AI-Driven 3D reconstruction and quality assessment for Cultural Heritage: first results from the HERITALISE project

Filiberto Chiabrando[1], Andrea Maria Lingua[2] , Alessio Martino[1], Francesca Matrone[2], Alessandra Spadaro[2]

[1] Laboratory of Geomatics for Cultural Heritage (LabG4CH), Department of Architecture and Design (DAD), Politecnico di Torino, Viale Pier Andrea Mattioli, 39, Torino (TO), Italy – filiberto.chiabrando@polito.it; alessio.martino@polito.it
[2] Geomatics Lab, Department of Environment, Land and Infrastructure Engineering (DIATI), Politecnico di Torino, Corso Duca degli Abruzzi, 24, Torino (TO), Italy – andrea.lingua@polito.it; francesca.matrone@polito.it; alessandra.spadaro@polito.it

Keywords: 3D Gaussian Splatting, cracks detection, statistical analysis, standard metrics, UAV, data fusion

Abstract

The accurate digital documentation of Cultural Heritage (CH) assets demands workflows capable of integrating heterogeneous, multiscale datasets while preserving both geometric fidelity and radiometric completeness. This paper presents the first results of the AI-based processing pipeline developed within the HERITALISE project (Horizon Europe, 2025–2028), applied to three multiscale case studies at the *Reggia di Venaria Reale* (Turin, Italy): an outdoor-indoor UAV photogrammetric survey, a kinematic SLAM acquisition of a contemporary sculpture garden, and a close-range dataset of an 18th-century decorative artefact. 3D Gaussian Splatting (3DGS) is evaluated as a novel view synthesis method across all three scenarios, demonstrating strong photorealistic rendering capabilities, particularly for complex material properties and geometrically challenging interiors, whilst highlighting current limitations for metric surveying applications. A two-stage crack detection workflow, combining tile-based text-prompted segmentation with SAM3 and multiview ray-based reprojection onto the reconstructed mesh, is validated on UAV imagery, achieving an 84.9% ray–mesh intersection rate. Finally, a standardised evaluation framework is proposed, encompassing adaptive, scale-dependent geometric and radiometric metrics organised into reference-based and no-reference assessment scenarios, aggregated into a transparent synthetic quality score with three adaptive quality classes. The proposed methodology contributes toward a reproducible, sensor-agnostic standard for the assessment of AI-generated CH documentation products.

1. Introduction

1.1 The HERITALISE project

Cultural Heritage (CH) objects encompass a vast array of tangible and intangible elements, including movable artifacts, architectural structures, archaeological sites, and natural environments, all of which can be captured in digital form.

Documenting, analysing, and preserving this diversity requires complex and heterogeneous datasets. Initial studies, such as VIGIE 2020/654 from the EU, represent a suitable starting point; however, no internationally recognised framework, methodology, or standard currently exists to define the quality, level of detail, completeness, or accuracy of CH digitisation processes, especially when dealing with AI (Artificial Intelligence) techniques.

The HERITALISE project (2025–2028), funded by the HORIZON Europe Programme under the call “European Collaborative Cloud for Cultural Heritage – Innovative tools for digitising cultural heritage objects” aims to investigate and develop advanced digitisation methods capable of accurately documenting and representing diverse CH assets, including both visible and non-visible characteristics, to support a comprehensive understanding of heritage components (Chiabrando et al., 2025).

In particular, Work Package 6 (WP6) focuses on the use of new methods for data processing and integration, supported by AI. More in detail, the objectives are related to:

- support the CH documentation and preservation with a multi-scale, multi-sensor and multi-temporal processing approach in 2D/3D;
- design a pipeline for data processing and fusion (visible/non-visible);
- define and test AI approaches using generative algorithms to improve the data completeness;
- develop a visualization system 2D/3D and 4D for temporal data, allowing experts/visitors to have different points of view and grades of interest, ensuring a complete knowledge of the object.

In the HERITALISE project, the activities related to WP6 began in September 2025.

Within this framework, the development of advanced methodologies for 2D and 3D data processing and post-processing, with the aim of enhancing the efficiency of workflows that integrate both visible and non-visible data sources, will allow to adopt a multi-scale, multi-sensor, and multi-temporal approach. The objective is to facilitate seamless data integration and deliver cost-effective solutions combining various digitisation approaches (e.g., LiDAR, photogrammetry, tomography, X-ray, etc.).

This integrated strategy supports the comprehensive digital reconstruction of CH objects, encompassing both internal and external features, without requiring physical disassembly. The process will further incorporate metadata (e.g., local measurement uncertainty) and paradata (e.g., digitisation conditions and environmental factors), alongside additional contextual information.

Moreover, the potential of AI-based generative algorithms will be investigated to improve data completeness and to support essential tasks such as data cleaning, classification, and annotation, which are critical for conservation and restoration activities, including the study of imperceptible attributes as material identification, structural damage assessment, and crack detection. In particular, advanced techniques such as Neural Radiance Fields (NeRF) and 3D Gaussian Splatting (3DGS) will be juxtaposed and employed to enhance the completeness and quality of digital representations of CH assets.

At the end, to assess the proposed methodologies and approaches, an evaluation of the accuracy and completeness of the so-generated data, through semi-automatic and automatic tests based on statistical tools, will also be developed. The underlying idea is to provide a standardized set of metrics for the assessment of the generated images and point clouds.

This contribution thus presents the initial results of the AI-based workflow proposed for the HERITALISE project. It begins from the data processing of the multiscale case studies with novel view synthesis methods as 3DGS, then it evolves into their analysis for the identification of cracks, and, in the end, the proposal of a new set of metrics to evaluate the so-generated data (completeness and metric accuracy).

1.2 State-of-the-art

The inspection and documentation of cultural and architectural heritage structures increasingly relies on the integration of data acquired at multiple spatial scales. Photogrammetric and LiDAR-based survey methodologies have been progressively combined to capture both the macro-scale geometry of entire structures and the fine-grained surface detail required for damage assessment (Remondino et al., 2023).

Uncrewed aerial vehicles (UAVs), terrestrial laser scanners, and close-range photogrammetric systems are routinely deployed in complementary configurations to maximise scene coverage while retaining the millimetre resolution needed to resolve crack networks and material discontinuities. On the other hand, X-rays, tomography, radar techniques, etc. could complement the complete digitisation of these 3D models.

However, the processing of such heterogeneous, multiscale datasets entails significant challenges related to co-registration, scale consistency, point cloud densification, different ground sampling distances, dissimilar sensor modalities, calibration and robust feature-matching strategies, etc.

In this framework, novel view synthesis methods could contribute to a faster, and easier data visualization and fruition. NeRF has emerged as a transformative technology for the digital documentation and preservation of architectural and cultural heritage (Remondino et al., 2023), enabling highly photorealistic 3D reconstructions from standard-oriented 2D images even in the presence of poor texture areas, reflective, or refractive surfaces. Unlike traditional photogrammetry or LiDAR, NeRF can reconstruct complex geometries and occluded areas, which frequently occur in historical architecture, while maintaining accurate point fidelity (Murtiyoso & Grussenmayer, 2023).

Conversely, 3D Gaussian Splatting (3DGS) has quickly emerged as a powerful technique, particularly for the digital visualization of cultural heritage. This method is especially valuable for the project's objectives, as it excels in capturing intricate architectural details and handling complex lighting conditions characteristic of heritage sites (Clini et al., 2024), which may arise during data acquisition. This technique is currently the most widely used for CH rendering and visualization.

Coupled with the 3D reconstruction, there is the semantic interpretation of the so-generated data. The automated identification of cracks in the built heritage sector constitutes an active and rapidly evolving research domain. Early deep learning approaches adapted convolutional neural networks (CNNs), originally developed for natural image classification, to the task of binary crack segmentation in 2D imagery (Zhang et al., 2016; Zou et al., 2019). Fully convolutional architectures, and in particular encoder-decoder networks such as U-Net (Ronneberger et al., 2015), have become the predominant choice for pixel-wise crack segmentation due to their ability to capture multi-scale contextual information while preserving spatial resolution through skip connections.

More recent investigations have explored the application of Vision Transformers (ViTs) and hybrid CNN-Transformer architectures to crack detection, leveraging global self-attention mechanisms to capture long-range spatial dependencies that are difficult to model with purely local convolutional operations (Dosovitskiy et al., 2020). Foundation models such as the Segment Anything Model (SAM) (Kirillov et al., 2023) have demonstrated remarkable zero-shot generalisation capabilities and have been adapted, through fine-tuning on domain-specific datasets, to the automated delineation of crack boundaries in structural imagery.

A persistent challenge in this field is the scarcity of large-scale, annotated training datasets for structural damage, particularly at the level of individual cracks with precise geometric labels. Domain adaptation and transfer learning strategies have been explored to mitigate the distributional gap between synthetic or laboratory-generated data and real-world inspection images (Dorafshan et al., 2018). Furthermore, the transition from 2D crack maps to 3D crack characterisation, encompassing width, length, depth, and orientation, remains an open problem.

Considering all the above-mentioned aspects, a unique workflow able to exploit the strengths of AI approaches, but at the same time, keeping a critical vision and able to evaluate the outputs, will be proposed in WP6 of the HERITALISE project.

1.3 The case study: the UNESCO site *Reggia di Venaria Reale*

As part of the HERITALISE demo site, different multi-scale datasets related to the *Reggia di Venaria Reale* (Turin, northwestern Italy) are used for the first tests. The complex of *La Venaria Reale* was conceived from 1658 as an estate for "leisure and hunting"; today, it is a museum that offers visitors magical atmospheres, cultural attractions, and leisure pursuits. It has been chosen for the purposes of this study since it offers the possibility to have together: *outdoor environments* with external facades and gardens, *indoor scenarios* with the private royal residence rooms and spaces, and *movable CH objects* as furniture or statues.

For the first tests, three different scenarios were employed:

1. a portion of the main central block of the residence, acquired with UAVs, related to the outdoor portion of the *Galleria Grande* facing the parterre garden of the Fluid sculptures (Figure 1a). The focus was mainly on the upper framing and the roof plan for the 3D reconstruction with 3DGS approaches;
2. a portion of the internal environment of the *Galleria Grande* at the first floor level (Figure 1b) for the automatic crack identification;
3. and a small object (*Piffetti dressing table casket*, Figure 1c), related to the permanent collection of the museum, for the 3D reconstruction of small museum objects.

https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-809-2026

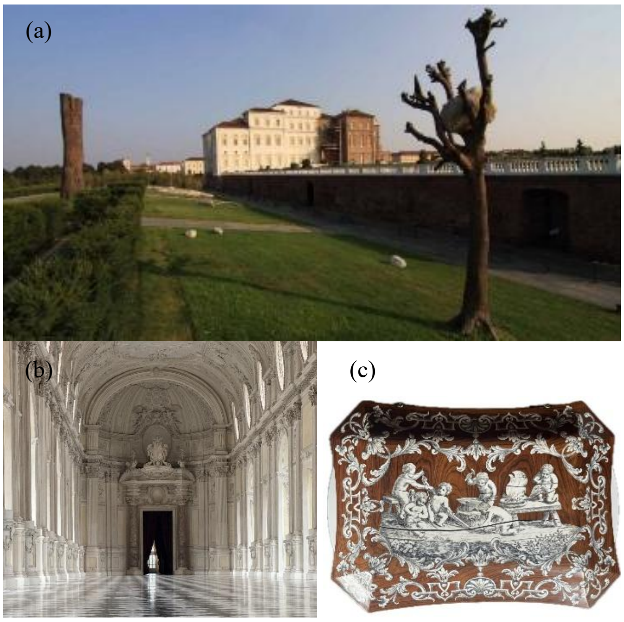

Figure 1. The scenarios selected for the first tests: the Garden of the Fluid sculptures with the main building block (a); the interior of Galleria Grande (b); the Piffetti dressing table casket (c).

2. Materials and methods

Figure 2 illustrates the steps of the proposed workflow: from the data acquisition and processing to the semantic interpretation and accuracy assessment of the data through a new set of standardized metrics.

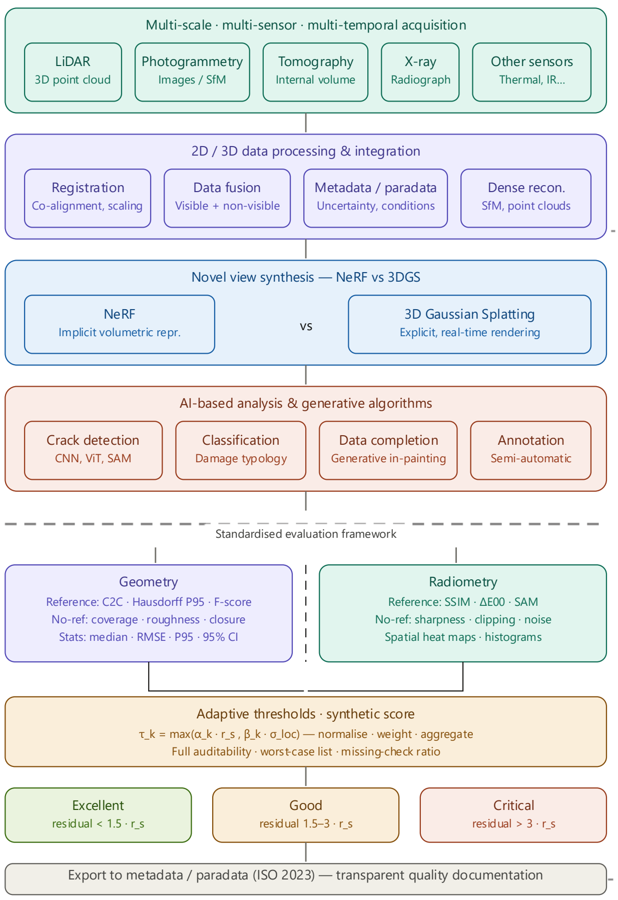

Figure 2. The overall workflow for WP6.

2.1 Data acquisition

To fully capture the immense spatial complexity of the Reggia di Venaria Reale, within the HERITALISE project, a comprehensive, multi-scalar data acquisition strategy was implemented. The methodology relied on the fusion of heterogeneous surveying techniques tailored to different levels of detail.

On a small scale, UAV photogrammetry and Simultaneous Localization and Mapping (SLAM) systems were deployed to rapidly map the vast architectural volumes, exterior landscapes, and complex interior trajectories across the entire estate. To ensure rigorous geometric accuracy for the primary architectural elements, these global datasets were integrated with a network of static Terrestrial Laser Scanning (TLS) setups.

Finally, to reconstruct the intricate morphological features of the estate's sculptural heritage, dedicated close-range terrestrial photogrammetry campaigns were conducted on smaller statues and individual artifacts. This integrated multi-sensor approach yielded a highly robust, multi-resolution dataset, providing the necessary foundation for advanced 3D processing and high-fidelity rendering.

However, for the specific purposes of this study, a targeted subset from this heterogeneous archive was selected to evaluate the potential of 3DGS (Table 1). By isolating these specific datasets, the research aims to assess the versatility and visual fidelity of 3DGS across varying spatial scales and instrumental origins.

Technique	Device	Main Technical Features	N°
UAV-based photogram.	DJI Mini 5 Pro	• 1” CMOS 50 MP sensor • f = 24 mm • 8192×6144 px	95 images 69’831 Tie points
SLAM + 360° camera	Stonex X70GO	• Range: 0.1-70 m • 200k pts/s • 1” 12 MP 210° FOV Color camera, f = 1.26 mm • 1” 12 MP 100° FOV Visual camera, f = 3.24 mm	38.5 M points
	Insta 360 X5	• 1/1.28” f = 6mm • 5.7 K+: 5760x2880 360° videos	3’795 images
vSLAM terrestrial photogrammetry	Stonex XVS	• Range: 0.4-40 m • 640x480 px, 65° FOV vSLAM camera • 5 MP 2448x2048 px, 89° FOV RGB camera	517 images 4.3 M points

Table 1. Summary of the selected data for 3DGS testing.

This encompasses three distinct acquisition scenarios and sensor modalities: an indoor UAV photogrammetric survey of the Galleria Grande captured with a DJI Mini 5 Pro consisting of 540 images of the internal walls acquired with perpendicular and oblique inclination and at various distances from the wall, to properly reconstruct the decorations and geometrical features of the Galleria; a kinematic SLAM trajectory navigating through the

Garden of Fluid Sculptures by Giuseppe Penone, utilizing a Stonex x70go equipped with an Insta360 X5 camera set at a calibrated distance in order to have a high-quality data as the base for 3DGS reconstruction; and a close-range terrestrial photogrammetry dataset of a wooden dressing table casket created in the 18th century by Pietro Piffetti, acquired via the Stonex XVS vSLAM (visual SLAM) system.

2.2 Data processing

All the datasets acquired and selected for this paper have been processed initially with the traditional SfM (Structure from Motion) approach or on the instrument's proprietary software, in order to generate a point cloud acting as the base for the 3DGS training process, which was carried out in different environments.

2.2.1 Interior Walls of the Galleria Grande: The interior of the Galleria Grande, a monumental Baroque corridor characterized by its vast scale, intricate stucco decorations, and iconic black-and-white checkered flooring, was digitized using a DJI Mini 5 Pro. To navigate the spatial complexity of the hall, the acquisition strategy employed a multi-angle approach: images were first captured parallel to the walls to establish a geometric baseline, followed by oblique shots designed to accurately resolve the protruding architectural reliefs and ornamental elements on the upper portion of the walls and on the two portals at the end of the gallery.

However, processing this dataset through traditional SfM pipelines, such as Agisoft Metashape (Figure 3), yielded significant camera alignment errors. These registration failures were primarily driven by the severe visual ambiguity of the environment's repeating features, namely the uniform floor tiles and the periodic architectural rhythms of the walls, increased by the absence of surveyed Ground Control Points (GCPs) to anchor the block adjustment.

Despite the lack of a complete global reconstruction, a robust, localized portion of the generated point cloud was successfully salvaged and utilized to initialize the subsequent 3D Gaussian Splatting models, as shown in Figure 4.

Figure 3. Dense point cloud generated using Agisoft Metashape on the portion of the Galleria Grande.

Figure 4. Comparison between the 3DGS models trained in Jawset Postshot (left) and in LichtFeld Studio (right).

2.2.2 Garden of the Fluid Sculptures: Transitioning to the exterior, the Garden of Fluid Sculptures by contemporary artist Giuseppe Penone presents a distinctly different digitization challenge. Located in the estate's lower park, the installation is an expansive, multi-sensory environment composed of highly organic shapes, including bronze trees, carved stone, water features, and integrated natural vegetation.

To efficiently capture this complex open-air setting, a kinematic acquisition was performed using a Stonex X70 GO mobile SLAM system equipped with an Insta360 X5 panoramic camera. This walking-trajectory approach allowed for continuous and rapid data collection along the garden's intricate pathways. To achieve the 3DGS reconstruction, the data processing adhered to a specific pipeline recommended by the manufacturer: the raw SLAM data and 360-degree imagery were initially processed in Stonex's proprietary GoPost software to resolve the precise trajectory, align the panoramic frames, and generate the foundational point cloud (Figure 5).

This spatially oriented dataset was subsequently exported to *Jawset Postshot*, where the final 3D Gaussian Splatting model was trained, effectively handling the highly irregular morphologies of Penone's artworks (Figure 6).

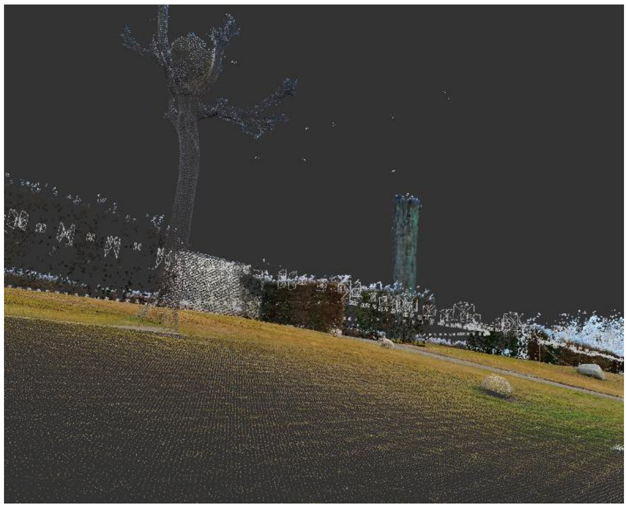

Figure 5. SLAM point cloud generated from the proprietary software GoPost of the Garden of the Fluid sculptures.

Figure 6. 3DGS model trained on Jawset Postshot on the Garden of the Fluid sculptures.

2.2.3 Piffetti Dressing Table Casket: To evaluate the performance on highly detailed, small-scale heritage artifacts, the study focused on the close-range documentation of a small wooden dressing table casket decorated by the renowned 18th-century cabinetmaker Pietro Piffetti. Given the object's intricate marquetry, diverse material properties, and elaborate surfaces, the acquisition was conducted using the Stonex XVS vSLAM (Visual SLAM) system. To ensure strict metric accuracy and facilitate the spatial orientation process, calibrated scalebars and coded targets were strategically distributed around the scene prior to scanning. These physical markers enabled the automatic detection and precise scaling of the model during the alignment phase.

The initial processing workload was streamlined through the manufacturer's ecosystem: the raw image datasets were uploaded and processed on Stonex's proprietary cloud service, which performed the primary feature matching and scaled 3D reconstruction. This point cloud (Figure 7) has been set as the base for the 3D Gaussian Splatting model in LichtFeld Studio (Figure 8).

Figure 7. Point cloud generated on the proprietary software 2fS Cloud of the Piffetti Dressing Table Casket.

Figure 8. 3DGS model trained in LichtFeld Studio of the Piffetti Dressing Table Casket.

2.3 Cracks detection

Crack detection was investigated through a two-stage workflow combining an image-based segmentation step, aimed at identifying candidate crack regions in high-resolution UAV images, and a multiview reprojection step, aimed at transferring this information to the 3D reconstructed surface.

In the first stage, crack candidates are extracted from undistorted images through a tiled text-prompted segmentation based on Meta SAM3: Segment Anything with Concepts (Carion et al., 2025). Since cracks are thin and localized features, each image is subdivided into overlapping tiles in order to preserve fine details and improve the detection of small discontinuities. For each tile, the prompt '*crack'* is used to generate candidate masks with an associated confidence score. These masks are then filtered and merged back into the original image space, producing a full-resolution output for each image. Rather than relying on a single binary result, the method combines the selected masks into a probability map and a related uncertainty map, so that both the spatial extent of the detected features and the confidence of the model response can be retained.

In the second stage, the image-based outputs are transferred to 3D through a multiview ray-based projection procedure. Camera calibration parameters derived from the SfM photogrammetric process are used to reconstruct the viewing geometry of each image and to intersect image rays with the triangulated mesh. The probability values associated with the detected pixels are then assigned to the corresponding 3D locations, generating a set of class points that describes the spatial distribution of crack evidence on the surface. A thresholded subset can also be extracted to retain only the most reliable responses, while view-dependent weighting is used to favour observations acquired under more suitable geometric conditions.

Although tested here on cracks, the same workflow can be extended to other subtle surface anomalies within the HERITALISE framework. In future developments, the derived 3D information could support both monitoring surface anomalies over time and integrating them with other diagnostic layers, thereby contributing to more informed conservation and documentation workflows.

2.4 Accuracy and completeness evaluation

The HERITALISE framework adopts a structured methodology for the evaluation of AI-derived products for CH documentation, with the dual objective of measuring both accuracy and completeness in a reproducible, scalable, and interpretable way (ISO, 2023). The proposed scheme is organised along two orthogonal dimensions, namely geometry and radiometry, and is further articulated according to two operational scenarios:

- *reference-based* evaluation, when a reliable ground-truth model or image dataset is available, and
- *no-reference* evaluation, when only internal consistency, acquisition metadata, and self-diagnostic indicators can be exploited (ISO, 2021; ISO, 2023).

In both cases, the assessment is performed at local and global scales, so that pointwise anomalies, spatially clustered degradations, and overall dataset performance can be jointly analysed.

A preliminary harmonization stage is required before any comparison. In the reference-based case, the generated product is first co-registered to the reference through accurate checkpoint-

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-809-2026 |

based or surface-based alignment procedures, and the geometric agreement is assessed according to established positional-accuracy principles. To reduce sampling bias and density-driven distortions, both datasets are regularized through controlled subsampling procedures, such as voxel-grid or Poisson-disk sampling, and unstable or non-overlapping regions are explicitly masked. The evaluation then relies on robust statistics rather than on a single residual descriptor: median, MAD (Median Absolute Deviation), IQR (Interquartile Range), RMSE, P95 percentiles, bias, and 95% confidence intervals are computed systematically in order to distinguish systematic effects, heavy-tailed residuals, and local outliers (JCGM, 2008). . This choice is particularly important in Cultural Heritage applications, where complex geometries, partial occlusions, heterogeneous materials, and multi-scale acquisitions may produce non-Gaussian error distributions.

For *geometric reference-based* evaluation, the framework includes a set of complementary checks addressing both metric agreement and spatial completeness. Depending on the available reference, the analysis may consider control-point residuals, cloud-to-cloud or cloud-to-mesh distances, robust Hausdorff descriptors (Aspert et al., 2002), local normal deviation, curvature consistency (Rusinkiewicz, 2004), and threshold-based completeness measures such as recall and F-score. In this context, the use of P95 rather than the maximum distance is preferred to limit the influence of isolated outliers, while local distance-based comparison strategies remain consistent with established 3D change-quantification and cloud-comparison approaches (Lague et al., 2013). Spatially distributed residuals are represented through 3D heat maps and histograms, allowing the identification of systematic deformation patterns, localized drift, or scale-dependent instabilities.

For *geometric no-reference* evaluation, the methodology instead focuses on internal quality proxies, including coverage ratio, hole ratio, local roughness, outlier rate, registration closure error, and internal precision estimated through Monte Carlo perturbation or subset-stability analysis. These indicators do not provide external accuracy in an absolute sense, but they are essential for identifying unreliable areas, low-support regions, and internal inconsistencies when no external benchmark is available (JCGM, 2008).

A similar logic is applied to the *radiometric component*. In the *reference-based radiometric case*, the comparison can be performed either in image space or on textured products, depending on the workflow. The proposed indicators include SSIM for structural similarity of images (Wang et al., 2004), ΔE00 for perceptual colour difference (Sharma et al., 2005), and SAM (Spectral Angle Mapper, Kruse et al., 1993) for multispectral or hyperspectral spectral agreement, and additional radiometric checks specifically relevant to imaging quality assessment in cultural heritage digitization workflows (ISO, 2021).

In the *no-reference radiometric* scenario, the framework relies on sharpness and blur diagnostics, clipping in highlights and shadows, effective usable dynamic range, noise outlier rate, overlap-based exposure consistency, multiview self-consistency, white-balance drift, and seam artefact detection again in line with established imaging-system quality analysis principles (ISO, 2021). Also in this case, the interpretation is not restricted to mean values, but is grounded on distribution-based descriptors, quantiles, confidence bounds, and spatial visualizations, so that both global degradation and localized radiometric failures can be explicitly documented (JCGM, 2008).

Finally, all indicators are converted into normalized sub-scores and aggregated into a synthetic score, while preserving full auditability of the evaluation. The aggregation is performed through a weighted scheme in which base indicators directly obtainable from standard outputs receive higher weights than advanced indicators requiring additional modelling or processing. However, the final synthetic score is never presented as an isolated number: it is always accompanied by the list of worst-performing indicators, the percentage of available versus missing checks, and any gate conditions or penalties applied because of registration failure, insufficient overlap, or incomplete radiometric support. This reporting strategy supports transparent quality documentation and direct export of quality descriptors to metadata and paradata structures (ISO, 2023). In this schema, a set of three adaptive quality classes (excellent, Good, and Critical) can be defined as a function of the effective scale of the survey rather than through fixed universal thresholds.

Let r_s denote the characteristic resolution of the product, expressed for example by the image GSD, the average point spacing, the mesh edge length, or the voxel size; when available, local precision estimates σ_{loc} derived from measurement uncertainty can also be incorporated. Thresholds are then defined as adaptive functions of scale:

$$\tau_k = \max\left(\alpha_k r_s, \beta_k \sigma_{loc}\right) \tag{1}$$

so that the same metric can be interpreted consistently across UAV blocks, close-range surveys, small artefacts, or architectural-scale models. For distance-like metrics, where lower values indicate better performance, a practical initialization may classify **Excellent** results for residuals below about (1-1.5) r_s, **Good** results (1.5-3) r_s and **Critical** results above $3r_s$, with coefficients to be refined according to sensor type and object complexity. For coverage- or similarity-based indicators, where higher values are better, the same adaptive logic is applied indirectly by computing the indicator within a scale-dependent neighbourhood or tolerance τ_k. In this way, the class assignment remains physically meaningful and comparable across datasets with different nominal resolutions. Each indicator is encoded in numerical form to ensure objective computation, comparability, and transparent aggregation, while a core subset of base indicators is defined to be systematically applied in all evaluation scenarios, independently of data type and workflow complexity

This reporting strategy avoids black-box interpretation and supports direct export of quality descriptors to metadata and paradata structures, in line with the HERITALISE objective of defining a transparent and standardized framework for the assessment of AI-supported digitization results. The final synthetic score is conceived as a composite indicator obtained through normalization, weighting, and aggregation of the selected quality metrics, following established methodological principles for composite-indicator construction and robustness analysis (Nardo et al., 2005; Munda and Nardo, 2005). Within HERITALISE, this score is introduced as an interpretative and comparative aid, not as a replacement for the explicit reporting of individual indicators, uncertainty bounds, and worst-case conditions.

3. Results

3.1 3D Gaussian Splatting

The visual assessment of the 3D Gaussian Splatting models demonstrates significant qualitative improvements over their foundational point clouds.

Regarding the dynamic SLAM dataset acquired in the Garden of Fluid Sculptures, the 3DGS rendering obtained through Jawset Postshot provides a continuous and highly photorealistic representation of the outdoor environment. When compared to the discrete, sparse nature of the raw SLAM point cloud, the Gaussian reconstruction of the complex bronze sculptures, boundary walls, and terrain exhibits a visibly closer correspondence to the real world, effectively resolving the irregular organic morphologies.

Similarly, in the challenging interior scenario of the Galleria Grande, the 3DGS approach showcased strong generative capabilities across different software implementations. While the traditional SfM point cloud suffered from severe data gaps, caused by alignment failures over repetitive architectural features without the aid of Ground Control Points, the trained Gaussian models successfully synthesized the missing visual information. By interpolating the available spatial data and camera views, both Jawset Postshot and LichtFeld Studio effectively filled the holes left by the SfM process, yielding visually complete renderings of the complex stucco decorations and marble columns.

However, a qualitative comparison between the two 3DGS outputs reveals subtle differences in how the algorithms handle empty space: while Jawset Postshot produced a notably clean and cohesive visual surface, LichtFeld Studio, despite maintaining strong textural sharpness, introduced minor floating artifacts in the peripheral areas of the scene. Nevertheless, both platforms definitively demonstrate the capacity of 3DGS to visually repair highly fragmented photogrammetric datasets.

Furthermore, in the close-range digitization of the Piffetti dressing table casket, the 3DGS approach demonstrated a distinct advantage in rendering complex material properties. Traditional mesh-based texturing often struggles with glossy finishes, either failing to reconstruct the geometry or incorrectly baking specular reflections into the static texture. In contrast, the view-dependent rendering capabilities inherent to Gaussian Splatting successfully captured the lucid, reflective interplay of light on the intricate ebony and ivory inlays, resulting in a remarkably photorealistic representation of the artifact's polished surfaces.

Nevertheless, it is crucial to clearly distinguish between visual fidelity and geometric accuracy. While the current iterations of 3D Gaussian Splatting yield exceptional results for web-based visualizations and immersive XR/VR applications, the underlying data structure is not yet mature enough for rigorous 3D metric surveying. The absence of a measurable, continuous surface topology means that these models cannot reliably support precise architectural measurements or scientific documentation.

Therefore, in the current digital heritage pipeline, 3DGS serves as a powerful engine for interactive presentation, while traditional photogrammetry and laser scanning remain strictly necessary for analytical survey tasks.

3.2 Cracks detection

A first experimental application of the proposed workflow was carried out on 16 UAV images, with a resolution of 40960x3072px, acquired using a DJI Mini 5 Pro over the Galleria di Diana in Venaria dataset. In the image domain, each image was processed through a tiled SAM3 segmentation scheme based on 35 tiles of 768 px, with 25% overlap, in order to preserve fine crack-like features while maintaining a manageable processing scale. On average, the image-based stage produced 71.2 candidate masks per image, with values ranging from 7 to 146, and an average mean score of 0.632. Overall, the extracted masks covered more than 1.13 million positive pixels, confirming the ability of the prompt-based approach to identify localized crack-like patterns within the dataset. The average processing time was 83.27 seconds per image.

The multiview reprojection step confirmed the possibility of transferring the image-based detections to object space. All 16 images were successfully used in the 3D processing stage, and the crack-related information was projected onto a mesh consisting of 705,144 vertices and 235,048 faces. In the thresholded configuration, 150 class points were obtained, representing the most reliable crack candidates on the surface. In the no-threshold configuration, 62,985 valid intersections were retrieved from 74,165 rays, corresponding to a hit rate of 84.9%, thus providing a denser spatial distribution of crack evidence.

These preliminary results confirm the feasibility of combining prompt-based image segmentation and 3D reprojection for the spatial localization of crack evidence on reconstructed surfaces, opening up further perspectives for documentation, interpretation, and monitoring.

A qualitative example of the proposed workflow is shown in Figure 10, where the detected crack masks are visualized on the original image, while the corresponding crack points projected onto the mesh highlight the spatial transfer of the image-based detections into object space.

Figure 10. Application of the crack detection workflow. Top left: original image with the detected crack masks overlaid in red. Bottom left: extracted masks. Right: crack points projected onto the 3D mesh.

4. Discussion

The results presented in this contribution offer a preliminary assessment of the three methodological pillars of the HERITALISE WP6 workflow: novel view synthesis via 3DGS, AI-driven crack detection, and standardised quality evaluation.

https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-809-2026

Regarding 3D Gaussian Splatting, the experiments across the three case studies confirm its considerable potential for photorealistic visualisation of CH assets at multiple scales. The method demonstrated a notable capacity to compensate for the geometric incompleteness inherent in traditional SfM pipelines, a limitation that became particularly evident in the visually ambiguous interior of the Galleria Grande, where repetitive architectural patterns and the absence of Ground Control Points caused systematic registration failures. In such contexts, 3DGS effectively acted as a generative completion layer, synthesising plausible surface information from available viewpoints. Similarly, the view-dependent rendering of complex material properties, such as the specular reflections of the Piffetti casket's polished inlays, represents a qualitative advantage over conventional mesh-based texturing approaches. Nevertheless, a fundamental distinction between visual fidelity and metric accuracy must be rigorously maintained: while 3DGS excels in immersive visualisation and XR/VR applications, its current data structure lacks the continuous, measurable surface topology required for scientific documentation and architectural metrology.

Currently, in the operational pipeline of HERITALISE, it therefore functions as a complementary representation layer rather than as a replacement for photogrammetric or laser scanning products.

Concerning crack detection, the two-stage workflow based on tiled SAM3 segmentation and multiview ray-based reprojection proved feasible for the spatial localisation of surface discontinuities on reconstructed meshes. The average confidence score of 0.632 per mask and the 84.9% ray–mesh intersection rate are encouraging for a zero-shot, prompt-driven approach applied to high-resolution UAV imagery, particularly given the absence of domain-specific fine-tuning. The variability in the number of candidate masks per image (ranging from 7 to 146) suggests, however, that the method's sensitivity to image content and acquisition geometry requires further investigation. The processing time of approximately 83 seconds per image also indicates that computational optimisation will be necessary for application at building or campus scale.

A key open challenge remains the transition from 2D probability maps to quantitative 3D crack characterization, encompassing geometric attributes such as width, continuity, and depth, which is essential for structural assessment purposes.

With respect to the evaluation framework, it is important to note that the quantitative assessment of the 3DGS outputs against a metrically controlled reference has not yet been carried out in this first phase of the project.

The adaptive threshold formulation introduced in Equation (1) is designed to ensure that the resulting scores remain physically meaningful and comparable across the different survey scales represented in the HERITALISE demo site.

5. Conclusions

This contribution has presented the initial results of the AI-based digitisation workflow developed within WP6 of the HERITALISE project, applied to multiscale case studies at the *Reggia di Venaria Reale*.

Three principal contributions have been demonstrated. First, 3D Gaussian Splatting has been evaluated across three acquisition scenarios, outdoor UAV photogrammetry, indoor kinematic SLAM, and close-range artefact digitization, confirming its effectiveness as a visualisation and data-completion tool, while also delineating its current limitations in metric accuracy.

Second, a two-stage crack detection pipeline combining tiled text-prompted segmentation with SAM3 and multiview 3D reprojection has been validated on real UAV imagery, producing spatially localised crack evidence on reconstructed meshes and demonstrating the feasibility of transferring image-level AI detections into object space.

Third, a structured evaluation framework has been proposed, organising geometric and radiometric quality indicators into reference-based and no-reference scenarios, underpinned by robust statistics and adaptive scale-dependent thresholds, and aggregated into a transparent synthetic quality score.

Several directions for future development are planned within the HERITALISE roadmap. The most immediate priority is the rigorous quantitative evaluation of the generated 3DGS models against TLS-derived ground-truth point clouds, applying the full suite of metrics defined in the proposed framework.

Regarding crack detection, future efforts will focus on extending the workflow to additional surface materials and damage typologies, improving computational efficiency for large-scale deployment, and developing the transition from probabilistic 2D masks to fully characterised 3D crack geometries, including width estimation and depth profiling.

Beyond these near-term objectives, the integration of non-visible data sources, including X-ray and computed tomography acquisitions, will be pursued to enable comprehensive volumetric documentation of CH objects without physical disassembly, in line with the broader HERITALISE objectives. This will be complemented by data fusion techniques capable of combining surface and sub-surface reconstructions into a unified metrically consistent model. Semantic segmentation pipelines oriented towards semi-automatic Scan-to-BIM workflows are also planned, providing structured digital representations suitable for conservation management and preventive monitoring.

Finally, the evaluation framework will be further refined and validated across a broader set of sensors, object types, and acquisition conditions, with the aim of establishing a reproducible, sensor-agnostic standard for the assessment of AI-supported CH digitisation products, contributing to the definition of internationally recognised quality guidelines currently absent from the field.

Acknowledgements

This study was carried out within the Horizon HERITALISE project. This manuscript reflects only the authors’ views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

References

Aspert, N., Santa-Cruz, D. and Ebrahimi, T., 2002, August. Mesh: Measuring errors between surfaces using the hausdorff distance. In Proceedings. IEEE international conference on multimedia and expo, Vol. 1, 705-708.

Carion, N., Gustafson, L., Hu, Y.T., Debnath, S., Hu, R., Suris, D., Ryali, C., Alwala, K.V., Khedr, H., Huang, A. and Lei, J.,

2025. Sam 3: Segment anything with concepts. arXiv preprint arXiv:2511.16719.

Chiabrando, F., Lingua, A., Spreafico, A., Sammartano, G., Matrone, F., Borras, M., Mendikute Garate, A., Miller, A., Ioannides, M., Siegkas, P. and Baker, D., 2025. HERITALISE. Project insights and initial developments. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 48, 269-276.

Clini, P., Nespeca, R., Angeloni, R. and Coppetta, L., 2024. 3D representation of Architectural Heritage: a comparative analysis of NeRF, Gaussian Splatting, and SfM-MVS reconstructions using low-cost sensors. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 48, 93-99.

Dorafshan, S., Thomas, R. J., & Maguire, M., 2018. Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. *Construction and Building Materials*, 186, 1031–1045.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. and Uszkoreit, J., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

ISO (International Organization for Standardization), 2021. Photography - Archiving systems - Imaging systems quality analysis, Part 1: Reflective originals. ISO 19264-1:2021. International Organization for Standardization, Geneva, Switzerland.

ISO (International Organization for Standardization), 2023. Geographic information - Data quality, Part 1: General requirements. ISO 19157-1:2023. International Organization for Standardization, Geneva, Switzerland.

Joint Committee for Guides in Metrology (JCGM), 2008. Evaluation of measurement data - Guide to the expression of uncertainty in measurement. JCGM 100:2008. BIPM, Sèvres, France.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y. and Dollár, P., 2023. Segment anything. In Proceedings of the IEEE/CVF international conference on computer vision, 4015-4026.

Kruse, F.A., Lefkoff, A.B., Boardman, J.W., Heidebrecht, K.B., Shapiro, A.T., Barloon, P.J., Goetz, A.F.H., 1993. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. Remote Sensing of Environment, 44(2-3), 145-163. doi.org/10.1016/0034-4257(93)90013-N.

Lague, D., Brodu, N., Leroux, J., 2013. Accurate 3D comparison of complex topography with terrestrial laser scanner: application to the Rangitikei canyon (N-Z). *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, 10-26. doi.org/10.1016/j.isprsjprs.2013.04.009.

Munda, G., Nardo, M., 2005. Constructing Consistent Composite Indicators: the Issue of Weights. EUR 21834 EN, European Commission, Joint Research Centre.

Murtiyoso, A. and Grussenmeyer, P., 2023. Initial assessment on the use of state-of-the-art nerf neural network 3d reconstruction for heritage documentation. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 48, pp.1113-1118.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffmann, A., Giovannini, E., 2005. Tools for Composite Indicators Building. EUR 21682 EN, European Commission, Joint Research Centre.

Remondino, F., Karami, A., Yan, Z., Mazzacca, G., Rigon, S. and Qin, R., 2023. A critical analysis of NeRF-based 3D reconstruction. Remote Sensing, 15(14), p.3585.

Riegler, G., Osman Ulusoy, A. and Geiger, A., 2017. Octnet: Learning deep 3d representations at high resolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, 3577-3586.

Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, Cham: Springer international publishing, 234-241.

Rusinkiewicz, S., 2004, September. Estimating curvatures and their derivatives on triangle meshes. In Proceedings. 2nd International Symposium on 3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004, pp. 486-493, IEEE.

Sharma, G., Wu, W., Dalal, E.N., 2005. The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. Color Research and Application, 30(1), 21-30. doi.org/10.1002/col.20070.

VIGIE 2020/654: European Commission: Directorate-General for Communications Networks, Content and Technology, Study on quality in 3D digitisation of tangible cultural heritage – Mapping parameters, formats, standards, benchmarks, methodologies, and guidelines – Executive summary, Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2759/581678

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4), 600-612. doi.org/10.1109/TIP.2003.819861.

Zhang, L., Yang, F., Zhang, Y.D. and Zhu, Y.J., 2016, September. Road crack detection using deep convolutional neural network. In 2016 IEEE international conference on image processing (ICIP), pp. 3708-3712, IEEE.

Zou, Q., Zhang, Z., Li, Q., Qi, X., Wang, Q. and Wang, S., 2018. Deepcrack: Learning hierarchical convolutional features for crack detection. IEEE transactions on image processing, 28(3), 1498-1512.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-809-2026 |