buildings represent the pinnacle of ancient Chinese construction techniques. In the realm of cultural heritage preservation, the application of deep learning has gradually expanded, demonstrating remarkable effectiveness. Point cloud serving as a crucial source for Dougong, encapsulates various information, enabling support for tasks like Dougong point cloud classification and completion. The quality of Dougong datasets directly impacts the outcomes of DNNs (deep neural networks), as they serve as the foundational data support for these models. The typical official-style Dougong, with its standardized and repetitive structural patterns, is highly suitable for training DNNs to accurately recognize and analyze these complex architectural elements. However, due to the inherent characteristics of Dougong, such as coplanarity and occlusion, acquiring point cloud data is challenging, resulting in poor data quality and organizational difficulties. To address this issue, our study adopts a multi-source data fusion approach to tackle the challenges of insufficient data quantity and poor data quality. Further, through data augmentation, we enhance the dataset's volume and generalize its characteristics. This effort culminates in the creation of the typical official-style Dougong Point Cloud Dataset (DG Dataset), poised to support deep learning tasks related to Dougong scenarios.

**Keywords** Dougong, Point cloud, Deep learning, Ancient wooden architecture

## Introduction

Ancient Chinese wooden architecture is a valuable cultural heritage that has stood the test of millennia, currently facing challenges such as natural aging and risk of disasters, necessitating ongoing conservation efforts. Traditional methods of architectural conservation are no longer sufficient to meet the increasing demands for protection, leading to a consensus among scholars on

the incorporation of digitalization in heritage conservation practices [1, 2]. Point clouds, emerging as the third type of spatial data following maps and images, play a crucial role by providing comprehensive and high-precision representations of the complex forms and structures of ancient buildings [3], thus becoming a primary data source for the preservation of these wooden structures.

The raw point cloud data necessitate processing tasks such as data cleaning and semantic segmentation to enable effective monitoring and management of the conservation and restoration of ancient Chinese wooden architecture. Advances in the computing field have demonstrated that deep learning offers highly automated data processing and precise recognition capabilities, which are extensively applicable in the cultural heritage context.

\*Correspondence:

Miaole Hou  
houmiaole@bucea.edu.cn

<sup>1</sup> School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 102616, China

<sup>2</sup> Key Laboratory for Architectural Heritage Fine Reconstruction and Health Monitoring, Beijing 100044, 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.

In recent years, research has primarily focused on enhancing the efficiency and accuracy of point cloud data processing in cultural heritage scenarios. The success of frameworks like PointNet, PointNet++ and others [4–9] have provided robust methods for handling point cloud data. Studies employing these frameworks have facilitated the denoising and completion of point cloud data by extracting features from well-maintained datasets of ancient Chinese wooden structures, training DNNs that improve the data's completeness and accuracy. These methodologies not only enhance data integrity but also provide a more reliable basis for conservation and restoration efforts, further advancing the preservation of ancient Chinese wooden architecture. In image-based studies, object detection and semantic segmentation of ancient Chinese wooden architecture are quite mature, enabling automatic recognition and segmentation of structures, damages, and other critical features, significantly more efficient than manual methods [10–13]. Additionally, Pierdicca [14] using DGCCN on the ArCH [16] dataset for semantic segmentation of European historical architectures like churches and monasteries offer insights into the application of point cloud segmentation in historic buildings.

Currently, DNNs are well-established and have been further applied in the domain of ancient Chinese wooden architecture preservation [15], yet challenges remain in dataset construction. The components of such architecture are diverse, with variations across different periods, necessitating extensive data to enhance generalizability. Some studies have constructed datasets for cultural heritage scenarios [16–19]. However, the breadth and depth of their applications have been adversely affected by the complexity of cultural heritage environments. The complex structures of ancient Chinese wooden architecture make high-quality data acquisition challenging, often failing to meet dataset standards. Furthermore, the diversity in data collection methods for this architecture leads to inconsistencies in datasets, complicating the construction and integration process. These issues result in scarce datasets, making training for ancient Chinese wooden architecture scenes exceptionally difficult and the lack of benchmarks complicates the evaluation of different models, limiting the development of deep learning in this field.

Given the severity and impact of these issues, this study aims to develop a method for constructing datasets of typical official-style ancient Chinese wooden architecture point clouds that is efficient and reliable, enabling the rapid, batch generation of high-quality point cloud data. These data will support DNNs crucial for tasks such as data processing and semantic segmentation. By creating a representative and diverse dataset of ancient Chinese

wooden architecture point clouds, this research seeks to address the current lack of benchmark datasets, fostering research and development in this field and offering more effective and sustainable solutions for the preservation and restoration of ancient Chinese wooden architecture. This paper selects the Dougong, a typical architectural component serving as a connection between the roof and columns characterized by traditional mortise and tenon joints, to establish a Dougong point cloud dataset Benchmark (DGPCD). This methodology is extendable to other components of ancient Chinese wooden architecture, aiding in the construction of benchmarks for these components.

The main contributions of this paper are: (i) The provision of a typical official-style Dougong point cloud benchmark, addressing the dataset gaps for this component and offering a method that can be applied to other ancient Chinese wooden architecture components, serving as a reference for constructing other component benchmarks; (ii) The development of an efficient and reliable method for rapidly generating high-quality point cloud data of ancient Chinese wooden architecture, ensuring data consistency by integrating data from multiple sources and enhancing the robustness and generalizability of the dataset across various point cloud scenarios.

## Related works

The existing datasets exhibit diverse sources, data formats, and organizational structures, typically organized based on specific deep learning tasks. Three-dimensional point cloud datasets are primarily acquired through RGB-D camera shooting, 3D laser scanning, and computer simulation.

### Acquisition via RGB-D camera shooting

RGB-D camera captures images integrating RGB and depth information, facilitating point cloud generation through depth reconstruction. For instance, ScanNet [20] employs RGB-D cameras to scan indoor environments, collecting data from 1513 scenes, encompassing 21 object categories, and comprising 2.5 million RGB-D images. NYU-Depth V2 [21] records video sequences of various indoor scenes using Kinect, including 464 scenes with 894 object categories and 1449 annotated RGB-D images, making it suitable for indoor scene semantic segmentation tasks.

### Acquisition via laser scanner scanning

Direct acquisition of point cloud data is achieved through 3D laser scanning. S3DIS [22] utilizes 3D laser scanning to obtain point cloud data for six areas, 13 semantic elements, and 11 scene types, providing commonly used data for indoor scene understanding and 3D semantic

segmentation. Semantic3D [23] employs static outdoor ground scanning to collect point clouds from various urban scenes like churches, streets, and railways, as well as natural scenes, facilitating outdoor natural scene semantic segmentation.

#### Acquisition via computer synthesizing

Computer simulation represents a rapid and efficient approach to generating point clouds. ModelNet [24] samples point clouds from over 10,000 CAD models, covering 40 categories such as airplanes, cars, and chairs, supporting tasks like point cloud classification and completion. ShapeNet [24], a large-scale 3D shape dataset with rich annotations covering 55 common object categories and around 51,300 unique 3D models, also obtains point clouds through CAD model sampling. It differs by having extensive annotations, enabling tasks like object point cloud part segmentation.

#### Acquisition via multiple sources

Due to the complexity of architectural heritage scenes and the diversity of tasks, point cloud datasets designed for them typically have one or more data sources. For instance, the WHU-TLS dataset [17–19] captures data from 1740 different environments using terrestrial laser scanners, specifically the VZ-400 and Leica P40, collecting over 311 million 3D points that include architectural heritage contexts, supporting tasks such as point cloud registration, semantic segmentation, and instance segmentation. Similarly, the ArCH (Architectural Cultural Heritage) dataset [16], which represents the inaugural benchmark for semantic segmentation of point clouds in historical architectural heritage, acquires data using diverse methods like RGB imaging, 3D laser scanning, and drone imaging. It includes 27 point cloud datasets, adhering to Level of Detail (LOD) standards, with 17 datasets accurately annotated for 10 categories, covering features like arches, columns, floors, doors/windows, walls, edges, stairs, archways, roofs, and other categories.

#### Dataset construction

In this study, we introduce a Dougong point cloud dataset suitable for a variety of deep learning applications. The construction of the dataset adhered to a methodical protocol encompassing data acquisition, preprocessing, annotation, and validation. Data were compiled from diversified sources, including real-world data acquisition and computer simulation. The raw data from these sources were subjected to preprocessing steps aimed at standardizing formats, eliminating noise, and rectifying inconsistencies. These steps comprised data cleaning, normalization, transformation, and feature extraction, ensuring uniformity and compatibility throughout the

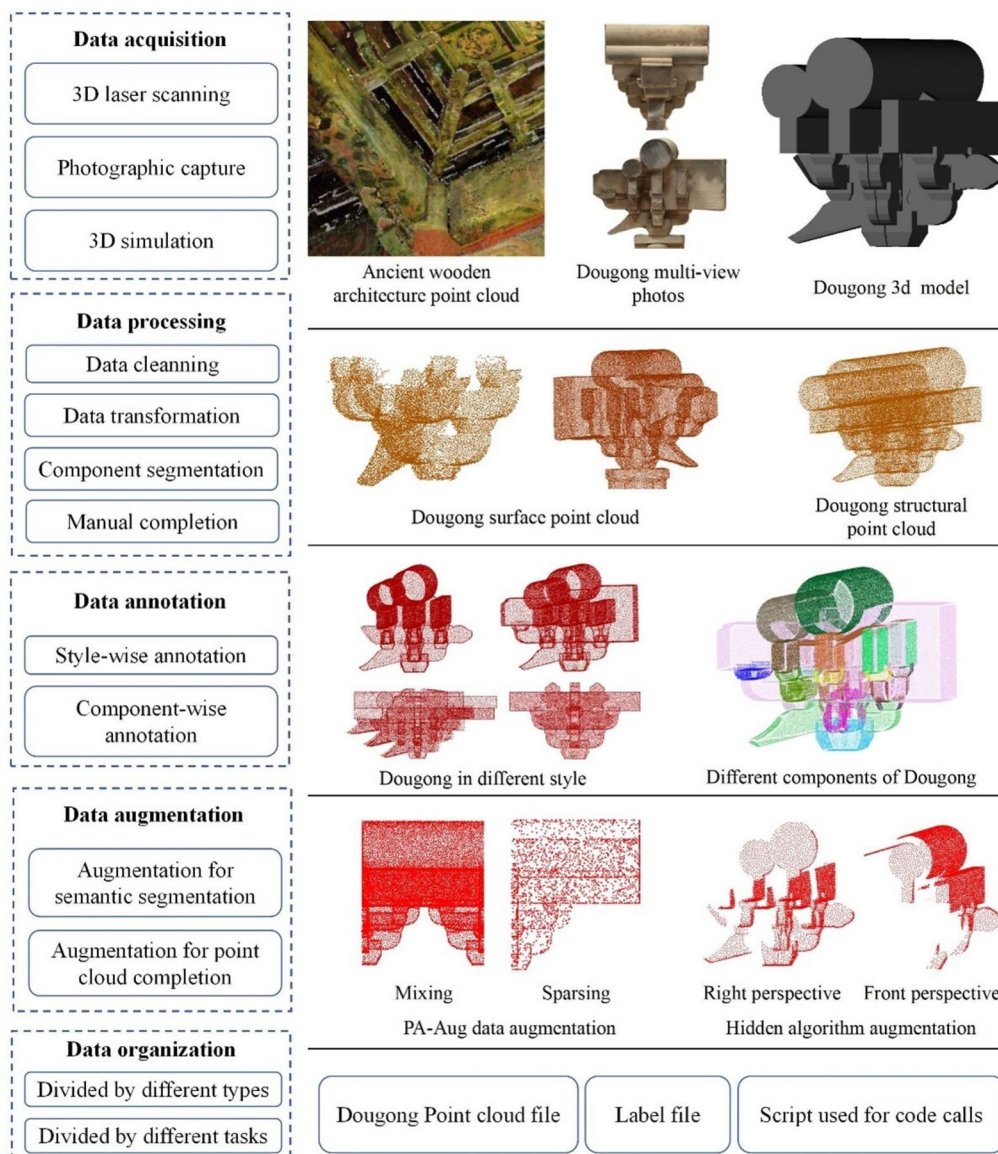
dataset. Annotations were applied to imbue the data with ground truth labels, critical for supervised learning tasks, involving both algorithmic assistance and manual input from domain experts. The dataset was subjected to stringent validation processes to confirm its quality, integrity, and applicability to deep learning research. These validation measures included cross-validation, inter-rater reliability assessments, and evaluations of algorithmic performance.

During the data acquisition phase, three methodologies were utilized: 3D laser scanning, multi-view photographic capture, and computer simulation. The latter two methods serve as complements to the former, offering benefits such as speed, cost-effectiveness, and suitability for conversion into point cloud data. The objective of the data processing phase is to transform the acquired data into Dougong point cloud data, ensuring compliance with the designated dataset format. Data annotation involves assigning semantic labels to various components of the Dougong within the point cloud data, primarily through manual efforts. We use data augmentation methods to enhance the interpretability of data and enhance the performance of the data set in deep learning tasks. Ultimately, the dataset is systematically organized to facilitate access to various task-specific subsets (Fig. 1).

#### Data acquisition

The primary data sources for this dataset were 3D laser scanning, photographic capture, and 3D simulation. These methods were selected due to their prevalence and relevance in the research and conservation of ancient Chinese wooden architecture. Utilizing these data sources allows for comprehensive coverage across different application scenarios, ensuring optimal performance of the trained models in real-world applications. This strategic integration of various data types significantly enhances the models' generalization capabilities and robustness across diverse operational scenarios.

3D laser scanning and photographic capture was primarily conducted on ancient architectural complexes located in Beijing, Shanxi, Liaoning, and other regions across China. These areas are noted for their rich historical and cultural heritage, exhibiting a strong continuity of historical artifacts. The Dougong structures in these regions, characterized by traditional craftsmanship, reflect a long-term evolutionary process. These structures encompass diverse types from various dynasties, thus effectively representing the principal types of Dougong found in China. Additionally, the dry climate and relatively isolated geographical locations of these regions have contributed to the minimal impact of warfare, preserving the ancient Chinese wooden architecture and mitigating common issues such as decay and cracking in



**Fig. 1** The main process of constructing a Dougong dataset

Dougong. This has enhanced the quality of the data collected. Simulation data were primarily generated through forward modeling based on the construction methods and component dimensions of Dougong as documented in “Yingzao Fashi”, “Qing Dynasty Construction Regulations”, “Dougong”, some authoritative texts covering various dynasties and forms of Dougong (Table 1 and Fig. 2).

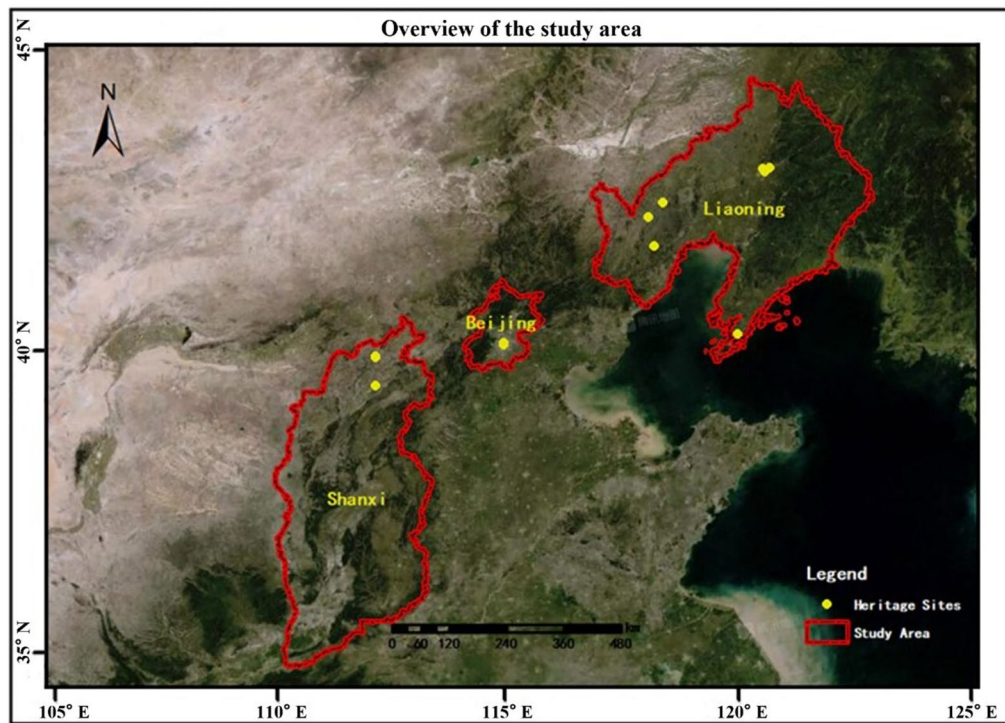
The 3D laser scanning equipment used is the Faro Focus 350 3D scanner. The data acquisition process follows the technical specifications outlined in the local standard DB11/T 1796–2020 for the collection of 3D information in ancient Chinese wooden architecture in Beijing.

Photographic capture utilizes the Sony a7 III. This data collection method is a quick, simple, and cost-effective approach. Since the Sony a7 III cannot record depth values, Dougong photographs are captured from multiple views. These photos are then processed using software to calculate depth information, ultimately generating a point cloud.

3D simulation involves referencing ancient Chinese wooden architectural drawings of Dougong, converting them into CAD drawings, modeling each component using modeling software, and assembling them into a complete 3D Dougong model. Various Dougong from different time periods and classifications are

**Table 1** Data source and quantity

Collection method	Data sources	Scene Types	Era	Number of Dougong collected
3D laser scanning, photography	Beijing	Palace	Ming and Qing	350
3D laser scanning, photography	Beijing	Palace	Ming and Qing	50
3D laser scanning, photography	Datong, Shanxi Province	Temple	Liao	100
Photography	Datong, Shanxi Province	Temple	Tang	45
Photography	Datong, Shanxi Province	Temple	Yuan	34
3D laser scanning, photography	Yingxian, Shanxi Province	Tomb	Liao	200
Photography	Shenyang, Liaoning Province	Palace	Qing	50
Photography	Shenyang, Liaoning Province	Tomb	Qing	50
Photography	Shenyang, Liaoning Province	Tomb	Qing	30
3D laser scanning, photography	Chaoyang, Liaoning Province	Temple	Qing	10
3D laser scanning, photography	Chaoyang, Liaoning Province	Temple	Qing	8
3D laser scanning, photography	Chaoyang, Liaoning Province	Temple	Yuan	12
3D laser scanning, photography	Liaoning Province	Temple	Qing	25
3D modeling	< Dougong >	–	Ming and Qing	20
3D modeling	< Qing Dynasty Construction Regulations >	–	Qing	10
3D modeling	< Yingzao Fashi >	–	Qing	5
3D modeling	< Construction Methods >	–	Song	5



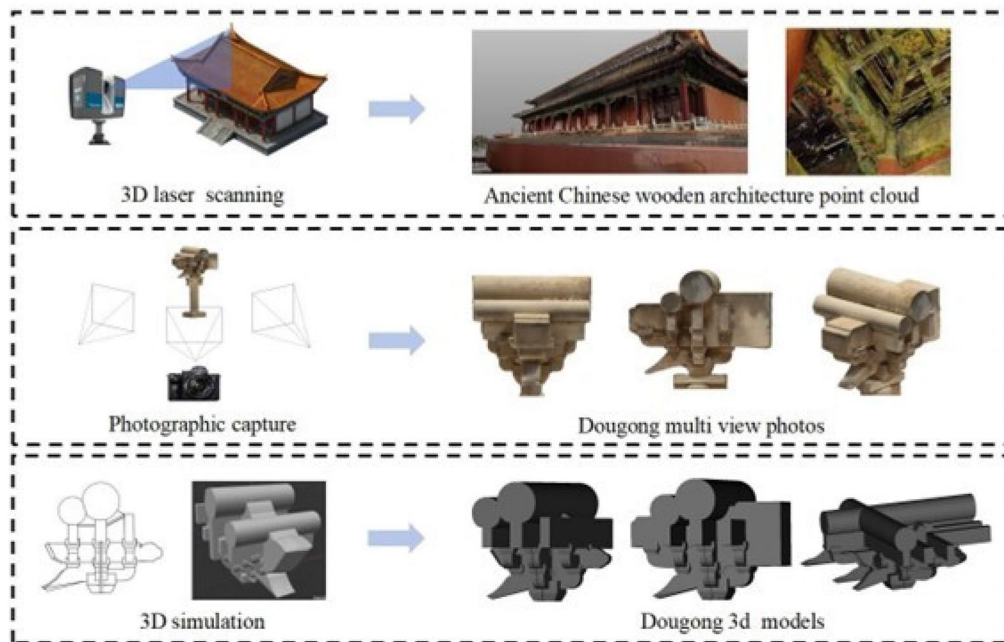
**Fig. 2** Field data collection distribution map

selected for modeling using 3ds Max software (Fig. 3 and Table 2).

**Data processing**



The primary objective of the data processing phase is to transform multi-source, multi-format data into

a unified, high-quality point cloud dataset. Given the diverse requirements in practical applications, the processing of Dougong point cloud data varies significantly. For some analyses or visualization tasks, only the surface point cloud data of Dougong is necessary, whereas for more comprehensive analyses, point cloud



**Fig. 3** Data collection results

**Table 2** Schematic diagram of data acquisition

Equipment name	R&D institutions	Parameters and value	Material object	
SONY Alpha 7III	Sony	Product type Full charge working time Image sensor Lens parameters Photo size Focal length		
		Professional full frame mirrorless camera 3 h 36×24 mm full frame CMOS with effective pixels of 40 million Horizontal: 74°–19°20' Vertical: 53°–13° 8192×5464 24–70 mm		
FAROs350	Faro	Product type Scanning element Pixel Optical resolution Maximum resolution Scanning range Scanning speed		
		3D scanner Laser module Up to 70 M color pixel 600×1200dpi Vertical resolution: 0.009° (40,960 3D pixels at 360°) Horizontal resolution: 0.009° (40,960 3D pixels at 360°) Scanning at a maximum distance of 330 m under direct sunlight 976,000 points/second		

data incorporating structural information is essential. Acknowledging this diversity, the dataset is designed to accommodate various needs by including both complete surface point cloud data and enriched point cloud data with structural details. This dual provision enhances the flexibility in data utilization, catering to both straightforward surface analyses and more intricate structural investigations.

#### Surface point cloud data processing

The surface point cloud of Dougong was derived from 3D laser scanning and multi-view photographic data. For the 3D laser scanning collected from historical architectural sites, a series of processing steps were implemented. Initially, multi-station data registration was conducted using the software SCENE. This was followed by the application of Cloud Compare for data cleaning and denoising to remove artifacts related to obstructions and reflections typically found in ancient Chinese wooden architecture. Manual segmentation was then performed to isolate the Dougong from the broader scene. Finally, surface reconstruction was applied to the isolated Dougong point cloud, addressing gaps from obstructions to yield a complete surface point cloud.

#### Structural point cloud data processing

Following the processing of the Dougong surface point cloud, semantic segmentation was conducted to differentiate the outer surfaces of various components. Expert knowledge was subsequently utilized to manually complete the point clouds for each component, aiming to restore their full morphology by addressing areas obscured from view. After this, the enhanced point clouds of individual components were integrated to generate a comprehensive Dougong point cloud, incorporating detailed internal structural information.

Owing to the inefficiency of manual segmentation, regional clustering [25] was implemented for the semantic segmentation of the Dougong point cloud. This technique involved the voxelization of the surface point cloud of Dougong and subsequently merging voxel data under multiple constraints to classify component point clouds. Regional clustering principally depends on smoothness and coplanarity constraints. Here, smoothness is defined by the formation of a continuous surface between two adjoining point cloud segments, characterized by a minimal angular discrepancy between their respective normal vectors  $\vec{x}_1$  and  $\vec{x}_2$ . This relationship is quantitatively described by the following formula:

$$\Delta Angle = \langle \vec{x}_1, \vec{x}_2 \rangle = \cos^{-1} (\vec{x}_1 \cdot \vec{x}_2) \quad (1)$$

Coplanarity describes the geometric relationship where two surfaces are aligned in the same plane, typically assessed by evaluating the proximity of point cloud segments along their normal vectors. A shorter distance between these segments suggests a higher degree of coplanarity. The calculation formula is as follows, where  $\vec{x}_1^C$  represents the distance from the suggested plane to the origin. Utilizing the centroid  $\vec{x}_1$ , the normal vector  $\vec{s}_1$ , and the distance from the centroid to the origin, the plane can be represented by the following expression:

$$\Delta Coplanarity = \left( \vec{x}_1^C - \vec{x}_{12}^D \right)^2 + \left( \vec{x}_2^C - \vec{x}_{21}^D \right)^2 \quad (2)$$

where  $\vec{x}_{12}^D$  can be represented by the following formula:

$$\vec{x}_{12}^D = \langle \vec{s}_1, \vec{s}_2 \rangle \quad (3)$$

Meanwhile, to enhance the accuracy of semantic segmentation, we incorporated prior knowledge to guide the merging of geometric primitives. We established the association between geometric primitives and components, with the extracted geometric primitives represented as follows:

$$S_i = \{type, norm, loc, area\} \quad (4)$$

$S_i$  represents the extracted geometric primitive surface, *type* denotes the type of the geometric primitive surface, *norm* indicates the orientation of the geometric primitive surface, *loc* represents the position of the geometric primitive surface, and *area* signifies the size of the geometric primitive. Based on the knowledge of the position, orientation, and size within the Dougong components, it is possible to determine whether the extracted geometric primitives are shared.

After segmenting the various components of the Dougong, their symmetrical and geometric properties were used to complete the point cloud data. Dougong is notably symmetrical, both within and between its components. This symmetry guided the restoration of the component point clouds. Additionally, by analyzing the dimensional and geometric relationships among the components, we estimated the sizes and shapes of some components based on the known dimensions of others. The integration of symmetry and size inference allowed for more accurate completion of the point clouds, especially for components where symmetry was insufficient. This method enabled the reconstruction of complete Dougong point cloud structures.

#### Data annotation

The Dougong point cloud contains three-dimensional coordinates (x, y, z) and color information (r, g, b). Due

to the complexity and disorder of color information in Dougong, extracting meaningful features is challenging and generally not utilized. When inputting the three-dimensional coordinate values into DNNs, the model can only comprehend the positional information of the point cloud, lacking an understanding of its overall category and component composition. This limitation restricts the original Dougong point cloud to tasks such as point cloud completion and renders it unsuitable for tasks such as type recognition and component segmentation. Therefore, this dataset employs manual annotation to define task-specific labels for various deep learning tasks, expanding the applicability of the dataset (Fig. 4).

**Dougong label definition**

The dataset defines two forms of Dougong labels, Dougong type labels and Dougong component labels, with each point cloud possessing both types of labels. Dougong type labels are annotated based on the type of Dougong, allowing DNNs to understand the overall type information of the Dougong point cloud for tasks such as Dougong type recognition. Dougong component labels are annotated based on different component types that constitute the Dougong, distinguishing point clouds of different components within a single Dougong model. This enables DNNs to comprehend the composition of the Dougong point cloud, making it suitable for Dougong component segmentation tasks.

*Style-wise annotation* In terms of type labels, references from books “Yingzao Fashi” and “Qing Gongbu engineering practices” were consulted. Based on the different positions of Dougong in ancient Chinese wooden architecture, they were categorized into four major types, Pingshenke

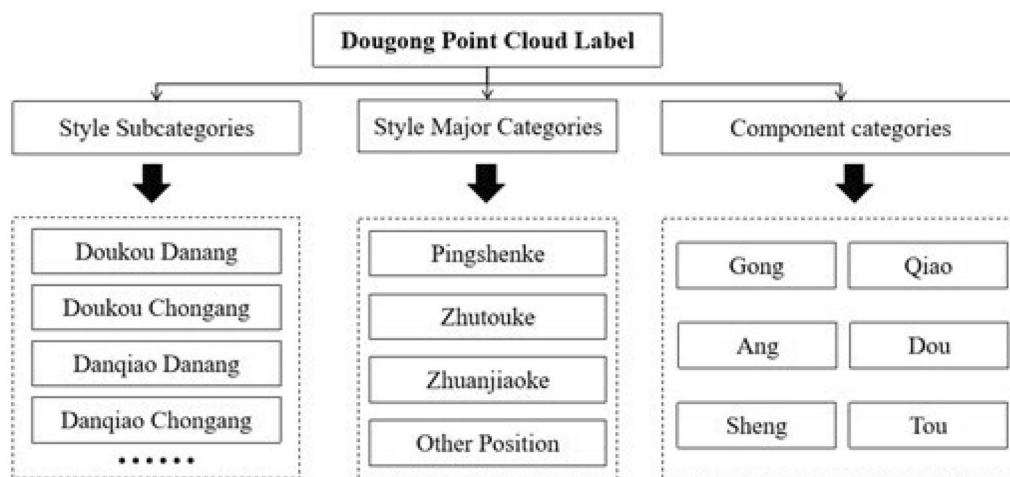
Dougong, Zhutouke Dougong, Jiaoke Dougong, and other position Dougong. Additionally, considering stylistic variations, Dougong was further classified into 28 subcategories (Fig. 5).

*Component-wise annotation* Dougong components are diverse, including Dou, Qiao, Gong, Sheng, Ang, Shua Tou, Cheng Tou. Depending on their position, size, and form, they are further divided into several subtypes. To prevent overly fine-grained divisions leading to overfitting in DNNs, the dataset categorizes Dougong component labels into six classes: Gong, Qiao, Ang, Dou, Sheng, Tou. Component labels provide high-resolution cognitive information about the composition of Dougong, enabling DNNs to deeply understand and differentiate point clouds of various components that constitute a complete Dougong (Fig. 6).

**Dougong label annotation**

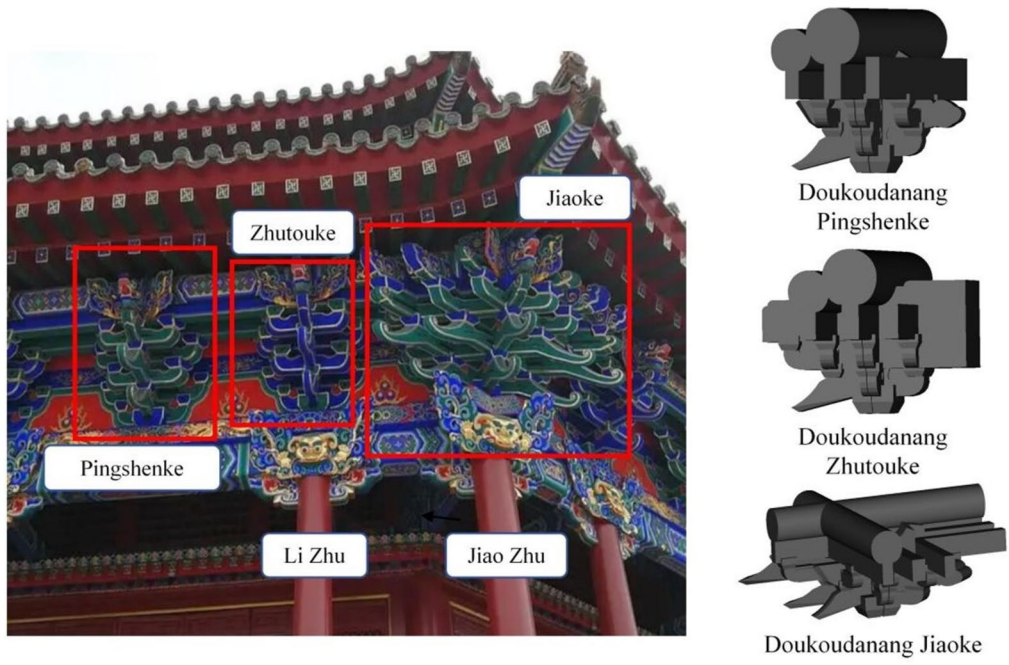
The annotation method for Dougong point cloud data is manual annotation. For Dougong type annotation, different categories are assigned corresponding numbers, and folders with the same numbers are created. Dougong of the respective types are placed in the corresponding folders, allowing the DNNs to understand the Dougong types.

The original point cloud is stored in txt format, where each line represents a point cloud with six columns of data, consisting of (x, y, z) three-dimensional coordinates and (r, g, b) intensity values. After removing color information, each line of the point cloud retains only three columns of three-dimensional coordinate values. We annotate each point to indicate the Dougong component it belongs to by adding a fourth column. Different

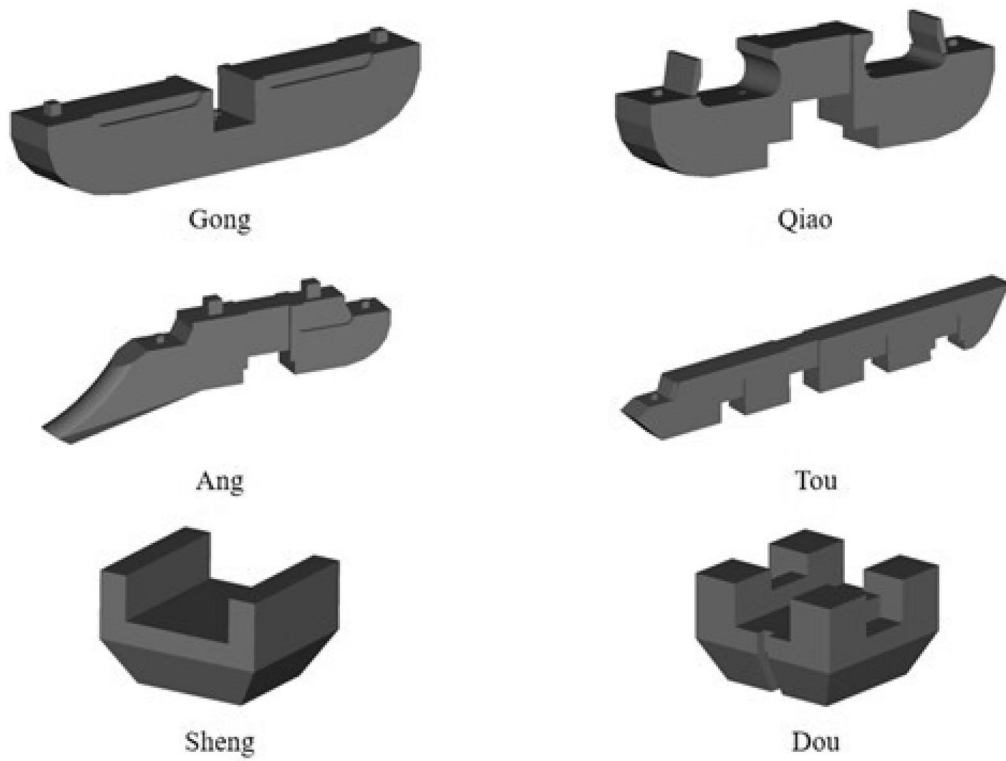


**Fig. 4** Labels contained in Dougong dataset





**Fig. 5** Dougong in different positions and its subcategories



**Fig. 6** Main types of Dougong components

values are used to distinguish different categories. For example, the value for Gong is 1, and for Qiao, it is 2. The specific process involves using the point cloud clipping tool in Cloud Compare software to extract point clouds of various components. Then, corresponding label values are added based on the label categories, completing the annotation of component point clouds. The DNNs can understand the component point cloud types based on these label values.

### Data augmentation

#### Data augmentation to enhance semantic information

The limited amount of data on Dougong, coupled with its intricate internal structure, poses a significant challenge for DNNs when applied directly to semantic segmentation tasks. Therefore, employing appropriate methods for data augmentation can facilitate the DNNs' understanding of the semantic attributes of Dougong. The PA-AUG (Part-Aware Data Augmentation) method [26] has proven to be effective in enhancing semantic information. This method significantly increases the data volume, enhances structural information, and improves the performance of the dataset in tasks such as object detection and type recognition networks. Inspired by this method, our dataset adopts a block-wise augmentation approach tailored to the characteristics of Dougong for data augmentation.

The Dougong point cloud  $PC$  can be represented as the union of foreground points  $FP$  and background points  $BP$ , as shown in Eq. (5). In Eq. (6),  $B$  represents the points within the partitioned bounding box,  $N$  is the number of partitioned boxes,  $P$  denotes internal points within the partition, and  $T$  indicates the position index of the partition within the box. The augmented foreground points  $FP_{aug}$  can be expressed as Eq. (7), where the bounding box and partition are denoted by  $B$  and  $P$ , respectively.

$$PC = FP \cup BP \quad (5)$$

$$FP = \cup_{i=1}^N B^{(i)}, B^{(i)} = \cup_{j=1}^T P_j^{(i)} \quad (6)$$

$$FP_{aug} = \cup_{i=1}^N \widehat{B}^{(i)}, \widehat{B}^{(i)} = \cup_{j=1}^T \widehat{P}_j^{(i)} \quad (7)$$

The Dougong point cloud is divided into four partitions, and each partition undergoes one of the following transformations such as random transformation, fusion, down sampling, or interpolation.

The “random transformation” refers to the replacement of points in the partition with points from the same category and partition position. Since Dougong within the same category often exhibit low similarity, applying transformations to large categories may lead

to data contamination. Therefore, transformations are exclusively performed on Dougong within small categories. “Mixing” involves blending points in the partition with points from the same category and partition position. Similar to transformations, fusion operations are restricted to small categories. “Down sampling” refers to the down sampling of points in the partition, with down sampling ratios of 0.8 and 0.5 selected for the Dougong point cloud. Down sampling simulates the phenomenon where the density of scanned points decreases with increasing distance during 3D laser scanning, enhancing the generalization of the data. “Noising” entails adding random noise to the partition, with noise points added at a proportion of 0.05 of the points within the partition. Interpolation simulates noise generated during the 3D laser scanning process due to occlusion and other factors, further improving data generalization.

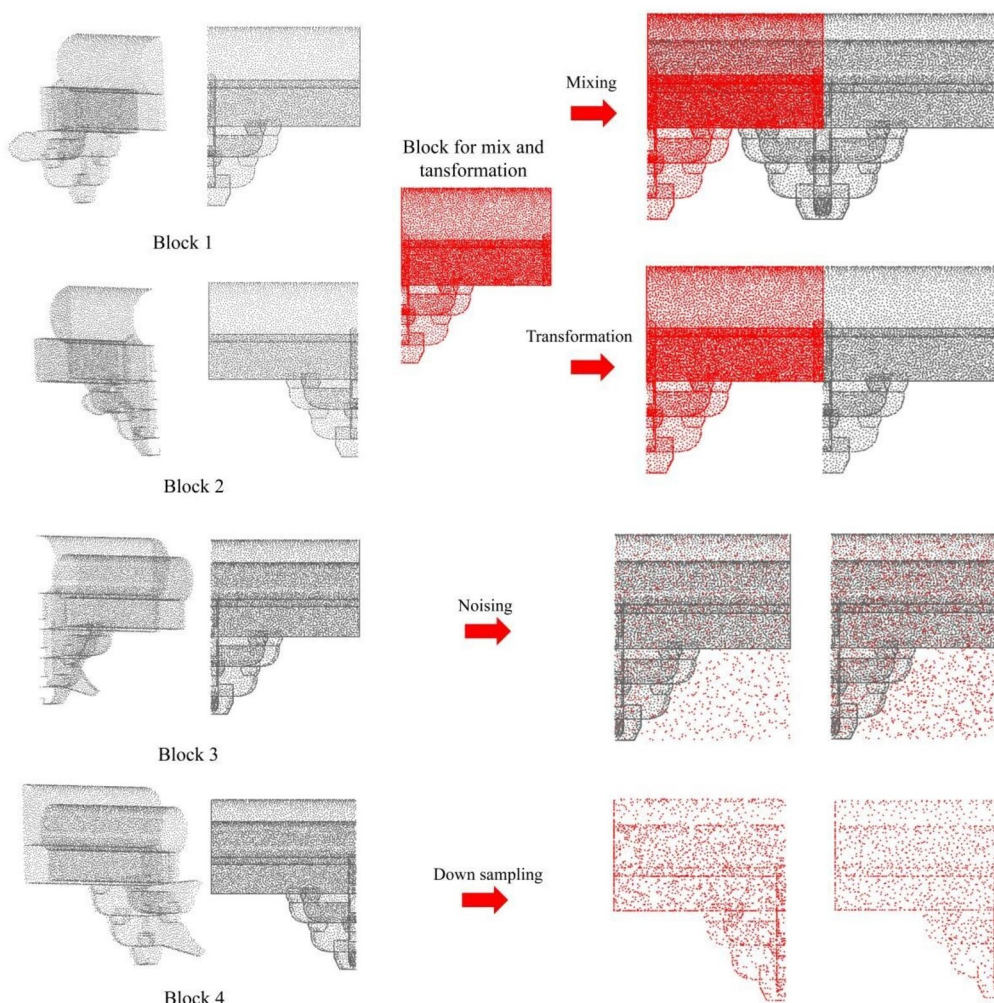
This data augmentation method tailored for Dougong point clouds effectively enhances DNNs' understanding of internal structural relationships, aligning with the goal of strengthening the model's capability to learn Dougong's internal structures and point cloud labels (Fig. 7).

#### Data augmentation to enhance geometric information

In the task of point cloud completion, DNNs need to learn from the missing points at various angles to simulate the limitations imposed by sensors under real-world conditions. This enables a better understanding of the overall structure of the data and improves the quality and accuracy of the completion results. We employed the hidden surface algorithm to downsample the Dougong sets, simulating the point cloud missing caused by occlusion during 3D laser scanning from different perspectives.

The Z-Buffer algorithm [27], a typical hidden surface algorithm, is employed for occlusion culling. This method maintains a Z-Buffer depth buffer storing depth information for each pixel, allowing the detection of occlusion situations. In point cloud occlusion culling, the three-dimensional point cloud is first projected onto a two-dimensional plane by transforming each point's three-dimensional coordinates (X, Y, Z) into two-dimensional coordinates (X, Y). Here, X and Y represent the pixel position on the plane, and Z represents the point's depth or distance. After projection onto the 2D plane, the Z-Buffer uses depth values to determine point overlap and coverage.

Utilizing the Z-Buffer algorithm simulates data gaps resulting from scans of Dougong using a 3D laser scanner in a real-world scenario. This simulated data serves as adversarial data to train DNNs to handle point cloud gaps from different perspectives. Following the requirements outlined in the Beijing Local Standard DB11/T 1796—2020 for the collection of 3D information in heritage



**Fig. 7** Block point cloud data augmentation diagram

architecture, the simulation process replicates scanning scenarios from different stations. By setting projection points and judging the depth values of each point, overlapping and covered points are removed, achieving occlusion culling on the complete Dougong point cloud data. This approach enhances the DNNs’ understanding of authentic scanned Dougong point cloud features, thereby improving the generalization of Dougong point cloud data (Fig. 8).

**Data organization**

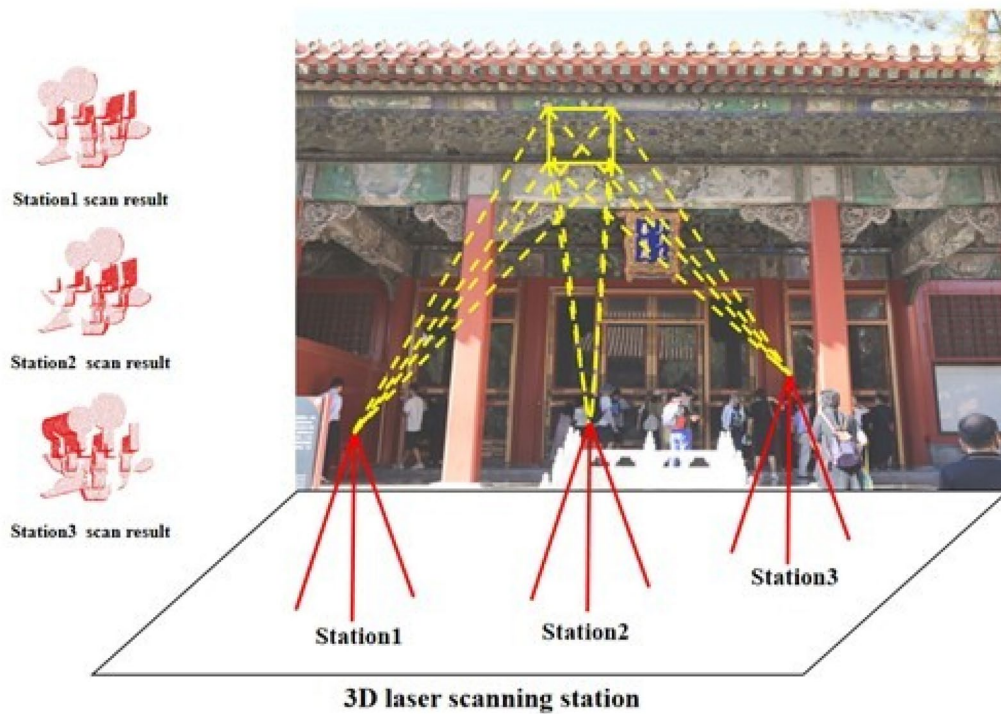
During data organization, we grouped Dougong components of the same type from different periods into a single category. Although these components exhibit slight variations in form, they possess considerable geometric and structural similarities when examined at the level of point cloud data. Organizing them together significantly enhances the DNNs’ generalization capability

for this category of Dougong, thereby improving task performance.

To further refine our data organization strategy, we differentiated between the surface point clouds and the structured point clouds of the Dougong components. This separation is designed to tailor the data architecture to support distinct tasks effectively. Additionally, label data are stored independently from the point cloud data to enhance data handling efficiency and facilitate more streamlined access during different analytical processes.

**Experiment and analysis**

To validate the usability of the Dougong point cloud data in this dataset, several typical DNNs were selected in this study. The dataset was used for training, and the performance of this dataset was tested in Dougong point cloud type classification and Dougong point cloud completion tasks. Specifically, PointNet and PointNet++ were



**Fig. 8** Hidden algorithm simulation scenario

chosen for Dougong point cloud type classification tasks, while PCN, 3D-Capsule, and PF-Net were selected for Dougong point cloud completion tasks (Table 3).

**Dougong point cloud classification task**

This study selected four DNNs, namely PointNet (vanilla), PointNet, PointNet++, and PointNet++ (with normal), without making modifications to the models. The evaluation focused on the application of the Dougong point cloud dataset in type classification. In this experiment, annotated Dougong point cloud data with labeled types were input into DNNs to train the models. Subsequently, the Dougong point clouds used for testing were input, and discriminative labels were obtained. The

type classification accuracy was determined by comparing the obtained labels with the ground truth labels.

PointNet was the first DNN to directly process 3D point clouds through convolution, demonstrating stability in extracting point set features even in the presence of fluctuations, noise, or missing data. PointNet++ is an improved version of PointNet, offering better extraction of fine local features in point clouds. Selecting these two DNNs allows the validation of the representation of overall and local features in the Dougong dataset while testing their performance in Dougong point cloud type classification tasks. PointNet, compared to PointNet (vanilla), and PointNet++ (with normal), compared to PointNet++, increase the complexity of the DNNs, enabling the extraction of more features. This helps verify

**Table 3** The DNNs selected for the experiment

Name	Year	Task	Number of citations
PointNet [4]	2017	Point cloud classification and segmentation	13,715
PointNet++ [5]	2018	Point cloud classification and segmentation	9616
PCN [10]	2018	Point cloud completion	735
3D-Capsule [28]	2019	Point cloud completion	373
PF-Net [9]	2021	Point cloud completion	345

whether the DNNs extract information from the point cloud models.

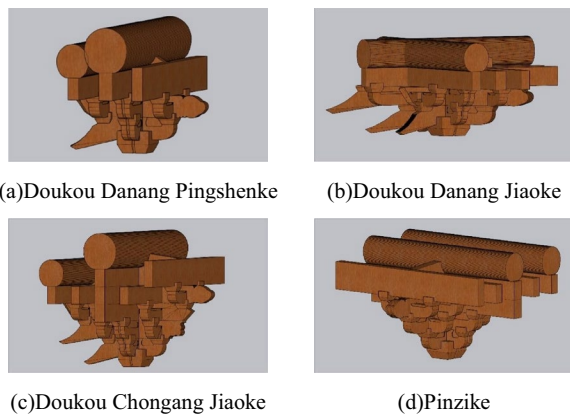
In this experiment, Dougong point clouds of four categories, including the simulated and real Dougong point clouds of Doukou Danang Pingshenke, Doukou Danang Jiaoke, Doukou ChongAng Jiaoke, and the Pinzike, were selected from the dataset. For each Dougong category, 150 simulated point cloud models and 100 real point cloud models were chosen as the training set, while 50 simulated point cloud models and 40 real point cloud models were used as the test set. To assess the performance of the point cloud dataset in representing Dougong features under different sampling point quantities and to evaluate the classification effect of DNNs on

Dougong point clouds, the input point cloud quantities were varied. Specifically, each point cloud model was sampled with 2048 and 4096 points, and separate training was conducted to obtain classification results (Fig. 9).

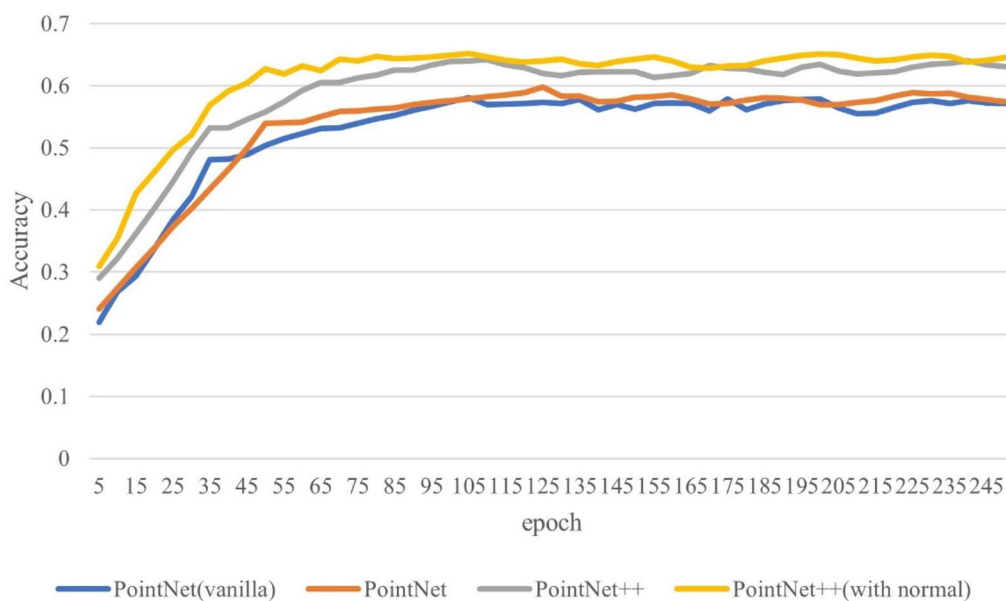
For the configuration of training parameters, the model training utilized a learning rate of 0.001, a batch size of 12, and 250 iterations during the training process. The cross-entropy loss function was employed, monitoring the model's performance on the validation set to prevent overfitting. The classification results were evaluated using accuracy as the metric to measure the overall classification accuracy on the test set, where higher accuracy values indicated more precise classification results.

Throughout the training process, the classification accuracy of each model increased with the growth of training batches. The training effectiveness was notably apparent in the first 100 epochs, with a significant increase in accuracy observed in the initial 50 training epochs. Beyond 100 epochs, the accuracy tended to stabilize. This outcome suggests that DNNs effectively learned from the input Dougong point cloud data and successfully classified the Dougong point cloud types. The successful classification results under the PointNet network indicate that Dougong point cloud data can effectively represent overall features, while the successful classification results under the PointNet++ suggest that Dougong point cloud data can effectively capture local features (Fig. 10).

The final results for the Dougong point cloud type classification task for the categories of “Doukou Danang Pingshenke,” “Doukou Danang Jiaoke,” “Doukou



**Fig. 9** Classification task Dougong type



**Fig. 10** Changes in accuracy during different DNNs training epoch

Chongang Pingshenke,” and “Pinzike” are presented in the table. Under the same input conditions, DNNs capable of extracting more features demonstrate better recognition performance. Among them, the simplest model, PointNet (vanilla), exhibits the lowest performance, while the most complex model, PointNet++ (with normal), achieves the best results. Across different inputs, except for PointNet (vanilla), datasets with more points show improved classification performance, with training on 4096 points outperforming the results obtained with 2048 points. These training outcomes align with those from publicly available datasets, indicating the good trainability of the Dougong dataset (Fig. 11).

**Dougong point cloud completion task**

This paper selects three common point cloud completion DNNs, namely PCN, 3D-Capsule, and PF-Net, without making any modifications, and evaluates their application on the Dougong point cloud dataset for point cloud completion. In this experiment, only Dougong point clouds with three-dimensional coordinates are input into the DNNs. The models learn the distribution characteristics of Dougong point clouds, and then input incomplete Dougong point clouds to estimate the complete point clouds. The effectiveness of point cloud completion is determined by comparing the distances between all real points and predicted points in the missing part of the point cloud.

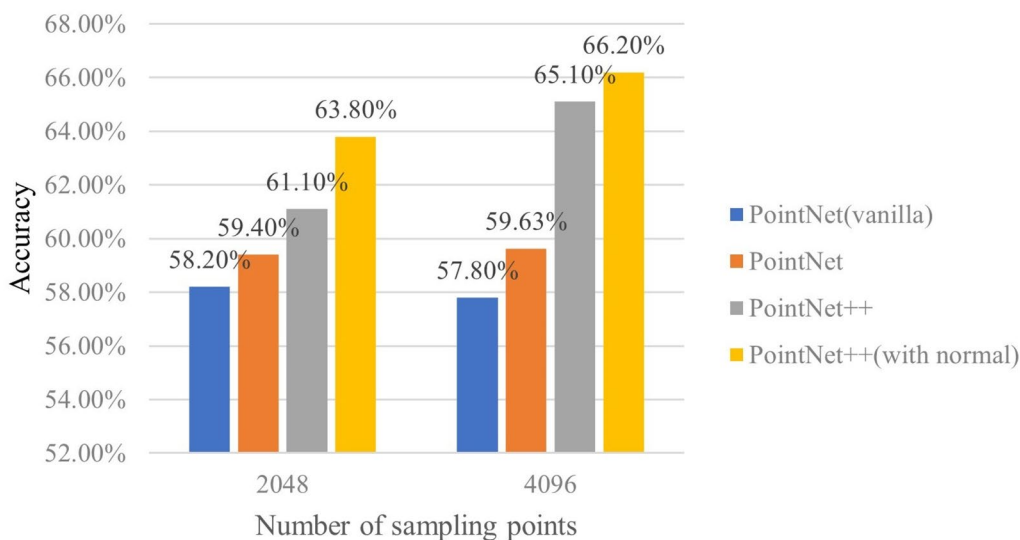
PCN, 3D-Capsule, and PF-Net are three types of DNNs that directly perform completion operations on the original point clouds without any structural assumptions (such as symmetry) or annotations on the underlying shapes (such as semantic classes). The input of

unannotated point clouds from the dataset reflects the learning status of the distribution characteristics of Dougong point clouds. If the Dougong dataset has learnable capabilities, then the trained models can predict and generate missing parts of point clouds from partial Dougong point clouds.

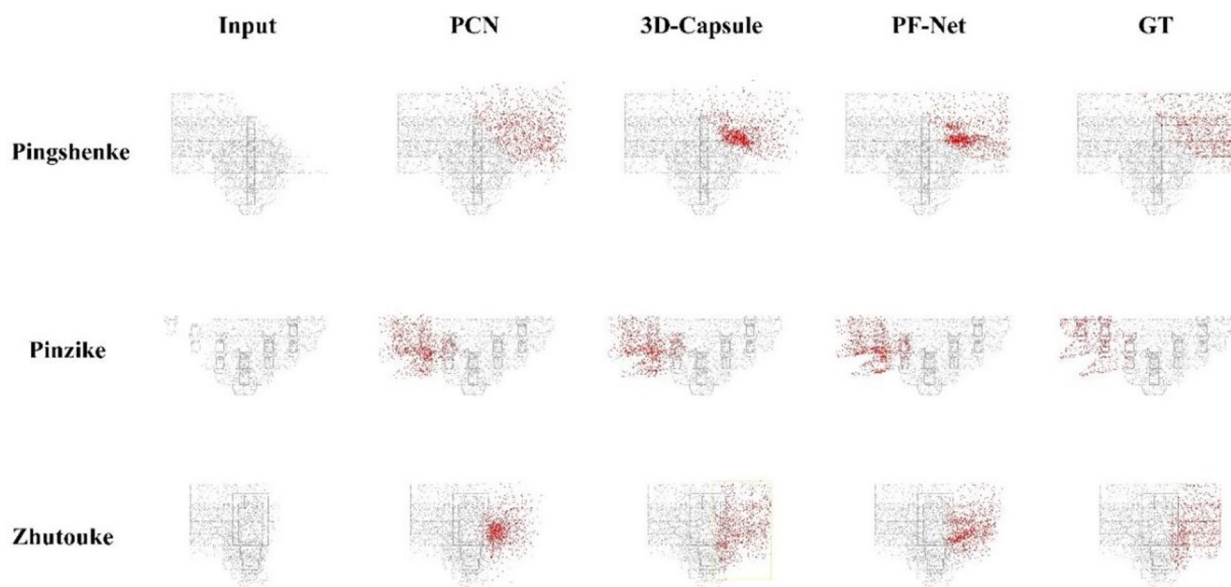
For this experiment, Dougong point clouds from three categories—Pingshenke, Zhutouke, and Pinzike are selected from the dataset. Each category has 250 simulated point cloud models and 50 real point cloud models for training, and 50 simulated point cloud models and 10 real point cloud models for testing. Each point cloud model is sampled with 4096 points, and the missing points are set to 1024 and 2048 points, respectively, for training. The model training uses a learning rate of 0.001, batch size of 24, and 400 iterations to prevent overfitting during the training process. Multi-stage completion loss functions and adversarial loss functions are employed.

After training with incomplete point clouds, the decoder outputs point clouds. PCN and 3D-Capsule output the entire Dougong point cloud, while PF-Net outputs the missing part of the Dougong point cloud. For comparison, the output point clouds of PCN and 3D-Capsule are truncated to facilitate a comparison of completion effectiveness (Fig. 12).

Observing the completion effects reveals variations in the learning performance of DNNs for different cases of missing Dougong point cloud data. In instances where the missing part of the Dougong point cloud has a relatively simple structure, all three DNNs exhibit good learning effects on the point cloud distribution. Notably, PCN produces the least satisfactory results for completing the point cloud of internal components, as it fails to



**Fig. 11** Average accuracy of classification tasks for different network types



**Fig. 12** Dougong point cloud completion results

accurately represent the structure and instead fills the missing region with scattered points. This is attributed to the lower complexity of the model, which struggles to predict complex point cloud structures effectively. Both 3D-Capsule and PF-Net achieve a certain degree of recovery for the Dougong structural point cloud, with PF-Net outperforming 3D-Capsule in recovery results. This suggests that DNNs capable of extracting more features can better restore point cloud structures.

To evaluate the completion effects for different training results, errors in Pred  $\rightarrow$  GT (Prediction to Ground Truth) and GT  $\rightarrow$  Pred (Ground Truth to Prediction) are utilized. The Pred  $\rightarrow$  GT error calculates the average squared distance from each point in the prediction to its closest point in the ground truth. Conversely, the GT  $\rightarrow$  Pred error calculates the average squared distance from each point in the ground truth to its closest point in the prediction (Table 4).

PCN exhibits a relatively poor completion performance. Possible reasons for this include the insufficient learning capacity of the PCN model for the complex geometric structures and local features of Dougong. 3D-Capsule performs better than PCN, but still faces certain difficulties in the Dougong completion task. PF-Net, being a more complex model, demonstrates significant advantages in handling Dougong point clouds, showcasing a better understanding and reconstruction of the geometric shape of Dougong. This aligns with the experimental results of the three DNNs on other public datasets, indicating that the Dougong point cloud dataset has good usability in point cloud completion tasks.

## Conclusion

In this study, we successfully proposed a rapid and high-quality method for constructing a benchmark for ancient Chinese wooden architectural components known as Dougong. This method, integrating 3D laser scanning, photography, and model data, significantly enhanced the generalizability of the dataset, addressing the challenges of difficult data collection and low data quality typically encountered in ancient architectural datasets. Additionally, we employed an algorithm-assisted manual annotation technique to overcome the complexities of semantic annotation for ancient architecture, thereby improving the efficiency and accuracy of semantic labeling. Targeted data augmentation was also implemented, ensuring robust performance across various tasks.

Given the absence of comparable existing datasets, we selected several widely used DNNs to conduct benchmark tests on our dataset. The results demonstrated good performance across these networks, validating its suitability and effectiveness for supporting deep learning tasks.

Furthermore, our dataset not only provides a reliable technical solution for the digital preservation of ancient architectural components but also offers valuable data resources for the restoration and study of ancient buildings. In the fields of cultural heritage preservation and ancient Chinese wooden architecture, this dataset has broad application prospects, including but not limited to automated damage detection, structural analysis, and 3D reconstruction of historical buildings.

**Table 4** Comparison of different DNNs completion results

Net	Dougong type	Missing points (piece)	Pred → GT (Unit distance*10 <sup>-3</sup> )	GT → Pred (Unit distance*10 <sup>-3</sup> )
PCN	Pingshenke	1024	4.263	3.878
		2048	6.125	5.365
	Zhutouke	1024	5.189	4.765
		2048	6.329	6.174
	Pinzike	1024	5.979	5.137
		2048	7.694	7.560
3D-Capsule	Pingshenke	1024	3.024	2.913
		2048	4.308	3.947
	Zhutouke	1024	4.970	4.828
		2048	5.601	5.132
	Pinzike	1024	4.698	4.175
		2048	5.031	4.803
PF-Net	Pingshenke	1024	2.156	1.901
		2048	2.981	2.841
	Zhutouke	1024	2.307	2.193
		2048	2.990	2.866
	Pinzike	1024	2.223	2.213
		2048	2.684	2.467

Moreover, the innovative methods and technologies used in the development of this dataset offer new tools and methodologies for researchers in similar fields, especially when dealing with objects that feature complex geometries and detailed characteristics. Future work may explore the application of these techniques to different types of cultural heritage objects, further advancing the digitalization and preservation of global cultural heritage.

In summary, our research not only successfully constructed a high-quality Dougong dataset but also demonstrated its extensive application potential through benchmark testing on multiple DNNs. We anticipate that this dataset construction method will play a significant role in future technological developments, academic research, and the preservation of cultural heritage.

#### Acknowledgements

The authors would like to acknowledge and thank the Palace Museum for its support of this work. The authors would also like to thank Dr. Su Yang who always give the authors the novel suggestion.

#### Author contributions

Conceptualization, Y.D and C.Z.; methodology, C.Z.; software, C.Z.; resources, Y.D.; writing—original draft preparation, C.Z.; writing—review and editing, Y.D, C.Z and M.H.; supervision, Y.D.; project administration, C.Z, Y.D. and M.H.; funding acquisition, M.H. All authors have read and agreed to the published version of the manuscript.

#### Funding

This research was funded by National Key Research and Development Program of China, Grant number 2022YFF0904300, National Natural Science Foundation of China, Grant number 42171356, 42171444, and 42301516, Beijing Municipal Education Commission- Municipal Education Commission Joint Fund Project, Grant number KZ202110016021, Research Project of

Beijing Municipal Education Commission—General Project of Science and Technology Plan, Grant number KM202110016005 and Beijing University of Civil Engineering and Architecture Special funds for basic scientific research business expenses of municipal universities, Grant number X20043.

#### Availability of data and materials

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to confidentiality.

#### Declarations

#### Competing interests

The authors declare that they have no competing interests.

Received: 26 January 2024 Accepted: 21 May 2024

Published online: 13 June 2024

#### References

- Liu J, Wu ZK. Rule-based generation of ancient Chinese architecture from the song dynasty. *J Comput Cult Herit (JOCCH)*. 2015;9(2):1–22. <https://doi.org/10.1145/2835495>.
- Hu Q, Wang S, Fu C, Ai M, Yu D, Wang W. Fine surveying and 3D modeling approach for wooden ancient architecture via multiple laser scanner integration. *Remote Sens*. 2016;8(4):270. <https://doi.org/10.3390/rs8040270>.
- Bisheng Y, Zhen D. Progress and perspective of point cloud intelligence. *Acta Geodaetica et Cartographica Sinica*. 2019;48(12):1575. <https://doi.org/10.1080/10095020.2023.2175478>.
- Qi CR, Su H, Mo K, Guibas LJ. Pointnet: deep learning on point sets for 3d classification and segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017;652–660. <https://doi.org/10.1109/cvpr.2017.16>.
- Qi CR, Yi L, Su H, Guibas LJ. Pointnet++: deep hierarchical feature learning on point sets in a metric space. *Adv Neural Inf Proc Syst*. 2017. <https://doi.org/10.48550/arXiv.1706.02413>.



6. Chen X, Ma H, Wan J, Li B, Xia T. Multi-view 3d object detection network for autonomous driving. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2017:1907–1915. <https://doi.org/10.48550/arXiv.1611.07759>.
7. Zhou Y, Tuzel O. Voxnet: end-to-end learning for point cloud based 3d object detection. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018:4490–4499. <https://doi.org/10.48550/arXiv.1711.06396>.
8. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431–3440. <https://doi.org/10.48550/arXiv.1605.06211>.
9. Yuan W, Khot T, Held D, Held D, Mertz C, Hebert M. Pcn: point completion network. 2018 international conference on 3D vision (3DV). IEEE. 2018:728–737. <https://doi.org/10.1109/3DV.2018.00088>.
10. Huang Z, Yu Y, Xu J, Ni F, Le X. Pf-net: Point fractal network for 3d point cloud completion. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020:7662–7670. <https://doi.org/10.1109/CVPR42600.2020.00768>.
11. Lin Y, Nie Z, Ma H. Structural damage detection with automatic feature-extraction through deep learning. *Comput-Aided Civil Infrastruct Eng.* 2017;32(12):1025–46. <https://doi.org/10.1111/mice.12313>.
12. Jian M, Weidong Y, Guoqi L. research on crack detection method of wooden ancient building based on YOLO v5. *J Shenyang Jianzhu Univ (Natural Science)*. 2021;37(05):927–34.
13. Malinverni ES, Pierdicca R, Paolanti M, Martini M, Morbidoni C, Matrone F, Lingua A. Deep learning for semantic segmentation of 3D point cloud. *Int Arch Photogramm Remote Sens Spat Inf Sci.* 2019;42:735–42. <https://doi.org/10.5194/isprs-archives-XLII-2-W15-735-2019>.
14. Pathak R, Saini A, Wadhwa A, Sharma H, Sangwan D. An object detection approach for detecting damages in heritage sites using 3-D point clouds and 2-D visual data. *J Cult Herit.* 2021;48:74–82. <https://doi.org/10.5194/isprs-archives-XLII-2-W15-735-2019>.
15. Pierdicca R, Paolanti M, Matrone F, Martini M, Morbidoni C, Malinverni ES, Frontoni E, Lingua AM. Point cloud semantic segmentation using a deep learning framework for cultural heritage. *Remote Sens.* 2020;12(6):1005. <https://doi.org/10.3390/rs12061005>.
16. Sun M. Completion and structural analysis of point clouds for wooden ancient building components. Beijing University of Civil Engineering and Architecture. 2023. <https://doi.org/10.26943/d.cnki.gbjzc.2022.000414>.
17. Matrone F, Lingua A, Pierdicca R, Malinverni ES, Paolanti M, Grilli E, Remondino F. A benchmark for large-scale heritage point cloud semantic segmentation. *Int Arch Photogramm Remote Sens Spat Inf Sci.* 2020;43:1419–26. <https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-1419-2020>.
18. Dong Z, Yang B, Liu Y, Liang F, Li B, Zang Y. A novel binary shape context for 3D local surface description. *ISPRS J Photogramm Remote Sens.* 2017;130:431–52. <https://doi.org/10.1016/j.isprsjprs.2017.06.012>.
19. Dong Z, Yang B, Liang F, et al. Hierarchical registration of unordered TLS point clouds based on binary shape context descriptor. *ISPRS J Photogramm Remote Sens.* 2018;144:61–79. <https://doi.org/10.1016/j.isprsjprs.2018.06.018>.
20. Dong Z, Liang F, Yang B, Xu Y, Zang Y, Li J, et al. Registration of large-scale terrestrial laser scanner point clouds: a review and benchmark. *ISPRS J Photogramm Remote Sens.* 2020;163:327–42. <https://doi.org/10.1016/j.isprsjprs.2020.03.013>.
21. Dai A, Chang A X, Savva M, Halber M, Funkhouser T, Niessner M. Scannet: Richly-annotated 3d reconstructions of indoor scenes. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017:5828–5839. <https://doi.org/10.48550/arXiv.1702.04405>.
22. Silberman N, Hoiem D, Kohli P, Fergus R. Indoor segmentation and support inference from rgbd images. *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Proceedings, Part V 12.* Springer Berlin Heidelberg. 2012:746–760. [https://doi.org/10.1007/978-3-642-33715-4\\_54](https://doi.org/10.1007/978-3-642-33715-4_54).
23. Armeni I, Sener O, Zamir A R, Jiang H, Brilakis I, Fischer M, Savarese S. 3d semantic parsing of large-scale indoor spaces. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016:1534–1543. <https://doi.org/10.1109/cvpr.2016.170>.
24. Hackel T, Savinov N, Ladicky L, Wegner J D, Schindler K, Pollefeys M. Semantic3d. net: a new large-scale point cloud classification benchmark. arXiv preprint arXiv. 2017;1704.03847. <https://doi.org/10.48550/arXiv.1704.03847>.
25. Wu Z, Song S, Khosla A, et al. 3d shapenets: a deep representation for volumetric shapes. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015:1912–1920. <https://doi.org/10.1109/CVPR.2015.7298801>.
26. Hao W, Dong Y, Hou M. Primitive Segmentation of Dougong Components Based on Regional Clustering. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* 2023;109-114. <https://doi.org/10.5194/isprs-annals-X-M-1-2023-109-2023>.
27. Choi J, Song Y, Kwak N. Part-aware data augmentation for 3d object detection in point cloud. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE. 2021;3391–3397. <https://doi.org/10.1109/IROS51168.2021.9635887>.
28. Greene N, Kass M, Miller G. Hierarchical Z-buffer visibility. Proceedings of the 20th annual conference on Computer graphics and interactive techniques. 1993;231–238. <https://doi.org/10.1145/166117.166147>.
29. Zhao Y, Birdal T, Deng H, Tombari F. 3D point capsule networks. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019;1009–1018. <https://doi.org/10.1109/cvpr.2019.00110>.

### Publisher's Note

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