drones

MDPI

Article

Cave of Altamira (Spain): UAV-Based SLAM Mapping, Digital Twin and Segmentation-Driven Crack Detection for Preventive Conservation in Paleolithic Rock-Art Environments

Jorge Angás [1,2,*], Manuel Bea [3], Carlos Valladares [3], Cristian Iranzo [4], Gonzalo Ruiz [5], Pilar Fatás [6], Carmen de las Heras [6], Miguel Ángel Sánchez-Carro [7], Viola Bruschi [7], Alfredo Prada [6] and Lucía M. Díaz-González [6]

1 Department of Ancient Sciences and Institute of Heritage and Humanities (IPH), ARAID–University of Zaragoza, 50009 Zaragoza, Spain
2 Escuela Técnica Superior de Ingenieros en Topografía, Geodesia y Cartografía, Universidad Politécnica de Madrid, 28031 Madrid, Spain
3 Department of Ancient Sciences and Institute of Heritage and Humanities (IPH), University of Zaragoza, 50009 Zaragoza, Spain; manubea@unizar.es (M.B.); cvalladares@unizar.es (C.V.)
4 Department of Geography and Spatial Management-Environmental Sciences Institute (IUCA), University of Zaragoza, 50009 Zaragoza, Spain; c.iranzo@unizar.es
5 Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50009 Zaragoza, Spain; gruiz@bifi.es
6 Museo Nacional y Centro de Investigación de Altamira, Ministerio de Cultura, 39330 Santillana del Mar, Spain; pilar.fatas@cultura.gob.es (P.F.); carmen.delasheras@cultura.gob.es (C.d.l.H.); alfredo.prada@cultura.gob.es (A.P.); lucia.diaz@cultura.gob.es (L.M.D.-G.)
7 Escuela de Ingenieros de Caminos, Canales y Puertos, Universidad de Cantabria, International Institute for Prehistoric Research of Cantabria (IIIPC), 39005 Santander, Spain; miguelangel.sanchez@unican.es (M.Á.S.-C.); viola.bruschi@unican.es (V.B.)
* Correspondence: j.angas@unizar.es

Highlights

What are the main findings?

- The UAV-based inspection enabled highly precise 3D reconstruction of the inaccessible rock wall above the La Hoya Hall, overcoming severe geometric, lighting and safety constraints that prevent conventional geomatics surveys.
- This study reports the first documented deployment of a LiDAR-SLAM confined-space UAV inside a Paleolithic World Heritage cave for structural monitoring, revealing active fractures, unstable blocks and sediment accumulations inaccessible to conventional methods.

What are the implications of the main findings?

- The integration of LiDAR–SLAM, videogrammetry, and deep learning–based crack detection demonstrates the potential of an integrated geomatics workflow to support the identification and assessment of geological instabilities in fragile subterranean environments under severe operational constraints.
- The incorporation of these datasets into a Digital Twin framework provides a structured basis for multitemporal analysis, expert-driven annotation, and informed decision-making, contributing to the development of long-term preventive conservation and monitoring strategies.

Check for updates

Academic Editor: Giordano Teza

Received: 1 December 2025
Revised: 15 January 2026
Accepted: 17 January 2026
Published: 22 January 2026
Corrected: 16 October 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) license.

Abstract

The Cave of Altamira (Spain), a UNESCO World Heritage site, contains one of the most fragile and inaccessible Paleolithic rock-art environments in Europe, where geomatics

documentation is constrained not only by severe spatial, lighting and safety limitations but also by conservation-driven restrictions on time, access and operational procedures. This study applies a confined-space UAV equipped with LiDAR-based SLAM navigation to document and assess the stability of the vertical rock wall leading to "La Hoya" Hall, a structurally sensitive sector of the cave. Twelve autonomous and assisted flights were conducted, generating dense LiDAR point clouds and video sequences processed through videogrammetry to produce high-resolution 3D meshes. A Mask R-CNN deep learning model was trained on manually segmented images to explore automated crack detection under variable illumination and viewing conditions. The results reveal active fractures, overhanging blocks and sediment accumulations located on inaccessible ledges, demonstrating the capacity of UAV-SLAM workflows to overcome the limitations of traditional surveys in confined subterranean environments. All datasets were integrated into the DiGHER digital twin platform, enabling traceable storage, multitemporal comparison, and collaborative annotation. Overall, the study demonstrates the feasibility of combining UAV-based SLAM mapping, videogrammetry and deep learning segmentation as a reproducible baseline workflow to inform preventive conservation and future multitemporal monitoring in Paleolithic caves and similarly constrained cultural heritage contexts.

Keywords: confined-space UAV; LiDAR SLAM; videogrammetry; structural monitoring; digital twin; rock wall stability in Paleolithic caves; cultural heritage conservation; Cave of Altamira; deep learning crack detection

1. Introduction

The Cave of Altamira (Santillana del Mar, Cantabria, Spain) (Figures 1 and 2) contains one of the most remarkable ensembles of Paleolithic rock art in Europe, featuring representations of horses, deer, polychrome bisons and hand stencils [1–8]. The site was inscribed on the UNESCO World Heritage List in 1985 for its Outstanding Universal Value. Its significance lies not only in the exceptional artistic quality of the paintings, among the earliest recognized as Paleolithic in Europe, but also in the remarkable preservation of the karstic environment in which they are embedded.

The inherent vulnerability of rock art makes structural monitoring a key tool for preventive conservation and long-term heritage management [9–13]. This is particularly critical in subterranean contexts, where confined and environmentally sensitive conditions amplify the challenges associated with documentation, monitoring and conservation efforts. Decorated caves are highly stable but sensitive micro-ecosystems that can be disrupted by variations in microclimate, ventilation patterns, CO_2 concentration, humidity or anthropogenic presence [14]. Such alterations favour microbial colonization, condensation processes and the physical degradation of pigments and rock surfaces. In caves with rock art, studies have shown that imbalance in environmental conditions facilitates microbial growth and biodeterioration, while the geological dynamics of karstic systems (such as collapses, structural instability or ground movement) further complicate preservation strategies. In the Cave of Altamira, these factors have historically required consolidation interventions and strict environmental monitoring protocols to ensure long-term conservation.

In December 2024, the geomatics recording of the rock wall at the entrance to the La Hoya Hall, located inside the Cave of Altamira, was carried out as part of the experimental project for documentation and monitoring within the cave.

This intervention forms part of a methodological proposal for conservation that relies on geomatics techniques, including videogrammetry, 3D laser scanning, and high-

resolution image acquisition. The action addresses a geological issue related to the need for structural control and monitoring of this vertical surface. The wall exhibits a marked state of deterioration, evidenced by the presence of cracks of varying lengths and dimensions, as well as by a substantial accumulation of sedimentary material along the different natural ledges that define it. Furthermore, the height of the wall and the configuration of these ledges constitute the main challenge for accurate documentation, as they hinder physical access and prevent direct observation.

Figure 1. Location of the study area on the PNOA-IGN cartographic basemap. Topographic projection of the cave plan adapted from the Spanish Ministry of Culture. Instituto Geográfico Nacional (IGN). (2025). *PNOA orthophoto map*. Centro Nacional de Información Geográfica.

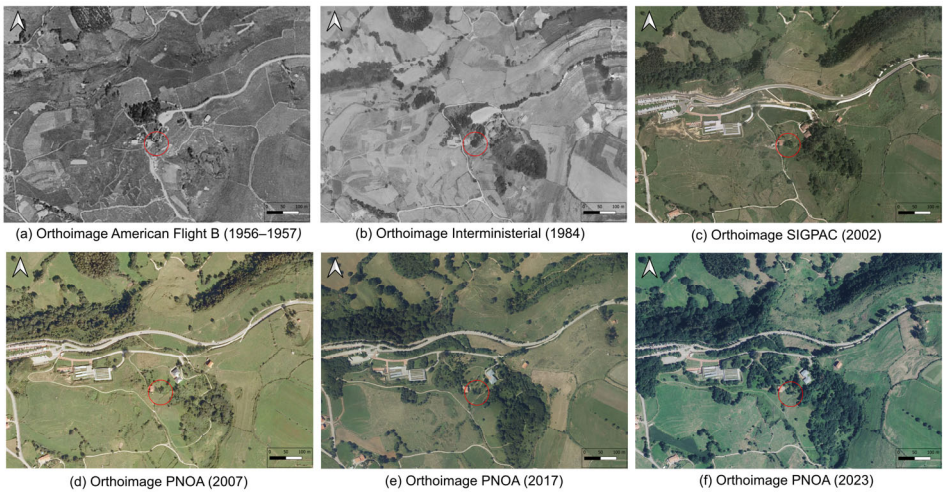

Figure 2. Landscape evolution of the cave area (red circle) from the mid-20th century to the present, derived from IGN base data. Instituto Geográfico Nacional (IGN). (2025). *PNOA orthophoto map*. Centro Nacional de Información Geográfica.

To overcome these constraints, we adopted a secure and controlled inspection strategy based on a UAV specifically designed for confined-space operations, originally developed for industrial inspection. The platform combines 4K RGB imaging with onboard ranging

and SLAM-based positioning, enabling the acquisition of video sequences that can be spatially referenced within a 3D reconstruction of the cave environment. In this setting, 3D videogrammetry has become a well-established approach for deriving metric 3D models from image sequences, with successful applications reported in biomechanics [15], biomedical engineering [16–18], sports performance monitoring [19], and the documentation of architectural heritage in narrow spaces [20].

In cultural heritage research, integrated 3D modelling solutions combining videogrammetry with V-SLAM or laser/spherical SLAM have been proposed for complex environments [21,22], and related workflows have demonstrated the potential of UAV video for detailed urban modelling [23] and its integration into HBIM environments [24]. In this study, the integration of videogrammetric outputs with SLAM-derived geometry supports the metric characterization of the wall and provides a baseline for future structural monitoring. Building on this acquisition framework, the extracted imagery was subsequently used to support automated crack mapping through a Mask R-CNN–based segmentation model [25].

Mask R-CNN has emerged as one of the most influential deep learning frameworks for instance segmentation and has seen growing adoption in crack detection across a range of scientific and engineering domains [26–30]. Its capacity to generate pixel-level segmentation masks while simultaneously performing object detection makes it particularly well suited to challenging environments such as Paleolithic caves, where cracks are often subtle, morphologically irregular, and embedded in noisy or textured backgrounds [31,32]. Building on the Faster R-CNN architecture, Mask R-CNN incorporates a dedicated mask prediction branch and typically employs a deep convolutional backbone (e.g., ResNet) combined with a Region Proposal Network, thereby enabling the precise localization and delineation of fine surface discontinuities [29].

Beyond Mask R-CNN, crack segmentation research has been largely driven by fully convolutional semantic segmentation architectures, particularly FCN-, U-Net- and DeepLab-based models. These approaches, often combined with multi-scale feature aggregation, deep supervision and enhanced feature reuse, have demonstrated strong performance in controlled engineering contexts such as pavements, tunnel linings and concrete infrastructures, where illumination conditions, imaging geometry and annotation protocols are comparatively homogeneous. Representative examples include DeepCrack, CrackSegNet and U-CliqueNet, as well as more recent lightweight and real-time variants designed to balance accuracy and computational efficiency [12,33–35]. In parallel, detection-based and instance-aware pipelines, including YOLO-derived architectures and hybrid detection–segmentation frameworks, have gained increasing attention due to their ability to support real-time inference and to delineate individual crack instances [36]. Recent developments also explore generative and data-augmentation-driven strategies, such as GAN- and diffusion-based models, to mitigate data scarcity and domain shift in crack segmentation tasks [37]. However, it is important to note that the performance reported in these studies is typically obtained under domain-specific benchmark conditions that differ substantially from subterranean cultural heritage environments. Paleolithic cave surfaces present extreme challenges, including low and highly heterogeneous illumination, irregular rock textures, complex relief and limited opportunities for systematic annotation, which hinder the direct transferability of existing architectures and evaluation metrics [38]. Within this context, Mask R-CNN was selected for the present study due to its instance segmentation capabilities, its flexibility under limited training data through transfer learning and its ability to localize individual fractures while preserving spatial coherence [19]. Accordingly, this work does not aim to compete quantitatively with state-of-the-art crack segmentation benchmarks, but rather to explore the feasibility and limitations of applying

deep learning–based crack detection under the severe constraints imposed by a protected subterranean heritage site.

Therefore, we present the results of applying a confined-space drone equipped with LiDAR-SLAM, integrating videogrammetry and automated crack segmentation to assess rock-surface stability and to put forward a new preventive-conservation approach. All the collected information was subsequently integrated into a digital twin of the cave through the DiGHER platform, in order to facilitate collaborative analysis.

2. Materials and Methodology

2.1. Study Area

The Cave of Altamira is located on the western flank of a small karstic valley developed in Cretaceous limestones. Extending for approximately 290 m, the cave comprises a main gallery formed by dissolution processes typical of the Cantabrian karst system, from which two major chambers branch off: the Polychrome Hall and La Hoya Hall, the latter situated at approximately 200 m from the cave entrance. A stable interior microclimate within the cave—characterized by low air circulation, high relative humidity, and limited thermal variability—creates highly sensitive environmental conditions that directly influence the preservation of both geological structures and Paleolithic parietal art [39].

The study area corresponds to the vertical rock wall that provides access to the La Hoya Hall, one of the deepest sectors of the cave (Figure 3). "Hoya" (like hoyo, joyo, or juyo) is a linguistic variant used in Cantabria to refer to a depression, cavity, or hole. It therefore indicates its position at a lower level than the rest of the cave. Even though it is the deepest gallery in the system, it lies only 14 m from the outer hillside (Figures 3 and 4). This wall is composed of stratified limestone exhibiting pronounced verticality and complex fracture patterns, including open joints, exfoliation surfaces, and detached blocks. The morphology of this sector promotes the accumulation of unconsolidated sediment on natural ledges, while its inaccessibility and confined geometry complicate direct observation and conventional surveying methods.

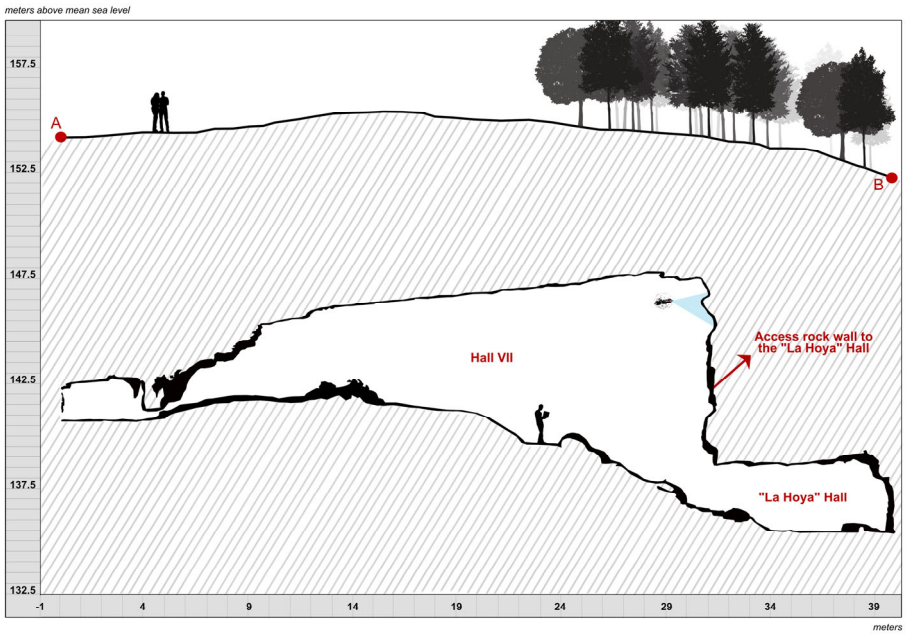

Figure 3. Correlation of the A–B cross-section (Figure 4) with the schematic representation of the terrain section of Hall VII and the rock-wall access to La Hoya Hall, generated through the integration of cartographic data, the second PNOA-LiDAR coverage (CNIG–IGN), and the LiDAR-SLAM point cloud collected by the confined-space UAV.

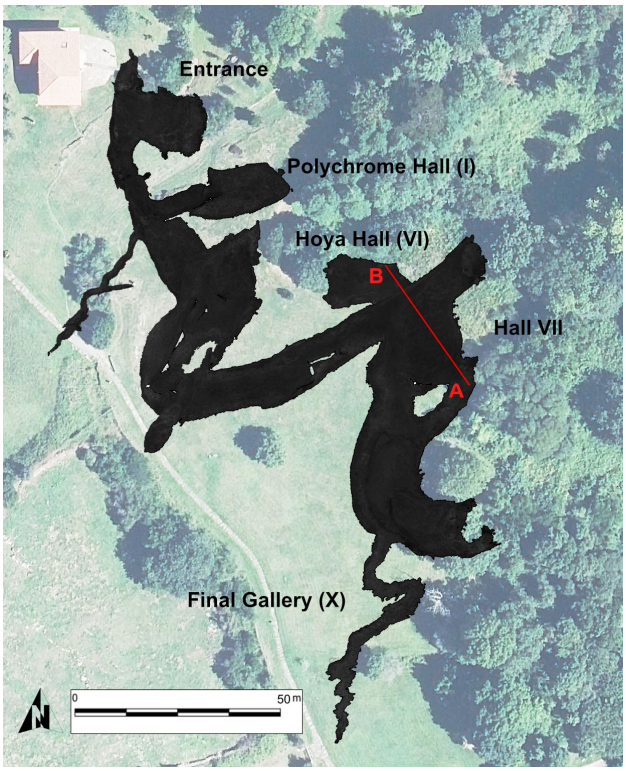

Figure 4. Location of the study area on the PNOA cartographic base and marked A–B section of the area corresponding to Hall VII and the wall above the access to the La Hoya Hall.

La Hoya Hall is classified as a critical zone within the conservation framework due to its geological instability and restricted accessibility. The combination of structural fragility, active cracking, and microenvironmental sensitivity necessitates non-invasive, high-resolution documentation and monitoring strategies. These conditions make the site particularly suitable for confined-space UAV operations integrating LiDAR SLAM and videogrammetric techniques, enabling detailed analysis of rock-wall stability and contributing to broader preventive conservation efforts within the cave.

2.2. Geological and Historical Background

The Cave of Altamira is located at the upper sector of the Santillana del Mar karst system, at an elevation of 159 m above sea level, and developed within Cenomanian–Turonian (Late Cretaceous) geological units. These units consist of well-stratified limestone and calcarenite beds of metric thickness, separated by thin marl–clay layers. This lithostratigraphic arrangement exerts a primary control on the morphology of the cave, particularly on the geometry of chambers and galleries and on the configuration of the ceilings, which commonly form flat slabs affected by hydroplastic deformation. In addition to stratification, the structural discontinuities of the rock mass constitute a second fundamental factor governing the development of the cavity. Fractures and joints define preferential water circulation pathways and play a key role in the mechanical behaviour and long-term stability of the galleries. Together, stratification and discontinuity networks determine both the internal morphology of the cave and its susceptibility to gravitational processes. Altamira is situated within the upper sector of a Pliocene karst system characterized by tabular limestone structures and fracture planes with a marked inclination. Under these geological conditions, the evolution of the cavity is dominated by gravitational collapse processes rather than by chemical dissolution. As a result, the karst system exhibits a progressive tendency toward structural degradation, representing a terminal geological state in which destructive processes prevail over sedimentation. Historical documentation since

the discovery of the cave confirms the persistent dynamics of this instability, with recorded collapse events dating back to at least 1924 and 1935 and continuing into the present. In response to this critical structural condition, corrective stabilization measures were implemented during the 1940s and 1950s. These interventions included the construction of artificial retaining walls to support ceiling strata threatened by collapse, the containment of potentially detachable blocks, and the improvement of visitor safety. At the same time, cracks and fractures were sealed using hydraulic mortar for superficial fissures and cement injections for deeper discontinuities. While these measures were effective in stabilizing the cavity from a structural standpoint, they also introduced significant microenvironmental modifications. These included alterations to ventilation patterns and drainage pathways, the isolation of previously interconnected sectors, disturbances to the microclimatic equilibrium, and the emergence of new water infiltration problems. Consequently, the present-day structural behaviour of the cave reflects the combined influence of its geological framework, long-term gravitational dynamics, and historical human interventions.

2.3. Evidence of Instability and Previous Monitoring at La Hoya Hall

To assess the condition of the rock mass hosting the cave galleries, a geological risk study was conducted in 2014 using geomechanical monitoring stations [12]. This study characterized the fracturing patterns and the state of discontinuities in La Hoya Hall, identifying a high overall level of instability and a particularly high to very high risk in the access zone (Figure 5). The rock mass exhibits a very high degree of fracturing, with fracture apertures exceeding 1 cm [40], low to medium persistence, visible water circulation along fracture planes, and clay-filled fissures. These discontinuities show a high degree of interconnection, generating decimetric blocks above the visitor pathway and several cantilevered blocks. According to the applied assessment system, this area was classified within the maximum level of point-type risk, requiring systematic monitoring of potential block movements. In response to this documented instability, two digital crackmeters with continuous recording and micron-level resolution were installed to monitor potential displacements. The sensors were mounted on metal supports fixed to the stairs leading to the chamber, preventing the transmission of vibrations to unstable blocks during installation (Figure 5). Each device incorporates an onboard data logger, enabling monthly data downloads with minimal intervention in the cave environment. Monitoring results revealed a close relationship between rainfall events and block displacement [41]. When accumulated rainfall exceeds approximately 250 mm, measurable movements are recorded after a delay of about 24 h, corresponding to the time required for water to percolate through discontinuities and reach clay-filled fractures. When rainfall remains below this threshold, the response is delayed or no significant displacement is detected. Periods of sustained heavy rainfall over several consecutive days result in continuous movements that cease only after prolonged dry intervals. Additional investigations aimed at characterizing the internal structural configuration of La Hoya Hall have been carried out using Ground Penetrating Radar (GPR) [39]. These surveys identified a high concentration of vertical and sub-vertical discontinuities, which can be grouped into three main categories: (i) deep developmental structures exceeding 3.80 m in depth, establishing direct hydraulic connections between the exokarst and endokarst; (ii) an interconnected system of joints forming a complex fracture network; and (iii) zones corresponding to planes of structural weakness associated with potential or incipient detachment processes. Taken together, these studies indicate that La Hoya Hall is affected by active geological processes of gravitational, hydrogeological and weathering origin. The documented instability, combined with ongoing deformation and water-driven dynamics, underpins the current management strategy based on preventive conservation principles, including continuous non-invasive monitoring, strict environmen-

tal control, and systematic high-resolution documentation aimed at the early detection of structural changes and volumetric evolution within the cave.

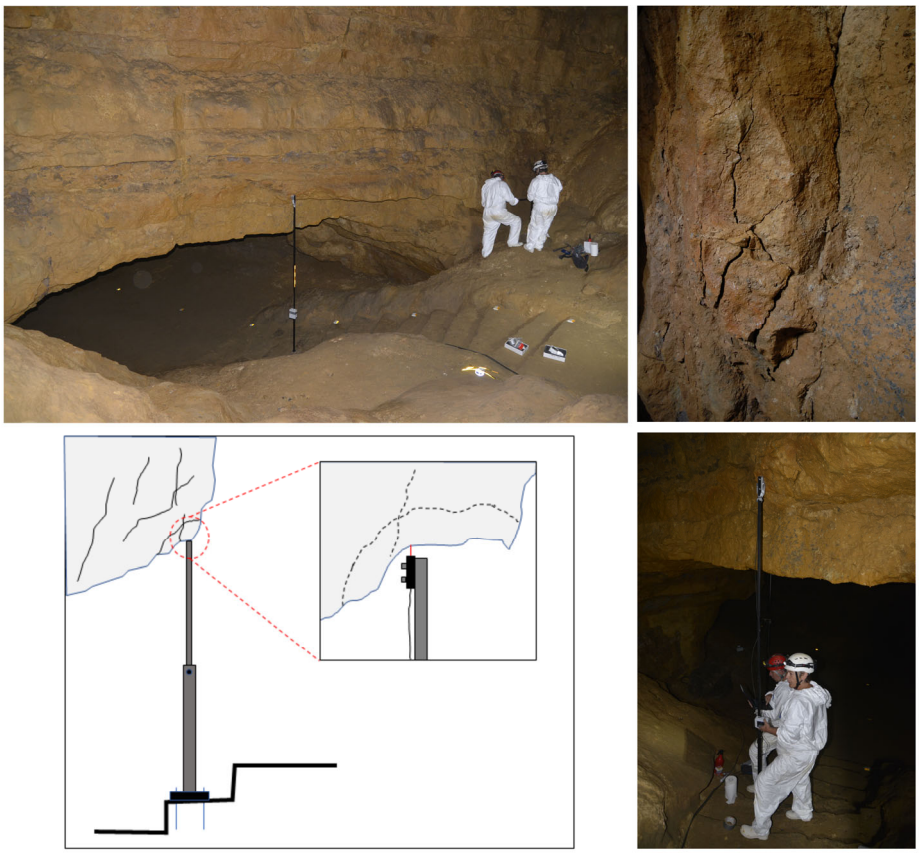

Figure 5. Top left: General view of the access area to the La Hoya Hall. **Top right**: detail of the decimeter-sized blocks cantilevered over the access. **Bottom left**: diagram showing the installation system. **Bottom right**: Image of the support containing the sensor at the top resting on the unstable blocks.

2.4. Description of the Study Wall and Site-Specific Constraints

La Hoya Hall was incorporated into the tourist route during the 1950s and 1960s, when access was facilitated by the construction of a stairway to overcome the 6 m vertical drop into the chamber (Figures 3 and 4). At the entrance, clear evidence of artificial modification is visible, as controlled blasting was used to regularize the opening, leaving characteristic fractures and blast marks on the surrounding rock surfaces. Above the lintel of the access to La Hoya Hall rises a nearly vertical rock wall approximately 12 m high. The application of this methodology allowed us to identify several previously unknown charcoal remains located on this lintel. These remains, some point-shaped and others linear in arrangement, of varying lengths, are currently under study and may correspond to cave passages and even Paleolithic graphic activity. This wall corresponds structurally to the overlying sector of the cave, within area VII, known as the "Great Hall", one of the largest chambers of the Altamira system. The wall documented in this study is a vertically oriented rock surface approximately 14 m wide, reaching a maximum height of 11.74 m, with a frontal extent of 7.96 m. Its morphology is strongly conditioned by lithostratigraphic alternation, resulting in an irregular surface with numerous projections and overhangs of variable dimensions, in some cases extending between 30 and 40 cm. These protrusions favour the accumulation of unconsolidated sedimentary material derived from erosion processes and previous rockfall events, constituting a significant risk factor due to the potential mobilization of these materials. In addition to its complex morphology, the wall

exhibits a high degree of fracturing, particularly in the sector corresponding to the direct access to La Hoya Hall. Although the main fracture affecting this area has previously been subjected to structural inspection, the micro-geological condition of the overlying zones remains largely unknown. It is likely that fractured and potentially unstable blocks are distributed across the entire vertical surface, but their identification through direct visual observation is severely limited. The substantial height of the wall, its vertical orientation, and the presence of the access infrastructure at its base impose significant constraints on conventional documentation techniques. The morphology and spatial configuration of the area hinder the application of standard terrestrial recording methods, such as digital-camera photogrammetry or terrestrial laser scanning, preventing comprehensive coverage and accurate assessment of the rock mass. These site-specific constraints highlight the need for alternative, non-contact documentation strategies capable of capturing high-resolution geometric and visual information in inaccessible and potentially hazardous conditions. These constraints directly condition the methodological approach adopted in this study and are explicitly addressed in the following section.

2.5. Methodological Challenges and Contributions

Based on the geological and structural setting of the Cave of Altamira (Section 2.2), the evidence of active instability at La Hoya Hall (Section 2.3), and the characteristics of the rock wall above its access (Section 2.4), this study is conditioned by a set of interrelated methodological challenges that directly shaped the proposed workflow. The first challenge is the acquisition of reliable and geometrically consistent 3D data in a confined subterranean environment characterized by restricted accessibility, safety constraints, and complex geometry. The considerable height and near-vertical orientation of the studied wall, together with its irregular morphology, overhangs, and local sediment accumulations, limit stable viewpoints and hinder the use of conventional terrestrial techniques. These difficulties are compounded by the absence of GNSS signals and by the need to minimize physical interaction and equipment footprint in a strictly protected heritage context. A second challenge, closely linked to conservation requirements, concerns the practical limitation of time and operational procedures inside the cave to reduce potential impacts on the microclimate and on fragile rock-art environments; this constrains acquisition duration, the possibility of repeated measurements, the use of illumination, and the overall completeness of coverage. A third challenge relates to the detection and characterization of fractures on a highly heterogeneous rock surface under severe visual constraints. The high degree of fracturing documented in the study area, combined with variable illumination, complex textures, low-contrast discontinuities, and partial occlusions, makes direct visual inspection and standard image-based approaches insufficient for systematic appraisal; identifying potentially unstable blocks and crack networks therefore requires methods capable of operating under visual noise and incomplete observations while preserving traceability and geometric consistency. The methodological contribution of this work lies in addressing these constraints through an integrated, non-contact workflow that combines confined-space UAV deployment, LiDAR–SLAM-based geometric reconstruction, close-range videogrammetry, and exploratory deep learning–based crack segmentation. The fusion of LiDAR-derived geometry with high-resolution visual information enables comprehensive documentation of inaccessible surfaces in GNSS-denied conditions, while the use of instance-based segmentation is introduced as a methodological test to evaluate the feasibility of semi-automated crack mapping under extreme subterranean imaging conditions in support of expert-led assessment.

Beyond acquisition and analysis, the workflow incorporates a digital-twin-oriented data organization strategy to facilitate access, traceability, and the systematic registra-

tion of baseline datasets for future multitemporal updates. Accordingly, this work provides a baseline documentation and feasibility assessment rather than a fully operational monitoring system: SLAM-derived products may be affected by trajectory-dependent drift and therefore should not be interpreted as millimetric global accuracy. The crack-segmentation results are presented as a proof-of-concept under limited site-specific training data and domain shift, intended to support expert-led interpretation and guide future model refinement.

2.6. Materials

Due to the spatial constraints and the specific characteristics of the wall (Section 2.5), the documentation work was carried out using a drone specifically designed to operate in confined environments. The system simultaneously records 4K RGB video and an onboard LiDAR–SLAM point cloud; the imagery is time-synchronized with the estimated pose, allowing frames to be spatially referenced within the SLAM reconstruction for inspection and subsequent videogrammetric processing. The platform used was a Flyability Elios 3 (Figure 6), equipped with a 4K Ultra HD RGB camera (3840 × 2160 px at 30 fps), a thermal camera, and an onboard LiDAR–SLAM module. The RGB camera is mounted on a support allowing up to 180° rotation around the X-axis (Table 1). It should be noted that the primary purpose of the camera is not to acquire individual photographs, but rather to continuously record video during the flight, enabling real-time documentation of the entire surface being surveyed.

Figure 6. Data acquisition and inspection workflow using the confined-space UAV, integrating LiDAR–SLAM navigation with simultaneous video capture for videogrammetry.

Its operational features also include the ability to regulate the four front LED lights with which it is equipped, reaching up to 16,000 lumens. In addition, the drone allows the alternate activation of the lateral lights, enabling the creation of different lighting configurations to enhance micro-reliefs and surface details on the inspected rock face.

The system is complemented by an Ouster OS0-32 LiDAR sensor with 32 channels (Ouster Inc., San Francisco, CA, USA) [42] which employs SLAM technology for mobile mapping (Figure 7). This sensor enables the generation of a three-dimensional point cloud during flight, with a ranging precision that varies between approximately ±0.8 cm at short distances and ±4 cm at longer ranges, depending on target distance and reflectivity (Figure 8). The Elios 3 platform was used to acquire LiDAR–SLAM data in confined underground conditions. It is important to distinguish between the onboard 3D Live Model, generated in real time for navigation, coverage verification and mission control, and the post-processed point cloud, which is intended for measurement-oriented analysis and downstream products (mesh generation, digital twin integration, etc.). In SLAM-based mapping, accuracy is primarily affected by drift, i.e., the cumulative error that builds up

along the trajectory, and therefore should be interpreted in terms of global accuracy rather than as a fixed absolute value independent of travel distance and loop-closure conditions. For each flight, the system-reported *Mean LiDAR point validity (%)* was recorded as a quality indicator representing the average proportion of LiDAR returns classified as valid (i.e., retained after internal filtering of noise/outliers) during onboard processing [42]. The device is particularly useful, as it allows the precise spatial location of each video frame to be registered within the point cloud generated during the inspection. Another notable feature is that it is an indoor drone equipped with anti-collision proximity sensors and a protective cage made of highly resistant materials such as carbon fibre. This cage prevents accidental impacts from compromising the stability of the drone or damaging critical components such as the propellers. The design, inherited from its original functional use in industrial inspection, is fundamental when operating in caves for heritage documentation. Its protective structure ensures safe operation for both personnel and the cave environment, making it exceptionally well suited for this type of application. During the inspection flights, proximity sensing supported a nominal stand-off distance of 30 cm from the rock surface, increasing operational safety and helping to maintain a more stable image acquisition geometry, which benefited the subsequent photogrammetric workflow.

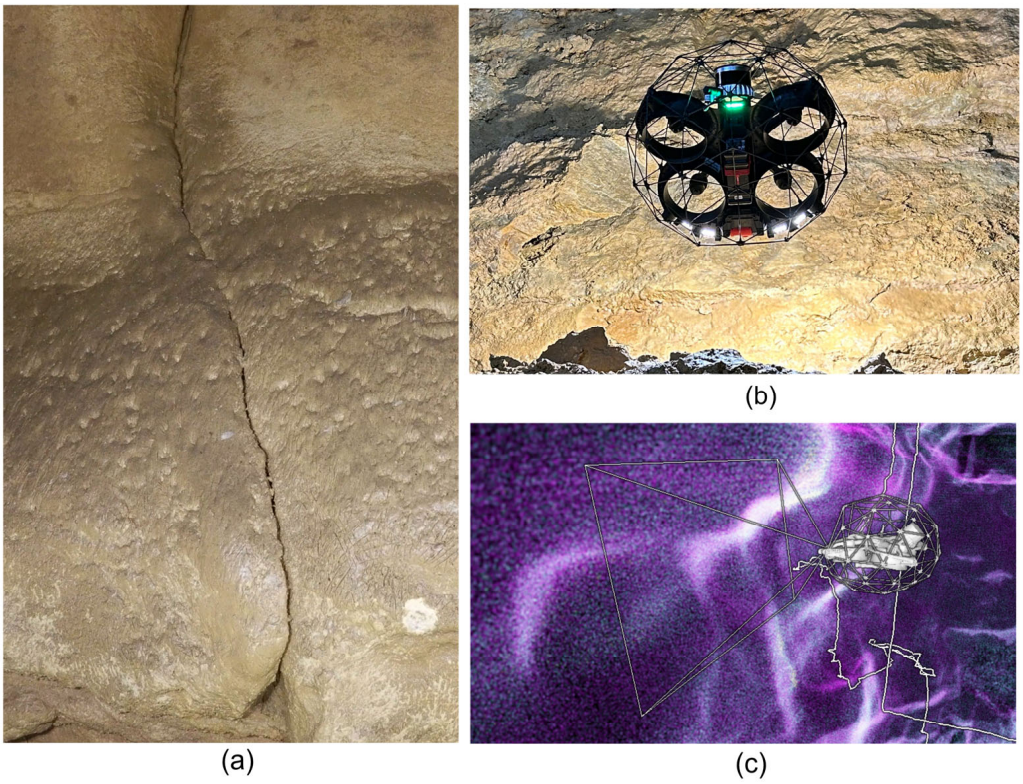

Figure 7. Confined-space UAV inspection with real-time, SLAM-based self-localization over the LiDAR point cloud. (**a**) Crack captured by the UAV RGB camera, illustrating a point of interest (POI) detected during the inspection and metrically assessed in real time. (**b**) Elios 3 UAV positioned above the crack shown in (**a**). (**c**) UAV pose and viewing frustum displayed in Flyability's point-cloud viewer, indicating the RGB camera field of view (FOV) over the point cloud.

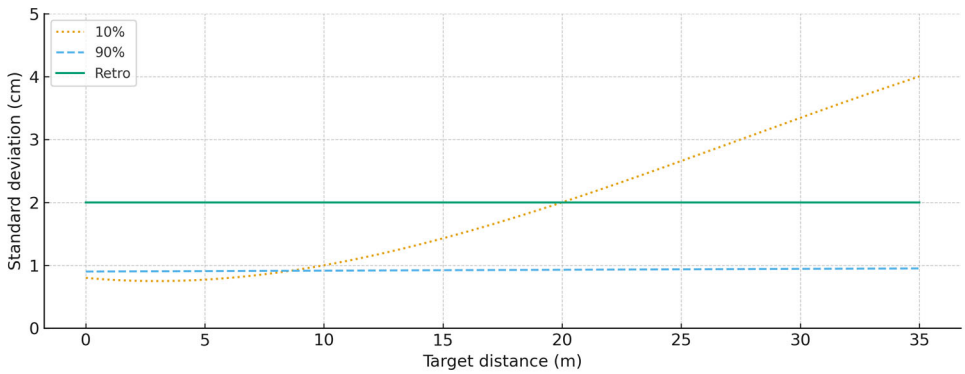

Figure 8. Range measurement precision (standard deviation, 1σ) of the Ouster OS0 Ultra-Wide View High-Resolution Imaging LiDAR as a function of target distance. Precision is computed as the standard deviation of 100 repeated range measurements on a static target, shown for Lambertian targets with 10% and 90% reflectivity and for a retroreflective target ("Retro"). Across the plotted range (0–35 m), typical 1σ precision spans approximately 0.8–4.0 cm. Adapted from [42].

Table 1. Characteristics of Elios confined-space drone and sensors.

	Elios 3
Manufacturer	Flyability (SA, Paudex, Switzerland)
Weight (g)	Approx. 1900 g includes battery, payload and protection
Max. payload (g)	2350 g
Power source	4350 mAh LiPo
Endurance (min)	9–12 min
Camera	2.71 mm focal length. Fixed focal
Thermal Camera	Sensor Lepton 3.5 FLIR
LiDAR Sensors	Ouster OS0-32 beams sensor [1]
Flight control sensors	IMU, magnetometer, barometer, LiDAR, 3 computer vision cameras and a ToF distance sensor

[1] See Figure 8 for detailed specifications.

To ensure adequate working conditions and, in particular, to guarantee optimal illumination during video capture in each flight, it was necessary to install auxiliary lighting points strategically distributed throughout the area. These spotlights were placed along the entire frontal zone facing the rock surface under study. In total, four light sources were positioned, each oriented directly towards the area to be documented. Their purpose was to uniformly illuminate the surface without oversaturating it, functioning exclusively as ambient lighting and significantly improving visibility conditions during the flights.

2.7. Methodology and Workflow Overview

The proposed methodology was designed to provide a detailed geometric and visual documentation of the access wall to the La Hoya Hall, establishing a baseline for monitoring fracture evolution and identifying potentially unstable blocks (Figure 9). Another key objective is the detailed observation and characterization of the material accumulated on the ledges formed along the rock surface. The aim is to identify each fragment, record its size, and analyze how it is arranged on its respective ledge, in order to better understand the erosion and accumulation processes affecting the wall. The inspection of the fracture network constitutes an additional fundamental aspect of the study, with priority given to achieving an accurate representation of fracture continuity and aperture, particularly those developing parallel to the rock surface. These fractures are of particular concern, as their distribution across the entire wall may intersect with other discontinuities, leading to the formation of detached rock blocks. In the context of a vertical wall, such blocks could be

left almost cantilevered, suspended over the access to the La Hoya area, representing a potential safety concern.

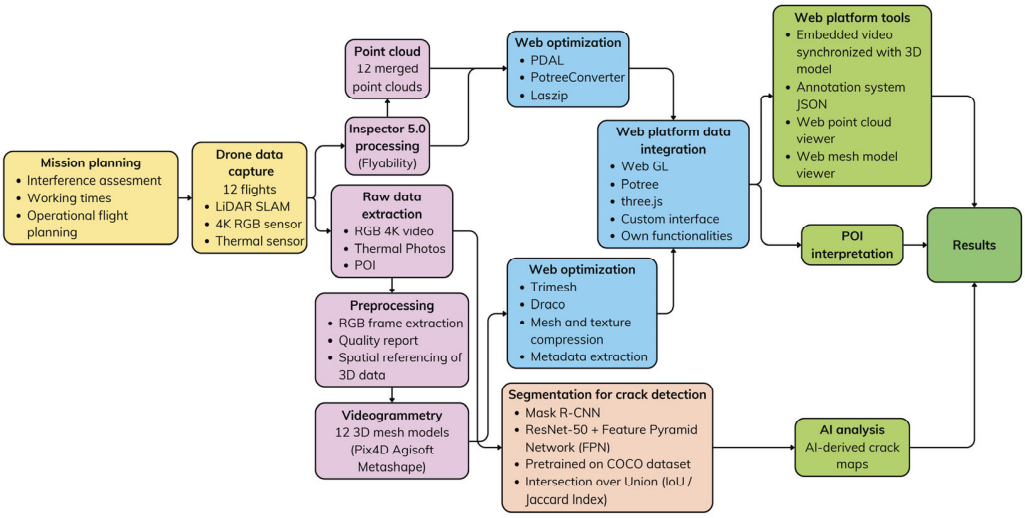

Figure 9. End-to-end workflow for confined-space UAV inspection and digital-twin delivery. Yellow blocks: mission planning and data capture (flights; LiDAR–SLAM and 4K RGB video). Purple blocks: processing and 3D reconstruction (Inspector processing, video/POI extraction, preprocessing, videogrammetry, and point-cloud merging). Blue blocks: web optimization and platform integration (e.g., Potree/three.js; synchronized video, visualization and annotation). Beige block: AI-based crack segmentation (Mask R-CNN) producing crack maps. Green blocks: POI interpretation and final outputs.

These objectives required high-definition image acquisition and integration into a 3D environment, which underpins the technical workflow developed in this study.

Although several documentation campaigns have been carried out to generate high-resolution orthophotos (with ground sampling distances down to 2 mm/pixel), the natural configuration of the rock surface and its considerable height limit the reliability of orthophoto generation (Figure 10). Orthophotos produced in previous campaigns are not optimal for this case, as they do not provide the viewing geometry or effective resolution necessary to reliably identify fissures or sediment accumulations on the ledges. This is largely due to distortions and aberrations introduced during orthophoto generation when they rely on automatic correlations derived from photogrammetric 3D models captured from lower vantage points. This acquisition strategy was developed within the DiGHER project, which provides a digital-twin web environment to integrate and explore multi-scale datasets (LiDAR, videogrammetry, photogrammetry, historical imagery, IoT sensors, etc.) for long-term monitoring and preventive conservation. Within this environment, 3D point clouds, meshes, time-stamped image sequences and semantic annotations can be jointly explored, compared over time and enriched by domain experts. The platform is conceived as a tool for long-term monitoring and preventive conservation—facilitating the early detection of structural or environmental changes—as well as for research and public dissemination through web-based visualization interfaces. Its design follows the FAIR (Findable, Accessible, Interoperable and Reusable) [43] data principles, so that datasets and derived products can be systematically stored, shared and reused. In this context, the data acquired during the monitoring of access to La Hoya Hall in the Cave of Altamira have been deployed within the DiGHER platform, so that researchers at the *Museo Nacional y Centro de Investigación de Altamira* can easily access them and use them in monitoring and research initiatives focusing on the cave itself.

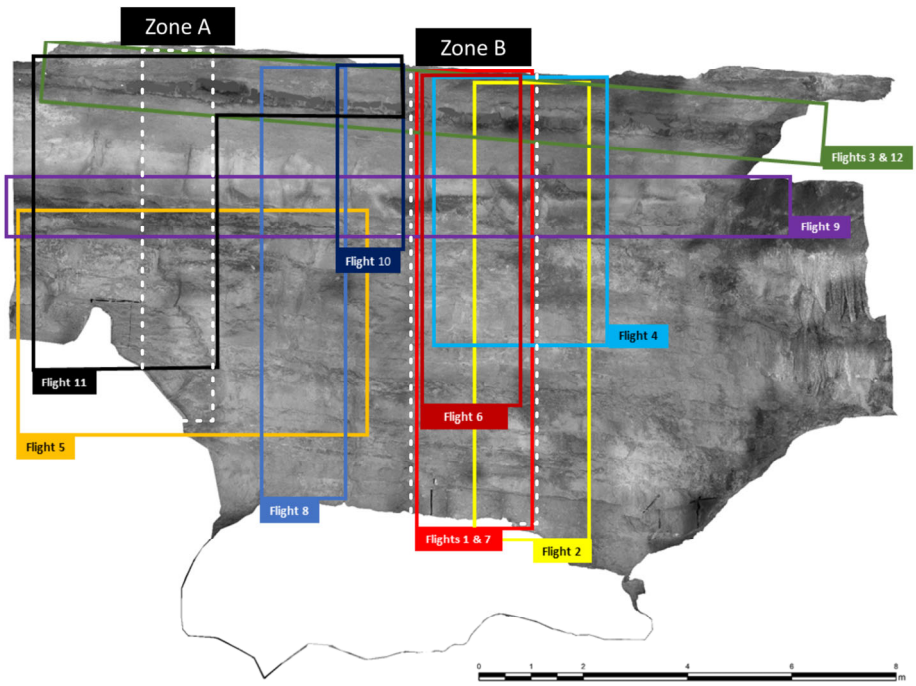

Figure 10. Orthophoto of the rock surface highlighting the sections inspected in each flight. Based on the topographic survey conducted by the *Museo Nacional y Centro de Investigación de Altamira*.

2.7.1. Crack Segmentation Workflow (Mask R-CNN)

Within this unified digital twin environment, the extracted image sequences served as the input dataset for the Deep Learning–based crack detection model. The segmentation workflow was based on the standard Mask R-CNN architecture [25] (Figure 11), consisting of a Backbone ResNet50 + FPN that extracts multi-scale feature maps through lateral 1×1 convolutions and top-down fusion. These feature maps feed into the Region Proposal Network (RPN), a fully convolutional module that predicts candidate object coordinates and objectness scores. RoI Align then accurately samples features for each proposal, preserving spatial alignment before forwarding them into two parallel heads: a classification branch with FC layers that outputs the object category and refined bounding-box coordinates, and a Fully Convolutional Network (FCN) branch that generates high-resolution instance masks. Together, these three branches (RPN proposals, category prediction, and pixel-level mask segmentation) enable end-to-end learning of detection and segmentation across objects of different scales.

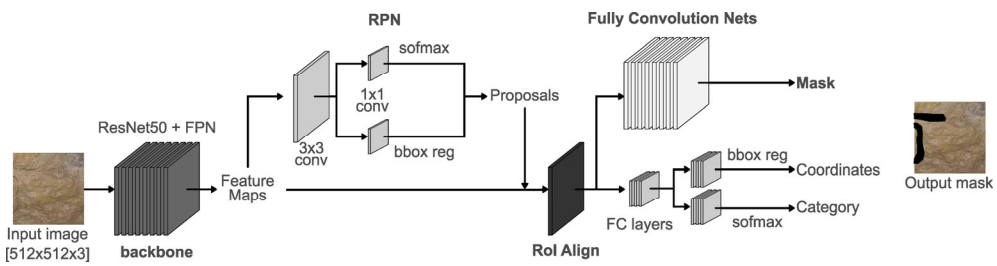

Figure 11. Mask R-CNN architecture.

Training deep learning models typically requires large, high-quality annotated datasets. In this study, we adopted a transfer-learning strategy by combining a photo-interpreted crack dataset with pre-training on the COCO dataset [44]. Transfer learning is particularly effective for specialized tasks such as crack detection, where annotated data are often limited. By initializing the Mask R-CNN model with weights learned from COCO, the network

benefits from general visual features (such as edges, textures, and geometric structures) thereby reducing computational cost and training time while improving convergence and performance [45].

The COCO dataset contains over 300,000 images across 80 object categories. Our photo-interpreted dataset consists of 1070 image tiles of size 512 × 512 pixels, each containing at least one annotated crack (Figure 12). The training set was constructed from tiles extracted from frames acquired during the first UAV flight (587 images), whereas the testing set comprises tiles from the fifth flight (483 images). This split helps ensure that the evaluation includes cracks captured under varying illumination conditions and different UAV–wall distances. To further increase data variability and mitigate overfitting, we applied a simple data augmentation strategy by horizontally flipping 50% of the training images.

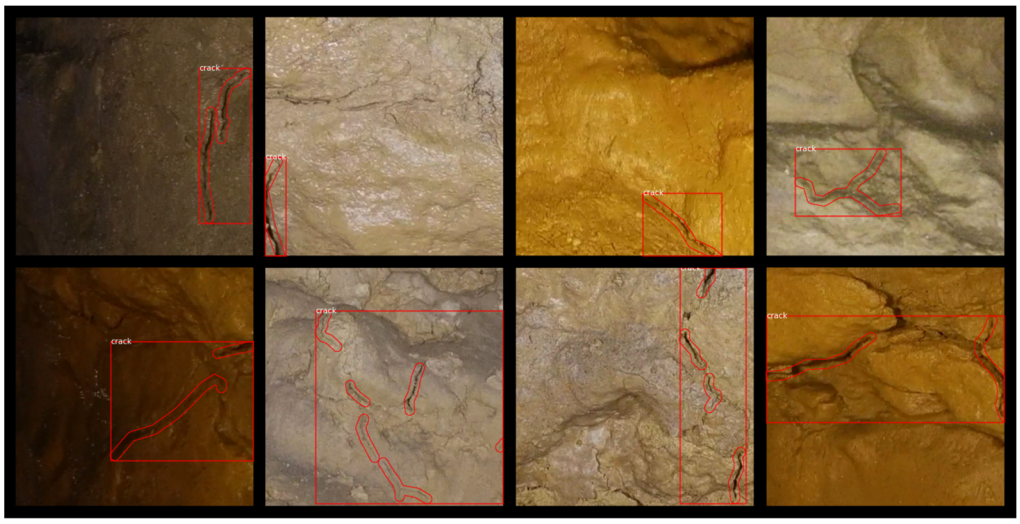

Figure 12. Examples of crack masks generated under varying illumination and viewing conditions in frames captured during the first flight.

Model performance was evaluated using the Intersection over Union (IoU) metric, which quantifies the overlap between the predicted segmentation mask and the ground-truth annotation. IoU (also known as the Jaccard Index) is defined as the ratio between the area of intersection and the area of union of the two masks. This metric calculates the area of the intersection between the prediction ρ and the ground-truth label l, and divides it by the area of their union (Equation (1)).

$$\mathrm{IoU}(\rho, l) = \frac{\mathrm{Area}(\rho \cap l)}{\mathrm{Area}(\rho \cup l)} \tag{1}$$

Higher IoU values indicate more accurate localization and delineation of cracks. This metric enables a consistent assessment of segmentation quality across varying illumination and flight conditions, providing a reliable indicator of the ability of the model to generalize. The IoU metric is rooted in the classical Jaccard similarity coefficient [46]. Model performance was assessed using mean Average Precision at 50% Intersection-over-Union (mAP@IoU = 50), a standard benchmark in instance segmentation. This metric computes the area under the precision–recall curve after determining whether a predicted mask sufficiently overlaps with a ground-truth annotation, where IoU $\geq$ 0.50 is considered a correct detection. By averaging precision scores across all classes and confidence thresholds, mAP@IoU = 50 quantifies both localization accuracy and segmentation quality, providing a standard summary of detection performance.

2.7.2. Mask R-CNN in Crack Detection: Capabilities and Limitations

Across heritage, archaeological, and engineering applications, Mask R-CNN has demonstrated high segmentation accuracy and robustness in the presence of complex visual clutter, outperforming or matching alternative approaches such as U-Net, YOLO-based detectors, and conventional feature-based methods [47,48]. In cultural heritage conservation, the model has been used to map deterioration patterns and structural pathologies with high precision, facilitating large-scale and automated condition assessments [32]. Similarly, in structural monitoring of concrete, metal, and composite materials, Mask R-CNN has been shown to improve crack detection rates and reduce false positives, particularly when integrated with domain-specific preprocessing and augmentation strategies [28,30].

Despite its advantages, several limitations must be acknowledged. The two-stage architecture is computationally intensive and requires substantial training data, often demanding extensive manual annotation and domain adaptation to achieve optimal performance [29,47]. Moreover, model effectiveness can be sensitive to hyperparameter choices, including learning rate schedules and augmentation pipelines, and thus requires careful tuning and validation [30].

Taken together, existing research indicates that Mask R-CNN provides a powerful and adaptable framework for automated crack detection, delivering the precision required for complex and delicate cultural heritage contexts, such as subterranean or rock-art sites. These characteristics make it a promising candidate for application to Paleolithic cave surfaces, where accurate, high-resolution mapping of crack networks is essential for documentation, conservation, and long-term monitoring.

2.7.3. Automated Point Cloud Comparison and Structural Assessment Algorithm

To quantify the structural evolution of the cave, the DiGHER platform incorporates a custom Cloud-to-Cloud (C2C) comparison module. A primary challenge in processing high-density LiDAR data is the prohibitive computational cost of brute-force Euclidean distance calculations (O(N^2)). To mitigate this bottleneck and keep processing feasible on standard server infrastructure, the reference point cloud is indexed using a KD-tree (k-dimensional tree). This spatial structure optimizes nearest-neighbour searches to an average complexity of O(log N) per query, allowing the overall runtime to scale quasilinearly (O(N log N)) rather than quadratically.

Furthermore, to accommodate the memory constraints of web-based deployment (see Section 3.3), the algorithm applies batch (chunk) processing. Instead of loading the full dataset into memory simultaneously—which may compromise system stability—the comparison iterates through the query cloud in bounded segments, maintaining stable memory usage and ensuring scalability as point densities increase. From an implementation perspective, the comparison module is built on an open-source Python (v3.10.12) ecosystem designed for scientific reproducibility and seamless integration with the DiGHER backend. The workflow relies on Laspy for the efficient ingestion and writing of standard ASPRS LAS files, ensuring the preservation of original radiometric and spectral metadata throughout the process. The core spatial optimization is powered by SciPy's cKDTree, which provides a high-performance C++ backend essential for executing the nearest-neighbour queries at scale, while NumPy enables the vectorized computation of Euclidean distances and displacement components without the overhead of explicit loops. Furthermore, the automated generation of analytical deliverables is orchestrated through Matplotlib (v3.10.3), used for rendering statistical distributions and vector plots, and ReportLab, which compiles these assets into the standardized PDF reports. This reliance on standard, community-validated scientific libraries not only ensures computational reliability but also facilitates future code maintenance and technology transfer.

Finally, the module addresses the interpretability gap often found in raw geomatics outputs by automatically generating standardized deliverables, including false-colour scalar-field point clouds (absolute displacement magnitude) and displacement vector visualizations (directionality). These visual products are coupled with auto-generated PDF reports containing statistical summaries (e.g., histograms and metrics) and processing metadata, improving consistency and traceability across monitoring epochs and supporting data-driven conservation decisions for the Cave of Altamira.

3. Results

This section presents the main outputs derived from the proposed workflow, focusing on the geometric, visual, and analytical results obtained from the UAV-based survey and subsequent processing.

3.1. Spatial Coverage, Geometric Completeness, and POI-Based Inspection

Two inspection sectors were defined on the target rock face (Figure 10) and are hereafter referred to as Zone A and Zone B. Zone A was the primary target in Flights 5, 8, 9 and 11, whereas Zone B was covered in Flights 1–4, 6, 7, 9, 10 and 12. The two sectors are not mutually exclusive: several flight paths partially overlapped both zones (e.g., Flights 3, 9 and 12), as trajectories were adapted in situ to follow curatorial guidance and to maximize coverage of the inspected rock face.

Zone A, located on the left portion of the rock surface and centred on a lower protruding area, covered approximately 1.5 × 7.0 m (10.5 m^2). Zone B, positioned in the central sector of the same rock face, encompassed a larger area of 10.0 × 2.0 m (20.0 m^2) to ensure complete documentation of the central panel. In total, 12 UAV flights were completed, with a mean duration of 6 min 30 s, producing ~80 min of recorded 4K video (Table 2). Although Zones A and B constituted the primary targets, the acquired video sequences also covered most of the rock face, as flight trajectories were continuously adjusted in situ following the instructions of the technical staff and curators of the *Museo Nacional y Centro de Investigación de Altamira.*

Table 2. Summary of flight duration and video-based frame extraction parameters used for videogrammetric reconstruction. For each flight, we report the number of points of interest (POIs), the frame sampling rate (FPS), the total number of extracted frames (image size in pixels), and the mean reprojection error obtained during photogrammetric processing.

Flight	Duration (min:s)	POIs (n)	FPS (frames/s)	Extracted Frames (px)	Mean Reprojection Error (px) [1]
1	5:20	12	3	796 (3840 × 2160)	0.21 (Pix4D)
2	6:21	7	3	807 (3840 × 2160)	1.32 (Metashape)
3	6:49	5	3	738 (3840 × 2160)	2.80 (Metashape)
4	8:00	5	3	921 (3840 × 2160)	0.21 (Pix4D)
5	6:43	8	3	920 (3840 × 2160)	1.23 (Metashape)
6	5:15	11	3	726 (3840 × 2160)	1.46 (Metashape)
7	6:22	13	2	765 (3840 × 2160)	1.34 (Metashape)
8	7:08	12	2	858 (3840 × 2160)	1.51 (Metashape)
9	7:21	11	2	885 (3840 × 2160)	0.21 (Pix4D)
10	6:35	3	2	649 (3840 × 2160)	0.21 (Pix4D)
11	7:35	3	2	807 (3840 × 2160)	1.49 (Metashape)
12	7:06	6	2	763 (3840 × 2160)	0.22 (Pix4D)

[1] Mean reprojection error reported by the photogrammetric software used for each flight (Pix4D 4.8.4 or Agisoft Metashape 2.1).

For videogrammetric processing, frames were sampled at 2–3 fps, yielding 649–921 extracted frames per flight (3840 × 2160 px) and flight-wise mean reprojection errors ranging from 0.21 to 2.80 px depending on the software pipeline (Pix4Dmapper v4.8.4 or Ag-

isoft Metashape v2.1; Table 2). In parallel, the integrated LiDAR–SLAM sensor generated 12 flight-wise point clouds with ~13.9–19.9 million retained points per flight, Mean LiDAR point validity values of 34.81–81.36% (system quality indicator), and mapped surface areas reported in Table 3. The exported 3D meshes (GLB) show flight-dependent complexity, with equivalent mean mesh edge lengths spanning 1.41–3.53 mm (Table 3), reflecting variations in stand-off distance, viewpoint, and the extent of non-clipped contextual geometry captured during each flight.

Table 3. Flight-wise LiDAR–SLAM and mesh summary. For each flight, we report the total number of LiDAR points retained after preprocessing, the Mean LiDAR point validity as a system quality indicator, the mapped surface area, and the resulting 3D mesh complexity (number of triangles in the exported GLB) together with the equivalent mean mesh edge length (mm).

Flight	Total Points (LiDAR)	Mean LiDAR Point Validity (%) [1]	Surface Area (m^2)	Number of Triangles (3D GLB)	Equivalent Mean Edge Length (mm)
1	14,455,526	81.36	30.238	12,170,988	2.40
2	17,098,313	46.95	11.426	10,837,834	1.56
3	17,306,464	34.81	6.498	3,625,552	2.03
4	19,442,588	53.22	23.384	4,825,395	3.35
5	17,301,812	73.66	8.247	9,507,768	1.41
6	13,879,239	67.41	22.208	9,321,681	2.35
7	17,767,492	70.95	13.440	4,701,385	2.60
8	19,139,884	65.31	21.900	6,570,000	2.40
9	18,831,330	43.15	18.826	4,847,764	2.99
10	17,057,076	50.95	23.997	4,877,966	3.37
11	19,868,249	35.04	16.129	6,198,498	2.45
12	16,673,922	40.16	15.514	4,939,250	2.69

[1] Mean LiDAR point validity (%) is a system quality indicator reporting the proportion of LiDAR returns classified as valid (non-noise/outlier) by the onboard processing.

Geometric completeness and interpretability were further assessed through synchronized exploration of all outputs in Inspector 5, which allows simultaneous visualization of the 4K video together with the spatial position (pose) of the corresponding frames within the LiDAR point cloud. This frame-to-geometry linkage supports the identification of occlusions (e.g., behind ledges/overhangs) and helps interpret local gaps in sampling in relation to viewpoint and illumination conditions.

A key outcome of this integrated inspection is the systematic registration of Points of Interest (POIs), recorded both in the video timeline and in the spatial model. Across the 12 flights, a total of 96 POIs were annotated (Table 2) to index and revisit relevant anomalies (e.g., cracks, material displacements, sediment accumulations, potentially unstable elements, and anthropogenic marks) within a single, spatially coherent reference. Inspector's POI-linked measurement tools further enable rapid metric checks on selected frames and point-cloud subsets. While such in-software measurements do not reach the precision of high-resolution close-range photography or terrestrial laser scanning, their value lies in providing actionable, spatially referenced observations that can be replicated in future inspection campaigns, supporting early risk screening and the planning of preventive conservation actions.

3.2. Videogrammetry Results: Flight-Wise Mesh Models

From the UAV video dataset, extracted frames were generated at 2–3 fps, yielding flight-specific frame sequences used for videogrammetric reconstruction. This workflow produced 12 textured mesh models (one per flight), providing a baseline 3D record of the rock surface above the access to La Hoya Hall for subsequent comparison and monitoring.

The resulting meshes translate close-range video observations into spatially coherent 3D surfaces, supporting interpretation of fine-scale features and documentation of areas that are difficult to inspect directly. First, video segments with stable camera motion and favourable illumination were retained, while segments affected by abrupt motion, motion blur, or LED glare were excluded. Frames were then extracted at 2–3 fps (depending on flight duration and speed), yielding a more uniform set of extracted frames across flights (Table 2). This sampling was selected to maintain sufficient inter-frame overlap while limiting the overall data volume. The extracted frames underwent basic preprocessing, including brightness and contrast adjustment, white balance correction, and—when necessary—removal of redundant frames or frames containing people or out-of-context artefacts. Each flight-specific dataset was then processed independently in Agisoft Metashape v2.1 and Pix4Dmapper v4.8.4 following a standardized Structure-from-Motion (SfM) and Multi-View Stereo (MVS) workflow. For each flight, processing produced (i) a dense point cloud containing approximately 50–100 million points and (ii) a mesh with ~10 million faces, on average. Textured surface meshes were generated by triangulation suitable for complex geometries, with the aim of preserving micro-relief features related to cracks and the overall geomorphology of the rock wall; textures were derived from the original extracted frames. Two photogrammetric pipelines were used to compare outputs and assess the sensitivity of reconstruction to the implemented SfM/MVS workflows. To enhance cross-model geometric consistency, metric references extracted from the LiDAR–SLAM point cloud and clearly identifiable structural features (edges, ledges, and major fractures) were employed as internal validation cues. The resulting checks support the quantitative assessment presented above and motivate the limitations addressed in the next section. Heterogeneous illumination and surface reflectance produced both overexposed and underexposed areas, which propagated into texture inconsistencies in the reconstructed meshes. Finally, the videogrammetric meshes were finely registered to the integrated LiDAR point cloud to mitigate local deviations associated with imaging constraints, leveraging the geometric stability of the LiDAR reference. The resulting integrated models combine LiDAR-based geometry with high-detail textures, supporting detection and measurement of discontinuities, overhanging blocks, and changes in sediment accumulation.

3.3. LiDAR–SLAM Point Clouds, Integrated Meshes, and Geometric Consistency Checks

Although the derived meshes provide very high geometric detail (i.e., dense triangulation and small equivalent edge length) (Table 3), resolution should not be conflated with accuracy. In LiDAR–SLAM mapping, global accuracy depends on drift accumulation along the travelled path and on the effectiveness of loop closures; therefore, accuracy is best described using global error metrics.

Flyability reports centimetre-level global-accuracy RMSE values when comparing Elios 3 SLAM point clouds to a TLS control in a controlled test environment (e.g., ~18.3 cm RMSE for FlyAware processing vs. ~3.5 cm RMSE for FARO Connect), highlighting that different processing pipelines and trajectory length can significantly influence global accuracy [49]. Therefore, in this study, the system-provided *Mean LiDAR point validity* is interpreted as a data-quality indicator (proportion of retained valid returns) rather than a direct measure of metric accuracy, while quantitative accuracy assessment should rely on independent checks (e.g., control measurements, C2C and C2M comparisons). The videogrammetric reconstruction was complemented with geometric data acquired by the onboard LiDAR–SLAM sensor, resulting in 12 flight-wise point clouds and an integrated 3D representation of the inspected volume. After basic cleaning to remove isolated points and acquisition artefacts, the point clouds were retained without clipping to the immediate wall vicinity, allowing the surveyed sector to be contextualized within Hall VII of the Cave of

Altamira. Because all LiDAR–SLAM outputs were produced in a common local coordinate system, the individual flight point clouds could be directly integrated into a coherent combined model of the inspected volume (Figure 10). The resulting dataset provides the geometric reference for subsequent analyses, including crack mapping and future multi-temporal comparisons, and it was deployed in the DiGHER platform to support interactive inspection and expert annotation by the conservation team. To evaluate the geometric consistency of the derived 3D products and to detect potential discrepancies between data sources, several checks were performed focusing on (i) the quality and internal coherence of the LiDAR–SLAM point clouds in Zones A and B and (ii) the agreement between the UAV point clouds and the videogrammetry-derived meshes. These verifications help identify local deviations related to occlusions, variations in stand-off distance, illumination conditions, and potential SLAM trajectory drift effects, and they provide a baseline for future multi-temporal comparison in support of preventive conservation. To this end, different inspection and geometric comparison procedures were applied using the Analysis module of Cyclone 3DR (2024.0.2.45638) and, complementarily, using a tool integrated in the DiGHER platform, based on the Open3D library. In Cyclone 3DR, Cloud-to-Cloud (C2C) comparisons were carried out between selected flight point clouds and between the integrated model and each individual cloud, in order to characterize spatial differences and identify isolated points or systematic discrepancies. Likewise, Cloud-to-Mesh (C2M) comparisons were performed between the LiDAR–SLAM geometry (considered as the primary geometric reference frame) and the videogrammetric meshes, with the aim of evaluating the consistency of the photogrammetric reconstructions and detecting local deformations or biases. In parallel, the DiGHER tool enabled these checks to be reproduced in a traceable manner within a collaborative working environment, through comparison routines implemented on top of Open3D (e.g., point-cloud distance computation). This facilitated interactive inspection by the technical team and the linkage of geometric discrepancies to visual evidence (frames/POIs). Overall, these checks provide a quality-control framework oriented toward interpretation and monitoring, and they prepare the dataset for more exhaustive future evaluations (e.g., zone-wise aggregated metrics, sector-based analyses, and multitemporal comparison). Complementary to the overlap-zone C2C checkpoint check between Flights 1 and 7 (Figure 13), we evaluated the consistency of the videogrammetry-derived surface reconstructions by performing a Mesh-to-Mesh (M2M) checkpoint comparison in Cyclone 3DR (Analysis module) (Figure 14; Table 4). This additional control helps localize potential discrepancies attributable to occlusions, variable stand-off distance, illumination-driven texture instability, and residual misregistration between the two flight-wise videogrammetric meshes.

Building on the local checkpoint-based checks, we further assessed the repeatability of the UAV-based LiDAR–SLAM mapping—without external GNSS corrections—through a systematic Cloud-to-Cloud (C2C) comparison using the automated workflow described in Section 2.7.3. Distances were computed as Euclidean nearest-neighbour point-to-point distances between a reference flight and repeated acquisitions within two main inspection sectors (Zone A and Zone B) (Figure 13). A conservative search radius of 5.0 m (max_distance) was adopted to prevent premature rejection of correspondences in areas affected by occlusions and incomplete overlap, ensuring that both fine-scale deviations and gross outliers could be detected. To avoid over-interpreting extreme values driven by non-overlapping regions or sparse geometry, we report not only the mean distance and standard deviation, but also robust indicators (median and upper percentiles, e.g., P95) and an inlier ratio within practical tolerances (e.g., $\leq$0.05 m and $\leq$0.10 m), which better reflect the dominant geometric consistency relevant for preventive conservation workflows. For Zone A, Flight 8 was selected as the reference geometry due to its optimal

coverage of the lower protruding sector. Comparisons were computed against Flights 5, 9, and 11. The results indicate a high degree of geometric coherence, with mean absolute distances ranging from 3.8 cm to 5.1 cm. The comparison between Flight 8 and Flight 11 showed the highest consistency, with a mean deviation of 3.79 cm. Vector analysis suggests that discrepancies are not systematic shifts, but rather distributed noise consistent with SLAM drift in feature-poor environments, with mean displacement vectors along X, Y, and Z remaining below 2 cm in most cases. In Zone B, encompassing the wider central panel, Flight 1 served as the reference for comparisons against Flights 4, 6, 7, 9, 10, and 12. The results highlight the variability of SLAM positioning over larger scan areas. Flight 7 exhibited the closest agreement with the reference, achieving a mean distance of 2.25 cm and a standard deviation of 6.25 cm, demonstrating the system's capability to achieve high repeatability under favourable conditions. Other comparisons in this zone generally remained within the 5–7 cm range for mean distance (e.g., Flight 12 vs. Flight 1: 5.9 cm mean). Overall, the zone-wise C2C results indicate centimetric repeatability across repeated LiDAR–SLAM flights, with mean/median distances typically in the ~2–7 cm range under favourable overlap and viewpoint conditions (Figure 13). Although this level of consistency does not match the millimetric precision achievable with static terrestrial laser scanning (TLS) under controlled surveying conditions, it provides a reliable geometric baseline for rapid, non-contact documentation and for identifying conservation-relevant changes at the scale of preventive conservation (e.g., decimetric block detachment or major sediment displacement). In this sense, the proposed UAV workflow is best interpreted as a baseline documentation and feasibility framework that prepares the dataset for more exhaustive future evaluations (e.g., denser control, zone-wise aggregation, and multitemporal comparisons) rather than as a fully operational monitoring system.

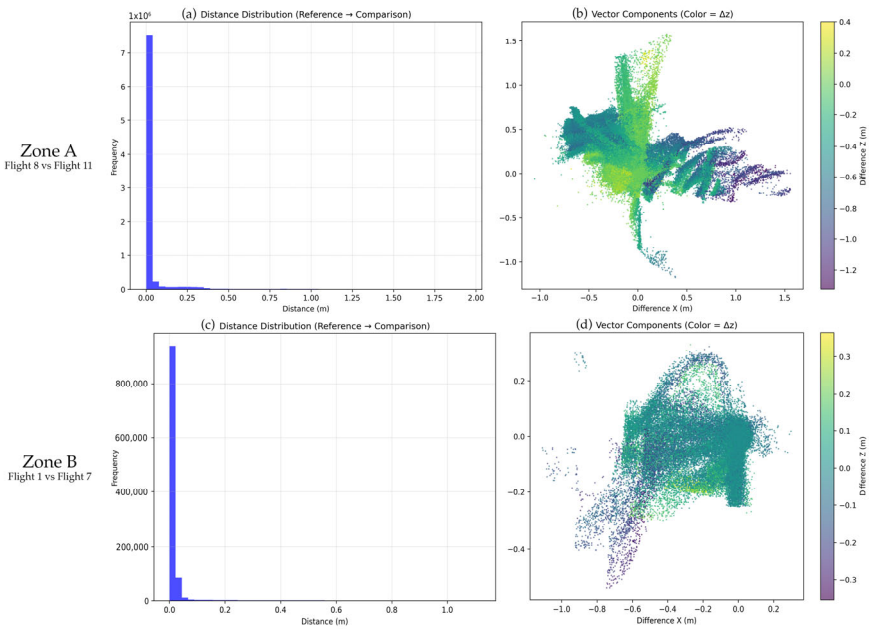

Figure 13. Statistical analysis of geometric repeatability for the representative cloud-to-cloud (C2C) comparisons described in the text. (**a**,**b**) Analysis of Zone A (Flight 8 vs. Flight 11). (**c**,**d**) Analysis of Zone B (Flight 1 vs. Flight 7). The histograms (**a**,**c**) illustrate the frequency distribution of Euclidean distances, showing a dominant alignment tendency in the sub-5 cm range (blue bars) despite the presence of outliers. The scatter plots (**b**,**d**) display the displacement vector components (Δx, Δy) coloured by vertical deviation (Δz). The dispersed nature of these point clouds supports the interpretation of discrepancies as non-systematic SLAM drift rather than rigid transformation errors. Note that while the search radius was 5.0 m, the plots focus on the relevant distribution range.

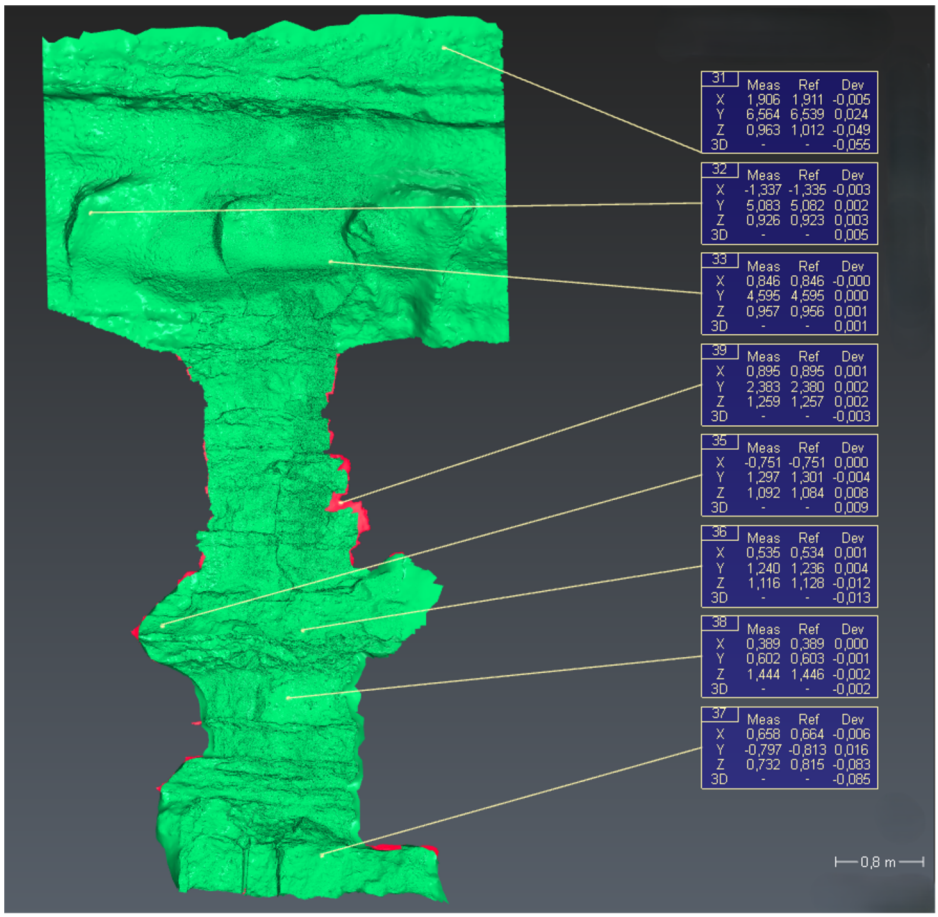

Figure 14. Mesh-to-Mesh (M2M) checkpoint comparison between the videogrammetry-derived mesh from Flight 7 (Meas) and the videogrammetry-derived mesh from Flight 1 (Ref), performed in the Analysis module of Cyclone 3DR (the non-overlapping surface is shown in red). View of the compared meshes and checkpoint locations (checkpoints #30–#39). (Signed Dev 3D values measured at checkpoints (8); absolute values (Dev 3D) are used to interpret deviation magnitudes.

Finally, an additional geometric-consistency check was performed by comparing the Flight 1 textured mesh (generated by videogrammetry) with a set of checkpoints sampled on the Flight 7 LiDAR–SLAM point cloud within the overlap area. This independent, point-based control evaluates to what extent the higher visual detail provided by videogrammetry (texture fidelity and apparent micro-relief) translates into metric agreement with the SLAM-derived geometry. In practice, the observed discrepancies are mainly attributable to (i) SLAM drift accumulation along the trajectory (and its sensitivity to loop-closure performance and the effective path) and (ii) local point-cloud noise in sectors affected by complex geometry, variations in stand-off distance, and less stable returns. Across the selected checkpoints, the mean 3D deviation was 6.15 cm, providing a practical estimate of repeatability at the local scale. Consequently, the improved visual "resolution" of the videogrammetric model does not necessarily imply a corresponding increase in geometric accuracy with respect to the LiDAR–SLAM representation.

Table 4. Mesh-to-Mesh (M2M) comparison at checkpoints between the videogrammetry-derived mesh from Flight 7 (Meas) and the videogrammetry-derived mesh from Flight 1 (Ref) (Cyclone 3DR, Analysis module). For each checkpoint, the 3D coordinates measured on the inspected mesh (Meas) are compared with those on the reference mesh (Ref). Dev X, Dev Y, and Dev Z report the signed coordinate residuals (Meas − Ref), and Dev 3D is the resulting 3D deviation at each checkpoint.

Checkpoint	Meas X (m)	Meas Y (m)	Meas Z (m)	Ref X (m)	Ref Y (m)	Ref Z (m)	Dev X (m)	Dev Y (m)	Dev Z (m)	Dev 3D (m)
Label #31	1.906	6.564	0.963	1.911	6.539	1.012	−0.005	0.024	−0.049	−0.055
Label #32	−1.337	5.083	0.926	−1.335	5.082	0.923	−0.003	0.002	0.003	0.005
Label #33	0.846	4.595	0.957	0.846	4.595	0.956	−0.000	0.000	0.001	0.001
Label #35	−0.751	1.297	1.092	−0.751	1.301	1.084	0.000	−0.004	0.008	0.009
Label #36	0.535	1.240	1.116	0.534	1.236	1.128	0.001	0.004	−0.012	−0.013
Label #37	0.658	−0.797	0.732	0.664	−0.813	0.815	−0.006	0.016	−0.083	−0.085
Label #38	0.389	0.602	1.444	0.389	0.603	1.446	0.000	−0.001	−0.002	−0.002
Label #39	0.895	2.383	1.259	0.895	2.380	1.257	0.001	0.002	0.002	−0.003

3.4. Exploratory AI-Based Crack Detection Results Under Cave Conditions

This subsection presents the results of an exploratory application of deep learning–based instance segmentation for crack detection under the extreme visual constraints of the cave environment. The proposed model achieved a mAP@IoU = 50 of 20% when validated against the flight 5 dataset. While the network successfully identifies the presence and general location of cave-wall cracks, this relatively low score reflects systematic over-segmentation: ground-truth annotations represent each digitized crack as a single continuous mask, whereas the model frequently predicts multiple smaller subcracks instead of the complete structure. These fragmented detections reduce IoU with the annotated region, lowering precision and thereby decreasing the overall mAP value, despite visually plausible crack localization.

Large cracks are generally well predicted, with the model capturing most of their spatial extent and overall morphology (Figure 15b). These major fractures present substantial depth, producing characteristic shadows that make them visually distinctive and easier for the network to detect. However, the model also generates some false positives by interpreting unrelated shadow patterns or lighting variations as cracks, as illustrated in Figure 15a. Furthermore, due to the low resolution of certain image tiles and the high density of fine fissures, many small cracks were not included during manual annotation. In these cases, the model occasionally identifies subtle subcracks overlooked by the expert, revealing both annotation limitations and the sensitivity of the model to minor structural discontinuities (Figure 15).

The POI frames processed by the model indicate that it can delineate substantial portions of the structural cracks on the inspected rock wall (Figure 16, Table 5). Because the network was trained using polygon masks, predictions often include a narrow buffer around the crack trace. Although smaller cracks (e.g., POI #2) can be detected, their continuity is only partially recovered and some segments remain undetected. Under this POI-based assessment, shadows and specular reflections did not systematically degrade the predictions, and no widespread increase in illumination-driven false positives was observed, although occasional shadow-related confusions may occur (e.g., Figure 15a).

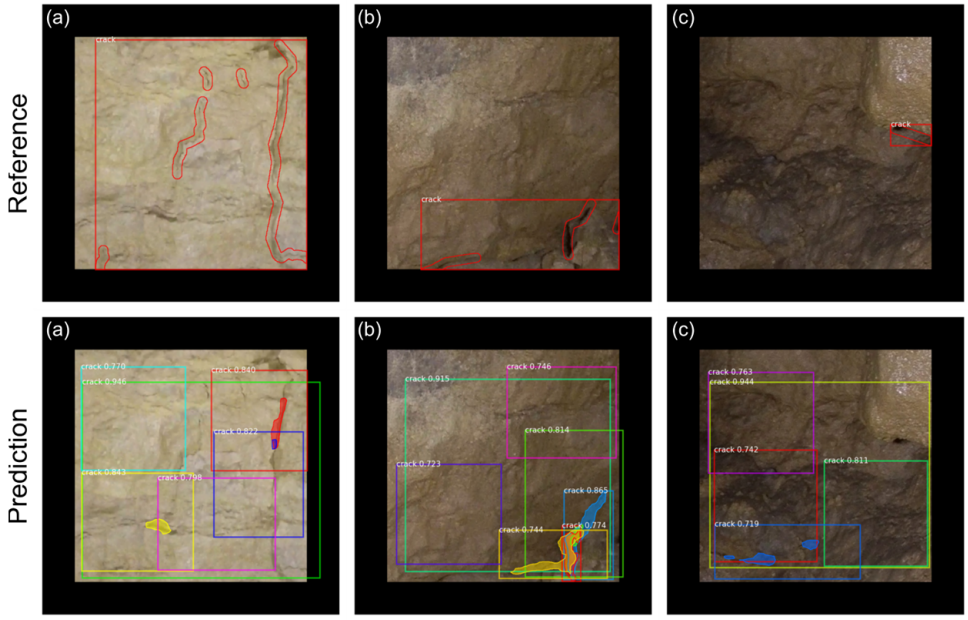

Figure 15. Qualitative comparison between ground-truth and predicted crack segmentations. The top row displays the manually annotated reference masks from the Flight 5 dataset, while the bottom row shows the corresponding predictions generated by the proposed model. Column (**a**) illustrates a shadow-induced false positive, where the network misinterprets illumination artefacts as cracks. Column (**b**) presents a correctly segmented large crack, demonstrating accurate localization and shape reconstruction. Column (**c**) shows a predicted crack that was not included in the expert annotations, highlighting the ability of the model to detect subtle fissures overlooked during manual labelling.

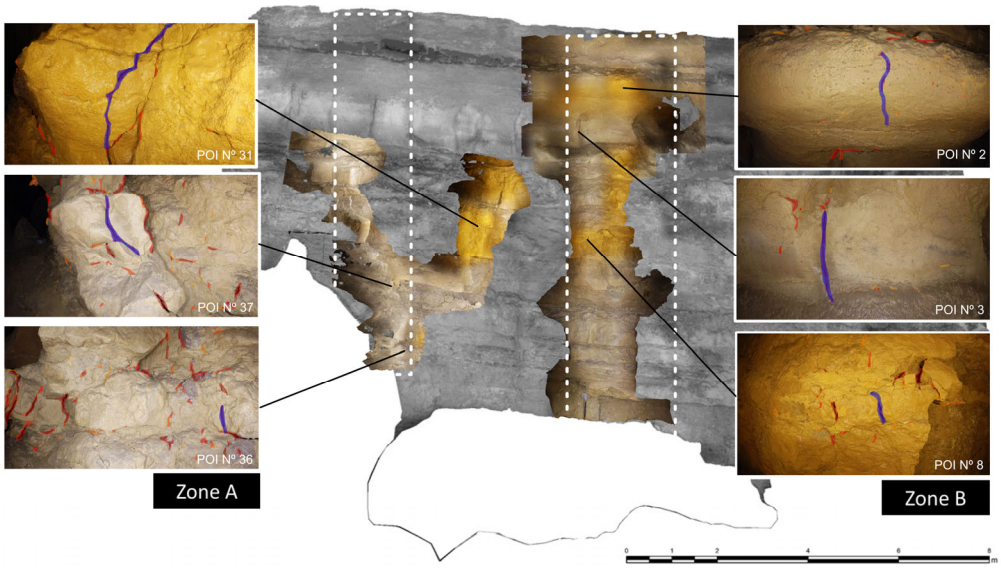

Figure 16. Selection of POIs showing crack measurements and their spatial localization on the cave wall within the 3D model of the documented zones. Segmentation model–predicted crack masks are colour-coded by predicted crack probability (red: higher probability; orange: lower probability). Blue overlays indicate the crack portions selected for length measurement (measured segments drawn on top of the predicted masks).

Table 5. Statistics aggregating crack probability for each of the masks composing the analyzed crack. The trained Mask R-CNN model can generate multiple masks to represent a single crack, with each mask associated with a probability value. The reported statistics are computed by aggregating all crack segments associated with each measured crack.

POI	Measurement (mm) [1]	Margin of Error (mm)	Mean Prob.	Q1 (25th)	Q2 (Median)	Max Prob.
2	264	±9	0.77	0.76	0.79	0.91
3	599	±15	0.81	0.78	0.82	0.92
8	101	±5	0.84	0.78	0.85	0.96
31	479	±17	0.82	0.77	0.83	0.90
36	260	±7	0.85	0.84	0.88	0.93
37	110	±11	0.82	0.76	0.83	0.92

[1] Crack lengths were measured on POI frames in Inspector 5 using the "Add Measurement" tool, which uses the distance-measurement sensor to perform pixel calibration and triangulation.

Overall, the obtained performance defines a feasibility baseline rather than an operational automated solution, highlighting both the potential and the current limitations of AI-based crack detection in subterranean heritage contexts. From a structural health monitoring perspective, crack length/extent provides a necessary baseline; however, safety assessment (e.g., defining 'safe' vs. 'critical') typically requires structural engineering interpretation and temporal indicators, such as crack propagation rates derived from repeated observations. Accordingly, the reported measurements should be interpreted as a feasibility baseline for future longitudinal monitoring, rather than a definitive safety classification.

4. Discussion

4.1. Analysis of the AI Model for Automated Crack Detection

A key component of this study is the integration of a deep learning approach for automated crack detection on high-resolution UAV imagery acquired within a confined subterranean environment. The performance of the Mask R-CNN model demonstrated clear potential but also highlighted important limitations primarily related to data quality and annotation completeness.

Reported performance metrics of crack segmentation models in the literature, including U-Net- and YOLO-based approaches, cannot be directly compared with the results obtained in this study, as they rely on different datasets, annotation strategies, imaging geometries and illumination conditions [36]. In many benchmark scenarios, data are acquired under controlled settings with homogeneous surfaces and extensive ground-truth annotations, which contrasts sharply with the constraints of a protected subterranean heritage environment [50]. Within this context, the relatively low mAP achieved in the present study reflects the combined impact of limited training data, highly variable illumination and complex rock textures, rather than deficiencies of the selected architecture alone [38]. Despite the low IoU values contributing to this mAP score, the model is nevertheless able to detect a large number of cracks and their approximate spatial extent. The reduced IoU mainly results from missed reference annotations and fragmented predictions, which penalize overlap-based metrics, even when the detected regions are visually plausible. At the same time, the model shows a limited tendency to misclassify strong shadows or illumination variations as cracks, indicating a degree of robustness to lighting-related artefacts. These observations reinforce the exploratory nature of the proposed approach and highlight the need for domain-specific datasets and tailored learning strategies to improve performance in future developments.

As shown in the results, the model performance is strongly influenced by the quality and completeness of the annotated dataset. Because many fine-scale cracks were not

digitized during manual labelling, the network lacks sufficient examples to learn their visual characteristics, leading to fragmented predictions and reduced mAP scores. Increasing the number of annotated masks (particularly for thin, low-contrast cracks) would provide a more representative training distribution and help the model better generalize across varying wall textures, lighting conditions, and crack morphologies.

To achieve this, a larger digitization effort is required, ideally incorporating systematic labelling of small fissures that are currently underrepresented. However, manually annotating these structures is time-consuming, subjective, and prone to omission due to low resolution and visual ambiguity. Future work may therefore benefit from semi-automatic or fully automatic labelling strategies (such as weak supervision, active learning, or self-training) to accelerate mask generation and reduce expert workload. Such approaches could expand the training dataset, enhance crack boundary precision, and ultimately improve both detection accuracy and segmentation consistency.

Despite these limitations, the Mask R-CNN architecture remains well suited for this type of analysis due to its ability to jointly perform object detection and pixel-level segmentation. Its region-proposal mechanism allows it to localize cracks of varying shapes and scales, while the parallel mask-prediction branch provides detailed delineation of crack boundaries even under heterogeneous imaging conditions. Moreover, Mask R-CNN is highly modular and can be adapted through improved backbones, feature-pyramid designs, and domain-specific augmentations, making it a robust foundation for future enhancements. With a more comprehensive annotated dataset, the inherent strengths of the model (particularly its capacity to capture fine structural details) could be fully leveraged to achieve substantially higher detection and segmentation performance.

4.2. Integration of Geospatial Data into a Digital Twin Framework: Infrastructure and Hierarchical Model

The architecture of the DiGHER platform was designed to address the challenge of data fragmentation in cultural heritage. While traditional repositories often isolate geometric data from metadata, our approach integrates a relational database (PostgreSQL (v14.19) [51]), with a hierarchical structure (municipalities-collections-items-visualizations). This relational consistency is critical for long-term preservation, ensuring that geometric data remains contextually linked to its heritage significance rather than existing as isolated files.

The choice of a client-server model based on Django (v5.1.1) [52] and a responsive frontend allows for a centralized management workflow, which is essential for multi-user environments like the Altamira research team. However, implementing such a structured system introduces a complexity trade-off compared to simpler, ad hoc file storage solutions. To mitigate this, user and permission management were tightly coupled with the data hierarchy. This ensures that while the system remains scalable and collaborative, it strictly adheres to the data security protocols required for sensitive heritage sites, balancing accessibility with the necessary restrictions on non-public archaeological data.

4.3. Point Cloud Integration and Web-Based Visualization

The integration of large-scale 3D point cloud data represents a significant technical challenge due to the massive volume of information involved. To enable efficient in-browser rendering without the barriers of heavy client-side software, the platform uses a system based on WebGL and JavaScript technologies (Figure 17). The core visualization engine is derived from Potree (v1.8.2) [53], which relies heavily on foundational libraries such as three.js for 3D rendering and proj4.js for geospatial coordinate transformations. This approach was selected specifically for its ability to handle multi-resolution octree structures, which allows for the dynamic streaming of billions of points. Unlike traditional

methods that might require full dataset downloads, this web-optimized strategy ensures that high-density datasets—such as the complex scans of Altamira—remain accessible and responsive, effectively bridging the gap between massive geospatial data and remote web accessibility.

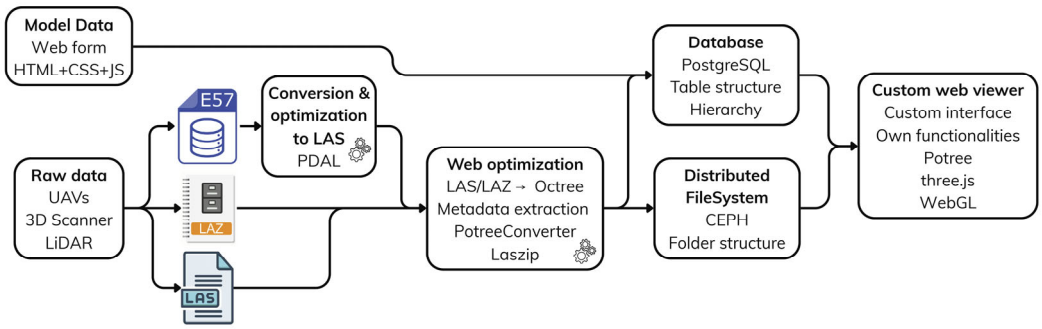

Figure 17. Schematic representation of the point cloud integration workflow in the DiGHER platform. Point cloud files in E57, LAZ, and LAS formats are processed and aligned on the server side and made accessible via an interactive visualization interface supporting measurement tools, semantic enrichment, and collaborative analysis.

From an operational perspective, ensuring data uniformity is as critical as visualization performance. Since heritage projects often generate data in heterogeneous formats like E57, relying on manual conversion by users can lead to inconsistencies and workflow bottlenecks. To address this, an automated backend pipeline using the Point Data Abstraction Library (PDAL (v2.3.0) [54] was established. By automatically standardizing incoming files into the optimized LAS/LAZ format required for further processing, the system significantly reduces the technical burden on researchers. This automation not only streamlines the upload process but also guarantees that all datasets adhere to a consistent internal structure essential for long-term digital preservation. Finally, the delivery mechanism required balancing high-speed data transfer with rigorous access control. Serving massive static assets directly through a web application framework (like Django) would introduce unacceptable latency, yet exposing them via a public web server would compromise the confidentiality of the site. The solution adopted involves a hybrid architecture using NGINX (v1.29.1) [55] as a reverse proxy coupled with an *X-Accel-Redirect* mechanism. This design effectively decouples permission logic from file serving: while Django validates user authority, NGINX handles the heavy lifting of streaming the binary data. This setup ensures that sensitive archaeological content is protected by robust authentication protocols while benefiting from the raw performance of an optimized static file server. Beyond visualization, DiGHER is conceived to support long-term preservation, traceability, and reuse of heterogeneous geomatics outputs by linking each 3D asset to its acquisition and processing context (e.g., mission, sensor, processing version, and spatial reference). This structured registration enables controlled updates of the digital twin while preserving provenance and contextual integrity, and it aligns with FAIR data principles by improving the findability, accessibility (under controlled access), interoperability, and reusability of digital documentation. This is particularly relevant for sensitive subterranean heritage sites where in situ access is constrained and remote expert review can support preventive conservation planning.

4.4. Mesh Integration and Semantic Enrichment

Beyond point clouds, the DiGHER platform also handles meshed 3D models—typically the outputs of photogrammetry or high-density laser scanning that have been processed

into 3D surfaces (meshes) with textures. Integrating these models required evaluating available technologies.

The suite of tools developed by the Smithsonian Institution [56], which are built on top of widely adopted web technologies like three.js and WebGL (Figure 18), provides rich interactivity, high-performance rendering, and extensible annotation functionalities through open-source components, forming a solid basis for building customized 3D heritage viewers. This option presented significant advantages and shaped the direction of the final implementation adopted in the DiGHER platform.

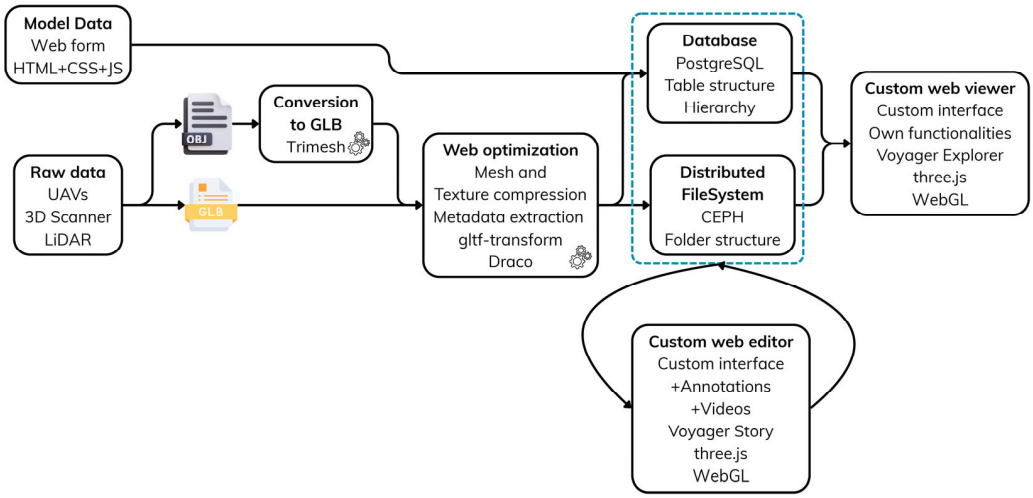

Figure 18. Integration and annotation workflow for meshed models in the DiGHER platform. Imported 3D assets (OBJ and GLB formats) are processed and aligned on the server, after which they are published in an interactive viewer that supports collaborative annotation, metadata structuring, and multi-format interoperability.

The integration of textured 3D meshes into a web environment required addressing the inherent trade-off between visual fidelity and network performance. Raw photogrammetric outputs (often in OBJ format) are typically too heavy for real-time streaming, which necessitates an optimization strategy. To resolve this, the platform implements an automated pipeline that converts assets into the glTF/GLB [57] (open standard for 3D scenes that is efficient for web delivery) using the Trimesh library (v4.6.6) [58]. A crucial architectural decision in this workflow was the application of Draco [59] compression via gltf-transform (v4.3.0) [60]. Although Draco introduces a lossy compression step, it reduces file sizes by up to 95% with negligible visual impact. This optimization is fundamental for the Digital Twin concept; it ensures that the high-resolution videogrammetric models of the Cave of Altamira are not merely static archival files, but fluid, interactive assets accessible even over standard network connections.

Furthermore, beyond mesh visualization (following the same general approach as the point-cloud viewer described above), the platform was designed to bridge the "semantic gap" that often limits the utility of 3D heritage data. Visualization alone allows for observation, but not interpretation. To overcome this, we adapted components from the Smithsonian Voyager suite, creating a customized viewer that prioritizes semantic enrichment over complex configuration. The implementation of a specialized annotation system allows researchers to link spatial coordinates directly to multimedia narratives.

A particularly innovative aspect of this semantic layer is the spatiotemporal linking capability, which connects specific points on the 3D rock wall to precise timestamps in the inspection videos. This feature transforms the 3D model from a passive representation into an index for the raw video data, allowing users to verify the context of a crack or

sediment accumulation instantly. By simplifying the interface for these annotations, the system lowers the barrier to entry for non-technical experts, encouraging the continuous enrichment of the Digital Twin with archaeological and geological knowledge.

4.5. Deployment Strategy and System Scalability

Deploying a complex platform like DiGHER that combines database, backend logic, and specialized 3D processing tools can be challenging (Figure 19). Replicating this mix of components across different environments becomes even more difficult. To mitigate this risk and ensure scientific reproducibility, the DiGHER platform adopts a full containerization strategy using Docker (Engine v24.0+) [61]. Rather than treating the software environment as a secondary concern, this approach encapsulates the entire runtime—including the Ubuntu LTS base, Python dependencies, and compiled binaries—into a single, immutable artefact. This decision moves the platform away from fragile, manual server configurations toward a deterministic deployment model, where the computational environment is as reproducible as the data itself.

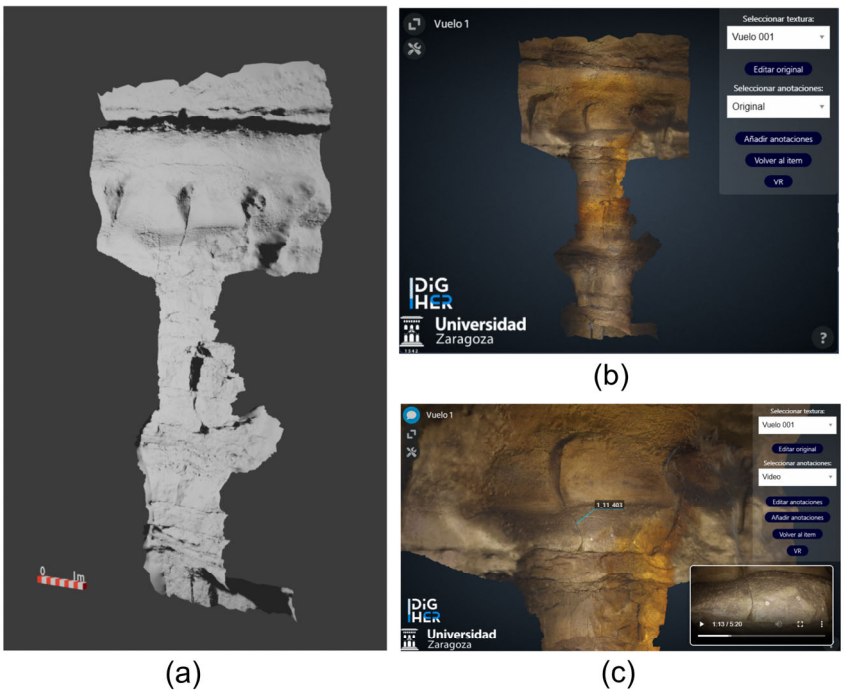

Figure 19. View of the processed platform displaying the meshed GLB models with annotations and synchronized video keyframes over the 3D model. (**a**) Mesh model of flight 1. (**b**) Visualization of the mesh model of flight 1 within the DiGHER interface. (**c**) Detail of the DiGHER mesh model viewer showing the synchronized display of the mesh model with the video captured from flight 1.

From a maintenance and scalability perspective, containerization addresses the long-term preservation challenges specific to digital heritage. Software obsolescence often renders digital archives inaccessible within a few years. By freezing the toolchain (including specific versions of processing libraries like PDAL or gltf-transform) within the container, we ensure that the Digital Twin remains functional and the processing pipeline remains verifiable in the future, regardless of updates to the underlying host operating system. Furthermore, the orchestration of the Django backend behind an NGINX reverse proxy within this containerized environment provides a production-ready architecture that creates a portable solution. This ensures that the platform can be easily replicated across different academic servers or collaborative research environments without requiring specialized

DevOps intervention, allowing researchers to focus on the heritage content rather than on resolving complex infrastructure constraints.

4.6. Overall Interpretation

Taken together, UAV-based LiDAR–SLAM mapping, videogrammetric reconstruction, automated crack segmentation, and digital-twin deployment provide an integrated framework to support structural inspection and monitoring in highly constrained subterranean environments. This workflow mitigates several practical limitations of conventional surveying in confined settings and establishes a coherent basis for repeatable documentation aimed at preventive conservation.

Regarding automated crack mapping, the current Mask R-CNN results—reported through standard instance-segmentation metrics (IoU and mAP@IoU = 0.50)—should be interpreted as a proof-of-concept, as performance is strongly influenced by the limited size and representativeness of the annotated dataset and by the challenging imaging conditions typical of rock-art caves (e.g., heterogeneous illumination and low-contrast discontinuities). Future work should therefore prioritize expanding and diversifying the annotations, improving model robustness, and increasing the temporal depth of the digital twin through repeated inspections. These steps will strengthen multi-temporal comparison and facilitate earlier identification of structural changes, contributing to the long-term safeguarding of the Paleolithic heritage preserved in the Cave of Altamira.

5. Conclusions

The use of a confined-space UAV equipped with LiDAR–SLAM and high-resolution imaging sensors provides a practical solution for documenting surfaces that are difficult or unsafe to access using conventional geomatics approaches under similarly restrictive conditions. Stable flight in complex geometries, together with 4K RGB imaging and onboard LiDAR–SLAM, enabled detailed documentation of the rock wall at the entrance to La Hoya Hall. To our knowledge, this study represents the first documented deployment of a LiDAR–SLAM–equipped confined-space UAV for structural monitoring inside a Paleolithic World Heritage cave. The results indicate that hybrid SLAM–videogrammetric workflows can mitigate the physical, geometric, and safety constraints typical of subterranean heritage settings and provide a repeatable baseline for future monitoring.

The synchronized analysis of videographic records with the LiDAR-derived point cloud supported an initial technical appraisal of the inspected surface, enabling the identification of elements relevant to stability assessment, including fractures, overhanging blocks, and accumulations of unconsolidated sediment. These observations provide a basis for future preventive conservation actions in the Cave of Altamira and are transferable to other confined or high-risk heritage environments requiring comparable monitoring standards.

Furthermore, the identification of possible graphic and transit evidence on this lintel demonstrates the potential application of this methodology for the prospecting, documentation and study of archaeological evidence in places that are difficult to access or whose conservation does not allow movement through them. However, this finding opens the door to its application in the prospecting and documentation of archaeological heritage not only in cave environments, but also in the open air, allowing for the optimization of prospecting work in environments that are difficult to access.

Beyond crack mapping, this hybrid UAV-based workflow can also support non-contact documentation and long-term monitoring of Paleolithic rock art, including both paintings and engravings. With appropriate adaptations in acquisition strategy and illumination control, it could likewise be transferred to open-air post-Paleolithic artistic contexts. High-resolution, spatially referenced imagery integrated within a 3D digital-twin environment

can facilitate repeatable visual inspection of motifs and surface condition (e.g., pigment loss, flaking, micro-detachment, or surface alteration), providing complementary evidence to inform conservation decisions in decorated cave settings.

Future campaigns should consolidate this hybrid methodology by prioritizing repeatability and consistent acquisition geometry across surveys. Where feasible, complementary high-density meshes in accessible areas could enrich both visual inspection and metric analysis. Establishing a periodic documentation schedule would facilitate multi-temporal tracking of fractures, sediment accumulations, and potentially unstable structural elements, strengthening preventive conservation planning.

All generated datasets—including 3D models, point clouds, and synchronized video—were integrated into the University of Zaragoza's DiGHER platform, enabling interactive web-based visualization, spatial synchronization between geometry and audiovisual records, and functions for measurement, collaborative annotation, and temporal comparison. This integration supports traceability, remote expert review, and long-term management of the digital twin as a shared baseline for decision-making. Therefore, this documentation supports long-term digital preservation within a FAIR-aligned geomatics framework and facilitates analysis, remote access, and the planning of future conservation actions by the *Museo Nacional y Centro de Investigación de Altamira* and other stakeholders.

Future work should focus on repeated UAV-based inspections to build multitemporal datasets suitable for detecting structural evolution at the scale required for preventive conservation. In parallel, expanding and refining annotated imagery is expected to improve the robustness of deep-learning crack segmentation under heterogeneous illumination and surface textures. Together, these steps will enhance early-warning capabilities and support more informed, data-driven strategies for the preventive conservation of Paleolithic rock-art caves, as well as for the systematic documentation and artistic analysis of the motifs.

Author Contributions: Conceptualization, J.A., M.B., C.V. and P.F.; methodology, J.A., M.B., C.V., C.I., G.R., M.Á.S.-C. and V.B.; software J.A., C.V., C.I. and G.R.; validation, J.A., M.B., C.V., C.I. and G.R.; formal analysis, J.A., M.B., C.V., C.I. and G.R.; investigation, J.A., M.B., C.V., C.I., G.R., P.F., M.Á.S.-C., V.B., A.P. and L.M.D.-G.; resources, J.A., P.F. and C.d.l.H.; data curation, J.A., M.B., C.V., C.I. and G.R.; writing—original draft preparation, J.A., M.B., C.V., C.I., G.R., M.Á.S.-C. and V.B.; writing—review and editing, J.A., M.B., C.V., C.I., G.R., P.F., C.d.l.H., M.Á.S.-C., V.B., A.P. and L.M.D.-G.; visualization, J.A., M.B., C.V., C.I. and G.R.; supervision, J.A., P.F. and C.d.l.H.; project administration, J.A., P.F. and C.d.l.H.; funding acquisition, J.A., P.F. and C.d.l.H. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the research project *DiGHER* (grant number CPP2022-009631), led by Jorge Angás and funded by the MCIU/AEI/10.13039/501100011033 and by the European Union through the "NextGenerationEU"/PRTR programme at the University of Zaragoza. *Museo Nacional y Centro de Investigación de Altamira* supported the archaeological data acquisition. This work was also benefitted from the collaboration of the research project *Searching for the Origins of Rock Art in Aragón (SEFORA)*, led by Manuel Bea financed by the Proy_H04_24, Government of Aragón. Cristian Iranzo worked on this paper through PhD research contract funded by Department of Science, University and Knowledge Society of the Government of Aragón (Spain). The APC was funded by SeGAP (University of Zaragoza).

Data Availability Statement: The research data supporting this publication are not publicly available. The data were collected by the University of Zaragoza (Spain) as part of the research and conservation studies of the Cave of Altamira. These data are kept in the *Museo Nacional y Centro de Investigación de Altamira* and at the SeGAP of the University of Zaragoza (Spain).

Acknowledgments: The authors express their gratitude to the open-source libraries and tools that made this work possible, all of which were used in full compliance with their respective Creative Commons and open-source licensing agreements. The authors gratefully acknowledge GIM Geomat-

ics for supplying additional topographic plans that supported the topographic verification process. We also thank Flyability for their technological support and the SeGAP of the University of Zaragoza for their collaboration.

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

Abbreviations

The following abbreviations are used in this manuscript:

C2C	Cloud-to-Cloud (point-cloud distance comparison)
C2M	Cloud-to-Mesh (point-cloud to surface/mesh distance comparison)
CLI	Command-Line Interface
CNIG	Centro Nacional de Información Geográfica
COCO	Common Objects in Context dataset
FAIR	Findable, Accessible, Interoperable and Reusable
FCN	Fully Convolutional Network
FPS	Frames per Second
FPN	Feature Pyramid Network
GLB	Binary form of glTF
glTF	GL Transmission Format
IGN	Instituto Geográfico Nacional
IMU	Inertial Measurement Unit
IoT	Internet of Things
IoU	Intersection over Union
JSON	JavaScript Object Notation
KTX2	Khronos Texture 2.0
M2M	Mesh-to-Mesh (surface-to-surface distance comparison)
MVS	Multi-View Stereo
NGINX	NGINX Web Server
PDAL	Point Data Abstraction Library
PNOA	Plan Nacional de Ortofotografía Aérea
POI	Point Of Interest
R-CNN	Region-based Convolutional Neural Network
ResNet	Residual Network
RoI	Region of Interest
RPN	Region Proposal Network
SfM	Structure from Motion
SLAM	Simultaneous Localization and Mapping
ToF	Time-of-Flight
YOLO	You Only Look Once

References

1. Fatás Monforte, P. Altamira, símbolo, identidad y marca. In *El Patrimonio Cultural Como Símbolo. Actas del Simposio Internacional*; Garrote, L., Ed.; Fundación del Patrimonio Histórico de Castilla y León: Valladolid, Spain, 2011; pp. 163–186.
2. de las Heras, C.; Montes, R.; Lasheras, J.A. Altamira: Nivel gravetiense y cronología de su arte rupestre. In *Pensando el Gravetiense: Nuevos Datos Para la Región Cantábrica en su Contexto Peninsular y Pirenaico*; de las Heras, C., Lasheras, J.A., Arrizabalaga, Á., de la Rasilla, M., Eds.; Monografías del Museo Nacional y Centro de Investigación de Altamira; Ministerio de Educación, Cultura y Deporte: Madrid, Spain, 2013; pp. 476–491.
3. Sanz de Sautuola, M. *Breves Apuntes Sobre Algunos Objetos Prehistóricos de la Provincia de Santander*; Imprenta y Litografía de Telesforo Martínez: Santander, Spain, 1880.
4. Lasheras, J.A. El arte paleolítico de Altamira. In *Redescubrir Altamira*; Lasheras, J.A., Ed.; Turner: Madrid, Spain, 2002; pp. 65–92.
5. Lasheras, J.A.; de las Heras, C.; Fatás Monforte, P. El nuevo museo de Altamira. *Boletín De La Soc. De Investig. Del Arte Rupestre De Boliv.* **2002**, *16*, 23–28.

6. Lasheras, J.A.; de las Heras, C. Estudio introductorio a Sanz de Sautuola, M. 1880. Breves apuntes sobre algunos objetos prehistóricos de la Provincia de Santander. In *Breves Apuntes Sobre Algunos Objetos Prehistóricos de la Provincia de Santander*; Botín, E., Ed.; Grupo Santander: Madrid, Spain, 2004.
7. Lasheras, J.A.; de las Heras, C.; Montes, R.; Rasines, P.; Fatás Monforte, P. *La Altamira del siglo XXI (el nuevo Museo y Centro de Investigación de Altamira)*; Museo Nacional y Centro de Investigación de Altamira: Patrimonio histórico de Castilla y León, Spain, 2002; pp. 23–34.
8. Fatás Monforte, P.; Lasheras Corruchaga, J.A. La cueva de Altamira y su museo/The cave of Altamira and its museum. *Cuad. De Arte Rupestre* **2014**, *7*, 25–35.
9. Sánchez-Moral, S.; Cuezva, S.; Fernández Cortés, Á.; Janices, I.; Benavente, D.; Cañaveras, J.C.; González Grau, J.M.; Jurado, V.; Laiz Trobajo, L.; Portillo Guisado, M.d.C.; et al. *Estudio Integral del Estado de Conservación de la Cueva de Altamira y Su Arte Paleolítico (2007–2009); Perspectivas Futuras de Conservación*; Ministerio de Educación Cultura y Deporte, Secretaría General Técnica: Madrid, Spain, 2014.
10. Sánchez, M.A.; Foyo, A.; Tomillo, C.; Iriarte, E. Geological Risk Assessment of the Area Surrounding Altamira Cave: A Proposed Natural Risk Index and Safety Factor for Protection of Prehistoric Caves. *Eng. Geol.* **2007**, *94*, 180–200. [CrossRef]
11. Sanchez-Moral, S.; Cuezva, S.; Garcia-Anton, E.; Fernandez-Cortes, A.; Elez, J.; Benavente, D.; Cañaveras, J.C.; Jurado, V.; Rogerio-Candelera, M.A.; Saiz-Jimenez, C. Microclimatic Monitoring in Altamira Cave: Two Decadesof Scientific Projects for Its Conservation. In *The Conservation of Subterranean Cultural Heritage*; CRC Press: Boca Raton, FL, USA, 2014.
12. de Guichen, G. *Programa de Investigación Para La Conservación Preventiva y Régimen de Acceso de La Cueva de Altamira (2012–2014)*; Ministerio de Cultura: Madrid, Spain, 2014; Volume 4.
13. Bayarri, V.; Prada, A.; García, F.; Ibáñez, M.; Benavente, D. Integration of Remote-Sensing Techniques for the Preventive Conservation of Paleolithic Cave Art in the Karst of the Altamira Cave. *Remote Sens.* **2023**, *15*, 1087. [CrossRef]
14. Bontemps, Z.; Crovadore, J.; Sirieix, C.; Bourges, F.; Gessler, C.; Lefort, F. Dark-Zone Alterations Expand throughout Paleolithic Lascaux Cave despite Spatial Heterogeneity of the Cave Microbiome. *Environ. Microbiome* **2023**, *18*, 31. [CrossRef]
15. Peña-Trabalon, A.; Moreno-Vegas, S.; Estebanez-Campos, M.B.; Nadal-Martinez, F.; Garcia-Vacas, F.; Prado-Novoa, M. A Low-Cost Validated Two-Camera 3D Videogrammetry System Applicable to Kinematic Analysis of Human Motion. *Sensors* **2025**, *25*, 4900. [CrossRef] [PubMed]
16. Matuzevičius, D.; Serackis, A. Three-Dimensional Human Head Reconstruction Using Smartphone-Based Close-Range Video Photogrammetry. *Appl. Sci.* **2022**, *12*, 229. [CrossRef]
17. Quispe-Enriquez, O.C.; Valero-Lanzuela, J.J.; Lerma, J.L. Craniofacial 3D Morphometric Analysis with Smartphone-Based Photogrammetry. *Sensors* **2024**, *24*, 230. [CrossRef] [PubMed]
18. Teixeira Coelho, L.C.; Pinho, M.F.C.; Martinez de Carvalho, F.; Meneguci Moreira Franco, A.L.; Quispe-Enriquez, O.C.; Altónaga, F.A.; Lerma, J.L. Evaluating the Accuracy of Smartphone-Based Photogrammetry and Videogrammetry in Facial Asymmetry Measurement. *Symmetry* **2025**, *17*, 376. [CrossRef]
19. Marčiš, M.; Fraštia, M.; Hideghéty, A.; Paulík, P. Videogrammetric Verification of Accuracy of Wearable Sensors Used in Kiteboarding. *Sensors* **2021**, *21*, 8353. [CrossRef]
20. Sun, Z.; Zhang, Y. Accuracy Evaluation of Videogrammetry Using a Low-Cost Spherical Camera for Narrow Architectural Heritage: An Observational Study with Variable Baselines and Blur Filters. *Sensors* **2019**, *19*, 496. [CrossRef]
21. Ortiz-Coder, P.; Sánchez-Ríos, A. An Integrated Solution for 3D Heritage Modeling Based on Videogrammetry and V-SLAM Technology. *Remote Sens.* **2020**, *12*, 1529. [CrossRef]
22. Pepe, M.; Alfio, V.S.; Costantino, D.; Herban, S. Rapid and Accurate Production of 3D Point Cloud via Latest-Generation Sensors in the Field of Cultural Heritage: A Comparison between SLAM and Spherical Videogrammetry. *Heritage* **2022**, *5*, 1910–1928. [CrossRef]
23. Alsadik, B.; Khalaf, Y.H. Potential Use of Drone Ultra-High-Definition Videos for Detailed 3D City Modeling. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 34. [CrossRef]
24. Currà, E.; D'Amico, A.; Angelosanti, M. HBIM between Antiquity and Industrial Archaeology: Former Segrè Papermill and Sanctuary of Hercules in Tivoli. *Sustainability* **2022**, *14*, 1329. [CrossRef]
25. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 386–397. [CrossRef] [PubMed]
26. Yu, Y.; Zhang, K.; Yang, L.; Zhang, D. Fruit Detection for Strawberry Harvesting Robot in Non-Structural Environment Based on Mask-RCNN. *Comput. Electron. Agric.* **2019**, *163*, 104846. [CrossRef]
27. Zhang, Q.; Chang, X.; Bian, S. Vehicle-Damage-Detection Segmentation Algorithm Based on Improved Mask RCNN. *IEEE Access Pract. Innov. Open Solut.* **2020**, *8*, 6997–7004. [CrossRef]
28. Wu, S.; Fu, F. Crack Control Optimization of Basement Concrete Structures Using the Mask-RCNN and Temperature Effect Analysis. *PLoS ONE* **2023**, *18*, e0292437. [CrossRef]
29. Ameli, Z.; Nesheli, S.J.; Landis, E.N. Deep Learning-Based Steel Bridge Corrosion Segmentation and Condition Rating Using Mask RCNN and Yolov8. *Infrastructures* **2023**, *9*, 3. [CrossRef]

30. Yang, F.; Huo, J.; Cheng, Z.; Chen, H.; Shi, Y. An Improved Mask R-CNN Micro-Crack Detection Model for the Surface of Metal Structural Parts. *Sensors* **2023**, *24*, 62. [CrossRef]
31. Bonhage, A.; Eltaher, M.; Raab, T.; Breuß, M.; Raab, A.; Schneider, A. A Modified Mask Region-Based Convolutional Neural Network Approach for the Automated Detection of Archaeological Sites on High-Resolution Light Detection and Ranging-Derived Digital Elevation Models in the North German Lowland. *Archaeol. Prospect.* **2021**, *28*, 177–186. [CrossRef]
32. Hatır, M.E.; İnce, İ.; Korkanç, M. Intelligent Detection of Deterioration in Cultural Stone Heritage. *J. Build. Eng.* **2021**, *44*, 102690. [CrossRef]
33. Liu, Y.; Yao, J.; Lu, X.; Xie, R.; Li, L. DeepCrack: A Deep Hierarchical Feature Learning Architecture for Crack Segmentation. *Neurocomputing* **2019**, *338*, 139–153. [CrossRef]
34. Li, G.; Ma, B.; He, S.; Ren, X.; Liu, Q.; Li, G.; Ma, B.; He, S.; Ren, X.; Liu, Q. Automatic Tunnel Crack Detection Based on U-Net and a Convolutional Neural Network with Alternately Updated Clique. *Sensors* **2020**, *20*, 717. [CrossRef] [PubMed]
35. Ren, Y.; Huang, J.; Hong, Z.; Lu, W.; Yin, J.; Zou, L.; Shen, X. Image-Based Concrete Crack Detection in Tunnels Using Deep Fully Convolutional Networks. *Constr. Build. Mater.* **2020**, *234*, 117367. [CrossRef]
36. Sohaib, M.; Arif, M.; Kim, J.-M.; Sohaib, M.; Arif, M.; Kim, J.-M. Evaluating YOLO Models for Efficient Crack Detection in Concrete Structures Using Transfer Learning. *Buildings* **2024**, *14*, 3928. [CrossRef]
37. Song, Y.; Su, Y.; Zhang, S.; Wang, R.; Yu, Y.; Zhang, W.; Zhang, Q.; Song, Y.; Su, Y.; Zhang, S.; et al. CrackdiffNet: A Novel Diffusion Model for Crack Segmentation and Scale-Based Analysis. *Buildings* **2025**, *15*, 1872. [CrossRef]
38. Yu, G.; Dong, J.; Wang, Y.; Zhou, X.; Yu, G.; Dong, J.; Wang, Y.; Zhou, X. RUC-Net: A Residual-Unet-Based Convolutional Neural Network for Pixel-Level Pavement Crack Segmentation. *Sensors* **2022**, *23*, 53. [CrossRef]
39. Bayarri, V.; Prada, A.; García, F. A Multimodal Research Approach to Assessing the Karst Structural Conditions of the Ceiling of a Cave with Palaeolithic Cave Art Paintings: Polychrome Hall at Altamira Cave (Spain). *Sensors* **2023**, *23*, 9153. [CrossRef]
40. *Rock Characterization, Testing and Monitoring: ISRM Suggested Methods*; Brown, E.T., Ed.; Pergamon Press: Oxford, UK, 1981.
41. Sánchez, M.A.; Bruschi, V.; Iriarte, E. Evaluación del riesgo geológico en las cuevas de Altamira y Estalactitas. In *Monografías del Museo y Centro de Investigación Altamira*; Ministerio de Cultura: Madrid, Spain, in press.
42. Ouster, Inc. *OS0 Ultra-Wide View High-Resolution Imaging Lidar: Datasheet (Rev7, v3.1)*; Ouster, Inc.: San Francisco, CA, USA, 2025.
43. Wilkinson, M.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; Bonino da Silva Santos, L.O.; Bourne, P.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef]
44. Lin, T.-Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. *arXiv* **2015**, arXiv:1405.0312. [CrossRef]
45. Jaikumar, P.; Vandaele, R.; Ojha, V. Transfer Learning for Instance Segmentation of Waste Bottles Using Mask R-CNN Algorithm. *arXiv* **2022**, arXiv:2204.07437. [CrossRef]
46. Jaccard, P. The Distribution of the Flora in the Alpine Zone. *New Phytol.* **1912**, *11*, 37–50. [CrossRef]
47. Luo, S.; Wang, H. Digital Twin Research on Masonry–Timber Architectural Heritage Pathology Cracks Using 3D Laser Scanning and Deep Learning Model. *Buildings* **2024**, *14*, 1129. [CrossRef]
48. Xu, X.; Zhao, M.; Shi, P.; Ren, R.; He, X.; Wei, X.; Yang, H. Crack Detection and Comparison Study Based on Faster R-CNN and Mask r-CNN. *Sensors* **2022**, *22*, 1215. [CrossRef] [PubMed]
49. Flyability. How Accurate Are the 3D Models You Can Make with FlyAware? Flyability Blog. Available online: https://www.flyability.com/blog/flyaware-accuracy (accessed on 3 January 2026).
50. Benedetto, A.D.; Fiani, M.; Gujski, L.M. U-Net-Based CNN Architecture for Road Crack Segmentation. *Infrastructures* **2023**, *8*, 90. [CrossRef]
51. PostgreSQL Global Development Group. *PostgreSQL; Relational Database Management System*; PostgreSQL Global Development Group: Santa Barbara, CA, USA, 2025.
52. Django Software Foundation. *Django Web Framework; Web Application Framework for Python*; Django Software Foundation: Huntersville, NC, USA, 2025.
53. Schütz, M. Potree: Rendering Large Point Clouds in Web Browsers. Ph.D. Thesis, Technische Universität Wien, Vienna, Austria, 2015.
54. PDAL Contributors PDAL: Point Data Abstraction Library; Point Cloud Processing Library; PDAL Project, 2025. Available online: https://pdal.io/en/2.9.0/project/contributors.html (accessed on 1 January 2026).
55. NGINX, Inc. NGINX Web Server 2025. Available online: https://nginx.org/ (accessed on 1 January 2026).
56. Smithsonian Institution Digitization Program Office. *Smithsonian Voyager/DPO-Voyager Tools; 3D Visualization and Annotation Toolkit*; Smithsonian Institution: Washington, DC, USA, 2025.
57. Khronos Group. *glTF: GL Transmission Format; 3D Scene and Model Transmission Format*; Khronos Group: Beaverton, OR, USA, 2025.
58. Trimesh Developers Trimesh: Python Library for Loading and Processing 3D Geometry. 2025. Available online: https://trimesh.org/ (accessed on 1 January 2026).
59. Google LLC. *Draco Compression Library; 3D Geometry Compression Library*; Google LLC: Mountain View, CA, USA, 2025.

60. McCurdy, D. glTF-Transform: Toolkit for glTF Optimization. 2025. Available online: https://gltf-transform.dev/ (accessed on 1 January 2026).
61. Docker, Inc. *Docker*; Containerization Platform; Docker, Inc.: Palo Alto, CA, USA, 2025.

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.