

NEURAL RADIANCE FIELDS (NeRF): REVIEW AND POTENTIAL APPLICATIONS TO DIGITAL CULTURAL HERITAGE

V. Croce ^{1,*}, G. Caroti ², L. De Luca ³, A. Piemonte ², P. Véron ⁴

¹ Dept. of Energy, Systems, Land and Construction Engineering, University of Pisa, 56122 Pisa, Italy - valeria.croce@ing.unipi.it
² Dept. of Civil and Industrial Engineering, University of Pisa, 56122 Pisa, Italy - gabriella.caroti@unipi.it, andrea.piemonte@unipi.it
³ UMR MAP 3495 CNRS/MC, Campus CNRS Joseph-Aiguier, 13402 Marseille, France - livio.deluca@map.cnrs.fr
⁴ LISPEN EA 7515, Arts et Métiers Institute of Technology, 13100 Aix-en-Provence, France - philippe.veron@ensam.eu

KEY WORDS: NeRF, Neural Radiance Fields, Cultural Heritage, photogrammetry, 3D reconstruction, Artificial Intelligence.

ABSTRACT:

Neural Radiance Fields (NeRF or NeRFs) are to date emerging as a novel method for synthesizing novel views of complex 3D scenes, leveraging an artificial neural network to optimize a volumetric scene function using a set of input views. We conduct a preliminary critical review of the scientific and technical literature on NeRFs, and we highlight possible applications of the latter in the Cultural Heritage domain, for the image-based reconstruction of 3D models of real, multi-scale objects, even in combination with the more well-established photogrammetric techniques. A comparison is made between NeRFs and photogrammetry in terms of operating procedures and outputs (volumetric renderings vs. point clouds or meshes). It is demonstrated that NeRFs could be conveniently used for rendering objects (sculptures, archaeological remains, sites, paintings etc.) that are challenging for photogrammetry, typically: i) metallic, translucent, and/or transparent surfaces; ii) objects that present homogeneous textures; iii) oclusions, vegetation, and elements of very fine detail.

1. INTRODUCTION

Neural Radiance Fields (NeRF or NeRFs) are a type of deep learning model that synthesizes novel views of an object from given multi-view images of a scene (Figure 1). The model was first presented by Mildenhall et al. (2020) (matthewtancik.com/nerf).

NeRF uses an artificial neural network to output volume density and view-dependent emitted radiance (i.e., the amount of light emitted or reflected by a surface). The neural network takes as input a single continuous 5D coordinate (spatial location x, y, z and viewing direction θ, ϕ) and outputs the volume density and view-dependent emitted radiance at that spatial location.

A key feature of NeRF is the ability to render high-quality, photo-realistic novel views with fine details and smooth transitions between different regions (Zhang et al., 2021).

A multi-scale architecture allows the model to learn and generate features at multiple scales simultaneously. NeRF also can handle occlusion and transparent objects, which makes them well-suited for tasks such as view synthesis and image-based rendering.

Between 2022 and 2023, since the large-scale implementation by Müller et al. (2022), NeRFs gained much attention in the Computer Vision field: the original article by Mildenhall et al. received more than 2500 citations and NeRFs have found interesting applications in various fields, including robotics (Adamkiewicz et al., 2022), industrial design (Mergy et al., 2021), autonomous navigation, medicine (Corona-Figueroa et al., 2022), 3D facial recognition (Guo et al., 2021) and human pose estimation (Su et al., 2021; Wang et al., 2022). Despite the ever-growing interest on the topic, the applications of NeRFs in Cultural Heritage are to date underdeveloped and understudied.

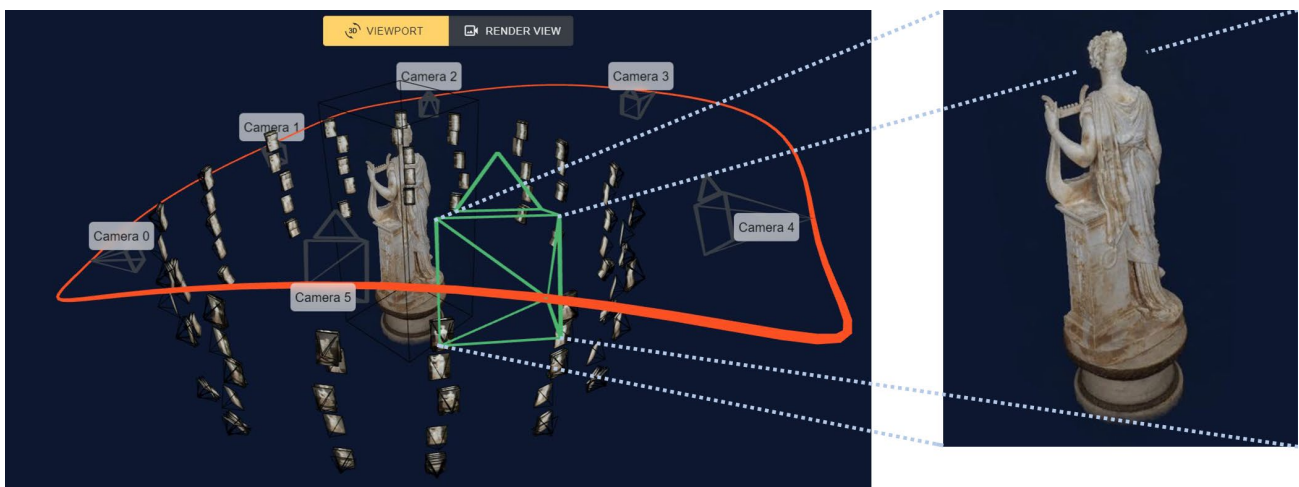


Figure 1. NeRF model and novel view synthesis for the Tersicore statue, Nerfstudio web viewer.

* Corresponding author

Moreover, the benefit of using NeRFs compared to more established image-based reconstruction techniques as photogrammetry are not yet adequately known or recognized. In tackling this issue, this paper reviews existing publications related to NeRF with the specific purpose of analysing implications of such techniques in state-of-the-art methods for multi-scale 3D modelling and graphics, in the digital heritage domain.

2. VIEW SYNTHESIS AND NEURAL RENDERING

The first advent of NeRFs is ascribed to the work by Mildenhall et al. (2020): starting from a set of images with known camera poses, views are synthesized by querying 5D coordinates along camera rays and classic volume rendering techniques are later used to project the output colours and densities into an image. In the basic form, a NeRF model represents a 3D static scene as a continuous 5D function, expressed as:

$$F_{\Theta}: (\mathbf{x}, \mathbf{d}) \rightarrow (\mathbf{c}, \sigma) \quad , \quad (1)$$

Where $\mathbf{x} = (x, y, z)$ are the in-scene coordinates, direction is expressed as the 3D Cartesian unit vector \mathbf{d} , $\mathbf{c} = (r, g, b)$ represents color values and σ stands for the volume density. (ϑ, φ) represent the azimuthal and polar viewing angles (viewing direction).

F_{Θ} is a Multi-Layer Perceptron (MLP), a feedforward artificial neural network, that outputs the colour information \mathbf{c} and the volume density σ (Figure 2). Although σ results to be independent of the viewing direction, the colour \mathbf{c} depends on both the viewing direction and the in-scene coordinate.

The models are trained per-scene, and the COLMAP pipeline (Schönberger et al., 2016; Schönberger and Frahm, 2016) is used to extract camera parameters and camera poses from the input image set. For each pixel in the image being synthesised, camera rays are marched through the scene and a set of sampling points is generated along each ray. For each sampling point, the known viewing direction and sampling locations are used to extract local colour and density through the MLP. 3D reconstruction and novel view synthesis are hence executed via volumetric rendering (Gao et al., 2022). In detail, given volume density and colour functions, volume rendering requires estimating the expected colour $C(r)$ of any camera ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$, with camera position \mathbf{o} and viewing direction \mathbf{d} . Looking at $T(t)$ as the probability that the ray travels from t_1 to t without hitting any other particle, $C(r)$ is expressed by equation:

$$C(r) = \int_{t_1}^{t_2} T(t) \cdot \sigma(\mathbf{r}(t)) \cdot c(\mathbf{r}(t), \mathbf{d}) \cdot dt \quad , \quad (2)$$

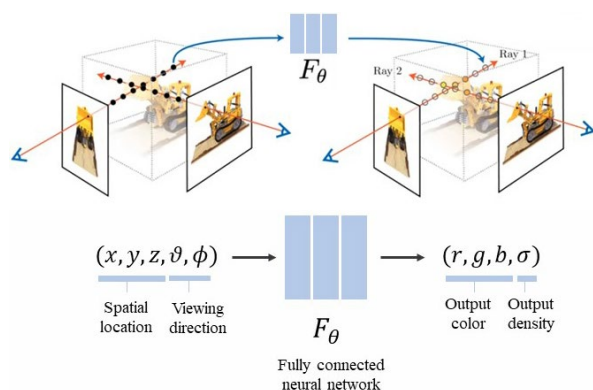


Figure 2. NeRF scene representation (adapted from Mildenhall et al., 2020).

Considering a non-deterministic stratified sampling approach, where the ray is divided into N equally spaced bins, a sample is uniformly drawn from each bin, so that equation (2) can be approximated as:

$$\hat{C}_{(r)} \approx \sum_{i=1}^N \alpha_i T_i c_i \quad , \quad (3)$$

2.1 NeRFs optimization

NeRF models employ: i) positional encoding to improve fine detail reconstruction and represent high-frequency functions, and ii) a hierarchical volume sampling strategy to allocate the MLP's performance towards areas of the scene with visible content (Mildenhall et al., 2020).

Significant implementations of the work by Mildenhall et al. (2020) include the view synthesis of dynamic scenes with objects in rigid or non-rigid motion (Chen and Tsukada, 2022; Attal et al., 2021; Pumarola et al., 2021), the anti-aliasing in rendering (Barron et al., 2021), the depth estimation (Li et al., 2021) and the reconstruction of dynamic fluids from sparse multi-view videos or images (Chu et al., 2022). The work by Barron et al. (2021) studied the incorporation of depth maps within the colour image set and proposed the so-called Mip-NeRF approach to model depth uncertainty through local sampling.

In the earliest version, NeRFs required extremely high computing power and long training times. The introduction of multi-resolution hash encoding in neural graphics by Müller et al., (2022), together with the advanced ray-tracing features delivered by GeForce RTX graphic cards, significantly reduced the processing time and capacity required for NeRF training (so that training neural graphics in seconds is to date possible).

2.2 Core implications of the theoretical formulation

Based on equation (3), colour can be expressed as the weighted combination of all alpha values α_i (transparency / opacity from alpha compositing at sample point i) and colour values c_i (colour evaluated at the sample point i) of points in a ray (Gao et al., 2022). In comparison with common ray tracing, NeRF leverage a probabilistic function to determine the expected value of colour along the ray (Figure 3). This has three main implications:

- The 3D reconstruction is provided in the form of a volumetric rendering. The NeRF model is neither a point cloud, nor a mesh, but it consists of continuous *voxels* made of shiny, transparent cubes;
- The scene representation is view-dependent, i.e., the colour of the object may change depending on the point of view, meaning that, for the case of non-diffusely reflecting surfaces (non-Lambertian materials), a different reflection might be shown depending on the point of view.
- NeRFs allow representation of real-world detailed scenes with complex occlusions, fine details and transient objects.

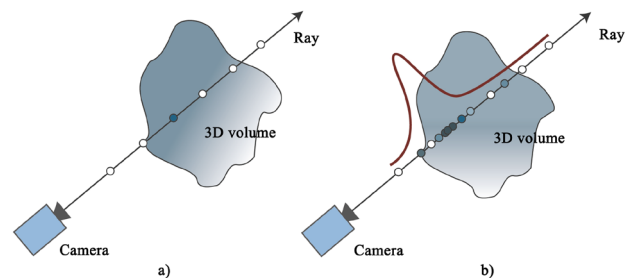


Figure 3. Common ray tracing (a) vs volume rendering (b).

2.3 Existing platforms and tools

The Instant NGP application, developed by Müller et al. (2022) for NVIDIA, accepts both photos and videos as input; the codebase is built on CUDA and Python 3.9, while the camera pose estimation is run through COLMAP. The graphical user interface can be launched via the Anaconda prompt.

Nerfstudio by Tancik et al. (2023) is a more recent Python framework allowing for end-to-end creation, training and testing of NeRFs. It integrates a real-time web viewer and supports multiple export modalities, including point clouds and meshes (Figures 4, 5). By default, Nerfstudio expects to apply a scaling factor to the input camera image resolution (*downscale*).

Luma AI by Luma (lumalabs.ai/) is a mobile NeRF capture platform for iOS, currently in beta, built up to integrate more user-friendly tools for neural rendering.

The first native volume rendering files compatible with the Unreal Engine video game and software development tool are to date emerging both in Nerfstudio, in the form of the plug-in Volinga (volinga.ai/), than in LumaAI.

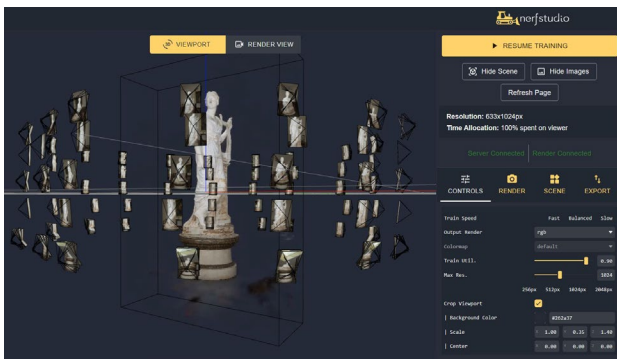


Figure 4. Rendering of the Tersicore statue visualized over Nerfstudio web viewer.

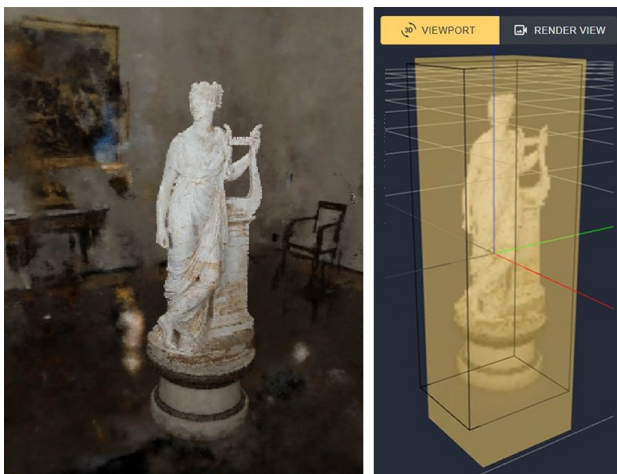


Figure 5. Bounding box for isolating portions of interest (to the right) from the neural rendering (to the left).

3. NERF APPLICATIONS IN DIGITAL HERITAGE

In the specific domain of Cultural Heritage, only few publications have, so far, explicitly identified NeRFs as a possible tool for virtual reconstruction, digital preservation and conservation of heritage objects and sites.

Condorelli et al. (2021) compared, focusing on the UNESCO site of Tour Saint-Jacques in Paris and using the same set of input images, the results of NeRFs with the photogrammetric

reconstruction via Structure-from-Motion (SfM) and Multi-View-Stereo (MVS). Their research, conducted with the early github implementation of the work by Mildenhall et al. (2020) (github.com/bmild/nerf), showed a significantly better result for MVS photogrammetry than for neural rendering, both in terms of reconstructed tower parts than in terms of computation time.

Later on, following the release of NVIDIA Instant NGP, Palestini et al. (2022) built NeRF models for few multi-scale cases of cultural interest: the façade of Santi Severino in Naples, the altar of the Church of S. Giustina in Padua, a bronze statue by artist A. Gonzales.

Despite these two attempts, the implications of the use of neural radiance fields in the Cultural Heritage domain still remain unclear and not exhaustively defined, especially compared to the widely used photogrammetry from SfM and MVS.

Other significant experiments, not specifically designed in the heritage field but potentially interesting for studies in this area, are listed in the following subparagraphs.

3.1 Neural 3D reconstruction *in-the-wild*

NeRFs have hitherto been shown to act in controlled, static environments from images acquired over a short period of time, whereby lighting effects are assumed nearly constant. Nerf in the Wild, or NeRF-W, introduced by Martin-Brualla et al. (2021), applies neural rendering to unconstrained image sets of a same artefact. The algorithm captures appearance variations of *in-the-wild* data, and the scene is decomposed so to evaluate transient elements and separate them from the static part of the scene. Large-scale image collections of tourist landmarks (taken *in-the-wild*, even available online), as the Brandenburg Gate in Berlin and the Trevi Fountain in Rome, were considered to test NeRF-W.

In a successive development of NeRF-W by Sun et al. (2022), appearance variations were modelled using appearance embedding, but the output was sought in meshes rather than in radiance fields. The modular approach by Kuang et al. (2022) first inferred the geometry of a real scene by neural rendering and then identified material properties and defined per-image lighting conditions to better relight and composite the captured scene.

Analogously, NeRFactor (Zhang et al., 2021) enabled free-viewpoint relighting to support shadow and material editing under arbitrary environment light changes.

At the territorial scale, Mari et al. (2022) developed Sat-NeRF for the reconstruction of *in-the-wild* satellite images. Another NeRF variant adapted to multi-date collections of satellite images is found in reference (Derksen and Izzo, 2021).

Taking into account photographs taken at different times, with different lighting and weather conditions, different temporal states and transient occluding objects such as pedestrians and cars, *in-the-wild* reconstruction allow the production of photorealistic and temporally consistent renderings from novel viewpoints.

However, main limitations of *in-the-wild* reconstructions are still identified in: i) the need to reconstruct initial camera poses for each training image (Kuang et al., 2022); ii) sensitivity to camera calibration errors, which can lead to blurry effects; iii) degradation of rendering quality in parts of the scene that are rarely visible in the training images (Martin-Brualla et al., 2021).

3.2 Semantic NeRFs

Recent developments in Cultural Heritage digitalization include the segmentation and classification of 3D models as point clouds (Croce et al., 2023) and meshes (Grilli and Remondino, 2019).

Class labeling can be attached to a geometric model to encode semantics, i.e., human-defined concepts, such as information on

architectural components, materials and degradation patterns within the digital representation (Croce et al., 2020).

The enrichment of NeRFs with semantic information has not been developed yet for cultural heritage: Zhi et al. (2021) extended NeRFs scene-specific representation to include semantic representations that were efficiently learned from partial sparse or noisy annotations of indoor scenes. Similarly, Pavlakos et al. (2022) relied on NeRF models for an accurate estimation of human pose and location. The recovered, semantically enriched 3D scene context was used to render novel views of the human localization within certain environments.

3.3. NERF rendering for Virtual and Extended reality

Most recent developments of NeRF include the navigation of neural renderings in Virtual Reality (VR) or Extended Reality (XR) applications (Deng et al., 2022; Park et al., 2022). NeRFs created with Instant NGP can be run in VR or AR modes, through compatible headsets or glasses (Li et al., 2022). However, limitations of these methods are the absence of direct interactions with individual objects of the 3D scene, the lack of real-time collision detection, the high latency and computational cost of renderings in medium- or large-sized scenes.

As novel view synthesis is a prerequisite to many VR and XR applications, Chiang et al. (2022) propose a method to control the *style* of a rendered 3D scene by enabling seamless switching between real-world scenes and virtual artistic styles, prior to VR and augmented reality (AR) applications.

4. CASE STUDIES AND EARLY RESULTS

The comparison between NeRFs and more established photomodelling techniques is essential to understand the extent to which neural rendering and novel view synthesis can enhance, or complement, existing techniques for cultural heritage digitization. Ongoing work is aimed at testing the advantages and disadvantages of NeRFs for the digital documentation of cultural heritage objects and sites, in comparison and combination with other existing, consolidated techniques such as photogrammetry. To this end, for the same set of images, taken with sufficient overlap, we evaluate photogrammetric reconstruction -with dense cloud and texture mesh extraction- on the one hand, and NeRF training on the other hand. The results of the two different methods are thereby aligned and compared (Figure 6).

A key initial assumption is that the camera poses are known on the input image set. The reconstruction of the camera orientation parameters via COLMAP is in fact common to both workflows. NeRFs are then trained to produce direct models (volumetric rendering) and some derived data, i.e., a point cloud and a textured mesh.

To export point clouds and meshes from NeRFs, we use the latest ns-export function of Nerfstudio: the marching cubes algorithm (Lorenzen and Cline, 1987) and the Poisson surface reconstruction (Kazhdan et al., 2006) are leveraged for mesh generation. Texture coordinates are later derived for the triangle meshes via the xatlas library (github.com/mworchel/xatlas-python). By deriving these outputs for the same synthetic datasets of different target objects, characterized by non-Lambertian materials or very fine details, we intend to compare photogrammetry and NeRFs, e.g., in terms of shape description and representation type (point cloud to point cloud or volumetric rendering to mesh).

For this reason, the multi-scale datasets considered are all chosen amongst cases of Cultural Heritage interest that are typically difficult to process with traditional photogrammetry: the Tersicore statue by the sculptor Antonio Canova, an eagle-shaped lectern of the 14th century and the Caprona Tower, near Pisa.

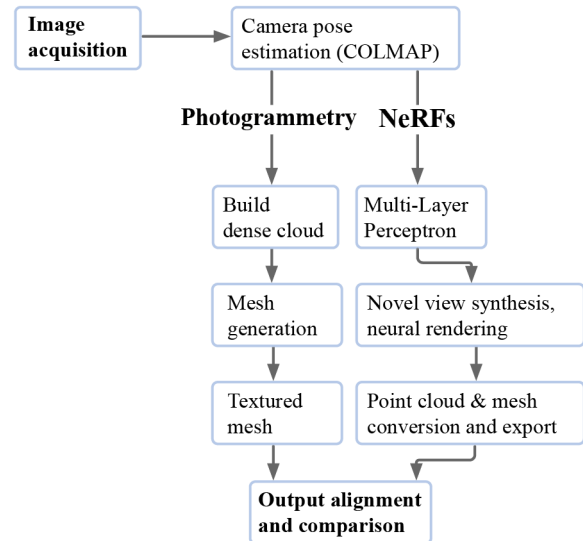


Figure 6. Proposed workflow for photogrammetry-to-NeRF output comparison.

The Tersicore statue. The Tersicore statue has a homogeneous colour, and the texture of its translucent material is difficult to render. The NeRF model (Figure 7) is generated using Nerfstudio and exported as a point cloud and a mesh from an input set of 233 images. The alignment between the NeRF-generated point cloud and the photogrammetric point cloud, realized via CloudCompare, allows an analysis of the cloud-to-cloud deviation (Figure 8) between the two outputs, on a scale from minimum (blue) to maximum values (red).

The results show that NeRF succeeds in describing, in a more complete manner, certain portions of the statue -as the head, the upper part of the pedestal, and the base- which were framed in a reduced number of images over the input dataset. The point cloud from NeRF is characterized by higher noise, which, however, disappears when looking at the starting volumetric rendering.

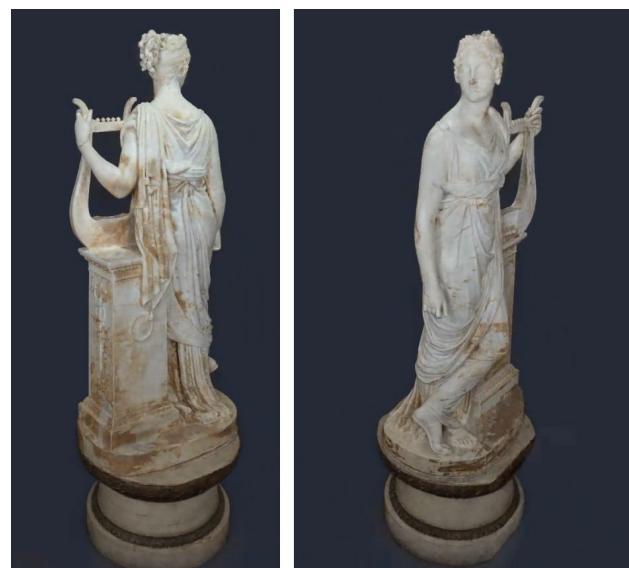


Figure 7. Neural renderings of the Tersicore dataset.

The bronze eagle-shaped lectern. By processing an input dataset of 254 images, NeRFs return color changes and reflections of the lectern's bronze material. The shiny, metallic effect, which is view-dependent (Figure 9), is much more

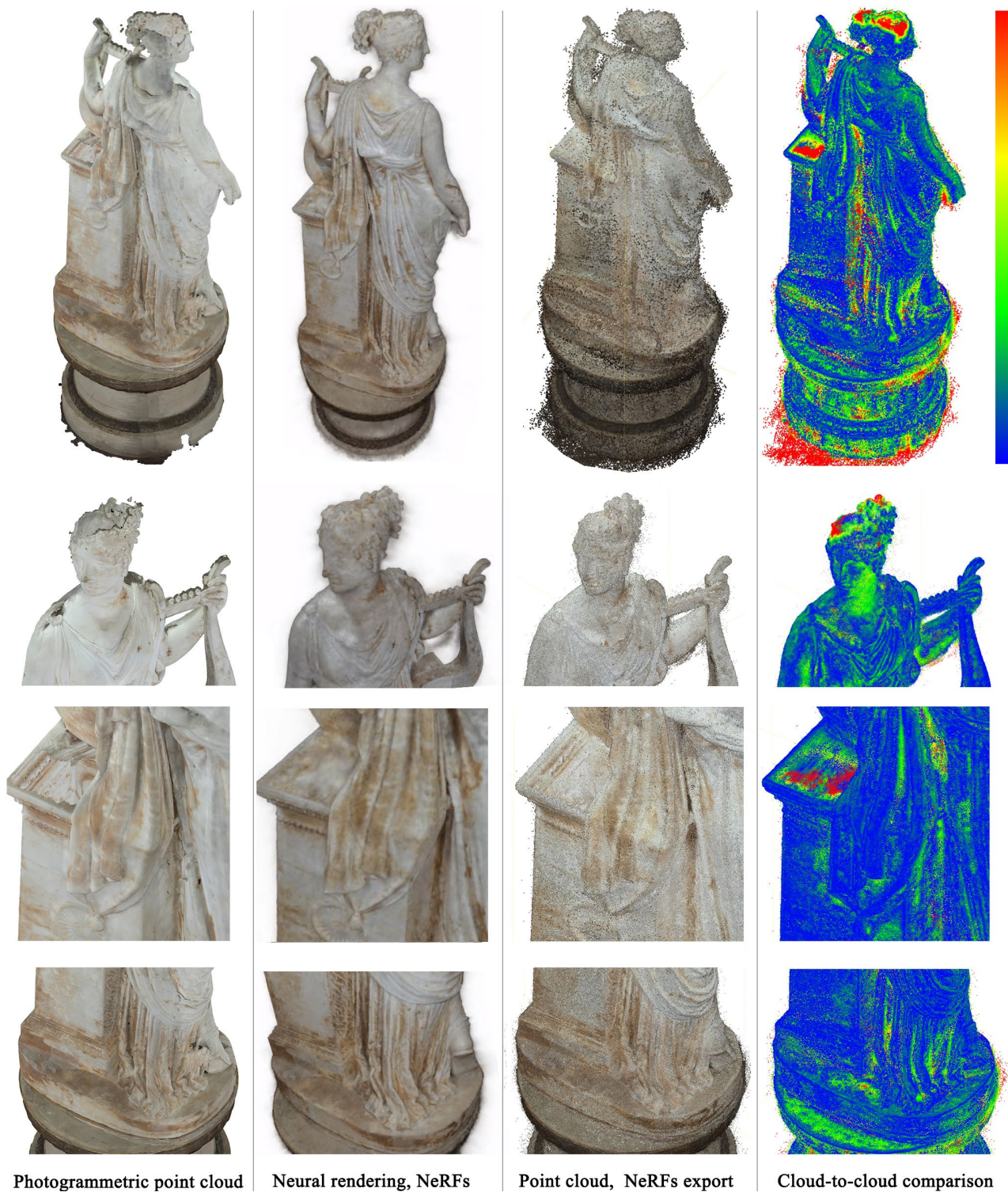


Figure 8. Cloud-to-cloud comparison for the Tersicore dataset.

consistent and realistic than the photogrammetric result. In contrast, the photogrammetric model is extremely flat and matte in terms of texture, making it difficult to tell whether the material is bronze or, e.g., wood (Figure 10).

Caprona Tower and its natural landscape. The challenge of the Caprona Tower dataset (Billi et al., 2023) is to integrate the natural environment around the tower, consisting of scattered vegetation and low shrubs, into the resulting model. We compare

the volumetric rendering with the photogrammetric mesh after processing a dataset of 124 drone images (Figures 11, 12).

In this case, there is a clear difference between the two types of rendering when it comes to the vegetation surrounding the tower: note the low bushes here and there, the presence of a few holes or excessively sharp sections in the mesh model that are fully returned in the volumetric rendering.

The difference can be seen in the tree to the left of the tower and in some of the low bushes.



Figure 9. View-dependent appearance of the NeRF model: the reflection of the eagle changes depending on the viewpoint.



Figure 10. Comparison between NeRF and photogrammetric textured mesh for the eagle-shaped lectern dataset.

This suggests the potential use of NeRFs in large-scale spatial mapping applications, even in emergency situations (Croce et al., 2021), to survey difficult to access landscapes or urban contexts characterized by dense vegetation.



Figure 11. Neural rendering of the Caprona Tower, on Nerfstudio web viewer.

5. DISCUSSION

Both photogrammetry and NeRF models allow the prediction of appearance and geometry from observed images.

At present, both techniques work on image sets with known camera poses. The camera parameters reconstruction phase is common to both methods, so that a set of images that cannot be correctly aligned by photogrammetry cannot be processed by neural rendering, at least for the time being. However, there have been several attempts at NeRF reconstruction with unknown and even randomly initialised camera poses (Meng et al., 2021).

Given the same resolution and size of the input images, NeRFs produce more complete models (with less loss of information) in the form of volumetric renderings. When compared to the result of point clouds or meshes obtained by photogrammetry, neural renderings could be meaningful for the survey of: i) objects with fine, volumetric details; ii) reflective or transparent surfaces; iii) occluding elements.

In photogrammetry, the appearance of real materials and surfaces does not depend on the point of view, while in NeRF models, the viewing direction is combined with location features so to predict the color from specific point of views and handle transparency and reflectivity.

As for photogrammetry, the 3D models produced by NeRFs are not to scale, unless they are appropriately combined with topographic surveys, e.g. from total stations or laser scanners.

In terms of outputs, NeRFs were specifically designed for novel view synthesis and volumetric rendering, and the shift from volumetric renderings to more conventional forms of representation as point clouds and meshes has not yet been extensively developed and studied; this prevents interoperability between the different model types. Conversion from NeRF to mesh (and vice versa, from mesh to NeRF) is indeed an active area of research.

6. CONCLUSIONS

This paper presented an overview of NeRFs, with a focus on possible applications and extensions of neural rendering in the domain of Cultural Heritage.

NeRF is a promising alternative to photogrammetry for image-based modelling, but combining the two techniques, also in terms of output, could be a game changer in overcoming some of the common problems of photogrammetry, including the difficulty of rendering homogeneous textures, transparent or translucent materials, and objects with extremely fine details.

Developments of this work could concern the conversion to other forms of digital representation, the comparison between NeRFs and photogrammetry in terms of processing time, resolution, interoperability. The model scalability in relation to information provided by metric surveys should even be assessed.

Finally, another interesting question is how NeRFs respond to the loss of information due to the reduced number or to the lower resolution of input images, compared to photogrammetry.

These aspects are currently under investigation.

AUTHOR CONTRIBUTIONS

The work is the result of an Italo-French collaboration between the home institutions of each author. V.C. is responsible for project conceptualization, execution of the experiments and drafting of the manuscript. G.C., L.D.L., A.P. and P.V. supervised all steps of the research, provided substantial insights and edits. The work has been funded, from January to February 2023, by the Joint LAB (CNR) project LIA Laboratoire International Associé (CNRS).



Figure 12. Comparison between NeRF model and photogrammetric textured mesh for the Caprona Tower.

REFERENCES

- Adamkiewicz, M., Chen, T., Caccavale, A., Gardner, R., Culbertson, P., Bohg, J., Schwager, M., 2022. Vision-Only Robot Navigation in a Neural Radiance World. *IEEE Robot. Autom. Lett.* 7, 4606-4613.
- 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
- Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P., 2021. Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, Montreal, QC, Canada, pp. 5835-5844.
- Billi, D., Rechichi, P., Montalbano, G., Croce, V., 2023. La Torre degli Upezzinghi a Caprona: analisi storico-archivistica e rilievo digitale per la documentazione dell’evoluzione temporale, in: Bevilacqua, M. G., Olivieri, D. (Eds.), *Defensive Architecture of the Mediterranean*, Proceedings of the International Conference on Fortifications of the Mediterranean Coast FORTMED 2023 Pisa, 23, 24 and 25 March 2023, XIII, 391-400.
- Chen, Q.-A., Tsukada, A., 2022. Flow Supervised Neural Radiance Fields for Static-Dynamic Decomposition, International Conference on Robotics and Automation (ICRA). 2022 IEEE International Conference on Robotics and Automation (ICRA), IEEE, Philadelphia, PA, USA, pp. 10641-10647.
- Chiang, P.-Z., Tsai, M.-S., Tseng, H.-Y., Lai, W.-S., Chiu, W.-C., 2022. Stylizing 3D Scene via Implicit Representation and HyperNetwork, in: 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), IEEE, Waikoloa, HI, USA, pp. 215-224.
- Chu, M., Liu, L., Zheng, Q., Franz, E., Seidel, H.-P., Theobalt, C., Zayer, R., 2022. Physics informed neural fields for smoke reconstruction with sparse data. *ACM Trans. Graph.* 41, 1-14.
- Condorelli, F., Rinaudo, F., Salvatore, F., Tagliaventi, S., 2021. A comparison between 3D reconstruction using NERF Neural Networks and MVS algorithms on Cultural Heritage images. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLIII-B2-2021, 565-570.
- Corona-Figueroa, A., Frawley, J., Taylor, S.B., Bethapudi, S., Shum, H.P.H., Willcocks, C.G., 2022. MedNeRF: Medical Neural Radiance Fields for Reconstructing 3D-aware CT-Projections from a Single X-ray, 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, Glasgow, Scotland, United Kingdom, pp. 3843-3848.
- Croce, V., Caroti, G., De Luca, L., Piemonte, A., Véron, P., 2020. Semantic annotations on heritage models: 2D/3D approaches and future research challenges. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLIII-B2-2020, 829-836.
- Croce, V., Caroti, G., Piemonte, A., De Luca, L., Véron, P., 2023. H-BIM and Artificial Intelligence: Classification of Architectural Heritage for Semi-Automatic Scan-to-BIM Reconstruction. *Sensors* 23, 2497.
- Croce, V., Diamantidis, D., Sýkora, M., 2021. Seismic damage evaluation and decisions on interventions supported by UAV-based surveys, in: Vayas, I., Mazzolani, F.M. (Eds.), *Protection of Historical Constructions. PROHITECH 2021, Lecture Notes in Civil Engineering*. Springer, Cham.
- Deng, N., He, Z., Ye, J., Duinkharjav, B., Chakravarthula, P., Yang, X., Sun, Q., 2022. FoV-NeRF: Foveated Neural Radiance Fields for Virtual Reality. *IEEE Trans. Visual. Comput. Graphics* 1-11.
- Derksen, D., Izzo, D., 2021. Shadow Neural Radiance Fields for Multi-view Satellite Photogrammetry, 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Nashville, TN, USA, pp. 1152-1161.
- 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.
- Grilli, E., Remondino, F., 2019. Classification of 3D Digital Heritage. *Remote Sensing* 11, 847.

- Guo, Y., Chen, K., Liang, S., Liu, Y.-J., Bao, H., Zhang, J., 2021. AD-NeRF: Audio Driven Neural Radiance Fields for Talking Head Synthesis, 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, Montreal, QC, Canada, pp. 5764-5774.
- Kazhdan, M., Bolitho, M., Hoppe, H., 2006. Poisson surface reconstruction. *Eurographics Symposium on Geometry Processing*.
- Kuang, Z., Olszewski, K., Chai, M., Huang, Z., Achlioptas, P., Tulyakov, S., 2022. NeROIC: neural rendering of objects from online image collections. *ACM Trans. Graph.* 41, 1-12.
- 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, 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, Montreal, QC, Canada, pp. 12558-12568.
- Li, K., Rolff, T., Schmidt, S., Bacher, R., Frintrop, S., Leemans, W., Steinicke, F., 2022. Immersive Neural Graphics Primitives.
- Lorenson, W.E., Cline, H.E., 1987. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. *SIGGRAPH Comput. Graph.* 21, 163-169.
- Mari, R., Facciolo, G., Ehret, T., 2022. Sat-NeRF: Learning Multi-View Satellite Photogrammetry With Transient Objects and Shadow Modeling Using RPC Cameras, 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, New Orleans, LA, USA, pp. 1310-1320.
- 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.
- Meng, Q., Chen, A., Luo, H., Wu, M., Su, H., Xu, L., He, X., Yu, J., 2021. GNeRF: GAN-based Neural Radiance Field without Posed Camera, 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Montreal, QC, Canada, pp. 6331-6341.
- Mergy, A., Lecuyer, G., Derksen, D., Izzo, D., 2021. Vision-based Neural Scene Representations for Spacecraft, 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Nashville, TN, USA, pp. 2002-2011.
- 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: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (Eds.), *Computer Vision - ECCV 2020, Lecture Notes in Computer Science*. Springer International Publishing, Cham, pp. 405-421.
- Müller, T., Evans, A., Schied, C., Keller, A., 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.* 41, 1-15.
- Palestini, C., Basso, A., Perticarini, M., 2022. Machine learning as an alternative to 3D photo-modeling employed in architectural survey and automatic design modelling. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLVIII-2/W1-2022, 191-197.
- Park, M., Yoo, B., Moon, J.Y., Seo, J.H., 2022. InstantXR: Instant XR Environment on the Web Using Hybrid Rendering of Cloud-based NeRF with 3D Assets. The 27th International Conference on 3D Web Technology, ACM, Evry-Courcouronnes France, pp. 1-9.
- Pavlakos, G., Weber, E., Tancik, M., Kanazawa, A., 2022. The One Where They Reconstructed 3D Humans and Environments in TV Shows. arXiv:2207.14279.
- Pumarola, A., Corona, E., Pons-Moll, G., Moreno-Noguer, F., 2021. D-NeRF: Neural Radiance Fields for Dynamic Scenes, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Nashville, TN, USA, pp. 10313-10322.
- Schönberger, J.L., Frahm, J.-M., 2016. Structure-from-Motion Revisited, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4104-4113.
- Schönberger, J.L., Zheng, E., Frahm, J.-M., Pollefeys, M., 2016. Pixelwise View Selection for Unstructured Multi-View Stereo, in: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), *Computer Vision - ECCV 2016*. Springer International Publishing, Cham, pp. 501-518.
- Su, S.-Y., Yu, F., Zollhoefer, M., Rhodin, H., 2021. A-NeRF: Articulated Neural Radiance Fields for Learning Human Shape, Appearance, and Pose. arXiv:2102.06199.
- Sun, J., Chen, X., Wang, Q., Li, Z., Averbuch-Elor, H., Zhou, X., Snavely, N., 2022. Neural 3D Reconstruction in the Wild, in: *Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings*. ACM, Vancouver BC Canada, pp. 1-9.
- 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. arXiv:2302.04264.
- Wang, K., Peng, S., Zhou, X., Yang, J., Zhang, G., 2022. NerfCap: Human Performance Capture With Dynamic Neural Radiance Fields. *IEEE Trans. Visual. Comput. Graphics* 1-13.
- Zhang, X., Srinivasan, P.P., Deng, B., Debevec, P., Freeman, W.T., Barron, J.T., 2021. NeRFactor: neural factorization of shape and reflectance under an unknown illumination. *ACM Trans. Graph.* 40, 1-18.
- Zhi, S., Laidlow, T., Leutenegger, S., Davison, A.J., 2021. In-Place Scene Labelling and Understanding with Implicit Scene Representation. arXiv:2103.15875.