Automated heritage building component recognition and modelling based on local features

Pang, Bo; Yang, Jian; Xia, Tian; Zhang, Anshan; Zhang, Kai; Xu, Qingfeng; Wang, Feiliang

DOI
10.1016/j.culher.2024.12.006

Publication date
2025

Document Version
Final published version

Published in
Journal of Cultural Heritage

Citation (APA)
Pang, B., Yang, J., Xia, T., Zhang, A., Zhang, K., Xu, Q., & Wang, F. (2025). Automated heritage building component recognition and modelling based on local features. *Journal of Cultural Heritage*, *71*, 252-264. https://doi.org/10.1016/j.culher.2024.12.006

Journal of Cultural Heritage 71 (2025) 252–264

Contents lists available at ScienceDirect

Journal of Cultural Heritage

journal homepage: www.elsevier.com/locate/culher

Original article

Automated heritage building component recognition and modelling based on local features

Bo Pang[a,c], Jian Yang[a,b,c,d,*], Tian Xia[e], Anshan Zhang[a,c], Kai Zhang[a,c], Qingfeng Xu[b], Feiliang Wang[a,c,*]

[a] Shanghai Key Laboratory for Digital Maintenance of Buildings and Infrastructure, School of Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, PR China
[b] Shanghai Key Laboratory of Engineering Structure Safety, SRIBS, Shanghai 200032, PR China
[c] State Key Laboratory of Ocean Engineering, Shanghai Jiao Tong University, Shanghai 200240, PR China
[d] School of Civil Engineering, University of Birmingham, Birmingham B15 2TT, UK
[e] Delft University of Technology, Delft, the Netherlands

ARTICLE INFO

Article history:
Received 11 June 2024
Accepted 9 December 2024

Keywords:
Heritage building
Point cloud
Semantic segmentation
Local descriptor
Automated modelling

ABSTRACT

The maintenance of buildings, underpinned by the digital twin technique, becomes integral to heritage conservation efforts. To achieve efficient modelling with minimal manual intervention, automated component recognition based on semantic segmentation of point clouds is imperative. Confronted by the challenges of the paucity of requisite datasets and the inherent geometric diversity of historical buildings, a two-step strategy including feature extraction and classification is proposed. First, an improved SHOT descriptor is proposed to extract discriminative features by defining a specific local reference system and concatenating support fields at different scales. The extracted features are then classified with a learning-based network, avoiding a feature learning process that relies on sufficient data. Experiments on real-world heritage point clouds yield 93.7% accuracy and an 80.0% mean-intersection-over-union (mIoU) when descriptors with radii of 0.3 m and 0.9 m are combined, surpassing computationally expensive deep learning networks and data-intensive unsupervised learning. A slight decrease in segmentation performance with random removal of points indicates the high robustness of the proposed method against data missing and sampling density changes. Additionally, a geometric modelling process with an error of less than 10% is introduced to achieve a direct transition from point cloud to model, contributing to the establishment of digital twins for heritage structures.

1. Introduction

Historical buildings, integral facets of immovable cultural heritage, serve as witness to urban development. The arisen cultural tourism, anchored in these legacies, has progressively become a local economic booster. However, the inexorable natural aging of building materials and the inevitable man-made vandalism accompanying heritage activities have engendered irreversible losses in structural performance. Unwarranted risks surface [1,2], necessitating a call for meticulous structural health monitoring, maintenance, and risk management of heritages.

The development of digital techniques, with heightened adaptability, has facilitated the implementation of information management frameworks in the scenario of heritage maintenance. Non-contact measurement and monitoring of the deformation are often necessary to avoid further damage, highlighting the importance of structural reconstruction and component recognition [3–6]. Notably, frameworks like HBIM (heritage building information model) [7–9] furnish semantic and geometric information, catering to the needs of managers, researchers, conservators, and restorers [10]. The established concept of digital twins (DT) also demonstrates the pivotal role of virtual entities in interactive risk management [11,12]. The three-dimensional (3D) point cloud obtained from laser scanning [13,14] or oblique photography [15,16] often serves as the primary data source for extracting geometric primitives and building structures, facilitating parametric modelling in such systems. Preprocessing of point clouds, a precursor to digital reconstruction, has become a critical prerequisite, imposing constraints on efficient and reliable modelling solutions. The semantic segmentation of point cloud is a technology that links the

* Corresponding authors.
E-mail addresses: j.yang.1@sjtu.edu.cn (J. Yang), wongfayeleung@sjtu.edu.cn (F. Wang).

https://doi.org/10.1016/j.culher.2024.12.006
1296-2074/

collected data with semantic labels used in 3D scene understanding and cognition [17]. The commonly used feature extraction-based geometric modelling, whether a developed representation algorithm [18,19] or mature commercial solutions [20], partially relies on manual processing by engineers. Therefore, an automated processing strategy for semantic segmentation and modelling is required to circumvent labour-intensive tasks and expedite the process. We attempt to achieve an automated scheme to obtain geometric parameters of heritage components required by the digital model.

The typical procedure to interpret building geometry is to acquire certain preprocessing steps, called semantic modelling, to extract a set of features which is then fed to a classification model that computes the most adequate class label [21]. The recognised components, presented by class labels, can be further geometrically modelled to provide necessary information for monitoring and maintenance activities. Features for semantic modelling could be acquired from context or end-to-end learning, resulting in different approaches.

The segmentation and classification are often sequential when features are obtained from contextual information of point clusters (i.e., the relationship between adjacent points). According to algorithms applied to divide the point cloud into clusters, noteworthy applications encompass region-growing, model fitting, and clustering-based algorithms. The basic conception of the region-growing algorithm is to merge spatial clusters exhibiting similarity in designated features. This similarity, gauged through criteria like normal vector, curvature, point-to-adjustment plane distance, or point-to-candidate point distance [22], facilitates the detection and segmentation of basic geometrical elements in heritage structures [23]. The features used for similarity identification can also be obtained from the deep learning framework, to realize unsupervised segmentation [24]. Model fitting aligns raw point clouds with geometric models utilizing sampling methods such as the Hough transform (HT) and random sample consensus (RANSAC). This yields the identification of geometric entities like lines, planes [25–27], cylinders [28], and spheres. However, computational expenses can burgeon when dealing with complex shapes in large datasets of practical heritages [29]. Model fitting is also the primary feature extraction tool in geometric modelling after segmentation. Clustering algorithms, such as K-means and mean shift, segment regions based on diverse geometric, colour, or reflection intensity features [30]. Similar to the region-growing approach, the deficiency for unsupervised clustering is the difficulty in achieving results that fully match practical component classification through suitable predefined parameters. By introducing predetermined human expertise, such as design regulations, primitive features extracted from algorithms could be aggregated to higher abstraction levels successively until ideal segmentation is accomplished [31]. Some hybrid approaches with region-growing and fitting algorithms were proposed to realize the segmentation of structural components in heritages [32].

In recent years, the emergence of end-to-end neural networks tailored for processing point clouds has garnered attention. The feature extraction in the usual workflow can be automatically implemented using a deep learning architecture, achieving a direct mapping from raw point data to semantic labels. Applied in processing bridge, building, and heritage point clouds, these networks are categorized into 3D convolution on voxels [33], 2D convolution on projected images [34,35], and feature learning on point sets [36,37]. After the learning-based local feature extraction, voxels or points are classified, and semantic annotations are given. PointNet++, a representative model, employs a hierarchical feature learning method, concurrently capturing local and global features. As mentioned above, the learned features can also be applied to local point clusters for classification, thus forming a hybrid method (such as the combination of convolution and region-growing in GrowSP [24]). Matching this approach, some attempts were made to identify meaningful features to recognize historical architectural elements [38]. Despite the efficacy of neural networks, the reliance on substantial labelled data poses a limitation in the early stages of training for learning-based models.

In the realm of heritage projects, the application of semantic segmentation presents unique challenges compared to its counterparts in construction scenes, traffic environments, and indoor settings. Requirements in point density are often aligned with the presentation demands of component details. Segmentation categories differ in morphology, such as the need to discern arches or vaults common in stone masonry is absent in East Asian timber structure scenes, which calls for specific research. These categories and the level of segmentation accuracy should also consider the application scenario of the converted digital models, like structural analysis or damage detection, to correctly handle the trade-offs in details of geometric features [39,40]. Besides, it is proper for segmentation algorithms to have the ability to train from small samples due to the lack of sufficient available heritage point cloud data. As widely adopted algorithms are primarily designed for large datasets like indoor scenes, the performance of such end-to-end algorithms on limited datasets needs to be further checked, and optimizations need to be made to adapt to such context.

To avoid a feature learning process that relies on sufficient data, a pragmatic two-step approach is adopted in this paper to achieve semantic segmentation. Algorithms are firstly leveraged to extract high-level manually defined features of each point as input for the classification network. Then point-wise labelling of features is implemented with a learning process in the second step. It is considered to yield satisfying segmentation results with limited training data and computational power, since the useful semantic information is directly obtained from pre-defined features, rather than extracting features from learning that rely on a large amount of training data. Within this paradigm, geometric [41,42], radiometric [43], and colour-related [44] context information has been used as high-level features. In some of the recent attempts [45], encoded features generated by the local descriptor were utilized. However, the work still targets common infrastructure and does not provide an automated modelling process. The ability of such a two-step approach to consistent expression in heritage scenarios also needs to be tested.

2. Research aim

With limited training data and variations in data quality, automated semantic segmentation of real-world point cloud data of heritages with high fidelity and efficiency has yet to be solved. In this paper, an automated process of semantic segmentation for heritage point clouds is proposed. Following the two-step strategy, an algorithm that can extract sufficient features from contextual information in the first step is the key to addressing the above-mentioned challenges. Taking the timber heritage buildings in East Asia as the research object, using the devised multi-scale local descriptor as a feature extractor and the neural network as a classifier, the proposed segmentation algorithm exhibits good robustness under varying point densities and integrity. Emphasizing accurate feature extraction becomes crucial for classification networks to yield better results, particularly when confronted with limited training samples. To form a complete process from scanning to digital models, the workflow of geometric modelling for labelled components is also involved. Consequently, the automated modelling from raw point clouds to digital twins of heritages could be performed to promote digital monitoring and management in heritage preservation.

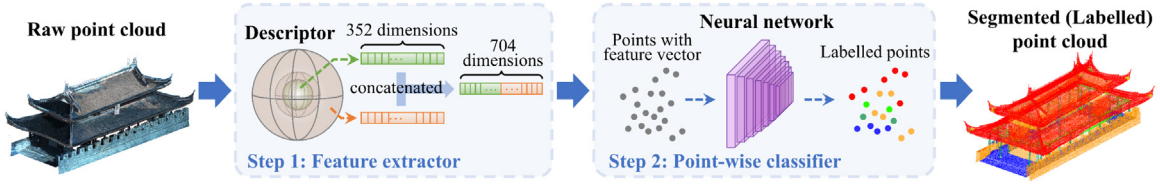

Fig. 1. Workflow of semantic segmentation.

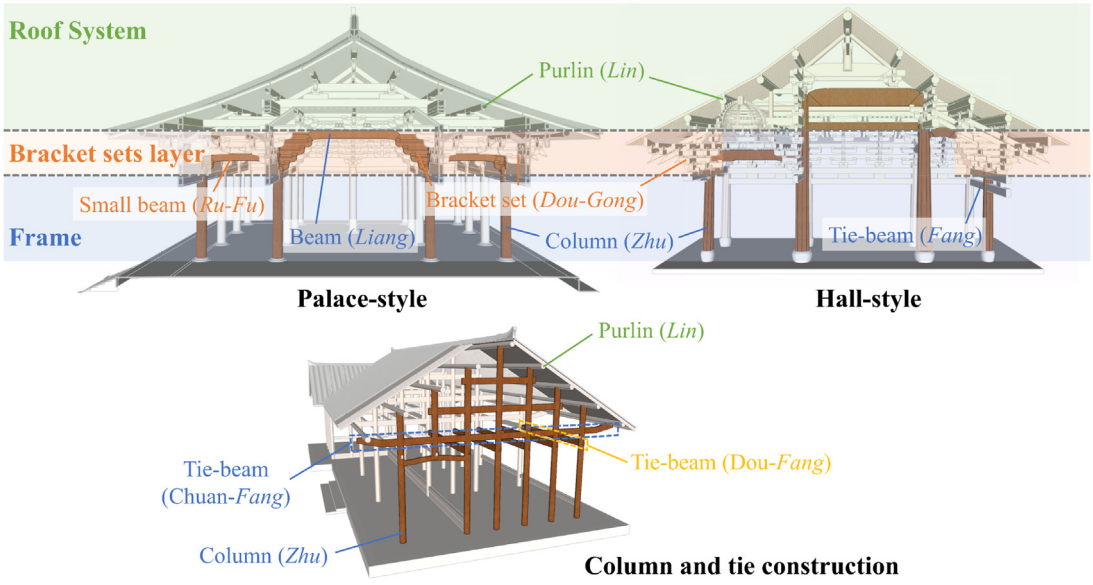

Fig. 2. Structural components in Chinese traditional architectures.

3. Materials and methods

In the first step of our semantic segmentation method, we opt for a designed descriptor to extract local features of point clusters. Then, a classification neural network is used to infer labels in the second step. The workflow of semantic segmentation is shown in Fig. 1. In the field of point cloud processing, a descriptor for a point is a feature vector calculated by a certain algorithm based on the distribution of neighbouring points. Such distinctive feature vectors are suitable as inputs for classification networks to infer point-wise labels, thereby achieving semantic segmentation. As target semantic labels in segmentation, typical components of Chinese traditional buildings are first defined in Section 3.1. The algorithm for computing multi-scale descriptors is introduced in Section 3.2, and details of label inference networks are introduced in Section 3.3. The practicability of the proposed descriptors and two-step strategy is confirmed in Section 4. An automated geometric modelling pipeline and its application are also introduced in Sections 3.4 and 5 to provide an application scenario for segmented point clouds.

3.1. Typical components of Chinese traditional buildings

The evolution of Chinese traditional architecture, particularly the construction techniques in the Song and Qing Dynasties, assumes pivotal significance in both archaeological and engineering realms. Insights from key technical standards, namely the "Treatise on Architectural Methods" (*Yin Zao Fa Shi*) [46] in the Song Dynasty and the "Code of Engineering Practice" (*Gong Cheng Zuo Fa Ze Li*) [47] in the Qing Dynasty, offer a detailed division of the structural form of the timber frame system. In the former paradigm, a classification distinguishes between palace-type and hall-type column-beam frames, emphasizing whether the bracket sets layer can be identified in the vertical direction. As shown in Fig. 2, the roof system in palace-type buildings is directly supported and connected to the frame through the bracket sets (*Dou-Gong*). The height of the interior and exterior columns is not equal in the hall-type frame, and the load-bearing beams are directly inserted into the column by mortise-tenon joints (*Sun-Mao*) [48]. A similar building classification based on the presence or absence of bracket sets was adopted in the Qing Dynasty, where bracket sets primarily serve ornamental purposes. Beyond the above two categories that can be regarded as the post and lintel construction, the column-and-tie construction emerges as a distinct structural form (see Fig. 2). In this system, the roof load directly transmits to the columns, leading to a dense column grid arrangement, utilizing the tie-beams (*Fang*) to connect columns under each purlin for support. From the above description, ground, columns, beams, walls, bracket sets, and roof systems constitute the six target labels for the semantic segmentation of the point cloud in this study. These categories could be regarded as the main structural components of Chinese timber buildings with topologically distinguishable.

3.2. Segmentation step 1: feature extraction with local descriptor

As the feature extractor in our proposed semantic segmentation process, a multi-scale local descriptor is designed. Considering the potential noise in practical point cloud data, we chose the encoding method of the Signature of Histograms of OrienTations (SHOT) [49] as a reference due to its ability to reduce noise effects [50]. SHOT records and counts the coordinate relationship between a point and its adjacent points into several histograms. The contextual information within the neighbourhood of this point (i.e., the support field) therefore is encoded into a feature vector. The division of support fields and the statistical strategy of histograms are

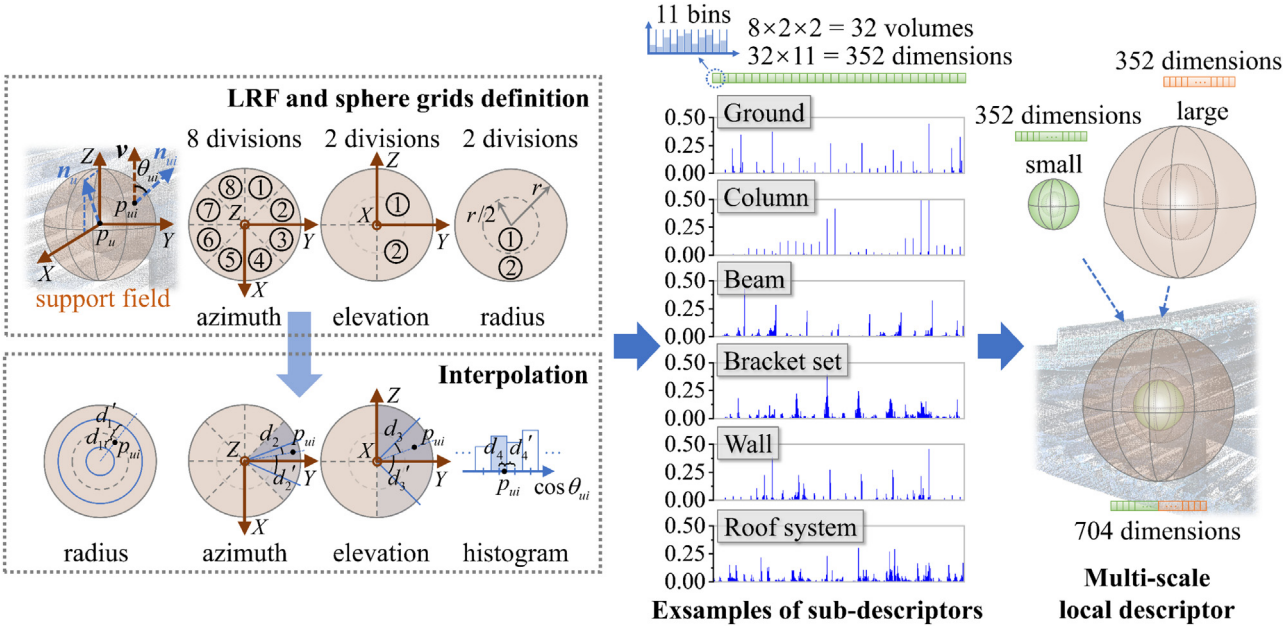

Fig. 3. Design of multi-scale descriptors.

inherited in this paper. Two improvements are made: (1) the local reference system (LRF) is specified according to the application scenarios for buildings, and (2) features of two different scales are concatenated. The resulting descriptor relies on the statistical strategy of orientation histograms and support fields of varying sizes, ensuring robustness and fidelity in extracting local features. Note that the size of the support field is determined by its radius, so the number of neighbouring points in the field varies. The detailed construction of our descriptor is defined as follows.

Assuming P_u is the u_{th} target point in the point cloud that needs to calculate its feature vectors, and $\mathbf{n}_u$ is the estimated normal vector of P_u. Within the nearest neighbours of P_u, the least-square plane fitting is implemented to find the local plane around this point, and the normal vector of this plane, which is the equivalent of $\mathbf{n}_u$, is estimated through principal component analysis (PCA) analysis. The neighbour radius for normal estimation was set to 0.1 m in the following implementation. We first introduce the LRF and sphere grids to define the geometric relationship between the P_u and its neighbouring points located in the support field. As given in Fig. 3, the Z-axis in LRF is parallel to the vertical direction $\boldsymbol{v}$ since this direction in building scenes is generally fixed to indicate elevation. To ensure the rotational invariance in the horizontal plane, the X-axis of LRF is along with the horizontal component of the $\mathbf{n}_u$, and the Y-axis is naturally the cross product of the Z-axis and X-axis. As for the sphere grids for histogram statistics, 2 radius divisions, 8 azimuth divisions, and 2 elevation divisions divide the support field into 32 volumes. Each volume in sphere grids has a histogram with 11 bins. Then, for the i_{th} neighbouring point P_{ui} of the target point P_u, the angle, θ_{ui}, between $\mathbf{v}$ and the normal of this neighbouring point, $\mathbf{n}_{ui}$, is chosen as the statistical variables to describe the geometric relationship between this two points. To improve computational efficiency, $\cos\theta_{ui}$ is calculated to reveal the value of θ_{ui}, since it can be easily obtained by $\mathbf{n}_{ui} \cdot \mathbf{v}$. After traversing all neighbouring points in the support field, counts are accumulated in bins according to $\cos\theta_{ui}$, forming the signature of histograms. Thus, a sub-descriptor, i.e., a feature vector, of the target point with 352 dimensions is formed. For example, the first 11 numbers of this vector represent the histogram of $\cos\theta_{ui}$ of all neighbouring points in the first volumes.

Since the local histograms are used here to represent features, it is inevitably affected by boundary effects. Therefore, for each point being accumulated into a specific local histogram bin, we perform interpolation with its neighbours to eliminate such effects (see Fig. 3). In particular, four factors, d_1 to d_4, are defined in the radius, azimuth, elevation, and histogram quantized domains, respectively. In radius, azimuth and elevation domains, d is the angular or Euclidean distance between the neighbouring point P_{ui} and the centre of volume. In histogram domains, d is the difference between the value of $\cos\theta_{ui}$ and the central value of the bin. Note that d in each domain is calculated in both the two nearest volumes or bins (i.e., d_1 and $d_1{}'$ in the radius domain shown in Fig. 3), and normalized by the size of the volume or the bin to meet the continuity. Subsequently, four factors are multiplied according to Eq. (1) to achieve the final interpolation, and the sub-descriptor is normalized by the L2 norm to address the non-uniform point density issue. Thus, points located on the boundary would be smoothly counted into adjacent volumes and bins. Examples of sub-descriptors corresponding to six categories in the practice dataset are given in Fig. 3. The discriminative ability of the designed descriptor is reflected through significant differences in feature vectors.

$$n = \prod_{i=1}^{4} (1 - d_i) \tag{1}$$

To enhance the perception of descriptors, two sub-descriptors with different support fields are concatenated, that is, two feature vectors of length 352 are concatenated into one feature vector of length 704, as shown in Fig. 3. In the small support field, the orientation distribution of close neighbours is described to separate the basic local geometric patterns, such as boundaries of structural components. The global topological relationships would be reserved in the large support field to locate similar components that are difficult to distinguish in the local area. By traversing all points in the point cloud, point-wise features are obtained as inputs for classification.

3.3. Segmentation step 2: point-wise classification with neural network

In our semantic segmentation method, the classification of the calculated descriptors is undertaken by a relatively simple neural network. Since descriptors with explicit topological relationships

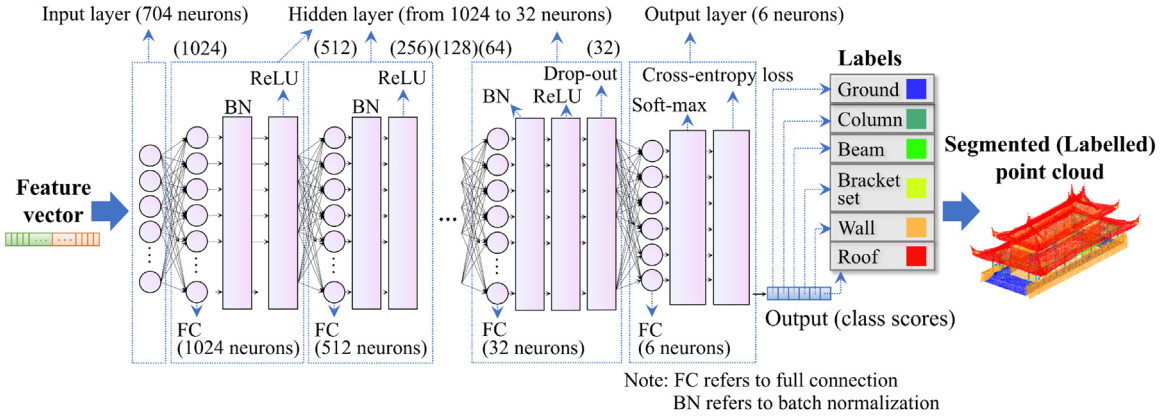

Fig. 4. Structures of the neural network.

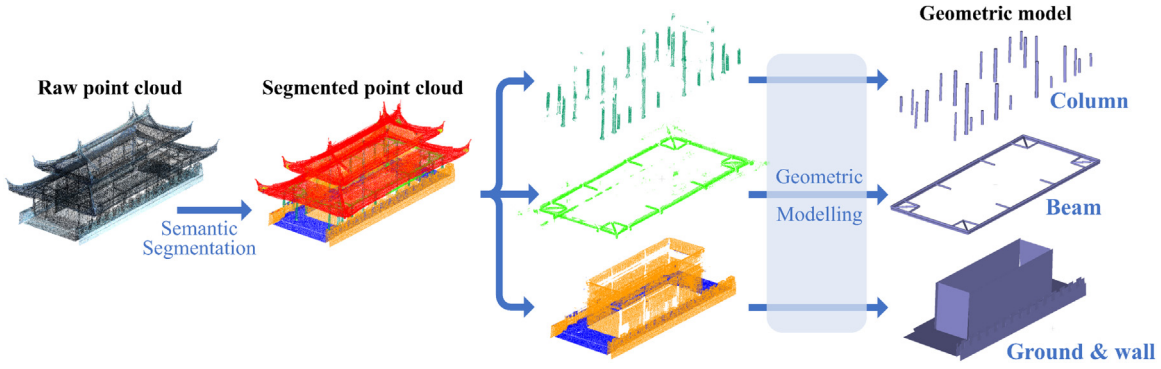

Fig. 5. Geometric modelling after segmentation.

and considerable perception have been obtained as features of point clusters, a simple classifier is sufficient to complete the training task with a small computational cost. The neurons adopted in this paper contain a linear function and a nonlinear function (i.e., activation function). The linear function is mathematically denoted as in Eq. (2). For the i_{th} neuron in the current layer, S_i is the result of the linear function, w_{ij} and x_j are the weight and the value of the j_{th} input respectively, b_i is the bias. The rectified linear unit (ReLU) is chosen as the activation function since it converges faster and avoids gradient vanishing [51]. The expression of ReLU is given as Eq. (3), where S_i and O_i are the input and output of this neuron, respectively. The outputs of neurons will be passed through connections and become the input of neurons in the next layer, and the n_{th} neuron in the next layer is illustrated given by Eq. (4), where O_n is the output of the neuron, w_{ni} is the weight of the i_{th} input, b_n is the bias. As for the output layer, the soft-max function, illustrated by Eq. (5), is taken as the nonlinear function, and the probability of each label (i.e., categories) could be obtained. Here, t_i is the result of the linear function of the i_{th} neuron in the output layer.

The structure of the neural network in this research is presented in Fig. 4. The number of inputs is the dimension of the descriptor. Six hidden layers gradually decreasing from 1024 to 32 neurons are adopted, and the number of outputs is 6 corresponding to the aforementioned categories of structural components. To mitigate over-fitting and enhance robustness in the feature space, batch normalization is implemented at each hidden layer, and a 50% drop-out is conducted after the last hidden layer. Such a drop-out technique is finished by randomly dropping neurons as well as their connections during training. During the training process using labelled data, the gradient descent and the standard cross-entropy loss are employed as the optimization method and loss function, respectively. Such loss function is defined as Eq. (6) and (7), where x is the input, y is the target, ω is the weight, N is the batch size. Weights are set to deal with unbalanced training datasets, as will be introduced later. Weights and biases of neurons are iteratively optimized through the partial derivatives of the loss function using the backpropagation training technique.

$$S_i = \sum w_{ij} x_j + b_i \tag{2}$$

$$O_i = f(S_i) = \max(0, S_i) \tag{3}$$

$$O_n = f\left(\sum w_{ni} f\left(\sum w_{ij} x_j + b_i\right) + b_n\right) \tag{4}$$

$$\mathrm{softmax}(t_i) = \frac{e^{t_i}}{\sum e^{t_i}} \tag{5}$$

$$l(x, y) = \sum_{n=1}^{N} \frac{1}{\sum_{n=1}^{N} w_{yn}} l_n \tag{6}$$

$$l_n = -\varpi_{yn} y_n \log(x_n) \tag{7}$$

3.4. After segmentation: geometric modelling

After the above two steps to achieve semantic segmentation of the raw data, different building components have been recognised as points with different labels. Further geometric modelling is expected to achieve the complete process of scanning to digital models, as shown in Fig. 5. Hence, an automatic process is proposed here to finish the modelling of each structural category for heritage buildings.

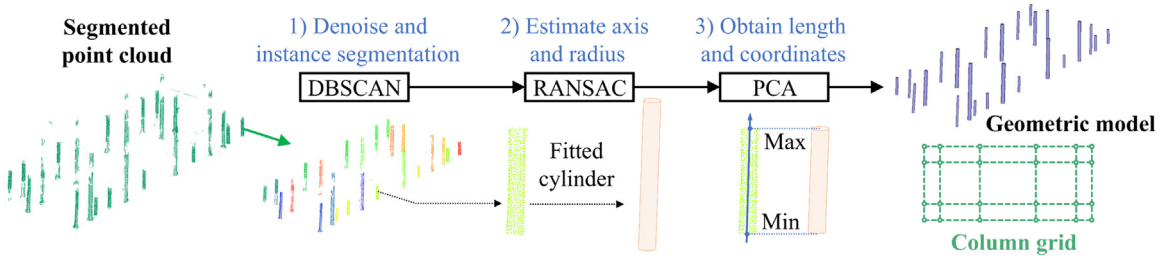

Fig. 6. Workflow of column modelling.

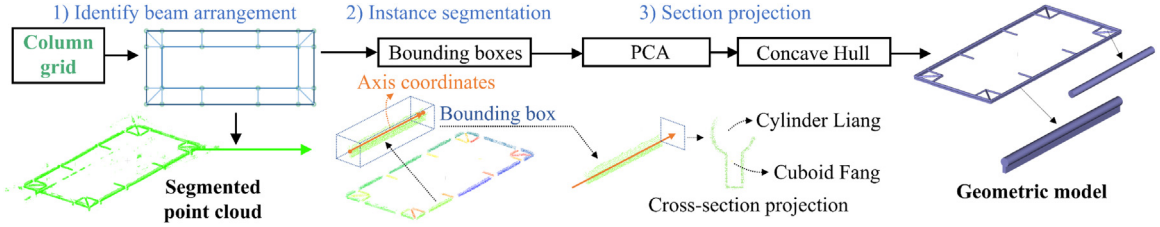

Fig. 7. Workflow of beam modelling.

3.4.1. Determination of the column grid and modelling of columns

Determining the column grid is the key to component positioning and is usually handled first. The workflow of column modelling is given in Fig. 6, and its details are as follows: 1) The density-based spatial clustering of applications with noise (DBSCAN), a clustering algorithm, is first introduced to separate each column and noise. DBSCAN finds core points of high density and expands clusters from them, thereby achieving instance segmentation. The parameters *eps* and *minPts* are crucial for clustering and are set to be 0.6 and 50 here. *eps* is the maximum relative distance between two points for one to be considered as in the neighbourhood of the other, and *minPts* is the minimum number of points in a neighbourhood for a point to be considered a core point. 2) Consider each column as a cylinder described by Eq. (8), estimate their axis vector and radii via RANSAC sampling and fitting. In Eq. (8), (x, y, z) is a point on the cylinder surface, (x_0, y_0, z_0) is a point on the axis, (a, b, c) is the axis vector, and r_0 is the cylinder radius. RANSAC identifies data whose distribution can be explained by some set of model parameters as interior and fits geometric parameters using inliers. The maximum distance from the model to a potential inlier point, d_{RANSAC}, is set to 0.02 m. 3) To obtain the length and the positioning coordinates of the column, the point cloud of each column is projected onto its axis using PCA to acquire the extreme coordinates in the axis direction, thereby the range of the column is determined. Although the length of the invisible part of the column could not be obtained, the known geometric parameters, i.e., axis vector, cylinder radius, coordinates, and visible length, are sufficient to locate the column grid and obtain the geometric model of columns. The column grid obtained here is the basic information for locating other components.

$$(x-x_0)^2+(y-y_0)^2+(z-z_0)^2-\left[a(x-x_0)+b(y-y_0)+c(z-z_0)\right]^2=r_0^2 \tag{8}$$

$$Ax+By+Cz+D=0 \tag{9}$$

3.4.2. Modelling of beams

Subsequently, the arrangement of beams can be estimated based on the obtained column grid as shown in Fig. 7. Its details are as follows: 1) Find the axis of all beams based on column coordinates. 2) Set appropriate cuboid bounding boxes along the axis coordinates to isolate individual beams, achieving instance segmentation. 3) In contrast to columns, timber beams are often composite due to the presence of Fangs, leading to an irregular cross-section. We use PCA to project the cross-section along the beam axis and the concave hull algorithm is applied to obtain the requisite dimensions of cross-sections.

3.4.3. Modelling of walls and ground

Walls and ground, with planar features, entail a straightforward task. The normal vectors of each plane are sequentially estimated following Eq. (9) using RANSAC. In Eq. (9), (x, y, z) is a point on the plane, (A, B, C) is the axis vector in Hessian paradigm, and D is a constant. d_{RANSAC} for RANSAC here is set to 0.15 m. By projecting all points onto the corresponding fitting plane, the boundaries of the plane can be determined using the convex hull algorithm. As for the roof system, it is difficult to identify certain structural components, such as purlins, due to the complete blocking, and the instance segmentation of the roof system is not considered in this paper. We hope to obtain this information in other approaches, rather than directly from the point cloud.

3.4.4. Extraction of symmetry plane of bracket sets

Complex types and configurations of bracket sets, compounded by a stepped structure, pose challenges for direct plane fitting. It is also not feasible to directly determine the centroid since the collected data is often incomplete. Their relatively small size also hinders feature extraction. Therefore, we adopt the process from key-point recognition, instance segmentation, to symmetry plane extraction (shown in Fig. 8) to obtain the centre coordinates of the bracket set, as described below: 1) Use PU-GCN [52] to up-sample the raw point cloud to a resolution of 15 mm to improve the accuracy of key-point recognition. In this network, the local point information from point neighbourhoods is encoded using a graph convolutional network to generate fine-grained details and perform the interpolation. 2) Calculate Harris key-points of the up-sampled point cloud and cluster them using affinity propagation or mean shift algorithms. The Harris detector determines geometric features of bracket sets, such as the corners and edges, by identifying regions of drastic changes in the point normal vector. The radius for normal estimation and the threshold to filter out weak corners are set to 0.07 m and 0.01, respectively. Examples of estimated key-points before and after up-sampling are

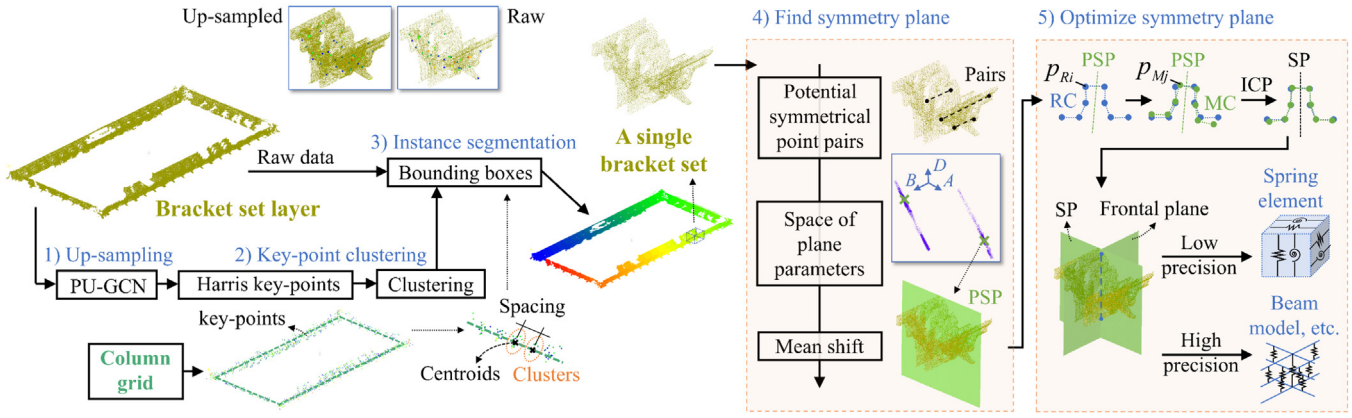

Fig. 8. Workflow of bracket sets symmetry plane extraction.

also given in Fig. 8, revealing that more details are recognised after up-sampling. 3) For each point cluster identified as a single instance, the centroids of clusters are projected onto the column grid to calculate the spacing between each component since bracket sets are often are uniformly spaced. The cuboid bounding boxes can be set accordingly to isolate individual bracket sets. The length of bounding boxes is the average spacing calculated above. Note that the instance segmentation is implemented on the raw data to reduce the computational complexity of subsequent steps.

Then, the precise coordinate of a bracket set is determined through the preliminary determination and optimization of the symmetry plane, as stated below: 4) For all point pairs in an instance point cloud, filter potential symmetrical pairs according to coordinates and normal vectors. Each pair can correspond to a potential symmetry plane with parameters *A, B, C*, and *D* expressed by Eq. (9). To determine a definite plane, place these plane parameters in a 3D space described by normalized *A, B*, and *D* (C is ignored since it is approximately 0), and employ the mean shift algorithm to find the region with the highest density. The coordinates corresponding to this region are the parameters of the preliminary symmetry plane PSP. 5) To further improve accuracy, optimization is executed by addressing registration errors. Given the raw point cloud RC, a mirror point cloud MC is derived based on the PSP. Traverse all points in RC, P_{Ri}, find the nearest point in MC, P_{Mj}, and calculate the distance between them, $d(P_{Ri}, P_{Mj})$. To avoid the effect of the local asymmetry caused by the incompleteness of the RC, a threshold, here we set to be 0.09 m, is defined to filter out pairs of points with a distance greater than this threshold. Take $d(P_{Ri}, P_{Mj})$ as the registration error, and employ an iterative closest point (ICP) approach to adjust the position of MC to minimize such error, resulting in the precise determination of the symmetry plane SP.

The geometric feature of bracket sets required for modelling is guided by precision considerations. Typical simplified calculation models with different accuracy have been summarized in [53]. In a numerical analysis with low precision, generally corresponding to the spring model, the information about the central axis coordinates is concerned. With a known column grid, the frontal plane aligned with the grid could be obtained, and the central axis of a bracket set is the intersection line of SP and the frontal plane. As for the precise analysis, specific details can also be derived by certain manual recognition, such as projecting onto symmetrical planes, thereby obtaining a 3D calculation model, such as the beam, truss, and corbel models listed in [53].

4. Semantic segmentation results and discussions

4.1. Datasets and implementation parameters

Two datasets of Chinese-style heritage buildings were collected to evaluate our proposed semantic segmentation method. In the first dataset, 4 buildings from open-source architectural cultural heritage point clouds datasets (i.e., ArCH dataset [54], Open Heritage 3D [55], and WHU-TLS Benchmark [56–58]) were involved. To obtain more point cloud data for accuracy purposes, 4 simulated point clouds were generated through ray tracing operations on 3D models with triangulated surfaces. Specifically, for each model, a circumscribed icosahedron composed of regular triangles is determined, and a series of virtual cameras are placed at the faces of the icosahedron. Point clouds from different viewpoints are captured by emitting virtual scanning rays from each virtual camera. An amalgamated cloud resembling practical scans is generated by aligning these point clouds through recorded camera poses. Of the 8 buildings, only KAS Pavilion-1 and KAS Pavilion-2 from ArCH retain their original labels, which include roofs, columns, and walls, aligning with the target labels of this research. The remaining buildings were manually labelled by us.

We collected the second dataset using the terrestrial laser scanner, including a timber building called Qiao-Lou, and the campus gate of Shanghai Jiao Tong University. The Qiao-Lou is located in the historic district Zi Town (*Zi-Cheng*) of Jiaxing City, China, and serves as a watchtower above the city wall since the last rehabilitation in 1908. The campus gate is a replica of the old one when the new campus was being constructed. The old gate with the style of the Qing Dynasty palace was built in 1934 and is considered a symbol of the University. Although the new gate is built of concrete, it is completely copied in appearance, maintaining typical Chinese architectural features. The point density was set to 3.1 mm over 10 m in two scans. Since it is difficult to obtain structural details of the roof by ground scans, unmanned aerial vehicle (UAV)-based oblique photography was also carried out to repair the holes caused by blocking and create a complete set of point clouds. Qiao-Lou and the campus gate were also manually labelled with 6 categories and labels of two datasets are marked with different colours in Fig. 9.

In the implementation of our semantic segmentation approach, the first dataset was used as the training set, while the second dataset functioned as the testing set. To address computer memory constraints, a voxel down-sampling with a size of 30 mm was performed on all data. Eventually, the training set contains 8 buildings with a total of 2247829 points and the test set contains

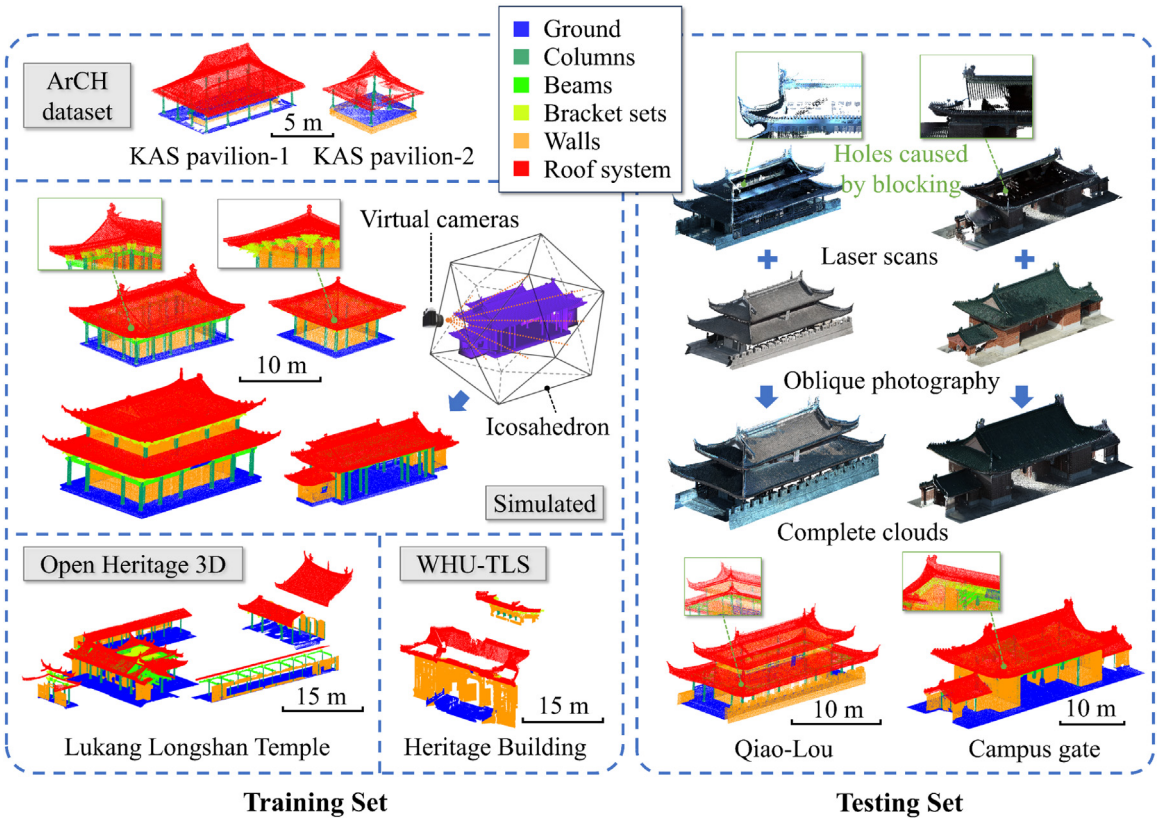

Fig. 9. Labelled datasets.

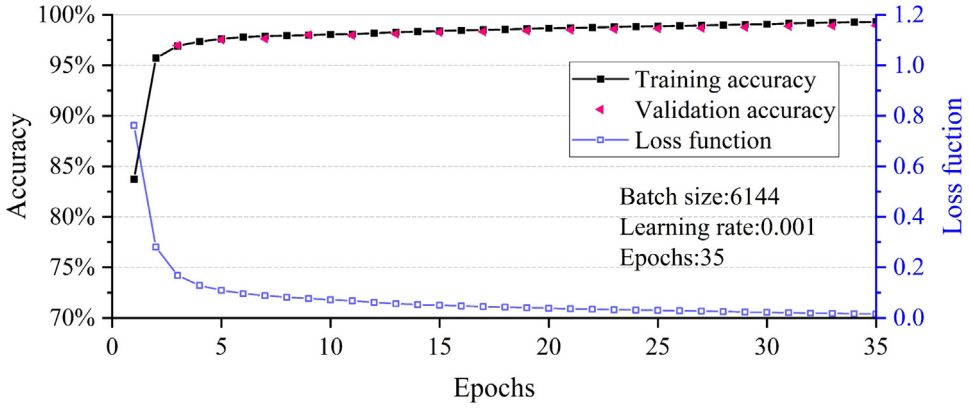

Fig. 10. Training process and parameters of the neural network.

2 buildings with a total of 2885076 points. Thirty percent of the points in the training set were randomly selected as the validation set during training. The output for classification comprised six categories, ground, columns, beams, bracket sets, walls, and roof system, with training weights (see Eq. (7)) adjusted to (2.67, 8.16, 6.65, 2.66, 2.96, 1.00) to mitigate the impact of imbalanced data during training. These weights were determined based on the ratio of points in each category, indicating that categories with fewer points in the training set will gain greater weight during training. Notably, the colour of scanned points was abandoned due to their susceptibility to lighting conditions, features of point clusters were derived directly from the coordinate information. The loss and accuracy curves when the radius of the descriptor is 0.9 m are given in Fig. 10 as an example of classifier training, and parameters in the experiment are also presented. It can be seen that when the epoch exceeds 25, the training and validation accuracy both increase slowly and exceed 98%, indicating that the parameters in the neuron have stabilized. Therefore, 35 serves as the number of training termination epochs for all models.

4.2. Implementation results

For evaluation metrics, mean intersection-over-union (mIoU) and overall accuracy (OA) were reported across all classes. If *TP* (true positive) is the number of correctly labelled points, *FP* (false positive) is the number of wrongly labelled points, *FN* (false negative) is the number of miss-labelled points, the IoU for each category can be given as Eq. (10), and the mIoU is the average of each category.

$$IoU = \frac{TP}{TP + FP + FN} \tag{10}$$

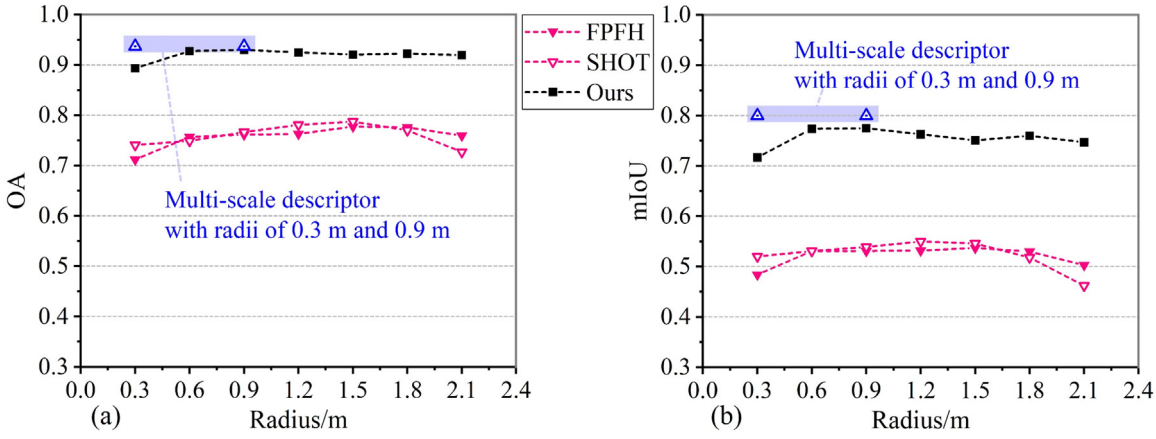

Fig. 11. Results of semantic segmentation of descriptors with different support radii: (a) OA results; (b) mIoU results.

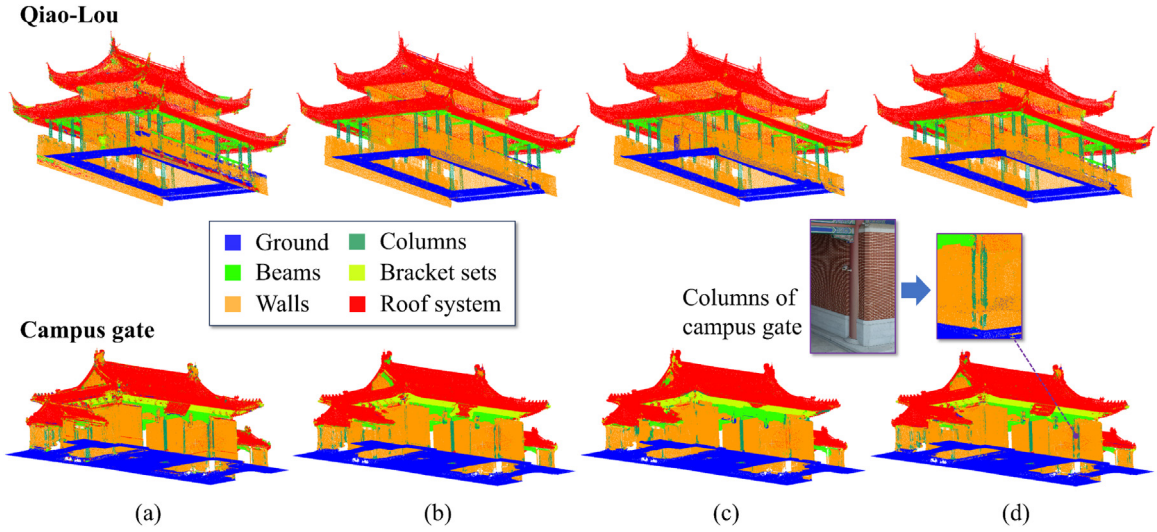

Fig. 12. Segmentation results of radius (a) 0.3 m; (b) 0.9 m; (c) 2.1 m; (d) 0.3 m and 0.9 m.

Firstly, the exploration of our single-scale descriptors, SHOT, and Fast Point Feature Histograms (FPFH) was conducted. Different descriptors with different support fields were calculated as inputs for the classifier proposed in Section 3.3. After training on the training set, different trained models were used for inference on the testing set, and the classification results are shown in Fig. 11. It is shown that the highest mIoU of these three descriptors is 0.77, 0.56, and 0.54 respectively, indicating that our descriptors perform better than other descriptors in the semantic segmentation accuracy. A trend of increasing and then decreasing can be observed in mIoU as the descriptor radius increases. The increase in mIoU reveals that more precise topological relationships are expressed along with the expansion of perception fields. A smaller field may result in points within the component being unable to classify correctly (Fig. 12 (a)). However, when the radius exceeds the general size of the structural component, more counted neighbouring points could make the extracted geometric features less distinguishable. The classification of points on the edge of components may be affected by surrounding components (Fig. 12 (c)). When the radius of the proposed single-scale descriptor is 0.9 m, a local maximum in mIoU is obtained, revealing a balance between accuracy and robustness.

Therefore, to fully represent the geometric features and improve the segmentation accuracy, the support radius of 0.3 m and 0.9 m were selected to form a multi-scale descriptor (blue points in Fig. 11), and the mIoU increased from 77.5% to 80.0%. The IoU results for each category are presented in Table 1. The segmentation result of columns in the campus gate is relatively poor due to the close connection between the columns and the infill wall in this building (Fig. 12 (d)), where the topological features of the column are similar to those of the wall. In contrast, when the column is exposed, as in Qiao-Lou, the results would be acceptable.

4.3. Discussion on different methodologies

Two end-to-end point-wise networks, GrowSP [24] and PointNet++ [36], were applied to perform semantic segmentation on our dataset for comparison. In the novel GrowSP network, 3D semantic elements are discovered via the growth of super-points. Point-wise features are learned from an encoder-decoder procedure, and the sizes of super-points progressively grow through clustering based on such features. The intuition of Pointnet++ came from the basic CNN structure where its lower-level neurons have smaller receptive fields whereas larger level has larger receptive fields. The hierarchical structure in this network is composed of a number of set abstraction levels to extract multi-level features. Here, our training set was used to train PointNet++ and perform inference on the testing set. On the contrary, the testing set was directly implemented in GrowSP, as it is an unsupervised method. Only the number of categories (i.e., 6) was given in GrowSP for unsupervised matching.

Table 1
IoU results of the testing set based on proposed descriptors and other methodologies.

Methodologies		IoUs for each category (Qiao-Lou / Campus gate)						mIoU
		Ground	Column	Beam	Bracket set	Wall	Roof	
Radius of proposed descriptors/m	0.3	0.71/0.89	0.61/0.37	0.56/0.66	-/0.75	0.74/0.78	0.92/0.88	0.71/0.72
	0.6	0.79/0.93	0.67/0.30	0.64/0.77	-/0.83	0.83/0.83	0.96/0.92	0.78/0.76
	0.9	0.74/0.94	0.64/0.28	0.69/0.72	-/0.87	0.82/0.82	0.96/0.95	0.77/0.76
	1.2	0.72/0.93	0.58/0.27	0.68/0.68	-/0.90	0.80/0.82	0.95/0.95	0.74/0.76
	1.5	0.71/0.93	0.58/0.26	0.64/0.64	-/0.87	0.78/0.82	0.95/0.96	0.73/0.75
	1.8	0.71/0.92	0.61/0.24	0.70/0.66	-/0.88	0.79/0.83	0.95/0.96	0.75/0.75
	2.1	0.69/0.92	0.57/0.21	0.70/0.65	-/0.85	0.77/0.84	0.95/0.96	0.74/0.74
	Multi	0.75/0.94	0.71/0.34	0.73/0.74	-/0.91	0.84/0.84	0.96/0.95	0.80/0.78
GrowSP [24]		0.01/0.01	0.09/0.07	0.16/0.03	-/0.01	0.02/0.01	0.04/0.02	0.06/0.03
PointNet++ [36]		0.26/0.03	0.01/0.01	0.04/0.06	-/0.24	0.64/0.73	0.65/0.31	0.32/0.22

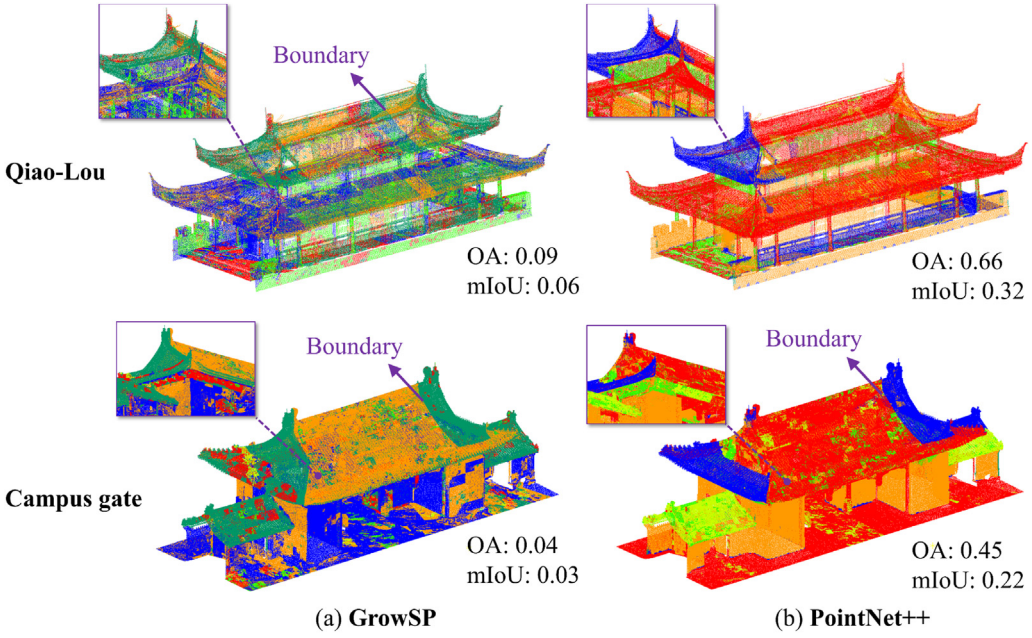

Fig. 13. Segmentation results of end-to-end networks (a) GrowSP; (b) PointNet++.

The results of these two networks are shown in Table 1 and Fig. 13. Due to the limited data, the discrimination of the features learned by the two networks is still restricted, resulting in unsatisfactory segmentations. The OAs and mIoUs of the two networks are less than 0.7 and 0.4, respectively, which is lower than the results obtained by our method. End-to-end networks rely on learning features with good discrimination from sufficient training samples. In the original paper of PointNet++ [36], an 84% accuracy for indoor scenes was obtained from a dataset with 1201 scenes, while the available models in this paper were only 8. Hence, with our dataset, points belonging to different categories are wrongly divided together using PointNet++, such as the confusion of ground and roof (see points marked in red in Fig. 13 (b)) due to similar normal vectors. The necessity of defining coordinate systems in small sample learning is therefore emphasized. Similar phenomena were also found in the results of GrowSP. The ground and walls were incorrectly divided together (see points marked in blue in Fig. 13 (a)) due to similar plane features. Some basic topological region divisions based on feature clustering could be seen with GrowSP. However, segmentation can hardly correspond to the true label due to the lack of a priori knowledge even though the geometric features such as Point Feature Histograms (PFH) were used to enhance the expression ability of its features. Such complete error in label mapping resulted in a nearly zero mIoU with GrowSP. However, even if the mismatch on this label is ignored, such segmentation with local irregular clustering is still unacceptable. The imbalance of data in categories also deteriorates the segmentation performance since limited information is learned in the category with fewer points. Besides, some unexpected boundaries on the roof deserve attention. As mentioned, in order to obtain a complete point cloud, different techniques are used to fill the holes caused by blocking, leading to certain differences in the topological relationship of points in these regions. The appearance of these boundaries indicates the high sensitivity of the two networks to the raw data. Therefore, our method is more accurate and applicable for the small sample learning scenario compared with other methodologies. The clear geometric meaning of the extracted features from the proposed descriptors makes the classifier more expressive and distinguishable, making them more suitable for complex scenes like heritage buildings.

4.4. Discussion on robustness of the proposed method

In practical heritage scanning, the point cloud that needs to be segmented is often a fusion of ground laser scanning and oblique photography, introducing challenges of spatial scale and density variations [43]. Since a robust result is desired with point clouds of different quality, the performance of the proposed method was evaluated by adjusting the integrity of point clouds in the testing set. The precision in estimating the surface normal, which is cru-

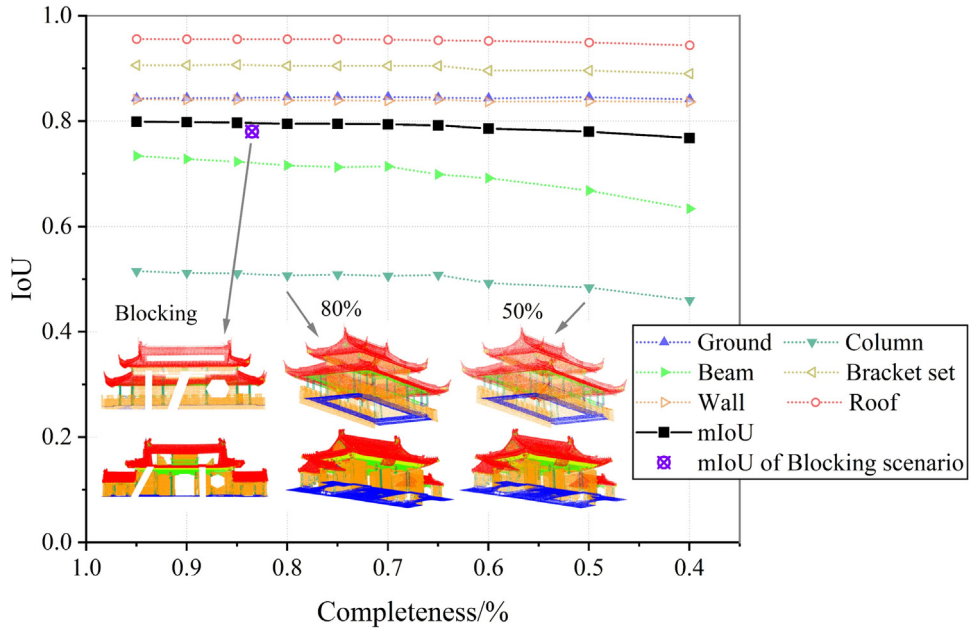

Fig. 14. IoUs of semantic segmentation with different levels of completeness.

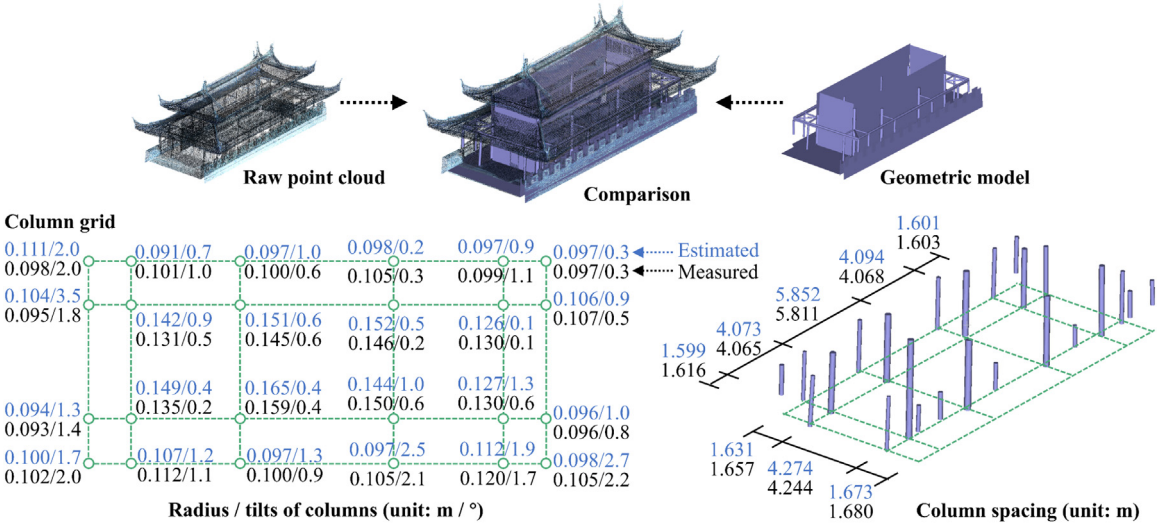

Fig. 15. Geometric modelling results.

cial for feature extraction, may be severely affected under different sampling conditions. By randomly dropping a certain percentage (90% to 60%) of points from raw data, the quantitative analysis on the relationship between the level of completeness and the segmentation results was conducted as given in Fig. 14. It is revealed that when 60% of points being removed, the mIoU slightly decreased from 0.80 to 0.76. To simulate the inevitable blocking problem in practical application, the removal of partial point clusters is also implemented. Points in various geometries were removed here to present the holes caused by blocking, and these removed points account for approximately 16.5% of the total number of points. In this blocking scene, the mIoU has decreased to 0.78.

It can be concluded that the segmentation is slightly affected by the integrity of point clouds. Due to the relatively small proportion of beam and column data, the reduction of point cloud density may be a concern for identifying such categories. The performance in the partially blocking scenario is slightly worse than those with similar completeness, which may be caused by the misclassification of the points at the edge of the holes. The topological relationship of points close to the holes will be significantly different from that of other complete regions. Therefore, as long as the topological structure of the point cloud is not fundamentally damaged, acceptable segmentation results could be achieved, which shows the good robustness of the proposed method.

5. Automatic geometric modelling results

To automatically acquire geometric models essential for BIM archives or numerical analysis, the proposed modelling process given in Section 3.4 is employed with Qiao-Lou, and the results are shown in Fig. 15. In the figure, the comparison between the estimated geometric parameters (marked by blue), and the on-site measured values (marked by black) is given. It is revealed that the error between the two is generally less than 10%, demonstrating satisfied modelling accuracy. With the small tilt of the column, the

estimation error is relatively significant since it is affected by the noise.

Since the bracket set is not involved in Qiao-Lou, we use the bracket set layer extracted from the campus gate to illustrate the evaluation results of the bracket set symmetry plane parameters. In this case, the evaluation of the average spacing between each bracket set is 1.197 m, while the measured value is 1.204 m, and the error between the two is 0.6%. When determining the symmetry plane parameters of each bracket set, the average $d(P_{Ri}, P_{Mj})$ corresponding to the preliminary symmetry plane PSP and the final symmetry plane SP are 22.6 mm and 21.4 mm, respectively, indicating an effective parameter optimization.

It should be emphasized that the large-scale geometric information of the structure, such as tilts of columns shown in Fig. 15, is focused on in the proposed modelling process. Although geometric details, such as local buckling or cracking of columns, may be generalized or simplified, the extracted information still plays a crucial role as the parameter for evaluating the overall structural load performance and the monitored indicator in structural health monitoring activities. When discussing a true-to-deformation geometric reconstruction, specific high-precision methods need to be developed, which is not covered in this paper.

6. Conclusions

This paper introduces a semantic segmentation method based on a novel multi-scale local descriptor to identify primary structural components in traditional Chinese buildings from point clouds. An automatic modelling procedure is also given to facilitate the transition from the classified point cloud to the required geometric model. The performance of the proposed approach was field validated using point cloud data collected from two buildings, from which the following conclusions can be drawn:

1. Using local descriptors to extract discriminative high-level features before learning-based classification can avoid a data-intensive feature learning process in structural component recognition. By defining a local reference system and concatenating support fields at different scales, the mIoU of semantic segmentation using the proposed descriptors has increased from 0.54 to 0.80 compared with the results obtained using FPFHs and SHOTs. The notable discrimination of the extracted features ensures the effective recognition of complex heritage components even with limited training samples, which is difficult for other end-to-end networks to achieve.
2. To achieve optimal performance, it is recommended to set the radii of the two-scale descriptors to 0.3 m and 0.9 m respectively in heritage building scenes. In this condition, the mIoU of segmentation decreases slightly to 0.76 with random removal of 60% points, demonstrating the high robustness of the proposed method against data missing and sampling density changes. The adaption of our process to the complexity of real-world scenarios is therefore highlighted.
3. Leveraging the proposed automatic geometric feature extraction process enables the generation of geometric models for heritage buildings with less than 10% evaluation error in geometric dimensions. This facilitates subsequent numerical analysis or digital twin construction, reducing the need for extensive manual operations.

It should be emphasized that the proposed strategy could be applied to other types of heritages, not only on Chinese-style timber buildings, by redefining the labels that need to be segmented and conducting corresponding network training. Still, the proposed method has some limitations. The performance of the proposed descriptor will decrease when the geometric features of the component are similar to those of the surrounding area (e.g., columns embedded in the wall as shown in Fig.10). The introduction of additional information, such as the picture colour, may potentially solve this problem in future research. Besides, the local descriptor is designed and proven to work well in outdoor scenes. For data containing indoor information, especially complex scenes with furniture, the proposed methodology needs to be adapted to eliminate the interference of objects unrelated to architecture.

Acknowledgements

The authors are grateful for the financial support of the Open Fund of Shanghai Key Laboratory of Engineering Structure Safety (Grant No. 2023-KF01), and the Research Project of China Railway Shanghai Group Co., Ltd. (Grant No. 2022158, 2022239).

References

[1] G. Ruggiero, R. Marmo, M. Nicolella, A methodological approach for assessing the safety of historic buildings' façades, Sustainability 13 (2021) 2812.
[2] M. Mishra, Machine learning techniques for structural health monitoring of heritage buildings: a state-of-the-art review and case studies, J. Cult. Herit. 47 (2021) 227–245.
[3] C.-X. Qu, J.-Z. Jiang, T.-H. Yi, H.-N. Li, Computer vision-based 3D coordinate acquisition of surface feature points of building structures, Eng. Struct. 300 (2024) 117212.
[4] G. Castellazzi, A.M. D'Altri, S. de Miranda, F. Ubertini, An innovative numerical modeling strategy for the structural analysis of historical monumental buildings, Eng. Struct. 132 (2017) 229–248.
[5] J. Martens, T. Blut, J. Blankenbach, Cross domain matching for semantic point cloud segmentation based on image segmentation and geometric reasoning, Adv. Eng. Inform. 57 (2023) 102076.
[6] K. Mirzaei, M. Arashpour, E. Asadi, H. Masoumi, A. Mahdiyar, V. Gonzalez, End–to-end point cloud-based segmentation of building members for automating dimensional quality control, Adv. Eng. Inform. 55 (2023) 101878.
[7] A. Khalil, S. Stravoravdis, D. Backes, Categorisation of building data in the digital documentation of heritage buildings, Appl. Geomat. 13 (2021) 29–54.
[8] G. Rocha, L. Mateus, J. Fernández, V. Ferreira, A scan-to-BIM methodology applied to heritage buildings, Heritage 3 (2020) 47–67.
[9] D.P. Pocobelli, J. Boehm, P. Bryan, J. Still, J. Grau-Bové, BIM for heritage science: a review, Herit. Sci. 6 (2018) 1–15.
[10] F. Poux, R. Billen, Voxel-based 3D point cloud semantic segmentation: unsupervised geometric and relationship featuring vs deep learning methods, ISPRS Int. J. Geo.-Inf. 8 (2019) 213.
[11] D.-G.J. Opoku, S. Perera, R. Osei-Kyei, M. Rashidi, Digital twin application in the construction industry: a literature review, J. Build. Eng. 40 (2021) 102726.
[12] S. Chen, G. Fan, J. Li, Improving completeness and accuracy of 3D point clouds by using deep learning for applications of digital twins to civil structures, Adv. Eng. Inform. 58 (2023) 102196.
[13] A. Murtiyoso, P. Grussenmeyer, D. Suwardhi, R. Awalludin, Multi-scale and multi-sensor 3D documentation of heritage complexes in urban areas, ISPRS Int. J. Geo.-Inf. 7 (2018) 483.
[14] M. Wojtkowska, M. Kedzierski, P. Delis, Validation of terrestrial laser scanning and artificial intelligence for measuring deformations of cultural heritage structures, Measurement 167 (2021) 108291.
[15] G.Y. Jeong, T.N. Nguyen, D.K. Tran, T.B.H. Hoang, Applying unmanned aerial vehicle photogrammetry for measuring dimension of structural elements in traditional timber building, Measurement 153 (2020) 107386.
[16] S.L. Chen, Q.W. Hu, S.H. Wang, H.J. Yang, A virtual restoration approach for ancient plank road using mechanical analysis with precision 3D data of heritage site, Remote Sens. 8 (2016) 828.
[17] Y. Xie, J. Tian, X. Zhu, A review of point cloud semantic segmentation. arXiv 2019, arXiv preprint arXiv:1908.08854.
[18] R. Qiu, Q.-Y. Zhou, U. Neumann, Pipe-run extraction and reconstruction from point clouds, in: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part III 13, Springer, 2014, pp. 17–30.
[19] F. Lafarge, R. Keriven, M. Brédif, Insertion of 3-D-primitives in mesh-based representations: towards compact models preserving the details, IEEE Trans. Image Process. 19 (2010) 1683–1694.
[20] Q. Lu, S. Lee, Image-based technologies for constructing as-is building information models for existing buildings, J. Comput. Civil Eng. 31 (2017) 04017005.
[21] M. Bassier, M. Vergauwen, F. Poux, Point cloud vs. mesh features for building interior classification, Remote Sens. 12 (2020) 2224.
[22] Z. Su, Z. Gao, G. Zhou, S. Li, L. Song, X. Lu, N. Kang, Building plane segmentation based on point clouds, Remote Sens. 14 (2021) 95.
[23] P.V. Paiva, C.K. Cogima, E. Dezen-Kempter, M.A. Carvalho, Historical building point cloud segmentation combining hierarchical watershed transform and curvature analysis, Pattern Recognit. Lett. 135 (2020) 114–121.

[24] Z. Zhang, B. Yang, B. Wang, B. Li, GrowSP: unsupervised semantic segmentation of 3D point clouds, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 17619–17629.
[25] J.S. Markiewicz, P. Podlasiak, D. Zawieska, A new approach to the generation of orthoimages of cultural heritage objects—integrating TLS and image data, Remote Sens. 7 (2015) 16963–16985.
[26] A.N. Andrés, F.B. Pozuelo, J.R. Marimón, A. de Mesa Gisbert, Generation of virtual models of cultural heritage, J. Cult. Herit. 13 (2012) 103–106.
[27] Y. Alshawabkeh, Linear feature extraction from point cloud using color information, Herit. Sci. 8 (2020) 1–13.
[28] H. Macher, T. Landes, P. Grussenmeyer, E. Alby, Semi-automatic segmentation and modelling from point clouds towards historical building information modelling, Digital Heritage, in: Progress in Cultural Heritage: Documentation, Preservation, and Protection: 5th International Conference, EuroMed 2014, Limassol, Cyprus, November 3-8, 2014. Proceedings 5, Springer, 2014, pp. 111–120.
[29] R. Lu, I. Brilakis, C.R. Middleton, Detection of structural components in point clouds of existing RC bridges, Comput.-Aided Civil Infrastruct. Eng. 34 (2019) 191–212.
[30] E. Quagliarini, P. Clini, M. Ripanti, Fast, low cost and safe methodology for the assessment of the state of conservation of historical buildings from 3D laser scanning: the case study of Santa Maria in Portonovo (Italy), J. Cult. Herit. 24 (2017) 175–183.
[31] V. Sanchez, A. Zakhor, Planar 3D modeling of building interiors from point cloud data, in: 2012 19th IEEE International Conference on Image Processing, IEEE, 2012, pp. 1777–1780.
[32] C. Pérez-Sinticala, R. Janvier, X. Brunetaud, S. Treuillet, R. Aguilar, B. Castañeda, in: Evaluation of Primitive Extraction Methods from Point Clouds of Cultural Heritage Buildings, Structural Analysis of Historical Constructions: An Interdisciplinary Approach, Springer, 2019, pp. 2332–2341.
[33] D. Maturana, S. Scherer, Voxnet: A 3d convolutional neural network for real–time object recognition, in: 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS), IEEE, 2015, pp. 922–928.
[34] H. Su, S. Maji, E. Kalogerakis, E. Learned-Miller, Multi-view convolutional neural networks for 3d shape recognition, in: Proceedings of the IEEE international conference on computer vision, 2015, pp. 945–953.
[35] E. Pellis, A. Murtiyoso, A. Masiero, G. Tucci, M. Betti, P. Grussenmeyer, 2D to 3D Label propagation for the semantic segmentation of Heritage building point clouds, in: XXIV ISPRS Congress "Imaging Today, Foreseeing Tomorrow", Commission II 2022 edition, 6–11 June 2022, Nice, France, ISPRS, 2022, pp. 861–867.
[36] C.R. Qi, L. Yi, H. Su, L.J. Guibas, Pointnet++: deep hierarchical feature learning on point sets in a metric space, advances in neural information processing systems, 30 (2017).
[37] W. Shi, R. Rajkumar, Point-gnn: graph neural network for 3d object detection in a point cloud, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 1711–1719.
[38] R. Pierdicca, M. Paolanti, F. Matrone, M. Martini, C. Morbidoni, E.S. Malinverni, E. Frontoni, A.M. Lingua, Point cloud semantic segmentation using a deep learning framework for cultural heritage, Remote Sens. 12 (2020) 1005.
[39] F. Hamid-Lakzaeian, Point cloud segmentation and classification of structural elements in multi-planar masonry building facades, Autom. Constr. 118 (2020) 103232.
[40] S. Yang, M. Hou, S. Li, Three-dimensional point cloud semantic segmentation for cultural heritage: a comprehensive review, Remote Sens. 15 (2023) 548.
[41] E. Valero, A. Forster, F. Bosché, E. Hyslop, L. Wilson, A. Turmel, Automated defect detection and classification in ashlar masonry walls using machine learning, Autom. Constr. 106 (2019) 102846.
[42] X. Xiong, A. Adan, B. Akinci, D. Huber, Automatic creation of semantically rich 3D building models from laser scanner data, Autom. Constr. 31 (2013) 325–337.
[43] E. Grilli, F. Remondino, Machine learning generalisation across different 3D architectural heritage, ISPRS Int. J. Geo.-Inf. 9 (2020) 379.
[44] V. Croce, G. Caroti, L. De Luca, K. Jacquot, A. Piemonte, P. Véron, From the semantic point cloud to heritage-building information modeling: a semiautomatic approach exploiting machine learning, Remote Sens. 13 (2021) 461.
[45] T. Xia, J. Yang, L. Chen, Automated semantic segmentation of bridge point cloud based on local descriptor and machine learning, Autom. Constr. 133 (2022) 103992.
[46] J. Li, Treatise on Architectural Methods, China: Royal Press, Kaifeng, 1100.
[47] The Department of Labor, Code of Engineering Practice, China: Royal Press, Beijing, 1734.
[48] Y. Lin, Q. Chun, C. Zhang, Y. Han, H. Fu, Research on seismic performance of traditional Chinese hall-style timber buildings in the Song and Yuan dynasties (960–1368 AD): a case study of the main hall of Baoguo Temple, J. Wood Sci. 68 (2022) 1–25.
[49] F. Tombari, S. Salti, L. Di Stefano, Unique signatures of histograms for local surface description, in: Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part III 11, Springer, 2010, pp. 356–369.
[50] A.E. Johnson, M. Hebert, Using spin images for efficient object recognition in cluttered 3D scenes, IEEE Trans. Pattern. Anal. Mach. Intell. 21 (1999) 433–449.
[51] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. (2012) 25.
[52] G. Qian, A. Abualshour, G. Li, A. Thabet, B. Ghanem, Pu-gcn: point cloud upsampling using graph convolutional networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 11683–11692.
[53] Y. Hua, Q. Chun, X. Jia, Simplified calculation model for typical Dou-Gong exposed to vertical loads, Buildings-Basel 12 (2022) 689.
[54] F. Matrone, A. Lingua, R. Pierdicca, E.S. Malinverni, M. Paolanti, E. Grilli, F. Remondino, A. Murtiyoso, T. Landes, A benchmark for large-scale heritage point cloud semantic segmentation, Int. Arch. Photogrammetry, Remote Sens. Spat. Inf. Sci. 43 (2020) 1419–1426.
[55] CyArk, China University of Technology (CUTe), CyArkLukang Longshan Temple-LiDAR-Terrestrial, Photogrammetry, 2019 Open Heritage 3D.
[56] Z. Dong, B. Yang, Y. Liu, F. Liang, B. Li, Y. Zang, A novel binary shape context for 3D local surface description, ISPRS-J. Photogramm. Remote Sens. 130 (2017) 431–452.
[57] Z. Dong, B. Yang, F. Liang, R. Huang, S. Scherer, Hierarchical registration of unordered TLS point clouds based on binary shape context descriptor, ISPRS-J. Photogramm. Remote Sens. 144 (2018) 61–79.
[58] Z. Dong, F. Liang, B. Yang, Y. Xu, Y. Zang, J. Li, Y. Wang, W. Dai, H. Fan, J. Hyyppä, Registration of large-scale terrestrial laser scanner point clouds: A review and benchmark, ISPRS-J. Photogramm. Remote Sens. 163 (2020) 327–342.