Machine and Deep Learning Implementations for Heritage Building Information Modelling: A Critical Review of Theoretical and Applied Research

ALEKSANDER GIL

Architecture and Built Environment, Northumbria University, Newcastle upon Tyne, United Kingdom, aleksander.gil@northumbria.ac.uk

YUSUF ARAYICI

Architecture and Built Environment, Northumbria University, Newcastle upon Tyne, United Kingdom, yusuf.arayici@northumbria.ac.uk

BIMAL KUMAR

Department of Architecture, University of Strathclyde, Glasgow, United Kingdom, b.kumar@strath.ac.uk

RICHARD LAING

Architecture and Built Environment, Northumbria University, Newcastle upon Tyne, United Kingdom, richard.laing@northumbria.ac.uk

Research domain and Problem: HBIM modelling from point cloud data has become a crucial research topic in the last decade since it is potentially considered the central data model paving the way for the digital heritage practice beyond digitization. Reality Capture technologies such as terrestrial laser scanning, drone-mounted LiDAR sensors and photogrammetry enable the reality capture with a sub-millimetre accurate point cloud file that can be used as a reference file for Heritage Building Information Modelling (HBIM). However, HBIM modelling from the point cloud data of heritage buildings is mainly manual, error-prone, and time-consuming. Furthermore, image processing techniques are insufficient for classification and segmenting of point cloud data to speed up and enhance the current workflow for HBIM modelling.

Due to the challenges and bottlenecks in the scan-to-HBIM process, which is commonly criticized as complex with its bespoke requirements, semantic segmentation of point clouds is gaining popularity in the literature.

Research Aim and Methodology: Therefore, this paper aims to provide a thorough critical review of Machine Learning and Deep Learning methods for point cloud segmentation, classification, and BIM geometry automation for cultural heritage case study applications.

Research findings: This paper files the challenges of HBIM practice and the opportunities for semantic point cloud segmentation found across academic literature in the last decade. Beyond definitions and basic occurrence statistics, this paper discusses the success rates and implementation challenges of machine and deep learning classification methods.

Research value and contribution: This paper provides a holistic review of point cloud segmentation and its potential for further development and application in the Cultural Heritage sector. The critical analysis provides insight into the current state-of-the-art methods and advises on their suitability for HBIM projects. The review has identified highly original threads of research, which hold the potential to significantly influence practice and further applied research.

ACM 1556-4673/2024/02-ART
http://dx.doi.org/10.1145/3649442

Additional Keywords and Phrases: HBIM, Point Cloud, Semantic Segmentation, Classification, Machine Learning, Deep Learning

1 INTRODUCTION

The continual growth of digital tools and workflows for designing, recording, and managing built assets has introduced a data-driven approach in the Cultural Heritage sector underpinned by the term Heritage Building Information Modelling (HBIM). Originating from Building Information Modelling (BIM), HBIM extends this technology's capabilities to the specific needs of historical buildings and targets enhancements in the Cultural Heritage sector[1]. The definition of this phenomenon has evolved from a 3D representation of a heritage asset generated by BIM software [2], [3], to a sophisticated 3D database solution that considers cultural values, local significance, and sustainable conservation strategy [4]. However, the evolution of theoretical concept has progressed at a much faster pace than the operational capability, which has resulted in a very complex, inefficient, and costly process of generating HBIM geometry.

Many criticise the initial HBIM geometry generation step from the captured point cloud as a bottleneck problem and underline the need for its automation [5], [6], [7]. Inspired by autonomous driving, computer-aided manufacturing and surveying sectors that also utilise point clouds, automated classification methods have been the most common hypothesis for addressing the articulated challenge. Machine Learning (ML), a discipline founded by [8], forms a critical subset of Artificial Intelligence (AI). It focuses on developing methodologies capable of "learning" from input data to enhance task performance across various domains. In essence, ML thrives on identifying patterns and making decisions with minimal human intervention.ML, along with its advanced counterpart, Deep Learning (DL), has demonstrated considerable success in automatically classifying text, images, and behavioural patterns. These technologies, particularly DL, which involves neural networks that simulate human brain functions, have revolutionized data interpretation and analysis.

However, current state-of-the-art ML and DL methods face significant challenges when applied to point cloud datasets in the context of heritage conservation, specifically Heritage Building Information Modelling (HBIM) [9]. The primary obstacle lies in the unstructured nature of these datasets. Heritage assets often exhibit considerable variability in materials, sizes, geometric configurations, and surface characteristics. This diversity, combined with the need for precise and homogeneous segmentation of elements within these datasets, presents a complex challenge for standard ML and DL algorithms. [10].

In this context, this paper will provide a critical review of the ML and DL methods in the literature, which have the potential or directly contribute to the automation of the HBIM generation process. The next section briefly describes challenges found in HBIM practice; then section 3 provides insight into our research methodology, followed by section 4 which focuses on machine learning methods and their application in heritage management and research. Section 5 is focused on the complex DL methods and their exploitation in the professional heritage sector. Lastly, sections 6 and 7 articulate a discussion leading to the paper's conclusion.

2 CHALLENGES IN SCAN TO BIM IN CULTURAL HERITAGE

The process of digitally documenting a heritage building will often be undertaken utilizing advanced surveying technologies (or photography-based 'structure from motion' techniques) which will ultimately generate 'point cloud' files as the critical reference for 3D modelling. Although over 6 million heritage assets were erected before 1919 in the UK alone, a limited number of standards and guidance regulates best-practices for transforming such data into a data-rich model (Scan-to-BIM). Furthermore, some attempts at standardizing concepts such as "Level of Detail" (LOD) or generating a standard library of parametric building components vary across studies and countries [11].

One of the most common criticisms found in this review of Scan-to-BIM in a cultural heritage context is the complex and laborious process of HBIM using standard BIM software [12]. In this context, many factors that contribute to the inefficiency of this process [13] [14] [15] [16] [17] [18] such as:

- Damaged, decayed, or weathered building elements
- The geometrical complexity of handcrafted building elements
- Interdisciplinary systems found following adaptive building reuse
- BIM software limitations in shape generation
- BIM software limitation in large-point cloud file handling
- Lack of Heritage building ontology in BIM software for semantic data mapping

It is underlined in the literature that the need for automation is crucial, to balance the clear benefits of digitisation against the costs of digital upskilling - or further interdisciplinary recruitment – within the cultural heritage sector itself.

3 RESEARCH METHODOLOGY

The semi-systematic review helps to identify theoretical approaches and gaps in the literature and verifies the progress of publications over time [19].Despite previous publications of similar literature review such as [20] or [21] our work bring novelty with its increased number of papers sampled, attention on the achieved results and in-depth analysis of publications. It proved an adequate research strategy as the research questions involve a theme explored by diverse disciplines. Bibliometric techniques and quantitative analysis allowed us to synthesize themes, entail assessments and critique perspectives of previous researchers with a similar scope of work [22]. Using Scopus as the main literature database and VOS viewer as the mapping tool, it was possible to identify the following research trends efficiently:

There is a steady increase in semantic point cloud segmentation publications (Figure 1)

The authors pioneering this field of research have strong working relationships (Figure 2)

Most of the research undertaken to this date originates from Europe.

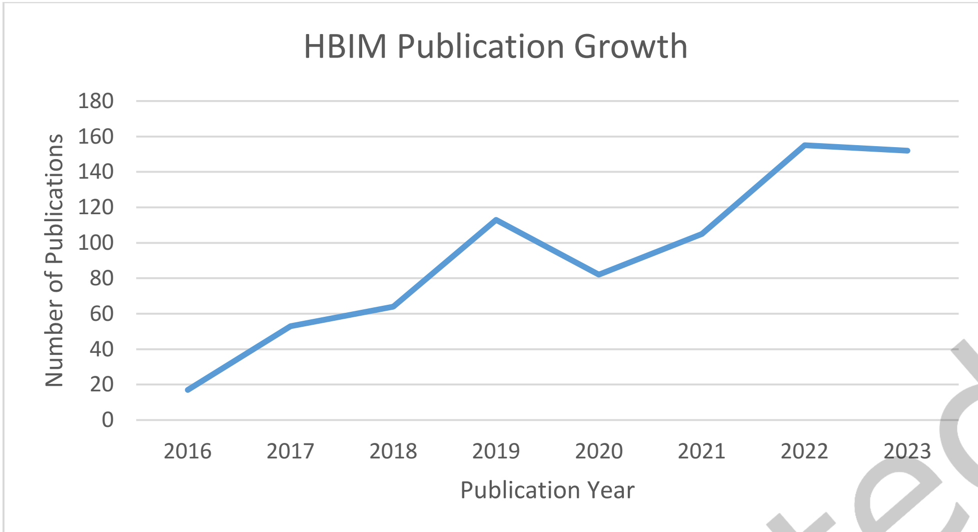


Figure 1: Increasing trend of research point cloud segmentation publications

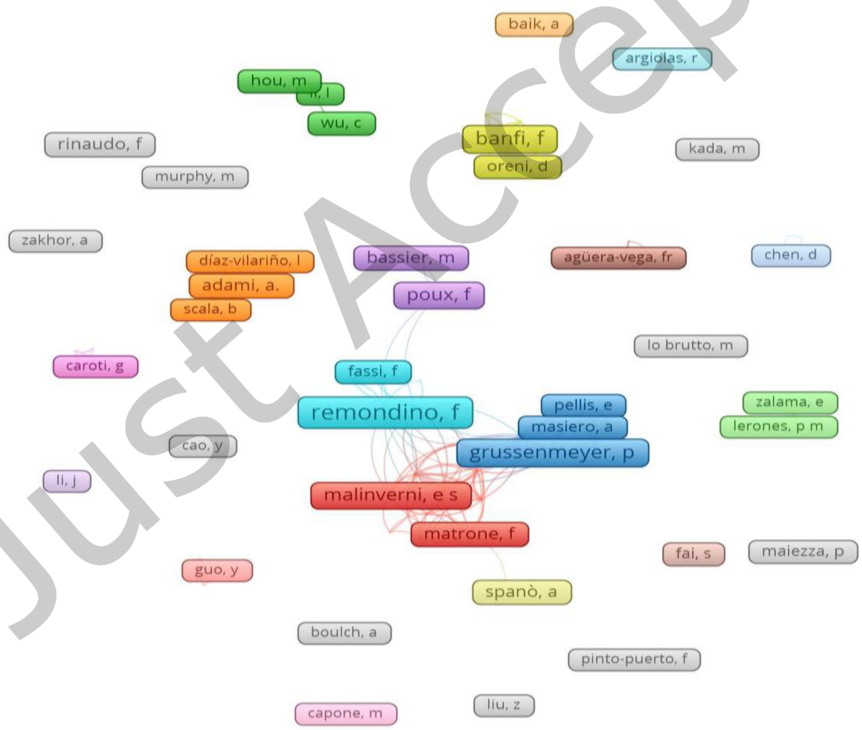


Figure 2: Concept map of the researchers in the literature and their linkups using VOSviewer

The literature data collected was analysed qualitatively via critical abstract review and NVivo keyword mapping. This allowed a refinement of the literature data, which was then reviewed and coded in relation to their relevance to Machine Learning, Deep Learning and Cultural Heritage.

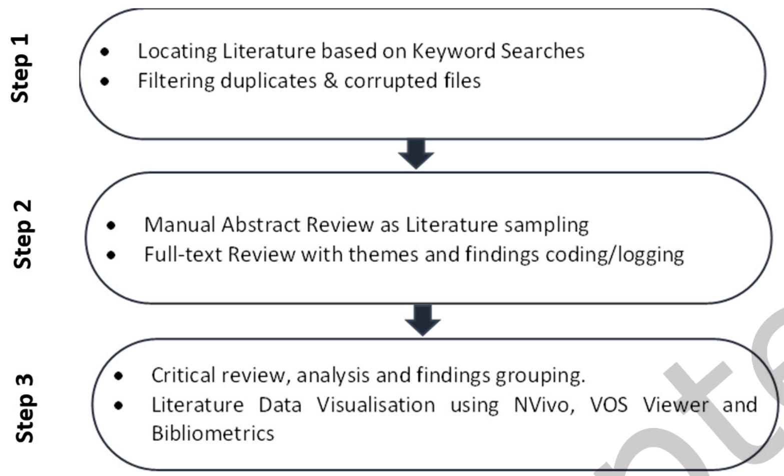


Figure 3: Research Process workflow

The abductive approach was adopted in the paper for the codification process while the grounded theory was formulated as the research strategy since the initial observations may be incomplete and novel insights from the literature could impact our preliminary theoretical assumptions [23].

4 ALGORITHMIC AND MACHINE LEARNING SEMANTIC SEGMENTATION METHODS

In the context of point clouds – a collection of geometrical primitives commonly complemented by colour, multispectral or intensity – it has been argued and demonstrated that ML can help with understanding their segmentation and classification methods, which is fundamental for automation experimentation. Segmentation refers to grouping or commonly called segmenting point clouds, that share similar characteristics, and classification assigns a label to a segment of point clouds according to specific criteria.

Prior to the advanced ML methods, statistical methods were used for data segmentation which can also be interpreted as algorithmic segmentation. These methods have been used individually and in combination to perform point cloud segmentation for as-built modelling [5]. Application of statistical and Machine Learning methods are critical for the success of computer-vision object detection, feature recognition or classification of diverse datasets [24], [25]. Machine Learning methods can be grouped into two major learning approaches: supervised learning and unsupervised learning. These are shown in Figure 4 and 5.

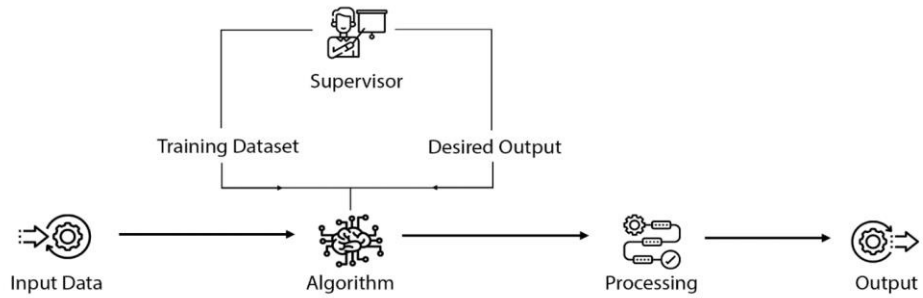


Figure 4: Machine Learning via Supervised Learning Approach

A supervised approach, where the semantic labels are learned from a user-annotated data sample and the trained model, is used for the classification of the entire dataset. This method comes with a degree of mandatory manual work, which is a standard practice in many domains, and it has a proven track record for accurate outcomes and gives the user with more control over the result.

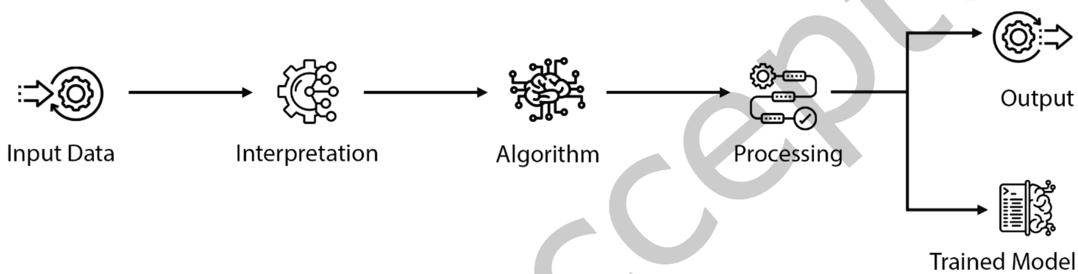


Figure 5: Machine Learning via unsupervised learning

In an unsupervised approach, the input data is split into segments based on the user-defined parameters that influence the algorithm. This method exposes a risk that the outcome might not align with users' intentions but if successfully conducted, it significantly reduces the amount of manual work required to obtain accurate results.

Regardless of the method, semantic segmenting of point clouds is challenging for practitioners and researchers due to their irregular file format and dataset size. Commonly millions to multiple billions of points are used to recreate a built space, and each point carries data features such as XYZ coordinates, RGB colour codes, intensity, omni-variance, planarity, linearity, sphericity, or verticality. As sampling reduces the dimensional detail required for feature recognition most methods utilize pooling methods to bypass the great computing power required for this task, which adds another layer of complexity to the overall workflow.

4.1 Statistical Methods

The earlier papers in the literature approached the segmentation of building features using edge-based, region-growing and model-fitting methods. A review by [26] provides a comprehensive list of relevant case studies such as powerline, wall, or surface point cloud classification. Edge-based segmentation detects the outlines of the borders of different regions and groups the points inside the boundaries to deliver final segments. The properties used by this method are normals, gradients and principal curvatures, which set the threshold for the segment depth map [27]. The region-growing segmentation method is based on the principle that one or more points grow around neighbouring points with similar characteristics such as surface or curvature. This method has a bottom-up and a top-down approach, which determines whether the algorithm grows or subdivides the assigned segments. Thus, initially developed by [28] but the best variation of this method was developed by [29], whereby colour properties were leveraged to achieve more accurate outcomes.

Model-fitting segmentation techniques were mainly encapsulated by two methods, Hough Transform (HT) and Random Sample Consensus (RANSAC) developed by [30], [31] respectively. The former detects planes, cylinders and spheres and the latter extracts shapes by randomly selecting minimum sample sets and testing the fit across the entire dataset. A comparison of both methods was undertaken by [32], which proved that RANSAC was more efficient and capable of processing larger datasets.[33] modified RANSAC to be less sensitive to noise and avoided under-segmentation. Further modifications such as M-estimator Sample and Consensus (MSAC) to address the cost function in linear regression and Maximum Likelihood Estimation Sample Consensus (MLESAC) to address the likelihood over the number of outliers, were also developed to enhance building feature detection [34], [35].

4.2 Supervised Machine Learning Methods

Linear Discriminant Analysis (LDA) is one of the oldest classifiers, which uses a linear transformation to calculate the directions of the axis that best distinguishes various classes [36]. Data points are projected onto the axis directions and the algorithm assigns classes to elements with similar trends, assuming that features are continuous and regularly distributed [37]. Logistic Regression is somewhat like LDA as it also establishes a linear transformation, but the key difference is that it compares the input variables against probabilities of the output categorical variable [38]. This method does not require input to be continuous or regularly distributed as the result focuses on the probability of belonging rather than the class itself.

Essentially, a support vector machine (SVM) algorithm aims to find a hyperplane in high dimensional feature space to solve some linearly inseparable problem by reviewing all possible hyperplanes against the maximum margin [39]. According to the studies by [40] and [41] this method is suitable for indoor and outdoor point cloud data and can be used for regression and classification studies. Although it is less popular in point cloud applications, the Naïve Bayes classifier is a probabilistic ML model based on the Bayesian Theorem (probability of an event occurring based on previous occurrences). This method examines the probabilistic relationship between a particular data point and class. Gaussian Naïve Bayes (GNB) is one of the most used extensions of this method, which assumes the probability of following gaussian distribution [42], [43]. [44] used this method for city-scale orthoimage classification, which was then translated onto the corresponding point cloud dataset. These methods are illustrated together in Figure 6.

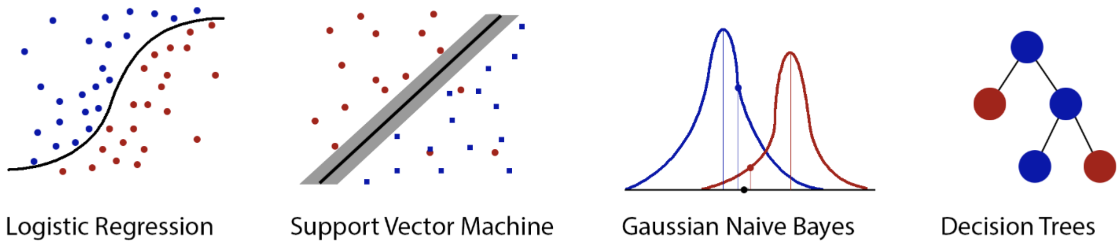


Figure 6: Supervised Machine Learning Methods

The decision Tree (DT) algorithm provides a method for supporting decisions and their possible consequences based on continually splitting the input data and calculating parameters for each part. This is a computationally efficient method. However, a major drawback of this method is that it can create complex trees due to small changes in data [45]. Random Forest (RF) is a method that creates a large collection of uncorrelated trees and then uses bootstrap aggregation (also known as bagging) to average them [46], which appears to improve accuracy and allows the model to score the estimation results leading to a winner-takes-all driven outcome [47]. This method is relatively simple to implement if compared to the other methods as it follows a similar logic to human thinking. Furthermore, it is found that it is the most common method used for point cloud classification in Cultural Heritage [37]. [48] proposed an improvement to this method in a point cloud context named 3DOR-Tree, which combines 3D node and octree with RF (Random Forest) to filter out empty nodes, resulting in a more efficient computation process for large irregular datasets.

4.3 Unsupervised Machine Learning Methods

K-means clustering is an unsupervised algorithm originally published by [49] that uses centroids as prototypes and similar attributes or characteristics to create clusters [25]. The main shortcoming of this method is the need for a defined number of desired clusters that enables the process of minimising the Euclidean, Manhattan, Squared or Cosine distance between the datapoint and the centroid.

Lloyd's algorithm is also very well-known implementation of k-means as it initially chooses the centroids as random, clusters surrounding points and follows through with recalculation of centroids until their location does not change [50]. [51] were one of the first to prove this method works on point cloud and [52] improved this method by incorporating a mean shift algorithm [53] to overcome the need to set a fixed number of clusters desired. [54] compared the Lloyds K-mean against more complex ANN methods using MATLAB on photogrammetry and Lidar point cloud datasets. Thus, the clustering results were satisfactory ANN proved to be more efficient and accurate. Overall, k-means is the most common unsupervised algorithm discussed in the literature, but some authors point out that it is only suitable for small datasets [55]. Figure 7 shows the K Means and Hierarchical clustering as the unsupervised Machine Learning Methods.

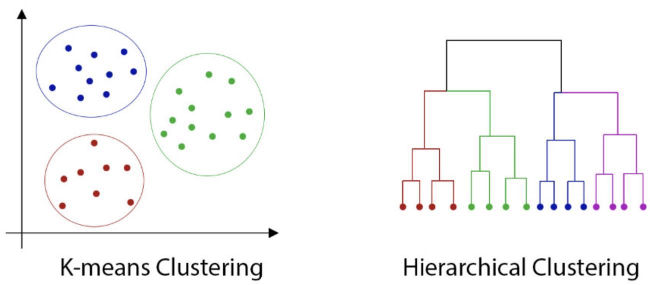


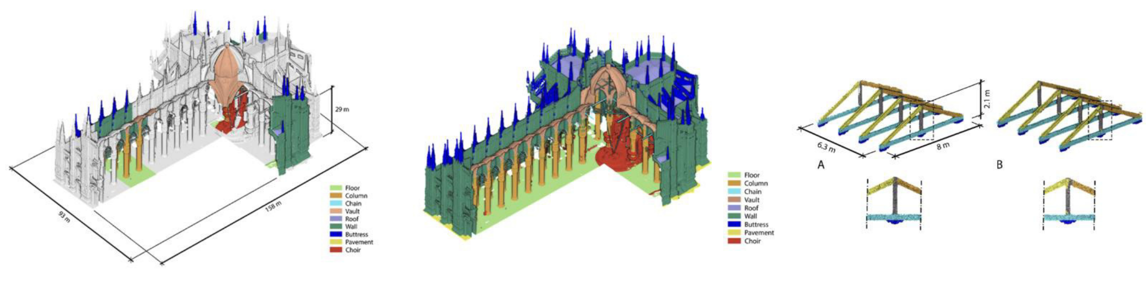
Figure 7: Unsupervised Machine Learning Methods

Hierarchical Clustering methods compute features for each data point based on geometrical and radiometric characteristics such as position, surface normal, surface fitting residuals or point reflectance [25]. These can be conducted following the agglomerative (bottom-up) or divisive (top-down) approach to establish the hierarchy across the dataset. A novel method of hierarchical clustering called Pairwise Linkage [55] was presented by whereby using the computation of closest neighbouring points, the issue of processing large and complex datasets captured by mobile, aerial, and terrestrial methods is resolved.

4.4 Semantic segmentation applications in Cultural Heritage

To resolve the semantic segmentation problem for point cloud datasets collected or derived from built heritage assets, it was found that a variety of mono and hybrid methods were tested on specific case studies and particular benchmark datasets. [56] combined the SVM classifier with extensive feature vectors and k-means to achieve accurate floor, wall, roof, ceiling, and beam segmentation on a sample of nine architectural archetypes, including a church, castle, houses, factories, and offices. This hybrid method initially transformed the point cloud into a planar mesh, which enabled a series of features to be computed and classified against 17 different classes. Although the model was trained in under 40 seconds and achieved 81% accuracy, better results would be predicted with an increased volume of labelled training data and reduced amount of clutter in the scanned datasets.

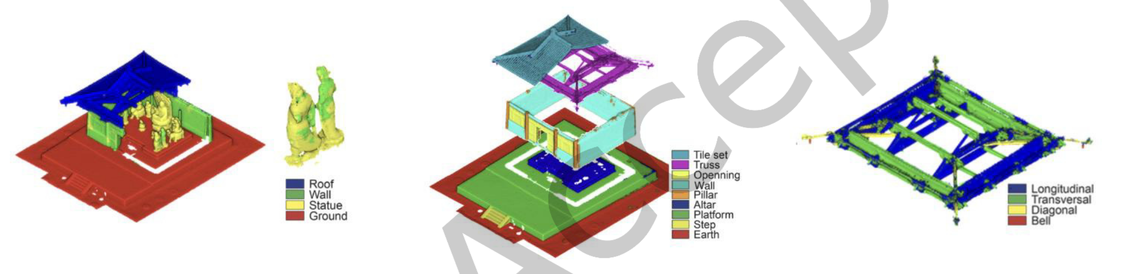
The application of Random Forest (RF) to a large complex dataset of the Milan Cathedral and Pomposa Abbey in a study by [57] exposed an accuracy drawback, which led to a proposal of a novel multi-level, multi-perception method (MLMR). A bespoke ontology was decided to develop for building elements that correlated to the three different subsamples of the input dataset. At each dataset scale, they applied RF, nearest neighbour algorithm and octree to isolate each building ontology class. Although this supervised method achieved 94%-97% accuracy for each subsample following a 5-minute training procedure, the study concluded that the RF classifier struggled to segment building features that combine similar segmentation classes correctly. Figure 8 illustrates the bespoke implementation of the Random Forest technique on a point cloud representation of Milan Cathedral.



Training Data Sample Level 1 Classification Results Level 3 Classification Results

Figure 8: Random Forest implementation with a bespoke ontology for point cloud segmentation [57]

[58] replicated this method on two buildings from the Wutai Monti UNESCO Heritage Site in China. They achieved an F1 result of 93%-96% depending on the subject case study building and reflected on the small building features that were missed in the classification results. The final recommendations from their study advised that extra attention should be devoted to the dataset quality and colour but generally, they believe this method is suitable for heritage building application. Figure 9 below shows the RF classification results on the cultural heritage case study point cloud.



Level 1 Classification Results Level 2 Classification Results Level 3 Classification Results

Figure 9: Random Forest classification results at three levels [58]

In a study by [59], RF and K-means clustering were tested to develop a heritage point cloud framework proven on four case studies varying in size, classification purpose and overall complexity. This framework showcases the use of RF on orthoimage to identify material types, the use of RF on point cloud to identify façade components and k-means clustering to analyse eroded and deteriorated surfaces. Although the results vary significantly across the different case studies, this framework paves the way for future studies and proves that 2D image classification and result projection onto 3D datasets is an efficient method especially if the image quality and uniform regions are optimized during the processing stages.

An unsupervised hierarchical clustering method proposed by [60] leverages advanced feature extraction methods and region-growing algorithms to automatically generate clusters before a supervised Random Forest workflow occurs. Grilli et al., (2021) applied this method to four case study datasets (including a point cloud representation of a heritage façade) and tested the unsupervised feature extraction against two other feature groups obtained directly from the input dataset. The result of this work proved that unsupervised clustering led to better RF classification and less computational power required although the heritage case study had the lowest performance with an F1 score of 80.55%.

[62] compared four different ML and DL classifiers on built heritage datasets, including random forest, one vs one classifier, 1D and 2D CNN and Bi-LSTM RNN. This study provided valuable insight into how different combinations of point features such as coordinates, radiometric values and geometric features affect the output accuracy and concluded that ML methods outperformed DL. Using the same RF method, [7] conducted further tests with various combinations of point features as the first part of their study. Figure 10 shows the outputs from the combined use of Random Forest and RANSAC methods for HBIM modelling from the point cloud dataset. This work was followed up by a subsequent study, [63] which documents a semi-automatic Scan-to-BIM reconstruction workflow that proves to be the current state-of-the-art in the HBIM research field.

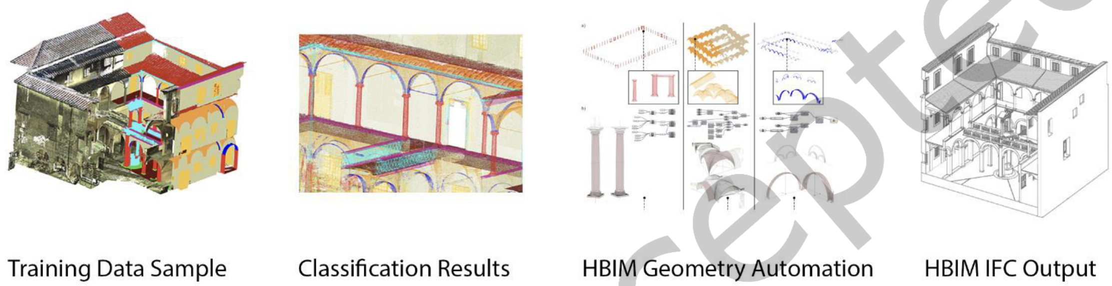


Figure 10: HBIM modelling from point cloud using Random Forest and RANSAC [7]

Their results proved that the best combination of features were nine geometric features (linearity, planarity, omni-variance etc.) and z coordinate, which achieved an F1 score of 98.81%. The second part of their study focused on HBIM generation automation, which was approached on a general and trivial basis depending on the building element. General elements were replicated algorithmically using parametric adaptive components and trivial elements leveraged a BIM library of components and primitive fitting using RANSAC.

5 DEEP LEARNING CLASSIFICATION METHODS

Deep learning (DL) is an evolution of ML that structures multiple algorithms into an input layer, hidden layers and an output layer, allowing the artificial neural network inspired by the human brain to learn and make intelligent decisions [62]. Although there are several deep learning model types tailored to specific purposes, it is found that the following three are most used for semantic point cloud classification:

ANN – Artificial Neural Network is a group of multiple neurons (such as linear regression) at each layer that only processes the inputs in a forward direction. Input information passes through various neurons and each layer learns certain weights, which leads to the desired outcome crafted by the author.

CNN – Convolution Neural Network utilises kernels as the building blocks for multilayer perception across various convolutional layers that can be entirely connected or pooled. These convolution layers create feature maps that record a region of an image, which is broken down into rectangles or pixels and sent out for non-linear processing.

RNN – Recurrent Neural Network is a group of multiple neurons that have a recurrent connection to the hidden layer. The output from each neuron is saved just like a memory cell and looped back into the model. This method has to capability to self-learn from incorrect perditions and backpropagate towards a correct prediction.

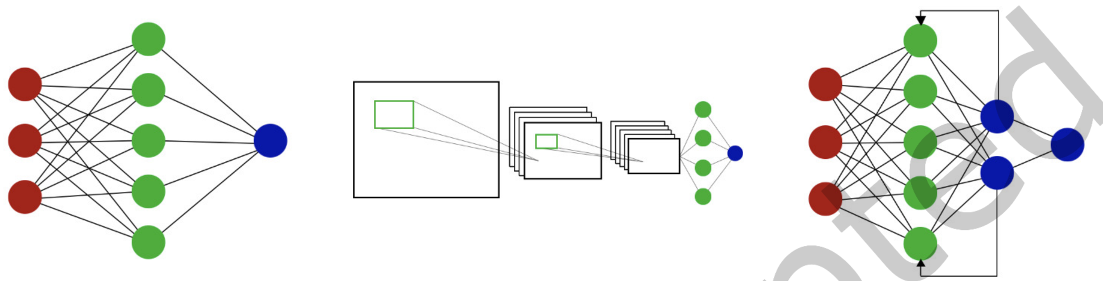


Figure 11: Three most common deep learning architectures

The success of DL methods found across computer vision studies has inspired researchers to tailor proven techniques to achieve the semantic segmentation of point clouds. However, due to the lack of structure and irregularity of point clouds, the application of DL on 3D data still faces significant challenges [64]. There is a diverse approach found in the deep learning methods for point cloud classification review, and literature regarding the way particular methods is grouped and discussed. In this paper, all techniques found in the literature will be presented by their network type without further grouping, such as direct/indirect [64] or point based/tree-based [65] as found in other review study publications.

5.1 Convolution Neural Networks & RGB-D Methods

The pioneering and award-winning networks such as AlexNet [66], Visual Geometry Group [67], GoogLeNet [68] and Microsoft ReNet [69] paved the way for computer vision using CNNs to classify image datasets. Although application of CNNs can be found across the available literature, such as building type recognition from urban façade image datasets [70] or heritage building defect classification linked directly to HBIM [71], generic CNN methods are not fit to process 3D datasets directly.

The advancements in RGB-D sensors introduced an additional feature for short-range indoor photographic data also referred to as 2.5D. Following the works of [72] [73] proposed a method where two CNNs were applied to this data type separately and fused to classify the subject class. Others proposed a CNN without fusion on RGB-D data by utilising a Laplacian Pyramid for feature extraction, which informed the super-pixels on the overall classification output [74], [75]. However, RGB-D sensors rely on infrared lights to project a pseudo-random dot matrix for stereo photography and are not typically suitable for outdoor applications. The prior methods set the foundation for a R-CNN method developed by [76] which was then applied to aerial laser scan point cloud generated raster maps [77].

The application of CNNs using 2.5 aerial datasets and additional features such as heat map data can also be found in a study by [78], whose novel CNN architecture allows for 2D classification and fusion onto the point cloud. Addressing the limitation of 2.5D input data, [79] proposed a method called SplatNet, which has a 2D-3D and a direct 3D variation that leverages CNN models and feature fusion for classification. Using a heritage façade as the input dataset, the 2D-3D variation that applies 2 layers of CNN proved to be most effective with an instance average IoU of 85.4%. Even more impressive results can be found in a study by [80], where a four-layer CNN process applied directly to 3D point cloud input data achieved 92.4% accuracy. In the context of 2.5D CNN methods applied to Cultural Heritage datasets, [81] proposed a pioneering method that achieves depth-based image classification from the ArCH benchmark dataset and applies a feature fusion process to classify the complementary point clouds. The results of this method achieved a global accuracy of 87%-90% and are presented in Figure 12.

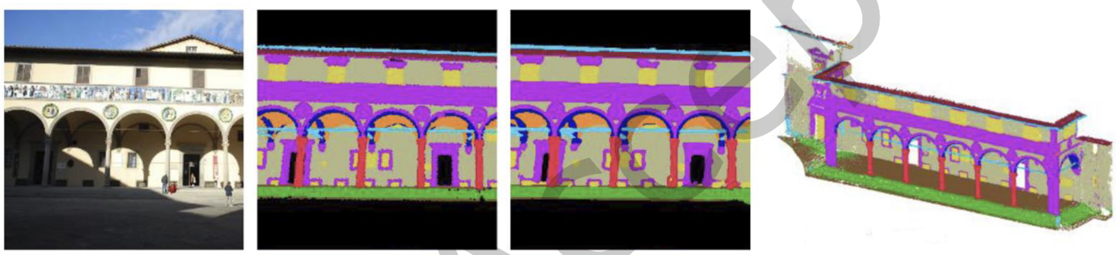


Figure 12: CNN method applied with a four-layer process in cultural Heritage [81]

5.2 Multi-View workflows & CNN Methods

The popularity and proven success of deep neural networks applied to 2D data with regular structure incepted a trend where 3D point cloud data is transformed into 2D views allowing for existing CNN methods to be exploited. Following this approach, [82] proposed MVCNN (Multi-View Convolutional Neural Networks), that uses a render engine to generate 2D images of the 3D sense to extract features via CNN. These generated images are then aggregated using a pooling layer. These features are processed by a secondary CNN to conduct segmentation and classification. Although this method does not consider spatial relationships, it achieves an accuracy of 87.2%. [79] developed a method called SnapNet, that takes a series of 2D snapshots of the 3D dataset. It is then pre-processed into a mesh and uses the 2.5D information of each snap for CNN classification. Classification results are enhanced by the depth feature and then mapped onto the 3D dataset with an overall top accuracy of 91%.

The application of a multi-view classification method in Cultural Heritage context can be found in a study by [83], where a cost-efficient processing method is applied to georeferenced, stereoscopic images to reconstruct heritage building facades. This method extracts façade features and plots them into a point cloud format that allows for meshing and texture projection. The texture and colour-driven feature bias found in this method was also addressed shortly after, improving the feature detection via depth maps. The deviation of final mesh accuracy was significantly reduced [84].

In contrast to the meshing workflow [85] addressed point cloud classification using a muti-view method on the ArCH benchmark dataset. As the benchmark contains 2D RGB-D images and point clouds, results from the available 2D data and 2D shots generated directly from the 3D datasets were directly compared. The 93.4% accuracy results initially obtained from the DeepLabv3+ CNN on the full image set classification gave a promising start to the study. However, in the following test conducted on other buildings, a lower quantity of data used for training and prediction images taken as snips of the point cloud proved that the accuracy decreased significantly due to lower image definition. Furthermore, it was found in their study that the back-projection of prediction image classification introduced significant data loss and mapping inaccuracies and the final point cloud only achieved an accuracy of 33.4% & 57.3%.

5.3 Volumetric Methods

Voxelization has been explored across the literature where the points are transformed into a regular volumetric occupancy grid to introduce a structure to a point cloud dataset. Using this structured grid data format as input, DL methods can be applied to achieve segmentation and classification. This method was explored by [86], who proposed VoxNet, which is a novel method that uses a 3D CNN to predict class labels directly from the occupancy of the 3D grid. Although this method solved the problem of point clouds' non-structure, it proved to be computationally demanding, inefficient during the training phase and subject to significant information loss. These limitations were addressed by [87] in their works, which resulted in PointGrid, a method that used the same voxel transformation but improved the 3DCNN to consider grid cells with fixed positions to classify higher geometrical detail. Overall, this method significantly improved the training time and computation issues, whilst maintaining an accuracy of 92%.

To address the sparsity of point clouds, [88] developed a spare convolutional network, which was proven to work in a segmentation task [89]. The main high computation limitation of this method was addressed by [90], which led to various spatial partitioning and sampling such as k-d tree and octree [91], [92], [93]. However, the drawback of these methods is that only voxel boundary is considered, and the geometric structure of the local region is overlooked. To resolve this limitation, [94] proposed SegCloud, a method which combines DL and ML methods (3D-FCNN, trilinear interpolation and conditional random fields) to classify point cloud data effectively. Further studies investigate alternative point cloud transformation methods such as 3DmFV-Net, which uses a modified 3D Fisher Vector representation as CNN input [95]. In contrast, [96] transformed the point cloud into 3D Hough space and, with the additional step of feature generation, applied a novel 3D CNN (2D&3DHNet) that achieved 97.6% accuracy on the Sydney Urban Object dataset.

5.4 *Point-based Methods*

The indirect approaches described in the previous parts of section 5 address the unstructured characteristic of point clouds by applying various transformations. On the other hand, the idea of a direct approach was also explored by multiple studies, where the proposed networks use the raw point data and its characteristics. The pioneering framework to directly learn from and classify point clouds, PointNet [97] addresses the problem of sparsity, permutation invariance and transformation invariance by processing the raw dataset, applying a multi-layer perception to extract independent point features, aggregating computed information via maximum pooling layer to obtain global features and spatially align the point clouds using a transformation matrix.

One of the major drawbacks of this method is that it fails to consider relationships between points and their local neighbourhood information, which leads to significant data loss when dealing with large datasets. The PointNet framework can be interpreted as a backbone for various methods published thereafter, which explore improvements or alternatives driven by four main approaches. The rest of this section will discuss these approaches and their application to Cultural Heritage found in the available literature.

5.4.1 Point Ordering Methods

A method proposed by [98] uses an X-Conv operator that transforms each of the input points by reassigning and weighting their individual features, ultimately changing the order of the full dataset. As this method assigns weights from original point features and does not apply any changes to the overall dataset, the network benefits greatly from using convolution kernels that are used on the X-transformed features to improve computing efficiency. Although this method beats some of the state-of-the-art competitors with an accuracy of 92.2% on the ModelNet40 benchmark its application to building point cloud data remains unexplored.

[99] take a different stance on point cloud ordering in their RSNet framework, where a slice pooling layer projects the irregular point features into vectors suitable for RNN processing. This type of neural network is designed for a structured sequence; the feature vectors can be interpreted as timestamps. These are used to exploit data relationships and assign novel features for classification, which is ultimately achieved by the slice de-pooling layer. Similarly, to slice pooling, SO-Net applies self-organising mapping (SOM) to fix the position of points and compute features suitable for deep learning [100]. Although this method made significant contributions and addressed computation challenges, using a pre-training auto-encoder has drawbacks in that the encoder is not powerful enough to capture fine-grained geometrical features [64].

5.4.2 Multi-scale Methods

Researchers often use CNN to extract features of objects, which are heavily influenced by the receptive fields that dictate how much neighbouring information is considered for classification. To avoid the issues driven by the size of receptive fields, multi-scale methods are continually explored across the literature. [101] developed PointNet++ as an improvement to their original pioneering framework, which introduces a sampling layer and a grouping layer before classifying the dataset using PointNet. The idea of the additional layers is to construct local regions using several points as centroids that are paired with a local region growing module. All regions have a certain overlap, which then allows for feature detection at various scales. [102] applied PointNet++ to four different Cultural Heritage case study datasets as shown in Figure 13, that initially had to be translated, scaled, and subsampled as the typical point cloud dataset used for Scan-to-HBIM is too heavy for this CNN. In their study, it was only possible to classify four classes (arc, column, wall and window) due to the limited amount of object commonality in the training data. Their results proved that this network was only achieve an average F1 score of 30.8%. On the other hand, implementing PointNet across a sample of 18 buildings in Gaziantep, Turkey were successful with 83.3% prediction accuracy and 95.14% training accuracy [103]. Their study addressed the computational issues by using individual rooms as inputs and they've generated synthetic training material from HBIM models that led to an improvement in classification results in comparison to only using laser scan building data.

Figure 13: PointNet++ implementation for point cloud classification [102]

Further examples of multi-scale approaches can be found in studies by [104] and [105]. The former study proposes a pointwise pyramid pooling module, which aggregates features of local neighbourhoods at different scales. Meanwhile, the RNN uses the computed features to learn the spatial context to achieve the fusion of semantic features with multiple levels. The latter method initially fuses the features learned at different scales, which is followed by aggregation and a combination of the local and global features. Finally, this workflow computes a winning score that proves to improve overall classification accuracy. Both methods show great potential and were tested on the S3DIS benchmark, achieving on-par results, but neither has been yet applied in the Cultural Heritage context.

5.4.3 Feature Fusion Methods

Combining features from different neural network layers or branches is an omnipresent module of Deep Learning methods. Although global and local feature fusion on raw point clouds found in PointNet networks is crucial to scene understanding, the same idea of feature fusion can be applied to the description of complex building elements assembled out of multiple individual shapes.

[106] used the idea of a scale-invariant feature transform module used in 2D CNN to develop the PointSIFT module, which encodes the information of eight main directions into a coding unit that is stacked several times to compute additional features. The network architecture of their study is based on a common three-tier downscaling and upscaling procedure paired with the PointSIFT module at each level, which proved to improve the overall accuracy by 10% whilst compared to PointNet on the

S3DIS dataset. Following a similar ideology, [107] developed a similarity group proposal network (SGPN) that computes three additional feature scores to all points based on their similarity matrix, confidence map and semantic segmentation map. This method uses a grouping procedure for all computed features to generate class results that are on par with point net in the context of interior building datasets.

Alternative methods to local and global feature fusion are also popular in the literature. A novel annular convolution module proposed by [108] extracts local neighbourhood features around each point using a K-NN search algorithm and a dilated ring technique. The local features are fused with the global features at the down-sampling and up-sampling stages of the network influencing the segmentation and classification outputs. In a study by [109] a hierarchal learning k-d tree structure is tested for feature fusion at a local and global level to encapsulate latent relations between regions.

This improved fusion method slightly outperforms both PointNet & PointNet++ on the S3DIS dataset. Further variations of point cloud feature fusion methods can be found in publications by [110], [111] and [112] but going beyond exploring novel features computation or fusion, the work of [113] improved the fundamental network structure, which reduces processing power required and outperforms the current state-of-the-arts. Although this is a very active area of research, no research is found for use in the Cultural Heritage point cloud as an input or case study.

5.4.4 Graph CNN Fusion Methods

The application of graphs used for establishing structure in sparse input datasets has a strong presence across the literature as it not only establishes relationships between neighbouring points but also considers boundary features. Thus, the CNN-inspired GCNN was proven to work on point cloud input datasets by [114], [115] proposed a significant advancement with their DGCNN architecture, where graphs used at each convolution layer are continually refined. This method is very similar to PointNet but it replaces multi-layer perceptions with edge convolutions that extract features of graph centre points, their edge vectors and KNN. In contrast, [116] retained the multi-layer perception and added graphical attention mechanisms, which are used to learn the local geometric information. DGCNN and GAPNet are proven to outperform PointNet achieving 92.2% and 92.4% accuracy respectively on smaller, object-scale dataset such as the ModelNet40. [117] also propose a novel approach by applying a superpoint graph method for semantic segmentation. Their method overcomes the challenges encountered when processing large point cloud datasets by introducing an unsupervised process of partitioning the input dataset into simple, yet meaningful shapes. The computed superpoint consists of seven novel features also using PointNet for classification, which lead to 85.5% accuracy on the S3DIS dataset, which consists of urban and rural scenes. Contrary to addressing large scenes or the entire built asset, classification of vault types is addressed using PointNet and DGCNN across three different experiments focused on medieval vaults [118]. Their approach utilised synthetically generated vaults generated by procedural modelling which were reverse engineered into point clouds and used for training data for the DNNs. This method achieved top accuracy of 69% on certain vault classes.

Application of edge convolutions has been explored beyond the foundation laid by [115], which led to DGCNN-mod taking into consideration of additional features such as point normal and HSV values

[119]. 3DLEBNet, a novel method proposed by [120] combines a DGCNN encoder, which computes a "codeword" feature with a folding decoder inspired by FoldingNet. Through a process of linking the codewords with the edge convolution outputs, semantic segmentation was achieved with an accuracy of up to 77% using the ArCH dataset as shown in Figure 14.

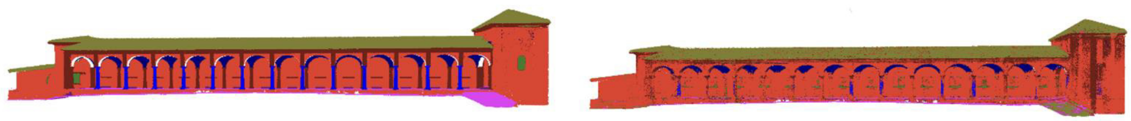


Figure 14: Semantic Segmentation with 3DLEBNet [120]

This can be interpreted as one of the DL state-of-the-art methods applied in the Cultural Heritage context. One of the most impressive aspects of this method is that it only requires 10% of the data for training compared to the other methods extensively such as PCNN, PointNet or DGCNN. Furthermore, 3DLEBNets ablation study explored the effectiveness of each part of the network architecture to summarise how the increased number of features improve the classification results while the increase in features should also be paired with the increase in training data used.

In parallel with the aim of this paper, [10] undertook a study that compares classification approaches using both ML and DL methods in the Cultural Heritage context. This study applies various ML & DL methods to the ArCH dataset and proposes DGCNN-mod-3Dfeat, which considers additional handcrafted features extracted using machine learning. Overall, the proposed method achieves an impressive 82.2% F1 score, which is only beaten by the ML RF classifier. This proves that including more point cloud data features for each point, such as omni-variance, improves the performance of DGCNN methods.

6 DISCUSSION

It is evident that the research on automating HBIM using AI semantic segmentation methods strongly correlates with the BIM adoption levels across geographical regions and the wealth of Cultural Heritage built assets. Research on AI for automated HBIM continually innovates to improve the state-of-the-art methods and pioneer new workflows and the validation of proposed methods to unalike Cultural Heritage building case studies currently shows a lack of geographical bias and potential for global adoption. The increasing complexity, variation and volume of research publications focused on the topic of point cloud classification in the Cultural Heritage context also confirms that this is only the beginning of resolving this common challenge well acknowledged globally.

When ML and DL evolutionary processes are compared, it will be seen that ML methods are more widely applied and currently offer more accurate point cloud segmentation. This critical review of ML and DL methods, as illustrated in Table 1, shows that the RF (Random Forest) classifier is the most popular method, which can be enhanced as experimented in the MLMR example, where the detail of

Cultural Heritage building elements is classified at different scale levels to enhance the overall performance.

Table 1: ML and DL methods experimented in the cultural heritage context.

Method	Authors	IoU	F1 Score	Accuracy
Mesh classifier (ML)	Bassier et al., (2017) [56]			81%
MLMR (ML)	Teruggi et al., (2020) [57]			94% / 97%
MLMR (ML)	K. Zhang et al., (2022) [58]		93% / 96%	
RF Classifier (ML)	Grilli et al., (2021) [61]		80.55%	
RF Classifier (ML)	Croce, Caroti, et al., (2021) [7]		98.81%	
SplatNet (DL)	Boulch et al., (2018) [79]	85.4%		
2D to 3D Label (DL)	Pellis et al., (2022)			87% / 90%
MVCNN (DL)	Pellis et al., (2022) [85]			33.4% / 57.3%
HPointNet++ (DL)	Malinverni et al., (2019) [102]		30.8%	
3DLEBNet (DL	Cao & Scaioni, (2021) [120]			67% / 77%
DGCNN-mod-3Dfeat (ML&DL)	Matrone et al., (2020) [10]		82.2%	

It is considered that this rational and systematic approach is not only found to be inspirational in the sense of how well it mimics human intelligence but also seems to have a strong potential to be aligned with BIM Level of Details and various building ontologies.

In terms of DL, the more limited segmentation results do seem to lead researchers on a path to experiment with a wider variety of methods and approaches. The fundamental PointNet and 2D pioneered CNN network architectures are found to be most influential for experimentation and paved the way for multi-view and graph-based methods that are already demonstrated in the Cultural Heritage context.

On the other hand, point ordering and voxelization methods are found to be commonly criticised for their high computational requirements and loss of information, which questions their suitability on larger input datasets with more complex classification needs. Although ML has superior results in comparison to DL, both have their strengths and weaknesses, and it would be incorrect to compare both methods equally. The improved results of ML come at the cost of more supervision and manual data labelling on input datasets. At the same time, ML offers more scope for customization of features and classes considered. Hence, it is very suitable for the Cultural Heritage domain. Furthermore, ML is found to be less dependent on training data size, generally easier to operate and take less time to train. Contrastingly DL methods require higher operational skillsets due to the more complex algorithm structure, which counterbalances in better interpretability and improved feature engineering possibilities. In general, DL offers a higher level of automation and once perfected, it could revolutionize the Scan-to-BIM process globally.

Regardless of the method applied for point cloud segmentation, training data and benchmarks tailored specifically for Cultural Heritage are currently limited. The only suitable dataset found across the literature, ArCH dataset, only contains 17 scenes labelled with 10 different classes [121]. This introduces a labour-intensive labelling stage and presents a significant challenge to network accuracy benchmarking for researchers.

Figure 15: Benchmark data for algorithm training in point cloud segmentation [121]

There is no superior approach to semantic point cloud segmentation, which subsequently limits the progress of HBIM automatic geometry generation. Out of the limited studies, the approach used by, [7] uses a library of parametric components, which are fitted to the point cloud segments with RANSAC to achieve the best outcome. Nevertheless, a lack of standard Cultural Heritage building ontology and specific limitations of BIM software APIs remains a challenge for future studies to address the full automation of the HBIM development process.

7 CONCLUSION

This paper presents a critical comprehensive review of existing point cloud segmentation methods applicable and applied across Cultural Heritage to automate the generation of Heritage Building Information Models. Firstly, a quantitative insight found across available literature is documented and the fundamental principles of machine learning methods are articulated. Several ML and HBIM studies are discussed regarding their context and segmentation results/success. This is followed by the description of the multiple DL types and techniques applied to point cloud files. Furthermore, their approach, results and potential for CH applications are reviewed.

The segmentation of Cultural Heritage point clouds is currently topical and thanks to the development of many improvements/advancements in ML & DL methods some early trends for success are emerging across the literature. Nonetheless, the current state-of-the-art still faces significant challenges, and the diversity of Cultural Heritage built assets still has not been tested on an international level. This paper reviewed the current research on AI-based HBIM automation in an exploratory nature and endeavours to inspire novel methods to leverage AI-based methods for HBIM automation.

Of course, the intellectual importance of understanding how (any) heritage-related research sits in relation to heritage itself has been explored. For example, a need is identified to better articulate the ways in which heritage studies have been served by strands of enquiry which exist outside heritage (but which may be relevant within), in heritage study and 'about' heritage (perhaps utilising metadata, or in applied HBIM, for example) [122]. In the context of heritage practice, as noted in the review, use of the outputs from digital laser scanning of objects being applied in cataloguing and representation processes it has become almost ubiquitous. On the other hand, using HBIM that same technology holds the potential to be utilised to help us understand the meaning of heritage to current generations, whilst similarly later recognising how the societal, economic, and practical forces that would have initially given rise to heritage objects, buildings or landscapes may be long gone or change [123].

The future work will focus on testing the discussed state-of-the-art classification methods on the UK case study Cultural Heritage dataset to formulate a novel AI-based implementation approach for point cloud semantic segmentation. The key lessons from this review paper will feed into the research to address the vital challenge of the laborious and expensive generation of HBIM so that digital heritage becomes a more meaningful and inclusive practice for heritage stakeholders.

REFERENCES

[1] R. Volk, J. Stengel, and F. Schultmann, "Building Information Modeling (BIM) for existing buildings - Literature review and future needs," *Automation in Construction*, vol. 38. pp. 109–127, Mar. 2014. doi: 10.1016/j.autcon.2013.10.023.

[2] D. Oreni, "From 3D Content Models to HBIM for Conservation and Management of Built Heritage," 2013.

[3] M. Murphy, E. McGovern, and S. Pavia, "Historic Building Information Modelling - Adding intelligence to laser and image-based surveys of European classical architecture," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 76, pp. 89–102, 2013, doi: 10.1016/j.isprsjprs.2012.11.006.

[4] Y. Arayici, J. Counsell, L. Mahdjoubi, G. Nagy, S. Hawas, and K. Dweidar, "Heritage Building Information Modelling," London, Feb. 2017.

[5] K. Pexman, D. D. Lichti, and P. Dawson, "Automated storey separation and door and window extraction for building models from complete laser scans," *Remote Sens (Basel)*, vol. 13, no. 17, Sep. 2021, doi: 10.3390/rs13173384.

[6] Y. Ji, Y. Dong, M. Hou, Y. Qi, and A. Li, "An extraction method for roof point cloud of ancient building using deep learning framework," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, International Society for Photogrammetry and Remote Sensing, Sep. 2021, pp. 321–327. doi: 10.5194/isprs-Archives-XLVI-M-1-2021-321-2021.

[7] V. Croce, G. Caroti, L. de Luca, K. Jacquot, A. Piemonte, and P. Véron, "From the Semantic Point Cloud to Heritage-Building Information Modeling: A Semiautomatic Approach Exploiting Machine Learning," 2021, doi: 10.3390/rs.

[8] W. S. Mcculloch, "A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY* n," 1943.

[9] L. Ceccarelli, M. G. Bevilacqua, G. Caroti, R. B. F. Castiglia, and V. Croce, "Semantic segmentation through Artificial Intelligence from raw point clouds to H-BIM representation," *DISEGNARECON*, vol. 16, no. 30, pp. 171–178, 2023, doi: 10.20365/disegnarecon.30.2023.17.

[10] F. Matrone, E. Grilli, M. Martini, M. Paolanti, R. Pierdicca, and F. Remondino, "Comparing machine and deep learning methods for large 3D heritage semantic segmentation," *ISPRS Int J Geoinf*, vol. 9, no. 9, Sep. 2020, doi: 10.3390/ijgi9090535.

[11] F. Banfi, "BIM orientation: Grades of generation and information for different type of analysis and management process," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, International Society for Photogrammetry and Remote Sensing, Aug. 2017, pp. 57–64. doi: 10.5194/isprs-archives-XLII-2-W5-57-2017.

[12] M. Capone and E. Lanzara, "SCAN-TO-BIM vs 3D IDEAL MODEL HBIM: PARAMETRIC TOOLS to STUDY DOMES GEOMETRY," in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Copernicus GmbH, Jan. 2019, pp. 219–226. doi: 10.5194/isprs-archives-XLII-2-W9-219-2019.

[13] P. Jouan and P. Hallot, "DIGITAL TWIN: A HBIM-BASED METHODOLOGY to SUPPORT PREVENTIVE CONSERVATION of HISTORIC ASSETS THROUGH HERITAGE SIGNIFICANCE AWARENESS," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, International Society for Photogrammetry and Remote Sensing, Aug. 2019, pp. 609–615. doi: 10.5194/isprs-archives-XLII-2-W15-609-2019.

[14] E. Valero *et al.*, "High level-of-detail BIM and machine learning for automated masonry wall defect surveying," in *35th International Symposium on Automation and Robotics in Construction and International AEC/FM Hackathon: The Future of Building Things, ISARC 2018*, International Association for Automation and Robotics in Construction I.A.A.R.C), 2018. doi: 10.22260/isarc2018/0101.

[15] C. Thomson and J. Boehm, "Automatic geometry generation from point clouds for BIM," *Remote Sens (Basel)*, vol. 7, no. 9, pp. 11753–11775, 2015, doi: 10.3390/rs70911753.

[16] S. Bruno, M. De Fino, and F. Fatiguso, "Historic Building Information Modelling: performance assessment for diagnosis-aided information modelling and management," *Automation in Construction*, vol. 86. Elsevier B.V., pp. 256–276, Feb. 01, 2018. doi: 10.1016/j.autcon.2017.11.009.

[17] D. Oreni, G. Karimi, and L. Barazzetti, "Applying bim to built heritage with complex shapes: The ice house of filarete's ospedale maggiore in milan, Italy," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, International Society

for Photogrammetry and Remote Sensing, Aug. 2017, pp. 553–560. doi: 10.5194/isprs-archives-XLII-2-W5-553-2017.

[18] L. Wang, Y. Liu, S. Zhang, J. Yan, and P. Tao, "Structure-aware convolution for 3D point cloud classification and segmentation," *Remote Sens (Basel)*, vol. 12, no. 4, Feb. 2020, doi: 10.3390/rs12040634.

[19] H. Snyder, "Literature review as a research methodology: An overview and guidelines," *J Bus Res*, vol. 104, pp. 333–339, Nov. 2019, doi: 10.1016/j.jbusres.2019.07.039.

[20] V. A. Cotella, "From 3D point clouds to HBIM: Application of Artificial Intelligence in Cultural Heritage," *Automation in Construction*, vol. 152. Elsevier B.V., Aug. 01, 2023. doi: 10.1016/j.autcon.2023.104936.

[21] N. Abreu, A. Pinto, A. Matos, and M. Pires, "Procedural Point Cloud Modelling in Scan-to-BIM and Scan-vs-BIM Applications: A Review," *ISPRS International Journal of Geo-Information*, vol. 12, no. 7. Multidisciplinary Digital Publishing Institute (MDPI), Jul. 01, 2023. doi: 10.3390/ijgi12070260.

[22] L. Groat and D. Wang, "Architectural Research Methods," 2013.

[23] B. Awuzie and P. McDermott, "An abductive approach to qualitative built environment research: A viable system methodological exposé," *Qualitative Research Journal*, vol. 17, no. 4, pp. 356–372, 2017, doi: 10.1108/QRJ-08-2016-0048.

[24] E. Grilli and F. Remondino, "Classification of 3D digital heritage," *Remote Sens (Basel)*, vol. 11, no. 7, Apr. 2019, doi: 10.3390/RS11070847.

[25] E. Grilli, F. Menna, and F. Remondino, "A review of point clouds segmentation and classification algorithms," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, International Society for Photogrammetry and Remote Sensing, Feb. 2017, pp. 339–344. doi: 10.5194/isprs-archives-XLII-2-W3-339-2017.

[26] S. Xia, D. Chen, R. Wang, J. Li, and X. Zhang, "Geometric Primitives in LiDAR Point Clouds: A Review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13. Institute of Electrical and Electronics Engineers, pp. 685–707, 2020. doi: 10.1109/JSTARS.2020.2969119.

[27] T. Rabbani, F. A. van den Heuvel, and G. Vosselman, "Segmentation of point clouds using smoothness constraint," 2006. [Online]. Available: https://www.researchgate.net/publication/228340970

[28] P. J. Besl and R. C. Jain, "Segmentation Through Variable-Order Surface Fitting," *IEEE Trans Pattern Anal Mach Intell*, vol. 10, no. 2, pp. 167–192, 1988, doi: 10.1109/34.3881.

[29] G. Vosselman *et al.*, "Recognising structure in laser scanner point clouds," 2004. [Online]. Available: https://www.researchgate.net/publication/228875768

[30] D. Ballard, "Generalizing the Hough Transform to detect arbitrary shapes," 1980.

[31] M. A. Fischler and R. C. Bolles, "Graphics and Image Processing Random Sample Consensus: A Paradigm for Model Fitting with Apphcatlons to Image Analysis and Automated Cartography," 1981.

[32] F. Tarsha-Kurdi, T. Landes, F. Tarsha-Kurdi, T. Landes, and P. Grussenmeyer, "Extended RANSAC algorithm for automatic detection of building roof planes from LiDAR DATA," 2008. [Online]. Available: https://www.researchgate.net/publication/228781826

[33] D. Chen, L. Zhang, P. T. Mathiopoulos, and X. Huang, "A methodology for automated segmentation and reconstruction of urban 3-D buildings from ALS point clouds," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 7, no. 10, pp. 4199–4217, Oct. 2014, doi: 10.1109/JSTARS.2014.2349003.

[34] K. Pleansamai and K. Chaiyasarn, "M-estimator Sample Consensus planar extraction from image-based 3D point cloud for building information modelling," *International Journal of GEOMATE*, vol. 17, no. 63, pp. 69–76, 2019, doi: 10.21660/2019.63.09667.

[35] P. H. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," 1996.

[36] R. A. Fisher, "THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS," *Ann Eugen*, vol. 7, no. 2, pp. 179–188, Sep. 1936, doi: 10.1111/j.1469-1809.1936.tb02137.x.

[37] C. Cabo, C. Ordóñez, F. Sáchez-Lasheras, J. Roca-Pardiñas, and J. de Cos-Juez, "Multiscale supervised classification of point clouds with urban and forest applications," *Sensors (Switzerland)*, vol. 19, no. 20, Oct. 2019, doi: 10.3390/s19204523.

[38] D. R. Cox, "The Regression Analysis of Binary Sequences," 1958. [Online]. Available: https://about.jstor.org/terms

[39] J. Zhang, X. Lin, and X. Ning, "SVM-Based classification of segmented airborne LiDAR point clouds in urban areas," *Remote Sens (Basel)*, vol. 5, no. 8, pp. 3749–3775, 2013, doi: 10.3390/rs5083749.

[40] W. C. Lu, X. B. Ji, M. J. Li, L. Liu, B. H. Yue, and L. M. Zhang, "Using support vector machine for materials design," *Adv Manuf*, vol. 1, no. 2, pp. 151–159, 2013, doi: 10.1007/s40436-013-0025-2.

[41] A. Adan and D. Huber, "3D reconstruction of interior wall surfaces under occlusion and clutter," in *Proceedings - 2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, 3DIMPVT 2011*, 2011, pp. 275–281. doi: 10.1109/3DIMPVT.2011.42.

[42] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, "Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier," *International Journal of Information Engineering and Electronic Business*, vol. 8, no. 4, pp. 54–62, Jul. 2016, doi: 10.5815/ijieeb.2016.04.07.

[43] W. Lou, X. Wang, F. Chen, Y. Chen, B. Jiang, and H. Zhang, "Sequence based prediction of DNA-binding proteins based on hybrid feature selection using random forest and Gaussian naïve Bayes," *PLoS One*, vol. 9, no. 1, Jan. 2014, doi: 10.1371/journal.pone.0086703.

[44] Z. Kang, J. Yang, and R. Zhong, "A Bayesian-Network-Based Classification Method Integrating Airborne LiDAR Data with Optical Images," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 10, no. 4, pp. 1651–1661, Apr. 2017, doi: 10.1109/JSTARS.2016.2628775.

[45] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning," Cambridge, 2016.

[46] L. Breiman, "Random Forests," 2001.

[47] D. Xue, Y. Cheng, X. Shi, Y. Fei, and P. Wen, "An Improved Random Forest Model Applied to Point Cloud Classification," in *IOP Conference*

Series: Materials Science and Engineering, Institute of Physics Publishing, Mar. 2020. doi: 10.1088/1757-899X/768/7/072037.

[48] J. Gong, Q. Zhu, R. Zhong, Y. Zhang, and X. Xie, "An efficient point cloud management method based on a 3D R-tree," *Photogramm Eng Remote Sensing*, vol. 78, no. 4, pp. 373–381, 2012, doi: 10.14358/PERS.78.4.373.

[49] J. MacQueen, "SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS," 1967.

[50] S. P. Lloyd, "Least Squares Quantization in PCM," 1982.

[51] G. Lavoué, F. Dupont, and A. Baskurt, "A new CAD mesh segmentation method, based on curvature tensor analysis," *CAD Computer Aided Design*, vol. 37, no. 10, pp. 975–987, Sep. 2005, doi: 10.1016/j.cad.2004.09.001.

[52] Y. Liu and Y. Xiong, "Automatic segmentation of unorganized noisy point clouds based on the Gaussian map," *CAD Computer Aided Design*, vol. 40, no. 5, pp. 576–594, May 2008, doi: 10.1016/j.cad.2008.02.004.

[53] D. Comaniciu, P. Meer, and S. Member, "Mean Shift: A Robust Approach Toward Feature Space Analysis," 2002.

[54] R. A. Kuçak, E. Özdemir, and S. Erol, "The segmentation of point clouds with K-Means and Ann (Artifical Neural Network)," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, International Society for Photogrammetry and Remote Sensing, May 2017, pp. 595–598. doi: 10.5194/isprs-archives-XLII-1-W1-595-2017.

[55] L. Xiaohu, Y. Jian, J. Tu, K. Li, L. Li, and Y. Liu, "PAIRWISE LINKAGE FOR POINT CLOUD SEGMENTATION," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. III–3, pp. 201–208, Jun. 2016, doi: 10.5194/isprsannals-iii-3-201-2016.

[56] M. Bassier, M. Vergauwen, and B. van Genechten, "AUTOMATED CLASSIFICATION of HERITAGE BUILDINGS for AS-BUILT BIM USING MACHINE LEARNING TECHNIQUES," in *26th International CIPA Symposium on Digital Workflows for Heritage Conservation 2017*, J. Hayes, C. Ouimet, S. Q. M, S. Fai, and L. Smith, Eds., Copernicus GmbH, 2017, pp. 25–30. doi: 10.5194/isprs-annals-IV-2-W2-25-2017.

[57] S. Teruggi, E. Grilli, M. Russo, F. Fassi, and F. Remondino, "A hierarchical machine learning approach for multi-level and multi-resolution 3d point cloud classification," *Remote Sens (Basel)*, vol. 12, no. 16, Aug. 2020, doi: 10.3390/RS12162598.

[58] K. Zhang, S. Teruggi, and F. Fassi, "MACHINE LEARNING METHODS FOR UNESCO CHINESE HERITAGE: COMPLEXITY AND COMPARISONS," in *9th International Workshop on 3D Virtual Reconstruction and Visualization of Complex Architectures, 3D-ARCH 2022*, L. Fregonese, F. Fassi, and F. Remondino, Eds., International Society for Photogrammetry and Remote Sensing, 2022, pp. 543–550. doi: 10.5194/isprs-archives-XLVI-2-W1-2022-543-2022.

[59] E. Grilli and F. Remondino, "Classification of 3D digital heritage," *Remote Sens (Basel)*, vol. 11, no. 7, Apr. 2019, doi: 10.3390/RS11070847.

[60] F. Poux, C. Mattes, and L. Kobbelt, "UNSUPERVISED SEGMENTATION of INDOOR 3D POINT CLOUD: APPLICATION to OBJECT-BASED CLASSIFICATION," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, International Society for Photogrammetry and Remote Sensing, Sep. 2020, pp. 111–118. doi: 10.5194/isprs-archives-XLIV-4-W1-2020-111-2020.

[61] E. Grilli, F. Poux, and F. Remondino, "Unsupervised object-based clustering in support of supervised point-based 3D point cloud classification," in *2021 24th ISPRS Congress Commission II: Imaging Today, Foreseeing Tomorrow*, N. Paparoditis, C. Mallet, F. Lafarge, M. Y. Yang, A. Yilmaz, J. D. Wegner, J. D. Wegner, F. Remondino, T. Fuse, and I. Toschi, Eds., International Society for Photogrammetry and Remote Sensing, 2021, pp. 471–478. doi: 10.5194/isprs-archives-XLIII-B2-2021-471-2021.

[62] E. Grilli, E. Özdemir, and F. Remondino, "Application of machine and deep learning strategies for the classification of heritage point clouds," in *ISPRS International GeoSpatial Conference 2019, Joint Conferences of 5th Sensors and Models in Photogrammetry and Remote Sensing, SMPR 2019 and 3rd Geospatial Information Research, GI Research 2019*, H. Arefi and S. S. M, Eds., International Society for Photogrammetry and Remote Sensing, 2019, pp. 447–454. doi: 10.5194/isprs-archives-XLII-4-W18-447-2019.

[63] V. Croce, G. Caroti, A. Piemonte, L. De Luca, and P. Véron, "H-BIM and Artificial Intelligence: Classification of Architectural Heritage for Semi-Automatic Scan-to-BIM Reconstruction," *Sensors*, vol. 23, no. 5, Mar. 2023, doi: 10.3390/s23052497.

[64] J. Zhang, X. Zhao, Z. Chen, and Z. Lu, "A Review of Deep Learning-Based Semantic Segmentation for Point Cloud," *IEEE Access*, vol. 7. Institute of Electrical and Electronics Engineers Inc., pp. 179118–179133, 2019. doi: 10.1109/ACCESS.2019.2958671.

[65] W. Liu, J. Sun, W. Li, T. Hu, and P. Wang, "Deep learning on point clouds and its application: A survey," *Sensors (Switzerland)*, vol. 19, no. 19. MDPI AG, Oct. 01, 2019. doi: 10.3390/s19194188.

[66] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," 2012. [Online]. Available: http://code.google.com/p/cuda-convnet/

[67] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.1556

[68] C. Szegedy *et al.*, "Going Deeper with Convolutions," Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.4842

[69] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2015, [Online]. Available: http://arxiv.org/abs/1512.03385

[70] S. Taoufiq, B. Nagy, and C. Benedek, "Hierarchynet: Hierarchical CNN-based urban building classification," *Remote Sens (Basel)*, vol. 12, no. 22, pp. 1–20, Nov. 2020, doi: 10.3390/rs12223794.

[71] F. Rodrigues, V. Cotella, H. Rodrigues, E. Rocha, F. Freitas, and R. Matos, "Application of Deep Learning Approach for the Classification of Buildings' Degradation State in a BIM Methodology," *Applied Sciences*, vol. 12, no. 15, p. 7403, Jul. 2022, doi: 10.3390/app12157403.

[72] R. Socher, B. Huval, B. Bhat, C. D. Manning, and A. Y. Ng, "Convolutional-Recursive Deep Learning for 3D Object Classification," 2012. [Online]. Available: www.socher.org.

[73] A. Eitel, J. T. Springenberg, L. Spinello, M. Riedmiller, and W. Burgard, "Multimodal Deep Learning for Robust RGB-D Object Recognition," Jul. 2015, [Online]. Available: http://arxiv.org/abs/1507.06821

[74] C. Couprie, C. Farabet, L. Najman, and Y. LeCun, “Indoor Semantic Segmentation using depth information,” Jan. 2013, [Online]. Available: http://arxiv.org/abs/1301.3572

[75] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, “Learning Hierarchical Features for Scene Labeling,” 2013.

[76] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” Nov. 2013, [Online]. Available: http://arxiv.org/abs/1311.2524

[77] J. Balado, P. Arias, L. Díaz-Vilariño, and L. M. González-Desantos, “Automatic CORINE land cover classification from airborne LIDAR data,” in *Procedia Computer Science*, Elsevier B.V., 2018, pp. 186–194. doi: 10.1016/j.procS.2018.07.222.

[78] X. Pan, L. Gao, A. Marinoni, B. Zhang, F. Yang, and P. Gamba, “Semantic labeling of high resolution aerial imagery and LiDAR data with fine segmentation network,” *Remote Sens (Basel)*, vol. 10, no. 5, May 2018, doi: 10.3390/rs10050743.

[79] A. Boulch, J. Guerry, B. le Saux, and N. Audebert, “SNAPNET: 3D POINT CLOUD SEMANTIC LABELING WITH 2D DEEP SEGMENTATION NETWORKS,” 2018.

[80] Y. Xu, T. Fan, M. Xu, L. Zeng, and Y. Qiao, “SpiderCNN: Deep Learning on Point Sets with Parameterized Convolutional Filters,” Mar. 2018, [Online]. Available: http://arxiv.org/abs/1803.11527

[81] E. Pellis, A. Murtiyoso, A. Masiero, G. Tucci, M. Betti, and P. Grussenmeyer, “2D TO 3D LABEL PROPAGATION FOR THE SEMANTIC SEGMENTATION OF HERITAGE BUILDING POINT CLOUDS,” in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, International Society for Photogrammetry and Remote Sensing, May 2022, pp. 861–867. doi: 10.5194/isprs-archives-XLIII-B2-2022-861-2022.

[82] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, “Multi-view Convolutional Neural Networks for 3D Shape Recognition,” 2015. [Online]. Available: http://vis-www.cs.umass.edu/mvcnn.

[83] K. Bacharidis, F. Sarri, V. Paravolidakis, L. Ragia, and M. Zervakis, “Fusing georeferenced and stereoscopic image data for 3D building Facade reconstruction,” *ISPRS Int J Geoinf*, vol. 7, no. 4, Apr. 2018, doi: 10.3390/ijgi7040151.

[84] K. Bacharidis, F. Sarri, and L. Ragia, “3D building façade reconstruction using deep learning,” *ISPRS Int J Geoinf*, vol. 9, no. 5, May 2020, doi: 10.3390/ijgi9050322.

[85] E. Pellis, A. Murtiyoso, A. Masiero, G. Tucci, M. Betti, and P. Grussenmeyer, “AN IMAGE-BASED DEEP LEARNING WORKFLOW FOR 3D HERITAGE POINT CLOUD SEMANTIC SEGMENTATION,” in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, International Society for Photogrammetry and Remote Sensing, Feb. 2022, pp. 429–434. doi: 10.5194/isprs-archives-XLVI-2-W1-2022-429-2022.

[86] D. Maturana and S. Scherer, “VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition,” 2015.

[87] T. Le and Y. Duan, “PointGrid: A Deep Network for 3D Shape Understanding,” 2018.

[88] B. Graham, “Spatially sparse convolutional neural networks,” 2014.

[89] F. Verdoja, D. Thomas, and A. Sugimoto, “Fast 3D point cloud segmentation using supervoxels with geometry and color for 3D scene understanding,” in *Proceedings - IEEE International Conference on Multimedia and Expo*, IEEE Computer Society, Aug. 2017, pp. 1285–1290. doi: 10.1109/ICME.2017.8019382.

[90] Y. Li, S. Pirk, H. Su, C. R. Qi, and L. J. Guibas, “FPNN: Field Probing Neural Networks for 3D Data,” 2016.

[91] G. Riegler, A. Osman Ulusoy, and A. Geiger, “OctNet: Learning Deep 3D Representations at High Resolutions,” 2017. [Online]. Available: https://3dwarehouse.sketchup.com

[92] M. Tatarchenko, A. Dosovitskiy, and T. Brox, “Octree Generating Networks: Efficient Convolutional Architectures for High-resolution 3D Outputs,” 2017. [Online]. Available: https://github.com/lmb-freiburg/ogn

[93] R. Klokov and V. Lempitsky, “Escape from Cells: Deep Kd-Networks for the Recognition of 3D Point Cloud Models,” 2017.

[94] L. Tchapmi, C. Choy, I. Armeni, J. Gwak, and S. Savarese, “SEGCloud: Semantic segmentation of 3D point clouds,” in *Proceedings - 2017 International Conference on 3D Vision, 3DV 2017*, Institute of Electrical and Electronics Engineers Inc., May 2018, pp. 537–547. doi: 10.1109/3DV.2017.00067.

[95] Y. Ben-Shabat, M. Lindenbaum, and A. Fischer, “3D Point Cloud Classification and Segmentation using 3D Modified Fisher Vector Representation for Convolutional Neural Networks,” Nov. 2017, [Online]. Available: http://arxiv.org/abs/1711.08241

[96] W. Song *et al.*, “2D&3DHNet for 3D Object Classification in LiDAR Point Cloud,” *Remote Sens (Basel)*, vol. 14, no. 13, Jul. 2022, doi: 10.3390/rs14133146.

[97] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, “PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation,” Dec. 2016, [Online]. Available: http://arxiv.org/abs/1612.00593

[98] Y. Li, †⇤ Rui Bu, M. Sun, W. Wu, X. Di, and B. Chen, “PointCNN: Convolution On X-Transformed Points,” 2018.

[99] Q. Huang, W. Wang, and U. Neumann, “Recurrent Slice Networks for 3D Segmentation of Point Clouds,” Feb. 2018, [Online]. Available: http://arxiv.org/abs/1802.04402

[100] J. Li, B. M. Chen, and G. H. Lee, “SO-Net: Self-Organizing Network for Point Cloud Analysis,” Mar. 2018, [Online]. Available: http://arxiv.org/abs/1803.04249

[101] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, “PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space,” Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.02413

[102] E. S. Malinverni *et al.*, “DEEP LEARNING for SEMANTIC SEGMENTATION of 3D POINT CLOUD,” in *International Archives of the*

Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, International Society for Photogrammetry and Remote Sensing, Aug. 2019, pp. 735–742. doi: 10.5194/isprs-archives-XLII-2-W15-735-2019.

[103] B. Haznedar, R. Bayraktar, A. E. Ozturk, and Y. Arayici, "Implementing PointNet for point cloud segmentation in the heritage context," *Herit Sci*, vol. 11, no. 1, Dec. 2023, doi: 10.1186/s40494-022-00844-w.

[104] J. Li, H. Huang, L. Du, and X. Zhang, "3D Recurrent Neural Networks with Context Fusion for Point Cloud Semantic Segmentation," 2018.

[105] Y. Ma, Y. Guo, Y. Lei, M. Lu, and J. Zhang, *3DMAX-Net: A Multi-Scale Spatial ContextualNetwork for 3D Point Cloud Semantic Segmentation*. 2018.

[106] M. Jiang, Y. Wu, T. Zhao, Z. Zhao, and C. Lu, "PointSIFT: A SIFT-like Network Module for 3D Point Cloud Semantic Segmentation," Jul. 2018, [Online]. Available: http://arxiv.org/abs/1807.00652

[107] W. Wang, R. Yu, Q. Huang, and U. Neumann, "SGPN: Similarity Group Proposal Network for 3D Point Cloud Instance Segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Dec. 2018, pp. 2569–2578. doi: 10.1109/CVPR.2018.00272.

[108] A. Komarichev, Z. Zhong, and J. Hua, "A-CNN: Annularly Convolutional Neural Networks on Point Clouds," Apr. 2019, [Online]. Available: http://arxiv.org/abs/1904.08017

[109] W. Zeng and T. Gevers, "3DContextNet: K-d Tree Guided Hierarchical Learning of Point Clouds Using Local and Global Contextual Cues," 2018.

[110] L. Liu, E. Chen, and Y. Ding, "TR-Net: A Transformer-Based Neural Network for Point Cloud Processing," *Machines*, vol. 10, no. 7, Jul. 2022, doi: 10.3390/machines10070517.

[111] Z. Zhang, T. Li, X. Tang, X. Lei, and Y. Peng, "Introducing Improved Transformer to Land Cover Classification Using Multispectral LiDAR Point Clouds," *Remote Sens (Basel)*, vol. 14, no. 15, Aug. 2022, doi: 10.3390/rs14153808.

[112] M. Gadelha, R. Wang, and S. Maji, "Multiresolution Tree Networks for 3D Point Cloud Processing," 2018.

[113] Q. Zheng, J. Sun, and W. Chen, "A Lightweight Network for Point Cloud Analysis via the Fusion of Local Features and Distribution Characteristics," *Sensors*, vol. 22, no. 13, Jul. 2022, doi: 10.3390/s22134742.

[114] Y. Zhang and M. Rabbat, "A Graph-CNN for 3D Point Cloud Classification," Nov. 2018, [Online]. Available: http://arxiv.org/abs/1812.01711

[115] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph Cnn for learning on point clouds," *ACM Trans Graph*, vol. 38, no. 5, Oct. 2019, doi: 10.1145/3326362.

[116] C. Chen, L. Z. Fragonara, and A. Tsourdos, "GAPNet: Graph Attention based Point Neural Network for Exploiting Local Feature of Point Cloud," May 2019, [Online]. Available: http://arxiv.org/abs/1905.08705

[117] L. Landrieu and M. Simonovsky, "Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs," Nov. 2017, [Online]. Available: http://arxiv.org/abs/1711.09869

[118] C. Battini, U. Ferretti, G. De Angelis, R. Pierdicca, M. Paolanti, and R. Quattrini, "Automatic generation of synthetic heritage point clouds: Analysis and segmentation based on shape grammar for historical vaults," *J Cult Herit*, vol. 66, pp. 37–47, Mar. 2024, doi: 10.1016/j.culher.2023.10.003.

[119] R. Pierdicca *et al.*, "Point cloud semantic segmentation using a deep learning framework for cultural heritage," *Remote Sens (Basel)*, vol. 12, no. 6, Mar. 2020, doi: 10.3390/rs12061005.

[120] Y. Cao and M. Scaioni, "3DLEB-net: Label-efficient deep learning-based semantic segmentation of building point clouds at LoD3 level," *Applied Sciences (Switzerland)*, vol. 11, no. 19, Oct. 2021, doi: 10.3390/app11198996.

[121] F. Matrone *et al.*, "A BENCHMARK for LARGE-SCALE HERITAGE POINT CLOUD SEMANTIC SEGMENTATION," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, International Society for Photogrammetry and Remote Sensing, Aug. 2020, pp. 1419–1426. doi: 10.5194/isprs-archives-XLIII-B2-2020-1419-2020.

[122] E. Waterton and S. Watson, "Framing theory: Towards a critical imagination in heritage studies," in *International Journal of Heritage Studies*, Sep. 2013, pp. 546–561. doi: 10.1080/13527258.2013.779295.

[123] S. Brand, *How buildings learn: what happens after they are built?* New York: Penguin, 1994.