Article

Assessment of 3D Model for Photogrammetric Purposes Using AI Tools Based on NeRF Algorithm

Massimiliano Pepe 1,*, Vincenzo Saverio Alfio 2 and Domenica Costantino 2

1 Department of Engineering and Geology (InGeo), “G. d’Annunzio” University of Chieti-Pescara, Viale Pindaro 42, 65127 Pescara, Italy
2 Dipartimento di Ingegneria Civile, Ambientale, del Territorio, Edile e di Chimica, Polytechnic University of Bari, Via E. Orabona 4, 70125 Bari, Italy; vincenzosaverio.alfio@poliba.it (V.S.A.); domenica.costantino@poliba.it (D.C.)

* Correspondence: massimiliano.pepe@unich.it

Abstract: The aim of the paper is to analyse the performance of the Neural Radiance Field (NeRF) algorithm, implemented in Instant-NGP software, for photogrammetric purposes. To achieve this aim, several datasets with different characteristics were analysed, taking into account object size, image acquisition technique and geometric configuration of the images. The NeRF algorithm proved to be effective in the construction of the 3D models; in other words, in Instant-NGP it was possible to obtain realistic 3D models in a detailed manner and very quickly, even in rather weak geometric configurations of the images. The performance obtained in the latter environment was compared with that achieved by two software packages, one widely used in the photogrammetric field, Agisoft Metashape, and one open source, Colmap. The comparison showed encouraging results in building 3D models, especially under weak geometry conditions; although, the geometric description of objects under point clouds or meshes needs improvement for use in the photogrammetric field.

Keywords: NeRF; photogrammetry; Instant-NGP; SfM; 3D models; C2C; AI

1. Introduction

The building of 3D models from images is a growing research topic thanks to the development of algorithms and tools that make the modelling process easy to automate, accurate and detailed. Indeed, the building of realistic 3D models involves numerous fields of application, such as cultural heritage [1–3], engineering [4], urban planning and architecture [5], medicine [6], virtual reality [7], etc.

In order to build realistic models, new algorithms and increasingly high-performance approaches are being developed every day. A widely used algorithm in building 3D models is the Scale Invariant Feature Transformation (SIFT), which is applied to extract connection points from multiple source images [8]; the success of this algorithm is due to the fact that SIFT works reliably under very different radiometric and geometric conditions [9]. However, SIFT imposes a great computational burden, especially for real-time systems such as visual odometry or low-power devices such as mobile phones; as a result, the scientific community has led an intensive search for substitutes with lower computational cost [10]. Another algorithm used to build 3D models is the Speeded Up Robust Features (SURF) that uses an integer approximation of the determinant of Hessian blob detector, which can be computed with three integer operations using a precomputed integral image [11]. An efficient alternative to the SIFT or SURF is the ORB [10], acronym of Oriented FAST and Rotated BRIEF [12], which is rotation invariant and resistant to noise and two orders of magnitude faster than the SIFT. These algorithms are widely used in the photogrammetric approach using Structure from Motion (SfM) in order to reconstruct the position of cameras and generates a sparse point cloud. In general, the SFM pipeline can be summarized in the...
following main steps [13,14]: (i) find interest points in each image; (ii) find candidate correspondences (match descriptors for each interest point); (iii) perform geometric verification of correspondences (RANdom SAmple Consensus-RANSAC and fundamental matrix); and (iv) solve for 3D points and camera that minimize reprojection error. SfM and Multi View Stereo (MVS) are a group of techniques that use stereo correspondence as their main cue and use more than two images [15] allowing for dense point clouds to be obtained and, of consequence, to be used with success in several applications [16,17].

More recent deep-learning-based novel view synthesis methods have the advantages of end-to-end training and producing distributed representations that can be easily used in other neural-net based downstream tasks. An example of this approach is the Neural Radiance Field (NeRF), which use volume rendering, i.e., a collection of methods used in computer graphics and scientific visualization to create a 2D projection from a discretely sampled 3D data set, with implicit neural scene representation via Multi-Layer Perceptrons (MLPs) [18]. NeRF involves training artificial intelligence (AI) algorithms to enable the creation of 3D objects from two-dimensional photos. This latter algorithm represents a scene using a fully connected (non-convolutional) deep network, whose input is a single continuous 5D coordinate (spatial location, \( x, y, z \) and viewing direction \( \theta, \phi \)) and whose output is the volume density and view-dependent emitted radiance at that spatial location [19]. Indeed, using a 5D vector-valued function as input, the algorithm outputs a single volume density and view-dependent RGB colour. The power of the neural field lies in the fact that it is able to produce different representations for the same point when viewed from different angles, and as a result, is able to capture various light effects, such as reflections and transparencies, making it ideal for rendering different views of the same scene. In this way, it is possible to obtain a much better representation than the voxel grid (a value on a regular grid in three-dimensional space) or mesh. Since NeRF was originally proposed, there has been a great deal of interest in this approach as demonstrated by the large number of papers that take into account this algorithm. Indeed, there are many variations to this algorithm that have improved the quality and speed of 3D model building. Meng et al., 2021 [20] introduce GNeRF, a framework to connect Generative Adversarial Networks (GAN) with NeRF reconstruction for the complex scenarios with unknown and even randomly initialized camera poses. In this latter approach, the authors identify a novel two-phases end-to-end framework; the first phase takes the use of GANs into a new realm for optimizing coarse camera poses and radiance fields jointly, while the second phase refines them with additional photometric loss. Muller et al., 2022 [21], were able to speed up training from hours to a couple of seconds; to achieve this, the authors use typical neural fields, also known as neural graph primitives, combined with a new representation of the input called multi-resolution hash coding that is able to utilize small neural networks and, consequently, reduce the total number of floating-point operations required. A comprehensive review of the contribution of NeRF in 3D Vision can be found in Gao et al., 2022 [22].

1.1. Background and Motivation

The comparison of 3D reconstructions obtained by means of AI, in particular NeRF Neural Networks, and by SfM and MVS algorithms is a subject of interest in the scientific community in order to assess the quality of NeRF in photogrammetric applications [23]. Condorelli et al., 2021 [24], show how NeRF networks, although computationally demanding, can be an interesting alternative or complementary methodology, especially in cases where classical photogrammetric techniques do not provide satisfactory results. In the test performed on Tour Saint Jacques, the difference between the mesh model generated with NeRF and one based on Colmap is a few centimetres. Murtyioso et al., 2023 [25], showed through the comparison made between two datasets how in terms of geometric accuracy, the NeRF is still far from meeting the requirements of documentation at a high level of detail since the amount of noise generated in the resulting point cloud is still too great. Vandenabeele et al., 2023 [26], discuss the results achieved by a metric comparison between
the model generated starting from Terrestrial Laser Scanning-TLS survey (Leica RTC 360) and another from images using the NeRF algorithm; taking into consideration the churches of Padua, Lisieux and Tournus, the authors showed that the differences between the two models was within 1.25 cm of 56%, 46% and 21%, respectively.

Therefore, taking into account the state-of-the-art NeRF applications in the photogrammetric field, the aim of the paper is to evaluate the performance of this algorithm in the construction of 3D models and make a comparison with the established approach in the literature based on the use of SfM and MVS algorithms. In particular, datasets containing images of complex objects, such as in the field of cultural heritage, are taken into consideration. In fact, the results in the literature are conflicting and do not clearly show the quality of 3D models for photogrammetric purposes. On the basis of the results obtained in this experimentation on several datasets, the authors suggest evaluating the qualities of the 3D models generated by the NeRF algorithm in terms of metric accuracy, texture and geometric continuity of the model.

1.2. Overview of the Paper

The main idea of the work concerns the identification of an analysis methodology and the selection of appropriate datasets for the evaluation for photogrammetric purposes of the NeRF algorithm implemented in Instant-NGP. The paper is organized in the following section: Section 2 describes the method used to investigate the quality of 3D models on three specific datasets while Section 3 reports the results of the experimentation. The discussion and conclusions are summarized at the end of the paper.

2. Materials and Methods

Simplicity, speed, open source, accuracy, and high automation play an important role in the choice of software. The Instant-NGP software, which uses the NeRF algorithm, was used to evaluate performance in the building of 3D models; in addition, the results obtained with this model were compared with two other software, Colmap and Agisoft Metashape. Instant-NGP software enables a voxel model, capable of representing complex real-world geometry and appearance as well as lending itself well to gradient-based optimisation using projected images. In this environment, MLP predicts the volume density $\sigma$ and the view-dependent emitted radiance or colour $c = [R, G, B]$ at that spatial location for volumetric rendering according the following relation [27]:

$$\begin{bmatrix} R, G, B, \sigma \end{bmatrix} = F_{\Theta}(x, y, z, \theta, \varphi)$$

(1)

where $\Theta = \{ Wi, bi \}$ represent weights and biases, respectively, of the MLP. Practically, the procedure to build the 3D model is divided in two main steps. In the first step, by the use of the algorithm called “Colmap2NeRF”, a file called “transforms” is generated. Indeed, Instant-NGP requires Colmap to determine the position of the cameras. To achieve this aim, it is necessary to run a specific script. In particular, once Anaconda is opened, the following script was used for all the datasets considered:

```
python colmap2nerf.py --colmap_matcher exhaustive --run_colmap --aabb_scale 8 --images images
```

The aabb_scale parameter specifies the extent of the scene, defaulting to 1; that is, the scene is scaled such that the camera positions are at an average distance of 1 unit from the origin. For small synthetic scenes such as the original NeRF dataset, the default aabb_scale of 1 is ideal and leads to the fastest training. By setting aabb_scale to a power greater than 2 (up to a maximum of 128), the NeRF model will extend the beams to a much larger selection rectangle. It should be noted that the choice of this parameter may slightly influence the training speed. The parameter specifies the extent of the scene, defaulting to 1; that is, the scene is scaled such that the camera positions are at an average distance of 1 unit from the origin. For small synthetic scenes such as the original NeRF dataset, the default aabb_scale of 1 is ideal and leads to fastest training. By setting aabb_scale to a power greater than 2 (up to a maximum of 128), the NeRF model will extend the beams to
a much larger selection rectangle. It should be noted that the choice of this parameter may slightly influence the training speed. In the second step, uploading the transformed JSON file in Instant-NGP, it is possible to visualise the construction of the model. In addition, in this latter environment, there is a tool capable of transforming the voxel into a 3D mesh model and select the region of interest.

Colmap [28] is an open-source software, end-to-end image-based 3D reconstruction pipeline, i.e., SFM and MVS. Colmap uses geometry-based MVS systems based on the incremental SFM pipeline (correspondence estimation, pose estimation, triangulation and bundle adjustment). Colmap was chosen since the NeRF acquires image positions from this software, which is open source and successfully used in many photogrammetric applications [29,30].

Agisoft Metashape software [31] is one of the most widely used commercial software packages for 3D model reconstruction based on SfM-MVS algorithms. To build a 3D model according to the mesh, the pipeline is [29]: (i) image acquisition; (ii) feature detection, matching, triangulation (or align photos); (iii) sparse reconstruction, bundle adjustment (or point cloud generation); (iv) dense correspondence matching (or dense cloud generation); and (v) mesh/surface generation. In particular, the processing of the images was realized using the following settings: alignment—high, building dense point cloud—medium, building mesh—arbitrary, and interpolation—disabled.

The main features of the software considered are shown in Table 1.

<table>
<thead>
<tr>
<th>Software</th>
<th>License Type</th>
<th>Point Cloud</th>
<th>Dense Cloud</th>
<th>Voxel</th>
<th>Mesh</th>
<th>Texture</th>
<th>Editing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colmap</td>
<td>GNU GPL v3</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Agisoft Metashape</td>
<td>Commercial</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Instant-NGP</td>
<td>Open Source</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes (minimal)</td>
</tr>
</tbody>
</table>

Nowadays, there are many and increasingly numerous datasets that the scientific community makes available in order to experiment with algorithms, techniques and the quality of photogrammetric models. The first dataset under consideration is called “fox” and aims to assess the quality of the 3D model using converging images of an object according to a semi-hemisphere; in particular, the dataset concerns 53 images of 1080 × 1920 pixels.

A second dataset called “Buddha” concerns images captured at 360 degrees around the small statue [32]; the dataset contains 67 images of 2736 × 1540.

A third dataset called “Tavole Palatine” concerns images acquired from an Unmanned Aerial Vehicle—UAV platform for the photogrammetric survey of an archaeological area [33]; the images were acquired using DJI Mavic 2 Pro equipped with a camera Hasselblad L1D-20c, with sensor CMOS 1” 20 MP (5464 × 3640 pixel). The developed methodology consists of two main steps: (i) data processing and modelling in order to analyse the quality of the 3D model (Section 3.3) and (ii) metric evaluation of a dataset (Section 3.4). The qualitative assessment of the 3D model according to the mesh is established taking into account the continuity of the model, texture, number and time required to generate the 3D model. The point clouds generated in Colmap and Agisoft Metashape are managed in this latter software in order to build a 3D mesh model. An intrinsic analysis on the metric quality of the model cannot be performed within Colmap and Instant-NGP since they are not equipped with tools, such as Agisoft Metashape, to provide model accuracy based on the use of Ground Control Points (GCPs). Therefore, in order to assess the metric quality, the point clouds generated with the different software are compared with that generated by Agisoft Metashape. However, before making the comparison, it is necessary to transform the model generated in Instant-NGP into mesh and then into point cloud, e.g., using specific tools implemented in Cloud Compare or Meshlab software [34]. In these
latter environments, it is possible to carry out an Helmert transformation, i.e., performing a roto translation with scale factor (conformal transformation) of the point cloud obtained from the mesh generated in Instant-NGP in an arbitrary coordinate system \((xyz)\) than the point cloud generated in Agisoft Metashape in the reference coordinate system \((XYZ)\):

\[
X = T + \lambda Rx
\]

where \(T\) is the vector of translations and \(R\) is the rotation matrix, defined as a function of three rotation parameters \([35]\). In order to determine the seven parameters, a number of seven equations must be known. These equations can be formed with the help of 3D coordinates of at least three points, known in both coordinate systems. This task can be performed on the point clouds using the Cloud-to-Cloud (C2C) tool implemented in Cloud Compare software that calculates the Euclidean distance between individual points in a reference point cloud and the respective nearest neighbour point (NN) in the target point cloud.

Data processing of the datasets were performed with a laptop Alienware m15 R4 and the main features of this device are: (i) CPU: Intel Core i7-10870H CPU @ 2.20 GHz, 2.21 GHz; (ii) GPU: Nvidia GeForce RTX 3070; (iii) RAM: 16 GB; and (iv) Storage: 1 TB SSD.

### 3. Results

#### 3.1. Dataset “Fox”

The first dataset taken into consideration is the one called “Fox”. Using the pipeline of each software chosen for analysis, it was possible to obtain the 3D models; indeed, Table 2 shows some features of the image processing and the images of the 3D model generated in each software.

**Table 2.** Features of the 3D processing obtained with the different software using images of “Fox” dataset.

<table>
<thead>
<tr>
<th></th>
<th>Colmap</th>
<th>Agisoft Metashape</th>
<th>Instant-NGP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Textured model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time for building mesh (s)</td>
<td>393 (building dense cloud) + 9 s (building mesh)</td>
<td>226 (building dense cloud) + 8 s (building mesh)</td>
<td>225 (Colmap2NERF) + 16 (volumetric model) + 2 (building mesh)</td>
</tr>
<tr>
<td>Dense point cloud</td>
<td>786,803</td>
<td>542,848</td>
<td>9,342,630</td>
</tr>
<tr>
<td>Mesh</td>
<td>1,787,782</td>
<td>1,320,445</td>
<td></td>
</tr>
</tbody>
</table>

As shown in Table 2, the quality of the NeRF processing is very accurate if compared with other software. Indeed, Agisoft Metashape is useful to reconstruct the geometry of the fox but with a lower quality than with Instant-NGP.

Even the Colmap software was not able to reconstruct the geometry of the object under consideration; indeed, from the dense point cloud, it was not possible to generate a mesh model in the part of the object under consideration.
3.2. Dataset “Tavole Palatine”

The processing taking into account rather large dimensions of the structure was conducted using images acquired from UAV platforms. The structure to be surveyed is rather complex and irregular; in addition, nadiral and oblique images were taken to cover the archaeological area in order to ensure adequate overlap and sidelap values. Figure 1 shows the 3D model and geometric configuration of the cameras within Instant-NGP software. On the left side, a toolbar is present that includes many features, such as:

- Comprehensive controls for interactively exploring neural graphics primitives (training, loss graph, etc.);
- Virtual reality (VR) mode for viewing neural graphics primitives through a virtual-reality headset;
- Saving and loading “snapshots” so you can share your graphics primitives on the internet;
- A camera path editor to create videos;
- NeRF->Mesh and SDF->Mesh conversion;
- Camera pose and lens optimization.

Figure 1. Screen shot of Instant-NGP where it is possible to note the camera network on the 3D model of the “Tavole Palatine” site and on the left side the toolbar containing the different elements for setting and managing the processing parameters.

Therefore, a significant challenge was to reconstruct a model in all its parts, especially in areas with a weak geometric configuration, such as the side parts of the columns and the connection of the columns with the architrave.

In Table 3, the results of the processing in the different software were reported. As shown in Figure 1 and Table 3, Instant-NGP was able to reconstruct every detail of the structure. In the other two software packages considered, there were some problems in the lintel pillar connection and in the case of Agisoft Metashape, even part of the column was not reconstructed accurately. In addition, the time in order to build the 3D model according to the mesh in Instant-NGP is faster than the other two software under consideration.

3.3. Dataset “Buddha”

The original dataset consists of 67 images around the small statue. However, in order to highlight the behaviour of the software with a weak geometry, the photos observing the upper part of the structure were removed. Consequently, the dataset was reduced to 28 images. Despite this weak configuration, Instant-NGP was capable of obtaining a
continuous model of the structure taken into consideration, also in the upper part of the statue (Figure 2). However, it should be noted that in the case of continuous acquisition according to the close-range photogrammetry technique, the algorithm generates voxels that do not allow a clear observation of the scene. Only through a manual process of eliminating this noise, is it possible to transform it into mesh.

Table 3. Features of the 3D processing obtained with the different software using images of “Tavole Palatine” dataset.

<table>
<thead>
<tr>
<th></th>
<th>Colmap</th>
<th>Agisoft Metashape</th>
<th>Instant-NGP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Textured model</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>Time (s)</td>
<td>1402 (building dense cloud) + 57 s (building mesh)</td>
<td>1219 (building dense cloud) + 59 s (building mesh)</td>
<td>240 (Colmap2NERF) + 44 (volumetric model) + 2 (building mesh)</td>
</tr>
<tr>
<td>Dense point cloud mesh</td>
<td>1,440,339</td>
<td>1,451,064</td>
<td>2,696,000</td>
</tr>
</tbody>
</table>

Figure 2. Camera network on statue of “Buddha” site using Instant-NGP.

The results of the processing obtained by the use of Colmap and Instant-NGP using only 28 images of the Buddha dataset are reported below (Table 4).

In this latter case, both Colmap and Agisoft Metashape were unable to reconstruct the upper part of the statue.

On the other hand, the quality of the model generated in Instant-NGP is inferior to that generated with Colmap and Agisoft Metashape; in fact, observing the face of the statue, it is easy to note that the 3D model is much more defined than that generated by Instant-NGP.

3.4. Quality Metric Comparison of the Three Datasets

The metric evaluation of the models generated by the different software was assessed by comparing the point clouds generated by several software. Instant-NGP software generates mesh which subsequently imported into Meshlab can be transformed into point cloud using the following steps: Filters > Sampling > Point Cloud Simplification.
Therefore, the point cloud of the “fox” dataset derived from Instant-NGP (Figure 3a) was compared with that generated in Agisoft Metashape (Figure 3b). The results of the comparison performed in Cloud Compare software are shown in Figure 3c. From Figure 3c, a significant difference between the two models emerges. In fact, Instant-NGP also generates points (erroneously) within the structure. This means that there are important differences; however, observing Figure 3c, it can be seen that the greatest differences between the two models are concentrated at points within the analysed model.

Therefore, the point cloud of the “fox” dataset derived from Instant-NGP (Figure 3a) was compared with that generated in Agisoft Metashape (Figure 3b). The results of the comparison performed in Cloud Compare software are shown in Figure 3c. From Figure 3c, a significant difference between the two models emerges. In fact, Instant-NGP also generates points (erroneously) within the structure. This means that there are important differences; however, observing Figure 3c, it can be seen that the greatest differences between the two models are concentrated at points within the analysed model.

<table>
<thead>
<tr>
<th>Time (s)</th>
<th>Colmap</th>
<th>Agisoft Metashape</th>
<th>Instant-NGP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dense point cloud</td>
<td>201 (building dense cloud) + 11 s (building mesh)</td>
<td>113 s (building dense cloud) + 9 s (building mesh)</td>
<td>46 (Colmap2NERF) + 14 (volumetric model) + 2 (building mesh)</td>
</tr>
<tr>
<td>Mesh</td>
<td>1,888,471, 2,926,632</td>
<td>826,138, 2,173,182</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>501,928</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Features of the 3D processing obtained with the different software using images of “Buddha” dataset.

Figure 3. Comparison of the point clouds of the “Fox” dataset: point cloud derived from Instant-NGP (a); point cloud generated in Agisoft Metashape (b) and metric comparison (c).
The quality of the 3D model used as reference was evaluated by the Total Error, i.e., the Root Mean Square Error (RMSE) for \( x, y, z \) coordinates, which can be calculated by following formula:

\[
\text{Total Error} = \sqrt{\frac{\sum_{i=1}^{n} (x_{i,\text{est}} - x_{i})^2 + (y_{i,\text{est}} - y_{i})^2 + (z_{i,\text{est}} - z_{i})^2}{n}}
\]

where \( x_{i} \) is the input value for \( x \) coordinate, \( y_{i} \) the input value for \( y \) and \( z_{i} \) the input value for \( z \) coordinate for the \( i \)-th camera position, while \((x_{i,\text{est}}, y_{i,\text{est}}, z_{i,\text{est}})\) corresponds to its estimated position. The RMSE error evaluated using the eight GCPs was 0.015 m. The comparison of the point cloud generated and based on the NeRF algorithm (Figure 4a) with the one generated by Agisoft Metashape (reference) (Figure 4b) showed a mean difference of 4 cm with an RMS of 10 cm (Figure 4c); in particular, about 88% of the difference is contained in 5 cm. The most important differences between the two models occur in the part between vegetation and small structures, while in the main structure (column and architrave), there are very small differences. This behaviour can be explained by the model generated in Instant-NGP being rather noisy in the definition of the terrain (model base) and, more in general, in a loss of detail definition in the various processes leading to the construction of the point cloud.

![Figure 4. Comparison of the point clouds of the “Palatine” dataset: point cloud derived from Instant-NGP (a); point cloud generated in Agisoft Metashape (b) and metric comparison (c).](image)

As far as the statue dataset is concerned, in order to compare point clouds, it was necessary to scale the model through the use of control points. Subsequently, the model generated in Instant-NGP (Figure 5a), appropriately scaled, was compared with that obtained in Agisoft Metashape (Figure 5b) and imported in Cloud Compare software in order to evaluate the distance between the two models (Figure 5c).

The point clouds show important differences because the model generated in the Instant-NGP environment contains voxels and, consequently, mesh and point clouds within the statue envelope. However, if the focus is only on the shell of the statue, then the results obtained are encouraging, even if the model definition is not high, as shown in Figure 5c.

![Figure 5. Comparison of the point clouds of the “Buddha” dataset: point cloud derived from Instant-NGP (a); point cloud generated in Agisoft Metashape (b) and metric comparison (c).](image)
4. Discussion

The Neural Radiance Field algorithm has proven to be an effective tool for the construction of 3D models. In fact, as shown in the different case studies, the Instant-NGP software made it possible to construct a complete 3D model even where very common software in the photogrammetric field such as Colmap and Agisoft Metashape failed. However, the definition of the 3D models generated in the Instant-NGP environment according to mesh or point clouds is not highly defined in detail compared to those generated in the SfM-MVS environment. For the definition of the quality of the model, several parameters must be taken into consideration. In the different case studies, as shown in Figure 6. From the analysis of Figure 6, it can be seen that the aabb_scale value is very influential for values 1 and 128, while optimal values were obtained for aabb_scale values of 8 and 16. In all cases examined, no computational problems were found but there was an increase in noise as the aabb_scale value decreases. In addition, a small noise on the basis of each investigated object was detected; however, this noise can be eliminated not only by choosing an appropriate aabb_scale value but also by graphically selecting the object. The analysed dataset with the highest noise was the Buddha dataset, i.e., the one characterised by a 360-degree acquisition around the object. In this last case, the elimination is rather complex in Instant-NGP; in fact, only after exporting the model created in mesh and importing it into an external editing software, was it possible to eliminate the noise around the object.

Furthermore, the generation of voxels and, in turn, mesh (or point cloud) within the model creates noise that does not allow easy quantification of the error with respect to a reference model. In fact, comparing the point clouds (C2C) obtained in the different case studies described in the paper, the verified distances are important in all datasets taken into consideration. However, taking into consideration only those points that belong to the domain of the reconstructed object, it can be seen that the results obtained from the metric comparison between 3D models are rather encouraging. Indeed, examining a section of the Buddha statue model, it can be seen that the points that generate a high error in the C2C comparison are the voxels that determine the filling of the analysed object (Figure 7a). Furthermore, by extracting two profiles from the point cloud (Figure 7b) generated in Instant-NGP (red colour) and the one generated in Agisoft Metashape (blue colour), it was possible to calculate some statistical parameters relating to the differences found along 13 sections. In particular, the maximum distance reached was 0.07 cm, with a mean value of 0.002 m and a standard deviation of 0.002 m.

![Figure 6. Impact of aabb_scale in 3D model varying from 1 to 128 in the different datasets.](image-url)
Figure 6. Impact of aabb_scale in 3D model varying from 1 to 128 in the different datasets.

Furthermore, the generation of voxels and, in turn, mesh (or point cloud) within the model creates noise that does not allow easy quantification of the error with respect to a reference model. In fact, comparing the point clouds (C2C) obtained in the different case studies described in the paper, the verified distances are important in all datasets taken into consideration. However, taking into consideration only those points that belong to the domain of the reconstructed object, it can be seen that the results obtained from the metric comparison between 3D models are rather encouraging. Indeed, examining a section of the Buddha statue model, it can be seen that the points that generate a high error in the C2C comparison are the voxels that determine the filling of the analysed object (Figure 7a). Furthermore, by extracting two profiles from the point cloud (Figure 7b) generated in Instant-NGP (red colour) and the one generated in Agisoft Metashape (blue colour), it was possible to calculate some statistical parameters relating to the differences found along 13 sections. In particular, the maximum distance reached was 0.07 cm, with a mean value of 0.002 m and a standard deviation of 0.002 m.

Figure 7. Comparison between 3D model of the Buddha generated by Instant-NGP and Agisoft Metashape along a section: C2C (a) and comparison of the two profiles (b).

In the case study of the photogrammetric survey from UAV, the model of the structure generated in Instant-NGP is comparable to that obtained in Agisoft Metashape and shows differences between the two models contained in a few centimetres.

5. Conclusions

The NeRF algorithm implemented in Instant-NGP proved to be a valid algorithm for constructing 3D models, especially in the reconstruction of scenes where there is a rather weak geometric configuration of the cameras. This algorithm is set to revolutionise the world of photogrammetry; however, some critical elements must be taken into consideration. In general, the quality of the 3D model is influenced by two aspects: the presence of diffuse noise, which is also present inside or below the structure to be investigated, and the mesh transformation. In the first case, part of the noise can be removed directly in Instant-NGP using a special tool; furthermore, using dedicated software, such as MeshLab, Cloud Compare, etc., it is possible to eliminate this noise semi-automatically once the transformation from voxel to mesh has been performed. The creation of such noise is particularly evident when objects are acquired by rotating 360 degrees around the object, as shown in the case study of the dataset of the “Buddha” statue. As far as the transformation from voxel to mesh is concerned, this process requires rather high PC performance and becomes unsustainable when the object to be investigated is large. Furthermore, a negative contribution is due to the creation of meshes even within the object, which do not make the process smooth and do not contribute to the morphological description of the structure to be investigated. In addition, the lack of tools for assessing model accuracy is a weakness. Indeed, if the 3D model generated in this way is very realistic, it does not allow a metric evaluation of the model in this environment. To scale and/or georeference the model, it is necessary to perform this operation in other software. For example, the Meshlab software allows the roto-translation operation on the mesh model; the Cloud Compare software allows the georeferencing of the point cloud. Therefore, the development of tools that allow GCPs to be placed directly on the images would give the user greater control over the accuracy and precision of the model.

With regard to the optimal aabb_scale values to be used in the construction of the 3D model, we recommend values of 8 or 16.

A relevant positive aspect of the approach described in the paper for managing photogrammetric data is the growth of open-source software packages; in fact, using Colmap, Instant-NGP, and Meshlab (or Cloud Compare) software, it is possible to obtain
detailed 3D models or to edit the point cloud or mesh. It is desirable in the future, to have the integration of the different functionalities in a single software that, through the use of specific tools, make the photogrammetric process quick, precise and simple. As companies and organisations seek to reduce costs and increase flexibility, open-source technology has rapidly gained traction and popularity in the field of spatial data management in recent years; this means that the development of tools in open software allows for improved functional and operational aspects that both the scientific and professional communities can benefit from.

Taking into account the potential of the NERF algorithm, especially in the field of Cultural Heritage, could be very successful. Indeed, in many contexts, where it is difficult to obtain optimal photogrammetric configurations, due to complex shapes or the complicated morphology of places, the use of AI-based software and tools such as NeRF algorithms could generate a huge advantage in a three-dimensional modelling context. Even with datasets that do not ensure an optimal geometric configuration of the scene to be surveyed, the application of such algorithms would enable the processing of high-performance 3D models, guaranteeing the metric and photorealistic quality of the asset or object being surveyed. For this reason, future research is aimed at developing tools capable of generating detailed, accurate and metrically valid 3D photogrammetric models for use in various fields of application.

**Author Contributions:** Conceptualisation, M.P., V.S.A. and D.C.; methodology, M.P., V.S.A. and D.C.; software, M.P., V.S.A. and D.C.; validation, M.P., V.S.A. and D.C.; formal analysis, M.P., V.S.A. and D.C.; data curation, M.P., V.S.A. and D.C.; and writing—review and editing, M.P., V.S.A. and D.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We want to thank the reviewers for their careful reading of the manuscript and their constructive remarks.

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

**References**


33. Pepe, M.; Alfio, V.S.; Costantino, D.; Scaringi, D. Data for 3D Reconstruction and Point Cloud Classification Using Machine Learning in Cultural Heritage Environment. Data Brief 2022, 42, 108250. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.