APPLICATION OF MACHINE AND DEEP LEARNING STRATEGIES FOR THE CLASSIFICATION OF HERITAGE POINT CLOUDS

E. Grilli [a,b], E. Özdemir [a], F. Remondino [a]

[a] 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy
Web: http://3dom.fbk.eu
[b] University of Bologna, Italy
Email: <grilli><eozdemir><remondino>@fbk.eu

KEY WORDS: Point Clouds, Classification, Machine Learning, Deep Learning, Cultural Heritage

ABSTRACT:

The use of heritage point cloud for documentation and dissemination purposes is nowadays increasing. The association of semantic information to 3D data by means of automated classification methods can help to characterize, describe and better interpret the object under study. In the last decades, machine learning methods have brought significant progress to classification procedures. However, the topic of cultural heritage has not been fully explored yet. This paper presents a research for the classification of heritage point clouds using different supervised learning approaches (Machine and Deep learning ones). The classification is aimed at automatically recognizing architectural components such as columns, facades or windows in large datasets. For each case study and employed classification method, different accuracy metrics are calculated and compared.
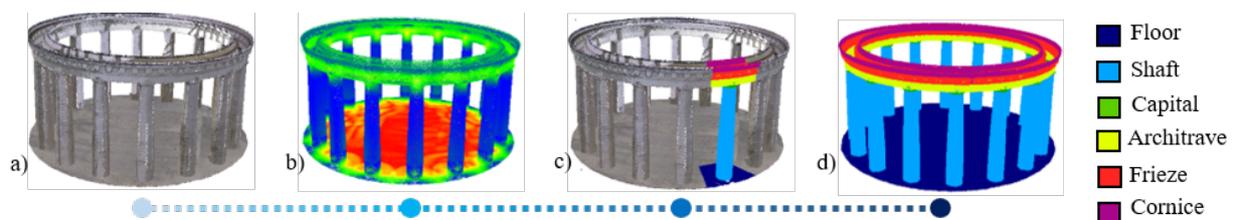
Figure 1. 3D point cloud classification process based on machine learning: surveyed point cloud (**a**), features extraction (**b**), manual annotation of a small portion to define training set and classes (**c**), final automated classification results (**d**).

1. INTRODUCTION

The 3D documentation of Cultural Heritage monuments and sites with point clouds or meshes, coming from photogrammetry and laser scanning surveys, is broadly diffuse. Given the recent evolution of technologies and digital tools, the need for automated and reliable methods to classify point clouds or meshes is becoming fundamental. Among the possible and interesting applications provided by the classification of heritage 3D data we can mention: identification and distinction of structural and decorative architectural elements, mapping of different states of conservation and materials, automatic recognition of similar architectural elements as a propaedeutic phase for Building Information Modelling (BIM), etc.

In the literature, different methods of data classification were proposed (Grilli et al., 2017) like edge and region-based approaches (applied initially for image segmentation) (Wang and Shang, 2009; Vo et al., 2015) or model-fitting approaches, based on the possibility to fit geometric primitives to the 3D shapes (Chen et al., 2014). With the advent of Artificial Intelligence (AI) solutions, further progress in automation and interesting results came out. In particular, Machine and Deep Learning (ML/DL) methods allowed the development of algorithms that let machines to take decisions based on empirical training data.

Deep Learning can be considered an evolution of Machine Learning. Its algorithms are structured in layers to create an artificial neural network that can learn and make intelligent decision on its own. The use of Machine Learning techniques for point cloud classification has been successfully investigated in the last decade in the geospatial environment (Guo et al., 2014; Niemeyer et al., 2014; Weinmann et al., 2015; Qi et al., 2017; Özdemir and Remondino, 2019a) while in the Cultural Heritage (CH) field it has only recently started to be explored (Poux et al., 2017; Grilli and Remondino, 2019).

The paper aims to explore the potential offered by Machine and Deep Learning approaches for the supervised classification of 3D heritage case studies (Figure 1).

In the paper, firstly, a literature review is presented. Secondly, different ML/DL point cloud classification approaches are presented and then experimented on two different case studies: the temple of Neptune in Paestum and some renaissance buildings with porticoes in Bologna. Classification results are finally presented and commented relying on confusion matrix scores.

2. RELATED WORKS

In recent years, significant progress has come out in automatic procedures for classification of point clouds or meshes thanks to the advent of Machine Learning approaches (Hackel et al., 2016; Weinmann et al., 2017; Wang et al., 2018).

Several benchmarks have been proposed in the Geomatics community, providing labelled terrestrial and airborne data on which users can test and validate their algorithms. Most of the available datasets provide classified natural, urban, and street scenes (e.g., www.semantic3d.net, www.cityscapes-dataset.com, etc.). While in those scenarios, the object classes and labels are almost defined (mainly ground, roads, trees, and buildings), the

identification of precise categories in the heritage field is much more complicated, as:

- for the same case study several classes can be identified based upon different purposes;
- not always a semantic architectural class is linked to a precise shape/colour.

Probably for these reasons, up to now, the only available databases of annotated heritage are with 2D images and refer only to building facades, e.g. eTRIMS (Korc and Forstner, 2009), Ecole Centrale Paris (ECP) Facades dataset (Teboul et al., 2010), CMP Facade Database (Tyleček and Šára, 2013). Despite this existing data shortage, different Machine Learning approaches were proposed in the architectural and heritage context. Oses et al. (2014) used different Machine Learning classifiers to perform an image-based delineation of masonry walls. Amato et al. (2015) used k-nearest neighbour (kNN) classification and landmark recognition techniques to address the problem of monument recognition in images. Convolutional Neural Networks (CNN) were applied for the first time to heritage scenarios in Llamas et al. (2016) and Llamas et al., (2017). CNNs are also used by Yasser et al. (2017) for visual categorization and to create a digital heritage search platform (ICARE) that allows users to archive digital heritage content and perform semantic queries over multimodal cultural heritage data archives.

In some cases, the classification is performed for annotation and restoration purposes, and the information is transferred from 2D to 3D (Campanaro et al., 2016; Grilli et al., 2018). The web platform Aioli (www.aioli.cloud) allows a semi-automatic annotation of 3D heritage, where 2D mapping data are in real-time displayed onto a 3D model (Roussel et al., 2019).

To the author's knowledge, there are no works applying Deep Learning methods for the classification of 3D architectural heritage.
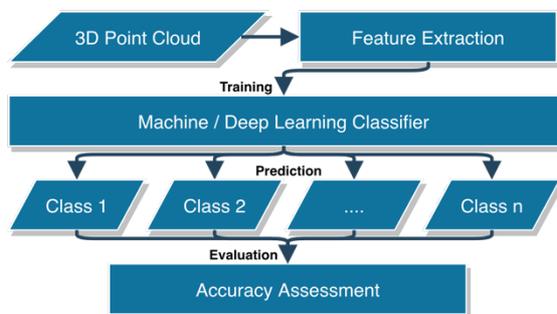
3. ANALYZED METHODS

Figure 2. Classification workflow.

The different ML/DL methods presented in this paper (Section 3.2-3.5) work directly on 3D point clouds and they are all supervised, as the input data contain associated labels (i.e. classes) information.

The classification processes (Figure 2) consist of four steps: feature extraction, feature selection, model training and prediction. To evaluate the performance of the classification methods for each case study a test set is taken into consideration. For every point of the dataset, the label predicted by the classifier is compared with the same manually annotated. Confusion matrices are then generated, and the following accuracy metrics are calculated for each class:

$$Precision = \frac{Tp}{Tp + Fp} \qquad (1)$$

$$Recall = \frac{Tp}{Tp + Fn} \qquad (2)$$

$$F1\ score = 2 * \frac{Recall * Precision}{Recall + Precision} \qquad (3)$$

$$True\ negative\ rate = \frac{Tn}{Tn + Fp} \qquad (4)$$

$$Balanced\ accuracy = \frac{True\ negative\ rate + Recall}{2} \qquad (5)$$

where Tp = true positive, Tn = true negative, Fp = false positive, Fn = false negative.

3.1 Features extraction and selection

For the training and classification goal, different sets of features are used (Figure 3), depending on the case study and the approach (ML / DL) (Özdemir and Remondino, 2019b). In case of heritage and architectural 3D data, we combined the use of:

- Decentralised coordinates: they are used to represent the local geometry around a point as a patch of k-number of nearest points. To decentralise the coordinates, the minimum x, y, z values are subtracted within each sequence and the sequences are sorted with respect to the decentralised coordinate values (Figure 3a).
- Radiometric values: the input data is a 3-band RGB colour space with an 8-bit radiometric resolution per band. The RGB values are re-scaled to have values between 0-1 (Figure 3b).
- Geometric features (Figure 3c): covariance features and others are described in Section 3.1.1.
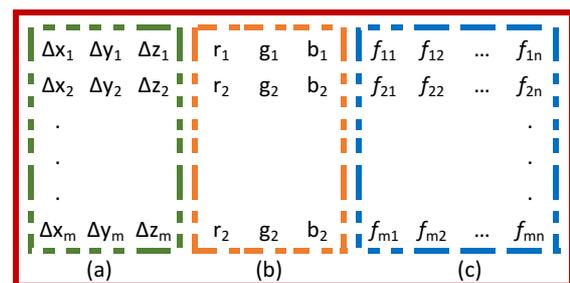

Figure 3. Data sequence used for the classification purpose with m number of points and n number of features: decentralised coordinates (**a**), radiometric values (**b**), geometric features (**c**).

Linearity	$L_\lambda = \frac{\lambda 1 - \lambda 2}{\lambda 1}$	(6)
Planarity	$P_\lambda = \frac{\lambda 2 - \lambda 3}{\lambda 1}$	(7)
Sphericity	$S_\lambda = \frac{\lambda 3}{\lambda 1}$	(8)
Omnivariance	$O_\lambda = \sqrt[3]{\prod_{j=1}^{3} \lambda j}$	(9)
Anisotropy	$A_\lambda = \frac{\lambda 1 - \lambda 3}{\lambda 1}$	(10)
Surface Variation	$C_\lambda = \frac{\lambda 3}{\Sigma \lambda}$	(11)
Verticality	$V = 1 - n_z$	(12)

Table 1. Considered geometric features.

3.1.1 Geometric features

The geometric features employed include (i) covariance features, (ii) normal based features (Verticality V), and (iii) height-based features (Z coordinates). The covariance features (also called eigenfeatures) are based on the covariance matrix (Cheata et al., 2009) computed within a local neighbourhood of a 3D point. The combinations of three eigenvalues $\lambda_i (\lambda_1 > \lambda_2 > \lambda_3)$ extracted from the covariance matrix hold a great potential to calculate local features and describe the shape of the neighbourhood (Blomey et

al., 2014). The measures of linearity L, planarity P and sphericity S provide information about the presence of a linear 1D structure, a planar 2D structure or a volumetric 3D structure. Further measures are provided by omnivariance O, anisotropy A and Local surface variation C_λ (Table 1).

Different strategies can be applied to identify local neighbourhoods for points belonging to a 3D point cloud (Weinmann et al., 2013). In our method (Grilli et al 2019), features are first calculated on spherical neighbourhoods at various radius sizes (multi-scale approach) using CloudCompare (Hackel et al., 2016). Then, features are examined to investigate whether some classes are particularly well described by features at specific scales. Finally, the optimum subset of features is selected in order to emphasize the differences between the classes we are interested in.

Unlike the conventional DL approaches, we provide handcrafted features (the same features for the ML methods) as input to the employed DL algorithms (Ozdemir and Remondino, 2019b).

3.2 Machine Learning approach - Random Forest

Random Forest (RF) is a supervised classification classifier (Breiman, 2001) that uses an ensemble of classification trees, gets a prediction from each tree, and selects the best solution by means of voting. Two parameters need to be set to produce the forest trees: the number of decision trees to be generated (N_{tree}) and the number of variables to be selected and tested for the best split when growing the trees (M_{try}) (Belgiu et al., 2016). We rely on the RF implementation available in the Scikit-learn Python library (version 0.21.1). During the training process, the Ntree and Mtry are tuned considering the best F1-score computed on the test set.

3.3 Machine Learning approach – OvO classifier

The One-versus-One (OvO) classifier converts a group of binary classifiers into a multiclass classifier. It works training the binary classifiers in a one vs. one trend. In the case of N possible classes, it trains N*(N-1)/2 binary classifiers, which are then employed for the identification of the classes on the test sample. In our tested we used the OvO classifier available in the dlib C++ library (King, 2009).

3.4 Deep Learning approach - 1D and 2D CNN

Two Convolutional Neural Networks (CNN) (Fukushima et al., 1980) methods are also proposed. CNN is a specific type of artificial neural network specialized in processing data that has a grid-like topology, such as an image. The layers of a CNN consist of an input layer, an output layer and a hidden layer that includes multiple convolutional layers, pooling layers, fully connected layers and normalization layers. The tested CNNs are:
- 1D CNN: it consists of 1 input layer, 2 convolutional layers, 3 dense layers, 1 maximum pooling layer, 1 global average pooling layer and 1 dropout layer.
- 2D CCN: it is composed of 1 input layer, 4 2D convolutional layers, 2 2D max pooling layer, 3 dropout layers, 1 flatten layer and 2 dense layers.

3.5 Deep Learning approach - Bi-LSTM

Recurrent Neural Network (RNN) (Rumelhart et al., 1988) is commonly used for modelling sequential data. The data is sequential if the building blocks in a dataset are not independent from each other. The most common application for RNN are handwriting or speech recognition, translation, etc. Our RNN consists of five layers: sequence input layer, Bidirectional Long Short-Term Memory layer (Bi-LSTM) with 200 hidden units, fully connected layer, softmax layer and classification

layer. We describe each point with a sequence that is generated with its surrounding points (i.e. each row in Figure 3 is a part of the sequence). These sequences are expected to represent the geometry around each point in a better way when compared to a single feature vector representation.

4. CASE STUDIES AND RESULTS

The aforementioned ML/DL classification approaches were applied to two different heritage datasets:

a) The Greek temple of Neptune in Paestum (Italy): it was built in the Doric order around 460 – 450 BC. It measures ca 24,5 x 60 m (Fig. 4) and the available point cloud is the result of a combined UAV and terrestrial photogrammetric survey (Fiorillo et al., 2013). With ML and DL approaches, the aim is to semantically segment the 3D data of the temple considering its Greek architectural elements.

Figure 4. Temple of Neptune in Paestum, Italy.

b) Building with porticos in Bologna (Italy): the historical porticos of Bologna (Fig. 5) were built during the 11th–20th centuries. We consider a portion of ca 85x6m, surveyed with photogrammetric techniques (Remondino et al., 2016). The ML/DL classification is aimed at a semantic annotation of the different architectural and decorative elements.

Figure 5. Building with porticoes in Bologna, Italy (Source: Google Street View).

For both case studies, the five classification approaches described in Section 3 are run, using different sets of features as input (Table 2). For the ML approaches and the 1D CNN all the points of the cloud are described by a feature vector that contains different geometric features chosen ad hoc. Regarding the 2D CNN and Bi-LSTM, in addition to these geometric features, the classification was led with, and without the decentralised coordinates.

	TEMPLE		PORTICOS	
RF	c		b + c	
OvO	c		b + c	
Conv 1D	c		b + c	
Conv 2D	a + c	c	a + b + c	b + c
BiLSTM	a + c	c	a + b + c	b + c

Table 2. Considered geometric features for each classification approach. As described in figure 3: a = decentralised coordinates, b = radiometric features, c = geometric features.

4.1 Temple of Neptune in Paestum

After applying a subsampling, to speed up the computational process, the data consist of some 2.2 million points (Figure 6).

Figure 6. Photogrammetric point cloud of the temple of Neptune in Paestum (ca 2,2 million points).

Ten different classes corresponding to the architectural elements of the temple were identified. Then a small portion of the entire data set was accurately manually annotated (Figure 7).

Grass
Crepidoma
Pavement
Shaft
Echinus
Abacus
Architrave
Frieze
Cornice
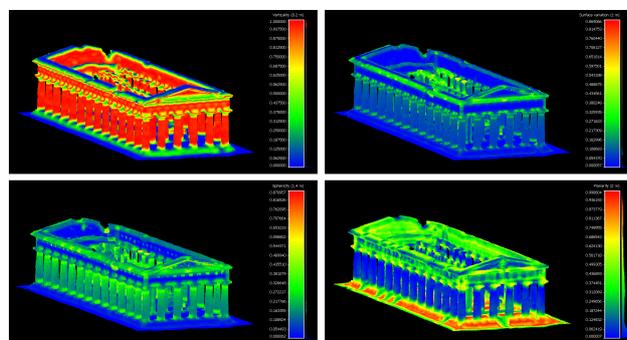Tympanum

Figure 7. Manually labelled temple.

Figure 8. Some relevant features computed on the Neptune's point cloud. From left to right respectively, clock-wise: verticality, surface variation, sphericity and planarity.

Covariance features (Table 1, Fig. 8) were then extracted at different neighbourhood sizes, correlated with the dimensions of the column orders. According to the selected geometric feature chosen and the used radius size (*r*), it is possible to highlight different architectural elements. The feature *Surface Variation*, for example, can emphasise the columns if extracted at a neighbourhood size *r* = radius of the columns. The *Planarity* extracted with the same radius distinguishes planar elements (e.g. facades, floors) from cylindrical ones (e.g. columns).

Using a combination of these features, both Machine and Deep Learning methods were trained to predict the labels on the entire dataset (Figure 8).
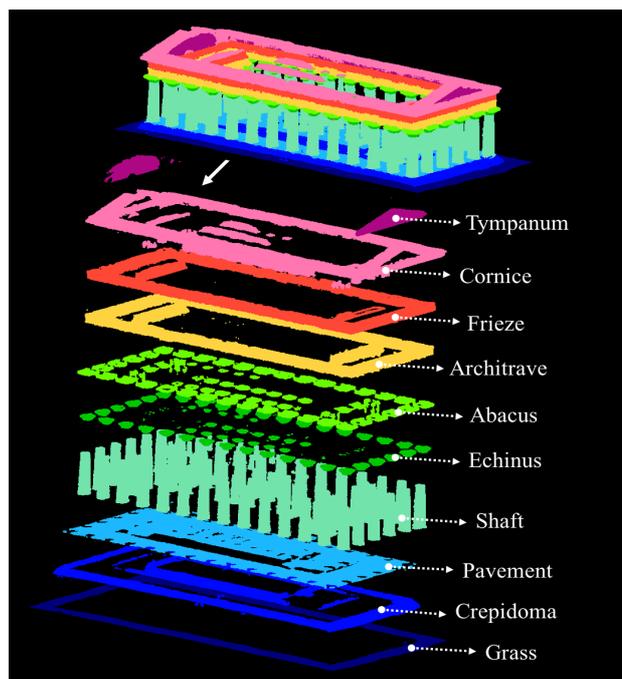

Figure 9. RF classification results and exploded view.

Table 3 reports the results obtained with the RF classifier, including confusion matrix and accuracy metrics. Each row of the matrix represents the instances in an actual class (ground truth), while each column represents the instances in a predicted class. In general, we notice that most of the classification errors are between classes with quite similar geometries as "Abacus" and "Architrave", and "Frieze", "Cornice" and "Tympanum. Table 4 shows in parallel the per-class F1-score results for each method applied. The F1-score averages are between the 86.69 % with Bi-LSTM and 92% with RF. Table 5 summarizes all the accuracy metrics reached with the different approaches. As we can observe from the diagram, higher levels of accuracy were achieved using the machine learning approaches (RF and OvO).

CLASS NAME	Grass	Crep.	Pavim.	Shaft	Echinus	Abacus	Architr.	Frieze	Cornice	Tymp.	Prec.	Recall	F1	True Neg.	Balanc. Acc.
Grass	56998	1647	0	0	0	0	0	0	0	0	97%	100%	99%	100%	100%
Crepidoma	67	38389	484	497	0	0	0	0	0	0	97%	93%	95%	100%	97%
Pavement	0	958	62211	993	0	0	0	0	0	0	97%	99%	98%	100%	99%
Shaft	0	54	125	169950	172	56	4	0	268	0	100%	98%	99%	100%	99%
Echinus	0	0	0	966	18717	988	4	0	40	0	90%	87%	88%	100%	93%
Abacus	0	0	0	63	2651	23941	3998	0	636	0	77%	88%	82%	99%	93%
Architrave	0	0	0	0	0	2140	50903	2600	1	0	91%	91%	91%	99%	95%
Frieze	0	0	0	0	0	0	1036	34609	334	33	96%	80%	87%	100%	90%
Cornice	0	42	0	90	67	223	90	4569	106415	2071	94%	95%	95%	99%	97%
Tympanum	0	0	0	0	0	0	4	1492	3841	21393	80%	91%	85%	99%	95%
AVERAGE											92%	92%	92%	99%	96%

Table 3. RF classification results: Confusion Matrix and per-class accuracy for the temple dataset.

CLASS / ALGORITHM	F1-Score						
	RF	OvO	CNN 1D	CNN 2D (a,c)	CNN 2D (c)	Bi-LSTM (a,c)	Bi-LSTM (c)
Grass	98.52%	99.05%	99.05%	99.08%	98.89%	98.20%	97.95%
Crepidoma	95.34%	95.42%	95.30%	96.22%	95.84%	92.48%	91.57%
Pavement	97.98%	97.84%	97.31%	97.89%	98.02%	96.68%	96.48%
Shaft	99.04%	98.73%	98.25%	98.30%	98.77%	97.60%	97.90%
Echinus	88.45%	84.40%	81.03%	79.54%	84.91%	78.15%	76.53%
Abacus	81.66%	80.64%	74.85%	75.93%	79.23%	77.86%	72.94%
Architrave	91.16%	91.90%	78.56%	80.22%	90.69%	89.10%	89.53%
Frieze	87.31%	87.74%	77.50%	66.97%	87.31%	78.13%	86.02%
Cornice	94.55%	93.90%	93.24%	92.73%	93.54%	87.76%	89.52%
Tympanum	85.19%	84.66%	85.99%	80.55%	85.31%	70.94%	74.14%
AVERAGE	91.92%	91.43%	88.11%	86.74%	91.25%	86.69%	87.26%

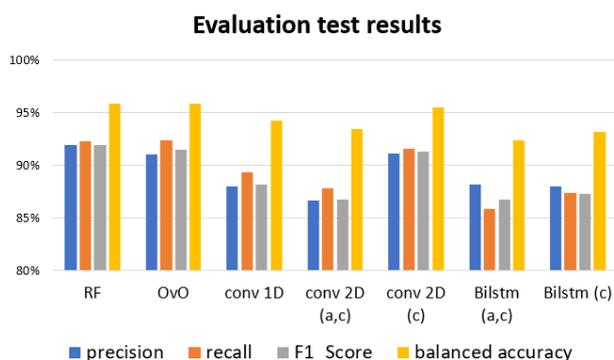Table 4. A summary of all tested ML/DL classification methods reporting per-class F1-score for the temple dataset.

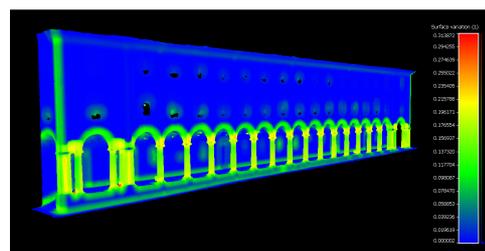Table 5. Summary of the classification results for the temple dataset achieved with the different ML/DL methods.

4.2 Renaissance buildings in Bologna

The photogrammetric point cloud of the porticos consists of ca 1 million points (Figure 10).

Figure 10. Photogrammetric point cloud of a renaissance building in Bologna (ca 1,1 million points).

A a small portion of the entire dataset was manually annotated with 14 classes (Figure 11) and some significant features were extracted (Figure 12). Then, the different classifiers were trained to classify the entire point cloud (Figure 13). Confusion matrixes and classification results obtained using the machine and deep learning approaches are showed below (Table 6-7-8). As for the previous case study, the best results were achieved applying machine learning approaches.

Figure 11. A portion of the Bologna dataset manually labelled with 14 classes.

Figure 12. Example of the Surface Variation feature extracted on the Bologna dataset.

Figure 13. RF classification results on the Bologna dataset.

Among the identified 14 classes, the ones with more classification errors were those with similar geometric properties. For example, the "road" was misclassified as "pavement " or also the "moulding" as "facade". This is probably due to the limited number of points in the annotations.
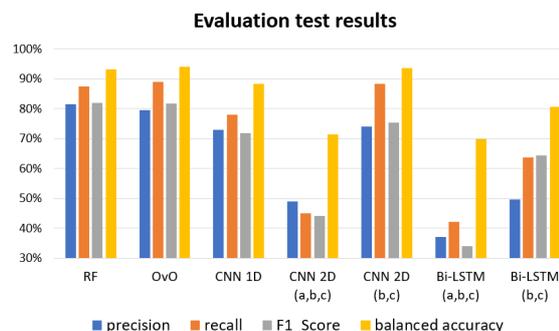

Table 6. Summary of the classification results for the porticos dataset achieved with the different ML/DL methods.

CLASS NAME	Road	Facade	Pav.	Base	Shaft	Capital	Arch.	Win/do or	Vault	Mouldi ng	Curtai n	Drainp. V	Cornic e	Drainp. H	Prec.	Recall	F1	True Neg.	Bal. Acc.
Road	1411	43	5105	0	0	0	0	0	0	0	0	30	0	0	21%	100%	35%	96%	98%
Facade	0	46838	1968	73	0	0	145	95	254	660	62	151	14	0	93%	91%	92%	96%	93%
Pav.	0	59	16695	282	0	0	0	21	1	1	0	25	0	0	98%	70%	82%	100%	85%
Base	0	8	17	1604	226	0	1	11	0	1	0	3	0	0	86%	80%	83%	100%	90%
Shaft	0	0	0	0	9143	0	0	0	0	0	0	9	0	0	100%	95%	97%	100%	97%
Capital	0	0	0	0	246	1101	178	0	73	0	0	22	0	0	68%	99%	81%	100%	99%
Arch	0	169	0	0	0	0	6179	0	757	5	0	0	0	0	87%	91%	89%	99%	95%
Win/do	0	27	0	0	0	0	0	2321	0	242	0	0	0	0	90%	74%	81%	100%	87%
Vault	0	1060	0	0	0	1	218	0	18331	0	0	0	0	0	93%	94%	94%	99%	97%
Moulding	0	2919	55	19	0	0	7	693	0	6526	205	3	0	0	63%	82%	71%	97%	89%
Curtain	0	49	0	0	0	0	0	0	0	198	3464	4	0	0	93%	92%	93%	100%	96%
Drainp.V	0	221	0	23	52	10	69	8	0	354	36	1150	47	0	58%	82%	68%	99%	91%
Cornice	0	118	0	0	0	0	0	0	0	0	0	2	4746	175	94%	98%	96%	100%	99%
Drainp.H	0	0	0	0	0	0	0	0	0	0	0	0	13	547	98%	76%	85%	100%	88%
AVERAGE															82%	87%	82%	99%	93%

Table 7. RF classification results: Confusion Matrix and per-class accuracy for the porticos dataset.

CLASS / ALGORITHM	F1-Score						
	RF	OvO	CNN 1D	CNN 2D (a,b,c)	CNN 2D (b,c)	Bi-LSTM (a,b,c)	Bi-LSTM (b,c)
Road	35.28%	46.54%	87.29%	0.00%	2.93%	3.40%	9.98%
Facade	92.05%	90.79%	88.61%	81.41%	91.35%	78.70%	83.68%
Pavement	81.59%	86.48%	89.20%	71.14%	80.47%	75.12%	78.50%
Base	82.85%	86.58%	65.96%	0.00%	83.48%	0.00%	0.00%
Shaft	97.17%	96.90%	93.23%	82.10%	97.90%	69.95%	82.28%
Capital	80.60%	82.90%	58.52%	0.00%	64.58%	0.00%	0.00%
Arch	88.86%	86.90%	67.05%	40.01%	88.48%	21.58%	53.06%
Window/door	80.89%	79.90%	61.48%	65.67%	77.52%	53.15%	68.97%
Vault	93.94%	92.90%	88.82%	84.33%	94.57%	80.30%	85.99%
Molding	70.88%	58.77%	48.14%	23.02%	66.26%	0.02%	0.43%
Curtain	92.60%	89.11%	81.15%	82.98%	95.71%	6.21%	90.02%
Drainp. V.	68.27%	62.63%	21.82%	0.00%	57.32%	0.00%	0.00%
Cornice	96.26%	95.78%	85.84%	87.32%	94.23%	86.57%	90.42%
Drainp. H.	85.34%	88.02%	69.26%	0.00%	60.29%	0.00%	0.00%
AVERAGE	81.90%	81.73%	71.88%	44.14%	75.37%	33.93%	45.95%

Table 8. A summary of all tested ML/DL classification methods reporting the per-class F1-score for the porticos dataset.
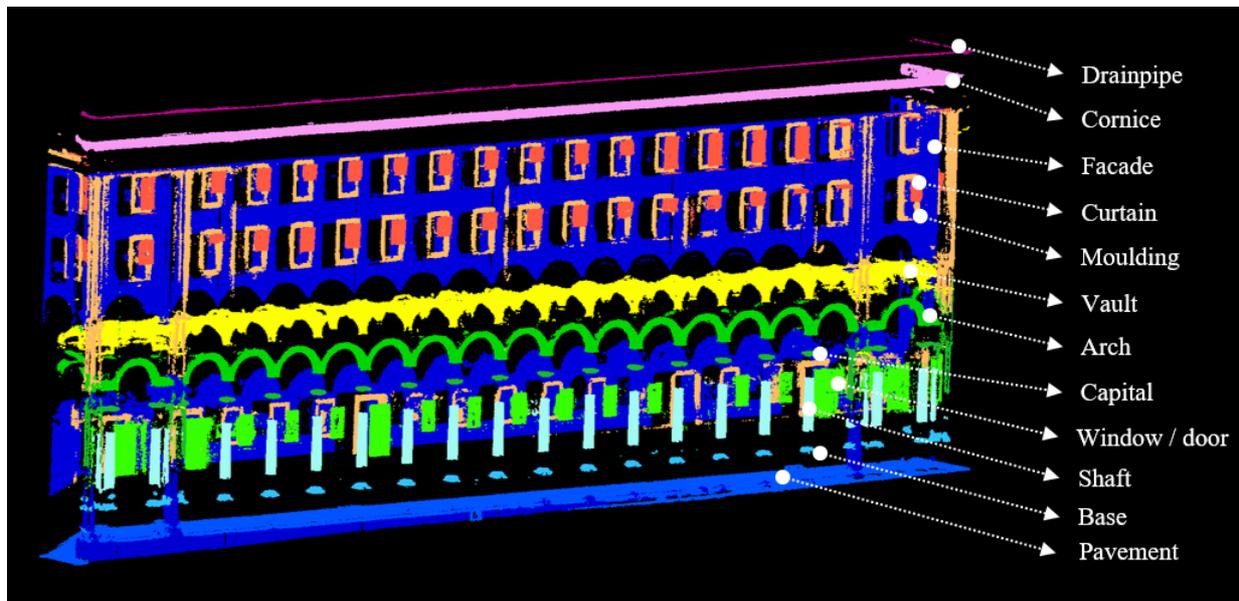

Figure 14. Exploded view of the Bologna dataset after the automated classification (vertical drainpipe and road classes are not visualized due to their low accuracy score).

5. CONCLUSIONS

The paper presented an evaluation of different ML/DL classification approaches to semantically segment point clouds of architectural and archaeological scenarios. From the summarized results in Tables 5 and 6, we can state that:

- ML approaches outperformed DL methods;
- for the Temple dataset, the F1-scores are between 86% and 95% while for the Bologna they change between 34% up to 82%. In our opinion, even if the features were handcrafted appositely for the case studies, the class structure complexity of the Bologna dataset caused lower accuracy metrics;
- based on the achieved accuracy metrics in the Bologna dataset, we can suggest that the used DL approaches are not suitable for this kind of dataset;
- in case of heritage datasets, the conventional use of decentralised coordinates for DL approaches reduced the overall accuracy. From detailed analysis of Table 4 and Table 8 we can observe that the accuracy decreases when the classes share the same geometry. This represents one of the most challenging point for heritage classification, as there's not always a correspondence between shape/colours and semantics for the architectural classes.
- as concern the training times, the used DL approaches took about 10 minutes on GPU, while the ML ones completed the training in less than a minute on a CPU.

As further evolutions we plan to explore new features and class structures to improve our classification results. Moreover, starting from the achieved classification results (Fig. 9 and 14), it would be interesting to develop a tool to assist the conversion of semantic point cloud to parametric 3D model (HBIM/BIM).

ACKNOWLEDGEMENTS

REFERENCES

Amato, G., Falchi, F. and Gennaro, C., 2015. Fast image classification for monument recognition. *Journal on Computing and Cultural Heritage (JOCCH)*, *8*(4), p.18.

Belgiu, M. and Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, pp.24-31.

Blomley, R., Weinmann, M., Leitloff, J., and Jutzi, B., 2014. Shape distribution features for point cloud analysis-a geometric histogram approach on multiple scales. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *2*(3), p. 9.

Breiman, L., 2001. Random forests. *Machine learning*, *45*(1), pp. 5-32.

Campanaro, D. M., Landeschi, G., Dell'Unto, N., and Touati, A.M. L. 2016. 3D GIS for cultural heritage restoration: A 'white box' workflow. *Journal of Cultural Heritage*, Vol. 18, pp.321-332.

Chehata, N., Guo, L. and Mallet, C., 2009. Airborne lidar feature selection for urban classification using random forests. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 38-3-W8, pp 207-2012.

Chen, D., Zhang, L., Mathiopoulos, P. T. and Huang, X., 2014. A Methodology for Automated Segmentation and Reconstruction of Urban 3-D Buildings from ALS Point Clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 10, Vol. 7, pp. 4199-4217.

Fiorillo, F., Fernández-Palacios, B. J., Remondino, F., and Barba, S., 2013. 3D Surveying and modelling of the Archaeological Area of Paestum, Italy. *Virtual Archaeology Review*, Vol. 4(8), pp. 55-60.

Fukushima, K., 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, *36*(4), pp.193-202.

Guo, Y., Bennamoun, M., Sohel, F., Lu, M., Wan, J. and Kwok, N.M., 2016. A comprehensive performance evaluation of 3D local feature descriptors. *International Journal of Computer Vision*, *116*(1), pp.66-89.

Grilli, E., Menna, F. and Remondino, F., 2017. A review of point clouds segmentation and classification algorithms. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, pp.339-346.

Grilli, E., Dininno, D., Petrucci, G. and Remondino, F., 2018. From 2D to 3D supervised segmentation and classification for cultural heritage applications. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, *42*(2), pp. 399-406.

Grilli, E. and Remondino, F., 2019. Classification of 3D Digital Heritage. *Remote Sensing*, *11*(7), p.847.

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

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

King, D.E. 2009. Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research* 10, pp. 1755-1758.

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

Llamas, J., Lerones, P. M., Zalama, E., and Gómez-García-Bermejo, J., 2016, October: Applying deep learning techniques to cultural heritage images within the INCEPTION project. In *Euro-Mediterranean Conference*, pp. 25-32. Springer, Cham.

Llamas, J., M Lerones, P., Medina, R., Zalama, E., and Gómez-García-Bermejo, J., 2017. Classification of architectural heritage images using deep learning techniques. *Applied Sciences*, 7(10), p. 992.

Niemeyer, J., Rottensteiner, F. and Soergel, U., 2014. Contextual classification of LiDAR data and building object detection in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 87, pp. 152-165.

Özdemir, E. and Remondino, F., 2019a. Classification of aerial point clouds with deep learning. *Int. Archives of the*

Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. *42-2-W13*, pp.103-110.

Özdemir, E. and Remondino, F., 2019b. Aerial point cloud classification with deep learning and machine learning algorithms. *Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Science.* Proceedings Geospatial Conference 2019, in press.

Oses, N., Dornaika, F., and Moujahid, A., 2014. Image-based delineation and classification of built heritage masonry. *Remote Sensing,* 6(3), pp 1863-1889.

Poux, F., Neuville, R., Hallot, P. and Billen, R., 2017. Point cloud classification of tesserae from terrestrial laser data combined with dense image matching for archaeological information extraction. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4*, pp.203-211.

Qi, C. R., Su, H., Mo, K., and Guibas, L. J., 2017. Pointnet: Deep learning on point sets for 3D classification and segmentation. Proc. *Computer Vision and Pattern Recognition* (CVPR), IEEE, 1(2), p.4.

Remondino, F., Gaiani, M., Apollonio, F., Ballabeni, A., Ballabeni, M. and Morabito, D., 2016. 3D documentation of 40 kilometers of historical porticoes-the challenge. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 41(B5), pp. 711-718.

Roussel R., Bagnéris M., De Luca L. and Bomblet P., 2019. A digital diagnosis for the <<autumn>> statue (Marseille, France): photogrammetry, digital cartography and construction of a thesaurus. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* Vol. XLII-2/W15.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 1988. Learning representations by back-propagating errors. *Cognitive modeling*, *5*(3), p.1.

Teboul, O., Kokkinos, I., Simon, L., Koutsourakis, P. and Paragios, N., 2013. Parsing facades with shape grammars and reinforcement learning. *IEEE transactions on pattern analysis and machine intelligence*, *35*(7), pp.1744-1756.

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

Vo, A. V., Truong-Hong, L., Laefer, D. F., and Bertolotto, M., 2015. Octree-based region growing for point cloud segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 104, pp. 88-100.

Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M. and Solomon, J.M., 2018. Dynamic graph CNN for learning on point clouds. *arXiv preprint arXiv:1801.07829.*

Wang, J. and Shan, J., 2009, March. Segmentation of LiDAR point clouds for building extraction. In *American Society for Photogramm. Remote Sens. Annual Conference, Baltimore, MD*, pp. 9-13.

Weinmann, M., Jutzi, B. and Mallet, C., 2013. Feature relevance assessment for the semantic interpretation of 3D point cloud data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *5*(W2), p.1.

Weinmann, M., Schmidt, A., Mallet, C., Hinz, S., Rottensteiner, F. and Jutzi, B., 2015. Contextual classification of point cloud data by exploiting individual 3D neigbourhoods. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-3 (2015), Nr. W4*, *2*(W4), pp.271-278.

Weinmann, M. and Weinmann, M., 2017. Geospatial Computer Vision Based on Multi-Modal Data - How Valuable Is Shape Information for the Extraction of Semantic Information? *Remote Sensing*, *10*(1), p.2.

Yasser, A.M., Clawson, K. and Bowerman, C., 2017, July. Saving cultural heritage with digital make-believe: machine learning and digital techniques to the rescue. In *Proceedings of the 31st British Computer Society Human Computer Interaction Conference*. BCS Learning & Development Ltd, pp 97-101.