Proceedings of the joint international event 9th ARQUEOLÓGICA 2.0 & 3rd GEORES, Valencia (Spain). 26–28 April 2021

Received: 27/11/2020
Accepted: 25/03/2021
DOI: https://doi.org/10.4995/Arqueologica9.2021.12101

GENERATIVE NETWORKS FOR POINT CLOUD GENERATION IN CULTURAL HERITAGE

Roberto Pierdicca[a,*], Marina Paolanti[b], Ramona Quattrini[a], Massimo Martini[b], Eva Savina Malinverni[a], Emanuele Frontoni[b]

[a] Dipartimento di Ingegneria Civile, Edile e dell'Architettura, Università Politecnica delle Marche, Via Brecce Bianche 1, 60131 Ancona, Italy. r.pierdicca@univpm.it; r.quattrini@univpm.it; e.s.malinverni@univpm.it

[b] Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, Via Brecce Bianche 1, 60131 Ancona, Italy. m.paolanti@univpm.it; m.martini@pm.univpm.it; e.frontoni@univpm.it

Abstract:

In the Cultural Heritage (CH) domain, the semantic segmentation of 3D point clouds with Deep Learning (DL) techniques allows to recognize historical architectural elements, at a suitable level of detail, and hence expedite the process of modelling historical buildings for the development of BIM models from survey data. However, it is more difficult to collect a balanced dataset of labelled architectural elements for training a network. In fact, the CH objects are unique, and it is challenging for the network to recognize this kind of data. In recent years, Generative Networks have proven to be proper for generating new data. Starting from such premises, in this paper Generative Networks have been used for augmenting a CH dataset. In particular, the performances of three state-of-art Generative Networks such as PointGrow, PointFLow and PointGMM have been compared in terms of Jensen-Shannon Divergence (JSD), the Minimum Matching Distance-Chamfer Distance (MMD-CD) and the Minimum Matching Distance-Earth Mover's Distance (MMD-EMD). The objects generated have been used for augmenting two classes of ArCH dataset, which are columns and windows. Then a DGCNN-Mod network was trained and tested for the semantic segmentation task, comparing the performance in the case of the ArCH dataset without and with augmentation.

Keywords: virtual archaeology, digital archaeology, cyber-archaeology, cultural heritage, documentation, 3D reconstruction

1. Introduction

The restoration, conservation, valorisation and cataloguing of architectural, archaeological and Cultural Heritage (CH) goods necessitate as preliminary operation the survey that allows analysing several characteristics of the object of study. With the survey, it is possible to identify, analyze and highlight the shape, structure, single elements, composition, relationships between the parts and the whole, material, state of conservation and any critical situations, historical evidence, temporal and spatial evolution.

The methodological operation of the traditional survey consists of a series of actions for which the final purpose is the reproduction of the object studied using a graphic representation. This representation must respond to certain aims, depending on the purpose of the survey, since there are several areas of investigation for an object as a historical or artistic document.

Traditionally, due to time and costs, to obtain a result suitable for the purposes of the survey, a preliminary phase was anticipated to the measurement phase in which it was decided which data to acquire, thus making a selection on which all the final results would have depended.

The significant convenience introduced by the detection methodology using new high-resolution technologies is that of having principles of objectivity and identifying a procedure that allows the acquisition of all possible data of the existing elements, considering successively the subjective phase.

The articulated shapes, often characterized by curved, concave and convex elements, rotated parts, whose position is often determined by innumerable alterations, as well as by the loss of finishes, have suggested the use of the laser scanner technology, to overcome the difficulties of a traditional survey with all the uncertainties and imprecisions deriving from the morphological conformation.

Lasers scanner allow acquiring information about the shape of an object through the points cloud which returns images from arbitrary points of view and in arbitrary light conditions (Blais, 2004).

* Corresponding Author: Roberto Pierdicca, r.pierdicca@univpm.it

The use of 3D point clouds in the field of CH have demonstrated considerable advantages, since the representation of CH artifacts through 3D data allows to perform several tasks. In fact, their rich informative virtual representation makes morphological analysis, map degradation and data enrichment. So, the management of cultural heritage goods is important to understand this data and to develop the available conservation strategy (Pierdicca et al., 2020).

Traditional methods are used in literature to deal with 3D point clouds using manual and time-consuming operations (Murtiyoso & Grussenmeyer, 2019; Grilli, Özdemir & Remondino, 2019; Spina, Debattista, Bugeja, & Chalmers 2011). However, most recent applications based on deep learning (DL) approaches have demonstrated high performance and more efficient ways to elaborate this data (Qi, Su, Mo, & Guibas 2017; Qi, Yi, Su, & Guibas 2017). Some examples are also in CH field where deep learning algorithms are used to semantically segment 3D point clouds to recognize historical architectural elements at an appropriate detail level (Pierdicca et al., 2020). The semantic segmentation task that classifies different parts associating to each part a label, is an approach most used in the cultural heritage field (Barsanti, Guidi & De Luca, 2017; Llamas, Lerones, Medina, Zalama, & Gómez-García-Bermejo, 2017; Grilli & Remondino, 2019).

The problem is that the training of deep learning networks requires a great amount of data mainly considering CH dataset with unique and unrepeatable elements. Moreover, the dataset has unbalanced classes and then the networks are not able to correctly recognize the objects (e.g. wall, roof, column, vault, etc.) due to the lack of data (Pierdicca et al., 2020).

For this reason, mainly in the last years, the use of Generative Networks have filled this gap since they are able to generate novel suitable data as well as learn (Arshad & Beksi, 2020).

In this context, the aim of this paper is to propose a framework based on DL that synthetically generates additional architectural elements to increase segmentation accuracy. To generate novel scenes, we use three different generative networks: PointGrow (Sun, Wang, Liu, Siegel, & Sarma, 2020), PointFlow (Yang et al., 2019), and PointGMM (Hertz, Hanocka, Giryes, & Cohen-Or, 2020). Moreover, to compare the best performances we train a novel Deep Neural Network, namely DGCNN-Mod that classifies the synthetically generate scenes (Pierdicca et al., 2020). The experiments have been performed to ArCH dataset described in (Matrone et al., 2020a). In contrast to many existing datasets, it has been manually labelled by domain experts, thus providing a more precise dataset.

The main contributions of this paper can be summarized as follows:

1) Generation of CH point clouds, useful for 3D documentation of documents and sites.
2) A generative based approach adapted to this domain.
3) A new set of synthetically generated architectural elements that is useful for data augmentation.

The paper is organized as follows. Section 2 provides a description of the approaches that were adopted for Point Clouds generations. Section 3 describes our approach to present the Point Clouds generation and a new challenging dataset for the CH domain. Section 4 offers an extensive comparative evaluation and a detailed analysis of our approach. Finally, Section 5 draws conclusions and discuss future directions for this field of research.

2. Related works

Synthetic data are used in several application fields. One of the pioneer works is proposed by Myers (1999) that demonstrates that the use of simulated data has significant advantages. This work introduces a simulator that describes and stochastically generates a sequence of DNA with different repeated structures.

Most recent work is proposed by that Fan, Su & Guibas (2017) that produces synthetic point clouds departing from a single 2D-dimensional image and reconstructing the 3D geometry of the complete object. To do this task they use an algorithm based on deep learning.

A 3D point cloud is automatically generated in the work of Li et al. (2016) and used to train a traditional random forest network. This approach is based on supervised learning to classify 3D real urban scenes.

The work of (Wu, Zhou, Zhao, Yue, & Keutzer, 2019) extracts synthetic point clouds of urban scenes (road scene) from the Grand Theft Auto famous videogame to augment a standard benchmark dataset (KITTI). This approach has the aim to increase the semantic segmentation task based on Convolutional Neural Networks.

Street scenes are also dealt in the work of (Wang, Zhuang, Gu, & Hu, 2019) where virtual LiDAR sensors are used to acquire synthetic point clouds. The work simulates several point clouds acquisition tools.

The urban context is also the object of the work proposed by Griffiths & Boehm (2019), where there is a need of a great amount of data to train a deep neural network to classify a 3D point cloud. The authors create a synthetic dataset and intend to demonstrate that the network trained with the dataset is able to generalize correctly the points cloud.

In the work of Chu, Sung, & Cho (2018) the authors propose a GAN to generate 3D points cloud departing from RGB-D images and corresponding to a single red, green, blue image. The method involves two phases. In the first phase, a generative adversarial network generates a depth image estimation from a single RGB image. In the second the 3D point cloud is calculated from the depth image. During the experimental phase, they demonstrate that the method provides high-quality 3D point clouds from single 2D images.

The work of Arshad & Beksi (2020) proposes a novel GAN that creates in an unsupervised way dense coloured 3D point clouds of various classes of objects. They propose a point transformer that using a graph convolution to increase in progress of the network, this to overcome the problem to acquire complex details with high resolution. The aim of the paper is to create a network able to produce coloured point clouds with fine details at multiple resolutions.

In the context of cultural heritage interesting is the work of Egiazarian et al. (2019) where they propose a model that combines the latent-space GAN and Laplacian GAN architectures to form a multi-scale model capable of generating 3D point clouds at augmenting levels of detail. During the experimental phase they demonstrate that the method is better than other object of comparison.

The work of Martínek, Lenc, & Král (2019) shows different training approaches with the task to classify optical character in historical documents. To solve the problem to have a great amount of annotated data, they summarize several methods to create a synthetic dataset. Moreover, they train a convolutional recurrent neural network using input a synthetic dataset and validate the approach using a real annotated dataset.

3. Materials and Methods

This section introduces the framework as well as the dataset used for the evaluation. The framework, as said in the introduction section, is depicted in Figure 1. The purpose of the proposed method is to improve the semantic segmentation of point clouds of an unbalanced dataset, generating new objects through generative approaches. Three generative networks have been trained for point cloud generation in CH Domain.

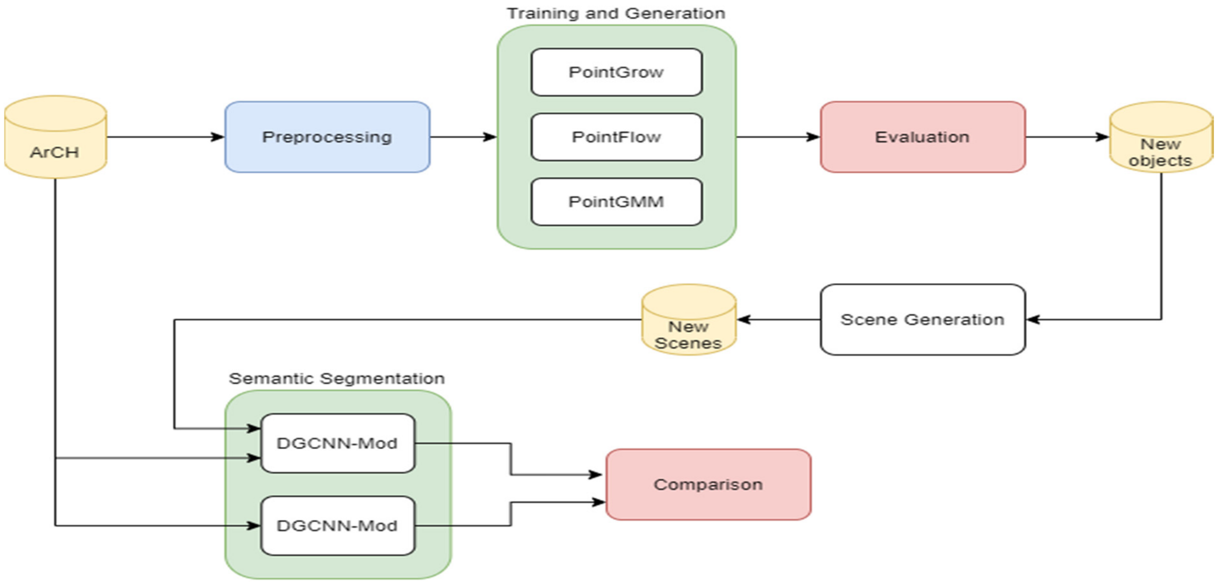


Figure 1: Workflow of the proposed approach.

An initial pre-processing phase is also introduced, to give in input proper objects to the networks, and a final phase for the comparison of two different pieces of training for the semantic segmentation of point clouds.

Further details are given in the following subsections. The framework is comprehensively evaluated on ArCH, a publicly available dataset.

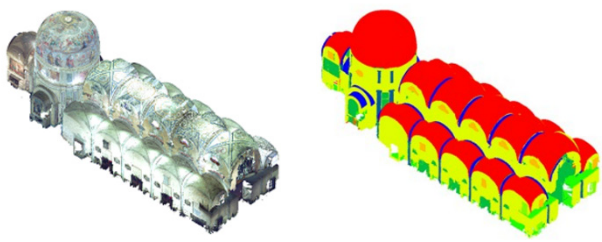
3.1. ArCH Dataset

The experiment section are based on the scenes of the ArCH benchmark (Matrone et al. 2020b), a group of architectural point clouds collected by several research groups and universities, with the purpose of sharing and labelling a large dataset of point clouds for training and testing learning approaches.

In literature, there are several benchmarks and datasets for point clouds classification and semantic segmentation, but the ArCH dataset is the only one specifically focused on the CH domain and with a higher level of detail.

The dataset is composed of 17 point clouds of architectural scenarios, divided into 15 scene for the training and 2 scenes for the final test.

This work is based on the tests carried out in (Matrone et al., 2020a), specifically the experiment of the Trompone's scene is reproduced using only the coordinates as a feature. Figure 2 shows this particular scene, representing original features and the relative ground truth.



(a) (b)

Figure 2: Trompone's scene from ArCH dataset: a) shows the scene features; b) shows the ground truth.

The objects of this dataset are given as input to generative approaches, in order to generate new objects and counterbalance the dataset.

However, this dataset does not provide labels of the instances but only for the classes. Then, every single object has been manually divided. ArCH dataset provides the ground truth for 10 classes, but only the classes that are more difficult to recognize has been selected, as indicated in (Matrone et al., 2020a). For this reason, only the objects regarding Column and Window classes have extrapolated from all scenes.

The dataset used to train generative approaches comprises:

- 234 columns;
- 390 windows.

In the last phase of the workflow, ArCH dataset is used to compare two different trainings: the first one using the original dataset and the last one using the increased dataset through the generated objects.

3.2. Generative Approaches for Point Cloud Generation

This subsection describes the generative approaches used to generate new ArCH dataset objects and counterbalance those classes that have only few instances.

Before the training of the networks, a pre-processing phase is performed on the data, consisting of 2 steps: subsampling and data normalization.

These networks can only take in input objects of 1024 points, so a random subsampling has been done for each instance of the dataset. Then, the objects are spatially centred at the (0,0,0) point and normalized to obtain values in the range of 0 and 1.

The pre-processed data are then used to train three different generative networks: PointGrow (Sun, Wang, Liu, Siegel, & Sarma, 2020), PointFlow (Yang et al., 2019), and PointGMM (Hertz, Hanocka, Giryes, & Cohen-Or, 2020).

These networks have been chosen because they are very recent state of the art approaches and have very good performance in point clouds generation, which are then used to improve related tasks such as classification and segmentation.

The first tested approach is PointGrow (Sun, Wang, Liu, Siegel, & Sarma, 2020), an autoregressive method for generating recurrently every point. This network estimates a conditional distribution of the point by considering all its preceding points. Taking into account the irregularity of the point clouds, the authors of this paper propose two point cloud-based self-attention modules for dynamically aggregating long-range dependencies from the other points. There are several ways to train these networks: in this work, the Unconditional PointGrow approach is used. To facilitate the generation process, training points are sorted according to their z coordinates; in this way, the shape should be encouraged to be generated mainly along its primary axis during the test phase.

The second network of our framework is called PointFlow (Yang et al., 2019). It is a variational auto-encoders (VAE) based approach, it is composed by three modules:

- An Encoder, that encodes a point cloud into a shape representation z;
- A Prior Module P(z) over shape representations z;
- A Decoder P(X|z) that models the distribution of points given the shape representation.

This particular network learns a two-level hierarchy of distributions: the first one is the distribution of shapes and the second one is the distribution of points given a shape. A continuous normalizing flow is used to learn these particular levels of distributions.

The third network, PointGMM (Hertz, Hanocka, Giryes, & Cohen-Or, 2020), is also composed of encoders and decoders: the first receives point clouds as input and generates a features map as an output, the second build a GMM representation from the previous latent vector. GMM is a Gaussian mixture model, usually used as an alternative representation for 3D objects. This approach learns a hierarchical GMM (hGMM), performing coarse-to-fine learning to improve performance, instead of a common single scale GMM.

All three approaches take the input data from the preprocessing phase, generate new objects, and then are compared according to appropriate metrics

3.3. Performance Evaluation

To evaluate the performance of the generative networks the following metrics have been adopted: Minimum Matching Distance based on Chamfer Distance (MMD-CD), Minimum Matching Distance based on Earth Mover's Distance (MMD-EMD) and Jensen-Shannon Divergence (JSD) (Achlioptas, Diamanti, Mitliagkas, & Guibas, 2018).

First of all, Chamfer Distance (CD) and the Earth Mover's Distance (EMD) should be introduced: they are two symmetric distance metrics to measure the distance between two points clouds. In our case, the two points clouds are the original and the generated ones.

Given two points clouds, S1 and S2, CD metric measures the squared distance between each point in S1 to its nearest neighbour in S2. So, the Chamfer Distance between S1 and S2 is defined in Eqs. (1) and (2):

$$d_{CD}(S_1, S_2) = \sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2^2 + \sum_{y \in S_2} \min_{x \in S_1} \|x - y\|_2^2 \quad (1)$$

Instead, the EMD is defined as:

$$d_{EMD}(S_1, S_2) = \min_{\varphi: S_1 \to S_2} \sum_{x \in S_1} \|x - \varphi(x)\|_2 \quad (2)$$

where

φ = bijection between S1 and S2

Then, a method to measure the similarity of A (set of generated point clouds) with respect to B (set of original point clouds) is needed. To this end, every point in B is matched with the closer one in A, by using minimum distance (MMD) and reporting the average of distances in the matching. CD and EMD metrics can be used as pointset distance for MMD, yielding MMD-CD and MMD-EMD (Achlioptas, Diamanti, Mitliagkas, & Guibas, 2018).

The last comparison metric is JSD, which measures if point clouds of A tend to occupy similar spaces as those of B, and the degree of this similarity. Given the empirical distributions (P_A, P_B), JSD metric is described as follow in Eq. (3):

$$JSD(P_A \,||\, P_B) = \frac{1}{2} D(P_A \,||\, M) + \frac{1}{2} D(P_B \,||\, M) \quad (3)$$

where

$M = \frac{1}{2}(P_A + P_B)$ and $D(||)$ the KL-divergence between the two distributions

Another state of the art method to evaluate the quality of the generated points clouds is to use PointNet (Qi, Su, Mo, & Guibas 2017) as a shape classifier (Sun, Wang, Liu, Siegel, & Sarma, 2020). In fact, if the generated objects will have very discriminating features for their relative class, it means that a classification model trained on the original objects should have good performance in classifying those generated, and vice-versa.

In this work, after generating 100 points clouds per category, two classification activities are conducted:

1) Training on original data and testing on generated shapes;
2) Training on generated shapes and testing on original data.

4. Results and Discussions

This section describes the configuration for network training and discusses the results of the various experiments.

The three generative networks were trained using Window and Column objects from the ArCH dataset. As described in Chapter 3.2, the objects were sub-sampled at 1024 points, then centred with respect to the point (0.0.0) and normalized with values between 0 and 1.

The training was done by splitting the dataset as follows:

- 80% of the objects for training;
- 20% of the objects for the final test.

A part of the training was used as the validation set.

The networks have been trained using the hyperparameters configuration described in the related papers.

The first experiment performed concerns the training of generative networks. Table 1 shows the comparison of the approaches in terms of JSD, MMD-CD and MMC-EMD metrics. The most satisfactory results are those obtained with PointFlow, which is the best approach for the generation of columns object, by obtaining lower JSD. Mand both MMD values. However, regarding windows objects, the two best methods for their generations seem to be PointGrow and PointGMM.

Table 1: Results of the generative approaches.

Class	*Model*	*JSD*	*MMD-CD*	*MMD-EMD*
	PointGrow	0.1941	0.0090	0.1352
Column	PointFlow	**0.0820**	**0.0078**	**0.1228**
	PointGMM	0.0929	0.0080	0.1421
	PointGrow	0.2588	**0.0061**	**0.1030**
Window	PointFlow	0.3014	0.0292	0.1613
	PointGMM	**0.1704**	0.0079	0.1052

The second experiment regards the classification performed with PointNet. The results are shown in Table 2: even in this case, PointFlow approach turns out to be the best generative method, as the classification accuracy for both SG and GS tests remain congruent.

Table 2: Classification accuracy using PointNet. SG: Training on ArCH Dataset and testing on generated shapes; GS: Training on generated shapes and testing on ArCH Dataset.

Model	*SG*	*GS*
PointGrow	0.5733	0.9350
PointFlow	0.7143	0.7150
PointGMM	0.6900	0.9200

From the results obtained, the PointFlow network has been chosen to generate column objects, while the PointGrow network has been used to generate Window objects.

Using CloudCompare software, it was possible to build 3 scenes containing these objects. To make the scene more realistic, portions of the wall were also added, so that the objects could be positioned consistently. Figure 3 shows the scenes created with the generated objects.

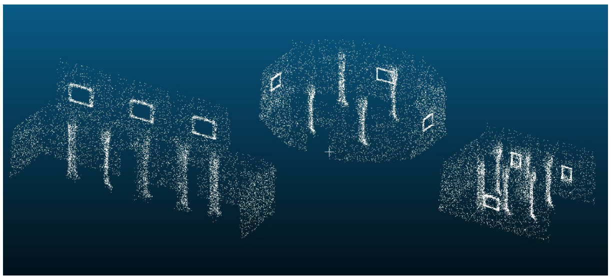


Figure 3: Scenes created by using generated objects.

Finally, these scenes are used to improve the training of the DGCNN-Mod network, for the semantic segmentation task on ArCH dataset. As said previously, experiments are conducted on the Trompone's scene.

Two different experiments are performed: the first regards the training by using original dataset, while the last regards the training with the ArCh dataset augmented by the generated scenes.

Table 3 and Table 4 show the results in terms of Precision, Recall and F1-Score, for every class of the Arch dataset. Performance of the DGCNN-Mod semantic segmentation on ArCH Dataset without the generated scenes are represented in Table 3. On the other hand, the performance of the DGCNN-Mod trained with the augmented dataset is shown in Table.

Table 3: Results of the DGCNN-Mod semantic segmentation on ArCH Dataset without the generated scenes.

Object	*Precision*	*Recall*	*F1-Score*
Arc	0.2904	0.1268	0.1765
Column	**0.9906**	**0.0856**	**0.1576**
Molding	0.5074	0.1465	0.2273
Floor	0.9485	0.9051	0.9263
Door-Win	**0.0974**	**0.0467**	**0.0631**
Wall	0.5867	0.2339	0.3344
Stairs	0.0424	0.0006	0.0011
Vault	0.9177	0.7408	0.8198
Forniture	0.3484	0.9241	0.5060

These experiments show that semantic segmentation performance regarding column class are improved. However, window segmentation has decreased. This is probably due to the fact that in the ArCH dataset windows are included in the same class as doors, because they are the classes with fewer elements. By increasing the number of windows, the network probably has more difficulty in classifying door points.

These findings deserve a more detailed explanation. First, looking at the column class, it is worth noting that despite the precision value is higher "without" than "with" the generated scenes, F1-score and Recall increase in this latter case (see the comparison between Tables 3 and 4); such values provide comforting insights from our experiments, since they are meaningful and opens to future generalizations of the method. This is not true for the class Door-Win, but the reason shall is that this class is merged among two objects (namely doors and windows), while in the benchmark dataset such classes were considered separately.

Table 4: Results of the DGCNN-Mod semantic segmentation on ArCH Dataset with the generated scenes.

Object	*Precision*	*Recall*	*F1-Score*
Arc	0.2284	0.4445	0.3018
Column	**0.6692**	**0.2131**	**0.3232**
Molding	0.4460	0.2488	0.3194
Floor	0.9159	0.9808	0.9473
Door-Win	**0.1207**	**0.0354**	**0.0547**
Wall	0.6646	0.5773	0.6179
Stairs	0.3583	0.1789	0.2386
Vault	0.9380	0.7460	0.8310
Forniture	0.5556	0.8286	0.6652

It is possible to obtain a further comparison through the graph of Figure 4. In this graph, the comparison of metrics for the entire dataset is shown. The results show that the accuracy of semantic segmentation has also improved thanks to the inclusion of these new generated scenes. It is fairly straightforward to deduce the motivation behind this increase in accuracy; data generation has been performed for those classes with a lower number of points available, that is to say unbalanced classes. Consequently, by balancing the dataset, the overall performances yield better results.

To the sake of completeness, the latest discussion is devoted to the generative model. The PointFlow model will be the one to design future experiments, as it gains stability for both SG and GS values. This means that the generator is able to reproduce the object regardless of the type of training-test method (see Table 2).

5. Conclusions and Future Works

The aim of this paper is to propose a framework based on DL methods able to semantically segment 3D points clouds of CH domain. Since the dataset is unbalanced, we used generative adversarial networks to increase the performance of the semantic segmentation. Experimental results demonstrated that augmenting the dataset, the rate of recognition of each element increases. In this paper, we worked with three generative networks, PointGrow, PointFlow and PointGMM for the generation of point clouds related to the cultural heritage domain. The main purpose was to obtain new scenes to improve the semantic segmentation task of the DGCNN-Mod network. The generated scenes provided good performance for the column segmentation task.

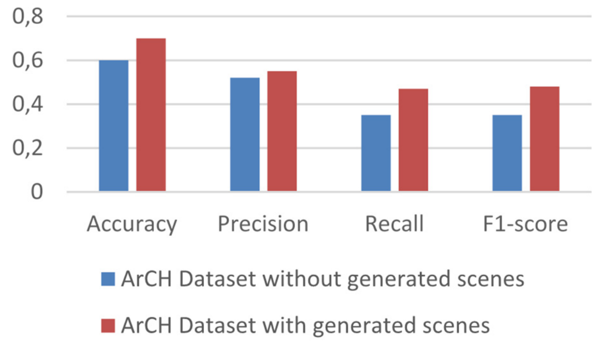


Figure 4: *Comparison of metrics considering ArCH dataset without and with generated.*

The automatic generation of 3D class object can revolutionize the cultural heritage domain in several aspects. (H)BIM, given the uniqueness of CH objects, requires time-consuming parametrization that are nowadays completely entrusted to human operation; our method opens up to great opportunities in terms of automatization. Moreover, as demonstrated by the literature, one of the main bottlenecks towards the full exploitation of DL methods for semantic segmentation is the lack of available and annotated datasets. Again, our method proves to be a robust alternative to manual and time consuming manual labelling.

A possible future development is to increase the points clouds, both for the training phase and to obtain more defined shapes. Another possible future development is to improve the fine-tuning technique on PointFlow freezing the first layers of the network. Moreover, another possible future development could be to implement additional features such as colours to make realistic the generated point clouds.

References

Achlioptas, P., Diamanti, O., Mitliagkas, I., & Guibas, L. (2018). Learning representations and generative models for 3d point clouds. In *International conference on machine learning* (pp. 40-49). PMLR.

Arshad, M., & Beksi, W. J. (2020). *A Progressive Conditional Generative Adversarial Network for Generating Dense and Colored 3D Point Clouds*. ArXiv, abs/2010.05391.

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.

https://doi.org/10.5194/isprs-annals-IV-2-W2-115-2017

Blais, F. (2004). Review of 20 years of range sensor development. *Journal of electronic imaging, 13*(1), 231-243. https://doi.org/10.1117/12.473116

Chu, P. M., Sung, Y., & Cho, K. (2018). Generative adversarial network-based method for transforming single RGB image into 3D point cloud. *IEEE Access, 7*, 1021-1029. https://doi.org/10.1109/ACCESS.2018.2886213

Egiazarian, V., Ignatyev, S., Artemov, A., Voynov, O., Kravchenko, A., Zheng, Y., & Burnaev, E. (2019). *Latent-Space Laplacian Pyramids for Adversarial Representation Learning with 3D Point Clouds.* arXiv preprint arXiv:1912.06466.

Fan, H., Su, H., & Guibas, L. J. (2017). A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 605-613).

Griffiths, D., & Boehm, J. (2019). SynthCity: A large scale synthetic point cloud. arXiv preprint arXiv:1907.04758.

Grilli, E., Özdemir, E., & Remondino, F. (2019). Application of machine and deep learning strategies for the classification of heritage point clouds. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*. https://doi.org/10.5194/isprs-archives-XLII-4-W18-447-2019.

Grilli, E., & Remondino, F. (2019). Classification of 3D digital heritage. *Remote Sensing, 11*(7), 847. https://doi.org/10.3390/rs11070847.

Hertz, A., Hanocka, R., Giryes, R., & Cohen-Or, D. (2020). PointGMM: a Neural GMM Network for Point Clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12054-12063). https://doi.org/10.1109/CVPR42600.2020.01207

Llamas, J., M Lerones, P., Medina, R., Zalama, E., & Gómez-García-Bermejo, J. (2017). Classification of architectural heritage images using deep learning techniques. *Applied Sciences, 7*(10), 992. https://doi.org/10.3390/app7100992.

Li, Z., Zhang, L., Zhong, R., Fang, T., Zhang, L., & Zhang, Z. (2016). Classification of urban point clouds: A robust supervised approach with automatically generating training data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10*(3), 1207-1220. https://doi.org/10.1109/JSTARS.2016.2628399

Martínek, J., Lenc, L., & Král, P. (2019). Training Strategies for OCR Systems for Historical Documents. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 362-373). Springer, Cham

Matrone, F., Grilli, E., Martini, M., Paolanti, M., Pierdicca, R., & Remondino, F. (2020a). Comparing Machine and Deep Matrone Learning Methods for Large 3D Heritage Semantic Segmentation. *ISPRS International Journal of Geo-Information, 9*(9), 535. https://doi.org/10.3390/ijgi9090535

Matrone, F., Lingua, A., Pierdicca, R., Malinverni, E. S., Paolanti, M., Grilli, E., Remondino, F. Murtiyoso & Landes, T. (2020b). A Benchmark for Large-Scale Heritage Point Cloud Semantic Segmentation. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 43*, 1419-1426. https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-1419-2020

Myers, G. (1999). *A dataset generator for whole genome shotgun sequencing*. In ISMB (pp. 202-210).

Murtiyoso, A., & Grussenmeyer, P. (2019). Automatic heritage building point cloud segmentation and classification using geometrical rules. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*. https://doi.org/10.5194/isprs-archives-XLII-2-W15-821-2019

Pierdicca, R., Paolanti, M., Matrone, F., Martini, M., Morbidoni, C., Malinverni, E. S., E. Frontoni & Lingua, A. M. (2020). Point Cloud Semantic Segmentation Using a Deep Learning Framework for Cultural Heritage. *Remote Sensing, 12*(6), 1005. https://doi.org/10.3390/rs12061005

Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 652-660). https://doi.org/10.1109/CVPR.2017.16

Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017). Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems*, 5099-5108.

Spina, S., Debattista, K., Bugeja, K., & Chalmers, A. (2011). Point cloud segmentation for cultural heritage sites. In *Proceedings of the 12th International conference on Virtual Reality, Archaeology and Cultural Heritage* (pp. 41-48). http://dx.doi.org/10.2312/VAST/VAST11/041-048

Sun, Y., Wang, Y., Liu, Z., Siegel, J., & Sarma, S. (2020). Pointgrow: Autoregressively learned point cloud generation with self-attention. In *The IEEE Winter Conference on Applications of Computer Vision* (pp. 61-70). https://doi.org/10.1109/WACV45572.2020.9093430

Wang, F., Zhuang, Y., Gu, H., & Hu, H. (2019). Automatic generation of synthetic LiDAR point clouds for 3-d data analysis. *IEEE Transactions on Instrumentation and Measurement, 68*(7), 2671-2673. https://doi.org/10.1109/TIM.2019.2906416

Wu, B., Zhou, X., Zhao, S., Yue, X., & Keutzer, K. (2019). Squeezesegv2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud. In *2019 International Conference on Robotics and Automation (ICRA)* (pp. 4376-4382). IEEE. https://doi.org/10.1109/ICRA.2019.8793495

Yang, G., Huang, X., Hao, Z., Liu, M. Y., Belongie, S., & Hariharan, B. (2019). Pointflow: 3d point cloud generation with continuous normalizing flows. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 4541-4550). https://doi.org/10.1109/ICCV.2019.00464