

# Comparative Evaluation of NeRF Algorithms on Single Image Dataset for 3D Reconstruction

Francesca Condorelli<sup>1</sup>, Maurizio Perticarini<sup>2</sup>

<sup>1</sup> Free University of Bozen, Italy – francesca.condorelli@unibz.it

<sup>2</sup> University of Padua, Italy – maurizio.perticarini@unipd.it

**KEYWORDS:** NeRF, 3D Reconstruction, Single-Images, Cultural Heritage, Videogames Environment

## ABSTRACT:

The reconstruction of three-dimensional scenes from a single image represents a significant challenge in computer vision, particularly in the context of cultural heritage digitisation, where datasets may be limited or of poor quality. This paper addresses this challenge by conducting a study of the latest and most advanced algorithms for single-image 3D reconstruction, with a focus on applications in cultural heritage conservation. Exploiting different single-image datasets, the research evaluates the strengths and limitations of various artificial intelligence-based algorithms, in particular Neural Radiance Fields (NeRF), in reconstructing detailed 3D models from limited visual data. The study includes experiments on scenarios such as inaccessible or non-existent heritage sites, where traditional photogrammetric methods fail. The results demonstrate the effectiveness of NeRF-based approaches in producing accurate, high-resolution reconstructions suitable for visualisation and metric analysis. The results contribute to advancing the understanding of NeRF-based approaches in handling single-image inputs and offer insights for real-world applications such as object location and immersive content generation.

## 1. INTRODUCTION

The reconstruction of three-dimensional scenes from a single image is a fundamental problem in computer vision with numerous practical applications ranging from robotics and autonomous navigation to augmented reality and virtual tourism. Despite its importance, the inherent challenges associated with inferring the complete 3D geometry and appearance from a single viewpoint make this task particularly difficult. Traditional methods often rely on geometric priors or features, limiting their ability to capture the rich details and structures present in real-world scenes. In recent years, the advent of deep learning techniques, particularly Neural Radiance Fields (NeRF) (Mildenhall, et al., 2020), has revolutionized the field of 3D reconstruction by offering a data-driven approach that directly models the volumetric scene representation. NeRF-based approaches demonstrate impressive capabilities in synthesizing photorealistic novel views and reconstructing detailed 3D scenes from a sparse set of input images. However, most existing studies focus on multi-view settings where multiple images of the same scene are available, neglecting the scenario where only a single image is provided.

From the perspective of metric surveying and architectural representation, especially in relation to 3D modelling of cultural heritage, exploring this open issue is of fundamental importance. Indeed, a dataset suitable for digitising cultural heritage is not always available, but this remains a key step in taking actions on the heritage itself. An example concerns the realisation of a 3D model of heritage that is not visible or not easily accessible, a situation in which it is impossible to perform an adequate metric survey and acquire data correctly. Another situation concerns the three-dimensional reconstruction of cultural heritage that no longer exists because it has been destroyed over time, for which it is necessary to process historical image and video datasets because there is no other way to be able to obtain metric data. A final very challenging situation involves the creation of heritage models that never existed but were only imagined such as illustrations, drawings, or plans.

On these three situations, the authors have already started research following two paths. The first is to modify photogrammetry algorithms to optimise the processing of data

consisting of images that are insufficient for standard photogrammetric processing and of low quality in terms of resolution and lack of metadata (Condorelli et al., 2019). The second way is to experiment with the use of different AI-based algorithms to be combined with photogrammetric algorithms. Some experiments have been carried out with the use of NeRF algorithms to improve the meshing phase, especially in cases where the source images lack certain views of the object or have reflective and metallic surfaces for which the photogrammetric processing fails (Condorelli et al., 2021; Condorelli et al., 2024). There are also interesting experiments with a method called Gaussian Splatting in the field of cultural heritage: a very useful tool for faster visualization of a product derived from Neural Radiance Field (Basso et al., 2024). Other experiments were carried out to improve and increase the size of the starting dataset using text-to-image algorithms (Condorelli 2023).

In this article, we would like to continue along this path, focusing mainly on the problem of obtaining 3D models from datasets formed from a single image.

It will therefore first present an in-depth study of the state of the art of the most recently developed algorithms, classifying them according to the methodology used. After this, the study focuses on the examination of NeRF-based algorithms for processing single images to obtain 3D model meshes. This methodology was applied to several case studies, mainly related to cultural heritage. This is followed by a discussion of the results obtained, highlighting strengths and weaknesses of the method, also in relation to a possible metric analysis of the results, when possible.

## 2. MOTIVATION OF THE RESEARCH: CULTURAL HERITAGE RECONSTRUCTION IN CHALLENGING CONTEXTS

This paper conducts a study of the state of the art of algorithms specifically tailored for single-image 3D reconstruction and presents first experiments of these algorithms on different datasets concerning the cultural heritage field. By leveraging a diverse set of single-image datasets, the aim is to provide comprehensive insights into the strengths and limitations of different AI based algorithms in the context of single-image 3D reconstruction tasks.

In particular, the study focuses on applications related to the context of cultural heritage. Often in this field there is a need to create digital copies quickly and expeditiously, without using highly expensive tools. NeRFs, as already experienced in previous research (Condorelli et al., 2021) are excellent algorithms for obtaining 3D models as an alternative to SfM algorithms, especially in those difficult cases where the latter can fail. This is the case with smooth, reflective surfaces (Croce et al., 2023), or symmetrical objects as we shall see in the cases examined in this research. Again, NeRFs are useful in cases of shots taken with 360 cameras, which speed up the survey process but at the same time reduce the quality of the texture if equirectangular images are used for processing. As we have already seen (Palestini et al., 2024), NeRFs can help to achieve a better result in this case as well. This is why we have chosen to apply them to the very challenging yet very common case of finding oneself with a small initial image dataset. This is in particular the case of cultural heritage that no longer exists for various reasons but still needs to be documented and thus digitised. In these cases, since it has not been possible to carry out a complete survey, there are only a few images available, sometimes just one, which are not suitable for processing for metric purposes with classic photogrammetry. Yet they can be a valuable source of information, especially in those cases where digitised 3D models of cultural heritage are used for visualisation purposes. In particular, as a research group we deal with heritage education through the creation of video games set in culturally rich architectural contexts. It is therefore crucial for our research to deal with innovative solutions to enhance existing algorithms for obtaining 3D models even in the most difficult cases.

### 3. STATE OF THE ART OF THE MOST RECENT SINGLE-SCENE RECONSTRUCTION ALGORITHMS USING AI

Regarding the synthesis of new views, there are several ongoing research efforts. Current methods are divided into different approaches: those utilizing multiple images, those using a single image, those leveraging 3D prompts or semantic information. As known, if images from various viewpoints of an object are available, the hypothesized 3D object can be useful in generating new views. This process can be achieved through multi-view geometry or depth maps. In this era dominated by artificial intelligence, the use of deep neural networks (DNNs) to comprehend spatial dimensions is proving to be the right path. Some studies have focused on creating synthetic images from noisy or gapped depth maps; others have concentrated on generating new perspectives from two or more images characterized by reduced baselines; while others have utilized volumetric rendering to generate implicit voxel clouds (thus not specifying the object's surface or its mathematical and geometric data), which are then used to generate novel images. Methods for generating view synthesis through a single image have been studied for some time, but following different approaches. One of these involves leveraging a large dataset of images paired with 3D and semantic information (depth maps and ground truth information), to train the algorithm on the 3D representation of the object indicated by the source image. This method is characterized by long processing times and the use of 3D acquisition technologies (depth cameras or lidar). Other studies use DNNs for view synthesis but follow an end-to-end process (from image to image), while others exploit learned embeddings that, through the learning phase, enable the algorithm to recognize patterns or classify objects in the image. Some of these studies have limitations because they focus only on certain classes of objects, or consider small viewpoint shifts—resulting in gaps and errors when the shift is larger—or focus on

specific movements such as forward movement (KITTI) (Chen et al., 2019).

Studies aiming to create a 3D object from a single image are related to so-called "zero-shot" processes, which refer to a model's ability to handle tasks for which it has not been explicitly trained. A zero-shot model is capable of creating an image from scratch, one it has never seen before during training, drawing from examples it has seen during training. Two of the most widely used DNN models in these contexts are CLIP (Contrastive Language-Image Pre-Training), a simplified version of ConVIRT trained from scratch, developed by OpenAI, and GAN networks such as DOGAN, CycleGAN, or BigGAN (Pavlo et al., 2023). Other models include ALIGN and BASIC, which have expanded the dataset size, including multilingual data. Additional examples are LiT, Flamingo, and LAION-5B (Schuhmann et al., 2022).

Here are some of the most recent studies:

**SynSin:** An end-to-end project that utilizes a single source image, the BigGAN model, a UNet for image segmentation, and a neural renderer for point cloud to generate other views (Wiles et al., 2020).

**pixelNeRF:** Uses a Convolutional Neural Network (CNN) to support a NeRF to learn a scene prior and generate new views from one or a few original views. The team experimented with its algorithm considering databases such as ShapeNet or DTU (Yu et al., 2021).

**WONDER 3D:** Proposes a cross-domain diffusion model that generates normal maps from different viewpoints and the corresponding-colored image. It utilizes a "geometry-aware normal fusion" algorithm that extracts high-quality surfaces from 2D representations from different viewpoints. It uses CLIP and two different domains: one composed of normal maps and one of diffuse maps (Long et al., 2023).

**MAKE IT 3D:** Uses a NeRF and SDS and has a creation scheme that unfolds in two phases: In the first phase, there is a coarse NeRF reconstruction generated from a single image. From this, novel views are generated from the reference view. In this phase, a descriptive text of the object is also inserted, resulting in a plausible 3D representation with errors in depth and geometry, which are subsequently corrected by a depth prior process. In the second phase, the NeRF is converted into a point cloud with a texture generated from multi-view RGBD based on the reference image. For point cloud rendering, a 2D UNet architecture is used (Tang et al., 2023).

**DreamCraft3D:** Utilizes a 2D image to generate the 3D geometry of the framed object, subsequently improving the texture. Concerning geometry, it uses a view-dependent diffusion model to enhance the process called "score distillation sampling"; for texture improvement—due to this process reducing its quality significantly—a method called "Bootstrapped Score Distillation" is employed. Subsequently, a customized diffusion model called Dreambooth is trained, providing guidance on how the scene or object in the source image should appear from other angles. The two processes reinforce each other: the optimized 3D scene helps improve the diffusion model specific to that scene, which in turn further enhances the quality of the 3D scene (Sun et al., 2023).

**ZERO 1 to 3:** Implements viewpoint control within a diffusion model (such as Stable Diffusion) that, drawing from datasets from the internet, is constrained to certain viewpoints. The model must understand details such as shadows and textures, as well as

the function and structure of the object. To enable this, the model uses two approaches: one based on CLIP combined with camera position information; one in which the source image is concatenated with the new ones—after a denoising process (U-Net)—to help the model not lose important details of the represented object. To reconstruct the 3D, it uses a method called Score Gradient Chaining (SJC) for model improvement, but to decrease the randomness of this improvement, it has implemented the DreamFusion method (which optimizes NeRF but leads to low-quality models and slow processing) (Liu et al., 2023).

**Magic3D:** Improves DreamFusion and is structured in two phases; in the first, using a process with a sparse 3D hash grid structure, an initial 3D model is obtained using a low-resolution diffusion model; in the second phase, the model is optimized and transformed into a textured mesh (Lin et al., 2023).

**LRM (Large Reconstruction Model):** Adopts a highly scalable transformer-based architecture with 500 million trainable parameters to directly predict a neural radiance field (NeRF) from the input image. End-to-end training on a multi-view dataset of approximately 1 million objects, derived from synthetic renderings (Objaverse) and real acquisitions (MVIImgNet), enables the model to learn generalizability and adapt to realistic lighting and noise conditions (Hong et al., 2024).

**TripoSr:** Based on the LRM architecture, it makes improvements to data handling and rendering, model design, and training techniques. Similarly to LRM, it leverages transformer architecture to reconstruct 3D objects from a single image. The image is processed by an encoder extracting global and local information, which is then transformed into a triplanar representation. This transformation is processed by a decoder considering internal relationships and global information of the image to predict RGB and density parameters for each point in 3D space in a triplanar NeRF model (Tochilkin et al., 2024).

#### 4. METHODS

The single-scene 3D reconstruction algorithms analysed are very recent research, some still in the pre-print phase or with the code not yet released or not allowing it to be replicated on custom datasets. This presented a limitation to our analysis as we could not test all the methods present in the state of the art. Our experimentation therefore focused only on methods that have released open access code and that allow it to be replicated on specific custom datasets. However, we intend to complete the research with an exhaustive analysis in the near future.

For our tests, we chose to use both the Threestudio platform (<https://github.com/threestudio-project/threestudio>), which includes the aforementioned state-of-the-art algorithms and directly the source code of Zero-1-to-3 (<https://github.com/cvlab-columbia/zero123>). It is designed for synthesising novel views of an object based on a single RGB image. It operates in an under-constrained setting, meaning it can generate new views even with limited input information.

The methodology involves training a conditional diffusion model on synthetic data to learn controls of the camera viewpoint, enabling the synthesis of new views of objects from single RGB images. Despite being trained on synthetic data, the model exhibits strong zero-shot generalisation to out-of-distribution datasets and in-the-wild images, including impressionist paintings. This suggests that the model can effectively synthesise new views even for objects it hasn't encountered during training.

The first step in implementing the algorithm is to prepare the input data. Images should be prepared as RGBA images or RGB images with a white background.

The live demo allows the control of camera rotation and thereby generates novel viewpoints of an object within a single image. It is based on Stable Diffusion. By using the learned controls of the relative camera viewpoint, the framework can reconstruct the 3D structure of objects depicted in the image, after running the training phase.

#### 5. CASE STUDY IN CULTURAL HERITAGE FIELD

The case studies were selected to analyse different situations in which it is possible to find when working on cultural heritage. Below, is a description of each category.

##### 1. Statues

Working in the field of cultural heritage it is possible to find the situation of having to recreate a digital copy of medium-large sized objects such as statues or similar. For this reason, the methodology was tested on two statues in Abruzzo (Italy). The first one is the sculpture "Dedalo" by the Abruzzo author Andrea Cascella, created in 1982 and part of the Walter Fontana collection in Milan. The second one is "Fanciulla", a statue, constructed on an urban scale, inspired by the "12 Fanciulle d'Abruzzo" by the Abruzzo artist Franco Summa. Are sculptures marked distinctly by the unique color grammar that Summa developed over the years.



Figure 1. Sculpture "Dedalo" by Andrea Cascella and "Fanciulla" by Franco Summa

##### 2. Existing architectures

Two existing architectures in Brixen (Italy) were analysed, the facade of the cathedral and the building of the Free University of Bozen/Bolzano. In the cathedral case study, metric measurements made with laser scanners were available to be used for the metric analysis of the results and to compare the reconstruction made from a single scene with reliable surveys.



Figure 2. Facade of the cathedral of Brixen and the building of the Free University of Bozen.

### 3. Famous Monuments

The reconstruction of a famous monument such as the Eiffel Tower in Paris was added to the analysis. The aim is to assess whether the presence of a world-famous architecture in the datasets used by the pre-trained neural networks on which the proposed method is based can improve the achievement of 3D reconstruction from a single scene.



Figure 3. Eiffel Tower

### 4. Architecture from historical photos

The processing of historical images, both photos and videos, is of fundamental importance for the documentation of heritage that no longer exists because it has been destroyed or disappeared over time, but which appears in them. For this reason, a case study previously analysed by the authors (Condorelli et al., 2021) on which several experiments were conducted is reported, namely that of the Tour Saint Jacques in Paris (France).



Figure 4. Tour Saint Jacques in Paris - France Paris Vue 1900

### 5. Illustrated architectures

Finally, it is also presented the case study of the illustration of the *Turris Babel* in Athanasius Kirker's *Archontology*, an ancient copy of which is in the library of the major seminary in Brixen (Italy). The imagery of the Tower of Babel can be found in various cultures and various descriptions and illustrations exist. It therefore stands as a known, if only imagined, piece of architecture, but for this reason of interest from the point of view of intangible cultural heritage.

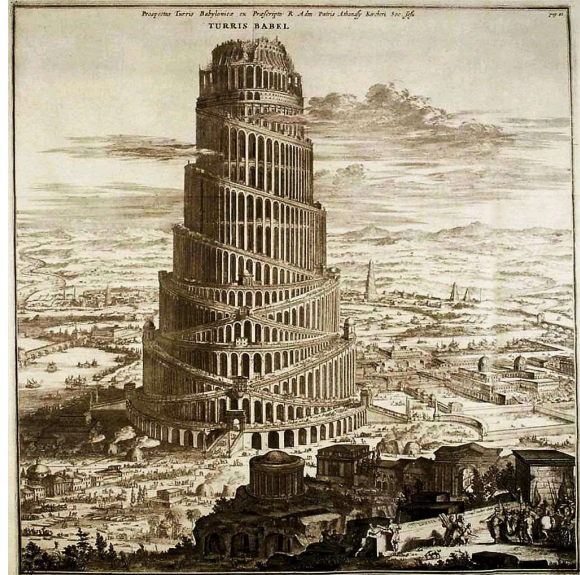


Figure 5. *Turris Babel* in Athanasius Kirker's *Archontology*

### 6. RESULTS

The methodology was applied to each dataset. Figure 6 shows the results of the RGBA images used for the generation of the novel viewpoints as basis of the subsequent 3D model reconstruction phase.

Already at this early stage it can be seen that the algorithm performs best when the objects to be reconstructed are symmetrical, because obviously the side views are similar to the front view. However, the results for non-symmetrical buildings show satisfactory results as the reconstruction of the side views is successful.

The metric comparison (Figure 7) shows that the model of the cathedral taken as a reference obtained from a previous standard photogrammetric survey and the one obtained with the single-view reconstruction methodology present a distance between the two meshes computed in cloud compare acceptable for the architectural scale in question.

Bringing the result of the exports in .obj format into Blender 3D software (Figure 8), it can be observed that, aesthetically, the models exhibit good mesh topology and texture at a fairly high resolution, but in some cases, they lack precision in detail. The mesh consists of triangular polygons, and the topology and number of polygons adapt where changes in surface curvature are encountered.

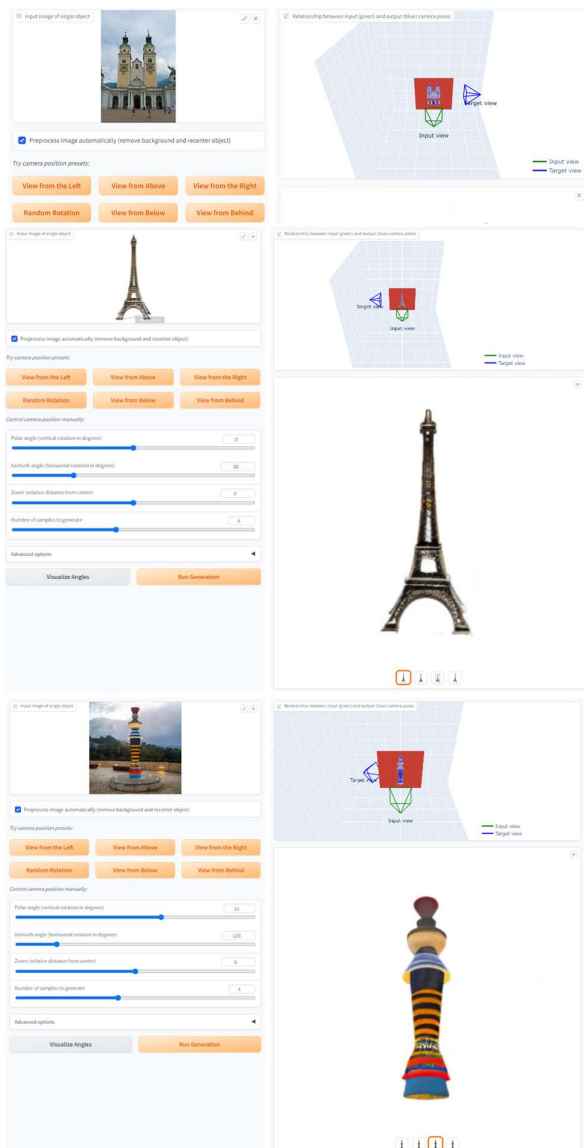


Figure 6. New scene reconstruction using Zero 1-2-3.

The texture paired with the .obj file, in high definition, sometimes exhibits errors in details, and there is a sort of ambient occlusion in areas corresponding to shaded parts. The only case where the texture appears confused is that of the Tower of Babel, perhaps due to the low-resolution source image or the type of image representation lacking clear contrasts. Through the Cycles rendering engine in Blender 3D, it was possible to analyze the meshes and shaders, as well as improve the overall appearance of the models by adding channels (such as bump). The meshes do not exhibit overlapping polygons, gaps, or n-polygons.



Figure 7. Metric comparison in cloud compare of the Cathedral of Brixen

## 7. CONCLUSIONS AND FUTURE WORKS

This paper analyses the latest algorithms that exploit artificial intelligence to obtain new views and 3D models from datasets consisting of a single image. The application of this method to the case studies under examination has made it possible to conduct the first experiments on critical situations in which cultural heritage may find itself, such as objects or buildings that are inaccessible, no longer visible, no longer exist and never existed because they were only illustrated. The results show the great potential of the applied method. Indeed, from the visualisation point of view, the results are of good resolution and quality, and the accuracy of the reconstruction of the other views is also faithful. When it was possible to conduct a metric analysis, the results were reliable from the point of view of accuracy and consistent with 3D models obtained by other methods such as laser scanning.

This research therefore opens up many avenues in the field of digitisation and 3D modelling of cultural heritage, and especially when combined with standard SfM algorithms it is possible to improve and boost the workflow especially in the most challenging cases such as those presented in this paper.

In conclusion, through this comparative study, we seek to contribute to a deeper understanding of the capabilities of NeRF-based approaches in handling single-image inputs, thereby advancing the state-of-the-art in 3D scene reconstruction from limited visual data. Moreover, the findings of this research can inform the development of more effective and practical solutions for real-world applications requiring single-image 3D reconstruction, such as scene understanding, object localization, and immersive content generation.

Future studies will focus on combining NeRF algorithms with prompt-to-image algorithms following the trend of recent research. Indeed, the greatest challenge to be faced in digitising cultural heritage is to train these algorithms on heritage data, as the existing datasets on which algorithms such as Stable Diffusion and the like have been driven lack such data. This work will necessarily require the collaboration of different disciplines working on heritage but will certainly open up new avenues of research leading to important developments in this field.

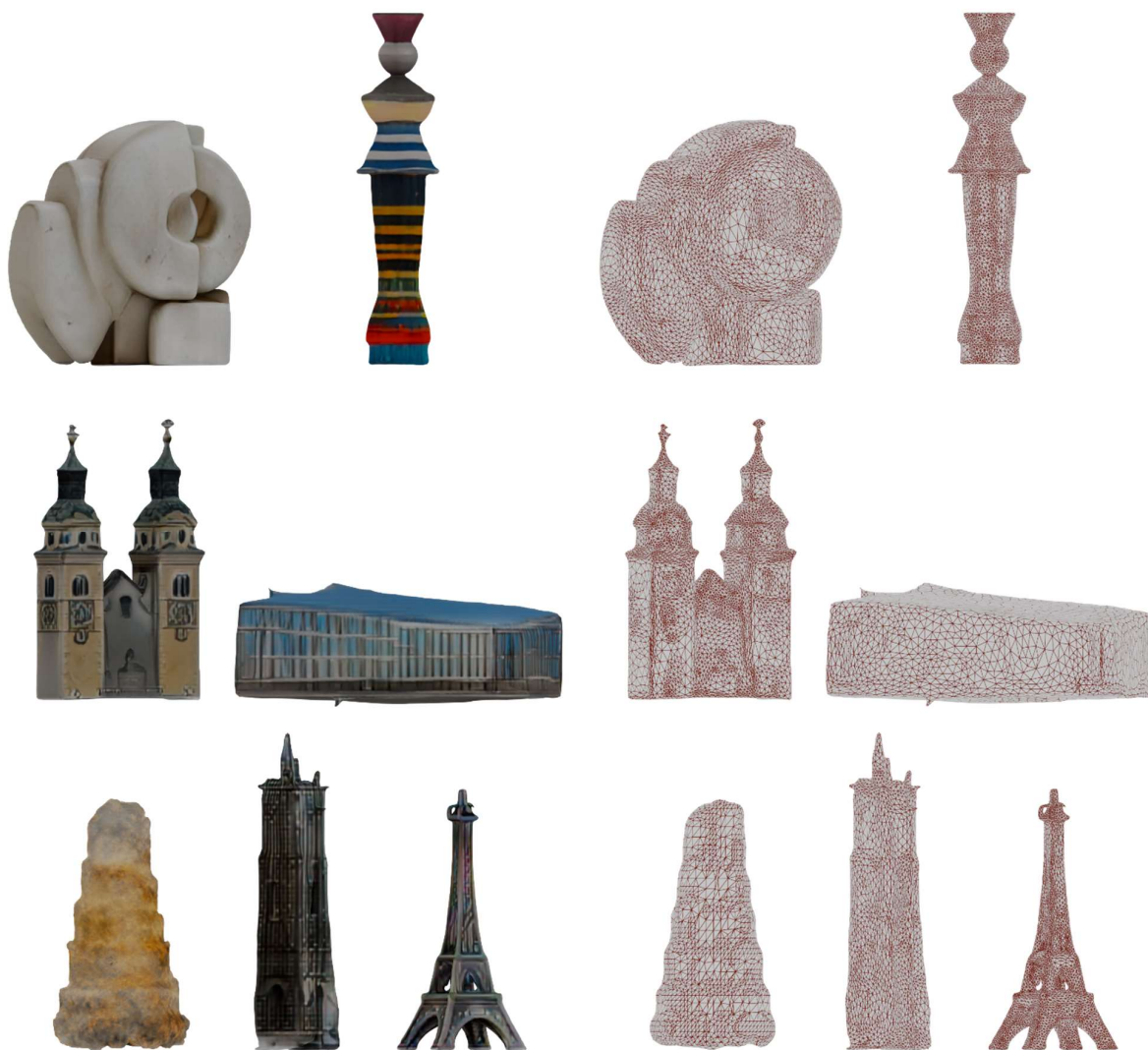


Figure 7. Rendering of the results in Blender 3D and the corresponding mesh wireframe

## REFERENCES

Basso, A., Condorelli, F., Giordano, A., Morena, S., and Perticarini, M. 2024: EVOLUTION OF RENDERING BASED ON RADIANCE FIELDS. THE PALERMO CASE STUDY FOR A COMPARISON BETWEEN NERF AND GAUSSIAN SPLATTING, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-2/W4-2024, 57–64, [https://doi.org/10.5194/isprs-archives-XLVIII-2-W4-2024-57-\(2024\)](https://doi.org/10.5194/isprs-archives-XLVIII-2-W4-2024-57-(2024)).

Cai, Shengqu and Obukhov, Anton and Dai, Dengxin and Van Gool, Luc, 2022: Pix2NeRF: Unsupervised Conditional p-GAN for Single Image to Neural Radiance Fields Translation, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2981–3990, (2022).

Chen, X., Song, J., & Hilliges, O. 2019: Monocular Neural Image Based Rendering with Continuous View Control. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (2019).

Condorelli, F. 2023: IMAGE RETRIEVAL FOR 3D MODELLING OF ARCHITECTURE USING AI AND PHOTOGRAMMETRY, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-M-2-2023, 441–446, [https://doi.org/10.5194/isprs-archives-XLVIII-M-2-2023-441-2023, \(2023\)](https://doi.org/10.5194/isprs-archives-XLVIII-M-2-2023-441-2023, (2023)).

Condorelli, F., Higuchi, R., Nasu, S., Rinaudo, F., and Sugawara, H. 2019: IMPROVING PERFORMANCE OF FEATURE EXTRACTION IN SFM ALGORITHMS FOR 3D SPARSE POINT CLOUD, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W17, 101–106, [https://doi.org/10.5194/isprs-archives-XLII-2-W17-101-2019, \(2019\)](https://doi.org/10.5194/isprs-archives-XLII-2-W17-101-2019, (2019)).

Condorelli, F. and Luigini, A. 2024: NERF-DRIVEN ALGORITHMS APPLIED ON 360 IMAGES FOR 3D MODELLING OF HERITAGE ASSET IN VIRTUAL ENVIRONMENT, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-2/W4-2024, 135–140, [https://doi.org/10.5194/isprs-archives-XLVIII-2-W4-2024-135-2024, \(2024\)](https://doi.org/10.5194/isprs-archives-XLVIII-2-W4-2024-135-2024, (2024)).

- 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. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 43(B2-2021), 565–570.
- Croce, V., Caroti, G., Luca, L. De, & Piemonte, A., 2023: Neural radiance fields (NeRF): review and potential applications to digital cultural heritage. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-M-2-2023, 453–460 (2023).
- Hong, Y., Zhang, K., Gu, J., Bi, S., Zhou, Y., Liu, D., Liu, F., Sunkavalli, K., Bui, T., & Tan, H. 2024: LRM: Large Reconstruction Model for Single Image to 3D. arXiv preprint arXiv:2311.04400.
- K. Gu, T. Maugey, S. Knorr and C. Guillemot, 2022: Omni-NeRF: Neural Radiance Field from 360° Image Captures, 2022 *IEEE International Conference on Multimedia and Expo (ICME)*, Taipei, Taiwan, pp. 1-6, doi: 10.1109/ICME52920.2022.9859817 (2022).
- Lin, C. H., Gao, J., Tang, L., Takikawa, T., Zeng, X., Huang, X., ... & Lin, T. Y. 2023: Magic3d: High-resolution text-to-3d content creation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 300-309) (2023).
- Liu, R., Wu, R., Van Hoorick, B., Tokmakov, P., Zakharov, S., & Vondrick, C. 2023: Zero-1-to-3: Zero-shot one image to 3d object. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 9298-9309) (2023).
- Lv, J.; Guo, J.; Zhang, Y.; Zhao, X.; Lei, B. 2023: Neural Radiance Fields for High-Resolution Remote Sensing Novel View Synthesis. *Remote Sens.* 15, 3920. <https://doi.org/10.3390/rs15163920>, (2023).
- Long, X., Guo, Y.-C., Lin, C., Liu, Y., Dou, Z., Liu, L., Ma, Y., Zhang, S.-H., Habermann, M., Theobalt, C., & Wang, W. 2023: Wonder3D: Single Image to 3D using Cross-Domain Diffusion. arXiv, eprint=2310.15008 (2023).
- Minjung Son and Jeong Joon Park and Leonidas Guibas and Gordon Wetzstein, 2023: SinGRAF: Learning a 3D Generative Radiance Field for a Single Scene, *CVPR*, (2023).
- 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. *Lecture Notes in Computer Science* (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12346 LNCS, 405–421 (2020).
- Palestini, C., Meschini, A., Perticarini, M., Basso, A., 2024: Neural Networks as an Alternative to Photogrammetry. Using Instant NeRF and Volumetric Rendering. In: Giordano, A., Russo, M., Spallone, R. (eds) *Beyond Digital Representation. Digital Innovations in Architecture, Engineering and Construction*. Springer, Cham. [https://doi.org/10.1007/978-3-031-36155-5\\_30](https://doi.org/10.1007/978-3-031-36155-5_30) (2024).
- Pavlo, D., Tan, D. J., Rakotosaona, M. J., & Tombari, F. 2022: Shape, Pose, and Appearance from a Single Image via Bootstrapped Radiance Field Inversion. arXiv, doi:10.48550/arXiv.2211.11674 (November 2022).
- Rematas, K., Martin-Brualla, R., & Ferrari, V., 2021: ShaRF: Shape-conditioned Radiance Fields from a Single View. *International Conference on Machine Learning* (2021).
- Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., ... & Jitsev, J. 2022: Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems*, 35, 25278-25294 (2022)
- Sun, J., Zhang, B., Shao, R., Wang, L., Liu, W., Xie, Z., & Liu, Y. 2023: DreamCraft3D: Hierarchical 3D Generation with Bootstrapped Diffusion Prior. ArXiv, abs/2310.16818 (2023).
- Tang, Junshu & Wang, Tengfei & Zhang, Bo & Zhang, Ting & Yi, Ran & Ma, Lizhuang & Chen, Dong. 2023: Make-It-3D: High-Fidelity 3D Creation from A Single Image with Diffusion Prior. 22762-22772. 10.1109/ICCV51070.2023.02086 (2023).
- Tochilkin, D., Pankratz, D., Liu, Z., Huang, Z., Letts, A., Li, Y., Liang, D., Laforte, C., Jampani, V., & Cao, Y.-P. 2024: TripoSR: Fast 3D Object Reconstruction from a Single Image. arXiv preprint arXiv:2403.02151 (2024).
- Trevithick, A., Chan, M., Stengel, M., Chan, E., Liu, C, Yu, Z., Khamis, S., Chandraker, S., Ramamoorthi, R., Nagano, R., 2023: Real-Time Radiance Fields for Single-Image Portrait View Synthesis. *ACM Trans. Graph.* 42, 4, Article 135, 15 pages. <https://doi.org/10.1145/3592460> (August 2023).
- Yu, A., Ye, V., Tancik M. and Kanazawa, A. 2021: "pixelNeRF: Neural Radiance Fields from One or Few Images," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, pp. 4576-4585, doi: 10.1109/CVPR46437.2021.00455 (2021).
- Xu, Dejia and Jiang, Yifan and Wang, Peihao and Fan, Zhiwen and Shi, Humphrey and Wang, Zhangyang, 2022: SinNeRF: Training Neural Radiance Fields on Complex Scenes from a Single Image, arXiv preprint arXiv:2204.00928 (2022).
- Wei Jiang, Kwang Moo Yi, Golnoosh Samei, Oncel Tuzel, Anurag Ranjan, 2023: NeuMan: Neural Human Radiance Field from a Single Video, <https://doi.org/10.48550/arXiv.2203.12575> (2023).
- Wiles, O., Gkioxari, G., Szeliski, R., & Johnson, J. 2020: Synsin: End-to-end view synthesis from a single image. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7467-7477) (2020).