

NEURAL RADIANCE FIELDS (NeRF) FOR MULTI-SCALE 3D MODELING OF CULTURAL HERITAGE ARTIFACTS

V. Croce^{1,*}, G. Forleo², D. Billi³, M. G. Bevilacqua², A. Piemonte³, G. Caroti³

¹ LISPEN EA 7515, Arts et Métiers Institute of Technology, 13100 Aix-en-Provence, France – valeria.croce@ensam.eu

² Dept. of Energy, Systems, Territory and Construction Engineering, University of Pisa, 56122 Pisa, Italy – marco.giorgio.bevilacqua@unipi.it

³ Dept. of Civil and Industrial Engineering, University of Pisa, 56122 Pisa, Italy – d.billi1@studenti.unipi.it, gabriella.caroti@unipi.it, andrea.piemonte@unipi.it

Commission II

KEY WORDS: NeRF, Photogrammetry, AI, Cultural Heritage, 3D Survey, Neural Radiance Fields.

ABSTRACT:

This research aims to assess the adaptability of Neural Radiance Fields (NeRF) for the digital documentation of cultural heritage objects of varying size and complexity. We discuss the influence of object size, desired scale of representation, and level of detail on the choice to use NeRF for cultural heritage documentation, providing insights for practitioners in the field. Case studies range from historic pavements to architectural elements or buildings, representing diverse and multi-scale scenarios encountered in heritage documentation procedures. The findings suggest that NeRFs perform well in scenarios with homogeneous textures, variable lighting conditions, reflective surfaces, and fine details. However, they exhibit higher noise and lower texture quality compared to other consolidated image-based techniques as photogrammetry, especially in case of small-scale artifacts.

1. INTRODUCTION

First introduced in 2020 (Mildenhall et al., 2020), Neural Radiance Fields (NeRF) allow for novel view synthesis and neural rendering from a given set of input images, properly acquired from multiple views of a same object. Since relying on Neural Networks, specifically on a Multi-Layer Perceptron (MLP), NeRF fall within the subset of Deep Learning algorithms. The basic concept is akin to photogrammetric surveying, since relying on the construction of a 3D model from multiple images of the same object.

The application of NeRF in the cultural heritage field, even though little explored so far, could streamline various processes related to digital documentation and preservation (Pepe et al., 2023), significantly benefiting efforts in preservation and conservation within this realm (Murtiyoso and Grussenmeyer, 2023).

In this study, we delve into the capabilities and limitations of NeRF for the 3D modeling of multi-scale heritage artifacts. Our goal is to evaluate the method across various potential scales of representation, specifically for cultural heritage documentation and dissemination purposes, and to compare its performance with that of photogrammetry.

2. RELATED WORK

A NeRF is a neural network used to synthesize novel views of complex 3D scenes by optimizing an underlying continuous volumetric scene function based on an initial set of sparse views (Tancik et al., 2023). MLPs are trained to generate 3D objects from two-dimensional images. A 5D-vector function is taken as input: the MLP takes a 5D coordinate, comprising spatial coordinates within the scene (x, y, z) and two angles, azimuthal and polar, defining the viewing direction (θ, ϕ). It outputs a volume density, denoted as σ , and an RGB color, dependent on the viewing direction.

Several variations on the original algorithm aimed at improving either inference speed alone or both training and inference speed. Notable implementations include the introduction of the Multi-resolution hash encoding technique for instant neural graphics primitives (Müller et al., 2022) and advancements like view synthesis over dynamic scenes (Attal et al., 2021), depth estimation (Li et al., 2021), and the incorporation of latent vectors -the latent conditional NeRFs by Liu et al. (2021), to control 3D scene composition, shape, and appearance.

The theoretical and operational principles of NeRF have been extensively reviewed by Gao et al. (2022), and their existing applications in cultural heritage have been previously described in (Croce et al., 2023), encompassing the potential for neural 3D reconstruction of in-the-wild scenes (Martin-Brualla et al., 2021), semantic structuring of models (Zhi et al., 2021), and rendering for Virtual and Extended Reality (Li et al., 2022).

Remondino et al. (2023) proposed various evaluation metrics, including noise level, surface deviation, geometric accuracy, and completeness, to assess NeRF-based reconstruction methods. However, the adaptability and flexibility of use of NeRF to represent cultural objects of varying size and complexity remains an open question (Croce et al., 2024). The issue is crucial in assessing the applicability of NeRF, as its effectiveness could be influenced by size and complexity of the objects being examined: small sculptures, artworks or architectural details, compared to larger monuments as historical buildings or archaeological sites.

3. CASE STUDIES

The selected cases aim to respond to a sufficiently diverse sample of scenarios that might be encountered in common heritage documentation practices. They are presented in Table 1 in descending order of expected level of detail and scale of representation. The table also indicates potential challenges that may arise from object characteristics, contextual factors, and acquisition schemes.

* Corresponding author



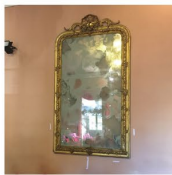
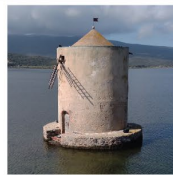

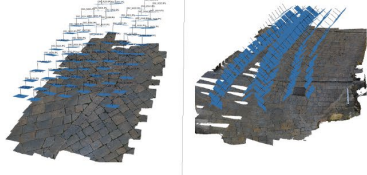


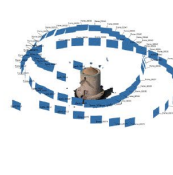
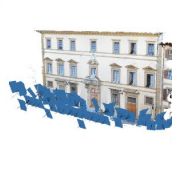
n.	1	2	3	4	5
Case study					
Acquisition Scheme					
Scale	1 : 5	1 : 10 / 1 : 20		1 : 50 / 1 : 100	
Objectives	Accurate description of the details and texture of the individual elements that make up the pavements. Detection of any irregularities in the "Volterra bench" stone and monitoring of its state of preservation.	Accurate description of the details and colorimetric characteristics of the object as whole. Legibility of the raised decorative apparatus and description of degraded areas.		Global description of the geometries, colorimetric features and texture of the objects. Recognizability of component elements, represented with a moderate level of detail.	

Table 1. Summary of the different case studies considered with related scale of representation.

Specifically, the case studies include:

- Two pavement laying patterns in Volterra, of Piazza dei Priori and of the ascent of Porta all'Arco, respectively,
- the 19th-century stove from the Aula Magna of Palazzo Boileau in Pisa,
- the 19th-century mirror from the Aula Magna of Palazzo Boileau in Pisa,
- the Spanish Mill of the Orbetello Lagoon,
- the façade of Palazzo Boileau in Pisa.

The Spanish Mill dataset is acquired by a DJI Mini 2 drone, while all other datasets are captured using fixed focal length cameras, positioned close to the subject of interest. They exhibit a range of required representation scales, determined by pre-defined requirements for redesign and description.

For the historical pavement laying patterns of Piazza dei Priori and the ascent of Porta all'Arco in Volterra, images are taken at a distance of about 60-70 cm from the ground; a representation scale of 1:5 is required to properly describe the texture of the individual stones as well as the irregularities of the material that compose the paving.

The stove and mirror are captured at a distance of ~1m, rotating around the object by around 160-170°; for the façade of Palazzo Boileau, the shooting distance is 3-4 m. In these two cases, a scale of at least 1:10-1:20 is demanded: for the legibility of the raised decorative apparatus in the case of the stove, and for the identification of the degraded areas of the reflective surface and the description of the frame decoration in the case of the mirror. The Spanish mill and the façade of Palazzo Boileau require a more global description of geometric and chromatic features of the object, hence a lower level of detail, at a 1:50 or 1:100 scale.

4. METHODS

We concentrate on elucidating the specific scenarios in which NeRF outperforms photogrammetry, emphasizing its strengths in rendering. The process starts with the acquisition of a series of

images, taken from various viewpoints of the same object with sufficient overlap between successive shots. These data undergo dual processing, employing both photogrammetry (via Agisoft Metashape) and NeRF construction (via Nerfstudio), following the method previously described in (Croce et al., 2024). Camera orientation parameters are derived in both cases through alignment on Metashape. Post-processing involves exporting point clouds and polygonal meshes for both photogrammetry and NeRF, with geo-referencing based on topographic surveying. Metric and visual comparisons are then performed using cloud-to-cloud or cloud-to-mesh techniques.

5. RESULTS AND DISCUSSION

Processing times are evaluated considering for each dataset the photogrammetric and NeRF-based processing (Figure 1).

The photogrammetric process includes alignment, dense point cloud and textured mesh generation; the NeRF processing comprises alignment, training and web viewer navigation, point cloud and mesh export.

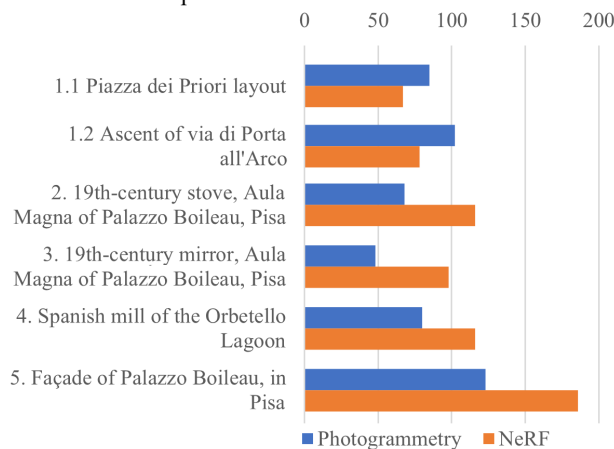


Figure 1. Processing times for the various case studies, presented in minutes.

4.1 Historical pavements in Volterra

A scale of 1:5 is required to accurately capture the details of individual paving stones, patterns and decorative inlays (Caroti, Piemonte & Ulivieri, 2023). In the case of the Piazza dei Priori

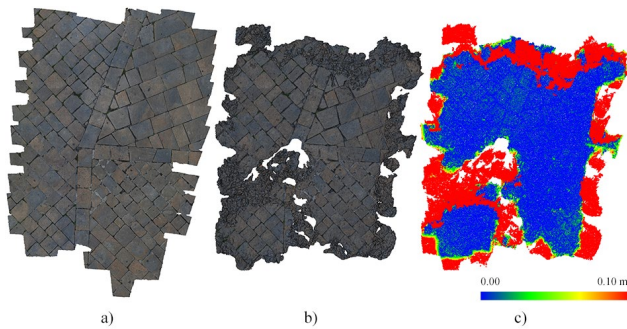


Figure 2. Piazza dei Priori layout: photogrammetric (a) and NeRF (b) meshes. Comparison between photogrammetric (reference) and NeRF point clouds (c).

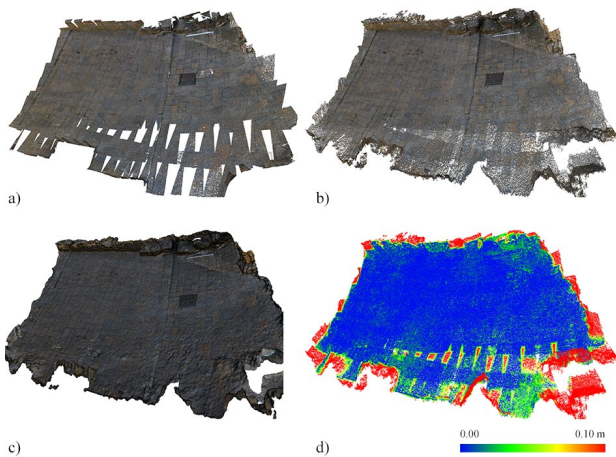


Figure 3. Ascent of via di Porta all'Arco: photogrammetric point cloud (a) and NeRF point cloud (b); the mesh generated by the NeRF (c) and cloud-to-cloud comparison (d).

layout, the NeRF model appears incomplete and exhibits lower texture quality, if compared to the photogrammetric model. Excessive noise is observed in both the central and peripheral parts of the scene, which is clearly highlighted in the cloud-to-cloud comparison (Figure 2). In the case of the ascent to the Porta all'Arco, on the other hand, the NeRF model showcases its strength at a scale of 1:5. It fills gaps in the photogrammetric model, attributed to a lower quantity of input images covering specific portions of the scene. The cloud-to-cloud comparison reveals the above-described differences (Figure 3).

4.2 19th-century stove, Aula Magna of Palazzo Boileau, Pisa

For the stove case, a scale of 1:10 or 1:20 would offer an accurate depiction of the object's details and colorimetric characteristics. The intricate details of the mouldings and the legibility of the raised decorative apparatus (relief decorative elements, upper cup) would significantly benefit from such a scale. With a cylindrical structure of diameter 0.8 m and overall height of 2.4 m, embedded within a niche, the stove is externally coated with *stucco*: the challenges in processing the data stem from the homogeneity of the material and the sub-optimal lighting conditions in the room. The non-neutral artificial lighting, coupled with the partial embedding of the object in the niche, complicates the data collection process.

The evaluation featuring a cloud-to-cloud comparison reveals that the NeRF model outperforms the photogrammetric model, especially in regions with inadequate lighting conditions behind the object, given its partial containment within the niche (Figure 4). At a scale of 1:10 or 1:20, the NeRF model excels in rendering view-dependent variations as evidenced by volumetric rendering. Such a scale would enhance the extreme realism of the stucco decorations on the stove achieved by NeRF, providing an accurate representation of the ornate features. However, the moderate noise of the generated point cloud, while not prohibitive, impacts mesh generation negatively (Figure 5).

4.3 19th-century mirror, Aula Magna of Palazzo Boileau, Pisa

The vertical planar mirror, nearly rectangular in shape with dimensions of 1.8×2.9m, is enclosed on the sides by a richly decorated frame. The reflectivity of the mirror, the gilding of the wooden frame, and the poor lighting conditions inside the room,

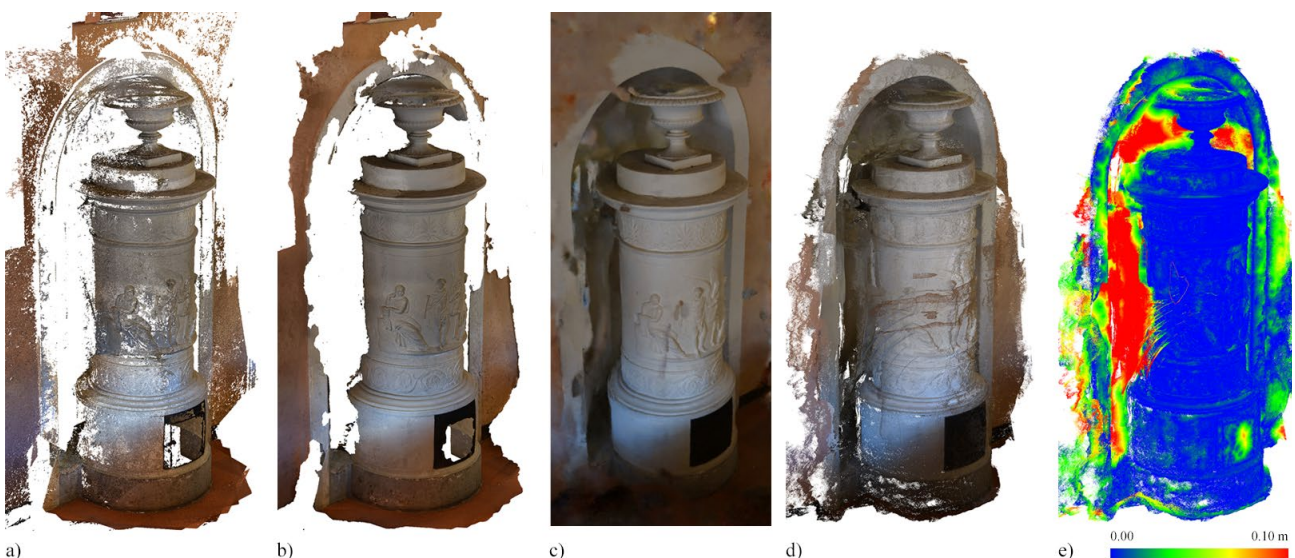


Figure 4. Stove: photogrammetric point cloud (a) and mesh (b); NeRF rendering (c) and NeRF-based point cloud (d); cloud-to-cloud comparison (e).



Figure 5. Stove details from the photogrammetric mesh (a) and the volumetric rendering generated by the NeRF (b).

exacerbated by non-neutral artificial lighting, pose serious challenges for data processing. These challenges, only partially mitigated by visible degradation signs on the mirrored surface serving as fixed reference points, underscore the complexity of capturing this object at a scale of 1:10 or 1:20.

Despite the inherent difficulties, the NeRF model results are geometrically more complete than the photogrammetric model, although notable gaps exist corresponding to sources of reflected light on the mirrored surface (a chandelier and two windows). Moreover, excessive noise and geometric errors are noted, such as the division or duplication (splitting) of parts of the model. For comparison, the photogrammetric output is shown in black and white with a red outline, while dark grey regions highlight missing areas in both models (Figure 6).

In terms of texture, the result achieved through photogrammetric processing appears valid (except for the lower part of the object), but a large portion of the mirror surface is not rendered. The NeRF model provides a less accurate description of the details of the outer frame, due to noise, but reveals some portions of the glass surface not captured in the photogrammetric model. From the comparison of Figure 6-c, overlaying the two results, it is evident that some areas (highlighted in dark grey) remain undescribed by either survey method.

4.4 The Spanish mill of the Orbetello Lagoon

The mill, featuring a diameter of approximately 6 m and a total height of 9.4 m, incorporates a system of two wooden blades, with only 2 remaining out of the original 4. Situated in a lagoon and encircled by water, the reflective nature of the surrounding body complicates processing. At a scale of 1:50 or 1:100, the point cloud extracted from the NeRF rendering, despite its overall completeness, inaccurately includes a substantial extra portion of the water mirror, not corresponding to reality. The reflective properties of the water introduce noise, impacting the geometric accuracy of the reconstruction in the immediate vicinity of the object (Figure 7).

In line with previously shown case studies, the texture quality of the NeRF model is limited compared to photogrammetry, especially if considering a 1:50 scale. However, lighting conditions and reflections are faithfully represented in the volumetric rendering.

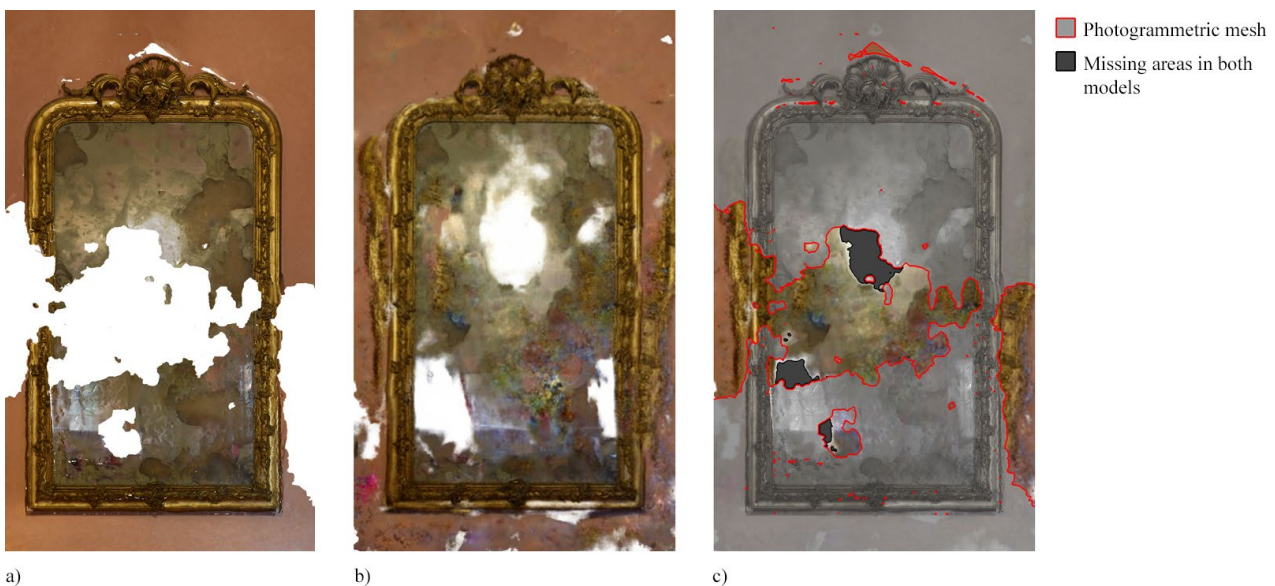


Figure 6. Mirror: photogrammetric mesh (a), the volumetric rendering generated by NeRF (b); superimposition of the two outputs (c).

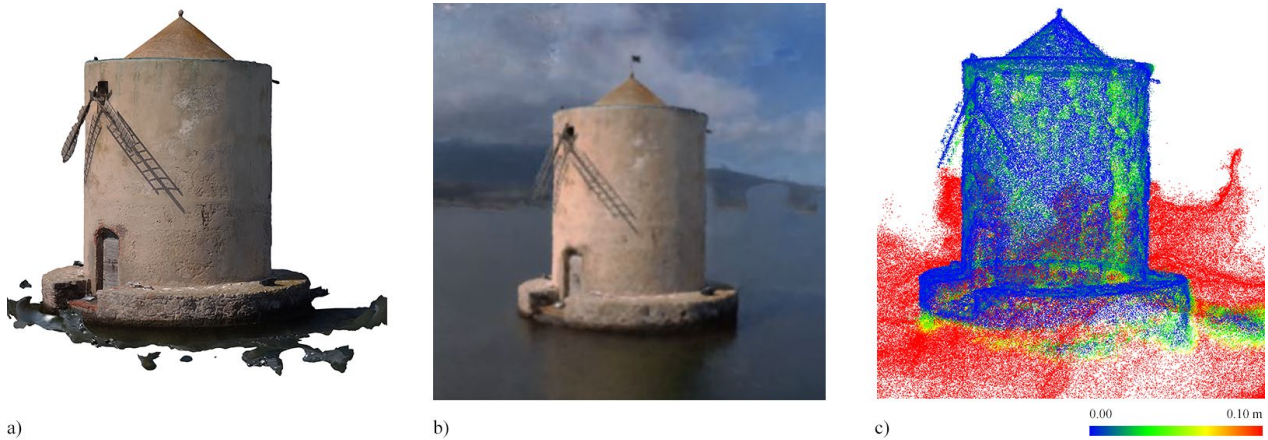


Figure 7. Spanish mill: photogrammetric mesh (a), volumetric rendering (b) and cloud-to-cloud comparison between the two point clouds (c).

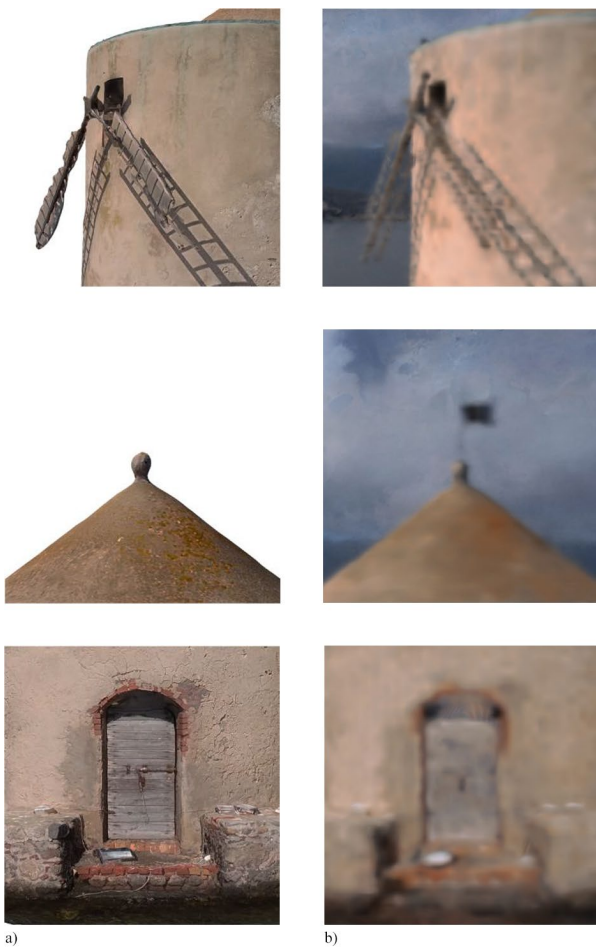


Figure 8. Spanish mill details from the photogrammetric mesh (a) and the volumetric rendering generated by the NeRF (b).

Additionally, the NeRF rendering proves to be comprehensive, capturing intricate details such as the mill blades and the flag at the top.

Notably, the flag, omitted in the generation of the photogrammetric mesh, consistently appears in NeRF outputs, whether in volumetric rendering or derived outputs, albeit partially in the latter case (Figure 8).

4.5 The façade of Palazzo Boileau, in Pisa

Dating back to the late sixteenth century, the symmetrical façade almost 30 m long is divided into three levels, featuring a double stringcourse and sill. The plaster is white and the decorative elements are crafted from *pietra serena*, a type of sandstone. The object itself poses no significant issues in terms of dataset processing.

At a scale of 1:50 or 1:100, the NeRF model proves to be more complete than its photogrammetric counterpart, despite being noisier. The photogrammetric point cloud, in fact, displays gaps in areas over protruding elements, near reflective window glazing, and at the poorly lit and glazed entrance. Filling these gaps during mesh generation introduces geometry errors, particularly around window frames, sills and at the junction between the tympanum and plastered surfaces. In contrast, NeRF's attempt to approximate these unseen regions introduces noise in both the volumetric rendering and the derived outputs. In addition, a lower texture quality hinders the legibility of fine details. The NeRF model may thus prove to be better suited to a 1:100 scale representation, where the demand for intricate details on the façade is less pronounced.

4. DISCUSSION

From the conducted tests, several noteworthy considerations can be derived, offering guidance to heritage experts considering the adoption of NeRF either as an alternative or in combination with photogrammetry. Specifically, with respect to the chosen representation scales:

- **At a representation scale of 1:5**, the NeRF model demonstrates reduced accuracy in capturing fine details. Although the texture quality is comparatively lower, NeRF proves valuable in delineating less visible regions of the input images, especially where reconstruction is conducted from a limited number of images. In other words, while the performance may be less precise in intricate details, NeRF's utility becomes pronounced in reconstructing portions with limited visual data or obscured visibility.
- **At a scale of 1:10 or 1:20**, the representation of the object's characteristics and intricate details, particularly in challenging datasets marked by uniform textures or difficult lighting conditions (attributable to

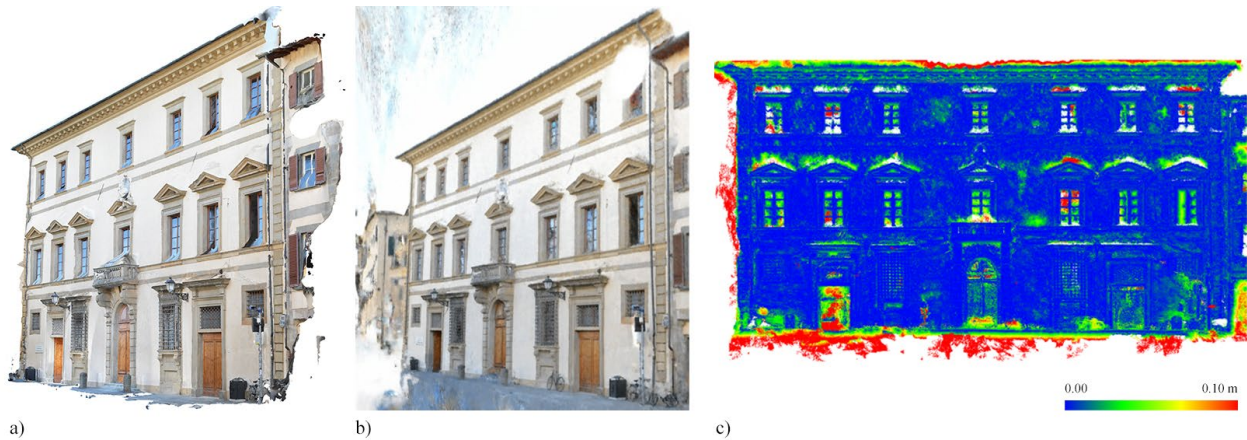


Figure 9. Façade of Palazzo Boileau: photogrammetric mesh (a) and volumetric rendering generated by NeRF (b); cloud-to-cloud comparison between the two point clouds (c).



Figure 10. Façade of Palazzo Boileau. Details of the photogrammetric mesh (above) and the volumetric rendering generated by the NeRF (below).

either the surrounding environment or the inherent material properties of the object), is noteworthy. This includes the rendering of view-dependent variations (the stucco decorations of the stove or the reflective surface of the mirror), and revealing segments of the object not captured in the photogrammetric model. However, it is important to note that the NeRF-based point cloud tends to be noisier.

- **At a scale of 1:50 or 1:100**, where the emphasis shifts towards a global description of geometries, the recognizability of component elements is maintained

in the NeRF model at a suitable level of detail, and the NeRF rendering results are more complete even in those areas that were less visible in the input images. However, the extracted point cloud is noisier than the photogrammetric one. While the model is more comprehensive, featuring gap-filling capabilities and capturing certain details absent in the photogrammetric model, it comes at the cost of increased noise.

It has also to be noted that the Spanish mill results demonstrate that, in scenarios involving the presence of reflective surfaces as water, the generated point cloud may misinterpret the presence of

additional portions that do not exist in reality. Addressing this issue could involve exploring alternative extraction algorithms for the point cloud derived from volumetric rendering.

Finally -and against the trend- the analysis of overall processing times shows a general advantage of photogrammetric processing. However, it is important to emphasize that the time parameter is highly dependent on the available computational capabilities. The tests presented were conducted using an NVIDIA GTX 1080 Ti GPU; however, Nerfstudio recommends using 3000 series or higher NVIDIA graphics cards for optimal performance. It is hoped that the use of such hardware components will significantly reduce training times, leading to a reassessment of the time parameter.

6. CONCLUSIONS

In this contribution, we evaluated the scale-dependent performance of NeRF models for Cultural Heritage. This scale dependency should be considered when determining the suitability of NeRF for specific applications.

The research highlights a trade-off between texture quality, geometric accuracy, gap-filling, output noise and object size, that should be taken into account when comparing NeRF models to photogrammetry. While NeRF excels at handling less visible regions and data completeness, the level of detail and clarity of surface textures is lower, especially at higher representation levels (scale 1:5). When dealing with datasets which can be challenging to capture through image-based surveys, such as those featuring complex elements like reflective surfaces or homogeneous textures, the results obtained on the mirror and stove dataset suggest the possible combination of NeRF and photogrammetry for a more comprehensive and detailed description of the surveyed objects.

Investigating optimal integration strategies to effectively leverage the strengths of both NeRF and photogrammetry could improve the overall performance and versatility of 3D reconstructions in Cultural Heritage. Future work will focus on refining more general guidelines for the selection of the appropriate scale when applying NeRF models in Cultural Heritage, even including the urban or regional scales.

AUTHOR CONTRIBUTIONS

V.C. is responsible for project conceptualization, paper preparation and writing. M.G.B., A.P., G.C., V.C. supervised all steps of the work, provided substantial insights and edits. The work is the result of a master thesis by author G.F., discussed at the University of Pisa under the joint supervision of V.C., M.G.B. and G.C.. The research was conducted and funded by the ASTRO Laboratory.

REFERENCES

Attal, B., Laidlaw, E., Gokaslan, A., Kim, C., Richardt, C., Tompkin, J., O'Toole, M., 2021. TöRF: Time-of-Flight Radiance Fields for Dynamic Scene View Synthesis. arXiv:2109.15271.

Caroti, G., Piemonte, A., and Ulivieri, D., 2023. Historic stone pavement photogrammetric 3D-survey: a way to get a catalogue to pass on ancient crafts. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-M-2-2023, 373–378. doi.org/10.5194/isprs-archives-XLVIII-M-2-2023-373-2023.

Croce, V., Billi, D., Caroti, G., Piemonte, A., De Luca, L., Véron, P., 2024. Comparative Assessment of Neural Radiance Fields and Photogrammetry in Digital Heritage: Impact of Varying Image

Conditions on 3D Reconstruction. *Remote Sensing*, 16(2), 301. doi.org/10.3390/rs16020301

Croce, V., Caroti, G., De Luca, L., Piemonte, A., Véron, P., 2023. Neural Radiance Fields (NeRF): Review and potential applications to Digital Cultural Heritage. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-M-2-2023, 453–460. doi.org/10.5194/isprs-archives-XLVIII-M-2-2023-453-2023.

Gao, K., Gao, Y., He, H., Lu, D., Xu, L., Li, J., 2022. NeRF: Neural Radiance Field in 3D Vision, A Comprehensive Review. arXiv:2210.00379.

Li, J., Feng, Z., She, Q., Ding, H., Wang, C., Lee, G.H., 2021. MINE: Towards Continuous Depth MPI with NeRF for Novel View Synthesis. arXiv:2103.14910.

Li, K., Rolff, T., Schmidt, S., Bacher, R., Frintrop, S., Leemans, W., Steinicke, F., 2022. Immersive Neural Graphics Primitives. arXiv:2211.13494.

Liu, S., Zhang, X., Zhang, Z., Zhang, R., Zhu, J.-Y., Russell, B., 2021. Editing Conditional Radiance Fields. arXiv:2105.06466.

Martin-Brualla, R., Radwan, N., Sajjadi, M.S.M., Barron, J.T., Dosovitskiy, A., Duckworth, D., 2021. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. arXiv:2008.02268.

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. arXiv:2003.08934.

Müller, T., Evans, A., Schied, C., Keller, A., 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41, 1–15. doi.org/10.1145/3528223.3530127.

Murtiyoso, A., 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.* XLVIII-M-2–2023, 1113–1118. doi.org/10.5194/isprs-archives-XLVIII-M-2-2023-1113-2023.

Pepe, M., Alfio, V.S., Costantino, D., 2023. Assessment of 3D Model for Photogrammetric Purposes Using AI Tools Based on NeRF Algorithm. *Heritage*, 6, 5719–5731. doi.org/10.3390/heritage6080301.

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. doi.org/10.3390/rs15143585

Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Kerr, J., Wang, T., Kristoffersen, A., Austin, J., Salahi, K., Ahuja, A., McAllister, D., Kanazawa, A., 2023. Nerfstudio: A Modular Framework for Neural Radiance Field Development. Proceedings of the SIGGRAPH'23: Special Interest Group on Computer Graphics and Interactive Techniques Conference, Los Angeles, CA, USA, 6–10 August 2023.

Zhi, S., Laidlow, T., Leutenegger, S., Davison, A.J., 2021. In-Place Scene Labelling and Understanding with Implicit Scene Representation. arXiv:2103.15875.