

RESEARCH

Open Access



# Implementing PointNet for point cloud segmentation in the heritage context

Bulent Haznedar<sup>1\*</sup>, Rabia Bayraktar<sup>2</sup>, Ali Emre Ozturk<sup>3</sup> and Yusuf Arayici<sup>4</sup>

## Abstract

Automated Heritage Building Information Modelling (HBIM) from the point cloud data has been researched in the last decade as HBIM can be the integrated data model to bring together diverse sources of complex cultural content relating to heritage buildings. However, HBIM modelling from the scan data of heritage buildings is mainly manual and image processing techniques are insufficient for the segmentation of point cloud data to speed up and enhance the current workflow for HBIM modelling. Artificial Intelligence (AI) based deep learning methods such as PointNet are introduced in the literature for point cloud segmentation. Yet, their use is mainly for manufactured and clear geometric shapes and components. To what extent PointNet based segmentation is applicable for heritage buildings and how PointNet can be used for point cloud segmentation with the best possible accuracy (ACC) are tested and analysed in this paper. In this study, classification and segmentation processes are performed on the 3D point cloud data of heritage buildings in Gaziantep, Turkey. Accordingly, it proposes a novel approach of activity workflow for point cloud segmentation with deep learning using PointNet for the heritage buildings. Twenty-eight case study heritage buildings are used, and AI training is performed using five feature labelling for segmentation namely, walls, roofs, floors, doors, and windows for each of these 28 heritage buildings. The dataset is divided into clusters with 80% training dataset and 20% prediction test dataset. PointNet algorithm was unable to provide sufficient accuracy in segmenting the point clouds due to deformation and deterioration on the existing conditions of the heritage case study buildings. However, if PointNet algorithm is trained with the restitution-based heritage data, which is called synthetic data in the research, PointNet algorithm provides high accuracy. Thus, the proposed approach can build the baseline for the accurate classification and segmentation of the heritage buildings.

**Keywords:** Deep learning, Artificial intelligence, Cultural heritage, Segmentation, 3D point cloud

## Introduction

Segmentation and classification of the building elements is critical in both research and practice. Thus, AI concepts such as deep learning have been developed, which have gained importance due to the increasing demand for Heritage Building Information Modelling (HBIM) from point cloud data.

In the literature, image processing techniques for point cloud segmentation incorporating voxelization [1],

region growing, brute force plane sweeps, Hough transforms [2], expectation maximisation techniques [3] are tested and implemented for the surface-based segmentation. Due to large amounts of data and extracting information from enormous datasets, these techniques were still not sufficient for point cloud segmentation. Thus, studies using deep learning approach with outstanding mechanism for point cloud segmentation have started to increase in recent years.

In recent years, deep learning studies on 3D Point Cloud have become a wide research area to determine whether deep learning shows the same success in irregular data. Studies on 3D Point Cloud can be based on 4 different methods: Voxelization-based [4,

\*Correspondence: haznedar@gaziantep.edu.tr

<sup>1</sup> Department of Computer Engineering, Gaziantep University, Gaziantep, Turkey

Full list of author information is available at the end of the article

5], multi-view-based [6, 7], graph-based [8–10] and set-based [11, 12]. OctNet [13] and Kd-Net [14], created by using the advantages of the voxelization-based method, are two different methods that reduce the computational cost. In these methods, the voxel, which is expressed as empty in the data allocated to the voxels, is not included in the calculation, thus saving both time and memory. The multi-view-based method [6, 7] defines the 3D point cloud as a series of images taken from different angles. The number of images taken from different angles, the image distribution, and radial distances between images are not at regular intervals. Therefore, different parameters are required for each study. It is often described as an indefinite method. Graph-based method [8–10] is a Convolutional Neural Network (CNN)-based method that processes the neighbourhoods of each point in the point cloud in planar space and then creates the final planar space graph.

Methods that require obtaining 2D images or scanning the entire point cloud in order to segment from 3D data are not cost and time effective. Therefore, there is a need for solutions that can be worked directly on the point cloud without pre-processing. In the part segmentation study by Yi et al. [15], a method for object segmentation was proposed over point cloud data belonging to 16 different categories containing different numbers of data. According to this method, different regions of the object were determined in each object category and the system was trained in this direction. Deep learning methods using a total of 95,000 data were supported by different framework methods and a structure called Scalable Active Framework was created. With this part segmentation method, an F1 score varying between 85% and 95% was obtained in 16 different categories.

PointNet [16], which is an end-to-end deep neural network architecture that allows working directly on the point cloud and can be used for classification, part segmentation and semantic segmentation, is one of the pioneering studies in this field. Using the PointNet architecture, the semantic segmentation performance was obtained 83.7%. The authors, who stated that PointNet could not capture local geometries over time, presented the PointNet++ [17] architecture as a new study. In this study, a hierarchical grouping was made to identify local features. More details on the point cloud can be captured using point-to-point metric calculations.

This paper aims to propose an approach for segmentation of point cloud data for heritage structures using the PointNet deep learning algorithm. There is currently a significant gap in research and practice on the automated segmentation of point cloud data for heritage building towards automated HBIM modelling. Previous research and literature review show that it is necessary

to future-proof digital records of historical buildings to ensure that their components can be reliably located through tagging, such as semantically recognizable doors, windows, and walls. However, within the field of document analysis and pattern recognition in cultural heritage, it is widely recognized that current analysis of pattern recognition and deep learning methods are inadequate for the analysis/recognition of degraded, information-rich historical buildings since most work in the literature has concentrated on relatively narrow scope objects, such as textual documents or small 3D artefacts rather than buildings.

Hence, this paper examines and proposes a segmentation approach using PointNet for heritage buildings point cloud data. In this study, classification and segmentation processes are performed on 3D point cloud data of the heritage buildings in Gaziantep in Turkey. In this process, the segmentation of the historical structure, which is the most comprehensive step to create a BIM model, is achieved using artificial intelligence and deep learning methods, and the results are examined.

### Related works

In this section, the studies that focused on similar methods as in this study related to the segmentation for point cloud data have been critically reviewed. In a study by Shen et al. [18], 3D point clusters are defined as 3D data stacks whose correlation can be calculated, which can respond jointly to neighbouring points and can learn. The two methods named Edge-Conditioned Convolution (ECC) [19] and Superpoint Graph (SPG) [20] are based on the graph-based method that proposes to create convolution filters using graph weights. Since these methods can only operate on predefined weights, they have been effective only on certain data structures. Therefore, it is not a recommended method in the literature.

According to Wang et al. [21], the set-based method can be applied directly to point-level data. However, it is a method that is not preferred in semantic segmentation studies since it ignores the neighbourhood relations that contain structural information between the points.

In a CNN-based study by Su et al. [22] for object identification, a network model trained with 2D images was created to describe 3D images. The dataset known as modelnet40 was used to train the created model and 90.1 ACC was obtained. Different from the Modelnet40 dataset, which is widely used in part segmentation and classification studies, the results obtained using the Stanford Large-Scale 3D Indoor Spaces (S3DIS) [23] dataset used in the studies [24–26] for structure segmentation are detailed in "Comparison with literature findings" section.

In a study by Hackel et al. [27], a trained network was created using different datasets for the classification and

segmentation of 3D point cloud data. In this study, unlike other studies cited as a reference, a confusion matrix was also included in the evaluation. Ma et al. [28] conducted a study in which PointNet and Dynamic Graph Convolutional Neural Network (DGCNN) architectures were used together for the semantic segmentation of BIM models and point cloud data in 2020. In their study, S3DIS dataset, which consists of undeformed data, was taken as reference. For the creation of the synthetic data from restitution information, one field out of six fields was selected in this dataset, and the synthetic data was produced using 44 rooms from the chosen area. The DGCNN algorithm outstripped the PointNet algorithm in both synthetic and real point cloud data for 12 classes as ceiling, floor, wall, beam, column, window, door, chair, table, bookcase, sofa, and board.

Stasinakis et al. [29] applied a method called Generative Adversarial Networks (GAN)-based Cascaded refinement network on fragmented archaeological objects. This method was performed for self-supervised data augmentation using high-level geometry techniques and achieved successful results.

Perez-Perez et al. [30] presented an approach called Scan2BIM-NET, which is a deep learning network model used in mechanical, structural, architectural, and component segmentation. In this approach, which can be processed with Point Cloud data, two CNNs and one Recurrent Neural Network (RNN) network were used. Operations were performed on 5 different classes, namely, beam, ceiling, column, floor, pipe, and wall. In the dataset used, the average accuracy value was obtained as 86.13%.

Pierdicca et al. [31] used a deep learning network that was trained using the Architectural Cultural Heritage (ArCH) dataset to achieve semantic. In this dataset, in addition to XYZ values, Hue Saturation Value (HSV) and Red-Green-Blue (RGB) values were used for training of the proposed model called DGCNN. In this respect, it differs from the point cloud features used in the literature. This method surpassed the PointNet architecture, which has become a reference for point cloud segmentation, with 74.8% precision 74.2% recall and 72.2% f1 score.

Matrone et al. [32] proposed a hybrid method combining DGCNN, DGCNN-Mod, DGCNN-3Dfeat used in the literature. When the results of these three methods are examined; DGCNN has alone 0.37 IoU and 0.79 f1-score, while DGCNN-Mod and 3Dfeat has 0.59 IoU and 0.91 f1 score. The results were obtained using the publicly available ArCH dataset.

Model definition, analysis and conservation steps, which are important factors affecting the success of the model in deep learning studies, must be completed

correctly. Teruggi et al. [33] presented a study recommending the use of machine learning methods with the multi-level and multi-resolution (MLMR) approach. In their study, two large-scale and complex datasets were used. According to the three-level classification results made with these datasets, an f1 score of over 90% was obtained at each level.

Croce et al. [34] used heritage-building information models based on semi-automatic methods for 3D reconstruction. In these methods, the correct conversion of semantic information, the correct application of feature selection methods, data marking and conversion to the HBIM model was considered. This is one of the examples of a hybrid method that combines ML and DL methods to generate geometry in Revit BIM software successfully and ultimately outputs HBIM in IFC format.

In a study by Rodrigues et al. [35], besides the segmentation methods used in the literature, anomaly detection studies were conducted from point cloud data using known architectures such as Resnet. After various augmentations applied on the data collected as an image, conversions from image data to point cloud data were made and integrated into the BIM model. This study can be considered a reference, but it lags behind with the 0.60 f1 score in the literature.

In cases where CNN networks are not effective in terms of both time and cost in large data sets, structures called transformers can be included in the network. Liu et al. [36] proposed an architecture called TR-Net in which classification and segmentation units are defined in a transformer consisting of encoder and decoder blocks. Global features obtained from the encoder are given as input to both classification and segmentation units. According to the studies on the benchmark data, TR-Net outperformed PointNet (83.77%) and PointNet++ (85.1%) in part segmentation with a mIoU value of 85.3%.

By taking into consideration latest in the literature about point cloud segmentation with AI, this paper proposes a novel approach for increasing the accuracy of segmentation with PointNet for point cloud data of heritage buildings. Next section provides the methodology and research design for the formulation of the proposed novel approach for point cloud segmentation with PointNet at higher accuracy.

## Materials and methods

### Research methodology: case study

Heritage buildings in Turkey at risk in Gaziantep are selected as case studies, provided by the Heritage Conservation Department of the Gaziantep Metropolitan Municipality, called KUDEB that is an active partner in the project as the end user. Thus, experts from KUDEB also validate the research outcomes and the related test

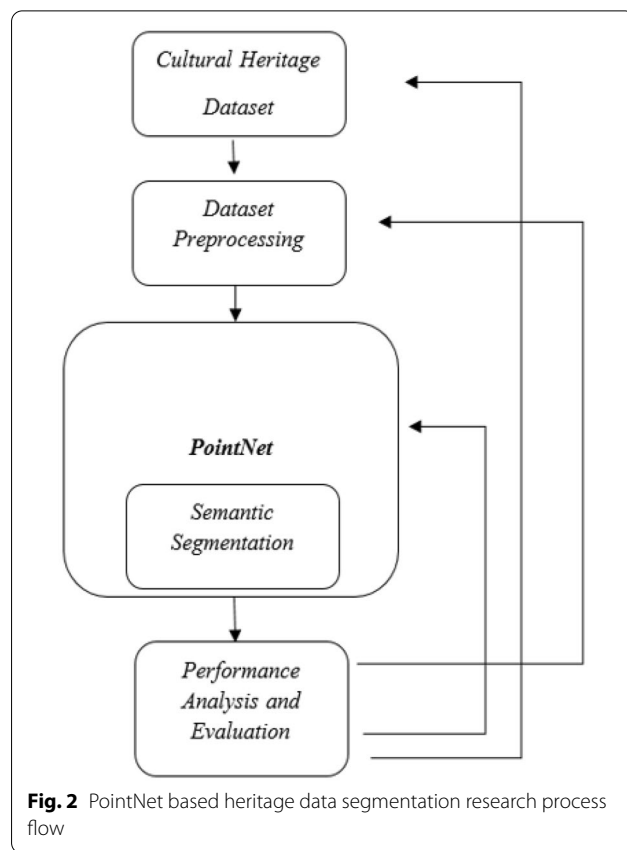
results. Images of the case studies are shown in Fig. 1. These mansions from the 16th century are the listed historic buildings in Gaziantep, reflecting the local character and identity, and their restoration has been recently completed by the Gaziantep municipality. Relevant documentations about their historical background, restitution records, restoration experience and challenges are recorded and available in KUDEB.

Point cloud data of the heritage buildings captured via terrestrial 3D laser scanner was used since it was more appropriate than airborne Light Detection and Ranging (LIDAR) in capturing the characteristic details of heritage buildings. These point cloud data will form the datasets, which will be provided by KUDEB for research and development. In this study, the segmentation study of historical-cultural structures in Gaziantep was made with our original data by improving the PointNet network [16]. Figure 2 shows the Deep Learning (DL) based research process flow.

The main problem articulated in the paper is the accurate classification and segmentation of the point cloud model for heritage buildings. Accordingly, the aim is set as the definition of a novel approach for accurate point cloud segmentation using PointNet by iterative experimentation and development. The main strategy for this is the surface-based segmentation because the intention is to categorise the mesh model of the building as: e.g., surfaces of walls, windows, doors, floors.

**Point cloud dataset**

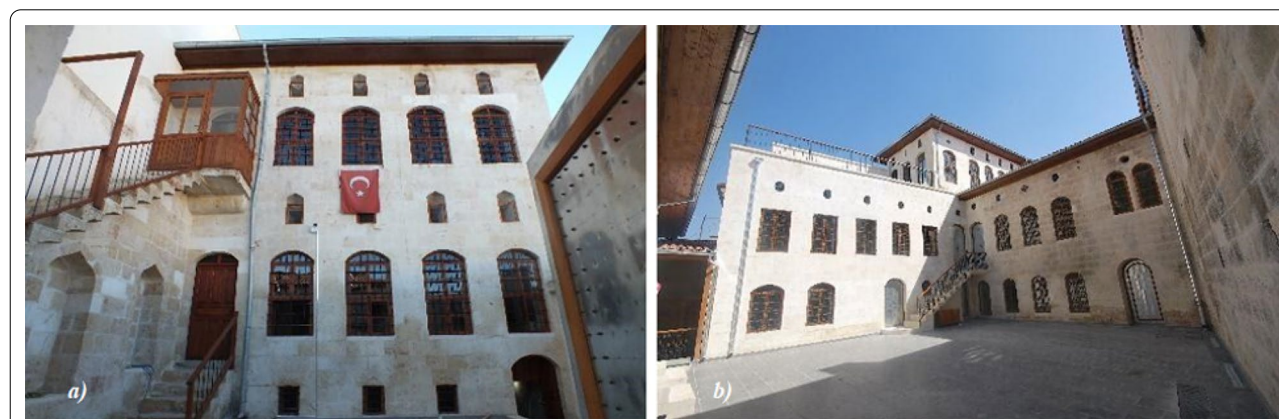
The HBIM dataset consists of 3D point clouds of historical buildings in the Gaziantep province. These data were obtained from the relevant institutions and organisations working on these structures in Gaziantep. Since the number of case study buildings was insufficient for training of the PointNet algorithm, building rooms were considered



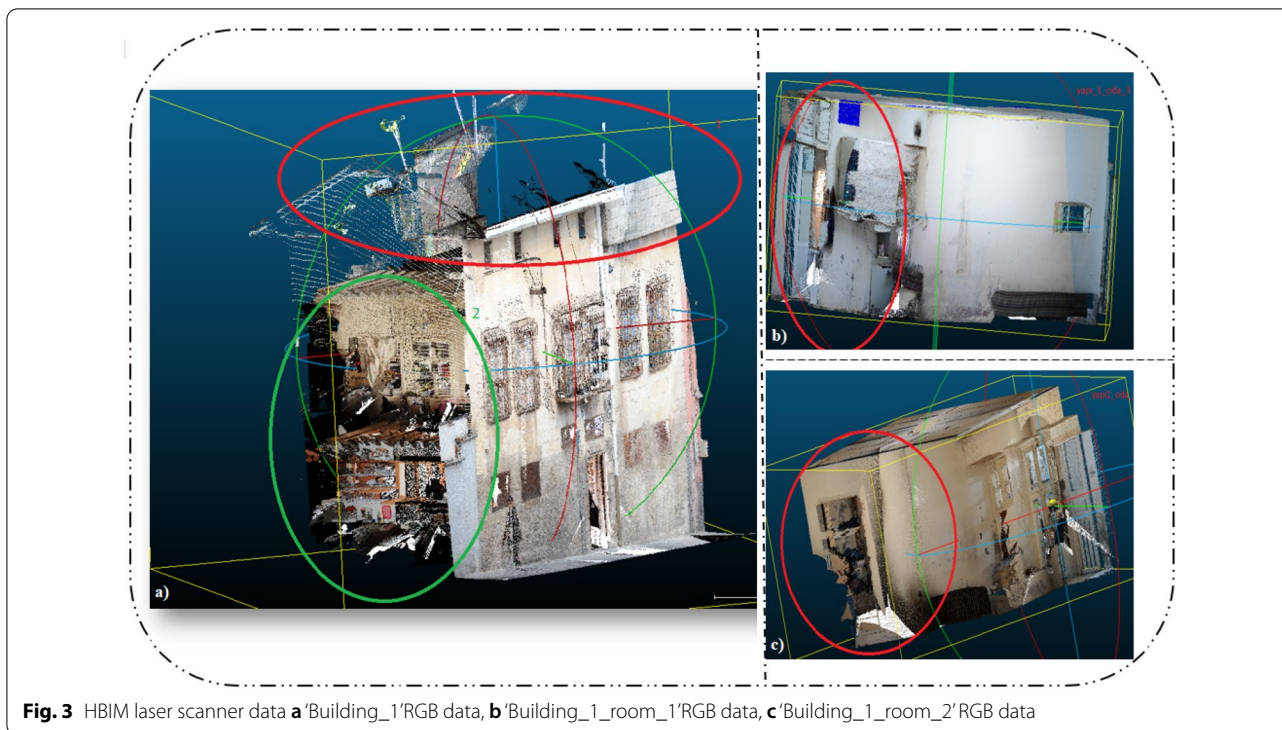
**Fig. 2** PointNet based heritage data segmentation research process flow

as the main dataset for training as this would increase the accuracy in AI training. In this way, 140 rooms were obtained from 19 historical buildings. The images of the laser scanning data of these buildings are presented in Fig. 3.

These 19 historical buildings with different numbers of rooms in each building consist of deformed point cloud data. Each room, which is processed as a single structure



**Fig. 1** Historical buildings in Gaziantep **a** Kozanlı Konak, **b** Eyüpoğlu Konak



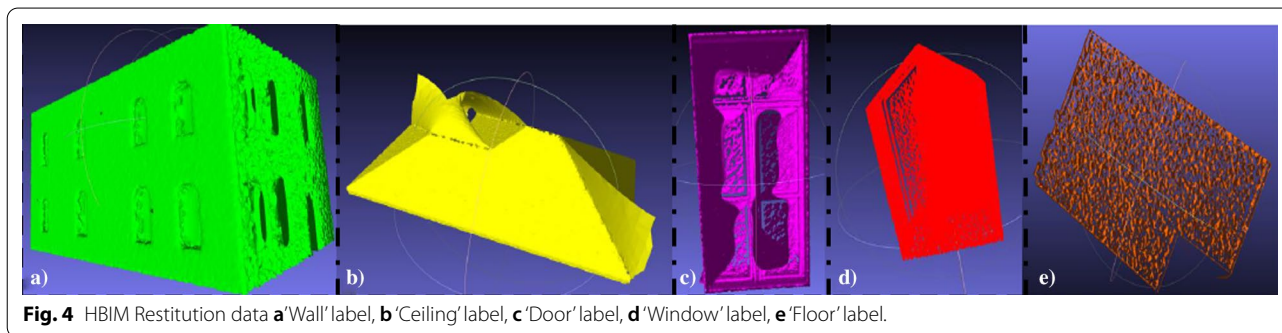
**Fig. 3** HBIM laser scanner data **a** 'Building\_1' RGB data, **b** 'Building\_1\_room\_1' RGB data, **c** 'Building\_1\_room\_2' RGB data

with the aim of increasing data, is separated from each other in terms of width, height and amount of deformation. For this reason, working on separate rooms didn't affect the model performance in terms of overfitting or underfitting. In addition, the other reason why the buildings are divided into rooms is that the existing cultural and historical building data [23, 32, and 33] do not match the deformed data discussed in this study and sufficient data cannot be obtained.

Using large number data and data diversity are important to achieve accurate results in training of deep learning models. However, HBIM dataset used in this study contains too many deformed building elements and the number of point cloud data is limited. For this reason, data generation from the restitution information of the

heritage case studies were carried out with the feedback method in a reverse engineering strategy. This reverse engineering process included the 3D BIM modelling from the restitution information, then conversion of the 3D BIM model to the 3D point cloud data for the training of the PointNet algorithm. BIM models were imported into the CloudCompare platform in FBX file format. The amount of data for the deep learning network was increased with the 11 restitution point cloud data structures (converted from 3D HBIM model to point cloud) were created and included in the system. The point cloud representations of the labels of the restitution data are given in Fig. 4.

A frequently mentioned concept to describe information richness of BIM objects is 'Level of Detail' or also



**Fig. 4** HBIM Restitution data **a**'Wall' label, **b**'Ceiling' label, **c**'Door' label, **d**'Window' label, **e**'Floor' label.

referred to as ‘Level of Development’ (LoD). LoDs allow to specify the amount of detail and generalization present in the 3D model. In this use case, the LoD of the synthetic data is an important factor as it contributes to the accuracy of the deep learning network. There are different levels of development in literature whose definitions differ in geometric accuracy, quality or completeness of semantic information. One of them, LoD200 is a design development of a product which contains geometry information [37–39]. Point cloud data contains precise geometric information such as width, length, height, and detail sizes in itself, but not semantic information and therefore the synthetic data was generated at LoD200 level like scan data. Some building examples obtained from the synthetic data generation process at LoD200 level used in DNN training are given in Table 1.

The synthetic data we call restitution data are produced by the feedback method, also known as reverse engineering. While performing the reverse engineering application, the point cloud was produced in 3 steps. These steps create 2D restitution information, 3D HBIM models from 2D restitution information, and convert these models to point clouds. This process uses survey and restitution data to train the deep learning network. Five labels for each room of the building, were determined as door, window, wall, floor, and ceiling, defined as unique building elements. The labelling process of 140 rooms and 11 restitution data used is shown in Fig. 5.

During the labelling process, the unique architectural features of these historical buildings are considered. Point cloud datasets are labelled with a point cloud processing software. First, a model was produced by giving coordinates to the corners of the labelled building elements, as in Fig. 5. However, it was determined that the model cannot be created in some building elements by only giving coordinates to the corner points. As a result, the second method was developed to perform the segmentation of building elements.

The second method is the process of location-based separation of the structural element to be segmented from the entire structure that has been laser-scanned. This process is performed by leaving the individual building elements in isolation from the whole building data and saving the isolated element as a separate file without changing or distorting its location and coordinates. The building elements were recorded by naming them according to the room and type. This way, a more accurate model was obtained by giving coordinates to each point of the defined elements.

The PointNet model trained with the classified data was implemented in the segmentation of the other point cloud data. The intersection over union (IoU) value compute method, known as the Jaccard index [40], was used

to measure and verify the performance of the segmentation process. The IoU value is a frequently used verification and measurement method of object detection [41], object segmentation [42], and definition of workspaces. This value measures the similarity between ground truth and model prediction.

The IoU calculation method is the intersection of the ground truth and the predicted area divided by the combination of these two areas, as shown in Eq. (1). Ground truth is the volume calculated using point cloud data of historical buildings.

$$\text{Score}(IoU) = \frac{\text{Area\_of\_overlap}}{\text{Area\_of\_union}} \quad (1)$$

### PointNet algorithm architecture


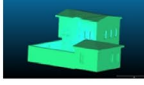


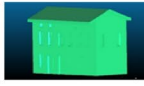











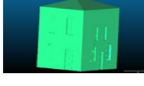
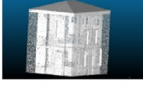



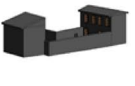
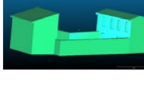





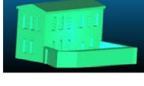





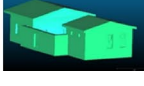

In the PointNet architecture given in Fig. 6, the input layer consists of a set of Multi-Layer Perceptrons (MLPs) that use the properties of point clouds. In the layer known as the Max Pooling layer, the symmetric properties of the input data are used, the input permutation calculations are made, and the global values of the data are calculated. Fully Connected Layers, known as the last layer, perform label prediction and classification.

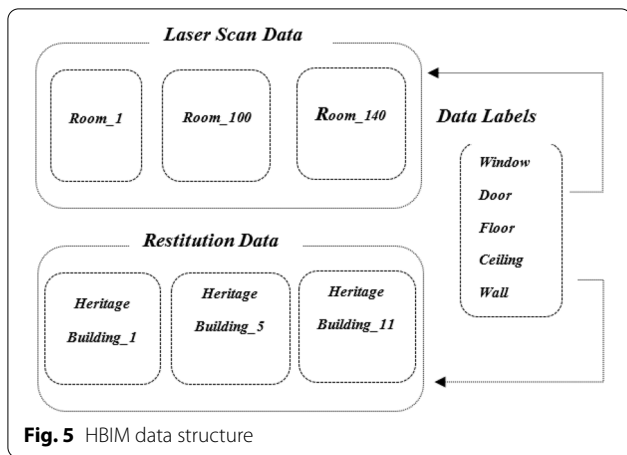
In the PointNet network, 3D data consisting of  $n$  points is taken as input. To transform the input data, the input transform and feature transform operations are performed, which enable the independent transformation of each point. The schema showing this transformation is given in Fig. 7. In the most general terms, PointNet takes a series of ( $x$ ,  $y$ , and  $z$ ) coordinate values, and each point in this coordinate array is in the form of labelled data. It is an integrated system that can classify and segment by calculations on coordinate values and determination of the surface normal values. Three basic modules make up this integrated system. These modules are explained by Qi et al. [16] as follows.

The Symmetry Function for Unordered Input module is described as ordering a set of irregular data in an understandable order, training the ordered data using the RNN network, and generating a new set of vectors using a symmetric function.

PointNet processes the  $n$  input data in an artificial neural network known as MLP to obtain regular data. After the input is transformed (64,64), it is passed through the MLP network again for the feature transformation (64,128,1024) and the input data is converted into regular information in  $n \times 1024$  dimension. It is proven in the literature that high performance is achieved with the use of RNN networks on 3D point cloud data. To create a suitable RNN network in the PointNet network, our input data must be based on a universal function. This function is shown in Eq. (2).

**Table 1** Some building examples obtained from the synthetic data generation process

Building No	HBIM Model (.rvt)	Solid Model (.fbx)	Point Cloud (.txt)
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			



**Fig. 5** HBIM data structure

$$f(\{x_1, x_n\}) \approx g(h(x_1), \dots, h(x_n)) \tag{2}$$

where,  $f : 2^{R^N} \rightarrow R, h : R^N \rightarrow R^k, g : R^k \times \dots \times R^k \rightarrow R$

An input dataset consisting of  $[f_1, \dots, f_k]$  can be used for training using an SVM (Support Vector Machine) or another classifier. However, a combination of local and global information must be used to perform point cloud segmentation.

PointNet has defined the module where it performs this operation as *Local and Global Information Aggregation*. Point features are extracted from the point inputs and a new operation is defined by using the global properties of each point in the network given in Fig. 6. In this way, combined properties consisting of new local and global information are defined for each point. Although the number of data does not change during segmentation, the input data containing more information are included in the model. Therefore, our chance of more accurate segmentation will be increased. The module called Joint Alignment Network (JAN) is included in the PointNet architecture so that the labels of the segmented point clouds are not lost after 3D grid or solid model

transformations, and to protect the segmentation. In this module, a transformation matrix is defined in a mini-network called T-Net for data transformation. This matrix is shown in Fig. 7.

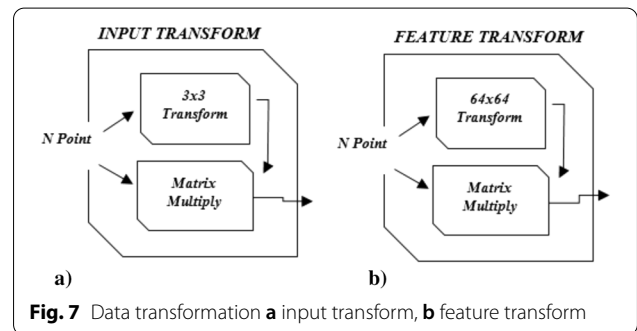
The size increases in the feature matrix due to this matrix transformation, causing the model optimization to take much more time. This issue was solved using the Softmax training function in the model. The feature transformation matrix is limited by the formula given in Eq. (3). In this way, a more stable and efficient network is obtained.

$$L_{reg} = \|I - AA^T\|_2^2 F, \tag{3}$$

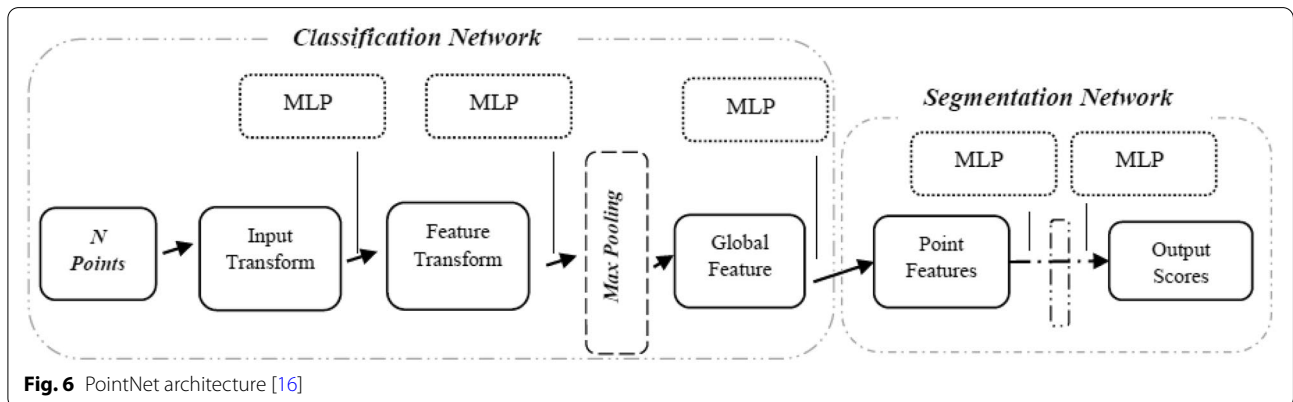
where,  $A$  is the feature alignment matrix predicted by a mini network.

**Classification and segmentation**

It is important that the data to be used in the training and testing of our deep learning network is obtained from LIDAR or laser scanning data. The *Point Interactions* operation states that if we want to obtain meaningful data from each point, the points should be evaluated together with their neighbourhoods. In the step called *Transformation invariance* shown in Fig. 8, MLP was used to increase the  $(x, y, z)$  coordinates of each point from 3

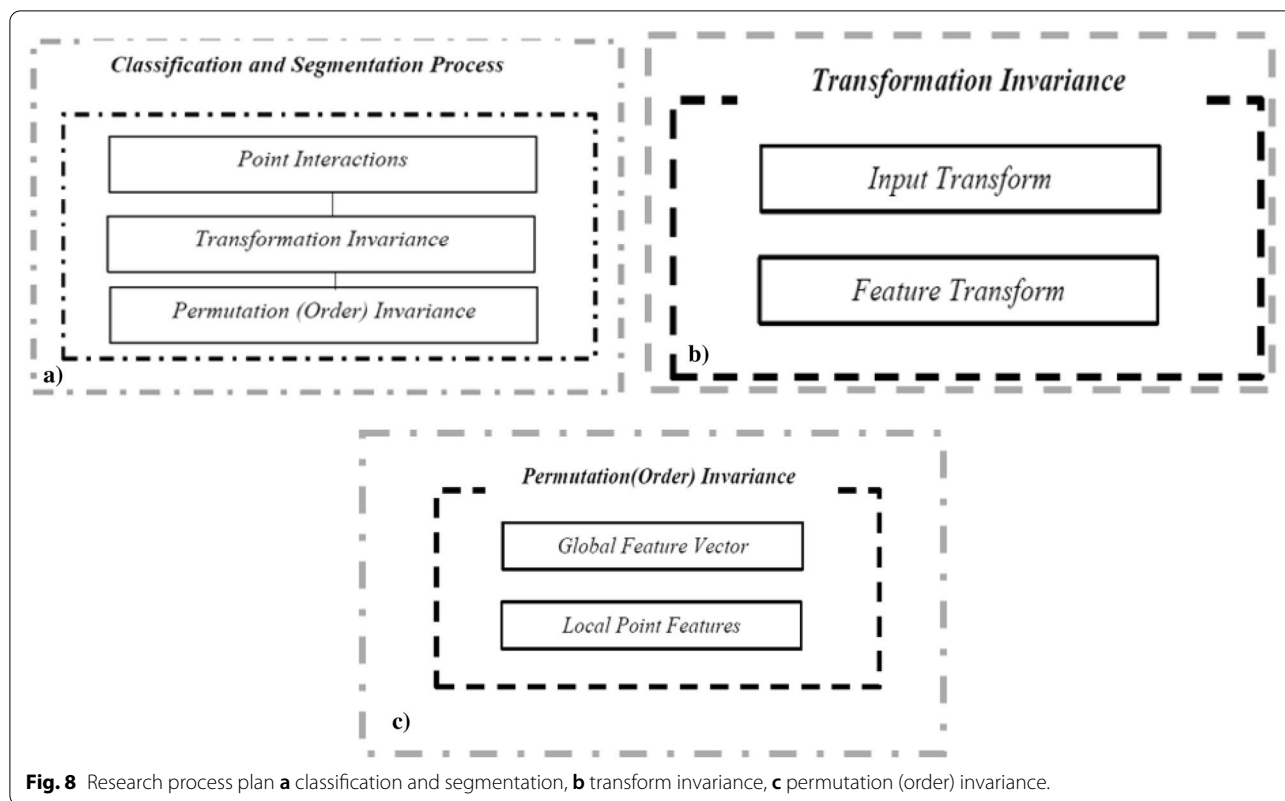


**Fig. 7** Data transformation **a** input transform, **b** feature transform



**Fig. 6** PointNet architecture [16]





dimensions to 64 dimensions and then from 64 dimensions to 1024 dimensions. These processes, detailed as Input transforms and Feature transforms in the previous sections, constitute the first stage of the classification process.

Deep learning architectures are used to directly consume point clouds and well respect the permutation invariance of points in the input capable of reasoning about 3D geometric data such as point clouds or meshes. In the step called Permutation invariance, presented in Fig. 8, an MLP network was used to obtain global features and Local Point Features. For an array containing N points, N! situation arises. N! cases must mean the same thing for a single point. Therefore, all probabilities must be based and fixed on a single function.

Global and local features are obtained as the output of the MLP network after fixation using a symmetric function. While global feature vectors are used in the classification, segmentation can be performed when used with local point features. The vector defined as  $R^{1088}$  for each point in the MLP network used in the segmentation process is converted into an array of  $nxm$  dimensions. Here, n is the number of points and m is the number of classes.

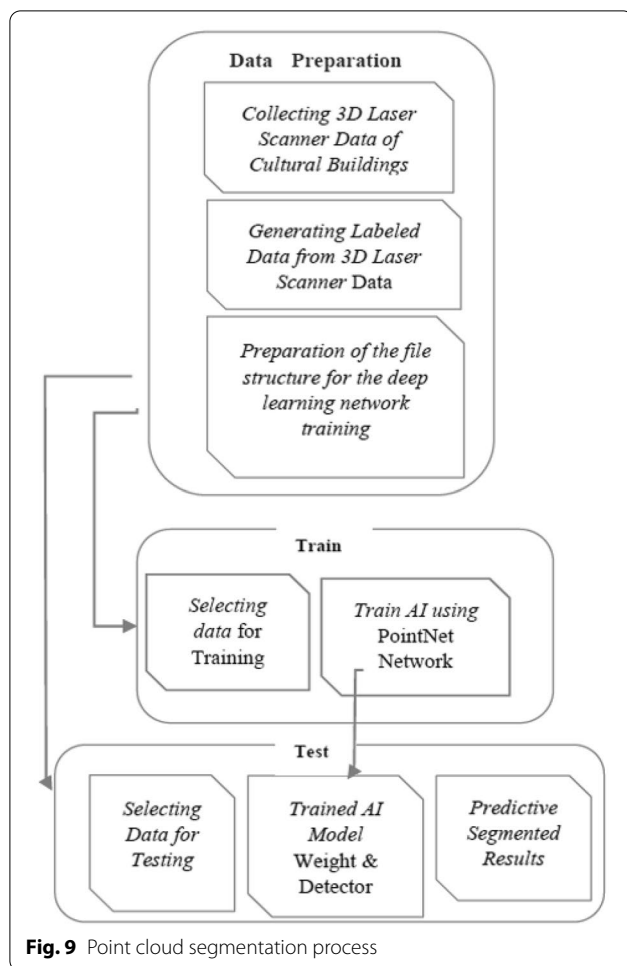
A point cloud dataset collected with 3D laser scanners was created. The objects of the dataset were labelled as doors, windows, walls, etc. This process was

a labour-intensive and manual. The dataset was divided into 3 groups for training, verification, and testing. This separation was done at 70%, 10% and 20%, respectively. Weights were created by training the training and validation dataset with the PointNet model. The test dataset with 20% rate was used to measure the test success of the trained model.

**Point cloud segmentation approach on heritage buildings with PointNet**

The point cloud dataset of Gaziantep historical buildings shown in Fig. 1 and the BIM object catalogue produced from the restitution information of historical buildings is used as Input Data for the training of the learning network. Process diagrams for processing a point cloud and performing its semantic segmentation are shown in Fig. 9. Also, Fig. 10 contains detailed information on how this process works in the HBIM integrated system.

*Input data—data preparation;* Heritage buildings scanned by 3D laser scanners were converted into point clouds data and a dataset was created. The collected 3D point cloud data was tagged and made ready for AI training. In addition, BIM models produced using the Revit program were converted into point cloud data. It was automatically tagged during the conversion process, making it ready for AI training.



*AI training:* The Point Cloud model was trained using the 3D point cloud dataset obtained from 3D the 3D scanner and 3D HBIM models. At the end of the training, historical buildings components such as doors, windows, walls, an AI-based classification weight file was obtained that recognizes the objects.

*Segmented point cloud—prediction:* The 3D point cloud data of a building scanned using a 3D laser scanner was classified with the AI decision system and the objects found on the building were classified.

### Experimental results

AI training is planned to be performed in two stages. In the first stage, labelling for data preparation will be made, and in the second stage, training and testing will be carried out using labelled data. At this stage, a significant part of the data set will be used for AI training. In the literature, 70–80% of the datasets were used for training and the remaining 20% to 30% were used for testing. Considering these rates, test and training sets will be used at the same intervals in this study. As it is known,

different functions and optimizers can be used in deep learning networks. In this study, Adam optimizer and ReLu activation function were used in the training of the model.

HBIM data consist of 19 structures with 140 rooms and 11 restitution structures. Data augmentation processes applied to increase the number of data and improve system performance are mentioned in the following sections. 3D objects belonging to HBIM data and images of these objects after segmentation are shown in Fig. 11.

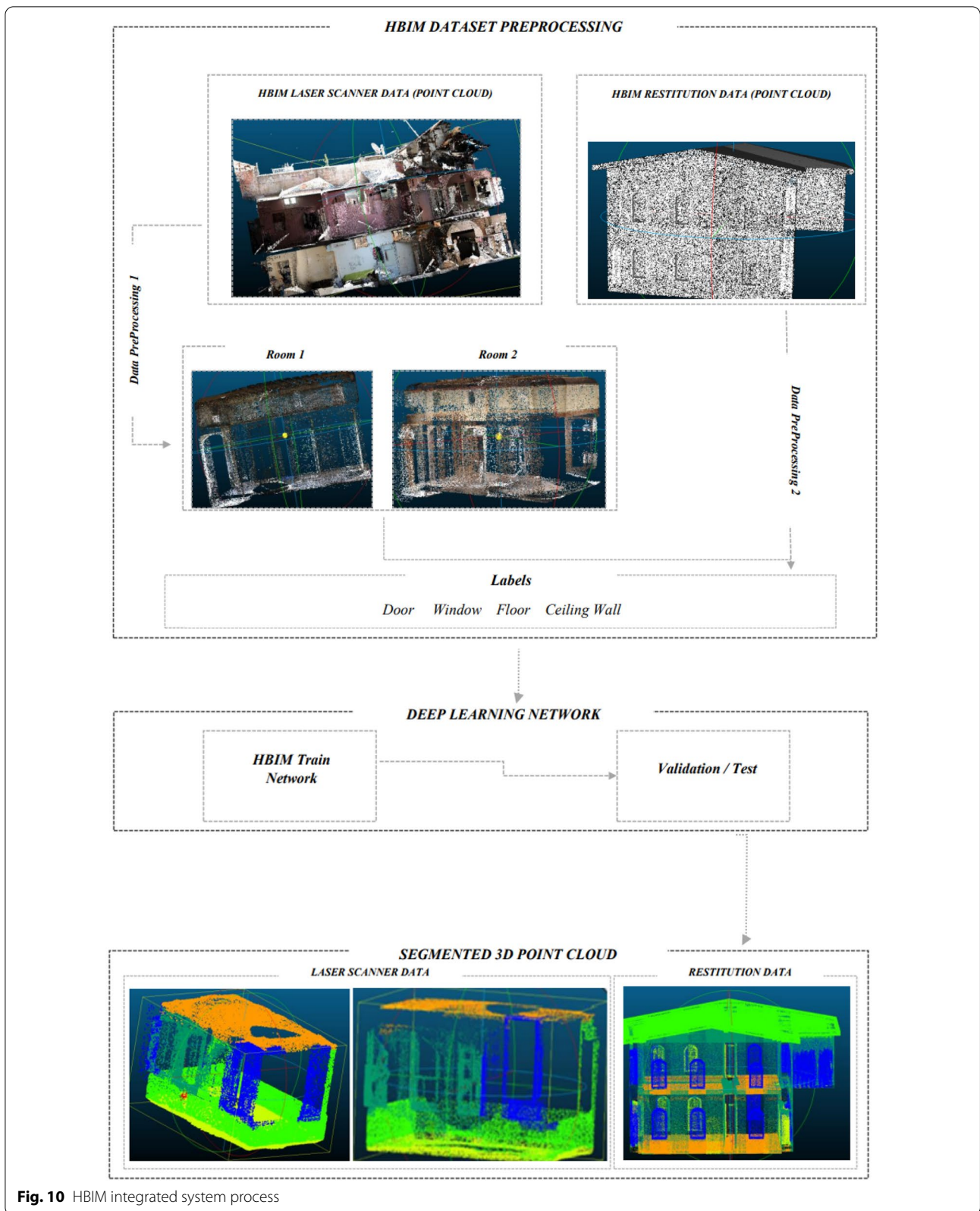
Laser scan data is named RGB data, and the result of the trained network is expressed as segmented data. The average accuracy value for these outputs is shown in Table 2. Additionally, the accuracy and loss values obtained from the results of the simulation using the data of historical buildings in Gaziantep and PointNet data obtained from Stanford University within the scope of the project are shown in detail in Table 2.

A segmented point set was obtained as the output from the test data used in the trained model. The results of all studies on the segmentation and classification of these data are given in Table 2. According to the results obtained using the original data, the model performance was 57.83% and lagged the performance obtained using PointNet.

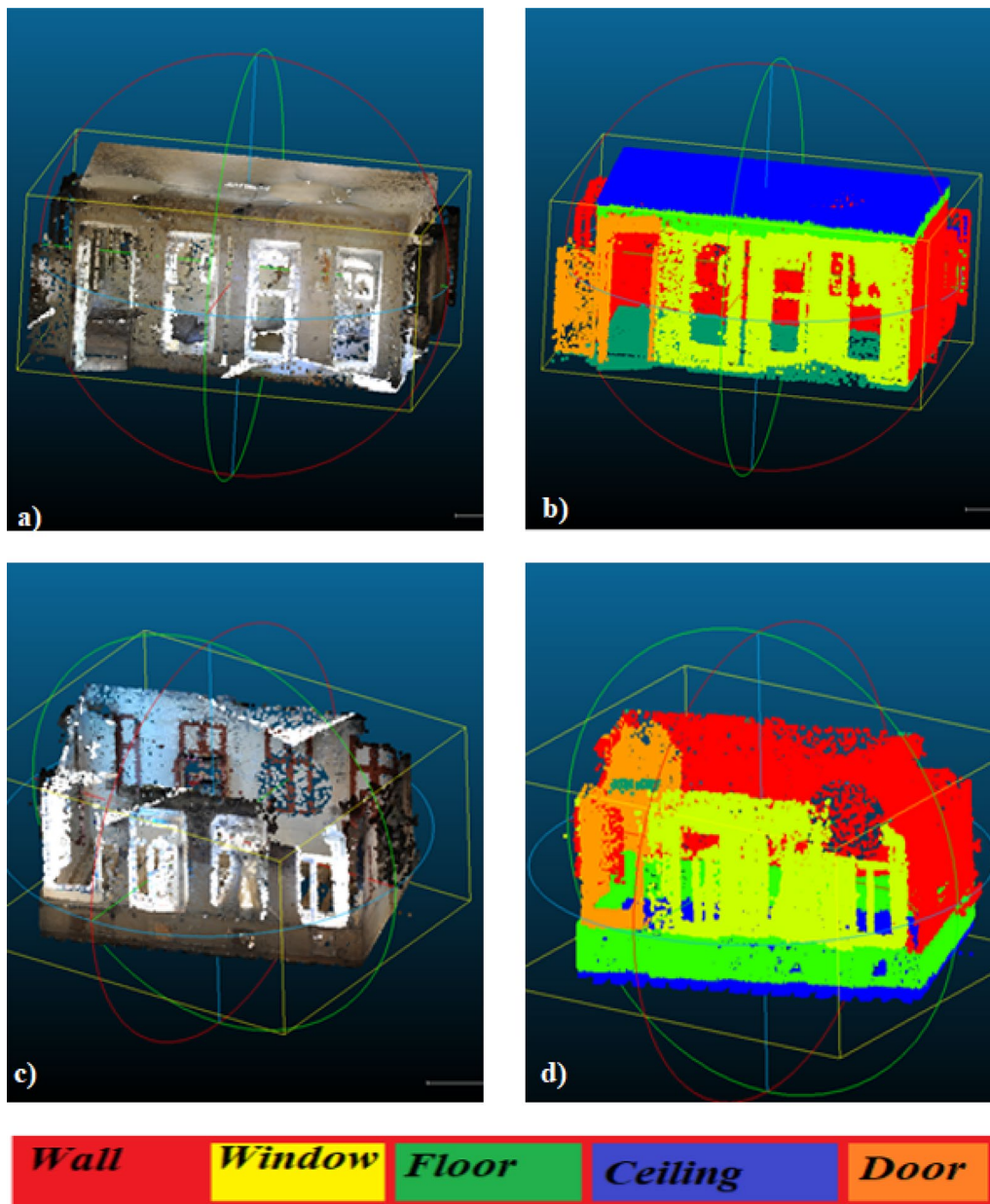
To increase the model performance, one building from the current PointNet data is included in the HBIM data set. The test accuracy has been reached 87.93%. It has been shown that the data whose coordinates can be calculated exactly increases the model performance. The accuracy and loss values obtained at each step during the training of the deep learning network are shown in Fig. 12.

In the studies to increase the Model Performance, restitution data suitable for the structure of Gaziantep historical buildings were created using the restitution data, and their segmentation and classification performances were calculated by including them in the model. After this experiment, the model performance reached 91.20% after restitution data were included in the training dataset. The purpose of creating restitution data is to determine its effect on increasing model accuracy. It has been observed that increasing the quality of the training data also increases the segmentation accuracy. The point to be considered here is to use the correct number of restitution data because the number of restitution data used can decrease the test accuracy while increasing the training accuracy. The results obtained from studies based on increasing model performance are shown in Table 3.

As can be seen, with the inclusion of restitution data in the training network, the test performance of the deformed data obtained from the laser scanner has been increased. Additionally, some of the restitution data were



**Fig. 10** HBIM integrated system process



**Fig. 11** HBIM segmented data **a** RGB 'Data\_1'; **b** Segmented RGB 'Data\_1'; **c** RGB 'Data\_2'; **d** Segmented RGB 'Data\_2'

used as test data and the same level of performance was obtained. The accuracy and loss graphs obtained using laser scanning and restitution data in the deep learning network are shown in Fig. 13.

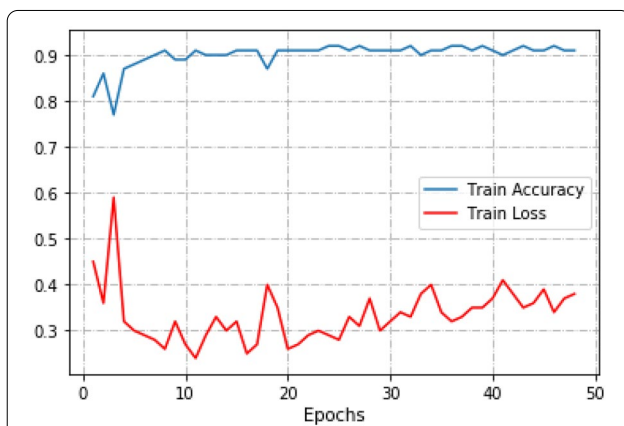
A few results of the segmentation made with restitution data created to support the original Gaziantep Cultural Heritage data are shown in Fig. 14. According to these results, 84.22% ACC was obtained from the data we named structure\_29, 85.89% test ACC from

structure\_30, and 77.23% test ACC from structure\_25. When only the given 3 structures were evaluated, the average test accuracy was 82.98%.

Segmentation results using Gaziantep original Cultural Heritage data are shown in Fig. 15. According to the segmentation result, 91.13% test ACC, 0.20 loss value was obtained using building\_1\_room\_1 and 90.70% test ACC, 41.06 loss value were obtained for building\_2\_room\_7.

**Table 2** Implementation results

Dataset	Building/ number of rooms	Dataset		Accuracy		Loss	
		Train	Test	Train	Test	Train	Test
<i>HBIM Gaziantep Cultural Heritage</i>	18 HBIM building /100 room	Area_1 Area_2 Area_3 Area_4 Area_5	Area_6 (Laser Scanner data)	0.7147	0.5783	1.1179	1.8003
<i>PointNet-Stanford</i>	6 PointNet building / 271 room	Area_1 Area_2 Area_3 Area_4 Area_5	Area_6	0.8102	0.8200	0.7072	0.7526
<i>HBIM Gaziantep Cultural Heritage</i>	18 HBIM building / 100 room + 1 PointNet building / 40 room	Area_2 Area_3 Area_4 Area_5 Area_6	Area_1 (Laser Scanner data)	0.9230	0.8793	0.1818	0.4410
<i>HBIM Gaziantep Cultural Heritage</i>	18 HBIM building/ 100 room + 1 PointNet building /40 room + 11 restitution Building	Building_6 Building_7 ... Building_21 Building_22	Building_30 (Restitution data) Building_29 (Restitution data) Building_25 (Restitution data)	0.9514	0.8652	0.3320	0.4150
				0.9514	0.8159	0.3320	0.7734
				0.9514	0.9120	0.3320	0.3250



**Fig. 12** Train metrics of HBIM network

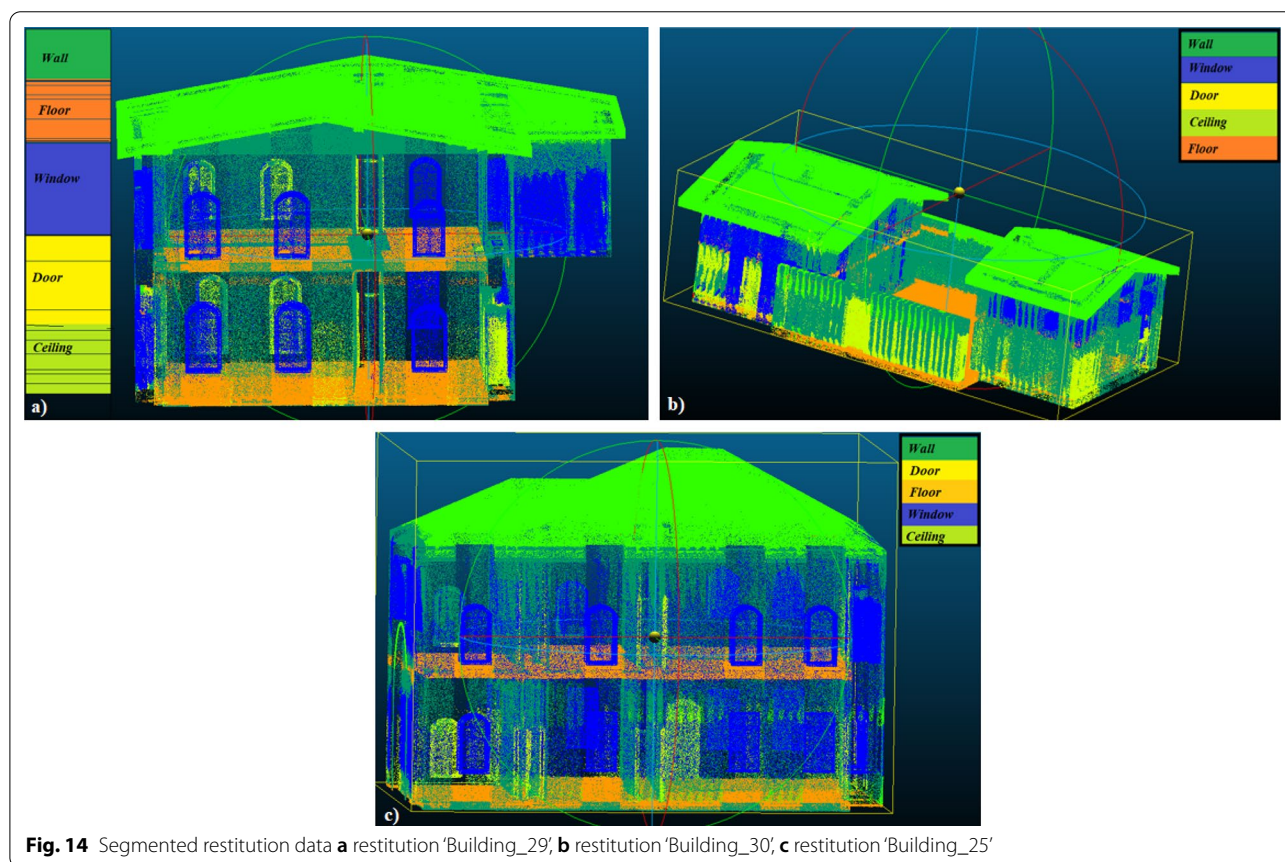
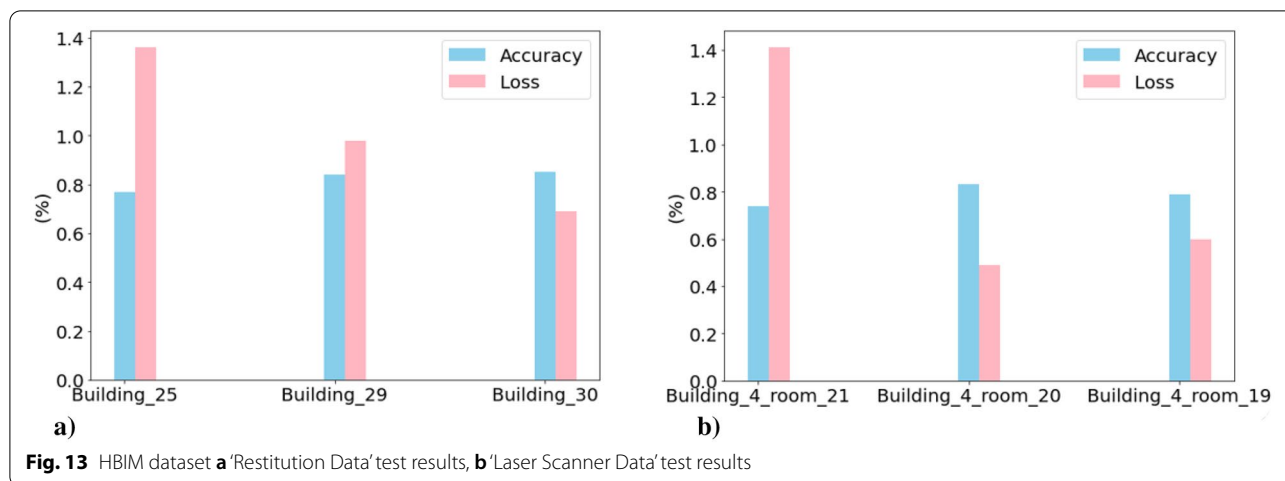
**Strength and weakness**

PointNet data is an open-source dataset presented using 271 rooms and 13 labels. We use this dataset with five labels in our study. The high number of data used in training the PointNet network and the high number of tags used in the original network (using 13 classes for segmentation) resulted in average performance of 78.62% in the PointNet study [16]. However, in our study, a Conference\_room of PointNet original data was segmented with five labels and 60% segmentation accuracy was then achieved. The expected result in the model output is segmenting the points shown in green as windows. But the model was predicted wrong and segmented as doors. This shows that even though the training performance of the model and the number of data is high, erroneous results can be obtained even with the most suitable PointNet data for the network.

In the original data of Gaziantep historical buildings used in this study, ruins and deformations have occurred over time. Coordinate losses have occurred in structures in the 1–18 range, which we call the original data due to these ruins and deformations. Due to these deformations and fractures, the test performance of the HBIM model was below average on some structures. Segmented HBIM data with 38.06% test success due to fracture-related losses. The Point Cloud data used for testing consist of completely missing coordinates. The fact that these coordinates are insufficient for the trained network directly affects the test result.

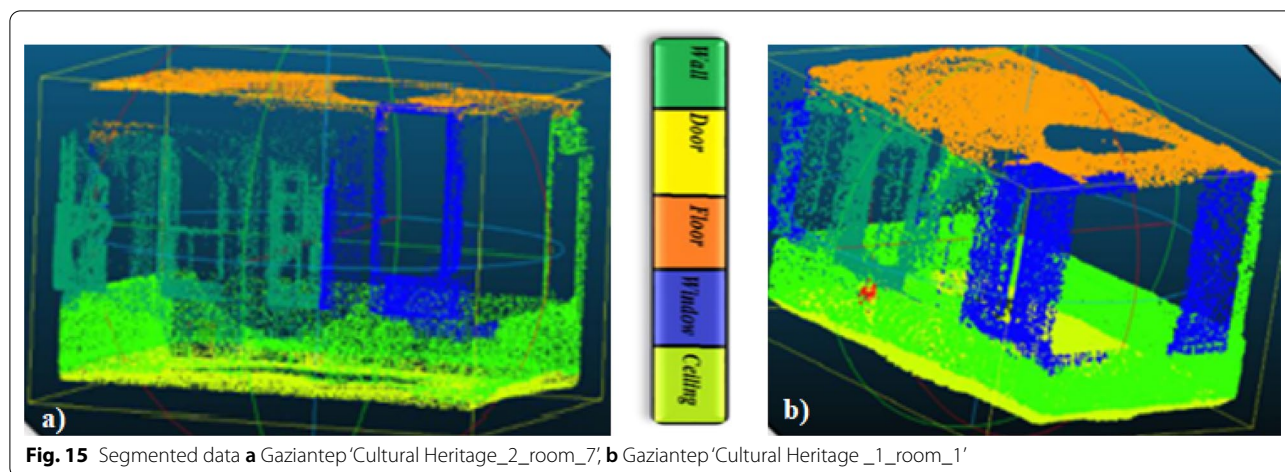
**Table 3** Comparison of the restitution data and Laser scanner data results

	Number of buildings	Test accuracy (%)	Test loss
Gaziantep cultural heritage Data	Building_4_room_21	71.40	1.41
	Building_4_room_20	83.30	0.49
	Building_4_room_19	79.56	0.60
Restitution data	Building_25	77.23	1.36
	Building_29	84.22	0.98
	Building_30	85.89	0.69



The HBIM dataset is a dataset created with the deformed data given in Figs. 11 and 15. With this aspect, it should be evaluated differently from the segmentation studies, examples of which we have seen in the literature. When working with these data, the expected result is that it has lower performance than the examples in the literature.

In this study, we proposed and implemented new methods to improve the accuracy of the segmentation results with PointNet deep learning. Most appropriate segmented data for the case study buildings were used to increase the training and test performance and to obtain the closest results to the truth. With the use of restitution data produced via reverse engineering



approach from the restitution data, the learning network was transformed into an integrated system consisting of both laser scan data of existing conditions and the restitution data obtained from produced by using the characteristics of historical buildings in Gaziantep. The results of the test obtained using the new data set consisting of laser scanning and restitution data as training data are detailed in Table 4.

The IoU value for each label of laser scanning and restitution data is given in Table 4. When these values are examined, it is seen that the IoU value of the window obtained from the laser scanning data alone is very low. However, significant increases were recorded in the IoU values obtained using laser scanning and restitution data together. The effect of restitution data on the IoU value of each label is shown in Fig. 16.

As mentioned in the previous sections, because the windows and doors are very similar both visually and in size in the deep learning network created, the desired results in these two labels could not be obtained. As seen in Fig. 16, when restitution data is used for AI

training, segmentation accuracy for windows and doors are relatively high and satisfactory.

**Comparison with literature findings**

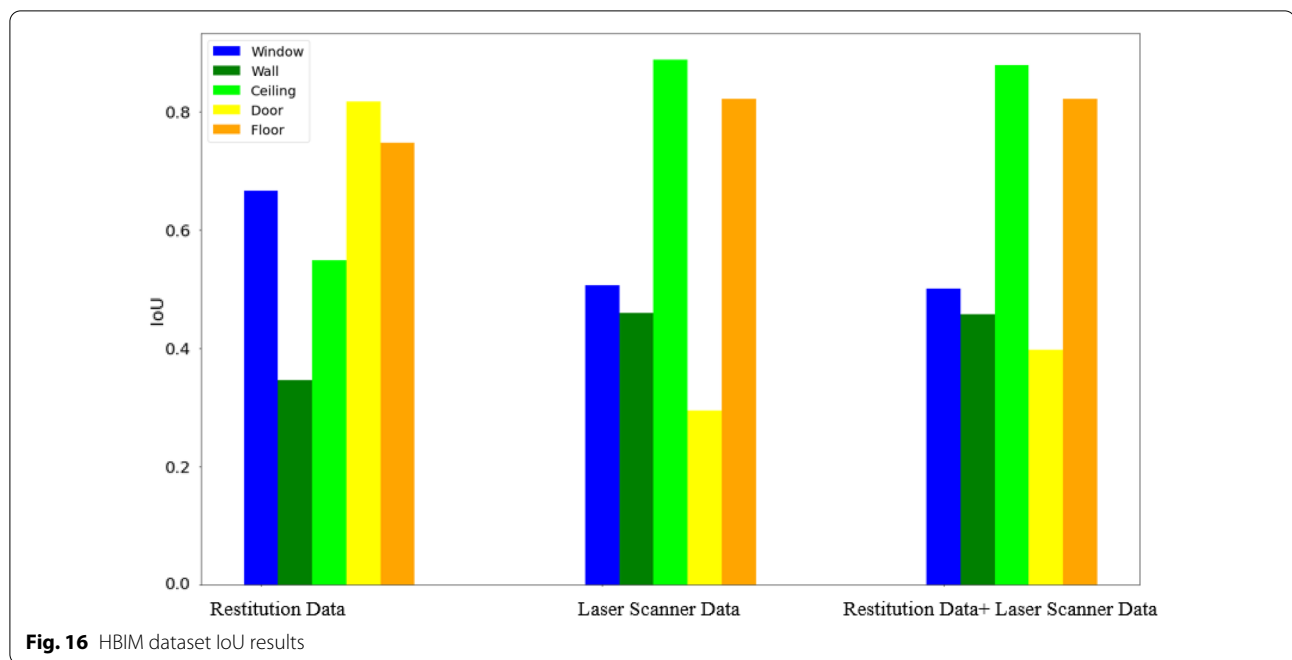
The common feature of these studies, which are referenced in the segmentation area and compared in Table 5, is that the data used are clear and clean. The results that can be obtained using the 3D point cloud datasets used in the references cited are predictable.

The dataset used in this study is 3D laser scanning data obtained from damaged historical buildings that were not used in the literature before. In addition, restitution models of damaged buildings were used, and data augmentation was performed. The HBIM model will have a unique place in the literature.

The success rate of the studies in the literature that make segmentation using the 3D Point cloud data set is listed in Table 5. It is aimed that the created list includes the comparison of accuracy values and studies using different networks with point cloud dataset. The studies in the list generally used the point cloud dataset, which

**Table 4** Comparison of the restitution data and Laser scanner data IoU results

Number of building	Types of data	Window (IoU)	Wall (IoU)	Ceiling (IoU)	Door (IoU)	Floor (IoU)	Average (IoU)
Building_25 Building_29 Building_30	Restitution data	0.5066	0.4588	0.8873	0.2948	0.8211	0.4948
Building_4_room_19 Building_4_room_20 Building_4_room_21	3D Laser Scanner data	0.0663	0.3458	0.5481	0.8165	0.7469	0.4206
Building_25 Building_29 Building_30 Building_4_room_19 Building_4_room_20 Building_4_room_21	Restitution Data + 3D Laser Scanner data	0.4999	0.4572	0.8782	0.3979	0.8208	0.5090



**Fig. 16** HBIM dataset IoU results

**Table 5** Comparison of studies

Year	Study	Method	Dataset	Performance accuracy (%)
2016	Generative and discriminative voxel modeling with convolutional neural networks	MVCNN [6]	ModelNet40	90.1
2016	Fast semantic segmentation of 3D point clouds with strongly varying density	TMLC-MSR [27]	TerraMobilita	90.28
2017	A scalable active framework for region annotation in 3D shape collections	Yi [15]	ModelNet40	81.4
2017	OctNet: learning deep 3D representations at high resolutions	Oct-Net [13]	ModelNet10	81.5
2017	Escape from cells: Deep Kd-networks for the recognition of 3D point cloud models	Kd-Net [14]	ModelNet40	82.3
2017	PointNet: deep learning on point sets for 3d classification and segmentation	PointNet [16]	ModelNet40	83.7
2017	PointNet++: deep hierarchical feature learning on point sets in a metric space	PointNet++ [17]	ModelNet40	90.7
2017	SEGCloud: semantic segmentation of 3D point clouds	SegCloud [24]	S3DIS	88.1
2017	Dynamic edge-conditioned filters in convolutional neural networks on graphs	ECC [19]	ModelNet40	82.4
2017	Unstructured point cloud semantic labeling using deep segmentation networks	SnapNet [25]	S3DIS	88.6
2017	Deep projective 3D semantic segmentation	DeePr3SS[26]	S3DIS	88.9
2018	Mining point cloud local structures by kernel correlation and graph pooling	KCNet [18]	ModelNet40	91.0
2018	Large-scale point cloud semantic segmentation with superpoint graphs	SPG [20]	S3DIS	85.5
2019	Graph attention convolution for point cloud semantic segmentation	GACNet [21]	S3DIS	87.79

is the output of the laser scanner device in the machine learning process. In our study, laser scanner data and synthetic point cloud data from the restitution HBIM models were used simultaneously in the machine learning process. In addition, the segmentation of the 3D point cloud dataset of historical heritage buildings that are not in good structural condition is the challenging part of the study. Table 5 has been created for comparison to determine the place of our study in the literature

and which gap it will fill. Thus, the most similar literature information was compared with our study.

As seen in Table 5, the literature used the point cloud data type and accuracy values ranged between 81.4% and 91.7%. Our study presents 95.14% training accuracy and 83.3% test accuracy. While the success achieved with the dataset consisting of 3D point cloud data type of structurally damaged buildings, an example of which is shown in Fig. 3, is 57.83%, this success has been increased to



83.3% with the restitution dataset. With this increase, the success of automating the pre-restoration processes by scanning historical heritage buildings with 3D laser scanners has been increased. For this reason, our study could be compared with successful studies that contributed to the literature.

## Conclusion

In the research reported in this paper, the scanned data from existing historical buildings, which are deteriorated and deformed, were used in the AI-based segmentation using PointNet. The results showed that 83.30% prediction and 95.14% training accuracy was achieved even though the scanned data did not contain sufficient information about the structure due to the deformations in the buildings. Segmentation of point cloud data for historic buildings can be challenging and AI-based algorithms can be insufficient due to these historic buildings' unique and deformed conditions. However, preparing training data set from the restitution information of the historic building that is called restitution data in this research helps significantly for high accurate segmentation. This restitution data and laser scanning data were used together for segmentation of five components (windows, doors, wall, ceiling and floor). The reason for five components is because the case study heritage buildings are deteriorated and deformed from which the segmented results were still satisfactory.

The results show that the combined use of restitution data and existing conditions data together would be the way forward for the point cloud segmentation with AI for heritage structures belonging to the same period. Therefore, the research will expand further by identifying other minor components in the case study buildings by preparing a training dataset for the algorithm towards enhanced and detailed segmentation with higher accuracy. In addition, PointNet++ [17], an improved system of PointNet [16], can provide better segmentation performance with proposed approach. Therefore, as an expansion from the current research, PointNet++ will also be considered to improve the segmentation as part of the research plan on R-CNN and Fast R-CNN networks to incorporate the unlabelled data into the HBIM network.

## Abbreviations

HBIM: Heritage Building Information Modelling; S3DIS: Stanford Large-Scale 3D Indoor Spaces Dataset; DL: Deep learning; IoU: The intersection over union; MLP: Multi-layer perceptron; SVM: Support Vector Machine; JAN: Joint Alignment Network.

## Acknowledgements

This paper is written within the TUBITAK 1001 project (Grant number: 119Y038) and the authors would like to acknowledge The Scientific and Technological Research Council of Turkey (TUBITAK) for their support.

## Author contribution

BH: supervision, methodology, writing, original draft preparation. RB: conceptualization, methodology, validation, software. AEO: supervision, conceptualization, methodology, validation, software. YA: writing, reviewing and editing. All authors read and approved the final manuscript.

## Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

## Availability of data and materials

The data will be available upon reasonable request.

## Declarations

### Competing Interests

The Authors declare no conflict of interests.

### Author details

<sup>1</sup>Department of Computer Engineering, Gaziantep University, Gaziantep, Turkey. <sup>2</sup>AI Enablement, Huawei Turkey R&D Center, Umraniye, İstanbul, Turkey.

<sup>3</sup>Department of Electric-Electronic Engineering, Hasan Kalyoncu University, Gaziantep, Turkey. <sup>4</sup>Department of Architecture and Built Environment, Northumbria University, Newcastle, UK.

Received: 22 August 2022 Accepted: 11 December 2022

Published online: 04 January 2023

## References

- Gonzalez RC, Woods RE. Digital image processing. Addison-Wesley Publishing Company; 1993.
- Dubb D, Zell A. Real-time plane extraction from depth images with the randomised hough transform. In: Proceedings of the IEEE international conference on computer vision workshops (ICCV Workshops). 2011. p. 1084–1091.
- Zhu H, Meng F, Cai J, Lu S. Beyond pixels: a comprehensive survey from bottom-up to semantic image segmentation and cosegmentation. *J Vis Commun Image Represent*. 2016;34(5):12–27.
- Truc L, Duan Y. PointGrid: a deep network for 3D shape understandings. In 2018 IEEE/CVF conference on computer vision and pattern recognition. 2018. p. 9204–9214.
- Wang P, Liu Y, Guo Y, Sun C, Tong X. O-CNN: octree-based convolutional neural networks for 3D shape analysis. *ACM Trans Graphics*. 2017;36(4):1–11.
- Qi CR, Su H, Nießner M, Dai A, Yan M, Guibas LJ. Volumetric and multi-view CNNs for object classification on 3D data. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2016. p. 5648–5656.
- Le T, Giang B, Duan Y. A multi-view recurrent neural network for 3D mesh segmentation. *Comput Graph*. 2017;66:103–12.
- Bronstein MM, Bruna J, LeCun Y, Szlam A, Vandergheynst P. Geometric deep learning: going beyond euclidean data. *IEEE Signal Process Mag*. 2017;34(4):18–42.
- Yi L, Su H, Guo X, Guibas L. SyncSpecCNN: synchronised spectral CNN for 3D shape segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 6584–6592.
- Niepert M, Ahmed M, Kutzkov K. Learning convolutional neural networks for graphs. In: Proceedings of the 33rd international conference on machine learning. 2016. p. 2014–2023.
- Xie Y, Tian J, Zhu XX. Linking points with labels in 3D: a review of point cloud semantic segmentation. *IEEE Geosci Remote Sens Mag*. 2020;8(4):38–59.
- Wang Z, Liu H, Yueliang Q, Xu T. Real-time plane segmentation and obstacle detection of 3D point clouds for indoor scenes. In: Fusiello A, Murino V, Cucchiara R, editors. European conference on computer vision (ECCV). 2012. p. 22–31.

13. Riegler G, Ulusoy AO, Geiger A. OctNet: learning deep 3D representations at high resolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2017. p. 6620–6629.
14. Klokov R, Lempitsky V. Escape from cells: deep kd-networks for the recognition of 3D point cloud models. In: Proceedings of the IEEE international conference on computer vision (ICCV). 2017. p. 863–872.
15. Yi L, et al. A scalable active framework for region annotation in 3D shape collections. *ACM Trans Graph*. 2016;35(6):1–12.
16. Charles RQ, Su H, Kaichun M, Guibas LJ. PointNet: deep learning on point sets for 3D classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 77–85.
17. Charles RQ, Yi L, Su H, Guibas LJ. PointNet+: deep hierarchical feature learning on point sets in a metric space. In: Proceedings of the 31st international conference on neural information processing systems. 2017. p. 5105–5114.
18. Shen Y, Feng C, Yang Y, Tian D. Mining point cloud local structures by kernel correlation and graph pooling. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. p. 4548–4557.
19. Simonovsky M, Komodakis N. Dynamic edge conditioned filters in convolutional neural networks on graphs. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2017. p. 29–38.
20. Landrieu L, Simonovsky M. Large-scale point cloud semantic segmentation with super point graphs. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. p. 4558–4567.
21. Wang L, Huang Y, Hou Y, Zhang S, Shan J. Graph attention convolution for point cloud semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2019. p. 10288–10297.
22. Su H, Maji S, Kalogerakis E, Learned-Miller E. Multi-view convolutional neural networks for 3D shape recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 945–953.
23. Retrieved from papers with code. <https://paperswithcode.com/dataset/s3dis>. Accessed 2022.
24. Tchapmi L, Choy C, Armeni I, Gwak J, Savarese S. SEGCloud: semantic segmentation of 3D point clouds. In: 2017 international conference on 3D vision (3DV). 2017. p. 537–547.
25. Boulch A, Saux BL, Audebert N. Unstructured point cloud semantic labeling using deep segmentation networks. In: Eurographics workshop on 3D object retrieval. 2017. p. 17–24.
26. Lawin FJ, Danelljan M, Tosteberg P, Bhat G, Khan FS, Felsberg M. Deep projective 3D semantic segmentation. In: Felsberg M, Heyden A, Krüger N, editors. Computer analysis of images and patterns. 2017. p. 95–107.
27. Hackel T, Wegner JD, Schindler K. Fast semantic segmentation of 3D point clouds with strongly varying density. *ISPRS Ann Photogramm Remote Sens Spat Inf Sci*. 2016;III–3:177–84.
28. Ma JW, Czerniawski T, Leite F. Semantic segmentation of point clouds of building interiors with deep learning: augmenting training datasets with synthetic BIM-based point clouds. *Autom Constr*. 2020;113:103144.
29. Stasinakis A, Chatzilaris E, Nikolopoulos S, Kompatsiaris I, Karolidis D, Touloumtzidou A, Tzetzis D. A hybrid 3D object auto-completion approach with self-supervised data augmentation for fragments of archaeological objects. *J Cult Herit*. 2022;56:138–48.
30. Perez-Perez Y, Golparvar-Fard M, El-Rayes K. Scan2BIM-NET: deep learning method for segmentation of point clouds for scan-to-BIM. *J Constr Eng Manag*. 2021;147(9):04021107.
31. Pierdicca R, Paolanti M, Matrone F, Martini M, Morbidoni C, Malinverni ES, Lingua AM. Point Cloud semantic segmentation using a deep learning framework for cultural heritage. *Remote Sens*. 2020;12(6):1005.
32. Matrone F, Grilli E, Martini M, Paolanti M, Pierdicca R, Remondino F. Comparing machine and deep learning methods for large 3D heritage semantic segmentation. *ISPRS Int J Geo-Inf*. 2020;9(9):535.
33. Teruggi S, Grilli E, Russo M, Fassi F, Remondino F. A hierarchical machine learning approach for multi-level and multi-resolution 3D point cloud classification. *Remote Sens*. 2020;12(16):2598.
34. Croce V, Caroti G, De Luca L, Jacquot K, Piemonte A, Véron P. From the semantic point cloud to heritage-building information modeling: a semiautomatic approach exploiting machine learning. *Remote Sens*. 2021;13(3):461.
35. Rodrigues F, Cotella V, Rodrigues H, Rocha E, Freitas F, Matos R. Application of deep learning approach for the classification of buildings' degradation state in a BIM methodology. *Appl Sci*. 2022;12(15):7403.
36. Liu L, Chen E, Ding Y. TR-Net: a transformer-based neural network for point cloud processing. *Machines*. 2022;10(7):517.
37. Morbidoni C, Pierdicca R, Paolanti M, Quattrini R, Mammoli R. Learning from synthetic point cloud data for historical buildings semantic segmentation. *J Comput Cult Herit*. 2020;13(4):1–16.
38. Mengqi Z, Yan T. Exploring spatiotemporal changes in cities and villages through remote sensing using multibranch networks. *Herit Sci*. 2021;9(1):1–15.
39. Dong Y, Li Y, Hou M. The point cloud semantic segmentation method for the Ming and Qing Dynasties' official-style architecture roof considering the construction regulations. *Int J Geo-Inf*. 2022;11(4):214.
40. Jaccard P. The distribution of the flora in the alpine zone. *New Phytol*. 1912;11(2):37–50.
41. Xu J, Ma Y, He S, Zhu J. 3D-GIoU: 3D generalised intersection over union for object detection in point cloud. *Sensors*. 2019;19(19):4093.
42. Hou F, Lei W, Li S, Xi J, Xu M, Luo J. Improved mask R-CNN with distance guided intersection over union for GPR signature detection and segmentation. *Autom Constr*. 2021;121(1):103414.

### Publisher's Note

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

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)