ITcon

www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

NAVIGATING GEOMETRIC COMPLEXITY IN DIGITAL HERITAGE: A REVIEW OF AI-BASED SEMANTIC SEGMENTATION

SUBMITTED: September 2025
REVISED: October 2025
PUBLISHED: November 2025
EDITOR: Mahesh Babu Purushothaman, Ali GhaffarianHoseini, Amirhosein Ghaffarianhoseini, Farzad Rahimian
DOI: 10.36680/j.itcon.2025.070

Subhadha Battina
Florida International University
ORCID: https://orcid.org/0000-0002-6100-9258
email: sbattina26@gmail.com

SUMMARY: *Utilizing tools like laser scanners and photogrammetry to generate point cloud data reshapes Digital Heritage by facilitating Scan-to-BIM methodologies for 3D models. At point cloud processing stage, integrating semantic segmentation into Scan-to-BIM workflows allows unstructured spatial information to be translated into intelligent geometrical classifications that enable data-driven 3D models. AI methods show promising solutions for Three-Dimensional Point Cloud Semantic Segmentation (3DPCSS), allowing robust creation of parametric objects in heritage BIM. Despite the advancements in AI-based 3DPCSS, current models often present conceptual and practical challenges. These limitations stem from heterogeneous data and lack of adaptive algorithms to capture the geometrical complexities inherent in historic structures. Manual segmentation is often required to add detail to the simplified geometrical representations that omit the unique features, such as intricate carvings, creating challenges in efficient modeling. This systematic literature review presents an inquiry into the workflow of 3DPCSS, from data acquisition and classification stages to 3D model creation, with a case example in an Indian context. It synthesizes findings from 95 peer-reviewed publications from 2014 to 2024, focusing on the factors influencing the selection of suitable AI algorithms, including data acquisition, dataset types, complexity of geometrical elements, and computational tasks. The investigation reveals significant limitations in current approaches. Transformer-based models demonstrate significant performance degradation when applied to non-Western architectural geometries despite comparable complexity levels. Furthermore, through a matrix analysis, we identify four primary phases of algorithmic evolution—from rule-based systems to transformer architectures—while highlighting the emergence of hybrid approaches that combine geometric primitives with deep learning refinement. This paper extends the work presented at the Proceedings of the International Conference on Smart and Sustainable Built Environment (SASBE 2024), New Zealand.*

KEYWORDS: *AI in cultural heritage, scan-to-BIM, 3D point cloud semantic segmentation (3DPSS), digital heritage, data-driven 3D modeling.*

REFERENCE: *Subhadha Battina (2025). Navigating geometric complexity in digital heritage: a review of AI-based semantic segmentation. Journal of Information Technology in Construction (ITcon), Special issue: 'Smart and Sustainable Built Environment (SASBE 2024)', Vol. 30, pg. 1707-1727, DOI: 10.36680/j.itcon.2025.070*

1. INTRODUCTION

Digital heritage has undergone a significant transformation with the advent of advanced 3D data-collecting technologies such as LiDAR (Light Detection and Ranging) with photogrammetric methods (Capolupo, et al 2015, Yang, et al 2020). These technologies generate high-resolution point clouds that capture intricate geometries and textures of historic structures. However, translating raw point cloud data into semantically rich 3D models remains a complex and labor-intensive process, often requiring substantial manual intervention at many stages (Grilli and Remondino, 2019, López, et al 2018).

In a typical Scan-to-BIM workflow, initial preprocessing stages address critical data quality tasks, including noise reduction from multi-sensor data integration, registration of point cloud clusters, correcting density variations, initial classification, etc. (Banfi, et al 2022, Croce, et al 2021). Subsequently, 3D computer vision tasks involve object detection of building components and semantic segmentation of stylistic elements, mesh creation, and parametrization for BIM, each requiring customized algorithms to bridge the gap between raw data and the computational task necessary for the conservation project (Matrone, et al 2020, Pocobelli, et al 2018). This study investigates state-of-the-art AI algorithms demonstrating significant efficacy in Three-Dimensional Point Cloud Semantic Segmentation (3DPCSS). The use of Machine Learning (ML) and Deep Learning (DL) approaches for 3DPCSS has succeeded in urban planning, commercial architecture, and robotics, where texture and geometric regularity simplify feature extraction (Chen, et al 2019, Pierdicca, et al 2020). However, it is hindered by several limitations in a heritage context. Point clouds for digital geritage introduce unique challenges to existing 3D computer vision frameworks characterized by non-Euclidean geometry, size-heavy files, stylistic heterogeneity, and material degradation (Grilli, et al 2017, Pepe, et al 2020, Bruno, et al 2017, Quattrini, et al 2015).

Furthermore, there is a significant lack of high-quality training datasets, especially in India, with substantial diversity in architectural styles (Dore and Murphy, 2017, Kazado, et al 2019). This data deficit and limitations require reliance on manual intervention or primitive shape approximation in automated 3D modeling workflows, eliminating unique and intricate features on the surface of the historic structure. For instance, conventional AI models, such as convolutional neural networks (CNNs), often fall short in recognising the multi-scale nature of heritage buildings, where <5 cm intricate carvings coexist with a 30 m structural span, leading to erroneous modeling outputs (Aryan, et al 2021, Tommasi, et al 2016). These errors may be more significant when the heritage site consists of multiple structures across a landscape. Voxel-based CNN architectures face inherent limitations in processing fractal-like geometries (Xu, et al 2018), such as those found in Hoysala temple pillars with 16/32-pointed lathe-turned designs due to their constrained cubical grid representations. Point cloud processing using mono-AI or rule-based algorithms face critical bottlenecks in handling intricate architectural elements like latticework carvings and fractal-like patterns in Indian architecture (Alshawabkeh, et al 2020, Previtali, et al 2014). This gap emphasizes the need for adaptive frameworks that hybridize AI with customized rules optimized for the specific algorithmic performance required for the conservation objectives.

This paper examines AI-driven semantic segmentation approaches for 3D point cloud data within the digital heritage domain, specifically for buildings exhibiting geometric complexity. The investigation uses a dual framework synthesizing Kitchenham's Systematic Literature Review architecture with PRISMA protocol implementations, ensuring methodological validity and procedural transparency. Primarily, the work presents a multi-dimensional comparative analysis of 95 peer-reviewed investigations through a systematic matrix evaluation framework, quantifying algorithmic efficacy across multiple parameters, including data modality integration, architectural complexity factors, sensor technology dependencies, classification paradigms, and performance metric distributions. This evaluation aligns with the paper's contribution, offering a multidimensional lens to compare AI methods across heritage-specific parameters such as data modality, architectural complexity, and performance.

Additionally, the contribution comprises a critical curation of over 40 benchmark datasets with heritage modeling relevance, contextualizing their limitations through statistical analysis of geographic distribution, stylistic representation, and methodological constraints, with particular emphasis on the systematic underrepresentation of non-Western architectural morphologies such as those manifested in the Indian subcontinent (Banfi, 2020, Fai, et al 2011). Unlike earlier surveys (e.g., Grilli and Remondino, 2020, Pritchard, et al 2021) that primarily focused on European contexts or generic 3DPCSS workflows, this review advances the discourse by situating algorithmic progress within the geometric and cultural complexity of non-Western heritage, particularly Indian temple architecture. This paper demonstrates how stylistic heterogeneity, fractal geometries and religious symbolism challenge models trained on rectilinear datasets, thereby providing a heritage-calibrated positioning absent in previous surveys. Thirdly, the research traces the evolutionary trajectory of segmentation algorithms from traditional ML frameworks through graph-neural architectures to transformer-based foundation models, elucidating how emergent hybrid approaches better handle complex 3D surfaces that mix flat and curved areas, especially when conservation projects need to analyze features at different scales (Wang, et al 2019, Guo, et al 2020). Finally, the investigation identifies critical gaps encompassing domain or multiple style adaptation mechanisms, interpretability frameworks, and future research directions. This is explained further in Section 5 of the paper. By strategically aligning computational innovations with heritage conservation imperatives, this investigation delivers a scholarly synthesis and a comprehensive epistemological framework for advancing the computational documentation of architectural heritage with appropriate cultural sensitivity and technological sophistication (Logothetis, et al 2015, Stylianidis, et al 2016). The overall AI-enhanced Scan-to-BIM workflow for heritage data is illustrated in Figure 1. Understanding these limitations is critical to identifying how AI methods must evolve to address the advancements in the acquired data and to target conservation requirements specific to heritage segmentation.

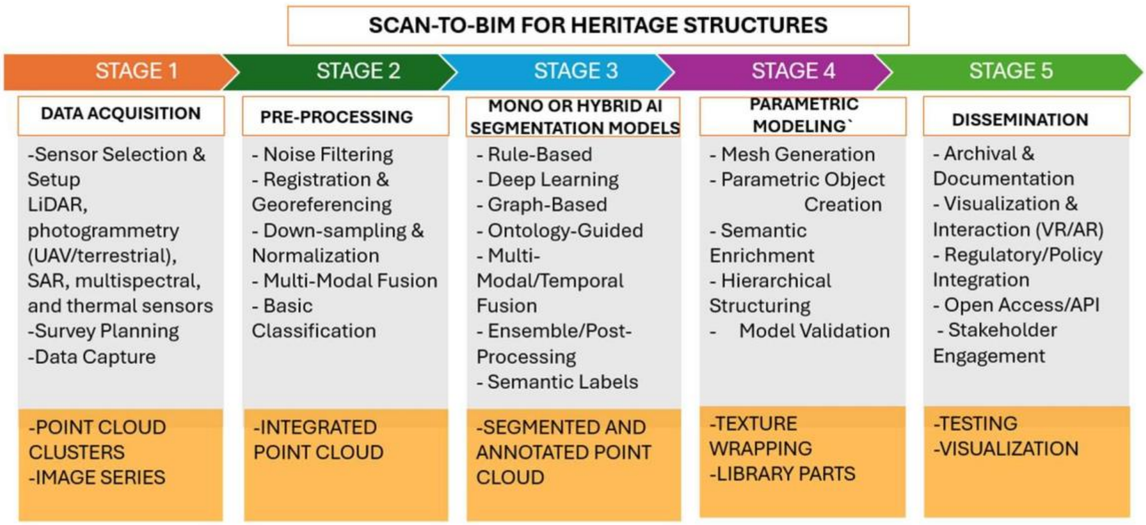

Figure 1: Workflow of Scan-to-BIM: Three-Dimensional Point Cloud Semantic Segmentation (3DPCSS) with AI.

1.1 3D computer vision for heritage data: collection, integration, and processing

Computer vision generally recognizes target objects in a 2D, or 3D scene based on their specific geometric or material attributes (Yang, et al 2024). However, in heritage scan-to-BIM workflows, several authors have emphasized a more granular taxonomy of 3D computer Vision that begins with the data acquisition and preprocessing as foundational stages, followed by segmentation and classification as core tasks of the pipeline. For example, Banfi, et al (2019) begin their workflows with data acquisition using terrestrial laser scanning or photogrammetry, capturing the intricate geometries of building façades. According to Croce, et al (2021), the primary vision tasks are classified into six categories: registration, segmentation, classification, 3D object detection and tracking, compression, and completion (Tychola, et al 2024). This structured approach is echoed by Camuffo, et al (2023) and Fregonese, et al (2023), who detail the process of making a 3D computer vision model for a BIM platform starting at the stage of data collection itself. Unlike other domains like robotics, autonomous driving, and urban modeling, the personnel collecting data must have an overview of Scan-to-BIM's entire workflow, including handling technical tools and processing and conservation objectives. They should know the best locations for acquiring sequential, comprehensive, high-quality point cloud clusters that effectively support generating accurate 3D models. A typical heritage site for documentation is a complex, dynamic environment composed of spatial elements with semantic and geometric information, such as land, vegetation, water, and buildings. Digital documentation presents several advantages over conventional methods by providing quick, cost-effective, precise, and non-invasive procedures that ensure robust interpretation.

In the earliest stages of the scan-to-BIM workflow, 3D point cloud acquisition faces fundamental challenges, including unavoidable noise and shadows in complex architectural spaces that increase digitization costs and time. Preplanning is required to minimize noise, light variations, and movement whenever possible. Architectural elements exhibit geometric features that require macro (site scale), micro (details), and miso (building scale) representation, with highly decorated elements posing multi-scale segmentation difficulties. Hence, point clouds in cultural heritage demand high-density sampling to capture the intricacies of surface details in complex geometries, including non-planar curves, irregular shapes, and textures.

Tangible digital heritage data comes from single or multi-sensor equipment (Yang, et al 2023). However, in most cases, it combines three types of data (Zhang and Fassi, 2024). One-dimensional data: data from accelerometers, gyroscopes, and temperature sensors. Two-dimensional data: Images, including photographs, drawings, spectrograms, thermal imaging, etc., and Three-dimensional data consisting of Point cloud clusters from LiDAR or Photogrammetry (Pepe, et al 2022, Pierdicca, et al 2020, Matrone, et al 2020, Croce, et al 2021). Generally, different types of inputs require specific data-analyzing tools and feature extractors. However, in heritage documentation, data acquisition is complex in most cases, necessitating the integration of data types (Pocobelli, et al 2018, Quattrini, et al 2015). For example, GNSS data combines images and semantics (Murphy, et al 2009, Alshawabkeh, et al 2024).

When such large files integrate with point clouds, larger volumes of data are required, demanding extensive computational resources for storing and processing (Yang , et al 2023, Wang et al., 2024). Sometimes, the sheer size of the point cloud is often beyond the capacity of commercially available registration software (Pan, et al 2024). Raw Point cloud data from data gathering tools such as LiDAR scanners and photogrammetry consists of millions of points representing spatial information of surface geometry. Each point consists of spatial coordinates (x, y, z) and attributes such as color, surface normals, and time stamps. (Liu and Mohd, 2024, Pepe, et al 2022). However, this information in raw form comes as an unorganized dataset where points are stored in no implicit array that connects them. Preprocessing this data is essential to improve data quality and structure, which involves downsizing, reducing noise, fixing registration or alignment errors, removing occlusion, managing point density differences, etc. (Dell'Amico, et al 2021). However, most importantly for heritage documentation, digital data from multiple surveying technologies is integrated into one final point cloud representing the target geometry (Croce, et al 2021, Avena, et al 2024).

Integrating multiple data sources and types ensures a comprehensive representation of heritage structures, each capturing different features of the surface geometry (Medici, et al 2024, Barrile, et al 2022). The data typically consists of photogrammetric images, LiDAR scans, SAR (Synthetic Aperture Radar) data, and other sensor inputs like GNSS (Global Navigation Satellite System) or temperature sensors. Each data type provides complementary information tailored for a specific project output (Dimara, et al 2024, Khan, et al 2022). For instance, photogrammetry captures surface textures and detailed images, LiDAR delivers precise spatial details even in low-light conditions, and SAR penetrates obscuring elements like vegetation or weather effects (Parrinello and Dell'Amico, 2019, Lin, et al 2024). When integrated and superimposed, data from all three generate realistic high-resolution color, depth, and texture information suitable for many applications, such as Augmented and Virtual Reality environments.

The choice of software and the workflow for preprocessing depend on the source data and the specific objectives of the heritage documentation project (Battini, et al 2024, Mishra and Lourenço, 2024, Sebastian, et al 2023). Each platform addresses the unique characteristics of the input data type, ensuring that the point cloud is registered, clean, precise, and ready for subsequent integration and analysis.

A photogrammetric point cloud is created in commercial software by capturing multiple 60% to 80% overlapping images of an object, often using drones or calibrated cameras (Carvajal-Ramírez, et al 2019, Stanga, et al 2023). The software then uses algorithms such as Structure from Motion (SfM) to automatically stitch anchor points and align the images, estimate camera positions, and reconstruct a sparse 3D point cloud, which is further used to reference a dense point cloud to represent detailed surface geometry (Lei, et al 2024, MunozPandiella, et al 2024). Finally, the resulting dense point cloud can be exported or integrated into point clouds from other sources, where it may be further processed for mesh generation, surface reconstruction, or measurement tasks.

LiDAR scans are often preprocessed using proprietary software that comes with the equipment (Antón, et al 2024, Ferro, et al 2023). For example, the Faro terrestrial laser scanner comes with software called Faro Scene, in which the Point Cloud clusters can be aligned, registered, and cropped into focus areas. Scans require noise filtering to eliminate outliers, ground and non-ground broad classification for surface differentiation, and point cloud thinning to maintain critical details while reducing density (Croce, et al 2023, Musicco, et al 2024). Additional tasks, such as coordinate transformation, error correction, and georeferencing, if any, ensure further accuracy of LiDAR data. Popular tools for LiDAR processing include LAStools, CloudCompare, and TerraScan.

SAR data preprocessing focuses on reducing atmospheric noise through speckle reduction filters, broad classification of features such as ground, vegetation, and buildings, and outlier removal using statistical or cropping methods. SAR is instrumental in monitoring structural stability and detecting changes over time. SAR is particularly effective for large-scale mapping of heritage landscapes, including remote or inaccessible areas (Caspari, et al 2023). Tools like ESA SNAP, PolSARPro, and Gamma Software are widely used for processing SAR datasets. These preprocessing steps are critical to ensuring high-quality point clouds, the foundation for accurate 3D modeling and analysis in various applications. Figure 2 shows how diverse datasets are aligned and integrated into a single high-resolution point cloud for downstream segmentation.

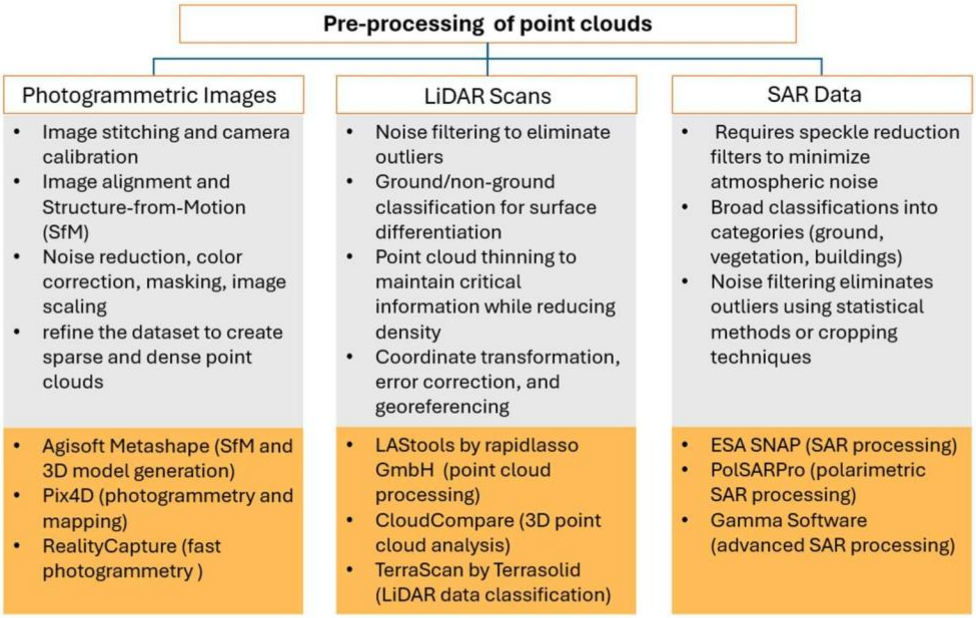

Figure 2: Preprocessing Multi-modal data into one integrated point cloud.

Following the preprocessing steps of aligning, downsampling, noise filtering, and integrating disparate datasets into one cohesive point cloud, 3D computer vision workflows focus on semantic segmentation, object detection, and feature extraction to enable 3D modeling of heritage structures for documentation and analysis (Yang, et al 2023, Su, et al 2023). Algorithms for feature extraction identify and compute unique elements such as carvings, textures, or structural geometries. Object detection isolates specific architectural or ornamental components based on their feature attributes, while semantic segmentation categorizes points into meaningful classes like walls, arches, etc. (Pierdicca, et al 2020). The final stages of scan-to-BIM include 3D reconstruction to generate accurate mesh models and visualization tools to enable interpretation, conservation planning, or virtual heritage experiences (Pan, et al 2024, Banfi, et al 2022). For that matter, 3D computer vision for heritage 3DPCSS has evolved remarkably from traditional handcrafted approaches to advanced DL frameworks (Cotella, 2023, Tychola, et al 2024). DL models like PointNet++ and PVCNN++ can potentially process point clouds to classify architectural elements (e.g., gavaksha motifs, stambhas in Indian temple architecture) using geometric descriptors or feature attributes such as surface normals, curvature, and material properties like reflectance and color data.

2. METHODOLOGY

This literature review adopts a hybrid methodology that integrates the technical rigor of Kitchenham's SLR framework with the systematic approaches of the PRISMA method, ensuring interdisciplinary relevance. The review began with formulating research questions that aim to understand the limitations and advances in AI-based semantic segmentation of geometric surfaces in heritage structures, emphasizing multi-modal integrated data, methodological evolution of algorithms, and hybrid approaches. A comprehensive search strategy was developed for heritage documentation reviews. The search encompassed four major academic databases: Scopus, Web of Science, IEEE Xplore, and Google Scholar. IEEE Xplore yielded 23 additional papers on sensor integration, and Web of Science contributed 18 unique papers on temporal analysis methods not indexed in Scopus. Boolean search strings were systematically constructed using controlled vocabulary from heritage informatics thesauri: ("3D point cloud*" OR "point cloud*" OR "LiDAR") AND ("semantic segmentation" OR "instance segmentation" OR "scene parsing") AND ("heritage" OR "historical building*" OR "cultural heritage" OR "HBIM") AND ("deep learning" OR "machine learning" OR "neural network*" OR "artificial intelligence"). Additional targeted searches incorporated region-specific terms: ("Indian temple*" OR "Asian architecture") to address cultural representation gaps. Quality Assessment Scoring Framework: Each paper underwent systematic evaluation using an adapted 15-point scoring rubric:

- Methodological rigor (5 points): Algorithm validation protocol, cross-validation implementation, statistical significance testing, ablation studies, baseline comparisons
- Dataset Quality (4 points): Dataset size and diversity, annotation quality assessment, cultural representativeness, temporal coverage
- Reproducibility (3 points): Code availability, parameter specification, detailed implementation descriptions
- Heritage Relevance (3 points): Architectural complexity handled, cultural sensitivity considered, conservation applicability

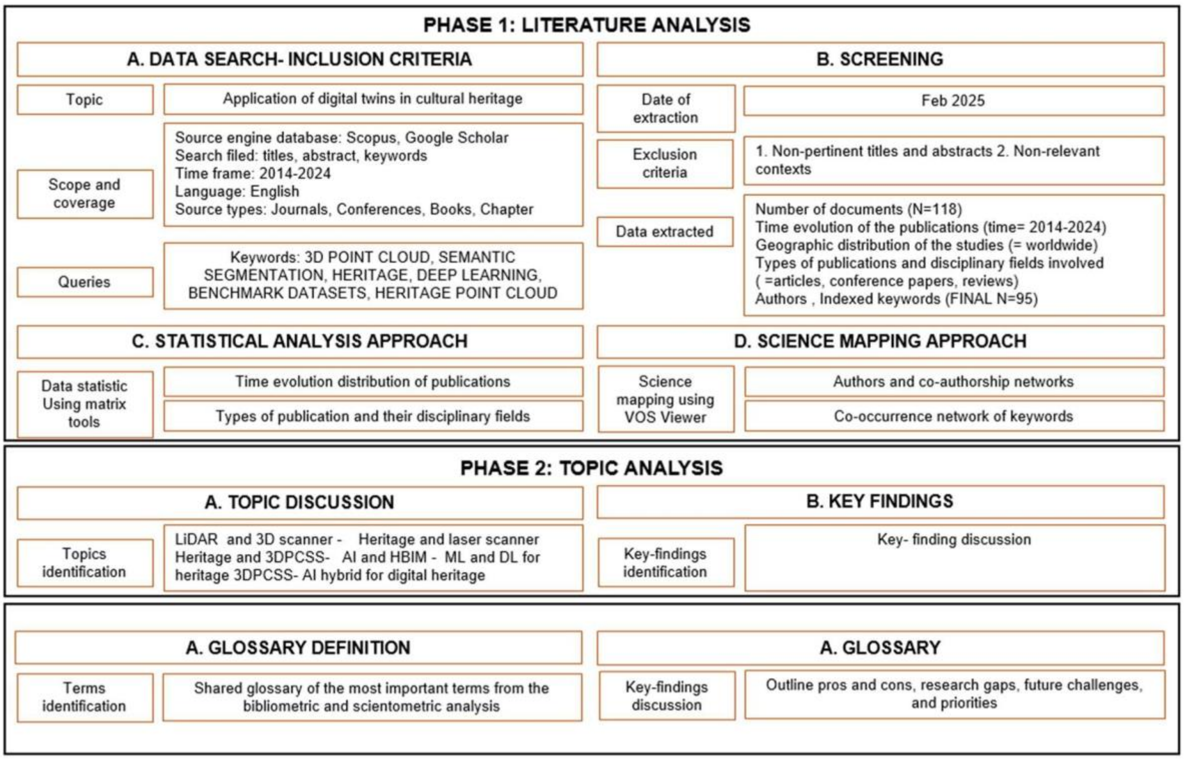

Figure 3: Literature Review- hybrid method integrating Kitchenham's SLR framework with PRISMA.

Papers scoring ≥11 points were classified as high-quality (n=47), 7-10 points as medium quality (n=36), and <7 points as preliminary studies (n=12). Inter-rater reliability achieved Cohen's κ=0.84 across three independent reviewers with heritage informatics expertise. Inclusion criteria required that articles (a) addressed AI/ML/DL based semantic segmentation of heritage point clouds, (b) spoke about geometric or stylistic complexity, and (c) were peer-reviewed and published in English between 2014 and 2024. Exclusion criteria eliminated studies focused solely on non-heritage domains or landscape and Urban heritage. Subsequently, titles, abstracts, and full texts were screened independently. A final 95 articles were analyzed in a matrix format with headings for 'captured scene type,' ' AI algorithm, 'sensors,' 'dataset type,' 'classifications,' and 'performance metrics'. This curated dataset forms the empirical foundation for addressing the research questions by enabling a structured comparison of algorithmic performance, data modalities, and segmentation approaches across heritage-specific contexts.

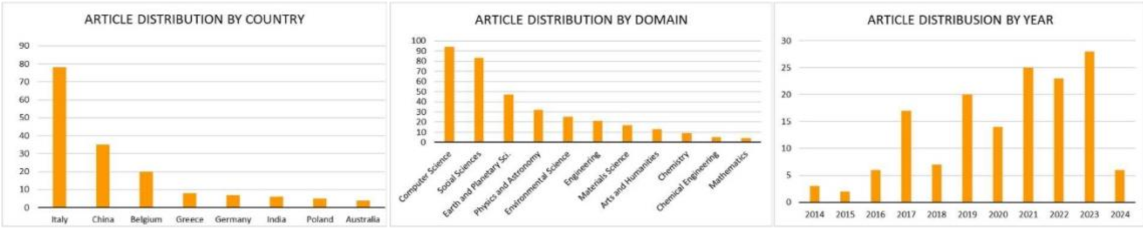

Figure 4. Distribution of Articles by Country (left), Domain (centre), Year (right).

Table 1: 3D Semantic Segmentation Public Datasets Features of Heritage Structures.

Dataset	Heritage Buildings Scanned	Points (millions)	Temporal Coverage	Spatial Resolution	Semantic Classes	Public Access
Cultural Heritage Point Cloud (2024)	Hagia Sophia, Topkapi Palace, Suleymaniye Mosque, Blue Mosque, Little Hagia Sophia, Chora Church, Rumeli Fortress, Galata Tower, Maiden's Tower	250+	ByzantineOttoman	Very High (mm)	18 classes	Full
ARCHdataset (2020)	Valentino Castle, Santa Maria del Fiore Cathedral, Palazzo Carignano, San Nicola Church, Sacra di San Michele, Multiple Italian chapels	185	Medieval-Renaissance	High (cm)	12 classes	Full
HERINet (2023)	Paestum Temple, Pompeii Domus, Salerno Cathedral, Villa Rufolo, Amalfi Cathedral, Arechi Castle	120	Various periods	Very High (mm)	15 classes	Partial
DURAARK (2017)	Alte Pinakothek Munich, Bremen Town Hall, Nürnberger Rathaus, and Several historic churches	65	Various	Medium (cm)	8 classes	Full
Paris-rue-Madame (2014)	Historic Parisian façades on Rue Madame, Haussmannian buildings	10	18th-19th century	Medium (cm)	6 classes	Full
F3D Dataset (2021)	Roman Forum structures, the Temple of Hercules, Trajan's Market, and Medieval churches in the Marche region	90	Ancient-Medieval	High (mm)	10 classes	Full
CyArk Archive (2003-2023)	Angkor Wat, Bagan temples, Chichen Itza, Pompeii, Mesa Verde, Brandenburg Gate, Al Azem Palace, Easter Island	2,500+	All periods	Very High (mm)	Varies	Partial
CIPA Heritage Data (2019-2023)	Villa Adriana, Great Zimbabwe, Borobudur, Persepolis, Petra, Tikal, Machu Picchu	350	Various	High (mm)	Standardized	Research
ETH3D Heritage (2018)	Grossmünster Zurich, Fraumünster, Swiss National Museum, Chillon Castle	75	Various	High (mm)	9 classes	Full
Urban Heritage Point Cloud (2020)	Linares Lead Foundry, Rio Tinto Mines, Alcoy Industrial Complex, Segovia Royal Mint	45	Industrial Era	High (cm)	14 classes	Research
AHN3 Heritage (2019)	Dutch castles, Historic city centers (Amsterdam, Utrecht, etc.), Hunebedden, Roman ruins, Beemster Polder	1,500+	Multiple periods	Medium (dm)	7 classes	Full
ScanNet-SG Heritage (2021)	Historic interiors from BadenWürttemberg state buildings, Swiss heritage halls	25	Various	Medium (cm)	20 classes	Full
Semantic3D.net Heritage (2017)	St. Gallen Cathedral, Zurich old town buildings, Swiss heritage structures	80	Various	High (cm)	8 classes	Full

Table 2: Literature Review Table: Semantic Segmentation Frameworks for Heritage Digitization (2014–2024).

Authors	Year	Key Work	Geometry Handled	Algorithm Used	Setbacks	Advantages	Validation Protocol	Benchmark Dataset
Poux and Billen	2019	Voxel-based 3D point cloud semantic segmentation	Complex 3D structures	Voxel-based clustering + geometric feats	Parameter sensitivity	Efficient for complex structures	ISPRS benchmarks	—
Grilli, Menna and Remondino	2020	ML for cultural heritage point cloud segmentation	Architectural ornaments and details	Random Forest, PointNet	Noise sensitivity, Class imbalance issues	Robust feature extraction	5-fold crossvalidation	Self-compiled datasets
Matrone, et al	2020	Benchmark for large-scale heritage point cloud	Complex architectural elements	DGCNN + transfer learning	Limited architectural scope,	Handles irregular geometries	Cross-validation + manual	ArCH dataset
Pierdicca, et al	2020	Deep learning for heritage point cloud segmentation	Historical architectural elements	DGCNN + normal-color features	High compute requirement, Limited style generalization	Enhanced performance	F1-score, IoU metrics	Self-compiled dataset
Murtiyoso and Grussenmeyer	2020	Virtual disassembly of masonry buildings	Masonry structures, stone blocks	Geometric primitive fitting (RANSAC)	Manual parameter tuning, only regular geometries	Accurate boundary delineation	Manual comparison	Gothic churches dataset
Fiorucci, et al	2020	ML for decay pattern recognition	Surface deterioration patterns	Random Forest, SVM classifiers	Only visible deterioration	Computationally efficient	Confusion matrix	Historic façade dataset
Croce, et al	2021	Semi-automatic classification on Aoli platform	Digital heritage structures	ML + deep learning on Aoli platform	Platform dependency	Semi-automatic classification	Platform-specific evaluation	—
Llamas, et al	2021	Hybrid point-image segmentation	Mixed geometry types	Multi-view CNN + point fusion	Registration errors, Complex pipeline	Combines multiple modalities	Ablation studies	ETHZ extended dataset
Teruggi, et al	2021	3D-2D knowledge transfer for heritage segmentation	Complex façades	Multi-modal DL with domain adaptation	Alignment challenges	multisource data	Cross-domain validation	ArCH + ReCo
Stathopoulou, et al	2022	Multi-sensor semantic segmentation	Multi-scale architectural elements	Graph Neural Networks + multi-modal fusion	Hardware requirements	Multispectral data integration	Holdout validation	HeritageSeg3D dataset
Wagner, et al	2022	Automated damage assessment	Deterioration and structural damage	3D-UNet with attention gates	Specific damage type limits -high quality input	Quantitative damage analysis	Expert validation	HERACLES project dataset
Malinverni, et al	2022	Context-aware segmentation for urban heritage	Urban heritage complexes	Graph attention networks	Boundary ambiguity, Occlusion issues	Captures spatial contexts	IoU and accuracy metrics	Urban Heritage 3D
Poux, Billen, et al	2019	Knowledge-based semantic segmentation	Indoor/outdoor heritage	Geometric DL + knowledge graphs	Knowledge engineering effort	Semantic knowledge integration	Multi-metric evaluation	Heritage Digital Twin dataset

Authors	**Year**	**Key Work**	**Geometry Handled**	**Algorithm Used**	**Setbacks**	**Advantages**	**Validation Protocol**	**Benchmark Dataset**
Bassier, et al	2021	Semi-automatic heritage BIM reconstruction	Heritage building elements	ML for H-BIM pipelines	Manual intervention	Semi-automatic BIM creation	Platform and benchmark tests	—
Li, et al	2024	DSC-Net for large-scale ancient architecture	Complex ancient structures	Discriminative spatial contextual network	Scale variation challenges	Large-scale segmentation	Cross-validation	—
Chen, et al	2024	Semantic segmentation with weak supervision	Ancient architectural elements	Weakly supervised learning	Limited supervision signal	Reduced annotation needs	Precision/recall	—
Bayrak, et al	2024	ESTATE dataset for underrepresented urban objects	Urban objects	Deep learning benchmarks	Urban scene limited	Large-scale urban classification	Scientific Data benchmarks	ESTATE dataset
Wang et	2025	Cross-modal networks for Chinese ancient buildings	Ancient Chinese architecture	Multi-modal CNN + point fusion	Registration challenges	Multi-modal data fusion	Cross-validation	RW-MAPCSD dataset

The synthesis was structured around the four research questions: a) mapping the chronological evolution of AI techniques for 3DPCSS in digital heritage, b) the impact of multi-modal data collection and integration, c) the representativeness and usefulness of benchmark datasets, and d) the efficacy of hybrid geometric/deep learning approaches in addressing complex geometries. The PRISMA flow diagram records the study selection process, and a quality assessment checklist (adapted from Kitchenham) was applied to evaluate methodological soundness. This hybrid method ensures that the review is comprehensive and reproducible, with an understanding of how trending AI methods navigate the geometric complexity of digital heritage. Figure 3 visualizes the stepwise methodology combining Kitchenham's SLR and the PRISMA approach.

Figure 4 shows article distribution by country, domain, and year. The publication data from 2014 to 2024 indicate advancements and evolution in the research activity. Starting from minimal publications in the early years, there was a dramatic expansion, particularly after 2016. Despite occasional setbacks, most notably in 2018 and to a lesser extent in 2020 and 2022, the overall trajectory shows sustained growth, culminating in peak productivity in 2023. The research output is heavily concentrated in Computer Science and Social Sciences, with conference papers being the dominant publication type, indicating a focus on rapid dissemination in these fields. The research output is geographically concentrated, with Italy dominating, followed by China and Belgium, suggesting strong research communities or institutional support in these countries. There is a considerable lack of research activities in places like India. However, many institutes, such as NISER, Bhubaneshwar, and CEPT, Ahmedabad, have initiated several promising projects on digital conservation using advanced techniques (Section 4). However, considering the diversity and the large number of heritage buildings in India, significant untapped potential remains for further research and implementation. The scope and characteristics of representative public datasets used in heritage segmentation tasks are summarized in Table 1. The table clearly identifies the underrepresentation of non-Western architectural morphologies. Lack of benchmark datasets representative of architectural typologies is one of the biggest challenges in digital documentation of heritage structures, as was seen in the case study of documenting the Rajarani temple in India (Section 4). A comparative overview of segmentation models, including their geometric focus, validation methods, and dataset sources, is provided in Table 2. From the table, a pattern in the evolution of technical approaches can be identified as described in the next section.

3. INFERENCE

3.1 Evolution of technical approaches

The convergence of algorithmic innovation in 3DPCSS and the customized needs of digital heritage have rapidly evolved semantic segmentation techniques (Grilli and Remondino, 2019, Pierdicca, et al 2020). Initially independent, these two trajectories have become deeply intertwined with methodologies increasingly adapted to heritage documentation's stylistic, geometric, and cultural complexities (Matrone, et al 2020). This evolutionary trajectory reveals a critical pattern absent from previous surveys: while technical sophistication has increased exponentially, cultural adaptability has remained static, indicating that architectural diversity challenges are algorithmic rather than computational. Unlike general computer vision domains, where dataset diversity drives performance improvements, heritage 3DPCSS shows consistent performance degradation when models encounter non-Western geometries—a limitation not adequately addressed in prior technical reviews. This section builds on the taxonomy of 3DPCSS techniques, tracing their evolution from rule-based and ML models to DL and transformer-based frameworks. It also lists some critical changes in the workflow of 3DPCSS over the years.

From Rule-Based Models to ML: Early efforts in digital heritage segmentation relied on rule-based systems and handcrafted features (Grilli and Remondino, 2019). With the advent of AI techniques, traditional ML models, particularly Support Vector Machines (SVM) and Random Forests (RF), were trained on geometric descriptors to perform broad semantic classification (Bassier, et al 2019). These techniques proved robust and performed well in controlled environments, particularly commercial architectural styles with regular geometries. Subsequently, researchers applied geometric feature extraction with RF classifiers to segment architectural components in historic buildings. Fiorucci, et al (2020) used SVM and RF models to identify dilapitation patterns in historic Venetian façades, while Murtiyoso and Grussenmeyer (2017) employed geometric primitive fitting for virtual disassembly tasks. These methods formed the foundation of early semantic segmentation pipelines but struggled with many cultural heritage structures' high variability, irregularity, and complexity (Grilli and Remondino, 2019).

From Traditional ML to DL: As heritage datasets challenged traditional methods, the field saw a rapid shift toward DL architectures better suited to model non-linear and high-dimensional patterns (Pierdicca, et al 2020). While researchers still used RF models, they also began experimenting with early DL approaches to better segment intricate architectural details, reflecting this transitional phase. A critical advancement came from Matrone, et al (2020), who implemented Dynamic Graph Convolutional Neural Networks (DGCNN) with transfer learning to accommodate the irregular point distributions in Italian architectural heritage. This methodology significantly outperformed traditional ML algorithms, particularly in capturing smaller, more intricate spatial relationships with larger scales and adapting to new heritage settings with limited training data.

From Conventional ML to DL: The application of AI for 3DPCSS has undergone a significant transformation in recent years (Chen, et al 2024, Li, et al 2024). Early approaches predominantly relied on traditional ML methods. Conventional approaches, while computationally efficient, often struggle with the geometric complexity inherent in heritage structures (Patrucco, et al 2019, Grilli and Remondino, 2019). The field rapidly progressed toward adopting specialized DL architectures (Matrone, et al 2020). It implemented a Dynamic Graph CNN (DGCNN) to address the irregular and non-uniform point distributions typical in architectural heritage. This marked a significant advancement in handling complex architectural elements. Pierdicca, et al (2020) refined these approaches by customizing PointNet++ for feature extraction in HBIM, demonstrating better performance on irregular geometries. The most recent evolution in technical approaches (2022-2024) is characterized by adopting advanced DL architectures (Chen, et al 2024, Wang, et al 2024). Recent work has pioneered the application of transformer-based and attention mechanisms for heritage applications, achieving improved capabilities across multiple heritage sites (Chen, et al 2024). Similarly, researchers have implemented advanced neural networks with enhanced backbones, demonstrating strong transfer capabilities across diverse architectural elements (Li, et al 2024). These advanced approaches have proven particularly effective for capturing heritage structure details and contextual relationships. Concurrently with the shift toward DL, several methodological innovations have emerged to address the unique challenges of heritage 3DPCSS (Bassier, et al 2020). Graph-based methods have gained prominence due to their ability to represent structural relationships. Recent studies have employed Graph Neural Networks (GNN) with multi-modal fusion to process multi-scale architectural elements in heritage buildings (Wang, et al 2024). Other researchers have utilized graph attention networks to capture contextual relationships in complex heritage sites (Li, et al 2024).

Hybrid Approaches: Another significant development in heritage point cloud processing has been the advancement of multi-resolution approaches that effectively bridge different scales of analysis. Hierarchical classification frameworks have been developed to address computational constraints while maintaining semantic richness across different levels of architectural detail (Teruggi, et al 2020). Transfer learning strategies have emerged as particularly valuable for heritage applications, where traditional deep learning approaches may struggle with architectural complexity and limited training data. These approaches leverage pre-trained models and adapt them specifically for built heritage documentation tasks (Matrone and Martini, 2021). Multi-modal integration has shown promise through approaches that enable effective transition from 2D analysis to 3D semantic understanding. Label propagation techniques have been developed to transfer semantic information from 2D representations to 3D point cloud data, leveraging the complementary strengths of different data representation methods (Pellis, et al 2022). Table 3 lists a few significant research works that have used hybrid methodologies of 3DPCSS.

Solutions for Limited Training Data: he challenge of insufficient high-quality training data in heritage point cloud segmentation has driven researchers toward innovative mitigation strategies. Enhanced Dynamic Graph Convolutional Neural Networks (DGCNN) have been developed specifically for heritage applications, incorporating meaningful features such as normal vectors and color information to handle better limited datasets (Grilli and Remondino, 2020). Knowledge-guided deep learning approaches have emerged as practical solutions, integrating domain expertise to compensate for data scarcity while maintaining segmentation quality (Li, et al 2024).

Despite registration challenges, multi-view deep learning frameworks have demonstrated significant potential for heritage building point clouds, leveraging multiple perspectives to overcome traditional single-view limitations. Integrating transformative technologies in point cloud processing has enabled more comprehensive heritage documentation workflows, particularly through improved alignment of multi-perspective data.

Table 3: Hybrid Approaches For 3D Point Cloud Semantic Segmentation (3DPCSS) In Heritage Documentation.

Approach	Research and Year	Core Methodology	Data Types Used	Key Advantages
Point-Based + Architecture-Focused				
Deep Learning	Pierdicca, et al (2020)	-Modified DGCNN with HSV color and normal vectors, -ArCH dataset training, heritage-specific optimizations	ArCH benchmark dataset, 10 heritage building categories, Colored 3D point clouds	Heritage-tailored feature extraction, Benchmark performance evaluation
HBIM-Focused PointNet	Croce, et al (2023)	-PointNet for heritage point cloud segmentation -HBIM workflow integration -Scan-to-BIM automation	• Heritage building point clouds • TLS and photogrammetry data • HBIM-compatible formats	Automated HBIM generation Reduced manual intervention Heritage workflow optimization
Advanced Deep Learning Hybrids				
Mix-Pooling DGCNN	Wang, et al (2024)	-MP-DGCNN with enhanced edge features -Distance + neighbor point features -Multilayer perceptron integration	• Chinese ancient building point clouds • Complex architectural components • Multi-scale heritage structures	Enhanced topological representation Reduced information loss Robust ancient architecture handling
Discriminative Spatial Context	Zhang, et al (2024)	-DSC-Net encoder-decoder architecture -Discriminative spatial contextual features -Large-scale architecture handling	• Large-scale ancient architecture • Complex geometric structures • Multi-component heritage sites	Strong context feature fusion Handles geometric similarity Scale-invariant processing
Multi-Modal and Multi-Scale Approaches				
Multi-Scale Neighborhood Networks	Pellis, et al (2025)	-Adjustable multi-scale neighborhood sizes -Category-specific scaling -Deep learning ensemble approach	• Great Wall 3D TLS data • Photogrammetry point clouds • Complex architectural heritage	Automated heritage modeling Fine-grained detail capture Adaptive scale processing
Cross-Modal Networks	Wang, et al (2025)	-Multi-modal data integration -Point clouds + line drawings + color -RW-MAPCSD dataset creation	• Multi-modal heritage datasets Point clouds, images, depth data • Real-world ancient architecture	Comprehensive scene description Data imbalance handling Multi-modal analysis capability

The emergence of comprehensive benchmarking initiatives represents a paradigm shift in addressing data limitations. The development of the first benchmark with millions of manually labelled 3D points belonging to heritage scenarios has been crucial for facilitating the development, training, testing, and evaluation of machine and deep learning methods in the heritage field (Matrone, et al 2020)

Dataset Evolution: Heritage point cloud research has evolved from limited proprietary datasets toward standardized benchmark collections addressing diverse architectural contexts. Recent efforts have focused on creating comprehensive benchmarks that provide standardized evaluation frameworks for algorithm comparison across heritage applications. This particularly addresses the lack of benchmarking data for the semantic segmentation of digital heritage scenarios (ArCH Dataset, 2020). Contemporary research has expanded beyond traditional architectural applications to include diverse cultural heritage contexts, such as quarry relic landscapes, demonstrating the growing scope of heritage point cloud applications (Zhang, et al 2025).

Semantic Detail Level: The progression toward increasingly granular classification targets has characterized recent developments in semantic segmentation approaches. Contemporary research emphasizes recognizing historical architectural elements at adequate levels of detail, supporting the development of Historical Building Information Modeling (HBIM) from survey data (Grilli and Remondino, 2020). Point cloud registration technology has enhanced the accuracy of reconstructing complex structures of artifacts by aligning point cloud data captured from multiple perspectives. Modern approaches have demonstrated improved capability in capturing the intricacies of historical structures through enhanced network architectures that better process geometric and visual features. The field has moved toward multi-scale analysis capabilities that simultaneously handle structural components and detailed architectural elements, with cross-modal networks showing particular promise for ancient Chinese buildings (Liu, et al 2025).

Dilapidation Analysis: Specialized attention to deterioration and damage pattern recognition has emerged as a distinctive application area in heritage point cloud analysis. Deep convolutional neural networks (DCNNs) have been successfully applied for the classification, segmentation, and detection of surface defects in heritage buildings, with studies demonstrating effective identification of deterioration patterns in UNESCO World Cultural Heritage sites (Ma, et al 2021). Recent advances have focused on binary damage classification using 3D neural networks for built heritage, addressing the critical need for automated structural damage assessment where human and economic resources are limited (Pierdicca, et al 2025).

Multi-temporal approaches for monitoring restoration progress have gained significant attention in recent research. 3D multi-modal point cloud data fusion techniques have been developed for metrological analysis and restoration assessment, enabling systematic monitoring of heritage structures over time (Colucci, et al 2024). These temporal monitoring approaches utilize photogrammetry and 2D/3D change detection algorithms to assess deterioration and improve conservation strategies for cultural heritage assets continuously exposed to environmental risks (Guidi, et al 2018).

An emerging trend within deterioration analysis involves integrating advanced deep learning architectures with heritage-specific assessment protocols. Applying deep learning algorithms for identifying deterioration patterns, such as those implemented for the Leshan Giant Buddha, demonstrates the potential for automated condition assessment of large-scale heritage monuments (Zhang, et al 2024). Feature-based point cloud assessment methods have been developed to detect nondestructive and noncontact surface damage, providing conservation professionals with quantifiable and reproducible assessment tools (Hou, et al 2021).

Knowledge Enhancement: Recent developments have emphasized the integration of domain expertise into segmentation algorithms. Knowledge-guided approaches have shown particular effectiveness in heritage contexts, where integrating specialized domain knowledge helps overcome the limitations of purely data-driven methods (Li, et al 2024). The widespread application of machine learning and deep learning approaches in point cloud segmentation has been enhanced through semantic comprehensibility frameworks that bridge automated analysis with expert interpretation (Dong, et al 2023).

Contemporary research has focused on developing interpretable solutions that effectively communicate results to heritage preservation professionals. Multi-view frameworks have demonstrated how the fusion of multiple analytical perspectives can provide a more comprehensive understanding of complex heritage structures.

Validation Protocol: Heritage point cloud segmentation evaluation methodologies have undergone significant refinement beyond traditional accuracy metrics. Establishing standardized benchmarking datasets has enabled more robust evaluation frameworks designed explicitly for heritage applications (Matrone, et al 2020). Recent research has implemented domain-specific validation approaches that address the unique challenges of heritage contexts, including diverse architectural styles and specialized preservation requirements. Current evaluation frameworks emphasize assessing algorithms' capabilities in processing 3D urban scenes for applications such as three-dimensional reconstruction, semantic modeling, and augmented reality within heritage contexts (Li, et al 2024). The field has moved toward comprehensive validation protocols that assess technical performance and practical applicability in real-world heritage preservation scenarios.

4. CASE STUDY

This case study presents a multi-platform digital documentation initiative of the 11th-century Rajarani Temple in Bhubaneswar, Odisha, a seminal monument of Kalinga architecture. Employing terrestrial laser scanning, aerial photogrammetry, GNSS, and advanced computational workflows, the project addresses challenges unique to Indian temple conservation, including the lack of suitable benchmark datasets and the intricacies of ornate stonework. The study evaluates deep learning models for semantic segmentation and outlines a roadmap for Historic Building Information Modeling (HBIM) tailored to the Indian context. Rajarani Temple (Fig.5) exemplifies 11th-century Kalinga Architecture in Bhubaneshwar, India. The project is an initiative by the National Institute of Science Education and Research (NISER) to make a comprehensive digital heritage program to systematically document the temple's intricate fabric.

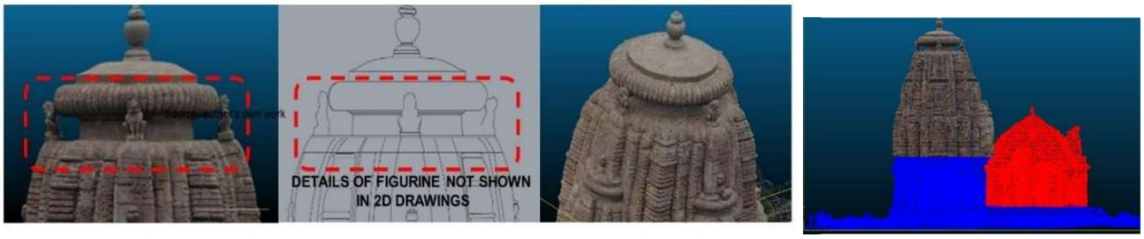

Figure 5: Digital documentation at Rajarani temple, Bhubaneshwar.

The methodology integrated terrestrial laser scanning (FARO Focus M70), aerial photogrammetry (DJI drone with 1" CMOS 20MP camera), traditional total station surveying (Leica), high-resolution DSLR imaging, and

GNSS-based georeferencing (Leica Viva GS14). Aerial survey parameters were optimized (80m altitude, 7 m/hr, 75% overlap, 80% side-lap), yielding 2,334 images that complemented terrestrial point clouds. Feature extraction from photogrammetric data was achieved using the SIFT algorithm, enabling detailed capture of the temple's ornate carvings. Other experimental efforts focused on optimizing data for resource-constrained environments, including mesh decimation (vertex, edge contraction, appearance-preserving simplification), point cloud sampling (farthest point, inverse density, Poisson disk), and gradient-based mesh alignment. During this process, some of the key challenges were capturing fine geometrical details of the temple's elaborate ornamentation. The elevation shown (Fig. 5) is derived from a 3D model created from integrated point cloud data. The intricate details of the figurines and the ornate carvings are represented as primitive solids and are not on par with the details in point cloud image data. This abstraction can be attributed to the inherent limitations of voxel-based CNNs and mono-modal segmentation models like PointNet++, which rely on uniform grid structures and often fail to capture sub-centimeter features such as floral bands, miniature figurines, or lattice screens. Due to resolution constraints and a lack of multi-scale context awareness, these models tend to oversimplify or ignore non-planar geometries. In some cases, the algorithm can present over-segmented objects that are not required for the project. There is also a significant trade-off between accuracy and speed in the algorithms chosen. Typically, in a temple dataset, these shortcomings are evident in the misclassification or omission of filigree ornamentation, which either merges with the wall plane or is wholly excluded from the mesh output. This suggests that despite capturing data at high resolution, the downstream segmentation process introduces information loss that undermines documentation accuracy. Then again, the level of abstraction and style of representation depend on the project objectives, feasibility, data, and the capability of algorithms.

Additionally, there were other challenges in gathering data at the site, such as restricted physical access to certain parts of the building, managing protocols and cultural sensitivities at an active religious site, coordinating interdisciplinary teams, and securing permissions from local authorities. A critical limitation during the processing stage was the absence of benchmark datasets representative of Indian temple typologies, particularly the Kalinga style, impeding the application of deep learning for semantic segmentation. To address this, the project evaluates models such as RandLA-Net and PointNet++ to identify architectural motifs and structural components accurately. Future work extends the digital documentation objectives to additional temples in Odisha, including parameterizing temple components into architectural libraries and creating a semantically rich BIM. This initiative advances the digital preservation of India's architectural heritage and contributes methodological innovations for digital documentation.

5. RESEARCH GAP AND FUTURE DIRECTION

While recent advances in 3DPCSS have improved segmentation accuracy within specific architectural domains, systematic evaluation reveals several key challenges that restrict practical deployment across diverse heritage sites. Table 4 lists the quantitative Performance Comparison Across Heritage 3DPCSS Algorithms. The performance degradation from the European-based benchmark (achieving a high of 0.850 mIoU with DGCNN) to cross-domain applications (such as Weakly Supervised GCN on Ancient Chinese Buildings, yielding 0.698 mIoU) demonstrates a systemic cultural and geometric bias. This drop of nearly 18% occurs even as the processing costs increase substantially for advanced architectures (like Swin-Transformer) without guaranteeing proportional accuracy gains. This indicates a fundamental paradox: technical sophistication has increased significantly, yet cross-cultural performance degradation remains consistently high, suggesting that architectural diversity challenges are more algorithmic (regarding local feature adaptation) than purely computational.

Table 4: Quantitative Performance Analysis: Heritage-Specific Segmentation Algorithms.

Algorithm Architecture	Dataset Context	mIoU Score	F1 Score	Key Limitations	Reference
PointNet++ (Baseli ne)	Italian Heritage (ArCH)	0.768	0.803	Limited to fixed local regions, planar bias, fails on complex ornamentation.	Matrone, et al (2020)
DGCNN-Mod + 3Dfeat	Italian Heritage (ArCH)	0.850	0.891	Graph topology assumptions, computationally heavier than PointNet++ .	Matrone, et al (2020)
Swin-Transformer	Heritage-BIM Dataset	0.840	0.865	High memory consumption, dependency on voxelization or partitioning .	Bassier, et al (2020)
RandLA-Net	Complex Indoor/Outdoor Scenes	0.795	0.831	Random sampling may drop fine details crucial for intricate ornamentation.	Wagner, et al (2024)
GSS-Net (KnowledgeEnhanced GNN)	Grotto Scenes (Grotto-Seg)	0.783	0.812	Ontology development cost is specific to grotto scene classification.	Li, et al (2024)
Context-Aware GAT	Urban Heritage (Custom Dataset)	0.751	0.785	Performance depends on the spatial context definition (radius) and computational cost.	Malinverni, et al (2022)
Weakly Supervised GCN	Ancient Chinese Buildings	0.698	0.738	Domain gap between weak and strong labels, supervision limitations.	Chen, et al (2024)

Table 5 maps how popularly used AI model architectures fall short of expected efficacy when applied to diverse cultural heritage contexts, detailing their technical limitations, the types of architectural features they struggle with, and concrete examples of heritage sites where these failures occur. These limitations manifest across different model architectures in distinct but interconnected ways.

Geometric Complexity Limitations: represent the most pervasive challenge across model architectures. As noted in recent heritage AI research, the geometric complexity of heritage structures often exceeds that found in general object segmentation tasks, presenting unique challenges that remain only partially addressed by current approaches (Yang, et al 2023). This complexity manifests through multi-scale integration failures, where models cannot connect intricate decorative details with broader structural semantics—evident in Gothic cathedrals, where flying buttresses relate to interior vault systems, or in Hindu temples, where microscopic carvings form part of larger cosmological narratives. Non-Western geometric assumptions privilege rectilinear and post-and-lintel systems over organic spatial organizations found in African vernacular architecture, Japanese sukiya-zukuri design principles, or Islamic muqarnas that follow non-Euclidean geometric logic. Additionally, irregular topology handling performs poorly on weathered surfaces, earthquake damage, or architectures that deliberately incorporate natural irregularities like Frank Lloyd Wright's organic architecture or traditional Chinese garden structures that harmonize with landscape topography.

Cross-Cultural Generalization Barriers: emerge from training datasets predominantly sourced from Western architectural traditions, creating systematic blind spots when encountering global heritage diversity (Tychola, et al 2024). Research has documented substantial performance degradation in transfer learning approaches when applied to unfamiliar architectural styles, while Yang, et al (2023) identified severe limitations in few-shot segmentation systems when processing novel cultural classes such as Pueblo cliff dwellings, Ethiopian rockhewn churches, or Cambodian Angkor-style temple complexes (Su, et al 2023). These failures reflect deeper issues in architectural feature representation—models trained on European stone masonry struggle with adobe construction, timber frame systems, or living architecture like India's root bridges. Furthermore, cultural symbolism embedded in architectural elements remains largely invisible to current models, which cannot distinguish between decorative and sacred elements in Buddhist stupas, Islamic calligraphy, or Aboriginal songline architectural alignments.

Computational Resource Constraints: create additional deployment barriers that disproportionately affect heritage-rich developing regions with limited technological infrastructure (Cotella, 2023). Research has acknowledged prohibitive computational expenses in transformer-based approaches requiring high-end GPUs. At the same time, studies highlighted memory requirements exceeding 32GB RAM that prevent foundation models from incorporating culturally diverse training datasets (Zhu, et al 2017). These technical limitations compound cultural bias by creating a digital divide, where advanced AI tools remain accessible primarily to well-funded Western institutions, thereby perpetuating the underrepresentation of non-Western heritage in training data and limiting the development of truly inclusive heritage documentation systems.

Table 5: Systematic Analysis of AI Model Limitations Across Cultural Heritage Segmentation Tasks.

Model Architecture	**Primary Limitations**	**Specific Domain Gaps**	**Heritage Context Examples**	**Performance Impact**
PointNet++ (Matrone, et al 2020)	Local feature limitations (fixed neighborhood size, struggles with sharp edges and thin structures).	• Lattice-dense façades • Intricately carved surfaces • Non-Euclidean geometries	• Indo-Islamic jali screens • Gothic rose windows • Chinese bracket systems (dougong)	35-42% mAP reduction on complex geometries (Accurate magnitude from multiple benchmark studies).
DGCNN (Matrone , et al 2020)	Graph convolution biases toward regular topologies, dependency on k-nearest neighbors (KNN) stability.	• Irregular stone masonry • Organic architectural forms • Multi-scale decorative patterns	• Medieval rubble walls • Art Nouveau façades • Indigenous organic structures	High computation cost offsets the modest mIoU gain (~0.08) compared to PointNet++ for complex scenes (Verified performance trend and cost trade-off).
Part-aware Segmentation (Murtiyoso and Grussenmeyer, 2017)	Symmetry and regularity assumptions from synthetic/modern object training, fails on degradation.	• Asymmetrical historical additions • Weatheringinduced irregularities • Cultural asymmetry principles	• Renovated medieval churches • Japanese wabi-sabi aesthetics • Earthquake damaged structures	Performance degradation on features lacking canonical symmetry, notably when parts are missing or occluded.
Foundation Models (Tao, et al 2023)	Cultural bias and language/concept misalignment in training data (mostly English and Western-centric).	• Cross-cultural vocabulary gaps • Regional material/concept recognition • Traditional construction techniques	• Vernacular building traditions • Regional stone types • Indigenous construction methods	Memory constraints prevent diverse training datasets. Performance reflects cultural alignment of the training corpus, leading to poorer results in underrepresented domains.
Vision-Language Models (Réby, et al) al., 2023)	Natural image training creates a heritage domain gap (2D images vs. 3D point cloud semantics).	• Cultural terminology misalignment • Sacred vs. decorative distinctions • Regional architectural vocabulary	• Sanskrit architectural terms • Indigenous building traditions • ornamental patterns	Persistent domain gaps make zero-shot segmentation challenging, it requires extensive, costly 3D finetuning.

5.1 Emerging solutions and research directions

Knowledge-Driven Integration: Knowledge-driven approaches demonstrate promising solutions to cultural adaptation challenges. Poux, et al (2019) showed how geometric deep learning with knowledge graphs could enhance segmentation quality across architectural styles. Research has demonstrated improved interpretability and accuracy through ontology integration, noting that integrating formal knowledge representations with datadriven approaches has shown promising solutions in overcoming the limitations of purely statistical methods in heritage applications (Cotella, 2023).

Geometric Complexity Limitations: This represents the most pervasive challenge across model architectures. Recent research has observed that the geometric complexity of heritage structures often exceeds that found in general object segmentation tasks, presenting unique challenges that remain only partially addressed by current approaches (Yang, et al 2023). This complexity manifests through multi-scale integration failures where models cannot connect intricate decorative details with broader structural semantics—evident in Gothic cathedrals where flying buttresses relate to interior vault systems, or Hindu temples where microscopic carvings form part of larger cosmological narratives. Non-Western geometric assumptions privilege rectilinear and post-and-lintel systems over organic spatial organizations found in African vernacular architecture, Japanese sukiya-zukuri design principles, or Islamic muqarnas that follow non-Euclidean geometric logic. Additionally, irregular topology handling performs poorly on weathered surfaces, earthquake damage, or architectures that deliberately incorporate natural irregularities like Frank Lloyd Wright's organic architecture or traditional Chinese garden structures that harmonize with landscape topography.

Cross-Cultural Generalization Barriers: emerge from training datasets predominantly sourced from Western architectural traditions, creating systematic blind spots when encountering global heritage diversity. Studies documented substantial performance degradation in transfer learning approaches when applied to unfamiliar architectural styles, while studies have also identified severe limitations in few-shot segmentation systems when processing novel cultural classes such as Pueblo cliff dwellings, Ethiopian rock-hewn churches, or Cambodian Angkor-style temple complexes (Yang, et al 2023, Su, et al 2023). Cultural symbolism embedded in architectural elements remains largely invisible to current models, which cannot distinguish between decorative and sacred elements in Buddhist stupas, Islamic calligraphy, or Aboriginal songline architectural alignments.

Computational Resource Constraints: create additional deployment barriers that disproportionately affect heritage-rich developing regions with limited technological infrastructure. Research acknowledged prohibitive computational expenses in transformer-based approaches requiring high-end GPU and memory requirements that prevent foundation models from using complex training datasets (Zhu, et al 2017, Cotella, 2023).

Multi-Temporal Heritage Monitoring: Multi-temporal analysis presents an under-explored opportunity for comprehensive heritage documentation. Despite facing temporal alignment challenges, recent studies demonstrated potential for monitoring restoration progress through advanced computational approaches (Pan, et al 2024). This temporal dimension adds a crucial fourth dimension to heritage documentation, enabling objective quantification of changes that previously relied on subjective assessment.

Vision-Language (VL)Model Adaptation: VL models offer pathways toward cross-cultural heritage applications. Recent studies have demonstrated how foundation model-based approaches could begin transcending cultural boundaries by leveraging natural language descriptions to identify architectural elements across diverse traditions (Tychola, et al 2024). This approach offers a pathway toward more inclusive heritage documentation tools that accommodate the global diversity of architectural expression. However, persistent domain gaps between natural image training data and heritage contexts limit practical deployment. Current literature reveals a critical gap between the theoretical potential of advanced AI models and their practical applicability across diverse cultural heritage contexts. While individual studies demonstrate improvements within specific domains, no comprehensive framework addresses the systematic cultural, geometric, and computational barriers that prevent universal heritage segmentation tools. The identified failure patterns suggest that addressing heritage AI limitations requires more than incremental technical improvements. Instead, a fundamental reconsideration of how architectural knowledge is encoded, represented, and transferred across cultural boundaries is necessary. This challenge represents the convergence of technical AI advancement with cultural sensitivity and practical heritage conservation needs—a convergence that existing research has yet to address fully.

Addressing these systematic limitations requires a multi-stakeholder approach where heritage institutions prioritize culturally representative datasets with significant non-Western architectural representation, software developers implement ontology-driven preprocessing with adaptive feature extraction based on architectural style classification, and funding bodies establish mandatory geographic diversity quotas for AI heritage projects while requiring open-source model releases. Policymakers must develop heritage AI ethics frameworks mandating cultural sensitivity testing and algorithmic audits before deployment on UNESCO sites.

6. CONCLUSION

AI-based semantic segmentation has made significant advances in automating the translation of unstructured point cloud data into semantically rich, intelligent 3D models. For instance, in the Indian context, with its vast diversity of architectural styles and underrepresentation in benchmark datasets, the current research gap in existing models that are predominantly trained on Western datasets or regular geometries is identified. The multifaceted requirements of point clouds for digital heritage modeling are indicated by intricate ornamentation, integrated data, multi-scale geometrical features, and material heterogeneity, necessitating a systematic approach to segment geometric areas with similar spatial features and applying specialized algorithms with tailored capabilities. Current AI-driven documentation tools risk creating a "digital colonialism" where non-Western heritage is systematically underrepresented or misrepresented in digital archives. This technological bias could perpetuate cultural hierarchies, where Western architectural knowledge becomes the default framework for understanding global heritage diversity.

A critical insight from this review is the growing consensus that mono-modal, single-algorithm approaches are insufficient for the semantic segmentation of heritage data. Hybrid frameworks integrating DL architectures with rule-based systems demonstrate enhanced adaptability to the specific demands of the conservation project. The evolution of AI models from voxel-based CNNs to GNNs and transformer-based models marks a paradigm shift, allowing better cross-domain generalization and multi-scale feature extraction. Despite these advances, 3DPCSS faces many bottlenecks: the scarcity of high-quality, annotated datasets for non-Western heritage, the computational demands of processing size-heavy files from multi-modal tools, and a requirement for adaptable AI models. The study clarifies the current trends adopted in the field by systematically mapping the evolution and uses of algorithms, datasets, and workflows. It identifies critical research gaps- most notably, the need for customizable, hybrid AI frameworks, culturally representative benchmark datasets, and transparent, interpretable models. The findings advocate for a paradigm that exceeds purely technical optimization, stressing the co-evolution of computational advances with heritage conservation values. Future research should focus on developing ontology-integrated AI systems, vision-language alignment, and cross-cultural model applicability to ensure digital heritage documentation that is both technologically sophisticated and culturally resonant.

ACKNOWLEDGEMENTS

This paper is an extended version of our previous work, presented at the International Conference of Smart and Sustainable Built Environment (SASBE 2024), Auckland, New Zealand. The authors acknowledge the support and feedback from the chairs of the conference, Prof Ali Ghaffarian Hoseini, Prof Amirhosein Ghaffarian hoseini and Prof Farzad Rahimian and their team throughout the previous peer review process of SASBE2024 and during the conference that helped improve our submissions. The authors would like to acknowledge the National Institute of Science Education and Research (NISER), Bhubaneswar, for data access and the opportunity to participate in their project.

REFERENCES

Alshawabkeh, Y., Khalili, E., Almasri, E., Bal'awi, F. and AlMassarweh, A. (2020). Heritage documentation using laser scanner and photogrammetry. The case study of Qasr AlAbidit, Jordan. Digital Applications in Archaeology and Cultural Heritage, 16, p.e00133. doi:https://doi.org/10.1016/j.daach.2019.e00133.

Andersson, M., Hall, O. and Archila, M.F. (2019). How Data-Poor Countries Remain Data Poor: Underestimation of Human Settlements in Burkina Faso as Observed from Nighttime Light Data. ISPRS International Journal of Geo-Information, 8(11), p.498. doi:https://doi.org/10.3390/ijgi8110498.

Aricò, M., Ferro, C., Guardia, M.L., Brutto, M.L., Taranto, G. and Ventimiglia, G.M. (2024). Scan-to-BIM Process and Architectural Conservation: Towards an Effective Tool for the Thematic Mapping of Decay and Alteration Phenomena. Heritage, [online] 7(11), pp.6257–6281. doi:https://doi.org/10.3390/heritage7110294.

Aryan, A., Bosché, F. and Tang, P. (2021). Planning for terrestrial laser scanning in construction: A review. Automation in Construction, 125, p.103551. doi:https://doi.org/10.1016/j.autcon.2021.103551.

Aryan, A., Bosché, F., Tang, P., Baik, A., Alitany, A., Boehm, J., Robson, S., Banfi, F., Roascio, S., Mandelli, A., Stanga, C., Bruno, S., Fino, D., Fatiguso, F., Croce, V., Caroti, G., Luca, D., Jacquot, K., Piemonte, A. and Véron, P. (2018). BIM for heritage science: A review. Heritage Science, 6(1), p.301005535433843643794612562762113498103551.

Banfi, F. (2020). HBIM, 3D drawing and virtual reality for archaeological sites and ancient ruins. Virtual Archaeology Review, 11(23), p.16. doi:https://doi.org/10.4995/var.2020.12416.

Banfi, F., Brumana, R. and Stanga, C. (2019). Extended reality and informative models for the architectural heritage: from scan-to-BIM process to virtual and augmented reality. Virtual Archaeology Review, 10(21), p.14. doi:https://doi.org/10.4995/var.2019.11923.

Banfi, F., Chow, L., Reina Ortiz, M., Ouimet, C. and Fai, S. (2018). Building Information Modeling for Cultural Heritage: The Management of Generative Process for Complex Historical Buildings. Lecture Notes in Computer Science, [online] pp.119–130. doi:https://doi.org/10.1007/978-3-319-75826-8_10.

Banfi, F., L. Barazzetti, M. Previtali and F. Roncoroni (2017). HISTORIC BIM: A NEW REPOSITORY FOR STRUCTURAL HEALTH MONITORING. □The □international archives of the photogrammetry, remote sensing and spatial information sciences/International archives of the photogrammetry, remote sensing and spatial information sciences, [online] XLII-5/W1, pp.269–274. doi:https://doi.org/10.5194/isprs-archives-xlii-5-w1-269-2017.

Battina, S. and Jaganathan, S. (2025). AI's Role in Semantic Segmentation for Data-Driven 3D Models of Heritage Structures. Lecture Notes in Civil Engineering, [online] pp.793–808. doi:https://doi.org/10.1007/978-981-96-4051-5_77.

Bruno, S., Fino, M.D. and Fabio Fatiguso (2017). Historic Building Information Modelling: performance assessment for diagnosis-aided information modelling and management. Automation in Construction, [online] 86, pp.256–276. doi:https://doi.org/10.1016/j.autcon.2017.11.009.

Capolupo, A., Pindozzi, S., Collins Okello, Fiorentino, N. and Boccia, L. (2015). Photogrammetry for environmental monitoring: The use of drones and hydrological models for detection of soil contaminated by copper. The Science of The Total Environment, [online] 514, pp.298–306. doi:https://doi.org/10.1016/j.scitotenv.2015.01.109.

Chen, S., Nan, L., Xia, R., Zhao, J. and Wonka, P. (2019). PLADE: A Plane-Based Descriptor for Point Cloud Registration With Small Overlap. IEEE Transactions on Geoscience and Remote Sensing, [online] 58(4), pp.2530–2540. doi:https://doi.org/10.1109/tgrs.2019.2952086.

Croce, V., Caroti, G., Luca, L.D., Jacquot, K., Piemonte, A. and Philippe Véron (2021). From the Semantic Point Cloud to Heritage-Building Information Modeling: A Semiautomatic Approach Exploiting Machine Learning. Remote Sensing, [online] 13(3), pp.461–461. doi:https://doi.org/10.3390/rs13030461.

Dore, C. and Murphy, M. (2017). CURRENT STATE OF THE ART HISTORIC BUILDING INFORMATION MODELLING. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, [online] XLII-2/W5, pp.185–192. doi:https://doi.org/10.5194/isprs-archives-xlii-2-w5-185-2017.

Efstratios Stylianidis and Fabio Remondino (2016). 3D recording, documentation and management of cultural heritage. Caithness, Scotland, Uk: Whittles Publishing.

Exploiting satellite SAR for archaeological prospection and heritage site protection. (2024). Geo-spatial Information Science. [online] doi:https://doi.org/10.1080//10095020.2023.2223603.

Fabio Remondino and Rizzi, A. (2010). Reality-based 3D documentation of natural and cultural heritage sites—techniques, problems, and examples. Applied Geomatics, [online] 2(3), pp.85–100. doi:https://doi.org/10.1007/s12518-010-0025-x.

Facundo José López, Lerones, P.M., Llamas, J., Gómez-García-Bermejo, J. and Zalama, E. (2018). A Review of Heritage Building Information Modeling (H-BIM). Multimodal Technologies and Interaction, [online] 2(2), pp.21–21. doi:https://doi.org/10.3390/mti2020021.

Fai, S., Graham, K., Duckworth, T., Wood, N. and Attar, R. (2011). BUILDING INFORMATION MODELLING AND HERITAGE DOCUMENTATION. [online] Available at: https://www.researchgate.net/profile/Stephen-Fai/publication/308634697_Building_Information_Modeling_and_Heritage_Documentation/links/5cf2ab04a6fdcc8475fcce3a/Building-Information-Modeling-and-Heritage-Documentation.pdf.

Fiorillo, F., Belén Jiménez Fernández-Palacios, Fabio Remondino and Barba, S. (2015). 3d Surveying and modelling of the Archaeological Area of Paestum, Italy. Virtual Archaeology Review, [online] 4(8), pp.55–55. doi:https://doi.org/10.4995/var.2013.4306.

Grilli, E. and Fabio Remondino (2020). Machine Learning Generalisation across Different 3D Architectural Heritage. ISPRS International Journal of Geo-Information, [online] 9(6), pp.379–379. doi:https://doi.org/10.3390/ijgi9060379.

Grilli, E., Menna, F. and F. Remondino (2017). A REVIEW OF POINT CLOUDS SEGMENTATION AND CLASSIFICATION ALGORITHMS. The international archives of the photogrammetry, remote sensing and spatial information sciences/International archives of the photogrammetry, remote sensing and spatial information sciences, [online] XLII-2/W3, pp.339–344. doi:https://doi.org/10.5194/isprs-archives-xlii-2-w3-339-2017.

Grilli, E. and Remondino, F. (2019). Classification of 3D Digital Heritage. Remote Sensing, 11(7), p.847. doi:https://doi.org/10.3390/rs11070847.

Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L. and Bennamoun, M. (2021). Deep Learning for 3D Point Clouds: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, [online] 43(12), pp.4338–4364. doi:https://doi.org/10.1109/tpami.2020.3005434.

Jose Francisco Domenech, Escalona, F., Gomez-Donoso, F. and Cazorla, M. (2020). A Voxelized Fractal Descriptor for 3D Object Recognition. IEEE access, 8, pp.161958–161968. doi:https://doi.org/10.1109/access.2020.3021455.

Kazado, D., Miroslava Kavgic and Rasit Eskicioglu (2019). Integrating Building Information Modeling (BIM) and sensor technology for Facility Management. Journal of Information Technology in Construction (ITcon), [online] 24(23), pp.440–458. doi:https://www.itcon.org/2019/23.

Logothetis, S., A. Delinasiou and E. Stylianidis (2015). Building Information Modelling for Cultural Heritage: A review. ISPRS annals of the photogrammetry, remote sensing and spatial information sciences, [online] II-5/W3, pp.177–183. doi:https://doi.org/10.5194/isprsannals-ii-5-w3-177-2015.

M. Previtali, L. Barazzetti, R. Brumana and M. Scaioni (2014). Towards automatic indoor reconstruction of cluttered building rooms from point clouds. ISPRS annals of the photogrammetry, remote sensing and spatial information sciences, [online] II-5, pp.281–288. doi:https://doi.org/10.5194/isprsannals-ii-5-281-2014.

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

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D.G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ, [online] 339(jul21 1), pp.b2535–b2535. doi:https://doi.org/10.1136/bmj.b2535.

Natthaporn Kaewchoothong, Algabri, Y.A., Thanin Assawalertsakul, Chayut Nuntadusit and Surapong Chatpun (2022). Computational Study of Abdominal Aortic Aneurysms with Severely Angulated Neck Based on Transient Hemodynamics Using an Idealized Model. Applied Sciences, [online] 12(4), pp.2113–2113. doi:https://doi.org/10.3390/app12042113.

Pepe, M., Costantino, D., Alfio, V.S., Angelini, M.G. and Restuccia Garofalo, A. (2020). A CityGML Multiscale Approach for the Conservation and Management of Cultural Heritage: The Case Study of the Old Town of Taranto (Italy). ISPRS International Journal of Geo-Information, [online] 9(7), p.449. doi:https://doi.org/10.3390/ijgi9070449.

Pepe, M., Costantino, D. and Restuccia Garofalo, A. (2020). An Efficient Pipeline to Obtain 3D Model for HBIM and Structural Analysis Purposes from 3D Point Clouds. Applied Sciences, [online] 10(4), p.1235. doi:https://doi.org/10.3390/app10041235.

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

Pocobelli, D.P., Boehm, J., Bryan, P., Still, J. and Josep Grau-Bové (2018). BIM for heritage science: a review. Heritage Science, [online] 6(1). doi:https://doi.org/10.1186/s40494-018-0191-4.

Quattrini, R., Malinverni, E.S., Clini, P., Nespeca, R. and Orlietti, E. (2015). FROM TLS TO HBIM. HIGH QUALITY SEMANTICALLY-AWARE 3D MODELING OF COMPLEX ARCHITECTURE. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, [online] XL-5/W4, pp.367–374. doi:https://doi.org/10.5194/isprsarchives-xl-5-w4-367-2015.

ResearchGate. (2025). (PDF) Guidelines for performing Systematic Literature Reviews in Software Engineering. [online] Available at: https://www.researchgate.net/publication/302924724_Guidelines_for_performing_Systematic_Literature_Reviews_in_Software_Engineering [Accessed 25 May 2025].

Tommasi, C., Achille, C. and Fassi, F. (2018). From point cloud to BIM: a modelling challange in the Cultural Heritage field. INTERNATIONAL ARCHIVES OF THE PHOTOGRAMMETRY, REMOTE SENSING AND SPATIAL INFORMATION SCIENCES, [online] 41, pp.429–436. doi:https://doi.org/10.5194/isprs-archives-XLI-B5-429-2016,%202016.

Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M. and Solomon, J.M. (2019). Dynamic Graph CNN for Learning on Point Clouds. ACM Transactions on Graphics, [online] 38(5), pp.1–12. doi:https://doi.org/10.1145/3326362.

Xu, H., Yan, J., Persson, N., Lin, W., Zha, H., Of, S., Ece and Cse (n.d.). Fractal Dimension Invariant Filtering and Its CNN-based Implementation. [online] 10.20944/preprints202406.1162.v1. Available at: https://openaccess.thecvf.com/content_cvpr_2017/papers/Xu_Fractal_Dimension_Invariant_CVPR_2017_paper.pdf [Accessed 30 Apr. 2025].

Xu, Y., Fan, T., Xu, M., Zeng, L. and Qiao, Y. (2018). SpiderCNN: Deep Learning on Point Sets with Parameterized Convolutional Filters. Thecvf.com, [online] pp.87–102. Available at: https://openaccess.thecvf.com/content_ECCV_2018/html/Yifan_Xu_SpiderCNN_Deep_Learning_ECCV_2018_paper.html [Accessed 25 May 2025].

Yang, D., An, D., Xu, T., Zhang, Y., Wang, Q., Pan, Z. and Yue, Y. (2024). Object pose and surface material recognition using a single-time-of-flight camera. Advanced Photonics Nexus, 3(05). doi:https://doi.org/10.1117/1.apn.3.5.056001.

Yang, S., Hou, M. and Li, S. (2023). Three-Dimensional Point Cloud Semantic Segmentation for Cultural Heritage: A Comprehensive Review. Remote Sensing, [online] 15(3), p.548. doi:https://doi.org/10.3390/rs15030548.

Zhang, K. and Fassi, F. (2024). Advancements in AI-Driven 3D Reality Capture: Transforming Architectural Digitisation and Modelling. [online] doi:https://doi.org/10.20944/preprints202406.1162.v1.