Machine Learning, Extraction and Classification of Memorial Objects from 3D Point Cloud Data

Nicholas Ian Arnold
School of Computing and Communications
Lancaster University

A thesis submitted for the degree of
Doctor of Philosophy

August, 2023

Dedicated to my beloved wife, Lily. Who's support has been my constant source of strength and motivation. Your unwavering belief in me has guided me through every challenge and triumph. I am deeply grateful for your boundless love, unwavering encouragement and never ending thoughtfulness.

Declaration

I declare that the work presented in this thesis is, to the best of my knowledge and belief and except where specific reference is made to the works of others, original and my own work. The material has not been submitted, either in whole or in part, for a degree at this, or any other university. This thesis does not exceed the maximum permitted word length of 80,000 words including appendices and footnotes, but excluding the bibliography.

Nicholas Ian Arnold

Machine Learning, Extraction and Classification of Memorial Objects from 3D Point Cloud Data

Nicholas Ian Arnold.
School of Computing and Communications, Lancaster University
A thesis submitted for the degree of *Doctor of Philosophy*. August, 2023

Abstract

This thesis resides at the intersection of machine learning (ML) and cultural heritage, addressing the unique challenges posed by the analysis of 3D point cloud data from historic sites. The convergence of these sites with the modern age necessitates innovative approaches to their documentation, interpretation and management. Advances in remote sensing technologies, like LiDAR, offer a solution by enabling non-invasive, high-resolution digital representations of these sites in the form of point clouds. However, effectively utilising this digital medium for mapping and inventory remains challenging. This thesis bridges the gap by researching, designing and implementing an automated object extraction solution using ML and artificial intelligence (AI) techniques, emphasising explainability to enhance human learning from AI decision-making.

The core pursuit of this thesis revolves around the automatic extraction and classification of objects from point cloud data within cultural heritage sites. A series of interconnected articles form the foundation of this exploration. The initial article introduces GeoPart-Transfer for automated extraction and labelling of memorial objects. The second article presents XPCC, a prototype-based classification and visualisation method for point clouds. It embraces interpretability and adaptability, permitting continuous learning without extensive retraining. The final article presents GeoPart-XPCC, a comprehensive framework applied to multiple scenes from various burial ground sites.

Throughout these articles, a review of related research is undertaken, while experimentation adds empirical weight to the findings. The experimental results illustrate how these methodologies enhance the usability of digital point cloud data, offering broad applicability across different scenarios. By addressing the challenge of automatic object extraction from point cloud data, this thesis brings practical and impactful contributions to the domains of cultural heritage, archaeology and AI. Through comprehensive research, innovation, and experimentation, it illuminates the potential of digital technologies in preserving our historical heritage. It also lays a foundation for future explorations to other domains.

Publications

Several publications, shown below, have been created directly from the thesis. Chapter 2 was published in the peer-reviewed Journal of Computer Applications in Archaeology. Chapter 3 chapter was published in the IEEE Transactions on Artificial Intelligence. Chapter 4 is to be submitted to IEEE Transactions on Neural Networks and Learning Systems. Appendix 1 was published in WIREs Data Mining and Knowledge Discovery.

N. I. Arnold, P. P. Angelov, T. Viney, and P. M. Atkinson (Apr. 2021). "Automatic Extraction and Labelling of Memorial Objects From 3D Point Clouds". English. In: *Journal of Computer Applications in Archaeology* 4.1, pp. 79–93. DOI: 10.5334/jcaa.66

N. I. Arnold, P. Angelov, and P. M. Atkinson (2022). "An Improved eXplainable Point Cloud Classifier (XPCC)". in: *IEEE Transactions on Artificial Intelligence*, pp. 1–1. DOI: 10.1109/TAI.2022.3150647

Under revision:
N. I. Arnold, P. Angelov, and P. M. Atkinson (2023). "An Explainable and Transferable AI Framework for the Extraction of Burial Ground Objects". English. In: *IEEE Transactions on Neural Networks and Learning Systems*. ISSN: 2162-237X

The following publication has been generated while developing this thesis. My overall contributions as an author were the conceptualisation; formal analysis; investigation; methodology; writing-review and editing. My direct contribution within this publication is the entirety of Section 4, Review of the state-of-the-art. This review, has helped to guide the development of the methods and algorithms contained within this thesis. See below:

P. P. Angelov, E. A. Soares, R. Jiang, N. I. Arnold, and P. M. Atkinson (2021). "Explainable artificial intelligence: an analytical review". In: *WIREs Data Mining and Knowledge Discovery* 11.5, e1424. DOI: https://doi.org/10.1002/widm.1424

Acknowledgements

I express my profound appreciation to Pete Atkinson and Plamen Angelov for their invaluable insights and encouragement throughout the course of my research journey. Your expertise, patience and mentorship have been instrumental in shaping the direction of this work.

I wish to extend my sincere thanks to everyone at the Lancaster School of Computing and Communications for their support and provision of resources, which have created a conducive environment for my research pursuits. Your dedication to a culture of academic excellence has significantly contributed to my development as a researcher.

I extend my thanks to Atlantic Geomatics (UK) Ltd. for their generous support and collaboration, including the provision of valuable datasets that enriched the empirical aspects of this study.

Lastly, I wish to acknowledge the encouragement of my family, friends, and loved ones, who have been a constant source of motivation throughout this journey.

Thank you all.

Contents

Chapter 1

Introduction

This thesis sits at the interesting crossroads of machine learning (ML) and artificial intelligence (AI) applied to 3D point cloud data and cultural heritage management. Advances in remote sensing technologies, such as terrestrial and airborne light detection and ranging (LiDAR) systems, have led to the development of versatile mobile sensors capable of creating digital representations of heritage sites (Leica, n.d.; A. F. Chase et al., 2011). The massive point datasets, called point clouds, generated by these systems represent surface features by the spatial position of points in three-dimensions and accurately preserve intrinsic information relating to geometric shape, surface and depth. The surface features, or objects, of interest may include buildings and their type, vehicles and their type, trees and their functional type and so on. These object types define the categories or classes that objects can be grouped into. Point clouds representing real world environments provide fine-resolution digital proxies and offer a valuable data source for cultural heritage management and related fields such as archaeology (Bewley, Crutchley, and Shell, 2005; Spina et al., 2011; Ø. Trier et al., 2016; Gallwey et al., 2019). An important task for every heritage management body is the mapping and management of such objects for the purposes of curation, interpretation and assessment. However, the manual extraction and labelling of memorial objects from these point clouds is a labour-intensive process. Traditionally, it is necessary for the objects to be observed, recorded and processed manually by an operator (Kramer, 2015). Therefore, it is of great significance to automate these tasks.

1.1 Objectives and Focus

The primary objective of this thesis is to develop a robust and efficient system for the detection, classification, and extraction of memorial objects, such as headstones and grave markers, from 3D point cloud scenes. The study aims to bridge the gap

between advanced computer vision techniques and the preservation and management of burial sites, thereby offering valuable insights for cultural heritage preservation and maintenance.

Cemeteries and other cultural heritage sites serve as important repositories of cultural heritage, documenting historical narratives, genealogy, and the evolution of societal norms. However, the accurate identification, cataloguing, and maintenance of memorial objects within vast cemeteries remain a challenging and lab or-intensive task. This thesis proposes innovative approaches that utilises 3D point cloud data captured through LiDAR or structured light scanning technologies.

Subsequently, the thesis focuses on the development of novel detection and classification methods specifically tailored for memorial objects. The research explores the integration of deep learning architectures, such as convolutional neural networks (CNN) and point-based networks, to analyse and classify different types of memorial objects accurately. Additionally, the project investigates the incorporation of explainable artificial intelligence (XAI) techniques to enhance the accuracy and interoperability of the classification process.

Furthermore, the research addresses the challenge of segmenting individual memorial objects from the 3D point cloud scenes. The proposed segmentation algorithm aims to extract precise and well-defined segments of memorial objects, thus facilitating better visualisation, analysis, and maintenance.

The ultimate goal of this thesis is to create an automated system capable of efficiently detecting, classifying, and segmenting memorial objects in large-scale 3D point cloud scenes. The developed system is expected to significantly reduce the manual effort involved in cemetery maintenance, providing valuable insights for historical research, cultural heritage preservation, and urban planning.

By combining computer vision techniques with the field of cultural heritage, this research strives to contribute to the emerging interdisciplinary area of digital archaeology and open new avenues for the preservation and appreciation of our collective past.

1.2 Overview of Point Cloud Technology

Point clouds are produced by scanning processes and, as such, typically only represent the surface of a physical object or an area of interest. The basic form of point cloud is referred to as 3D point clouds, as each point is represented by its three-dimensional vector of coordinates (x, y, z). However, additional dimensional data may also be added. Observational data, such as colour values, (R, G, B), are typically added to aid in visualisation and sometimes used in processing. A point cloud may also contain surface normals (i.e., a vector that is perpendicular to the tangent plane line at a particular point on the underlying surface). Visually, the orientation of the surface

normals are used to apply the shading and other visual effects from light sources in computer graphics applications. The magnitude of the reflected signal from laser systems is known as intensity and is also commonly used.

Three common approaches to generating point cloud data are depth cameras, photogrammetry and LiDAR systems.

- Depth cameras are a special class of digital camera which combine a colour image sensor and a depth sensor. The colour image sensor captures red, green and blue (RGB) values, whereas the depth sensor captures per-pixel depth information (Endres et al., 2014). For this reason, they are often denoted as RGB-D cameras, where D is *depth*. By mapping the RGB images with the depth information, this technology can create a coloured point cloud.

- Photogrammetry is a method to obtain reliable information about physical objects and their environment from overlapping photographs of an object or area and converting them into a 3D digital model (Mikhail, Bethel, and McGlone, 2001).

- Light detection and ranging, or LiDAR, is a method wherein light pulses are measured to determine the distances and dimensions of physical objects or areas. The LiDAR sensor emits a pulse of light towards a target which is then reflected from the surface. A built-in receiver then detects the backscattered signal and calculates the distance light has travelled. The 3D position of where the signal was returned can then be calculated using the known position of the sensor, the direction in which the light was sent and, as previously mentioned, the distance. The inherent output format of LiDAR is a 3D point cloud. The point clouds derived from LiDAR are typically coloured by the elevation (z value) or the intensity value. RGB values can also be added by colourising the point cloud from a digital orthomosaic (Wolf, Dewitt, and Wilkinson, 2014). A significant benefit of LiDAR is that it uses an active sensor, meaning that it does not depend on auxiliary illumination such as the Sun. Therefore, it is possible to collect data through cloud cover and at night. Several common types of LiDAR system are as follows.

  - Airborne laser scanning (ALS) is a system of LiDAR which is installed to an aircraft, such as an aeroplane, helicopter or unmanned aerial vehicle (UAV), to create a point cloud model of the landscape. The use of an aircraft makes it possible to quickly cover a large footprint. This type of LiDAR is useful for creating detailed and accurate digital elevation models (Atkinson and Tatnall, 1997). Furthermore, the true ground surface can be extracted from the unprocessed point cloud to create a digital terrain model (Lloyd and Atkinson, 2006). Flying at higher altitudes usually reduces the

accuracy and point density of the point cloud but has the added benefit of covering even larger areas more quickly. Conversely, the increased height of airborne LiDAR results in poor coverage of oblique (vertical) faces of structures. This makes it ineffective at capturing complex environments with many small vertical objects, such as burial grounds.

- Terrestrial laser scanning (TLS) is a LiDAR system that is either stationary or mobile but that is connected to the Earth in some way. The stationary variety is commonly used as a survey method, where the LiDAR is fixed to a stationary device, such as a tripod, connected to the ground (Pfeifer and Briese, 2007). TLS can also be used to scan the interiors of structures (S. Lee, Majid, and Setan, 2013).
- Mobile LiDAR is a subdivision of TLS; however, the system is attached to a moving mount and the point cloud is then acquired along the path of travel. In comparison to ALS, Mobile LiDAR has a direct view of oblique surfaces but cannot easily capture the tops of structures. As the system is moving, the point density is more variable compared to the stationary TLS and the ALS systems. It is therefore common for two or more scanners to be used in combination to produce more even coverage. Mobile LiDAR are typically mounted to vehicles and used to scan roadways and cityscapes (Guan et al., 2016). However, the introduction of handheld, wearable and other mobile systems has made it possible to apply this technology to locations such as cultural heritage sites (Rodríguez-Gonzálvez et al., 2017). The experiments presented in this paper were performed on point cloud data produced from mobile LiDAR scanners.

1.3 Computation Challenges and Hardware Constraints

Processing 3D point clouds can pose significant computational demands, especially when dealing with complex and dense scenes, particularly in tasks such as object detection and classification. These tasks involve handling vast amounts of data and complex algorithms, necessitating substantial computational power. Graphics processing units (GPU) are commonly employed in such work due to their parallel processing capabilities, which accelerate deep learning algorithms effectively. However, at the outset of this research, access to a GPU was limited, and the computations primarily relied on a 2.7GHz Dual-Core central processing unit (CPU) with 8GB of DDR3 random access memory (RAM).

The absence of GPU resources significantly influenced the research's direction and methodology. Traditional deep learning models and large-scale neural networks,

which typically require extensive memory and processing power, were not viable options without GPUs. Consequently, the research focused on exploring innovative and resource-efficient approaches to achieve object detection and classification in 3D point cloud scenes. As a result of the GPU limitations, interactions with deep learning models were confined to CPU inference only.

The absence of GPU resources proved to be a pivotal aspect of this research, prompting the development of methods applicable in resource-constrained environments where GPU availability may be limited. Emphasising lightweight algorithms, transfer learning, and explainable AI techniques, the research produced solutions that are less reliant on powerful GPUs, thus enhancing their suitability for implementation in such settings.

Although GPU-accelerated computers became available two years into the thesis, the research direction had already been established, and the benefits of earlier limitations were apparent. The work carried out without GPU access underscored the adaptability and practicality of the developed methods, enabling their application in various contexts facing similar constraints. This adaptability bolsters the applicability of the research in areas where access to high-end hardware is limited, ultimately contributing to the accessibility and broader utility of the developed techniques in cultural heritage preservation, archaeological studies, and related fields.

1.4 Deep Learning on 3D Point Cloud Data

AI is an academic discipline devoted to the research of machine systems which can perceive their environment and act rationally towards a goal. The rationality is defined as drawing conclusions based on a system of logic following reason (Dick, 2019). Machine learning is strongly related to AI and is most often considered a subfield (Langley, 2011). It is a broad term for the use and development of methods that leverage information from sample data (i.e., learn) to make predictions or decisions. The exact scope of machine learning is still being explored (Alpaydin, 2010). However, the ability to make optimised predictions and decisions offers significant advantages to enabling automation and problem solving.

Deep learning is a part of the family of terms used to describe machine learning and describes methods that are composed of multiple layers. Most modern deep learning models are classified as artificial neural networks (ANN) (LeCun, Bengio, and Hinton, 2015). ANN are computing models based on a collection of connected units called artificial neurons. They are loosely modelled after the biological neural network of animal brains (LeCun, Bengio, and Hinton, 2015). Each connection receives a real number as a signal, processes it, and then messages the neurons connected to it. The output of each neuron is computed by some non-linear function of the sum of its inputs (Atkinson and Tatnall, 1997). ANNs can usually be represented as graphs

and are typically organised into multiple layers and consist of the input layer, zero or more hidden layers, and the output layer. Connections between the neurons are called edges and typically have a weight associated with them. The weights control the strength of the signal at a connection and are adjusted by the learning process.

Convolutional neural networks (CNNs) are a type of ANN that use convolutional layers to systematically slide a learnable convolution matrix, or kernel, across an input (LeCun, Bengio, and Hinton, 2015). This aggregates information from adjacent entities into features that are then passed to the next layer. These layers progressively extract multi-level features from a raw input. Low-level features can, in the case of a pixel-wise input from a 2D image, include edges and lines; whereas high-level features (also called latent features) describe more abstract characteristics (Alzubaidi et al., 2021; Simonyan, Vedaldi, and Zisserman, 2013a). CNNs have been applied to a considerable number of applications within image processing (Egmont-Petersen, Ridder, and Handels, 2002) and natural language processing (Zulqarnain et al., 2020), and more recently in the fields of cultural heritage and archaeology (Olivier and W. V.-v. d. Vaart, 2021).

As an abstract data structure, point clouds are an unordered set of continuous points in space and, as such, are not confined to a countable set of locations and have no specific order (Qi, Su, Mo, et al., 2016). The effect of this is that permutations to the order of points will yield the same point cloud representation. This property is often called 'permutation-invariance' (Qi, Su, Mo, et al., 2016). However, the point locations are not isolated. There is an important interaction between neighbouring points, that is, they form a meaningful subset Qi, Yi, et al., 2017a. The point cloud's structure is, instead, defined by the distance metric from the original (real world) data space; the neighbouring points are neighbouring locations. This contrasts with 2D images, where moving the order of a pixel would alter the representation, and each pixel has explicitly defined neighbouring pixels as the data are restricted to a grid. It is a non-trivial task to apply machine learning and AI to point cloud data, as it is this structured format that ML and AI approaches typically rely on to input data into models (Boulch, 2019). As such, previous archaeological and cultural heritage studies do not use 3D point clouds directly and instead derive 2D visualisations that are more immediately interpretable, such as digital elevation maps (W. B. V.-v. d. Vaart and Lambers, 2019). These are then used as the input to traditional ML and AI models such as CNN (Guyot, Hubert-Moy, and Lorho, 2018). However, by not using point clouds directly, these approaches fail to take advantage of the intrinsic spatial information present in true 3-dimensional data.

Recently, deep learning (DL) methods were introduced that are adapted to the properties of 3D point cloud data (Bronstein, Bruna, et al., 2017). These deep learning methods can be divided broadly into three types: multi-view (Su, Maji, et al., 2015), volumetric (Qi, Su, Niessner, et al., 2016) and point-based (Qi, Su, Mo, et al., 2016)

approaches. Methods that fall under the multi-view category project the point cloud into multiple 2D representations (i.e., images) of the object or scene which are then input into image-based deep learning methods such as traditional CNN. Therefore, multi-view methods, by definition, do not act directly on the point clouds and so cannot be used to retain the 3D spatial information. Volumetric methods use voxel representations, such as occupancy grids (Maturana and Scherer, 2015), to impose a structure. However, this accrues significant overhead to the computational cost of processing and can include bias due to the grid-axis alignment. This thesis makes use of point-based methods, which are capable of learning directly on the point cloud structure without intermediate representations. As classifiers, these models use error-correction to learn an embedding for each point and aggregate this information into a global shape descriptor. Prominent network architectures for point-based deep learning methods include graph convolution networks (Landrieu and Simonovsky, 2018), pointwise multi-layer perceptron (MLP) neural networks (Qi, Su, Mo, et al., 2016) and kernel point CNNs(Thomas et al., 2019).

Artificial Intelligence has strong roots and significant contributions from academic research and institutions, however, it is not limited to the confines of academia. AI is a diverse and interdisciplinary field that spans beyond traditional academic settings and has far-reaching applications across various industries and sectors. As research into AI has progressed and technology advanced, it has found practical applications in real-world scenarios, beyond the academic realm. As such, its development is driven by collaboration between academia, industry, and various communities. AI is now an integral part of numerous industries, including healthcare, finance, automotive, retail, entertainment, and more. It has been integrated into various products, services, and systems, transforming the way we interact with technology and influencing almost every aspect of modern life.

This thesis uses a transfer learning approach to overcome the problem of applying ML and AI directly to point clouds of cultural heritage scenes. Transfer learning is used to adapt a ML model trained on one data source (i.e., the domain), by transferring information related to another data source (Weiss, Khoshgoftaar, and D. Wang, 2016). Importantly, this allows for the training data and application data to be from different, but related, domains. Point-based ANN are trained to learn discriminative latent features from the input point clouds representing general (non-memorial type) objects. These networks are then frozen and transferred to another learner to perform classifications, that is, predict the categories (class labels) for a given example of input data.

1.5 Cultural Heritage Sites as an Application

Several characteristics of cultural heritage sites make them an interesting application case, as the design of a process suitable for the automatic extraction and labelling of objects is not immediately obvious. The sites' age, location and cultural setting affect the physical environment and appearance. Moreover, the objects contained within them will likely reflect these characteristics and may appear in different styles, designs or materials. For example, cultural heritages sites, such as burial grounds, and especially those in the United Kingdom (UK), can date back hundreds of years and are often planted with intentional or voluntary forms of vegetation, such as gardens, trees, grass and shrubbery. Over time, the condition of objects may degrade, and the vegetation can become more prominent. Older sites may still be in use today as either active sites or tourist attractions and might contain examples of both traditional objects and more modern ones. Additionally, sites located in different geographical regions might have unique traits or variation in the typical objects found locally. Inversely, sites from the same region might exhibit differences based on cultural aspects such as religion and economics. In the context of developing an automated solution for the extraction and classification of memorial objects from these sites, the above characteristics pose significant challenges.

This thesis develops a proof-of-concept approach applicable to the National Burial Ground Survey project being conducted by *Atlantic Geomatics (UK) Limited* for the Church of England. The survey is a nationwide digital churchyard mapping project that will see all of England's some 19,000 Anglican burial grounds digitally scanned and mapped over a seven-year period. The scans are performed using a wearable mobile LiDAR platform to generate fine-resolution point clouds. The current approach to processing the point cloud data relies on a user to manually identify and segment the memorial objects from each scene. This approach has a clear limitation: the point cloud dataset is massive, containing many millions of points, and processing these data manually is a very time-consuming task. Furthermore, to do so requires expert knowledge of the scenes, objects of interest and how to identify them, as well as of the software and tools necessary to do so. Moreover, it can be difficult for even an expert human operator to identify objects heavily occluded by another object or dense vegetation. Similarly, it can be difficult to identify objects concealed by earthen ground surfaces. For example, grave kerbs, which form a rectangular edge around a grave plot, can be flush with the ground surface and makes them difficult to perceive. Software tools can be used to support such manual operations. However, their deployment to real-world datasets has limitations. For example, Atlantic Geomatics reported difficulty applying tools to burial grounds on sloped surfaces or hills. Therefore, automation of the mapping process in such a way as to overcome the current limitations of manual processing is critical to achieving this

goal.

In developing this thesis, careful consideration was taken to define the requirements, scope, and limitations of the project, as this was crutial to defining the boundaries and expectations of the research. While the ultimate goal is to automate the extraction and classification of memorial objects from 3D point cloud scenes, it is essential to carefully delineate the project's scope to ensure feasibility and practicality. The primary focus of the project is on detecting and classifying memorial objects, such as headstones and grave markers, using specialised ML and AI techniques. This choice is rooted in the significance of cultural heritage preservation and site maintenance. By automating the extraction and classification of memorial objects, archaeologists, historians, and site caretakers can benefit from efficient and accurate data analysis, facilitating the identification and documentation of these objects for further research and preservation efforts.

However, while there was potential to extend the project's scope to include extraction of further details, such as text extraction from gravestones, doing so would introduce non-trivial challenges and complexities. Text extraction from 3D point clouds involves additional specialised techniques and could lead to a significant increase in computational complexity. Additionally, the accuracy of text recognition in irregular and varying point cloud data might not be as reliable as conventional 2D text recognition methods. This challenge is further compounded by the variable states of repair in which many memorial objects are found. Moreover, the preservation of textual content requires careful consideration of privacy concerns and ethical implications, as gravestones often contain personal information about the deceased.

Considering these factors, it is reasonable for the project to conclude at the point of memorial object extraction without further detail extraction. Focusing on memorial object detection and classification aligns with the primary objectives of the research, offering practical and meaningful outcomes for the preservation and understanding of cultural heritage sites. The potential complexities and ethical considerations associated with text extraction can be addressed in future research, as a separate and specialised endeavour, allowing for more focused investigations and advancements in the respective field. By adhering to a well-defined scope, the project maintains a strong and coherent focus on its primary objectives, ensuring the development of a robust and effective solution for memorial object extraction from 3D point cloud scenes.

1.6 Explainable AI for Cultural Heritage

The importance of explainability in object detection of cultural heritage objects from 3D point clouds arises from the intricate nature and critical significance of preserving historical sites and cultural artefacts. Precise identification of objects

like archaeological artefacts and memorial markers holds paramount importance for archaeologists, historians, and site caretakers. However, the utilisation of deep learning models for object detection often presents a "black box" problem, lacking transparency in decision-making, thereby raising concerns about the model's reliability and trustworthiness. This lack of interpretability hampers its practical implementation in real-world scenarios and obstructs the opportunity to learn valuable insights from the model's operations.

The scope of explainability in this thesis aims to develop a transparent mechanism that sheds light on the rationale behind the model's object detection predictions. By incorporating explainability techniques, users can gain a comprehensive understanding of how the classification model identifies and classifies cultural heritage objects from point cloud data. This understanding proves critical for domain experts to validate the accuracy of label predictions, assess potential biases, and promote confidence in the model's outcomes.

However, the application of explainability to 3D point cloud data presents several challenges and limitations. Unlike traditional 2D image data, point clouds lack a well-defined structure, making it more intricate to interpret the model's decisions based on individual pixels or voxels. Additionally, the complexity of objects and their contextual variations within scenes further complicates the explainability process. Moreover, achieving explainability without compromising detection accuracy and computational efficiency requires a delicate balance that necessitates careful consideration.

To address the need for explainability in object detection of cultural heritage objects from 3D point clouds, a prototype-based explainable classifier has been selected as the preferred approach. Prototype-based classifiers represent objects in a high-dimensional space using prototypes or centroids that encapsulate essential characteristics of each class. This approach is well-suited to offer interpretable insights, allowing users to comprehend how the model associates new data points with the learned prototypes.

By utilising a prototype-based explainable classifier, users can visualise the prototypes, serving as representative points for each object class, thereby aiding in understanding decision boundaries and object characteristics. The transparency of the classifier allows for manual inspection and manipulation of prototypes, facilitating fine-tuning of the model's decision-making process based on user expertise. Furthermore, this approach inherently handles variable states of repair and preservation of cultural heritage objects by capturing essential object features instead of relying solely on pixel-level details.

Consequently, the prototype-based explainable classifier approach stands as a fitting solution to address the necessity for transparency and interpretability in the object detection of cultural heritage objects from 3D point clouds. It

empowers domain experts to validate and refine the model's predictions, ensuring the preservation and understanding of invaluable cultural artefacts and historical sites.

The graphical user interface (GUI) presented in Appendix B aims to bridge the gap between machine learning and non-expert users by providing a proof-of-concept for an intuitive and user-friendly application for the Geometric-Partitioning and eXplainable Point Cloud Classifier framework, that will be in introduced in Chapter 4 as GeoPart-XPCC. As the adoption of machine learning and AI methods continues to expand across various domains, user-oriented applications become crucial to make these powerful tools accessible to a wider audience. This GUI serves as a proof-of-concept for the GeoPart-XPCC framework, enabling users to interact with the methodology without the need for extensive knowledge in machine learning or programming. In this section, we introduce the key features and functionalities of the GUI, which can potentially open doors for broader utilisation of the GeoPart-XPCC framework by researchers, practitioners, and cultural heritage experts alike.

1.7 Aims and Significance

In this thesis, several detailed methodologies involving specialised ML and AI techniques are developed to automate the extraction and labelling of objects from 3D point cloud scenes (i.e., object extraction) with a particular focus on cultural heritage sites. The body of the thesis is made up of a series of related articles, each relating to the design, development and implementation of solutions to this application, and experimentation to explore them. In the context of point cloud data, object extraction consists of finding the points belonging to an object and partitioning them from the scene. It shares similarities with object detection and semantic segmentation. Object detection involves finding the location and label of objects in a scene using a bounding box. However, object detection does not isolate the object from the scene, and the objects typically have a spatial context (Thomas, 2019). Similarly, semantic segmentation is a further development that gives a label to each point of a scene to partition it semantically. However, semantic segmentation does not distinguish between individual objects (i.e., multiple objects in a single category are treated as one entity) (Guo et al., 2018). Instead, in the case of object extraction, each individual object is identified as an instance belonging to a category and the result of performing object extraction is a collection of sub-set point clouds, where each (ideally) contains only one object. This task is very similar to 3D instance segmentation, as is sometimes used in the literature. However, the distinction made in this thesis is that object extraction targets one or more class categories of interest found in the scene, rather than all.

Aims were identified that make applying machine learning and AI methods to the domain of cultural heritage management and related fields, such as archaeological and

historic sites, unique. These are outlined below:

1. Decrease the time needed for a user to complete the task of extracting and labelling of 3D objects for the purpose of mapping cultural heritage sites. Efficiently reducing the time required for users to complete object extraction and labelling tasks is essential to enhance productivity and streamline archaeological and heritage preservation workflows. By developing automated methods that accurately identify and classify objects, manual intervention can be minimised, resulting in faster data processing and mapping, which is critical for time-sensitive applications and large-scale heritage site documentation projects.

2. Create a method that can correctly extract and classify an object better than the state-of-the-art methods. Advancing the accuracy of object extraction and classification methods beyond the state-of-the-art is crucial to improve the reliability of heritage site mapping and analysis. By achieving higher classification accuracy, the developed method can provide more trustworthy results, instilling greater confidence in the detected objects' identities and attributes. This improvement has significant implications for archaeological studies, cultural heritage management, and research, as it enhances data quality and facilitates more informed decision-making.

3. Use point clouds directly to take full advantage of the spatial information encoded in the 3D data. Directly leveraging the inherent spatial information within 3D point clouds is vital to capture and preserve the fine-grained details of cultural heritage objects and their surroundings accurately. By harnessing point cloud data without intermediate conversions or data loss, the developed method can maintain the rich geometric and topological properties of the objects, leading to more accurate and comprehensive representations of the heritage sites.

4. Address the need for ground references and annotated point cloud data of the objects of interest, specifically, cultural heritage scenes and objects, to be used as training data. This is a major challenge towards the application of ML methods, as the data does not already exist. The lack of annotated and ground-referenced point cloud data is a significant challenge in applying machine learning methods to cultural heritage object detection. Addressing this limitation by developing innovative techniques to create ground truth annotations and training data is crucial for the successful implementation of machine learning models. The availability of curated and annotated data enables the development of accurate and robust classifiers and contributes to the broader application of AI methods in cultural heritage research.

5. Objects or classes of interest should be able to be added easily to the classification model without imposing significant costs in time or effectiveness.

It is necessary to process scenes from different scanning technologies and with varying point densities at both the object and scene level. It is also entirely possible for one scene to contain unique objects of interest. The above necessitates the need for adaptable solutions. Designing an adaptable classification model that allows easy addition of new object classes is essential to accommodate the dynamic nature of cultural heritage sites. As new objects or classes of interest are discovered, integrating them into the model without incurring substantial retraining costs or compromising performance ensures the scalability and longevity of the developed solution, supporting continuous updates and refinements as new archaeological findings emerge.

6. The predictions and decision-making process should be clear and human understandable. Inherently explainable models enable this form interaction without an operator having extensive knowledge of the underlying algorithms. Developing interpretable and explainable models is crucial to enable meaningful interactions between users and the system. By providing clear insights into the model's decision-making process, operators can perform diagnostics and updates to enhance the model's performance and adapt it to specific applications. This transparency empowers non-expert users, such as archaeologists and site caretakers, to validate the model's outputs and make informed decisions, leading to increased trust and acceptance of AI methods in heritage site mapping and preservation efforts.

1.8 Dataset Description

The dataset utilised in this research comprises 3D point clouds representing five cultural heritage sites of burial grounds found across England. These sites were thoughtfully selected to encompass diverse archaeological characteristics and variations in environmental conditions. The dataset was generously provided by the burial ground management system team at Atlantic Geomatics (UK) Limited, and it includes operator-labelled mappings for accurate ground truth annotations.

The point clouds in the dataset were acquired using the Leica Pegasus capture platform, with the exception of Scene 11, which was collected using UAV-mounted sensors and photogrammetry techniques. The variation in data acquisition methods introduces diversity in the dataset, capturing distinct characteristics and attributes of each scene.

The point cloud scenes in this thesis encompass a diverse array of objects commonly found in cultural heritage sites and burial grounds. These scenes feature various elements, including benches, trees, and other vegetation. The presence of buildings and roads, along with pathways, adds to the contextual setting of the

scenes. Fences delineate the boundaries of the sites, offering unique geometric features for analysis. Additionally, the scenes incorporate significant cultural heritage objects, such as headstones, tomb slabs, crosses, and sarcophagi. It is important to note that the majority of these scenes originate from different locations scattered across England, capturing a wide range of architectural styles and cultural artefacts, enriching the dataset with substantial diversity and representing various cultural heritage contexts.

Further details on the point cloud datasets can be found in Chapters 2, 3, and 4.

1.8.1 Data Resolution

A critical consideration in this research is the trade-off between data resolution and computational complexity. The point density of the point cloud scenes varies, contingent upon the distance of the scanner from the surface of the environment being scanned. Higher point densities yield greater precision and richer detail, presenting potential benefits for object detection and classification tasks. Nonetheless, it simultaneously increases the computational demands of the algorithms, as the sheer volume of data to be processed escalates.

In addressing this trade-off, two distinct point reduction techniques were utilised: random sub-sampling and structured sub-sampling. Random sub-sampling involves uniformly reducing the number of points in the dataset, thereby decreasing data resolution. On the other hand, structured sub-sampling retains specific patterns of points while reducing the overall count, aiming to preserve critical information while effectively reducing computational complexity.

The choice between these point reduction techniques necessitated careful consideration of the research objectives and computational constraints. The selection of an appropriate point reduction strategy significantly influences the performance and efficiency of the object detection and classification algorithms. Each chapter of research within this thesis took this balance into consideration and was optimised based on the nature of the algorithm and the computational resources available during its undertaking.

1.8.2 Exclusion of RGB Image Data

Apart from capturing point cloud data, the Leica Pegasus capture platform also acquires RGB image data of the scene, which it uses internally to colour the corresponding point cloud data. However, the research in this thesis predominantly excludes the use of this colour image data. The decision to exclude RGB image data collected alongside the Point Cloud data was driven, in part, by the research aim, which primarily focuses on analysing and processing Point Cloud data in isolation.

While RGB images can offer valuable contextual information to complement the 3D point data, their inclusion could introduce challenges in terms of practicality and consistency.

One of the primary reasons for this exclusion was to ensure that the research remains relevant and applicable to a wide range of Point Cloud datasets. While some Point Cloud datasets may have accompanying RGB images, it is not a universal standard, and not all datasets are likely to include RGB data. By focusing solely on Point Cloud data, the research aims to address the common case scenario and establish methodologies that are widely applicable and not dependent on additional data modalities.

Moreover, the quality and usability of the RGB images can be variable, making their integration challenging. Factors such as weather conditions, scanning speed, and equipment settings may result in under or over-exposed images or images that are out of focus. These limitations can undermine the reliability and consistency of the data and may not provide substantial value in cases where the RGB information is not of sufficient quality.

Additionally, it is non-trivial to combine point cloud and RGB image classifications due to the inherent differences in data formats and processing techniques. Integrating these modalities poses complex challenges in terms of data alignment, feature extraction, and model integration. Successfully combining point cloud and RGB image classifications requires careful consideration of their complementary nature and the development of specialised methodologies to effectively leverage the strengths of both data types.

By excluding RGB images, this research adopts a more straightforward and consistent approach, streamlining the analysis, and enabling a robust comparison of results across different Point Cloud datasets. This decision allows the research to focus on the intrinsic complexity and challenges associated with processing and analysing 3D Point Cloud data, which aligns with the primary objectives of the study. In summary, the deliberate exclusion of RGB image data ensures the research's relevance, practicality, and applicability to a broad range of Point Cloud datasets, avoiding reliance on additional data modalities that may not always be available or may introduce complexities in processing and analysis.

1.8.3 Data Availability and Utilisation Strategies

The availability of labelled data in this research was limited, posing a significant challenge for training and evaluating machine learning models. The process of manually annotating cultural heritage objects in 3D point cloud scenes is a time-consuming and labour-intensive task, which results in a scarcity of labelled data. However, the presence of labelled data is crucial for supervised learning approaches,

enabling the models to learn from ground truth annotations and make accurate predictions.

Given the scarcity of labelled data, the research employed various strategies to maximise its utility. Transfer learning was utilised to leverage pre-trained models from related tasks or domains with abundant data. By fine-tuning these models on the limited labelled data available, the research aimed to adapt the models to the specific object detection and classification task in cultural heritage sites. This approach facilitated the effective use of labelled data, reducing the demand for extensive annotations while still achieving respectable performance.

In contrast, a relatively substantial amount of unlabelled data was available. Unlabelled data, though not directly useful for supervised learning, plays a vital role in unsupervised and semi-supervised learning approaches. Unsupervised learning methods, such as clustering, were utilised to discover patterns and groupings in the unlabelled data, assisting in understanding the underlying structure of the point cloud scenes.

The integration of labelled and unlabelled data in this research aimed to optimise the use of available resources, ensuring that the models benefit from the limited annotated data to leverage the generation of labels for the unlabelled data to gain insights into the data distribution and enhance generalisation capabilities.

1.9 Thesis outline

Chapter 2, establishes a method for automating the extraction and labelling of memorial objects within 3D point cloud representations. This methodology involves two distinct machine learning operations, segmentation and classification. The segmentation process employs a conditional multi-scale partitioning approach to extract memorial objects based on geometric shape attributes. Subsequently, a convolutional neural network is utilised to extract latent feature descriptors, which are then passed to a MLP classifier.

In Chapter 3, an explainable point cloud classifier (XPCC) tailored for 3D point clouds is proposed. The XPCC directly processes unordered and unstructured point sets. It leverages local densities and global multivariate generative distributions, embracing an anthropomorphic machine learning approach for comprehensive, interpretable object-based classification (P. P. Angelov and Gu, 2018). XPCC exhibits continuous learning without requiring complete retraining, transferability across domains, and adaptability for specific applications. Its computational efficiency stems from recursive calculations. Additionally, we introduce the Compound Prototype Cloud (CPC), a prototype-based visual representation. XPCC's prototype objects are combined to create prototypical class representations in the feature space, enabling human-understandable insights into model explanation and object contributions to

classification.

Chapter 4 introduces the GeoPart-XPCC framework, designed for the autonomous extraction and labelling of memorial objects from cultural heritage scenes. The framework involves a two-step process. Initially, a robust machine learning algorithm is employed for segmentation, tailored to handle fine-resolution point clouds with varying point density on a large scale. This approach builds upon the segmentation method presented in Chapter 2. Subsequently, a transfer learning-supported explainable neural network is utilised for efficient and accurate classification. This methodology is an enhanced version of the XPCC algorithm discussed in Chapter 3. The framework's effectiveness is evaluated through experiments conducted on 11 scenes derived from five distinct burial ground sites situated in the United Kingdom.

Chapter 5 discusses the implications of the methods proposed in previous chapters and how this thesis fits within the fields of AI and cultural heritage management. Possible areas for future studies are also discussed.

Chapter 6 concludes the thesis with a closing summary of the overall findings, achievements and contributions of this research.

Appendix A provides an analytical review of the current state-of-the-art in relation to the explainability of artificial intelligence in the context of recent advances in ML and DL. It defines key terminology and taxonomy related to explainable artificial intelligence.

Appendix B provides an overview of a GUI for the GeoPart-XPCC framework. This GUI was developed as a proof-of-concept user interface (UI) for processing point cloud scenes representing burial grounds. It enables a user-operator to perform model training, segmentation and classification, as well as visualise and manipulate the XPCC method's prototype structure.

1.10 Thesis Contributions

This thesis seeks to address methodological shortcomings in ML and AI and how they can be applied to cultural heritage management. Current methods of object extraction from cultural heritage point cloud data rely on manual processing or limited semi-autonomous tools such as purpose-built spatial feature descriptors and recognition patterns. Manual processing massive point cloud data is time consuming, and it is not trivial to find the most effective combination of features or patterns for a specific task. It is, therefore, difficult to generalise such approaches across applications and expert knowledge of both the domain and processes is often required. By developing advanced methods in ML and AI, it becomes possible to improve upon this standard to produce robust and efficient end-to-end automated solutions. Contributions detailed within this thesis are both methodological and application focused. It is motivated by both a desire to improve upon the current state-of-the-art in ML and AI applied

to point clouds and to develop a proof-of-concept approach to the extraction and classification of memorial objects from 3D cultural heritage scenes. The contributions of the chapters are as follows.

Chapter 2: Of the few previous related studies available in the literature surrounding object extraction applied to cultural heritage sites, none focus on object extraction and classification directly from the 3D point cloud scenes. As such, the research conducted in Chapter 2 is the first research of its kind on the extraction of memorial objects directly from large scale 3D point cloud data. The contributions of this chapter are that it:

i Presents the application of extracting and classifying memorial objects from burial ground cultural heritage sites.

ii Designs a methodology to address the challenges associated with processing point clouds directly, through segmentation and classification processes.

iii Proposes a conditional multi-scale partitioning scheme to improve the detection of ground level objects.

iv Describes a transfer learning classification approach to leverage the discriminative power of CNNs while mitigating their limitations.

Chapter 3: Few studies have investigated explainable methods for the classification of objects from 3D point clouds. The XPCC method offers several advantages over previous explainable and non-explainable methods. The main contributions of this chapter are summarised as follows:

i An explainable point cloud classifier network XPCC is proposed that addresses the lack of transparent object classification algorithms for 3D point cloud data.

ii It is specific to point cloud data and offers several layers of human-interpretable explainability.

iii A new prototype-based visual representation is proposed that explores explanations within the 3D space.

iv An evaluation of the proposed classification network that demonstrates an improvement in classification accuracy compared to existing methods.

Chapter 4: Previous studies utilise point clouds derived from airborne LiDAR, whereas the proposed GeoPart-XPCC framework is suitable for the fine-resolution and variable density point clouds derived from mobile LiDAR. The main contributions of this chapter are the following:

i A framework for the autonomous extraction and labelling of memorial objects from cultural heritage scenes called GeoPart-XPCC is proposed.

ii Application of the framework at scale to five cultural heritage sites, with the sites divided into 11 separate scenes.

iii Within-sample accuracy results of the framework.

iv Out-of-sample accuracy results demonstrating the domain transferability of the framework.

Appendix A: Appendix A presents an analytical review of the current state-of-the-art in explainable artificial intelligence. The contributions of this Appendix are:

i Analysis on the trend use of explainable AI.

ii Further taxonomy of explainable AI.

iii Investigation of novelties in terms of explainable AI.

iv An analysis involving the Caltech-101 benchmarking dataset.

Appendix B: An interface allows non-experts to interact with the methodology effortlessly, without the requirement of extensive technical knowledge. The GUI included in Appendix B serves as a user-oriented application for the GeoPart-XPCC framework, addressing the need for increased usability in machine learning and AI methods. With this GUI, users can efficiently segment a point cloud scene, train the XPCC classifier, and effectively manipulate and visualise the prototypes. By providing a proof-of-concept for the GeoPart-XPCC framework, this GUI offers a valuable contribution towards making the methodology more accessible and practical for a wider audience, including researchers, practitioners, and cultural heritage experts.

1.11 Background Concepts

This section is a reference to the methods and methodology mentioned or used in this thesis that are not explicitly detailed in either the main introduction or the later chapters. A brief description of each item is given to provide context. Methods developed during the research are omitted, as later chapters describe these in detail.

1.11.1 Data Pre-Processing

Data pre-processing is a term that relates to the manipulation, transformation or scaling of data to aid an algorithm's performance.

1.11.1.1 Min-Max Normalisation

Min-max normalisation, or min-max scaling, is an operation to rescale the range of values to a specified range. A range of [0, 1] is used in this thesis. The formula for this min-max transformation is given as:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

where x is the original value and x' is the normalised value. After the transformation, all values will be between 0 and 1.

1.11.1.2 L2 Normalisation

The L2 norm, also known as the Euclidean norm, is a method for calculating the magnitude of a vector. Within this thesis, the L2 norm is used as a regularisation method to improve the convergence speed of algorithms. It is calculated as the square root of the sum of the squared vector values and represented with the notation $L2(x) = ||x||^2$.

1.11.2 Algorithms and Heuristic

This subsection provides brief descriptions of the algorithms and heuristics used within this thesis.

1.11.2.1 Greedy Algorithm

A greedy algorithm follows the heuristic of making locally optimal choices at each stage of the decision-making process. That is, the algorithm selects the best option available at the current moment and does not consider the overall optimal solution.

1.11.2.2 Incremental Algorithm

An incremental algorithm adapts to the changes in the input by recomputing only those outputs which depend on the changed data.

1.11.2.3 Feedforward Algorithm

A feedforward algorithm is one where the information moves in only one direction within a directed acyclic graph. That is, the information does not travel in a cycle or loop.

1.11.2.4 Supervised Learning

Supervised learning builds a mathematical model using annotated (labelled) data. That is, the data, known as training data, contain both the input and the desired output. Using these training data, a supervised learning algorithm learns a function that can be used to predict the output associated to new inputs.

1.11.2.5 Classification

Classification is a concept within supervised learning where the goal is to predict the categories (class labels) for a given example of input data.

1.11.2.6 Unsupervised Learning

Unsupervised learning finds patterns in a stream of input. The input is a set of data that contains only the sample itself with no label associated.

1.11.2.7 Clustering

Clustering, or cluster analysis, is the assignment of observations into subsets without the use of labelled training data. These subsets are called clusters and can be seen as the unsupervised equivalent of classification. Samples belonging to the same cluster are similar according to some predefined metric, whereas samples belonging to different clusters are dissimilar according to the same metric.

1.11.2.8 Point Cloud Segmentation

Segmentation, as it relates to 3D point clouds, is a process which can be supervised or unsupervised. It is the process of clustering point clouds into multiple homogeneous regions based on some predefined similarity metric (Guinard and Landrieu, 2017).

1.11.3 Traditional Classification Methods

In the context of this thesis "traditional classification methods" are machine learning models that contrast with representational learning models, such as deep learning models. The defining difference between traditional and representational learning classification approaches is the way that features are extracted from the data. For representational learning classification methods, both feature extraction and classification are performed by the same model (Alpaydin, 2010). Conversely, for traditional classification methods, feature extraction is performed as a separate process and then passed into the classification model (Alzubaidi et al., 2021). Several traditional classification methods are detailed briefly in the following subsections.

1.11.3.1 Support Vector Machine

A support vector machine is a supervised machine learning algorithm. In this thesis, SVMs are used for classification tasks. The objective of the SVM algorithm is to construct a hyperplane or set of hyperplanes that best separate the data points into classes (Hastie et al., 2009). The hyperplane is the decision boundary in N-dimensional space. In the case of the linear model, N is the number of features input to the model. The position of the hyperplane is determined by maximising the functional margin (i.e., the distance to the nearest training data point of any class) (Fradkin and Muchnik, 2015). The data points nearest to the hyperplane are designated the support vectors. Therefore, maximising the margin of the classifier is the inverse of minimising the number of support vectors. If the input data are not linearly separable within the original dimensional space, a transformation can be applied (Hastie et al., 2009). This transformation maps the original space into a higher-dimensional space, where it is, presumably, then possible to linearly separate the data. Furthermore, the so-called kernel trick can be applied to construct a non-linear classifier. In this case, the input to the kernel function is the data in the original (lower dimensional) space and returns the dot product of the transformed vectors in the new (higher dimensional) space. The use of the kernel trick makes the use of higher dimensional spaces practical in terms of computational costs (Koutroumbas and Theodoridis, 2008).

1.11.3.2 *k*-Nearest Neighbours

The k-nearest neighbours (k-NN) function is a non-parametric supervised learning algorithm that works off the assumption that similar data points are found near each other (Shakhnarovich, Darrell, and Indyk, 2008). Predictions are made by calculating the distances between the test data and all training data points. In this thesis, k-NN is used for classification using the standard Euclidean distance. In the simple case of k=1, the classifier assigns a data point to the class of its closest neighbour within the feature space. When $k >= 1$, the classifier assigns a data point to a class based on a majority vote and counts how many of the k-neighbours belongs to each class. The class with the largest number of representatives within those k-neighbours is then assigned (Shakhnarovich, Darrell, and Indyk, 2008).

1.11.3.3 Decision Tree Classifier

A decision tree classifier is a supervised learning approach algorithm. Intuitively, decision trees are constructed by continually dividing the features present in the training set into a binary (yes/no) decision, resulting in a tree structure. The tree begins with the root node and is split into subsets which represent the successor children. This process is repeated until the subset at a node has all the same values

as the target variable, or when a split no longer adds value to the predictions (Hastie et al., 2009).

1.11.3.4 Random Forest Classifier

A random forest classifier is an ensemble learning method that operates by combining the output of many decision trees. Each tree within the ensemble is comprised of samples drawn by random sampling with replacement from the training dataset. The output of the random forest classifier is the class selected by most trees (Hastie et al., 2009).

1.11.4 MLP & Activation Functions

A Multilayer Perceptron is a fundamental artificial neural network architecture that consists of multiple layers of interconnected nodes, known as neurons. Each neuron in a layer is connected to every neuron in the subsequent layer, forming a network structure. MLPs are extensively used for various machine learning tasks, including classification, regression, and pattern recognition (Hornik, Stinchcombe, and White, 1990).

The core components of an MLP include an input layer, one or more hidden layers, and an output layer. Neurons within each layer apply a weighted sum of their inputs, followed by a nonlinear activation function, which introduces the model's nonlinearity and expressive power. This activation function enables the network to capture complex relationships in the data.

The following gives the definitions of several common activation functions used in this thesis.

Sigmoid The sigmoid activation function is defined as the following:

$$\frac{1}{1+e^{-x}}$$

Tanh The hyperbolic tangent activation function, or tanh as it is commonly referred to, is defined as the following:

$$\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$$

ReLU The Rectified linear unit, or ReLU as it is commonly referred to, is defined as the following:

$$\max 0, x$$

During training, an optimisation algorithm adjusts the weights associated with each connection in the network to minimise a chosen loss function. This process, known as backpropagation, iteratively updates the weights using gradients computed through the chain rule of calculus Werbos, 1990.

The depth and width of hidden layers, as well as the choice of activation functions and optimisation algorithms, influence the model's capacity and learning capabilities. MLPs are capable of learning intricate patterns in data but may suffer from overfitting if not properly regularized.

1.12 Literature Review

Primary literature was reviewed throughout the chapters of the completed thesis. Chapter 2 includes a review of the literature regarding machine learning techniques, including transfer learning, to automate the detection of archaeological and cultural heritage objects. Chapter 3 includes a review of related research around deep learning for point cloud object classification, explainable deep learning on point clouds and prototype learning. Chapter 4 includes an additional review of the literature regarding conventional and deep learning methods for object extraction from cultural heritage point cloud data and approaches to mitigate the necessity of large-scale training sets. Appendix A consists of a review of the literature on explainable AI; included is an analytical comparison of the state-of-the-art and a consolidation of the taxonomy with respect to XAI. The reviewed literature in this thesis contributes insights into the body of knowledge surrounding machine learning and artificial intelligence methods for the extraction and classification of objects from 3D point clouds that represent the real world.

1.12.1 Memorial Object Extraction

Memorial object extraction from cultural heritage sites poses several challenges that researchers need to address. One significant problem involves the diversity of memorial objects themselves, including variations in shapes, sizes, materials, and levels of degradation. This variability makes it challenging to develop a one-size-fits-all approach for object extraction. Additionally, cultural heritage sites often have complex and cluttered backgrounds, leading to difficulties in distinguishing memorial objects from the surrounding environment. Fiorucci et al., 2020 explore the interaction between machine learning and cultural heritage, catering to both cultural heritage practitioners looking to leverage ML techniques and ML practitioners seeking to apply their expertise to cultural heritage applications.

Another challenge is the availability and quality of data. Cultural heritage sites are often unique and irreplaceable, making it difficult to obtain comprehensive and

diverse training datasets for model development if the objects have been damaged. (Hassan and Fritsch, 2020) explores the integration of point clouds for the preservation of cultural heritage objects. The study by (Ren et al., 2022) investigates hole-filling techniques for restoring missing areas in culturally significant objects.

Chapters 2 and 4 provide in-depth reviews of methodologies pertaining to cultural heritage object detection from point clouds, as well as other significant applications of point cloud data within the cultural heritage domain, in their respective "Related Work" sections. These chapters offer in-depth insights into the diverse methods and techniques employed in the context of cultural heritage preservation and analysis, shedding light on their effectiveness, limitations, and contributions to the field.

1.12.2 Object Extraction in 3D Point Cloud Data

The extraction of objects from 3D point cloud data is a fundamental and challenging task with broad applications in various domains, including cultural heritage preservation, urban planning, robotics, and autonomous navigation. Object extraction from point cloud data involves identifying and segmenting individual objects or structures of interest from the surrounding environment. The 3D nature of point cloud data captures rich spatial information, making it an essential data source for accurately delineating and analysing complex three-dimensional scenes. As the field of computer vision and machine learning advances, numerous algorithms and techniques have been proposed to tackle the intricacies of object extraction in point cloud data. This section of the literature review critically examines the current state-of-the-art methodologies, discussing their strengths, weaknesses, and suitability for different applications. By exploring the existing body of research, this section aims to identify gaps and challenges that lay the foundation for the subsequent discussion of novel and innovative approaches to address the complexities of object extraction in 3D point cloud data.

Object extraction from 3D point cloud data presents several challenges that researchers and practitioners need to overcome:

Complex and Diverse Scenes: 3D point cloud data can represent complex and diverse scenes with a wide range of object shapes, sizes, orientations, and densities. This variability makes it challenging to design algorithms that can accurately identify and extract different types of objects.

Noise and Artefacts: Point cloud data captured from real-world environments can contain noise, anomalies, and missing points due to sensor limitations, occlusions, and other environmental factors. These imperfections can negatively impact the accuracy of object extraction algorithms.Perhaps unsurprisingly, de-noising point clouds has been a topic of interest (Dinesh et al., 2018; Rakotosaona et al., 2019; Luo and W. Hu, 2021; W. Hu, Q. Hu, et al.,

2021). Furthermore, (W. Hu, Gao, et al., 2019; Pistilli et al., 2020) underscore the efficacy of graph-based approaches in managing noisy data. De-noising methods can be employed to pre-process point cloud objects before classification, although this may lead to a trade-off in computational performance.

In contrast, other approaches seek to build methods that are robust to the noise directly. In their work, (Zhu et al., 2020) introduce a 3D shape signature that incorporates symmetry, convex hull, and Chebyshev polynomial fitting. This approach aims to approximate angle-radius function coefficients, enabling the construction of a 3D object detection network that is robust to noise and sparsity in the data.

Likewise, TANet (Liu et al., 2019) introduces what the authors call a triple attention module that combines channel-wise, point-wise, and voxel-wise attention to enhance target information while suppressing unstable points. A Coarse-to-Fine Regression module improves localisation accuracy without excessive computation. The proposed method showcases advancements in addressing noise and object detection challenges in point clouds.

Large-Scale Data: Efficiency concerns arise when dealing with massive point cloud datasets in scenarios such as cultural heritage sites and urban environments, which may involve extensive 3D scenes. Computational challenges are posed by the need to process and analyse these datasets efficiently. To address scalability issues, various approaches have been proposed, ensuring the effectiveness of point cloud analysis on larger spatial scales (Gunji et al., 2016; Landrieu and Simonovsky, 2017; Landrieu and Simonovsky, 2018; Q. Hu et al., 2019; Zeng et al., 2022).

Occlusion and Clutter: In scenes, objects often face occlusion from other objects or get obscured by clutter, demanding precise object identification. This challenge involves effectively managing occlusion and discerning between object and background points. A noteworthy solution is ImVoteNet (Qi, Chen, et al., 2020), a multi-modal 3D detection architecture that marries 2D votes extracted from images and 3D votes from point clouds, augmenting resilience against noise and data gaps. Moreover, multi-view CNNs have demonstrated their ability to handle occlusion and clutter (Pang and Neumann, 2016).

Segmentation and Grouping: Separating individual objects from a point cloud requires effective segmentation and grouping algorithms. Objects may consist of disconnected parts or be in close proximity to other objects, making it necessary to accurately segment and group points belonging to the same object.

Point cloud segmentation stands as one of the most extensively explored areas within point cloud analysis. Researchers have dedicated significant efforts to

develop algorithms and techniques that address the challenges posed by the inherent complexity, noise, and variability present in point cloud data (B. Wu, Wan, et al., 2018; B. Wu, Zhou, et al., 2019; Graham, Engelcke, and Maaten, 2017; Boulch et al., 2018; Qi, Su, Mo, et al., 2016; Qi, Yi, et al., 2017a; Thomas et al., 2019). These segmentation methods aim to enhance the usability of point cloud data, thereby contributing to the broader field of point cloud analysis and its applications across diverse domains. However, the majority of these methods are primarily evaluated on benchmark datasets, posing challenges in their direct application to real-world data.

1.12.3 Explainable AI

Explainable AI (XAI) has emerged as a crucial and evolving field in the realm of artificial intelligence and machine learning, aiming to shed light on the opaque decision-making processes of complex models (Bishop, 2006; Goodfellow et al., 2014). As AI technologies permeate various industries and critical applications, the need for transparent and interpretable AI systems becomes increasingly vital. XAI techniques strive to provide human-understandable explanations for the predictions and outcomes generated by AI models, fostering trust, accountability, and comprehension in AI-enabled systems. As the thesis seeks to develop adaptable and interpretable solutions for object extraction within diverse cultural heritage environments, XAI provides a bridge between sophisticated AI methods and human understanding.

The early phases of AI encompassed methods such as decision trees, symbolic AI, expert systems, and ANNs, which exhibited greater interpretability and self-explainability than modern counterparts such as SVMs and more advanced ANNs (Hearst et al., 1998). In recent times, the focus on explainability has intensified due to the increasing integration of AI and ML into human-centric applications. The transition from accuracy-centric to explainability-centric approaches has been propelled by the realisation that complex "black box" models can lead to detrimental outcomes (Stilgoe, 2020; Pasquale, 2016; Rudin, 2019).

In light of these challenges, the central question is not merely about the possibility of explainable AI solutions, but rather about attaining highly accurate XAI solutions that rival DL's accuracy. The subsequent sections provide insights into these dynamics, emphasising the interplay between AI's historical roots, ML's evolution, and the need for interpretable and accountable AI solutions in an era dominated by complex models and vast datasets.

This section of the literature review delves into the diverse methodologies and advancements in the realm of Explainable AI, exploring the principles, approaches used to achieve model interpretability. This section aims to offer an overview of XAI to establish a contextual framework for the subsequent methods developed in this

thesis. By exploring various XAI techniques and their attributes, this overview sets the stage for the approaches and insights presented in the following sections.

SHapley Additive exPlanation (SHAP): (Lundberg and S.-I. Lee, 2017) offers a game-theoretic framework to explain predictions made by machine learning models. It draws inspiration from cooperative game theory, representing the relationship between features and model predictions as a coalition game. Each feature acts as a player, contributing to the game's outcome. The key concept here is the Shapley value, which measures the average contribution of a feature across all possible feature combinations.

This approach ensures both local and global interpretability by consistently attributing feature importance. It allows comparisons between explanations, as the average prediction is fairly distributed among Shapley values. This feature enables users to contrast different explanations across instances and understand their varying contributions.

However, SHAP's interpretability might be compromised if the model isn't inherently additive. In cases where predictive models exhibit non-independent pay-off splits, interpreting Shapley values becomes challenging. Furthermore, while SHAP is intended to be model-agnostic, implementing it efficiently across various model types can be intricate.

SHAP's game-theoretic foundation and consideration of feature interactions make it a promising technique for explaining machine learning predictions. Its focus on ensuring coherent and consistent interpretations aligns with the growing need for transparent AI models.

Local Interpretable Model-Agnostic Explanations (LIME): (Dieber and Kirrane, 2020) is a widely recognised technique for generating locally optimized explanations for predictions made by complex machine learning models, regardless of their underlying architecture.

LIME operates by creating a simplified surrogate model that approximates the behaviour of the black-box model in the vicinity of a specific prediction. It achieves this by perturbing the input data and observing the resulting changes in predictions. These perturbations are usually applied to small patches of contiguous superpixels in the case of image data. By training the surrogate model on perturbed instances of the original input, LIME learns how changes in these interpretable aspects of the data affect the predictions.

The rationale behind LIME is that by focusing on locally explainable behaviour, users can gain insights into which specific features or aspects of the input are driving the model's decision in a given instance. While LIME can provide

useful insights into individual predictions, its effectiveness might be influenced by the choice of perturbation parameters. Additionally, LIME's explanations are limited to the specific local context and may not provide a holistic understanding of the model's overall behaviour or feature interactions.

LIME serves as a valuable tool for generating interpretable explanations for individual predictions, making it suitable for cases where understanding the decision-making process on a per-instance basis is essential. However, it is important to consider its limitations and ensure that the perturbation parameters are chosen thoughtfully to provide meaningful insights.

Sensitivity Analysis and Layer-wise Feature Relevance: Sensitivity analysis (Barredo Arrieta et al., 2020b) and layer-wise feature relevance propagation (Tritscher et al., 2020) are two approaches in the realm of explainable AI that aim to shed light on the importance of input features and their impact on model predictions.

Sensitivity Analysis: Sensitivity analysis focuses on understanding how changes in input features influence model outputs. This method involves perturbing individual or groups of features and observing the resulting effects on predictions. By quantifying the change in output due to variations in inputs, sensitivity analysis provides insights into feature importance and their contributions to decision-making. However, this approach may lack granularity in explaining complex interactions among features.

Layer-wise Feature Relevance Propagation: Layer-wise feature relevance propagation takes advantage of the layered structure in neural networks. It aims to uncover the contribution of each input feature to the model's output by tracing the flow of relevance from the output layer back to the input. This technique assigns relevance scores to features in each layer, highlighting their impact on the final decision. However, this method may not provide a comprehensive understanding of feature interactions and dependencies within the model's architecture.

Both sensitivity analysis and layer-wise feature relevance propagation contribute to explainability by providing insights into feature importance. However, they may fall short in capturing the intricate relationships among features, particularly in complex models. These methods serve as stepping stones toward more comprehensive and holistic explanations of machine learning predictions.

Global Attribution Mappings (GAMs): (Ibrahim et al., 2019) offer a method to explain neural network predictions on a global scale, especially across different subpopulations. GAMs formulate attributions as weighted conjoined rankings, enabling the representation of how various features contribute to predictions.

By creating a rank distance matrix between features and applying a clustering algorithm, GAMs group similar local feature importances and summarize patterns in clusters. This approach allows for meaningful exploration of features among different subpopulations of samples.

Gradient-based Saliency Maps and Deep Attribute Maps: Gradient-based saliency maps (Simonyan, Vedaldi, and Zisserman, 2013b) and deep attribute maps (Ancona et al., 2017) are techniques that aim to visualise the importance of input features in influencing a neural network's predictions. Saliency maps render the absolute value of gradients with respect to input features as a heatmap, highlighting influential areas. Deep attribute maps multiply the gradient of the output by the input to create explanations in the form of heatmaps, indicating contributions to the decision-making process. However, both methods have limitations, such as sensitivity to noisy gradients and difficulty in explaining why models produce similar or different results. Grad-CAM (Selvaraju et al., 2016), Grad-CAM++ Chattopadhyay et al., 2017, and Eigen-CAM (Muhammad and Yeasin, 2020) all fall under the category of this specific type of XAI method.

Concept Activation Vectors (CAVs): (B. Kim et al., 2017) provide a means of globally explaining the internal states of a neural network by mapping abstract latent features extracted by the network to human-understandable concepts. CAVs measure the degree to which these abstract features align with chosen human-understandable concepts. For instance, given an image of one class, ascertain the corresponding area on another image that exhibits the closest similarity. CAVs offer insights into the reasoning behind the model's decisions by linking abstract features to understandable concepts, thereby allowing identification of defects in the model's decision-making process. However, CAVs depend on the unique meaningfulness of chosen concepts for each class. If certain concepts are present in multiple classes, the distinctiveness of explanations may diminish. This limitation can affect the quality and clarity of the explanations provided by CAVs.

Nevertheless, XAI presents several challenges and issues that researchers are actively addressing. One primary problem lies in achieving a balance between model accuracy and interpretability. Complex models that achieve high accuracy often sacrifice transparency, making it challenging to understand their decision-making processes. Additionally, the effectiveness and reliability of XAI methods across different domains and data types remain a concern. XAI techniques developed for one type of model or dataset might not readily generalise to other contexts. Another challenge involves the trade-off between local and global explanations. While some

XAI methods provide insights at a local level (explaining individual predictions), extending these explanations to a global level (understanding model behaviour across a dataset) can be complex. Lastly, addressing human cognitive biases and preferences in the design of XAI methods is an ongoing concern, as interpretations should align with human intuition and understanding. Currently, the field of explainable methods for point cloud object extraction is in its early stages of development. For a comprehensive overview of the progress made in this area, please refer to the "Related Work" section in Chapter 4.

Chapter 2

Automatic Extraction and Labelling of Memorial Objects from 3D Point Clouds: GeoPart-Transfer

The introductory section of the thesis provide a foundation for the exploration of automated object extraction and classification within the context of cultural heritage and archaeology. Furthermore, it establishes the objectives and scope of the research, highlighting the integration of point cloud technology, deep learning methodologies, and explainable AI techniques. Computational challenges and hardware considerations relevant to the study are discussed, emphasising the necessity of efficient and scalable solutions.

Furthermore, the significance of cultural heritage sites as an application domain is outlined, underscoring the value of innovative technologies in preserving and interpreting historical artefacts. The introduction highlights the aims of the thesis, including the development of adaptable and interpretable solutions that cater to the unique challenges posed by diverse cultural heritage environments.

Overall, the initial sections of the thesis lay a strong foundation by elucidating the context, challenges, and goals of the study. This foundation serves as a springboard for the subsequent chapters, including the focus of the paper presented in this chapter, which introduces the GeoPart-Transfer framework for automated object extraction and classification from 3D point cloud scenes of cultural heritage sites.

This chapter presents a comprehensive methodology for the automatic extraction and labelling of memorial objects from cultural heritage sites, represented as 3D point cloud scenes. The research addresses the challenges of utilising point clouds directly in real-world applications, especially in contexts where extensive

training data and computational resources are limited. The proposed methodology leverages transfer learning and CNNs to achieve accurate segmentation and semantic labelling of memorial objects. The process begins with a conditional multi-scale partitioning scheme to segment objects based on their geometric shape characteristics. Subsequently, high-level latent feature descriptors are extracted using a CNN pretrained on diverse 3D object models from a standard dataset (e.g., ModelNet). These descriptors enable the training of a multilayer perceptron for semantic labelling. The chapter highlights three key contributions: the direct operation on point clouds for cultural heritage and archaeology applications, the effective use of CNNs while mitigating their limitations, and addressing the unique challenges posed by point cloud data while capitalising on their 3D nature. Experimental results demonstrate the efficacy of the proposed methodology in accurately extracting and labelling grave marker objects from cultural heritage sites.

The significance of this chapter within the overarching context of the thesis is its role in addressing the challenges associated with automated object extraction and classification in the domain of cultural heritage and archaeology. As an integral part of the research journey, this chapter takes a decisive step toward fulfilling the thesis's objectives of developing innovative solutions that harmonise point cloud technology, deep learning techniques, and explainable AI principles for the benefit of cultural heritage preservation and management.

The GeoPart-Transfer framework serves as a critical milestone by introducing a methodological approach to autonomously extracting and labelling memorial objects from 3D point cloud scenes. Building upon the groundwork established in the preceding sections of the thesis, this chapter dives into the intricate details of the GeoPart-Transfer methodology. By incorporating a two-part process involving robust machine learning algorithms for segmentation and the application of an explainable neural network for classification, the chapter addresses the challenges posed by variable point density and intricate object attributes inherent to cultural heritage sites.

This chapter's contributions are highly aligned with the thesis's overarching goals of enabling efficient and interpretable solutions for object extraction in complex environments. It is a vital bridge that connects theoretical foundations with practical implementation, demonstrating the feasibility and efficacy of the proposed approach through experimentation on a diverse set of cultural heritage scenes. The GeoPart-Transfer framework, with its adaptable and accurate nature, manifests the thesis's commitment to developing solutions that empower practitioners and researchers in the field of cultural heritage.

In essence, GeoPart-Transfer substantiates the thesis's trajectory by presenting a concrete methodological solution that not only addresses the challenges of automated object extraction and classification but also reinforces the thesis's mission to leverage

modern technologies for the preservation and understanding of cultural heritage.

journal of computer applications in archaeology

Automatic Extraction and Labelling of Memorial Objects From 3D Point Clouds

RESEARCH ARTICLE

NICHOLAS ARNOLD
PLAMEN ANGELOV
TIM VINEY
PETER ATKINSON

*Author affiliations can be found in the back matter of this article

ubiquity press

ABSTRACT

This research addresses the problem of automatic extraction of memorial objects from cultural heritage sites represented as scenes of 3D point clouds. Point clouds provide a fine spatial resolution and accurate proxy of the real world. However, how to use them directly is not always obvious. This is especially true for applications where extensive training data or computational resources are not available. In this paper, we present a methodology for automatic segmentation and labelling of cultural heritage objects from 3D point cloud scenes. The proposed methodology is based on machine learning techniques and, in particular, makes use of the concept of transfer learning. Memorial objects are segmented from the scene based on their geometric shape characteristic through a conditional multi-scale partitioning scheme. Then, high-level latent feature descriptors are extracted by a convolutional neural network pre-trained on different 3D object models from a standard dataset (e.g., ModelNet). Based on these descriptors, a classification model (multilayer perceptron) is trained and applied to obtain semantic labels. Experiments demonstrated that the proposed methodology is effective for the extraction and labelling of grave marker objects from cultural heritage sites.

CORRESPONDING AUTHOR:
Nicholas Arnold
Lancaster University, GB
n.arnold@lancaster.ac.uk

KEYWORDS:
3D; Point Cloud; Transfer Learning; Cultural Heritage Management; Object Extraction

TO CITE THIS ARTICLE:
Arnold, N, Angelov, P, Viney, T and Atkinson, P. 2021. Automatic Extraction and Labelling of Memorial Objects From 3D Point Clouds. *Journal of Computer Applications in Archaeology*, 4(1), 79–93. DOI: *https://doi.org/10.5334/jcaa.66*

1 INTRODUCTION

Historic, cultural heritage and archaeological sites can be interpreted as hierarchical organisations of objects. The process of mapping and keeping an inventory of physical objects is fundamental to site conservation, management and analysis. Traditionally, objects are observed physically and recorded manually by an operator. Recent advances in remote sensing technologies, such as light detection and ranging (LiDAR) and digital photogrammetry, make it possible to instead create digital representations in the form of 3D point clouds. Indeed, 3D scanning technologies are becoming both more affordable and more versatile (Chase, Chase and Chase 2017; Favorskaya and Jain 2017; Royo and Ballesta-Garcia 2019). LiDAR based sensor hardware is appearing in both wearable systems and handheld devices. Additionally, photogrammetry software allows 2D images to be stitched together into a 3D point cloud scene; with this there is the potential to turn any camera into a proxy 3D sensor. These offer non-invasive, fine resolution alternatives to manual recording. As a result, 3D point cloud data are becoming a valuable resource for the fields of archaeological and cultural heritage. However, the question is then how to design an automated methodology for extracting, labelling and organising objects from these point clouds; especially one that is suitable for the real-world context.

Despite the adoption of digital technology, it remains a time-consuming task for an operator to find and label each object. Machine learning techniques that seek to automate object detection in point cloud data have been recently proposed; the most notable of which are built around convolutional neural networks (CNN) (Bello, Yu and Wang 2020). These supervised networks are composed of sequential layers wherein increasingly complex features are extracted. Specifically, the convolutional layers slide systematically a learnable convolution matrix, or kernel, across an input. This aggregates information from adjacent entities into features that are then passed to the next layer. Provided with a large training set of labelled data, CNNs are capable of generating discriminative high-level features (Bello, Yu and Wang 2020). The problem is that point clouds are an unusual data type. They are an unordered set of points in space and represent the external surface of the sampled object or scene. Each point is a vector denoting its x, y, z coordinates, and, depending on the specific acquisition technology, the points may contain additional observed information such as colour or intensity. Moreover, CNNs cannot easily take unstructured point clouds as input.

Many practical problems, and, in particular, cultural heritage and archaeological applications, often have limited access to labelled data, and in some cases the required data may be entirely non-existent, a necessary component for training CNNs. Training of CNNs is also computationally demanding. Moreover, the addition of new validation data requires a complete retrain of the network. Therefore, it is not immediately clear how to take advantage of point clouds in real world applications such as these. To this end, we present a methodology suitable for the automatic extraction and identification of objects from cultural heritage sites. We highlight how point cloud data can be used directly to map and extract objects from archaeological and cultural heritage contexts without the need to first rasterise or convert to another representation (e.g., digital elevation, surface or terrain models). We apply and validate the proposed methodology for the task of locating, extracting and labelling grave markers from cultural heritage sites.

Grave marker detection is a relatively unexplored area. Notably, to the best of our knowledge, this is the first research of its kind on the extraction of grave marker objects directly from 3D point cloud data. Grave markers can be made from many different material components and take on a multitude of different sizes and shapes depending on their location, environment, age, condition and cultural background. Therefore, a highly generalisable methodology is necessary for their detection. While not related directly to the proposed methodology, LiDAR data have been used previously to aid in cemetery surveys (Weitman 2012). Additionally, point clouds have been used to represent memorial object models; for example Jaklič et al. 2015 reconstructed sarcophagi from a point cloud data representation of a sunken Roman shipwreck. Zacharek et al. 2017 presented a low-cost approach to the collection of 3D grave marker models. To look below ground, Cannell et al. 2018 used ground penetrating radar and geochemical analysis to explore an unmarked graveyard at a medieval church site in Norway.

A recent focus in the literature has been on applying machine learning techniques to automate the detection of archaeological and cultural heritage objects. Initial methods were concerned mainly with 2D representations (Chase, Chase and Chase 2017). However, the growing presence of LiDAR as a digital surveying tool, as well as its integration into platforms such as geographic information system (GIS) software, has made point cloud analysis an important subject of research (Chase, Chase and Chase 2017). A natural step has been to ask how point cloud data can be paired with machine learning to benefit the archaeological and cultural heritage fields. Point cloud derived representations, such as surface and terrain models, have been used in conjunction with machine learning algorithms to automate the detection of barrows (Kramer 2015; Sevara et al. 2016) and Neolithic burial mounds (Guyot, Hubert-Moy and Lorho 2018), as well as in the detection of sub-surface archaeological structures (Fryskowska et al. 2017). While more traditional machine learning techniques are often employed, neural networks have been considered as well. Kazimi et al. 2018 demonstrated a CNN using LiDAR derived digital terrain maps to examine historic mining

regions in Germany. However, by not using point clouds directly, these approaches fail to take advantage of the innate dimensional information inherent to 3D.

A limitation of traditional supervised machine learning processes is that they are domain-specific; that is, they make predictions through learned properties determined by the data with which they are trained. In contrast, transfer learning allows a machine learning model trained in one domain to be reapplied to another domain of similar data. Transfer learning is one possible solution for areas where limited training data are available. Within the context of machine learning applied to cultural heritage and archaeology, the transfer learning concept has been applied to images from remote sensing surveys (Trier et al. 2016; Trier, Cowley and Waldeland 2019; Zingman et al. 2016). More recently, Verschoof-van der Vaart and Lambers 2019 explored transfer learning with region-based CNNs to detect barrows, Celtic fields and charcoal kilns from LiDAR-derived 2D images.

This paper presents a novel methodology to address the automatic extraction and identification of object instances within cultural heritage sites represented as 3D point clouds. The contributions of this methodology are: (i) how to operate directly on point clouds in a way suitable for real world application to cultural heritage and archaeology contexts, (ii) how to use the discriminative power of CNNs while mitigating their limitations and (iii) how to address the inherent challenges associated with point cloud data while benefiting from their 3D nature. Additionally, we propose a conditional multi-scale partitioning scheme within the methodology to ensure ground level objects are detected. In contrast to previous methodologies applied in a cultural heritage and archaeology context, the methodology presented in this paper involves methods applied directly to the 3D point cloud data rather than first transforming them to another structure.

The remainder of this paper is organised as follows. The methods and structure of the proposed methodology are detailed in Section 2; Section 2.1 details the segmentation process and extraction of geometric features and Section 2.2 describes the approach used for classification. Section 3 details the experimental results and Section 4 discusses the findings along with suggestions for future research. Finally, Section 5 concludes the paper.

2 METHODOLOGY

In the proposed methodology we consider the input point cloud P as a set of 3D points {p_i| i = 1, ..., n} such that $p_i \in P$ where n is the total number of points in P. The points P_i are a vector of coordinates (x, y, z). The input P represents scenes of cultural heritage sites and is assumed to have been registered and pre-processed to remove outlier and duplicate points. The goal of the methodology is to partition the scene into segments $S = \{S_1, ..., S_h\}$ and provide a label L for each from a set of semantic classes C. To do so, it is comprised of two steps: segmentation and classification.

1. Segmentation is an unsupervised process that seeks to partition the point cloud scene into regions based on the continuity and homogeneity of the properties. We define the regions as local neighbourhoods and compute features that describe their similarities. We further, embed this information into an attributed graph structure and approximate the segments with smooth pre-defined shapes by a generalised minimal partition model, a type of loss function (Landrieu and Simonovsky 2017). This serves two purposes. First, the resulting segments effectively represent the objects contained within the scene, either in parts or as a whole. Second, by considering segments rather than individual points, the classification task is made easier as there is guaranteed to be fewer segments than points. See ***Figure 1*** for an illustration of the segmentation method.
2. For classification, we explore the idea of transfer learning and pre-train the ConvPoint network (Boulch 2019) on generic object models from the ModelNet40 (Wu et al. 2015) dataset. Transfer learning allows us to take advantage of the discriminative power of the CNN and couple it with the flexibility from more classic models for per-class training. We apply each partitioned segment to the pre-trained network to generate a set of high-level abstract features. These features represent a global descriptor for the segments and are input into a multilayer perception (MLP) network classifier to predict the class labels. See ***Figure 4*** for illustration of the classification method.

2.1 POINT CLOUD SEGMENTATION

In general, point cloud data are unstructured. That is, there is no defined neighbourhood to connect each point in space. This is in contrast to 2D images, where each pixel sits on a grid and has explicit neighbouring pixels. To extract meaningful features then, some form of structure must be imposed and designed specific to point cloud data. The common approach is to avoid processing the 3D data directly, instead rasterising the point cloud into multiple 2D representations (Su et al. 2015). An alternative is for the points to be placed within volumetric containers such as voxels (Qi et al. 2016b).

Methods such as Spin Images (Johnson and Hebert 1999), kernel signatures (Aubry, Schlickewei and Cremers 2011; Bronstein and Kokkinos 2010), and inner-distance descriptors (Ling and Jacobs 2007) use a local estimate of the underlying surface around the point. Recent kernel methods build on this.

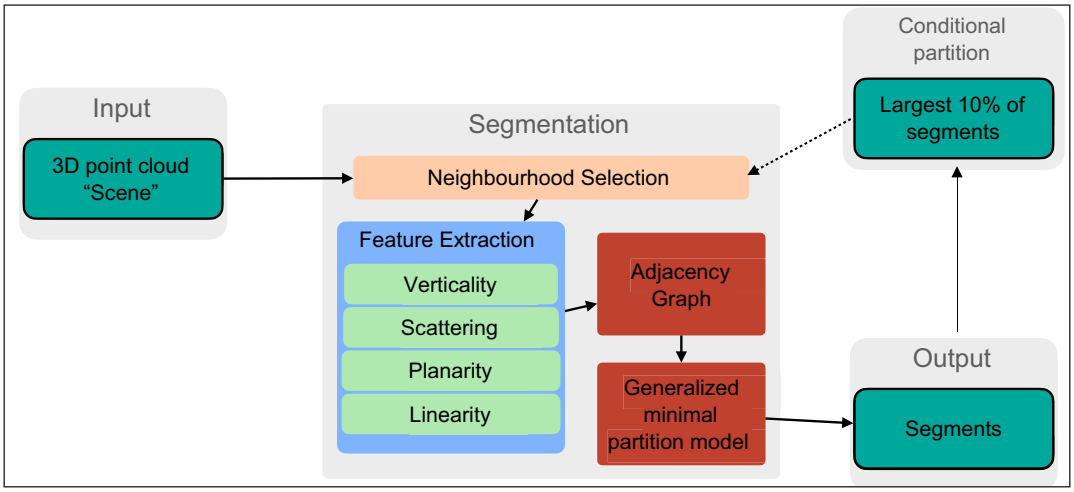


Figure 1 Concept of the proposed segmentation methodology. Solid arrows represent the flow of processes; dashed arrows represent conditional processes. A complex 3D point cloud scene is taken as input and divided into multiple segments based on the set of features extracted in relation to the points' local neighbourhood. These segments are subsets of the original scene, themselves being point clouds, and are assumed to represent the objects contained in the scene.

Sparse kernels (Graham 2015, 2015), deformable kernels (Dai et al. 2017; Su et al. 2018), and continuous kernels (Boulch 2019) have been used to achieve leading accuracy scores for object recognition and semantic segmentation against benchmarks.

Recently, many researchers have begun to explore structure applied more directly to 3D point clouds, such as tree-based (Klokov and Lempitsky 2017; Riegler, Ulusoy and Geiger 2017), graph-based (Simonovsky and Komodakis 2017) and set-based approaches (Qi et al. 2016a, 2017). A more traditional, but equally direct solution, is to perform a point-wise search and define a structure based on the points' local neighbourhoods.

The term 'segmentation' in the context of point cloud data means the partitioning of spatial regions within the scene, based on some criteria. We can distinguish between two classes of point cloud segmentation problems commonly found in the literature. The first class of problems is to segment the scene based on some geometric similarity or characteristic and can be seen as the inference of object detection or localisation. The second class of problem is semantic segmentation, a fine-grained instance of classification. This segmentation performs point-wise classification, where individual points are provided with a label. Thereby, the scene is partitioned based on semantic similarity.

A simple form of the first class of segmentation problem is to partition the foreground and background of a scene (Dohan, Matejek and Funkhouser 2015; Golovinskiy, Kim and Funkhouser 2009). Because this type of segmentation represents regions of similarity, it is used regularly as precursor to object classification (Golovinskiy, Kim and Funkhouser 2009; Shapovalov, Velizhev and Barinova 2010). Spina et al. 2011 demonstrated this type of point cloud segmentation in a cultural heritage context. Similar to Hackel, Wegner and Schindler 2016 and Guinard and Landrieu 2017 our method concerns the use of local point neighbourhoods by which we extract features to represent local regions of the scene. This method was chosen in contrast to semantic segmentation, which would require a network to be trained for specific terrain types as well as objects.

We consider a point-wise search to define local neighbourhoods. One such strategy is to search using a fixed-radius r, whereby a spherical (Lee and Schenk 2002) or cylindrical (Filin and Pfeifer 2005) representation is used to define the neighbourhood. Another is to consider the k-nearest neighbours around each point, based on some form of distance metric. This typically involves 2D (Niemeyer, Rottensteiner and Soergel 2014) or 3D (Jonathan et al. 2001) distances. As noted by Weinmann et al. 2015, for this solution to remain practical across varying scene types, search-based solutions require some form of optimization. This is either in terms of r or k, respectively. We define the points' local neighbourhoods through a k-nearest neighbour search in Euclidean space and optimise k based on eigenentropy, as advocated by Guinard and Landrieu 2017. This approach is suited to different point densities and gives more precise control over neighbourhood size (Weinmann et al. 2015).

2.1.1 Feature extraction

In this section, we present the features and algorithms used for the segmentation process.

The first stage of the proposed methodology is the segmentation process. Here, features that characterise the local dimensionality of the scene are extracted. For each point p_i, the k-nearest neighbouring points in the point cloud P are selected and the covariance matrix of their positions is calculated. From this we obtain the set of eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and corresponding eigenvectors u_1, u_2, u_3. To determine the optimal size for k, a specific

energy function, the same as in (Weinmann et al. 2015), is used to minimise the eigenentropy E of the vector $(\lambda_1/\Lambda, \lambda_2/\Lambda, \lambda_3/\Lambda)$:

$$E = -\sum_{i=1}^{3} \lambda_i / \Lambda \ln(\lambda_i / \Lambda),$$

with $\Lambda = \sum_{i=1}^{3} \lambda_i$. This results in neighbourhoods which have maximum homogeneity or minimum disorder of points within the neighbourhood. The size of k is varied between $k_{min} = 10$ and $k_{max} = 100$ in increments of 1 (i.e., $\Delta k = 1$).

Using the eigenvalues, we construct a set of features $f_i \in R^4$, which characterise the neighbourhood's local dimensionality and geometry. We use linearity, planarity, scattering (Demantké et al. 2012) and verticality (Guinard and Landrieu 2017):

$$\begin{aligned} Linearity &= \lambda_1 - \lambda_2 / \lambda_1, \\ Planarity &= \lambda_2 - \lambda_3 / \lambda_1, \\ Sphericity &= \lambda_3 / \lambda_1, \\ Verticality &= \sum_{j=1}^{3} \lambda_j \left| \left\langle [0,0,1], u_j \right\rangle \right|. \end{aligned}$$

The first three features are often referred to as dimensionality. Linearity describes how well the neighbourhood represents a 1-dimensional straight line, while planarity describes how well it fits to a 2-dimensional plane. Similarly, sphericity (also referred to as scattering in the literature) measures how well the neighbourhood resembles a sphere. Verticality indicates the geometric orientation of the neighbourhood; for example, $Verticality_{min} = 0$ represents a horizontal orientation whereas $Verticality_{max} = 1$ represents a vertical orientation (Guinard and Landrieu 2017). Examples of these features can be seen in ***Figure 2***.

2.1.2 Adjacency graph structure

A graph structure can be used to capture how different entities are related to one another. The graph nodes (or vertices as they are sometimes called) represent a singular entity, while edges connecting nodes represent the relationship between entities. The edges may be either directed, such that they can be traversed only in a single direction, or undirected, such that they can be traversed in either direction. Graphs are commonly used in machine learning to represent probabilistic models. For example, Bayesian networks, Markov random fields (MRF), and conditional random fields (CRF), are all graphical models. Additionally, graphical models may also be used as the basis of a graph CNN, a generalisation of convolution operations to arbitrarily structured graphs (Landrieu and Simonovsky 2017).

Applied to point clouds, graphical models can be used as both a structure and for data analysis (Bronstein et al. 2017). Niemeyer et al. 2011 proposed graphical models to encode the spatial relationship between points into a graph structure called an adjacency graph. Furthermore, they showed how point cloud density and number of adjacent points affect this construction. Regarding this, they concluded that a larger neighbourhood has

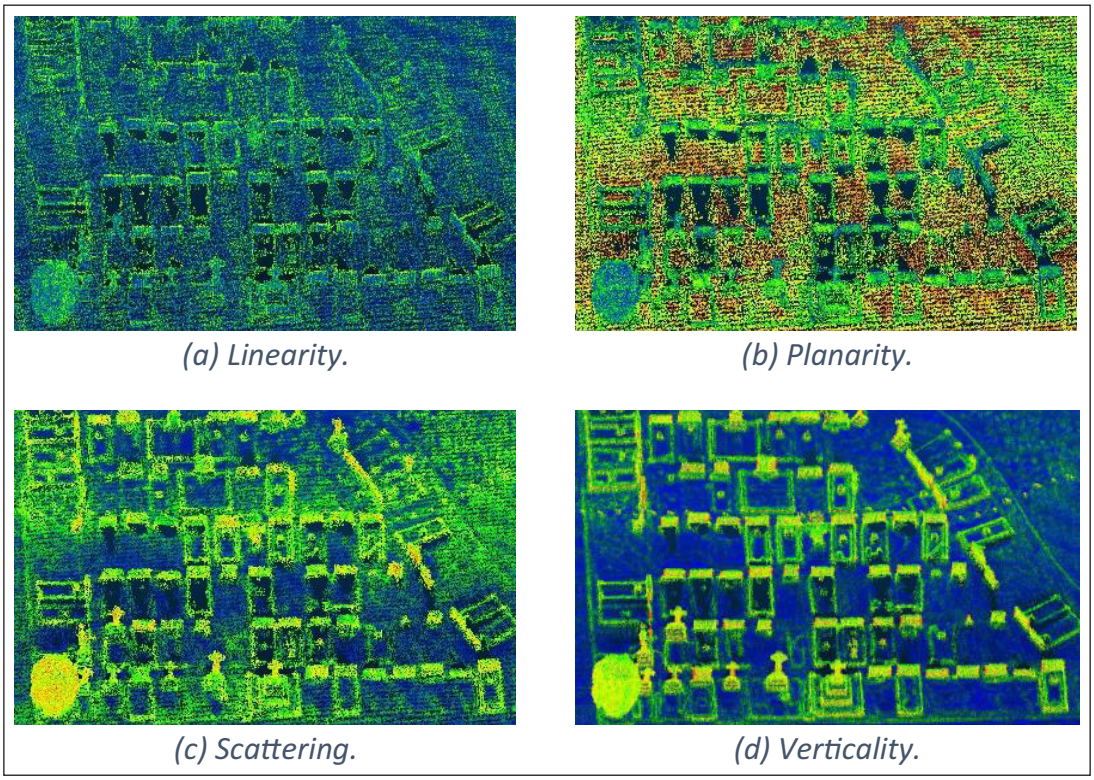


Figure 2 Geometric features shown for an example image containing grave markers: **(a)** linearity, **(b)** planarity, **(c)** scattering and **(d)** verticality. Point cloud data provided by Atlantic Geomatics (UK) Limited.

the potential to better represent adjacency, albeit at a significant computational trade-off. Refining this conclusion, Guinard and Landrieu 2017 advocated a graph that represents the adjacency of the 10 nearest points. Note that the neighbourhood of points represented in the adjacency graph is different to the neighbourhood used for feature extraction.

To encode the spatial relationship between points, the point cloud is represented using an undirected adjacency graph $G_{nn} = (V, E_{nn})$. The set of nodes $V = \{V_1, ..., V_n\}$ is constructed from each point in the point cloud, whereby each point p_i is represented by its associated features vector f_i, and the edges E_{nn} encode the adjacency relationship of the 10 nearest neighbour points (Niemeyer et al. 2011). Segmentation is then a process in which the graph is split optimally into non-overlapping connected components. These splits are computed using the l_0-cut pursuit algorithm (Landrieu and Obozinski 2017) and defined as the vector $g^* \in R^{4 \times n}$ which minimises the following generalised minimum partition model:

$$g^* = \arg\min \sum_{i \in V} \| g_i - f_i \|^2 + \rho \sum_{(i,j) \in E_{nn}} w_{i,j} \left[g_i - g_j \neq 0 \right],$$

with g as the variable value used to determine the optimal minimisation. The Iverson bracket $[\cdot]$ yields 0 if the internal expression is true, and 1 everywhere else. The edge weight $w \in R_+^{|E|}$ is chosen to be linearly decreasing with respect to the edge length and factor ρ is the regularisation strength, which determines the coarseness of the resulting partition (Landrieu and Simonovsky 2017). This formulation ensures that the resulting point cloud segments correspond to similar values of f without the need to define a maximum size for the segments. The point cloud segments are represented as the set $S = \{S_1, ..., S_h\}$, where h is the number of segments returned by the cut adjacency graph. For clarity, the segments are the non-overlapping connected components. The segments are subsets of the original point cloud and the number of points vary per segment, see ***Figure 3***.

To increase the chances of finding smaller objects that may have been missed in the initial segmentation, a conditional multi-scale partitioning scheme is proposed. This secondary conditional partition considers only the largest 10% of planar segments. These are passed through the segmentation process again, with the neighbourhood for feature extraction adjusted to within a radius defined by the point density. If new components are found, then they are added to the set of segments. Otherwise, the segment is assumed to have continuous local shape and considered to be a single segment. This process ensures ground level objects are detected in large segments of ground level points.

2.2 CLASSIFICATION

Using the 3D point clouds directly (instead of converted 2D representations) in the classification method is essential to the real-world applicability, efficiency and generalisability of the overall methodology. Conversions to other representations would not only result in information loss but would require an in-between representation, such as mesh models (Su et al. 2015). This conversation alone remains a difficult task when applied to fine resolution real-world 3D point cloud

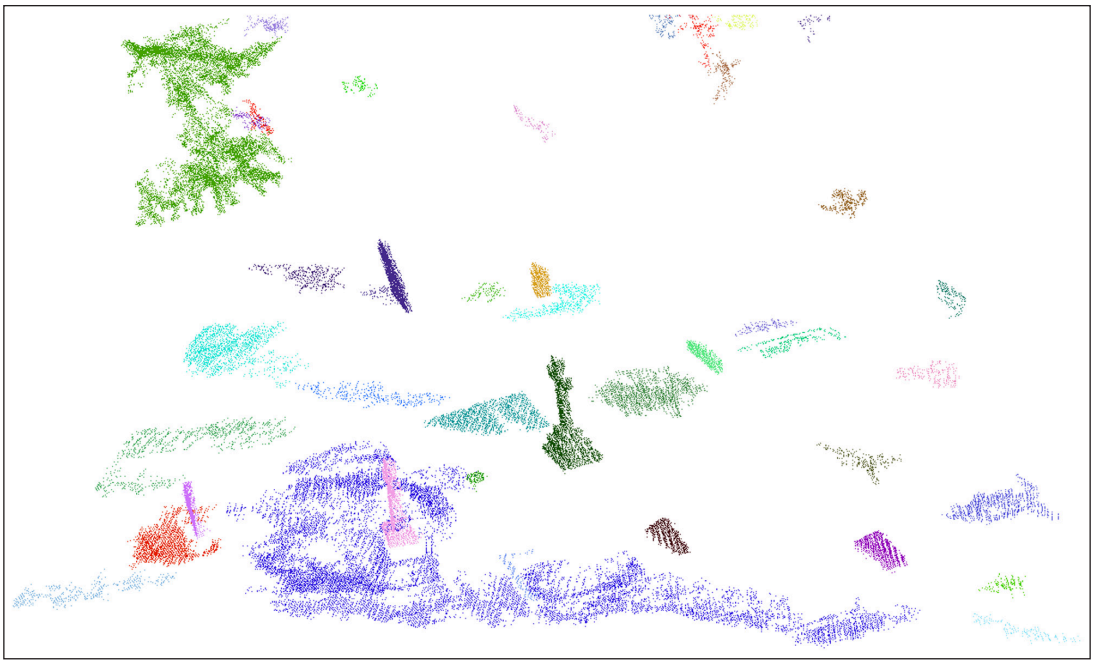


Figure 3 An example of a partitioned scene. The segments are assigned a colour randomly for demonstration purposes. Point cloud data provided by Atlantic Geomatics (UK) Limited.

data. Additionally, 2D multi-view methods are sensitive to viewpoint selection and occlusions within query instances. The real-world extracted objects are likely to contain noise from background objects (i.e., vegetation) and registration artefacts. While more recent multi-view 2D methods achieve leading accuracy scores on benchmarks, they rely on observational colour (*RGB*) information (Yu, Meng and Yuan 2018). Many fine-resolution point cloud datasets (especially those from LiDAR sensors) do not include this information as it requires specialised equipment and processing to collect and register the colour dimensions to the spatial points. Even if they are included, several factors present during the scanning process (e.g., glare, moisture, motion blur, camera focus, etc.) can contribute to inconsistent *RGB* values. This is not to say that *RGB* or other multispectral data should not be used when available, but that to remain generally applicable, the classification method should ingest directly the point cloud data and not be dependent on any additional observed features.

For the classification sequence the transfer learning paradigm is followed. The ConvPoint network is pre-trained using generic object models from the ModelNet40 benchmark. An example of the adapted ConvPoint network is provided in ***Figure 5***. The ConvPoint CNN was chosen because of its flexibility. It does not require a set input size and is robust to the permutation, scale and translation of the input 3D point cloud (Boulch 2019). As the name suggests, the Modelnet40 dataset is a collection of 3D models from across 40 different object

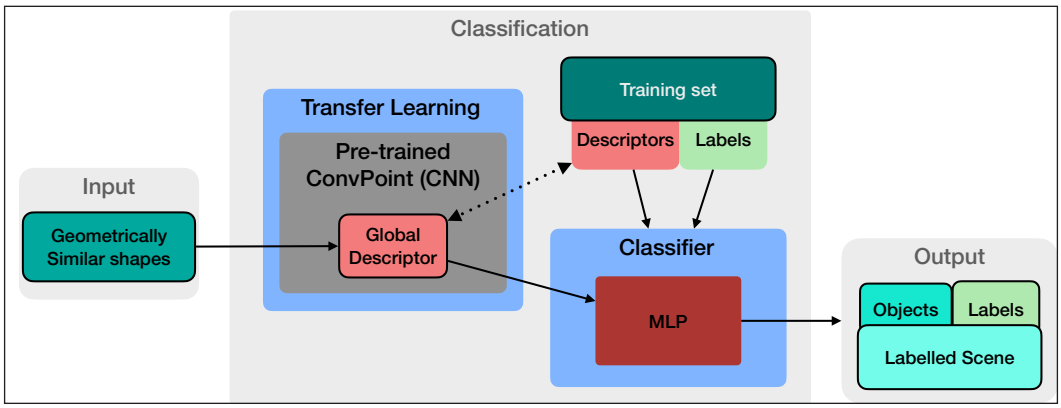


Figure 4 Illustration of the proposed classification methodology. The segments produced by the segmentation methodology are used as input; each is a 3D point cloud. Solid arrows represent the flow of processes. The dotted arrow indicates that the descriptors in the training set are created using the same pre-trained CNN model. The output to this process is a set of labels that relate the input point cloud segments to their predicted class.

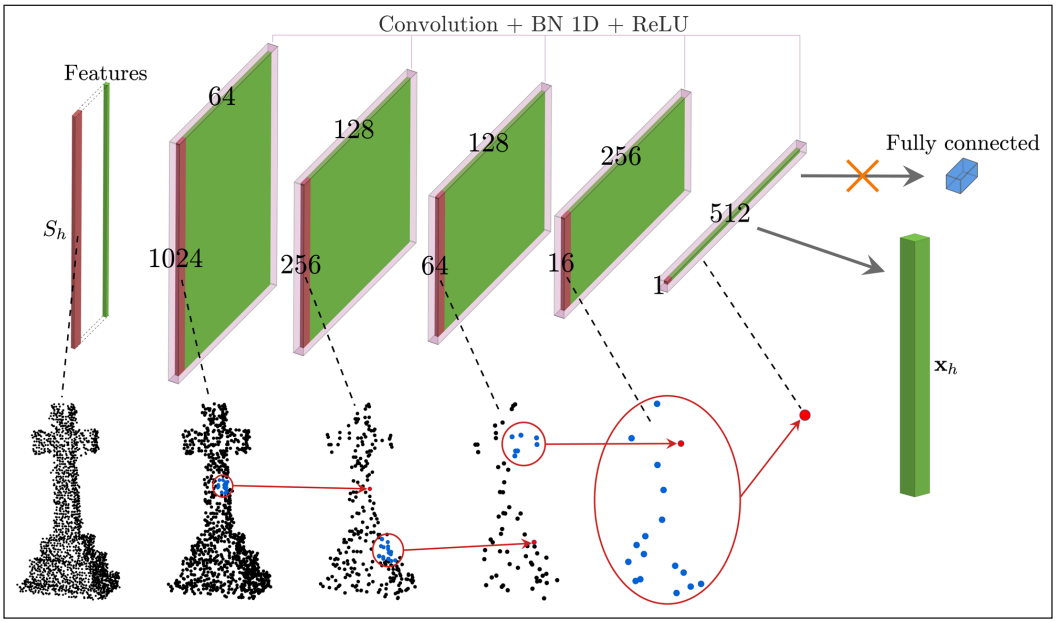


Figure 5 The ConvPoint CNN adapted as classification feature extractor. The CNN is composed of five convolutional layers. Each consists of a convolution operation, one-dimensional batch normalisation and uses the rectified linear unit for the activation function. At each layer, the number of points per object is reduced, which inversely increases the number of descriptive features. The features from the last layer are used for classification, rather than continue to the fully connected output layer of the original CNN.

classes. It is important to note that the classes available in ModelNet40 do not include directly any of the target labels for classification (e.g., cultural heritage objects). ConvPoint directly ingests the spatial coordinate points through an adaptation of discrete kernel convolutions to be continuous. A simple MLP learns a dense geometrical weighting function that independently distributes the input points onto a kernel. At each layer, the convolution operation effectively mixes the estimation in the feature space and geometrical space. The derived kernel is then an explicit set of points associated with weights. Normalisation is added according to the input set size (Boulch 2019). The final fully connected layer of the pre-trained network is used to leverage the weighted layers as a fixed feature extractor.

Using the last convolutional layer, we compute a 1 by 512 feature vector x_h, where $\{x\_h \mid 1, ..., h\}$ is a global descriptor for each segment S_h. This vector is used to define an abstract feature space which is optimised for the separation of the training objects. Transfer learning, as a concept, assumes that this feature space can also be used to separate the new test objects.

A simple MLP was trained to learn the difference in the feature space, therefore, leverage the knowledge learned to classify new data and apply semantic labels *L*. The MLP itself is formulated as one hidden layer with 100 units, one output layer and uses the logistic sigmoid activation function,

$$f(x) = 1/\left(1 + exp(-x)\right).$$

The features in *x* are assumed to be normally distributed and as such each is standardised by setting the features' mean at 0 and scaling to unit variance of 1; e.g., compute the standard score $z = (x - mean(x))/std(x)$ per feature. In doing so, we found this to increase classification accuracy results by at least 5%. See ***Figure 5*** for an illustration of the transfer learning procedure. We test and compare a variety of supervised classifiers for the classification task, which can be found in Section 3.

3 EXPERIMENTAL RESULTS AND ANALYSIS

3.1 DATASETS AND EVALUATION

3.1.1 ModelNet10

The Princeton ModelNet project provides a collection of 3D CAD object models split into two benchmarks: a 40-class subset and 10-class subset known as ModelNet40 and ModelNet10, respectively. The ModelNet10 data set was used to analyse the performance of the proposed transferred ConvPoint global descriptor in the classification process. The dataset was divided into training and validation sets. The CAD models were converted into 3D point clouds by randomly sampling points along the model surfaces. ***Table 1*** shows summary statistics for all 10 classes and their corresponding training and test samples.

CLASS NUMBER	CLASS NAME	TRAINING	TEST	TOTAL
1	bathtub	106	50	156
2	bed	515	100	615
3	chair	889	100	989
4	desk	200	86	286
5	dresser	200	86	286
6	monitor	465	100	565
7	nightstand	200	86	286
8	sofa	680	100	780
9	table	392	100	492
10	toilet	344	100	444

Table 1 Classification index for the ModelNet10 dataset; including number of training and test samples.

3.1.2 Cultural Heritage Scenes

Two separate cultural heritage sites represented as 3D point cloud scenes were chosen for the evaluation of the proposed methodology applied to real-world data. The digitised cultural heritage sites were provided by the burial ground management system team at Atlantic Geomatics (UK) Limited. Scene 1 is a burial ground from Adlington civil parish in North West England. The scene is a subset of a much larger scene; the same large scene from which the classification training data were acquired. Scene 2 is a separate dataset: a burial ground located in Staines-upon-Thames in South East England. It is not taken from a larger scene. The scenes were collected by a terrestrial LiDAR sensor platform with a relative accuracy of 2 to 3 cm. In the analysis, four separate semantic classes of objects were targeted; memorial objects (grave markers such as headstones, stone crosses, sarcophagus, etc.), infrastructure (buildings, walls, gates, street poles, etc.), vegetation (tall grasses, shrubs, trees, canopy leaves, etc.) and ground (grass terrain, roads, paths, etc.).

3.1.3 Evaluation metrics

Following the standard convention from the field of machine learning, Precision, Recall and F_1-score, along with their the macro- and weighted-average variations, were used as evaluation metrics. In a classification context, these metrics are ratios defined with respect to the number of true positives *TP*, false positives *FP* and false negatives *FN* returned per class. The recall, defined as $TP/(TP + FN)$, indicates the classifier's ability to find all positive samples. Likewise, precision is the fraction $TP/(TP + FP)$ that reflects the ability to return more relevant results then irrelevant ones. The F_1 is a measure of the classifier's accuracy. Formulated as F_1 = 2(Precision Recall/(Precision + Recall)), it considers both precision

and recall. All three metrics produce a score in the range [0, 1], reaching their worst value at 0 and best value at 1. The macro-average variation is then the mean of all scores divided by the number of classes. Similarly, the weighted-average is the score of each class weighted by the number of samples from that class.

3.1.4 Processing Platform

Experiments were run on a Unix machine with 2.7 GHz Intel Core i5, 16GB RAM and SSD. The combined process of segmentation and classification had an average run-time of 25 minutes for a point cloud of roughly 7 million points.

Several factors present during the scanning process can contribute to inconsistent observational colour (*R*, *G*, *B*) values. Furthermore, while some processes of point cloud acquisition, such as photogrammetry, inherently provide data as *R*, *G*, *B*, specialised equipment and processing is needed to register the colour dimensions to the point clouds generated by LiDAR sensors. As a result, many fine spatial resolution point cloud data sets do not include this information. In the proposed methodology we restrict the points to contain only the *x*, *y*, *z* coordinate information.

3.2 ANALYSIS OF TRANSFERRED DESCRIPTOR

3.2.1 Comparison of Classification Algorithms within the Proposed Methodology

Baselines are used generally to determine how well an algorithm performs. Thus, it can be a problem when a baseline for the specific domain does not exist. Applying the intuition behind transfer learning we conducted an initial experiment to gauge the effectiveness of our approach. We chose to assess a variety of supervised classification algorithms from the Scikit-learn Python package (Pedregosa et al. 2011) and test each against the ModelNet10 benchmark. A 'best case' for the baseline can be provided with the test data matching the data used to train the CNN feature extractor. To better reflect the objects recovered from the segmentation process, we chose to vary the number of points sampled, per model, to between 32 and 2048 points. The ModelNet10 point clouds were then given a global descriptor set using the adapted ConvPoint network. This allowed us to explore how different classifiers interact with the data and determine the most appropriate approach for classification.

We interpret from the results in ***Table 2*** that the multi-layer perceptron (MLP) network implementations are the most promising among the tested methods, although the linear support vector machine (SVM) achieves similar scores, placing it behind the MLP by as little as 1% in the majority of metric categories. Within this experiment we investigated and compared the behaviours of different MLP activation functions. In particular, the weighted-average F_1-score for the MLP with sigmoid activation performed particularly well, with an increase of at least 1–2% over the other MLP formulations. Consequently, this translates to a 7% and 5% F1-score increase over the next most accurate methods behind the SVM; random forest and *k*-nearest neighbours classifiers, respectively. The Gaussian Naive Bayes and decision tree classifiers performed the least accurate, where the MLP(sigmoid) had an increase of 21% and 19%, respectively.

3.2.2 Handcrafted global descriptors versus the transfer learning approach

The comparison of the proposed transfer learning global descriptor and commonly applied global features available in the open source Point Cloud Library (Rusu and Cousins 2011) are shown in ***Table 3***. These global

CLASSIFIER	PRECISION	RECALL	F1-SCORE
k-Nearest Neighbours	0.82/0.82	0.80/0.82	0.80/0.81
Gaussian Naive Bayes	0.71/0.74	0.68/0.70	0.68/0.70
Linear SVM	**0.87**/0.86	0.85/0.86	0.86/0.86
Random Forest	0.85/0.84	0.81/0.83	0.82/0.83
Decision Tree	0.67/0.68	0.66/0.68	0.66/0.67
MLP (sigmoid)	**0.87/0.87**	**0.86/0.87**	**0.87/0.87**
MLP (tanh)	**0.87/0.87**	**0.86/0.87**	0.86/**0.87**
MLP (relu)	0.86/0.86	0.85/0.86	0.85/0.86

Table 2 Results of the experiment to determine a baseline for classification. The values before the slash represent the macro-average score and after the slash the weighted-average score. The largest value for each metric is shown in bold.

DESCRIPTOR	PRECISION	RECALL	F1-SCORE
VFH(Rusu et al. 2010)	0.71/0.71	0.69/0.70	0.70/0.70
CVFH(Aldoma et al. 2011)	0.67/0.67	0.65/0.65	0.64/0.64
ESF(Wohlkinger and Vincze 2011)	0.05/0.05	0.17/0.19	0.07/0.08
GASD(Silva do Monte Lima and Teichrieb 2016)	0.81/0.80	0.79/0.80	0.79/0.80
Proposed transferred descriptor	**0.87/0.87**	**0.86/0.87**	**0.87/0.87**

Table 3 Classification results comparing global descriptors from the Point Cloud Library to the proposed transferred global descriptor on the ModelNet10 dataset. The values before the slash represent the macro-average score and after the slash the weighted-average score. The largest value for each metric is shown in bold.

descriptors are so called "handcrafted" and as such, have been designed specifically to encode certain aspects of a point cloud. This is in contrast to the transferred ConvPoint feature descriptor, which instead learned what features to encode from training data. The handcrafted descriptors were applied to classification of the ModelNet10 dataset. This experiment used the same methodology as described in Section 2.2, with their substitution as the global descriptor. The MLP(sigmoid) classifier was used as the classification model. It can be seen that the transferred descriptor from the ConvPoint network outperformed all handcrafted global descriptors across all three evaluation metrics, achieving a 7% increase in the weighted-average F1-score over the next best global descriptor implementation.

3.3 EVALUATION ON CULTURAL HERITAGE SITES

3.3.1 Comparison of Classifier Models with Real World Data

Real world data are often of varying quality and measured point clouds of real-world surfaces are no exception. They are commonly affected by variations in point density and objects of interest can occasionally become occluded during the scanning process. Therefore, we compare and investigate the behaviours of different classification models when applied to real world data of cultural heritage sites (***Figure 6***), although this should apply generally to any real-world dataset.

The classification results obtained for real world data from a pre-segmented cultural heritage scene are presented in ***Table 4***. Analysis of these scores indicates that the MLP classifier outperformed the other tested methods; thus, the MLP demonstrated an ability to handle the real-world data. This is in support of the earlier assessments of the MLP classifiers. It is interesting to note that, given real world data, the random forest model performed closely with the linear SVM, and in fact achieved the largest macro-average precision score; this is in contrast to the earlier experiment. Based on these results, we concluded that an MLP with sigmoid activation function is the most suitable, of the tested classifiers, for use within the proposed methodology.

CLASSIFIER	PRECISION	RECALL	F1-SCORE
k-Nearest Neighbours	0.89/0.89	0.76/0.89	0.79/0.88
Gaussian Naive Bayes	0.64/0.78	0.65/0.75	0.62/0.75
Linear SVM	0.86/0.90	0.82/0.90	0.84/0.89
Random Forest	**0.91**/0.89	0.76/0.89	0.80/0.88
Decision Tree	0.72/0.82	0.72/0.81	0.71/0.81
MLP (sigmoid)	0.88/**0.91**	**0.83/0.91**	**0.85/0.91**
MLP (tanh)	0.87/**0.91**	**0.83/0.91**	0.84/**0.91**
MLP (relu)	0.87/**0.91**	**0.83**/0.90	**0.85**/0.90

Table 4 Comparison of different classification methods applied to memorial objects from a pre-segmented scene. Performance is evaluated using precision, recall and F_1-score; the values before the slash represents the macro-average score and after the slash the weighted-average score. The largest score for each metric is shown in bold.

3.3.2 Evaluation of Methodology on Cultural Heritage Scenes

We applied the methodology to two separate cultural heritage scenes, the results of which are shown in ***Tables 5*** and ***6***. The same training data were used in both scenes to train the MLP classifier. For both scenes, the proposed approach achieved a weighted average of at least 91% across all metrics. In general, classification of memorial objects was highly accurate, with an F_1-score of 92% and 95% for scenes 1 and 2, respectively. The results from scene 2 illustrates how the proposed methodology generalises to a different dataset without retraining, even across two different and spatially distant regions. Classification accuracies for memorial, vegetation and infrastructure objects were similar to the scores in scene 1. However, there is a decrease when correctly determining segments that contain ground points. This is likely explained by the difference in terrains (e.g., slopes, flats, hills, etc.) between scenes; without the addition of these landscape characteristics the classification model can struggle to accommodate these changes.

	PRECISION	RECALL	F1-SCORE
Memorial	0.95	0.88	0.92
Infrastructure	0.56	0.83	0.67
Vegetation	0.94	0.94	0.94
Ground	0.85	0.94	0.89
macro avg.	0.83	0.90	0.90
weighted avg.	0.92	0.91	0.91

Table 5 Precision, recall and F1-score of MLP classification applied to Scene 1. Scores are an average result after 100 runs.

	PRECISION	RECALL	F1-SCORE
Memorial	0.94	0.95	0.95
Infrastructure	0.73	0.63	0.68
Vegetation	0.91	0.91	0.91
Ground	0.82	0.83	0.83
macro avg.	0.85	0.83	0.84
weighted avg.	0.91	0.91	0.91

Table 6 Precision, recall and F1-score of MLP classification applied to Scene 2. Scores are an average result after 100 runs.

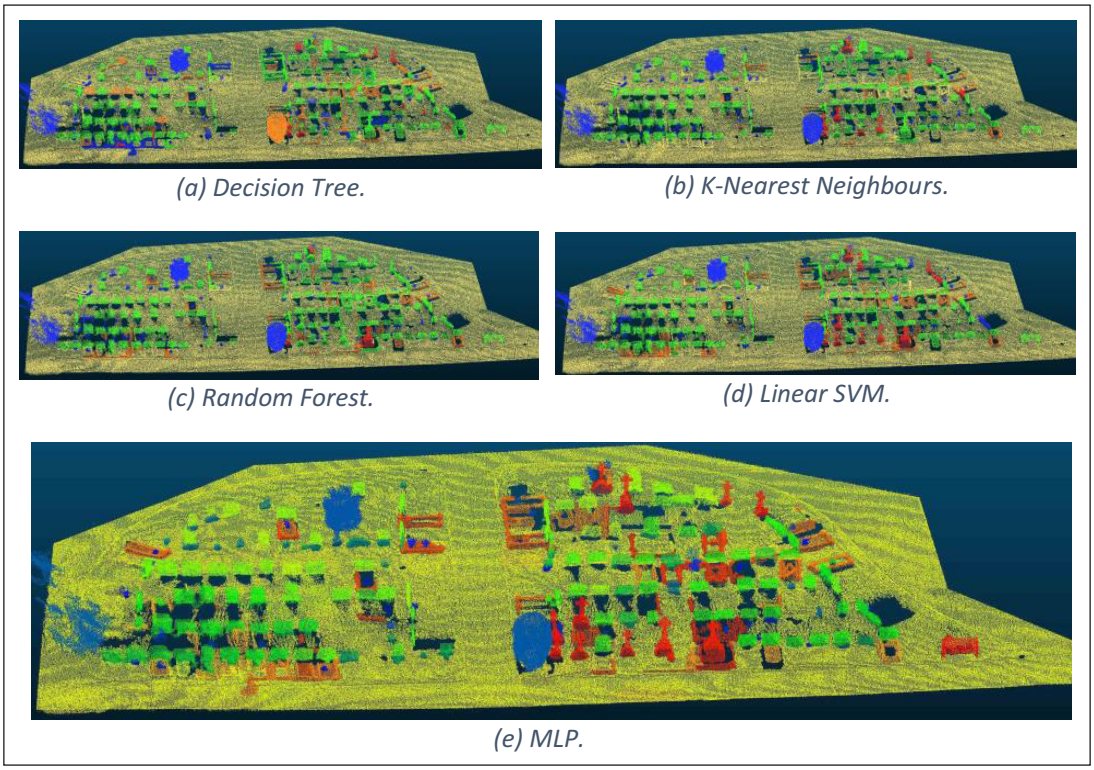


Figure 6 Examples of a classified region from different classifier methods. Ground points are represented in yellow; vegetation is in blue; infrastructure is in orange and memorial objects are marked in green and red showing two sub-class identifications (headstone and cross). Point cloud data was provided by Atlantic Geomatics (UK) Limited.

4 DISCUSSION

Misclassification of the cultural heritage data lies in the infrastructure objects class. This is, in part, to be expected as memorial objects are often subjective. Cultural heritage sites normally contain various items of street furniture, and those in and of themselves might be a type of monument, e.g., a bench object may be classified into either the memorial or infrastructure class depending on semantics alone, with no visually distinct reasoning. The same can be true for trees and shrubs. The experiment classification index adheres strictly to memorial and non-memorial objects based on the manually labelled scene, which does not take this into account. This raises questions of how to impose semantic meaning to objects with little or no visually discerning attributes.

Pre-segmentation can also influence the classification results. It is possible for buildings and walls to be partitioned into smaller parts which share characteristics with headstone monuments or even vegetation, thus, potentially resulting in misclassification. In this sense, classification results are contingent on the quality of the segmentation process. An alternative to an object-wise segmentation would be to use a region-growing or point-wise algorithm. However, point density and noise have a direct impact on the time complexity of such methods. As a result, they have a limited ability to segment large-scale point clouds (Landrieu and Simonovsky 2017). Experiments showed that the proposed methodology is capable of running on a personal computer. However, we note that RAM capacity was a limiting factor. Considerably large point clouds may need to be divided manually into smaller regions beforehand or else down sampled, provided that there is no great loss in visual representation. The objects in question should be easily identified by an operator.

In the future, we are interested in exploring how a more fine-grained classification could be achieved *within* the object classes. For example, many different burial ground monument types are found in a single cemetery. Additionally, grave markers from different geographical areas, different time periods and coming from different cultures, likely appear distinct and different from one another. With the general public becoming more interested in family ancestry and genealogy, there is a real need for this information to be available and to be provided at scale. Similarly, we are interested in ways to incorporate new object variations, unknown objects and additional classes to the methodology. The grounds, building infrastructure and serviceable equipment, etc., are all objects of importance for the maintenance of cultural heritage sites. It is, therefore, of value if the classification model does not have to be completely retrained each time a new variation is seen. Furthermore, based on the results of the transfer learning approach within this research, we are motivated to explore the use of various point cloud specific neural networks as feature extractors and evaluate their relative performances.

5 CONCLUSIONS

We presented a new methodology for the automatic identification and extraction of objects from 3D point cloud representations of cultural heritage sites. This methodology addressed how point cloud data can be used directly to map and extract objects from archaeological and cultural heritage contexts, without the need to rasterise or transform the data into another representation beforehand. Benchmarking exercises established that, compared to several classification methods, the proposed methodology achieves a statistically higher accuracy for both artificial and real-world datasets. We applied the methodology to the task of locating, extracting and labelling grave marker objects from two cultural heritage sites. The results demonstrated that the proposed approach can leverage transfer learning to separate objects from the scene and distinguish between multiple classes. We believe that this is the first time that such a methodology has been developed for the automatic and direct extraction and labelling of memorial objects from cultural heritage sites using 3D point cloud data.

ACKNOWLEDGEMENTS

The authors thank members at Atlantic Geomatics (UK) Limited for providing access to the raw LiDAR data and reference datasets.

FUNDING INFORMATION

This research was supported by Cumbria Innovations Platform (CUSP) at Lancaster University and the European Regional Development Fund (ERDF).

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Nicholas Arnold *orcid.org/0000-0003-3968-6233*
Lancaster University, GB

Plamen Angelov *orcid.org/0000-0002-5770-934X*
Lancaster University, GB

Tim Viney
Atlantic Geomatics, GB

Peter Atkinson *orcid.org/0000-0002-5489-6880*
Lancaster University, GB

REFERENCES

Aldoma, A, Vincze, M, Blodow, N, Gossow, D, Gedikli, S, Rusu, RB and **Bradski, G.** 2011. CAD-model recognition and 6DOF pose estimation using 3D cues. In: *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. November 2011. pp. 585–592. DOI: *https://doi.org/10.1109/ICCVW.2011.6130296*

Aubry, M, Schlickewei, U and **Cremers, D.** 2011. *The wave kernel signature: A quantum mechanical approach to shape analysis*. 1 November 2011. pp. 1626–1633. DOI: *https://doi.org/10.1109/ICCVW.2011.6130444*

Bello, SA, Yu, S and **Wang, C.** 2020. Review: deep learning on 3D point clouds. *arXiv:2001.06280 [cs]*. DOI: *https://doi.org/10.3390/rs12111729*

Boulch, A. 2019. ConvPoint: continuous convolutions for cloud processing. *arXiv:1904.02375 [cs]*. DOI: *https://doi.org/10.1016/j.cag.2020.02.005*

Bronstein, MM, Bruna, J, LeCun, Y, Szlam, A and **Vandergheynst, P.** 2017. Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4): 18–42. DOI: *https://doi.org/10.1109/MSP.2017.2693418*

Bronstein, MM and **Kokkinos, I.** 2010. Scale-invariant heat kernel signatures for non-rigid shape recognition. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. June 2010. pp. 1704–1711. DOI: *https://doi.org/10.1109/CVPR.2010.5539838*

Cannell, RJS, Gustavsen, L, Kristiansen, M and **Nau, E.** 2018. Delineating an Unmarked Graveyard by High-Resolution GPR and pXRF Prospection: The Medieval Church Site of Furulund in Norway. *Journal of Computer Applications in Archaeology*, 1(1): 1–18. DOI: *https://doi.org/10.5334/jcaa.9*

Chase, ASZ, Chase, DZ and **Chase, AF.** 2017. LiDAR for Archaeological Research and the Study of Historical Landscapes. In: Masini, N and Soldovieri, F (eds.), *Sensing the Past. Geotechnologies and the Environment*. Cham: Springer International Publishing. pp. 89–100. DOI: *https://doi.org/10.1007/978-3-319-50518-3_4*

Dai, J, Qi, H, Xiong, Y, Li, Y, Zhang, G, Hu, H and **Wei, Y.** 2017. Deformable Convolutional Networks. *arXiv:1703.06211 [cs]*. DOI: *https://doi.org/10.1109/ICCV.2017.89*

Demantké, J, Mallet, C, David, N and **Vallet, B.** 2012. Dimensionality Based Scale Selection in 3d Lidar Point Clouds. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII-5/W12: 97–102. DOI: *https://doi.org/10.5194/isprsarchives-XXXVIII-5-W12-97-2011*

Dohan, D, Matejek, B and **Funkhouser, T.** 2015. Learning Hierarchical Semantic Segmentations of LIDAR Data. In: *2015 International Conference on 3D Vision*. October 2015 pp. 273–281. DOI: *https://doi.org/10.1109/3DV.2015.38*

Favorskaya, MN and **Jain, LC.** 2017. Overview of LiDAR Technologies and Equipment for Land Cover Scanning. In:

Favorskaya, MN and Jain, LC (eds.), *Handbook on Advances in Remote Sensing and Geographic Information Systems: Paradigms and Applications in Forest Landscape Modeling. Intelligent Systems Reference Library*. Cham: Springer International Publishing. pp. 19–68. DOI: https://doi.org/10.1007/978-3-319-52308-8_2

Filin, S and **Pfeifer, N.** 2005. Neighborhood Systems for Airborne Laser Data. *Photogrammetric Engineering & Remote Sensing*, 71: 743–755. DOI: https://doi.org/10.14358/PERS.71.6.743

Fryskowska, A, Kedzierski, M, Walczykowski, P, Wierzbicki, D, Delis, P and **Lada, A.** 2017. Effective Detection of Sub-Surface Archeological Features from Laser Scanning Point Clouds and Imagery Data. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W5: 245–251. DOI: https://doi.org/10.5194/isprs-archives-XLII-2-W5-245-2017

Golovinskiy, A, Kim, VG and **Funkhouser, T.** 2009. *Shape-based recognition of 3D point clouds in urban environments*. In: September 2009 IEEE. pp. 2154–2161. DOI: https://doi.org/10.1109/ICCV.2009.5459471

Graham, B. 2015. Sparse 3D convolutional neural networks. *arXiv:1505.02890 [cs]*. DOI: https://doi.org/10.5244/C.29.150

Guinard, S and **Landrieu, L.** 2017. Weakly Supervised Segmentation-Aided Classification of Urban Scenes from 3d Lidar Point Clouds. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-1/W1: 151–157. DOI: https://doi.org/10.5194/isprs-archives-XLII-1-W1-151-2017

Guyot, A, Hubert-Moy, L and **Lorho, T.** 2018. Detecting Neolithic Burial Mounds from LiDAR-Derived Elevation Data Using a Multi-Scale Approach and Machine Learning Techniques. *Remote Sensing*, 10(2): 225. DOI: https://doi.org/10.3390/rs10020225

Hackel, T, Wegner, JD and **Schindler, K.** 2016. *Contour Detection in Unstructured 3D Point Clouds*. June 2016 IEEE. pp. 1610–1618. DOI: https://doi.org/10.1109/CVPR.2016.178

Jaklič, A, Erič, M, Mihajlović, I, Stopinšek, Ž and **Solina, F.** 2015 Volumetric Models from 3d Point Clouds: The Case Study of Sarcophagi Cargo from a 2nd/3rd Century Ad Roman Shipwreck Near Sutivan on Island BračCroatia. *Journal of Archaeological Science*, 62: 143–152. DOI: https://doi.org/10.1016/j.jas.2015.08.007

Johnson, AE and **Hebert, M.** 1999. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5): 433–449. DOI: https://doi.org/10.1109/34.765655

Jonathan, E, Roberts, C, Presentations, S, Linsen, L and **Prautzsch, H.** 2001. *Local Versus Global Triangulations*.

Kazimi, B, Thiemann, F, Malek, K, Sester, M and **Khoshelham, K.** 2018. *Deep Learning for Archaeological Object Detection in Airborne Laser Scanning Data*. 1 September 2018 p.

Klokov, R and **Lempitsky, V.** 2017. Escape from Cells: Deep Kd-Networks for the Recognition of 3D Point Cloud Models. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. October 2017 Venice: IEEE. pp. 863–872. DOI: https://doi.org/10.1109/ICCV.2017.99

Kramer, I. 2015. *An archaeological reaction to the remote sensing data explosion. Reviewing the research on semi-automated pattern recognition and assessing the potential to integrate artificial intelligence*. PhD Thesis.

Landrieu, L and **Obozinski, G.** 2017. Cut Pursuit: fast algorithms to learn piecewise constant functions on general weighted graphs. *SIAM Journal on Imaging Sciences*, 10(4): 1724–1766. DOI: https://doi.org/10.1137/17M1113436

Landrieu, L and **Simonovsky, M.** 2017. Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs. *CoRR*, abs/1711.09869. DOI: https://doi.org/10.1109/CVPR.2018.00479

Lee, I and **Schenk, A.** 2002. Perceptual organization of 3D surface points. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34.

Ling, H and **Jacobs, DW.** 2007. Shape Classification Using the Inner-Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2): 286–299. DOI: https://doi.org/10.1109/TPAMI.2007.41

Niemeyer, J, Rottensteiner, F and **Soergel, U.** 2014. Contextual classification of lidar data and building object detection in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87: 152–165. DOI: https://doi.org/10.1016/j.isprsjprs.2013.11.001

Niemeyer, J, Wegner, JD, Mallet, C, Rottensteiner, F and **Soergel, U.** 2011. Conditional Random Fields for Urban Scene Classification with Full Waveform LiDAR Data. In: Stilla, U, Rottensteiner, F, Mayer, H, Jutzi, B and Butenuth, M (eds.), *Photogrammetric Image Analysis. Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer. pp. 233–244. DOI: https://doi.org/10.1007/978-3-642-24393-6_20

Pedregosa, F, Varoquaux, G, Gramfort, A, Michel, V, Thirion, B, Grisel, O, Blondel, M, Prettenhofer, P, Weiss, R, Dubourg, V, Vanderplas, J, Passos, A, Cournapeau, D, Brucher, M, Perrot, M and **Duchesnay, E.** 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12: 2825–2830.

Qi, CR, Su, H, Mo, K and **Guibas, LJ.** 2016a. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *arXiv:1612.00593 [cs]*.

Qi, CR, Su, H, Niessner, M, Dai, A, Yan, M and **Guibas, LJ.** 2016b. Volumetric and Multi-View CNNs for Object Classification on 3D Data. *arXiv:1604.03265 [cs]*. DOI: https://doi.org/10.1109/CVPR.2016.609

Qi, CR, Yi, L, Su, H and **Guibas, LJ.** 2017. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. *arXiv:1706.02413 [cs]*. DOI: https://doi.org/10.1109/CVPR.2017.701

Riegler, G, Ulusoy, AO and **Geiger, A.** 2017. OctNet: Learning Deep 3D Representations at High Resolutions. *arXiv:1611.05009 [cs]*. DOI: https://doi.org/10.1109/CVPR.2017.701

Royo, S and **Ballesta-Garcia, M.** 2019. An Overview of Lidar Imaging Systems for Autonomous Vehicles. *Applied Sciences*, 9(19): 4093. DOI: https://doi.org/10.3390/app9194093

Rusu, RB, Bradski, G, Thibaux, R and **Hsu, J.** 2010. Fast 3D recognition and pose using the Viewpoint Feature Histogram. In: *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. October 2010. Taipei: IEEE. pp. 2155–2162. DOI: *https://doi.org/10.1109/IROS.2010.5651280*

Rusu, RB and **Cousins, S.** 2011. 3D is here: Point Cloud Library (PCL). In: *2011 IEEE International Conference on Robotics and Automation*. May 2011. pp. 1–4. DOI: *https://doi.org/10.1109/ICRA.2011.5980567*

Sevara, C, Pregesbauer, M, Doneus, M, Verhoeven, G and **Trinks, I.** 2016. Pixel versus object—A comparison of strategies for the semi-automated mapping of archaeological features using airborne laser scanning data. *Journal of Archaeological Science: Reports*, 5: 485–498. DOI: *https://doi.org/10.1016/j.jasrep.2015.12.023*

Shapovalov, R, Velizhev, E and **Barinova, O.** 2010. Nonassociative markov networks for 3d point cloud classification. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII, Part 3A*. 2010. pp. 103–108.

Silva do Monte Lima, JP and **Teichrieb, V.** 2016. An Efficient Global Point Cloud Descriptor for Object Recognition and Pose Estimation. In: *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. October 2016. pp. 56–63. DOI: *https://doi.org/10.1109/SIBGRAPI.2016.017*

Simonovsky, M and **Komodakis, N.** 2017. Dynamic Edge-Conditioned Filters in Convolutional Neural Networks on Graphs. *arXiv:1704.02901 [cs]*. DOI: *https://doi.org/10.1109/CVPR.2017.11*

Spina, S, Debattista, K, Bugeja, K and **Chalmers, A.** 2011. *Point Cloud Segmentation for Cultural Heritage Sites*. 1 January 2011. pp. 41–48. DOI: *https://doi.org/10.2312/VAST/VAST11/041-048*.

Su, H, Jampani, V, Sun, D, Maji, S, Kalogerakis, E, Yang, M-H and **Kautz, J.** 2018. SPLATNet: Sparse Lattice Networks for Point Cloud Processing. *arXiv:1802.08275 [cs]*. DOI: *https://doi.org/10.1109/CVPR.2018.00268*

Su, H, Maji, S, Kalogerakis, E and **Learned-Miller, E.** 2015. Multi-view Convolutional Neural Networks for 3D Shape Recognition. *arXiv:1505.00880 [cs]*. DOI: *https://doi.org/10.1109/ICCV.2015.114*

Trier, ØD, Cowley, DC and **Waldeland, AU.** 2019. Using deep neural networks on airborne laser scanning data: Results from a case study of semi-automatic mapping of archaeological topography on Arran, Scotland. *Archaeological Prospection*, 26(2): 165–175. DOI: *https://doi.org/10.1002/arp.1731*

Trier, Ø, Salberg, A-B, Pilø, L, Tonning, C, Johansen, H and **Aarsten, D.** 2016. *Semi-automatic mapping of cultural heritage from airborne laser scanning using deep learning*. 17 April 2016.

Verschoof-van der Vaart, WB and **Lambers, K.** 2019. Learning to Look at Lidar: The Use of R-CNN in the Automated Detection of Archaeological Objects in Lidar Data from the Netherlands. *Journal of Computer Applications in Archaeology*, 2(1): 31–40. DOI: *https://doi.org/10.5334/jcaa.32*

Weinmann, M, Schmidt, A, Mallet, C, Hinz, S, Rottensteiner, F and **Jutzi, B.** 2015 Contextual Classification of Point Cloud Data by Exploiting Individual 3d Neigbourhoods. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*. March 2015. Copernicus GmbH. pp. 271–278. DOI: *https://doi.org/10.5194/isprsannals-II-3-W4-271-2015*

Weitman, S. 2012. *Using Archaeological Methods in Cemetery Surveys with Emphasis on the Application of Lidar*. Georgia Southern University.

Wohlkinger, W and **Vincze, M.** 2011. Ensemble of shape functions for 3D object classification. In: *2011 IEEE International Conference on Robotics and Biomimetics*. December 2011. pp. 2987–2992. DOI: *https://doi.org/10.1109/ROBIO.2011.6181760*

Wu, Z, Song, S, Khosla, A, Yu, F, Zhang, L, Tang, X and **Xiao, J.** 2015. 3D ShapeNets: A deep representation for volumetric shapes. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2015. pp. 1912–1920. DOI: *https://doi.org/10.1109/CVPR.2015.7298801*

Yu, T, Meng, J and **Yuan, J.** 2018. Multi-view Harmonized Bilinear Network for 3D Object Recognition. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 2018. Salt Lake City, UT: IEEE. pp. 186–194. DOI: *https://doi.org/10.1109/CVPR.2018.00027*

Zacharek, M, Delis, P, Kedzierski, M and **Fryskowska, A.** 2017. Generating Accurate 3d Models of Architectural Heritage Structures Using Low-Cost Camera and Open Source Algorithms. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-5/W1: 99–104. DOI: *https://doi.org/10.5194/isprs-archives-XLII-5-W1-99-2017*

Zingman, I, Saupe, D, Penatti, OAB and **Lambers, K.** 2016. Detection of Fragmented Rectangular Enclosures in Very High Resolution Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8): 4580–4593. DOI: *https://doi.org/10.1109/TGRS.2016.2545919*

TO CITE THIS ARTICLE:
Arnold, N, Angelov, P, Viney, T and Atkinson, P. 2021. Automatic Extraction and Labelling of Memorial Objects From 3D Point Clouds. *Journal of Computer Applications in Archaeology*, 4(1), 79–93. DOI: *https://doi.org/10.5334/jcaa.66*

Submitted: 30 September 2020 Accepted: 23 March 2021 Published: 23 April 2021

Chapter 3

An Improved eXplainable Point Cloud Classifier (XPCC)

The preceding chapters of this thesis have illuminated the challenges and innovative methodologies involved in automated object extraction and classification from 3D point cloud scenes within the domain of cultural heritage and archaeology. Chapter 2 introduced the GeoPart-Transfer framework, which addressed the complexities of utilising point clouds in real-world applications and demonstrated the potential of convolutional neural networks (CNNs) in object segmentation and labelling.

In the pursuit of advancing the state-of-the-art in this field, Chapter 3 introduces the eXplainable Point Cloud Classifier (XPCC) method. This chapter underscores the critical need for explainability in AI models applied to cultural heritage management and preservation. XPCC not only inherits the robustness of CNNs but also introduces interpretability to the decision-making process.

The relevance of Chapter 3 lies in its crucial role as a bridge between the GeoPart-Transfer framework and the overarching objectives of the thesis. While GeoPart-Transfer laid the groundwork for object extraction and classification, XPCC extends this foundation by introducing an explainable methodology that empowers practitioners to comprehend and learn from the classification outcomes.

This chapter delves into the core of XPCC, presenting its design principles, features, and experimental validation. The chapter begins by introducing the concept of explainability and its importance within the context of cultural heritage and archaeology. It elaborates on how XPCC integrates interpretability with high classification accuracy, filling a critical gap in the current landscape of object extraction methods. The XPCC melds local densities and global multivariate generative distributions, ensuring both comprehensive and interpretable object-based categorisation. The intrinsic recursion of XPCC contributes to computational efficiency, enabling continuous learning. Uniquely tailored for 3D data, the method

introduces three novel layers within the architecture, augmenting feature extraction, global compound prototype weighting, and SoftMax functionality.

The experimentation phase is showcased through benchmarking XPCC on the ModelNet40 dataset, demonstrating its effectiveness in classification accuracy and explainability over existing approaches. These results validate XPCC's contribution to the domain of cultural heritage and underline its alignment with the thesis's mission to empower cultural heritage management through advanced technological solutions.

In summary, this chapter outlines the XPCC classifier's inception, detailing its design principles, benefits, and performance. This work not only enriches the toolbox of 3D object classification but also aligns seamlessly with the thesis's mission to empower cultural heritage management through advanced technological solutions.

An Improved Explainable Point Cloud Classifier (XPCC)

Nicholas I. Arnold, Plamen Angelov, *Fellow, IEEE*, and Peter M. Atkinson

***Abstract*—Classification of objects from 3-D point clouds has become an increasingly relevant task across many computer-vision applications. However, few studies have investigated explainable methods. In this article, a new prototype-based and explainable classification method called eXplainable point cloud classifier (XPCC) is proposed. The XPCC method offers several advantages over previous explainable and nonexplainable methods. First, the XPCC method uses local densities and global multivariate generative distributions. Therefore, the XPCC provides comprehensive and interpretable object-based classification. Furthermore, the proposed method is built on recursive calculations, thus, is computationally very efficient. Second, the model learns continuously without the need for complete retraining and is domain transferable. Third, the proposed XPCC expands on the underlying learning method explainable deep neural networks (xDNN), and is specific to 3-D. As such, the following three new layers are added to the original xDNN architecture: 1) the 3-D point cloud feature extraction, 2) the global compound prototype weighting, and 3) the SoftMax function. Experiments were performed with the ModelNet40 benchmark, which demonstrated that XPCC is the only one to increase classification accuracy relative to the base algorithm when applied to the same problem. In addition, this article proposes a novel prototype-based visual representation that provides model- and object-based explanations. The prototype objects are superimposed to create a prototypical class representation of their data density within the feature space, called the compound prototype cloud. They allow a user to visualize the explainable aspects of the model and identify object regions that contribute to the classification in a human-understandable way.**

***Impact Statement*—The classification of 3D point cloud data has become a significant topic in recent years, in part because of the popularisation of various unmanned robotics, augmented reality, and 3D mapping software. Such applications often involve decisions direct consequences to individuals and society, yet very little research has been done towards explainable 3D classification algorithms. This article proposes an inherently explainable prototype-based classification and visualisation method for 3D point cloud objects. Experiments demonstrate that the proposed method is not only competitive with the state-of-the-art, but that it is also transferable and improves accuracy over the base algorithm.**

***Index Terms*—3-D, AI, classification, deep learning, explainable, point cloud data.**

Manuscript received August 24, 2021; revised November 15, 2021 and January 8, 2022; accepted February 8, 2022. This work was supported in part by the Cumbria Innovations Platform (CUSP), Lancaster University. This article was recommended for publication by Associate Editor Manuel Roveri upon evaluation of the reviewers' comments. *(Corresponding author: Nicholas I. Arnold.)*

The authors are with the School of Computing and Communications Lancaster University, Lancaster University, LA1 4YW Lancaster, U.K. (e-mail: n.arnold@lancaster.ac.uk; p.angelov@lancaster.ac.uk; pma@lancaster.ac.uk).

Digital Object Identifier 10.1109/TAI.2022.3150647

I. Introduction

CLASSIFICATION of 3-D point cloud data has become an important research goal in response to the widespread adoption of 3-D sensor technologies, such as LiDAR and RGB-depth cameras. The 3-D point cloud data are a sparse collection of unordered coordinates in 3-D space. They offer a fine-grained representation of real-world objects and accurately preserve intrinsic geometric 3-D shape, surface, and depth information. Point clouds have become an increasingly relevant data structure across a range of computer-vision applications, including remote sensing, autonomous driving, and robotics. In many cases such applications may have important real-world consequences (e.g., misperception of the environment may lead to the collision of an autonomous vehicle). Therefore, it is critical that a 3-D object classification model is not only efficient in terms of accuracy and speed, but that the model's decisions can be understood intuitively and interpreted correctly by a human. In this article, we propose a new prototype-based classification method called eXplainable point cloud classifier (XPCC) for object classification of 3-D point cloud objects. For clarity, this article adopts the terminology supported by [1] to describe explainable machine-learning models.

Early methods for 3-D point cloud object classification rely on handcrafted features extracted directly from local neighborhood regions or through estimation of the surface around each point. For example, the fast point feature histogram (FPFH) [2] algorithm encodes the local geometric shape based on the normal angle between points and their neighbors, Rusu *et al.* [3] used binning to extend the FPFH into a global object descriptor and Johnson and Hebert [4] built a histogram of point locations summed along the bins of an accumulator constructed around each point to create an image. These features are designed to be invariant to shape transformations. Classification based on these handcrafted features can then be achieved through classical supervised machine-learning algorithms. These handcrafted features are explainable at a human level precisely because they were handcrafted to describe specific (local or global) properties of the shape. However, it is not trivial to find the most effective feature combination for a specific task. In this regard, the move to features learned through artificial neural network (NN) algorithms, such as convolutional neural networks (CNN) and other deep neural networks (DNN), was a breakthrough. The XPCC method benefits from the transfer learning paradigm [5] to incorporate learned features while retaining human interpretability.

Deep learning applied to 3-D point clouds is far from being straightforward, as point cloud data are unordered and nonstructured [6]. That is, there are no defined neighborhoods to connect each point in space. This contrasts with 2-D images, where each pixel sits on a grid and has explicitly defined neighboring pixels. Recently, however, deep learning classifiers, such as PointNet [7] and its derivatives [8], have been proposed and adapted to the specific properties of 3-D point cloud data. These classifiers learn an embedding for each point and aggregate this information into a global shape descriptor. Classification is then achieved by feeding the global descriptor into several fully connected layers. By removing these last fully connected layers, the XPCC method uses a fixed pretrained CNN to act directly on the point cloud data by extracting a global feature vector from the 3-D point cloud objects. The choice of fixed feature extractor is not linked to any method in particular, it is, therefore, modular and can updated as research into DNN on point clouds progress.

Several characteristics of deep learning algorithms limit their wider real-world application. First, DNNs are domain-specific; that is, they make classifications through learned properties as determined by the data on which they were trained. The addition of new classes or even additional data that do not follow the same statistical characteristics as the training data requires a complete retrain of the network. Second, training of DNNs is computationally demanding and requires substantial numbers of training data, computational resources, and time. While several 3-D point cloud benchmark datasets have been published, the classes available are far from exhaustive. Therefore, training on atypical and uncommon classes is problematic. XPCC overcomes these limitations through both task domain transfer and learning domain transfer. Adding a new class requires only training the model on the new data samples, rather than a complete retrain, and the new classes are not required to be known to the feature-extraction method. Furthermore, classification can be achieved with only a few training samples per class.

Deep learning methods, such as CNNs, involve and require a large number (millions or more) of model parameters (network weights), which have no direct link to the physicality of the problem. In addition, the architectures of DNNs, such as CNNs involve several *ad hoc* decisions about the number and type of layers, stride, and kernel size. Due to this complexity, and the opacity of the link between the inputs (point cloud coordinates) and the output (class label), such solutions are considered as "black-box" [9]. The output is a multilevel embedded function (i.e., a function of a function of a function) of inputs. This makes it difficult to explain the cause—effect relationship and the intuition of how the final decision is arrived at to a human user. A lack of transparency is a particular drawback in the case of 3-D point cloud-object classification. 3-D point clouds are often used for real-world applications and the actions informed by the point-cloud data have the potential to adversely affect results or endanger human life, if an incorrect decision is made (e.g., in selfdriving cars). As algorithmic decisions become more consequential to individuals it becomes crucial that the algorithms are explainable in human terms. Efforts in explainable AI have focused on explaining deep learning methods [10], [11], but very little has been done to introduce explainability specifically to point-set learning on 3-D point clouds.

This research builds on the image-based xDNN framework [12] by extending it to object classification on 3-D point sets. It is specific to point cloud data and offers several layers of human-interpretable explainability. By design, the XPCC internal architecture is algorithmically transparent, simple, and, thus, easy to explain to a human user; the prototypes are the highly representative data samples and are learned incrementally after the first encounter of a specific class. The proposed method is noniterative; instead, XPCC is an incremental, greedy learning algorithm that selfdevelops autonomously. It evolves the internal structure with the addition of new prototypes that reflect the changes of the data pattern represented by the local data density. In this study, we use a fixed KP-CNN, pretrained on the ModelNet40 benchmark without limitation to the generality of the proposed concept. Experiments show that the proposed XPCC is not only explainable and computationally more efficient than the state of the art in explainable point-set deep-learning classifiers, but also superior in terms of classification accuracy. To the best of our knowledge, XPCC is the only explainable point-set classifier that achieves a higher overall accuracy compared to the benchmark deep networks used.

The main contributions of this article are summarized as follows.

1) A novel XPCC network is proposed that addresses the lack of transparent object-classification algorithms for 3-D point cloud data.
2) A new prototype-based visual representation is proposed that explores explanations within the 3-D space.
3) An evaluation of the proposed classification network to improve classification accuracy over existing methods.

The rest of this article is organized as follows. In Section II, a brief overview of relevant related work is provided. Section III details the proposed XPCC classifier and compound prototype clouds (CPC) method. Then, in Section IV, we describe the experiments conducted and analyze the results. Finally, Section V concludes this article.

II. Related Work

A. Deep Learning for Point Cloud Object Classification

Recently, deep learning classifiers were proposed that adapt to the properties of 3-D point cloud data. These classifiers use error-correction to learn an embedding for each point and aggregate this information into a global shape descriptor. Hard classification is performed by feeding the global descriptor into several fully connected layers. Deep learning classifiers on point set data can be divided broadly into three types. The first of these is multiview approaches, a technique pioneered by MVCNN [13], whereby the 3-D object is projected into multiple 2-D representations. However, it is difficult to design an efficient and robust strategy for choosing viewpoints. Recently, Wei *et al.* [14] utilized a graph convolutional network to

optimize viewpoint sampling. The second type is volumetric-based methods. These methods divide the point cloud into voxels; for example, VoxNet [15] structures point-cloud data into a volumetric occupancy grid as input to a 3-D CNN. Originally, volumetric methods were limited to point clouds with a relatively small number of points. However, octree structures have been used to reduce memory usage and increase the computational speed [16]. Nevertheless, volumetric approaches suffer from undesirable bias due to grid axis alignment and it is not clear if the advantages to processing 3-D data directly in this manner are worth the additional overhead accrued [17]. The third type of classifiers is point-based methods. These methods are capable of learning directly on the point-cloud structure without intermediate representations. This contrasts with the previous two types, where the point cloud data are converted and structured to apply mature 2-D or 3-D CNNs: a process that inherently results in information loss. Prominent network architectures for the point-based methods include graph convolution networks [18], [19], pointwise multilayer perceptron (MLP)-type NN [7], [8], and kernel point CNN (KP-CNN) [20], [21].

A main goal of the method proposed in this article was to explore explanations uniquely possible within the 3-D space. Therefore, the multiview and volumetric approaches were not used for feature extraction. Instead, a pretrained kernel point convolutional (KPConv) network, KP-CNN was used in the 3-D feature-extraction layer. Unlike grid convolution, the KPConvs define continuous convolution kernels composed of a series of kernel points with weights. Specifically, the weights for neighboring points are related to the spatial distribution with respect to the center point; formulated as an optimization problem [20].

B. Explainable Deep Learning on Point Clouds

Previous literature on explainable deep learning on point-cloud data focuses on techniques to better understand the representations learned by the network. Charles *et al.* [7], [20], [22] demonstrated how to visualize information learned by the point-cloud-based NN through projecting back a coloring based on the level of activation of the point functions onto the input point cloud. In addition, Charles *et al.* [7] used *t*-SNE to embed point-cloud global features into a 2-D space and visualize the correlation between the point clouds. Zhang *et al.* [22] modified the PointNet network to create class-attentive mappings and specified in [23] the model agnostic 3-DCAM; however, these representations are not always intuitive to nonexperts. Zhang *et al.* [24] proposed a two-stage method of local-to-global attributes for explainable point-cloud classification. Specific to kernel point-based methods, in [20] the effective receptive field is computed as the gradient of kernel point responses to measure the influence of each input point in relation to the result at a particular location. PointMask [25] introduces a differentiable layer before the encoder that learns to mask out points by maximizing mutual information between masked points and the class labels. Shen *et al.* [26] demonstrated visualizations using kernel correlation as an affinity measure between two different point sets: the neighboring points and kernel points. These explanation techniques are limited primarily to *posthoc* interpretations and, in comparison to their base architecture, the explainability negatively impacts classification accuracy.

In contrast to the abovementioned, the proposed method is explainable by design, and an overall increase in accuracy over deep-learning methods. Furthermore, previous explainable point-set classification methods do not use the 3-D medium itself, beyond color visualization, for explanations. The proposed CPC is a novel approach to using the inherent nature of 3-D to offer explication not possible in 2-D.

C. Prototype Learning

Prototype-based methods learn a set of highly representative samples (i.e., the prototypes), which themselves represent the probability distribution in the feature space [27]. One or more prototypes represent each class in the dataset, with new samples assigned to a class based on a similarity metric. Prototype-based models have long demonstrated their high efficiency and versatility in classification problems and have an obvious interpretation. These models can be manipulated through the addition, removal, or adaptation of prototypes. This makes them well suited for incremental learning. Examples of well-known prototype-based approaches are the learning vector quantization model [28], the radial basis function (g) network [29], Gaussian mixtures, and the selforganizing map (SOM) [30] model. In addition, *k*-means, support vector machine (SVM), and particle filtering (sequential Monte-Carlo methods), may be considered as kinds of prototype-based methods. The *k*-nearest neighbors (kNN) algorithm is related to many prototype-learning methods. However, because the classic technique for *k*NN stores all data rather than selective exemplars, it is only loosely considered a prototype-based method and is, strictly speaking, not a learning method. The XPCC can, like other prototype-learning methods, be viewed as a type of feedforward NN. In particular, it is similar to SOM in that the proposed method does not use error-correction learning. Instead, a type of greedy competitive learning is applied. XPCC is based on local densities and empirically-derived global multivariate generative distributions.

III. Proposed Method

In the following sections, we consider the input to the XPCC classification method as a set of $\mathcal{N}$ point cloud objects, $\mathcal{O} = \{\mathcal{O}_i \mid i = 1, \ldots, \mathcal{N}\}$, and the output is the set of predicted labels. Each object is a separate point cloud, and these, in turn, represent the shape of each object as a set of 3-D (x, y, z) coordinates. Depending on the specific acquisition technology, the points may contain additionally observed information, such as color and intensity. The proposed method classifies based only on the objects' shape and, thus, only the coordinates are taken as the input. This means that the classification will not be affected by inconsistent color values between different objects. As such, object point-cloud models are considered to contain only one entity. Functionally, point-cloud objects do not need to have the same number of points. The architecture of the XPCC is

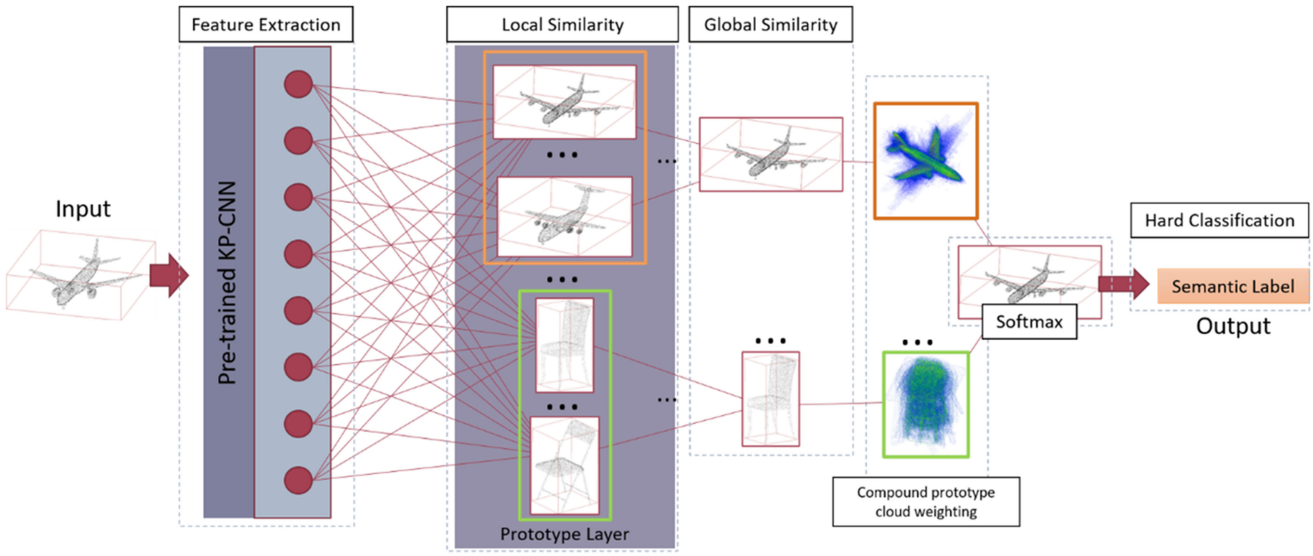


Fig. 1. XPCC classification architecture represented as network layers. First, the point-cloud object is input to the feature-extraction layer and the pretrained CNN is used to generate a global descriptor that encodes the global shape of the object into a feature vector. Second, the local-similarity layer compares the object's global descriptor against the prototypes of each class. This is called the prototype layer in this figure and is represented by the colored rectangles within the local-similarity layer. Third, the similarity score for each class is extracted by the global-similarity layer as the classes' most similar prototype. Fourth, the classes' similarity score is weighted by the object's similarity to the respective classes' compound prototype. Last, the SoftMax layer normalizes the output of the previous layer to a probability distribution over the predicted output classes. Hard classification is then performed on the output to produce the predicted class label.

structured as a series of layers within a neural network, as illustrated in Fig. 1.

A. Feature Extraction

Feature extraction encodes the global shape of each point-cloud object into a global descriptor. To do this, the layers of a pretrained CNN are used as a fixed feature extractor, and the global descriptor obtained as the feature vector computed from the final fully connected layer. The set of feature vectors define a feature space that is optimized for the separation of the training objects. From the transfer learning concept, it is assumed that this feature space can also effectively separate objects from a different domain. The features extracted by the DNN are denoted as $X = \mathrm{DNN}(\mathcal{O}) \in \mathbb{R}^{\mathcal{N} \times d_f}$, where $\mathrm{DNN}(\cdot)$ is the fixed DNN and d_f is the dimension of the extracted features. In the case of the KP-CNN, $d_f = 1024$. We use $\mathbf{x}_i$ to refer to an object represented as its feature vector extracted by the fixed DNN.

The point-based KP-CNN with rigid KPConv blocks [20] pretrained on the ModelNet40 [31] dataset is used as the feature extractor to produce a 1-by-1024-dimensional global feature vector per object. It incorporates a method to perform down-scaling or upscaling to the input point clouds, as required, so that they do not need to have the same number of points. This also means that the network remains robust to varying point densities, particularly across objects from different scenes or scanning technologies. Furthermore, it has been shown that the KP-CNN identifies simple geometric structures (lines, planes, and spherical regions) at lower layers of the network, and more complex characteristics at further layers [20]. Not only does this provide some transparency to the CNN, but is also an important indication of generalizing to object types that are not in the CNN training data.

The feature vectors are individually scaled to their unit norm (i.e., $L2$ normalization is performed on each element)

$$\mathbf{x}_i = \frac{\mathbf{x}_i}{\max(\|\mathbf{x}_i\|, \epsilon)} \tag{1}$$

where ϵ is a small constant.

B. Training

Training the XPCC starts by performing a filtering operation where the prototypes are identified from the training data. This is done directly through a noniterative "one pass" process; the prototypes themselves are the most representative training samples belonging to a particular class. Thus, metaparameters for the XPCC are trained per-class; all the calculations are performed separately for each class and can be performed simultaneously (in parallel). Each class's parameters are initialized with the first observation sample of that class

$$k \leftarrow 1, M \leftarrow 1, \mu \leftarrow x_1, p_1 \leftarrow x_1, N_1 \leftarrow 1, r_j \leftarrow r^* \tag{2}$$

where k is the current instance (number of training samples seen), M is the number of prototypes identified for a class, μ is the recursive global mean of all data samples as yet observed, p_1 is the first prototype $\{p_j \,|\, j = 1, \ldots, M\}$, N_j is the number of member points around each prototype, and r_j is the radius of the area of influence of the corresponding prototype. r^* is the initial degree of similarity of the prototype member space and is defined as

$$r^* = \sqrt{2\,(1 - cos\,(30°))} = \left\| \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|} - \frac{p_i}{\|p_i\|} \right\|. \tag{3}$$

After initialization, the parameters of each class are updated recursively, absorbing only the samples that belong to them. The following pseudocode demonstrates the process for each

new training sample of the class type. First, the instance count is updated through

$$k \leftarrow k + 1 \tag{4}$$

then μ is updated as follows:

$$\mu \leftarrow \frac{\mu\,(k-1) + \mathbf{x}_i}{k}. \tag{5}$$

Data samples (i.e., the objects) that are closer to the global mean have higher density values. Therefore, the data density indicates how strongly data samples influence one another in the data space. The density function is defined as a Cauchy function [32]

$$D\,(x_i) = \frac{1}{\frac{1+\|\mathbf{x}_i-\mu\|^2}{\sigma}} \tag{6}$$

where μ is the global mean and $\sigma = 1 - \|\mu\|^2$.

The prototypes are determined by partitioning the labeled training-data based on the data density and area of influence within the latent feature space. The prototypes are the local peaks of the data density in the feature space for their corresponding class. It is important to note that the prototypes are independent from each other, such that the addition of a new prototype does not influence the already existing prototypes.

$$\begin{aligned} &\alpha = \max D(p_j) \\ &\beta = \min D(p_j) \\ &j^* = \arg\min(\mathbf{x}_\mathrm{i} - \mathrm{p_j}) \end{aligned} \tag{7}$$

$$\begin{aligned} &\textbf{IF } \mathrm{D}(\mathbf{x}_\mathrm{i}) > \alpha \textbf{ OR } \mathrm{D}(\mathbf{x}_\mathrm{i}) < \beta \textbf{ OR } (\|\mathbf{x}_\mathrm{i} - \mathrm{p_{j^*}}\| \leq \mathrm{r_{j^*}}) \\ &\textbf{THEN } \text{add a new prototype} \end{aligned} \tag{8}$$

The prototypes with maximum and minimum density, α and β, respectively, and the index j^*, which denotes the prototype closest to the current sample, are used to control the addition of new prototypes [32]. If either of the first two conditions in (8) is met, or if the sample lies outside the area of influence of the closest prototype [the third condition in (8)], then the new data sample is added as a new prototype to its respective class. If the conditions are not met, then new samples are assigned as a support member of the prototype nearest in the feature space.

The prototype is then updated recursively as follows [32]

$$p_{j^*} \leftarrow \frac{p_{j^*}\,(N_j - 1) + \mathbf{x}_i}{N_j}. \tag{9}$$

The support, or the number of data samples associated with a certain prototype, is updated $S \leftarrow S + 1$ and the radius is updated recursively using the following [32]:

$$r_{j^*} \leftarrow \frac{r_{j^*} + (1 - p_{j^*})}{2}. \tag{10}$$

After the initial training process, the model can learn continuously by absorbing new training samples of previously seen classes or of new unseen classes. Alternatively, a user can manipulate and fine-tune the model through the addition, removal, or adaption of prototypes manually.

Explanation of prototypes can be represented in the form of linguistic logical IF...THEN rules where the density *D* can be seen as a fuzzy degree of membership [12]. All rules, per class, can be combined using the logical OR operator (see Fig. 2).

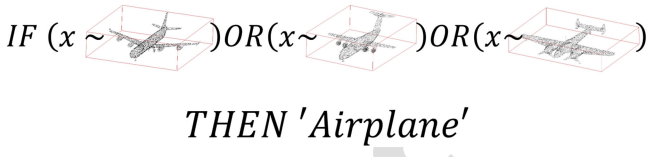


Fig. 2. Visual illustration of linguistic IF...THEN rules, where $\sim$ stands for similarity and x is the queried sample. If the sample is within the degree of similarity for the set of prototypes belonging to a class, then the class label is applied.

C. Compound Prototype Cloud

After training, the XPCC model contains several class containers, one per class (n.b. that the containers are referred to as "data clouds" in [12]. However, we call them class containers to avoid confusion with 'point clouds'). Each container consists of the per-class metaparameters, including the prototypes identified for that class. The number of prototypes is much less than the number of training samples of that type that were seen (i.e., $P \ll N$). The prototypes themselves represent the 3-D point-cloud object. The CPCs are composed of all point cloud-based prototypes superimposed and, as a result, one such aggregated prototype per class. The CPC creation process is as follows. Starting with the first identified prototype as the reference object, perform principal component analysis to obtain the three main directions of the point cloud and deduce the main axis. Second, take the extent along the main axis to scale the clouds to the reference (reference objects are chosen as the initial prototype). Then, perform a fine registration using a point-to-point iterative closest point algorithm [33].

D. Classification Algorithm

In this section, we describe the classification procedure for the XPCC method. The principle of the XPCC-classification approach is based on the intuition that people learn by comparing similarity between objects, but only remember a few distinct objects during decision making (i.e., the prototypes)—the so-called anthropomorphic approach to machine learning [34]. If a new object is encountered, a person is likely to assume that it belongs to the class which it most closely resembles. Following this logic, the learning of the proposed method revolves around the position and properties of the prototypes in the feature space.

Given a new test sample $\mathbf{x}_t$, the proposed XPCC method first finds the local similarity to each class's prototypes

$$S_j = \mathrm{Similarity}(\mathbf{x}_t, p_j) = \frac{1}{(1 + \|\mathbf{x}_t - p_j\|^2)} \tag{11}$$

and then determines the global similarity to each class as $\mathrm{S}_j^* = \max \mathrm{S}_j$. With these, the proposed method then performs a global weighting (13), where $\odot$ is the Hadamard product, based on similarity between the new sample and the CPC feature vector. The CPC feature vector is estimated as the recursively updated

global mean of the class, that is

$$S_{\mathrm{CPC}} = \mathrm{Similarity}(\mathbf{x}_i, \mathrm{CPC}_j) \tag{12}$$

$$\gamma_i = \mathrm{S}_{\mathrm{CPC}} \odot \mathrm{S}_j^*. \tag{13}$$

In the proposed method, the following SoftMax function is used to normalize the output of the previous layer to a probability distribution over the predicted output classes:

$$\gamma_i^* = \frac{\exp(\gamma_i)}{\sum_k \exp(\gamma_k)}. \tag{14}$$

Finally, the hard classification is conducted through the $\arg\max(\gamma_{\mathrm{i}}^*)$ function, thus providing the label of the most likely class.

IV. Experiments

We compared the proposed XPCC classifier against both classical machine-learning classifiers and state-of-the-art point-set learning classifiers. Unless and otherwise stated, the term XPCC refers specifically to the use of the proposed method with the KP-CNN feature-extractor network. In the experiment comparing the proposed method against classical machine-learning algorithms, the

$$\mathrm{Precision} = \mathrm{TP}/\left(\mathrm{TP} + \mathrm{FP}\right)$$

$$\mathrm{Recall} = \mathrm{TP}/\left(\mathrm{TP} + \mathrm{FN}\right)$$

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

were used as evaluation metrics, where TP, FP, and FN are the number of true positives, false positives, and number of false negatives, respectively. For all experiments we also give the overall accuracy

$$OA = \frac{TN + TP}{TP + FP + TN + FN}$$

where TN is true negative. For our experiments comparing the proposed method against point-set deep-learning algorithms, we also use mean accuracy ($m\mathrm{Acc}$),

$$m\mathrm{Acc} = \frac{1}{M}\sum_{i=1}^{M} \mathrm{acc_i}$$

where acc_i is the accuracy for the samples class i and M is the total number of samples.

The Princeton ModelNet project provides a collection of synthetic 3-D CAD-object models split into two benchmarks: a 40-class subset and 10-class subset known as ModelNet40 and ModelNet10, respectively. We believe ModelNet is the only publicly available 3-D object benchmark specific to object classification. Other 3-D model and point-cloud datasets exist, but they are either a collection of objects without a defined test/train split [35], have very few models per class [36], or are intended for different tasks like semantic segmentation [37]–[39], or 3-D object detection [40]. The ScanObjectNN [41] benchmark is an interesting dataset made up of real-world object models; but at the time of writing, it is not available for public access. It will be a topic of future consideration. To evaluate the XPCC we used the ModelNet40 shape-classification benchmark. There are 12,311 CAD models from 40 categories of human-made objects, split into 9843 for the training set and 2468 for the testing set. The CAD models were converted into 3-D point clouds by sampling points randomly along the model surfaces.

Experiments with the XPCC were run on a Unix machine with 3.6 GHz AMD Ryzen 5 3600 6-Core CPU, 16 GB RAM, and SSD. The KP-CNN feature-extractor network was trained using an NVIDIA 2070S GPU. Training the KP-CNN took five hours. Once trained, the fixed network acts as a generalized point-cloud feature extractor. In our comparison studies, we also evaluated the PointNet++ network as a feature extractor; this was trained under the same conditions. The points are restricted to contain only the (x, y, z) coordinate information. Training of and inference with the XPCC is highly parallelizable and can be conducted on either CPU or GPU hardware.

A. CPC-Demonstration

The motivation for the CPC is to represent what the model has learned by visualizing the per-class parameters as a system of superimposed prototypes in 3-D space. This is comparable to the idea of an object that a human might imagine when thinking of a particular object class. For example, if asked to think of an "airplane", a cylindrical capsule shape with wings and tail rudder is likely to come to mind. Although this is the general shape, there are other aspects that might be different depending on the person's experiences with the object. For the airplane example, attributes, such as the wing positions, angle, or number, might vary and additionally the plane might have propellers or jet engines. Much like the human idea of an object, the CPC contains these aspects as physical options. We illustrate the CPC representation in Fig. 3. By encoding the object's similarity (the normalized data density) in the feature space as a color, the CPC indicates the areas that contribute to the classifier's decisions when examining new objects. As can be seen in Fig. 3, green and red regions are areas with greater density within the model's decision space and correspond to the areas of an object that contribute to the model's knowledge of that class and, ultimately, the decision-making process.

B. Comparative Results With Classical Methods

In this section, we compared the proposed XPCC method against classical machine-learning approaches to classification including *k*NN, SVM, decision trees, random forest, MLP, and Gaussian Naive Bayes. All methods were implemented using the scikit-learn Python library [42]; however, we cite the underlying algorithm or library if applicable. The *k*NN method is formulated as a *brute force* problem. Five variations of SVM classifier are reported. L-SVM is underpinned by [43] and utilizes squared-hinge loss to train one-versus-rest classifiers. Conversely, C-SVM is underpinned by [44] and

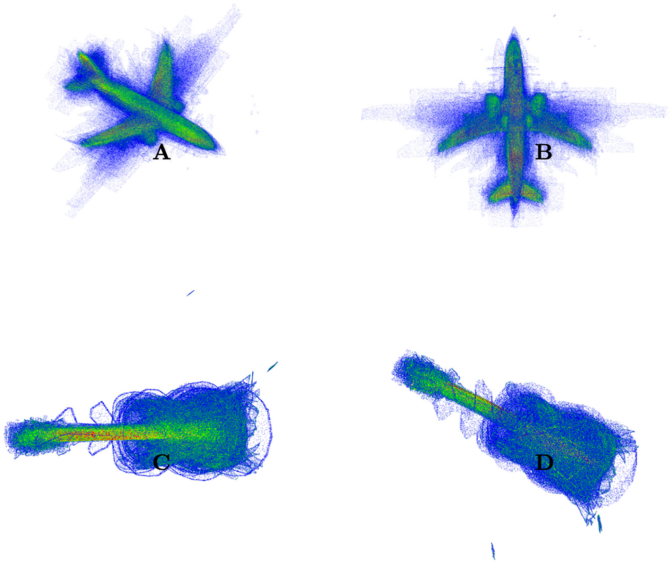


Fig. 3. Visual depiction of the compound prototype clouds for selected ModelNet40 classes. A and B are the CPC for class airplane. C and D are the CPC for the class guitar. The color represents the distribution of confidence per data point extracted from the data alone. Blue areas are those with low confidence, green areas are those with medium confidence, and red areas are those with high confidence.

TABLE I
PERFORMANCE COMPARISON WITH CLASSICAL CLASSIFIERS ON THE MODELNET40 BENCHMARK

Method	Metric			
	Accuracy	Precision	Recall	F1
*k*NN (brute-force) [42]	90.04	88.23	87.64	87.78
L-SVM (hinge-loss) [43]	90.36	87.89	87.27	87.25
C-SVM (linear-kernel) [44]	91.17	88.16	88.51	88.67
C-SVM (poly-kernel) [44]	91.05	88.97	88.03	88.35
C-SVM (rbf-kernel) [44]	91.00	89.04	88.39	88.53
C-SVM (sigmoid-kernel) [44]	90.88	87.63	88.49	87.88
Decision Tree (gini) [42]	81.97	76.89	77.77	77.00
Decision Tree (entropy) [42]	78.72	73.67	73.15	72.75
Random Forest (gini) [42]	90.43	87.79	87.71	87.61
Random Forest (entropy) [42]	90.55	88.69	88.09	88.25
MLP (relu) [42]	90.15	87.38	87.05	86.93
MLP (tanh) [42]	90.07	87.86	87.81	87.69
MLP (sigmoid) [42]	90.76	87.62	88.32	87.79
Gaussian Naive Bayes [45]	87.76	84.78	84.91	84.44
XPCC (ours)	**91.82**	**92.10**	**91.82**	**91.83**

Note: All classical methods were implemented with the scikit-learn [42] Python library. The classical explainable classifiers are trained with the same feature vectors as XPCC, generated by the fixed-CNN feature extractor. Normalization was performed as described in (1) for all experiments.

trains one-versus-one classifiers; results are reported using the linear, polynomial (poly), RBF, and sigmoid kernels. Two variations of the decision-tree and random-forest algorithms are recorded, the first using the gini-impurity and the second using entropy-impurity. In addition, results are compared against three variations of MLP. Each consist of the standard three-layer configuration and use the ReLU, TanH, sigmoid activation functions, respectively. Optimization of the MLP is performed using [46]. Lastly, we compare against a Gaussian Naive Bayes classifier. Further specifications of the comparative classifier implementations can be found in [42].

From Table I, it is clear that the XPCC method achieved higher scores across all metrics in comparison to these classical approaches. The proposed method increased the overall accuracy by 0.65 percentage points (p.p.), increased F1-score by 3.16 p.p., and increased recall by 3.30 p.p., compared to the subsequently leading C-SVM with linear-kernel method. Similarly, precision was increased by 3.06 p.p. over the subsequently leading C-SVM with RBF kernel. The proposed method shares the ability to perform online machine learning with the Gaussian Naive Bayes method. A comparison of these two methods shows that the XPCC method increases accuracy by 4.06 p.p., precision by 7.32 p.p., recall by 6.91 p.p., and F1-score by 7.39 p.p. Furthermore, the XPCC method does not incur performance or stability overhead when performing online learning because it is built on recursive calculations.

C. Comparative Results With State of the Art

We performed a classification test on the ModelNet40 benchmark and compared the XPCC classifier against the state of the art in explainable point-set learning algorithms, including PointHop, CLAIM, and PointMask. In addition, we compared the proposed method against baseline deep-learning algorithms including PointNet, PointNet++, and the base KP-CNN. The results from this experiment can be found in Table II.

Prior XPCCs use either PointNet or PointNet++ as their base method. These have an overall accuracy score of 89.2 p.p. and 90.7 p.p., respectively, (as presented on the ModelNet40 benchmark rankings). However, compared to the base methods the mechanisms by which the explainable methods provide explanation result in a decrease in overall accuracy: PointHop incurs a 0.6 p.p. decrease from PointNet and 2 p.p. decrease from PointNet++, CLAIM sustains a 2.1 p.p. decrease from PointNet and 3.6 p.p. decrease from PointNet++, and PointMask experiences a 7 p.p. decrease from PointNet and 8.5 p.p. decrease from PointNet++. With our implementation of the KP-CNN, the base network alone achieved baseline accuracies of 91.80% (overall accuracy) and 88.75% (mean accuracy). In comparison with this baseline the XPCC produced on average a 0.02% increase to overall accuracy and a 0.12 p.p. increase to mean accuracy. Therefore, we believe XPCC to be the only explainable point-set classifier that leads to an increase in accuracy relative to the base algorithm. In comparison with the previous explainable approaches, the proposed method increased classification accuracy by 2.7 p.p. (versus PointHop), 4.6 p.p. (versus CLAIM), and 9.64 p.p. (versus PointMask).

As a further experiment for accuracy, we reversed the test/train split (i.e., train on 2468 samples and test on 9843). The XPCC achieved an overall accuracy of 96.97% and mean accuracy of 95.46%. This would be far too few training samples to produce accurate results with the deep learning point-set learning methods. This demonstrates that the proposed method can achieve high accuracy with significantly less training data, although we do note that this particular experiment negates any difficulties purposefully encoded within the test set of the benchmark.

TABLE II
PERFORMANCE COMPARISON AGAINST STATE OF THE ART AND EXPLAINABLE POINT-SET DEEP LEARNING METHODS ON THE MODELNET40 BENCHMARK

Method	OA	mAcc	Training Time	Device	# of parameters	Transparent	Reproducibility	Retraining
PointNet[7]	89.2	86.2	5+ h	GPU	3.5 M	No	No	Yes
PointNet++[8]	90.7	–	5+ h	GPU	1.5 M	No	No	Yes
PointHop[24]	88.65	83.3	∼20 m	CPU	–	Yes	No	Yes
CLAIM[47]	87.1	–	–	GPU	–	Yes	No	Yes
PointMask[25]	82.18	–	–	–	–	Yes	–	–
KP-CNN (rigid KPConv)[20]	91.80*	88.75	5+ h	GPU	14.3 M	No	No	Yes
xDNN[12]	89.42	86.37	∼7s	CPU	$\mathbf{P} \times 2$	Yes	Yes	No
XPCC & KPConv (ours)	91.82	**88.87**	**∼2s (GPU) ∼6s (CPU)**	CPU/GPU	$\mathbf{P} \times 2$	Yes	Yes	No
XPCC & PointNet++ (ours)	**92.18**	88.43	**∼2s (GPU) ∼4s (CPU)**	CPU/GPU	$\mathbf{P} \times 2$	Yes	Yes	No

Note: The highest result for each accuracy metric listed is in bold, where OA is the overall accuracy and mAcc is the mean accuracy. Training time is the approximate time needed to train a method. The device column reports the device typically required by the algorithm. In the number of parameters column, **P** is the number of prototypes; M is million. For both XPCC and xDNN there are two parameters (μ and σ) per prototype. The transparent metric is conditional on the network's ability to be interpreted or if it is a "black box" method. The reproducibility metric is conditional on, if given the same training data, the network will always conclude the same predictions. The retraining is conditional on if the network must be completely retrained in order to add a new training sample. Metrics that are not reported by or cannot be understood from the source literature are marked using the "–" notation.
* The authors of KPConv record an accuracy score of 92.9 in their publication, however, we were unable to reproduce this score; this is not unique to our experience [41]. Our retrained KP-CNN achieves an average accuracy score of 91.8.

D. Analysis

Improvements to benchmark results are often marginal and relying on these scores alone can sometimes ignore other aspects of an algorithm, which may be equally beneficial. As such, we stress that accuracy is only one aspect of explainable classifiers. Specifically, it has been shown that scores over 91% on the ModelNet40 benchmark are sensitive to optimizations that do not necessarily transfer to real world experiments [17], [41]. To highlight this effect, we also pretrained a PointNet++ network as feature extractor using the optimized data augmentation described in [17]. In terms of overall accuracy, this formulation of the XPCC does perform better than the KPConv feature extractor on the benchmark's defined test/train split. However, the XPCC & PointNet++ configuration was less accurate under the mean accuracy metric, which takes class balance into consideration. As shown in the "Domain Transfer" section, the features extracted by the KP-CONN are demonstrated to be discriminative when applied generally. For this reason, we opted for the more sophisticated KP-CNN as feature extractor.

We list the number of parameters for all methods (if available) in Table II. The deep learning methods that rely on error-correction learning (e.g., backpropagation with gradient descent) have millions of parameters, a characteristic that extends to their derivative explainable methods. In comparison, the xDNN and XPCC methods have two parameters for each prototype, where the number of prototypes for our proposed method is around 10% of the number of training samples seen for a class.

The XPCC is significantly faster than all other methods. Training took on average six seconds on a CPU and two seconds on a GPU. This is thanks to the highly parallelizable structure of the model: all the calculations are performed separately for each class and, thus, can run simultaneously. In relation to the other methods this translates into the XPCC being at least 9000 times faster to train than methods that took five hours, and 600 times faster than methods that took 20 minutes to train. This, compounded with the fact that the XPCC does not need to be completely retrained to add new classification types (i.e., it learns continuously), demonstrates the efficiency and practical applicability.

E. Domain Transfer

The XPCC method incorporates two varieties of domain transfer: task-wise domain transfer and learning domain transfer.

Task-wise domain transfer is intrinsic to the method. The prototype-based internal structure of XPCC allows for a trained model to be transferred to a new domain without a complete retrain. Specifically, adding a new class requires training the model on only the new data samples, rather than a complete retrain. For example, if an XPCC model has been trained to classify chairs and tables, the model only needs to be updated with training samples of a new class, such as televisions. Similarly, the model does not need to forget previously learned classes when transferred.

To investigate the effectiveness of the learning domain transfer and the approach to transfer learning, we assessed the transferability of the fixed-CNN feature extractor in the XPCC method. This experiment was conducted using the ModelNet10 benchmark, by examining accuracy results when applied on classes of objects not seen by the feature-extraction network. The fixed-CNN feature-extractor network was trained without access to one of the object classes. Then, the XPCC model was trained on the full dataset (including that object class) and validated using the ModelNet10 train-test data split. In this way, only the XPCC model had any knowledge of the hidden class and, therefore, not reliant on the fixed-CNN having knowledge of that class of object. This process was systematically repeated for each class within the ModelNet10 benchmark, and the average accuracy score recorded. In this experiment, the XPCC and KP-CNN feature extractors achieved an overall accuracy score of 91.38% for the hidden classes. We performed the same experiment with the XPCC & PointNet++ feature extractor. This configuration achieved only an overall accuracy score of 89.96% for the hidden class, giving a 1.42 p.p. decrease in accuracy; these results suggest that the PointNet++ feature extraction method is not as discriminative when applied generally. For this reason,

we selected the more sophisticated KP-CNN feature extractor, which makes the XPCC method more viable for training and classifying atypical 3-D point cloud object model classes.

F. Error Analysis

From examining the similarity scores between negatively predicted samples and prototypes, it is clear there were some limitations to the proposed method. A primary bottleneck of the method is discerning between objects that share many similar characteristics. As such, errors in the classification are predominately from objects with subjective classification labels, such as the distinction between a glass and vase, or table and desk. Other errors were from samples whose shapes are similar, such as bench and sofa. This appears as a trend across explainable 3-D methods on the ModelNet benchmarks and suggests that more contextual information is needed beyond geometric shape. In addition, semantic limitations present within the ModelNet40 benchmark are arguably hard for a human to discern between. For example, shapes that have different semantic labels but are geometrically very similar, such as the difference between a "flower pot" (with a plant in it) and a "plant" (in a pot).

V. Conclusion

We proposed a new classification method, the XPCC, for object classification within 3-D point clouds. The proposed method was algorithmically and structurally transparent, learns continuously, without the need to be completely retrained at the addition of new classes, and offers several layers of human-interpretable explainability. This article also presented a novel technique to visualize the explainable aspects of the model, called CPC. The technique was unique to 3-D point clouds and prototype-based learning and represented what the model has learned. Specifically, it identified object regions, which contribute to the classification. Experiments showed that the proposed classifier method was computationally efficient, trainable on thousands of samples in seconds, and is competitive with the state of the art in point-set deep-learning classifiers in terms of classification accuracy. Furthermore, the proposed method was the only explainable point set classifier that achieved higher accuracy compared to the base network used. A limitation of the proposed method was that classification relies on point clouds containing only one object. In our future work, we will focus on applying the method to real-world data and extending the method to other point-cloud specific objectives, such as object detection within a scene.

References

[1] A. Barredo Arrieta *et al.*, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020.

[2] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2009, pp. 3212–3217.

[3] R. B. Rusu, G. Bradski, R. Thibaux, and J. Hsu, "Fast 3D recognition and pose using the viewpoint feature histogram," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2010, pp. 2155–2162.

[4] A. E. Johnson and M. Hebert, "Surface matching for object recognition in complex three-dimensional scenes," *Image Vis. Comput.*, vol. 16, no. 9, pp. 635–651, Jul. 1998.

[5] S. Bozinovski, "Reminder of the first paper on transfer learning in neural networks, 1976," *Informatica*, vol. 44, no. 3, pp. 291–302, Sep. 2020.

[6] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. R. Salakhutdinov, and A. J. Smola, "Deep sets," *Adv. Neural Inf. Process. Syst.*, vol. 30, no. 11, pp. 3394–3404, 2017.

[7] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 77–85.

[8] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet : Deep hierarchical feature learning on point sets in a metric space," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5105–5114.

[9] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Mach. Intell.*, vol. 1, no. 5, pp. 206–215, May 2019.

[10] E. A. Soares, P. P. Angelov, B. Costa, M. Castro, S. Nageshrao, and D. Filev, "Explaining deep learning models through rule-based approximation and visualization," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 8, pp. 2399–2407, Aug. 2021.

[11] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K. Müller, "Evaluating the visualization of what a deep neural network has learned," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 11, pp. 2660–2673, Nov. 2017.

[12] P. Angelov and E. Soares, "Towards explainable deep neural networks (xDNN)," *Neural Netw.*, vol. 130, pp. 185–194, Oct. 2020.

[13] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3D shape recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 945–953.

[14] X. Wei, R. Yu, and J. Sun, "View-GCN: View-based graph convolutional network for 3D shape analysis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1847–1856.

[15] D. Maturana and S. Scherer, "VoxNet: A 3D convolutional neural network for real-time object recognition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2015, pp. 922–928.

[16] P.-S. Wang, Y. Liu, Y.-X. Guo, C.-Y. Sun, and X. Tong, "O-CNN: Octree-based convolutional neural networks for 3D shape analysis," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–11, Jul. 2017.

[17] A. Goyal, H. Law, B. Liu, A. Newell, and J. Deng, "Revisiting point cloud shape classification with a simple and effective baseline," in *Proc. Mach. Learn. Res.*, 2021, pp. 3809–3820.

[18] Q. Xu, X. Sun, C.-Y. Wu, P. Wang, and U. Neumann, "Grid-GCN for fast and scalable point cloud learning," *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 5661–5670.

[19] L. Landrieu and M. Simonovsky, "Large-scale point cloud semantic segmentation with superpoint graphs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4558–4567.

[20] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. Guibas, "KPConv: Flexible and deformable convolution for point clouds," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6410–6419.

[21] A. Boulch, "ConvPoint: Continuous convolutions for point cloud processing," *Comput. Graph.*, vol. 88, pp. 24–34, May 2020.

[22] B. Zhang, S. Huang, W. Shen, and Z. Wei, "Explaining the PointNet: What has been learned inside the pointNet?," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 71–74.

[23] Y. Cao, M. Previtali, and M. Scaioni, "Understanding 3D point cloud deep neural networks by visualization techniques," *ISPRS - Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. XLIII-B 2-2020, pp. 651–657, Aug. 2020.

[24] M. Zhang, H. You, P. Kadam, S. Liu, and C.-C. J. Kuo, "PointHop: An explainable machine learning method for point cloud classification," *IEEE Trans. Multimedia*, vol. 22, no. 7, pp. 1744–1755, Jul. 2020.

[25] S. A. Taghanaki, K. Hassani, P. K. Jayaraman, A. H. Khasahmadi, and T. Custis, "PointMask: Towards interpretable and bias-resilient point cloud processing," in *Proc. Workshop Human Interpretability Mach. Learn.*, 2020, pp. 246–252.

[26] Y. Shen, C. Feng, Y. Yang, and D. Tian, "Mining point cloud local structures by kernel correlation and graph pooling," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4548–4557.

[27] J. Bien and R. Tibshirani, "Prototype selection for interpretable classification," *Ann. Appl. Statist.*, vol. 5, pp. 2403–2424, 2011.

[28] T. Kohonen, J. Kangas, J. Laaksonen, and K. Torkkola, "LVQPAK: A software package for the correct application of learning vector quantization algorithms," in *Proc. IJCNN Int. Joint Conf. Neural Netw.*, 1992, pp. 725–730.

[29] M. J. Er, S. Wu, J. Lu, and H. L. Toh, "Face recognition with radial basis function (RBF) neural networks," *IEEE Trans. Neural Netw.*, vol. 13, no. 3, pp. 697–710, May 2002.
[30] T. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, Sep. 1990.
[31] Z. Wu *et al.*, "3D shapeNets: A deep representation for volumetric shapes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1912–1920.
[32] P. P. Angelov and X. Gu, *Empirical Approach to Machine Learning* (Studies in Computational Intelligence Series). Berlin, Germany: Springer, 2019.
[33] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.
[34] P. P. Angelov and X. Gu, "Toward anthropomorphic machine learning," *Computer*, vol. 51, no. 9, pp. 18–27, 2018.
[35] A. X. Chang *et al.*, "ShapeNet: An information-rich 3D model repository," Dec. 2015, *arXiv:1512.03012*. [Online]. Available: https://shapenet.org/publications
[36] M. Deuge, A. Quadros, C. Hung, and B. Douillard, "Unsupervised feature learning for classification of outdoor 3D scans," *Australas. Conf. Robot. Automat.*, 2013, pp. 350–358.
[37] H. Caesar *et al.*, "nuScenes: A multimodal dataset for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11618–11628.
[38] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, "ScanNet: Richly-annotated 3D reconstructions of indoor scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 2432–2443.
[39] T. Hackel, N. Savinov, L. Ladicky, J. D. Wegner, K. Schindler, and M. Pollefeys, "Semantic3d.net: A new large-scale point cloud classification benchmark," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. IV-1/W1, pp. 91–98, May 2017.
[40] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 3354–3361.
[41] M. A. Uy, Q.-H. Pham, B.-S. Hua, D. T. Nguyen, and S.-K. Yeung, "Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1588–1597.
[42] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
[43] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, p. 1871–1874, Jun. 2008.
[44] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, May 2011.
[45] H. Zhang, "The optimality of Naive Bayes," in *Proc. FLAIRS Conf.*, 2004, pp. 562–567.
[46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017, *arXiv:abs/1412.6980*. [Online]. Available: https://dblp.org/db/conf/iclr/iclr2015.html
[47] S. Huang, B. Zhang, W. Shen, and Z. Wei, "A CLAIM approach to understanding the PointNet," in *Proc. 2nd Int. Conf. Algorithms, Comput. Artif. Intell.*, 2019, pp. 97–103.

Chapter 4

An Explainable and Transferable AI Framework for the Extraction of Burial Ground Objects: GeoPart-XPCC

This chapter introduces the GeoPart-XPCC framework, a two-part process comprising segmentation and classification. The segmentation phase employs a robust machine learning algorithm tailored for large-scale, fine-resolution point clouds, addressing variable point density. This is followed by the application of an explainable neural network, facilitated by transfer learning, for efficient and accurate classification. The experimentation phase is substantiated through the analysis of 11 scenes from five diverse burial ground sites in the United Kingdom. Results showcase GeoPart-XPCC's efficiency, adaptability, and remarkable accuracy.

The GeoPart-XPCC framework stands as a significant advancement within the broader context of the thesis, building upon the foundation laid by two pivotal predecessors: GeoPart-Transfer and XPCC. Each of these components has contributed distinct dimensions to the overarching exploration of automated object extraction and classification from cultural heritage point clouds.

The GeoPart-Transfer, discussed in previous sections of the thesis, set the stage by introducing a method for the autonomous extraction and labelling of objects from cultural heritage scenes. This approach, founded on segmentation and classification tasks, aimed to transcend the limitations of existing methodologies by utilising an MLP (Multi-Layer Perceptron) classifier. This innovation marked a departure from earlier techniques that relied on spatial descriptors or predefined patterns, offering a more adaptable approach. However, GeoPart-Transfer retained certain limitations, including its choice of classifier and potential improvements in adaptability.

In parallel, the XPCC (eXplainable Point Cloud Classifier) component emerged as an innovative solution to classification challenges in 3D point cloud data. This neural network-based classifier, distinguished by its explainable nature, exhibited remarkable proficiency in classification accuracy while offering insights into decision-making processes. The XPCC thus aligned closely with the overarching thesis goal of combining the power of deep learning with human-understandable explanations.

In this context, the GeoPart-XPCC framework effectively bridges the contributions of GeoPart-Transfer and XPCC. GeoPart-XPCC, by replacing the MLP classifier of GeoPart-Transfer with the XPCC classifier, amalgamates the strengths of both predecessors. This union results in a comprehensive framework that not only autonomously extracts and labels memorial objects from cultural heritage scenes but does so with an explainable, interpretable classification methodology. This distinctive blend signifies an important step forward in cultural heritage preservation and management, as it combines accuracy and efficiency with transparency in decision-making.

Furthermore, the introduction of GeoPart-XPCC does not only serve as an upgrade to GeoPart-Transfer but also signifies a strategic incorporation of XPCC's explainability and adaptability into the methodology. The adaptive capabilities of XPCC complement the adaptable nature that GeoPart-Transfer was already designed to possess. This integration underscores the framework's ability to excel across diverse cultural heritage sites, accommodating variations in object styles, materials, and surroundings.

In essence, the GeoPart-XPCC framework emerges as a synthesis of two critical facets of the thesis: the adaptive and object extraction methodology from GeoPart-Transfer and the explainable, accurate classification prowess of XPCC. This amalgamation propels the thesis's overarching mission of revolutionising cultural heritage preservation through the fusion of cutting-edge technology and human-understandable decision-making processes.

GeoPart–XPCC: An Explainable and Transferable AI Framework for the Extraction of Burial Ground Objects

Nicholas I. Arnold, Plamen P. Angelov, *Fellow, IEEE*, and Peter M. Atkinson

***Abstract*—Recent advances in LiDAR technology have made the creation of fine-resolution 3D point cloud scenes of cultural heritage and archaeology sites possible. However, the challenge of how to automatically extract and label objects of interest from these point clouds remains. This paper introduces a novel framework for the autonomous extraction and labelling of memorial objects from cultural heritage scenes; called GeoPart-XPCC. In this framework, extraction is performed in two parts. First, segmentation is performed using a robust machine learning algorithm that is suitable for large-scale fine-resolution point clouds with varying point density. Second, an explainable neural network supported by a transfer learning process is used to perform classification quickly and accurately. Experiments were performed on 11 scenes from five different burial ground sites located in the United Kingdom. These experiments demonstrate that the GeoPart-XPCC framework is quick to train and apply, easily adaptable and accurate. It is shown that the framework could be trained on part of a scene and perform prediction on the remainder, as well as be trained on one scene and predict on a different scene. Previous studies utilise point clouds derived from airborne LiDAR, whereas this framework is suitable for the fine-resolution and variable density point clouds derived from mobile LiDAR. Compared with other methods, the experiments show that the GeoPart-XPCC is more accurate and robust to changes in domain. The proposed framework is, therefore, widely applicable.**

***Index Terms*—XAI, ML, Classification, Segmentation, 3D, Point Cloud, Cultural Heritage, Transfer Learning**

I. Introduction

BURIAL grounds in the United Kingdom, such as cemeteries and graveyards, are cultural heritage sites and provide an interesting application case for automated approaches to extract and classify objects from 3D point cloud scenes. Recent advances in point cloud sensor technologies, for example, light detection and ranging (LiDAR) and photogrammetry, have led to the development of versatile mobile LiDAR sensors capable of creating 3D point clouds of heritage sites. Point clouds representing real world environments provide fine-resolution digital proxies for the multitude of landscape objects covered within each scene, and offer a valuable data source for cultural heritage management [1]–[4]. However, the extraction and classification of the objects of interest is a necessary step for the operational use of these data, for example, for curation, interpretation and assessment purposes. In this paper, the landscape objects of interest are memorial objects. Performing manual extraction and labelling of memorial objects is labour-intensive and automation of these tasks is, therefore, desirable

Several characteristics of burial grounds make the design of a process suitable for automated extraction and labelling not immediately obvious. The burial grounds' age, location, and cultural setting affect the physical environment of sites and the appearance of objects contained within them. Memorial objects might appear as having different styles and designs and may be made from different materials depending on these factors. For example, many burial grounds, especially those in the UK, can date back hundreds of years and are often planted with gardens, trees and other forms of vegetation. Over time, the condition of memorial objects may degrade, and the vegetation can become more prominent. Furthermore, many of these older sites are still in use today. The effects of this are that a burial ground might contain examples of both traditional memorials and more modern ones, such as headstones and memorial benches, respectively. Additionally, burial grounds located in different geographical regions might have unique traits or variation in the typical memorial objects found locally. Inversely, sites from the same region might exhibit differences based on cultural aspects such as religion and economics. In the context of developing an automated solution for the extraction and classification of memorial objects from these sites, the above characteristics pose significant challenges.

Conventional methods of object extraction from cultural heritage point cloud data, such as object-based image analysis [5] and template matching [6], rely on purpose-built spatial descriptors or recognition patterns, both of which must be pre-defined. These approaches are, therefore, only semi-autonomous, and difficult to generalise across applications [7]. To address these limitations, a variety of methods based on machine learning were developed for object extraction and the mapping of archaeological sites [8]–[11], and cultural heritage sites by extension [12]–[15]. The most prolific approach amongst these methods is the application of deep neural networks (DNN), specifically, convolutional neural networks (CNN) [16]–[19]. DNNs are supervised computational models that are composed of several processing layers that learn increasingly complex latent features [20]. Most archaeological and cultural heritage studies do not use 3D point clouds directly and, instead, derive visualisations that are more immediately interpretable to the human eye [21]. These representations, such as 2D digital elevation maps, are then used as the input to the DNN models. However, by not using point clouds directly, these approaches fail to take advantage of

the innate spatial information present in true 3-dimensional data and, potentially, introduce viewpoint bias.

Recently, point-based deep learning methods, such as PointNet [22] and PointNet++ [23], were introduced that are adapted to the properties of 3D point cloud data [24]. These neural networks are capable of learning directly on the point cloud structure without intermediate representations. As classifiers, these models use error-correction to learn an embedding for each point and aggregate this information into a global shape descriptor. Hard classification is performed by feeding the global descriptor into several fully connected layers. Prominent network architectures for point-based deep learning methods include graph convolution networks [25], [26], pointwise multi-layer perceptron (MLP) neural networks [22], [23] and kernel point CNNs [27], [28]. A point-based method is used in this paper for the extraction of latent features directly from 3D point clouds, similar to [12].

The number of data required for training traditional deep learning models is variable. In most cases, training these models requires many thousands of data per classification [29]. In applications where extensive datasets of annotated training data do not exist, as is the case for memorial objects. This need for training data, thus, can create an obstacle to implementation. The common approach to overcoming this obstacle is data synthesis or augmentation [30]. However, the design and execution of synthesising training data is itself time-consuming and often requires expert knowledge of both the application domain and the model itself. Even with an adequate pool of training data, if an object is significantly different from those seen during training the model may fail to label it correctly. Adding a new class to the model would, therefore, require many examples of the new class and the model to be completely retrained. For this reason, we argue that methods which are inherently adaptable are more suitable.

Transfer learning is another strategy for overcoming the restrictions imposed by training a DNN. This strategy applies a pretrained model from a source domain to initialise a targeted domain model [31]. Transfer learning has gained attraction in recent years, in part because of its flexibility and wide adaptability. It has found various uses within the context of machine learning approaches applied to cultural heritage and archaeology [12], [32]–[35]. In this paper, we take 3D point cloud data directly as input and train an explainable classifier as the transfer learning model. The use of interpretable and explainable models is valuable beyond high stake decisions [36]. The ability to discern why a decision or prediction was made, wrong or right, is highly useful. With this knowledge it becomes possible to perform diagnostics of the model and make corrections.

This paper presents a framework for the automatic extraction and labelling of memorial objects from 3D point cloud scenes that is adaptable, explainable and efficient. It can, therefore, be applied at scale. This paper serves as an improvement to the system of methods developed in our earlier research, GeoPart-Transfer [12]. This updates the methodology, wherein segmentation and classification are performed in two separate tasks, allowing adoption of the eXplainable Point Cloud Classifier (XPCC) [37] in place of the MLP classifier as used in the original paper. As such, the proposed framework is named GeoPart-XPCC. The adoption of XPCC enables a user to quickly add new objects for classification. Through a set of experiments, we demonstrate that the GeoPart-XPCC framework is efficient in terms of accuracy and is adaptable. The experiments illustrate that the proposed method can be applied to both within-sample and out-of-sample predictions, meaning that it can train one part of a scene and perform prediction on the remainder, as well as train on one scene and predict on an entirely difference scene with no significant change in accuracy. Comparative results between GeoPart-Transfer and the proposed GeoPart-XPCC are also presented.

Previous studies in the literature, utilised point clouds derived from airborne LiDAR systems (ALS), but very few studies focused on those created by mobile LiDAR. This is unsurprising, as ALS data representing large portions of the Globe are readily available from sources such as the USA National Science Foundation (NSF)-funded OpenTopography project [38]. However, ALS is limited in its ability to scan vertical surfaces and robustly capture small objects; qualities that are necessary for creating highly detailed representations of complex environments with many small vertical objects, such as burial grounds. In contrast, mobile LiDAR systems have a direct view of vertical faces and can provide a much finer, although more varied, spatial resolution with respect to the ground and objects. Mobile LiDAR was deployed typically in automotive scenarios. However, the introduction of handheld and wearable systems has made it possible to apply this technology to locations such as cultural heritage sites. The experiments presented in this paper were performed on point cloud data produced from mobile LiDAR scanners. Furthermore, many of the ALS techniques are not suitable for the fine-resolution point clouds produced by mobile LiDAR. The GeoPart-XPCC was designed specifically for dense mobile LiDAR point clouds.

Contributions of this paper include the following:

1) A novel GeoPart-XPCC framework for the autonomous extraction and labelling of memorial objects from cultural heritage scenes.
2) Application of the framework to five cultural heritage sites, with the sites divided into 11 separate scenes.
3) Within-sample accuracy results of the framework.
4) Out-of-sample accuracy results demonstrating domain transferability of the framework.

The remainder of this paper is organised as follows. Section 2 details the methods used in the proposed framework. Section 2.1 covers segmentation and Section 2.2 the classification method. Section 3 presents the experimental results. Section 3.1 includes descriptions of the cultural heritage site datasets used and the experimental design. Sections 3.2 and 3.3 report the results of the within-sample and out-of-sample experiments, respectively. We provide a discussion of these results in Section 4, including examination of comparative results between GeoPart-Transfer and

GeoPart-XPCC. Finally, in section 5 a conclusion is given.

II. Methods

We consider a 3D point cloud P as a set of 3D points $p_n \mid n = 1, \ldots, N$ corresponding to a scanned scene of a burial ground. The number of points in the cloud is denoted as N. The objective is to obtain a classification of the points in P as either belonging, or not belonging, to a memorial object and to assign each point to a homogeneous cluster of points that represent the object to which they belong. It is assumed that the point cloud scenes are scaled to a known unit size and have been pre-processed to remove statistical outliers and duplicate points.

III. Segmentation

Segmentation is implemented to divide the scene into geometrically simple regions. In this section, we describe the segmentation process of the GeoPart-XPCC framework. We expand on the segmentation methodology as described in [12]. A radius k-neighbourhood search is applied instead of the traditional k-nearest neighbour algorithm. As shown by [39], radius neighbourhood search is more robust when applied in a non-uniform sampling setting. A global energy model [40] is used to divide the scene based on the geometric complexity of local areas. Starting with the entire scene, areas that are more complex are divided progressively into less complex regions. By using geometric complexity rather than physical size, large shapes that are geometrically simple, such as walls and roads, can be obtained as a single continuous segment. Accordingly, the segments are not defined or limited by the number of points contained within them.

A set of features that characterise the local dimensionality of the scene is first computed, see Table I. For each point p_n, the nearest neighbouring points in the point cloud P are selected as the 45 closest neighbouring points within a set ball-radius of 1 m. If there are not enough points within this radius, then the next closest points are used. Second, the covariance matrix of their positions is calculated. From this, the set of eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and corresponding eigenvectors u_1, u_2, u_3 are obtained. The neighbourhood size is chosen such that it minimises the eigentropy E of the vector $(\lambda_1, \lambda_2, \lambda_3)$ where $\Lambda = \sum_{i=1}^{3} \lambda_i$ [41]. The set of features $f_n \in \mathbb{R}^4$ is constructed from the eigenvalues. This feature vector characterises the local dimensionality and geometry of the point and its neighbours. The type of dimensionality features used includes linearity, planarity, and scattering [42]. Linearity describes how well the neighbourhood represents a straight line; planarity describes how well the neighbourhood represents a plane; sphericity measures the scattering of points in relation to a spherical shape. Additionally, the verticality feature is used, which indicates the geometric orientation of the neighbourhood [40]. A verticality value of 0 represents a horizontal orientation, whereas a verticality value of 1 represents a vertical orientation.

The spatial relationship between points is represented as a nearest neighbour adjacency graph $G_{nn} = (V, E_{nn})$ of the point cloud, where $V = \{V_1, \ldots, V_N\}$ is the set of feature vectors f_n representing its associated point p_n, and E_{nn} is the corresponding edges to the nearest 14 neighbouring points (i.e., $nn = 14$). The adjacency relationship between a node and its neighbours is encoded as the weight of the edges. The relation between point density and the number of edges is chosen to be computationally efficient and accuracy as advocated by [43]. Segmentation is performed by splitting the graph into non-overlapping connected components defined as the solution of the following Potts energy model:

$$argmin \sum_{n \in V} \left\| g_n - f_n \right\|^2 + \rho \sum_{(i,j) \in E_{\text{nn}}} w_{i,j} [g_i - g_j \neq 0]. \quad (1)$$

The variable g is used to determine the optimal value; that is, the minimisation of the problem. The Iverson bracket $[\cdot]$ yields 0 if the internal expression is true and 1 everywhere else. The edge weight $w \in \mathbb{R}_+^{|E|}$ is chosen to decrease linearly with respect to the edge length, and factor ρ is the regularisation strength, which determines the coarseness of the resulting partition [26]. The minimised high-order function is nonconvex and noncontinuous, meaning that the problem cannot be solved easily for large point clouds. However, the l_0 "cut pursuit" algorithm [44] is used to find an approximate solution quickly, regardless of the number of points. The conditional multi-scale partitioning filter from GeoPart-Transfer is then applied to increase the chances of finding smaller objects that have been missed in the initial segmentation [12]. The point cloud segments are represented as the set $S = \{S_i \mid i = 1, \ldots, h\}$, where h is the number of segments returned by the cut adjacency graph and partitioning filter.

TABLE I
Local Dimensional Geometric Features.

Property	Formula
Linearity	$\frac{\lambda_1 - \lambda_2}{\lambda_1}$
Planarity	$\frac{\lambda_2 - \lambda_3}{\lambda_1}$
Sphericity	$\frac{\lambda_3}{\lambda_1}$
Verticality	$\sum_{j=1}^{3} \left\| \langle [0,0,1], u_j \rangle \right\|$

IV. Classification

In this paper, the XPCC method is applied within the proposed framework to perform classification. The input to the GeoPart-XPCC classification method is taken as the set of point cloud segments, S, from the segmentation process. Functionally, the point cloud segments do not need to have the same number of points. The output of the classification step is a set of predicted labels $L = \{L_n \mid n = 1, \ldots, N\}$ where each point in the point cloud is assigned a class label. The classification considers only each segment's geometric shape. That is, the x, y, z coordinate points are taken as the only input. Colour and other observational features are not used. Therefore, the classification is not affected by

inconsistencies caused by external factors present during data collection, such as different lighting or times of day. The GeoPart-XPCC classification model is a non-iterative prototype-based feedforward neural network which directly takes the point coordinates as input. An illustration of the GeoPart-XPCC classifier can be seen in Figure 1. Algorithms for training the model and performing predictions are detailed in the following sections.

A. Latent Feature Extraction

The first layer of the GeoPart-XPCC model is the feature extraction layer. This layer encodes the shape of each segment of the point cloud scene into a global descriptor. To do this, a transfer learning concept is applied, and the layers of a pre-trained DNN are used as a fixed feature extractor. The global descriptor is obtained as the vector of latent features computed from the final, fully connected layer of this DNN. The latent feature space is optimised for the separation of the training data, and it has been shown that this feature space can also effectively separate objects from a different domain [37]. In contrast to [12], the DNN chosen is the point-based KP-CNN with rigid Kernel Point Convolution (KPConv) blocks [27] pre-trained on the ModelNet40 [45] dataset. Crucially, the KP-CNN network is rotation-invariant, accepts variable cloud sizes as input and remains robust to varying point densities. These characteristics are important to remain generalisable to objects from different scenes or scanning technologies. The latent feature extractor produces a 1-by-1024 dimensional global feature vector; one per segment.

The set of latent feature vectors extracted by the DNN are denoted as $X = \mathrm{DNN}(S) \in \mathbb{R}^{N \times d_f}$, where $\mathrm{DNN}(\cdot)$ is the fixed DNN and d_f is the dimension of the extracted features. In the case of the KP-CNN, $d_f = 1024$. The latent feature vector representing a given segment is denoted as $\mathbf{x}_i$, where i is the index of the corresponding segment. *L2* normalisation is computed for each individual feature vector, $\mathbf{x}_i = \frac{\mathbf{x}_i}{\max(\|\mathbf{x}_i\|, \epsilon)}$, where ϵ is a small constant value, as is standard.

B. Training

The principle behind the GeoPart-XPCC classifier is the so-called anthropomorphic approach to machine learning [46]; that is, the intuition that people learn by comparing similarity between objects, but only remember a few distinct objects during decision-making (i.e., the prototypes). If a new object is encountered, a person is likely to assume that it belongs to the class which it most closely resembles. Following this logic, the learning of the proposed method revolves around the position and properties of the prototypes in the feature space.

The GeoPart-XPCC classification model consists of independently trained structures called *class containers*. Meta-parameters for the GeoPart-XPCC classifier are trained per-class and calculations are performed separately for each one. For efficiency, each class container can be trained simultaneously in parallel. The class containers are trained using a filtering algorithm to identify the prototypes from within the training data; the prototypes being the most representative training samples for that class. This algorithm is non-iterative and can be performed in one-pass. The meta-parameters for any given class container are initialised starting with the first training observation belonging to it,

$$k \leftarrow 1, M \leftarrow 1, \mu \leftarrow x_1, \varphi_1 \leftarrow x_1, N_1 \leftarrow 1, r_j \leftarrow r^*, \tag{2}$$

where k is the current instance (number of training data seen), M is the number of prototypes identified for a class ($M \ll N$), μ is the recursive global mean of all training data as yet observed, φ_1 is the first prototype $\{\varphi_j \mid j = 1, \ldots, M\}$, N_j is the number of member points around each prototype, and r_j is the radius of the area of influence of the corresponding prototype. The element r^* is the initial degree of similarity of the prototype member space and is computed as:

$$r^* = \frac{x_i}{\|x_i\| - \frac{\varphi_i}{\|\varphi_i\|}}. \tag{3}$$

As new training data are seen, the meta-parameters for any given class container are updated. The instance count is first updated, $k \leftarrow k+1$. Then, the global mean is updated, $\mu \leftarrow \frac{(\mu(k-1)+x_i)}{k}$. With the updated global mean, it is then possible to calculate the data density. Density indicates how strongly data samples influence one another within the data space [46]. Thus, samples with a higher density value are closer to the global mean. The density function is defined as the following Cauchy function [46]:

$$D(x_i) = \frac{1}{\left(1 + \left\|x_i - \mu\right\|^2\right)/\sigma} \tag{4}$$

where $\sigma = 1 - \left\|\mu^2\right\|$.

The prototypes are the local peaks of the data density within the feature space [46]. Accordingly, the prototypes are determined by filtering the training data according to their density and area of influence. The prototypes are independent from each other, such that the addition of a new prototype does not influence existing ones. The prototypes with the maximum and minimum density, $\alpha = D(p_j)_{max}$ and $\beta = D(p_j)_{min}$, respectively, and the index $j^* = argmin\left(x_i - p_j\right)$, which denotes the prototype closest to the current observation, are used to control the addition of new prototypes [47]. If the density of the current observation is greater than α or less than β, the observation is added as a new prototype to its respective class. The observation may also be added as a prototype if it lies outside the area of influence of the closest already existing prototype. When none of these conditions is met, the new training observation is assigned as a member of the nearest prototype in the feature space; the prototype is then updated recursively as follows [47]:

$$\phi_{j^*} \leftarrow \frac{\varphi_{j^*}\left(N_j - 1\right) + x_i}{N_j}. \tag{5}$$

The number of data associated with a certain prototype is increased accordingly, $S \leftarrow S + 1$.

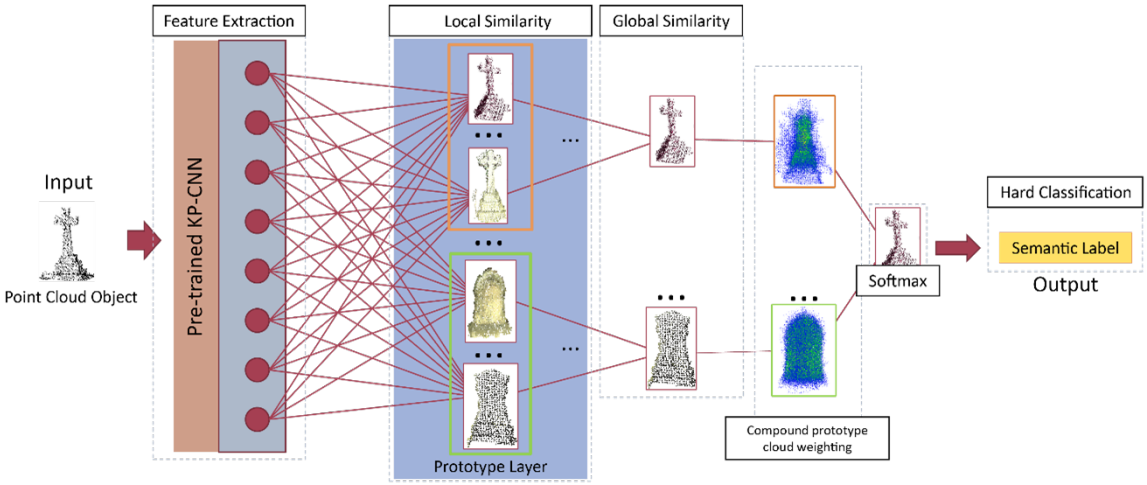


Fig. 1. Illustration of the GeoPart-XPCC classifier.

After the initial training process, the model can continue to learn by absorbing new training data. As the updates are performed recursively, the model retains information about previously seen data samples with no retraining required. Similarly, entirely new classes can be added with the addition of a new class container. Alternatively, it is possible for a user to manipulate and fine-tune the model through the manual addition, removal or adaption of prototypes. Pseudo-code of the training algorithm is given in Figure 2.

C. Prediction

The GeoPart-XPCC prediction architecture is represented as a series of feedforward network layers. For any given input data not seen during training, S_t, a classification label can be predicted. In the first layer, the feature extraction layer, the point cloud segments are input as described in Section 1.2. The output of this layer is the corresponding latent feature vector x_t that encodes the global shape of the object. The second layer is the prototype layer, wherein the local similarity between the previous layer's output and each class's prototypes is calculated. Similarity is a metric defined as:

$$\text{Similarity}\left(x_t, p_j\right) = \frac{1}{1 + \left\| x_t - p_j \right\|^2}. \tag{6}$$

The output of the prototype layer is the tensor $\mathbb{S}_j$ which contains the similarity of the input data observation with respect to the prototypes of all classes. In the third layer, denoted the Global Similarity layer, the similarity score for each class is extracted as the most similar prototype from each class. That is, the output of this layer is the tensor $\mathbb{S}_j^* = \max \mathbb{S}_j$. In the fourth layer, the global similarity score of each class is weighted by the input sample's similarity to the respective class's compound prototype. The compound prototype (CPC) is estimated as the recursively updated global mean of the class, $\mathbb{S}_{CPC} = \text{Similarity}(x_t, CPC_j)$. The output of this layer is $\mathbb{S}_j^* = \mathbb{S}_j^* \odot \mathbb{S}_{CPC}$, where $\odot$ is the Hadamard product. In the last layer, the Softmax function is used.

$$\gamma_i^* = \frac{\exp\left(\gamma_i\right)}{\sum_k \exp\left(\gamma_k\right)}. \tag{7}$$

This normalizes the output of the previous layer to a probability distribution over the predicted output classes. Hard classification is performed using the $argmax\gamma_i^*$ function applied on the output to produce the predicted class label for each segment. This label is then assigned to each point within the segment and, thus, achieves the goal of each point being assigned a label. That is, the network output is $L = \{L_n \mid n = 1, \ldots, N\}$. The classifier labels the scene as multiple subclasses of memorial objects, and these labels are then aggregated into a binary classification of memorial object and non-memorial object.

V. Experimental Results

VI. Data scenes and experimental design

The GeoPart-XPCC framework was validated on five different cultural heritage sites of burial grounds found across England. These sites are represented as 3D point clouds and were divided to create Scenes 1 to 11. The digitised cultural heritage sites and operator-labelled mappings were provided by the burial ground management system team at Atlantic Geomatics (UK) Limited. The point clouds were collected using the Leica Pegasus capture platform [48], with the exception of Scene 11. Scene 11 was collected using UAV mounted sensors and photogrammetry. The point density of the point cloud scenes varies, with respect to the distance of the scanner and the surface of the environment being scanned. Figure VI-A-VI-E show the scenes in RGB values for visualisation purposes only, and these properties are not used as input to either the segmentation or classification algorithms within the proposed framework.

For all experiments, the precision, recall, accuracy and F1-score were used as evaluation metrics for point-wise

Fig. 2. XPCC Training Process

for `Each Class` **do**
  Initialise using the first feature vector sample x_i
  $k \leftarrow 1$, ▷ current instance (number of training samples seen)
  $M \leftarrow 1$, ▷ number of prototypes identified
  $\mu \leftarrow x_1$, ▷ recursive global mean of all data samples observed
  $p_1 \leftarrow x_1$, ▷ the first prototype $\{p_j \mid j = 1, \ldots, M\}$
  $N_1 \leftarrow 1$, ▷ number of member points around each prototype
  $r_j \leftarrow r^*$ ▷ radius of the area of influence of the corresponding prototype
  for $i = 2$, ... **do** ▷ Read x_i
    $D(x_i)$ ▷ Calculate Similarity of x_i
    $\{D(p_j) \mid j = 1, ..., M\}$ ▷ Calculate Similarity of Prototypes
    $\alpha \leftarrow \max D(p_j)$
    $\beta \leftarrow \min D(p_j)$
    if $D(\mathbf{x}_i) > \alpha$ **OR** $D(\mathbf{x}_i) < \beta$ **OR** $(\|\mathbf{x}_i - p_{j^*}\| \leq r_{j^*})$ **then**
      Add x_i as new prototype
    else
      $j^* \leftarrow \arg\min(\mathbf{x}_i - p_j)$ ▷ Search for the closest prototype
      $p_{j^*} \leftarrow \frac{p_{j^*}(N_j - 1) + \mathbf{x}_i}{N_j}$ ▷ Update the prototype
      $S \leftarrow S + 1$ ▷ Update count of support members
      $r_{j^*} \leftarrow \frac{r_{j^*} + \left(1 - p_{j^*}\right)}{2}$ ▷ Update the radius of influence
    end if
  end for
end for

qualitative results. Intuitively, accuracy is the fraction of predictions that were correct out of all observations. Similarly, the precision reports what proportion of positive identifications were correct. Recall reports the proportion of real positives that were identified correctly. The F1-score is the harmonic mean of precision and recall and is calculated as

$$F_1 = 2 \times \frac{\text{recall} \times \text{precision}}{\text{recall} + \text{precision}}. \tag{8}$$

A perfect F1-score of 1 indicates a perfect precision and recall and implies that the framework performed identically to the user operator, regarding the task of extracting and labelling memorial objects. Both the macro average and weighted average of the accuracy, precision, recall and F1-score are reported. The macro average does not consider label imbalances, whereas the weighted average accounts for the number of true instances for each label. As such, the weighted average for F-score can result in a value that is not between precision and recall. All metrics used reach their minimum and maximum values at 0 and 1, respectively. Satisfactory results are those which are visually comparable to the operator-labelled point clouds.

Experiments were run on a Unix machine with 3.4 GHz AMD Ryzen 9 5950X 16-Core CPU, NVIDIA 2070 SUPER GPU, 32GB RAM and SSD. The fixed KP-CNN feature extractor network was trained using the NVIDIA 2070 SUPER GPU. Training the KP-CNN on the ModelNet40 dataset took around 5 hours. This only needs to be performed once. After training, the fixed network acts as a generalised point cloud feature extractor. The times elapsed to perform the segmentation and classification tasks are presented in Table VI-E, along with the number of points in each scene.

A. Adlington Cemetery, Adlington, Lancashire

Scenes 1 and 2 are taken from the point cloud of Adlington Cemetery. The cemetery is located in the North West region of England, in the Borough of Chorley in Lancashire. Scene 1 consists mainly of lawn areas and roads. There are several instances of medium and large trees, and thick vegetation is found in proximity to the perimeter wall. Scene 2 is similar in composition, although many more instances of medium trees exist. A chapel and gate building, along with cars parked on the road can be found in Scene 1. Both scenes contain instances of headstones, crosses, obelisks, ledger slabs and kerbs. Training data are taken from the Adlington Scene for the model results in the out-of-sample experiments section.

B. Church of St. Nicholas graveyard, High Bradfield, Yorkshire

The graveyard of the Church of St. Nicholas is in the town of High Bradfield in the county of South Yorkshire, England. The original scene is split to form Scene 3 and Scene 4. Both scenes are composed of the traditional monumental style of cemetery, where monuments are in tight proximity and contain many instances of headstones and ledger slabs. Dense vegetation and large trees are present around the edges of both scenes. The church building is a Grade One listed building and is found within Scene 3. Several instances of stone chests and grave kerbs are found in

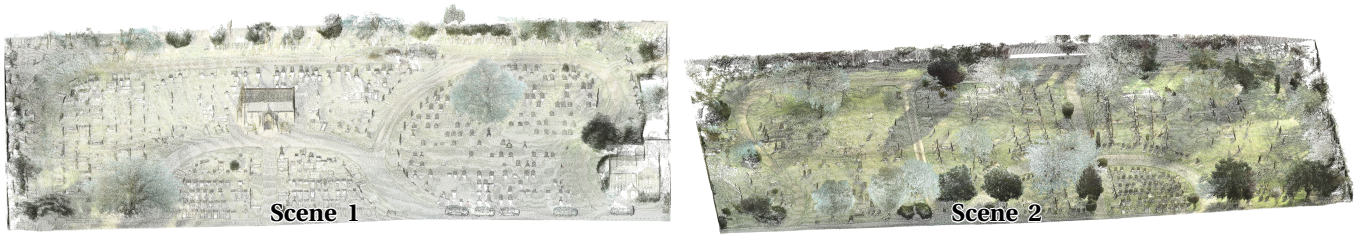


Fig. 3. Illustration of Scenes 1 and 2 from the Adlington Cemetery cultural heritage site.

the area immediately surrounding the church building. In contrast to the other cultural heritage sites used in this paper, this graveyard is situated on a hill rather than in a flat area.

C. Staines Cemetery, Staines-upon-Thames, Surrey

Staines Cemetery is in the market town of Staines-upon-Thames, Surrey in the South East region of England. The cemetery contains multiple sections of lawn, a dedicated garden as well as roads and a chapel building. Several types of monuments are located within the cemetery, including headstones, ledger slabs, grave kerbs, obelisks and stone chests. The original scene is split into three sub-scenes: Scenes 5 - 7. The chapel building is located within Scene 6. All scenes contain instances of vegetation and trees.

D. Peter's Church Churchyard, Heversham, Cumbria

The churchyard of St. Peter's Church is located in the village of Heversham, Cumbria in the North West Region of England. The original scene is split into three sub-scenes: Scenes 8 – 10. The Church building is a Grade II* listed building which dates to the 12th century. The church building is present in Scene 9. Scene 10 contains part of an outbuilding and Scene 8 contains a prominent perimeter wall. Given its age, the churchyard contains many different styles of memorial objects. The churchyard is laid similarly to the Church of St. Nicholas graveyard, in the traditional monumental style, although the monuments have not been placed as close together. Most of the monuments are either headstones or stone ledgers.

E. Churchyard of the Church of All Saints, Lullington, Somerset

Scene 11 is a scan of the churchyard belonging to the Church of All Saints, in the village of Lullington, Somerset in the South West region of England. The church building dates to the 12th century and is a Grade I listed building. The churchyard is similar in layout and monument types to the other scenes of contemporary churchyards (Scenes 3 and 9). The scene was collected from a UAV, rather than from the mobile terrestrial LiDAR platform. Notably, the aerial platform meant that the upper regions of the church building could be scanned, in contrast to examples such as Scenes 3 and 9. However, an artefact of the aerial view is that the areas under dense vegetation are occluded and this results in missing data. This can be seen in the lower portion of the scene in Figure VI-E.

TABLE II
ELAPSED TIME TAKEN TO PERFORM THE SEGMENTATION AND CLASSIFICATION METHODS FOR EACH SCENE ROUNDED TO MINUTES (M), SECONDS (S), AND SUB-SECONDS. THE NUMBER OF POINTS IN EACH SCENE IS ALSO INCLUDED. CLASSIFICATION WAS PERFORMED ON A NVIDIA 2070 SUPER GPU.

Scene Number	Segmentation Time	Classification Time	Size of Point Cloud
1	7 m 46 s	1.3 s	2,653,259
2	17 m 41 s	2.4 s	5,350,073
3	10 m 43 s	1.5 s	3,571,623
4	19 m 14 s	2.7 s	6,771,377
5	12 m 24 s	1.5 s	3,806,240
6	18 m 28 s	2.6 s	5,841,840
7	16 m 18 s	2.5 s	5,350,073
8	3 m 14 s	1.1 s	1,229,821
9	6 m 59 s	1.3 s	2,221,683
10	6 m 39 s	1.3 s	2,065,839
11	3 m 9 s	1.2 s	2,342,077

VII. WITHIN-SAMPLE ACCURACY

The following section details the experiments for within-sample accuracy. These experiments demonstrate the ability of the framework to learn from the memorial objects present in part of a scene and then autonomously find those remaining. In this experiment, both the training and testing data come from the same cultural heritage site. As such, results are reported using the cultural heritage site name in addition to the specific scenes. Quantitative results for all experiments are given in Table V and visual results are presented in Figure 7. Training the GeoPart-XPCC classifier took around 3 seconds on average for each experiment using the GPU.

TABLE III
QUANTITATIVE RESULTS FOR THE WITHIN-SAMPLE EXPERIMENTS

	Scene Name		Precision	Recall	F1-score
1		Accuracy			0.95
2	Scene 1	Macro Avg.	0.86	0.83	0.85
3		Weighted Avg.	0.95	0.95	0.95
4		Accuracy			0.93
5	Scene 4	Macro Avg.	0.88	0.81	0.84
6		Weighted Avg.	0.93	0.93	0.93
7		Accuracy			0.97
8	Scene 7	Macro Avg.	0.86	0.90	0.88
9		Weighted Avg.	0.97	0.97	0.97
10		Accuracy			0.90
11	Scene 8	Macro Avg.	0.75	0.86	0.79
12		Weighted Avg.	0.93	0.90	0.91

The within-sample accuracy experiment results for the point cloud of Adlington Cemetery are shown in Table III rows 1-3. Visual results are shown in Figure 7 (a & b). In

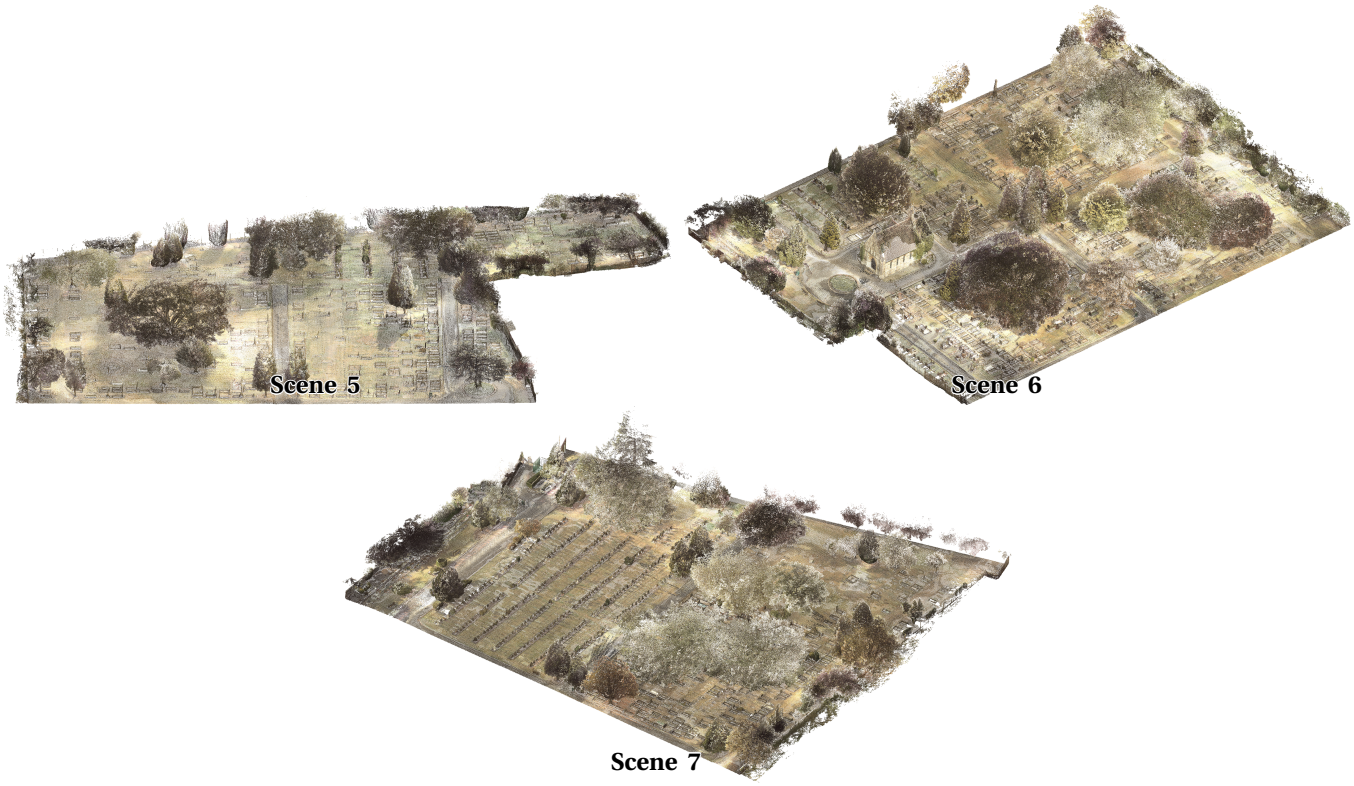


Fig. 4. Illustration of Scenes 5, 6, and 7 from the Staines Cemetery cultural heritage site.

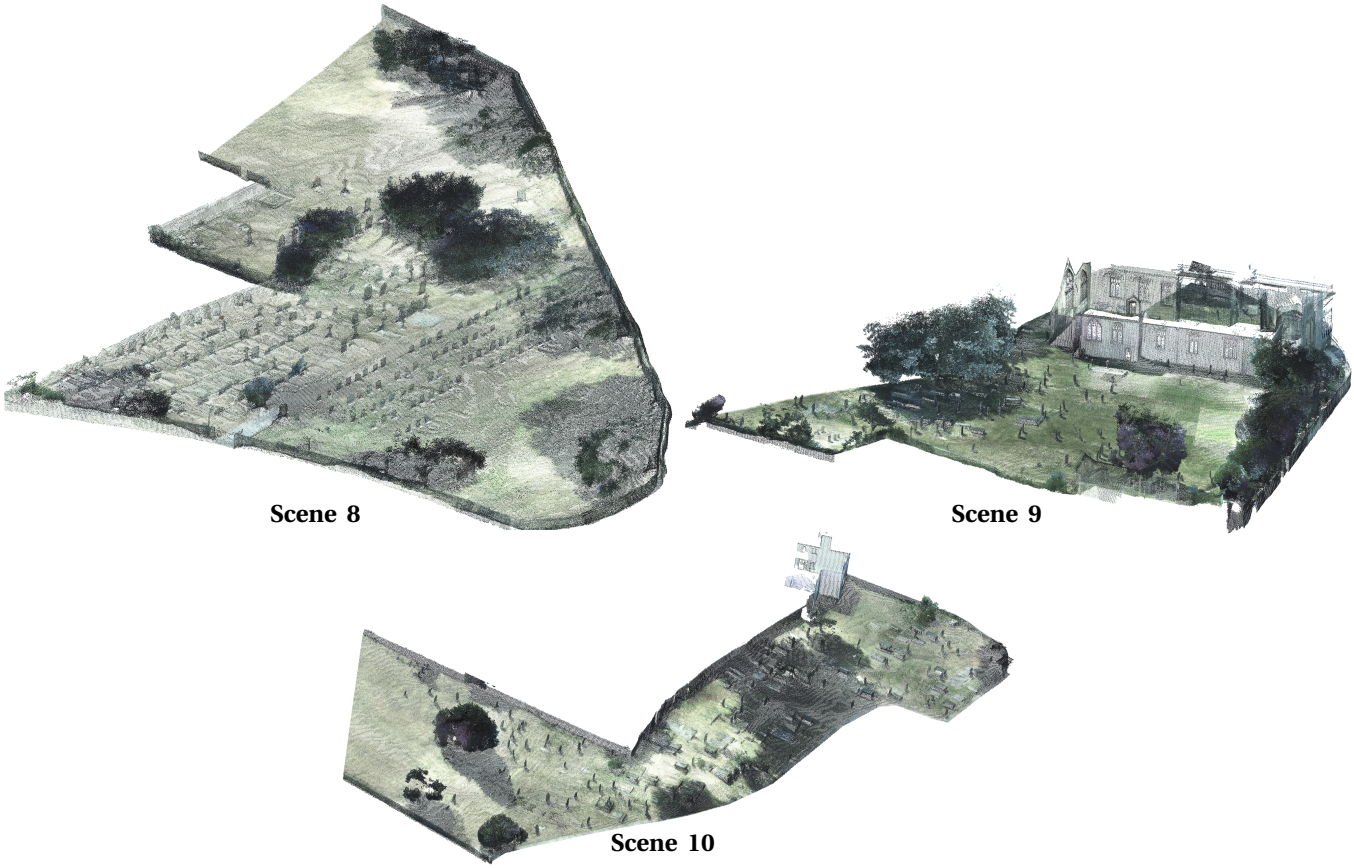


Fig. 5. Illustration of Scenes 8, 9 and 10 from the St. Peter's Church churchyard cultural heritage site.

Fig. 6. Illustration of Scene 11 of the Church of All Saints churchyard.

this experiment, the Scene 2 operator-labelled points were used as training data and the GeoPart-XPCC framework was applied to Scene 1. The framework achieved a weighted average F1-score of 0.95. Similarly, the scores for the weighted average precision and recall were both 0.95. The macro average for precision, recall and F1-score were 0.86, 0.83 and 0.85, respectively. Overall accuracy was 0.95.

Experimental results for the Church of St. Nicholas graveyard are given here. For this experiment, the operator-labelled point cloud for Scene 3 was used as training data, and the proposed framework was applied to Scene 4. The quantitative results can be seen in Table III rows 4-6, and visual results can be seen in Figure 7 (c & d). The accuracy score was 0.93, and the reported weighted F1-score achieved was 0.93. For the precision metrics, macro and weighted average, the scores were 0.88 and 0.93, respectively. The scores for the recall metrics achieved were 0.81 for the macro average and 0.93 for the weighted average. The macro average F1-score was 0.84.

The within-sample experiment for Staines Cemetery used the operator-labelled memorial objects from Scenes 5 and 6 to train the classifier. Prediction was performed on Scene 7. Results can be seen in Table III (rows 7-9) and Figure 7 (e & f). The metric scores achieved by the proposed framework for this experiment are as follows. The macro average scores for precision, recall and F1-score were 0.86, 0.90 and 0.88, respectively. The weighted precision, recall and F1-score were 0.97 for all. An overall accuracy of 0.97 was achieved.

For the experiment on St. Peter's Church churchyard cultural heritage site, operator-labelled point clouds of Scenes 9 and 10 were used as training data and the GeoPart-XPCC framework was then applied to Scene 8. Results can be seen in Table III rows 10-12. A visual comparison against the operator-labelled points for Scene 8 can be seen in Figure 7 (g & h). The framework achieved an overall accuracy of 0.90, a weighted F1-score of 0.91, a weighted precision score of 0.93, and a weighted recall score of 0.90. Similarly, the framework achieved a macro average F1-score of 0.91, a macro average precision score of 0.75, and a macro average recall score of 0.86.

VIII. Out-of-sample Accuracy

The experiments for out-of-sample accuracy indicate domain transferability. That is, they demonstrate the ability of the framework to learn from the memorial objects present in one scene and then autonomously extract and label memorial objects in completely different scenes. Training data are from a different point cloud scene than the test data. The GeoPart-XPCC's classifier was trained on the operator-labelled data from Scenes 1 and 2 from the Adlington Cemetery cultural heritage site. The same classifier was used for all out-of-sample experiments; training took 2.6 seconds on a GPU. Quantitative results for all experiments are presented in Table V and visual results are shown in Figures 8 and 9.

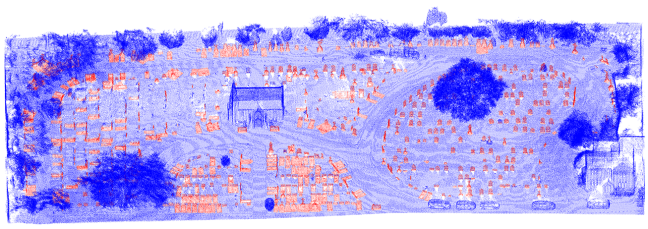
In terms of accuracy, the GeoPart-XPCC framework achieved a score above 0.90 in all, but one of the out-of-sample experiments. Scenes 11 and 3 had the highest

TABLE IV
Comparison of weighted average F1-scores for the proposed GeoPart-XPCC and the GeoPart-Transfer MLP-sigmoid method. Both tests use the same segmentation proposed GeoPart-XPCC segmentation method for fairness.

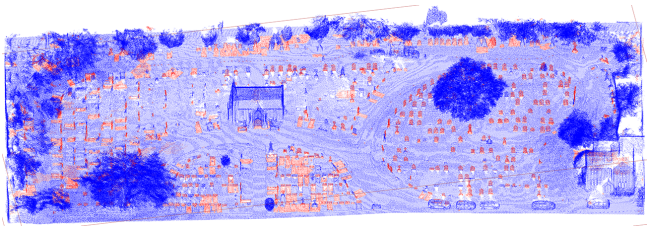
Scene Name	GeoPart-Transfer (MLP-σ)	GeoPart-XPCC
Scene 1	0.91	0.95
Scene 4	0.86	0.93
Scene 7	0.91	0.97
Scene 8	0.72	0.91

TABLE V
Quantitative results for the out-of-sample experiments

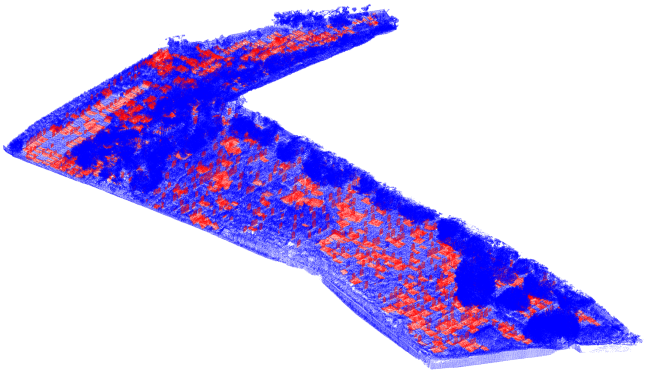
	Scene Name		Precision	Recall	F1-score
1		Accuracy			0.95
2	Scene 3	Macro Avg.	0.93	0.76	0.82
3		Weighted Avg.	0.95	0.95	0.94
4		Accuracy			0.91
5	Scene 4	Macro Avg.	0.9	0.68	0.74
6		Weighted Avg.	0.91	0.91	0.89
7		Accuracy			0.93
8	Scene 5	Macro Avg.	0.78	0.85	0.81
9		Weighted Avg.	0.94	0.93	0.94
10		Accuracy			0.92
11	Scene 6	Macro Avg.	0.8	0.8	0.8
12		Weighted Avg.	0.92	0.92	0.92
13		Accuracy			0.93
14	Scene 7	Macro Avg.	0.76	0.84	0.79
15		Weighted Avg.	0.94	0.93	0.94
16		Accuracy			0.94
17	Scene 8	Macro Avg.	0.86	0.79	0.82
18		Weighted Avg.	0.93	0.94	0.93
19		Accuracy			0.78
20	Scene 9	Macro Avg.	0.66	0.82	0.68
21		Weighted Avg.	0.9	0.78	0.82
22		Accuracy			0.91
23	Scene 10	Macro Avg.	0.85	0.85	0.85
24		Weighted Avg.	0.91	0.91	0.91
25		Accuracy			0.96
26	Scene 11	Macro Avg.	0.63	0.86	0.68
27		Weighted Avg.	0.98	0.96	0.97



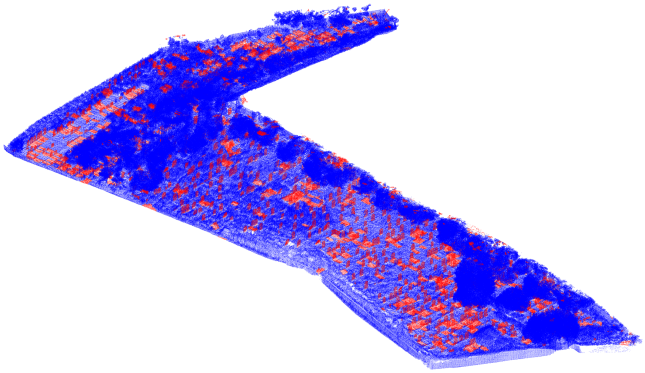
(a) Scene 1 operator-Labelled point cloud



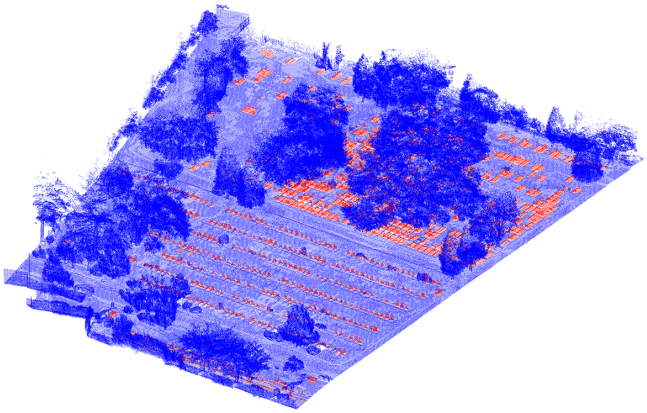
(b) Scene 1 GeoPart-XPCC predicted objects



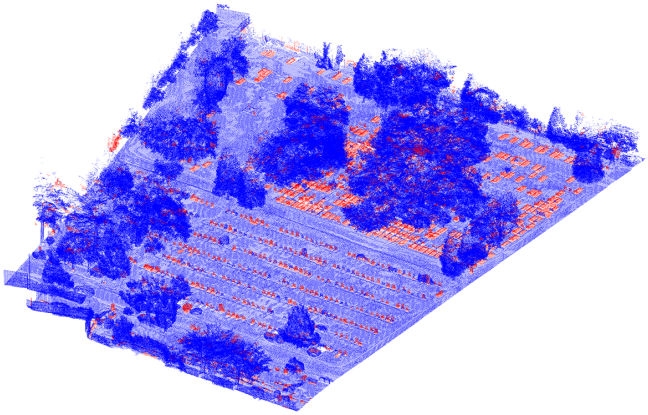
(c) Scene 4 operator-Labelled point cloud



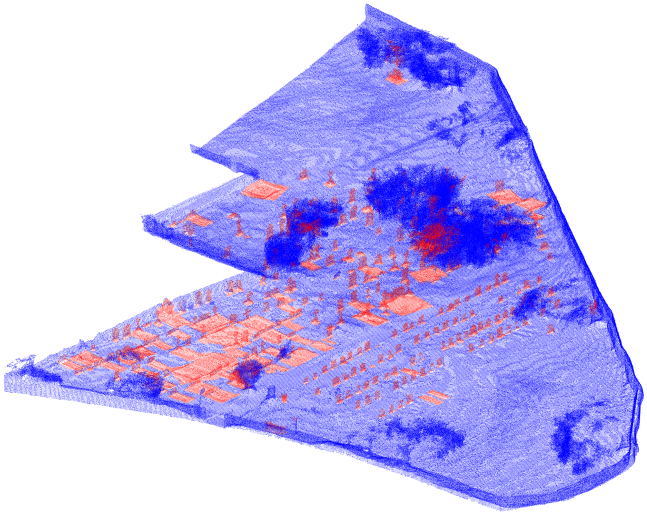
(d) Scene 4 GeoPart-XPCC predicted objects



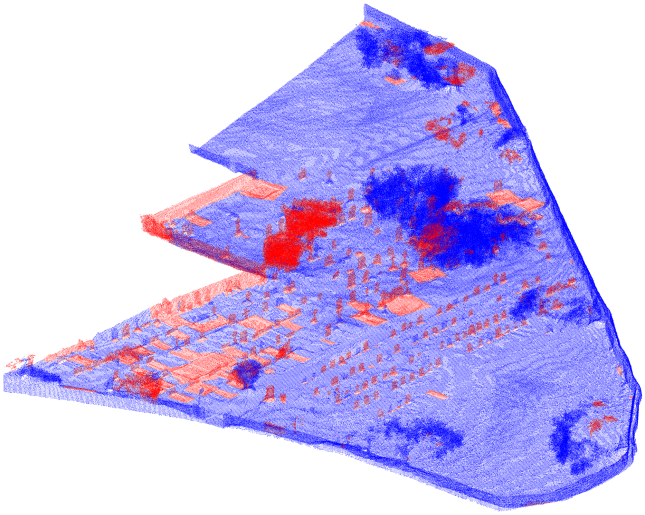
(e) Scene 7 operator-Labelled point cloud



(f) Scene 7 GeoPart-XPCC predicted objects



(g) Scene 8 operator-Labelled point cloud



(h) Scene 8 GeoPart-XPCC predicted objects

Fig. 7. Visual illustrations of point cloud Scenes 1, 4, 7, and 8, from the within-sample experiments. Points belonging to the memorial objects are coloured in red; the non-memorial points are colored in blue. The sub-figures to the left of the centre line (a, c, e, and g) are the operator-labelled point cloud scenes. The sub-figures to the right of the centre line (b, d, f, and h) are the GeoPart-XPCC predicted point cloud scenes. The operator-labelled and predicted illustrations of each scene are paired per row.

accuracy, with scores of 0.96 and 0.95, respectively. The score for Scene 8 was the next highest, at 0.94. An accuracy score of 0.93 was achieved by the framework for Scenes 5 and 7, a score of 0.92 for Scene 6, and scenes 4 and 10 had scores of 0.91. Lastly, the experiment on Scene 9 resulted in an accuracy score of 0.78.

The precision score identifies the frequency with which the proposed framework was correct when predicting a positive class. From the out-of-sample experiments, the highest macro average precision achieved by the GeoPart-XPCC was 0.93 on Scene 3. Scores of 0.90, 0.86, 0.85, and 0.8 were the results for this metric on Scenes 4, 8, 10 and 6, respectively. When applied to Scenes 5 (0.78), 7 (0.76), and 9 (0.66), all had macro average precision scores under 0.8, and the lowest was 0.63 on Scene 11. The weighted average precision takes class imbalances into consideration. The lowest weighted precision score from the out-of-sample experiments was 0.90 on Scene 9. Applying GeoPart-XPCC on Scenes 4 and 10 resulted in a weighted precision score of 0.91, and Scenes 6 and 8 had scores of 0.92 and 0.93, respectively. The next highest weighted scores were 0.95 on Scene 3 and 0.94. The highest weighted precision score achieved for this set of experiments was 0.98 for Scene 11.

Recall is an indication of how many predictions the GeoPart-XPCC framework was able to correctly identify, out of all possible positive labels. The lowest macro recall score for the out-of-sample experiments was 0.68 on Scene 4, followed by 0.76 on Scene 3 and 0.79 on Scene 8. The framework achieved a score above 0.8 for the following scenes: Scene 6 (0.80), Scene 9 (0.82), Scene 7 (0.84), Scene 10 (0.85), Scene 5 (0.85), and Scene 11 (0.86). From the out-of-sample experiments, the lowest weighted average recall was on Scene 9, with a score of 0.78. On all other experiments, from this section, the framework achieved weighted precision scores higher than 0.90. For this metric, the results show that for Scenes 4 and 10, the framework achieved a score of 0.91 for both. In ascending order, the framework achieved the following: a score of 0.92 on Scene 6, a score 0.93 for both Scenes 7 and 5, a score of 0.94 on Scene 8, and a score of 0.95 for Scene 3. Finally, the out-of-sample experiment with the highest weighted average recall score was 0.96 and was achieved on Scene 11.

The F1-score is a combination of the information provided by precision and recall; thus, it is a measurement of the accuracy of the proposed framework with respect to the balance of these metrics. For the out-of-sample experiments, the following F1-scores were the result of applying the GeoPart-XPCC framework to the memorial site scenes. For the macro average: a score of 0.68 was achieved on both Scenes 11 and 9, followed by a score of 0.74 on Scene 4, then 0.79 on Scene 7, 0.8 on Scene 6, 0.81 on Scene 5, 0.82 on Scenes 3 and 8, and finally, a score of 0.85 on scene 10. For the weighted average: a score of 0.82 was achieved on Scene 9, then 0.89 on Scene 4, 0.91 on Scene 10, 0.92 on Scene 6, 0.93 on Scene 8, followed by a score of 0.94 on Scenes 7, 8, and 9. Lastly, a score of 0.97 was the result on Scene 11, and was the highest F1-score achieved by the framework for the out-of-sample experiments.

TABLE VI
COMPARISON OF WEIGHTED AVERAGE F1-SCORES RESULTS FROM THE PROPOSED GEOPART-XPCC AND THE GEOPART-TRANSFER MLP-SIGMOID METHOD. BOTH TESTS USE THE SAME SEGMENTATION PROPOSED GEOPART-XPCC SEGMENTATION METHOD FOR FAIRNESS.

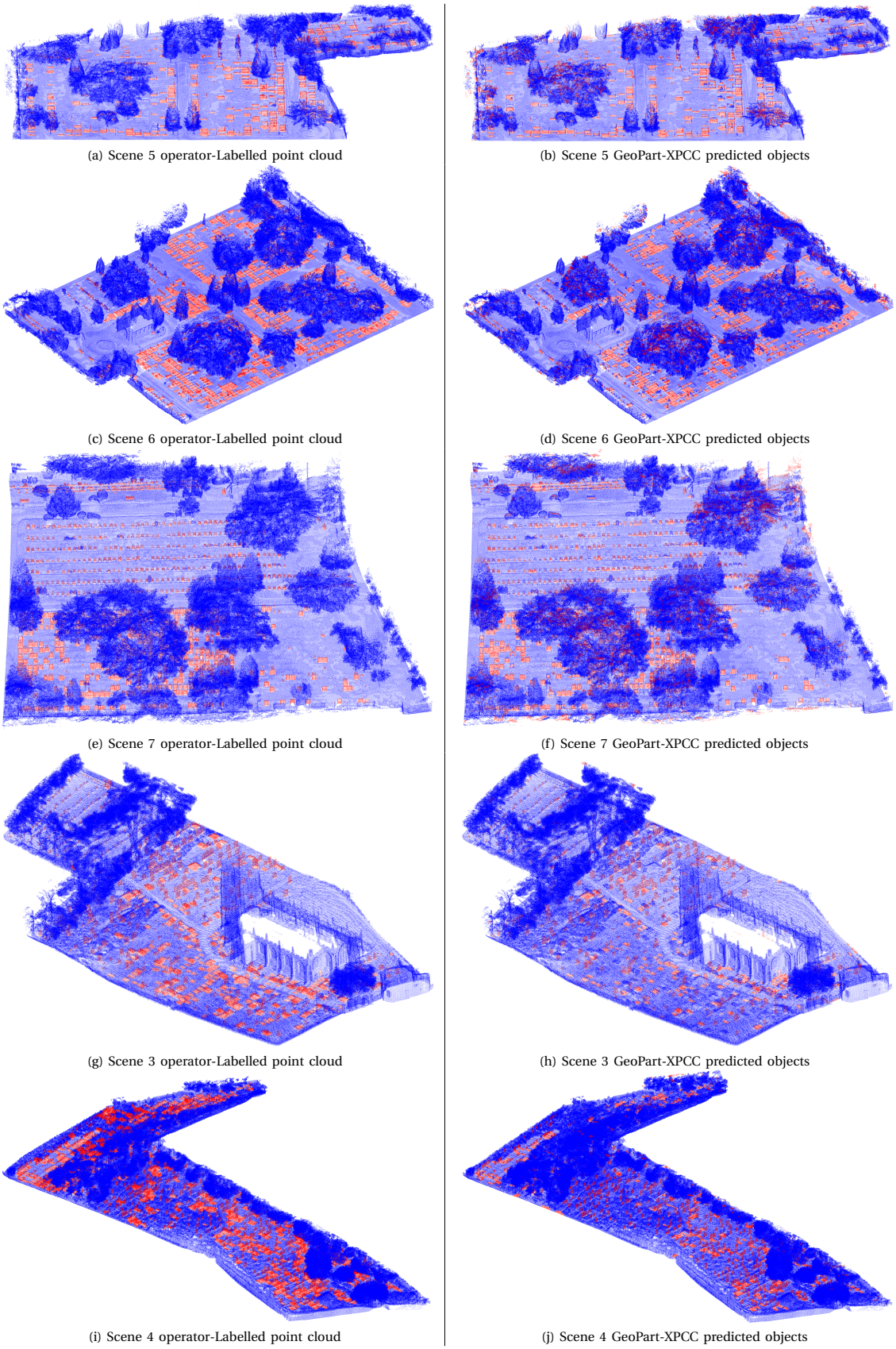
Scene Name	GeoPart-Transfer (MLP-σ)	GeoPart-XPCC
Scene 3	0.90	0.94
Scene 4	0.71	0.89
Scene 5	0.91	0.94
Scene 6	0.89	0.92
Scene 7	0.89	0.94
Scene 8	0.86	0.93
Scene 9	0.57	0.82
Scene 10	0.87	0.91
Scene 11	0.66	0.97

The last out-of-sample experiment provides a demonstration of the adaptability of the proposed framework. New data samples, and even classes, can be added to the GeoPart-XPCC classification algorithm after the initial training process has been completed. A new class-container was added to the classifier for the purpose of extracting and labelling the church and chapel buildings found throughout the cultural heritage sites. Visual results can be seen in Figure 10. The new class-container was trained using the chapel and gatehouse building from the Adlington Cemetery memorial site, Figure 10 (a). Both buildings are found in Scene 1. The classifier was then applied to Scenes 3, 6, 9 and 11, all of which contain a building of interest. There was no significant difference to the metrics reported between this experiment and the previous out-of-sample experiments.

IX. DISCUSSION

In the within-sample experiments, the visual and quantitative results demonstrate that the GeoPart-XPCC framework was satisfactory in the task of learning from the memorial objects present in one part of a scene, and then autonomously extracting and correctly labelling memorial objects from the remainder. A clear example of this can be seen in Figure 8 (e & f), for Scene 7 from the Staines Cemetery cultural heritage site. A visual comparison between the operator-labelled point cloud and the GeoPart-XPCC predicted point cloud shows only minor differences between them. This assertion is further supported by the quantitative results, as the proposed framework achieved a score of 0.97 across all weighted average metrics for this scene. A further testament to the framework's ability to perform this task is from the within-sample results for Scene 1. The visual result for this scene shows that most of the memorial objects were found; even those nested deeply within vegetation.

The task of performing out-of-sample prediction is, admittedly, a more challenging one than within-sample prediction. Nevertheless, the results from the out-of-sample experiments illustrate GeoPart-XPCC's ability to accomplish this task without a significant reduction in accuracy. The average accuracy from the within-sample experiments was 0.93, whereas the average accuracy from the out-of-sample



(a) Scene 5 operator-Labelled point cloud

(b) Scene 5 GeoPart-XPCC predicted objects

(c) Scene 6 operator-Labelled point cloud

(d) Scene 6 GeoPart-XPCC predicted objects

(e) Scene 7 operator-Labelled point cloud

(f) Scene 7 GeoPart-XPCC predicted objects

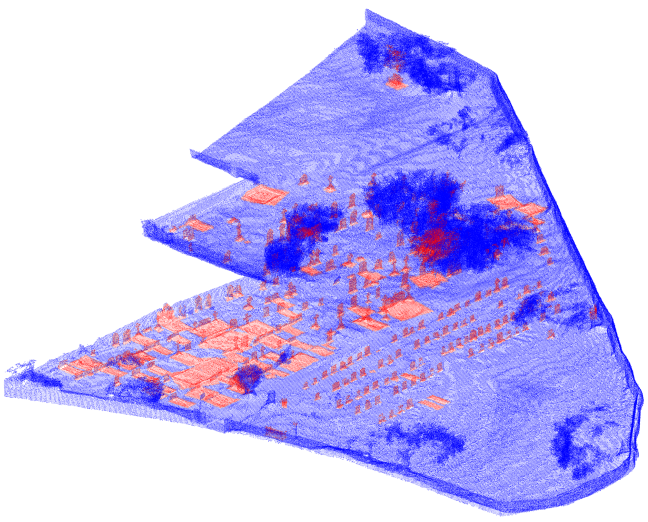
(g) Scene 3 operator-Labelled point cloud

(h) Scene 3 GeoPart-XPCC predicted objects

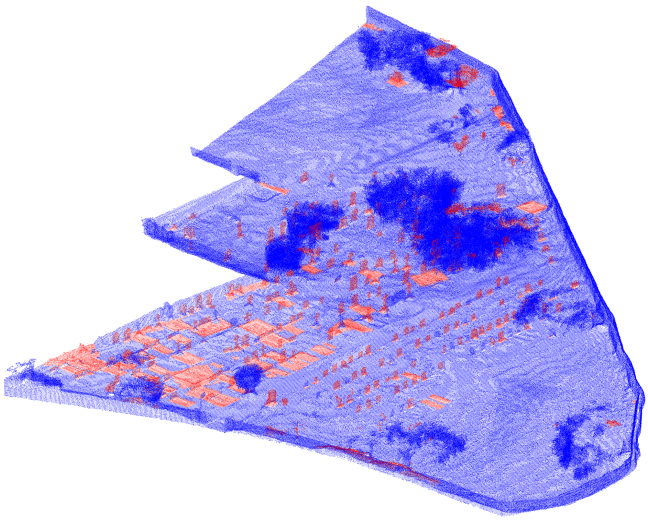
(i) Scene 4 operator-Labelled point cloud

(j) Scene 4 GeoPart-XPCC predicted objects

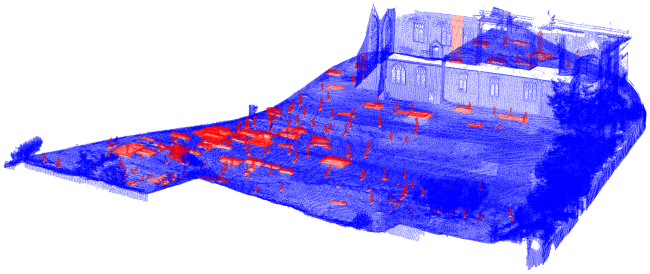
Fig. 8. Visual illustrations of point cloud Scenes 3, 4, 5, 6, and 7, from the out-of-sample experiments. Points belonging to the memorial objects are coloured in red; the non-memorial points are colored in blue. The sub-figures to the left of the centre line (a, c, e, g, and i) are the operator-labelled point cloud scenes. The sub-figures to the right of the centre line (b, d, f, h, and j) are the GeoPart-XPCC predicted point cloud scenes. The operator labelled and predicted illustrations of each scene are paired per row.



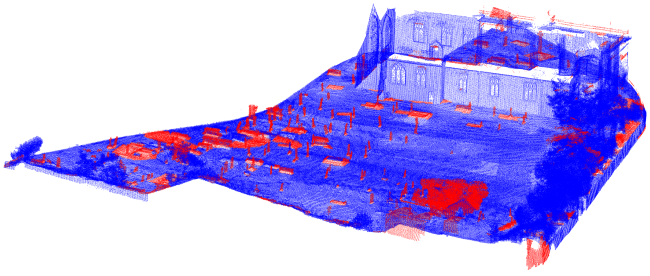
(a) Scene 8 operator-Labelled point cloud



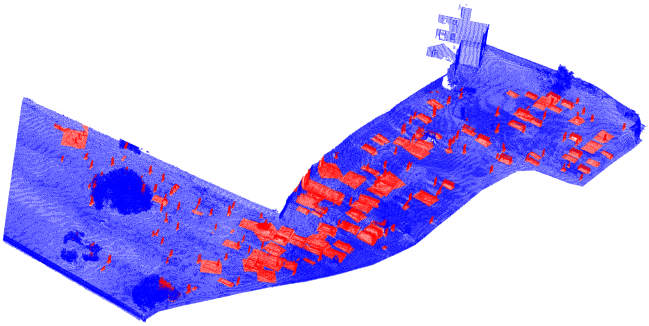
(b) Scene 8 GeoPart-XPCC predicted objects



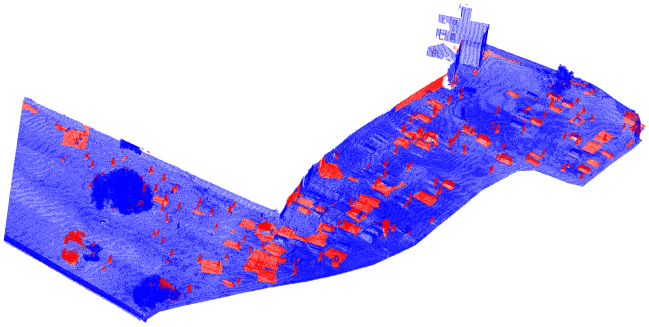
(c) Scene 9 operator-Labelled point cloud



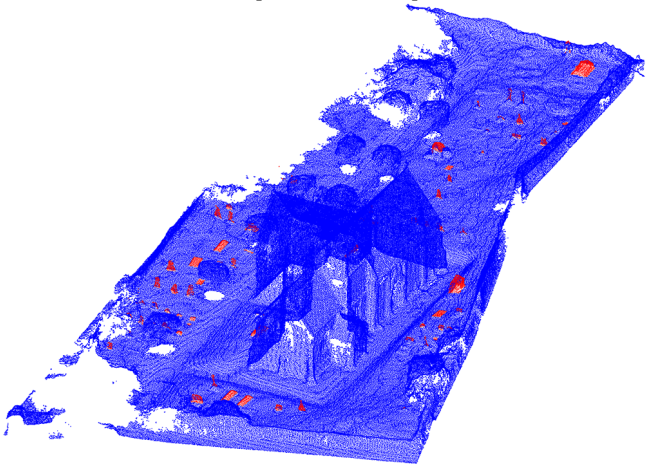
(d) Scene 9 GeoPart-XPCC predicted objects



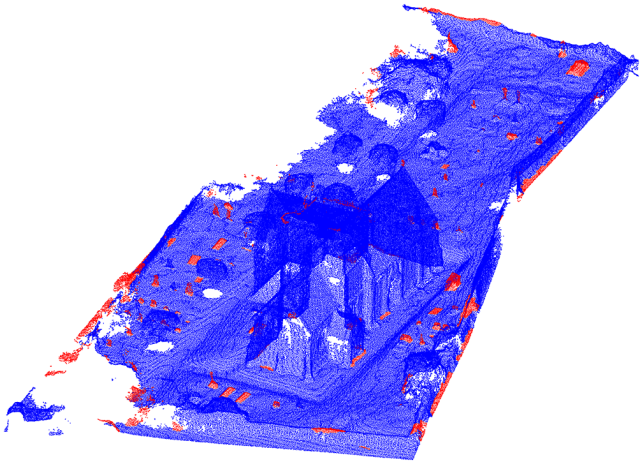
(e) Scene 10 operator-Labelled point cloud



(f) Scene 10 GeoPart-XPCC predicted objects



(g) Scene 11 operator-Labelled point cloud



(h) Scene 11 GeoPart-XPCC predicted objects

Fig. 9. Visual illustrations of point cloud Scenes 8, 9, 10, and 11 from the out-of-sample experiments. Points belonging to the memorial objects are coloured in red; the non-memorial points are colored in blue. The sub-figures to the left of the centre line (a, c, e, and g) are the operator-labelled point cloud scenes. The sub-figures to the right of the centre line (b, d, f, and h) are the GeoPart-XPCC predicted point cloud scenes. The operator labelled and predicted illustrations of each scene are paired per row.

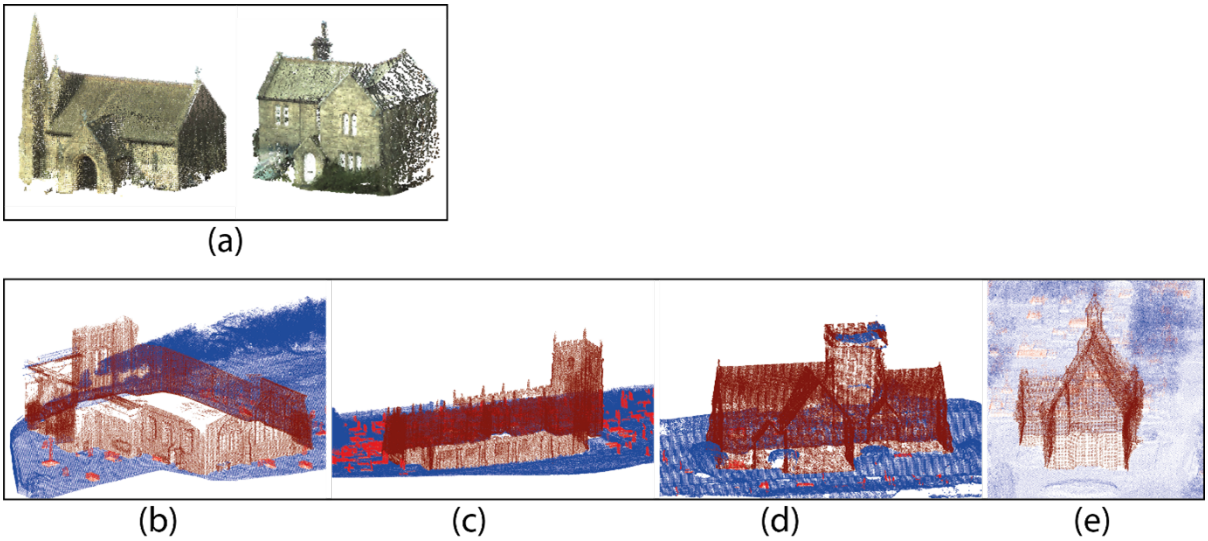


Fig. 10. Visual illustrations of the classification results after the GeoPart-XPCC framework was trained to identify Church and chapel buildings.

experiments was 0.92. Similarly, the average of the weighted F1-scores from the within-sample experiment was 0.94 as compared to 0.92 for the out-of-sample experiments.

The GeoPart-XPCC framework was also shown to be robust to unknown objects. In the within-sample experiment on the Church of St. Nicholas graveyard, the training scene, Scene 3, contains many examples of large trees, whereas the test scene has more examples of dense vegetation and medium-sized trees. Additionally, the training scene has a steeper sloped hill and contains fewer memorial objects. Regardless of this, the proposed framework achieved a weighted average F1-score of 0.93 for Scene 4, and the visual results, Figure 8 (c & d), show that the classifications were not adversely affected in most instances. The within-sample results for Scene 1 also highlight the robustness of the GeoPart-XPCC framework. Importantly, non-memorial objects such as cars, trees and much of the building infrastructure were not mistakenly classified as memorial objects. In terms of further applications of within-sample tasks, a human operator would need to label manually only a portion of a scene to apply the framework. This portion can contain fewer memorial objects and be spatially smaller than the remainder. As shown in the above, the framework would then be capable of classifying the rest of the scene, even if a difference in terrain or unknown objects is present.

From Table VI-E, it can be seen that the longest time period taken to perform the segmentation task was 19 minutes and 14 seconds. The time needed to train the classifier and perform the classification tasks was negligible. As such, the entire framework can be applied for under 20 minutes for a point cloud containing up to approximately 6 million points. Given these results, it is evident that the GeoPart-XPCC can reduce the human-hours needed to extract and label memorial objects. Multiple instances of GeoPart-XPCC could be run on a distributed architecture to process multiple point cloud scenes at once. Therefore, GeoPart-XPCC has the potential to be a valuable tool to increase productivity.

From the within-sample experiment on Adlington Cemetery, a visual examination of Scene 1 results, Figure 8 (a & b), illustrates some limitations of the proposed framework. The front door of the chapel building was miss-identified as a memorial object and, indeed, it is similar in appearance to the larger headstones. This type of miss-identification is difficult to deal with, as the definition of what is and what is not a memorial is subjective. The algorithm has, in fact, been taught that many objects that look like the door are memorial objects, and therefore the assumption that the door would be another object as well is understandable. Additionally, there are a few cases where the headstone of a memorial object was found correctly, but the plinth on which it sits was missed, or else, only part of it was identified. In this case, it can be said that the location and presence of the memorial object was correctly found, albeit only part of it. The ability to encode knowledge of spatial relationships between objects is a possible solution to address these limitations and is a topic of consideration for future research.

The ability to understand why a model arrived at a certain classification decision, and then correct the error, is invaluable for diagnostic purposes. It is clear from the visual results for Scene 8, from the St. Peter's Church Churchyard memorial site, that the framework wrongly classified several areas within the scene during the within-sample experiment. While most memorial objects were found, notably, parts of the perimeter wall and a small tree were also identified as false positives. A major benefit of the GeoPart-XPCC framework is the ability to examine the reasoning behind classification decisions. Doing so revealed that the prototypes most like those wrongly identified were examples of memorial objects very near to a wall or were surrounded by vegetation. By removing these prototypes, the GeoPart-XPCC framework was able to correctly identify the objects. The initial weighted F1-score of the within-

sample results on Scene 8 was a result of 0.91. After correction, the experimental weighted F1-score increased to 0.93, comparable with the results for Scene 4.

Spatially, the Adlington Cemetery cultural heritage site is located around 80 km away from both the Church of St. Nicholas graveyard and the St. Peter's Church Churchyard sites. It is around 290 km from the Churchyard of the Church of All Saints site, and 322 km from the Staines Cemetery site. In terms of composition, design and age, however, the Adlington Cemetery site is most like the Staines Cemetery site. The out-of-sample experiments, therefore, validate the proposed framework's ability to learn from one scene, and then be applied to scenes from different areas to that on which it has been trained. Furthermore, experimental results on Scene 11 indicate the ability of the framework to be applied to a scene acquired with a different scanning technology. Remarkably, for weighted scores, the framework achieved a 0.97 for F1-score, 0.96 for recall, 0.98 for precision. These high values may be attributed to the relative simplicity of the scene compared to the other cultural heritage sites. The above benefits of the GeoPart-XPCC framework are important in terms of further applications. More specifically, it shows that the method is transferable and the training scene does not need to be from the same area.

Of the cultural heritage sites used in this study, both Adlington Cemetery (Scenes 1 and 2, used for training in the out-of-sample experiments) and the St. Peter's Church churchyard (Scenes 8, 9, and 10) are located in the North West region of England. A comparison between the visual results from both the within-sample and out-of-sample experiments for Scene 8, seen in Figure 8 (h) and Figure 10 (f), reveals that several areas that were identified as false positives in the within-sample experiment were instead predicted as true negatives in the out-of-sample experiment. Quantitatively, the weighted average F1-score for the out-of-sample experiment was higher than in the within-sample experiment (0.93 and 0.91, respectively). The framework performs better for Scene 8 when trained on nearby Adlington Cemetery than when trained on a portion of St Peter's Church churchyard, highlighting similarities that must exist between the two cultural heritage sites. It may, therefore, be possible to use the GeoPart-XPCC framework to conduct analysis regarding the differences and similarities found amongst memorial objects between different regions. However, such a study is out of the scope of the current paper.

A significant advantage of the GeoPart-XPC framework is its adaptability. A demonstration of this is seen in Figure 11, illustrating the results of the experiment to extract and label the church and chapel buildings. Visual results show clearly that the buildings were labelled correctly. Training the updated model is a sequential (online) operation, where only the new samples are processed. With respect to further applications, this means that a user can add in new classes and class samples easily, without the model needing to be completely retrained.

Comparative results between the GeoPart-XPCC and the GeoPart-Transfer classification methods are shown in Table IV and Table VI for both the within-sample and out-of-sample experiments, respectively. Results for the GeoPart-Transfer were achieved using the MLP with sigmoid activation function method advocated by [12]. Preliminary experiments showed that the MLP-sigmoid method performed the best in terms of weighted average F1-score in comparison to the other methods discussed in that paper. Traditional DNN, such as the raw KP-CNN, cannot be sufficiently trained using the training data available within the experiments. As such, comparative results could not be achieved using these methods. In both experiments and across all scenes, the proposed GeoPart-XPCC achieved a higher weighted average F1-score and demonstrates its comparatively high accuracy relative to the MLP-sigmoid method. Notably, the latter method achieved a weighted average F1-score of 0.66 on the out-of-sample experiment for scene 11, compared to the score of 0.97 achieved by the proposed GeoPart-XPCC method. The MLP-sigmoid method results were likely affected by anomalies caused by the change in sensor technology between the training and testing data, whereas GeoPart-XPCC was better able to adapt. However, the exact reason for the MLP-sigmoid method having a decreased accuracy cannot be easily explained because of its nature as a black-box classifier. The same would be true for any traditional black-box DNN. In contrast, the similarity between the test objects and the prototypes within the GeoPart-XPCC classification model can be directly examined; thus, providing explanation.

X. Conclusion

Point clouds representing real world environments of cultural heritage sites provide a valuable data source for cultural heritage management. However, the manual extraction and labelling of memorial objects from these point clouds is a labour-intensive process. Therefore, it is of great significance to automate these tasks. To this end, this paper proposed a novel GeoPart-XPCC framework for the automatic extraction and labelling of memorial objects from 3D point cloud scenes. Experiments for both within-sample and out-of-sample tasks were performed on 11 point cloud scenes. The scenes were derived from mobile 3D LiDAR scans of five cultural heritage sites across the UK. The GeoPart-XPCC framework achieved high accuracy scores across these experiments. These results demonstrated that the framework could be trained on part of a scene and perform prediction on the remainder, as well as be trained on one scene and predict on a different scene. Comparative results show that GeoPart-XPCC is not only more accurate, but that it is more robust to changes in domains between training and testing data and provides diagnostic benefits by way of being explainable. This paper shows that the proposed framework is quick to train and apply, easily adaptable and widely applicable.

REFERENCES

[1] R. H. Bewley, S. P. Crutchley, and C. A. Shell, "New light on an ancient landscape: Lidar survey in the stonehenge world heritage site," *Antiquity*, vol. 79, no. 305, pp. 636–647, Sep. 2005.

[2] A. F. Chase *et al.*, "Airborne lidar, archaeology, and the ancient maya landscape at caracol, belize," *Journal of Archaeological Science*, vol. 38, no. 2, pp. 387–398, Feb. 2011.

[3] D. H. Evans *et al.*, "Uncovering archaeological landscapes at angkor using lidar," *Proceedings of the National Academy of Sciences*, vol. 110, no. 31, pp. 12 595–12 600, Jul. 2013.

[4] T. Inomata *et al.*, "Monumental architecture at aguada fénix and the rise of maya civilization," *Nature*, vol. 582, no. 7813, p. 7813, Jun. 2020.

[5] T. Freeland *et al.*, "Automated feature extraction for prospection and analysis of monumental earthworks from aerial lidar in the kingdom of tonga," *Journal of Archaeological Science*, vol. 69, pp. 64–74, May 2016.

[6] O. D. Trier and L. H. Pilø, "Automatic detection of pit structures in airborne laser scanning data," *Archaeological Prospection*, vol. 19, no. 2, pp. 103–121, 2012.

[7] R. Opitz and J. Herrmann, "Recent trends and long-standing problems in archaeological remote sensing," *Journal of Computer Applications in Archaeology*, vol. 1, no. 1, May 2018.

[8] I. Kramer, "An archaeological reaction to the remote sensing data explosion. reviewing the research on semi-automated pattern recognition and assessing the potential to integrate artificial intelligence," Ph.D. dissertation, 2015.

[9] K. Lambers, W. B. Verschoof-van der Vaart, and Q. P. J. Bourgeois, "Integrating remote sensing, machine learning, and citizen science in dutch archaeological prospection," *Remote Sensing*, vol. 11, no. 7, Jan. 2019.

[10] A. Guyot, L. Hubert-Moy, and T. Lorho, "Detecting neolithic burial mounds from lidar-derived elevation data using a multi-scale approach and machine learning techniques," *Remote Sensing*, vol. 10, no. 2, p. 225, Feb. 2018.

[11] M. Niculiță, "Geomorphometric methods for burial mound recognition and extraction from high-resolution lidar dems," *Sensors*, vol. 20, no. 4, Jan. 2020.

[12] N. Arnold, P. Angelov, T. Viney, and P. Atkinson, "Automatic extraction and labelling of memorial objects from 3d point clouds," *Journal of Computer Applications in Archaeology*, vol. 4, no. 1, Apr. 2021.

[13] J. Gallwey, M. Eyre, M. Tonkins, and J. Coggan, "Bringing lunar lidar back down to earth: Mapping our industrial heritage through deep transfer learning," *Remote Sensing*, vol. 11, no. 17, Jan. 2019.

[14] S. Weitman, "Using archaeological methods in cemetery surveys with emphasis on the application of lidar," 2012, [Online]. Available: https://digitalcommons.georgiasouthern.edu/etd/627.

[15] A. Jaklič *et al.*, "Volumetric models from 3d point clouds: The case study of sarcophagi cargo from a 2nd/3rd century ad roman shipwreck near sutivan on island brač, croatia," *Journal of Archaeological Science*, vol. 62, pp. 143–152, Oct. 2015.

[16] A. S. Z. Chase, D. Z. Chase, and A. F. Chase, "Lidar for archaeological research and the study of historical landscapes," in *Sensing the Past*, N. Masini and F. Soldovieri, Eds. Springer International Publishing, 2017, vol. 16, pp. 89–100.

[17] C. Sevara *et al.*, "Pixel versus object—a comparison of strategies for the semi-automated mapping of archaeological features using airborne laser scanning data," *Journal of Archaeological Science: Reports*, vol. 5, pp. 485–498, 2016.

[18] A. Fryskowska *et al.*, "Effective detection of sub-surface archeological features from laser scanning point clouds and imagery data," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W5, pp. 245–251, Aug. 2017.

[19] B. Kazimi *et al.*, "Deep learning for archaeological object detection in airborne laser scanning data," Sep. 2018.

[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[21] v. Kokalj and M. Somrak, "Why not a single image? combining visualizations to facilitate fieldwork and on-screen mapping," *Remote Sensing*, vol. 11, no. 7, Jan. 2019.

[22] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," *arXiv preprint arXiv:1612.00593*, 2016.

[23] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 5105–5114.

[24] S. A. Bello, S. Yu, and C. Wang, "Review: Deep learning on 3d point clouds," *arXiv preprint arXiv:2001.06280*, 2020.

[25] Q. Xu, X. Sun, C.-Y. Wu, P. Wang, and U. Neumann, "Grid-gcn for fast and scalable point cloud learning," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[26] L. Landrieu and M. Simonovsky, "Large-scale point cloud semantic segmentation with superpoint graphs," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4558–4567.

[27] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. Guibas, "Kpconv: Flexible and deformable convolution for point clouds," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 6410–6419.

[28] A. Boulch, "Convpoint: Continuous convolutions for point cloud processing," *Computers & Graphics*, vol. 88, pp. 24–34, May 2020.

[29] R. L. Figueroa, Q. Zeng-Treitler, S. Kandula, and L. H. Ngo, "Predicting sample size required for classification performance," *BMC Medical Informatics and Decision Making*, vol. 12, no. 1, p. 8, Feb. 2012.

[30] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, Jul. 2019.

[31] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, p. 9, May 2016.

[32] O. Trier *et al.*, "Semi-automatic mapping of cultural heritage from airborne laser scanning using deep learning," Apr. 2016.

[33] O. D. Trier, D. C. Cowley, and A. U. Waldeland, "Using deep neural networks on airborne laser scanning data: Results from a case study of semi-automatic mapping of archaeological topography on arran, scotland," *Archaeological Prospection*, vol. 26, no. 2, pp. 165–175, 2019.

[34] I. Zingman *et al.*, "Detection of fragmented rectangular enclosures in very high resolution remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4580–4593, Aug. 2016.

[35] W. B. V. der Vaart and K. Lambers, "Learning to look at lidar: The use of r-cnn in the automated detection of archaeological objects in lidar data from the netherlands," *Journal of Computer Applications in Archaeology*, vol. 2, no. 1, Mar. 2019.

[36] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, May 2019.

[37] N. I. Arnold, P. Angelov, and P. M. Atkinson, "An improved explainable point cloud classifier (xpcc)," *IEEE Transactions on Artificial Intelligence*, pp. 1–1, 2022.

[38] "About | opentopography," https://opentopography.org/about, accessed: Jun. 12, 2022.

[39] P. Hermosilla, T. Ritschel, P.-P. Vázquez, A. Vinacua, and T. Ropinski, "Monte carlo convolution for learning on non-uniformly sampled point clouds," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 6, pp. 235:1–235:12, Dec. 2018.

[40] S. Guinard and L. Landrieu, "Weakly supervised segmentation-aided classification of urban scenes from 3d lidar point clouds," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-1/W1, pp. 151–157, May 2017.

[41] M. Weinmann, A. Schmidt, C. Mallet, S. Hinz, F. Rottensteiner, and B. Jutzi, "Contextual classification of point cloud data by exploiting individual 3d neighbourhoods," in *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. II-3-W4, Mar. 2015, pp. 271–278.

[42] J. Demantké, C. Mallet, N. David, and B. Vallet, "Dimensionality based scale selection in 3d lidar point clouds," in *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVIII-5/W12, Sep. 2012, pp. 97–102.

[43] J. Niemeyer, J. D. Wegner, C. Mallet, F. Rottensteiner, and U. Soergel, "Conditional random fields for urban scene classification with full waveform lidar data," in *Photogrammetric Image Analysis*. Springer, 2011, pp. 233–244.

[44] L. Landrieu and G. Obozinski, "Cut pursuit: Fast algorithms to learn piecewise constant functions on general weighted graphs," *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1724–1766, 2017.

[45] Z. Wu *et al.*, "3d shapenets: A deep representation for volumetric shapes," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1912–1920.

[46] P. P. Angelov and X. Gu, "Toward anthropomorphic machine learning," *Computer*, vol. 51, no. 9, pp. 18–27, Sep. 2018.

[47] ——, *Empirical Approach to Machine Learning*. Springer International Publishing, 2019.
[48] "Digitising environments: A map of the future," https://leica-geosystems.com/en-GB/products/mobile-mapping-systems/stories-insights/digitising-environments, accessed: Jun. 13, 2022.

Chapter 5

Discussion

Historic, cultural heritage and archaeological sites stand as a reminder of those who came before us. In many ways, these sites represent the collective story of humanity, and the very concept of memorialising our loved ones is something fundamentally human. Burial grounds are a unique combination of culture, architecture, sculpture, art, landscape and wildlife. These sites are a finite resource and the loss of one, in whole or in part, risks permanent erasure from our understanding of the past, both as individuals and as a society. Thus, cultural heritage is evolving with the advent of new digitising technologies that enrich our ability to preserve these valuable resources. However, prior to this thesis, the cultural heritage community had limited understanding of how to apply modern techniques in ML and AI. As such, this thesis tackled head-on the methodological shortcomings in ML and AI applied to 3D point cloud data and provided solutions to overcome existing barriers to their application to cultural heritage management.

As interdisciplinary research, this thesis is a contribution to the application of ML and AI to digital cultural heritage and archaeology, as well as a contribution to ML and AI methods on 3D point clouds. As a complete body of work, it seeks to amplify the useability of point cloud data. This chapter discusses the implications and limitations of this research and how this thesis fits into the bigger picture. Additionally, this chapter also discusses potential directions for future studies and exploration.

5.1 Implications and Versatility

This thesis presents a methodology for the automatic extraction and classification of memorial objects from burial ground sites. It also presents the automatic classification of memorial objects, in general, directly in the 3D data space. Neither of these tasks have been achieved before. By extracting and labelling the objects within a site completely in 3D space, a unique catalogue and map of the objects is effectively

created. This has immediate benefits towards the preservation and documentation of cultural heritage sites. Firstly, there is the speed of processing. The GeoPart-XPCC framework can be applied in under 20 minutes to a point cloud containing approximately 6 million points. Furthermore, it could be possible to run multiple instances of GeoPart-XPCC on a distributed architecture; allowing multiple point cloud scenes to be processed at once. Given these results, it is evident that the research presented in this thesis can reduce the human-hours needed to extract and label memorial objects. Therefore, GeoPart-XPCC has the potential to be a valuable tool to increase productivity with respect to extracting and labelling memorial objects from point cloud data. Secondly, having a digital 3D map of a space allows for information to be linked with an object in a tangible way. This Information could be to do with maintenance schedules or locations of utilities, important persons or future planning. Moreover, it creates a digital record of the objects' physical appearance and location, which can be used for restoration purposes if an object becomes damaged or lost.

Machine learning classification of 3D point clouds is still in its infancy, and as such, there are many areas yet unexplored. Explainable models typically classify less accurately than their black box alternatives (see Appendix A). The ModelNet40 benchmark dataset was used to compare the XPCC method proposed in Chapter 3 with the state-of-the-art methods for both black box point set learning and explainable point set learning. The XPCC method is an explainable method that increases the classification accuracy over existing explainable methods, while still being comparable to the state-of-the-art black box methods by within 1-2 pp. This implies no considerable loss in classification accuracy by using the XPCC method, while still gaining the benefit of an explainable model, namely the improved interpretability of model classification decisions. The benefits of an explainable method pertaining to cultural heritage applications was discussed in Chapter 4: namely, the injection of diagnostic abilities and intuition into the model's performance. However, these properties also make the XPCC and GeoPart-XPCC methods suitable for applications where classification accuracy and explainablity are critical, such as self-driving vehicles and medical image analysis.

Benchmarking on performance metrics provides only half of the picture. Accordingly, the research and developments in this thesis provide benefits above performance. The discussion in Chapter 3 highlights the ability of the XPCC method to be transferable amongst applications that use 3D point cloud data. For example, the XPCC could be trained to classify street furniture, such as tables and chairs, in one setting, and then additively trained to classify headstones and crosses for another setting. There is no inherent reason within the model to remove the original two classes. Moreover, the entire process of adding or removing classes, or even individual objects, can be achieved without retraining the entire model. For example, if the

model was initially trained to extract headstones and stone grave slabs, but then became desirable to also extract crosses. To do so using traditional DNN architecture would require the entire network to be retrained. Whereas the XPCC method allows for the model to learn incrementally and requires only to see the new samples. The same is true if the model was trained to recognise one subset of crosses and it became desirable to recognise another (e.g., Celtic crosses). This means that the model can adapt to new tasks and applications; even as new developments or goals are set out. Thus, within the field of 3D point cloud data, the XPCC and its derivative GeoPart-XPCC, are impactful in their ability to be explainable, accurate and adaptable.

5.1.1 Applicability to Cultural Heritage Projects

The Church of England's National Burial Ground Survey project is a nationwide endeavour to digitally map all 19,000 Anglican burial grounds located in England over a seven-year period. Looking back at the out-of-sample and within-sample experiments from Chapter 4, the results presented there demonstrate that the GeoPart-XPCC framework is robust and flexible. Specifically, it can be trained on part of one scene and perform predictions on the remainder. Similarly, it can also be trained on one scene and perform predictions on a different scene. Furthermore, it is explainable, accurate in terms of classification predictions and robust to changes in domains between training and testing data. These characteristics make it quick to train and apply, easily adaptable and, therefore, widely applicable. This has important implications, as the point cloud burial ground datasets used in this thesis are only a small subset of burial grounds in the UK alone. The characteristics listed above, enable GeoPart-XPCC to be adaptable to new object classes and variations in objects so that it may be adaptable to any burial ground site regardless of where it is located. As such, this thesis is directly applicable to the Church of England's National Burial Ground Survey project.

5.2 Importance of Real-world Benchmarking

Chapter 3 mentions the lack of real-world benchmark datasets of 3D point cloud objects for classification purposes. Most of the datasets publicly available are generated synthetically (A. X. Chang et al., 2015). While these are useful for a proof-of-concept approach, they are not comparable regarding the level of challenge that is inherent to real-world data (Uy et al., 2019). There does exist some real-world 3D point cloud datasets, but very few of these are designed specifically for benchmarking classification methods (see Chapter 3, IV). As such, many of the related literature reviewed coming directly from the fields of AI and DL had a clear weakness: they were not vetted on a real-world classification benchmark. Therefore, it would be

highly beneficial to publish annotated real-world benchmarks. It would be of interest to produce a formal version of the graveyard object dataset that was produced as part of this thesis. This would require the datasets to be made public, however. As discussed in Chapters 2 and 5, the segmentation method can support the unsupervised generation of annotated training data. This could be used to enable the building of robust real-world datasets from other data sources.

5.3 Overcoming Vegetation

The presence of substantial vegetation in cultural heritage sites presented a significant challenge during the course of this thesis. To address this issue, various methods were investigated to filter out the vegetation from the point cloud scenes before the segmentation process. However, removing vegetation while preserving memorial objects proved to be a complex task and incurred substantial computational costs, making these methods impractical for the objectives of this research. As a result, the focus of the study shifted towards minimising the impact of vegetation noise on the classification process and developing a classification method that could handle the presence of vegetation within objects effectively. Chapters 4 and 5 detail the methodologies specifically devised to tackle the challenges introduced by vegetation in burial ground scenes. The developed approaches aimed to provide robust and accurate classification results even in the presence of dense vegetation, ensuring the reliable extraction and preservation of memorial objects from the point cloud data.

Comparative results between the GeoPart-XPCC and the GeoPart-Transfer methods are shown in Chapter 4 Tables 3 and 5 for both the within- and out-of-sample experiments, respectively. These results highlight the improvements made in Chapter 4 over those presented in Chapter 2. Specific to the challenge of minimising the noise introduced by vegetation, Chapter 2 used a k-nearest (k being an integer) neighbourhood search in its segmentation method. In comparison, a significant change to the Chapter 4 segmentation method was the adoption of a k-radius (k being a distance). The radius neighbourhood search is more robust when applied in a non-uniform sampling setting (Hermosilla et al., 2018), as is typical of mobile LiDAR datasets. This makes sense, as the nearest points may be any distance from a query point, and it is only guaranteed that those neighbours are the closest according to the metric. By introducing the requirement of points falling within a certain radius, the points that are closer to the query point are effectively given a larger weight. The result is that the neighbourhood around the query point is better represented. This means that segments are more tightly packed and, therefore, include less points taken from nearby objects, such as vegetation.

5.3.1 Impact of Segmentation Neighbourhoods on GeoPart-Transfer and GeoPart-XPCC

In addressing the potential concerns about the comparability of the two methods due to the introduction of a different segmentation approach, it is crucial to recognise that the overall goal was to enhance the effectiveness of the segmentation process and subsequently improve the accuracy of the classification results. The implementation of the k-radius approach, which offered improved robustness to the presence of vegetation, allowed both GeoPart-Transfer and GeoPart-XPCC to be more adaptive and successful in handling complex and diverse cultural heritage scenes.

It is important to note that the GeoPart-Transfer segmentation was performed using the updated segmentation method utilised by GeoPart-XPCC, which employed a k-radius neighbourhood search, instead of the k-nearest neighbourhood search used in the earlier GeoPart-Transfer method. This change was made to address the challenges posed by the presence of large amounts of vegetation within the cultural heritage scenes, as discussed in Chapter 4. As such, the GeoPart-Transfer and GeoPart-XPCC classification methods were compared fairly, and both were assessed using the same benchmark datasets. While the changes in the segmentation approach may have introduced some variability in the results, they were essential to address the challenges posed by vegetation in the scenes, which are common in cultural heritage environments.

Table 5.1: Comparative benchmark for within-sample accuracy evaluating k-NN in segmentation.

Scene Name	GeoPart-Transfer (k-NN)	GeoPart-XPCC
Scene 1	0.82	0.88
Scene 4	0.74	0.91
Scene 7	0.80	0.91
Scene 8	0.66	0.87

The comparison of GeoPart-Transfer and GeoPart-XPCC using the k-nearest neighbors (k-NN) and k-radius search, respectively, is presented through two benchmark results tables: Table 5.1 for within-sample metrics and Table 5.2 for out-of-sample metrics. These tables provide valuable insights into the impact of the segmentation neighbourhood approach on the performance of the classifiers.

In Table 5.1, the weighted average F1-scores for both GeoPart-Transfer and GeoPart-XPCC are displayed for several scenes. Notably, the scores for GeoPart-XPCC are consistently higher across all scenes compared to GeoPart-Transfer. This difference in performance can be attributed to the adoption of the k-radius approach, which offers improved robustness to the presence of vegetation and other

Table 5.2: Comparative benchmark for out-of-sample accuracy evaluating k-NN in segmentation.

Scene Name	GeoPart-Transfer (k-NN)	GeoPart-XPCC
Scene 3	0.78	0.90
Scene 4	0.61	0.84
Scene 5	0.80	0.91
Scene 6	0.75	0.89
Scene 7	0.72	0.90
Scene 8	0.69	0.88
Scene 9	0.51	0.78
Scene 10	0.71	0.85
Scene 11	0.60	0.93

complex objects within the cultural heritage scenes. As a result, GeoPart-XPCC demonstrates superior adaptability and success in handling diverse and challenging scene characteristics.

Similarly, Table 5.2 further underscores the advantages of the k-radius search used by GeoPart-XPCC. The F1-scores for both methods are again shown for different scenes, and once more, GeoPart-XPCC outperforms GeoPart-Transfer in all cases. This consistent trend demonstrates the effectiveness of the k-radius approach in the GeoPart-XPCC classifier.

Overall, the comparison presented in these two tables emphasizes the significance of employing the k-radius search in GeoPart-XPCC, which has led to notable improvements in the accuracy and robustness of the classifier when dealing with real-world cultural heritage data. The choice of segmentation neighbourhood has played a crucial role in enhancing the effectiveness of the classifiers, making the GeoPart-XPCC method a preferred choice for accurate and reliable classification tasks in 3D point cloud data.

5.4 Limitations and Future Directions

Despite the significant advancements and contributions made in this thesis, it is essential to acknowledge the limitations inherent in the proposed methods. While the GeoPart-XPCC framework demonstrates remarkable accuracy and interpretability in the classification of 3D point cloud data, there are specific challenges that warrant consideration. The primary limitation lies in the ability of the XPCC method to distinguish between physically similar objects with different semantic meanings, a scenario that is not uncommon in real-world applications. Additionally,

the classification of classes with overlapping characteristics, such as distinguishing between benches and sofas, remains a challenge. Moreover, the effectiveness of the methods may be influenced by variations in point cloud data quality, particularly in scenes with large amounts of vegetation, where noise introduced by the vegetation can impact classification results. These limitations, while addressed to the best extent possible within the scope of this thesis, provide valuable opportunities for further research and improvement to ensure the applicability and robustness of the methods in a wider range of real-world scenarios.

The project also assumes that high-quality and accurately annotated ground truth data are available for training and evaluation purposes. Gathering such data for cultural heritage sites can be challenging, and the accuracy of the classification results heavily depends on the quality and representativeness of the training datasets. Moreover, the point cloud data should be of sufficient quality and resolution to capture the intricate details of memorial objects and the surrounding environment. Low-quality or sparse data may lead to reduced classification performance.

5.4.1 Contextual Linking of Prototypes

One possible route for exploration is to link prototypes together by learning meta-characteristics that encode information about relationships between objects in the scene. For instance, the model could learn that headstones are typically located on a segment belonging to the 'ground' class type and that they may or may not appear near trees or flowers. By utilising contextual information, the model could improve classification accuracy. An approach for this could be based on graphical learning, leveraging techniques from (Bronstein, Bruna, et al., 2017). Such an adaptation aligns with the anthropomorphic approach to machine learning, where context clues aid in predictions (P. P. Angelov and Gu, 2018). Moreover, contextual linking of prototypes opens the possibility of achieving panoptic segmentation, which combines instance segmentation (detecting and masking each distinct object of interest) with semantic segmentation (providing a label for each point in the scene). Panoptic segmentation offers a more comprehensive understanding of the scene (Kirillov et al., 2019).

5.4.2 Hierarchical Classifications and X-Class Discovery

Future research should explore hierarchical classifications, particularly object part segmentation, to produce point-level semantic annotations indicating micro-level parts of an object, complementing macro-level class annotations. Embedding XPCC's 'class-containers' within each other would enable micro-level object part classifications as sub-classes of the macro-level classification, offering a more detailed representation. The embeddings could be extended to the CPCs, allowing variable-grain visualisation

of regions contributing to the classification decision. Additionally, enabling X-class discovery (P. Angelov and E. Soares, 2021) would permit the XPCC classifier to learn entirely new classes or subclasses of objects during the prediction phase, reducing errors arising from unknown objects in the domain.

5.4.3 Cross-Site Comparisons and Multi-Modal Inputs

Another intriguing application involves comparing differences and similarities between cultural heritage and archaeological sites from different locations based on prototypical representations. Discovering similarities between sites could aid in cases where no training data is available, using previously discovered object classes as proxies to train the feature extraction model. Exploring methods for enabling XPCC to perform classification from multi-modal inputs (e.g., 2D images and 3D point clouds) would allow for inspections on sites not scanned in 3D, broadening the range of potential applications.

5.4.4 Linking Metadata to Memorials

As mentioned in earlier discussions, the linkage of memorials to the information they record is of significant importance in the context of cultural heritage preservation and management. While the focus of this thesis has primarily been on the automatic extraction and classification of memorial objects from 3D point cloud data, the potential for linking external information to the models is an intriguing area for future investigation.

One potential approach to achieve this linkage is through the integration of metadata within the training and classification process. Metadata could include information such as historical context, inscription details, and other relevant attributes associated with the memorial objects. By incorporating this external information as additional input features during the training phase, the models can learn to associate the geometric characteristics of the memorial objects with the corresponding metadata, thus enabling the linkage between the objects and their recorded information.

Moreover, as advancements in natural language processing (NLP) and knowledge graph technologies continue to progress, there could be an opportunity to leverage these methods to link external information to the models post-hoc. For instance, by using NLP techniques, textual information from inscriptions or historical records could be extracted and semantically analysed. The resulting knowledge graphs could then be integrated into the models, enabling the models to retrieve relevant contextual information for each detected memorial object during the classification process

Another potential avenue for achieving linkage is through the adoption of multi-modal learning techniques. By incorporating multiple data modalities, such as images,

textual information, or contextual data, into the classification process, the models can be enriched with a more comprehensive understanding of the memorial objects and their associated information.

However, it is essential to acknowledge that integrating external information into the models also presents challenges. Ensuring data consistency, accuracy, and availability of external information is crucial to prevent any potential biases or misinterpretations. Moreover, privacy and data ethics considerations should be addressed when dealing with sensitive historical information.

While the current research has primarily focused on the automatic extraction and classification of memorial objects from 3D point cloud data, the potential for linking external information to the models offers an exciting direction for future research. By incorporating metadata, leveraging NLP and knowledge graph technologies, and exploring multi-modal learning approaches, the models can become more than just classifiers; they can serve as intelligent tools to preserve, enrich, and link cultural heritage information to memorial objects, contributing to a more comprehensive understanding and appreciation of our shared past.

5.4.5 Expanding to Aerial LiDAR Datasets

This thesis presents the use of fine-resolution mobile LiDAR for cultural heritage management, as was discussed in detail in Chapter 4. An appealing application for future research would be to apply the GeoPart-XPCC method to discovering objects of interest from aerial LiDAR datasets, as such data are readily available from sources such as the USA National Science Foundation (NSF) funded OpenTopography project. Doing so would allow for many more applications to be explored that otherwise do not have the same fine-resolution mobile LiDAR datasets. An obvious extension of this research would be to look for prehistoric burial mounds or tumuli within landscapes. However, the range of applications would not need to be limited to only cultural heritage and archaeology. It could also be possible to map geographic features, crops or industrial building surveys, for example.

5.4.6 Future Research Outside of Cultural Heritage

This research has primarily focused on the analysis of memorial objects in cultural heritage scenes, but its implications extend far beyond this domain. The methodological advancements made in this thesis are applicable to a wide range of applications, some of which have been previously discussed. One such application lies in autonomous robotic systems that utilise LiDAR for positional awareness and environment interaction, exemplified by Boston Dynamics' Spot and Atlas platforms (Bouman et al., 2020).

The methods developed and presented in this thesis offer more than just a means of classifying burial ground objects in 3D point cloud data. The XPCC approach, in particular, holds the potential for broader applicability in various non-cultural heritage contexts. Although the primary focus of this research is on cultural heritage site mapping, the underlying principles and advantages of the XPCC classifier can be extrapolated to other domains where object detection and classification from point cloud data are essential.

One potential application lies in urban planning and infrastructure management. Point cloud data obtained from LiDAR scans or photogrammetry can be leveraged to identify and classify various urban elements, such as buildings, roads, sidewalks, street furniture, and vegetation. The XPCC classifier's ability to capture and represent object prototypes can enable efficient and interpretable classification of these elements, facilitating the development of accurate 3D models for urban planning and infrastructure analysis. Furthermore, the ability to quickly add and subtract classes means that the model can be better adapted to suit new applications.

Another area where the XPCC approach could prove valuable is in environmental monitoring and natural resource management. Remote sensing techniques, including airborne and satellite-based LiDAR, are increasingly used to study forests, wetlands, and other natural landscapes. By applying the XPCC classifier to point cloud data derived from these environments, it may be possible to detect and classify various vegetation types, track changes in vegetation density, and assess ecosystem health and biodiversity.

Additionally, the XPCC classifier's interpretable nature can be advantageous in safety-critical applications, such as autonomous vehicles and robotics. Point cloud data obtained from sensors on these platforms can be analysed using the XPCC approach to identify and classify obstacles, road markings, and traffic signs, enabling safe and reliable navigation in complex environments.

It is essential to acknowledge that these potential applications are speculative at this stage and require real-world investigation and validation. Conducting experiments in diverse environments and datasets beyond cultural heritage sites would be necessary to assess the generalisation and performance of the XPCC classifier in non-cultural heritage contexts.

Despite this speculative nature, the advantages offered by the XPCC approach in terms of interpretability, efficiency, and performance make it a promising candidate for various object detection and classification tasks in point cloud data. As the field of 3D point cloud analysis continues to advance, exploring and adapting the XPCC approach to other domains could lead to significant advancements in a wide range of applications, beyond cultural heritage mapping, further expanding the impact and relevance of this research.

5.5 Bridging the Gap for Non-Expert Users

The development of a GUI for partitioning and classifying objects in a point cloud is essential to facilitate the accessibility and usability of the proposed methods within this research thesis (Opitz and Herrmann, 2018). Point cloud analysis, particularly object segmentation and classification, poses unique challenges that demand specialised expertise and technical knowledge, making it challenging for non-technical end users to adopt and apply the developed methodologies effectively. By incorporating a user-friendly GUI, the complex processes involved in object partitioning and prototype-based classifier interpretation can be streamlined and made accessible to a broader audience. Appendix B contains an example of one such GUI, called the *GP-XPCC app*. The GUI can empower non-technical users, such as archaeologists, heritage site managers, and researchers, to leverage the full potential of the methods developed herein, enabling them to make informed decisions based on the results obtained.

Outlined here is the design considerations and features of the *GP-XPCC app*, emphasising its role in bridging the gap between technical expertise and practical applicability, ultimately broadening the reach and impact of this research.

1. **Enhanced Usability for Non-Technical Users:** The primary motivation behind incorporating a GUI is to democratise the use of object partitioning and classification methods for non-technical end users. The *GP-XPCC app* abstracts complex technicalities, allowing users to execute partitioning and classification tasks efficiently without the need for specialised training or extensive domain knowledge. This includes: performing unsupervised object extraction and segmentation via GeoPart-Transfer, train the XPCC classifier, perform object classification and inspection of the trained XPCC classifier.

2. **Iterative Refinement and Feedback:** With the GUI's aid, non-technical end users can engage in an iterative process of refinement, evaluating and validating the partitioning and classification results. The GUI allows users to provide feedback, correct misclassifications, and refine the training data, thereby improving the overall accuracy and adaptability of the methods over time.

3. **Interactive Visualisation and Manipulation:** Through the *GP-XPCC app*, users can manipulate the partitioned object prototypes, visualise their latent spatial relationships, and adjust classification labels, providing a sense of control and confidence in the outcomes. The ability to manually edit and manipulate prototypes within the classifier enhances the user's interpretability and trust in the system's decisions as well as error correction.

4. **Program Efficiency:** The GUI application was developed using C++ as its foundation, supplemented with an embedded Python3 interpreter, prioritising efficiency at its core. To optimise Python code execution, a significant emphasis was placed on incorporating just-in-time compilation and parallel processing. By leveraging Python's multiprocessing module and the Numba Python package, computationally expensive Python-based operations, such as for-loops, were accelerated to enhance overall performance. This optimisation was particularly crucial, give the computationally demanding nature of point cloud data processing tasks, which could be time-consuming with conventional serial processing in Python. When possible, operations were deferred C++ source code. To further boost efficiency and enable support for both CPU and GPU acceleration, the PyTorch library was integrated into the application. These considerations make it possible to perform tasks efficiently even on machines that do not have GPU accelerated hardware.

5. **Improved Interpretability of Prototype-based Classification:** The inclusion of the XPCC in the GUI facilitates the interpretation of object classifications. Users can inspect the prototype samples representing different object categories, gaining insights into the classifier's decision-making process. Such interpretability is crucial, especially when dealing with sensitive or critical applications where the understanding of classifier outcomes is essential for informed decision-making.

6. **Integration with Standard Point Cloud Viewers:** The GUI omits direct visualisation of point clouds, segmentation masks, and classification results. In a previous version, these features were present, but they were later excluded to leverage the use of more robust and widely supported point cloud viewer applications, such as CloudCompare, for this purpose. As a result, the *XPCC-GUI app* exports its results into formats compatible with these external visualisation tools. One exception to this approach is the basic visualisation window within the *GP-XPCC app*, which allows users to view prototype objects directly within the application. This feature was kept as to minimise the need to switch applications when interacting with the XPCC classifier.

The inclusion of a well-designed GUI for the proposed methods not only simplifies the usage for non-technical users but also empowers them to actively participate in the classification process, making it more intuitive, interpretable, and versatile. This integrated approach contributes to the effectiveness and widespread adoption of the developed techniques for object partitioning, classification, and interpretability in point cloud data analysis. Future research or work could be done to further develop this application in a robust UI computer application; this would allow for the research to be adopted by a wider audience.

5.6 Synthesis of Framework Benefits

Throughout the thesis, various alternative frameworks, including the GeoPart-Transfer and GeoPart-XPCC methodologies, have been discussed in detail, showcasing their relative advantages and limitations. However, it is important to provide a concise and coherent synthesis of this information to help readers better understand the unique contributions and benefits of the XPCC-based frameworks.

The XPCC based frameworks stand out for the innovative use of class-containers, which are prototypical representations capturing essential characteristics of object classes. This approach offers several key benefits. Firstly, the use of class-containers enables a compact and interpretable representation of object classes, making the XPCC classifier highly transparent and explainable. Secondly, the XPCC framework demonstrates remarkable adaptability, allowing for incremental learning and the addition of new classes without retraining the entire model. This flexibility is advantageous in real-world scenarios where datasets are subject to change and expansion.

Another notable strength of the XPCC frame lies in its ability to handle complex and diverse point cloud scenes with large amounts of vegetation, common in cultural heritage sites. The introduction of the k-radius segmentation approach ensures robustness to the presence of vegetation, enhancing the classifier's accuracy in such challenging environments.

5.7 Reflective Assessment of Framework Adoption

While the benefits of the XPCC-based frameworks have been demonstrated through rigorous experimentation, a more reflective assessment of its adoption by other users is essential to foster wider acceptance and utilisation. The potential user base for the XPCC framework extends beyond the academic and research community to various stakeholders, including cultural heritage experts, practitioners, hobbyists, and industries involved in 3D point cloud analysis.

To facilitate the adoption of the XPCC framework, it is important to address potential barriers and challenges. Firstly, the implementation and integration of the XPCC classifier and the GeoPart-XPCC framework into existing workflows and software environments may require technical expertise and effort. Thus, providing user-friendly documentation, tutorials, and software libraries and tools could assist potential users in adopting the framework more effectively. For example, releasing the XPCC classifier as a Python package compatible with industry standard packages such as Scikit-Learn, PyTorch, and Keras would enhance its accessibility and ease of integration.

Selecting an appropriate classification framework for 3D point cloud data analysis

involves considering various factors, including dataset characteristics, classification requirements, and interpretability needs. The XPCC classifier offers unique features that can be advantageous in specific scenarios. To aid users in identifying when and how to apply the XPCC classifier effectively, the following general guidelines are proposed:

- **Dataset Characteristics** The first step in deciding whether the XPCC classifier is suitable for a given application is to assess the dataset characteristics. Consider the complexity and diversity of objects present in the point cloud data. The XPCC classifier is particularly beneficial in handling scenes with objects of varying shapes, sizes, and orientations. Additionally, if the dataset contains scenes with significant vegetation or clutter, the XPCC's robustness to noise introduced by vegetation makes it a strong candidate for classification in such environments.

- **Classification Requirements** The second aspect to consider is the specific classification requirements of the application. If interpretability and explainability are essential, the XPCC classifier's interpretable nature provides valuable insights into the reasoning behind classification decisions. This feature is particularly useful in applications where human-readable explanations of model predictions are vital for decision-making and analysis.

  Moreover, the XPCC classifier's adaptability allows for easy incorporation of new object classes. If the dataset requires continuous updates or expansion to accommodate new object categories, the XPCC classifier's ability to quickly adapt to changes offers a significant advantage.

- **Interpretability Needs** In cases where interpretability is crucial for user trust and acceptance, the XPCC classifier's prototype-based approach can be highly beneficial. Prototypes provide visual representations of object classes, enhancing the model's transparency and enabling users to understand the reasoning behind classification results. Additionally, the use of the CPC can be used to highlight commonalities amongst the classified objects. If the application domain requires clear and interpretable classification outcomes, the XPCC classifier should be considered.

- **Performance and Efficiency** Efficiency is an important consideration, particularly for large-scale or real-time applications. The XPCC classifier's efficiency in terms of training and prediction time allows for rapid analysis of point cloud scenes. For applications requiring timely and accurate results, the XPCC classifier's performance can be advantageous.

In conclusion, the decision to apply the XPCC classifier should be based on a thorough assessment of dataset characteristics, classification requirements, and interpretability needs. Its adaptability, interpretability, and robustness to noisy environments make it well-suited for diverse applications. By carefully considering these guidelines, users can leverage the strengths of the XPCC classifier to achieve efficient and accurate 3D point cloud data analysis in their specific domains.

5.8 Additional Discussion by Chapter

This section provides an opportunity to delve into aspects specific to each chapter that could not be fully explored within the scope of the original papers. These discussions encompass valuable insights, supplementary analyses, and further context that enhance the comprehensiveness of the research presented in this thesis. Due to word limits in the original papers or considerations of maintaining a focused narrative, certain aspects had to be omitted from the main discussions. By offering a more comprehensive examination of aspects from each chapter, this section enriches the understanding of the methodologies and their applications, providing readers with a deeper insight into the research findings and discoveries. The chapter-specific discussions aim to highlight noteworthy points that contribute to the overall significance and impact of the presented work, enhancing the value and relevance of this research thesis.

5.8.1 Chapter 2 Discussion

5.8.1.1 Detection of Prone Gravestones

The automatic detection and classification of ground-level objects, specifically prone gravestones, present both successes and challenges in this research. Prone gravestones, due to their similar appearance to grave slabs, are detected fairly well since both objects largely exhibit planar regions. The detection process can accurately identify planar surfaces, leading to satisfactory recognition of these ground-level objects. However, a notable limitation arises from the lack of distinction between prone gravestones and grave slabs within the training data. Consequently, the current classification model does not differentiate between the two, resulting in the misclassification of prone gravestones as grave slabs. This indicates that while the detection of planar regions is effective, further improvements are necessary to distinguish between these specific types of ground-level objects. Addressing this issue is crucial to enhance the accuracy and specificity of the classification model, allowing for more precise identification and labelling of prone gravestones in cultural heritage sites represented as 3D point cloud scenes.

5.8.1.2 Inclusion of Text or Carvings

Incorporating text or other carvings into the object classification process presents an intriguing avenue for further enhancing the XPCC classifier's capabilities. The current methodology focuses on the extraction and labelling of memorial objects based on geometric shape characteristics, which has proven effective for various cultural heritage objects, including headstones and grave markers. However, text and carvings on these objects hold valuable contextual and historical information that can enrich the classification results and contribute to a more comprehensive understanding of the site.

To integrate text and carvings into the XPCC classifier, one approach could involve manually adding prototypes representing different types of inscriptions or carvings. These prototypes would act as representative points for the corresponding classes, capturing the visual features and characteristics associated with different types of text or carvings. By incorporating such prototypes, the XPCC classifier could learn to recognise and distinguish various inscriptions and carvings present on memorial objects, further refining its object classification abilities.

Manually adding prototypes would enable the classifier to learn from a limited number of annotated examples, significantly reducing the need for extensive training data. Moreover, it would facilitate interpretability, as the prototypes can be visually inspected and modified by domain experts to ensure accuracy and alignment with the site's historical and cultural context.

However, it is essential to consider the challenges that come with integrating text and carvings into the classification process. The diverse and intricate nature of inscriptions and carvings on memorial objects may require specialised techniques for feature extraction and representation. Additionally, the integration process should be approached with caution, as it may introduce complexities in the decision-making process and potential biases if not handled appropriately.

Nonetheless, the inclusion of text and carvings holds significant promise in providing a more comprehensive analysis of cultural heritage sites. By enriching the XPCC classifier with this additional information, the research can contribute to a more detailed and nuanced understanding of memorial objects' significance and historical context, further advancing the preservation and documentation of cultural heritage for future generations. Further investigations and experiments in this direction could unlock valuable insights and expand the classifier's applicability in broader archaeological and heritage management contexts.

5.8.1.3 Rationale for Selecting a Graph-Cut Method

The choice of graph-cut as the approach for object extraction and segmentation in this task is well-justified due to its effectiveness in handling complex and interconnected

structures within 3D point cloud scenes. Graph-cut techniques have been extensively applied in computer vision and image processing, and their adaptability to point cloud data has shown promising results in various applications (Landrieu and Obozinski, 2017; Landrieu and Simonovsky, 2017; Simonovsky and Komodakis, 2017).

One key advantage of graph-cut is its ability to leverage both local and global information in the data, enabling the incorporation of contextual cues and spatial relationships among neighbouring points. In the context of cultural heritage site mapping, this is particularly relevant as the point cloud scenes often contain a multitude of interconnected objects, such as headstones, grave markers, and other structural elements. Graph-cut's capability to model the relationships between neighbouring points allows for more accurate and coherent segmentation of individual objects, ensuring that each object is delineated precisely (Landrieu and Simonovsky, 2018).

Furthermore, graph-based approach is well-suited for handling the inherent noise and irregularities often present in point cloud data (W. Hu, Gao, et al., 2019; Pistilli et al., 2020). As 3D scanning processes can introduce imperfections and outliers, robustness to noise is critical for achieving reliable object extraction results. Graph-cut's formulation as an energy minimisation problem enables it to effectively handle noise and uncertainties, leading to more robust and stable object segmentation.

The interpretability of graph-cut results is another compelling aspect of this approach. Graph-cut algorithms provide clear and intuitive outputs, with explicit delineation of object boundaries and segmentation masks. This interpretability is crucial in cultural heritage applications, where the extracted objects' accuracy and visual representation are of great importance for archaeological analysis and documentation.

While graph-cut demonstrates several advantages for object extraction in point cloud data, it is essential to acknowledge its limitations. In some cases, graph-cut may struggle with handling more complex scenes, where object shapes are intricate, or object boundaries are not well-defined. Additionally, graph-cut's computational complexity can be significant, which may become a challenge when dealing with large-scale point cloud scenes or resource-constrained environments.

Notwithstanding these limitations, the suitability of graph-based methods for the task of segmentation is evident, given its robustness, interpretability, and ability to capture both local and global contextual information. By effectively segmenting and extracting cultural heritage objects from 3D point cloud scenes, graph-cut contributes significantly to the broader objectives of this research, enabling precise and reliable preservation and documentation of cultural heritage sites. The successful application of graph-cut further validates its viability as an appropriate approach for object extraction in the context of archaeological studies and cultural heritage management.

5.8.1.4 Clarification on Training Data

In the Evaluation of Methodology on Cultural Heritage Scenes in Chapter 2, it is important to clarify that both the training set and Scene 1 are derived from a larger point cloud of the same cultural heritage site. The training set comprised labelled data representing various objects of interest within the site. However, these subsets are carefully selected from distinct spatial regions, ensuring that there is no overlap in data points between the two scenes. The training set was specifically used to perform a within-sample test on Scene 1, assess the methodology's performance on familiar ground, where the algorithm had access to labelled examples similar to those encountered during training. Subsequently, an out-of-sample test was conducted on scene 2, which represents a completely different cultural heritage site. Scene 2 was not part of the training set and presented unseen objects and environmental conditions. This out-of-sample evaluation provided a robust assessment of the methodology's generalisation capability and its ability to adapt to novel contexts, demonstrating its potential applicability beyond the training data and the specific cultural heritage site used for training.

5.8.1.5 SVM Kernel Choice

The decision to not use the kernel trick in the Support Vector Machine for the experiments performed in Chapter 2 is grounded in empirical results and the objective of feature descriptor discrimination. A comprehensive set of experiments was conducted, comparing the performance of different kernel functions, including linear, polynomial, radial basis function, and sigmoid kernels. Notably, the linear kernel exhibited superior accuracy compared to the other tested kernels. While these results were not explicitly reported in the paper, they played a pivotal role in shaping the SVM approach. The choice of the linear SVM was thus deemed appropriate, as it demonstrated the capability to effectively differentiate the global feature descriptors without the need for non-linear transformations. Utilising the linear SVM avoids the computational overhead associated with the kernel trick and emphasises the practicality of the linear kernel's ability to discern and discriminate feature representations accurately, contributing to the overall efficiency and effectiveness of the methodology.

5.8.2 Chapter 3 Discussion

5.8.2.1 Distinctiveness of the XPCC method

The XPCC method introduced in Chapter 3 is a novel approach to object classification in 3D point cloud data, setting it apart from other explainable methods explored in

Chapter 3. Unlike alternative techniques, XPCC acts as an object classification head for a pre-trained neural network, effectively replacing the final classification layer used during training. This unique design enables the seamless integration of interpretable prototype-based classification within the existing network architecture, distinguishing it from methods that function as explainable feature extraction networks. In contrast, PointHop adopts a local-to-global attribute building process, creating new attributes that consider the relationships between multiple points, and leverages the Saab transform for dimension reduction to control attribute growth.

Both XPCC and PointHop demonstrate interpretability, but their scopes and levels of explainability vary. XPCC's interpretability primarily lies in the visual inspection of prototypes, facilitating users' comprehension of the learned representations and decision boundaries. However, it may not offer a complete explanation of the feature extraction process, as it operates on top of the extracted features from the pre-trained network. Similarly, PointHop provides interpretability as a feature extractor, allowing insights into feature visualisation on the original input point cloud. Nonetheless, it does not entirely explain the classification decisions made by the subsequent classification process.

Despite this distinction, the significance of explainable feature extractors such as PointHop should not be underestimated. Their ability to extract and represent local-to-global attributes in an interpretable manner enhances understanding of feature extraction from 3D point cloud data. Combining the capabilities of explainable feature extractors with the XPCC classifier represents a promising avenue for future research. For instance, applying XPCC on top of the PointHop architecture could potentially provide more comprehensive insights into object classification decisions, further enhancing the overall interpretability and performance of the classification system. This avenue of exploration holds promise in leveraging the strengths of both XPCC and explainable feature extractors, ultimately contributing to the advancement of explainable and effective object classification in 3D point cloud data.

5.8.2.2 Data Selection

Chapter 3 of this thesis deviates from using cultural heritage data and instead focuses on benchmarking datasets to emphasise the methodological aspects of the proposed approach. The rationale behind this decision is to provide a comprehensive evaluation and validation of the method in controlled and well-established settings, ensuring its robustness and applicability across various scenarios beyond cultural heritage contexts. By concentrating solely on benchmarking datasets, Chapter 3 aims to showcase the XPCC method's generalisation capability and its effectiveness in handling diverse point cloud scenes, devoid of any specific domain bias.

The decision to focus on benchmarking datasets aligns with the aim of establishing

the proposed method as a versatile and widely applicable solution for object detection and classification in 3D point cloud data. Utilising benchmarking datasets allows for direct comparisons with state-of-the-art methods, objectively evaluating the performance of the proposed approach against well-known and established baselines. This facilitates a fair assessment of the method's strengths and weaknesses, showcasing its competitive edge and potential advantages in a broader context.

Moreover, benchmarking datasets serve as a common ground for researchers and practitioners in the field of 3D point cloud analysis. By demonstrating the effectiveness of the proposed method on benchmarking datasets, the research contributes to the collective knowledge and understanding of the state-of-the-art in object detection and classification, fostering advancements in the field at large.

Despite not using cultural heritage data in Chapter 3, its relevance to the thesis as a whole remains crucial. The chapter serves as a solid foundation for the subsequent research in chapter 4 that focus on the application of the proposed method to cultural heritage sites. By establishing the method's effectiveness and generalisability on benchmarking datasets, Chapter 2 builds confidence in the method's capabilities, setting the stage for its successful adaptation and implementation in the unique and complex environments of cultural heritage sites.

The decision to focus on benchmarking datasets in Chapter 3 was a strategic choice to emphasise the methodological aspects of the proposed approach and demonstrate its versatility beyond cultural heritage contexts. By providing a comprehensive evaluation and validation, the chapter establishes the method's robustness and applicability, contributing to the advancement of 3D point cloud analysis as a whole. Ultimately, this approach ensures that the XPCC method is well-equipped to address the specific challenges and requirements of cultural heritage site mapping discussed in subsequent chapters, further solidifying the thesis's overall contribution to the field of 3D point cloud analysis and cultural heritage preservation.

5.8.2.3 Ablation Study

The ablation study results presented in Table 5.3 showcase the incremental contributions of different elements to the performance of the XPCC method in the context of two baseline models: KPConv and PointNet++. For the KPConv baseline, incorporating 3D feature extraction with the XPCC-base (comprising only the local and global similarity layers followed by *argmax*) yielded a marginal accuracy decrease of 0.84 percentage points, revealing the nuances of prototype-based representation adaptation. Notably, the incorporation of the Compound Prototype Cloud resulted in a substantial accuracy enhancement of 0.61 percentage points, signifying its role in capturing object data density nuances for refined classification. The subsequent inclusion of the SoftMax layer contributed an additional accuracy

Table 5.3: Ablation Study of XPCC Method.

Experiment	Accuracy (%)
Baseline (KPConv)	91.8
– 3D Feature Extraction + XPCC-base	90.96 (-0.84)
– Compound Prototype Cloud	91.57 (+0.61)
– SoftMax Layer	91.82 (+0.25)
Baseline (PointNet++)	90.7
– 3D Feature Extraction + XPCC-base	89.62 (-0.93)
– Compound Prototype Cloud	92.05 (+1.04)
– SoftMax Layer	92.18 (+0.13)

boost of 0.25 percentage points suggesting improved class separation for enhanced object classification within 3D point clouds.

Similarly, within the PointNet++ baseline, the integration of XPCC-base with 3D feature extraction led to a slight accuracy decrease of 0.93 percentage points, reiterating the importance of feature representation. The assimilation of the Compound Prototype Cloud showcased a notable accuracy improvement of 1.04 percentage points, highlighting its ability to enable the model to grasp class-specific attributes, enhancing classification accuracy. Additionally, the introduction of the SoftMax layer generated a modest accuracy increase of 0.13 percentage points, reaffirming the influence of activation mechanism adjustments on classification performance.

In conclusion, the ablation study outcomes accentuate the incremental roles of individual components in shaping the XPCC method's classification accuracy. The Compound Prototype Cloud emerges as an especially impactful addition, spotlighting the model's capacity to leverage object- and class-oriented explanations for augmented performance. This study imparts valuable insights into the significance of different components and their collective impact on the XPCC method's resilience and interpretability.

5.8.2.4 XPCC as a Classification Head Layer for Neural Networks

The XPCC classifier has demonstrated its effectiveness in classifying point cloud objects in various applications, as seen in the XPCC paper and the GeoPart-XPCC framework. Building upon this success, the XPCC classifier can also serve as an effective classification head layer for neural networks in the domain of deep learning. In this context, a classification head layer refers to the final layer of a neural network responsible for making the class predictions based on the features extracted by the

preceding layers. By incorporating the XPCC classifier as a classification head layer, neural networks can benefit from its interpretability, transfer learning capabilities, and robustness to domain shifts, holds promise for extending its capabilities to diverse domains.

When used as a classification head layer, the XPCC classifier can be employed in transfer learning scenarios, where it is fine-tuned on specific datasets to adapt to new object classes or domains. This transfer learning approach allows the XPCC classifier to leverage its prototypical representations learned from one dataset and efficiently classify objects in new datasets without starting the training process from scratch. This adaptability is especially valuable in scenarios with limited annotated data, enabling the model to generalise effectively to new object classes or scenes.

The XPCC classifier's interpretable nature is a key advantage when used as a classification head layer in deep learning and explainable deep learning frameworks. The feature extraction process based on prototypical representations provides a transparent and intuitive understanding of how the model makes decisions. This interpretability is crucial for applications where model decisions must be explained and justified, such as in critical systems, healthcare, and autonomous vehicles. By using the XPCC classifier as the classification head layer, deep learning models can inherit this explainable characteristic, enhancing the transparency and trustworthiness of the entire pipeline.

Moreover, the XPCC classifier's efficient and lightweight design makes it suitable for real-time applications, even in resource-constrained environments. In deep learning frameworks, the XPCC classification head layer can be combined with other layers to build end-to-end pipelines for point cloud analysis. The ability to perform classification directly from point cloud data within the deep learning model streamlines the workflow and avoids the need for additional complex preprocessing steps, enhancing the overall efficiency and performance of the system. By extending the use of the XPCC classifier as a classification head layer in deep learning, researchers can explore its capabilities across various domains beyond cultural heritage sites.

5.8.3 Chapter 4 Discussion

5.8.3.1 Impact of Scene Features on Classification Scores

The variations in scores across different scenes in our study are attributed to the distinct characteristics of each scene, as well as the specific strengths and limitations of the classification methods employed. While it may be challenging to explain all differences between scenes and classification methods, several generalisable observations can be made.

Firstly, the scene-specific features play a crucial role in influencing the classification

scores. Each scene contains unique variations in object density, point cloud resolution, object size, and spatial distribution, among others. These variations can impact the performance of the classification methods differently. For instance, scenes with dense and cluttered point clouds may pose challenges for some methods that rely heavily on local point features, resulting in lower classification accuracy compared to scenes with sparse and well-separated objects.

Secondly, the complexity and diversity of object classes present in each scene can influence the classification performance. Scenes with a wide range of object categories and variations may require classifiers with stronger generalisation capabilities and the ability to handle intra-class variances effectively. The GeoPart-XPCC method's prototype-based classification has demonstrated advantages in handling such complexities, as it can better capture the essence of object classes and handle variations within categories.

While it is challenging to provide a complete explanation for all differences between scenes and classification methods, the observations made above provide insights into the factors contributing to variations in classification scores. These findings underscore the importance of carefully considering scene-specific features and selecting appropriate classification methods based on the characteristics and requirements of the 3D point cloud data. By understanding these nuances, researchers and practitioners can make informed decisions in choosing the most suitable approach for object classification in different real-world scenarios.

5.8.3.2 Unidentified Memorial Objects

The results of the object extraction process in Chapter 2 demonstrated a high success rate in detecting and locating memorial objects within the 3D point cloud scenes. The method performed well in identifying objects even when they were deeply nested within vegetation, indicating its effectiveness in handling challenging scenes with occlusions and clutter.

However, a few instances of objects were not detected, and this can be attributed to certain limitations of the segmentation process. In some cases, the segmentation algorithm was overly aggressive, resulting in the partitioning of objects into smaller fragments, which were then not recognised as complete memorial objects. This issue highlights the importance of striking a balance between accurate segmentation and preserving the contextual information necessary for proper classification.

Furthermore, in areas with significantly lower point density, some memorial objects were missed during the detection process. The lower density may have resulted in incomplete or sparse representations of objects, making them difficult to distinguish from the background noise. This emphasises the significance of data quality and completeness in achieving accurate and comprehensive object extraction results.

Therefore, while the object extraction method demonstrated high accuracy in detecting memorial objects within the 3D point cloud scenes, certain challenges remain, such as aggressive segmentation leading to fragmented object representations and lower point density affecting object detection. These limitations suggest opportunities for future research and development of more advanced and context-aware segmentation techniques, as well as the consideration of data quality and density in improving object detection performance.

5.8.3.3 Ablation Study

In this ablation study, we assess the contributions of key elements to the performance of the GeoPart-XPCC method using the baseline classification GeoPart-Transfer (k-NN + ConvPoint + MLP-σ). We investigate the impact of incorporating two distinct components: KPCONV and XPCC. The table below presents the accuracy values achieved by the GeoPart-XPCC method and its ablated versions. The conducted ablation study was based on the accuracy averaged over a 10-fold cross-validation using all point cloud scenes as provided in Chapter 4.

Table 5.4: Ablation Study of GeoPart-XPCC Method.

Experiment	Accuracy (%)
Baseline (GeoPart-Transfer)	80.50
– k-NN + KPConv + MLP-σ	82.48 (+1.98)
– k-NN + ConvPoint + XPCC	82.60 (+2.10)
– k-NN + KPConv + XPCC	84.05 (+3.55)
– Radius-NN + ConvPoint + MLP-σ	86.95 (+6.45)
– Radius-NN + KPConv + MLP-σ	87.11 (+6.61)
– Radius-NN + ConvPoint + XPCC	87.80 (+7.30)
– Radius-NN + KPConv + XPCC	88.61 (+8.11)

In the ablation study presented in Table 5.4, the contributions of different elements to the performance of the GeoPart-XPCC method were systematically evaluated. The baseline accuracy achieved using the GeoPart-Transfer method is 80.50%. By gradually introducing enhancements, consistent improvements in accuracy can be observed across the experiments.

Starting with k-NN + KPConv + MLP-σ, there is an observed performance improvement of 1.98 percentage points. . This result suggests that the utilization of KPConv, a robust kernel point convolution operation, aids in better feature extraction and representation, enhancing the classification performance.

Moving on to k-NN + ConvPoint + XPCC, results in an accuracy increase of 2.10 percentage points. This demonstrates that the incorporation of the XPCC method

enhances object classification through explainable neural networks. This improvement reinforces the importance of leveraging the XPCC model to achieve accurate and interpretable results.

By introducing k-NN + KPConv + XPCC, a boost of 3.55 percentage points over the baseline is observed. This emphasises the combined strength of KPConv and XPCC, which together provide an effective means of handling complex point cloud data and extracting meaningful features for improved classification accuracy.

The transition to the use of the Radius-NN, as found in Radius-NN + ConvPoint + MLP-σ leads to a remarkable increase of 6.45 percentage points in accuracy. It should be noted that the GeoPart-Transfer baseline is analogous to a combination of k-NN + ConvPoint + MLP-σ. This observation underlines the significance of integrating and transition from k-NN to Radius-NN in the segmentation process.

Continuing with the Radius-NN + KPConv + MLP-σ configuration, we observe a notable accuracy improvement of 6.61 percentage points. Notably, by introducing the Radius-NN method alongside KPConv, this combination yields an additional increase of 4.63 percentage points compared to the same configuration without Radius-NN. This emphasises the pivotal role of KPConv in facilitating precise feature extraction, thereby significantly enhancing the overall performance of the model.

Moving to Radius-NN + ConvPoint + XPCC, we see a substantial accuracy increase of 7.30 percentage points. This emphasises the successful integration of XPCC with the ConvPoint module, which improves classification accuracy by effectively capturing spatial relationships within the point cloud.

Finally, the combination of Radius-NN + KPConv + XPCC yields the highest accuracy of 88.61%, representing an 8.11 percentage point increase over the baseline, a more than 10% increase to performance. Moreover, this configuration showcases a 0.81 percentage point augmentation compared to the ConvPoint feature extractor. This result validates the effect of incorporating KPConv and XPCC in the method, demonstrating their collective capability in achieving superior object classification performance.

The results underscore the substantial impact of individual components on the overall performance of the GeoPart-XPCC method. The addition of KPConv and Radius-NN, along with the choice of classification models, contributes significantly to the accuracy gains observed in each configuration. These findings provide valuable insights for designing and optimising the GeoPart-XPCC method for the extraction of cultural heritage objects from point cloud data.

Chapter 6

Conclusion

This thesis makes distinct contributions in two main aspects: the application of the developed methodologies to cultural heritage sites and the technical contributions in the form of novel methodologies. In terms of the application, the thesis showcases the practical application of point clouds to burial ground sites, demonstrating the extraction of memorial objects from these 3D point cloud scenes. Additionally, it presents the successful implementation and evaluation of the GeoPart-Transfer methodology, which allows for the accurate and efficient classification of memorial objects in point cloud data. Furthermore, the thesis introduces and evaluates the GeoPart-XPCC framework, which builds upon the GeoPart-Transfer method and incorporates the XPCC classifier to provide an explainable classification approach for cultural heritage sites.

On the other hand, the technical contributions of the thesis lie in the formulation and development of the novel methodologies. Specifically, the thesis introduces the GeoPart-Transfer methodology, a point cloud segmentation and classification approach that leverages the concept of transfer-learning for improved object representation and classification without the reliance on 2D imagery. Additionally, the thesis presents the XPCC classifier, which extends the traditional point cloud classification by incorporating the concept of prototype-based class-containers, enabling enhanced interpretability and explainability of the classification decisions. Finally, the GeoPart-XPCC framework represents a significant technical contribution, as it combines the strengths of the GeoPart-Transfer and XPCC methods, offering an integrated and robust solution for accurate and interpretable classification in cultural heritage scenes.

It is essential to recognise that while the technical contributions lay the foundation for broader applications beyond cultural heritage sites, the application-specific contributions demonstrate the practicality and significance of the methodologies in a real-world context. By focusing on cultural heritage, the thesis showcases the potential of the methods and algorithms introduced in it to contribute meaningfully

to the preservation, documentation, and analysis of historical sites, contributing to the broader field of cultural heritage management and archaeology.

By delineating the contributions in terms of the application and technical methodologies, the thesis effectively separates the achievements related to the cultural heritage domain from the novel technical advancements, providing a clear and comprehensive understanding of the contributions made to both the specific application and the broader field of point cloud analysis.

This thesis provides solutions to several research questions towards the development of methods for the extraction and classification of objects from 3D point clouds.

1. **How to create an automated solution for the extraction and classification of memorial objects from 3D Point cloud representations of cultural heritage sites?**

   This thesis addresses this question with the introduction of the GeoPart-Transfer methodology. GeoPart-Transfer is an approach wherein ML and AI techniques are used to automate the processes of segmentation and classification of objects from 3D point clouds. Using the XPCC, a prototype-based explainable deep neural network for performing classifications on 3D point cloud data, the GeoPart-XPCC was also developed as a framework tailored to the extraction of memorial objects from scenes of 3D point cloud scenes of burial grounds that is adaptable, explainable and efficient. Thus, the methods developed in this thesis provide fast and scalable automated solutions that are applicable to the extraction of memorial objects from 3D scans of cultural heritage sites.

2. **How to manage object occlusions?**

   The methods developed in this thesis address this question by operating directly on the 3D point clouds. In doing so, all decision-making takes place within the 3D data space or in higher latent feature spaces. In these data spaces, the concept of a directional view of an object is non-existent. Thus, the occlusions appearing in lower dimensional representations, such as 2D views of the data, are eliminated.

3. **How to apply ML directly to the 3D point cloud data?**

   The GeoPart-Transfer, XPCC and GeoPart-XPCC methods introduced in this thesis solve this problem with the use of a transfer learning approach. Point-based ANNs that are capable of learning directly on the point cloud structure without intermediate representations are pre-trained and used as fixed global feature extractors. Benchmarking exercises in Chapter 2 established that this approach of using transfer learning global descriptor achieved a statistically higher accuracy, with a 22.5% average percentage difference in F1-scores,

compared to hand-crafted feature descriptors. Similarly, the proposed approach to classification achieved increases of roughly 9% and 6% compared with seven traditional classification methods applied to both artificial and real-world datasets, respectively.

4. **Related to both the second and third research questions, how to apply ML in a context (i.e., cultural heritage management) with little training data?**

   To this end, the transfer learning approach introduced in this thesis serves two purposes. The point based ANNs used as fixed global feature extractors were first pre-trained on a large dataset of generic objects. Once trained, these feature extractors produce vectors that define a feature space optimised for the separation of the training objects. Results in Chapters 2, 3, and 4 demonstrated that this feature space enables the GeoPart-Transfer, XPCC, and GeoPart-XPCC methods to use the pre-trained feature space to separate new objects (i.e., memorial objects) without the need for an extensive amount of domain specific training data. It was shown in Chapter 2 experiments that classification can be performed effectively with only tens of objects per class. Moreover, the unsupervised segmentation process can be used to accelerate the generation of annotated ground reference data from the cultural heritage sites.

5. **How to perform these tasks on sloping and hilly burial ground scenes?**
   How to discover ground-level objects (i.e., grave kerbs)?
   The fifth and sixth research questions are application-led, and both answered using the same solution. A multi-scale partitioning scheme was introduced in this thesis. This method checks if the large partitions, seen as one homogeneous object at the global level, are also homogeneous at various local scales. The strength of the segmentation algorithm is inverse to the scale, that is, the smaller the sale, the more aggressive the segmentation is. This method ensures that even flat low-lying objects obscured in earthen surfaces are found. The approach was introduced in the GeoPart-Transfer segmentation process and later expanded on in the GeoPart-XPCC segmentation process, with the adoption of a hybrid radius neighbourhood search. The combination of the multi-scale partitioning and the new neighbourhood definition ensures that burial grounds located on hills can be processed. Both Scenes 3 and 4 are from a memorial site located on the side of a hill. The GeoPart-XPCC method achieved a classification accuracy of 93% on the within-sample experiment for Scene 4 and accuracies of 91% and 95% for Scenes 3 and 4, respectively, for the out-of-sample experiment. This, provides a solution to the sixth research question.

6. **How to correctly extract and classify memorial objects above the state-of-the-art?**

   This is one of the central aims of this thesis, namely, in comparison with the leading non-explainable point-set deep learning classifier, XPCC was found to be within a 0.4% percentage difference in terms of accuracy. Furthermore, the XPCC & PointNet++ classifier achieved a 3.3% increase in accuracy over the base PointNet++ network. Comparative results show that GeoPart-XPCC is, on average, 14% more accurate and more robust to changes in domains between training and testing data compared to GeoPart-Transfer. Experiments in Chapter 4 were performed at scale, on 11 point cloud scenes derived from mobile LiDAR scans of five cultural heritage sites across England. The algorithm achieved accuracy scores above 90% on all scenes tested for the within-sample experiment results. Similarly, the algorithm achieved accuracy scores above 90% in all, but one of the out-of-sample experiments. These experimental results demonstrated that the methods developed in this thesis are computationally efficient, trainable on thousands of samples in seconds, and explainable, thus, capable of supporting accurate and informed decision-making.

7. **How to adapt to different objects, classes and domains of data?**

   The XPCC is algorithmically and structurally transparent and offers several layers of human-interpretable explainability. Moreover, it can learn continuously, without the need to be completely retrained at the addition of new classes or class samples. Experiments in Chapter 4 demonstrated that the GeoPart-XPCC framework can be trained efficiently on one scene and perform prediction on a different one, and can also be trained on part of a scene and perform prediction on the remainder. Experiments showed that the algorithm is flexible to process data from different scanning technologies in addition to scenes and objects with varying point densities. This thesis also presents a technique to visualise the explainable aspects of the XPCC model called CPC. The technique is unique to 3D point clouds and prototype-based learning. The CPC represents what the model has learnt, and specifically identifies the object regions which influence classification decisions. The GeoPart-XPCC introduces the ability to perform diagnostics and allows users to make informed decisions to update the model manually to suit changes in the application.

In summary, this completed thesis presents the design and implementation of a solution for the automatic extraction of 3D objects from cultural heritage sites. The methods presented within this thesis increase performance above the state-of-the-art in respect to classification accuracy, explainablity, adaptability and speed of

implementation. The methods have the potential to significantly decrease the hours needed for an operator to perform the tasks of mapping and labelling cultural heritage objects. Furthermore, they do so in a transparent and human understandable way that is robust to changes in the data sources and input domain. This enables the models to be meaningfully updated to better fit the application. Thus, this thesis contributes to the knowledge of object extraction from 3D point clouds, as well as contributes to the field of cultural heritage management. As evident by the further research proposed, this thesis has very wide applicability and many avenues for further progression in knowledge.

Appendix A

Explainable artificial intelligence: an analytical review

Received: 31 January 2021 | Revised: 15 April 2021 | Accepted: 10 June 2021

DOI: 10.1002/widm.1424

OVERVIEW

WIREs DATA MINING AND KNOWLEDGE DISCOVERY WILEY

Explainable artificial intelligence: an analytical review

Plamen P. Angelov[1,2] | **Eduardo A. Soares**[1,2] | **Richard Jiang**[1,2] | **Nicholas I. Arnold**[1,3] | **Peter M. Atkinson**[2,3]

[1]School of Computing and Communications, Lancaster University, Lancaster, UK

[2]Lancaster Intelligent, Robotic and Autonomous Systems (LIRA) Research Centre, Lancaster, UK

[3]Lancaster Environment Centre, Lancaster University, Lancaster, UK

Correspondence
Plamen P. Angelov, School of Computing and Communications, Lancaster University, Lancaster LA1 4WA, UK.
Email: p.angelov@lancaster.ac.uk

Edited by: Sushmita Mitra, Associate Editor and Witold Pedrycz, Editor-in-Chief

Abstract

This paper provides a brief analytical review of the current state-of-the-art in relation to the explainability of artificial intelligence in the context of recent advances in machine learning and deep learning. The paper starts with a brief historical introduction and a taxonomy, and formulates the main challenges in terms of explainability building on the recently formulated National Institute of Standards four principles of explainability. Recently published methods related to the topic are then critically reviewed and analyzed. Finally, future directions for research are suggested.

This article is categorized under:

Technologies > Artificial Intelligence

Fundamental Concepts of Data and Knowledge > Explainable AI

KEYWORDS

black-box models, deep learning, explainable AI, machine learning, prototype-based models, surrogate models

1 | INTRODUCTION

Artificial intelligence (AI) and machine learning (ML) have demonstrated their potential to revolutionize industries, public services, and society, achieving or even surpassing human levels of performance in terms of accuracy for a range of problems, such as image and speech recognition (Mnih et al., 2015) and language translation (Young et al., 2018). However, their most successful offering in terms of accuracy—deep learning (DL) (LeCun et al., 2015)—is often characterized as being "black box" and opaque (Pasquale, 2015; Rudin, 2019). Indeed, such models have a huge number (many millions or even a billion) of weights (parameters) which are supposed to contain the information learned from training data. Not only is the number of these weights very large, but their link to the physical environment of the problem is extremely hard to isolate. This makes explaining such forms of AI to users highly problematic. Using opaque, "black box" models is especially problematic in highly sensitive areas such as healthcare and other applications related to human life, rights, finances, and privacy. Since, the applications of advanced AI and ML, including DL, are now growing rapidly, encompassing the digital health, legal, transport, finance, and defense sectors, the issues of transparency and explainability are being recognized increasingly as critically important. For example, a search in Google Trends (https://trends.google.com/trends/) reveals that in the last decade publications using the terms "DL" and "explainable AI" (XAI) both grew significantly, but while the curve for DL is now in a saturation stage over the last 3 years or so, the curve for XAI is growing exponentially starting precisely 3 years ago when the saturation in regards to

DL started to be observed as illustrated by Figure 1. This is not coincidental and demonstrates that as the huge increase in interest towards DL starts to saturate, interest towards XAI is gearing up trying to address open research and acceptability questions (Arrieta et al., 2020). In this paper, this trend is analyzed and, further, a taxonomy of XAI is provided. Moreover, we also investigate novelties in terms of XAI and include an analysis involving the Caltech-101 benchmarking dataset. Following a brief historical and critical state-of-the-art review we also consider some applications for which explainability is critically important. The paper concludes with a discussion.

2 | BRIEF HISTORICAL PERSPECTIVE

AI was closely linked to both ML and to logic and symbolic forms of reasoning from its inception in the middle of the 20th century (Samuel, 1959; Smolensky, 1987). ML and data-driven statistical techniques gained momentum in recent years due to an unprecedented increase in the number and complexity of data available (now the majority of data are unstructured, with many more images/videos as well as text/speech in comparison to the 20th century) (Bishop, 2006; Goodfellow et al., 2014).

Historically, the first methods of AI, such as decision trees (Quinlan, 1990), symbolic AI (Smolensky, 1987), expert systems, fuzzy logic, and automated reasoning (Robinson & Voronkov, 2001), as well as some forms of artificial neural networks (ANNs), for example, radial-basis function (RBF) architectures and linguistic, prototype-based, representations were significantly more interpretable and self-explainable than the more recent and more efficient forms such as support vector machines (SVMs) (Hearst et al., 1998) and most other forms of ANNs.

In the last few years, explainability has become an important issue not only for scientists, but also for the wider public including, regulators, and politicians. As AI and ML (and, especially, DL) become more wide spread and intertwined with human-centric applications, and algorithmic decisions become more consequential to individuals and society, attention has shifted back from accuracy to explainability (Angelov & Soares, 2020; Core et al., 2006; Pedreschi et al., 2019). Complex and "black box" (Pasquale, 2015; Rudin, 2019) types of models can easily fool users (Nguyen et al., 2015) and, in turn, this can lead to dangerous or even fatal consequences (Stilgoe, 2020). Opening the "black box" is critically important not only for acceptability within society, but also for regulatory purposes. (In 2019 the US Congress passed the Algorithmic Accountability Act (MacCarthy, 2019) and the EU enshrined the right for an explanation to the consumer (Core et al., 2006; Goodman & Flaxman, 2017; Pedreschi et al., 2019).) The current data-rich environment brought the temptation to take shortcuts from raw data to solutions using a very large number of abstract, purely numerical parameters (Angelov & Soares, 2020; Rudin, 2019; Stock & Cisse, 2018), without providing a deep insight into, and understanding of, the underlying dependencies, causalities, and internal model structures. The issue of explainability is an open research question for some of the most successful (in terms of accuracy) forms of ML such as SVMs, DL, and many of the ANNs (Bishop, 2006), as Figure 2 illustrates.

In the above context, the main question is not so much: *Can we get an XAI solution?*, but *Can we get a highly accurate XAI solution comparable to the accuracy that DL would provide*? Table 1 illustrates some results for the Caltech-101

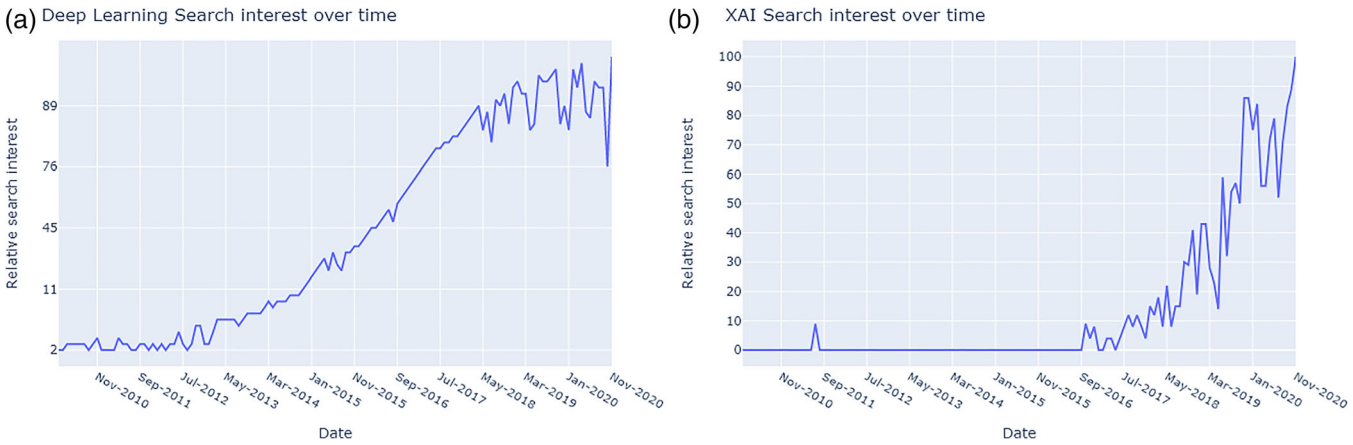


FIGURE 1 Illustrates the interest evolution towards two terms according to Google Trends: (a) deep learning (DL), (b) explainable artificial intelligence (XAI)

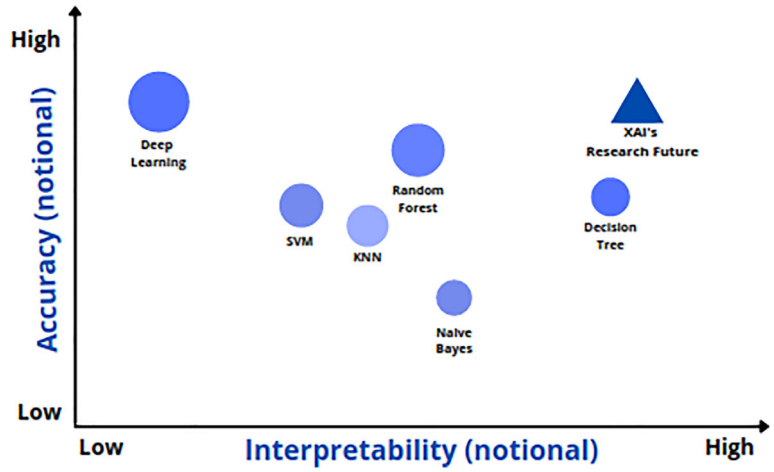


FIGURE 2 Accuracy vs. interpretability for different machine learning models

TABLE 1 Performance comparison for the Caltech-101 dataset

Method	Accuracy	Time (s)	#Parameters	Interpretability
xDNN (Angelov & Soares, 2020)	**94.31%**	362	**4 per class**	**High**
VGG-16 (Simonyan & Zisserman, 2014)	90.32%	18,332	138.000.000	Very low
ResNet-50 (He et al., 2016)	90.39%	12,540	23.000.000	Very low
Random forest (Breiman, 2001)	87.12%	412	~20,000	Medium
SVM (Hearst et al., 1998)	86.64%	783	~15,000	Low
*k*NN (Peterson, 2009)	85.65%	**221**	~300 and all data	Low
Decision tree (Quinlan, 1996)	86.42%	236	~5 rules per class	**High**
Naive Bayes (Rish, 2001)	54.84%	323	409,700	Medium

benchmarking dataset considering different methods and evaluation metrics (best results are highlighted with bold fonts).

AI has been closely related to automated reasoning and mimicking human intelligence from its inception (Angelov & Gu, 2018). ANNs, as a branch of AI that is closely related to ML, went through a roller coaster cycle of development starting in the middle of the past century while the Second World War was still raging with the introduction of the computational perceptron—the model of a single neural cell (neuron) by Warren McCulloch and Walter Pitts in 1943 (Bien & Tibshirani, 2011). At the time that this theory was introduced it was very close to the biological original that inspired it—the human brain. With the power that comes from networking multiple (i.e., millions or even billions of) neurons comes predictive power, but also a complexity that tends to opaqueness. Historically, ANNs went through their first boom during the 1980s–1990s when one of the main achievements was the introduction of popular efficient supervised learning methods, such as the back-propagation algorithm by Werbos (1990), and architectures such as multi-layer perceptrons (MLPs), RBF, and so on. The power of ANNs is in the layered architecture which, in effect, performs a series of embedded mathematical transformations—somewhat analogous to the Russian dolls called "matryoshka":

$$y_p = f_n\left(\sum_q\left(w_{pq}f_{n-1}\left(\ldots f_2\left(\sum_i\left(w_{ri}f_1\left(\sum_j\left(w_{ij}x_j\right)\right)\right)\right)\right)\right)\right), \tag{1}$$

where y_p denotes the pth output; w represents the weights; j is the number of inputs/features and i is the number of neurons at a hidden layer; f_1 denotes the activation function of the input layer; f_2 denotes the activation function of a hidden layer; f_n denotes the activation function of the output layer.

MLPs have one or more hidden layers. ANNs with multiple hidden layers are now called "deep." ML occurs by modifying the weights of the links between the nodes of neighboring layers. The good news is that the huge number of weights allows one to model highly complex functions that map the inputs (in the case of classification—features) to the output(s)—(in classification—class labels). It has been proven theoretically that an MLP can approximate arbitrarily well any nonlinear function (Hornik et al., 1990). Indeed, ANNs are highly parametric models. In addition, the layered structure allows the learning to be performed using error back-propagation with the relatively simple and reliable gradient family of methods combined with conveniently described single neuron activation functions, such as Gaussian or linear. The bad news is that the potentially huge number of weights (unknown parameters) requires a lot of training data, time, computational resources, and makes the overall model hard to explain because the link between the weights and the physicality of the original problem is broken. In comparison, humans can learn from very few examples, or even a single example, and understand the concept behind the classification decision as well as explain and articulate it to others. The AI systems of the future will have to emulate such capabilities (Angelov & Gu, 2018).

Historically, interest towards ANNs dipped around the mid-90s leading to the so-called *AI winter* and attention moved away. Even the success of IBMs Deep Blue chess playing computer program that was able to beat the world's best chess player Gary Kasparov was not able to revitalize AI (Campbell et al., 2002). However, when in 2015 the UK company DeepMind that defeated the world champion of the game Go with its AlphaGo (Chang et al., 2016) using a deep form of ANN, interest in ANNs and AI, as well as ML, returned with full force. It does have to be acknowledged that problems such as (so-called fair) games are well-structured logical problems and, therefore, easier to learn in comparison to the much more open-ended real-life problems that are of interest to various industrial and defense applications. For example, autonomous driving or aerial systems, financial, health, legal, and other real-life problems are more complex, unpredictable, and uncertain (Nilsson, 2014).

3 | XAI TAXONOMY

In the literature, a variety of terms exist to indicate the opposite of the "black box" nature of some of the AI and ML, and especially DL, models. We distinguish the following terms:

- *Transparency*: a model is considered to be transparent if, by itself, it has the potential to be understandable. In other words, transparency is the opposite of "black-box" (Adadi & Berrada, 2018).
- *Interpretability*: is defined as the capacity to provide interpretations in terms that are understandable to a human (Gilpin et al., 2018).
- *Explainability*: is related with the notion of explanation as an interface between *humans* and an AI system. It comprises AI systems that are accurate and comprehensible to humans (Gilpin et al., 2018).

Although these terms are similar in their semantic meanings, they confer different levels of AI to be accepted by humans. For more details, the ontology and taxonomy of XAI at a high level can be detailed as below:

- *Transparent model*: Typical transparent models (Adadi & Berrada, 2018) include *k*-nearest neighbors (*k*NN), decision trees, rule-based learning, Bayesian network, and so on. The decisions from these models are often transparent, although transparency, as a property, is not sufficient to guarantee that a model will be readily explainable.
- *Opaque model*: Typical opaque models (Pasquale, 2015; Rudin, 2019) include random forest, neural networks, SVMs, and so on. Although these models often achieve high accuracy, they are not transparent.
- *Model agnostic*: Model-agnostic XAI approaches (Dieber & Kirrane, 2020) are designed with the purpose of being generally applicable. As a result, they have to be flexible enough, so that they do not depend on the intrinsic architecture of the model, thus, operating solely on the basis of relating the input of a model to its outputs.
- *Model-specific*: Model-specific XAI approaches often take advantage of knowing a specific model and aim to bring transparency to a particular type of one or several models (Bach et al., 2015).
- *Explanation by simplification*: By simplifying a model via approximation (Tritscher et al., 2020), we can find alternatives to the original models to explain the prediction we are interested in. For example, we can build a linear model or a decision tree around the predictions of a model, using the resulting model as a surrogate to explain the more complex one.

- *Explanation by feature relevance*: This idea is similar to simplification. Roughly, this type of XAI approaches attempts to evaluate a feature based on its average expected marginal contribution to the model's decision, after all possible combinations have been considered (Chen et al., 2019; Pedreschi et al., 2019).
- *Visual explanation*: This type of XAI approach is based on visualization (Chattopadhay et al., 2018). As such, the family of data visualization approaches can be exploited to interpret the prediction or decision over the input data.
- *Local explanation*: Local explanations (Selvaraju et al., 2017) approximate the model in a narrow area, around a specific instance of interest, and offer information about how the model operates when encountering inputs that are similar to the one we are interested in explaining.

The ML literature predominantly uses the term "interpretability" as opposed to "explainability," but according to Burkart and Huber (2020), *interpretability* itself is insufficient as it does not cover all possible problems associated with understanding "black-box" models. To gain the trust of users, and acquire meaningful insights about the causes, reasons, and decisions of "black-box" approaches, explainability is required rather than simple interpretability. Although, explainable models are interpretable by default, the opposite is not always true. The existing literature (Adadi & Berrada, 2018) divides XAI taxonomy by:

- *Scope* (local (Bach et al., 2015; Selvaraju et al., 2017) and global (Chen et al., 2019; Pedreschi et al., 2019)).
- *Usage* (post hoc, e.g., surrogate models (Dieber & Kirrane, 2020; Pedreschi et al., 2019; Tritscher et al., 2020) and intrinsic to the model architecture, e.g., explainable-by-design (Soares, Angelov, Biaso, et al., 2020; Soares, Angelov, Costa, et al., 2020)).
- *Methodology* (focused on the features (Chen et al., 2019; Selvaraju et al., 2017) or on the model parameters (Dieber & Kirrane, 2020)).

In recognition of the growing importance of this topic, NIST published in August 2020 *Four principles of XAI* (Phillips et al., 2020), which define the following fundamental principles which an AI must honor to be considered an XAI as follows:

- *Explanation*: this principle states that an AI system must supply evidence, support; or reasoning for each decision made by the system.
- *Meaningful*: this principle states that the explanation provided by the AI system must be understandable by, and meaningful to, its users. As different groups of users may have different necessities and experiences, the explanation provided by the AI system must be fine-tuned to meet the various characteristics and needs of each group.
- *Accuracy*: this principle states that the explanation provided by the AI system must reflect accurately the system's processes.
- *Knowledge limits*: this principle states that AI systems must identify cases that they were not designed to operate in and, therefore, their answers may not be reliable.

Figure 3 depicts the ontology of the XAI taxonomy. Transparent models can easily achieve explainability, while opaque models require post hoc approaches to make them explainable. The categories of post hoc approaches are illustrated accordingly.

4 | REVIEW OF THE STATE-OF-THE-ART

Current research on XAI is still mostly limited to sensitivity analysis (Arrieta et al., 2020), layer-wise feature relevance propagation and attribution (Tritscher et al., 2020), local pseudo explanations by LIME (Dieber & Kirrane, 2020), game-theoretic Shapley additive explanations (Chen et al., 2019), gradient-based localization, and Grad-CAM (Selvaraju et al., 2017) or surrogate models. In this section, some of the more widely used methods are outlined.

4.1 | Features-oriented methods

SHapley Additive exPlanation (SHAP) (Lundberg & Lee, 2017) is a game-theoretic approach to explain ML predictions. SHAP seeks to deduce the amount each feature contributed to a decision by representing the features as players in a

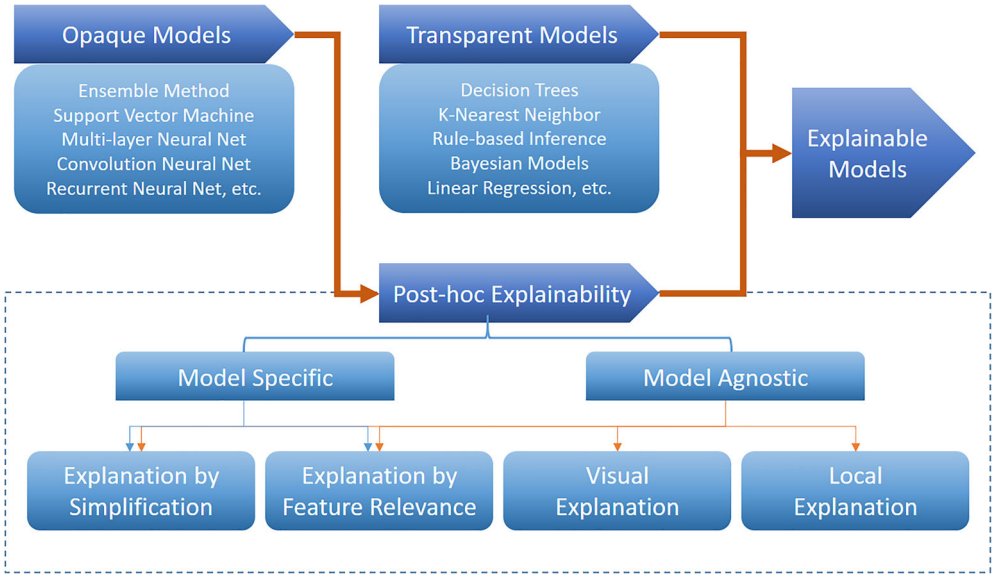


FIGURE 3 The high-level ontology of explainable artificial intelligence approaches

coalition game. The payoff of the game is an additive measure of importance, the so called Shapley value, which represents the weighted average contribution of a particular feature within every possible combination of features. As such, local and global interpretations of a model are consistent and the average prediction is fairly distributed across all Shapley values, meaning that contrasting comparisons between explanations are possible. However, if the model is not additive then interpretation of the Shapley values is not always transparent, as predictive models may have non independent pay-off splits. Furthermore, while SHAP can be considered model agnostic, optimized implementations of the SHAP algorithm to all model types is not immediately straight forward or efficient.

Class activation maps (CAMs) are specific to CNNs. CAMs represent the per-class weighted linear sum of visual patterns present at various spatial locations in an image (Zhou et al., 2016). More formally, global average pooling is applied to the final convolutional feature map in a network, before the output layer. These pooled feature maps are then used as the input features to a fully connected layer and output through a loss function. By projecting the weights of the output back to the previous convolutional layer, the areas in the input image with greater influence over the CNNs' decision are highlighted per-class and visible through a heatmap representation. CAMs cannot be applied to pre-trained networks and networks that do not adhere to the specified fully convolutional network architecture. Additionally, spatial information can be lost by the fully connected layer and map scaling. Two generalizations of the base CAM model, Grad-CAM (Selvaraju et al., 2017) and Grad-CAM++ (Chattopadhay et al., 2018), try to further increase the explainability of CNNs.

Gradient-weighted class activation mapping (Grad-CAM) (Selvaraju et al., 2017) generalizes CAM to any arbitrary CNN architecture and without retraining. The gradients for any target class are fed into the final convolutional layer and an importance score computed in respect to the gradients. As with other methods, a heatmap representation of the Grad-CAM indicates which regions of the input image were most important in the CNN's decisions. However, Grad-CAM produces only coarse-grained visualizations and cannot explain multiple instances of the same object in an image. Grad-Cam++ (Chattopadhay et al.) considers the weighted average of the gradients to overcome these drawbacks.

Feature oriented methods provide insights into where a decision is taking place in terms of the input, but fall short of a human level explanation of how and why the model came to those decisions. Consequently, a human could not exactly reproduce the explanations rendered by the model.

4.2 | Global methods

For features with precise semantic definitions, global attribution mappings (GAMs) (Ibrahim et al., 2019) can explain a neural network's predictions on a global level, across subpopulations, by formulating attributions as weighted conjoined rankings. The advantages are that different subpopulations can be captured through a tuneable granularity parameter.

GAMs find a pair-wise rank distance matrix between features and a K-medoids clustering algorithm used to group similar local feature importances into clusters. The medoid of each cluster then summarizes the pattern detected in each cluster as a global attribution. This approach is therefore relevant to feature exploration among different subpopulations of samples.

Gradient-based saliency maps (Simonyan et al., 2013) are a visualization technique which render the absolute value of the gradient (in respect to the input features) of the majority predicted class as a normalized heatmap. The pixels with a high activation are highlighted and correspond to areas that are most influential (i.e., salient). The method's explanation lies in the ability for a user to look at what features in the image are being used in the classification decision. However, the absolute value means that gradients of neurons with negative input are suppressed when propagating nonlinear layers. As with feature-oriented methods, gradient-based saliency maps do little to communicate decisions beyond model diagnostics.

In Ancona et al. (2018), deep attribute maps are presented as a technique for rendering the explainability of gradient-based methods. Importantly, the proposed framework illustrates evaluations between different saliencey-based explanation models. Simply, the gradient of the output is multiplied by the respective input to generate an explanation of a model's prediction in the form of a heatmap. Red and blue colors indicate positive and negative contributions, respectively, to the output decision. Explanations are sensitive to noisy gradients and variations in the input. Deep attribute maps alone cannot explain why two models produce similar or different results.

4.3 | Concept models

Concept activation vectors (CAVs) were introduced by Kim et al. (2021), a technique to explain globally the internal states of a neural network by mapping human understandable features to the high-level latent features extracted by the neural network. As such, CAVs represent the degree to which these abstract features point towards a set of human understandable concepts chosen by a user. Of course, a certain amount of human bias is imposed, but by explaining the associated concept it becomes possible to determine any defects in the decision-making process the model has learned; for instance, if certain characteristics are mistakenly seen as important. Subsequently, automatic concept based explanations (Ghorbani et al., 2019) extract CAVs automatically without human supervision, thereby removing human bias. Instead of being chosen, the human understandable concepts are segmented at various spatial resolutions from in-class images. Nevertheless, concept-based methods are reliant on the concepts being uniquely meaningful to the class, and the effectiveness of explanation is adversely affected if a chosen concept is commonly present in multiple classes.

4.4 | Surrogate models

Local interpretable model-agnostic explanations (LIME) (Dieber & Kirrane, 2020) is a model-agnostic technique to create locally optimized explanations of ML models. LIME trains an interpretable surrogate model to learn the local behavior of a global "black box" model's predictions. For image classification, an input image is divided into patches of contiguous superpixels (i.e., an image object) and a weighted local model is then trained on a new set of permuted instances of the original image (i.e., some superpixels are turned to gray). The intuition is then that by changing aspects of the input data that are human understandable (spatial objects) and learning the differences between those perturbations and the original observations, one can learn what about the input contributed to each class score. However, these explanations are not always informative or reliable at a human level if the parameters that control the perturbations are chosen based solely on heuristics.

4.5 | Local, pixel-based methods

Layer-wise relevance propagation (LRP) (Bach et al., 2015) uses predefined propagation rules to provide an explanation of a multilayered neural network's output in respect to the input. The method renders a heatmap, thereby providing insight into which pixels contributed to the model's prediction and the extent to which they did. Accordingly, LRP highlights positive contributions to a network's decision. While LRP can be applied to an already trained network, this process is post hoc and therefore provides only a simplified distillation of the features' role in the decision and is only

applicable if the network implements backpropagation. DeconvNet (Noh et al., 2015) uses a semantic segmentation algorithm which learns a deconvolution network and, therefore, provides insights about pixel contribution during the classification process. Similarly, a deep belief network (Hinton et al., 2006) was proposed to improve the interpretability of traditional neural networks.

4.6 | Human-centric methods

The above methods, despite their advantages, do not provide clear explanations understandable to humans. They rather "barely scratch the surface" of the "black box" aiming for "damage limitation" with post hoc hints about the features (attribute allocation) or localities within an image. This is radically different from the way people reason and make decisions, make associations, evaluate similarities, and draw an analogy that can be articulated in court or to another expert (e.g., in medicine, finance, law or other area). The aforementioned methods do not answer the fundamental questions of model structure and parameters relating to the nature of the problem and completely ignore reasoning.

Recently, in Angelov and Soares (2020) a cardinally different approach to explainability was proposed which treats it as a human-centric (anthropomorphic) phenomena rather than reducing it to statistics. Indeed, humans compare items (e.g., images, songs, and movies) in their entirety and not per feature or pixel. People use similarity to associate new data with previously learned and aggregated prototypes (Bien & Tibshirani, 2011) while statistics is based on averages (Bishop, 2006).

5 | EXPLAINABILITY-CRITICAL APPLICATIONS

The frequency and importance of algorithms in applications have lead regulators and official bodies to develop policies that provide clearer accountability for algorithmic decision-making. One such example is the European Union's General Data Protection Right, which some have interpreted as a "Right to Explanation" (Goodman & Flaxman, 2017). Although the extent of this right is in dispute, the discourse around such topics has reinforced that automated systems must avoid inequality and bias in decisions. Furthermore, they must fulfill the requirements for safety and security in safety-critical tasks. Consequently, there has been a recent explosion of interest in XAI models in different areas. Recently, it has been reported that XAI has been applied in several critical domain applications such as medicine (Holzinger et al., 2017), the criminal justice system (Dressel & Farid, 2018), and autonomous driving (Cysneiros et al., 2018).

In the medical domain there is a growing demand for AI approaches, most notably during the COVID-19 pandemic. However, AI applications must not only perform well in terms of classification metrics, but need also to be trustworthy, transparent, interpretable, and explainable, especially for clinical decision-making (Holzinger et al., 2017). Soares, Angelov, Biaso, et al. (2020), for example, offered an explainable DL approach for COVID-19 identification via computed tomography (CT) scans. The proposed approach was reported to surpass mainstream DL approaches such as ResNet (He et al., 2016), GoogleNet (Szegedy et al., 2015), and VGG-16 (Simonyan & Zisserman, 2014) in terms of accuracy, F1 score and other statistical metrics of performance, but critically, this approach is based on prototypes which, in this case, represent a CT scan that a radiologist can clearly understand. The prototypes are examples of CT scans of patients with or without COVID. This approach can be expanded readily to include more classes, such as "mild" or "severe" COVID, and so on, or go to the level of superpixels as in Tetila et al. (2020). Furthermore, the proposed deep neural network has a clear and explainable architecture (with each layer having a very clear meaning and using visual images of CT scans so the decision can easily be visualized).

Couteaux et al. (2019) proposed an explainable DeepDream approach where the activation of a neuron is maximized by performing gradient ascent of a given image. The method has output curves that show the evolution of the features during the maximization. This favors the visualization and interpretability of the neural network and was applied for tumor segmentation from liver CT scans (Couteaux et al., 2019).

Another example application of XAI is the criminal justice system. In some countries such as the United States automated algorithms are being used to predict where crimes will most likely occur, who is most likely to commit a violent crime, who is likely to fail to appear at their court hearing, and who is likely to re-offend at some point in the future (Dressel & Farid, 2018). One such widely used criminal risk assessment tool is the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS). Although the data used by COMPAS do not include an individual's race,

TABLE 2 XAI critical applications—summary

Method	Application
Holzinger et al. (2017)	Medicine
Dressel and Farid (2018), Soares and Angelov (2019)	Criminal justice system
Soares, Angelov, Biaso, et al. (2020)	COVID-19 identification
Couteaux et al. (2019)	Tumor segmentation
Mathews (2019)	NLP
Smith-Renner et al. (2019), Xie and Philip (2018)	Anomaly and fraud detection
Soares et al. (2019)	Novelty detection
Soares, Angelov, Costa, et al. (2020)	Autonomous vehicles

other aspects of the data may be correlated to race that can lead to racial biases in the predictions. Therefore, explanations of such critical decisions are necessary to favor fairness and reduce racism during the decisions (Dressel & Farid, 2018). As discussed by Soares and Angelov (2019), prototype-based algorithms can be a solution to reduce bias and favor fairness as one can check and balance the prototypes generated to guarantee a fairer decision. Moreover, the approach proposed in Soares and Angelov (2019) also provides human explainable rules to assist specialists during decision-making.

Applications based on NLP also benefit from XAI. Mathews (2019) presented an explainable approach for tweet data classification based on LIME. XAI techniques for anomaly and fraud detection are also explored by different authors as a means of enhancing users' trust (Smith-Renner et al., 2019; Xie & Philip, 2018).

Another application domain in which there is a growing number of applications and interest towards XAI is defined as autonomous systems (these may be airborne, maritime or land-based individual vehicles with a control system or swarms). Self-driving vehicles, for example, are automated systems that are expected to be used in possibly an unknown environment (Das & Rad, 2020). In this context, the trust and acceptance of such systems require transparency, in contrast to "black-box" solutions. For example, a recent crash (on 18 March 2018) by an autonomous car owned by Uber led to the operator being charged with negligent homicide (Stilgoe, 2020) two and a half years later. It is, therefore, critically important (not only from the point of view of public perception and trust which can make or break market perspectives, but also from a purely regulatory and legal perspective) to have transparent, interpretable, and explainable, non-"black-box" models in use. This can lead to more reliable systems which are necessary to guarantee safety and meet regulations (Das & Rad, 2020). Recently, examples of prototype-based approaches were published in which XAI was used for understanding the visual scene (Soares et al., 2019) and the situation awareness of a self-driving car on a highway/motorway/autobahn through the so-called vector of affordance indicators (relative velocities and distances to the neighboring vehicles) (Soares, Angelov, Costa, et al., 2020). Not only were the accuracy, F1 score and other statistical measures reported to be comparable with, or surpass conventional DL methods, but the model was clearly explainable to a human in the form of linguistic rules and visual means. Moreover, for cases when the situation on the road is deemed to be generated from a class that was never used in training (a completely new type of scene) it was reported that conventional DL methods can make an incorrect prediction with a high confidence, which may have very damaging consequences for autonomous vehicles, passengers, legal outcomes, and trust. Instead, Soares et al. (2019) proposed a self-evolving approach, which can pro-actively learn from new situations due to its prototype nature, and also provide explainable rules. These safety mechanisms are very important for critical applications such as autonomous driving. Table 2 summarizes the applications mentioned in this section:

6 | FURTHER DISCUSSION

XAI aims to help humans to understand why a machine decision has been reached and whether or not it is trustworthy. Consequently, XAI is inevitably a paradigm on how to bridge machine intelligence and human intelligence, with the goal being to enable and widen the acceptance of AI systems by human subjects. In this sense, XAI can be interpreted as "AI for people."

6.1 | Critical importance of XAI

Even though intelligent systems offer great possibilities, the research initiative of XAI raises concerns of giving such intelligent systems too much power without the ability to explain the decision-making process lying underneath such complex systems to domain experts (e.g., medics, lawyers, financial experts, etc.) in terms, and in a form, understandable to them. This not only helps understand specific decisions made by such systems, but also encourages researchers to create more human-like (anthropomorphic) solutions as well as inspiring the study and increased understanding of the brain as a natural information processing phenomenon. Moreover, since machines are taking over the decision process in many daily situations, user rights have to be protected. Intelligent machines still mostly cannot process abstract information or real-world knowledge unless it is converted to a form understandable by the algorithm (features, outputs, and labels).

The above critical issue has become extremely important in many AI application areas. For example, the decision from an automated diagnosis system may influence the treatment plan of a patient, and doctors need to understand why such a decision was made and evaluate the underlying risks. If we consider farming-assisting autonomous drones, the farmers need to know why, when, and where drones decide to perform automated spraying of water or pesticides. Thus, a trustworthy XAI system becomes a critical prerequisite for AI to be applied to practically any real-world problem. Much research is now being focused on how to handle such kinds of problems.

6.2 | Bridge the gap between DL and neuroscience via XAI

DL as the state-of-the-art AI technique has its roots in the emulation of the human brain. To make deep neural networks explainable, an ultimate goal is to find a way to match human intelligence and find a way to build a human-made "brain" that can interpret the neuronal activities in the human brain or at least, at a functionally higher level, map the deep architectures to the layered information processing units in the brain.

There are two important differences between the features of current mainstream DL and the human brain. First, the human brain is more like an analogue circuit without the ability to store high precision parameters. Second, neurons in the human brain are highly interconnected instead of the carefully "handcrafted" architectures of the current mainstream DL. It is curious, therefore, that the mainstream DL literature is very critical of so-called "handcrafted" features (Goodfellow et al., 2014), but is slow to accept that the architectures it is pushing forward are "handcrafted", highly problem-specific and with multiple meta-parameters such as stride, kernel sizes, number of layers, and so on.

With the above concerns, XAI can help bridge the gap between DL and neuroscience in a mutually beneficial way. On one side, neuroscience and psychology can help build rationalized XAI models that are more easily understood by humankind (Byrne, 2019; Taylor & Taylor, 2020). On the other side, XAI models derived from deep neural networks can also help in understanding the mechanisms of intelligence in the human brain (Fellous et al., 2020; VU et al., 2018). The ultimate goal of XAI could be redefined as the pursuit of fully understanding how human intelligence originates from neurons.

6.3 | Future directions

One promising direction for future research is to focus on prototype-based models (Angelov & Gu, 2018; Angelov & Soares, 2020) rather than on abstract and highly embedded architectures. Prototype-based models are not new as such (Bien & Tibshirani, 2011)—starting with the simplest (and highly efficient example of *k*NN), through RBF types of ANNs and IF...THEN rules. The power of prototype-based models was noted by Tibshirani in (Bien & Tibshirani, 2011), but so far these were not developed in the context of DL where they can combine a deeper architecture with a clearly explainable form of representation. Despite its efficiency, the *k*NN method is, strictly speaking, not a learning method, because it requires *all* the data to be available and stored. Some sparsity is needed which can result from simple unsupervised forms of learning such as clustering or more complex end-to-end auto-encoders. There is an established misconception that the only form of learning is parametric learning through optimization (minimization) of a cost (or loss) function. In fact, people learn by acquiring prototypes from data samples using similarity. Following this logic, the learning in prototype-based models revolves around the position and properties of the prototypes in the feature/data space as opposed to the parameters/weights-centered approach that dominates the mainstream. In addition, there is a

principle difference between similarity and statistical learning (i.e., the two alternative approaches to evaluate the difference and divergence between two data items). Similarity can be defined over a pair of data items/samples while statistical measures require a large (theoretically infinite) number of independent data observations.

Another promising direction is to build Turing's type-B random machines (or unorganized machines) (Jiang & Crookes, 2019; Webster, 2012), also random Boltzmann machines, which can possibly lead to a generalized AI. The inclusion of new neuro-scientific findings into XAI models will make research on XAI more rationalized, and vice versa: such a cross-disciplinary exploitation will make XAI not only meaningful for AI researchers but also help solve century-old challenges on how to understand human intelligence, ultimately. Open research questions in this area include: (i) how best to determine the network/model architecture?; (ii) how best to extract and represent features?; (iii) what are the best distance metrics and what are the implications?; (iv) which is the best optimization method?; and (v) how to determine the best set of prototypes that represent the data best (if a prototype-based method is being used)?

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

AUTHOR CONTRIBUTIONS

Plamen Angelov: Conceptualization; data curation; formal analysis; methodology; supervision; writing-original draft; writing-review & editing. **Eduardo Soares:** Conceptualization; data curation; formal analysis; methodology; software; validation; visualization; writing-review & editing. **Richard Jiang:** Conceptualization; formal analysis; investigation; methodology; writing-review & editing. **Nicholas I. Arnold:** Conceptualization; formal analysis; investigation; methodology; writing-review & editing. **Peter M. Atkinson:** Conceptualization; formal analysis; investigation; methodology; writing-review & editing.

ORCID

Plamen P. Angelov https://orcid.org/0000-0002-5770-934X
Eduardo A. Soares https://orcid.org/0000-0002-2634-8270
Richard Jiang https://orcid.org/0000-0003-1721-9474
Nicholas I. Arnold https://orcid.org/0000-0003-3968-6233
Peter M. Atkinson https://orcid.org/0000-0002-5489-6880

RELATED WIREs ARTICLE

Causability and explainability of artificial intelligence in medicine

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138–52160.

Ancona, M., Ceolini, E., Öztireli, C., & Gross, M. (2018). *Towards better understanding of gradient-based attribution methods for deep neural networks*. http://arxiv.org/abs/1711.06104

Angelov, P., & Soares, E. (2020). Towards explainable deep neural networks (xDNN). *Neural Networks*, *130*, 185–194.

Angelov, P. P., & Gu, X. (2018). Toward anthropomorphic machine learning. *Computer*, *51*, 18–27.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., & Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One*, *10*, e0130140.

Bien, J., & Tibshirani, R. (2011). Prototype selection for interpretable classification. *The Annals of Applied Statistics*, *5*, 2403–2424.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Burkart, N., & Huber, M. F. (2020). *A survey on the explainability of supervised machine learning*. arXiv preprint arXiv:2011.07876.

Byrne, R. M. (2019). Counterfactuals in explainable artificial intelligence (XAI): Evidence from human reasoning. In *Proceedings of the twenty-eighth international joint conference on artificial intelligence (IJCAI-19)* (Vol. 1, pp. 6276–6282).

Campbell, M., Hoane, A. J., Jr., & Hsu, F.-H. (2002). Deep blue. *Artificial Intelligence*, *134*, 57–83.

Chang, H. S., Fu, M. C., Hu, J., & Marcus, S. I. (2016). Google deep mind's alphago. *OR/MS Today*, *43*, 24–29.

Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018). Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE Winter conference on applications of computer vision (WACV)* (pp. 839–847).

Chen, H., Lundberg, S., & Lee, S.-I. (2019). *Explaining models by propagating Shapley values of local components*. arXiv preprint arXiv: 1911.11888.

Core, M. G., Lane, H. C., Van Lent, M., Gomboc, D., Solomon, S., & Rosenberg, M. (2006). Building explainable artificial intelligence systems. In *AAAI* (pp. 1766–1773).

Couteaux, V., Nempont, O., Pizaine, G., & Bloch, I. (2019). Towards interpretability of segmentation networks by analyzing DeepDreams. In *Interpretability of machine intelligence in medical image computing and multimodal learning for clinical decision support* (pp. 56–63). Springer.

Cysneiros, L. M., Raffi, M., & do Prado Leite, J. C. S. (2018). Software transparency as a key requirement for self-driving cars. In *2018 IEEE 26th international requirements engineering conference (RE)*. IEEE (pp. 382–387).

Das, A., & Rad, P. (2020). *Opportunities and challenges in explainable artificial intelligence (XAI): A survey*. arXiv preprint arXiv:2006.11371.

Dieber, J., & Kirrane, S. (2020). *Why model why? Assessing the strengths and limitations of lime*. arXiv preprint arXiv:2012.00093.

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, *4*, eaao5580.

Fellous, J. M., Sapiro, G., Rossi, A., Mayberg, H., & Ferrante, M. (2020). Explainable artificial intelligence for neuroscience: Behavioral neurostimulation. *Frontiers in Neuroscience*, *13*, 1346.

Ghorbani, A., Wexler, J., Zou, J., & Kim, B. (2019). *Towards automatic concept-based explanations*. http://arxiv.org/abs/1902.03129

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*. IEEE (pp. 80–89).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, *3*, 2672–2680.

Goodman, B., & Flaxman, S. (2017). European union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine*, *38*, 50–57.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their Applications*, *13*, 18–28.

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*, 1527–1554.

Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). *What do we need to build explainable AI systems for the medical domain?* arXiv preprint arXiv:1712.09923.

Hornik, K., Stinchcombe, M., & White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks*, *3*, 551–560.

Ibrahim, M., Louie, M., Modarres, C., & Paisley, J. (2019). Global explanations of neural networks: Mapping the landscape of predictions. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, AIES '19*. Association for Computing Machinery (pp. 279–287). https://doi.org/10.1145/3306618.3314230.

Jiang, R., & Crookes, D. (2019). Shallow unorganized neural networks using smart neuron model for visual perception. *IEEE Access*, *7*, 152701–152714.

Kim, B., Wattenberg, M., Gilmer, J., Cai, C., Wexler, J., Viegas, F., & Sayres, R. (2021). *Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (TCAV)*. http://arxiv.org/abs/1711.11279

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4765–4774 https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

MacCarthy, M. (2019). *An examination of the algorithmic accountability act of 2019*. Available at SSRN 3615731.

Mathews, S. M. (2019). Explainable artificial intelligence applications in NLP, biomedical, and malware classification: A literature review. In *Intelligent computing-proceedings of the computing conference* (pp. 1269–1292). Springer.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*, 529–533.

Nguyen, A., Yosinski, J., & Clune, J. (2015). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 427–436).

Nilsson, N. J. (2014). *Principles of artificial intelligence*. Morgan Kaufmann.

Noh, H., Hong, S., & Han, B. (2015). Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision* (pp. 1520–1528).

Pasquale, F. (2015). *The black box society*. Harvard University Press.

Pedreschi, D., Giannotti, F., Guidotti, R., Monreale, A., Ruggieri, S., & Turini, F. (2019). Meaningful explanations of black box AI decision systems. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, pp. 9780–9784).

Peterson, L. E. (2009). K-nearest neighbor. *Scholarpedia*, *4*, 1883.

Phillips, P. J., Hahn, C. A., Fontana, P. C., Broniatowski, D. A., & Przybocki, M. A. (2020) *Four principles of explainable artificial intelligence*.
Quinlan, J. R. (1990). Decision trees and decision-making. *IEEE Transactions on Systems, Man, and Cybernetics*, *20*, 339–346.
Quinlan, J. R. (1996). Learning decision tree classifiers. *ACM Computing Surveys (CSUR)*, *28*, 71–72.
Rish, I. (2001). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence* (Vol. 3, pp. 41–46).
Robinson, A. J., & Voronkov, A. (2001). *Handbook of automated reasoning* (Vol. 1). Gulf Professional Publishing.
Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*, 206–215.
Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, *3*, 210–229.
Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618–626).
Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). *Deep inside convolutional networks: Visualising image classification models and saliency maps*. arXiv:1312.6034 [cs].
Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556.
Smith-Renner, A., Rua, R., & Colony, M. (2019). Towards an explainable threat detection tool. In *IUI workshops*.
Smolensky, P. (1987). Connectionist AI, symbolic AI, and the brain. *Artificial Intelligence Review*, *1*, 95–109.
Soares, E., & Angelov, P. (2019). *Fair-by-design explainable models for prediction of recidivism*. arXiv preprint arXiv:1910.02043.
Soares, E., Angelov, P., Biaso, S., Froes, M. H., & Abe, D. K. (2020). SARS-Cov-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-Cov-2 identification. *medRxiv*.
Soares, E., Angelov, P., Costa, B., & Castro, M. (2019). Actively semi-supervised deep rule-based classifier applied to adverse driving scenarios. In *2019 international joint conference on neural networks (IJCNN)*. IEEE (pp. 1–8).
Soares, E. A., Angelov, P. P., Costa, B., Castro, M., Nageshrao, S., & Filev, D. (2020). Explaining deep learning models through rule-based approximation and visualization. *IEEE Transactions on Fuzzy Systems*, *1*, 1–10.
Stilgoe, J. (2020). Who killed Elaine Herzberg? In *Who's driving innovation?* (pp. 1–6). Springer.
Stock, P., & Cisse, M. (2018). ConvNets and ImageNet beyond accuracy: Understanding mistakes and uncovering biases. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 498–512).
Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–9).
Taylor, J. E. T., & Taylor, G. W. (2020). Artificial cognition: How experimental psychology can help generate explainable artificial intelligence. *Psychonomic Bulletin and Review*, *28*, 6276–6282.
Tetila, E., Bressem, K., Astolfi, G., Sant'Ana, D. A., Pache, M. C., & Pistori, H. (2020). System for quantitative diagnosis of COVID-19-associated pneumonia based on superpixels with deep learning and chest CT. *ResearchSquare, 1, 1-13*. https://doi.org/10.21203/rs.3.rs-123158/v1
Tritscher, J., Ring, M., Schlr, D., Hettinger, L., & Hotho, A. (2020). Evaluation of post-hoc XAI approaches through synthetic tabular data. In *International symposium on methodologies for intelligent systems* (pp. 422–430). Springer.
VU, M., Adalı, T., Ba, D., Buzsáki, G., Carlson, D., Heller, K., Liston, C., Rudin, C., Sohal, V., Widge, A., Mayberg, H., Sapiro, G., & Dzirasa, K. A. (2018). A shared vision for machine learning in neuroscience. *Journal Neuroscience*, *18*, 1601–1607.
Webster, C. S. (2012). Alan turing's unorganized machines and artificial neural networks: His remarkable early work and future possibilities. *Evolutionary Intelligence*, *5*, 35–43.
Werbos, P. J. (1990). Backpropagation through time: What it does and how to do it. *Proceedings of the IEEE*, *78*, 1550–1560.
Xie, S., & Philip, S. Y. (2018). Next generation trustworthy fraud detection. In *2018 IEEE 4th international conference on collaboration and internet computing (CIC)*. IEEE (pp. 279–282).
Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, *13*, 55–75.
Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). *Learning deep features for discriminative localization*.

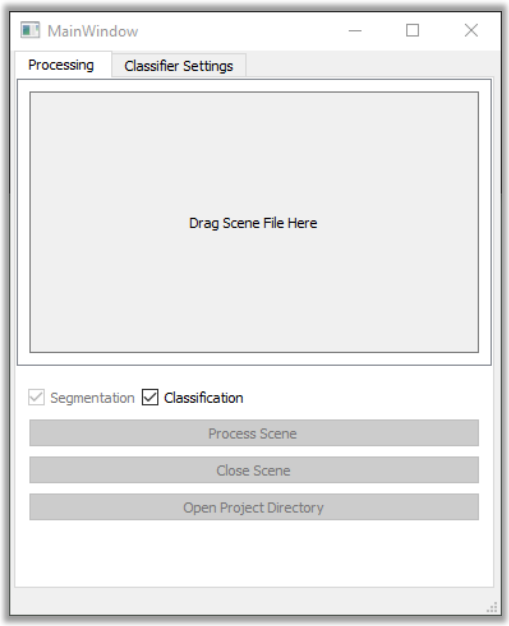
How to cite this article: Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: an analytical review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *11*(5), e1424. https://doi.org/10.1002/widm.1424

Appendix B

GeoPart-XPCC GUI

UI Overview

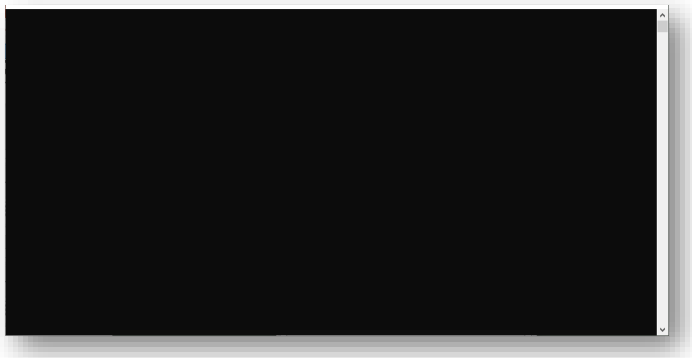
Main Window



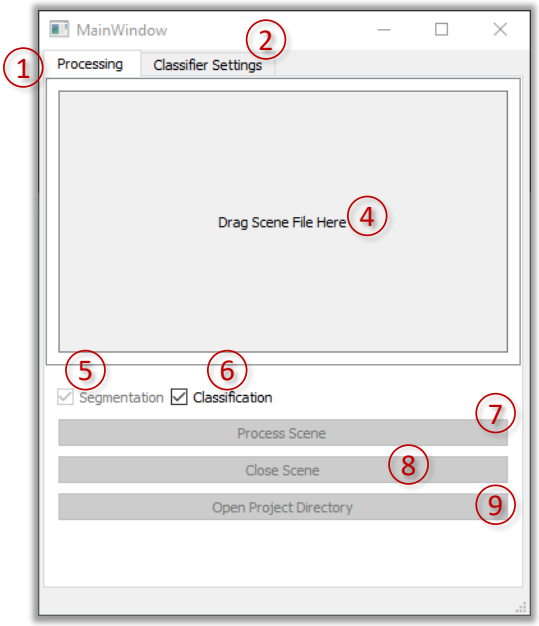
Console Window

A textual GUI console window is provided and prints the standard output and standard error messages from the underlying processes.

NB: closing the console window will also close the main window.

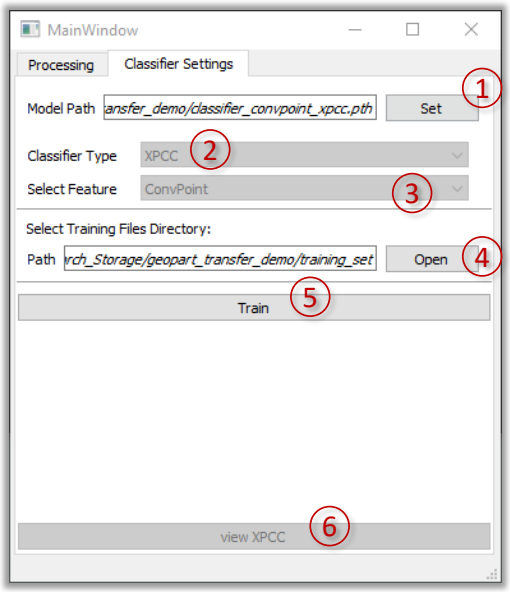


Main Window



1. Processing tab
2. Classifier Settings tab
   - Ensure that the Model Path is correct before processing a scene.
3. Scene drop area
   - Can load pointcloud files or filestore directory
   - Supported pointcloud file formats:
     - .las, .laz, .ply, Ascii (.xyz, .xyzrgb)
4. Toggle Segmentation Process
   - Enabled by default. Required if the scene is a pointcloud file. If the scene type is a filestore directory that has already been segmented, the segmentation is optional.
5. Toggle Classification Process
   - Enabled by default.
6. Process Scene
   - Enabled after scene has been placed in drag area.
7. Close Scene
   - Enabled after scene has been placed in drag area.
   - Closes the current scene loaded into the drag area.
8. Open Project Directory
   - Enabled after scene has been placed in drag area.
   - Opens the directory that contains the currently loaded pointcloud file and filestore.

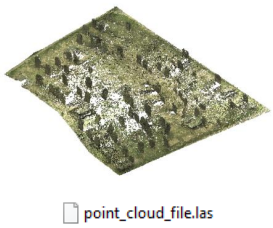
Classifier Settings



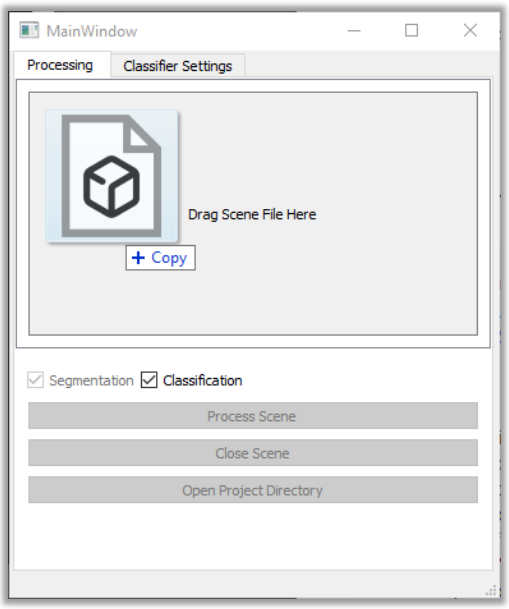
1. Classification Model Path
   i. Specifies the path of the classification model stored on disk.
   ii. Can either specify a trained model or the path to where a new model should be saved.
2. Classifier Type
   i. Options are: XPCC and MLP.
   ii. Can either be selected manually or inferred from the file name.
   iii. If training a new model, the classifier type will be appended to the name if it is not specified: e.g., `classifier_convpoint` will be saved as `classifier_convpoint_xpcc` if the XPCC classifier type is selected.
   iv. If inferred, the field cannot be changed unless the model type is removed from the model path name: e.g., by removing 'xpcc' from `classifier_convpoint_xpcc.pth` -> `classifier_convpoint.pth`
3. Classification Feature Type
   i. Options are: ConvPoint
   ii. Can either be selected manually or inferred from the file name.
   iii. If training a new model, the classifier feature type will be appended to the name if it is not specified: e.g., 'classifier' will be saved as `classifier_convpoint` if the ConvPoint classification feature type is selected.
   iv. If inferred, the field cannot be changed unless the classification feature type is removed from the model path name: e.g., by removing 'convpoint' from `classifier_convpoint_xpcc.pth` -> `classifier_xpcc.pth`
4. Training Files Directory.
   i. Specify the directory that contains the training files.
5. Train a classifier model.
   i. It is necessary to set the model path and training files directory prior to training.
6. View XPCC the XPCC model.

Scene Processing

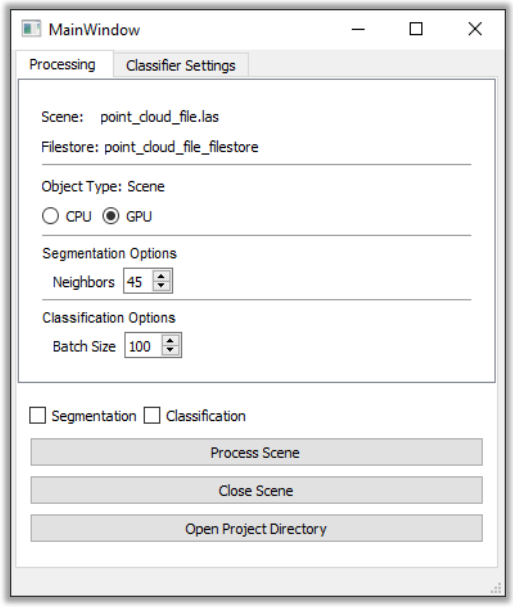
Click and drag the scene file into the drag scene region of the main window.

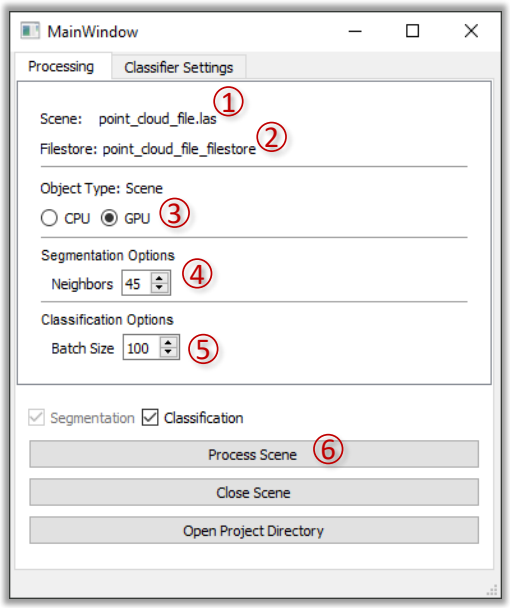


point_cloud_file.las



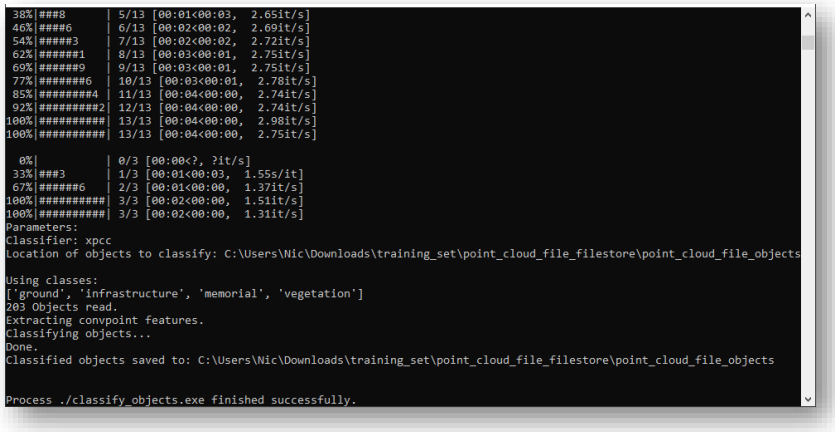
The UI will change and display the process settings. Click the ‘Close Scene’ button to unload the scene.



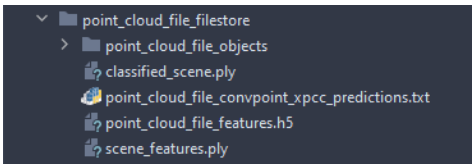


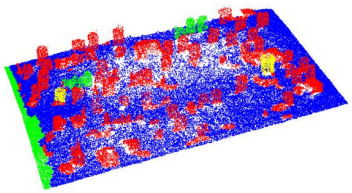
1. Scene file name
2. Scene Filestore name (location is relative to the scene file)
3. Hardware settings: CPU / GPU
   i. CPU is enabled by default. Most processes are multi-threaded or run in parallel.
   ii. If a GPU supported by PyTorch is installed (e.g., NVIDIA) the GPU option is enabled.
4. Select the number of neighbouring points to use for the segmentation process. Default is 45. Increasing the number of neighbours can result in more precise scene segmentations but increases computational load.
5. The number of scene segments to classify in each batch. Default is 100. Batch size is limited by RAM and CPU/GPU resources.
6. Click the 'Process Scene' button to begin the specified processes to run.

The console window will update with information about processes.



Scene File Store





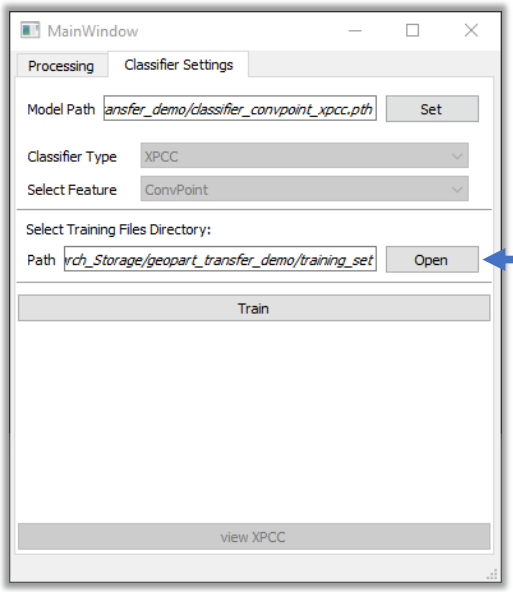
- The topmost directory in the _filestore will be named: `<pointcloud filename>_filestore` and is located in the same directory as the original pointcloud file.
- The file `<pointcloud filename>_objects` will contain the scene segments
- The file `classified_scene.ply` contains the classified scene and can be viewed in a point cloud viewing software such as CloudCompare. To support the majority of viewing software, the **classification indices** are stored in the **'red' vertex** of the .ply file. For CloudCompare, this can be specified as a scalar field upon loading the file.
- The file `<model classifier name>_prediction.txt` contains the classifier predictions used for calculating classification accuracies.
- The file `<pointcloud filename>_features.h5` contains the point cloud features saved in the Hierarchical Data Format.
- The file `scene_features.ply` contains the point cloud features stored as scalar values for the original point cloud scene and can be viewed in a point cloud viewing software such as CloudCompare.

  The ply vertex to feature mappings are:
  - Vertex – red -> Linearity
  - Vertex – green -> Planarity
  - Vertex – blue -> Sphericity
  - Vertex – nz -> Planarity

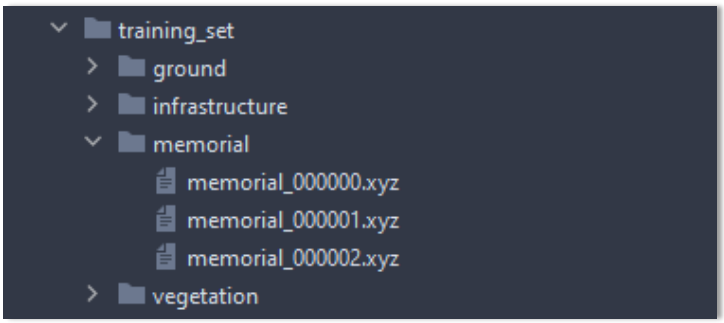
Classifier Training

Train a new classifier

1. Set the model path to a new file name.
2. Select or specify classifier type.
3. Select or specify feature type.
4. Open the Training files directory.
5. Click ‘Train’ button.

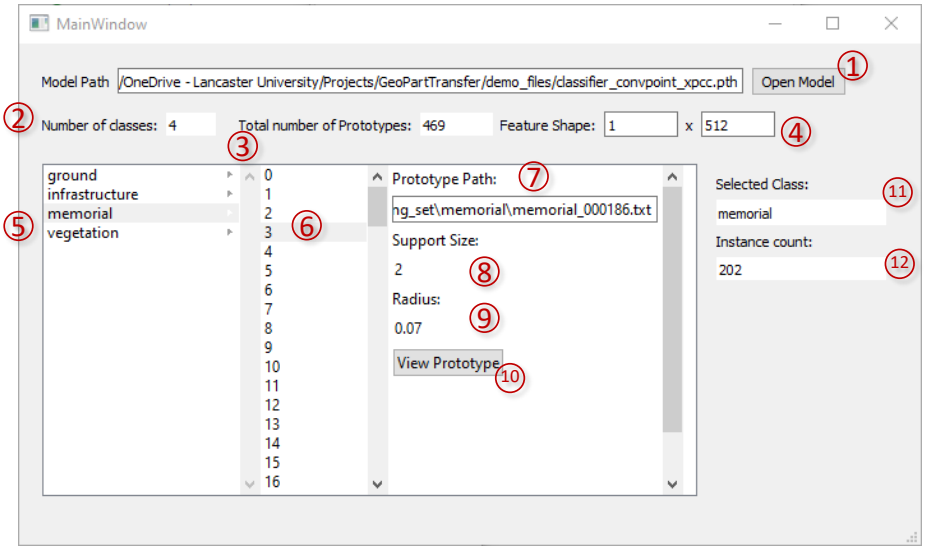


Training directory layout



- [Folder] TrainingSet
  - -- [Folder] Class1
    - --- [File] Object1.xyz
    - --- [File] Object2.xyz
    - --- [File] Object3.xyz
    - --- [File] …
  - -- [Folder] Class2
  - -- [Folder] Class3
  - -- [Folder] …

XPCC Viewer



1. Open a different model path
2. The number of classes the model has been trained with.
3. Total number of prototypes created by the model.
4. The shape of the classification feature.
5. List view of class types.
6. List view of prototypes for the selected class.
7. Path of the prototype object. (These values are stored in the model at training time)
8. Number of training objects similar to the prototype.
9. Radius of influence within the feature space for the prototype.
10. View the prototype (as specified by the Prototype Path)
11. The currently selected class.
12. Number of objects seen during training for that class.

References

Aldoma, A. et al. (Nov. 2011). "CAD-model recognition and 6DOF pose estimation using 3D cues". In: *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 585–592. DOI: 10.1109/ICCVW.2011.6130296.

Alpaydin, E. (2010). *Introduction to machine learning*. eng. MIT Press. ISBN: 978-0-262-01243-0. URL: http://archive.org/details/introductiontoma00alpa_0.

Alzubaidi, L. et al. (2021). "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions". In: *Journal of big Data* 8.1. Publisher: Springer, pp. 1–74.

Ancona, M. et al. (2017). "A unified view of gradient-based attribution methods for Deep Neural Networks". In: *CoRR* abs/1711.06104. arXiv: 1711.06104. URL: http://arxiv.org/abs/1711.06104.

Angelov, P. and E. Soares (Oct. 2020). "Towards explainable deep neural networks (xDNN)". en. In: *Neural Networks* 130, pp. 185–194. ISSN: 0893-6080. DOI: 10.1016/j.neunet.2020.07.010. URL: http://www.sciencedirect.com/science/article/pii/S0893608020302513.

— (Nov. 1, 2021). "Detecting and learning from unknown by extremely weak supervision: exploratory classifier (xClass)". In: *Neural Computing and Applications* 33.22, pp. 15145–15157. ISSN: 1433-3058. DOI: 10.1007/s00521-021-06137-w. URL: https://doi.org/10.1007/s00521-021-06137-w.

Angelov, P. P. and X. Gu (Sept. 2018). "Toward Anthropomorphic Machine Learning". In: *Computer* 51.9. Conference Name: Computer, pp. 18–27. ISSN: 1558-0814. DOI: 10.1109/MC.2018.3620973.

Angelov, P. P., E. A. Soares, et al. (2021). "Explainable artificial intelligence: an analytical review". In: *WIREs Data Mining and Knowledge Discovery* 11.5, e1424. DOI: https://doi.org/10.1002/widm.1424.

Angelov, P. P. and X. Gu (2019). *Empirical Approach to Machine Learning*. en. Studies in Computational Intelligence. Springer International Publishing. ISBN: 978-3-030-02383-6. DOI: 10.1007/978-3-030-02384-3. URL: https://www.springer.com/gp/book/9783030023836.

Arnold, N. I., P. Angelov, and P. M. Atkinson (2022). "An Improved eXplainable Point Cloud Classifier (XPCC)". In: *IEEE Transactions on Artificial Intelligence*, pp. 1–1. DOI: 10.1109/TAI.2022.3150647.

— (2023). "An Explainable and Transferable AI Framework for the Extraction of Burial Ground Objects". English. In: *IEEE Transactions on Neural Networks and Learning Systems*. ISSN: 2162-237X.

Arnold, N. I., P. P. Angelov, et al. (Apr. 2021). "Automatic Extraction and Labelling of Memorial Objects From 3D Point Clouds". English. In: *Journal of Computer Applications in Archaeology* 4.1, pp. 79–93. DOI: 10.5334/jcaa.66.

Atkinson, P. M. and A. R. L. Tatnall (Mar. 1997). "Introduction Neural networks in remote sensing". In: *International Journal of Remote Sensing* 18.4. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/014311697218700, pp. 699–709. ISSN: 0143-1161. DOI: 10.1080/014311697218700. URL: https://doi.org/10.1080/014311697218700.

Aubry, M., U. Schlickewei, and D. Cremers (Nov. 2011). "The wave kernel signature: A quantum mechanical approach to shape analysis". In: pp. 1626–1633. DOI: 10.1109/ICCVW.2011.6130444.

Barredo Arrieta, A. et al. (June 2020a). "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI". en. In: *Information Fusion* 58, pp. 82–115. ISSN: 1566-2535. DOI: 10.1016/j.inffus.2019.12.012. URL: https://www.sciencedirect.com/science/article/pii/S1566253519308103.

— (2020b). "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI". In: *Information Fusion* 58, pp. 82–115. ISSN: 1566-2535. DOI: https://doi.org/10.1016/j.inffus.2019.12.012. URL: https://www.sciencedirect.com/science/article/pii/S1566253519308103.

Bello, S. A., S. Yu, and C. Wang (Jan. 2020). "Review: deep learning on 3D point clouds". In: *arXiv:2001.06280 [cs]*. arXiv: 2001.06280. URL: http://arxiv.org/abs/2001.06280.

Besl, P. J. and N. D. McKay (Feb. 1992). "A Method for Registration of 3-D Shapes". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14.2, pp. 239–256. ISSN: 0162-8828. DOI: 10.1109/34.121791. URL: https://doi.org/10.1109/34.121791.

Bewley, R. H., S. P. Crutchley, and C. A. Shell (Sept. 2005). "New light on an ancient landscape: lidar survey in the Stonehenge World Heritage Site". en. In: *Antiquity* 79.305. Publisher: Cambridge University Press, pp. 636–647. ISSN: 0003-598X, 1745-1744. DOI: 10.1017/S0003598X00114577. URL: https://www.cambridge.org/core/journals/antiquity/article/abs/new-light-on-an-ancient-

landscape-lidar-survey-in-the-stonehenge-world-heritage-site/D65EA1D86301827E17559610FDA3460B.

Bien, J. and R. Tibshirani (2011). "Prototype selection for interpretable classification". In: *The Annals of Applied Statistics* 5, pp. 2403–2424.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag. ISBN: 0387310738.

Bogacz, B. and H. Mara (Sept. 2020). "Period Classification of 3D Cuneiform Tablets with Geometric Neural Networks". In: *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pp. 246–251. DOI: 10.1109/ICFHR2020.2020.00053.

Boulch, A. (Apr. 2019). "ConvPoint: continuous convolutions for cloud processing". In: *arXiv:1904.02375 [cs]*. arXiv: 1904.02375. URL: http://arxiv.org/abs/1904.02375.

— (May 2020). "ConvPoint: Continuous convolutions for point cloud processing". en. In: *Computers & Graphics* 88, pp. 24–34. ISSN: 0097-8493. DOI: 10.1016/j.cag.2020.02.005. URL: https://www.sciencedirect.com/science/article/pii/S0097849320300224.

Boulch, A. et al. (Apr. 2018). "SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks". In: *Computers & Graphics* 71, pp. 189–198. ISSN: 0097-8493. DOI: 10.1016/j.cag.2017.11.010. URL: http://www.sciencedirect.com/science/article/pii/S0097849317301942.

Bouman, A. et al. (Oct. 2020). "Autonomous Spot: Long-Range Autonomous Exploration of Extreme Environments with Legged Locomotion". In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). ISSN: 2153-0866, pp. 2518–2525. DOI: 10.1109/IROS45743.2020.9341361.

Bozinovski, S. (Sept. 2020). "Reminder of the First Paper on Transfer Learning in Neural Networks, 1976". en. In: *Informatica* 44.3. Number: 3. ISSN: 1854-3871. DOI: 10.31449/inf.v44i3.2828. URL: https://www.informatica.si/index.php/informatica/article/view/2828.

Breiman, L. et al. (1983). "Classification and Regression Trees". In.

Bronstein, M. M., J. Bruna, et al. (July 2017). "Geometric deep learning: going beyond Euclidean data". In: *IEEE Signal Processing Magazine* 34.4, pp. 18–42. ISSN: 1053-5888. DOI: 10.1109/MSP.2017.2693418. URL: http://arxiv.org/abs/1611.08097.

Bronstein, M. M. and I. Kokkinos (June 2010). "Scale-invariant heat kernel signatures for non-rigid shape recognition". In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1704–1711. DOI: 10.1109/CVPR.2010.5539838.

Caesar, H. et al. (2020). "nuScenes: A Multimodal Dataset for Autonomous Driving". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. DOI: 10.1109/cvpr42600.2020.01164.

Cannell, R. J. S. et al. (Mar. 2018). "Delineating an Unmarked Graveyard by High-Resolution GPR and pXRF Prospection: The Medieval Church Site of Furulund in Norway". In: *Journal of Computer Applications in Archaeology* 1.1, pp. 1–18. ISSN: 2514-8362. DOI: 10.5334/jcaa.9. URL: http://journal.caa-international.org/articles/10.5334/jcaa.9/.

Cao, Y., M. Previtali, and M. Scaioni (Aug. 2020). "Understanding 3d Point Cloud Deep Neural Networks by Visualization Techniques". en. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLIII-B2-2020, pp. 651–657. ISSN: 2194-9034. DOI: 10.5194/isprs-archives-XLIII-B2-2020-651-2020. URL: https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLIII-B2-2020/651/2020/.

Chang, A. X. et al. (Dec. 2015). "ShapeNet: An Information-Rich 3D Model Repository". In: *arXiv:1512.03012 [cs]*. arXiv: 1512.03012. URL: http://arxiv.org/abs/1512.03012.

Chang, C.-C. and C.-J. Lin (May 2011). "LIBSVM: A Library for Support Vector Machines". In: *ACM Trans. Intell. Syst. Technol.* 2.3. ISSN: 2157-6904. DOI: 10.1145/1961189.1961199. URL: https://doi.org/10.1145/1961189.1961199.

Chase, A. S. Z., D. Z. Chase, and A. F. Chase (2017). "LiDAR for Archaeological Research and the Study of Historical Landscapes". In: *Sensing the Past: From artifact to historical site*. Ed. by N. Masini and F. Soldovieri. Cham: Springer International Publishing, pp. 89–100. ISBN: 978-3-319-50518-3. DOI: 10.1007/978-3-319-50518-3_4. URL: https://doi.org/10.1007/978-3-319-50518-3_4.

Chase, A. F. et al. (Feb. 2011). "Airborne LiDAR, archaeology, and the ancient Maya landscape at Caracol, Belize". en. In: *Journal of Archaeological Science* 38.2, pp. 387–398. ISSN: 0305-4403. DOI: 10.1016/j.jas.2010.09.018. URL: https://www.sciencedirect.com/science/article/pii/S0305440310003286.

Chattopadhyay, A. et al. (2017). "Grad-CAM++: Generalized Gradient-based Visual Explanations for Deep Convolutional Networks". In: *CoRR* abs/1710.11063. arXiv: 1710.11063. URL: http://arxiv.org/abs/1710.11063.

Crammer, K. and Y. Singer (2001). "On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines". In: *Journal of Machine Learning Research* 2.Dec, pp. 265–292. ISSN: ISSN 1533-7928. URL: https://jmlr.csail.mit.edu/papers/v2/crammer01a.

Dai, A. et al. (July 2017). "ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. ISSN: 1063-6919, pp. 2432–2443. DOI: 10.1109/CVPR.2017.261.

Dai, J. et al. (Mar. 2017). "Deformable Convolutional Networks". In: *arXiv:1703.06211 [cs]*. arXiv: 1703.06211. URL: http://arxiv.org/abs/1703.06211.

Demantké, J. et al. (Sept. 2012). "Dimensionality Based Scale Selection in 3d Lidar Point Clouds". In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXVIII-5/W12, pp. 97–102. ISSN: 1682-1777. DOI: 10.5194/isprsarchives-XXXVIII-5-W12-97-2011. URL: http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXXVIII-5-W12/97/2011/.

Deuge, M. et al. (2013). "Unsupervised feature learning for classification of outdoor 3D Scans". In: *Australasian Conference on Robotics and Automation, ACRA*.

Dick, S. (2019). "Artificial intelligence". In: Publisher: PubPub.

Dieber, J. and S. Kirrane (2020). "Why model why? Assessing the strengths and limitations of LIME". In: *CoRR* abs/2012.00093. arXiv: 2012.00093. URL: https://arxiv.org/abs/2012.00093.

Dinesh, C. et al. (2018). *Fast 3D Point Cloud Denoising via Bipartite Graph Approximation & Total Variation*. arXiv: 1804.10831 [eess.SP].

Dohan, D., B. Matejek, and T. Funkhouser (Oct. 2015). "Learning Hierarchical Semantic Segmentations of LIDAR Data". In: *2015 International Conference on 3D Vision*, pp. 273–281. DOI: 10.1109/3DV.2015.38.

Egmont-Petersen, M., D. de Ridder, and H. Handels (2002). "Image processing with neural networks—a review". In: *Pattern recognition* 35.10. Publisher: Elsevier, pp. 2279–2301.

Endres, F. et al. (Feb. 2014). "3-D Mapping With an RGB-D Camera". In: *IEEE Transactions on Robotics* 30.1. Conference Name: IEEE Transactions on Robotics, pp. 177–187. ISSN: 1941-0468. DOI: 10.1109/TRO.2013.2279412.

Evans, D. H. et al. (July 2013). "Uncovering archaeological landscapes at Angkor using lidar". In: *Proceedings of the National Academy of Sciences* 110.31. Publisher: Proceedings of the National Academy of Sciences, pp. 12595–12600. DOI: 10.1073/pnas.1306539110. URL: https://www.pnas.org/doi/full/10.1073/pnas.1306539110.

Fan, R.-E. et al. (June 2008). "LIBLINEAR: A Library for Large Linear Classification". In: *J. Mach. Learn. Res.* 9, pp. 1871–1874. ISSN: 1532-4435. DOI: 10.5555/1390681.1442794.

Favorskaya, M. N. and L. C. Jain (2017). "Overview of LiDAR Technologies and Equipment for Land Cover Scanning". In: *Handbook on Advances in Remote Sensing and Geograic Information Systems: Paradigms and Applications in Forest Landscape Modeling*. Ed. by M. N. Favorskaya and L. C. Jain. Intelligent Systems Reference Library. Cham: Springer International Publishing, pp. 19–68. ISBN: 978-3-319-52308-8. DOI: 10.1007/978-3-319-52308-8_2. URL: https://doi.org/10.1007/978-3-319-52308-8_2.

Figueroa, R. L. et al. (Feb. 2012). "Predicting sample size required for classification performance". In: *BMC Medical Informatics and Decision Making* 12.1, p. 8. ISSN: 1472-6947. DOI: 10.1186/1472-6947-12-8. URL: https://doi.org/10.1186/1472-6947-12-8.

Filin, S. and N. Pfeifer (June 2005). "Neighborhood Systems for Airborne Laser Data". In: *Photogrammetric Engineering & Remote Sensing* 71, pp. 743–755. DOI: 10.14358/PERS.71.6.743.

Fiorucci, M. et al. (2020). "Machine Learning for Cultural Heritage: A Survey". In: *Pattern Recognition Letters* 133, pp. 102–108. ISSN: 0167-8655. DOI: https://doi.org/10.1016/j.patrec.2020.02.017. URL: https://www.sciencedirect.com/science/article/pii/S0167865520300532.

Fradkin, D. and I. B. Muchnik (2015). "Support Vector Machines for Classification". In.

Freeland, T. et al. (May 2016). "Automated feature extraction for prospection and analysis of monumental earthworks from aerial LiDAR in the Kingdom of Tonga". en. In: *Journal of Archaeological Science* 69, pp. 64–74. ISSN: 0305-4403. DOI: 10.1016/j.jas.2016.04.011. URL: https://www.sciencedirect.com/science/article/pii/S0305440316300206.

Fryskowska, A. et al. (Aug. 2017). "Effective Detection of Sub-Surface Archeological Features from Laser Scanning Point Clouds and Imagery Data". In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-2/W5, pp. 245–251. DOI: 10.5194/isprs-archives-XLII-2-W5-245-2017.

Gallwey, J. et al. (Jan. 2019). "Bringing Lunar LiDAR Back Down to Earth: Mapping Our Industrial Heritage through Deep Transfer Learning". en. In: *Remote Sensing* 11.17. Number: 17 Publisher: Multidisciplinary Digital Publishing Institute, p. 1994. ISSN: 2072-4292. DOI: 10.3390/rs11171994. URL: https://www.mdpi.com/2072-4292/11/17/1994.

Geiger, A., P. Lenz, and R. Urtasun (June 2012). "Are we ready for autonomous driving? The KITTI vision benchmark suite". In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. ISSN: 1063-6919, pp. 3354–3361. DOI: 10.1109/CVPR.2012.6248074.

Golovinskiy, A., V. G. Kim, and T. Funkhouser (Sept. 2009). "Shape-based recognition of 3D point clouds in urban environments". In: IEEE, pp. 2154–2161. ISBN: 978-1-4244-4420-5. DOI: 10.1109/ICCV.2009.5459471. URL: http://ieeexplore.ieee.org/document/5459471/.

Goodfellow, I. J. et al. (June 2014). "Generative Adversarial Networks". In: *arXiv e-prints*, arXiv:1406.2661, arXiv:1406.2661. DOI: 10.48550/arXiv.1406.2661. arXiv: 1406.2661 [stat.ML].

Goyal, A. et al. (2021). *Revisiting Point Cloud Shape Classification with a Simple and Effective Baseline*. arXiv: 2106.05304 [cs.CV].

Graham, B. (Aug. 2015). "Sparse 3D convolutional neural networks". In: arXiv: 1505.02890. URL: http://arxiv.org/abs/1505.02890.

Graham, B., M. Engelcke, and L. van der Maaten (Nov. 2017). "3D Semantic Segmentation with Submanifold Sparse Convolutional Networks". In: *arXiv:1711.10275 [cs]*. arXiv: 1711.10275. URL: http://arxiv.org/abs/1711.10275.

Guan, H. et al. (2016). "Use of mobile LiDAR in road information inventory: A review". In: *International Journal of Image and Data Fusion* 7.3. Publisher: Taylor & Francis, pp. 219–242.

Guinard, S. and L. Landrieu (May 2017). "Weakly Supervised Segmentation-Aided Classification of Urban Scenes from 3d Lidar Point Clouds". en. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-1/W1, pp. 151–157. ISSN: 2194-9034. DOI: 10.5194/isprs-archives-XLII-1-W1-151-2017. URL: http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-1-W1/151/2017/.

Gunji, N. et al. (2016). "3D object recognition from large-scale point clouds with global descriptor and sliding window". In: *2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 721–726. DOI: 10.1109/ICPR.2016.7899720.

Guo, Y. et al. (2018). "A review of semantic segmentation using deep neural networks". In: *International journal of multimedia information retrieval* 7.2. Publisher: Springer, pp. 87–93.

Guyot, A., L. Hubert-Moy, and T. Lorho (Feb. 2018). "Detecting Neolithic Burial Mounds from LiDAR-Derived Elevation Data Using a Multi-Scale Approach and Machine Learning Techniques". In: *Remote Sensing* 10.2, p. 225. DOI: 10.3390/rs10020225. URL: https://www.mdpi.com/2072-4292/10/2/225.

Hackel, T., N. Savinov, et al. (May 2017). "Semantic3d.net: A New Large-Scale Point Cloud Classification Benchmark". en. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-1/W1, pp. 91–98. ISSN: 2194-9050. DOI: 10.5194/isprs-annals-IV-1-W1-91-2017. URL: http://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/IV-1-W1/91/2017/.

Hackel, T., J. D. Wegner, and K. Schindler (June 2016). "Contour Detection in Unstructured 3D Point Clouds". In: IEEE, pp. 1610–1618. ISBN: 978-1-4673-8851-1. DOI: 10.1109/CVPR.2016.178. URL: http://ieeexplore.ieee.org/document/7780547/.

Hassan, A. and D. Fritsch (2020). "Integration of Laser Scanning and Photogrammetry in 3D/4D Cultural Heritage Preservation – A Review". In: URL: https://api.semanticscholar.org/CorpusID:220513782.

Hastie, T. et al. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer.

Hearst, M. et al. (1998). "Support vector machines". In: *IEEE Intelligent Systems and their Applications* 13.4, pp. 18–28. DOI: 10.1109/5254.708428.

Hermosilla, P. et al. (Dec. 2018). "Monte Carlo convolution for learning on non-uniformly sampled point clouds". In: *ACM Transactions on Graphics* 37.6, 235:1–235:12. ISSN: 0730-0301. DOI: 10.1145/3272127.3275110. URL: https://doi.org/10.1145/3272127.3275110.

Hornik, K., M. Stinchcombe, and H. White (1990). "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks". In: *Neural Networks* 3.5, pp. 551–560. ISSN: 0893-6080. DOI: https://doi.org/10.1016/0893-6080(90)90005-6. URL: https://www.sciencedirect.com/science/article/pii/0893608090900056.

Hu, Q. et al. (2019). "RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds". In: *CoRR* abs/1911.11236. arXiv: 1911.11236. URL: http://arxiv.org/abs/1911.11236.

Hu, W., X. Gao, et al. (2019). "Feature Graph Learning for 3D Point Cloud Denoising". In: *CoRR* abs/1907.09138. arXiv: 1907.09138. URL: http://arxiv.org/abs/1907.09138.

Hu, W., Q. Hu, et al. (2021). "Dynamic Point Cloud Denoising via Manifold-to-Manifold Distance". In: *IEEE Transactions on Image Processing* 30, pp. 6168–6183. DOI: 10.1109/tip.2021.3092826. URL: https://doi.org/10.1109%2Ftip.2021.3092826.

Huang, S. et al. (Dec. 20, 2019). "A CLAIM Approach to Understanding the PointNet". In: *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*. ACAI 2019. New York, NY, USA: Association for Computing Machinery, pp. 97–103. ISBN: 978-1-4503-7261-9. DOI: 10.1145/3377713.3377740. URL: https://doi.org/10.1145/3377713.3377740.

Ibrahim, M. et al. (2019). "Global Explanations of Neural Networks: Mapping the Landscape of Predictions". In: *CoRR* abs/1902.02384. arXiv: 1902.02384. URL: http://arxiv.org/abs/1902.02384.

Inomata, T. et al. (June 2020). "Monumental architecture at Aguada Fénix and the rise of Maya civilization". en. In: *Nature* 582.7813. Number: 7813 Publisher: Nature Publishing Group, pp. 530–533. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2343-4. URL: https://www.nature.com/articles/s41586-020-2343-4.

Jaklič, A. et al. (Oct. 2015). "Volumetric Models from 3d Point Clouds: The Case Study of Sarcophagi Cargo from a 2nd/3rd Century Ad Roman Shipwreck Near Sutivan on Island Brač, Croatia". In: *Journal of Archaeological Science* 62, pp. 143–152. ISSN: 0305-4403. DOI: 10.1016/j.jas.2015.08.007. URL: http://www.sciencedirect.com/science/article/pii/S0305440315002447.

Johnson, A. E. and M. Hebert (July 1998). "Surface matching for object recognition in complex three-dimensional scenes". en. In: *Image and Vision Computing* 16.9, pp. 635–651. ISSN: 0262-8856. DOI: 10.1016/S0262-8856(98)00074-2. URL: https://www.sciencedirect.com/science/article/pii/S0262885698000742.

— (May 1999). "Using spin images for efficient object recognition in cluttered 3D scenes". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21.5, pp. 433–449. ISSN: 0162-8828. DOI: 10.1109/34.765655.

Jonathan, E. et al. (Aug. 2001). "Local Versus Global Triangulations". In.

Kazimi, B. et al. (Sept. 2018). "Deep Learning for Archaeological Object Detection in Airborne Laser Scanning Data". In.

Kim, B. et al. (2017). "Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV)". In: *International Conference on Machine Learning*. URL: https://api.semanticscholar.org/CorpusID:51737170.

Kingma, D. P. and J. Ba (2017). *Adam: A Method for Stochastic Optimization*. arXiv: 1412.6980 [cs.LG].

Kirillov, A. et al. (2019). "Panoptic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9404–9413.

Klokov, R. and V. Lempitsky (Oct. 2017). "Escape from Cells: Deep Kd-Networks for the Recognition of 3D Point Cloud Models". In: *2017 IEEE International Conference on Computer Vision (ICCV)*. event-place: Venice. IEEE, pp. 863–872. ISBN: 978-1-5386-1032-9. DOI: 10.1109/ICCV.2017.99. URL: http://ieeexplore.ieee.org/document/8237361/.

Kohonen, T. (Sept. 1990). "The self-organizing map". In: *Proceedings of the IEEE* 78.9. Conference Name: Proceedings of the IEEE, pp. 1464–1480. ISSN: 1558-2256. DOI: 10.1109/5.58325.

Kohonen, T. et al. (1992). "LVQPAK: A software package for the correct application of Learning Vector Quantization algorithms". In: *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks*. Vol. 1, 725–730 vol.1. DOI: 10.1109/IJCNN.1992.287101.

Kokalj, Ž. and M. Somrak (Jan. 2019). "Why Not a Single Image? Combining Visualizations to Facilitate Fieldwork and On-Screen Mapping". en. In: *Remote Sensing* 11.7. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute, p. 747. ISSN: 2072-4292. DOI: 10.3390/rs11070747. URL: https://www.mdpi.com/2072-4292/11/7/747.

Koutroumbas, K. and S. Theodoridis (2008). *Pattern recognition*. Academic Press.

Kramer, I. (2015). "An archaeological reaction to the remote sensing data explosion. Reviewing the research on semi-automated pattern recognition and assessing the potential to integrate artificial intelligence". PhD thesis.

Lambers, K., W. B. Verschoof-van der Vaart, and Q. P. J. Bourgeois (Jan. 2019). "Integrating Remote Sensing, Machine Learning, and Citizen Science in Dutch Archaeological Prospection". en. In: *Remote Sensing* 11.7. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute, p. 794. ISSN: 2072-4292. DOI: 10.3390/rs11070794. URL: https://www.mdpi.com/2072-4292/11/7/794.

Landrieu, L. and G. Obozinski (2017). "Cut Pursuit: fast algorithms to learn piecewise constant functions on general weighted graphs". In: *SIAM Journal on Imaging Sciences* Vol. 10.No. 4, pp. 1724–1766. URL: https://hal.archives-ouvertes.fr/hal-01306779.

Landrieu, L. and M. Simonovsky (2017). "Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs". In: *CoRR* abs/1711.09869. URL: http://arxiv.org/abs/1711.09869.

— (June 2018). "Large-Scale Point Cloud Semantic Segmentation with Superpoint Graphs". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. ISSN: 2575-7075, pp. 4558–4567. DOI: 10.1109/CVPR.2018.00479.

Langley, P. (Mar. 2011). "The changing science of machine learning". en. In: *Machine Learning* 82.3, pp. 275–279. ISSN: 1573-0565. DOI: 10.1007/s10994-011-5242-y. URL: https://doi.org/10.1007/s10994-011-5242-y.

LeCun, Y., Y. Bengio, and G. Hinton (May 2015). "Deep learning". en. In: *Nature* 521.7553. Number: 7553 Publisher: Nature Publishing Group, pp. 436–444. ISSN: 1476-4687. DOI: 10.1038/nature14539. URL: https://www.nature.com/articles/nature14539.

Lee, I. and A. Schenk (Jan. 2002). "Perceptual organization of 3D surface points". In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 34.

Lee, S., Z. Majid, and H. Setan (2013). "3D data acquisition for indoor assets using terrestrial laser scanning". In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. II-2 W* 1. Publisher: Citeseer, pp. 221–226.

Leica, G. (n.d.). *Digitising environments: A map of the future*. en-GB. URL: https://leica-geosystems.com/en-GB/products/mobile-mapping-systems/stories-insights/digitising-environments.

Ling, H. and D. W. Jacobs (Feb. 2007). "Shape Classification Using the Inner-Distance". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.2, pp. 286–299. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2007.41.

Liu, Z. et al. (2019). "TANet: Robust 3D Object Detection from Point Clouds with Triple Attention". In: *CoRR* abs/1912.05163. arXiv: 1912.05163. URL: http://arxiv.org/abs/1912.05163.

Lloyd, C. D. and P. M. Atkinson (2006). "Deriving ground surface digital elevation models from LiDAR data with geostatistics". In: *International Journal of Geographical Information Science* 20.5. Publisher: Taylor & Francis

_eprint: https://doi.org/10.1080/13658810600607337, pp. 535–563. DOI: 10.1080/13658810600607337. URL: https://doi.org/10.1080/13658810600607337.

Lundberg, S. M. and S.-I. Lee (2017). "A Unified Approach to Interpreting Model Predictions". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS'17. Long Beach, California, USA: Curran Associates Inc., pp. 4768–4777. ISBN: 9781510860964.

Luo, S. and W. Hu (2021). "Score-Based Point Cloud Denoising". In: *CoRR* abs/2107.10981. arXiv: 2107.10981. URL: https://arxiv.org/abs/2107.10981.

Maturana, D. and S. Scherer (Sept. 2015). "VoxNet: A 3D Convolutional Neural Network for real-time object recognition". en. In: *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Hamburg, Germany: IEEE, pp. 922–928. ISBN: 978-1-4799-9994-1. DOI: 10.1109/IROS.2015.7353481. URL: http://ieeexplore.ieee.org/document/7353481/.

Meng Joo Er et al. (May 2002). "Face recognition with radial basis function (RBF) neural networks". In: *IEEE Transactions on Neural Networks* 13.3. Conference Name: IEEE Transactions on Neural Networks, pp. 697–710. ISSN: 1941-0093. DOI: 10.1109/TNN.2002.1000134.

Mikhail, E. M., J. S. Bethel, and J. C. McGlone (2001). *Introduction to modern photogrammetry*. John Wiley & Sons.

Muhammad, M. B. and M. Yeasin (2020). "Eigen-CAM: Class Activation Map using Principal Components". In: *CoRR* abs/2008.00299. arXiv: 2008.00299. URL: https://arxiv.org/abs/2008.00299.

Niculiță, M. (Jan. 2020). "Geomorphometric Methods for Burial Mound Recognition and Extraction from High-Resolution LiDAR DEMs". en. In: *Sensors* 20.4. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute, p. 1192. ISSN: 1424-8220. DOI: 10.3390/s20041192. URL: https://www.mdpi.com/1424-8220/20/4/1192.

Niemeyer, J., F. Rottensteiner, and U. Soergel (Jan. 2014). "Contextual classification of lidar data and building object detection in urban areas". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 87, pp. 152–165. ISSN: 0924-2716. DOI: 10.1016/j.isprsjprs.2013.11.001. URL: http://www.sciencedirect.com/science/article/pii/S0924271613002359.

Niemeyer, J., J. D. Wegner, et al. (2011). "Conditional Random Fields for Urban Scene Classification with Full Waveform LiDAR Data". In: *Photogrammetric Image Analysis*. Ed. by U. Stilla et al. Lecture Notes in Computer Science. event-place: Berlin, Heidelberg. Springer, pp. 233–244. ISBN: 978-3-642-24393-6. DOI: 10.1007/978-3-642-24393-6_20.

Olivier, M. and W. V.-v. d. Vaart (Dec. 8, 2021). "Implementing State-of-the-Art Deep Learning Approaches for Archaeological Object Detection in Remotely-Sensed Data: The Results of Cross-Domain Collaboration". In: *Journal of*

Computer Applications in Archaeology 4.1. Number: 1 Publisher: Ubiquity Press, pp. 274–289. ISSN: 2514-8362. DOI: 10.5334/jcaa.78. URL: http://journal.caa-international.org/articles/10.5334/jcaa.78/.

Omohundro, S. M. (1989). *Five Balltree Construction Algorithms*. Tech. rep.

OpenTopography — About (n.d.). URL: https://opentopography.org/about.

Opitz, R. and J. Herrmann (May 2018). "Recent Trends and Long-standing Problems in Archaeological Remote Sensing". eng. In: *Journal of Computer Applications in Archaeology* 1.1. Number: 1 Publisher: Ubiquity Press, pp. 19–41. ISSN: 2514-8362. DOI: 10.5334/jcaa.11. URL: http://journal.caa-international.org/articles/10.5334/jcaa.11/.

Pang, G. and U. Neumann (2016). "3D point cloud object detection with multi-view convolutional neural network". In: *2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 585–590. DOI: 10.1109/ICPR.2016.7899697.

Pasquale, F. (2016). *Black Box Society: The Secret Algorithms that Control Money and Information*. eng. First Harvard University Press paperback edition. Cambridge, Massachusetts: Harvard University Press. ISBN: 9780674368279. URL: https://worldcat.org/title/946975299.

Pedregosa, F. et al. (2011a). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

— (2011b). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Pfeifer, N. and C. Briese (2007). "Geometrical aspects of airborne laser scanning and terrestrial laser scanning". In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36.3. Publisher: ISPRS Vienna, Austria, pp. 311–319.

Pistilli, F. et al. (2020). "Learning Graph-Convolutional Representationsfor Point Cloud Denoising". In: *The European Conference on Computer Vision (ECCV)*.

Qi, C. R., X. Chen, et al. (2020). "ImVoteNet: Boosting 3D Object Detection in Point Clouds with Image Votes". In: *CoRR* abs/2001.10692. arXiv: 2001.10692. URL: https://arxiv.org/abs/2001.10692.

Qi, C. R., H. Su, K. Mo, et al. (Dec. 2016). "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation". In: *arXiv:1612.00593 [cs]*. URL: http://arxiv.org/abs/1612.00593.

Qi, C. R., H. Su, M. Niessner, et al. (Apr. 2016). "Volumetric and Multi-View CNNs for Object Classification on 3D Data". In: *arXiv:1604.03265 [cs]*. URL: http://arxiv.org/abs/1604.03265.

Qi, C. R., L. Yi, et al. (June 2017a). "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space". In: *arXiv:1706.02413 [cs]*. URL: http://arxiv.org/abs/1706.02413.

Qi, C. R., L. Yi, et al. (Dec. 2017b). "PointNet++: deep hierarchical feature learning on point sets in a metric space". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., pp. 5105–5114. ISBN: 978-1-5108-6096-4.

Rakotosaona, M. et al. (2019). "POINTCLEANNET: Learning to Denoise and Remove Outliers from Dense Point Clouds". In: *CoRR* abs/1901.01060. arXiv: 1901.01060. URL: http://arxiv.org/abs/1901.01060.

Raschka, S. (2018). "STAT 479: Machine Learning Lecture Notes". en. In: *Machine Learning*, p. 23.

Ren, Y. et al. (2022). "Multi-Scale Upsampling GAN Based Hole-Filling Framework for High-Quality 3D Cultural Heritage Artifacts". In: *Applied Sciences* 12.9. ISSN: 2076-3417. DOI: 10.3390/app12094581. URL: https://www.mdpi.com/2076-3417/12/9/4581.

Riegler, G., A. O. Ulusoy, and A. Geiger (Apr. 2017). "OctNet: Learning Deep 3D Representations at High Resolutions". In: *arXiv:1611.05009 [cs]*. arXiv: 1611.05009. URL: http://arxiv.org/abs/1611.05009.

Rodríguez-Gonzálvez, P. et al. (2017). "Mobile LiDAR system: New possibilities for the documentation and dissemination of large cultural heritage sites". In: *Remote Sensing* 9.3. Publisher: MDPI, p. 189.

Royo, S. and M. Ballesta-Garcia (Jan. 2019). "An Overview of Lidar Imaging Systems for Autonomous Vehicles". In: *Applied Sciences* 9.19, p. 4093. DOI: 10.3390/app9194093. URL: https://www.mdpi.com/2076-3417/9/19/4093.

Rudin, C. (May 2019). "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead". en. In: *Nature Machine Intelligence* 1.5, pp. 206–215. ISSN: 2522-5839. DOI: 10.1038/s42256-019-0048-x. URL: https://www.nature.com/articles/s42256-019-0048-x.

Rusu, R. B., G. Bradski, et al. (Oct. 2010). "Fast 3D recognition and pose using the Viewpoint Feature Histogram". In: *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. event-place: Taipei. IEEE, pp. 2155–2162. ISBN: 978-1-4244-6674-0. DOI: 10.1109/IROS.2010.5651280. URL: http://ieeexplore.ieee.org/document/5651280/.

Rusu, R. B., N. Blodow, and M. Beetz (May 2009). "Fast Point Feature Histograms (FPFH) for 3D registration". In: *2009 IEEE International Conference on Robotics and Automation*. ISSN: 1050-4729, pp. 3212–3217. DOI: 10.1109/ROBOT.2009.5152473.

Rusu, R. B. and S. Cousins (May 2011). "3D is here: Point Cloud Library (PCL)". In: *2011 IEEE International Conference on Robotics and Automation*, pp. 1–4. DOI: 10.1109/ICRA.2011.5980567.

Samek, W. et al. (Nov. 2017). "Evaluating the Visualization of What a Deep Neural Network Has Learned". In: *IEEE Transactions on Neural Networks and Learning*

Systems 28.11. Conference Name: IEEE Transactions on Neural Networks and Learning Systems, pp. 2660–2673. ISSN: 2162-2388. DOI: 10.1109/TNNLS.2016.2599820.

Selvaraju, R. R. et al. (2016). "Grad-CAM: Why did you say that? Visual Explanations from Deep Networks via Gradient-based Localization". In: *CoRR* abs/1610.02391. arXiv: 1610.02391. URL: http://arxiv.org/abs/1610.02391.

Sevara, C. et al. (2016). "Pixel versus object—A comparison of strategies for the semi-automated mapping of archaeological features using airborne laser scanning data". In: *Journal of Archaeological Science: Reports* 5, pp. 485–498.

Shakhnarovich, G., T. Darrell, and P. Indyk (2008). "Nearest-neighbor methods in learning and vision". In: *IEEE Trans. Neural Networks* 19.2, p. 377.

Shapovalov, R., E. Velizhev, and O. Barinova (2010). "Nonassociative markov networks for 3d point cloud classification." In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII, Part 3A*, pp. 103–108.

Shen, Y. et al. (June 2018). "Mining Point Cloud Local Structures by Kernel Correlation and Graph Pooling". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. ISSN: 2575-7075, pp. 4548–4557. DOI: 10.1109/CVPR.2018.00478.

Shorten, C. and T. M. Khoshgoftaar (July 2019). "A survey on Image Data Augmentation for Deep Learning". In: *Journal of Big Data* 6.1, p. 60. ISSN: 2196-1115. DOI: 10.1186/s40537-019-0197-0. URL: https://doi.org/10.1186/s40537-019-0197-0.

Silva do Monte Lima, J. P. and V. Teichrieb (Oct. 2016). "An Efficient Global Point Cloud Descriptor for Object Recognition and Pose Estimation". In: *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 56–63. DOI: 10.1109/SIBGRAPI.2016.017.

Simonovsky, M. and N. Komodakis (Aug. 2017). "Dynamic Edge-Conditioned Filters in Convolutional Neural Networks on Graphs". In: *arXiv:1704.02901 [cs]*. arXiv: 1704.02901. URL: http://arxiv.org/abs/1704.02901.

Simonyan, K., A. Vedaldi, and A. Zisserman (Apr. 19, 2013a). "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps". In: *arXiv:1312.6034 [cs]*. arXiv: 1312.6034. URL: http://arxiv.org/abs/1312.6034.

— (2013b). "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps". In: *CoRR* abs/1312.6034. URL: https://api.semanticscholar.org/CorpusID:1450294.

Soares, E. A. et al. (2020). "Explaining Deep Learning Models Through Rule-Based Approximation and Visualization". In: *IEEE Transactions on Fuzzy Systems*.

Conference Name: IEEE Transactions on Fuzzy Systems, pp. 1–1. ISSN: 1941-0034. DOI: 10.1109/TFUZZ.2020.2999776.

Spina, S. et al. (Jan. 2011). "Point Cloud Segmentation for Cultural Heritage Sites." In: pp. 41–48. DOI: 10.2312/VAST/VAST11/041-048.

Stilgoe, J. (2020). "Who Killed Elaine Herzberg?" In: *Who's Driving Innovation? New Technologies and the Collaborative State*. Cham: Springer International Publishing, pp. 1–6. ISBN: 978-3-030-32320-2. DOI: 10.1007/978-3-030-32320-2_1. URL: https://doi.org/10.1007/978-3-030-32320-2_1.

Su, H., V. Jampani, et al. (Feb. 2018). "SPLATNet: Sparse Lattice Networks for Point Cloud Processing". In: *arXiv:1802.08275 [cs]*. arXiv: 1802.08275. URL: http://arxiv.org/abs/1802.08275.

Su, H., S. Maji, et al. (2015). "Multi-view convolutional neural networks for 3d shape recognition". In: *Proc. ICCV*.

Taghanaki, S. A. et al. (July 2020). "PointMask: Towards Interpretable and Bias-Resilient Point Cloud Processing". In: *arXiv:2007.04525 [cs]*. arXiv: 2007.04525. URL: http://arxiv.org/abs/2007.04525.

Thomas, H. (2019). "Learning new representations for 3D point cloud semantic segmentation". en. In: p. 179.

Thomas, H. et al. (Oct. 2019). "KPConv: Flexible and Deformable Convolution for Point Clouds". In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. ISSN: 2380-7504, pp. 6410–6419. DOI: 10.1109/ICCV.2019.00651.

Trier, Ø. et al. (Apr. 2016). "Semi-automatic mapping of cultural heritage from airborne laser scanning using deep learning". In.

Trier, Ø. D., D. C. Cowley, and A. U. Waldeland (2019). "Using deep neural networks on airborne laser scanning data: Results from a case study of semi-automatic mapping of archaeological topography on Arran, Scotland". In: *Archaeological Prospection* 26.2, pp. 165–175. DOI: https://doi.org/10.1002/arp.1731. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/arp.1731.

Trier, Ø. D. and L. H. Pilø (2012). "Automatic Detection of Pit Structures in Airborne Laser Scanning Data". en. In: *Archaeological Prospection* 19.2. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/arp.1421, pp. 103–121. ISSN: 1099-0763. DOI: 10.1002/arp.1421. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/arp.1421.

Tritscher, J. et al. (2020). "Evaluation of Post-hoc XAI Approaches Through Synthetic Tabular Data". In: *Foundations of Intelligent Systems*. Ed. by D. Helic et al. Cham: Springer International Publishing, pp. 422–430. ISBN: 978-3-030-59491-6.

Uy, M. A. et al. (2019). "Revisiting Point Cloud Classification: A New Benchmark Dataset and Classification Model on Real-World Data". In: *International Conference on Computer Vision (ICCV)*.

Vaart, W. B. V.-v. d. and K. Lambers (Mar. 2019). "Learning to Look at LiDAR: The Use of R-CNN in the Automated Detection of Archaeological Objects in LiDAR Data from the Netherlands". en. In: *Journal of Computer Applications in Archaeology* 2.1. Number: 1 Publisher: Ubiquity Press, pp. 31–40. ISSN: 2514-8362. DOI: 10.5334/jcaa.32. URL: http://journal.caa-international.org/articles/10.5334/jcaa.32/.

Verschoof-van der Vaart, W. B. and K. Lambers (Mar. 2019). "Learning to Look at Lidar: The Use of R-CNN in the Automated Detection of Archaeological Objects in Lidar Data from the Netherlands". In: *Journal of Computer Applications in Archaeology* 2.1, pp. 31–40. ISSN: 2514-8362. DOI: 10.5334/jcaa.32. URL: http://journal.caa-international.org/articles/10.5334/jcaa.32/.

Wang, P.-S. et al. (July 2017). "O-CNN: Octree-based Convolutional Neural Networks for 3D Shape Analysis". In: *ACM Transactions on Graphics* 36.4. arXiv: 1712.01537, pp. 1–11. ISSN: 0730-0301, 1557-7368. DOI: 10.1145/3072959.3073608. URL: http://arxiv.org/abs/1712.01537.

Wei, X., R. Yu, and J. Sun (June 2020). "View-GCN: View-Based Graph Convolutional Network for 3D Shape Analysis". en. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE, pp. 1847–1856. ISBN: 978-1-72817-168-5. DOI: 10.1109/CVPR42600.2020.00192. URL: https://ieeexplore.ieee.org/document/9156567/.

Weinmann, M. et al. (Mar. 2015). "Contextual Classification of Point Cloud Data by Exploiting Individual 3d Neigbourhoods". In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. II-3-W4. Copernicus GmbH, pp. 271–278. DOI: https://doi.org/10.5194/isprsannals-II-3-W4-271-2015. URL: https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/II-3-W4/271/2015/.

Weiss, K., T. M. Khoshgoftaar, and D. Wang (May 2016). "A survey of transfer learning". In: *Journal of Big Data* 3.1, p. 9. ISSN: 2196-1115. DOI: 10.1186/s40537-016-0043-6. URL: https://doi.org/10.1186/s40537-016-0043-6.

Weitman, S. (May 2012). "Using Archaeological Methods in Cemetery Surveys with Emphasis on the Application of Lidar". In: *Georgia Southern University*. URL: https://digitalcommons.georgiasouthern.edu/etd/627.

Werbos, P. (1990). "Backpropagation through time: what it does and how to do it". In: *Proceedings of the IEEE* 78.10, pp. 1550–1560. DOI: 10.1109/5.58337.

Wohlkinger, W. and M. Vincze (Dec. 2011). "Ensemble of shape functions for 3D object classification". In: *2011 IEEE International Conference on Robotics and Biomimetics*, pp. 2987–2992. DOI: 10.1109/ROBIO.2011.6181760.

Wolf, P. R., B. A. Dewitt, and B. E. Wilkinson (2014). *Elements of Photogrammetry with Applications in GIS*. McGraw-Hill Education.

Wu, B., A. Wan, et al. (2018). "SqueezeSeg: Convolutional Neural Nets with Recurrent CRF for Real-Time Road-Object Segmentation from 3D LiDAR Point Cloud". In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1887–1893. DOI: 10.1109/ICRA.2018.8462926.

Wu, B., X. Zhou, et al. (2019). "SqueezeSegV2: Improved Model Structure and Unsupervised Domain Adaptation for Road-Object Segmentation from a LiDAR Point Cloud". In: *2019 International Conference on Robotics and Automation (ICRA)*, pp. 4376–4382. DOI: 10.1109/ICRA.2019.8793495.

Wu, Z. et al. (June 2015). "3D ShapeNets: A deep representation for volumetric shapes". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1912–1920. DOI: 10.1109/CVPR.2015.7298801.

Xu, Q. et al. (2020). "Grid-GCN for Fast and Scalable Point Cloud Learning". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. DOI: 10.1109/cvpr42600.2020.00570.

Yu, T., J. Meng, and J. Yuan (June 2018). "Multi-view Harmonized Bilinear Network for 3D Object Recognition". en. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT: IEEE, pp. 186–194. ISBN: 978-1-5386-6420-9. DOI: 10.1109/CVPR.2018.00027. URL: https://ieeexplore.ieee.org/document/8578125/.

Zacharek, M. et al. (May 2017). "Generating Accurate 3d Models of Architectural Heritage Structures Using Low-Cost Camera and Open Source Algorithms". In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-5/W1, pp. 99–104. ISSN: 2194-9034. DOI: 10.5194/isprs-archives-XLII-5-W1-99-2017. URL: https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-5-W1/99/2017/.

Zaheer, M. et al. (2017). "Deep Sets". en. In: *Advances in Neural Information Processing Systems* 30. URL: https://papers.nips.cc/paper/2017/hash/f22e4747da1aa27e363d86d40ff442fe-Abstract.html.

Zeng, Z. et al. (2022). *LACV-Net: Semantic Segmentation of Large-Scale Point Cloud Scene via Local Adaptive and Comprehensive VLAD*. arXiv: 2210.05870 [cs.CV].

Zhang, B. et al. (2019). "Explaining the PointNet: What Has Been Learned Inside the PointNet?" In: *CVPR Workshops*.

Zhang, H. (2004). "The Optimality of Naive Bayes". In: *FLAIRS Conference*.

Zhang, M. et al. (July 2020). "PointHop: An Explainable Machine Learning Method for Point Cloud Classification". In: *IEEE Transactions on Multimedia* 22.7. arXiv: 1907.12766, pp. 1744–1755. ISSN: 1520-9210, 1941-0077. DOI: 10.1109/TMM.2019.2963592. URL: http://arxiv.org/abs/1907.12766.

Zhu, X. et al. (2020). "SSN: Shape Signature Networks for Multi-class Object Detection from Point Clouds". In: *CoRR* abs/2004.02774. arXiv: 2004.02774. URL: https://arxiv.org/abs/2004.02774.

Zingman, I. et al. (Aug. 2016). "Detection of Fragmented Rectangular Enclosures in Very High Resolution Remote Sensing Images". In: *IEEE Transactions on Geoscience and Remote Sensing* 54.8, pp. 4580–4593. ISSN: 1558-0644. DOI: 10.1109/TGRS.2016.2545919.

Zulqarnain, M. et al. (2020). "A comparative review on deep learning models for text classification". In: *Indones. J. Electr. Eng. Comput. Sci* 19.1, pp. 325–335.