

# GEOMETRIC PRIMITIVE EXTRACTION FROM SEMANTICALLY ENRICHED POINT CLOUDS

S. Kyriakaki-Grammatikaki<sup>1</sup>, E. K. Stathopoulou<sup>1,2</sup>, E. Grilli<sup>2</sup>, F. Remondino<sup>2</sup>, A. Georgopoulos<sup>1</sup>

<sup>1</sup> Laboratory of Photogrammetry, School of Rural, Surveying and Geomatics Engineering, National Technical University of Athens, Greece - Email: stavroula.kiriakaki@yahoo.gr, drag@central.ntua.gr

<sup>2</sup> 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy  
Web: <http://3dom.fbk.eu> – Email: (estathopoulou, grilli, remondino)@fbk.eu

## Commission II

**KEY WORDS:** 3D point cloud, supervised learning, segmentation, primitive extraction, HBIM, cultural heritage

### ABSTRACT:

3D point clouds are robust representations of real-world objects and usually contain information about the shape, size, position and radiometry of the scene. However, unstructured point clouds do not directly exploit the full potential of such information and thus, further analysis is commonly required. Especially when dealing with cultural heritage objects which are, typically, described by complex 3D geometries, semantic segmentation is a fundamental step for the automatic identification of shapes, erosions, etc. This paper focuses on the efficient extraction of semantic classes that would support the generation of geometric primitives such as planes, spheres, cylinders, etc. Our semantic segmentation approach relies on supervised learning using a Random Forest algorithm, while the geometric shapes are identified and extracted with the RANSAC model fitting algorithm. In this way the parametric modelling procedure in a HBIM environment is easily enabled. Our experiments show the efficient label transferability of our 3D semantic segmentation approach across different Doric Greek temples, with qualitatively and quantitatively evaluations.

## 1. INTRODUCTION

Data acquisition techniques such as photogrammetry and laser scanning commonly generate 3D point clouds to describe the surface of a real-world object or a scene. Such representations are mathematically expressed by matrices of unorganised points; the rows of these matrices correspond to the total number of points contained in the 3D cloud and the columns contain point-level information such as coordinates, normal direction, colour, intensity etc. This kind of data can be sufficient for general 3D recording purposes, yet it does not provide any semantically meaningful attribute of the scene as such. Higher-level semantic information is indeed crucial towards scene understanding and analysis and, hence, semantic segmentation has become a powerful tool for 2D (Marmanis et al., 2016; Chen et al., 2018; Kirillov et al., 2019) as well as 3D (Blaha et al., 2016; Armenti et al., 2017; Stathopoulou et al., 2021) data analysis. Particularly in the field of cultural heritage, semantic segmentation has found applicability towards the identification and grouping of points with similar attributes with respect to their geometry, colour, material etc. (Grilli and Remondino, 2020; Croce et al., 2021). In various mapping and cultural heritage applications, this semantically enhanced data is further associated with external metadata, i.e. additional specific information regarding the scene, contributing towards a more holistic data representation (Carboni et al., 2016). Moreover, in the cultural heritage domain, the use of Historic Building Information Modelling (HBIM) has become a research topic of great interest as it is able to model the state of complex historic structures throughout their life cycle by deconstructing and analysing their different components and details, and several studies have worked towards this scope (Barazzetti, 2016; Murtiyoso and Grussenmayer, 2019; Yang et al., 2020a). As a result, HBIM could act as a multi-dimensional tool and process essential for the management, preservation, restoration and dissemination of cultural heritage.

### 1.1. Aim of the paper

This paper introduces an integrated pipeline which, starting from a point cloud generated by passive or active sensors, delivers

primitive models that will enable and facilitate an automated modelling procedure in HBIM. Point groups with similar attributes are first automatically segmented using supervised machine learning methods. Subsequently, geometric primitives are fitted on each semantic group, using the RANSAC model fitting algorithm (Figure 1). The extracted primitive shape information can be used as guidance in an HBIM environment for parametric shape generation in order to reduce manual labour in the scan-to-BIM pipeline and create parametric objects from point clouds (Murphy et al., 2013; Macher et al., 2017; Capone and Lanzara, 2019). Our work focuses on ancient Greek temples of Doric order. For this reason, the temple of Hephaestus in the ancient Agora of Athens is selected as a case study. Our segmentation step follows Grilli et al. (2019), since that work proved to be effective in a similar case study, the temple of Neptune in Paestum, Italy, also known as the 2<sup>nd</sup> temple of Hera. Compared to previous works, we aim to evaluate the transferability and generalisation of a 3D semantic segmentation method to similar architectures, yet exhibiting different details and characteristics, while being independent from the acquisition method.

## 2. RELATED WORKS

**3D Semantic segmentation.** Semantic segmentation in the 3D space can be defined as the process of assigning a semantically meaningful label to each point. Semantic segmentation can separate, group, highlight and extract clusters of points with similar attributes and consequently objects across a point cloud. Although semantic segmentation algorithms of 2D images have been proven to be robust enough in recent years, with high-performance scores using machine and deep learning algorithms (Plath et al., 2009; Long et al., 2015; Chen et al., 2017), the respective algorithms for 3D data are still an open challenge. Segmentation of 3D data indeed requires complex mathematical operations and huge computational power, while the available training data is not as extended as their 2D equivalents. Acknowledging this open issue, the community has conducted extensive research on this topic in the last years using simple statistical or machine learning algorithms. Earlier works studied

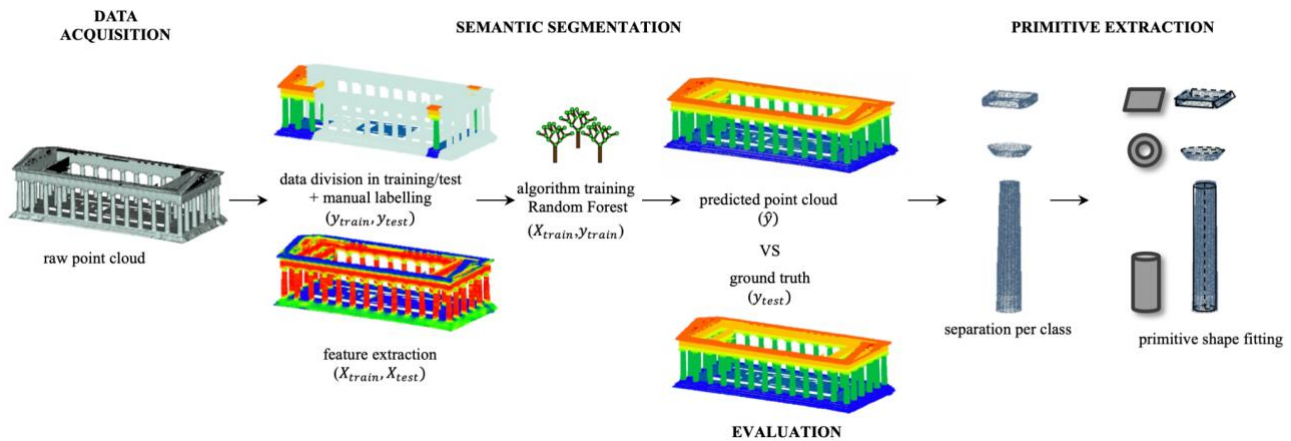


Figure 1. The proposed pipeline integrates 3D semantic segmentation and primitive shape extraction with model fitting to facilitate the inclusion of 3D surveying data into HBIM solutions.

airborne LiDAR data for road extraction (Boyko and Funkhouser, 2011), forest mapping (Dalponte et al., 2012) or urban mapping applications (Rottensteiner and Briese, 2002). The features', or descriptors', design is a step of utmost importance in semantic segmentation and research has been focused on the optimal design and combination of them along with the optimal neighbourhood definition (Weinmann et al., 2013; Hackel et al., 2016; Weinmann et al., 2017). With the seminal work of (Qi et al., 2017a; Qi et al., 2017b), deep neural networks also started to be used for the semantic segmentation in 3D. For airborne data, segmentation based on deep learning has several successful implementations (Yousefhusien et al., 2018; Huang et al., 2020; Özdemir et al., 2021). However, in the field of cultural heritage, the number of such applications is limited, mainly due to the complexity of the scenarios and the few available data (Fiorucci et al., 2020). Thus, most works focused on segmentation based on simple geometric features (Murtiyoso and Grussenmeyer, 2019) or supervised machine learning methods (Grilli et al., 2019; Grilli and Remondino 2020). Some first attempts at using deep learning algorithms for cultural heritage scenes have also been recently employed (Matrone et al., 2020; Pierdicca et al., 2020).

**Geometric model fitting.** Geometric primitive extraction can be defined as the process of grouping points that belong to the same primitive shape. Geometric primitives refer to the simplest shapes appearing in the 3D space such as planes, cubes, spheres, cylinders, cones, tori etc. and are primarily easy to understand and parameterise. Based on these ubiquitous primitives, complex man-made objects can be explained, decomposed and simplified, facilitating the identification of an object's form in an abstract way. Simple plane primitives can be detected in point clouds using 3D connected components analysis based on topology (Verma et al., 2006) or region-growing methods (Huber et al., 2011). The Hough transform (Hough 1962) is another, more complex method that enables the detection of both planes and higher-order shapes in 3D point clouds (Vosselman et al., 2004; Tarsha-Kurdi et al., 2007; Bosché et al., 2015). A particularly popular method due to its robustness to outliers is the Random SAmple Consensus (RANSAC) by (Fischler and Bolles, 1981) used for aerial (Tarsha-Kurdi et al., 2007; Xu et al., 2016), terrestrial (Yang et al., 2020b), as well as indoor data to facilitate the Scan-to-BIM process (Jung et al., 2014; Thomson and Boehm, 2015; Hong et al., 2015). However, complex curved surfaces may need more elaborated parametrisation approaches such as NURBS curve fitting (Barazzetti, 2016). Croce et al. (2021) propose an HBIM-oriented semi-automatic procedure combining semantic segmentation and primitive extraction.

Similarly to this work, we employ semantic segmentation, focusing on transferability across case studies. Moreover, in order to reduce manual labour, we implement an automatic geometric primitive extraction module and integrate it in the pipeline, aiming to facilitate parametric shape generation in HBIM.

### 3. METHODOLOGY

In this paper, we propose an operational pipeline that delivers geometric primitives from a raw, unstructured 3D point cloud (Section 3.1), enabling thus, HBIM modelling. To this end, we consider two interlinked steps that involve the enrichment of the raw 3D point clouds with semantic information (Section 3.2), and the subsequent extraction of modelling parameters to create geometric primitives (Section 3.3).

#### 3.1 Datasets

The employed 3D point clouds refer to ancient Greek temples of the Doric order (Figure 2): the temple of Neptune in Paestum, Italy (Temple 1) and the temple of Hephaestus in Athens, Greece (Temple 2). These complex and large structures can be decomposed according to their different architectural components, each of which has its own specific semantic property and geometry. The two temples were surveyed with different techniques, namely photogrammetry for Temple 1 (ca 0.01 m average point spacing) and laser scanning for Temple 2 (ca 0.03 m average point spacing). Both Doric temples have similar structural and architectural elements. However, the temple of Neptune (Temple 1) is approximately twice the size of the temple of Hephaestus (Temple 2). In addition, the material used for their construction is different, i.e., limestone for Temple 1 and marble for Temple 2, thus leading to varying decay states, which eventually affects the form of the architectural elements. Only in the temple of Hephaestus, the cella is preserved with its walls still standing. The eastern wall has been slightly altered, as a result of a later intervention, and is covered by the original roof. It is the best-preserved Doric temple in the world.

#### 3.2 Point cloud segmentation using machine learning

The segmentation methodology is based on Grilli et al. (2019), where supervised machine learning techniques were applied to infer per-point labels on the point cloud of the temple of Neptune in Paestum. Following Weinmann et al. (2015), the supervised semantic segmentation of point clouds consists of: (1) class determination, (2) data annotation/labelling, (3) point neighbourhood selection, (4) feature extraction, (5) algorithm training and (6) label inference and evaluation.

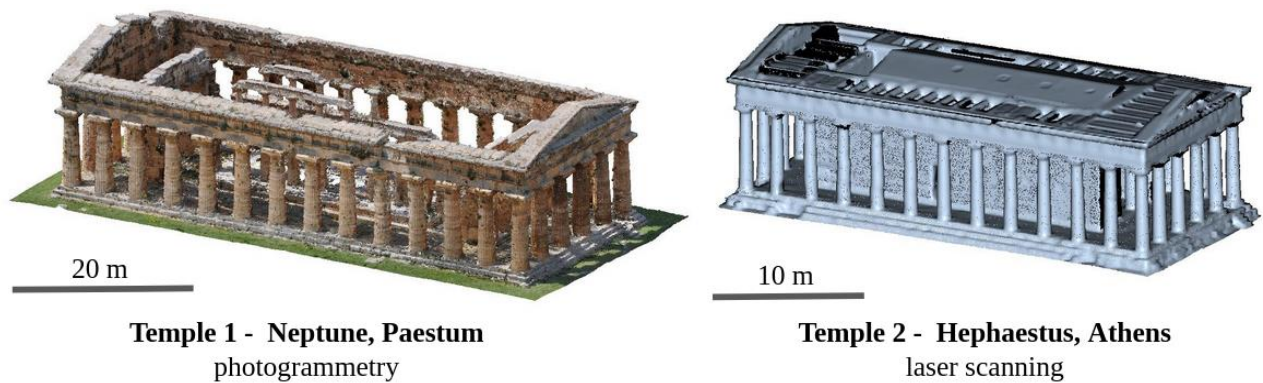


Figure 2. The two ancient Greek temples, both of the Doric order. The temple of Neptune in Paestum, Italy (Temple 1) is almost twice the size of the temple of Hephaestus in Athens, Greece (Temple 2). The two 3D point clouds were derived with different acquisition methods and present, to some extent, diverse characteristics in their element structure, fine details, material and state of erosion. As the internal wall and roof of Temple 2 are not present in Temple 1, they were excluded from the processing presented in the paper.

For the specific case study of Temple 2, the selected classes follow Grilli et al. (2019) and are based on the various architectural components of an ancient Greek temple: floor, crepidoma, column, echinus, abacus, architrave, frieze, cornice and tympanum. Only the classes that are common to both temples have been considered for consistency. Thus, the internal wall and the roof of Temple 2 were excluded from the segmentation procedure. Subsequently, Temple 2 point cloud is manually annotated, and these labels  $y$  are used as ground truth (GT) information for our evaluation. A fundamental step for the method is the appropriate definition of the neighbourhood for each point using a search radius (Hackel et al., 2016). Indeed, as proven in Grilli et al. (2019), the radius scale is proportional to the column's dimensions. Therefore, the covariance features, derived from the covariance matrix of the 3D point coordinates in a given neighbourhood are calculated and extracted in a radius equal to either the radius or the diameter of the column. These covariance features are expressed as combinations of the eigenvalues and eigenvectors of the covariance matrix for each point and include planarity, surface variation, sphericity, omnivariance, anisotropy, linearity as formulated in Hackel et al., (2016). Other extracted geometric features that are proven to be useful in such methods include normal based features, such as verticality, and height-based features, such as the Z coordinate. Feature extraction is performed on the entire point cloud and defines a feature matrix  $X$ . The geometric features extracted on Temple 2 are visualized in Figure 3.

The 3D semantic segmentation was based on supervised machine learning techniques, namely the Random Forest classifier (RF), as implemented in the *scikit-learn* library (Pedregosa et al. 2011). RF is based on a large number of decision trees and has been proven to work efficiently in similar tasks, including generalization (Grilli and Remondino, 2020). The dataset is subsequently divided into a training and a test set: the training set consisting of both the labels  $y_{train}$  and their respective features  $X_{train}$  are used to train the algorithm. The test set, including exclusively the pre-calculated features  $X_{test}$ , is used for label inference, based on the prediction model of the RF algorithm. In this way, each point of the test set is assigned to a class label, resulting in a semantically enriched point cloud  $\hat{y}$  (Figure 1). The evaluation of the method is performed by comparing the prediction model with the manually annotated (GT) model  $y_{test}$ . The comparison result is visualised with the confusion matrix, which analyses the amount of correct and false predictions. Furthermore, based on the confusion matrix, the overall accuracy, precision, recall and F1-score metrics are calculated (Hackel et al. 2016; Weinmann, 2015). The overall accuracy measures the overall ability of the model to correctly assign a

label to all points; the precision represents the performance of the model in relation to false positives, while the recall is in relation to the false negatives; the F1-score measures the performance of the model by taking into account both precision and recall values. In this work, emphasis is also given to learning transferability as an attempt to examine a more general and objective approach reducing the bias that exists when knowledge (training) and application (testing) data derive from the exact same point cloud. In this regard, a second experiment was designed, where Temple 1 serves as training set and Temple 2 as test set.

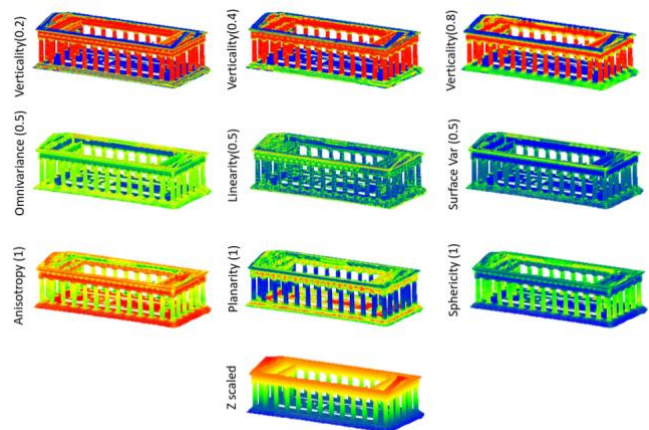


Figure 3: Feature extraction on the entire dataset ( $X_{train}$ ,  $X_{test}$ ) for the temple of Hephaestus (Temple 2). Please note that internal wall and roof were excluded from the processing as these two classes are not available in Temple 1.

### 3.3 Extraction of fitting parameters for geometric primitives

Primitive model fitting refers to the task of assigning a geometric primitive to describe a group of points. In our approach, the RANSAC algorithm is used due to its robustness to outliers and ease of implementation. RANSAC, being a stochastic method, is used to estimate the model parameters given a certain data set in an iterative way. The underlying principle is to fit various potential random hypotheses and iteratively find the model that best describes the given dataset, i.e. reach a level of agreement with most of the data samples, also known as consensus. In more detail, the algorithm first finds a minimum set of samples that are sufficient to describe the model. Afterward, it counts the percentage of the dataset adequately represented by this model and finally keeps the hypothesis that best fits the dataset. For every successful fitting, the model parameters are known as well as the number of inliers and outliers. The inliers correspond to



the subset of points that support the model, while the outliers to the subset of points that does not.

In our approach, after assigning a label to every point (Section 3.2), each semantic class defines a subset of 3D points with common geometric attributes. These subsets may contain one or more instances of the same class: e.g. the label "columns" contains several independent instances, whereas the class "floor" represents only one instance. Each class is analysed separately and it is considered as one or a collection of primitives with an a-priori defined geometry as a plane, cylinder, tori etc. The correspondence between primitive geometries and semantic classes is summarised in Table 1.










| class                                                                                            | primitive       |
|--------------------------------------------------------------------------------------------------|-----------------|
|  crepidoma (1)  | plane           |
|  floor (2)      | plane           |
|  column (3)     | cylinder / cone |
|  echinus (4)    | torus / sphere  |
|  abacus (5)     | plane           |
|  architrave (6) | plane           |
|  frieze (7)     | plane           |
|  cornice (8)    | plane           |
|  tympanum (9)   | plane           |

Table 1. The used class labels and the respective primitive shapes chosen for their representation.

For the geometric primitive fitting, we follow the Efficient RANSAC of Schnabel et al. (2007), as implemented in the open-source CGAL library (CGAL, 2022): starting from the segmented point cloud and some basic primitives (Table 1), fitting parameters can be automatically computed (Figure 4).

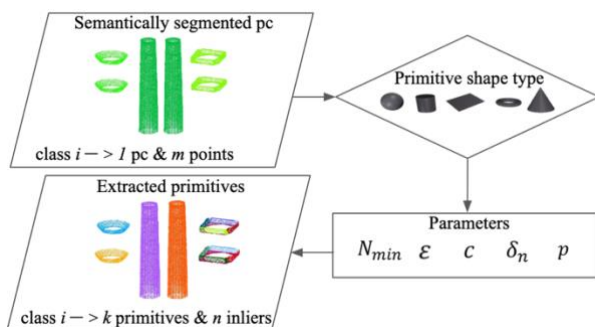


Figure 4. Process to derive the fitting parameters and create geometric primitives using CGAL library.

Efficient RANSAC depends on a set of parameters, namely the minimum number of points required for a primitive  $N_{min}$ , the maximum distance of the points to the primitive  $\epsilon$ , the sampling resolution meaning the distance between neighbouring points  $c$  the maximum normal deviation  $\delta_n$ , and the probability  $p$  that no better candidate was overlooked during sampling. Parameters  $\epsilon$  and  $c$  play a significant role to shape detection as slightly different settings may significantly change the results. In order to efficiently detect the primitives, these parameters should be tuned according to the specific scenario. Additional parameters that correspond exclusively to primitives like the sphere and the cylinder could include their minimum and maximum radius. In our approach, we use an adaptive strategy, based on the average spacing of our 3D data. In particular, according to a general rule of thumb,  $\epsilon$  is set approximately equal to the average spacing

and  $c$  defined as  $c = 5 * \epsilon$ . This adaptive parameter setting allows for a more automatic primitive shape extraction.

To enable the comparison between our method and a commonly used semi-automatic method, we also perform experiments using the qRansacSD plugin which is the implementation of the RANSAC algorithm in CloudCompare (CloudCompare, 2022).

#### 4. EXPERIMENTS AND RESULTS

Two different experiments were designed. In the first experiment both training and test data are derived from the exact same dataset, i.e. Temple 2 (*Experiment A*). The segmentation results (Figure 5) achieved an overall accuracy of ca. 98% (Table 2). The class "tympanum" is observed to have the lowest F1-score, in contrast to the class "column", which reaches the highest one. This can be explained due to the fact that the "tympanum" has the fewest number of 3D points affecting the neighbourhood computation and thus feature extraction. On the contrary, the "column" class includes the geometrically most distinctive instances. Comparing these results with to those of Grilli et al., (2019) for Temple 1, it can be observed that we achieve very similar values of overall accuracy, accuracy, recall and F1 score.

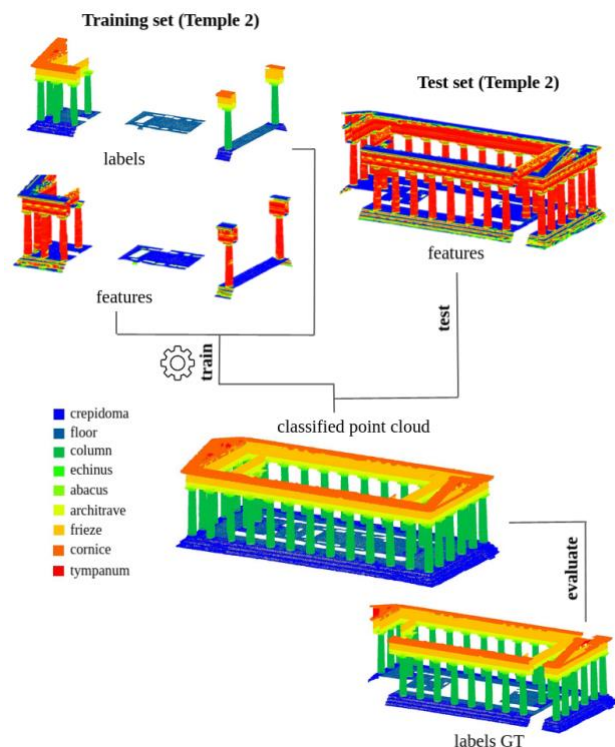


Figure 5: Experiment A. Both training and test sets are selected from Temple 2 (Temple of Hephaestus). GT labels are used for evaluation.

In order to reduce the manual labour for the preparation of labels, we investigated the transferability and generalisation of the method to similar architectures (*Experiment B*). The trained model of Temple 1 was thus used to infer the labels on Temple 2. The segmentation results (Figure 6) reached an overall accuracy of ca. 78% (Table 3). The predominant problems mainly concerned the classes "echinus", "abacus", "architrave" and "tympanum". Our intuition is that this derives from the fact that the two temples have similar, yet not identical characteristics. As expected, the best-predicted class is "column".

| class          | (1)           | (2)           | (3)            | (4)          | (5)           | (6)           | (7)            | (8)            | (9)          | prec. | recall      | F1   |
|----------------|---------------|---------------|----------------|--------------|---------------|---------------|----------------|----------------|--------------|-------|-------------|------|
| (1)            | <b>445010</b> | 15772         | 1446           | 0            | 0             | 0             | 0              | 0              | 0            | 0.95  | 0.96        | 0.96 |
| (2)            | 17902         | <b>146777</b> | 119            | 1            | 0             | 0             | 0              | 0              | 0            | 0.90  | 0.89        | 0.90 |
| (3)            | 4305          | 162           | <b>1181043</b> | 2699         | 0             | 0             | 0              | 0              | 0            | 1.00  | 0.99        | 0.99 |
| (4)            | 0             | 0             | 4359           | <b>37566</b> | 3008          | 3             | 0              | 0              | 0            | 0.89  | 0.84        | 0.86 |
| (5)            | 0             | 0             | 0              | 1710         | <b>110763</b> | 6063          | 0              | 0              | 0            | 0.92  | 0.93        | 0.93 |
| (6)            | 0             | 0             | 1              | 0            | 6269          | <b>461947</b> | 6986           | 0              | 0            | 0.92  | 0.97        | 0.95 |
| (7)            | 0             | 0             | 0              | 0            | 0             | 33614         | <b>2135445</b> | 8668           | 25           | 0.98  | 0.98        | 0.98 |
| (8)            | 0             | 0             | 0              | 0            | 0             | 0             | 27336          | <b>1788426</b> | 3528         | 0.99  | 0.98        | 0.98 |
| (9)            | 0             | 0             | 0              | 0            | 0             | 0             | 0              | 15369          | <b>19947</b> | 0.85  | 0.56        | 0.68 |
| Over. accuracy |               |               |                |              |               |               |                |                |              |       | <b>0.98</b> |      |

Table 2: Confusion matrix for experiment A: training and test sets are selected from Temple 2.

| class          | (1)           | (2)           | (3)            | (4)      | (5)          | (6)           | (7)            | (8)            | (9)         | prec. | recall      | F1   |
|----------------|---------------|---------------|----------------|----------|--------------|---------------|----------------|----------------|-------------|-------|-------------|------|
| (1)            | <b>522753</b> | 8734          | 45875          | 0        | 0            | 0             | 0              | 0              | 0           | 0.86  | 0.91        | 0.88 |
| (2)            | 86826         | <b>172204</b> | 2740           | 0        | 0            | 0             | 0              | 0              | 0           | 0.95  | 0.66        | 0.78 |
| (3)            | 201           | 0             | <b>1344526</b> | 344      | 50174        | 79155         | 0              | 0              | 0           | 0.97  | 0.91        | 0.94 |
| (4)            | 0             | 0             | 0              | <b>0</b> | 2776         | 52718         | 0              | 0              | 0           | 0.00  | 0.00        | 0.00 |
| (5)            | 0             | 0             | 0              | 0        | <b>20284</b> | 133309        | 0              | 0              | 0           | 0.24  | 0.13        | 0.17 |
| (6)            | 0             | 0             | 0              | 0        | 12267        | <b>222177</b> | 346753         | 0              | 0           | 0.46  | 0.38        | 0.42 |
| (7)            | 0             | 0             | 0              | 0        | 0            | 144           | <b>1812298</b> | 939172         | 147         | 0.84  | 0.66        | 0.74 |
| (8)            | 0             | 0             | 0              | 0        | 0            | 0             | 0              | <b>2365496</b> | 7166        | 0.71  | 1.00        | 0.83 |
| (9)            | 0             | 0             | 0              | 0        | 0            | 0             | 0              | 49709          | <b>8255</b> | 0.53  | 0.14        | 0.22 |
| Over. accuracy |               |               |                |          |              |               |                |                |             |       | <b>0.78</b> |      |

Table 3: Confusion matrix for experiment B: the trained model of Temple 1 is transferred to Temple 2 to inference labels.

It can be stated that both 3D segmentation results have a satisfactory degree of success. As expected, more accurate results are generated when both training and test set are derived from the same point cloud, rendering this bias in knowledge useful.

Regarding primitive extraction, the results for Temple 2 are shown in Table 4, Figure 7 and 8. Generally, the extracted shapes adequately represent the expected simplified geometry. In average for all classes, approximately 4% of the points do not fit the extracted primitives meaning that only a small number of points were excluded from the model as outliers. Such a low percentage is expected for noise-free point clouds acquired with laser scanning. It is observed that the most accurately extracted primitives relate to the classes that are easily differentiated and consist of many singular instances (e.g. "column", "echinus"). Indeed, primitive extraction in these cases is equivalent to instance segmentation. On the other hand, noisier classes or classes with more complex geometry with fine details and/or varying point density are more challenging (i.e. the "crepidoma", the "floor" and the "cornice") as shown in Figure 8. For these particular classes the default RANSAC parameters demonstrated good performance. For the classes "abacus", "architrave", "frieze" and "tympanum" the default values lead either to over-segmentation or inadequate detection; thus manual parameter tuning was required, following the empirical rules described in Section 3.2 and considering the particular characteristics of each class i.e. minimum points belonging to each shape, average spacing and normal deviation. The achieved results in CGAL are very similar to the primitives attainable in CloudCompare (Figure 8), demonstrating a general consistency in the number of detected shapes and outlier percentage (Table 4). To be mentioned that this was the best achieved results in CloudCompare, after manual parameter tuning for each class.

Once the fitting parameters are retrieved, they can be used to automatically model geometric primitives in HBIM environment (Figure 9): a cylinder is represented by its center, a direction and

a radius; a torus by the symmetry axis, a center and the major and minor radii; a plane is represented by its normal vector and the distance to the origin.

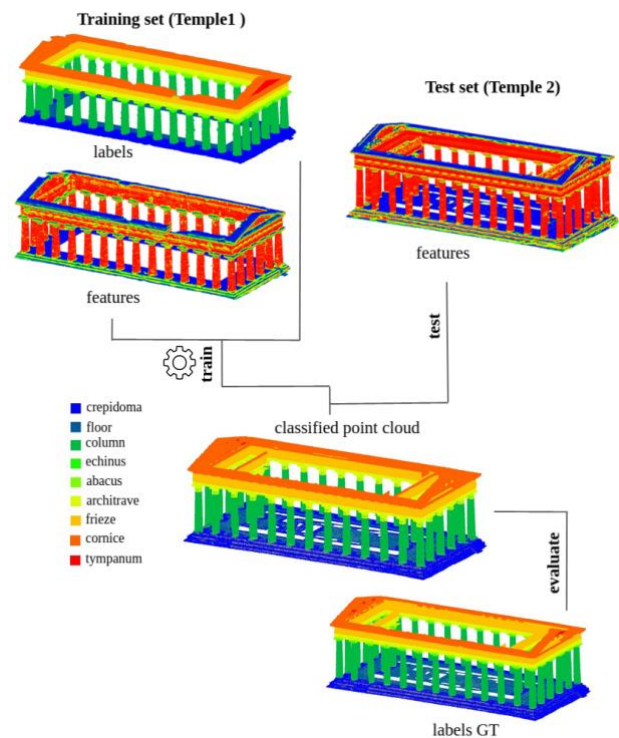


Figure 6: Experiment B. The trained model of Temple 1 (Temple of Neptune) is used to inference labels on Temple 2 (Temple of Hephaestus). GT labels are used for evaluation.

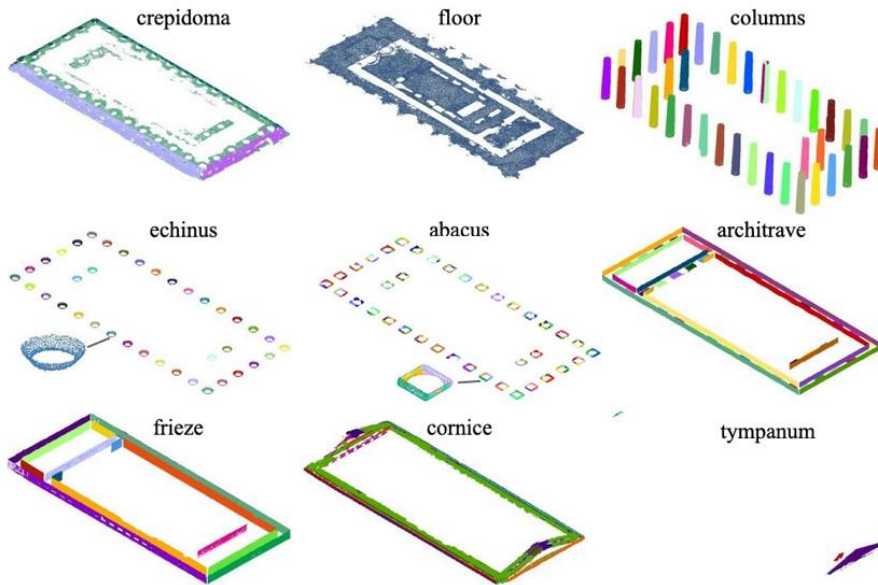


Figure 7. Isolated classes or instances corresponding to the identified primitives in CGAL.

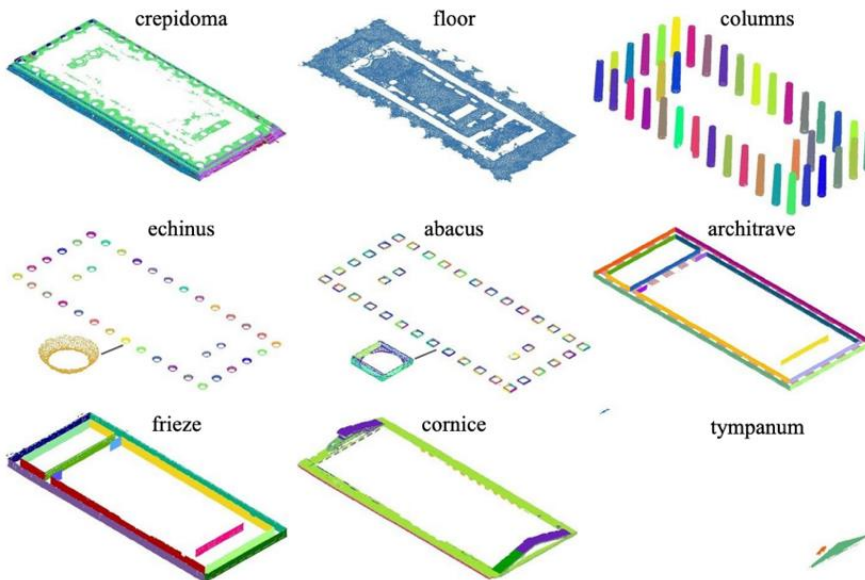


Figure 8. Isolated classes or instances corresponding to the identified primitives in CloudCompare.

| class | CloudCompare |         | CGAL      |         |
|-------|--------------|---------|-----------|---------|
|       | No. prim.    | inliers | No. prim. | inliers |
| (1)   | 16           | 94.77%  | 7         | 91.89%  |
| (2)   | 1            | 96.65%  | 1         | 99.43%  |
| (3)   | 38           | 98.42%  | 40        | 99.92%  |
| (4)   | 38           | 98.01%  | 38        | 98.91%  |
| (5)   | 191          | 96.89%  | 191       | 96.95%  |
| (6)   | 13           | 97.66%  | 56        | 93.44%  |
| (7)   | 14           | 95.54%  | 20        | 97.03%  |
| (8)   | 13           | 96.35%  | 14        | 92.58%  |
| (9)   | 4            | 96.75%  | 3         | 98.22%  |

Table 4: Number of detected primitives and percentage of inliers using CloudCompare and CGAL.

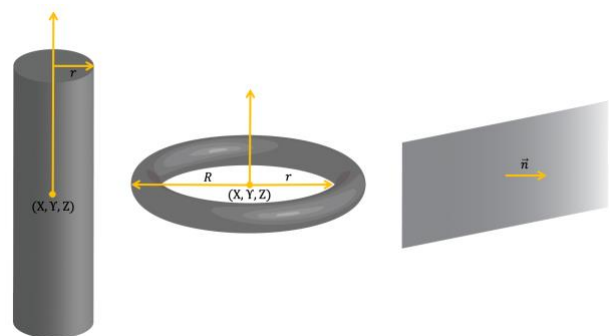


Figure 9: Considered geometric primitives (cylinder, torus and plane) and their modelling parameters.



## 5. DISCUSSION AND CONCLUSIONS

The suggested work for generating HBIM-friendly shapes relies on two main pillars: (i) supervised semantic segmentation of point clouds and (ii) automatic derivation of fitting parameters to create primitive shapes from identified classes or instances. Our framework provides an opportunity for understanding, interpretation and more complete visualisation of a point cloud, regardless of its acquisition method. Semantic segmentation refers to the assignment of a meaning to the 3D data discriminating the different elements. The detection and extraction of geometric primitives from the data discriminates the different entities and shapes of these elements. Per-point semantic labelling, in particular, can be very demanding and time-consuming due to the complexity of the architectural structures. However, for a successful segmentation, only a representative part of the dataset, approximately 25%, is required for the training, limiting the user's intervention. Respectively, only a small number of geometric features, namely 10, is required, limiting the computational effort. The label transferability between similar case studies has given promising results. Indeed, high percentage scores are reached for the segmentation of most of the structural elements on the test set for Experiment 2. However, the two case studies are not identical and thus have different geometric characteristics that negatively affect the label inference in the test set.

Regarding primitive extraction, the most critical step is the configuration of the RANSAC parameters. Its parameters cannot be easily predetermined and are to be adapted each time according to the given data in order to achieve satisfying results. Our implementation, based on the CGAL library, seems to give promising results allowing for automation in most of the classes. The successful derivation of fitting parameters to create shape primitives implies the isolation of each class or even instance belonging to a common mathematical model. As a result, the different elements of the analysed architecture, although composed of complex elements and details, are described by specific geometric parameters, enabling the modelling procedure in an HBIM environment.

## ACKNOWLEDGMENTS

Part of the presented research work was performed within the framework of the "Artificial Intelligence for Cultural Heritage" (AI4CH - <https://ai4ch.fbk.eu/>) project, a joint Italy–Israel lab which was funded by the Italian Ministry of Foreign Affairs and International Cooperation (MAECI).

## REFERENCES

- Armeni, I., Sax, S., Zamir, A.R. and Savarese, S., 2017. Joint 2d-3d-semantic data for indoor scene understanding. *arXiv preprint arXiv:1702.01105*.
- Barazzetti, L., 2016. Parametric as-built model generation of complex shapes from point clouds. *Advanced Engineering Informatics*, 30(3), pp.298-311.
- Blaha, M., Vogel, C., Richard, A., Wegner, J.D., Pock, T. and Schindler, K., 2016. Large-scale semantic 3d reconstruction: an adaptive multi-resolution model for multi-class volumetric labeling. *Proc. IEEE CVPR*, pp. 3176-3184.
- Boyko, A. and Funkhouser, T., 2011. Extracting roads from dense point clouds in large scale urban environment. *ISPRS J. of Photogrammetry and Remote Sensing*, 66(6), pp. S2-S12.
- Bosché, F., Ahmed, M., Turkan, Y., Haas, C.T. and Haas, R., 2015. The value of integrating Scan-to-BIM and Scan-vs-BIM

techniques for construction monitoring using laser scanning and BIM: The case of cylindrical MEP components. *Automation in Construction*, 49, pp.201-213.

Capone, M., Lanzana, E., 2019. Scan-to-BIM vs 3D ideal model HBIM: parametric tools to study domes geometry. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42-2/W9.

Carboni, N., Bruseker, G., Guillem, A., Bellido Castañeda, D., Coughenour, C., Domajnko, M., de Kramer, M., Ramos Calles, M. M., Stathopoulou, E. K., Suma, R., 2016. Data provenance in photogrammetry through documentation protocols. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. III-5, 57–64

Chen, Y., Wang, Y., Lu, P., Chen, Y. and Wang, G., 2018. Large-scale structure from motion with semantic constraints of aerial images. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, pp. 347-359. Springer, Cham.

Chen, L.C., Papandreou, G., Schroff, F. and Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

CGAL, Computational Geometry Algorithms Library, 2022. <https://www.cgal.org>

CloudCompare (version 2.11 Alpha) [GPL software], 2022. <http://www.cloudcompare.org/>

Croce, V., Caroti, G., De Luca, L., Jacquot, K., Piemonte, A. and Véron, P., 2021. From the Semantic Point Cloud to Heritage-Building Information Modeling: A Semiautomatic Approach Exploiting Machine Learning. *Remote Sensing*, 13(3), p.461.

Dalponte, M., Bruzzone, L. and Gianelle, D., 2012. Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. *Remote Sensing of environment*, 123, pp.258-270.

Grilli, E., Farella, E.M., Torresani, A. and Remondino, F., 2019. Geometric feature analysis for the classification of cultural heritage point clouds. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 42-2/W15, pp. 541–548.

Grilli, E., Remondino F., 2020. Machine Learning Generalisation across Different 3D Architectural Heritage. *ISPRS International Journal of Geo-Information*, Vol. 9(6), pp. 379.

Fiorucci, M., Khoroshiltseva, M., Pontil, M., Traviglia, A., Del Bue, A. and James, S., 2020. Machine learning for cultural heritage: A survey. *Pattern Recognition Letters*, 133, pp.102-108.

Fischler, M.A. and Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), pp.381-395.

Hackel, T., Wegner, J. D., Schindler, K., 2016. Fast segmentation of 3D point clouds with strongly varying density. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. III-3, pp. 177-184.

Hong, S., Jung, J., Kim, S., Cho, H., Lee, J. and Heo, J., 2015. Semi-automated approach to indoor mapping for 3D as-built building information modeling. *Computers, Environment and Urban Systems*, 51, pp.34-46.

Hough, P.V., 1962. *Method and means for recognizing complex patterns*. U.S. Patent 3,069,654.

- Huang, R., Xu, Y., Hong, D., Yao, W., Ghamisi, P. and Stilla, U., 2020. Deep point embedding for urban classification using ALS point clouds: A new perspective from local to global. *ISPRS Journal of Photogrammetry and Remote Sensing*, 163, pp.62-81.
- Huber, D., Akinci, B., Oliver, A.A., Anil, E., Okorn, B.E. and Xiong, X., 2011. Methods for automatically modeling and representing as-built building information models. In *Proceedings of the NSF CMMI Research Innovation Conference* (Vol. 856558). NSF.
- Jung, J., Hong, S., Jeong, S., Kim, S., Cho, H., Hong, S. and Heo, J., 2014. Productive modeling for development of as-built BIM of existing indoor structures. *Automation in Construction*, 42, pp. 68-77.
- Kirillov, A., He, K., Girshick, R., Rother, C. and Dollár, P., 2019. Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9404-9413.
- Long, J., Shelhamer, E. and Darrell, T., 2015. Fully convolutional networks for semantic segmentation. *Proc. IEEE CVPR*, pp. 3431-3440.
- Macher, H., Landes, T., Grussenmeyer, P., 2017. From point clouds to building information models: 3D semi-automatic reconstruction of indoors of existing buildings. *Appl. Sci.*, 7, 1030.
- Marmanis, D., Wegner, J.D., Galliani, S., Schindler, K., Datcu, M. and Stilla, U., 2016. Semantic segmentation of aerial images with an ensemble of CNSS. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2016, 3, pp.473-480.
- Matrone, F., Grilli, E., Martini, M., Paolanti, M., Pierdicca, R. and Remondino, F., 2020. Comparing machine and deep learning methods for large 3D heritage semantic segmentation. *ISPRS International Journal of Geo-Information*, 9(9), p.535.
- Murphy, M., McGovern, E., and Pavia, S., 2013. Historic Building Information Modelling -Adding intelligence to laser and image based surveys Elsevier. *ISPRS J. of Photogrammetry and Remote Sensing*, 76, pp. 89-102.
- Murtiyoso, A. and Grussenmeyer, P., 2019. Point cloud segmentation and semantic annotation aided by GIS data for heritage complexes. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 42, pp. 523-528.
- Özdemir, E., Remondino, F. and Golkar, A., 2021. An Efficient and General Framework for Aerial Point Cloud Classification in Urban Scenarios. *Remote Sensing*, 13(10), p.1985.
- Plath, N., Toussaint, M. and Nakajima, S., 2009. Multi-class image segmentation using conditional random fields and global classification. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 817-824.
- Qi, C.R., Su, H., Mo, K. and Guibas, L.J., 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proc. IEEE CVPR*, pp. 652-660.
- Qi, C.R., Yi, L., Su, H. and Guibas, L.J., 2017. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *arXiv preprint arXiv:1706.02413*.
- Pierdicca, R., Paolanti, M., Matrone, F., Martini, M., Morbidoni, C., Malinverni, E.S., Frontoni, E. and Lingua, A.M., 2020. Point cloud semantic segmentation using a deep learning framework for cultural heritage. *Remote Sensing*, 12(6), p.1005.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, pp.2825-2830.
- Rottensteiner, F. and Briese, C., 2002. A new method for building extraction in urban areas from high-resolution LIDAR data. In *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, Vol. 34, No. 3/A, pp. 295-301.
- Schnabel, R., Wahl, R., Klein, R., 2007. Efficient RANSAC for point cloud shape detection. *Computer Graphics Forum*, Vol. 26(2), pp. 214-226.
- Stathopoulou, E.K., Battisti, R., Cernea, D., Remondino, F. and Georgopoulos, A., 2021. Semantically Derived Geometric Constraints for MVS Reconstruction of Textureless Areas. *Remote Sensing*, 13(6), p.1053.
- Tarsha-Kurdi, F., Landes, T., Grussenmeyer, P., 2007. Hough Transform and extended RANSAC algorithms for automatic detection of 3D building roof planes from lidar data, *The International Archives of Photogrammetry, Remote Sensing*, Vol. 36, pp. 407-412.
- Thomson, C. and Boehm, J., 2015. Automatic geometry generation from point clouds for BIM. *Remote Sensing*, 7(9), pp.11753-11775.
- Verma, V., Kumar, R. and Hsu, S., 2006. 3D building detection and modeling from aerial lidar data. *Proc. IEEE CVPR*, Vol. 2, pp. 2213-2220. IEEE.
- Vosselman, M.G., Gorte, B.G.H., Sithole, G., Rabbani, T., 2004. Recognising structure in laser scanning point clouds. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 36.
- Weinmann, M., Jutzi, B., Mallet, C. 2013. Feature relevance assessment for the semantic interpretation of 3D point cloud data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II-5/W2, pp. 313-318.
- Weinmann, M., Urban, S., Hinz, S., Jutzi, B., Mallet, C., 2015. Distinctive 2D and 3D features for automated large-scale scene analysis in urban areas. *Comput. Graph.*, 49, 47-57.
- Weinmann, M., Weinmann, M., Mallet, C., Brédif, M., 2017. A Classification-Segmentation Framework for the Detection of Individual Trees in Dense MMS Point Cloud Data Acquired in Urban Areas. *Remote Sensing*, Vol. 9(3), pp. 277.
- Xu, B., Jiang, W., Shan, J., Zhang, J. and Li, L., 2016. Investigation on the weighted ransac approaches for building roof plane segmentation from lidar point clouds. *Remote Sensing*, 8(1), p.5.
- Yang, L., Cheng, J.C. and Wang, Q., 2020. Semi-automated generation of parametric BIM for steel structures based on terrestrial laser scanning data. *Automation in Construction*, 112, p.103037.
- Yang, X., Grussenmeyer, P., Koehl, M., Macher, H., Murtiyoso, A. and Landes, T., 2020. Review of built heritage modelling: Integration of HBIM and other information techniques. *Journal of Cultural Heritage*.
- Yousefhusien, M., Kelbe, D.J., Ientilucci, E.J. and Salvaggio, C., 2018. A multi-scale fully convolutional network for semantic labeling of 3D point clouds. *ISPRS J/ of Photogrammetry and Remote Sensing*, 143, pp.191-204.