

GRAPH CNN WITH RADIUS DISTANCE FOR SEMANTIC SEGMENTATION OF HISTORICAL BUILDINGS TLS POINT CLOUDS

C. Morbidoni*, R. Pierdicca, R. Quattrini, E. Frontoni

Università Politecnica delle Marche, Ancona, Italy - (c.morbidoni, r.pierdicca, r.quattrini, e.frontoni)@staff.univpm.it

KEY WORDS: Semantic Segmentation, 3D Point Clouds, Terrestrial Laser Scanning, Deep Learning, Radius distance, Dynamic Graph Convolutional Neural Networks, EdgeConv, Historical Buildings

ABSTRACT:

Point clouds obtained via Terrestrial Laser Scanning (TLS) surveys of historical buildings are generally transformed into semantically structured 3D models with manual and time-consuming workflows. The importance of automatizing this process is widely recognized within the research community. Recently, deep neural architectures have been applied for semantic segmentation of point clouds, but few studies have evaluated them in the Cultural Heritage domain, where complex shapes and mouldings make this task challenging. In this paper, we describe our experiments with the DGCNN architecture to semantically segment historical buildings point clouds, acquired with TLS. We propose a variation of the original approach where a radius distance based technique is used instead of K-Nearest Neighbors (KNN) to represent the neighborhood of points. We show that our approach provides better results by evaluating it on two real TLS point clouds, representing two Italian historical buildings: the Ducal Palace in Urbino and the Palazzo Ferretti in Ancona.

1. INTRODUCTION

Built Cultural Heritage management and preservation requires the creation of accurate and rich digital representations of historical buildings. Laser scanning has become a widely used technique to obtain accurate digital representations of 3D scenes and is commonly adopted by architects, archaeologists and scholars. The obtained digital representations come in the form of point clouds, which, however, lack of semantic information and are often insufficient to conduct further analysis and studies. More informative and structured representations are needed, and are usually derived relying on knowledge representation standards like the Building Information Model (BIM), as done, for example in (Quattrini et al., 2017a) and (Quattrini et al., 2017b), where annotated BIM models drive a semantic aware user interface to meaningfully explore architectural heritage. The process of transforming a point cloud to BIM is referred to as Scan-to-BIM and is usually carried out manually through a careful and time-consuming work.

For this reason, the interest in finding viable ways to partially automate the Scan-to-BIM process is gaining growing interest within the research community. A central problem in this context is that of separating a point cloud into components that represent single classes of objects of interest. In the case of historical buildings, it is desirable to automatically segment the 3D point cloud into specific architecture elements, referring to robust and consolidated thesaurus. The interest in this kind of task is witnessed by a number of recent research efforts, which attempt at addressing it with algorithmic workflows (Murtiyoso, Grussenmeyer, 2020a) or leveraging machine learning methods based on hand-crafted features (Grilli et al., 2019b).

The task of classifying points in a point cloud according to some types of 3D objects is accounted as semantic segmentation and is of interest to a variety of research areas, as, for example, autonomous driving and robot scene interpretation. The recent advances in deep learning (DL), have driven the

research in this area and in the last years several deep neural architectures have been proposed that attempt at semantically segment 3D point clouds, directly operating on the coordinate representation of 3D scenes. Examples are PointNet (Qi et al., 2017a), the pioneer study in this area, its subsequent improvement, Pointnet++ (Qi et al., 2017b), the Dynamic Graph Convolutional Neural Network (DGCNN) architecture proposed in (Wang et al., 2019), which attempts at generalizing the previous approaches into a common formal framework, and PointCNN (Yu et al., 2019).

While such neural architectures have been extensively tested on standard benchmark datasets of indoor scenes (as the S3DIS dataset (Armeni et al., 2017)), few attempts were done to investigate their use in historical buildings TLS point clouds (Malinverni et al., 2019), (Pierdicca et al., 2020a). While classic ML approaches often require feature engineering on a case by case basis (Grilli et al., 2019b), DL approaches attempt at automatically extracting (hidden) feature, thus possibly providing a uniform method that can be applied to a variety of case studies.

In this paper we describe our experiments with the DGCNN architecture to semantically segment historical buildings TLS point clouds. We propose a variation of the original approach (Wang et al., 2019) where a radius distance based technique is used instead of K-Nearest Neighbors (KNN) to represent the neighborhood of points. We show that our approach provides better results by evaluating it on two real TLS point clouds, representing two Italian historical buildings: the Ducal Palace in Urbino and the Palazzo Ferretti in Ancona.

2. RELATED WORKS

In this section we provide an overview of the state of the art in deep learning approaches for semantic classification of dense point clouds. Then we restrict to the Cultural Heritage field and overview some of the Machine Learning (ML) efforts in achieving semantic segmentation of architectural elements.

* Corresponding author

DL algorithms have been used in several application domains to semantically segment 3D point clouds (Xie et al., 2020, Zhang et al., 2019). However, the cultural heritage environment is still lacking in studies that use this methodology in the processing of 3D point clouds. A significant advantage is that this large amount of raw data represented by the 3D point clouds can be managed directly by the DL algorithms without using intermediate processing to obtain a more regular structure before classification. In this regard, the first algorithm was presented in the work (Qi et al., 2017a) in which the points belonging to the cloud are considered individually. Subsequently, an advanced version of this approach was presented (Qi et al., 2017b), which also exploits the local information of the points considering the nearby points and obtaining better results in terms of classification. A first attempt to use PointNet++ architecture to semantically segment 3D point clouds of CH dataset is presented by Malinverni et al. (Malinverni et al., 2019). The work aims at demonstrating the effectiveness of the method even in an area that has not been yet explored, considering a set of CH data suitably created and annotated manually by the domain experts. The Point Clouds Convolutional Neural Network (PCNN) implemented in the work of (Atzmon et al., 2018) is a novel architecture that uses a Convolutional Neural Network (CNN) to elaborate 3D point clouds. It is based on two operators: extension and restriction, one the opposite of the other, to classify the 3D points cloud. Novel point-based approaches to semantic segmentation have been recently proposed, often extending the previous architectures proposed in this field. According to (Guo et al., 2020), they can be divided into Point-wise MLP architectures, e.g. (Jiang et al., 2018), named PointSIFT and inspired by the 2D shape descriptor SIFT, Convolution based, as PointConv (Boulch, 2020), and Graph Based.

Dynamic Graph Convolutional Neural Network (DGCNN) is one of the first approaches in the latter category. DGCNN (Wang et al., 2019) is based on the EdgeConv operation. EdgeConv is a model that creates edge features and describes the neighbourhood relation among points, without generating point features from their embeddings. This module is designed to be invariant to permutation and ordering of neighbours. DGNN architecture was also used in CH context in the work proposed by (Pierdicca et al., 2020b). The aim is to make a semantic segmentation of 3D Point Clouds using an augmented DGCNN model by adding features like normal and colour. The advantages is to better manage CH elements with complex geometries, structures extremely variables and defined with a high level of detail. They make also a comparison between other DL methods.

In (Grilli et al., 2019c), the authors, taking into account the benefits of using ML and DL technologies, performed a comparison of performance in the classification of two different CH datasets. They highlight that the use of ML approaches (Random Forest and One-versus-One) allows to obtain excellent performance in terms of classification, even if there is no correlation between the characteristics. Although ML techniques are less recent than DL techniques, there are a limited number of applications in the literature that use ML-based methods to semantically segment 3D point clouds in the CH domain. However, according to the study proposed by (Grilli, Remondino, 2019), these methods have made great progress in this direction. After exploring the applicability of supervised ML approaches to cultural heritage, the authors propose a standardized pipeline with reference to the different case studies. In this context, the work proposed by (Oses et al., 2014) has two main objectives: the first is to provide a framework that extracts geometric prim-

itives from a masonry image and second to make an extraction and a selection of statistical features for the automatic clustering of masonry. They make a combination between an image processing and ML methodologies for the classification of masonry walls and then compare the performances between five different ML algorithms in the classification task. In this approach the main problem is that each block of the wall is not separately characterized. To overcome this limitation the work of (Riveiro et al., 2016) presents a new automatic segmentation algorithm of masonry blocks departing from a 3D point cloud acquired by LiDAR technology. The image processing is based on an optimization of the watershed algorithm, that is used to improve segmentation algorithms in other works (Barsanti et al., 2017, Poux et al., 2017).

In their research Grilli et al., (Grilli et al., 2018) use the UV maps of 3D models of DCH goods as input of supervised ML classification algorithms. In order to verify the efficiency of this method, the authors make a comparison of several classifier considering three different case studies. The research of (Grilli et al., 2019a) is another application of a supervised ML classifier (RF) used to classify a 3D points cloud of CH field. The authors evaluate the relationship between covariance features and architectural elements, in particular determining a relation between the feature search radii and the size of the element. Successively, the work (Grilli, Remondino, 2020) makes an analysis of the previous approach demonstrating its ability to generalise between different never seen architectural scenes. The research conducted by (Murtiyoso, Grussenmeyer, 2020b) has the aim to help the manual point clouds labeling of large training data set required from ML algorithms. Moreover, the authors introduce a series of functions that allow the automatic processing of some problems of segmentation and classification of point clouds for CH goods. Due to the complexity of the problem, the project considers only some important classes but it is suitable for different types of heritage.

3. MATERIALS AND METHODS

3.1 Dataset

In the framework of the on going project CIVITAS (ChaIn for the excellence of reflective societies to exploit digital cultural heritage and museumS), the 3dimensional digitization of the Ducal Palace at Urbino is carried out mainly based on TLS and photographic data capturing.

At the present survey phase, the total of acquired points is 1.790 mln (Nespeca, 2018, Clini et al., 2020), the comprehensive model for this large-complex building serves to develop innovative researches and tests facing the challenges in CH field. Although the whole numerical model refers to the complete Palace, in the current experiment only the part of the point cloud related to the Courtyard of Honour was exploited (Table 1).

The point cloud of the Ducal Palace is a main reference data set, acquired with different sensors and technologies, the current work refers to point cloud carried out by two laser scanners (Leica ScanStation C10 and a Leica ScanStation P40) mounted on a tripod. This acquisition was performed by setting different levels of resolution according to the complexity of the rooms and their decorative elements (see Figure 1), optimizing scanning times. Considering that one of the challenges of the project was to validate workflows from Scan to segmented, informed and semantically reality-based models, significant parts of the point cloud were selected and compared. In

Ducal Palace at Urbino		
	Whole point cloud	Annotated point cloud (Courtyard of Honor)
Number of points	1224 mln	17.3 mln
max alignment error	0.006 m	0.008 m
min alignment error	0.001 m	0.003 m
RMS	0.005 m	0.004 m

Palazzo Ferretti		
	Whole point cloud	Annotated point cloud
Number of points	1.273 mln	8.4 mln
Number of points	1273 mln	3.8 mln
max alignment error	0.008 m	0.005 m
min alignment error	0.003 m	0.001 m
RMS	0.007 m	0.007 m

Table 1. Number of points, max/min alignment error and RMS of the two TLS point clouds considered in this study.



Figure 1. Two views of the point cloud of the Laurana's Courtyard of Honour: general perspective and detail, in which the more fine architectural features are readable

order to verify the effectiveness of the automatic classification, the Laurana's Courtyard of Honour was chose as testing area. Future work, on the side of data acquisition, is planned using different mapping systems (such as Kaarta) and a photogrammetric flight with UAVs in order to cover the inaccessible parts and roofs of the building. In fact, a secondary goal of the project was to test strategies, improving and comparing LODs and required resources/tools in the different phases of work. Data obtained by Terrestrial Laser Scanner (TLS) and by photogrammetry are currently often integrated in order to obtain models able to serve for different purposes, with detailed analysis of some portions and users interactions with elements of interest.

Another dataset selected for testing the neural network, was the Palazzo Ferretti's point cloud considering its architectural style and the presence of significant mouldings and architectural classes related to the Renaissance Age. In fact, in the Scan to BIM process the recognition of very refined details with complex characteristics is a major obstacle. The TLS survey operations of Palazzo Ferretti was carried out in 2014 for developing structural analysis. It consisted in 49 stations (58 scans) for the external survey of the building plus 20 stations and scans for

the inner survey, involving the ground and the first floor. The obtained point cloud after alignment and cleaning operations is made up of 1.2 billion points (see table 1). Finally, before feeding the data to the neural networks we reduced the point clouds, by randomly sampling points from surfaces with a density of 1000 points per m^2 , obtaining approximately 17.3 mln point points for the Ducal Palace courtyard and 8.4 mln points for Palazzo Ferretti.

3.2 DGCNN with radius-based EdgeConv

The DGCNN model, introduced in (Wang et al., 2019), is based on the EdgeConv operation, which is implemented with a Multilayer Perceptron (MLP) fed with the *edge features* of a point: the distance vectors between the point and its neighbour points. In this way the learned point features take into account its surrounding points, thus being able to learn different shapes and to map them to different types of objects.

In the original design (Wang et al., 2019), the edge features are calculated from the K-Nearest Neighbours (KNN) of a point. As suggested by recent studies (Hermosilla et al., 2018), this method might not be optimal when the point cloud is characterized by non uniform density, as generally happens for TLS. In Figure 3 we show a fragment of the point cloud of the Ducal Palace of Urbino, where two windows were captured with different density. Using KNN the point neighbourhood cover a small area when the local density is high and a smaller one when points are more scattered. This means that different edge features can be derived from two point that belong to the same class (window). Aiming at overcoming this issue, we propose a variation of the DGCNN architecture (that we call *RadDGCNN* in this paper), where points neighborhood is based on the radius distance. As the neighbour points cover the same area for all points, independently from the local density, more representative features could possibly be learned. In our solution, for each point, all the points at distance $D < R$ are selected, then K points are randomly sampled and used as input to the EdgeConv operation.

The deep learning architecture used in our experiments is depicted in Figure 2. The input point cloud is processed by three EdgeConv operations, extracting hierarchical local features of points. Then the output of each EdgeConv for each point is concatenated and global features are learnt by a MLP layer, to finally produce a category prediction for each point in input. The architecture differs from the baseline DGCNN architecture in the EdgeConv block, based on radius distance, and in the adoption of a pseudo-random rotation.

As the DGCNN architecture suffers from sensibility to the orientation of shapes to be learned, in our experiments we use rotation to augment the dataset. In particular, we rotate each block around the up direction only and we perform, for each point, a rotation of a multiple of 90 degrees and a random one. This is motivated by simple considerations on the domain, where these constraints clearly reflect domain rules (e.g. walls are often positioned at 90 degrees from each other, architectural elements have always the same orientation with respect to the up direction). In this way we aim at optimizing the training phase, letting the system learn from meaningful variations of the input data.

4. EXPERIMENTS AND RESULTS

In our experiments we evaluated the DGCNN and RadDGCNN approaches on two point clouds: Ducal Palace of Urbino (PDU)

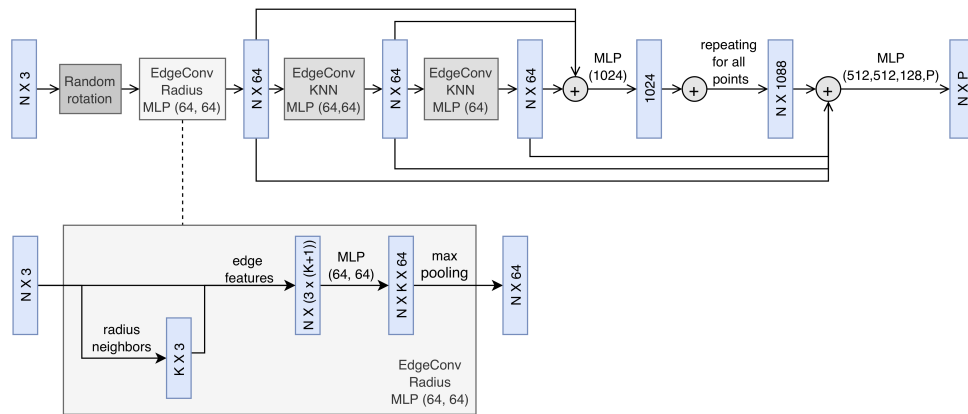


Figure 2. The architecture of the the RaddGCNN used in our experiments.

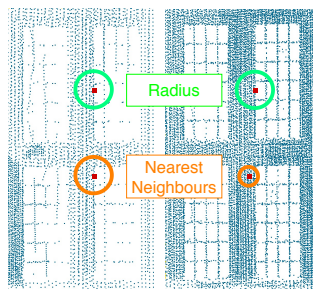


Figure 3. KNN vs. Radius-based neighborhood

and Palazzo Ferretti (PF), using half of the scene to train the network and the remaining for testing. In both cases we segmented the whole scene into blocks of dimension 1x1x1 meters, and for each block we sampled 4096 points. Such a scene segmentation step, usually applied in literature to experiment with DGCNN and other point based deep networks, is required because the MLP network needs a fixed number of points for each data example to be processed and, furthermore, processing the whole scene as a single data example would be too computationally expensive. The number of points for each class obtained after the segmentation and sampling for the two test sets used in our experiment is shown in Figure 4

The number of points representing neighbourhoods (K) was set to 20, as done in (Wang et al., 2019). In the case of Rad-DGCNN, the radius R was experimentally set to 0.1. Each point was represented with a 9 dimension vector encoding original XYZ coordinates, point color (expressed in HSV format) and normalized XYZ coordinates within each block. Regarding DGCNN network hyperparameter, we adopted the same setting used in (Wang et al., 2019) and trained the networks for 200 epochs.

Results are reported in 3 and 4, where the proposed architecture is compared with the DGCNN architecture adopted in (Wang et al., 2019). As shown, the use of radius-based EdgeConv provides better segmentation accuracy in almost all considered classes, leading to a overall accuracy increase of 1.8% in the case of PDU, and 1% in the for PF.

As an additional experiment, we tried to combined the two approaches using a simple multi-classifier architecture, where each point is processed by the two trained networks and the best output probability is selected (*Combined* column in the tables). In the PDU experiment, the combined model slightly

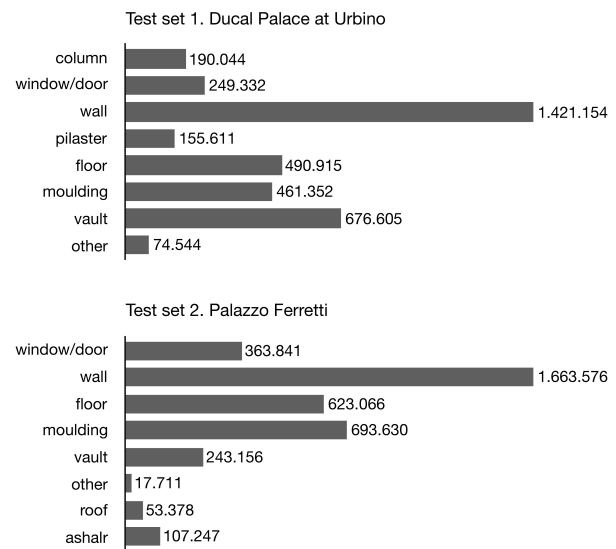


Figure 4. Number of points per class in the two test point clouds.

improves mean IoU and accuracy for some classes (e.g. pillar and floor), while is overcome by RaddGCNN with respect to other classes, e.g., *column*, thus leaving the overall accuracy almost unchanged. In the PF experiment, however, the improvement provided by the combined approach is larger in overall accuracy (+1.7%) as well as in mean IoU (+1.9%). This indicates that a simple combination of pre-trained models can be, in some cases, effective in boosting results.

In general, the results obtained are sensibly better in the PDU experiment (as also intuitively shown in Table 2). This is possibly due to the PDU scene being more symmetric if compared to the PF scene, thus resulting in more regular shapes to be learned. We can see that, as expected, more regular, repeatable or recognizable shapes, like wall and columns are better recognized. However, we think results are promising even in highly diverse shapes as those belonging to moulding, doors and windows.

In Figure 5 and 6 we report the confusion matrix of the classification results using the combined method on the PDU and PF scene, respectively. As one can see, in both the scenes, a number of classes (e.g. windows/doors, moulding and other) are often confused with walls. A possible reason for this is that the dataset is highly unbalanced and the wall class has lot more

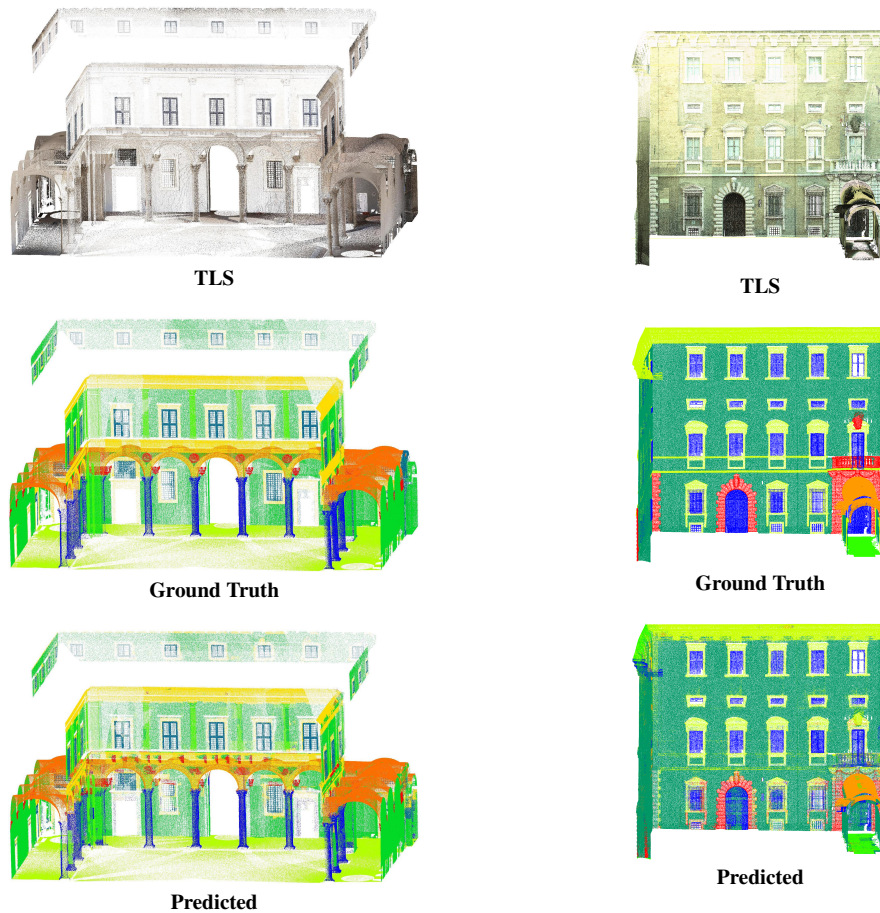


Table 2. Visualization of the result of the automatic segmentation of Ducal Palace at Urbino courtyard (left) and a view of the Palazzo Ferretti facade (right).

classes	DGCNN				RadDGCNN				Combined			
	precision	recall	f1-score	IoU	precision	recall	f1-score	IoU	precision	recall	f1-score	IoU
column	0.7454	0.6751	0.7085	0.5486	0.8424	0.8601	0.8511	0.7409	0.7766	0.7866	0.7816	0.6415
window/door	0.9154	0.6821	0.7817	0.6416	0.8912	0.7024	0.7856	0.6470	0.9267	0.7016	0.7986	0.6647
wall	0.8441	0.9207	0.8808	0.7869	0.8942	0.9084	0.9012	0.8202	0.8677	0.9239	0.8949	0.8098
pillar	0.8636	0.6649	0.7513	0.6017	0.7836	0.7094	0.7446	0.5931	0.8733	0.7001	0.7772	0.6356
floor	0.9580	0.9943	0.9758	0.9681	0.9117	0.9864	0.9476	0.9528	0.9679	0.9969	0.9822	0.9650
moulding	0.8258	0.7296	0.7747	0.6322	0.8339	0.8005	0.8169	0.6904	0.8577	0.7702	0.8116	0.6829
vault	0.9371	0.9640	0.9504	0.9054	0.9294	0.9744	0.9513	0.9072	0.9407	0.9722	0.9562	0.9160
other	0.6781	0.7742	0.7230	0.5661	0.6825	0.7568	0.7177	0.5597	0.7308	0.7946	0.7614	0.6147
micro avg			0.8726				0.8907				0.8894	
macro avg	0.8477	0.8008	0.8193	0.7063	0.8519	0.8383	0.8431	0.7389	0.8677	0.8308	0.8455	0.7413
weighted avg	0.8729	0.8726	0.8698		0.8898	0.8907	0.8892		0.8897	0.8894	0.8874	

Table 3. Results of the semantic segmentation of the Ducal Palace at Urbino scan.

classes	DGCNN				RadDGCNN				Combined			
	precision	recall	f1-score	IoU	precision	recall	f1-score	IoU	precision	recall	f1-score	IoU
ashlar	0.4712	0.2445	0.3220	0.1919	0.5557	0.3553	0.4335	0.2767	0.5772	0.3174	0.4096	0.2575
roof	0.9658	0.9833	0.9745	0.9502	0.9999	0.9725	0.9860	0.9724	0.9996	0.9812	0.9903	0.9808
window/door	0.6929	0.6564	0.6742	0.5085	0.8016	0.6201	0.6992	0.5376	0.7975	0.6811	0.7347	0.5806
wall	0.8231	0.8744	0.8480	0.7361	0.8353	0.8723	0.8534	0.7443	0.8377	0.8924	0.8642	0.7609
floor	0.9272	0.9655	0.9460	0.8975	0.9117	0.9864	0.9476	0.9004	0.9361	0.9932	0.9638	0.9302
moulding	0.8118	0.6520	0.7232	0.5664	0.7374	0.7153	0.7262	0.5701	0.7985	0.7073	0.7501	0.6002
vault	0.6420	0.8573	0.7342	0.5800	0.7103	0.7897	0.7479	0.5973	0.7255	0.8501	0.7829	0.6432
other	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
micro avg			0.8058				0.8152				0.8325	
macro avg	0.6667	0.6542	0.6527	0.5538	0.6940	0.6639	0.6742	0.5748	0.7090	0.6778	0.6870	0.5942
weighted avg	0.8021	0.8058	0.7999		0.8090	0.8152	0.8097		0.8266	0.8325	0.8267	

Table 4. Results of the semantic segmentation of the Palazzo Ferretti scan.

a) Palazzo Ducale

classes	DGCNN				RadDGCNN			
	precision	recall	f1-score	IoU	precision	recall	f1-score	IoU
accuracy			0.8769				0.8893	
macro avg	0.8423	0.8084	0.8221	0.7101	0.8433	0.8366	0.8377	0.7321

b) Palazzo Ferretti

classes	DGCNN				RadDGCNN			
	precision	recall	f1-score	IoU	precision	recall	f1-score	IoU
micro avg			0.8040				0.8338	
macro avg	0.6644	0.6404	0.6374	0.5493	0.7890	0.6928	0.7145	0.6086

Table 5. Example of results with different values of K and R (K=40, R=0.2), for the two scenes (Ducal Palace, a), and Palazzo Ferretti, b))

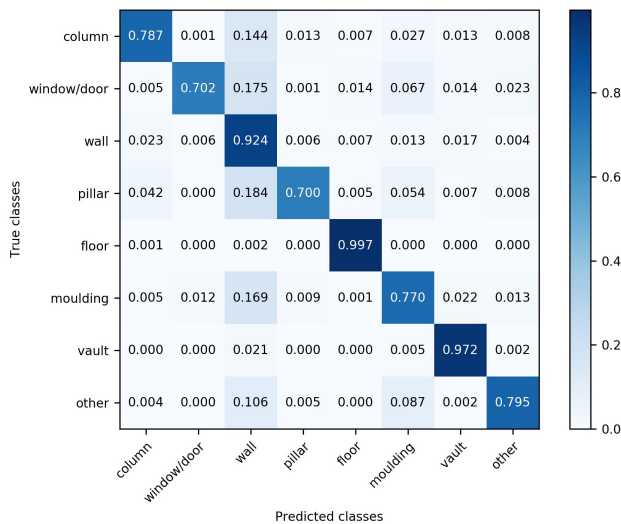


Figure 5. Confusion matrix of the combined approach on the PDU scene.

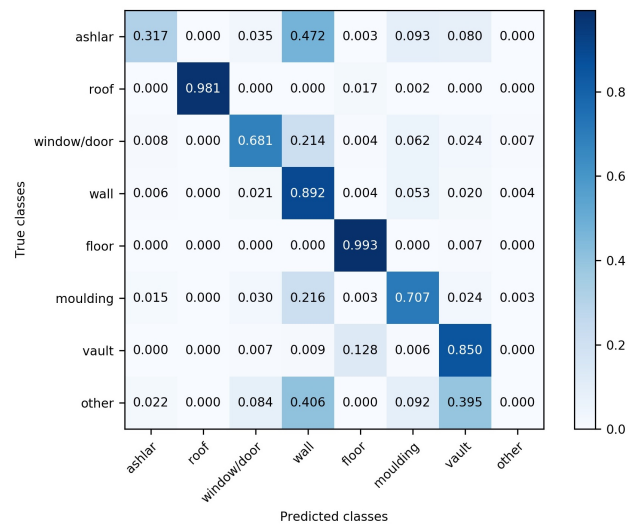


Figure 6. Confusion matrix of the combined approach on the PF scene.

points if compared to other classes. In the PDU test dataset, there are around 1.4 million points belonging to the class *wall*, while only 460 k points for *mouldings* and 74 k for the class *other*. In the case of the PF scene, we can observe that the classifier fails in detecting ashlar and often classify it as wall. This is somehow expected, as ashlar are in fact similar to walls, thus more examples of the class would be probably needed to properly train the network. We also note that the *other* class, which simply collect elements that are not classified, is, as expected, hard to recognize (especially in the PF scene), as it lacks of distinct characteristics and a sufficient number of data examples.

Finally, we would like to point out that, beside networks regular hyper-parameters tuning, the choice of neighbourhood to be considered in EdgeConv operation, might lead to different results. In the case of DGCNN, as the learning is driven by KNN, the value of K can be changed to spatially enlarge or restrict the neighbourhood used to extract hidden features from points. In the case of Radius distance, one can leverage a different parameter, R , which allows to choose a specific neighbourhood radius. While investigating the optimal choices based on the considered point clouds is out of the scope of this study, we show as an example, in Table 5, the results obtained with a different choice of parameters, where the point neighborhoods have been extended to consider more surrounding points. Specifically, we set K to 40 for both DGCNN and RadDGCNN, and R to 0.2 for RadDGCNN. While in the case of the PDU, where results were already better, there is no significant difference, on the PF scene the accuracy of our method increases (approximately by

2%). An interesting direction to investigate is that of understanding how to optimally tune such parameters, e.g. driven by the peculiarities of the specific data at hand.

5. CONCLUSIONS AND FUTURE WORKS

In this paper we focused on evaluation of DGCNN based techniques in the historical building domain. We therefore evaluated our method under a simple and directly measurable condition: we learn from a portion of a scene and we predict the remaining part. While this setting allows us to draw preliminary conclusions regarding the ability of semantically segmenting architectural objects, this is surely a simplified situation.

Even if the setting can be of a certain interest in practice, as automatically process a partially annotated scene can alone sensibly reduce annotation time consumption, more generalization has to be achieved by predicting totally unknown scenes. However, this requires enough annotated point cloud data to properly train the network, which can be not trivial to obtain. An interesting research direction is that of using synthetic data to form a critical mass of annotated point clouds to learn from. This idea was recently explored by our research group in (Morbidoni et al., 2020).

In conclusion, we think that results provided by our experiments are promising and motivate further studies in this direction. Future works include to deeply investigate the application of more solid domain grounded constraints and data processing techniques, as well as the possibility to build such knowledge into

the deep learning models. An other aspect to be addressed is that of understanding the importance of single features. For example, in the present experiments we used HSV encoded color. Further experiments are planned to quantify its effective contribution to the the classification results, and to measure performances in situations where color is not available.

ACKNOWLEDGEMENTS

This research is partially funded by the CIVITAS (ChaIn for excellence of reflective societies to exploit digital cultural heritage and museumS) project (Clini et al., 2020). The authors would like to thank Ordnance Survey GB (www.ordnancesurvey.co.uk) and ISpatial (www.ispatial.com/) for sponsoring the publication of this paper

REFERENCES

- Armeni, I., Sax, A., Zamir, A. R., Savarese, S., 2017. Joint 2D-3D-Semantic Data for Indoor Scene Understanding. *ArXiv e-prints*.
- Atzmon, M., Maron, H., Lipman, Y., 2018. Point convolutional neural networks by extension operators. *arXiv preprint arXiv:1803.10091*.
- Barsanti, S. G., Guidi, G., De Luca, L., 2017. Segmentation of 3D Models for Cultural Heritage Structural Analysis—Some Critical Issues. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, 115.
- Boulch, A., 2020. ConvPoint: Continuous convolutions for point cloud processing. *Computers and Graphics*, 88, 24-34.
- Clini, P., Quattrini, R., Bonvini, P., Nespeca, R., Angeloni, R., Mammoli, R., Dragoni, A. F., Morbidoni, C., Sernani, P., Mengoni, M., Leopardi, A., Silvestrini, M., Gambelli, D., Cori, E., Gallegati, M., Tamperi, M., Fraticelli, F., Acciarri, M. C., Mandolesi, S., 2020. Digit(al)isation in Museums: Civitas Project – AR, VR, Multisensorial and Multiuser Experiences at the Urbino's Ducal Palace. *Virtual and Augmented Reality in Education, Art, and Museumspp*, 194-228.
- Grilli, E., Dinunno, D., Marsicano, L., Petrucci, G., Remondino, F., 2018. Supervised segmentation of 3d cultural heritage. *2018 3rd Digital Heritage International Congress, DigitalHERITAGE, held jointly with 2018 24th International Conference on Virtual Systems & Multimedia, VSM 2018*, IEEE, 1–8.
- Grilli, E., Farella, E., Torresani, A., Remondino, F., 2019a. Geometric features analysis for the classification of cultural heritage point clouds. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.
- Grilli, E., Özdemir, E., Remondino, F., 2019b. Application of machine and deep learning strategies for the classification of heritage point clouds. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(4/W18), 447-454.
- Grilli, E., Özdemir, E., Remondino, F., 2019c. Application of machine and deep learning strategies for the classification of heritage point clouds. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 447–454.
- Grilli, E., Remondino, F., 2019. Classification of 3D Digital Heritage. *Remote Sensing*, 11(7), 847.
- Grilli, E., Remondino, F., 2020. Machine Learning Generalisation across Different 3D Architectural Heritage. *ISPRS International Journal of Geo-Information*, 9(6), 379.
- Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M., 2020. Deep Learning for 3D Point Clouds: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-1.
- Hermosilla, P., Ritschel, T., Vázquez, P.-P., Vinacua, A., Ropinski, T., 2018. Monte Carlo convolution for learning on non-uniformly sampled point clouds. *ACM Transactions on Graphics*, 37(6).
- Jiang, M., Wu, Y., Zhao, T., Zhao, Z., Lu, C., 2018. Pointsift: A sift-like network module for 3d point cloud semantic segmentation.
- Malinverni, E., Pierdicca, R., Paolanti, M., Martini, M., Morbidoni, C., Matrone, F., Lingua, A., 2019. DEEP LEARNING FOR SEMANTIC SEGMENTATION OF 3D POINT CLOUD. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.
- Morbidoni, C., Pierdicca, R., Paolanti, M., Quattrini, R., Mammoli, R., 2020. Learning from Synthetic Point Cloud data for historical buildings semantic segmentation. *ACM Journal on Computing in cultural Heritage*. In press.
- Murtiyoso, A., Grussenmeyer, P., 2020a. Virtual disassembling of historical edifices: Experiments and assessments of an automatic approach for classifying multi-scalar point clouds into architectural elements. *Sensors*, 20(8).
- Murtiyoso, A., Grussenmeyer, P., 2020b. 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.
- Nespeca, R., 2018. Towards a 3D digital model for management and fruition of Ducal Palace at Urbino. An integrated survey with mobile mapping. *SCIRES-IT*, 8(2), 1-14.
- Oses, N., Dornaika, F., Moujahid, A., 2014. Image-based delineation and classification of built heritage masonry. *Remote Sensing*, 6(3), 1863–1889.
- Pierdicca, R., Paolanti, M., Matrone, F., Martini, M., Morbidoni, C., Malinverni, E., Frontoni, E., Lingua, A., 2020a. Point cloud semantic segmentation using a deep learning framework for cultural heritage. *Remote Sensing*, 12(6).
- Pierdicca, R., Paolanti, M., Matrone, F., Martini, M., Morbidoni, C., Malinverni, E. S., Frontoni, E., Lingua, A. M., 2020b. Point Cloud Semantic Segmentation Using a Deep Learning Framework for Cultural Heritage. *Remote Sensing*, 12(6), 1005.
- Poux, F., Neuville, R., Hallot, P., Billen, R., 2017. Point cloud classification of tesserae from terrestrial laser data combined with dense image matching for archaeological information extraction. *International Journal on Advances in Life Sciences*, 4, 203–211.

- Qi, C. R., Su, H., Mo, K., Guibas, L. J., 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 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. *Advances in neural information processing systems*, 5099–5108.
- Quattrini, R., Pierdicca, R., Morbidoni, C., 2017a. Knowledge-based data enrichment for HBIM: Exploring high-quality models using the semantic-web. *Journal of Cultural Heritage*, 28, 129–139.
- Quattrini, R., Pierdicca, R., Morbidoni, C., Malinverni, E. S., 2017b. CONSERVATION-ORIENTED HBIM. THE BIMEXPLORER WEB TOOL. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42.
- Riveiro, B., Lourenço, P. B., Oliveira, D. V., González-Jorge, H., Arias, P., 2016. Automatic morphologic analysis of quasi-periodic masonry walls from LiDAR. *Computer-Aided Civil and Infrastructure Engineering*, 31(4), 305–319.
- Wang, Y., Sun, Y., Liu, Z., Sarma, S., Bronstein, M., Solomon, J., 2019. Dynamic graph Cnn for learning on point clouds. *ACM Transactions on Graphics*, 38(5).
- Xie, Y., Tian, J., Zhu, X., 2020. A Review of Point Cloud Semantic Segmentation. *IEEE Geoscience and Remote Sensing Magazine*.
- Yu, Y., Li, F., Zheng, Y., Han, M., Le, X., 2019. Clustering-enhanced pointcnn for point cloud classification learning. 2019-July.
- Zhang, J., Zhao, X., Chen, Z., Lu, Z., 2019. A Review of Deep Learning-based Semantic Segmentation for Point Cloud. *IEEE Access*.