POLITECNICO DI TORINO
Repository ISTITUZIONALE

A BENCHMARK FOR LARGE-SCALE HERITAGE POINT CLOUD SEMANTIC SEGMENTATION

Original
A BENCHMARK FOR LARGE-SCALE HERITAGE POINT CLOUD SEMANTIC SEGMENTATION / Matrone, F., Lingua, A., Pierdicca, R., Malinverni, E.S., Paolanti, M., Grilli, E., Remondino, F., Murtiyoso, A., Landes, T.. - In: INTERNATIONAL ARCHIVES OF THE PHOTOGRAMMETRY, REMOTE SENSING AND SPATIAL INFORMATION SCIENCES. - ISSN 2194-9034. - ELETTRONICO. - XLIII-B2-2020:(2020), pp. 1419-1426. [10.5194/isprs-archives-xliii-b2-2020-1419-2020]

Availability:
This version is available at: 11583/2844376 since: 2020-09-08T09:39:38Z

Publisher:
Copernicus Publications

Published
DOI:10.5194/isprs-archives-xliii-b2-2020-1419-2020

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
default_article_editorial [DA NON USARE]

-

(Article begins on next page)

25 November 2026

A BENCHMARK FOR LARGE-SCALE HERITAGE POINT CLOUD SEMANTIC SEGMENTATION

F. Matrone [1], A. Lingua [1,2], R. Pierdicca [3], E.S. Malinverni [3], M. Paolanti [4], E. Grilli [5], F. Remondino [5], A. Murtiyoso [6], T. Landes [6]

[1] DIATI, Politecnico di Torino, Torino, Italy - Email: <francesca.matrone><andrea.lingua>@polito.it
[2] PIC4SeR, Politecnico Interdepartmental Center for Service Robotics, Politecnico di Torino, Torino, Italy
[3] DICEA, Università Politecnica delle Marche, Ancona, Italy - Email: <r.pierdicca><e.s.malinverni>@staff.univpm.it
[4] DII, Università Politecnica delle Marche, Ancona, Italy - Email: m.paolanti@staff.univpm.it
[5] 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy – Email: <grilli><remondino>@fbk.eu
[6] Photogrammetry and Geomatics Group, ICube Laboratory UMR 7357, INSA Strasbourg, France – <arnadi.murtiyoso><tania.landes>@insa-strasbourg.fr

Commission II, WGII/8

KEYWORDS: benchmark, 3D heritage, point cloud, semantic segmentation, classification, machine learning, deep learning

ABSTRACT:

The lack of benchmarking data for the semantic segmentation of digital heritage scenarios is hampering the development of automatic classification solutions in this field. Heritage 3D data feature complex structures and uncommon classes that prevent the simple deployment of available methods developed in other fields and for other types of data. The semantic classification of heritage 3D data would support the community in better understanding and analysing digital twins, facilitate restoration and conservation work, etc. In this paper, we present the first benchmark with millions of manually labelled 3D points belonging to heritage scenarios, realised to facilitate the development, training, testing and evaluation of machine and deep learning methods and algorithms in the heritage field. The proposed benchmark, available at http://archdataset.polito.it/, comprises datasets and classification results for better comparisons and insights into the strengths and weaknesses of different machine and deep learning approaches for heritage point cloud semantic segmentation, in addition to promoting a form of crowdsourcing to enrich the already annotated database.

1. INTRODUCTION

The growing ease of point cloud acquisition, especially due to the developments of automated image-based solutions, SLAM methods and laser scanning systems, has created an increasing interest of the scientific community towards the use, interpretation and direct exploitation of point clouds for many different purposes.

Consequently, in the Cultural Heritage (CH) field, HBIM (Historical Building Information Modeling) has gained particular attention from experts, since it allows to manage architectural heritage data, in both geometrical and informative ways (Bruno and Roncella, 2018). As it is well-known, whether point clouds provide a needful starting point, the process of developing HBIM models is still entrusted on manual operation; experts are claimed at handling large and complex datasets, without the aid of any automatic or semi-automatic method to recognise and reshape 3D elements (Bitelli et al., 2017). Obviously, this process is very time consuming and brings to the waste of information, given the unavoidable simplification exerted.

In this scenario, the comeback of Deep Learning (DL) in several research fields has been overwhelming (Griffiths and Boehm, 2019). Deep Neural Networks (DNNs) settled as the more efficient technology for learning-based tasks (Paolanti et al., 2019; Bello et al., 2020). However, despite DNNs proved to be very promising for handling and recognising 3D data (Wang et al., 2019), for CH, manual operations look more trustworthy, at least to capture the real estate from point clouds (Murtiyoso and Grussenmeyer, 2019). There are many reasons for such scepticism; first of all, CH goods have complex geometries, which can be described only with a high level of detail. Moreover, the irregular shapes joined with the uniqueness of objects, make supervised learning techniques arduous for 3D data.

Besides the intrinsic complexity of 3D data, especially if compared with 2D ones (e.g. images or trajectories), there are other limitations that are hampering the exploitation of DNNs for CH; on one hand, the lack of training data, on the other, the computational effort. While this latter is going to be overcome by continuous technological advancements, enabling a system to learn from a labelled dataset, and generalise on unseen scenes, is still far. The manual annotation is expensive and time-consuming (even if more reliable), and exists a sort of reticence to share 3D data with the research community.

With the main purpose of investing much more effort on these research lines, the authors provide a large dataset of CH architectures, which aspires to become the reference benchmark in the field. To the best of our knowledge, it is the first point cloud dataset specifically released for the CH domain, which comprises data collected with both TLS and photogrammetric surveys, even providing the semantic ground truth annotation.

This paper aims to present a new 3D point cloud classification benchmark dataset (named *ArCH dataset*[1] - *Architectural Cultural Heritage*) with millions of manually labelled points belonging to heritage scenarios. The realised benchmark originates from the collaboration of different universities and research institutes (Politecnico di Torino, Università Politecnica delle Marche, FBK Trento, Italy, and INSA Strasbourg, France). It is unique as it offers, for the first time to the research community, annotated point clouds describing heritage scenes. These point clouds, labelled with 10 classes, are meant to

[1] http://archdataset.polito.it/

facilitate the development, training, testing and evaluation of machine learning algorithms as well as its subset of deep learning methods in the heritage field. For a more profitable use of this benchmark, aside from free download of all data, we provide public results of the submitted approaches, providing rankings about the most performing ones.

2. PREVIOUS WORKS

Several benchmarks have been proposed in the Geomatics community; their value is priceless. In fact, labelled 3D data enable users to test and validate their algorithms, beside improving the training phase for both machine and deep learning approaches. Among the existing benchmarks, it is worth to cite ModelNet 40 (Wu et al., 2015) with more than 100k CAD models of objects, mainly furniture, from 40 different categories; KITTI (Geiger et al., 2013) that includes camera images and laser scans for autonomous navigation; Sydney Urban Objects (De Deuge et al., 2013) dataset acquired in urban environments with 26 classes and 631 individual scans; Semantic3D (Hackel et al., 2017) with urban scenes such as churches, streets, railroad tracks and squares; S3DIS (Armeni et al, 2016) that includes mainly office areas and the Oakland 3-D Point Cloud dataset (Munoz et al., 2009) consisting of labelled laser scanner 3D point clouds, collected from a moving platform in an urban environment. Besides, it is worth mentioning other specific datasets, such as iQmulus (Vallet et al., 2015), The Cityscapes Dataset (Cordts et al., 2016), Paris-rue-Madame (Serna et al., 2014), Paris-Lille-3D (Roynard et al., 2018), 3DOMcity (Özdemir et al., 2019) and MiMAP (Wang et al., 2018) for BIM feature extraction.
Most of these datasets collect data from urban environments with point clouds composed of around 100k points.
In these scenarios, the object classes and labels are fairly general and almost standard (e.g. ground, roads, vehicles, vegetation, buildings etc.). On the other hand, in the heritage field, the identification of precise categories is much more complicated. Several peculiar classes could be identified in the same dataset. Shape and colour are not always linked to a specific semantic class, and objects belonging to the same class could have completely different shapes, in addition to complex geometries. Moreover, to date, there are still no published datasets focusing on immovable cultural assets with an adequate level of detail.
Up to now most of the available datasets of annotated architectural heritage consists of 2D images, such as the Ecole Centrale Paris (ECP) Facades dataset (Teboul et al., 2010), eTRIMS (Korc and Forstner, 2009), and CMP Facade Database (Tyleček and Šára, 2013), which all present datasets of manually annotated facade images from different cities around the world and diverse architectural styles. Still, in 2D, there is the work conducted by Llamas et al. (2017), where for the first time Convolutional Neural Networks (CNN) were applied to heritage scenarios. The authors also released a dataset with more than 10k images including categories like Altar, Apse, Belltower, Column, Dome (inner and outer), Flying buttress, Gargoyle, Stained glass, and Vault.
In this context, several researchers have started to approach the topic of semantic segmentation of cultural heritage (CH) point clouds within the machine and deep learning framework (Grilli et al., 2019a; Kharroubi et al., 2019; Murtiyoso and Grussenmeyer, 2020; Pierdicca et al., 2020). However, the lack of an appropriate 3D heritage dataset does not allow an effective comparison between methods and results.
Precisely for this reason, we propose ArCH dataset that can stimulate the scientific community on these challenging issues.

3. DATASET

The dataset is composed of 17 annotated and another 10 non-annotated point clouds, the latter of which could be labelled by users and added to the main dataset.
Many of the scenes included in the ArCH benchmark are part (or a candidate) of the UNESCO World Heritage List (WHL):

- the chapel of the Strasbourg Cathedral inside the Grande Île, inscribed in 1988;
- the courtroom of the Valentino's Castle (VAL) included in the "Residences of the Royal House of Savoy" from 1997;
- the Sacro Monte of Varallo (SMV) and Ghiffa (SMG) part of the wider site of "Sacri Monti of Piedmont and Lombardy" from 2003;
- St. Pierre church located inside the Neustadt inscribed in 2017;
- the porticoes of Bologna presented as a candidate in 2020.

Other scenes are nevertheless part of historical built heritage and represent various historical periods and architectural styles. This difference could constitute a drawback in the definition of the dataset classes, as it introduces elements of inhomogeneity within the same classes. However, providing the neural network with differing elements improves its ability to generalise among various CH case studies.
Among the labelled scenes of the benchmark, 15 scenes are available for training and 2 for testing. They all include churches, chapels, porticoes, loggias, pavilions and cloisters. The 2 test scenes (named A and B) have different characteristics:
- the first (*A_SMG_portico*) represents a simple, almost symmetrical building on one level and with more standard and repetitive geometric elements (Figure 1);

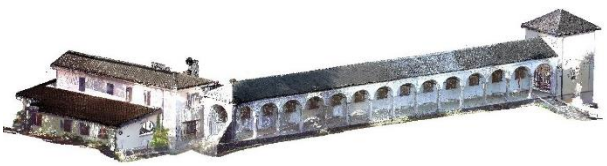

Figure 1. Point cloud of the portico of the Sacro Monte di Ghiffa (SMG).

- the second (*B_SMV_chapel_27to35*) represents a complex, non-symmetrical building, structured on two levels, surveyed both indoor and outdoor, with different types of vaults, stairways and windows (Figure 2).



Figure 2. Point cloud of the second test scene, the chapels from 27 to 35 of the Sacro Monte of Varallo (north and south views).

These two test scenes were chosen to (i) simplify the comparisons of the results, (ii) assess the effectiveness of the proposed algorithms and (iii) try to highlight the generalisation and learning capability of the networks not only on a relatively simple scene but also on a complex one.

TRAINING

Name	Number of points	Scene	Data acquisition	Number of classes (excluded Other)	Subsampling (cm)
1_TR_cloister	15,740,229	Indoor/outdoor	TLS + UAV	8/9	1
2_TR_church	20,862,139	Indoor	TLS	8/9	1
3_VAL_room	4,188,066	Indoor	TLS	6/9	1
4_CA_church	4,850,807	Outdoor	TLS + UAV	6/9	1
5_SMV_chapel_1	3,783,412	Outdoor	TLS + UAV	9/9	1
6_SMV_chapel_2to4	6,326,871	Indoor/outdoor	TLS + UAV	9/9	1
7_SMV_chapel_24	3,571,064	Outdoor	TLS + UAV	9/9	1
8_SMV_chapel_28	3,156,753	Outdoor	TLS + UAV	9/9	1
9_SMV_chapel_10	2,193,189	Indoor/outdoor	TLS + UAV	6/9	1
10_SStefano_portico_1	3,783,699	Outdoor	Terrestrial photogrammetry	8/9	1
11_SStefano_portico_2	10,047,392	Outdoor	Terrestrial photogrammetry	8/9	1
12_KAS_pavillion_1	598,384	Indoor/outdoor	TLS	4/9	1
13_KAS_pavillion_2	325,822	Indoor/outdoor	TLS	4/9	1
14_TRE_square	10,045,227	Outdoor	Terrestrial photogrammetry	8/9	1.5
15_OTT_church	13,264,040	Indoor/outdoor	TLS	9/9	1.5

TEST

Name	Number of points	Scene	Data acquisition	Number of classes (excluded Other)	Subsampling
A_SMG_portico	16,165,924	Outdoor	TLS + UAV	9/9	1
B_SMV_chapel_27to35	16,200,442	Indoor/outdoor	TLS + UAV	9/9	1

Table 1. Main features of the ArCH dataset. The number of classes in each scene could be helpful for choosing the Validation scene while training phases.

3.1 Data acquisition

The 3D data composing the benchmark (Table 1) are challenging, not only due to their size (up to $\approx 4 \cdot 108$ points per scan) but also because of their high measurement resolution and high density of the final point cloud. Most of the scenes are obtained through the integration of different point clouds, acquired with different sensors (cameras, scanners) and platforms (UAVs, etc.) and after an appropriate accuracy evaluation.
The employed terrestrial laser scanners include a FARO Focus 3D X 130 and 120 and a Riegl VZ-400. The photogrammetric surveys of the Sacro Monte of Varallo were performed with a Nikon D880E whereas for Bologna and Trento a Nikon D3100 and D3X were employed, respectively. A UAV platform was equipped with a SONY Ilce 5100L whereas the DJI Phantom 4 Pro has its integrated camera.

3.2 Data pre-processing

The collected point clouds were initially pre-processed to make the cloud structures more homogeneous (Table 1). The cloud pre-processing was performed in CloudCompare and followed 3 steps:

- spatial translation;
- subsampling;
- choice of features.

The *spatial translation* of the point clouds was necessary because of the georeferencing of the scenes. The coordinate values had too many digits to be processed by the neural networks, so the coordinates were truncated and every single scene was spatially moved close to the system origin (0,0,0).
The *subsampling* operation became necessary due to the high number of points (mostly redundant) in each scene (> 20M points). The option of random subsampling was discarded because it would have limited the test repeatability, therefore other two methods were tested: octree- and space-based subsampling.
From the comparison of the results coming from the application of the octree- and space-based subsampling, we opted for the second option. The variation in the test results was 1%, therefore the uniformity and simplicity of setting were preferred.
As far as the space-based method is concerned, a minimum space of 0.01 m between points was set; in this way, a high level of detail is ensured, but at the same time it is possible to considerably reduce the number of points and the size of the file, in addition to regularise the geometric structure of the point cloud.
In the DL framework, the *feature selection* is subject to two different approaches. The first one consists in selecting as few features as possible and letting the neural network just learn from them. The second, mainly used for smaller datasets, foresees the selection of specific handcrafted features, thus facilitating the learning task and improving the overall performances, though increasing computational times. In this case, most of the features are usually handcrafted for specific tasks (Zhang et al. 2019) and can be subdivided and classified into intrinsic and extrinsic, or also used for local and global descriptors (Han et al., 2018; Weinmann et al., 2015). The local features define the statistical properties of the local neighbourhood geometric information, while the global features describe the whole geometry of the point cloud. The most used properties are the local ones, such as eigenvalues based descriptors, 3D shape context, etc. Nevertheless, we provide only common intrinsic features, in order to allow users to find the most appropriate combinations. The only features calculated are the normals.
The point normals are computed on CloudCompare, most of the time with a plane local surface model and oriented with a minimum spanning tree with Knn=10. The orientation of the normals was then checked in MATLAB®.
Hence, the point cloud structure is *x, y, z, r, g, b, label, Nx, Ny, Nz*.

4. CLASS DEFINITION

Through the automatic recognition of architectural elements, the authors would like to support and speed up the process of reconstructing 3D geometries for HBIM models. In this context, it is essential to choose classes for our benchmark that are already available in object-oriented software or the underlying standards. In this way, the output labels of the neural network correspond exactly to the BIM categories and, once the geometry has been reconstructed, it will be possible to associate its information directly to the specific classes.

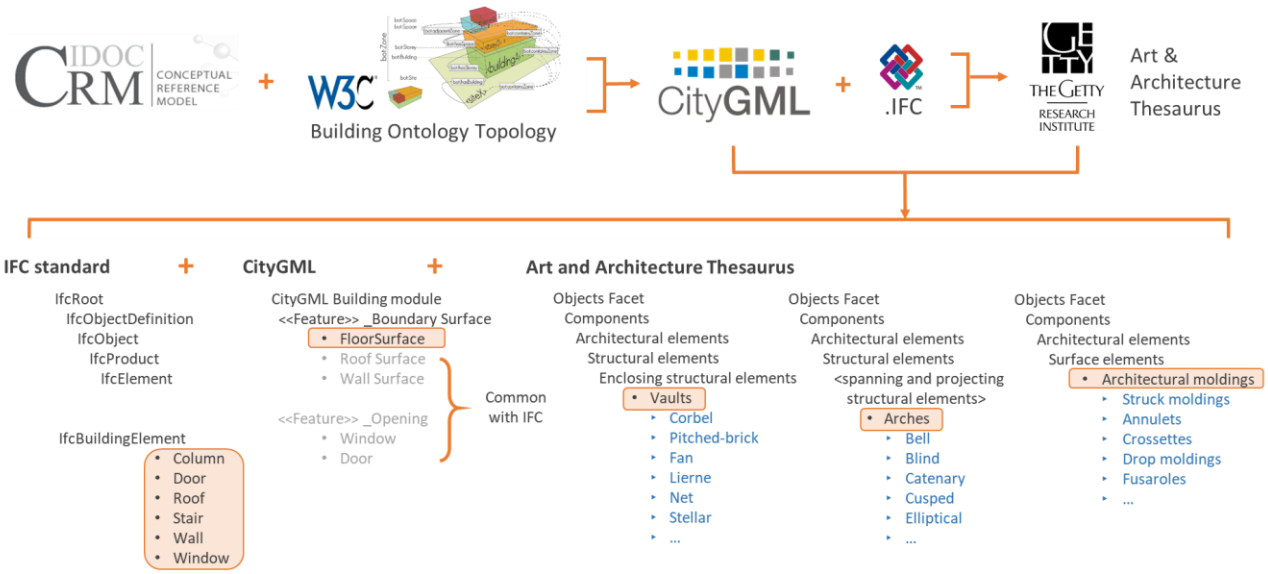

Figure 3. Updated class selection for point cloud annotation.

In the current state of the art, some works have already associated semantics, based on taxonomies and ontologies, to heritage elements (Mallik and Chaudhury, 2012) or HBIM models (Quattrini et al., 2017; Yang et al., 2019). However, there are still no studies that combine the semantic of the BIM domain with the automatic recognition of architectural elements through the DL techniques.
In addition to the standard which the BIM domain relies on (the IFC), other standards have been investigated, to have an overall and multi-level view of the cultural assets. In particular, the standards investigated are those aimed at describing:

- buildings and their surrounding context, such as BOT (Building Topology Ontology) and CityGML (Geography Markup Language);
- cultural assets themselves, such as the CIDOC-CRM (International Commitee for Documentation-Conceptual Reference Model) and the AAT (Art and Architecture Thesaurus) of the Getty Institute.

By semantically organising the data, they can be managed with a common vocabulary and the subdivision into classes is therefore not arbitrary but objective and standardised, equal for all the users and referring to an already codified lexicon. Thus, a unified method for the classification of the architectural elements was developed (Malinverni et al., 2019).
The concept of Level of Detail (LOD) derives from the CityGML data model and allows to describe an object according to different scales of representation, in which both the geometries represented and the information inserted range from the general to the particular. We have therefore applied this concept to the semantic segmentation of our point clouds: at first, we tried to understand at which level of detail the point clouds are segmented and, subsequently, the corresponding classes have been identified in the aforementioned standards.
In CityGML, the LOD 0 describes a regional and landscape scale, the LOD 1 the region or city, the LOD 2 the city districts, the LOD 3 and 4 the architectural models respectively with the outdoor and indoor elements.
If we consider some literature examples about point cloud classification in the geospatial field using NNs (Landrieu and Simonovsky, 2018; Hackel et al., 2017), we can assert that the level of detail reached till now is between LOD 1 and 2. Among the almost identified standard classes (i.e. vegetation, roads, buildings, etc.), the individual architectural elements are still missing.

Semantic annotation of the point clouds according to a CityGML LOD 3/4 has been therefore defined.
In particular, in the CityGML, the LOD 3 foresees the realisation of a detailed architectural model and its scheme has the insertion of objects as doors and windows.
The classes identified are within "Feature"_*Boundary Surface* 'Floor', 'Roof' and 'Wall' and within "*Feature"_Openings* 'Window' and 'Door'.
Regarding the IFC standard, the category that contains the architectural elements is *IfcBuildingElement*, a subclass of *IfcElement*. In this category, several architectural elements can describe a building, but just some of these are common in the DCH domain and some other are too specific for the new construction or for a specific construction technique. The classes identified are, therefore: 'Column', 'Door', 'Roof', 'Stair', 'Wall' and 'Window', two of which already in common with the CityGML data model.
Moreover, as the classes included in these two standards are not enough to describe properly a CH, the AAT was perused and, within the *Architectural elements* class and *Structural elements* category, the 'Vaults', 'Arches' classes have been taken into account, whereas from *Surface elements* 'Moldings' have been selected.
Following some studies and results of classification with the 3D features (Grilli et al., 2019b), it was decided to change the classification proposed in (Malinverni et al., 2019; Pierdicca et al., 2020), separating the class of columns and half-pilasters and inserting the latter in the new class 'Moldings' where there are also cornices and eaves.
With this purpose, 9 classes have been selected (Figure 3), plus another one defined as 'Other', containing all the points not belonging to the previous classes (e.g. paintings, altars, benches, statues, waterspouts...).
These classes have been used for the point clouds labelling (Figure 4). Nevertheless, the possibility of further extending this scheme for a higher Level of Detail (LOD 4/5), to be exploited for Instance Segmentation, is planned. Interested readers can deep this topic in (Mo et al., 2019).



Figure 4. Examples of labelled point clouds of the ArCH dataset.

4.1 Guidelines for annotation

Once the classes were chosen, given the heterogeneity of the architectural elements, some guidelines were defined for the annotation of the point clouds. These guidelines for the dataset annotation allow other researchers to contribute to expanding the datasets (Figure 5). The dataset has been labelled with common point cloud processing software as CloudCompare. However, on the benchmark page, an in-house web annotation tool built upon the Semantic-segmentation-editor web application is also available for users.
Considering each class, excluding the standard ones of walls, floors, roofs and stairs, the guidelines followed for the annotation have been:

- **Columns**. In this class, only stand-alone columns or pillars have been inserted, both with circular and square sections. As mentioned above, the half-pilasters or half-columns leaning on the walls have been included in the Moldings class.
- **Moldings**. Stuccos and all other types of moldings, like windows, doors or decorative moldings have been included in this class, in addition to the previously cited half-pilaster and half-columns (Figure 5). More generally, everything that protrudes from the masonry falls into this class.



Figure 5. Example of moldings (purple) and columns (red).

- **Doors and windows** have been combined into one class, given their reduced number of points and their similar geometry.
- **Vaults**. Every type of vault (barrel, cross, dome ...) has been included in this class. If the individual vaults were divided by protruding arches with respect to the vault itself then they were interrupted, otherwise a unique annotation has been kept.
- **Arches**. This class includes both the arches on the facade and those that divide one vault from another, but only if they are jutting (Figure 6).
- **Other**. Everything that does not fall within the previous classes has been included here. This class has the sole purpose of maintaining some architectural or furnishing elements (downpipes, benches, balustrades ...) which could be useful in the future and which, at the same time, help in the general understanding of the point cloud. For training and test phases, it is recommended to exclude this class, as it could adversely affect the loss function, the general performances of the neural networks or any other algorithms used.

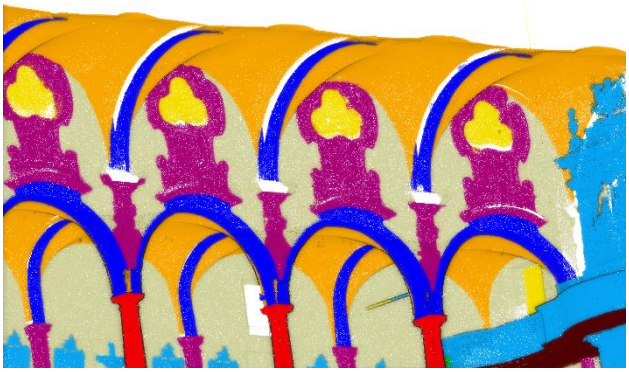

Figure 6. Examples of arches (blue) at a different height from the vaults (orange).

5. AIMS OF THE BENCHMARK AND EVALUATION

The benchmark is available at http://archdataset.polito.it/ and is divided into two sections:

- the point clouds already labelled for the training phases;
- the point clouds for the testing/evaluation.

In this way, the proposed benchmark could be used to train and evaluate state-of-the-art and new classification/segmentation methods. Furthermore, the users have the possibility to choose arbitrarily the scenes useful for their purposes.
The benchmark activity will also offer an evaluation of the performances of the segmentation methods. If authors will submit the predicted results for a given point cloud (ideally for all), we will automatically compare the achieved results with the ground truth ones and provide results in terms of Overall Accuracy, F1 Score, Precision, Recall and Intersection over Union (IoU).

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (1)$$

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \quad (2)$$

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \quad (3)$$

$$\text{F1 score} = 2 \times \frac{\text{Precision x Recall}}{\text{Precision} + \text{Recall}} \quad (4) \qquad \text{IoU} = \frac{|A \cap B|}{|A \cup B|} = \frac{|I|}{|U|} \quad (5)$$

Currently, the performances of state-of-art point cloud semantic segmentation networks are reported for PointNet (Qi et al., 2017a), PointNet++ (Qi et al., 2017b), PCNN (Atzmon et al., 2018), DGCNN (Wang et al., 2019) and modified DGCNN (Pierdicca et al., 2020). These DNNs are evaluated on 11 scenes out of 17 available.
A critical issue to be mentioned is the balancing of the classes (Figure 7). In fact, some of them, both for training and test, have a higher number of points and this can negatively affect the performance of the network and the various metrics.

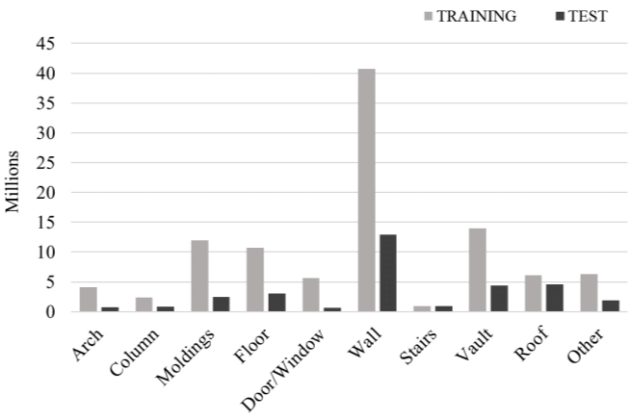

Figure 7. Number of points per class.

6. CONCLUSIONS

This paper describes *ArCH* benchmark, conceived for 3D point cloud semantic segmentation. The platform provides researchers with millions of points, labelled according to a defined standard, together with a generalised evaluation framework. The dataset comprises both annotated and not annotated point clouds, and we invite the research community in contributing to this tricky but essential task. Hopefully, in the upcoming months, the benchmark will become the reference source for testing and sharing new results and frameworks towards the end of automatizing object recognition for complex architectures. Some previous studies have demonstrated that CNN methods offer reliable strategies for 3D CH data classification. But it is fair to state that, conversely to other research domains, CH still presents several bottlenecks, which lead to the conclusion that, up to now, did not emerge an outperforming method. By providing open dataset and open source code, we foresee to infer a baseline for future implementations, as far as new algorithms will be developed in the near future. The class balancing, the heterogeneity of the architectural elements and the complexity of the scenes are currently the main drawbacks and open issues.
We are confident the benchmark meets the needs of the research activities in the heritage field and becomes a central resource for the development of new, efficient and accurate methods for classification of 3D heritage. The benchmark will strongly contribute to add the body of knowledge for semantic segmentation of CH good through automatic, supervised learning-based methods.

ACKNOWLEDGEMENTS

Authors would like to thank the scholarship holders Ilaria Bonfanti, Valeria De Ruvo, Emanuele Pontoglio and Gloria Rizzo of the Geomatics Lab of the Politecnico di Torino for their help in labelling the point clouds.
Moreover, sincere thankfulness goes to the VR Lab and the G4CH Lab of Politecnico di Torino (DISEG) and to the "Ente Gestore dei Sacri Monti" with its director dott. Elena De Filippis for allowing the publication of their point clouds. Thanks also to HM Sultan Sepuh XIV Arief Natadiningrat for having graciously accepted to share the point cloud of the Kasepuhan Palace. The authors would also like to thank the city of Ottmarsheim for its support.

REFERENCES

Armeni, I., Sener O., Zamir, A. R., Jiang, H., Brilakis, I., Fischer, M., Savarese, S., 2016. 3D semantic parsing of large-scale indoor spaces, Proc. IEEE CVPR, pp. 1534–1543.

Atzmon, M., Maron, H., Lipman, Y., 2018. Point convolutional neural networks by extension operators. *arXiv:1803.10091*.

Bello, S. A., Yu, S., Wang, C., 2020. Review: deep learning on 3D point clouds. *arXiv:2001.06280v1*.

Bruno, N., Roncella, R. A, 2018. Restoration oriented HBIM system for Cultural Heritage documentation: The case study of Parma cathedral. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci, 42, pp. 171-178.

Bitelli, G., Dellapasqua, M., Girelli, V., Sanchini, E., Tini, M., 2017. 3D Geomatics Techniques for an integrated approach to Cultural Heritage knowledge: The case of San Michele in Acerboli's Church in Santarcangelo di Romagna. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci., 42, pp. 291-296.

Wang, C.; Dai, Y.; Elsheimy, N.; Wen, C.; Retscher, G.; Kang, Z.; Lingua, A. Progress on ISPRS benchmark on multisensory indoor mapping and positioning, Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci, 42, pp. 1709-1713.

CloudCompare, 2019. CloudCompare 3D point cloud and mesh processing software Open Source Project. cloudcompare.org (last access: 01/05/2020).

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. and Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding. In *Proc. IEEE CVPR*, pp. 3213-3223.

De Deuge, M., Quadros, A., Hung, C., Douillard, B., 2013. Unsupervised feature learning for classification of outdoor 3D scans. In: *Australasian Conference on Robotics and Automation*, Vol. 2, p. 1

Geiger, A., Lenz, P., Stiller, C., Urtasun, R., 2013. Vision meets robotics: The KITTI dataset. *International Journal of Robotics Research*, 32(11), pp. 1231–1237.

Griffiths, D., Boehm, J., 2019. A Review on Deep Learning Techniques for 3D Sensed Data Classification. *Remote Sensing*, 11(12):1499

Grilli, E., Özdemir, E., & Remondino, F., 2019a. Application of Machine and Deep Learning strategies for the classification of heritage point clouds. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-4/W18, pp. 447-454.

Grilli, E., Farella, E. M., Torresani, A., and Remondino, F., 2019b. Geometric features analysis for the classification of cultural heritage point clouds. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W15, pp. 541–548.

Hackel, T., Savinov, N., Ladicky, L., Wegner, J.D., Schindler, K. and Pollefeys, M., 2017. Semantic3D. net: A new large-scale point cloud classification benchmark. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. IV-1-W1, pp. 91-98.

Han, X., Jin, J.S., Xie, J., Wang, M., Jiang, W., 2018. A comprehensive review of 3D point cloud descriptors. *ArXiv*, abs/1802.02297.

Kharroubi, A.,.Hajji, R., Billen, R., Poux, F., 2019. Classification and integration of massive 3D points clouds in a Virtual Reality

(VR) environment. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W17, pp. 165-171.

Korc, F. & Förstner, W., 2009. eTRIMS Image Database for interpreting images of man-made scenes. Dept. of Photogrammetry, University of Bonn, *Tech. Rep.* TR-IGG-P-2009-01.

Landrieu, L., Simonovsky, M., 2018. Large-scale point cloud semantic segmentation with superpoint graphs. Proceedings of *Proc. IEEE CVPR*, pp. 4558-4567.

Llamas, J., Lerones, P. M., Medina, R., Zalama, E., Gómez-García-Bermejo, J., 2017. Classification of Architectural Heritage Images Using Deep Learning Techniques. *Applied Sciences*, 9, Volume 7, p. 992.

Malinverni, E.S., Pierdicca, R., Paolanti, M., Martini, M., Morbidoni, C., Matrone, F., Lingua, A., 2019. Deep learning for semantic segmentation of point clouds. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* Vol. XLII-2/W15, pp. 735–742.

Mallik, A., Chaudhury, S., 2012. Acquisition of multimedia ontology: an application in preservation of cultural heritage. *International journal of multimedia information retrieval*, (1)4, 249-262.

Mo, K., Zhu, S., Chang, A. X., Yi, L., Tripathi, S., Guibas, L. J., & Su, H., 2019. Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3D object understanding. Proc. IEEE CVPR, pp. 909-918.

Munoz, D., Bagnell, J. A., Vandapel, N., Hebert, M., 2009. Contextual classification with functional max-margin markov networks. Proc. IEEE CVPR, pp. 975–982.

Murtiyoso, A.; Grussenmeyer, P., 2019. Point Cloud Segmentation and Semantic Annotation Aided by GIS Data for Heritage Complexes. In Proceedings of the 8th International Workshop 3D-ARCH "3D Virtual Reconstruction and Visualization of Complex Architecture", Bergamo, Italy, 6–8 February 2019, pp. 523–528.

Murtiyoso, A., Grussenmeyer, P., 2020. Virtual Disassembling of Historical Edifices: Experiments and Assessments of an Automatic Approach for Classifying Multi-Scalar Point Clouds into Architectural Elements. *Sensors*, 20(8), 2161.

Özdemir, E., Toschi, I., Remondino, F., 2019. A multi-purpose benchmark for photogrammetric urban 3D reconstruction in a controlled environment. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XLII-1/W2, pp. 53–60.

Paolanti, M., Romeo, L., Martini, M., Mancini, A., Frontoni, E., Zingaretti, P., 2019. Robotic retail surveying by deep learning visual and textual data. Robotics and Autonomous Systems, 118, pp. 179-188.

Pierdicca, R., Paolanti, M., Matrone, F., Martini, M., Morbidoni, C., Malinverni, E.S., Frontoni, E., Lingua, A.M., 2020. Point Cloud Semantic Segmentation Using a Deep Learning Framework for Cultural Heritage. *Remote Sens*., 12, 1005.

Quattrini, R., Pierdicca, R., Morbidoni, C., 2017. Knowledge-based data enrichment for HBIM: Exploring high-quality models using the semantic-web. *Journal of Cultural Heritage*, 28, pp. 129-139.

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017a. 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., Guibas, L.J., 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *arXiv:1706.02413*.

Roynard, X., Deschaud, J.E. and Goulette, F., 2018. Paris-Lille-3D: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification. *The International Journal of Robotics Research*, 37(6), pp.545-557.

Semantic Segmentation Editor: web labeling tool for camera and LIDAR data. Available online: https://github.com/Hitachi-Automotive-AndIndustry-Lab/semantic-segmentation-editor. (last access on 25/04/2020).

Serna, A., Marcotegui, B., Goulette, F. and Deschaud, J.-E., 2014. Paris-rue-madame database: a 3d mobile laser scanner dataset for benchmarking urban detection, segmentation and classification methods. *Proc. 4th ICPRAM Conference*.

Teboul, O., Kokkinos, I., Simon, L., Koutsourakis, P., Paragios, N., 2013. Parsing Facades with Shape Grammars and Reinforcement Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7, Volume 35, pp. 1744-1756.

Tyleček, R. and Šára, R., 2013, September. Spatial pattern templates for recognition of objects with regular structure. In *German Conference on Pattern Recognition*, pp. 364-374. Springer, Berlin, Heidelberg.

Vallet, B., Brédif, M., Serna, A., Marcotegui, B. and Paparoditis, N., 2015. TerraMobilita/iQmulus urban point cloud analysis benchmark. *Computers & Graphics*, 49, pp.126-133.

Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M., 2019. Dynamic graph CNN for learning on point clouds. ACM Trans. Graph. *arXiv:1801.07829v2*.

Wang, C., Hou, S., Wen, C., Gong, Z., Li, Q., Sun, X., Li, J., 2018. Semantic Line Framework-based Indoor Building Modeling using Backpacked Laser Scanning Point Cloud, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 143, pp. 150-166.

Weinmann, M., Jutzi, B., Hinz, S., & Mallet, C., 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, pp. 286-304.

Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J., 2015. 3Dshapenets: A deep representation for volumetric shapes. *Proc. IEEE CVPR*, pp. 1912– 1920.

Yang, X., Lu, Y.C, Murtiyoso, A., Koehl, M., Grussenmeyer, P., 2019. HBIM modeling from the surface mesh and its extended capability of knowledge representation. *ISPRS International Journal of Geo-Information*, (8)7, 301.

Zhang, K., Hao, M., Wang, J., Silva, C.W., & Fu, C., 2019. Linked Dynamic Graph CNN: Learning on Point Cloud via Linking Hierarchical Features. *ArXiv*, abs/1904.10014.