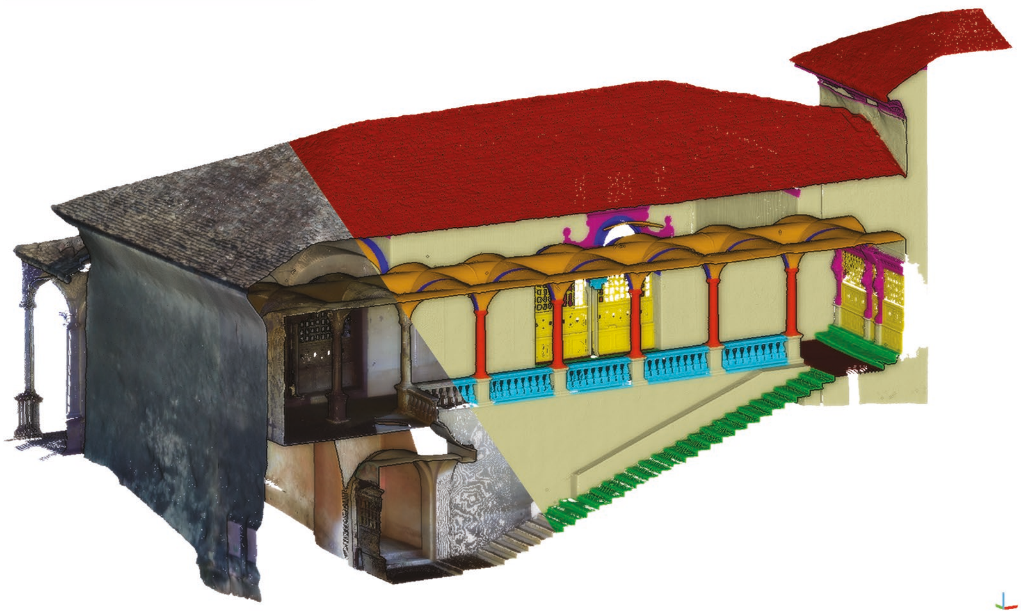
Deep Semantic Segmentation of Cultural Built Heritage Point Clouds: Current Results, Challenges and Trends

Francesca Matrone

Abstract

In the digital Cultural Heritage domain, the ever-increasing availability of 3D point clouds provides the opportunity to rapidly generate detailed 3D scenes to support the restoration, conservation, maintenance and safeguarding activities of built heritage. The semantic enrichment of these point clouds could support the automatization of the scan-to-BIM processes. In this framework, the use of Artificial Intelligence techniques for the automatic recognition of architectural elements from point clouds can thus provide valuable support.
The described methodology allows increasing the Level of Detail in the semantic segmentation of built heritage point clouds compared to the current state-of-the-art through deep neural networks.
The main outcome is therefore the first application of DL framework for CH point clouds, with the subsequent implementation of the selected neural network (the DGCNN) for the semantic segmentation task. These results also permit to evaluate the pros and cons of this approach, along with future challenges and trend

Keywords
semantic segmentation, point clouds, deep neural networks, cultural heritage.



doi.org/10.3280/oa-845-c193

Introduction

2018 was the *European Year of Cultural Heritage* and, within the related *European Framework for Action on Cultural Heritage*, the innovation and the use of digital technologies to enhance access to cultural heritage (CH) creating digital contents have been highly incentivized, recalling also what stated during the *Council of Europe Framework Convention on the Value of Cultural Heritage for Society* (Faro Convention of 2005).
In this sense, geomatics can bring significant benefits to the CH field. Among the main ones, it allows digital data production to catalogue and preserve historical memory, for the analysis and conservation of assets (movable and immovable), for their fruition even remotely or for the safeguarding in conditions of risk or vulnerability. In particular, point clouds are an increasingly used tool for asset management, and their always greater involvement is mainly due to the latest developments of faster and more efficient acquisition tools such as Mobile Mapping Systems (MMSs). Combining these systems with more consolidated techniques, as terrestrial laser scanning or aerial photogrammetry, allows the acquisition of massive amounts of data, sometimes even excessive. In fact, the generated point clouds are usually subsampled, filtered and post-processed for an effective use and to simplify their management.
This element has created an increasing interest of the scientific community towards the use, interpretation and direct exploitation of point clouds, contributing to the widespread usage of this 3D data also in other sectors such as autonomous navigation, robotics and bioinformatics. Related to this type of data, a new trend has recently emerged in information technology: the *semantic segmentation* of point clouds through artificial intelligence techniques such as Machine and Deep Learning (ML/DL). This tendency allows point clouds to be used as a basis for 3D modelling or as a support for semantic data processing. In the geomatics field, the subdivision of the point clouds into predefined categories (for an architectural or urban/regional scale) entails various tasks: speeding up the reconstruction of 3D models, automating analysis in GIS environments, supporting 3D city modelling, and so on. In particular, it could also be beneficial for the semantic enrichment of HBIM (Historic Building Information Modelling) and to speed up the reconstruction of parametric objects, whose scan-to-BIM process is still entrusted with manual operations. Experts are yet claimed at handling large and complex datasets without the aid of any automatic or semi-automatic method to recognize and reshape 3D elements. These operations are usually time-consuming and, as mentioned above, involve the waste of a large amount of data, given the unavoidable simplification exerted since the objects can be described through a few relevant points or contours.
In this scenario, the comeback of DL has been overwhelming [Griffiths, Boehm 2019a, pp. 1-29], and deep Neural Networks (NNs) settled as the more efficient technology for learning-based tasks [Bello et al. 2020, pp. 1-34]. However, although artificial NNs proved to be very promising for handling and recognizing 3D data, for CH, manual operations still look more trustworthy. There are many reasons for such skepticism; first of all, CH assets 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 unsupervised 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 deep NNs for CH; on the one hand, the lack of training data, on the other, the computational effort.
To foster research in this direction, it has been implemented an automatic semantic segmentation workflow along with the setting of a newly created database to be used as a common base. Besides, a Level of Detail (LoD) higher than the one achieved in the state-of-the-art for the semantic segmentation of point clouds [Boulch et al. 2018, pp. 189-198; Landrieu, Simonovsky 2018, pp. 4558-4567; Weinmann et al. 2015, pp. 271-278] was likewise required. In fact, among the usual and general categories as building, vegetation, street or vehicle, the *building* class lacks further detail, e.g. *roof, column, stair, arch, floor* or *vault* (Fig. 1) which have been here investigated.

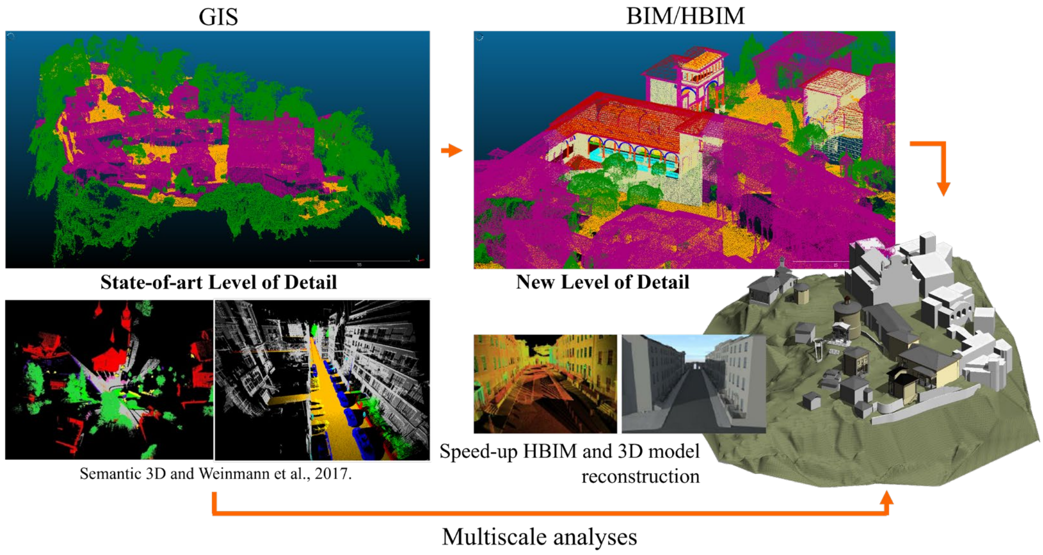


Fig. 1. Comparison between the state-of-the-art LoD for semantic segmentation and the proposed one. On the bottom, the HBIM models obtained from the point clouds.

The research questions this study tried to address are:
• is it possible to operate a multiscale point cloud semantic segmentation from urban to architectural scale? And which LoD should be selected?
• are DL techniques suitable for the CH domain, where the standardization of the elements, which should help automatic recognition, is almost absent?
• what are the pros and cons of the proposed methodology?

Point Cloud Semantic Segmentation Approaches

When dealing with point clouds, it has to be taken into account that they are geometric structures of irregular nature, characterized by the lack of a grid, with a high variability of density, unordered and invariant to transformation and permutation [Zaheer et al. 2017, pp. 3392-3402], making the use of DL techniques not straightforward and even more challenging when dealing with digital CH oriented datasets.

Method	Drawback	Pro
Multiview-based	• Limitations and loss in geometric structures • Laborious to choose enough appropriate viewpoints for the multi-view projection • Information bottleneck: limited exploitation of the potential of 3D data • Duplication of raw data	• Solution to the structuring problems of point cloud data • Easy application of CNN that proved to achieve excellent results
Voxel-based	• High memory and computational power required • Introduction of discretization artefacts • Data loss	• Ordered grid, maintaining the continuous properties and the tridimensionality of the point cloud • Exploitation of the potential of 3D data
Point-based	• Moderate computational power required • Scaling to a larger scene is still unexplored	• Bypassed the structuring problems of point cloud data • Direct exploitation of 3D data, particularly useful for the scan-to-BIM process • Latest trend of using graph NN can help integrating prior knowledge into the model
All	• Large amount of data required for training the model • Mainly focused on indoor scenes or for aerial LiDAR point clouds • Not designed for CH domain • Lack of a comprehensive and labelled CH dataset of point clouds	• Process automation and unsupervised learning in many cases

Tab. 1. Main weaknesses and strengths provided by the exploitation of point clouds with DL techniques.

On the other side, using point clouds make it possible to automate the recognition of the various architectural elements in the object-oriented software and overcome some limitations given by the use of 2D images, such as data incompleteness (given by the lack of three-dimensionality), lighting problems or possible occlusions.

Currently, point cloud semantic segmentation approaches can be divided into three categories [Zhang et al. 2019, pp. 179118-179133]:
• *Multiview-based*: they rely on the creation of a set of images from point clouds, on which Convolutional Neural Networks (CNN) can be applied;
• *Voxel-based*: they consist in the rasterization of point clouds in voxels, which allow having an ordered grid of point clouds, while maintaining the continuous properties and the third dimension, thus permitting the application of CNNs;
• *Point-based*: the classification and semantic segmentation are performed by applying feature-based approaches, directly exploiting the point clouds.
The main strengths and weaknesses of the point clouds exploitation methods, defined on the basis of the literature review, are reported in Tab. 1.

Selection of neural networks for testing

Based on the considerations presented in Table 1, the point-based methods were chosen for the semantic segmentation in the CH domain, even if recent studies apply also other approaches [Pellis et al. 2022, pp. 429-434]. In particular, among the networks proposed in the state-of-the-art, the deep NNs selected for this study have been: PointNet [Qi et al. 2017a, pp. 77-85], its extensions PointNet++ [Qi et al. 2017b, pp. 5100-5109], Point CNN [Atzmon et al. 2018, pp. 1-14] and the *Dynamic Graph CNN*, *DGCNN* [Wang et al. 2019, pp. 1-12]. The latter addresses many shortcomings of the previous works and consumes point clouds through graph structures.
With respect to these four deep NNs, the DGCNN has proved to achieve good results with the proposed dataset; therefore, it has been deepened and modified for the purposes of this research. PointNet++ and Point CNN (PCNN), on the other side, were less generalizable, and they seemed to work well mainly with small datasets and simple classes as in the case of ScanNet. Their results have been described in [Pierdicca et al. 2020, pp. 1-23].

Dataset

From the state-of-the-art investigation, it emerged that there are few datasets specific for some CH areas, such as [Korc, Förstner 2009; Teboul et al. 2011, pp. 2273-2280; Tyleček, Šára 2013, pp. 364-374]; nevertheless, they only provide bidimensional data, and they mainly consist of manually annotated façade images from different cities around the world and diverse architectural styles.
Regarding 3D data, an interesting project named OpenHeritage 3D has been proposed to provide open access to 3D CH datasets and foster community collaboration. However, only not labelled point clouds are available.
Precisely for these reasons, it has not been possible to identify one suitable dataset; hence an *ad hoc* one has been created.
The created dataset constitutes a new 3D large-scale benchmark for heritage point clouds (named *ArCH dataset – Architectural Cultural Heritage*) with millions of manually labelled points belonging to heritage scenarios [Matrone et al. 2020, pp. 1419-1426].
It has been made available for the scientific community, and it originates from the collaboration of different universities and research institutes (Politecnico di Torino, Università Politecnica delle Marche, FBK Trento, Italy, and INSA Strasbourg, France), offering for the first time, annotated point clouds describing heritage scenes.

Methodology

In the following section, the DL framework based on the DGCNN and its implementation [Pierdicca et al. 2020, pp. 1-23] will be only briefly outlined, giving more prominence to overall discussions and considerations on the method, highlighting the main results obtained with the relative pros and cons and the challenges to be faced for the next future.
Generally speaking, a symmetrical scene was chosen to perform the preliminary tests to set the network (Fig. 2, part 1); then its generalization capability was tested, training it from scratch on multiple scenes (Fig. 2, part 2), finally, the best configurations were tested on the whole ArCH dataset (Fig. 2, part 3).
The achieved results are compared with the Ground Truth (GT) in terms of *Overall Accuracy* (OA), *F1-Score*, *Precision*, *Recall* and *mean Intersection over Union* (mIoU).

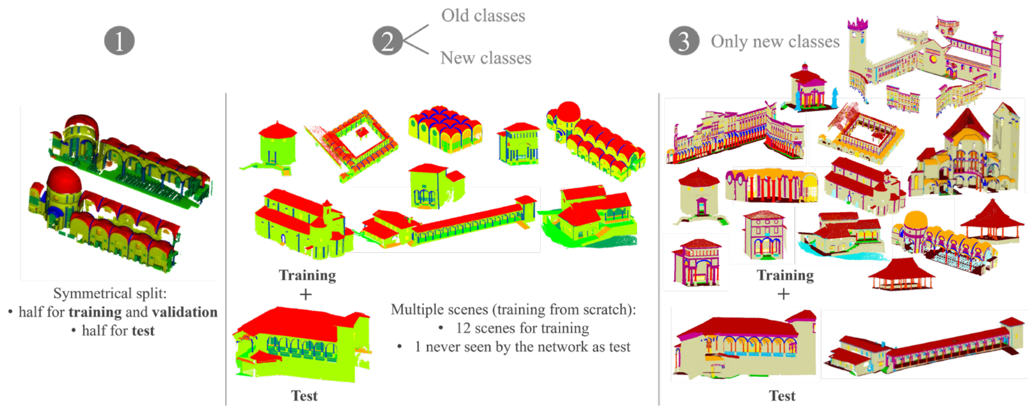


Fig. 2. Overview of the tests subdivision.

k-NN and Hyperparameters Setting

Once chosen the DGCNN, it was necessary to adapt it to the ArCH dataset.
To do this, a symmetrical point cloud was first selected from the dataset to efficiently carry out the preliminary tests on the network with moderate calculation times and computational power. In fact, the scene was split into two parts along the symmetry axis: one half was used as a training/validation set and the other half for testing. This method allowed setting the hyperparameters and provided the basis for all subsequent tests.
In the original DGCNN the analysis of the scenes takes place through endless *blocks*, interspersed with a certain *stride* (Fig. 3).
A certain number of subsampled points to be used as input for the network is then defined within each block. Therefore, it has been assessed whether different types of subsampling of the initial point clouds (*octree* or *space*) could affect the final performances. As a result, *block size*, *stride* and *number of subsampled points* were the first hyperparameters to be tested.

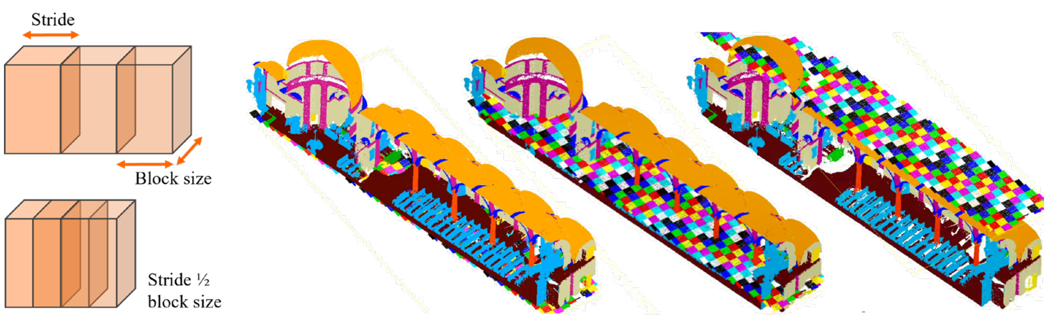


Fig. 3. The relation between stride and block size and footprint of the endless blocks along the half scene of the Trompone church to highlight the scene subdivision.

Since the ArCH dataset's point clouds also contain the radiometric component (expressed as RGB) and normal vectors, these values were also used as input for the network. In the original DGCNN at the input layer, k-NN is fed with normalized points coordinates only, while in this proposal all the available features were used. A scene block is thus introduced into the DNN, composed of 12 features for each point: *x y z* coordinates, *x' y' z'* normalized coordinates, 3 colour features (RGB or its conversion into Hue Saturation Value channels – HSV – or L*a*b*) and *Nx Ny Nz* normal vectors. This architecture was named DGCNN-Mod (Modified).

Class Imbalance and DGCNN Implementation

The results of the preliminary tests highlighted a relevant issue of class imbalance. In fact, all evaluated approaches failed in recognizing classes with low support, as *doors*, *windows* and *arches*. Besides, for these classes, high variability in shapes across the dataset was noticed [Pierdicca et al. 2020, pp. 1-23], and this element probably contributed to the networks' poor accuracy.
To remedy the class imbalance, several different approaches have been proposed in the literature, e.g. [Buda et al. 2018, pp. 249-259; Pouyanfar et al. 2018, pp. 112-117; Ando, Huang 2017, pp. 770-785; Griffiths, Boehm 2019b, pp. 981-987]. Among these proposals, the change of the loss function and data augmentation techniques, focused only on the minor classes, have been selected. In particular, according to the work of [Lin et al. 2020, pp. 318-327], a new type of loss has been chosen and implemented in the DGCNN: the focal loss. It is designed to solve the issue of the imbalance down-weighting the classes containing more examples to target the training on the categories with fewer samples.
It was then decided to help the network with ad hoc features to discriminate the classes. Based on the insights of [Grilli et al. 2019, pp. 541-548; Weinmann et al. 2015, pp. 271-278], a few 3D features were introduced to evaluate whether their contribution could be similar to that produced with ML classifier as RF. These 3D features derive from a compound of eigenvalues ($\lambda 1>\lambda 2>\lambda 3$), and they can describe and emphasize the different architectural elements in a particularly explicit way. Those selected have been *verticality* (f1 and f3), *omnivariance* (f2), *surface variation* (f4), *planarity* (f5) and *z value* (f6) (Fig. 4), so the new complete input data is ordered as follow: *x, y, z, R, G, B, f1, f2, f3, f4, f5, f6, Nx, Ny, Nz*.

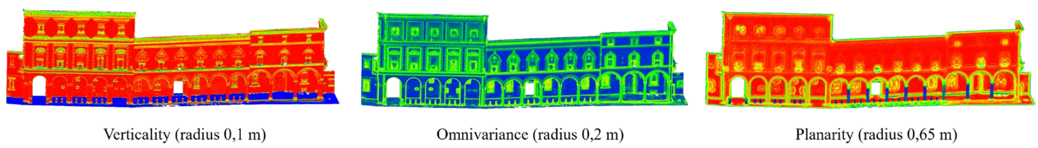


Fig. 4. Example of 3D features extracted and relative radius.

With these new 3D features, the performance of the DGCNN-Mod is compared with two novel versions of this network: the DGCNN-3Dfeat and the DGCNN-Mod+3Dfeat. In particular, the DGCNN-3Dfeat adds to the k-NN only the 3D features; whereas, for a complete comparison, the DGCNN-Mod+3Dfeat comprises all the 18 available features [Matrone et al. 2020b, pp. 1-22].
The positive insights of 3D features and eigenvalues led to consider the option of concatenating them to those learned from the network so that they could be available in the last layer before the semantic segmentation task. This procedure, theoretically, should lead to using the features with their informative contents as they are, and not reworked by the deep NN, adding new info to what has already been learned, and improving the model convergence. Based on [Huang et al. 2017], a new structure has been thus created to concatenate the initial 3D features with the last layer, defined as skip connection.
Besides, a DL approach can also be improved by using particular data augmentation

techniques on the training data. This solution is quite common with images, where colour space augmentations, random part exclusion, geometric transformations, kernel filters and so on can be applied and could also be used to prevent class imbalance. Many of the usual techniques cannot be chosen with point cloud data, but there are other methods, where the point cloud is augmented on-the-fly. In this case, rotation, clipping, spatial shifting, jittering and scaling strategies have been implemented along with transfer learning techniques [Matrone, Martini 2021, pp. 73-84].

Results and Discussions

The results obtained allowed an increased Level of Detail in the semantic segmentation of built heritage point clouds.
Specifically, the literature review has made it possible to identify several criticalities in the application of the DL framework to the CH domain, in particular:
• the scarce development of DL techniques for this domain and, even less, applied to heritage point clouds;
• the lack of an adequate LoD for the semantic segmentation of CH point clouds;
• the absence of a dataset consistent with the aims of this research.
Precisely for these reasons, the ArCH dataset was expressly created to provide a starting point for future developments in this field; however, it does not constitute a sufficiently exemplary dataset of the multitude of architectural cultural elements, very variable and different from each other across the various architectural lexicons.
For the annotation of this dataset, a Level of Detail equal to 3 was selected, according to the CityGML standard. This LoD improves the one present in the state-of-the-art, but on the other hand, it can be further increased only in proportion to the size of the datasets available for the DL techniques.
The results of the tests performed highlighted the importance of introducing normal vectors (and even more their correct orientation) and the radiometric component. Concerning the latter point, the test performed to investigate the relation between the colour channels and the individual classes showed that the use of HSV led to a slight increase in the performances. Regarding the subsampling method, the variation in the results between octree and space-based methods was about 1% of OA in favour of the octree. In this case, the immediacy of the space-based method was chosen for the following tests, even if with slightly lower results.
With this configuration, 73% of OA was obtained for the symmetrical scene and 83% of OA for the tests conducted with part of the benchmark scenes as a training set. These first results showed a good recognition of those classes represented by more points in the training set, and a significant criticality for the categories with fewer points. Class balancing has, therefore, turned out to be one of the main issues to be addressed. The introduction of focal loss, to overcome it, did not guarantee overwhelming results. In fact, the arch class was the only one to improve its metrics, while for the other classes with fewer points (column, door/window, stair and molding) common pattern could not be identified.
The introduction of 3D features is the element that, most of all, boosted the performances of the network: in the symmetrical scene, it led to an increase of about 10% of OA, while with an unseen scene of about 3%. Although the gain in OA is smaller in the second case, if the classes with a low number of points are considered, it emerges that almost all of them improve their metrics with the use of 3D features. Considering the F1-score: arch + 1%, column + 43%, door/window + 16%, stair + 14% are obtained. Therefore, it can be said that their introduction, associated with the use of focal loss, has led to the expected results.
The skip connection's introduction, to further improve the model convergence, resulted not very effective in terms of OA, but useful for discriminating some specific classes such as molding and door/window. Comparing the F1-score between two tests with and without skip connection a + 6% for the molding and + 2% for the door/window are recorded.

The data augmentation approach has confirmed a viable path for point clouds, even if, as described for the previous tests, it was impossible to identify a common pattern: some classes are discriminated better than others alternately, depending on the combinations of hyperparameters used.
Broadly speaking, to pursue DL generalization, the classic solution of expanding and implementing the training set is still the most suitable one, but the lack of additional datasets remains a compelling criticality for future developments.
In conclusion, after choosing the approach and the deep NN to be implemented, it has been possible to step from an initial 56.1% of OA up to a final 86.3% (86.6% with the whole ArCH dataset) (Fig. 5).

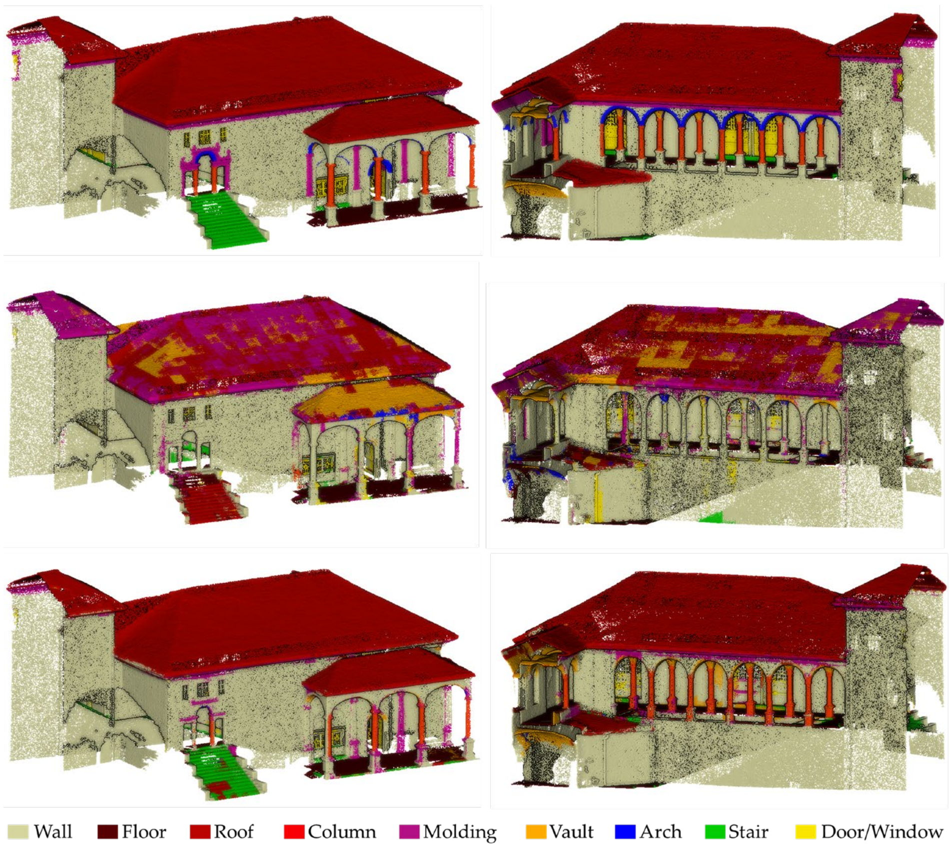


Fig. 5. Ground truth (on the top), the prediction of the first test with the DGCNN (in the middle) and the final test with the DGCNN-Mod+3Dfeat (on the bottom). The represented point cloud is one of the two test scenes of the ArCH dataset, and it belongs to the Sacro Monte of Varallo (SMV).

Conclusions: Future Perspectives and Challenges

Recalling the initial research questions it can be stated that it is certainly possible to increase the Level of Detail for the deep semantic segmentation of point clouds representing buildings or architectural assets. In particular, to date, it is feasible to reach a LoD equal to 3 with point-based approaches.
Although the CH domain is characterized by patchy elements and, consequently, poorly standardized, the creation of a new annotated dataset has provided the basis for the application of DL techniques even on CH point points. Currently, given the lack of labelled data, ML classifiers (such as Random Forest) are an excellent alternative, but they do not define the winning solution.
Based on the obtained results, future developments of this topic may consist in the integration of an ontological structure or taxonomy within the neural network, in order to guide and eventually correct it in the learning phases and the automatic training data generation to increase the dataset size.

The methodology proposed includes, among its strengths, the possibility of guaranteeing unsupervised learning, thus limiting the manual intervention of operators in processes such as the scan-to-BIM. In addition, feature engineering is less time-consuming with respect to other classifiers since the neural network can automatically learn the discriminating features. Finally, a good generalization and a high tuning capacity of the hyperparameters is also guaranteed.
The weaknesses, on the other hand, reside in the training set size dependencies, training time (strongly dependent on the hardware used and still higher than traditional classifiers) and unbalancing of the classes. This last aspect is undoubtedly a challenge for deep learning techniques applied to point clouds and, even more, for the CH field. As for the other disadvantages mentioned above, they will be partially solved in the next future with the continuous technological developments, which will assure ever higher computational powers. Regarding the scarcity of labelled datasets and tests, it should be noted that recent studies e.g. [Wysocki et al. 2022, pp. 529-536; Cao, Scaioni 2022, pp. 1-22; Pellis et al. 2022, pp. 429-434] are focusing on this topic, thus contributing: i) to the diffusion of the dataset, ii) to its extension and iii) to the improvement of NNs' performances trained on built CH point clouds. Likewise, a further research trend linked to the semantic segmentation task consists in using the output of artificial intelligence algorithms as input for additional processing, e.g. scan-to-BIM procedures [Croce et al. 2021, pp. 1-34] and/or mixed reality [Teruggi et al. 2021, pp. 155-162].
The scientific community's increasing and compelling interest in point cloud processing and semantic segmentation certainly leaves hope for excellent future prospects.out skip connection a + 6% for the molding and + 2% for the door/window are recorded.
The data augmentation approach has confirmed a viable path for point clouds, even if, as described for the previous tests, it was impossible to identify a common pattern: some classes are discriminated better than others alternately, depending on the combinations of hyperparameters used.
Broadly speaking, to pursue DL generalization, the classic solution of expanding and implementing the training set is still the most suitable one, but the lack of additional datasets remains a compelling criticality for future developments.
In conclusion, after choosing the approach and the deep NN to be implemented, it has been possible to step from an initial 56.1% of OA up to a final 86.3% (86.6% with the whole ArCH dataset) (Fig. 6).

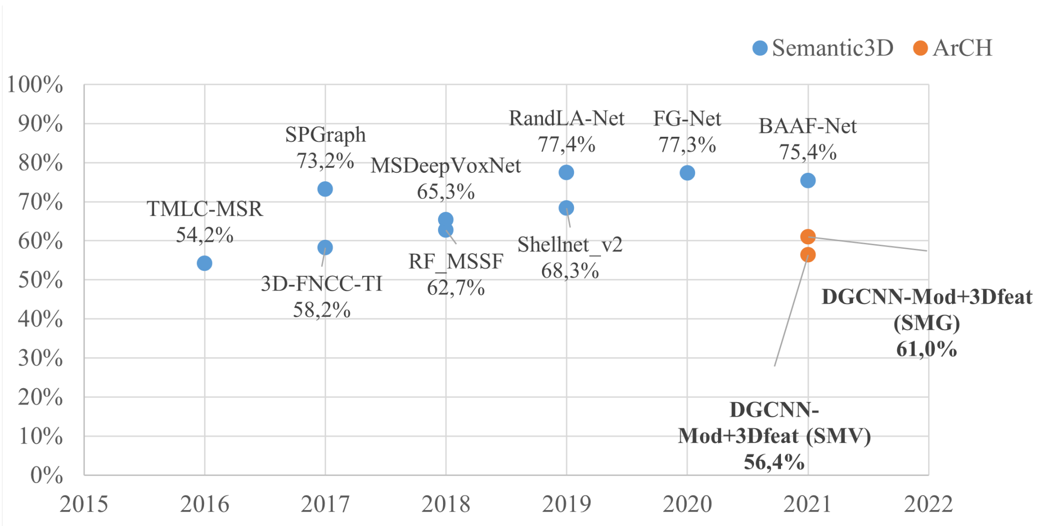


Fig. 6. Comparison with the state-of-the-art performances of Semantic3D benchmark compared to ArCH in terms of mIoU. SMV and SMG correspond to the two test point clouds of the ArCH dataset.

Conclusions: Future Perspectives and Challenges

Recalling the initial research questions it can be stated that it is certainly possible to increase the Level of Detail for the deep semantic segmentation of point clouds representing buildings or architectural assets. In particular, to date, it is feasible to reach a LoD equal to 3 with point-based approaches.
Although the CH domain is characterized by patchy elements and, consequently, poorly standardized, the creation of a new annotated dataset has provided the basis for the application of DL techniques even on CH point points. Currently, given the lack of labelled data, ML classifiers (such as Random Forest) are an excellent alternative, but they do not define the winning solution.
Based on the obtained results, future developments of this topic may consist in the integration of an ontological structure or taxonomy within the neural network, in order to guide and eventually correct it in the learning phases and the automatic training data generation to increase the dataset size.
The methodology proposed includes, among its strengths, the possibility of guaranteeing unsupervised learning, thus limiting the manual intervention of operators in processes such as the scan-to-BIM. In addition, feature engineering is less time-consuming with respect to other classifiers since the neural network can automatically learn the discriminating features. Finally, a good generalization and a high tuning capacity of the hyperparameters is also guaranteed.
The weaknesses, on the other hand, reside in the training set size dependencies, training time (strongly dependent on the hardware used and still higher than traditional classifiers) and unbalancing of the classes. This last aspect is undoubtedly a challenge for deep learning techniques applied to point clouds and, even more, for the CH field. As for the other disadvantages mentioned above, they will be partially solved in the next future with the continuous technological developments, which will assure ever higher computational powers. Regarding the scarcity of labelled datasets and tests, it should be noted that recent studies e.g. [Wysocki et al. 2022, pp. 529-536; Cao, Scaioni 2022, pp. 1-22; Pellis et al. 2022, pp. 429-434] are focusing on this topic, thus contributing: i) to the diffusion of the dataset, ii) to its extension and iii) to the improvement of NNs' performances trained on built CH point clouds. Likewise, a further research trend linked to the semantic segmentation task consists in using the output of artificial intelligence algorithms as input for additional processing, e.g. scan-to-BIM procedures [Croce et al. 2021, pp. 1-34] and/or mixed reality [Teruggi et al. 2021, pp. 155-162].
The scientific community's increasing and compelling interest in point cloud processing and semantic segmentation certainly leaves hope for excellent future prospects.

References

Ando Shin, Huang, Chun Yuan (2017). Deep Over-sampling Framework for Classifying Imbalanced Data. In Ceci Michelangelo, Hollmén Jaakko, Todorovski Ljupčo, Vens Celine, Džeroski Sašo (eds). *Machine Learning and Knowledge Discovery in Databases. Lecture Notes in Artificial Intelligence*, 10534. Cham: Springer, pp. 770-785.

Atzmon Matan, Maron Haggai, Lipman Yaron (2018). Point convolutional neural networks by extension operators. In *ACM Transactions on Graphics*, 37(4), 2018, pp.1-14.

Bello Saifullahi Aminu, Yu Shangshu, Wang Cheng, Adam Jibril Muhmmad, Li Jonathan (2020). Review: Deep learning on 3D point clouds. In *Remote Sensing*, 12(11), 2020, pp.1-34.

Boulch Alexandre, Guerry Joris, Le Saux Bertrand, Audebert Nicolas (2018). SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks. In *Computers and Graphics (Pergamon)*, 71, 2018, pp. 189-198.

Buda Mateusz, Maki Atsuto, Mazurowski, Maciej (2018). A systematic study of the class imbalance problem in convolutional neural networks. In *Neural Networks*, 106, 2018, pp. 249-259.

Cao Yuwei, Scaioni Marco (2021). 3DLEB-Net: Label-Efficient Deep Learning-Based Semantic Segmentation of Building Point Clouds at LoD3 Level. In *Applied Sciences*, 11(19), 2021, art. 8996.

Croce Valeria, Caroti Gabriella, De Luca Livio, Jacquot Kevin, Piemonte Andrea, Véron Philippe (2021). From the semantic point cloud to heritage-building information modeling: A semiautomatic approach exploiting machine learning. In *Remote Sensing*, 13(3), 2021, art. 461.

Griffiths David, Boehm Jan (2019a). A Review on deep learning techniques for 3D sensed data classification. In *Remote Sensing*, 11(12), 2019, pp. 1-29.

Griffiths David, Boehm Jan (2019b). Weighted point cloud augmentation for neural network training data class-imbalance. In *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 42(2/W13), 2019, pp. 981-987.

Grilli Eleonora, Farella Elisa Mariarosaria, Torresani Alessandro, Remondino Fabio (2019). Geometric Features Analysis for the Classification of Cultural Heritage Point Clouds. In *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W15, 2019, pp. 541-548.

Korc Filip, Förstner Wolfgang (2009). *eTRIMS Image Database for Interpreting Images of Man-Made Scenes*. Technical Report TR-IGG-P-2009-01.

Landrieu Loic, Simonovsky Martin (2018). Large-Scale Point Cloud Semantic Segmentation with Superpoint Graphs. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, pp. 4558-4567.

Lin Tsung-Yi, Goyal Priya, Girshick Ross, He Kaiming, Dollar Piotr (2020). Focal Loss for Dense Object Detection. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 2020, pp. 318-327.

Matrone Francesca, Grilli Eleonora, Martini Massimo, Paolanti Marina, Pierdicca Roberto, Remondino Fabio (2020b). Comparing machine and deep learning methods for large 3D heritage semantic segmentation. In *ISPRS International Journal of Geo-Information*, 9 (9), 2020, art. 535.

Matrone Francesca, Lingua Andrea, Pierdicca Roberto, Malinverni Eva Savina, Paolanti Marina, Grilli Eleonora, Remondino Fabio, Murtiyoso Arnadi, Landes Tania (2020a). A benchmark for large-scale heritage point cloud semantic segmentation. In *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B2-2020, 2020, pp. 1419-1426.

Matrone Francesca, Martini Massimo (2021). Transfer learning and performance enhancement techniques for deep semantic segmentation of built heritage point clouds. In *Virtual Archaeology Review*, 12 (25), 2021, pp. 73-84.

Pellis Eugenio, Murtiyoso Arnadi, Masiero Andrea, Tucci Grazia, Betti Michele, Grussenmeyer Pierre (2022). An image-based deep learning workflow for 3D heritage point cloud semantic segmentation. In *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVI-2/W1-2022, 2022, pp. 429-434.

Pierdicca Roberto, Paolanti Marina, Matrone Francesca, Martini Massimo, Morbidoni Christian, Malinverni Eva Savina, Lingua Andrea (2020). Point cloud semantic segmentation using a deep learning framework for cultural heritage. In *Remote sensing*, 12(6), 2020, pp. 1-23.

Pouyanfar Samira, Tao Yudong, Mohan Anup, Tian Haiman, Kaseb Ahmed, Gauen Kent, Dailey Ryan, Aghajanzadeh Sarah, Lu Yung-Hsiang, Chen Shu-Ching, Shyu Mei-Ling (2018). Dynamic Sampling in Convolutional Neural Networks for Imbalanced Data Classification. In *Proceedings – IEEE 1st Conference on Multimedia Information Processing and Retrieval, MIPR 2018*, pp. 112-117.

Qi Charles Ruizhongtai, Su Hao, Mo Kaichun, Guibas Leonidas John (2017a). PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 77-85.

Qi Charles Ruizhongtai, Yi Li, Su Hao, Guibas Leonidas John (2017b). PointNet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems*, 2017-Decem, pp. 5100-5109.

Teboul Olivier, Kokkinos Iasonas, Simon Loic, Koutsourakis Panagiotis, Paragios Nikos (2011). Shape grammar parsing via reinforcement learning. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2273-2280.

Teruggi Simone, Grilli Eleonora, Fassi Francesco, Remondino Fabio (2021). 3D surveying, semantic enrichment and virtual access of large cultural heritage. In *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, VIII-M-1-2021, 2021, pp. 155-162.

Tyleček Radim, Šára Radim (2013). Spatial pattern templates for recognition of objects with regular structure. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8142 LNCS, pp. 364-374.

Wang Yue, Sun Yongbin, Liu Ziwei, Sarma Sanjay Emani, Bronstein Michael, Solomon Justin (2019). Dynamic Graph CNN for Learning on Point Clouds. In *Computer Vision and Pattern Recognition*, 1(1), 2019, pp. 1-12.

Weinmann Martin, Schmidt Alena, Mallet Clément, Hinz Stefan, Rottensteiner Franz, Jutzi Boris (2015). Contextual classification of point cloud data by exploiting individual 3D neigbourhoods. In *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, II-3/W4, 2015, pp. 271-278.

Wysocki Olaf, Hoegner Ludwig, Stilla Uwe (2022). TUM-façade: reviewing and enriching point cloud benchmarks for façade segmentation. In *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVI-2/W1-2022, 2022, pp. 529-536.

Zaheer Manzil, Kottur Satwik, Ravanbhakhsh Siamak, Póczos Barnabas, Salakhutdinov Ruslan, Smola Alexander (2017). Deep sets. In *Advances in Neural Information Processing Systems*, 2017-Decem(ii), pp. 3392-3402.

Zhang Jiaying, Zhao Xiaoli, Chen Zheng, Lu Zhejun (2019). A Review of Deep Learning-Based Semantic Segmentation for Point Cloud. In *IEEE Access*, 7, 2019, pp. 179118-179133.

Authors
Francesca Matrone, Dept. of Environment, Land and Infrastructure Engineering, Politecnico di Torino, francesca.matrone@polito.it.

 Isbn 9788835141945