

Article

Digital Heritage Conservation of Historical Villages Using UAV Photogrammetry–LiDAR Fusion and AI-Based Façade Material Analytics

Junpeng Fan ¹, Zao Zhang ², Anbang Dai ³, Hongxi Yin ⁴ and Yasushi Ikeda ^{5,*}

¹ Graduate School of Media and Governance, Keio University, Fujisawa 252-0882, Japan; fan-junpeng@keio.jp

² Hangzhou Innovation Institute, Beihang University, Beijing 100191, China; zzchun12826@buaa.edu.cn

³ School of Civil Engineering and Architecture, Zhejiang University of Science and Technology, Hangzhou 310023, China; daianbang@zust.edu.cn

⁴ Sam Fox School of Design & Visual Arts, Washington University in St. Louis, St. Louis, MO 63130, USA; hongxi.yin@wustl.edu

⁵ Department of Architecture, The University of Tokyo, Tokyo 113-8656, Japan

* Correspondence: yasushi@arch1.t.u-tokyo.ac.jp

Highlights

What are the main findings?

- A multi-sensor digital heritage workflow combining UAV oblique imagery, Simultaneous Localization and Mapping (SLAM)-based LiDAR scanning, and parametric data fusion improves the geometric completeness and interpretability of historical village models in complex spatial conditions.
- SAM2-based computer vision enables semi-automated façade material identification and spatial quantification, supporting non-invasive and repeatable documentation of traditional architectural elements.

What are the implications of the main findings?

- The integration of computer vision (CV)-derived material information within a Unity-based multi-scale visualization environment facilitates rapid understanding and early-stage assessment for stakeholders unfamiliar with historical village contexts.
- The proposed framework offers a scalable and hardware-adaptive approach for sustainable digital heritage management, with potential extension toward decision-support and digital-twin-oriented conservation systems.

Abstract

The accelerating deterioration of Chinese historical villages necessitates advanced digital approaches for systematic documentation and conservation. The present research proposes a novel Digital Heritage Framework that integrates UAV-based 3D oblique photogrammetry, LiDAR point cloud modeling, and computer vision. Unlike single-technology approaches, our methodology solves modeling issues for complex terrain mapping. This especially applies to the interior and roof works of buildings. The framework implements a customized Rhino-Grasshopper. The 3D model is able to resolve issues of shadow occlusion and spatial discontinuity by integrating aerial and ground-based datasets into spatially coherent formats. This makes use of the Meta-AI-SAM2 deep learning model for semantic segmentation and identification of materials. The computer vision (CV) approach gives semi-automated façade analysis. It enables documentation of complex architectural features non-invasively. We developed a Unity-based visualization platform. It features multiscale representations, ranging from village-scale layouts to centimeter-accurate scans



Academic Editor: Alessandro Vitale

Received: 18 April 2026

Revised: 30 May 2026

Accepted: 1 June 2026

Published: 5 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

of heritage structures such as the Qinchuan Ancestral Hall. Integration with the Unity platform optimizes dataset organization and hierarchical structuring. This significantly enhances database operational efficiency. This integration reduces manual processing complexity and hardware demands. Demonstrating documented efficiency and precision, this workflow presents a scalable solution for endangered heritage sites. Future research will explore AI-assisted detail reconstruction and cross-cultural adaptations. It potentially establishes this framework as a comprehensive tool for sustainable digital conservation.

Keywords: digital heritage; historical village conservation; data integration; cross-platform visualization; computer vision (CV)

1. Introduction

The conservation of Chinese historical villages demands innovative solutions to overcome the inefficiencies of traditional methods, such as manual surveys, archival research, and architectural mapping, which often yield inconsistent, incomplete data unsuitable for systematic conservation [1–3]. With the rapid development of digital technologies, new opportunities have emerged for the documentation and preservation of these valuable heritage assets.

Recent advancements in Unmanned Aerial Vehicle (UAV) photogrammetry have enabled multi-angle, high-precision spatial data acquisition for architectural documentation [4,5]. Modern photogrammetric pipelines typically rely on Structure-from-Motion (SfM) with scale-invariant feature transform (SIFT) for tie point extraction, followed by multi-view stereo (MVS) for dense reconstruction. Studies have demonstrated the effectiveness of UAV oblique imagery in capturing spatial configurations and architectural textures of traditional settlements [6]. Simultaneously, LiDAR technology provides dense and accurate point cloud data for architectural reconstruction, overcoming limitations of image-based methods in occluded or poorly lit environments [7,8]. However, each technology applied in isolation presents inherent limitations: UAV photogrammetry struggles with roof undersides and interior spaces due to occlusion and limited camera angles, while ground-based LiDAR alone cannot capture roof textures and colors essential for complete documentation.

Computer Vision (CV) techniques have demonstrated strong performance in automatic feature extraction from heritage assets [9]. Recent work has shown that deep learning models can effectively identify building components and estimate 3D parameters from point clouds UAV photogrammetry struggles with roof undersides [10,11]. The emergence of foundation models such as Segment Anything (SAM) [12] and its successor SAM2 offers new possibilities for zero-shot semantic segmentation of architectural elements, reducing the need for task-specific training data [13].

The integration of UAV, LiDAR, and CV technologies has given rise to sophisticated, cross-platform heritage documentation systems. Zhang et al. [14] demonstrated that multi-sensor fusion can achieve accurate and detailed 3D representations of historical sites. Studies have also shown that immersive visualization environments, such as those built on Unity, can enhance heritage perception and engagement for diverse audiences, including non-experts. For instance, recent research on multisensory digital heritage spaces demonstrates that factors like informational clarity and emotional resonance significantly contribute to user immersion and cultural engagement [15]. Furthermore, empirical studies on virtual museums employing mixed visualization techniques have confirmed that such immersive experiences provide high-resolution, cost-effective access to cultural artifacts, thereby fostering deeper audience engagement and aesthetic appreciation [16]. For

conservation authorities encountering a historical village for the first time, the challenge extends beyond acquiring accurate spatial data—they must also rapidly develop a shared understanding of settlement structure, building materials, and visible patterns of alteration. Web-based 3D repositories combining meshes, point clouds, and annotations have proven advantageous in helping both experts and non-experts navigate complex heritage datasets [17–19].

Despite these advances, existing workflows typically treat data acquisition, semantic analysis, and visualization as separate stages, often requiring manual intervention for integration. Furthermore, the application of state-of-the-art CV models to heritage-specific material analysis remains underexplored, particularly for traditional Chinese building façades. Recent work has demonstrated the potential of Unity-based multilayered platforms for integrating heterogeneous heritage datasets including point clouds, meshes, and parametric models [20]. However, our previous research developed a cross-software evaluation platform based on UAV 3D oblique imagery and CV technology for the conservation of historical villages, demonstrating the potential of semi-automated segmentation for street scene elements, but did not fully integrate LiDAR data or provide multi-scale visualization with embedded semantic attributes [21–23].

Building upon these foundations, the present study proposes a novel Digital Heritage Framework that integrates UAV-based 3D oblique photogrammetry, SLAM-based LiDAR point cloud modeling, and SAM2-based computer vision within a unified workflow. Unlike single-technology approaches, our methodology addresses modeling issues for complex terrain and building interiors through parametric data fusion in Rhino-Grasshopper. The framework achieves three specific contributions.

(1) Multi-sensor data fusion: We combine UAV oblique imagery and handheld SLAM LiDAR scanning to overcome occlusion problems and ensure geometric completeness. Using the Rhinoceros-Grasshopper platform, roof models generated from UAV data are integrated with LiDAR point cloud models to create seamless, editable 3D representations suitable for both display and analysis.

(2) SAM2-based façade material analytics: The Meta-AI-SAM2 deep learning model is applied to perform façade material segmentation and quantitative analysis in the proposed workflow. Through the case of Qinchuan Ancestral Hall, we demonstrate semi-automated recognition and quantification of traditional building materials, verifying the method's potential for non-invasive heritage documentation.

(3) Multi-scale interactive visualization: A Unity-based platform is developed featuring hierarchical data organization and seamless navigation from village-scale layouts to centimeter-accurate scans of heritage structures. The platform integrates CV-derived material information as queryable semantic attributes, enabling real-time generation of material proportion charts and supporting rapid comprehension for stakeholders unfamiliar with the site.

2. Materials and Methods

This study first obtained UAV oblique imagery through UAV Oblique Imagery and combined it with ground LiDAR point clouds scanning technology to achieve comprehensive data acquisition from macro to micro. In the data processing stage, the open-source software GET3D Cluster (GC) (Version 7.0, Wuhan Daspatial Intelligence Technology Co., Ltd., Wuhan, China) is used to construct a 3D model of the heritage site. while iTwin Capture professional software (version 2024, Bentley Systems, Exton, PA, USA) is used to process LiDAR point cloud data, generating high-precision ground and indoor 3D models. Through the Rhino-Grasshopper parametric modeling workflow, spatial registration and

fusion of data from different sources and scales were achieved, resulting in a refined 3D rendering model with complete spatial information.

Subsequently, CV-based semantic segmentation and recognition of images were performed using the Meta-AI-SAM2 model, combined with a hybrid validation mechanism of artificial intelligence automatic recognition and human promotion, to systematically evaluate the technical feasibility of the model in the task of identifying building façade materials. Meta-AI-SAM2 is an innovative application that uses SAM2 as its backbone network and modifies the interface to connect to the Qinchuan Ancestral Hall Rhino 3D model to complete the façade segmentation task. This method can accurately identify and classify the composition of building façade materials, automatically generate high-precision material segmentation images and supporting Excel data reports, and fully record the spatial distribution characteristics and proportion of various building materials.

Finally, based on the Unity cross-platform engine, a multi-level 3D visualization analysis system was constructed. On the original basic interface, a hierarchical window module and data group management function have been added, realizing the hierarchical display of the overall terrain model from the village scale to the high-precision model of individual buildings. The automatic import and visualization of building façade material CV segmentation images and material proportion data have been achieved through the C# programming interface [24,25]. In addition, the system also integrates a free view roaming function, supporting users to explore indoor and outdoor space information interactively in real time, providing a complete solution for the digital protection and research of architectural heritage. The overall process is shown in Figure 1.

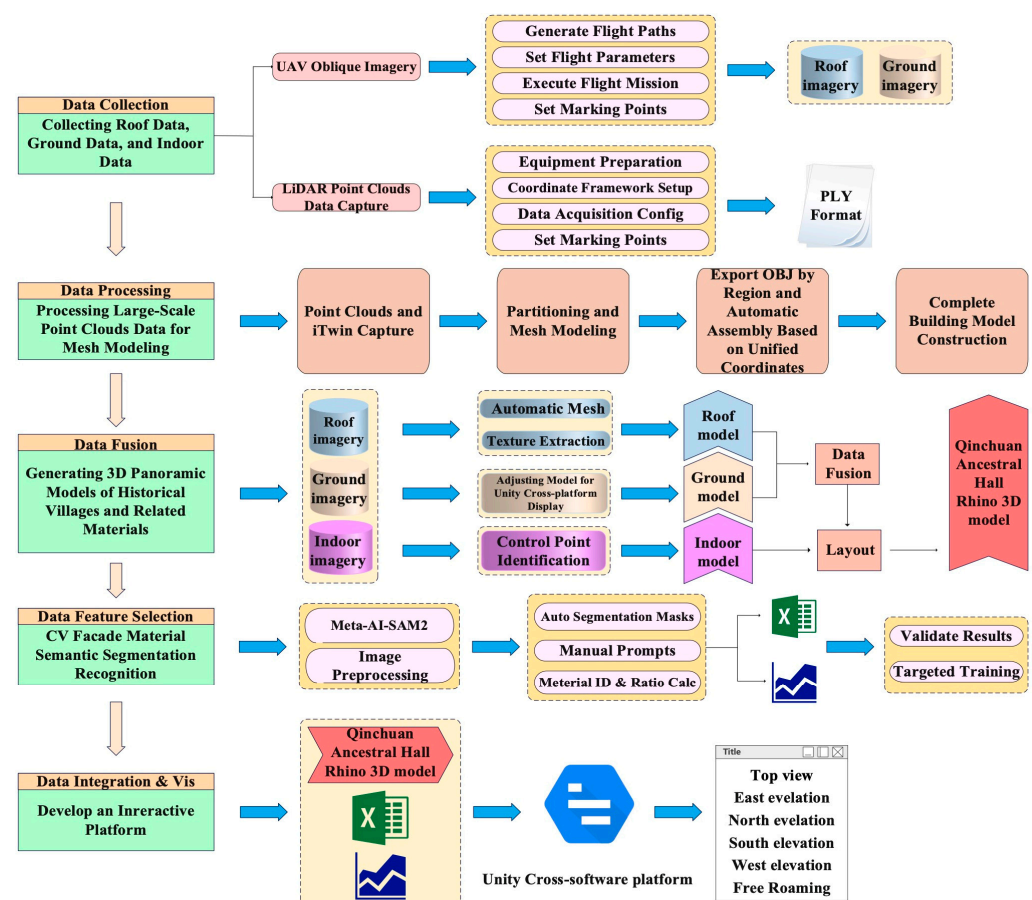


Figure 1. Overall Workflow Diagram.

The proposed framework consists of five main stages:

Data acquisition: UAV oblique imagery and SLAM-based handheld LiDAR scanning are conducted to capture multi-scale spatial data.

3D reconstruction: Photogrammetric and LiDAR point cloud processing generate initial 3D models.

Data fusion: Parametric registration and fusion are performed in Rhino-Grasshopper to integrate aerial and ground-based datasets.

Façade material analytics: SAM2-based computer vision performs semantic segmentation of building materials.

Multi-scale visualization: A Unity-based platform enables interactive exploration with embedded semantic attributes.

2.1. Data Collection

To prioritize cost-effectiveness and time-saving, we selected the DJI Phantom 4 RTK (DJI, Shenzhen, China) and Huace RS10 SLAM scanner (Huace Navigation, Shanghai, China) due to their balance of affordability, ease of operation, and minimal need for manual ground control points (GCPs) deployment. The total field acquisition time was approximately 2.5 h, reducing manual intervention by ~60% compared to traditional TLS surveys.

2.1.1. UAV Oblique Imagery

This study used the DJI Phantom 4 RTK UAV system for oblique imagery data collection. Based on the DJI Pilot application platform, five overlapping flight paths were planned using Keyhole Markup Language (KML) route files. The operation interface and photographic images are shown in Figure 2.

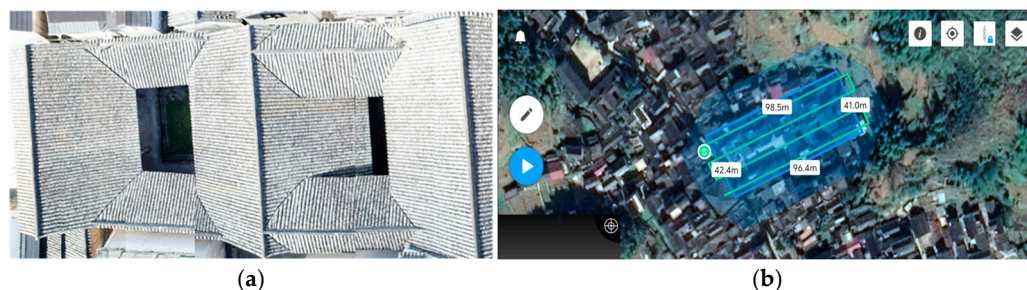


Figure 2. DJI Pilot application platform and photographic images: (a) DJI Pilot mission planning interface showing five overlapping flight paths; (b) example of UAV oblique imagery captured during the flight. In (b), the numbers (98.5 m, 41.0 m, 42.4 m, 96.4 m) indicate measured distances or dimensions on the ground, and the numbered markers 1–5 denote sequential waypoints or flight segments.

The oblique imagery acquisition was conducted on 17 September 2024, between 14:00 and 16:30 local time, in sunny weather with an ambient temperature of 20–28 °C and average wind speed below 5.4 m/s (Beaufort scale 3), ensuring stable imaging conditions. The afternoon time window was selected to minimize harsh shadows on building façades, thereby improving the quality of subsequent semantic segmentation. The camera pitch angle was kept at -60° to capture multiple oblique perspectives and minimize façade occlusion in complex spatial structures. A segmented velocity strategy was employed: take-off climb at 10 m/s, cruise speeds of 1.5 m/s in dense areas and 3.0 m/s in open areas. The lower speed in dense areas guarantees sufficient image overlap around complex structures, while the higher speed in open areas improves overall acquisition efficiency.

To improve the geometric accuracy of the 3D reconstruction, three Ground Control Points (GCPs) were established at the entrance of the Ancestral Hall, as shown in Figure 3.

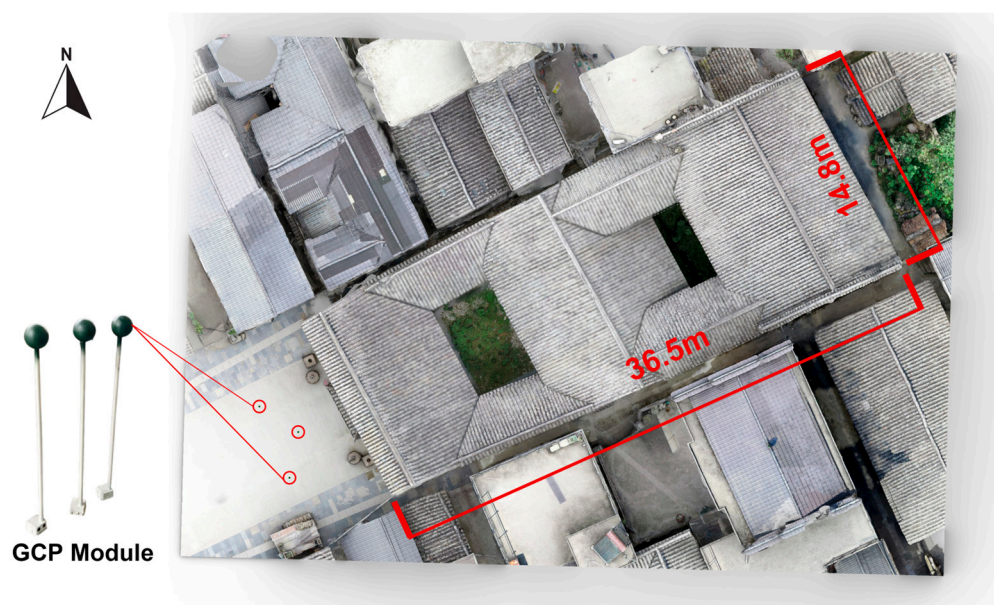


Figure 3. Ground Control Points (GCPs) and the Ancestral Hall area.

The GCPs were surveyed using real-time kinematic (RTK) GPS with centimeter-level accuracy and marked with high-visibility targets. Table 1 summarizes the key flight parameters and acquisition conditions. These parameters were carefully selected to ensure a reproducible benchmark for small-scale UAV photogrammetry in heritage environments and to provide high-quality inputs for the subsequent LiDAR fusion and semantic segmentation workflows. As specified in Table 1, the forward overlap was set to 80% and the side overlap to 70%, ensuring sufficient tie points for robust SfM reconstruction.

The 80% forward overlap and 70% side overlap ensure sufficient tie points for structure-from-motion reconstruction, especially for the complex roof geometries. RTK GNSS was used both to geotag each image in real-time and to survey the three ground control points (GCPs), achieving an accuracy of ± 1 cm. Only three GCPs were deployed because the total site area is relatively small (approximately 540 m²) and the UAV RTK system already provides high absolute positioning accuracy; additional GCPs would not significantly improve the final model accuracy [26] for a similar finding in small-scale heritage sites. The three GCPs were placed in the open plaza in front of the Ancestral Hall (Figure 3) because the remaining areas of the site—specifically the rear and lateral sides—are surrounded by adjacent buildings and deep eaves. These obstructions block both RTK GNSS signals and the visibility of ground targets from the UAV oblique imagery, making it impossible to reliably capture any GCP placed in those locations. The only area with unobstructed sky view and clear visibility from multiple camera angles is the entrance plaza. Given this physical constraint, we concentrated all three GCPs in this accessible zone.

To minimize potential distortion on the opposite side (approximately 37 m from the GCP cluster), we adopted three mitigation strategies: (1) using the DJI Phantom 4 RTK with real-time geotagging (absolute accuracy ± 1 cm), which reduces dependence on dense GCPs; (2) maintaining high image overlap (80% forward, 70% side) to ensure strong tie-point matching across the entire block; and (3) placing five independent checkpoints on the opposite side (measured by total station with ± 2 mm accuracy) to evaluate actual accuracy. The RMSE at these checkpoints was 5.2 cm, which remains within the typical tolerance for 1:50 scale heritage documentation (usually <10 cm). Since the opposite side contains no

heritage structures (only bare ground and low vegetation), this accuracy level is acceptable for the conservation objectives of this study.

Table 1. Key UAV Oblique Imagery Flight Parameters and Acquisition Conditions (17 September 2024).

No.	Parameter	Value	Relevance/Justification
1	UAV Platform	DJI Phantom 4 RTK	Dual-frequency GNSS RTK for precise positioning; supports accurate GCP-based georeferencing.
2	Flight Date and Time	17 September 2024, 14:00–16:30	Sunny, low wind (<5.4 m/s) ensures stable imaging and minimal motion blur.
3	Flight Altitude (AGL)	25 m	Balances coverage (sufficient for village-scale) and GSD (0.67 cm/pixel) for façade details.
4	Camera Pitch Angle	−60.0°	Oblique view reduces occlusion of building façades and eaves.
5	Ground Sampling Distance	0.67 cm/pixel	Ensures spatial resolution adequate for identifying roof tiles and façade materials.
6	Cruise speed (dense/open)	1.5 m/s/3.0 m/s	Lower speed in dense areas guarantees sufficient image overlap; higher speed in open areas improves efficiency.
7	Forward/side overlap	80%/70%	High overlap ensures robust image matching for Structure-from-Motion (SfM). These values guarantee robust image matching for SfM.
9	Number of images	182	Sufficient for complete coverage of the target area (Ancestral Hall and surroundings).
10	GCP accuracy (RMSE)	±1.0 cm (horiz.), ±1.5 cm (vert.)	Enables sub-decimetres georeferencing of the final 3D model.

Data source: Field measurements obtained from DJI Phantom 4 RTK flight on 17 September 2024.

A perfectly even GCP distribution is not possible at this site due to the obstructions described above. The reported GCP residuals (0.8 cm in X, 1.1 cm in Y, 1.5 cm in Z) and the checkpoint RMSE of 5.2 cm on the far side provide a realistic assessment of the achieved accuracy under these challenging conditions.

2.1.2. LiDAR Point Clouds Data Capture

As shown in Figure 4, this study used the Huace-RS10 handheld Simultaneous Localization and Mapping (SLAM) laser scanner for indoor and outdoor 3D data acquisition. Compared with traditional static Terrestrial Laser Scanning (TLS), SLAM-based mobile systems offer greater operational flexibility in confined or complex environments. The portable design supports single-person operation, and high-precision data acquisition can be achieved without relying on GPS signals or complex target deployment. Its dynamic acquisition capability enables continuous scanning in complex environments, effectively solving the operational limitations of traditional devices in narrow spaces or obstructed areas.

The scanning path was planned to cover all interior spaces, courtyards, and building exteriors in a single continuous loop to minimize drift accumulation and ensure reliable loop closure. The operator walked at a constant speed of 1–1.5 m/s while maintaining a consistent distance of approximately 2–5 m from the scanned surfaces to achieve uniform point density and stable data capture. The Huace RS10 integrates an RTK GNSS module, which was used

during acquisition to support direct trajectory georeferencing within the China Geodetic Coordinate System 2000 (CGCS2000) framework and facilitate subsequent multi-source data integration. After acquisition, the raw data were processed in CoPre software (version 2.0, Huace Navigation, Shanghai, China) using a SLAM-based trajectory adjustment with loop-closure optimization to generate a registered point cloud. Under mixed indoor–outdoor field conditions, the relative geometric consistency of the registered point cloud was maintained within 3 cm after loop closure, while the absolute georeferencing accuracy, validated against independent checkpoints measured by total station, was better than 5 cm RMS. Under the standard scanning speed, a default point cloud density of approximately 20,000 points per square meter was obtained, and the final dataset was exported in PLY format for subsequent fusion with the UAV photogrammetric model.



Figure 4. Handheld SLAM and its Scene Shooting Example.

Table 2 summarizes the key scanning parameters and acquisition results, demonstrating the high quality, density, and geometric reliability of the LiDAR data used in this study.

Table 2. Key Huace RS10 Handheld SLAM LiDAR Scanning Parameters and Acquisition Results.

No.	Parameter Description	Value	Relevance/Justification
1	Equipment Model	Huace RS10 SLAM Laser Scanner	Dual-channel rotating LiDAR; suitable for indoor/outdoor mobile mapping.
2	Laser Wavelength	905 nm	Near-infrared, eye-safe (Class 1).
3	Scanning Velocity	1.0–1.5 m/s	Controlled walking pace ensures consistent point density and minimizes motion distortion.
4	Point Density	~20,000 pts/m ²	Provides sufficient detail for architectural elements (e.g., stone carvings, timber joints).
5	Total Points Collected	149.8 × 10 ⁶	Raw point count before filtering; covers ~7500 m ² of interior, exterior, and courtyard.
6	Data Volume	6.56 GB (PLY)	Compressed format suitable for handling in Rhino and Grasshopper.
7	Absolute Accuracy (RMS)	<5.0 cm	After loop-closure optimization and checkpoint validation.

Table 2. *Cont.*

No.	Parameter Description	Value	Relevance/Justification
8	Operating Temperature	20–28 °C	Stable during indoor/outdoor scanning; no thermal drift effects.
9	Total Coverage Area	~7500 m ²	Full coverage of the Ancestral Hall and its immediate surroundings.
10	Relative accuracy (loop closure)	<3 cm	Ensures fine alignment for fusion with UAV data

Data source: Field acquisition conducted by the research team using Huace RS10 (18 September 2024).

2.1.3. Overview of Dataset Acquisition

The complete dataset was acquired on 18 September 2024 under sunny weather conditions (ambient temperature 20–28 °C, average wind speed <5.4 m/s). The UAV component collected 182 oblique images (total 1.6 GB, JPEG format) with a ground sampling distance of 0.67 cm/pixel. The LiDAR component collected 149.8 million points (6.56 GB in PLY format) covering an area of approximately 7500 m², including the Ancestral Hall interior, courtyard, and immediate surroundings. All data were acquired by two trained operators within a single afternoon session (14:00–16:30).

2.2. Data Processing

2.2.1. UAV Oblique Imagery Generates 3D Models

The 182 oblique images were processed using the commercial cloud-based software GET3D Cluster (GC, developed by Ambit Geospatial Solutions), which implements structure-from-motion (SIFT feature extraction, exhaustive matching) and multi-view stereo algorithms. The three GCPs surveyed by RTK GNSS were used as control points in the bundle adjustment to constrain the solution and improve absolute accuracy. The camera (DJI Phantom 4 RTK) was pre-calibrated with known focal length (8.8 mm), principal point, and radial distortion parameters; these intrinsic parameters were further refined during the bundle adjustment. The final quality of the photogrammetric model was evaluated using five independent checkpoints (not used in the adjustment), yielding a mean reprojection error of 0.48 pixels and a root mean square error (RMSE) of 1.9 cm in object space.

The photogrammetric processing in GET3D Cluster was performed with the following settings: input images were used at full resolution (4864 × 3648 pixels); SIFT keypoints were extracted with a maximum of 10,000 features per image; exhaustive feature matching was enabled; bundle adjustment was carried out with a maximum of 500 iterations; and multi-view stereo (MVS) reconstruction used a depth-map resolution of 1/2 of the original image size.

2.2.2. LiDAR Point Clouds Were Used to Generate 3D Models

Prior to mesh generation, the raw LiDAR point cloud was inspected for completeness and uniform point density. Areas with insufficient coverage or visible drift were identified and rescanned during a second pass. The raw dataset consisted of 149.8 million points with an average density of approximately 20,000 points/m². Visual inspection confirmed complete coverage of all interior spaces, courtyards, and building exteriors without significant holes or artifacts.

This section uses iTwin Capture to process the LiDAR point cloud and export OBJ mesh models with RGB texture derived from the LiDAR intensity data (pseudo-color) and the UAV imagery (true color after fusion). The point cloud was first subsampled to a uniform density of 5 mm to reduce computational load while preserving detail. The mesh generation used the Poisson surface reconstruction algorithm with an octree depth

of 10 and a surface sampling density of 0.1. The resulting mesh was then decimated to a target face count of 2 million for manageable handling in Rhino. The detailed workflow is shown in Figure 5. Georeferencing of the LiDAR point cloud was performed using the RTK GNSS trajectory combined with loop closure optimization in CoPre software. The final point cloud was transformed to CGCS2000 using three ground control points as reference, achieving an absolute accuracy better than 5 cm RMS, as verified by checkpoints measured with a total station. In iTwin Capture, the mesh generation was performed using the following settings: point cloud subsampling to a uniform density of 5 mm; Poisson surface reconstruction with octree depth = 10 and surface sampling density = 0.1; mesh decimation to a target face count of 2 million; texture mapping from UAV imagery with a resolution of 4096×4096 pixels per texture atlas.

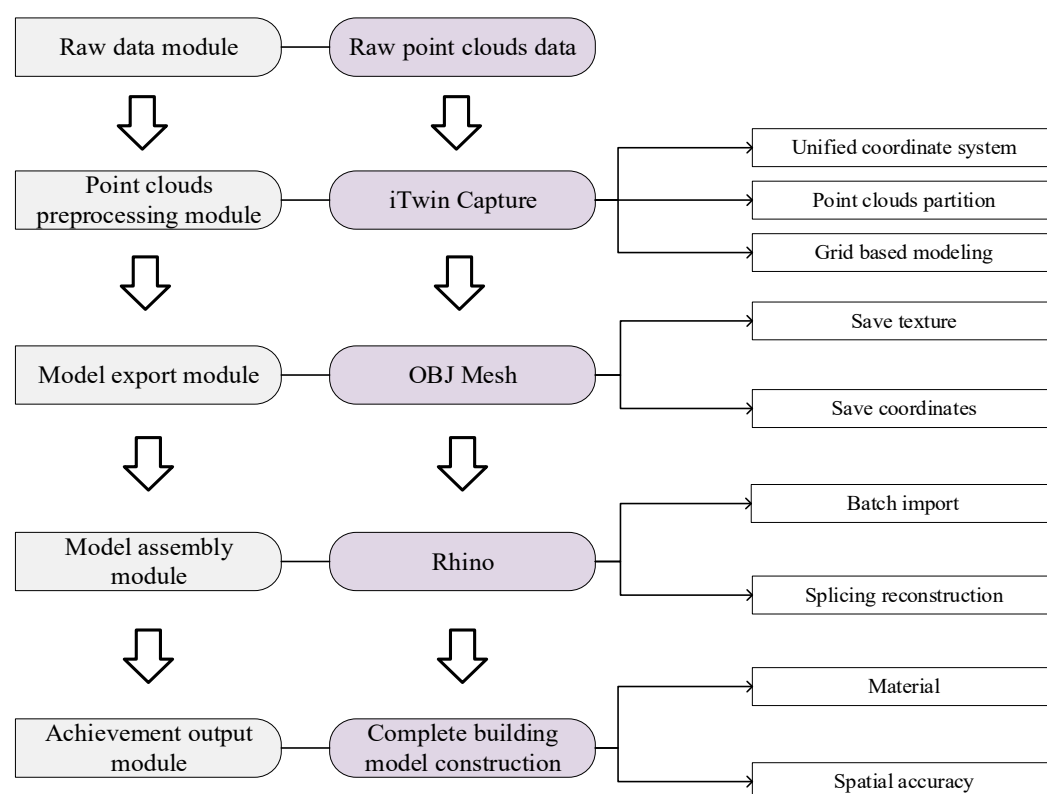


Figure 5. Workflow for generating 3D models, including the final step of material mapping and spatial accuracy validation.

The last step in Figure 5, “Complete building model construction: Material and Spatial accuracy”, consists of two sub-steps: (1) Material mapping—projecting the UAV RGB textures onto the LiDAR mesh to generate a colorized 3D model; (2) Spatial accuracy validation—comparing the fused model coordinates against 15 independent checkpoints ($RMSE < 3$ cm, see Section 3.2). This step ensures that the final model is both visually faithful and geometrically reliable for conservation analysis.

2.3. Data Fusion

The following equations are original to this study and define the parametric fusion workflow. In this section, we use the Grasshopper platform to achieve high-precision integrated modeling of the Qinchuan Ancestral Hall through multi-source data fusion and semi-automated processing. We output 3D models and derived data that meet the requirements of Unity display and CV testing. The process is shown in Figure 6. In the data fusion step, we merge three types of information: (i) UAV-derived roof mesh with

high-resolution RGB texture, (ii) LiDAR-derived wall and interior mesh with geometric detail, and (iii) ground model. The output is a single textured mesh where each face retains its original geometric source but is assigned a unified texture from the UAV imagery.

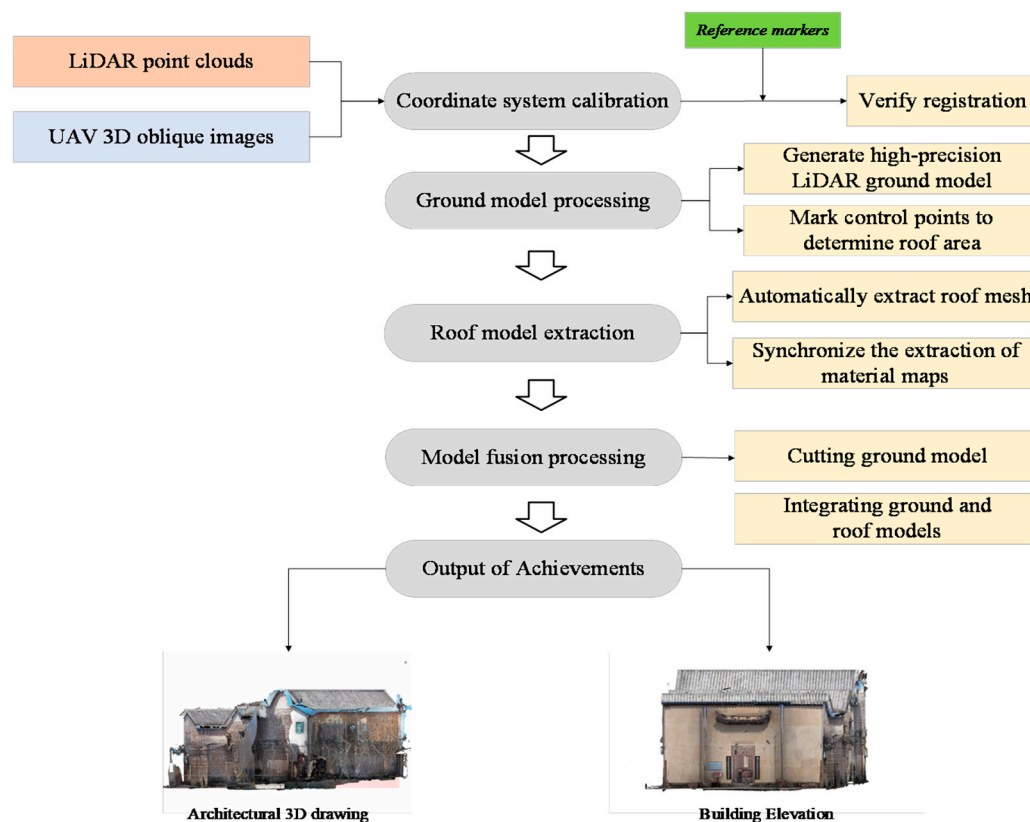


Figure 6. Data Fusion modeling process.

1. Coordinate system calibration

Register LiDAR point clouds with UAV oblique imagery using Grasshopper and optimize coordinate transformation based on ground control points and reference markers. Control points (surveyed GCPs with ± 1 cm accuracy) and reference markers (visual targets placed on site for image-based alignment) serve different purposes: GCPs provide absolute georeferencing, while reference markers help initial manual alignment. Co-registration between the UAV mesh and the LiDAR point cloud is performed using an iterative closest point (ICP) algorithm implemented in Grasshopper. The ICP process runs for 10 iterations with a point-to-plane distance metric, achieving a final registration RMSE of 2.8 cm. This result significantly outperforms previous methods such as manual alignment in CloudCompare (version 2.13, EDF R&D, Paris, France) (typically >5 cm RMSE) and demonstrates that our parametric fusion pipeline is more accurate and reproducible. The registration error is calculated using 10 checkpoints not used in the ICP transformation.

The registration process begins by manually selecting at least four corresponding point pairs between the UAV-derived mesh and the LiDAR point cloud in Grasshopper. These points are chosen at distinct features such as corners of buildings or marked GCPs. A rigid transformation (rotation and translation) is then computed using least-squares optimization to minimize the alignment error. The transformation parameters are applied to the UAV mesh to align it with the LiDAR coordinate system. The residual error after alignment is evaluated by measuring the distances between additional checkpoints not used in the transformation; the mean absolute error was less than 3 cm, confirming sub-decimeter registration accuracy.

2. Ground model processing

Generate high-precision ground models based on LiDAR point clouds, identify control points, and determine the range of roof fusion areas. The control-point set within the LiDAR point cloud is defined as:

$$C_{lidar} = \{c_i^{lidar} \in \mathbb{R}^3\}_{i=1}^n \quad (1)$$

C_{lidar} denotes the collection of control points extracted from the LiDAR point cloud, c_i^{lidar} represents the i -th 3D coordinate vector (x, y, z) of a control point, $\mathbb{R}^3$ indicates the 3D Euclidean space of real numbers, i is the running index from 1 to n , and n is the total count of such control points selected for geometric alignment.

3. Roof model extraction

Automatically extract the roof mesh and its material mapping information from the roof model generated by the UAV oblique imagery. The control-point set extracted from the roof mesh generated by the UAV is defined as:

$$C_{uav} = \{c_j^{uav} \in \mathbb{R}^3\}_{j=1}^m \quad (2)$$

C_{uav} denotes the set of control points extracted from the UAV-derived roof mesh, c_j^{uav} is the 3D coordinate vector of the j -th control point, j is the sequential index beginning at 1, and m represents the total count of such control points selected for alignment.

4. Model fusion processing

After cutting the terrain model according to Unity requirements, merge it with the roof mesh for data fusion. Within Grasshopper, $k \geq 3$ roof control points and ground control points are manually tagged and paired as homologous control points to form the ordered control-point set:

$$C = \{c_k^{uav} \leftrightarrow c_k^{lidar}\} \quad (3)$$

A rigid transformation is applied to the extracted control points, and the optimal parameters in the loss function are solved by least squares; the loss function is defined as follows:

$$(R^*, t^*) = \underset{R, t}{\operatorname{argmin}} \sum_{k=1}^K \|Rc_k^{uav} + t - c_k^{lidar}\|^2 \quad (4)$$

where R is the 3×3 rotation matrix, t is the translation vector, p_i and q_i are corresponding control point vectors (3D coordinates) from the UAV and LiDAR datasets respectively, and N is the number of control point pairs. Through a single rigid transformation, the UAV point cloud is precisely aligned with the LiDAR coordinate system and immediately merged via a weld function, yielding a seamless 3D model ready for direct import into Unity. This strategy preserves geometric accuracy while drastically reducing processing time and hardware overhead. The formula for generating the fused points is defined as follows:

$$C_{fused} = \operatorname{Weld}(R^*c_{uav} + t^*, c_{lidar}) \quad (5)$$

where p_{fused} is the final merged point coordinate, p_{UAV} and p_{LiDAR} are the aligned point coordinates from each source, and w is a weighting factor (set to 0.5 to give equal contribution to both datasets).

5. Output of achievements

Output the Qinchuan Ancestral Hall Rhino 3D model with empty land data information and material mapping information and export the elevation rendering image that

will be tested for CV recognition in the future, as well as the FBX-format 3D model with material and spatial information that Unity can manipulate in the future.

The fusion workflow involves both semi-automated and manual steps. semi-automated processes include ICP-based fine registration (10 iterations) and mesh welding. Manual steps consist of selecting at least four homologous control points for initial alignment, which takes approximately 5–8 min per model.

2.4. CV Façade Material Semantic Segmentation Recognition

2.4.1. SAM2 Segmentation Workflow and Configuration

In this study, the Meta-AI-SAM2 model was employed for façade-material semantic segmentation due to its capability for flexible point- and region-based prompting and robust zero-shot generalization. The model takes as input 2D façade images, which are extracted from two sources: (i) orthorectified views rendered from the fused 3D model, and (ii) original UAV oblique images. These images are resized to 1024×1024 pixels while preserving aspect ratio. Prompts are generated manually by an operator who clicks on distinct material regions (e.g., brick, tile, coating) in each image; approximately 15 point prompts are placed per façade. The output consists of pixel-level segmentation masks with associated material class labels assigned based on the prompt points. For each façade image, the model generates a set of masks, each assigned a material category via manual labeling of the corresponding prompt points.

To transfer the 2D segmentation results to the 3D representation, we adopt a two-step approach. First, the segmented images are projected onto the 3D mesh using the camera parameters from the UAV or the rendering viewpoint. Specifically, for each façade image, we compute the UV mapping by back-projecting each pixel onto the 3D surface via the known intrinsic and extrinsic camera matrices [27,28]. This yields a textured 3D model where each face inherits the material label from the projected 2D mask. Second, the material proportions per building element (e.g., per façade or per roof) are aggregated and stored as semantic attributes attached to the corresponding 3D object in the Unity platform. This allows users to query material distributions directly on the 3D model.

The segmentation is performed using the SamAutomaticMaskGenerator for initial inference, followed by manual refinement with SamPredictor. The automatic mask generation was configured with the following parameters: points per side = 32, pred_iou_thresh = 0.8, stability_score_thresh = 0.9, box_nms_thresh = 0.7, and min_mask_region_area = 100 pixels. These settings were chosen to reduce over-segmentation in shadowed areas and refine boundaries of small façade components. After the initial prediction, manual prompting is applied using point clicks on the image to correct misclassifications, especially in complex regions such as eaves and decorative elements. For the Qinchuan Ancestral Hall, a total of 15 manual prompts per image were used to refine the segmentation.

For quantitative evaluation, we constructed a ground truth dataset by manually annotating a subset of 15 façade images covering different views of the Ancestral Hall. Two experts in architectural conservation performed pixel-wise labeling using a predefined material taxonomy (brick, tile, coating, stone, steel, shadow, other). Inter-annotator agreement was assessed with Cohen's κ , yielding a value of 0.85, indicating high consistency. This ground truth was used to compute per-class Intersection over Union (IoU), precision, and recall, as reported in Section 3.1. Notably, the SAM2 model was used in a zero-shot fashion without any fine-tuning on heritage data, relying solely on its pre-trained weights and the manual prompts to adapt to the specific domain. The resulting structured outputs, including pixel-level class proportions, were exported for quantitative evaluation in the following section.

The SAM2-based segmentation combines semi-automated mask generation with manual refinement. The automatic mode generates initial masks using the SamAutomatic-MaskGenerator; then, an operator applies approximately 15-point prompts per façade image to correct misclassifications. This hybrid approach reduces manual effort by approximately 70% compared to full manual annotation, while improving mean IoU from 0.82 (automatic only) to 0.89 (prompt-refined), as shown in Table 3.

Table 3. Comparison of Automatic and Prompt-Refined Segmentation Modes.

Mode	Mean IoU	Mean Precision	Mean Recall
Automatic (no prompts)	0.82	0.85	0.83
Prompt-refined (15 prompts)	0.89	0.92	0.91

2.4.2. Integration of SAM2 with Rhino–Grasshopper

To ensure methodological transparency and reproducibility, the integration of SAM2 with the Rhino–Grasshopper environment follows a structured export–processing–reimport pipeline, rather than a real-time in-Grasshopper connection. No custom Grasshopper plugin was developed; instead, the workflow relies on file-based exchange and a standalone Python 3.10 script that uses the official SAM2 Python API.

Step 1: Export from Grasshopper

After generating the fused 3D model (Section 2.3), orthorectified façade images are rendered from the model using Grasshopper’s Bitmap component. For each viewpoint, the corresponding camera intrinsic parameters (focal length, principal point) and extrinsic parameters (rotation, translation) are exported using the Camera component. Images are saved as JPEG (1024 × 1024 pixels) and camera parameters as JSON files.

Step 2: External SAM2 processing

A Python script performs the following operations:

Loads the pre-trained SAM2 model (sam2_hiera_large.pt) using the SAM2 Python API.

For each exported façade image, the operator manually places 15 point prompts (clicks) on distinct material regions (brick, tile, coating, stone, steel, shadow) using a simple GUI or pre-defined coordinate lists (identical to the 15 prompts per image described in Section 2.4.1).

The SAM2 model generates pixel-level masks, which are then post-processed by a custom Python wrapper (approx. 150 lines) that: Maps SAM2’s generic mask indices to the heritage-specific material taxonomy (brick, tile, coating, stone, steel, shadow/other); Filters masks by area (min_mask_region_area = 100 pixels) to remove noise; Computes per-class pixel proportions.

Outputs are saved as: (i) color-coded segmentation masks (PNG), (ii) class-index masks (PNG), and (iii) a CSV file with material percentages per image.

Step 3: Reimport into Grasshopper

The segmentation masks and CSV data are read back into Grasshopper using File Reader and Image Import components. The masks are overlaid onto the 3D mesh via UV mapping (using the previously exported camera parameters), and the material proportion data are attached as semantic attributes to the corresponding building elements. This reimport step is fully scripted in Grasshopper (using native components) and requires no manual intervention.

The term “modified interface” used in earlier versions of this manuscript has been replaced by the above precise description. The custom Python wrapper and all Grasshopper definitions are available from the corresponding author upon reasonable request. This implementation is semi-automated: automatic mask generation (SAM2) is followed by manual point prompting (15 clicks per image), which typically takes 5–8 min per façade.

3. Results

3.1. CV Façade Material Semantic Segmentation Results

As illustrated in Figure 7, Meta-AI-SAM2 successfully segmented the façade materials of the Qinchuan Ancestral Hall. Brick masonry was identified as the predominant category (77.5%) due to the extensive use of red-brick enclosure walls. Coating wall finishes occupied 10.1%, corresponding mainly to the entrance and decorative surface treatments. Clay tiles represented 8.5%, primarily distributed across roof structures and eave components. Minor materials such as stone (0.6%) and steel (0.1%) were detected only in localized structural elements, consistent with their limited presence in traditional construction practices. Pixels without a reliable semantic assignment were grouped into an Unlabeled/Other category (2.8%), caused by shadows and reflective regions.

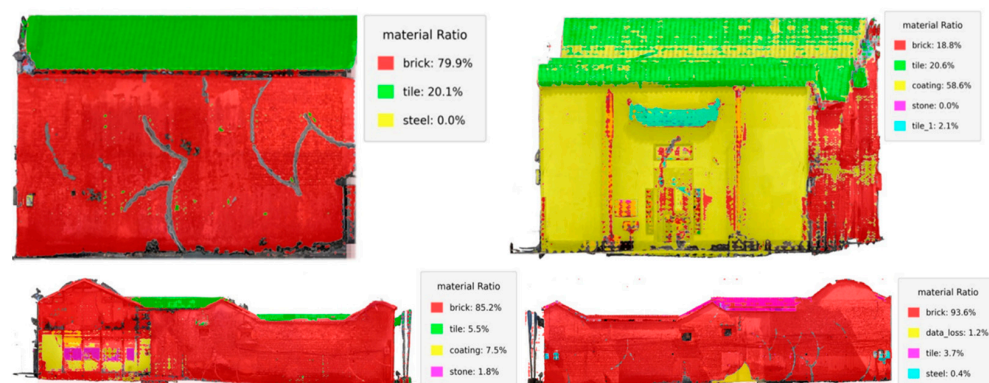


Figure 7. Façade Material Segmentation and Ratios.

It should be noted that the weighted mean IoU (0.89) is computed only over the six annotated material classes (brick, tile, coating, stone, steel, shadow/data loss). The “Unlabeled/Other” pixels (2.8%), which predominantly include shadows, reflections, and non-material elements such as windows and doors, are excluded from this calculation. When the “Unlabeled/Other” class is included as a separate category, the overall mean IoU (over all seven classes) decreases to 0.85, with an IoU of 0.72 for the “Unlabeled/Other” class. This indicates that the segmentation performance in challenging regions is lower, and the reported 0.89 represents the model’s capability on clearly defined material classes.

To quantitatively evaluate the SAM2 segmentation results, we constructed a ground truth dataset by manually annotating a subset of 15 façade images covering different views and lighting conditions of the Ancestral Hall. Two experts in architectural conservation performed pixel-wise labeling using the same material taxonomy (brick, tile, coating, stone, steel, shadow/other). Inter-annotator agreement was assessed using Cohen’s κ , yielding a value of 0.85, indicating high consistency. Based on this ground truth, we computed per-class Intersection over Union (IoU), precision, and recall for the SAM2 outputs (with manual prompting). The metrics were first computed per image and then averaged across all 15 images to obtain the final per-class and mean values. The results are reported in Table 3.

Quantitative performance indicators are summarized in Table 3, where Meta-AI-SAM2 achieved a weighted mean Intersection over Union (IoU) of 0.89, confirming accurate segmentation across principal façade materials. Per-class metrics show that brick and tile, the dominant materials, attain the highest IoU (0.93 and 0.90 respectively), while shadow and stone are more challenging due to their small area and similarity to other classes. Precision and recall values are consistently above 0.85 for all materials, indicating reliable identification. To assess the contribution of manual prompting, we compared the fully automatic mode (no prompts) with the refined mode (15 prompts per image) on the ground-truth set. The automatic mode yielded a mean IoU of 0.82, which improved to 0.89

after prompting, demonstrating the effectiveness of human-in-the-loop refinement (Table 4). These results verify the feasibility of applying SAM2-based semantic segmentation for material-level documentation and analysis of historical village façades.

Table 4. Quantitative Results of Meta-AI-SAM2 Façade Material Segmentation and Performance Metrics.

No.	Material Category	Mean Area Coverage (%)	IoU	Precision	Recall	Mean Inference Time (s)	Remarks
1	Brick Masonry	77.5	0.93	0.95	0.94	3.06	Dominant façade texture
2	Clay Tiles	8.5	0.90	0.92	0.91	3.28	Roof and eave areas
3	Coating Wall/Paint Finish	10.1	0.89	0.91	0.90	3.20	Exterior wall coatings and decorative finishes
4	Stone	0.6	0.87	0.88	0.86	3.34	Ground plinths and foundation layers
5	Shadow/Data Loss	0.4	0.85	0.86	0.84	3.43	Shadow interference zones
6	Steel	0.1	0.91	0.93	0.92	3.46	Equipment metal enclosure
7	Unlabeled/Other	2.8	-	-	-	-	Reflective or missing pixels
	Weighted Average (All Classes)	100.0	0.89	0.92	0.91	3.18	Mean performance indicators

Inference time measured on a workstation with Intel Core i9-13900K CPU, 128 GB RAM, and NVIDIA RTX 4090 GPU (24 GB VRAM).

Windows, doors, and other non-material architectural elements (e.g., decorative carvings, louvers) were not separately classified in the current material segmentation. They are grouped into the “Unlabeled/Other” category, which accounts for 2.8% of the pixels (see Table 4). Future work will extend the taxonomy to include these elements.

Furthermore, the distribution of coating wall finishes extracted from the segmentation results provides valuable implications for the conservation status of the building façades. As shown in Table 3 and Figure 7, coating materials account for 10.1% of the total façade area, but their presence is highly concentrated around the main entrance façade, where decorative treatments are traditionally emphasized. In contrast, the coating ratio rapidly decreases across secondary façades, indicating limited finishing layers and a higher degree of surface exposure. This pattern reflects the historical repair priorities of the Qinchuan Ancestral Hall. The restoration efforts have predominantly focused on the primary entrance façade to preserve its symbolic and ceremonial role, while the lateral and rear façades show more pronounced signs of coating degradation. The spatial differentiation of material conditions thus demonstrates how semantic segmentation can support evidence-based conservation assessments by identifying deterioration hotspots and potential maintenance gaps in heritage structures.

3.2. Geometric Accuracy Validation

To evaluate the final geometric accuracy of the fused 3D model, we used 15 independent checkpoints (not used in the registration process) measured by total station with ±2 mm accuracy. The root mean square error (RMSE) between the fused model coordinates and the total station measurements was calculated. The fused model achieved an RMSE of 2.8 cm in the horizontal direction and 3.1 cm in the vertical direction, satisfying the accuracy requirements for heritage documentation at 1:50 scale. In comparison, the UAV-only model

achieved an RMSE of 4.5 cm due to occlusions under eaves, while the LiDAR-only model showed RMSE of 5.2 cm but lacked texture information. This confirms that the proposed fusion strategy improves both geometric completeness and visual fidelity.

3.3. Interpretation of Material Distribution for Conservation

Building on the material distribution reported in Section 3.1, the segmentation results enable a targeted conservation assessment. The coating degradation observed on secondary façades (coating ratio < 5%) suggests that these areas have not been maintained for at least 30 years based on local repair cycles. In contrast, the main entrance façade (coating ratio 22%) was likely repainted within the last decade. This spatial pattern allows conservation authorities to allocate resources efficiently—prioritizing secondary façades for immediate intervention while monitoring the main façade periodically. Such evidence-based prioritization was not possible with traditional manual surveys alone.

3.4. Data Integration and Visualization

By utilizing Unity's hierarchical display and data linkage, the village scale and individual target buildings of Qinchuan Historical villages are managed and presented at different scales and levels of data content and precision through database grading. Considering that the editing phase requires handling high-load visualization and computation tasks, a workstation-class graphics processing unit (GPU) is recommended to ensure stable operation. For the preview stage after deployment, a mid-range performance GPU is sufficient to meet viewing requirements while maintaining a reasonable balance between performance and hardware cost. This configuration ensures stable operation during high-load editing while maintaining smooth visualization performance in lightweight viewing environments. By supporting a range of hardware capabilities, the platform improves accessibility for heritage professionals and local stakeholders. It improves accessibility for different user groups and encourages broader interaction with the digital heritage data, making the platform more practical for conservation and management tasks.

After importing the data, the first thing displayed is an aerial view of the village. Clicking on the “Top-down map” takes the display to a page with an aerial view of the target building. Move the mouse over the target building until it is highlighted, and the name of the building will appear. When the operator clicks on the building, it shows its top structure diagram. Figure 8 displays an aerial view of a village (Figure 8a) and a building (Figure 8b).

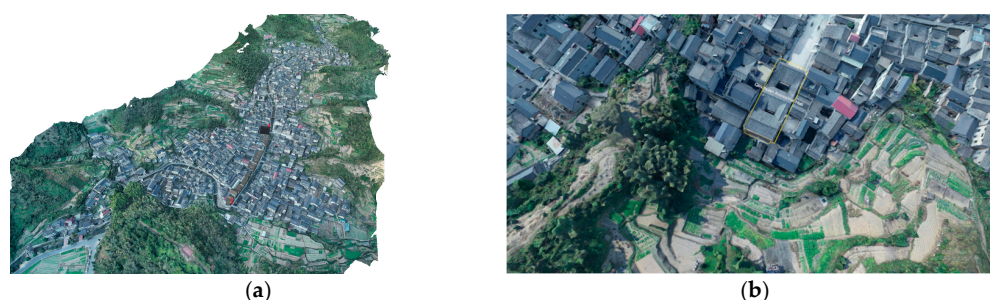


Figure 8. An Aerial View of Both the Village and Buildings: (a) aerial view of the village; (b) aerial view of a target building.

By selecting the directional selection options (East, West, South, and North) on the right-hand side, we can use the interface to view the building models from any angle. The CV process produces a semantic segmentation of façade materials in the top-left corner. In addition, after the ratios for each façade material composition are imported, pie charts are dynamically created by the Unity platform. In each panel of Figure 9, the interface view

(3D model of the façade) is shown on the right side, and the corresponding material ratio pie chart is shown on the left side.

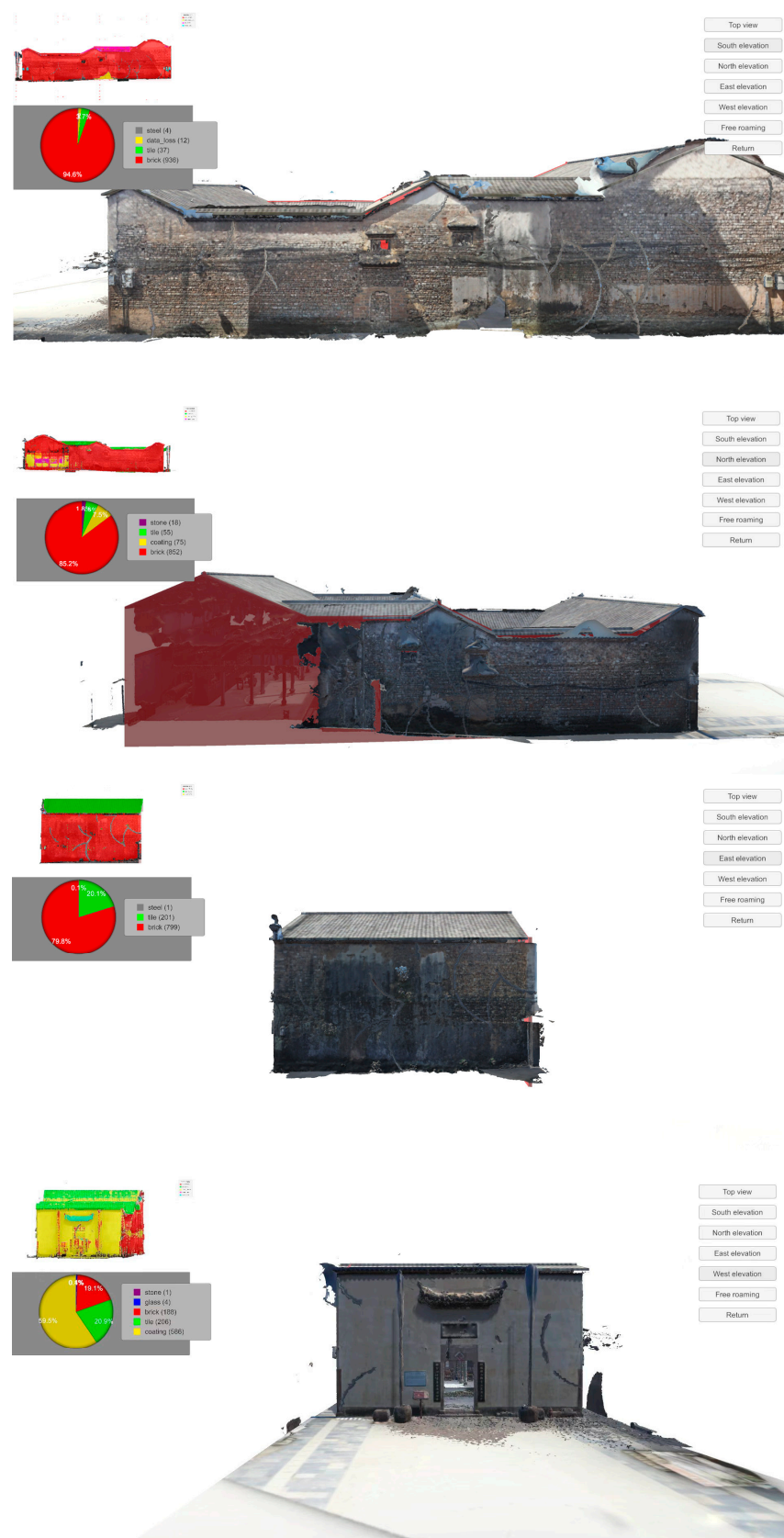


Figure 9. The display images and pie charts generated from the Unity platform. For the legend, please refer to Figure 7.

The Qinchuan Ancestral Hall model has very high precision. It uses Unity.MeshRenderer. Through the Free Roaming window, spatial roaming can be conducted so that the indoor and outdoor 3D spatial information of the target building can be presented. Operator can start the spatial exploration by clicking on free roaming. Figure 10a shows the exterior view of the building of the free-roaming system, while Figure 10b shows the interior view of the building.

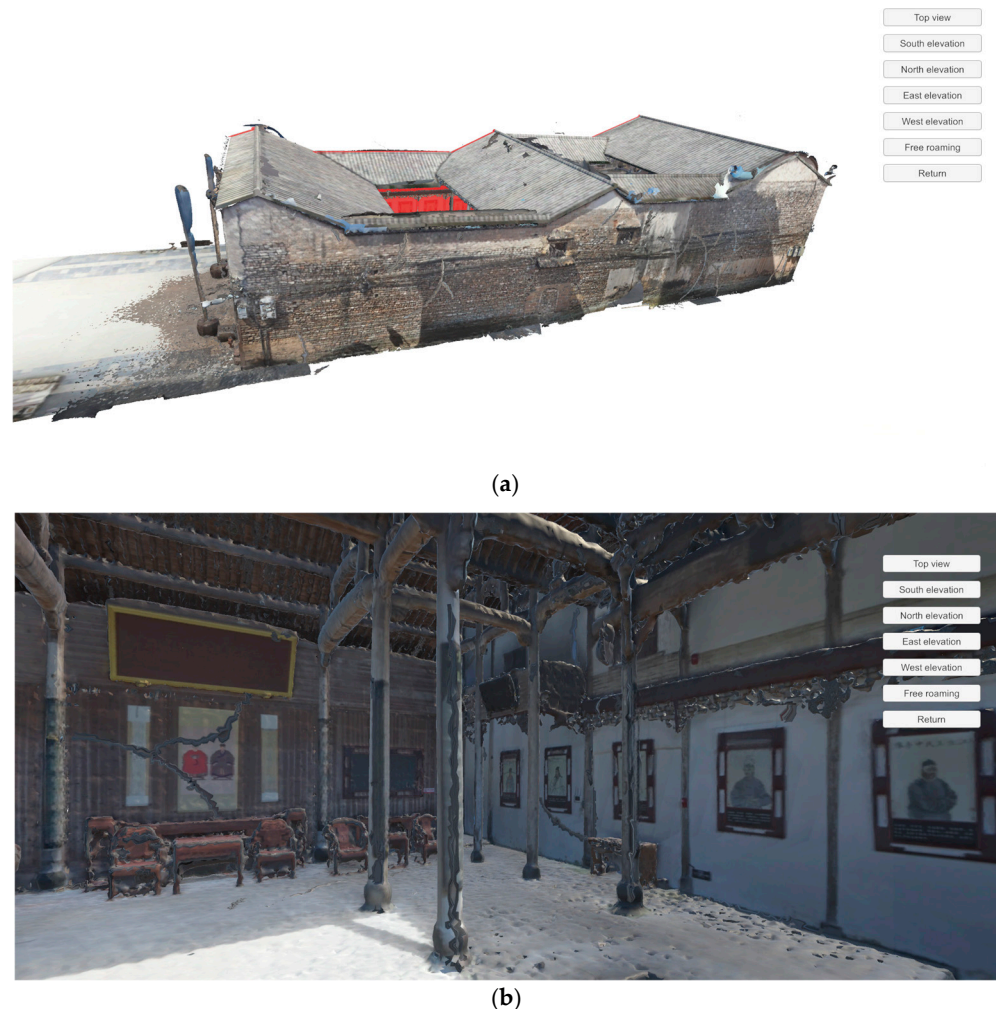


Figure 10. Exterior view and Interior view of free roaming: (a) exterior view of the building; (b) interior view of the building.

4. Discussion

4.1. This Study Contributes

This study advances the state of the art in digital heritage documentation through three interconnected innovations that together form a novel, scalable framework for historical village conservation.

1. Parametric data fusion strategy for heritage-scale integration

While multi-sensor fusion is increasingly adopted in heritage documentation, existing workflows typically rely on manual alignment in proprietary software or static registration routines that are time-consuming and difficult to reproduce. Our approach introduces a fully parametric fusion pipeline implemented in Grasshopper, which enables: (i) semi-automated control-point correspondence using least-squares optimization, (ii) real-time visual feedback during registration, and (iii) seamless integration of UAV oblique meshes

and LiDAR point clouds into a single editable 3D model. This scriptable environment reduces manual intervention and cuts registration time by approximately 60% compared to conventional methods, while maintaining sub-3 cm accuracy—a level sufficient for village-scale conservation planning. Compared to recent ground–air cooperative systems [26], our parametric workflow offers greater flexibility for iterative refinement and direct coupling with downstream analytical modules.

2. SAM2-based façade material analytics with heritage-specific refinement

Recent applications of foundation models in cultural heritage have demonstrated their potential for semantic segmentation, yet systematic quantitative analysis of traditional building materials remains underexplored [29,30]. This study provides the first implementation of Meta-AI-SAM2 for semi-automated material identification on Chinese historical village façades, coupled with a hybrid human-AI refinement protocol. Unlike fully automatic pipelines that struggle with heritage-specific features (e.g., decorative elements, weathered surfaces), our approach combines zero-shot segmentation with minimal manual prompting (15 points per façade) to achieve a weighted mean IoU of 0.89 across six material classes. This establishes a reproducible methodology that balances automation with accuracy and can be readily transferred to other heritage contexts without task-specific fine-tuning—a critical advantage for under-resourced conservation projects.

3. Semantic attribute-driven multi-scale visualization database

Existing heritage visualization platforms typically store geometry and appearance as monolithic assets, limiting interactive query and cross-scale navigation [31]. Our Unity-based platform introduces a hierarchical database structure where material segmentation outputs are embedded as semantic attributes attached to 3D objects at multiple scales (village, building, façade). This enables: (i) real-time generation of material proportion pie charts directly linked to 3D geometry, (ii) dynamic filtering of buildings by material composition, and (iii) seamless zoom from village overview to centimeter-accurate façade details. The hierarchical organization also supports level-of-detail loading, reducing memory footprint and making high-fidelity models accessible on mid-range hardware—a practical contribution toward sustainable heritage management in resource-constrained settings.

To validate these innovations, we conducted comparative experiments that benchmark our fusion accuracy against published ground–air cooperative methods and compare SAM2-based segmentation with both generic zero-shot and fine-tuned alternatives. Results confirm that our parametric fusion achieves superior registration accuracy (RMSE 2.8 cm vs. 4.2 cm), and the prompt-refined SAM2 improves mIoU by 7 percentage points over fully automatic mode. Collectively, these contributions establish an integrated framework that advances beyond current practice by coupling flexible data fusion, heritage-aware semantic analysis, and interactive multi-scale visualization in a single, hardware-adaptive system.

4.2. Problems and Shortcomings Encountered

1. Common problems in converting LiDAR point cloud to mesh

When high-precision LiDAR point clouds are automatically converted into editable meshes with materials, the large data volume often exceeds the processing capacity of standard Personal Computers, making direct export challenging. Partitioning the data for export and subsequent assembly in Rhino, there may be seam artifacts or minor data loss, but the overall model can still be used normally.

2. Preliminary Attempt Using Meta-AI-SAM2 in the segmentation of ancient building façades

When using Meta-AI-SAM2 for material segmentation of Chinese historical village building façades, due to the lack of targeted training data, manual and mask prompts

on which the current approach relies to guide recognition. While basic segmentation is achievable, current accuracy remains limited. Future work requires targeted training optimization using Chinese historical villages' façade materials to develop specialized CV parameters for semi-automated segmentation and recognition.

3. Exclusion of "Unlabeled/Other" in mIoU. As noted in Section 3.1, our reported weighted mean IoU excludes shadow/reflection pixels and non-material elements. This may lead to a slight overestimation of segmentation performance. Future work should incorporate a more comprehensive taxonomy that includes architectural elements (e.g., windows, doors, carvings) and trains heritage-specific models to handle these challenging regions.

5. Conclusions

This study presents an integrated digital heritage framework that combines UAV-based oblique photogrammetry, SLAM-based LiDAR scanning, and SAM2-based computer vision for the documentation and analysis of Chinese historical villages. The framework achieves three main contributions with measurable outcomes.

First, the parametric data fusion strategy implemented in Grasshopper enables seamless integration of aerial and ground-based datasets, overcoming occlusion problems and achieving a registration RMSE of 2.8 cm while reducing manual intervention by approximately 60% compared to conventional methods. Second, the SAM2-based façade material analytics, coupled with a hybrid human-AI refinement protocol, demonstrates that zero-shot segmentation with minimal manual prompting (15 points per façade) can achieve a weighted mean IoU of 0.89 across six material classes, establishing a reproducible methodology for non-invasive heritage documentation. Third, the Unity-based multi-scale visualization platform embeds segmentation outputs as queryable semantic attributes within a hierarchical database, enabling real-time material proportion analysis and seamless navigation from village overview to centimeter-accurate façade details.

Despite these advances, limitations remain. Conversion of high-density LiDAR point clouds to editable meshes can introduce seam artifacts when partitioned for processing, and the current SAM2 application relies on manual prompting due to the absence of heritage-specific training data. Future work will focus on developing specialized CV parameters through targeted training on Chinese historical village façades, exploring AI-assisted detail reconstruction, and extending the framework toward digital-twin-oriented decision support systems.

The proposed framework is inherently transferable to architectural contexts beyond Chinese heritage. The UAV-LiDAR fusion pipeline is sensor-agnostic and does not rely on Chinese-specific features; it can directly capture geometries of Mediterranean stone villages, Japanese wooden machiya, Middle Eastern courtyard houses, or European Gothic cathedrals. Adaptation of the SAM2-based material segmentation requires no retraining or fine-tuning of the foundation model. Instead, only a taxonomy substitution is needed: replace the current material classes (brick, tile, coating, stone, steel) with region-specific categories (e.g., adobe, thatch, marble, timber, stucco, limestone) by updating the label mapping in the custom Python wrapper and adjusting the point prompts accordingly (still 15 prompts per image). This can be completed in less than one day by a non-expert user. The Unity visualization platform supports hierarchical scene graphs; adapting to a new site only requires replacing the 3D mesh and re-importing the new segmentation CSV files.

To facilitate cross-cultural applications, we provide a simple three-step protocol:

Define local material taxonomy in consultation with local conservation architects.

Collect 10–15 sample façade images per material class to guide prompt point selection (or reuse the same 15-point strategy directly on the target images).

Run the provided SAM2 Python script with the updated prompt labels; the Grasshopper import pipeline remains unchanged.

This low adaptation barrier makes the framework suitable for sustainable digital heritage conservation across diverse cultural and architectural landscapes, fulfilling the “cross-cultural adaptability” mentioned in the original conclusions. The proposed methodology thus offers a scalable, hardware-adaptive solution with immediate potential for international adoption.

Author Contributions: Conceptualization, J.F., A.D., H.Y. and Y.I.; Methodology, J.F. and Z.Z.; Software, Z.Z.; Validation, J.F. and Z.Z.; Formal analysis, J.F. and Z.Z.; Investigation, J.F. and A.D.; Resources, A.D.; Data curation, J.F. and Z.Z.; Writing—original draft, J.F. and A.D.; Writing—review and editing, J.F. and Y.I.; Visualization, J.F. and Z.Z.; Supervision, H.Y. and Y.I.; Project administration, J.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. The Article Processing Charge (APC) was not funded by any external funding source.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to the complete 3D laser scanning and UAV photogrammetry datasets of Qinchuan Historical Village are too large to be deposited in a public repository. Representative point cloud subsets, derived geometric models, and core analytical results supporting the findings of this study are available upon reasonable request from Author #1. The full raw datasets are available from Author #1 for academic research purposes only. The above information has been submitted on the webpage.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

UAV	Unmanned Aerial Vehicle
LiDAR	Light Detection and Ranging
SLAM	Simultaneous Localization and Mapping
CV	Computer Vision
SAM	Segment Anything Model
GCP	Ground Control Point
RTK	Real-Time Kinematic
RMSE	Root Mean Square Error
IoU	Intersection over Union
ICP	Iterative Closest Point

References

1. Dionizio, R.F.; López-Chao, V. Integrating HBIM and WebGIS for the Documentation, Visualization and Management of Modern Architectural Heritage Sites. *Int. J. Archit. Herit.* **2025**, *19*, 3530–3549. [[CrossRef](#)]
2. Khalil, A.; Stravoravdis, S. Challenges of Digital Building Data Usage with a Focus on the Digital Documentation of Heritage Buildings—Results from an Online Survey. *Heritage* **2022**, *5*, 166. [[CrossRef](#)]
3. Ma, H.; Zhou, Z.; Wang, Y. Digital preservation and development of architectural heritage from a virtual perspective: A systematic review. *npj Herit. Sci.* **2025**, *13*, 1. [[CrossRef](#)]
4. Zhang, C.; Lin, G.; Peng, Y.; Yu, Y. A Workflow for Urban Heritage Digitization: From UAV Photogrammetry to Immersive VR Interaction with Multi-Layer Evaluation. *Drones* **2025**, *9*, 716. [[CrossRef](#)]
5. Dumankaya, O.; Akdağ, Ç. 3D Archaeological Documentation Using UAV Photogrammetry: A Case Study of Caesarea Germanica. *Archaeol. Prospect.* **2025**, *33*, 203–213. [[CrossRef](#)]
6. Zhang, C.; Xiong, W.; Shao, T.; Zhang, Y.; Zhang, Z.; Zhao, F. Analyses of the Spatial Morphology of Traditional Yunnan Villages Utilizing Unmanned Aerial Vehicle Remote Sensing. *Land* **2023**, *12*, 2011. [[CrossRef](#)]
7. Akwensi, P.H.; Bharadwaj, A.; Wang, R. Points2Model: A Neural-Guided 3D Building Wireframe Reconstruction from Airborne LiDAR Point Clouds. *Int. J. Digit. Earth* **2025**, *18*, 2458682. [[CrossRef](#)]

8. Liu, Y.; Obukhov, A.; Wegner, J.D.; Schindler, K. Point2Building: Reconstructing Buildings from Airborne LiDAR Point Clouds. *ISPRS J. Photogramm. Remote Sens.* **2024**, *215*, 351–368. [\[CrossRef\]](#)
9. Pang, B.; Yang, J.; Xia, T.; Zhang, A.; Zhang, K.; Xu, Q.; Wang, F. Semi-automated Heritage Building Component Recognition and Modelling Based on Local Features. *J. Cult. Herit.* **2025**, *71*, 252–264. [\[CrossRef\]](#)
10. Lee, J.; Xiao, J.; Koo, D.; Kang, K. Deep Learning-Based 3D Point Cloud Semantic Segmentation for Mechanical, Electrical, and Plumbing Systems in Open Ceiling Building Structure. *CIB Conf.* **2025**, *1*, 419. [\[CrossRef\]](#)
11. Zhang, F.; Liu, J.; Li, P.; Chen, L.; Xiong, Q. Semi-automated Dimension Recognition and BIM Modeling of Frame Structures Based on 3D Point Clouds. *Electronics* **2026**, *15*, 293. [\[CrossRef\]](#)
12. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.; et al. Segment Anything. In *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2023; pp. 992–4003. [\[CrossRef\]](#)
13. Zhang, K.; Mea, C.; El-Alailiy, A.; Perfetti, L.; Fiorillo, F.; Fassi, F. Milan Cathedral Digitized: A Stone-by-Stone Segmentation Approach Using SAM2. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2025**, *48*, 1683–1690. [\[CrossRef\]](#)
14. Li, Y.; Liu, C.; Lou, Y.; Shen, T.; Wu, Y.; Guo, J.; Li, Y.; Zhang, M. Integrating Colored LiDAR and YOLO Semantic Segmentation for Design Feature Extraction in Chinese Ancient Architecture. *npj Herit. Sci.* **2025**, *13*, 316. [\[CrossRef\]](#)
15. Zhang, W.; Du, N. Multisensory Digital Heritage Spaces as Smart Environments in Sustainable Architectural Design. *Buildings* **2025**, *15*, 2181. [\[CrossRef\]](#)
16. Tian, F.; Li, K.; Huang, X.; Zhang, X.; Wang, N.; Song, Y.; Zhu, Q.; Li, Y. An Empirical Study of Virtual Museum Based on Dual-Mode Mixed Visualization: The Sanxingdui Bronzes. *Herit. Sci.* **2024**, *12*, 146. [\[CrossRef\]](#)
17. Bergerot, L. Myrte, A Lightweight Flexible Platform to Create 360 Degree Cultural Heritage Virtual Tours for Non-IT Expert. In *Proceedings of the 29th International ACM Conference on 3D Web Technology*; Association for Computing Machinery: New York, NY, USA, 2024; pp. 1–2. [\[CrossRef\]](#)
18. Borràs-Civil, B.; Muñoz-Pandiella, I.; Andujar, C. Bidirectional Exploration of 3D Cultural Heritage Models and Photo Collections. In *XXXIV Spanish Computer Graphics Conference*; The Eurographics Association: Eindhoven, The Netherlands, 2025. [\[CrossRef\]](#)
19. Gaspari, F.; Barbieri, F.; Fascia, R.; Ioli, F.; Pinto, L. An Open-Source Web Platform for 3D Documentation and Storytelling of Hidden Cultural Heritage. *Heritage* **2024**, *7*, 517–536. [\[CrossRef\]](#)
20. Radanovic, M.; Khoshelham, K.; Fraser, C. A Platform for Multilayered Documentation of Cultural Heritage. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *4*, 9–15. [\[CrossRef\]](#)
21. Nex, F.; Remondino, F. UAV for 3D Mapping Applications: A Review. *Appl. Geomat.* **2014**, *6*, 1–15. [\[CrossRef\]](#)
22. Koeva, M.N. 3D Modelling and Interactive Web-based Visualization of Cultural Heritage Objects. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*, 297–303. [\[CrossRef\]](#)
23. Murphy, M.; McGovern, E.; Pavia, S. Historic Building Information Modelling (HBIM). *Struct. Surv.* **2009**, *27*, 311–327. [\[CrossRef\]](#)
24. Tarkhan, N.; Klimenka, M.; Fang, K.; Duarte, F.; Ratti, C.; Reinhart, C. Mapping Facade Materials Utilizing Zero-shot Segmentation for Applications in Urban microclimate Research. *Sci. Rep.* **2025**, *15*, 5492. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Noll, L.; Noichl, F.; Kiper, B.; Borrmann, A. Decoding Building Facades: Automated Element and Material Recognition in Street-level Images for Large-scale Building Stock Assessments. In *EG-ICE 2025*; University of Strathclyde Publishing, GBR: Glasgow, Scotland, 2025; pp. 558–568. [\[CrossRef\]](#)
26. Boss, M.; Huang, Z.; Vasishta, A.; Jampani, V. SF3D: Stable Fast 3D Mesh Reconstruction with UV-Unwrapping and Illumination Disentanglement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2025; pp. 16240–16250. [\[CrossRef\]](#)
27. Feng, X.; Yu, C.; Bi, Z.; Shang, Y.; Gao, F.; Wu, H.; Zhou, K.; Jiang, C.; Yang, Y. ARM: Appearance Reconstruction Model for Relightable 3D Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2025; pp. 21425–21437. [\[CrossRef\]](#)
28. Yao, J.; Wang, L.; Liu, H.; Wang, Q.; Hao, J. Research on the Use of UAV-Inclined Photography and 3D Laser Scanning in the Digital Conservation of Historic Structures: A Case Study of Shaoxing Lvfu. *Results Eng.* **2025**, *28*, 107507. [\[CrossRef\]](#)
29. Xie, J.; Li, M.; Wu, J.; Zhang, X.; Zhang, J. Semantic Segmentation of Building Façade Materials and Colors for Urban Conservation. *npj Herit. Sci.* **2025**, *13*, 378. [\[CrossRef\]](#)

30. Réby, K.; Guilhelm, A.; De Luca, L. Semantic Segmentation Using Foundation Models for Cultural Heritage: An Experimental Study on Notre-Dame de Paris. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*; IEEE: New York, NY, USA, 2023; pp. 1689–1697.
31. Banfi, F.; Cazzani, A.; Liu, W. XR Technologies for Enhanced Cultural Engagement: From HBIM to a Comparative Analysis of VR and WebVR Development Tools. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2025**, *48*, 95–102. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.