

Article

# Classification of 3D Digital Heritage

Eleonora Grilli <sup>1,2,\*</sup>  and Fabio Remondino <sup>1</sup>

<sup>1</sup> 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Via Sommarive 18, 38121 Trento, Italy; remondino@fbk.eu

<sup>2</sup> Department of Architecture, Alma Mater Studiorum—University of Bologna, Viale del Risorgimento 2, 40136 Bologna, Italy

\* Correspondence: grilli@fbk.eu

Received: 21 February 2019; Accepted: 26 March 2019; Published: 8 April 2019



**Abstract:** In recent years, the use of 3D models in cultural and archaeological heritage for documentation and dissemination purposes is increasing. The association of heterogeneous information to 3D data by means of automated segmentation and classification methods can help to characterize, describe and better interpret the object under study. Indeed, the high complexity of 3D data along with the large diversity of heritage assets themselves have constituted segmentation and classification methods as currently active research topics. Although machine learning methods brought great progress in this respect, few advances have been developed in relation to cultural heritage 3D data. Starting from the existing literature, this paper aims to develop, explore and validate reliable and efficient automated procedures for the classification of 3D data (point clouds or polygonal mesh models) of heritage scenarios. In more detail, the proposed solution works on 2D data (“texture-based” approach) or directly on the 3D data (“geometry-based approach) with supervised or unsupervised machine learning strategies. The method was applied and validated on four different archaeological/architectural scenarios. Experimental results demonstrate that the proposed approach is reliable and replicable and it is effective for restoration and documentation purposes, providing metric information e.g. of damaged areas to be restored.

**Keywords:** classification; segmentation; cultural heritage; machine learning; random forest

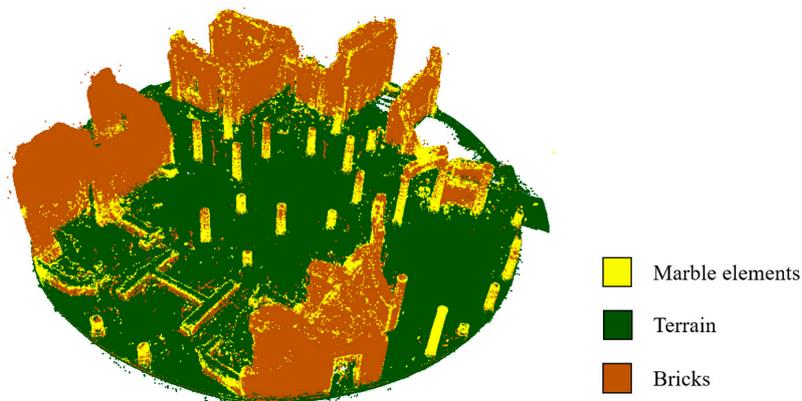
## 1. Introduction

The generation of 3D data of heritage sites or monuments, being point clouds or polygonal models, is altering the approach that cultural heritage specialists use for the analysis, interpretation, communication and valorization of such historical information. Indeed, 3D information allows one, e.g.; to perform morphological measurements, quantitative analysis, and information annotation, as well as produce decay maps, while enabling easy access and study of remote sites and structures.

The management of architectural heritage information is considered crucial for a better understanding of the heritage data as well as for the development of targeted conservation policies and actions. An efficient information management strategy should take into consideration three main concepts: segmentation, structuring the hierarchical relationships and semantic enrichment [1]. The demand for automatic model analysis and understanding is continuously increasing. Recent years have witnessed significant progress in automatic procedures for segmentation and classification of point clouds or meshes [2–4]. There are multiple studies related to the segmentation topic, mainly driven by specific needs provided by the field of application (Building Information Modeling (BIM) [5], heritage documentation and preservation [6], robotics [7] autonomous driving [8], urban planning [9], etc.).

In the cultural heritage field, the identification of different components (Figure 1) in point clouds and 3D meshes is of primary importance because it can facilitate the study of monuments and

integrating them with heterogeneous information and attributes. However, it remains a challenging task considering the complexity and high variety that heritage case studies can have.

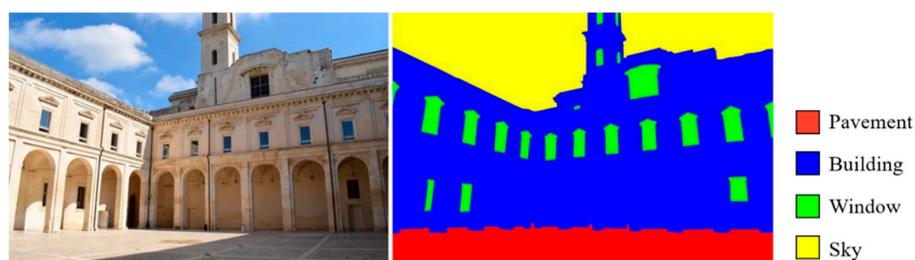


**Figure 1.** Example of a segmented and classified point cloud.

The research presented in this article was motivated by the need to identify and map different states of conservation phases or employed materials in heritage objects. Towards this direction, we developed a method: (i) to document and retrieve historical and architectural information; (ii) to distinguish different constructing techniques (e.g. types of opus, etc.); and (iii) to recognize the presence of existing restoration evidence. The retrieval of such information in historic buildings by traditional methods (e.g.; manual mapping or simple visual inspection by experts) are considered time-consuming and laborious procedures [10]. The aim of our research was not to develop a new algorithm for the classification of 3D data (point clouds or polygonal mesh models) of heritage scenarios, but to explore the applicability of supervised machine learning approaches to an almost unexplored field of application (i.e.; cultural heritage) proposing a reliable and efficient pipeline that can be standardized for different case studies.

## 2. State of the Art: 2D/3D Segmentation and Classification Techniques

Both image and cloud segmentation are fundamental tasks in various application, such as object detection [11], medical analyses [12], license plate and vehicle recognition [13], classification of microorganisms [14], fruit recognition [15] and many more [16]. Segmentation is the process of grouping data (e.g.; images, point clouds or meshes) into multiple homogeneous regions with similar properties [17] (Figure 2). These regions are homogeneous with respect to some criteria, called features, that constitute a characteristic property or set of properties which is unique, measurable and differentiable. In the case of 2D imagery, features refer to visual properties such as size, color, shape, scale patterns, etc., while, for 3D point cloud data, they typically result from specific geometric characteristic of the global or local 3D structure [18]. Typically, in 3D data, surface normals, gradients and curvature in the neighborhood of a point are used.

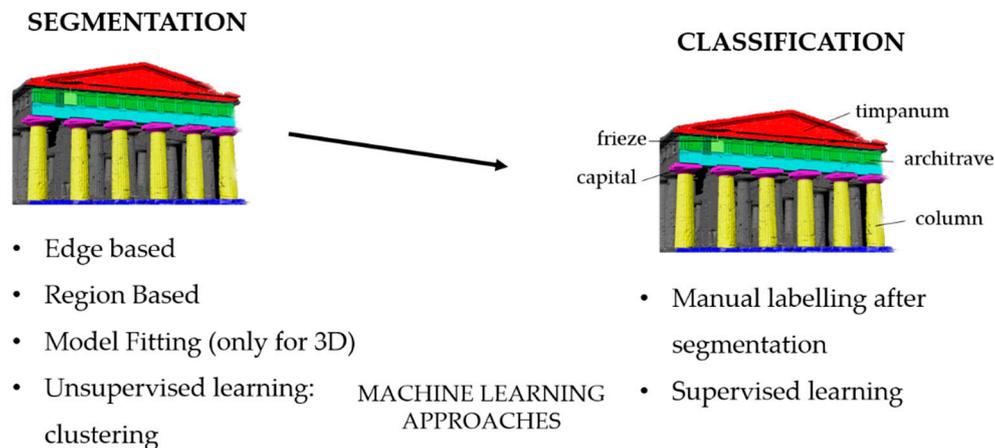


**Figure 2.** An example of image segmentation [19].

Once 2D or 3D scenarios have been segmented, each group can be labeled with a class giving the parts some semantics, hence classification is often called semantic segmentation or pixel/point labeling.

### 2.1. Segmentation Methods

The image segmentation topic has been widely explored [20] and current state-of-the-art techniques include edge-based [21,22] and region-based approaches [23] and clustering technique [24–27] (Figure 3).



**Figure 3.** Synthetic representation of the segmentation and classification methods. The 3D model represents the classification of the archaeological elements of the Neptune temple in Paestum [26].

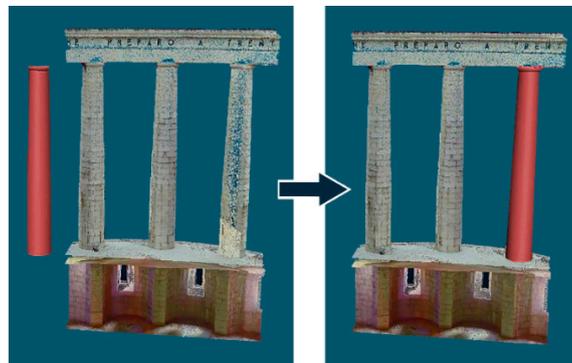
Except for the model fitting approach, most of the point cloud data segmentation methods have some root to image segmentation. However, due to the complexity and variety of point clouds caused by irregular sampling, varying density, different types of objects, etc.; point cloud segmentation and classification are more challenging and still very active research topics.

Edge-based segmentation methods have two main stages [27]: (i) edge detection to outlines the borders of different regions; and [28]: (ii) grouping of points inside the boundaries to deliver the final segments. Edges in a given depth map are defined by the points where changes in the local surface properties exceed a given threshold. The most used local surface properties are normals, gradients, principal curvatures or higher-order derivatives. Methods based on edge-based segmentation techniques were reported by Bhanu et al. [29], Sappa and Devy [30], and Wani and Arabnia [31]. Although such methods allow a fast segmentation, they may produce inaccurate results due to noise and uneven density of point clouds, situations that commonly occur in point cloud data. In 3D space, such methods often detect disconnected edges making the identification of closed segments difficult without a filling or interpretation procedure [32].

Region-based methods work with region growing algorithms. In this case, the segmentation starts from one or more points (seed points) featuring specific characteristics and then grows around neighboring points with similar characteristics, such as surface orientation, curvature, etc. [27,33]. The initial algorithm was introduced by Besl et al. [34], and several variations are presented in the literature [35–40]. In general, the region growing methods are more robust to noise than the edge-based ones because they utilize global information [41]. However, these methods are sensitive to: (i) the location of initial seed regions; and (ii) inaccurate estimations of the normals and curvatures of points near region boundaries.

The model fitting approach is based on the observation that many man-made objects can be decomposed into geometric primitives such as planes, cylinders and spheres (Figure 4). Therefore, primitive shapes (such as cylinders, cubes, spheres, etc.) are fitted onto 3D data and the points that best fit the mathematical representation of the fitted shape are labeled as one segment. As part of the model fitting category, two widely employed algorithms are the Hough Transform (HT) [42] and the Random Sample Consensus (RANSAC) [43]. Model fitting methods are fast and robust with outliers, although

a set of dedicated primitives is necessary. As it falls short for complex shapes or fully automated implementations, the use of the richness of surface geometry through local descriptors provides a better solution [44]. In the architectural field, details cannot always be modeled into easily recognizable geometric shapes. Thus, while some entities can be characterized by geometric properties, others are more readily distinguished by their color content [45].



**Figure 4.** Segmentation of 3D point cloud by geometric primitive fitting.

Machine learning approaches are described in detail in the next sections. It should be noted that, while clustering algorithms (unsupervised technique) belong to segmentation methods, after applying supervised machine learning methods, the 3D data are not only segmented but also classified.

Several benchmarks have been proposed in the research community, providing labeled terrestrial and airborne data on which users can test and validate their own algorithms. However, most of the available datasets provide classified natural, urban and street scenes [46–52]. 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 complex (several classes could be used to identify and describe the same building characteristics, based on different purposes). For these reasons, to the authors' knowledge, in the literature, there is currently a lack of benchmarks of labeled images or 3D heritage data.

## 2.2. Machine Learning for Data Segmentation and Classification

2D and 3D data classification are gaining interest and becoming an active field of research connected with machine learning application ([53–57] and [58–61], respectively).

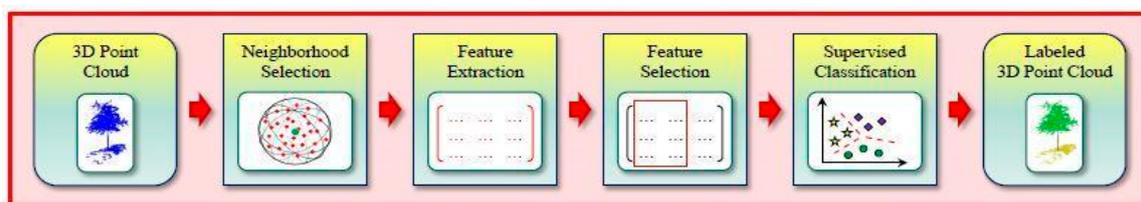
Machine learning (including deep learning) is a scientific discipline concerned with the design and development of Artificial Intelligence algorithms that allow computers to make decisions based on empirical and training data. Broadly, there are three types of approach with machine learning algorithms:

- A supervised approach is where semantic categories are learned from a dataset of annotated data and the trained model is used to provide a semantic classification of the entire dataset. If for the aforementioned methods the classification is a step after the segmentation, when using supervised machine learning methods, the class labeling procedure is planned before to segment the model. Random forest [62], described in detail at Section 4.2, is one of the most used supervised learning algorithms for classification problem [63,64].
- An unsupervised approach is where the data are automatically partitioned into segments based on a user-provided parameterization of the algorithm. No annotations are requested but the outcome might not be aligned with the user's intention. Clustering is a type of unsupervised machine learning that aims to find homogeneous subgroups such that objects in the same group (clusters) are more similar to each other than those in other groups. K-Means is a clustering algorithm that divides observations into  $k$  clusters using features. Since we can dictate the number of clusters,

it can be easily used in classification where we divide data into clusters that can be equal to or more than the number of classes. The original K-means algorithm presented by MacQueen et al. [65] has been then largely exploited for image and point clouds by various researchers [66–70].

- An interactive approach is where the user is actively involved in the segmentation/classification loop by guiding the extraction of segments via feedback signals. This requires a large effort from the user side but it could adapt and improve the segmentation result based on the user's feedback.

Feature extraction is a prerequisite for image/cloud segmentation. Features play an important role in these problems and their definition is one of the bottlenecks of machine learning methods [71,72]. Weinmann et al. [73] discussed the suitability of features that should privilege quality over quantity (Figure 5). High quality features allow better interpreting the models and enhancing algorithm performance with respect to both the speed and accuracy. This shows a need to prioritize and find robust and relevant features to address the heterogeneity in images or point clouds.



**Figure 5.** Framework for 3D scene analysis: a 3D point serves as input and the output consists of a semantically labeled 3D point cloud [18].

In both 2D and 3D segmentation/classification, approaches can be combined to exploit the strength of a method and bypass the weakness of others [41,74]. The success of these “hybrid methods” depends on the success of the underlying approaches.

### 3. Segmentation and Classification in Cultural Heritage

In the field of cultural heritage, processes such as segmentation and classification can be applied at different scales, from entire archaeological sites and landscapes to small artifacts.

In the literature, different solutions are presented for the classification of architectural images, using different techniques such as pattern detection [75], Gabor filters and support vector machine [76], K-means algorithms [77], clustering and learning of local features [78], hierarchical sparse coding of blocks [79] or CNN deep learning [16,80].

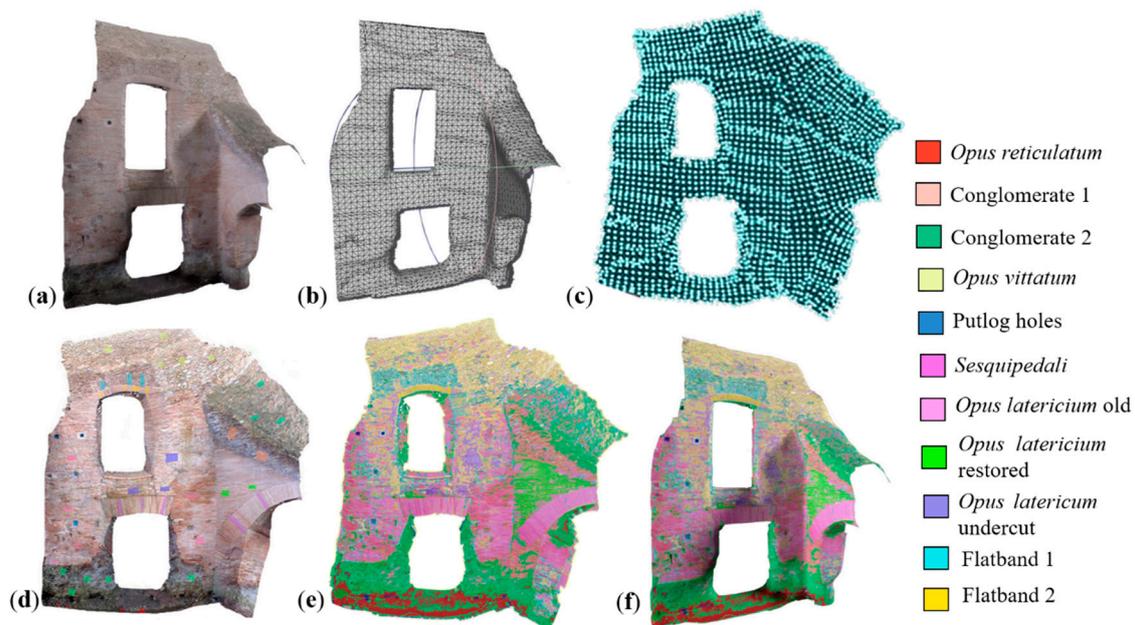
Many experiments were also carried out on 3D data at different scales [6,81,82]. Some works aim to define a procedure for the integration of architectural 3D models within BIM [1,5,83]. In many others, the classification is conducted manually for annotation purposes ([www.aioli.cloud](http://www.aioli.cloud)). In the NUBES project, for example, 3D models are generated from 2D annotated images. In particular, the NUBES web platform [84] allows the displaying and cross-referencing of 2D mapping data on the 3D model in real time, by means of structured 2D layer, such as annotations concerning stone degradation, dating and material. Apollonio et al. [85] used 3D models and data mapping on 3D surfaces in the context of the restoration documentation of Neptune's Fountain in Bologna. Campanaro et al. [86] realized a 3D management system for heritage structures by exploiting the combination of 3D visualization and GIS analysis. The 3D model of the building was originally split into architectural sub-elements (facades) to add color information projecting orthoimages by means of planar mapping techniques (texture mapping). Sithole [87] proposed an automatic segmentation method for detecting bricks in masonry walls, working on the point clouds and assuming that mortar channels are reasonably deep and wide. Oses et al. [76] used machine learning classifiers, support vector machines and classification trees for the masonry classification. Riveiro et al. [88] suggested an algorithm for the segmentation of bricks in point cloud built on a 2.5D approach and creating images based on the intensity attribute of LiDAR sensors. Recently, Messaoudi et al. [89] developed a correlation pipeline for the integration of

semantic, spatial and morphological dimension of a built heritage. The annotation process provides a 3D point-based representation of each 2D region.

#### 4. Project's Methodology

Considering the availability and reliability of segmentation methods applied to (2D) images and the efficiency of machine learning strategies, a new methodology was developed to assist cultural heritage experts analyze digital 3D data. In particular, the approach presented hereafter relies on supervised and unsupervised machine learning methods for segmenting texture information of 3D digital models. Starting from colored 3D point clouds or textured surface models, our pipeline relies on the following steps:

- Create and optimize models, orthoimages (for 2.5D geometries) and UV maps (for 3D geometries) (Figure 6a–c).
- Segment the orthoimage or the UV map following different approaches tailored to the case study (clustering, random forest) (Figure 6d–e).
- Project the 2D classification results onto the 3D object space by back-projection and collinearity model (Figure 6f).



**Figure 6.** Schematic representation of the developed segmentation and classification methodology: 3D model of a portion of Circus Maximus Cavea in Rome, Italy, (a); 3D model after re-meshing (b); UV map (c); manually identified training areas on the unwrapped texture (d); supervised classification results (e); and re-projection of the classification results onto the 3D model (f).

##### 4.1. Image Preparation

The method works on the texture information of a 3D model. The texture is prepared according to the geometry and complexity of the considered 3D object:

- Planar objects (e.g. walls, Section 5.1): The object orthophoto is created and the procedure classifies and finally re-maps the information onto the 3D geometry.
- Regular objects (e.g. building or other 3D structures with certain level of complexity fit into this category, Section 5.2): Instead of creating various orthoimages from different points of view, unwrapped texture (UV maps) are generated and classified. To generate a good texture image to be classified, we followed these steps:

1. Remeshing: Beneficial to improve the quality of the mesh and to simplify the next steps.
  2. Unwrapping: UV maps are generated using Blender, adjusting and optimizing seam lines and overlap (Figure 6c) to facilitate the subsequent analysis with machine learning strategies. This correction is made commanding the UV unwrapper to cut the mesh along edges chosen in accordance with the shape of the case study [90].
  3. Texture mapping: The created UV map is then textured (Figure 6d) using the original textured polygonal model (as vertex color or with external texture). This way the radiometric quality is not compromised despite the remeshing phase.
- Complex objects (e.g.; monuments or statue, Section 5.3): When objects are too complex for a good unwrap, the classification is done directly on the texture generates as output during the 3D modeling procedure.

When we consider color image segmentation, choosing a proper color space becomes an important issue [91]. This is because different color spaces present color information in different ways that make certain calculations more convenient and also provide a way to identify colors that is more intuitive. Several color representations are currently in use in color image processing. The most common is the RGB, but also HSV and  $L^*A^*B^*$  are frequently chosen color spaces [92,93]. In the RGB color space, for example, shadowed areas will most likely have very different characteristics than areas without shadows. In the HSV color space, the hue component of areas with and without shadow are more likely to be similar: the shadow will primarily influence the value, or the saturation component, while the hue—indicating the primary “color” without its brightness and diluted-ness by white/black—should not change so much. Another popular option is LAB color space—where the AB channels represent the color and Euclidean distances in AB space—better match the human perception of color. Again, ignoring the L channel (Luminance) makes the algorithm more robust to lighting differences.

#### 4.2. Supervised Learning Classification

The 2D classification method relies on different machine learning models embedded in WeKa [94] coupled with the Fiji distribution of ImageJ, an image processing software that exploits WeKa as an engine for machine learning models [95]. The method combines a collection of machine learning algorithms (random tree, support vector machine, random forest, etc.) with a set of selected image features to produce pixel-based segmentations. All the available classifiers are based on a decision tree learning method. In this approach, during the training, a set of decision nodes over the values of the input features (e.g. “feature x is greater than 0.7?”) are built and connected to each other in a tree structure.

This structure, as a whole, represents a complex decision process over the input features. The result of this decision is a value for the label that classifies the input example. During the training phase, the algorithm learns these decision nodes and connects them.

Among the different approaches, we achieved the best results in terms of accuracy exploiting the random forest method (Section 5.1) [62]. In this approach, several decision trees are trained as an ensemble, with the mode of all the predictions that is taken as the final one. This allows us to overcome some typical problems in decision tree learning, such as overfitting the training data and learning uncommon irregular patterns that may occur in the training set. This behavior is mitigated by the random forest procedure by randomly selecting different subsets of the training set and, for each of these subsets, a random subset of input features. At the same time, for each of these subsets of training examples and features, a decision tree is learned. The main intuition between these procedures is called “feature bagging”, where some features are very strong predictors for the output class. Such features will be likely to be selected in many of the trees, causing them to become correlated.

For each case study, the random forest was trained giving in input the manually annotated model’s textures. More specifically, not all the pixels of those images were manually annotated with its corresponding label, but just some significant and well distributed portions (e.g., see Figure 8b).

The first time the training process starts, the features of the input image are extracted and converted to a set of vectors of float values (Weka input). This step can take some time depending on the size of the images, the number of features and the number of cores of the machine where the classification is running. The feature calculation is done in a completely multi-thread fashion. The features are calculated only the first time it is trained after starting the plugin or after changing any of the feature options. In the case of color (RGB) images, the hue, saturation and brightness are also part of the features.

#### 4.3. Unsupervised Learning Classification

The unsupervised segmentation approach is performed using the k-means clustering plugin of ImageJ or Fiji [96]. The algorithm performs pixel-based segmentation of multi-band images. Each pixel in the input image is assigned to one of the clusters. Values in the output image represent the cluster number to which original pixel is assigned. Before starting the elaboration, the operator decides the number K of classes the image will be divided into and the cluster center tolerance.

#### 4.4. Evaluation Method

To assess the performance of the classification, small sections of the entire datasets were trained and then compared quantitatively with the ground truths. More specifically, we relied on the precision, recall and F1 score calculated for each class—computed for each point by comparing the label predicted by the classifier with the same manually annotated—and on the overall accuracy, which is useful to evaluate the overall performance of the classifier.

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

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

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

$$\text{Overall accuracy} = \frac{\text{number of correct predictions}}{\text{total number of predictions}} \quad (4)$$

where, for each considered class, (true positive),  $Tn$  (true negative),  $Fp$  (false positive), and  $Fn$  (false negative) come from the confusion matrix, which is commonly used to evaluate machine learning classifiers.

Once considerable levels of overall accuracy were reached (>70%), the classification was extended to the whole datasets, which were evaluated qualitatively. In this way, the performance of the model was assessed measuring the performance against another set of images, different from the ones used during the training phase, so that the capabilities of the model to generalize over unseen data could be effectively measured.

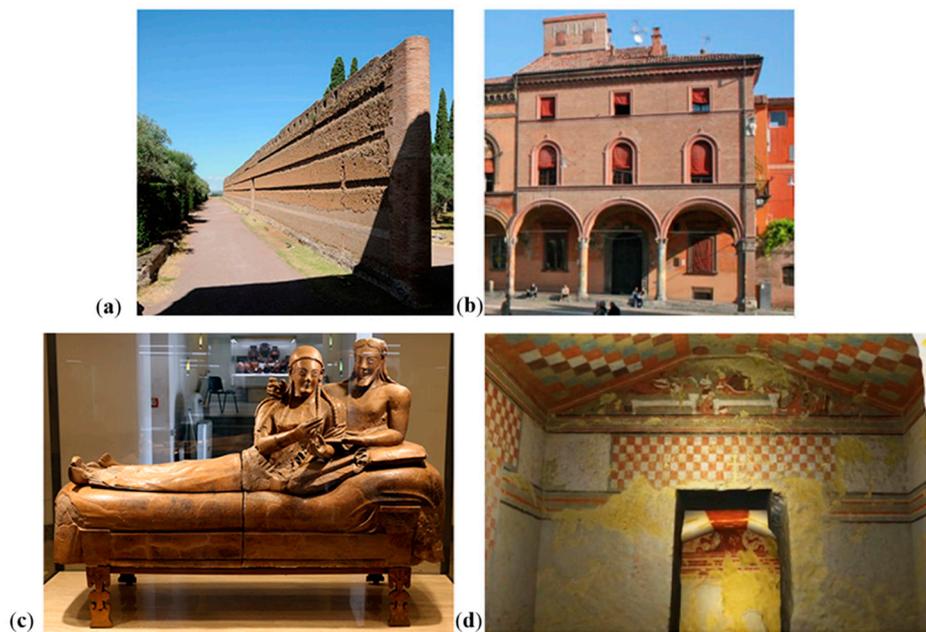
## 5. Test Objects and Classification Results

The proposed methodology was applied to and tested on various archaeological and architectural scenarios to prove its replicability and reliability with different 3D objects.

In particular, these case studies were considered:

- The Pecile's wall of Villa Adriana in Tivoli: it is a 60 m L × 9 m H wall (Figure 7a) with holes on its top meant for the beams of a roof. The digital model of the wall was classified to identify the different categories of opus (roman building techniques), distinguishing original and restored parts.
- Part of a renaissance portico located in the city center of Bologna: It spans ca. 8 m L × 13 m H × 5 m D (Figure 7b). The classification aimed to identify principal parts and architectural elements;

- The Sarcophagus of the Spouses (Figure 7c): It is a late 6th century BC Etruscan anthropoid sarcophagus, 1.14 m high by 1.9 m wide, made of terracotta, which was once brightly painted [97]. The classification aimed at identifying surface anomalies (such as fractures and decays) and quantifying the percentage of mimetic cement used to assemble the sarcophagus.
- The Bartoccini's Tomb in Tarquinia (Figure 7d): the tomb, excavated in the hard sand in the 4th century, has four rooms—a central one (ca. 5 m × 4 m) and three later rooms (ca. 3 m × 3 m)—all connected through small corridors. The height of the tomb rooms does not exceed 3 m and it is all painted with a reddish color and various figures. The aim was to automatically identify the still painted areas on the wall and the deteriorated parts.

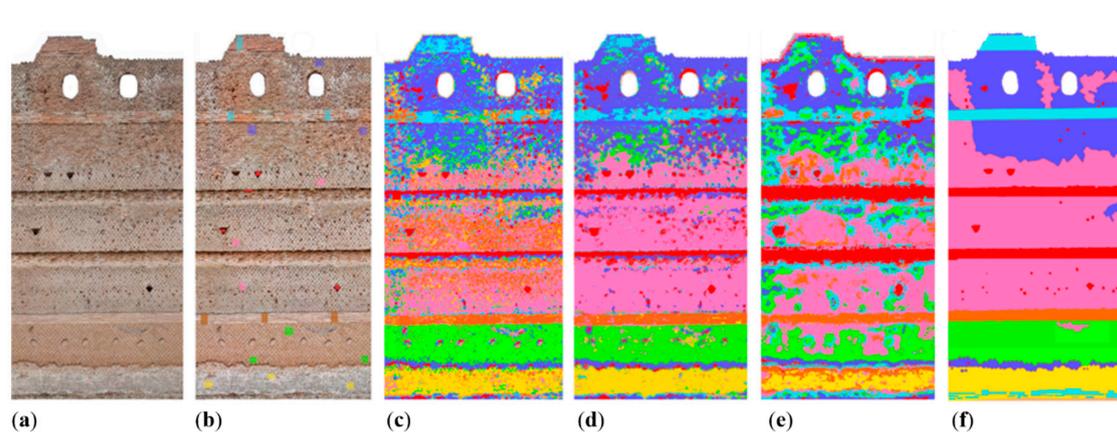


**Figure 7.** The case studies of the work to validate the semantic classification for analyses and restoration purposes: Pecile's wall of Villa Adriana in Tivoli, Italy (a); Renaissance building in Bologna, Italy (b); Sarcophagus of the Spouses, National Etruscan museum of Villa Giulia in Rome, Italy (c); Etruscan tomb in Tarquinia, Italy (d).

For the porticoes and the Bartoccini's tomb, a supervised segmentation approach directly on the 3D models was also applied.

### 5.1. The Pecile's Wall

To define the correctness of the developed approach, different analyses were conducted on this first case study. Only a portion of the Pecile's wall (4 m length × 9 m height) was considered at first (Figure 8). On this portion of the wall's orthoimage, different training processes were run using different image scales to identify the solution that best fit this case study, taking into account the classification aims.



**Figure 8.** Orthoimage of a portion of Pecile’s wall (4 m length  $\times$  9 m height) exported at 1:20 scale (a); corresponding training samples in the image (b); and classification results obtained at different scales: scale 1:10 (c); scale 1:20 (d); scale 1:50 (e); and ground truth (f).

At a 1:10 scale, the results present an over segmentation. Using a 1:50 scale, many details were lost, identifying only some macro-areas. The scale 1:20 (normally used for restoration purposes as it allows to distinguish bricks) turned out to be the optimal choice. It allowed the capture of the details but could not consider the cracks of the mortar between the bricks (Figure 8d). Given the manually selected training classes (seven classes), different classifiers were trained and evaluated. Table 1 reports the overall accuracy results for all tested classifiers run on the orthoimage at scale 1:20.

**Table 1.** Accuracy results and elapsed time for various classifier applied to an orthoimage at 1:20 scale.

Classifier	Overall Accuracy	Time
j48	0.44	22 s
Random Tree	0.46	15 s
RepTREE	0.47	33 s
LogitBoost	0.52	20 s
Random Forest	0.57	23 s
Fast Random Forest	0.70	120 s

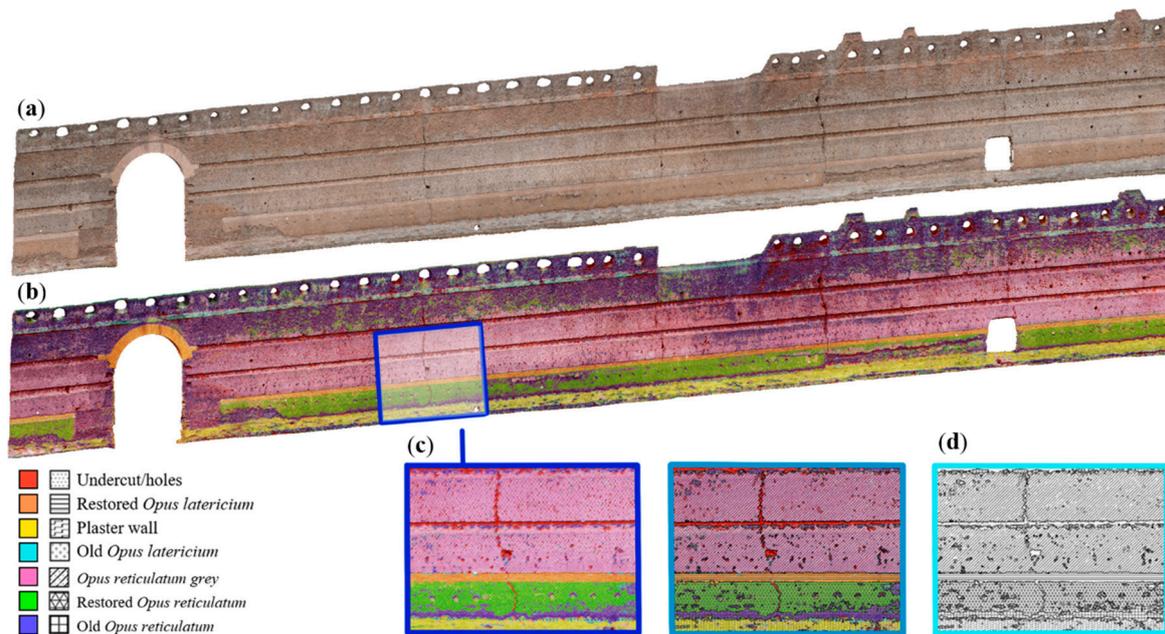
Moreover, we report the time elapsed for each algorithm, considering that creating the classes and the training data took around 10 min and the feature stack array required 14 min.

Out of all the tests performed with the different algorithms, the best overall accuracy obtained was 70% using a random forest classifier. To better identify the classification errors, a confusion matrix was used (Table 2). From the table analysis, it was possible to understand that most errors in classification were in those classes where an overlap of plaster was present on the surface of the opus. However, it is believed that an expert should not consider the accuracy percentage absolute without previous verification. Comparing the segmentation handled by the operator and by the algorithm, it was found that the supervised method allowed the identification of more details and differences in the material’s composition. In fact, it could not only distinguish the classes, but also identify the presence of plaster above the wall surface. This is an important advantage for the degradation analysis.

**Table 2.** Normalized confusion matrix to analyze the results of the supervised classification of a portion of Pecile’s wall at scale 1:20. On the left of the table are reported precision, recall and F1 calculated for each category.

TRUE LABEL								Precision	Recall	F1
	Undercut/holes	Restored <i>Opus latericium</i>	Plaster wall	Old <i>Opus latericium</i>	<i>Opus reticulatum</i> grey	Restored <i>Opus reticulatum</i>	Old <i>Opus reticulatum</i>			
Undercut/holes	0.64	0.01	0	0.07	0.08	0.02	0.13	0.56	0.67	0.6
Restored <i>Opus latericium</i>	0.2	0.73	0.02	0.02	0.18	0.01	0	0.85	0.63	0.72
Plaster wall	0.02	0.03	0.75	0.04	0.12	0.02	0.03	0.91	0.74	0.82
Old <i>Opus latericium</i>	0.08	0.02	0.01	0.43	0.2	0.05	0.21	0.56	0.43	0.49
<i>Opus reticulatum</i> grey	0.06	0.05	0.01	0.08	0.66	0.05	0.08	0.47	0.67	0.55
Restored <i>Opus reticulatum</i>	0.02	0.01	0.01	0.01	0.07	0.83	0.06	0.77	0.82	0.8
Old <i>Opus reticulatum</i>	0.15	0.01	0.02	0.12	0.08	0.09	0.52	0.5	0.52	0.51

Starting from this result the training dataset was applied to a larger part of the wall (Figure 9b). To classify 540 m<sup>2</sup> of surface the process took about 1 h. Considering that the operator took 4 h just for classifying a smaller part (24 m<sup>2</sup>), the supervised technique could obtain a more accurate result in a shorter time. The classification results can also be used to create the most commonly requested map for restoration purpose, with dedicated symbols/legend (Figure 9d).



**Figure 9.** The original (a); and classified (b) model of the Pecile’s wall long ca 60 m. A closer view is also reported to better show the classification results with random colors (c); or dedicated symbols (d).

### 5.2. Bologna’s Porticoes

The historical porticoes of Bologna were built during the 11th–20th centuries and can be regarded as unique from an architectural viewpoint in terms of their authenticity and integrity. Thanks to their great extension, permanence, use and history, the porticoes of Bologna are considered of outstanding universal value. They span approximately 40 km, mainly in the historic city center of Bologna, and they represent a high-quality architectural work.

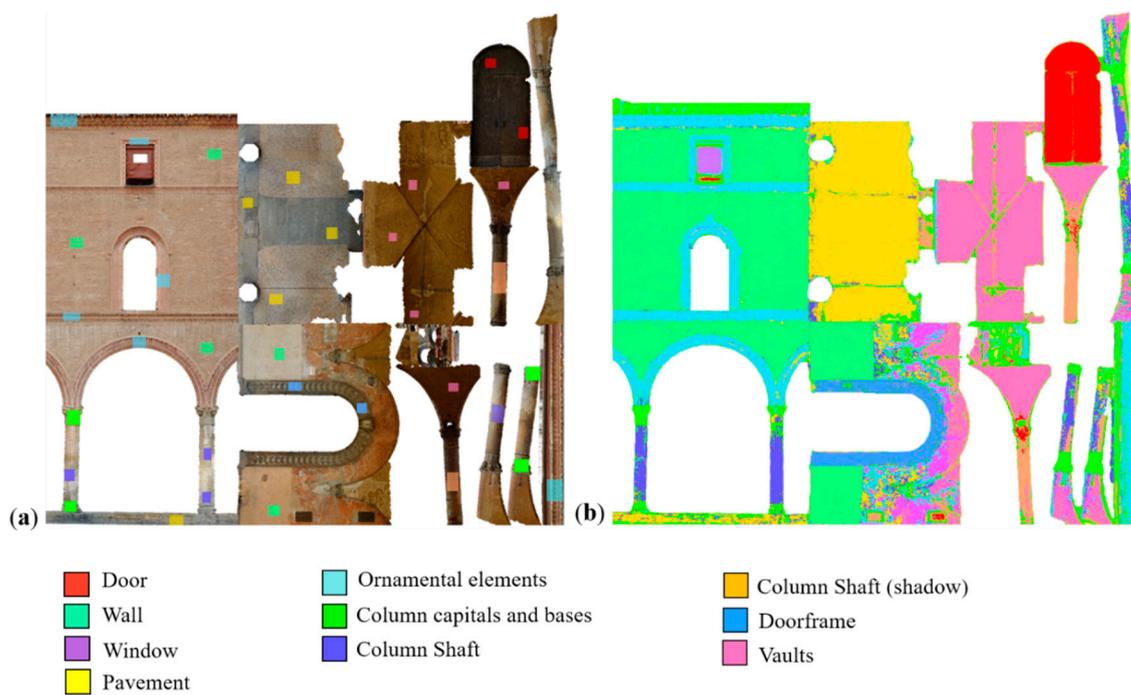
Such structures combine variegated geometric shapes, different materials and many architectural details such as moldings and ornaments. According to the different classification requirements, the aim of the task could be the identification of:

- different architectural elements;
- diverse materials (bricks vs. stones vs. marble); and

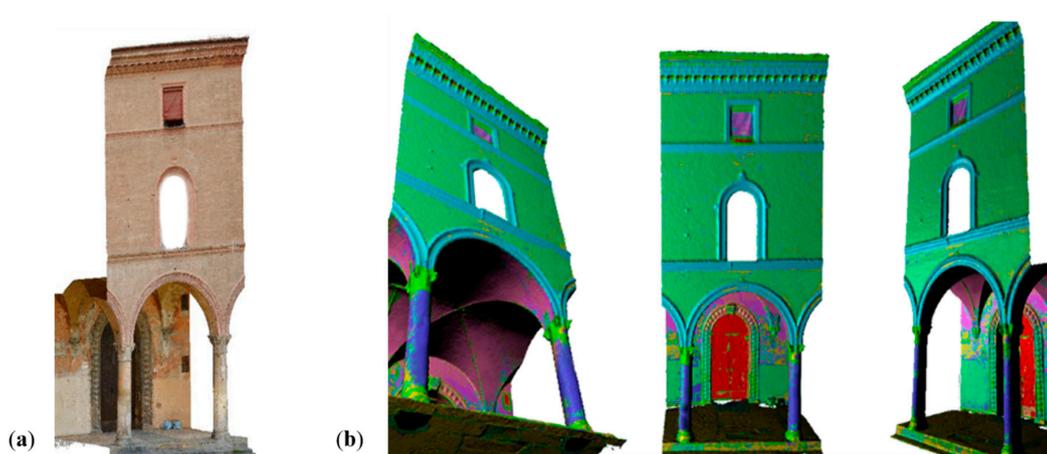
- categories of decay (cracks vs. humidity vs. swelling).

### 5.2.1. Classification of 2D Data

Starting from the available 3D data of the porticoes [98], the texture of the 3D digital porticoes was unwrapped (Figure 10) and used to manually identify training patches and classes (10). The results (Figure 11), based on fast random forest model/classifier, show many classification errors under the porticoes, where the plaster is not homogeneous and presents different decays. In this case, a solution might be to create many different classes according to the number of decay categories or to apply, as a post processing phase, algorithms to make more uniform the areas with small spots.



**Figure 10.** Manually identified training areas on the unwrapped texture of the porticoes portion (a); and classification results based on the selected 10 classes (b).

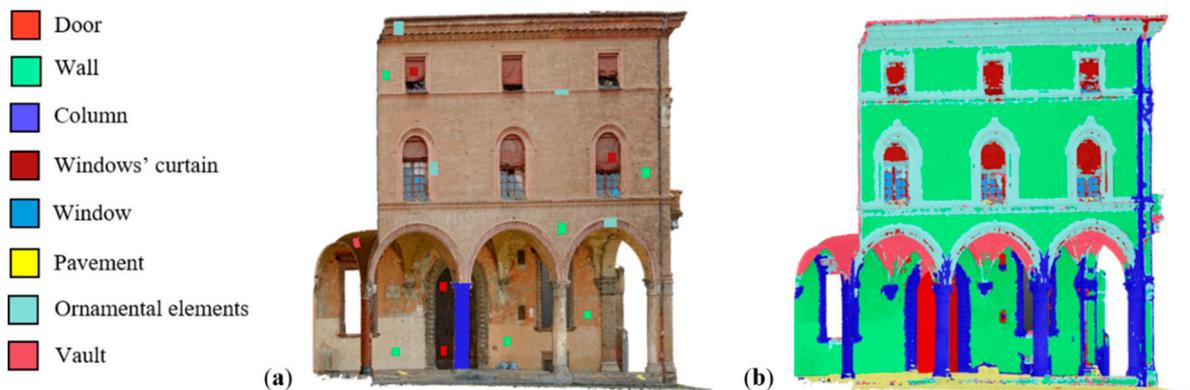


**Figure 11.** 3D model (a); and from 2D to 3D classification results of historical porticoes in Bologna (b).

### 5.2.2. Classification of 3D Data

Considering the classification errors of the facades, due to the heterogeneous surfaces and the presence of different decays, a supervised classification approach was performed based on the

Computational Geometry Algorithms Library (CGAL) [99] and the Random Forest Template Library (ETHZ Random Forest, 2018) [100]. The classification consists of three steps: feature computation, model training and prediction. To characterize the model, a combination of features considering both geometry and color factors was selected: distance to plane, eigenvalues of the neighborhood, elevation, verticality and HSV channels. The training dataset of correct labels was manually annotated on small portions of the entire 3D model. To conclude that phase, a classifier was defined and trained using random forest: from the set of values taken by the features at an input item, it measured the likelihood of this item belonging to one label or another. Finally, once the classifier was generated, the prediction process was performed on the entire point cloud (Figure 12).



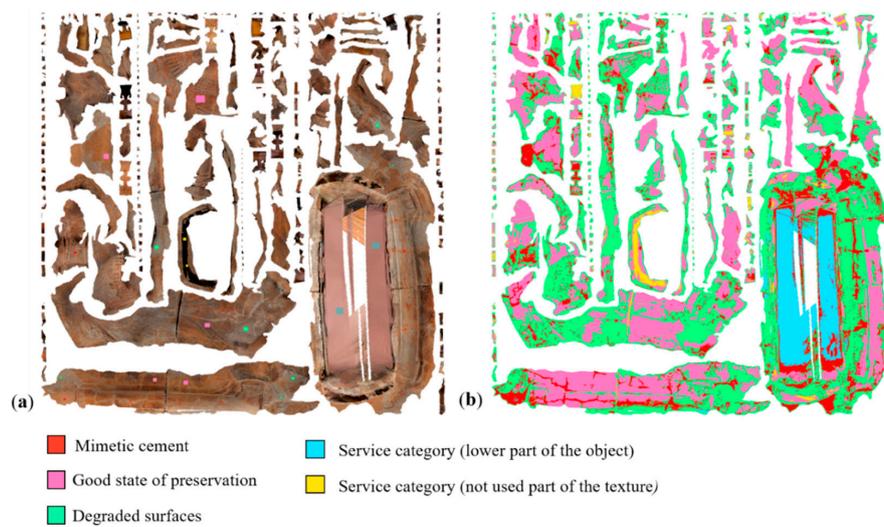
**Figure 12.** 3D model (a); and classification results of historical porticoes in Bologna (b).

As shown in the figure, the classification problems on the walls were solved, but there were some false positive due to the classification of drain pipes as columns (vertical pipes) and vaults (horizontal pipes). However, it was not considered a classification error, but an error in the training phase, as a class was not assigned to these features.

Considering the results obtained working onto the 3D models and the suitability of the case study (40 km of similar buildings), as a future works the authors aim to extend the classification to a bigger portion of the porticoes of Bologna.

### 5.3. The Etruscan Sarcophagus of the Spouses

The Etruscan masterpiece “Sarcophago degli Sposi” was found in 1881 in the Banditaccia necropolis in Tarquinia (Italy). The remains were found broken into more than 400 pieces. The sarcophagus was then reassembled and joined using a mimetic cement to fill the gaps among the different pieces. In 2013, digital acquisitions and 3D modeling of the sarcophagus, based on different technologies (photogrammetry, TOF and triangulation-based laser scanning) were conducted to deliver highly detailed photo-realistic 3D representation of the Etruscan masterpiece for successive multimedia purposes [101]. Using the high-resolution photogrammetric 3D model (5 million triangles), the segmentation task aimed to detect the surface anomalies (fractures and decays) and to test the reliability of the method on a heritage objects with a more complex topology and few chromatic differentiations on the texture. For the training set, three main categories, and two accessories ones, were identified (Figure 13a).



**Figure 13.** Manually identified training areas on the unwrapped texture of the sarcophagus (a); and related classification result on the texture (b).

The manual identification of the necessary patches took about 15 min and was accomplished with the support of restoration experts. The sustaining legs of the sarcophagus were excluded from the classification, as they are the only parts where pigment decorations are clearly visible, thus their analysis was outside the segmentation scope. After the patch identification, the model took some 2 h of processing to classify the entire texture (Figure 13b), which was then mapped onto the available 3D geometry (Figures 14 and 15). The segmented 3D model highlighted every single detail of the masterpiece assembly; fractures were distinguished from engraving; and the different grades of conservation were also identified. The classification output also allowed calculating the percentage that each label occupied. From the results, we have that the 12% of the entire surface of the object (i.e. 3D model) is composed by mimetic cement. As the overall surface of the 3D model is 6.8 m<sup>2</sup>, it means that approximately 0.8 m<sup>2</sup> are reconstructed parts.



**Figure 14.** Texturized (a) and segmented (b) 3D model of the Sarcophagus of the Spouses.



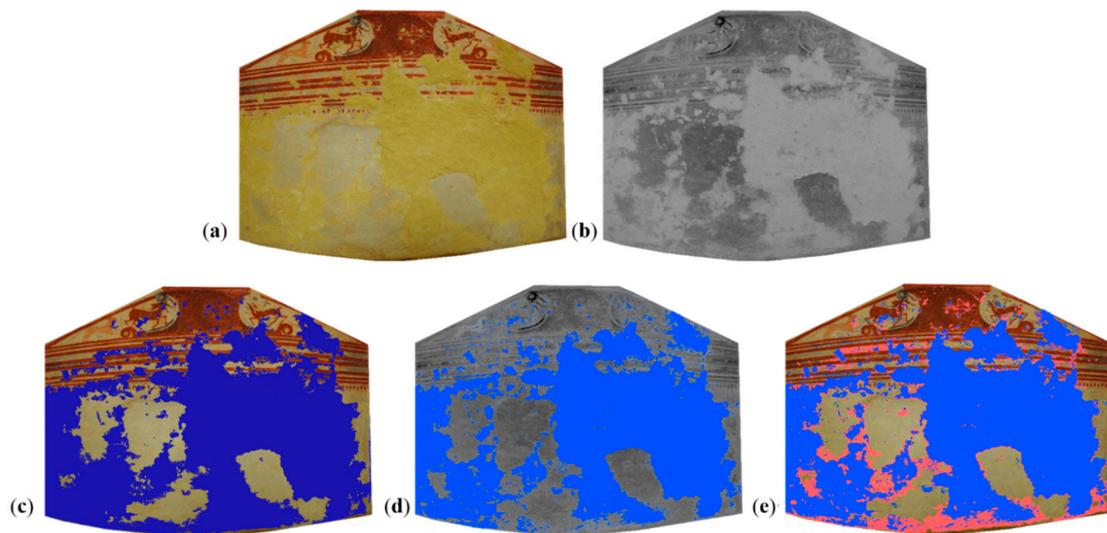
**Figure 15.** Mimetic cement areas of the sarcophagus of the Spouses highlighted in red (ca. 12% of the model).

#### 5.4. The Etruscan Bartoccini's Tomb in Tarquinia, Italy

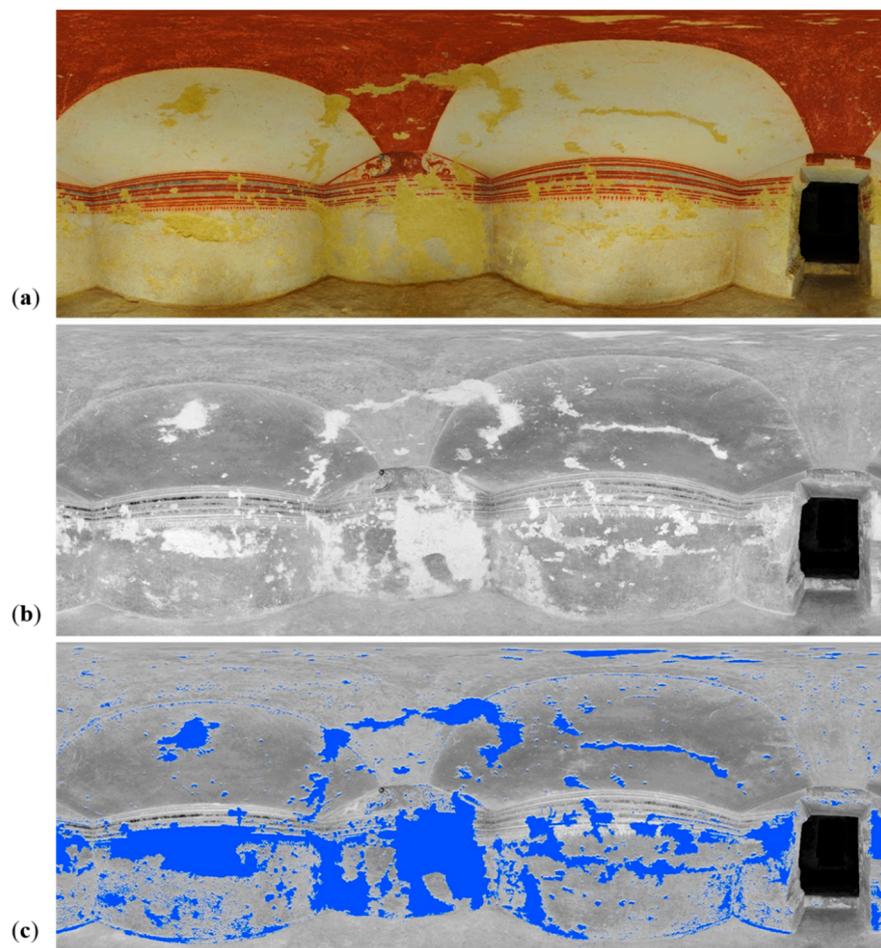
Tarquinia was one of the most ancient cities of the Etruscan civilization. The necropolis, situated in the areas of Monterozzi and Calvario, is composed of some 6000 tombs, 60 of which are decorated with paintings. The Bartoccini tomb, dated to around the 4th century B.C.; was discovered in 1959. Combined TOF scanning and panoramic photographic surveys were carried out to obtain the complete 3D model (3 million triangles) [102]. The TOF range data were used to derive the geometry of the tomb, while the panoramic image was used to derive the photo realistic high-resolution texture. In-house developed algorithms were used to project the panoramic image onto the 3D geometry and then extract a high-resolution texture. As over the centuries, the tomb has suffered from erosion caused by various reasons such as infiltration, seasoning, aging, etc. The aim of the segmentation was the automated identification of the deteriorated surfaces' area on the painted walls. To reach this goal, different strategies were tested, on both 2D (Section 5.4.1) and 3D (Section 5.4.2) data. Quantification analyses are presented in Section 5.4.3.

##### 5.4.1. Classification of 2D Data

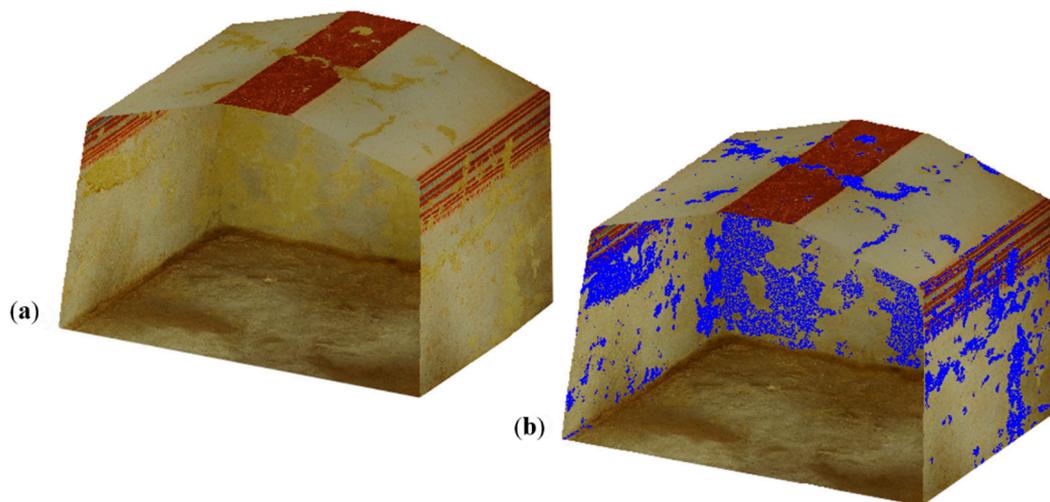
Given the texture of the tomb, a clustering process was chosen instead of manually training various classes. K-means clustering (Section 4.3) was performed to generate a pixel-based segmentation of the panoramic images. To optimize the work time in a trial phase, only one wall was analyzed, was then the segmentation was extended to the whole model. To avoid segmentation errors, and thus achieve better results, the image was transformed from RGB to Lab\* color space (Figure 16b). The obtained results (Figure 16d) were compared with the ground truth (manually segmented) data (Figure 16c), achieving an overall accuracy of 91.15%. The clustering method was then applied to the entire panoramic texture of the tomb's room (Figure 17) and finally mapped onto the 3D geometry (Figure 18).



**Figure 16.** Texture of a wall of the Bartoccini's tomb with RGB (a) and Lab\* colors (b); ground truth of eroded areas (c); clustering results (d); and marking (pink color) of not identified eroded areas by the automatic clustering method (e).



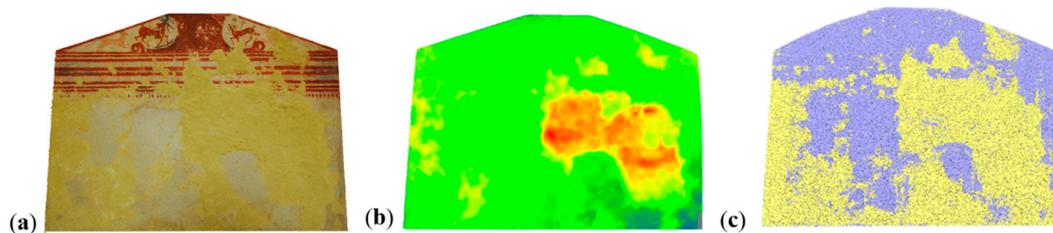
**Figure 17.** Panoramic image of one room of the tomb (a); Lab\* color space (b); and automatically identified eroded areas by clustering segmentation (c).



**Figure 18.** Section of the tomb (a); and visualization of the 2D segmentation results mapped on the 3D model (b).

5.4.2. Classification of 3D data

From the 3D geometry of a tomb’s wall, a plane fitting procedure allowed extracting a depth map of the wall and identifying the eroded surfaces (Figure 19b), i.e. those below the ideal fitted surface. Although only the damaged areas below a certain depth variation threshold were identified, the volume of the eroded wall could still be calculated (Table 3).

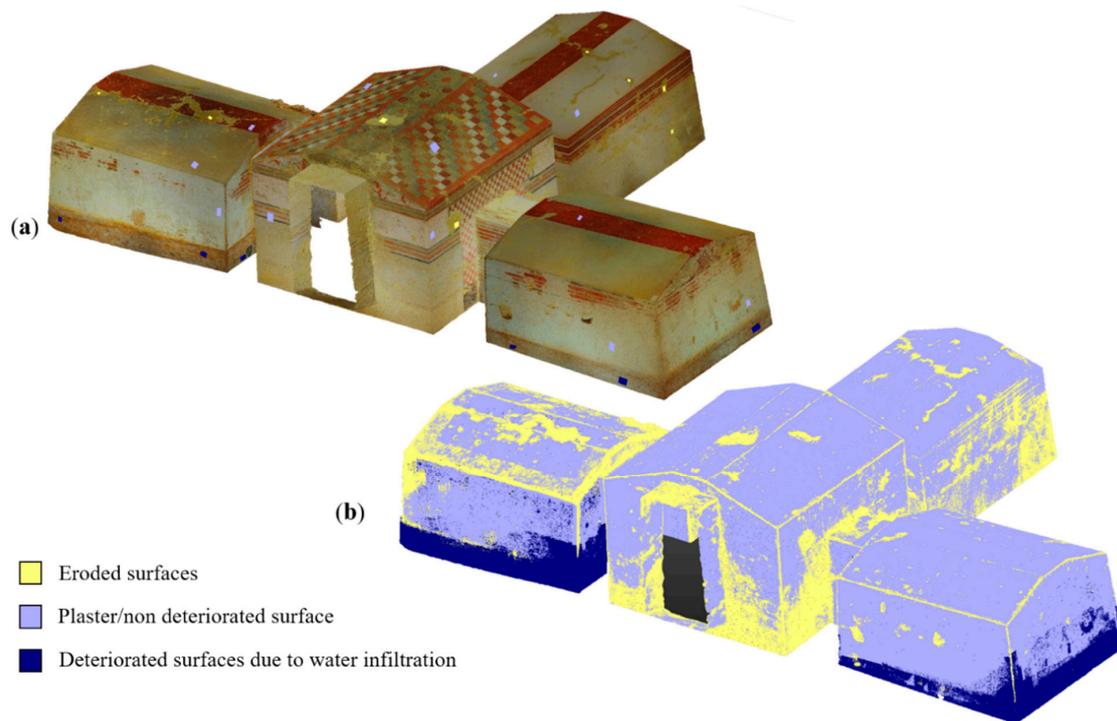


**Figure 19.** A wall of the Bartoccini’s tomb (a); depth map (b); and 3D supervised segmentation results (c).

**Table 3.** Computed eroded areas (and volumes) with the aforementioned approaches.

Classifier		% Eroded Surfaces	Area/Volume	Time for Elaborations
2D Manual classification		58%	2,8 m <sup>2</sup>	30 min
2D Unsupervised clustering		49%	2,4 m <sup>2</sup>	5 min
3D Supervised segmentation		56%	2,7 m <sup>2</sup>	10 min
Depth map		40%	1,96 m <sup>2</sup> / 0,1m <sup>3</sup>	10 min

Consequently, as for the porticoes case study, a supervised classification approach was achieved. The correct labels (three classes) were annotated in small, well-distributed portions, and then the prediction process was performed on the entire 3D model (Figure 20).



**Figure 20.** Texturized (a); and classified (b) 3D model of the Bartoccini's Tomb in Tarquinia.

#### 5.4.3. Quantification Analyses

For all the aforementioned experiments, the identified eroded surfaces were estimated as a percentage of the total area of interest. In particular, for the approaches based on 2D images, the percentage was calculated as a comparison between the number of pixels classified as eroded and the total number of pixels in the segmented image. On the other hand, in the case 3D data, the percentage resulted as the number of points belonging to the deteriorated areas over the total number of 3D points. Table 3 reports the eroded areas computed for one wall of the tomb (4.9 m<sup>2</sup>), with both percentages and square meter areas shown. This type of metric results could be beneficial for monitoring and restoration purposes.

## 6. Conclusions and Future Works

This paper presents a pipeline to classify 3D heritage data, either working on the texture or directly on the 3D geometry, depending on the needs and scope of the classification. With the proposed methods, archaeologists, restorers, conservator and scientists can automatically annotate 2D textures of heritage objects and visualize them onto 3D geometries for a better understanding. The existence of a vast variety of building techniques and the usage of diverse ornamental elements introduces an extra barrier in generalizing segmentation techniques to heritage case studies. Moreover, a monument can be subject to different types of degradation depending on its exposure under various conditions, hence increasing the efficiency of the classification tasks. A machine learning-based approach becomes beneficial for speeding up classification tasks on large and complex scenarios, provided that the training datasets are sufficiently large and diverse.

The advantages of the proposed method are:

- shorter time to classify objects with respect to manual methods (Table 1);
- over-segmentation results useful for restoration purposes to detect small cracks or deteriorated parts;
- replicability of the training set for buildings of the same historical period or with similar construction material (e.g. roman walls);

- visualization of classification results onto 3D models from different points of view, using unwrapped textures;
- possibility to compute absolute and relative areas of each class (Table 3), useful for analysis and restoration purposes; and
- applicability of the pipeline to different kinds of heritage buildings, monuments or any other kind of 3D models.

On the other hand, lesson learned and open critical issues can be summarized as:

- Difficult identification of the classes of analysis case by case (e.g. problems with the drainpipes classified as columns): the choice of the right classes during the training phase becomes fundamental.
- Misinterpretation of the shadows can introduce errors in the classification: the use of different color spaces from the classic RGB one, e.g. HSV and Lab, makes the lighting differences less problematic during the segmentation phase.
- Over-segmentation results in many classes, commonly useless in semantic analysis: this implies the need to make the regions more uniform in a post processing phase.

For future works, the authors aim to work across different heritage buildings, improving the generalization of the classification of some basic classes (e.g. windows, doors, and columns). To do that, it will be necessary to increase the number of labeled images and exploit more complex machine learning algorithms, in particular deep neural networks.

We will also tackle the objective of increasing the homogeneity of the segmentation to minimize and ideally avoid any post-processing phase.

Finally, we will work on new case studies, applying both techniques presented (from 2D to 3D and directly on 3D), to better understand the advantages of each method with respect to the other.

**Author Contributions:** E.G. designed the project objectives, performed the experiments and wrote the manuscript. F.R. supervised all the steps of the research, provided substantial insights and edits.

**Funding:** This research received no external funding.

**Acknowledgments:** The authors would like to acknowledge Accademia Adrianea and University of Bologna (Dept. Architecture) for providing the datasets of Pecile's wall in Villa Adriana and Bologna's porticoes. We are also thankful to Soprintendenza per i Beni Archeologici dell'Etruria Meridionale and the Etruscan National of Villa Giulia in Rome for the possibility to work on the Sarcophagus and Bartoccini's tomb case studies.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Saygi, G.; Remondino, F. Management of architectural heritage information in BIM and GIS: State-of-the-art and future perspectives. *Int. J. Herit. Digit. Era* **2013**, *2*, 695–713. [[CrossRef](#)]
2. Hackel, T.; Wegner, J.D.; Schindler, K. Fast semantic segmentation of 3D point clouds with strongly varying density. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *3*, 177–184. [[CrossRef](#)]
3. Weinmann, M.; Weinmann, M. Geospatial Computer Vision Based on Multi-Modal Data—How Valuable Is Shape Information for the Extraction of Semantic Information? *Remote Sens.* **2017**, *10*, 2. [[CrossRef](#)]
4. Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S.E.; Bronstein, M.M.; Solomon, J.M. Dynamic graph CNN for learning on point clouds. *arXiv*, 2018; arXiv:1801.07829.
5. Macher, H.; Landes, T.; Grussenmeyer, P. Point clouds segmentation as base for as-built BIM creation. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *2*, 191. [[CrossRef](#)]
6. Barsanti, S.G.; Guidi, G.; De Luca, L. Segmentation of 3D Models for Cultural Heritage Structural Analysis—Some Critical Issues. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *4*, 115. [[CrossRef](#)]
7. Maturana, D.; Scherer, S. Voxnet: A 3D convolutional neural network for real-time object recognition. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 922–928.

8. Wang, L.; Zhang, Y.; Wang, J. Map-Based Localization Method for Autonomous Vehicles Using 3D-LIDAR. *IFAC-PapersOnLine* **2017**, *50*, 276–281.
9. Xu, S.; Vosselman, G.; Oude Elberink, S. Multiple-entity based classification of airborne laser scanning data in urban areas. *ISPRS J. Photogramm. Remote Sens.* **2014**, *88*, 1–15. [[CrossRef](#)]
10. Corso, J.; Roca, J.; Buill, F. Geometric analysis on stone façades with terrestrial laser scanner technology. *Geosciences* **2017**, *7*, 103. [[CrossRef](#)]
11. Cheng, G.; Han, J. A survey on object detection in optical remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2016**, *117*, 11–28. [[CrossRef](#)]
12. Shen, D.; Wu, G.; Suk, H.I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [[CrossRef](#)]
13. Li, H.; Shen, C. Reading car license plates using deep convolutional neural networks and LSTMs. *arXiv*, 2016; arXiv:1601.05610.
14. Li, C.; Shirahama, K.; Grzegorzec, M. Application of content-based image analysis to environmental microorganism classification. *Biocybern. Biomed. Eng.* **2015**, *35*, 10–21. [[CrossRef](#)]
15. Dubey, S.R.; Dixit, P.; Singh, N.; Gupta, J.P. Infected fruit part detection using K-means clustering segmentation technique. *Ijimai* **2013**, *2*, 65–72. [[CrossRef](#)]
16. Llamas, J.; Lerones, P.M.; Medina, R.; Zalama, E.; Gómez-García-Bermejo, J. Classification of architectural heritage images using deep learning techniques. *Appl. Sci.* **2017**, *7*, 992. [[CrossRef](#)]
17. Grilli, E.; Menna, F.; Remondino, F. A review of point clouds segmentation and classification algorithms. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *42*, 339–344. [[CrossRef](#)]
18. Weinmann, M. *Reconstruction and Analysis of 3D Scenes: From Irregularly Distributed 3D Points to Object Classes*; Springer: Cham, Switzerland, 2016.
19. Stathopoulou, E.K.; Remondino, F. Semantic photogrammetry: boosting image-based 3D reconstruction with semantic labeling. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *422*, 685–690. [[CrossRef](#)]
20. Yuheng, S.; Hao, Y. Image Segmentation Algorithms Overview. *arXiv*, 2017; arXiv:1707.02051.
21. Al-Amri, S.S.; Kalyankar, N.V.; Khamitkar, S.D. Image segmentation by using edge detection. *Int. J. Comput. Sci. Eng.* **2010**, *2*, 804–807.
22. Kaur, J.; Agrawal, S.; Vig, R. A comparative analysis of thresholding and edge detection segmentation techniques. *Int. J. Comput. Appl.* **2012**, *39*, 29–34. [[CrossRef](#)]
23. Schoenemann, T.; Kahl, F.; Cremers, D. Curvature regularity for region-based image segmentation and inpainting: A linear programming relaxation. In Proceedings of the International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 17–23.
24. Chitade, A.Z.; Katiyar, S.K. Colour based image segmentation using k-means clustering. *Int. J. Eng. Sci. Technol.* **2010**, *2*, 5319–5325.
25. Saraswathi, S.; Allirani, A. Survey on image segmentation via clustering. In Proceedings of the 2013 International Conference on Information Communication and Embedded Systems (ICICES), Chennai, India, 21–22 February 2013; pp. 331–335.
26. Fiorillo, F.; Fernández-Palacios, B.J.; Remondino, F.; Barba, S. 3D Surveying and modelling of the Archaeological Area of Paestum, Italy. *Virtual Archaeol. Rev.* **2013**, *4*, 55–60. [[CrossRef](#)]
27. Naik, D.; Shah, P. A review on image segmentation clustering algorithms. *Int. J. Comput. Sci. Inf. Technol.* **2014**, *5*, 3289–3293.
28. Rabbani, T.; Van Den Heuvel, F.; Vosselman, G. Segmentation of point clouds using smoothness constraint. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2006**, *36*, 248–253.
29. Bhanu, B.; Lee, S.; Ho, C.C.; Henderson, T. Range data processing: Representation of surfaces by edges. In Proceedings of the 8th International Conference on Pattern Recognition, Paris, France, 27–31 October 1986; pp. 236–238.
30. Sappa, A.D.; Devy, M. Fast range image segmentation by an edge detection strategy. In Proceedings of the IEEE 3rd 3D Digital Imaging and Modeling, Quebec City, QC, Canada, 28 May–1 June 2001; pp. 292–299.
31. Wani, M.A.; Arabnia, H.R. Parallel edge-region-based segmentation algorithm targeted at reconfigurable multiring network. *J. Supercomput.* **2003**, *25*, 43–62. [[CrossRef](#)]
32. Castillo, E.; Liang, J.; Zhao, H. Point cloud segmentation and denoising via constrained nonlinear least squares normal estimates. In *Innovations for Shape Analysis*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 283–299.

33. Jagannathan, A.; Miller, E.L. Three-dimensional surface mesh segmentation using curvedness-based region growing approach. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 2195–2204. [[CrossRef](#)]
34. Besl, P.J.; Jain, R.C. Segmentation through variable order surface fitting. *IEEE Trans. Pattern Anal. Mach. Intell.* **1988**, *10*, 167–192. [[CrossRef](#)]
35. Vosselman, M.G.; Gorte, B.G.H.; Sithole, G.; Rabbani, T. Recognising structure in laser scanning point clouds. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2004**, *46*, 33–38.
36. Belton, D.; Lichti, D.D. Classification and segmentation of terrestrial laser scanner point clouds using local variance information. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2006**, *36*, 44–49.
37. Klasing, K.; Althoff, D.; Wollherr, D.; Buss, M. Comparison of surface normal estimation methods for range sensing applications. In Proceedings of the IEEE International Conference on Robotics and Automation, Kobe, Japan, 12–17 May 2009; pp. 3206–3211.
38. Xiao, J.; Zhang, J.; Adler, B.; Zhang, H.; Zhang, J. Three-dimensional point cloud plane segmentation in both structured and unstructured environments. *Robot. Auton. Syst.* **2013**, *61*, 1641–1652. [[CrossRef](#)]
39. Vo, A.V.; Truong-Hong, L.; Laefer, D.F.; Bertolotto, M. Octree-based region growing for point cloud segmentation. *ISPRS J. Photogramm. Remote Sens.* **2015**, *104*, 88–100. [[CrossRef](#)]
40. Liu, Y.; Xiong, Y. Automatic segmentation of unorganized noisy point clouds based on the gaussian map. *Comput. Aided Des.* **2008**, *40*, 576–594. [[CrossRef](#)]
41. Vieira, M.; Shimada, K. Surface mesh segmentation and smooth surface extraction through region growing. *Comput. Aided Geom. Des.* **2005**, *22*, 771–792. [[CrossRef](#)]
42. Ballard, D.H. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognit.* **1991**, *13*, 183–194. [[CrossRef](#)]
43. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [[CrossRef](#)]
44. Poux, F.; Hallot, P.; Neuville, R.; Billen, R. Smart point cloud: Definition and remaining challenge. In Proceedings of the 11th 3D Geoinfo Conference, Athens, Greece, 20–21 October 2016.
45. Barnea, S.; Filin, S. Segmentation of terrestrial laser scanning data using geometry and image information. *ISPRS J. Photogramm. Remote Sens.* **2013**, *76*, 33–48. [[CrossRef](#)]
46. 2D Semantic labelling. Available online: <http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html> (accessed on 28 March 2019).
47. Deepglobe. Available online: <http://deepglobe.org/challenge.html> (accessed on 28 March 2019).
48. Mapping challenge. Available online: <https://www.crowdai.org/challenges/mapping-challenge> (accessed on 28 March 2019).
49. Large-Scale Semantic 3D Reconstruction. Available online: <https://www.grss-ieee.org/community/technical-committees/data-fusion/data-fusion-contest/> (accessed on 28 March 2019).
50. Large-Scale Point Cloud Classification Benchmark. Available online: <http://www.semantic3d.net/> (accessed on 28 March 2019).
51. KITTI Vision Benchmark Suite. Available online: <http://www.cvlibs.net/datasets/kitti/> (accessed on 28 March 2019).
52. Semantic, instance-wise, dense pixel annotations of 30 classes. Available online: <https://www.cityscapes-dataset.com/> (accessed on 28 March 2019).
53. Noh, H.; Hong, S.; Han, B. Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1520–1528.
54. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)]
55. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A review on deep learning techniques applied to semantic segmentation. *arXiv*, 2017; arXiv:1704.06857.
56. Perez, L.; Wang, J. The effectiveness of data augmentation in image classification using deep learning. *arXiv*, 2017; arXiv:1712.04621.
57. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [[CrossRef](#)]
58. Guo, B.; Huang, X.; Zhang, F.; Sohn, G. Classification of airborne laser scanning data using JointBoost. *ISPRS J. Photogramm. Remote Sens.* **2014**, *92*, 124–136. [[CrossRef](#)]

59. Niemeyer, J.; Rottensteiner, F.; Soergel, U. Contextual classification of LiDAR data and building object detection in urban areas. *ISPRS J. Photogramm. Remote Sens.* **2014**, *87*, 152–165. [[CrossRef](#)]
60. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. Pointnet: Deep learning on point sets for 3D classification and segmentation. In Proceedings of the Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; Volume 1, p. 4.
61. Weinmann, M.; Schmidt, A.; Mallet, C.; Hinz, S.; Rottensteiner, F.; Jutzi, B. Contextual classification of point cloud data by exploiting individual 3D neighbourhoods. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *2*, 271–278. [[CrossRef](#)]
62. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
63. Bassier, M.; Van Genechten, B.; Vergauwen, M. Classification of sensor independent point cloud data of building objects using random forests. *J. Build. Eng.* **2019**, *21*, 468–477. [[CrossRef](#)]
64. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [[CrossRef](#)]
65. MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, June 1967.
66. Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **2010**, *31*, 651–666. [[CrossRef](#)]
67. Burney, S.A.; Tariq, H. K-means cluster analysis for image segmentation. *Int. J. Comput. Appl.* **2014**, *96*. [[CrossRef](#)]
68. Dhanachandra, N.; Manglem, K.; Chanu, Y.J. Image segmentation using K-means clustering algorithm and subtractive clustering algorithm. *Procedia Comput. Sci.* **2015**, *54*, 764–771. [[CrossRef](#)]
69. Zhang, C.; Mao, B. 3D Building Models Segmentation Based on K-means++ Cluster Analysis. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *42*. [[CrossRef](#)]
70. Hétroy-Wheeler, F.; Casella, E.; Boltcheva, D. Segmentation of tree seedling point clouds into elementary units. *Int. J. Remote Sens.* **2016**, *37*, 2881–2907. [[CrossRef](#)]
71. Guo, Y.; Bennamoun, M.; Sohel, F.; Lu, M.; Wan, J.; Kwok, N.M. A comprehensive performance evaluation of 3D local feature descriptors. *Int. J. Comput. Vis.* **2016**, *116*, 66–89. [[CrossRef](#)]
72. Georganos, S.; Grippa, T.; Vanhuyse, S.; Lennert, M.; Shimoni, M.; Kalogirou, S.; Wolff, E. Less is more: Optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. *GISci. Remote Sens.* **2018**, *55*, 221–242. [[CrossRef](#)]
73. Weinmann, M.; Jutzi, B.; Mallet, C. Feature relevance assessment for the semantic interpretation of 3D point cloud data. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, *5*, W2. [[CrossRef](#)]
74. Aijazi, A.K.; Serna, A.; Marcotegui, B.; Checchin, P.; Trassoudaine, L. Segmentation and Classification of 3D Urban Point Clouds: Comparison and Combination of Two Approaches. In *Field and Service Robotics*; Springer: Cham, Switzerland, 2016; pp. 201–216.
75. Mathias, M.; Martinovic, A.; Weissenberg, J.; Haegler, S.; Van Gool, L. Automatic architectural style recognition. *ISPRS-Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2011**, *3816*, 171–176. [[CrossRef](#)]
76. Oses, N.; Dornaika, F.; Moujahid, A. Image-based delineation and classification of built heritage masonry. *Remote Sens.* **2014**, *6*, 1863–1889. [[CrossRef](#)]
77. Shalunts, G.; Haxhimusa, Y.; Sablatnig, R. Architectural style classification of building facade windows. In Proceedings of the International Symposium on Visual Computing, Las Vegas, NV, USA, 26–28 September 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 280–289.
78. Zhang, L.; Song, M.; Liu, X.; Sun, L.; Chen, C.; Bu, J. Recognizing architecture styles by hierarchical sparse coding of blocklets. *Inf. Sci.* **2014**, *254*, 141–154. [[CrossRef](#)]
79. Chu, W.T.; Tsai, M.H. Visual pattern discovery for architecture image classification and product image search. In Proceedings of the 2nd ACM International Conference on Multimedia Retrieval, Hong Kong, China, 5–8 June 2012; p. 27.
80. Llamas, J.; Lerones, P.M.; Zalama, E.; Gómez-García-Bermejo, J. Applying deep learning techniques to cultural heritage images within the INCEPTION project. In Proceedings of the Euro-Mediterranean Conference, Nicosia, Cyprus, 31 October–5 November 2016; Springer: Cham, Switzerland, 2016; pp. 25–32.
81. Manferdini, A.M.; Remondino, F.; Baldissini, S.; Gaiani, M.; Benedetti, B. 3D modeling and semantic classification of archaeological finds for management and visualization in 3D archaeological databases. In Proceedings of the International Conference on Virtual Systems and MultiMedia (VSMM), Limassol, Cyprus, 20–25 October 2008; pp. 221–228.

82. Poux, F.; Neuville, R.; Hallot, P.; Billen, R. Point cloud classification of tesserae from terrestrial laser data combined with dense image matching for archaeological information extraction. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *4*, 203–211. [[CrossRef](#)]
83. De Luca, L.; Buglio, D.L. Geometry vs. Semantics: Open issues on 3D reconstruction of architectural elements. In *3D Research Challenges in Cultural Heritage*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 36–49.
84. Stefani, C.; Busayarat, C.; Lombardo, J.; Luca, L.D.; Véron, P. A web platform for the consultation of spatialized and semantically enriched iconographic sources on cultural heritage buildings. *J. Comput. Cult. Herit.* **2013**, *6*, 13. [[CrossRef](#)]
85. Apollonio, F.I.; Basilissi, V.; Callieri, M.; Dellepiane, M.; Gaiani, M.; Ponchio, F.; Rizzo, F.; Rubino, A.R.; Scopigno, R. A 3D-centered information system for the documentation of a complex restoration intervention. *J. Cult. Herit.* **2018**, *29*, 89–99. [[CrossRef](#)]
86. Campanaro, D.M.; Landeschi, G.; Dell’Unto, N.; Touati, A.M.L. 3D GIS for cultural heritage restoration: A ‘white box’ workflow. *J. Cult. Herit.* **2016**, *18*, 321–332. [[CrossRef](#)]
87. Sithole, G. Detection of bricks in a masonry wall. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2008**, XXXVII, 1–6.
88. Riveiro, B.; Lourenço, P.B.; Oliveira, D.V.; González-Jorge, H.; Arias, P. Automatic Morphologic Analysis of Quasi-Periodic Masonry Walls from LiDAR. *Comput.-Aided Civ. Infrastruct. Eng.* **2016**, *31*, 305–319. [[CrossRef](#)]
89. Messaoudi, T.; Véron, P.; Halin, G.; De Luca, L. An ontological model for the reality-based 3D annotation of heritage building conservation state. *J. Cult. Herit.* **2018**, *29*, 100–112. [[CrossRef](#)]
90. Cipriani, L.; Fantini, F. Digitalization culture vs. archaeological visualization: Integration of pipelines and open issues. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *42*, 195. [[CrossRef](#)]
91. Bora, D.J.; Gupta, A.K. A new approach towards clustering based color image segmentation. *Int. J. Comput. Appl.* **2014**, *107*, 23–30.
92. Jurio, A.; Pagola, M.; Galar, M.; Lopez-Molina, C.; Paternain, D. A comparison study of different color spaces in clustering based image segmentation. In Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Dortmund, Germany, 28 June–2 July 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 532–541.
93. Sural, S.; Qian, G.; Pramanik, S. Segmentation and histogram generation using the HSV color space for image retrieval. In Proceedings of the 2002 International Conference on Image Processing, Rochester, NY, USA, 22–25 September 2002; Volume 2, p. II.
94. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. *Data Mining: Practical Machine Learning Tools and Techniques*; Morgan Kaufmann: Burlington, MA, USA, 2016.
95. Image processing package Imagej. Available online: <http://imagej.net/Fiji> (accessed on 25 March 2019).
96. Imagej K-means plugin. Available online: <http://ij-plugins.sourceforge.net/plugins/segmentation/k-means.html> (accessed on 25 March 2019).
97. Kleiner, F.S. *A History of Roman Art*; Cengage Learning: Boston, MA, USA, 2016.
98. Remondino, F.; Gaiani, M.; Apollonio, F.; Ballabeni, A.; Ballabeni, M.; Morabito, D. 3D documentation of 40 kilometers of historical porticoes—the challenge. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*. [[CrossRef](#)]
99. Giraudot, S.; Lafarge, F. Classification. In *CGAL User and Reference Manual*, 4.14 ed; CGAL Editorial Board, 2019.
100. ETHZ Random Forest code. Available online: [www.pr.s.igp.ethz.ch/research/Source\\_code\\_and\\_datasets.html](http://www.pr.s.igp.ethz.ch/research/Source_code_and_datasets.html) (accessed on 25 March 2019).
101. Menna, F.; Nocerino, E.; Remondino, F.; Dellepiane, M.; Callieri, M.; Scopigno, R. 3D digitization of an heritage masterpiece—a critical analysis on quality assessment. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*. [[CrossRef](#)]
102. Jiménez Fernández-Palacios, B.; Rizzi, A.; Remondino, F. Etruscans in 3D—Surveying and 3D modeling for a better access and understanding of heritage. *Virtual Archaeol. Rev.* **2013**, *4*, 85–89. [[CrossRef](#)]

