Comparative Assessment of Neural Radiance Fields and Photogrammetry in Digital Heritage: Impact of Varying Image Conditions on 3D Reconstruction

Valeria Croce, Dario Billi, Gabriella Caroti, Andrea Piemonte, Livio De Luca, Philippe Véron

remote sensing

MDPI

Article

Comparative Assessment of Neural Radiance Fields and Photogrammetry in Digital Heritage: Impact of Varying Image Conditions on 3D Reconstruction

Valeria Croce [1], Dario Billi [2], Gabriella Caroti [2,*], Andrea Piemonte [2], Livio De Luca [3] and Philippe Véron [1]

[1] LISPEN EA 7515, Arts et Métiers Institute of Technology, 13100 Aix-en-Provence, France; valeria.croce@ensam.eu (V.C.); philippe.veron@ensam.eu (P.V.)
[2] Department of Civil and Industrial Engineering, ASTRO Laboratory, University of Pisa, 56122 Pisa, Italy; dario.billi@studenti.unipi.it (D.B.); andrea.piemonte@unipi.it (A.P.)
[3] UMR MAP 3495 CNRS/MC, Campus CNRS Joseph-Aiguier, 13402 Marseille, France; livio.deluca@map.cnrs.fr
* Correspondence: gabriella.caroti@unipi.it

Abstract: This paper conducts a comparative evaluation between Neural Radiance Fields (NeRF) and photogrammetry for 3D reconstruction in the cultural heritage domain. Focusing on three case studies, of which the Terpsichore statue serves as a pilot case, the research assesses the quality, consistency, and efficiency of both methods. The results indicate that, under conditions of reduced input data or lower resolution, NeRF outperforms photogrammetry in preserving completeness and material description for the same set of input images (with known camera poses). The study recommends NeRF for scenarios requiring extensive area mapping with limited images, particularly in emergency situations. Despite NeRF's developmental stage compared to photogrammetry, the findings demonstrate higher potential for describing material characteristics and rendering homogeneous textures with enhanced visual fidelity and accuracy; however, NeRF seems more prone to noise effects. The paper advocates for the future integration of NeRF with photogrammetry to address respective limitations, offering more comprehensive representation for cultural heritage preservation tasks. Future developments include extending applications to planar surfaces and exploring NeRF in virtual and augmented reality, as well as studying NeRF evolution in line with emerging trends in semantic segmentation and in-the-wild scene reconstruction.

Keywords: NeRF; cultural heritage; 3D reconstruction; photogrammetry; 3D surveying

Citation: Croce, V.; Billi, D.; Caroti, G.; Piemonte, A.; De Luca, L.; Véron, P. Comparative Assessment of Neural Radiance Fields and Photogrammetry in Digital Heritage: Impact of Varying Image Conditions on 3D Reconstruction. *Remote Sens.* **2024**, *16*, 301. https://doi.org/10.3390/rs16020301

Academic Editor: Fabio Remondino

Received: 25 November 2023
Revised: 29 December 2023
Accepted: 9 January 2024
Published: 11 January 2024

1. Introduction

The process of reconstructing and digitally documenting heritage artifacts and scenes holds great significance in studying, enhancing, and preserving tangible cultural heritage (CH). This significance is demonstrated by the extensive digitization campaigns that have been undertaken, the remarkable advancements in active and passive sensor-based survey techniques, and various established methodologies for three-dimensional reconstruction [1]. Artificial intelligence, virtual reality, and extended reality technologies are currently being extensively applied to three-dimensional data for a wide range of tasks related to the large-scale analysis [2], valorization [3,4], and communication [5,6] of cultural heritage objects. These technologies are designed to cater to the needs of both experts and the general public for a series of applications: preserving CH via the creation of digital replicas [7]; supporting restoration and conservation activities [8,9]; and education and tourism and cultural promotion (making CH more accessible, e.g., by virtual tours, interactive exhibits, and educational materials) [10].

In this context, the rendering of material and surface characteristics of existing heritage objects and sites, faithfully representing the actual shape and color of the latter, is of

paramount significance. Image-based reconstruction techniques have been studied for this purpose, and Structure-from-Motion (SfM) and Multi-View-Stereo (MVS) photogrammetry is considered as a consolidated methodology in this field, currently complementing active survey methods, such as Light Detection and Ranging (LiDAR) [11].

Photogrammetry is employed at multiple levels and scales of documentation to produce point clouds, textured 3D meshes, and orthophotos and is used to extract metric information through control points. These several outputs can be harnessed for further semantic enrichment, in-depth analysis, precise measurements, animations, and immersive visualizations [12].

With recent advancements in the realm of artificial intelligence applied to digital representation [5,13,14] and following Mildenhall's pioneering work [15], the introduction of Neural Radiance Fields (NeRFs) might appear to be a promising alternative to photogrammetry for 3D reconstruction from images. NeRF rely on a fully connected neural network to generate novel views of 3D objects and scenes based on a series of overlapping images with known camera poses. Initially designed for the specific task of novel view synthesis [16], NeRFs optimize an underlying continuous volumetric scene function that allows the generation of neural renderings.

NeRF models are photo-realistic [16], and they have achieved the visual quality required for the reproduction of real shapes and appearance of a 3D scene. However, the interest in applying NeRFs compared to other well-established technologies like photogrammetry has still not yet been recognized to its full potential. Even though the starting point is still a series of overlapping images, NeRFs use neural networks to create so-called radiance fields instead of relying on the reconstruction of geometrical relations between an image and the 3D world space [17]. The output, provided in the form of a neural rendering, can be turned into common, consolidated 3D objects as point clouds or meshes, e.g., via the marching cubes algorithm [18].

1.1. Background and Motivation

Three-dimensional reconstruction methods have constituted a major endeavor within the cultural heritage domain. They can be broadly divided into contact and non-contact sensors.

Among non-contact (passive) sensors, photogrammetry relies on capturing and analyzing data from photographs and does not involve actively emitting any form of energy or signals to gather information about the objects or the environment being studied. To date, it appears to be a consolidated technique for three-dimensional (3D) model reconstruction of features or topographies from overlaying two-dimensional photographs taken from various locations and points of views [19].

Common photogrammetric software, such as MicMac, RealityCapture 1.3.1 and Agisoft Metashape 2.1.0 enable the reconstruction of 3D data in the form of point clouds, polygonal meshes, and textured meshes. Ortho-rectified imagery and bi-dimensional maps can even be derived from a photogrammetric model and properly geo-referenced based on their combination with traditional topographic surveying methods [11]. Photogrammetry is widely applied in many domains ranging from industrial design [20] to archeology [21], architecture [22] and agriculture [23], geology [24], etc. However, photogrammetric reconstruction techniques face many common limitations in cases where objects exhibit challenging optical properties (such as absorptivity, intense reflectivity, or extensive scattering) [25], variations in lighting conditions, such as shadows, glare or inconsistent lighting, uniform or repetitive textures [11], and complex shapes or geometries [26]. In this context, NeRFs have emerged as cutting-edge technology that hold great promise for addressing some of these inherent limitations [16]. They rely on a supervised neural network model (Figure 1), the Multi-Layer Perceptron (MLP), i.e., a feedforward (non-convolutional) deep network that consists of fully connected neurons with a nonlinear kind of activation function.

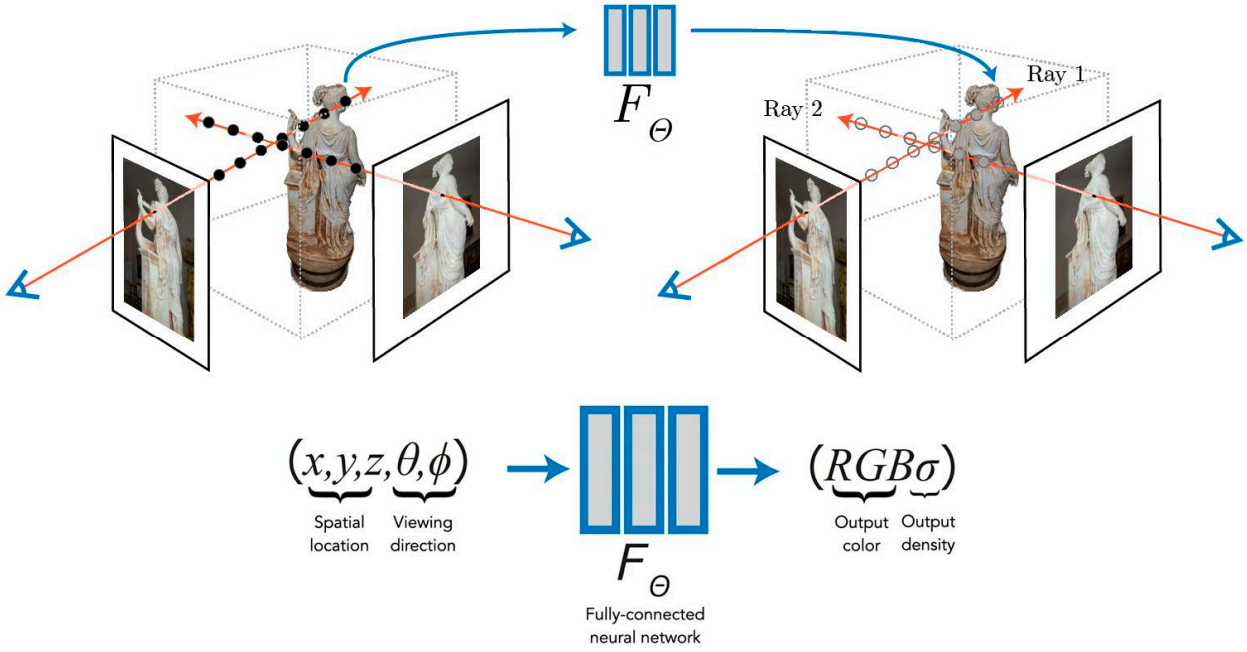


Figure 1. NeRF training overview adapted from the original NeRF paper by Mildenhall et al. [15].

The process essentially consists of three phases:

- First, as in photogrammetry, images are oriented in 3D space.
- Then, the sampled points, characterized by their three spatial dimensions and viewing direction, are processed by the MLP, resulting in color and volume density information as output.
- Finally, volumetric rendering techniques, such as ray tracing [15] or cone tracing [27], are used to synthesize the output information from the previous phase to produce new views (novel view synthesis).

In the formulation of NeRFs, the scene is intended as a continuous 5D function that associates to every point in space to the corresponding radiance emitted in each direction. The function maps spatial coordinates within the scene (x, y, z) and two viewing angles, azimuthal and polar, which specify the viewing direction (ϑ, φ), to output the volume density σ and a RGB color dependent on the viewing direction. The volume is intended as a differential opacity indicative of the radiance accumulated by a ray passing through each point. Considering the Multi-Layer Perceptron F_θ, the following formulation applies:

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

where $\mathbf{x}$ = (x, y, z) are the in-scene coordinates, and the direction is expressed as a 3D Cartesian unit vector $\mathbf{d}$ (ϑ, φ); $\boldsymbol{c}$ = (r, g, b) represents the color values; and σ indicates the volume density. Although σ is independent of the viewing direction, $\boldsymbol{c}$ depends on both the viewing direction and in-scene coordinates.

The theoretical formulation of NeRFs has the following consequences: (i) the 3D scene is represented as a neural rendering and not—at least initially—as a point cloud or mesh; (ii) the representation of the scene is view-dependent, which results in more realistic variations in color and illumination both with respect to lighting conditions and the handling of reflective surfaces.

Since their introduction in 2020, NeRFs have seen numerous implementations and have been extended to various fields of research, including urban mapping [28,29], robotics [30,31], autonomous driving [32], the simulation of climate effects [33], industrial design [34], and human pose estimation [35]. Mazzacca et al. [36] proposed several methods such as noise level, geometric accuracy, and the number of required images (i.e., image baselines) to evaluate NeRF-based 3D reconstructions.

In the specific field of cultural heritage, the first experiments began with the work of Condorelli and Rinaudo [37], and they have continued over the past three years, incorporating the developments and improvements that NeRFs have undergone since their earliest versions. Interest in the application of NeRFs to the cultural heritage sector is on the rise. Murtiyoso and Grussenmeyer [17] empirically demonstrated that NeRFs exhibit accelerated processing times in contrast to conventional MVS methods, but they noted a trade-off in geometric precision and the level of detail documentation. To assess the geometric quality of outputs, two heritage objects are subjected to evaluation, considering the mean error and standard deviation values across the resulting point clouds. Vandenabeele et al. [38] investigated the use of NeRFs for crowd-sourced surveying for building archaeology over unconstrained photo collections. Balloni et al. [39] conducted a comparative analysis between NeRFs and photogrammetry using the single case study of a statue. They employed terrestrial laser scanning (TLS) acquisition as the ground truth and compared the cloud-to-mesh distance and roughness values. The ethical implications of embodying new rendering methods based on AI in CH conservation practices were also discussed by [40].

In a previous contribution [41] presented at the CIPA 2023 Conference on Documenting, Understanding, Preserving Cultural Heritage, for the 'AI and NeRF for 3D reconstruction' session, we laid the groundwork for a comparative analysis between NeRFs and photogrammetry. The focus encompassed operational procedures and output modalities (volumetric renderings vs. point clouds or meshes). We posited that NeRFs offer a practical solution for rendering challenging objects, such as sculptures, archaeological artifacts, sites, and paintings, especially those featuring metallic, translucent, or transparent surfaces, homogeneous textures, or intricate details susceptible to occlusions, vegetation, and fine elements. However, the way NeRFs address the loss of information due to the reduced number or lower resolution of input images compared to photogrammetry still deserves further investigation.

1.2. Research Aim

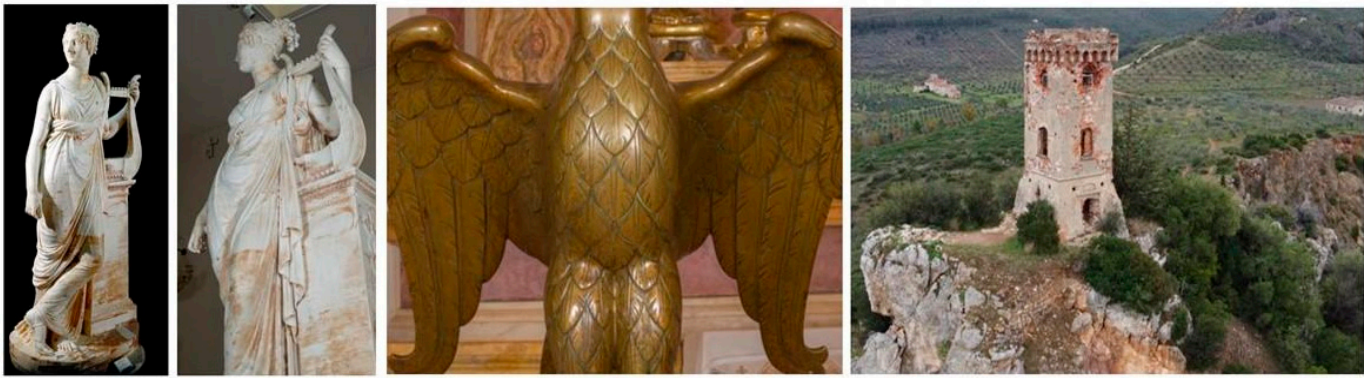
The aim of this work is to assess the intrinsic advantages or limitations of NeRFs compared to photogrammetry and the possible benefits of integrating the two methods for the digital 3D reconstruction of cultural heritage objects. Considering the same set of input images with known camera poses, we aimed to compare the two techniques in terms of quality and consistency of the results, handling of challenging scenes (e.g., objects with reflective, metallic or translucent surfaces), realistic renderings of the detected objects, processing time and impact of image resolution and numbering on the accuracy and fidelity of the 3D reconstruction.

2. Case Studies

For the case studies (Figures 2 and 3), we take into account objects of various scales, ranging from a small museum artifact to a monumental statue up to an urban context. The three specific cases presented in this contribution are related to datasets that are typically difficult to handle when using photogrammetry due to the different peculiarities of the objects' surfaces, e.g., very fine details or specific material characteristics (homogeneity, roughness, brilliance, chromaticity, luminosity, and hardness). Some of these objects feature non-Lambertian surfaces, meaning that they do not reflect light uniformly in all directions and exhibit angle-dependent reflectance properties. Further cases are still being implemented at the time of writing.

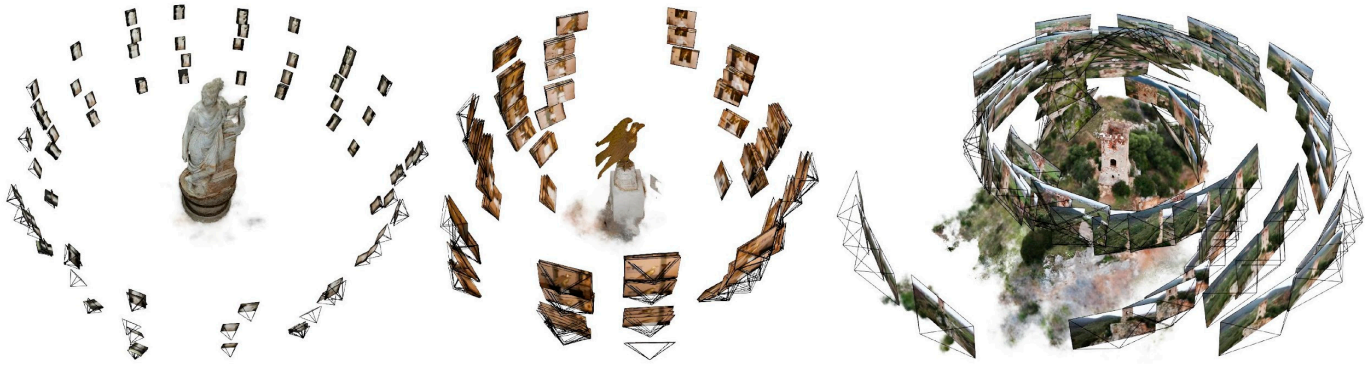
The pilot case study chosen to test the approach is the statue of Terpsichore, the masterpiece of the Italian artist and sculptural exponent of Neoclassicism, Antonio Canova (Figure 2a). The statue represents the muse of dance and choral singing, Terpsichore, as can be deduced from the lyre placed on the high pedestal. The work has intricate details, such as the strings and soundbox of the lyre or the drapery of the robe, and has specific material characteristics that are difficult to reproduce, such as the homogeneity of the color and the white and pink veining of its material. The survey was carried out by ACAS3D

Digital Solutions, a spinoff of the University of Pisa, using a Nikon D850 camera equipped with a 50 mm fixed focal length lens and operating in aperture priority mode (f/13). The original dataset consists of 233 images, each with a resolution of 2752 × 4128 pixels.



(**a**) (**b**) (**c**)

Figure 2. The three case studies analyzed the following: (**a**) Terpsichore by Antonio Canova, (**b**) an eagle-shaped lectern, and (**c**) Caprona Tower.



(**a**) (**b**) (**c**)

Figure 3. Image acquisition pattern displayed for the neural renderings of the following: (**a**) the Terpsichore, (**b**) the eagle-shaped lectern, and (**c**) Caprona Tower.

The second case study considered is that of the eagle-shaped lectern (254 images) that presents specular reflections and other variations in appearance based on the direction of incoming light and the viewing angle due to the metallic, shiny surface of the bronze (Figure 2b). The survey was carried out by ACAS3D Digital Solutions, a spinoff of the University of Pisa, using a Nikon D850 camera equipped with a 50 mm fixed focal length lens and operating in aperture priority mode (f/13).

Finally, the Caprona Tower dataset, composed of 124 images acquired via a drone, presents a wider context scale comprising the natural environment surrounding the tower, which is composed of trees and shrubs (Figure 2c). The photogrammetric reconstruction of this study object, together with the historical remarks on the tower, were previously documented in reference [42]. The images were captured from a video recorded by author D.B. in 2021 using a DJI Mavic Mini 2 drone equipped with a 4k resolution camera at 30 frames per second.

3. Methodology

The present study aimed to test the validity and potential advantages of using NeRFs in comparison to well-established image-based reconstruction techniques, with photogrammetry being the foremost among them. For a direct comparison between NeRFs and photogrammetry, this research employed the same set of synthetic images.These images

depict objects with non-Lambertian surfaces or composed of very fine materials. In this dataset, following standard photogrammetric practices, the object of study was captured through a sequence of successive shots, ensuring more than 70% overlap. Over these initial datasets, two distinct methodologies have been employed (Figure 4):

- **Photogrammetric Procedure.** This involved estimating the camera orientation parameters for sparse point cloud construction, generating a dense point cloud, creating a mesh, and extracting the textures. The software used for this task is Agisoft Metashape 2.1.0, and the alignment, dense point cloud, mesh and texture generation phases are run in high quality mode.
- **NeRF-Based Reconstruction.** This method requires the camera pose estimate to be known. With this input, a Multi-Layer Perceptron is trained for novel view synthesis, and the neural rendering (volumetric model) is generated. For the latter part, the Nerfstudio Application Programming Interface by Tancik et al. was used [43]. By default, this application applies a scaling factor to the images to reduce their dimensions and expedite the training process (*downscaling*).

As the alignment procedure is common to both processes and is executed following photogrammetry principles, an initial assumption is made about the dataset: the images used for reconstructing the 3D scene can be aligned, which is to say that the estimation of camera poses is feasible using the input images. In both cases, camera orientation parameters are calculated using Agisoft Metashape 2.1.0 software. The camera data file in .xml format, exported from Metashape, is then imported into Nerfstudio [43], serving as the initial data for the 3D scene reconstruction. It is important to note that for this step to work, all images are required to be of the same sensor type. By way of comparison, since the output of photogrammetry cannot be directly confronted with neural rendering, an additional conversion phase is introduced.

NeRFstudio allows the export of NeRFs to point clouds and meshes via the *ns-export* function. The marching cubes algorithm [18] and the Poisson surface reconstruction method [44] are used for this purpose, while texture coordinates are derived via the *xatlas* library (https://github.com/mworchel/xatlas-python, accessed date: 8 January 2024). Referring to known measurements from topographic surveys, the point clouds obtained through photogrammetry and NeRFs are individually scaled and aligned.

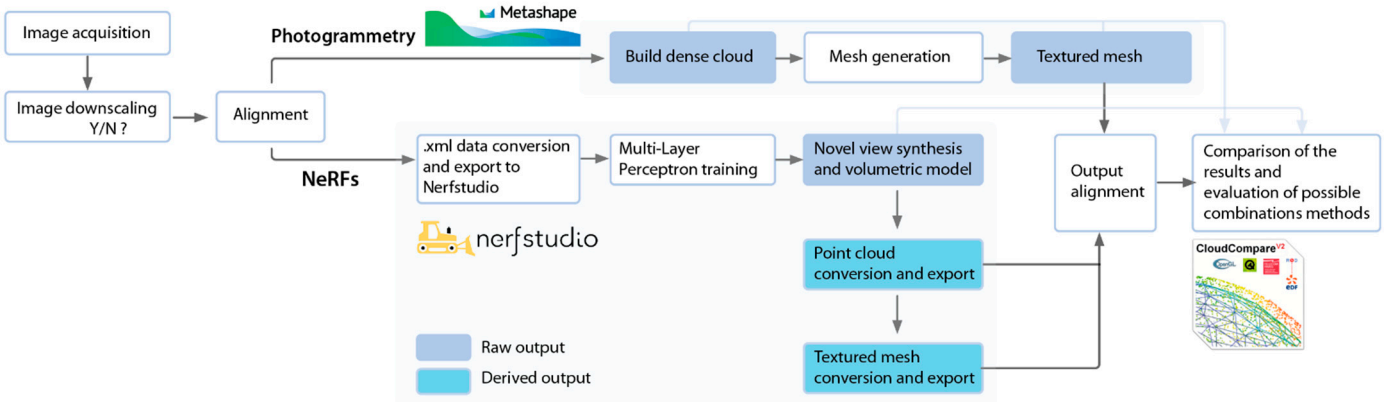
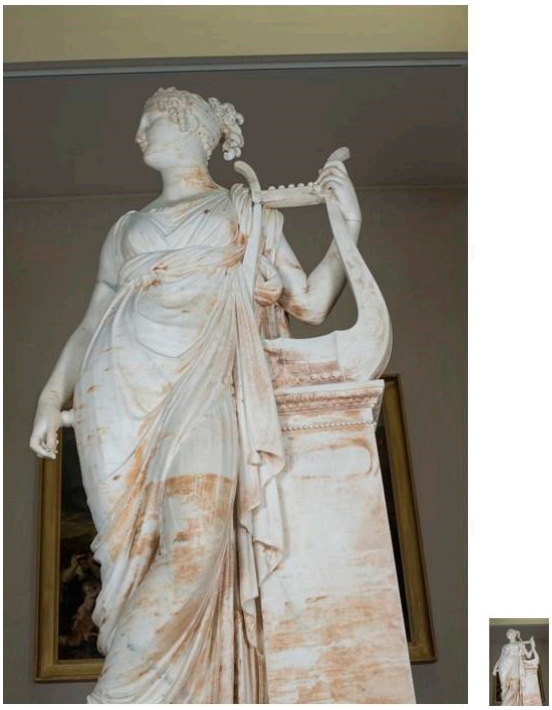


Figure 4. Overview of the proposed methodology.

Subsequently, they are processed to derive a cloud-to-cloud comparison using the CloudCompare software (CloudCompare 2.12.4, 2023). Various configurations of both input image sizes and quantities are considered in this phase to determine whether reducing the resolution or the number of images leads to a decrease in the quality of the 3D reconstruction. For the pilot case of the Terpsichore statue, we consider that the production process of volumetric models using NeRFs includes, by default, a downscaling factor, i.e., a reduction in the image size by a factor of 3 (Figures 5 and 6) and that a lower number of images

results in a reduction in training time. Four different dataset configurations were tested, as follows:

1. First: 233 photos, no downscale;
2. Second: 233 photos with a downscale of factor 3 (3×);
3. Third: a reduced dataset of 116 photos (~50% of the input dataset) with no downscale;
4. Fourth: a reduced dataset of 116 photos with downscale 3×.



(**a**) (**b**)

Figure 5. Comparison of image sizes between initial dataset ((**a**), 2752 × 4128 pixels) and reduced dataset downscaled by a factor of 3 ((**b**), to the left, 344 × 516 pixels).

(**a**) (**b**)

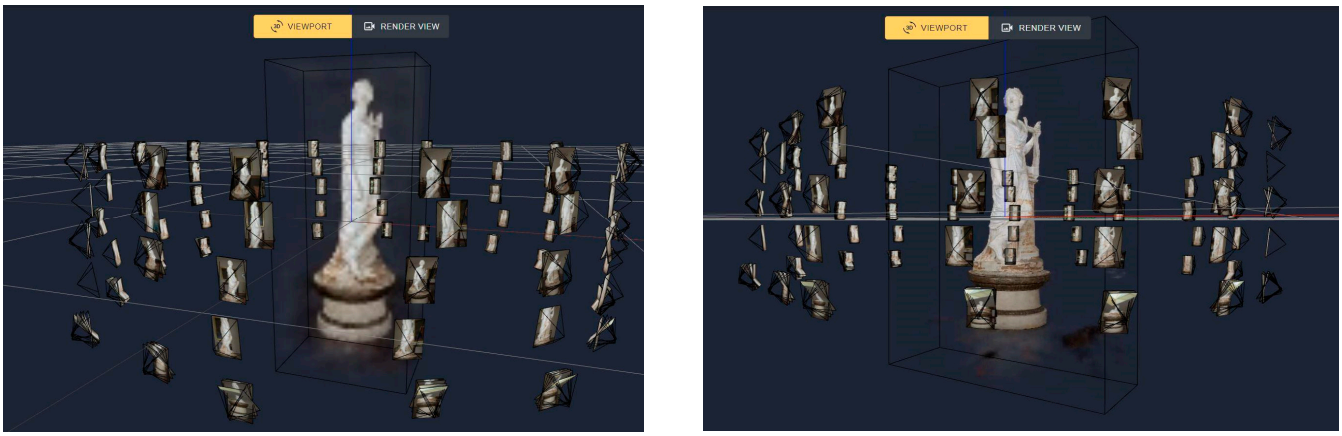
Figure 6. Comparison of image resolution between initial dataset ((**a**), 2752 × 4128 pixels) and reduced dataset downscaled by a factor of 3 ((**b**), 344 × 516 pixels).

It has to be noted that, in combinations 3 and 4, the reduction in dataset size is accomplished by systematically removing images one by one.

4. Results

For each of the four combinations considered, we proceeded by drawing up synoptic tables showing the comparison between the point cloud and the mesh acquired from

photogrammetry and between the neural rendering, the point cloud, and the mesh extracted from Nerfstudio (Figures 7–9). The point cloud extracted using NeRFstudio and the photogrammetric point cloud were later aligned on CloudCompare and compared with each other (Figure 10) to derive an analysis of deviation between the minimum values (in blue) and the maximum values (in red).

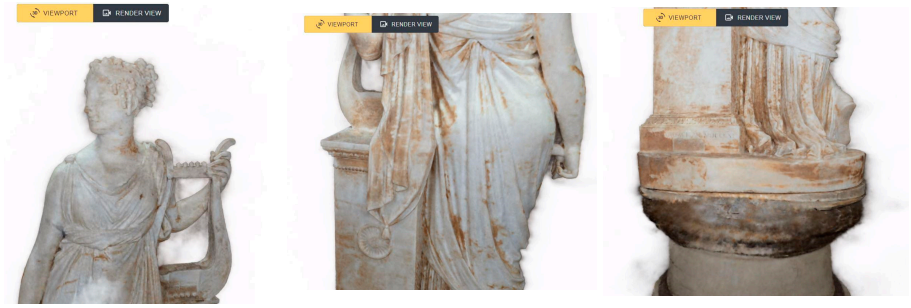


(**a**) (**b**)

Figure 7. Neural rendering with training in progress (**a**) and training completed (**b**).

(**a**) (**b**) (**c**)

Figure 8. Neural renderings of the Terpsichore statue from different points of view ((**a**,**b**)—back, (**c**)—front), case 1.



(**a**) (**b**) (**c**)

Figure 9. Several details of the neural renderings of the Terpsichore statue, case 4 ((**a**)—head, (**b**)—upper part of the pedestal and back drapery, (**c**)—lower par of the pedestal and back drapery).

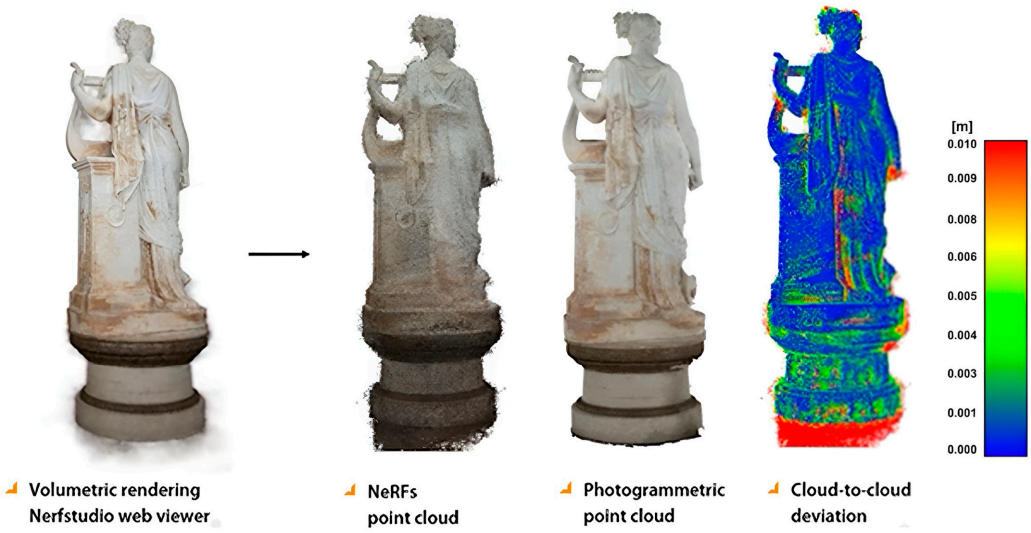


Figure 10. Scheme of the comparison between NeRF and photogrammetric point cloud.

The complete synoptic tables for all four datasets are presented in Figures 11–14, respectively. First, each synoptic table illustrates a general overview of the whole result of the point cloud and mesh from photogrammetry; the neural rendering, extracted point cloud, and extracted mesh from NeRFs; and the cloud-to-cloud comparison between NeRFs and photogrammetry. In addition, detailed, zoomed-in views of specific areas where the differences among the various tests are most pronounced are presented for each result. These areas of focus include the head, pedestal, base, and rear of the statue, with specific attention to details of the *lira*.

In detail, the synoptic tables provided in Figures 11–14 reveal that as image resolution and the number of available images decrease (provided image alignment remains feasible), NeRFs exhibit significantly less quantitative information loss compared to photogrammetry. This phenomenon becomes particularly pronounced when examining specific details such as the statue's head (present in only a few initial dataset images) and the upper and lower sections of the pedestal. Additionally, a noticeable enhancement in rendering very fine lyre details is observed when transitioning from case 1 to case 4.

The deviation analysis shows a general range from 0 mm (blue points) to a maximum value of 10 mm (red points). In the first dataset (Figure 11), the maximum deviation occurs at the top of Terpsichore's head and at the upper and lower sections of the pedestal. Moreover, it can be noted that the point cloud obtained from NeRFs exhibits more noise, with deviations spread across all surfaces. This is particularly noticeable where the deviation is minimal (1 mm in blue), but there appear isolated points with medium (4 to 7 mm in green-yellow) and high (1 cm in red) deviations.

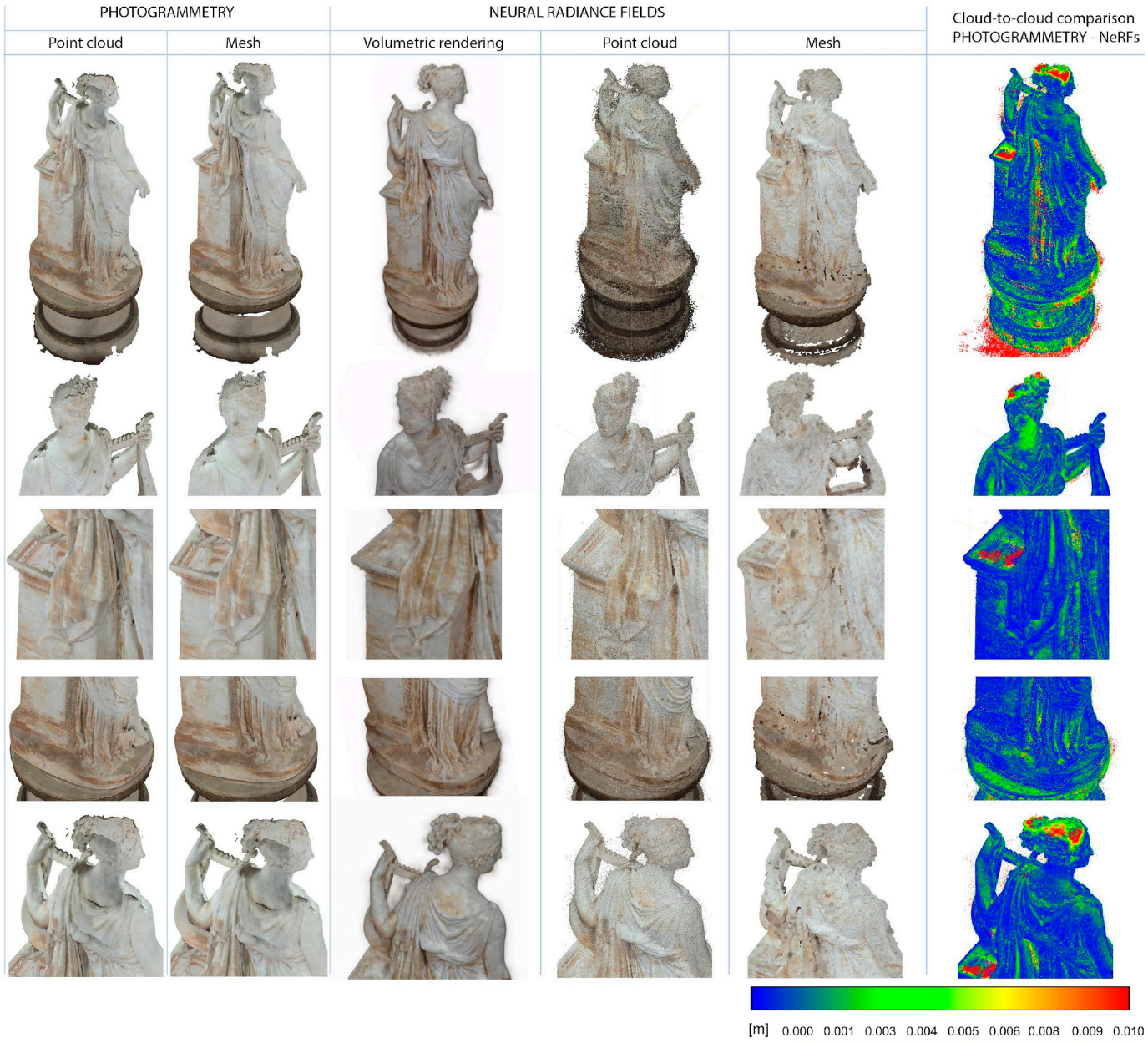


Figure 11. Analysis of the deviation between NeRF and photogrammetric processing results. Dataset 1: 233 images, no downscale.

In the second dataset (Figure 12), a maximum deviation of 1 cm is found at the head of Terpsichore and the lower section of the pedestal. Compared to the previous case, deviations between 1 and 4 mm are present throughout the entire statue, while the lyre and the pedestal maintain 1 mm deviations.

For the third dataset (Figure 13), maximum deviations of 1 cm shift to the lower part of the pedestal concentrated between the square of the pedestal and the drapery. The deviations at the head of Terpsichore decrease to an average of 3 to 8 mm.

For the last case (Figure 14), higher deviations, reaching a value of 1 cm, are observed at the head of the statue as well as in the lower part of the pedestal; indeed, these parts appear to be entirely missing in the photogrammetric results. Meanwhile, average deviations of 4 to 5 mm are found along the entire drapery. The consideration of the observed noise in the NeRF point cloud, as anticipated for the first case, is relevant for all the other cases mentioned above. This can also be appreciated by a visual comparison between the volumetric rendering and the photogrammetric mesh.

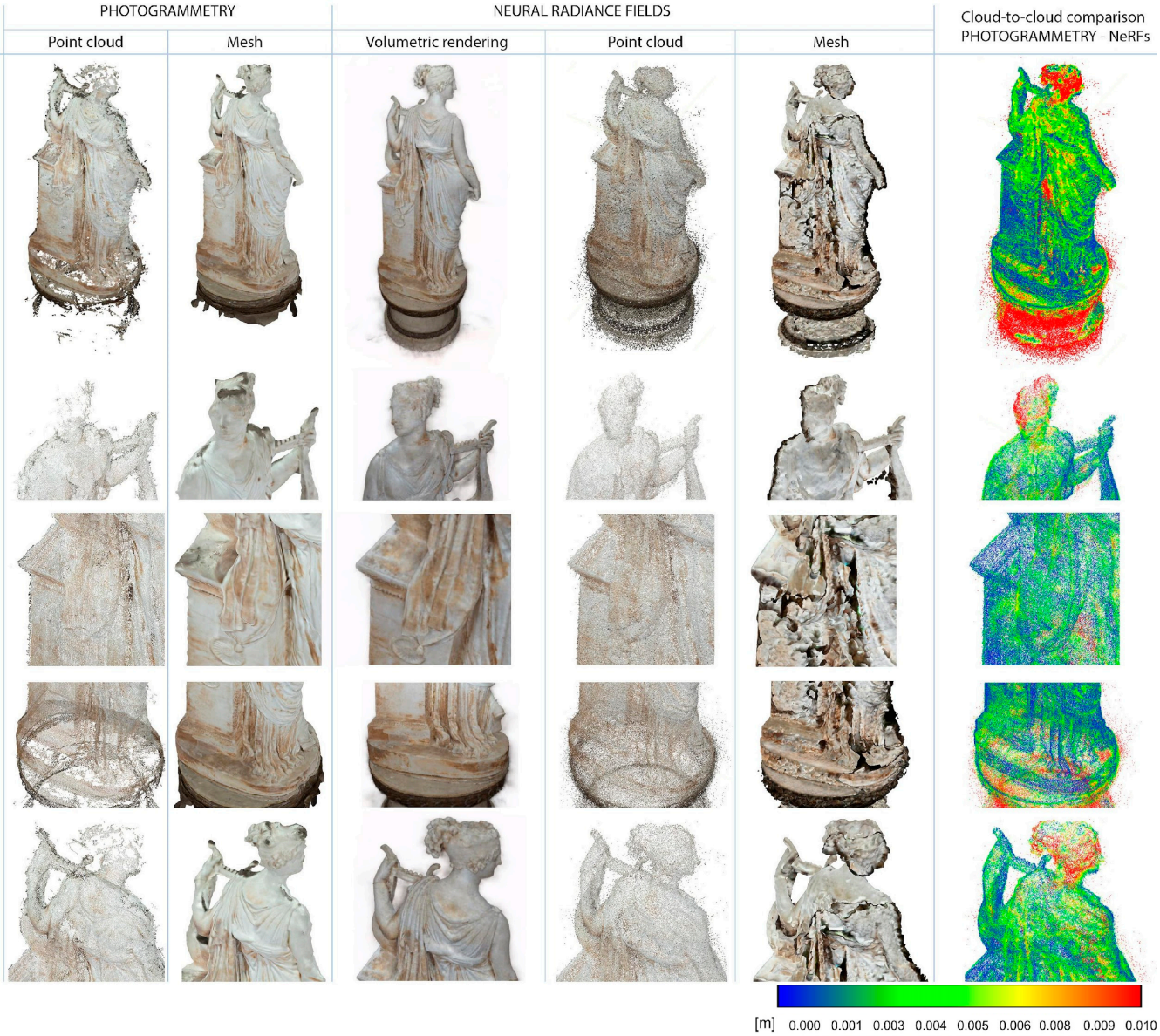


Figure 12. Analysis of the deviation between NeRF and photogrammetric processing results. Dataset 2: 233 images with 3× downscale.

Overall, the results of the pilot case study suggest the following:

- Compared to photogrammetry, NeRFs may offer the ability to handle reduced image data or reduced resolution of the images, with lower quantitative information loss. For the 3rd and 4th cases analyzed, indeed, NeRFs capture details, such as the head and lower pedestal, which are absent in the photogrammetric output. This is true, however, if the reconstruction of camera poses is possible over the reduced datasets;
- NeRF neural renderings more faithfully reproduce the statue's material texture compared to the textured mesh obtained through photogrammetry.
- However, NeRFs are more prone to noise, and for higher-resolution datasets, they may encounter challenges in capturing specific fine details compared to photogrammetry.

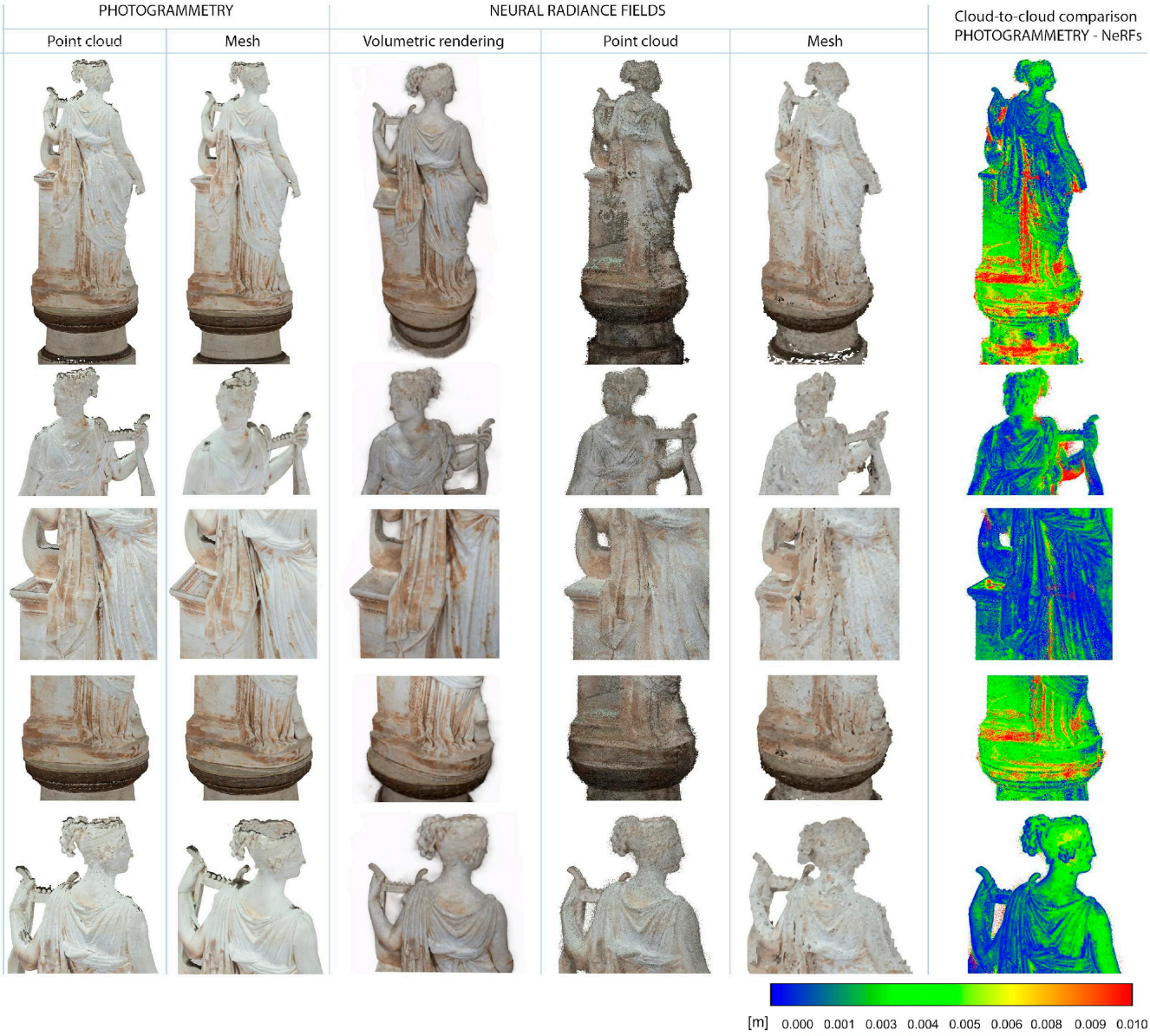


Figure 13. Analysis of the deviation between NeRF and photogrammetric processing results. Dataset 3: 116 images with no downscale.

To underscore the stark contrast in texture rendering between photogrammetry and NeRFs, the example of the eagle-shaped lectern vividly demonstrates the concept of view dependency achieved through the use of NeRFs. In the series of images provided as the input, the illumination on the eagle's chest varies dramatically. Examining the results obtained via photogrammetry, such variations in appearance cannot be observed. NeRFs, in this instance, excel in providing color changes and reflections that faithfully mimic the actual bronze surface, creating a highly realistic representation; as in Figure 15, the illumination on the eagle's chest varies dramatically. The images below, taken from different viewpoints, clearly show how the rendering of the raptor's metallic surface changes depending on the view direction. In stark contrast, the photogrammetric model appears quite matte in terms of texture (Figures 16 and 17). It does not convey the material's true nature, making it difficult to distinguish whether the surface is made of bronze or wood, illustrating the limitations of photogrammetry in capturing the nuanced material properties of the subject.

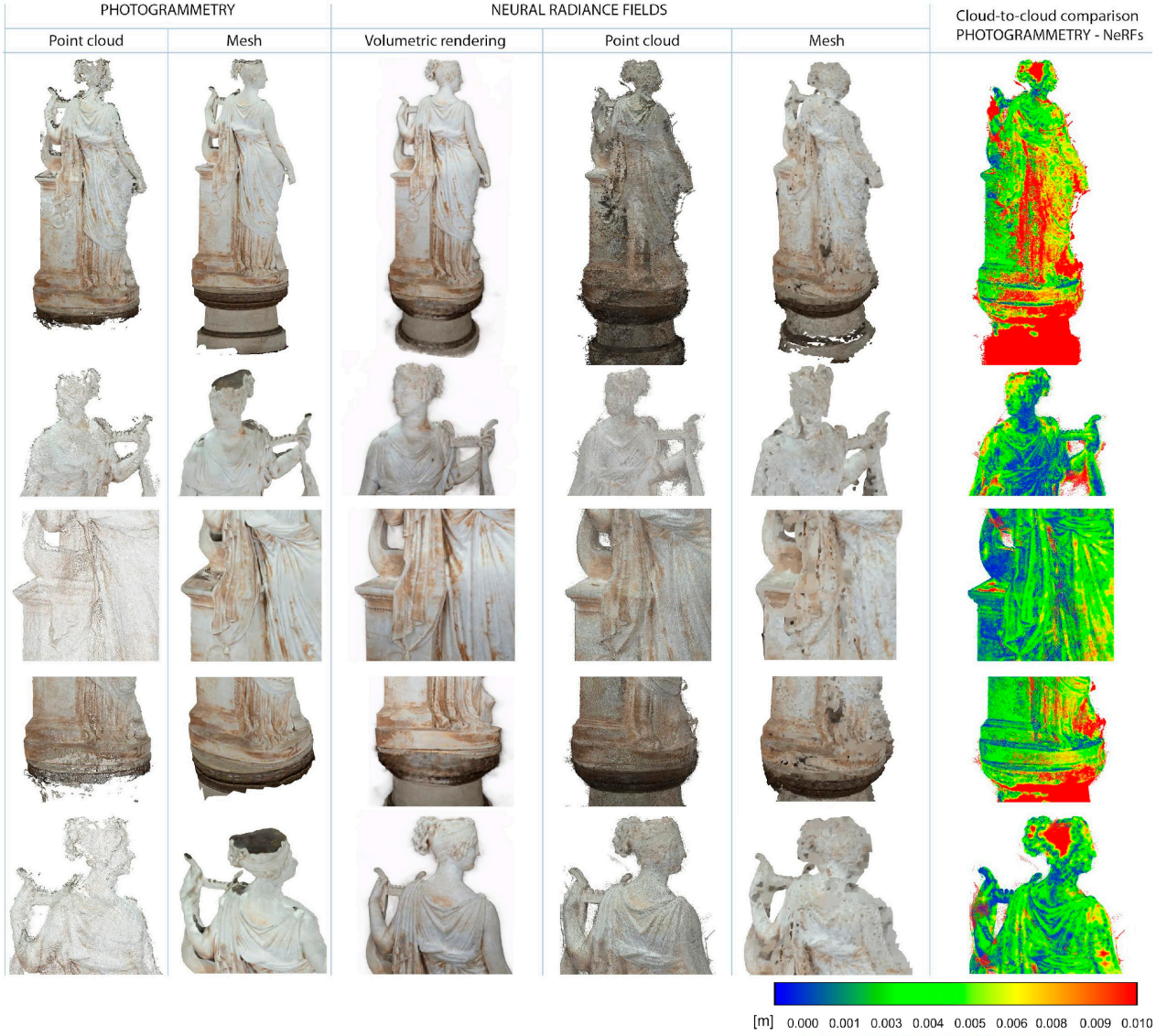


Figure 14. Analysis of the deviation between NeRF and photogrammetric processing results. Dataset 4: 116 images with 3× downscale.

Figure 15. From left to right: the right side of the eagle's chest is bathed in light, while in the second image, the reflective part shifts towards the left. In the third image, the light centers and descends lower on the sculpture.

Figure 16. Photogrammetry and NeRF comparison results of the eagle-shaped lectern.

Figure 17. Photogrammetry and NeRF comparison results of the eagle-shaped lectern.

Concerning the case study of Caprona Tower, when comparing the results obtained through the use of NeRFs with those produced via photogrammetry, a striking disparity becomes evident in terms of representing the vegetation surrounding the tower (Figures 18 and 19). Notably, there are low bushes scattered about, and certain parts of the mesh model exhibit holes or pronounced sharpness, which are rendered exceptionally well-defined in terms of volumetric representation. This discrepancy is particularly noticeable when observing the tree on the left side of the tower and some of the lower bushes.

Figure 18. Photogrammetry and NeRF comparison results for Caprona Tower.

Figure 19. Photogrammetry and NeRF comparison results for Caprona Tower.

However, it is essential to acknowledge that while NeRFs excel in capturing the volumetric details of natural elements, such as the aforementioned vegetation, they may fall short in representing the finer details of the tower's masonry texture. In contrast, photogrammetry excels in capturing the intricate texture of the tower's facades.

Given this disparity in strengths and weaknesses, a possible solution may involve integrating the two models. Combining the precision of NeRFs in representing natural elements with the texture detail captured by photogrammetry could lead to a more comprehensive and accurate overall representation of the scene. This fusion of techniques could strike a balance between volumetric accuracy and textural fidelity, offering a more holistic view of the subject matter.

5. Discussion

Considering the outcomes of our investigation, the discussion comparing NeRFs and photogrammetry can be framed around several key factors, including model description, data processing time, and possible conversion to other forms of representation. In any case, a shared characteristic between these two techniques is represented by the reconstruction of internal and external camera orientation parameters. As a result, ensuring that the model

maintains metric correctness and accurate scaling necessitates processes like georeferencing or the utilization of known measurements. The main differences between the two methods are explained below.

Model Description. The comparisons between these two techniques reveal that, with respect to characterizing the shape of models, NeRFs introduce a unique capability. It allows the appearance of materials to be contingent on the observer's perspective, an attribute that proves exceptionally powerful when dealing with surfaces that are not opaque but rather transparent or reflective and when handling textures that are uniformly consistent. These are challenging scenarios for standard photogrammetry to replicate. Moreover, the results on the pilot case of the Tersicore dataset demonstrate that NeRF models restitute finer details and parts as the number, or the resolution of the input images decreases.

Representation. A notable difference in the model output between NeRF and photogrammetry lies in their representation: NeRFs generate a continuous volumetric model, while photogrammetry produces a dense cloud or mesh-based model with discrete textured surfaces. This contrast in representation can significantly influence the suitability of each method for specific use cases, although many techniques for extracting meshes or point clouds from neural rendering are being proposed [42].

Data Processing Times. When it comes to the time required for processing data, photogrammetry typically demands a longer duration to generate a textured mesh from the same set of images. In contrast, NeRF training, although not instantaneous, is relatively swift, and can be usually completed in approximately 30 min to an hour contingent on the initial image file size. Furthermore, NeRFs offer the capability to explore the model even before training reaches 100%. However, it is important to note that NeRFs face difficulties when processing very high-resolution images.

Sensitivity to reduction in number and resolution of images and trade-off with noise. NeRFs are generally less sensitive to a reduced number of input images or to image downscaling than photogrammetry. In other words, NeRF models can maintain a more consistent level of detail in the 3D reconstruction, even with a smaller image dataset. However, it is important to note that as the number and resolution of images decrease, the NeRF point cloud can become noisier.

Conversion to Other Representations and Exporting Extension. A distinction between these methods pertains to their openness. Standard photogrammetry, particularly when based on Structure-from-Motion (SfM), adopts an open system approach. This allows for relatively straightforward conversion to alternative forms of representation and the exportation of models to various file formats. On the other hand, NeRFs operate as closed systems, making the conversion to alternative representations and the export to different file extensions still challenging.

In conclusion, it must be noted that, as for the case of the Caprona Tower case study, a combination of NeRF and photogrammetry might offer a complementary approach, leveraging the strengths of each method for more comprehensive and versatile 3D modeling solutions (e.g., better texture description of planar surfaces for photogrammetry vs. better description of reflecting objects and finer details as vegetation for NeRF). This hybrid approach could potentially yield superior results by addressing the limitations of each technique and opening up new possibilities for diverse applications in fields such as computer vision, augmented reality, and environmental mapping.

6. Conclusions

In this paper, we provide a comparative evaluation of NeRF and photogrammetry for applications in the cultural heritage domain. Our processing results, based on the pilot case study of the Terpsichore statue, indicate that as the input data and image resolution decrease on the same initial set of images, NeRFs exhibits better preservation of both completeness and material description compared to photogrammetry. Consequently, we recommend the use of NeRFs for datasets with limited images or low resolution, provided that the camera poses for the initial image set are known. This approach holds great

significance in scenarios requiring large-scale area mapping conducted via aerial surveying, especially during emergency situations in which the swift acquisition of extensive survey data is critical and resources and accessibility are constrained. Based on the results of the three case studies considered, a comparison between photogrammetry and NeRF is conducted in Section 5 in terms of the following: model description and representation, data processing times, sensitivity to reduction in number and resolution of images and trade-off with noise, conversion to other representations, and exporting extension.

Despite being a more recent technique still under development compared to the well-established photogrammetry, NeRFs display greater potential in terms of describing material characteristics. This technology may find applications in enhancing the representation of view-dependent materials and objects with intricate details, especially in cases of a limited number of input images. Furthermore, the rendering property of NeRF may appear particularly useful for materials featuring homogeneous textures.

In the future, the integration of NeRF with photogrammetry could complement the limitations of each method, resulting in a more comprehensive and accurate representation of cultural heritage objects to be employed by restorers and conservators in digital or physical preservation protocols. The findings of this research could integrate frameworks and findings of recent studies of the European Commission Directorate-General for Communications Networks, Content and Technology [45].

Our future developments encompass extending these findings to other case studies to validate and expand the range of possibilities for NeRF applications in cultural heritage, architecture, and industrial design. Specifically, we are interested in exploring applications on planar surfaces such as building facades, objects with homogeneous textures like statues or bronze furnishings, and mechanical components constructed from steel or other materials with reflective surfaces.

The very recent applications for creating models categorized into classes through semantic labels (semantic segmentation) [46] or for the reconstruction of in-the-wild scenes [47] further suggest possible developments in this context. The exploration of NeRFs in terms of virtual and augmented reality applications, following the works by Deng et al. [48], is also the subject of ongoing work.

Author Contributions: Conceptualization, V.C.; methodology, V.C.; validation, V.C., D.B., G.C., A.P.; formal analysis, V.C., D.B.; investigation, V.C., D.B.; resources, G.C., A.P.; data curation, V.C., D.B.; writing—original draft preparation, V.C. (Sections 1–6), D.B. (Sections 3 and 4); writing—review and editing, All authors; visualization, V.C., D.B.; supervision, A.P., G.C., L.D.L., P.V.; project administration, A.P., G.C., L.D.L., P.V.; funding acquisition, G.C., A.P., L.D.L. All authors have read and agreed to the published version of the manuscript.

Funding: The work has been supported, by the Joint LAB project LIA Laboratoire International Associé, funded by the French CNRS (National Center for Scientific Research), type: Coordination Action, from January to February 2023. From March 2023 to date, the work was funded by the ASTRO Laboratory of the University of Pisa with funding from the research grants: Università di Pisa—Dipartimento di Ingegneria Civile e Industriale—N°Ordine: 2636/22, Pisa: 07/11/2022, Università di Pisa—Dipartimento di Ingegneria Civile e Industriale—N°Ordine: 1476/23, Pisa: 01/06/2023.

Data Availability Statement: The data that support the findings of this study are available from the author, V.C., upon reasonable request.

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

References

1. Moyano, J.; Nieto-Julián, J.E.; Bienvenido-Huertas, D.; Marín-García, D. Validation of Close-Range Photogrammetry for Architectural and Archaeological Heritage: Analysis of Point Density and 3d Mesh Geometry. *Remote Sens.* **2020**, *12*, 3571. [CrossRef]
2. Wojtkowska, M.; Kedzierski, M.; Delis, P. Validation of Terrestrial Laser Scanning and Artificial Intelligence for Measuring Deformations of Cultural Heritage Structures. *Meas. J. Int. Meas. Confed.* **2021**, *167*, 108291. [CrossRef]

3. Díaz-Rodríguez, N.; Pisoni, G. *Accessible Cultural Heritage through Explainable Artificial Intelligence*; Association for Computing Machinery: New York, NY, USA, 2020; pp. 317–324.
4. Škola, F.; Rizvić, S.; Cozza, M.; Barbieri, L.; Bruno, F.; Skarlatos, D.; Liarokapis, F. Virtual Reality with 360-Video Storytelling in Cultural Heritage: Study of Presence, Engagement, and Immersion. *Sensors* **2020**, *20*, 5851. [CrossRef]
5. Fiorucci, M.; Khoroshiltseva, M.; Pontil, M.; Traviglia, A.; Del Bue, A.; James, S. Machine Learning for Cultural Heritage: A Survey. *Pattern Recognit. Lett.* **2020**, *133*, 102–108. [CrossRef]
6. Pedersen, I.; Gale, N.; Mirza-Babaei, P.; Reid, S. More than Meets the Eye: The Benefits of Augmented Reality and Holographic Displays for Digital Cultural Heritage. *J. Comput. Cult. Herit.* **2017**, *10*, 11. [CrossRef]
7. Trunfio, M.; Lucia, M.D.; Campana, S.; Magnelli, A. Innovating the Cultural Heritage Museum Service Model through Virtual Reality and Augmented Reality: The Effects on the Overall Visitor Experience and Satisfaction. *J. Herit. Tour.* **2022**, *17*, 1–19. [CrossRef]
8. Gros, A.; Guillem, A.; De Luca, L.; Baillieul, É.; Duvocelle, B.; Malavergne, O.; Leroux, L.; Zimmer, T. Faceting the Post-Disaster Built Heritage Reconstruction Process within the Digital Twin Framework for Notre-Dame de Paris. *Sci. Rep.* **2023**, *13*, 5981. [CrossRef]
9. Croce, V.; Caroti, G.; De Luca, L.; Jacquot, K.; Piemonte, A.; Véron, P. From the Semantic Point Cloud to Heritage-Building Information Modeling: A Semiautomatic Approach Exploiting Machine Learning. *Remote Sens.* **2021**, *13*, 461. [CrossRef]
10. Bekele, M.K.; Pierdicca, R.; Frontoni, E.; Malinverni, E.S.; Gain, J. A Survey of Augmented, Virtual, and Mixed Reality for Cultural Heritage. *J. Comput. Cult. Herit.* **2018**, *11*, 7. [CrossRef]
11. Bevilacqua, M.G.; Caroti, G.; Piemonte, A.; Terranova, A.A. Digital Technology and Mechatronic Systems for the Architectural 3D Metric Survey. *Intell. Syst. Control Autom. Sci. Eng.* **2018**, *92*, 161–180. [CrossRef]
12. Rea, P.; Pelliccio, A.; Ottaviano, E.; Saccucci, M. The Heritage Management and Preservation Using the Mechatronic Survey. *Int. J. Archit. Herit.* **2017**, *11*, 1121–1132. [CrossRef]
13. Croce, V.; Caroti, G.; Piemonte, A.; De Luca, L.; Véron, P. H-BIM and Artificial Intelligence: Classification of Architectural Heritage for Semi-Automatic Scan-to-BIM Reconstruction. *Sensors* **2023**, *23*, 2497. [CrossRef] [PubMed]
14. Condorelli, F.; Rinaudo, F. Cultural Heritage Reconstruction from Historical Photographs and Videos. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *XLII-2*, 259–265. [CrossRef]
15. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *arXiv* **2020**, arXiv:2003.08934. [CrossRef]
16. Gao, K.; Gao, Y.; He, H.; Lu, D.; Xu, L.; Li, J. NeRF: Neural Radiance Field in 3D Vision, a Comprehensive Review. *arXiv* **2022**, arXiv:2210.00379.
17. Murtiyoso, A.; Grussenmeyer, P. Initial Assessment on the Use of State-of-the-Art NeRF Neural Network 3D Reconstruction for Heritage Documentation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *XLVIII-M-2-2023*, 1113–1118. [CrossRef]
18. Lorensen, W.E.; Cline, H.E. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. In Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques, Anaheim, CA, USA, 27–31 July 1987; Association for Computing Machinery: New York, NY, USA, 1987; pp. 163–169.
19. Yastikli, N. Documentation of Cultural Heritage Using Digital Photogrammetry and Laser Scanning. *J. Cult. Herit.* **2007**, *8*, 423–427. [CrossRef]
20. James, D.W.; Belblidia, F.; Eckermann, J.E.; Sienz, J. An Innovative Photogrammetry Color Segmentation Based Technique as an Alternative Approach to 3D Scanning for Reverse Engineering Design. *Comput. Aided Des. Appl.* **2017**, *14*, 1–16. [CrossRef]
21. Fiz, J.I.; Martín, P.M.; Cuesta, R.; Subías, E.; Codina, D.; Cartes, A. Examples and Results of Aerial Photogrammetry in Archeology with UAV: Geometric Documentation, High Resolution Multispectral Analysis, Models and 3D Printing. *Drones* **2022**, *6*, 59. [CrossRef]
22. Caroti, G.; Piemonte, A. Integration of Laser Scanning and Photogrammetry in Architecture Survey. Open Issue in Geomatics and Attention to Details. *Commun. Comput. Inf. Sci.* **2020**, *1246*, 170–185. [CrossRef]
23. Guimarães, N.; Pádua, L.; Marques, P.; Silva, N.; Peres, E.; Sousa, J.J. Forestry Remote Sensing from Unmanned Aerial Vehicles: A Review Focusing on the Data, Processing and Potentialities. *Remote Sens.* **2020**, *12*, 1046. [CrossRef]
24. Haneberg, W.C. Using Close Range Terrestrial Digital Photogrammetry for 3-D Rock Slope Modeling and Discontinuity Mapping in the United States. *Bull. Eng. Geol. Environ.* **2008**, *67*, 457–469. [CrossRef]
25. Nicolae, C.; Nocerino, E.; Menna, F.; Remondino, F. Photogrammetry Applied to Problematic Artefacts. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *XL-5*, 451–456. [CrossRef]
26. Ippoliti, E.; Meschini, A.; Sicuranza, F. Digital Photogrammetry and Structure From Motion for Architectural Heritage: Comparison and Integration between Procedures. In *Geospatial Intelligence: Concepts, Methodologies, Tools, and Applications*; IGI Global: Hershey, PA, USA, 2019; Volume 2, pp. 959–1018. ISBN 978-1-5225-8055-3.
27. Barron, J.T.; Mildenhall, B.; Tancik, M.; Hedman, P.; Martin-Brualla, R.; Srinivasan, P.P. Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; IEEE: Montreal, QC, Canada, 2021; pp. 5835–5844.
28. Derksen, D.; Izzo, D. Shadow Neural Radiance Fields for Multi-View Satellite Photogrammetry. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021; IEEE: Nashville, TN, USA, 2021; pp. 1152–1161.

29. Semeraro, F.; Zhang, Y.; Wu, W.; Carroll, P. NeRF Applied to Satellite Imagery for Surface Reconstruction. *arXiv* **2023**, arXiv:2304.04133.
30. Kerr, J.; Fu, L.; Huang, H.; Avigal, Y.; Tancik, M.; Ichnowski, J.; Kanazawa, A.; Goldberg, K. Evo-NeRF: Evolving NeRF for Sequential Robot Grasping of Transparent Objects. In Proceedings of the 6th Conference on Robot Learning, Auckland, New Zealand, 14–18 December 2022.
31. Zhou, A.; Kim, M.J.; Wang, L.; Florence, P.; Finn, C. NeRF in the Palm of Your Hand: Corrective Augmentation for Robotics via Novel-View Synthesis. *arXiv* **2023**, arXiv:2301.08556.
32. Adamkiewicz, M.; Chen, T.; Caccavale, A.; Gardner, R.; Culbertson, P.; Bohg, J.; Schwager, M. Vision-Only Robot Navigation in a Neural Radiance World. *IEEE Robot. Autom. Lett.* **2022**, *7*, 4606–4613. [CrossRef]
33. Li, Y.; Lin, Z.-H.; Forsyth, D.; Huang, J.-B.; Wang, S. ClimateNeRF: Extreme Weather Synthesis in Neural Radiance Field. *arXiv* **2023**, arXiv:2211.13226.
34. Mergy, A.; Lecuyer, G.; Derksen, D.; Izzo, D. Vision-Based Neural Scene Representations for Spacecraft. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021; IEEE: Nashville, TN, USA, 2021; pp. 2002–2011.
35. Gafni, G.; Thies, J.; Zollhöfer, M.; Nießner, M. Dynamic Neural Radiance Fields for Monocular 4D Facial Avatar Reconstruction. *arXiv* **2023**, arXiv:2012.03065.
36. Mazzacca, G.; Karami, A.; Rigon, S.; Farella, E.M.; Trybala, P.; Remondino, F. NeRF for Heritage 3D Reconstruction. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *XLVIII-M-2-2023*, 1051–1058. [CrossRef]
37. Condorelli, F.; Rinaudo, F.; Salvadore, F.; Tagliaventi, S. A Comparison between 3D Reconstruction Using NERF Neural Networks and MVS Algorithms on Cultural Heritage Images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *XLIII-B2-2021*, 565–570. [CrossRef]
38. Vandenabeele, L.; Häcki, M.; Pfister, M. Crowd-Sourced Surveying for Building Archaeology: The Potential of Structure From Motion (SFM) and Neural Radiance Fields (NERF). *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *XLVIII-M-2-2023*, 1599–1605. [CrossRef]
39. Balloni, E.; Gorgoglione, L.; Paolanti, M.; Mancini, A.; Pierdicca, R. Few Shot Photogrametry: A Comparison Between Nerf and Mvs-Sfm for the Documentation of Cultural Heritage. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *48M2*, 155–162. [CrossRef]
40. Pansoni, S.; Tiribelli, S.; Paolanti, M.; Stefano, F.D.; Frontoni, E.; Malinverni, E.S.; Giovanola, B. Artificial Intelligence and Cultural Heritage: Design and Assessment of an Ethical Framework. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *XLVIII-M-2-2023*, 1149–1155. [CrossRef]
41. Croce, V.; Caroti, G.; De Luca, L.; Piemonte, A.; Véron, P. Neural Radiance Fields (NERF): Review and Potential Applications to Digital Cultural Heritage. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *XLVIII-M-2-2023*, 453–460. [CrossRef]
42. Billi, D.; Rechichi, P.; Montalbano, G.; Croce, V. La Torre Degli Upezzinghi a Caprona: Analisi storico-archivistica e rilievo digitale per la documentazione dell'evoluzione temporale. In *Defensive Architecture of the Mediterranean, Proceedings of the International Conference on Fortifications of the Mediterranean Coast FORTMED 2023, Pisa, Italy, 23–25 March 2023*; Bevilacqua, M.G., Ulivieri, D., Eds.; Pisa University Press: Pisa, Italy, 2023; Volume XIII, pp. 391–400.
43. Tancik, M.; Weber, E.; Ng, E.; Li, R.; Yi, B.; Kerr, J.; Wang, T.; Kristoffersen, A.; Austin, J.; Salahi, K.; et al. Nerfstudio: A Modular Framework for Neural Radiance Field Development. In Proceedings of the SIGGRAPH'23: Special Interest Group on Computer Graphics and Interactive Techniques Conference, Los Angeles, CA, USA, 6–10 August 2023.
44. Kazhdan, M.; Bolitho, M.; Hoppe, H. Poisson Surface Reconstruction. In Proceedings of the Eurographics Symposium on Geometry Processing, Sardinia, Italy, 26–28 June 2006.
45. European Commission. *Directorate-General for Communications Networks, Content and Technology, Study on Quality in 3D Digitisation of Tangible Cultural Heritage—Mapping Parameters, Formats, Standards, Benchmarks, Methodologies, and Guidelines—Executive Summary*; Publications Office of the European Union: Luxembourg, 2022; Available online: https://data.europa.eu/doi/10.2759/581678 (accessed on 22 December 2023).
46. Zhi, S.; Laidlow, T.; Leutenegger, S.; Davison, A.J. In-Place Scene Labelling and Understanding with Implicit Scene Representation. *arXiv* **2021**, arXiv:2103.15875.
47. Martin-Brualla, R.; Radwan, N.; Sajjadi, M.S.M.; Barron, J.T.; Dosovitskiy, A.; Duckworth, D. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. *arXiv* **2021**, arXiv:2008.02268.
48. Deng, N.; He, Z.; Ye, J.; Duinkharjav, B.; Chakravarthula, P.; Yang, X.; Sun, Q. FoV-NeRF: Foveated Neural Radiance Fields for Virtual Reality. *IEEE Trans. Visual. Comput. Graph.* **2022**, *28*, 3854–3864. [CrossRef]