The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume L-4/W2-2026
21st International 3D GeoInfo Conference 2026, 28–29 September 2026, Sofia, Bulgaria

Comparison of AI-driven 3D Reconstruction Methods for Cultural Heritage Modeling

Alper Tunga Akın [1], Bura Adem Atasoy [1], Fatih Terzi[2], Ziya Usta [3]

[1] Karadeniz Technical University, Faculty of Engineering, Department of Geomatics Engineering, 61080 Ortahisar/Trabzon, Türkiye - {alpertunga; bugratasoy}@ktu.edu.tr
[2] Karadeniz Technical University, Graduate School of Natural and Applied Sciences, 61080 Ortahisar/Trabzon, Türkiye - fatihterzi@ktu.edu.tr
[3] Department of Geomatics Engineering, Artvin Coruh University, Artvin 08100, Türkiye - ziyausta@artvin.edu.tr

Keywords: 3D Reconstruction, Cultural Heritage, NeRF, 3D Gaussian Splats, SAM 3D

Abstract

Artificial intelligence-driven 3D reconstruction offers new opportunities for documenting cultural heritage under limited and low-cost image-acquisition conditions. This study compares Neural Radiance Fields (NeRF), 3D Gaussian Splatting (3DGS), and SAM 3D for reconstructing Saint Michael Church in Trabzon, Türkiye, using imagery representative of a short amateur UAV survey. From a UAV dataset of 178 images, 50 images were used to train NeRF and 3DGS, while SAM 3D was evaluated using four cardinal-view images. The methods were assessed through rendering speed, PSNR, SSIM, LPIPS, Chamfer mean, Chamfer RMSE, model size, and visual inspection against a photogrammetric reference model. 3DGS achieved the strongest overall performance among the scene-based methods, reaching 31.6 FPS, a PSNR of 29.71 dB, an SSIM of 0.920, and an LPIPS of 0.070, while also providing better average geometric proximity than NeRF. SAM 3D produced the lowest Chamfer mean and RMSE values despite using only four images. However, visual inspection revealed hallucination-like alterations in architecturally significant elements, including pillar dimensions, arch counts, and roof overhangs. The findings indicate that, in this case, low geometric distance does not necessarily imply documentary reliability. Under the tested configuration and with sufficient image overlap, 3DGS provides the most balanced solution for visually faithful and interactive heritage representation. SAM 3D remains promising for extremely sparse imagery, but its inferred geometry requires explicit verification. Future work may investigate multiview and video-based SAM 3D workflows and hallucination-resistant art-historical constraints.

1. Introduction

The preservation of cultural heritage increasingly relies on high-fidelity digital documentation and virtual reconstruction. Historically, architectural and archaeological fields have employed established methods like Structure-from-Motion (SfM) photogrammetry and Terrestrial Laser Scanning (TLS) to capture the geometry and surface textures of historical artifacts, monument complexes, and urban environments. Although these traditional approaches can provide metrically reliable reference data, they face limitations such as computational bottlenecks in generating dense point clouds, strict hardware requirements, and challenges in reconstructing surfaces with properties like high reflectivity, transparency, or homogeneous textures (Croce et al., 2024). For example, traditional photogrammetry uses SfM algorithms to reconstruct 3D geometry from 2D images by matching features across multiple views. It performs well on richly textured objects with clear features but demands substantial computational resources, labor, and expertise, especially for large-scale or complex heritage sites (Tunalıoğlu et al., 2025).

Conversely, more recent artificial intelligence (AI)-driven techniques, such as NeRF (Neural Radiance Fields) and its variants, offer promising alternatives by leveraging neural networks to implicitly represent 3D scenes, and synthesize novel views from posed images (Mildenhall et al., 2020; Remondino et al., 2023). Alternatively, 3DGS (3D Gaussian Splatting) represents a departure from the neural representation used by NeRF. It represents scenes as a set of 3D Gaussians, optimized for efficient real-time rendering and offering a balance between visual fidelity and computational efficiency (Kerbl et al., 2023; Chen & Wang, 2024). Besides, SAM 3D is a generative model for visually grounded 3D object reconstruction that predicts object geometry, texture, and layout from a single image (SAM 3D Team et al., 2025).

Consequently, the advent of these AI-driven 3D reconstruction methodologies has promised a new paradigm for spatial documentation of cultural heritage. These methods differ in their representation and intended use; NeRF and 3DGS reconstruct scenes from multiple images, whereas SAM 3D generates an object-level 3D representation from a single image.

However, the practical significance of this transition becomes more apparent when cultural heritage documentation must be conducted under constrained acquisition conditions. Professional surveys generally rely on systematically planned image networks, sufficient multi-view overlap, stable illumination, and calibrated or high-quality sensors (Nocerino et al., 2014). Such conditions cannot always be achieved for geographically remote, physically inaccessible, endangered, or rapidly deteriorating heritage assets. In these cases, the available data may instead consist of a limited number of photographs contributed by visitors or volunteers, or images acquired during a short flight using a lightweight consumer-grade UAV. Previous research has demonstrated that cultural heritage digitization can be broadened through crowdsourced photogrammetry using publicly accessible imaging devices (Ch'ng et al., 2019). Similarly, lightweight and low-cost UAV systems can provide a viable basis for architectural and archaeological documentation, although their metric reliability may be constrained by the technical characteristics of their onboard sensors (Adami et al., 2019).

The limited and irregular nature of these image collections presents a particularly demanding test for contemporary reconstruction methods. NeRF-based approaches are known to suffer from geometric ambiguity, overfitting, and rendering artifacts when trained on only a small number of views, although sparse-view regularization strategies have partially mitigated these limitations (Niemeyer et al., 2022; Yang et al., 2023). Similarly, conventional 3DGS can lose geometric consistency as the number of observations decreases, despite offering substantially faster training and rendering and benefiting from additional depth constraints in few-shot settings (Li et al., 2024). SAM 3D provides a different perspective by introducing learned geometric and appearance priors that may support object-level reconstruction when conventional multi-view constraints are insufficient. It therefore remains unclear which of these methods is most suitable when the objective is not to process an extensive professional survey but to recover a useful heritage representation from sparse, heterogeneous, and potentially non-professional imagery.

Accordingly, this study addresses the following primary research question: When a cultural heritage object is documented using only a limited number of images or a short hobby-level UAV flight, which of NeRF, 3DGS, and SAM 3D provides the most accurate, complete, and computationally feasible 3D representation? A related question is whether volunteer-contributed photographs and amateur UAV imagery, or uncalibrated and sparse image capturing in other words, can serve as viable inputs for cultural heritage reconstruction and, if so, which methodological approach is best suited to transform such data into usable digital heritage models.

To answer these questions, this paper evaluates the efficacy of various AI-driven methodologies for the three-dimensional (3D) reconstruction of cultural heritage objects, focusing on the comparative analysis of NeRF, 3DGS, and SAM 3D. Rather than assuming an extensive and professionally planned image-acquisition campaign, the investigation specifically examines their performance under data-constrained conditions representative of volunteer contributions and short amateur UAV flights. For this experiment, uncalibrated conditions were imitated by downsampling a calibrated flight in terms of overlaps and image counts. The methods are comparatively assessed in terms of geometric accuracy, visual fidelity, computational demands, and their ability to preserve heritage-relevant characteristics. Through this comparison, the study seeks to determine which technique produces the most balanced result under limited observations, and to assess the potential of reduced UAV image sets for cultural heritage documentation.

2. Study Area and Data Source

The study area is Saint Michael Church, situated in the Akçaabat district of Trabzon (Figure 1) and built by the Empire of Trebizond during the 13th–14th centuries. This asset ranks among the important cultural monuments in the historical and cultural landscape of the Black Sea region (Ballance, 1960).

The images used to produce the photogrammetric reference model were acquired during a UAV flight conducted at an altitude of 56.4 meters, achieving 80% frontal and 70% side image overlap across 178 images. The resulting dataset features a ground sampling distance (GSD) of 8.74 mm/pixel and an initial point density of 2.79 points/cm² in the resulting photogrammetric model. The UAV dataset used in this study was previously documented by Atasoy et al. (2026) and was adapted here for the comparative experiments.

For the AI-based reconstruction experiments, a subset of 50 images was selected from the complete dataset of 178 images and used to train the 3DGS and NeRF models. In contrast, the SAM 3D experiment used only four images, each capturing the test object from one of the four cardinal directions: north, east, south, and west, thus, four SAM 3D outputs were generated independently from images capturing the object from the north, east, south, and west directions.

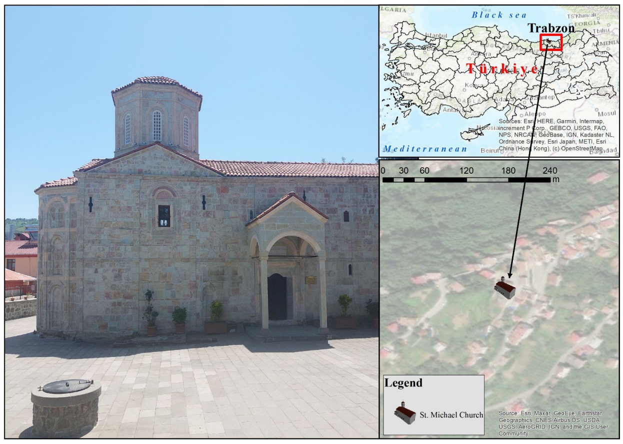

Figure 1. Location of the study object: Saint Michael Church (coordinates: 41.019° N, 39.562° E) in Akçaabat, Trabzon, Türkiye (Map from ArcGIS World Imagery, Sources: Esri, Maxar, Earthstar Geographics, and the GIS User Community).

3. Experimental Setup

To evaluate the performance of these 3D reconstruction methodologies, an experimental setup was designed, focusing on specific metrics relevant to cultural heritage documentation. NeRF, 3DGS, and SAM 3D were benchmarked using nerfstudio for NeRF and 3DGS and Meta's SAM 3D Playground for SAM 3D, using images from Saint Michael Church.

Key comparisons spanned training duration and computational efficiency (GPU/CPU hours), rendering speed for VR/AR heritage exhibits (FPS), and reconstruction quality via PSNR, SSIM, and Chamfer distances against the photogrammetric reference model.

The NeRF and 3DGS experiments were conducted on a workstation equipped with an Intel Xeon W processor, NVIDIA RTX A4 GPU support (20GB DDR6 VRAM), 32GB+ DDR4 RAM, NVMe SSD storage. The workstation ran Windows 11 with CUDA 12.1+, PyTorch 2.1, and nerfstudio in Docker containers.

4. Results

4.1. Experimental comparison

3DGS and NeRF were trained using the same subset of 50 images selected from the complete collection of 178 UAV images. SAM 3D was evaluated using only four images representing the north, east, south, and west views of Saint Michael Church. Because SAM 3D operates as a pretrained image-to-3D reconstruction model rather than a scene-specific radiance-field training framework, its training time, rendering speed, and image-based rendering metrics were not directly comparable with those of 3DGS and NeRF.

https://doi.org/10.5194/isprs-archives-L-4-W2-2026-9-2026 |

The training durations of 3DGS and NeRF were different under the tested settings. 3DGS completed training in approximately 23 min 30 sec, whereas NeRF required approximately 33 min 05 sec. NeRF therefore required 9 min 35 sec longer than 3DGS. Besides, a marked difference emerged during rendering. 3DGS achieved 31.61 FPS, compared with only 0.33 FPS for NeRF, corresponding to an approximately 96-fold rendering-speed advantage. Thus, only the 3DGS implementation reached a frame rate suitable for interactive navigation and potentially real-time visualization.

The quantitative results of the experiments, including training times, rendering speeds, image-quality metrics, and geometric accuracy parameters for all three tested methods, are comprehensively summarized in Table 1.

4.2. Image-based reconstruction quality

3DGS produced better values than NeRF in the image-based evaluation. It achieved a PSNR of 29.71 dB, compared with 17.40 dB for NeRF, representing a difference of 12.32 dB. The SSIM value similarly increased from 0.429 for NeRF to 0.920 for 3DGS. This indicates that the images rendered by 3DGS showed greater structural and local similarity to the reference views.

The LPIPS score, for which lower values indicate greater perceptual similarity, was 0.070 for 3DGS and 0.255 for NeRF. 3DGS therefore produced an approximately 72.5% lower perceptual error. Taken together, the PSNR, SSIM, and LPIPS results show that 3DGS produced better visual-based results than NeRF when reconstructing the church from the limited 50-image dataset. The advantage was not confined to rendering speed; it also extended to the visual consistency of the reconstructed scene. The 3DGS checkpoint occupied approximately 295.65 MB, whereas the NeRF checkpoint required 167.94 MB. 3DGS therefore incurred an approximately 76% larger storage requirement. This result indicates a trade-off between storage efficiency and practical visualization performance: NeRF produced a smaller checkpoint, but this benefit was accompanied by substantially lower rendering speed and poorer image-quality metrics.

4.3. Geometric comparison

The geometric evaluation produced a different ordering from the image-based comparison. SAM 3D achieved the lowest Chamfer mean, at 0.117, followed by 3DGS at 0.167 and NeRF at 0.180. Relative to 3DGS and NeRF, the SAM 3D Chamfer mean was approximately 29.7% and 35.0% lower, respectively. SAM 3D also obtained the lowest Chamfer RMSE, with a value of 0.872, compared with 0.965 for 3DGS and 0.960 for NeRF.

Between the two radiance-field methods, 3DGS achieved a Chamfer mean approximately 7.5% lower than that of NeRF. However, its Chamfer RMSE was slightly higher. This combination suggests that 3DGS provided better average geometric proximity, while the two methods remained similar in terms of larger geometric deviations. The marginally higher RMSE of 3DGS may indicate that a limited number of reconstructed areas contained greater local errors despite its better overall average.

Method	Input images	**Training time**	**FPS**	**PSNR**	**SSIM**	**LPIPS**	**Evaluated Points**	**Chamfer mean**	**Chamfer RMSE**	**Model size**
3DGS	50	23 min 30 sec	31.615	29.714	0.920	0.070	77,926	0.167	0.965	295.65 MB
NeRF	50	33 min 05 sec	0.328	17.398	0.429	0.255	100,000	0.180	0.960	167.94 MB
SAM 3D	4	N/A	N/A	N/A	N/A	N/A	65,604	0.117	0.872	N/A

Table 1. Evaluation metrics of the experiment.

The geometric result obtained by SAM 3D is notable because it was produced from only four cardinal-view images. Nevertheless, this outcome should not be interpreted as demonstrating that SAM 3D was categorically superior to the other methods. SAM 3D and the radiance-field methods performed different reconstruction tasks and were evaluated using different numbers of input images. Moreover, PSNR, SSIM, LPIPS, FPS, and scene-training duration were unavailable for SAM 3D. Its performance can therefore be compared with the other methods only in terms of the reported point-cloud distance metrics.

https://doi.org/10.5194/isprs-archives-L-4-W2-2026-9-2026

The geometric comparisons were conducted using a maximum of 100,000 reference points. The evaluated prediction point counts also differed: 77,926 points were used for 3DGS, 100,000 for NeRF, and 65,604 for SAM 3D. These differences should be considered when interpreting the Chamfer results because point density and spatial sampling can affect point-to-point distance metrics.

4.4. Visual Interpretation

Figure 2 presents the outputs generated by each reconstruction technique alongside the photogrammetric reference model from five viewpoints: north, south, east, west, and top. Although SAM 3D achieved the lowest Chamfer mean and RMSE values, the visual assessment does not indicate the same level of architectural fidelity. This discrepancy can be attributed to the prior-driven and generative nature of the method. The predominantly planar and orthogonal wall surfaces of the church can maintain relatively small point-to-point distances even when individual architectural elements are incorrectly proportioned, simplified, or inferred.

Several localized discrepancies are visible in the SAM 3D reconstruction. The two pillars framing the northern entrance appear thicker than those in the photogrammetric reference model and the outputs produced by 3DGS and NeRF. On the southern elevation, the bright section beneath the nave dome is represented by three arches, whereas five arches are visible in the reference model and preserved by the other two techniques. The roof overhang surrounding the nave dome is also slightly enlarged in the SAM 3D output, while 3DGS and NeRF more closely maintain the corresponding distance from the wall surfaces.

SAM 3D additionally produced a more complete representation of the rear elevation, particularly in the west view, than the other methods. However, this completion does not necessarily originate from direct visual evidence. It reflects the method's ability to infer unobserved geometry from the visible parts of the object and from geometric priors learned during pretraining. Although this capability improves apparent completeness, it can also introduce hallucination-like effects, as demonstrated by the incorrectly reconstructed arches, enlarged roof overhangs, and altered pillar dimensions. The visual comparison, therefore, shows that a geometrically plausible and numerically close reconstruction may still contain architecturally significant inaccuracies.

5. Discussion

The experimental results indicate that geometric accuracy, visual fidelity, and practical rendering performance represent distinct dimensions of reconstruction quality. Consequently, the method with the lowest point-cloud distance cannot automatically be considered the most reliable technique for cultural heritage documentation.

Among the two scene-based reconstruction methods trained with the same 50-image subset, 3DGS provided the strongest overall performance. It achieved a PSNR of 29.71 dB, an SSIM of 0.920, and an LPIPS score of 0.070, compared with 17.40 dB, 0.429, and 0.255, respectively, for NeRF. 3DGS also rendered the reconstructed scene at approximately 31.6 FPS, whereas NeRF achieved only 0.33 FPS. These findings indicate that 3DGS produced a substantially more visually consistent representation while also supporting interactive navigation. The geometric comparison produced a less straightforward ranking. SAM 3D achieved the lowest Chamfer mean and RMSE, with values of 0.117 and 0.872, respectively. 3DGS obtained a Chamfer mean of 0.167 and an RMSE of 0.965, while the corresponding NeRF values were 0.180 and 0.960. Nevertheless, the visual assessment revealed that SAM 3D altered several architecturally meaningful components. The relatively low Chamfer values may therefore reflect the accurate recovery of the church's coarse massing, including its approximately orthogonal walls and overall volume, rather than the faithful reconstruction of individual details.

This distinction is particularly important in cultural heritage applications. Chamfer distance primarily measures the spatial proximity between two-point sets. It does not determine whether the correct number of arches was reconstructed, whether a pillar retains its original proportions, or whether a roof overhang corresponds to the observed structure. A method may therefore produce a globally plausible surface while modifying features that are historically, architecturally, or diagnostically important. For this reason, geometric distance metrics should be complemented by visual inspection, feature-level comparison, and, where possible, expert architectural assessment.

The completion of the rear elevation by SAM 3D illustrates both the benefit and the risk of generative reconstruction. When only a few views are available, learned priors can provide surfaces for regions that are absent or weakly represented in the input images. This capability may be useful for visualization, preliminary interpretation, or the rapid creation of approximate models. However, the completed regions should not be regarded as measured documentation unless they can be confirmed using additional images, survey data, historical records, or expert examination. In heritage documentation, inferred geometry should ideally be identified and stored separately from directly observed geometry.

The SAM 3D experiment also used four deliberately selected cardinal views rather than an uncontrolled collection of crowdsourced photographs. Its results therefore demonstrate the potential of a minimal but spatially structured image set, not the reliability of arbitrary crowdsourced imagery. The north, south, east, and west views provided broad coverage of the external form and likely contributed to the recovery of the church's overall massing. Images contributed by visitors may contain more severe occlusion, repeated viewpoints, inconsistent framing, variable illumination, and incomplete coverage.

NeRF was less effective under the tested 50-image configuration. Its image-based metrics indicate that the available views were insufficient for reproducing the appearance of the church with the consistency achieved by 3DGS. Its low rendering speed also limited, under the tested settings, its suitability for interactive heritage presentation. However, this finding applies to the tested implementation, dataset, and configuration and should not be generalized to every NeRF variant. Sparse-view regularization, depth supervision, alternative camera trajectories, or additional optimization may improve NeRF performance.

Figure 2. Visual samples from experiments and the photogrammetric reference model.

In relation to the primary research question, the results suggest that 3DGS is the most balanced option when a short amateur UAV flight can provide a moderately sized set of overlapping images. It combines high visual fidelity, reasonable geometric proximity, and interactive rendering performance. When only a very small number of well-distributed images are available, SAM 3D can produce a complete and geometrically plausible initial representation. Nevertheless, its inferred components require verification because apparent completeness may be achieved through generative reconstruction rather than direct observation.

The findings therefore do not identify a universally superior method. Instead, they indicate a task-dependent selection. 3DGS appeared more suitable for scene-level visualization and digital presentation based on a short image sequence, whereas SAM 3D appeared more appropriate for generating an approximate proxy from extremely sparse observations. A combined workflow may also be considered, in which 3DGS represents the observed scene and SAM 3D is used only to propose possible geometry for insufficiently captured regions. Any generatively completed component should, however, remain explicitly labelled as inferred rather than documented.

6. Limitations

This study has several limitations that constrain the generalization of its findings. First, all experiments were conducted on a single cultural heritage object. Saint Michael Church contains a particular combination of masonry textures, façade articulation, roof geometry, surrounding vegetation, terrain, and occlusions. The observed ranking may change for monuments with reflective materials, homogeneous walls, intricate sculptures, extensive interior spaces, severe deterioration, or substantially different scales.

The image dataset did not represent a truly uncontrolled crowdsourced or social-media collection. The images originated from the same heritage documentation context and were selected for the experiments. Consequently, variability in camera models, image resolution, compression, exposure, illumination, weather, capture dates, and contributor behaviour was not systematically examined. The study therefore evaluates data-constrained acquisition rather than fully crowdsourced reconstruction.

The prediction point clouds contained different numbers and distributions of points. Although point sampling was limited to a maximum of 100,000 points, the 3DGS and SAM 3D outputs contained fewer usable points than the NeRF output. Since Chamfer distance is affected by point density and sampling distribution, repeated uniform or surface-area-weighted resampling should be used in future experiments.

The model file size was available only for 3DGS and NeRF and should not be interpreted as total computational memory consumption. Runtime GPU memory, peak VRAM, energy use, preprocessing duration, export duration, and hardware-specific performance were not recorded. Editability, user interaction, web-delivery latency, and the effectiveness of inpainting or component manipulation were also not experimentally measured. Claims concerning these properties should therefore be reserved for further study rather than presented as findings of the current experiment.

7. Conclusion

This study investigated whether AI-driven reconstruction methods can produce useful cultural heritage representations from a limited number of images representative of a short, low-cost, or hobby-level acquisition campaign. NeRF, 3DGS, and SAM 3D were evaluated using imagery of Saint Michael Church and compared in terms of visual quality, rendering performance, geometric proximity, training duration, and model size, where the corresponding metrics were available.

Among the two scene-level radiance-field methods, 3DGS produced the strongest overall performance. It substantially outperformed NeRF in PSNR, SSIM, LPIPS, and rendering speed, while also obtaining a lower mean Chamfer distance. Its approximately 31.6 FPS rendering rate demonstrates that a 3DGS-based representation can support interactive heritage visualization even when trained from a reduced 50-image UAV dataset. The cost of this performance was a larger checkpoint size, although 3DGS completed training 9 min 35 sec faster than NeRF.

SAM 3D achieved the lowest Chamfer mean and RMSE using only four cardinal-view images. This result indicates the potential of pretrained image-to-3D models for producing approximate geometric representations when multi-view coverage is extremely limited. Nevertheless, the absence of comparable rendering and image-quality metrics and the potential contribution of learned geometric priors prevent SAM 3D from being identified as the overall best-performing method.

The findings therefore lead to a conditional rather than universal answer to the research question. For a reduced UAV image set that provides several overlapping views, 3D Gaussian Splatting is the most balanced of the tested methods for generating a visually convincing, geometrically reasonable, and interactively renderable heritage representation. When only a very small number of well-distributed photographs is available, SAM 3D can provide a promising initial geometric proxy, but its inferred regions must be treated cautiously and verified against observed evidence. Under the current sparse-input configuration, NeRF was less suitable because it combined low rendering speed with weaker visual-quality results.

The study consequently supports the use of amateur and potentially volunteer-contributed imagery, provided that the images retain adequate spatial distribution and coverage. However, the viability of unrestricted crowdsourced data has not yet been demonstrated. Future research should test these findings across multiple heritage objects and actual crowdsource image collections. Controlled experiments should vary the number, spatial distribution, quality, and overlap of the input images, while combining global distance metrics with component-level dimensional checks, semantic correctness, architectural feature counts, and expert assessment. Particular attention will be given to extending SAM 3D toward multi-view reconstruction workflows in which outputs derived from multiple viewpoints are jointly constrained and integrated rather than generated independently. The study will also investigate methods capable of producing 3D models from temporally continuous inputs, such as short videos or video-like image sequences, to determine whether sequential visual evidence can improve geometric consistency and reduce uncertainty in occluded regions. In parallel, the work will explore the development of hallucination-resistant, art-historical constraints that incorporate prior knowledge about architectural typologies, stylistic conventions, construction logic, symmetry, and historically plausible component arrangements. Such constraints could help distinguish defensible geometric completion from visually plausible but historically unsupported generation, thereby improving the documentary reliability of AI-based cultural heritage reconstruction.

Acknowledgements

The authors would like to thank the Scientific and Technological Research Council of Türkiye (TÜBİTAK) for the financial support provided under the BİDEB 2224-A Travel Fellowship for International Scientific Events Program (Grant number: 1919B022606051).

References

Adami, A., Fregonese, L., Gallo, M., Helder, J., Pepe, M., & Treccani, D. (2019). Ultra light UAV systems for the metrical documentation of cultural heritage: Applications for architecture and archaeology. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W17*, 15–21. https://doi.org/10.5194/isprs-archives-XLII-2-W17-15-2019

Atasoy, B. A., Akın, A. T., Terzi, F., & Yuka, A. (2026). A digital layer of memory: 3D documentation, art historical interpretation, and web-based presentation of the Saint Michael Church. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48, 57-61. https://doi.org/10.5194/isprs-archives-XLVIII-4-W18-2025-57-2026

Ballance, S. (1960). The Byzantine churches of Trebizond. *Anatolian Studies, 10*, 141–175. https://doi.org/10.2307/3642433

Chen, G., & Wang, W. (2024). A survey on 3D Gaussian Splatting. *arXiv*. https://doi.org/10.48550/arXiv.2401.03890

Ch'ng, E., Cai, S., Zhang, T. E., & Leow, F.-T. (2019). Crowdsourcing 3D cultural heritage: Best practice for mass photogrammetry. *Journal of Cultural Heritage Management and Sustainable Development, 9*(1), 24–42. https://doi.org/10.1108/JCHMSD-03-2018-0018

Croce, V., Forleo, G., Billi, D., Bevilacqua, M. G., Piemonte, A., & Caroti, G. (2024). Neural Radiance Fields (NeRF) for multi-scale 3D modeling of cultural heritage artifacts. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-2/W4-2024*, 165–171. https://doi.org/10.5194/isprs-archives-XLVIII-2-W4-2024-165-2024

Fassi, F., Fregonese, L., Ackermann, S., & De Troia, V. (2013). Comparison between laser scanning and automated 3D modelling techniques to reconstruct complex and extensive cultural heritage areas. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-5/W1*, 73–80. https://doi.org/10.5194/isprsarchives-XL-5-W1-73-2013

Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3D Gaussian Splatting for real-time radiance field rendering. ACM *Transactions on Graphics*, 42(4), 139-1. https://doi.org/10.1145/3592433

Li, J., Zhang, J., Bai, X., Zheng, J., Ning, X., Zhou, J., & Gu, L. (2024). DNGaussian: Optimizing sparse-view 3D Gaussian radiance fields with global-local depth normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 20,775–20,785).

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). NeRF: Representing scenes as neural radiance fields for view synthesis. In Computer Vision – ECCV 2020 (pp. 405–421). https://doi.org/10.1007/978-3-030-58452-8_24

Niemeyer, M., Barron, J. T., Mildenhall, B., Sajjadi, M. S. M., Geiger, A., & Radwan, N. (2022). RegNeRF: Regularizing Neural Radiance Fields for view synthesis from sparse inputs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5,480–5,490). https://doi.org/10.1109/CVPR52688.2022.00540

Nocerino, E., Menna, F., & Remondino, F. (2014). Accuracy of typical photogrammetric networks in cultural heritage 3D modeling projects. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-5*, 465–472. https://doi.org/10.5194/isprsarchives-XL-5-465-2014

Remondino, F., Karami, A., Yan, Z., Mazzacca, G., Rigon, S., & Qin, R. (2023). A critical analysis of NeRF-based 3D reconstruction. *Remote Sensing, 15*(14), 3585. https://doi.org/10.3390/rs15143585

Tunalıoğlu, N., Erdoğan, B., Öcalan, T., Arıcan, D., Altuntaş, C., Arda, T., Doğan, A. H., & Arıkan, E. Z. (2025). Reconstruction of historical buildings using social media data: A case study from the World Cultural Heritage List. *International Journal of Engineering and Geosciences, 11*(1), 21–39. https://doi.org/10.26833/ijeg.1635501

Yang, J., Pavone, M., & Wang, Y. (2023). FreeNeRF: Improving few-shot neural rendering with free frequency regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8,254–8,263). https://doi.org/10.1109/CVPR52729.2023.00798

SAM 3D Team, Chen, X., Chu, F.-J., Gleize, P., Liang, K. J., Sax, A., Tang, H., Wang, W., Guo, M., Hardin, T., Li, X., Lin, A., Liu, J., Ma, Z., Sagar, A., Song, B., Wang, X., Yang, J., Zhang, B., Dollár, P., Gkioxari, G., Feiszli, M., & Malik, J. (2025). SAM 3D: 3Dfy Anything in Images. arXiv. https://doi.org/10.48550/arXiv.2511.16624