Digital Applications in Archaeology and Cultural Heritage 35 (2024) e00371

Contents lists available at ScienceDirect

Digital Applications in Archaeology and Cultural Heritage

journal homepage: www.elsevier.com/locate/daach

Comparative assessment of Neural Rendering methods for the 3D reconstruction of complex heritage sites in the inner areas of the Marche region - Italy

Emanuele Balloni [a,*], David Ceka [a], Roberto Pierdicca [b], Marina Paolanti [c], Adriano Mancini [a], Primo Zingaretti [a]

[a] *Department of Information Engineering (DII), Università Politecnica Delle Marche, 60131, Ancona, Italy*
[b] *Department of Construction, Civil Engineering and Architecture (DICEA), Università Politecnica Delle Marche, Via Brecce Bianche, 12, 60131, Ancona, Italy*
[c] *Department of Political Sciences, Communication and International Relations, University of Macerata, Via Don Minzoni 22/A, 62100, Macerata, Italy*

1. Introduction

The attempt to accurately reproduce real-world objects and environments in three-dimensional digital form has always been a significant challenge in the field of Digital Cultural Heritage (DCH) Cotella (2023), Abergel et al. (2023). Traditional techniques such as Structured Light Scanning, Photogrammetry and Stereo Multi-View, while advanced for their time, were often limited in their ability to manage complex structures and capture the rich details of intricate surfaces Cianci and Colaceci (2023). Despite their benefits, these techniques are limited by the diverse materials, properties and large-scale intricate details of heritage sites, posing significant challenges in achieving comprehensive digital representations.

Artificial Intelligence (AI) approaches to 3D reconstruction have brought a paradigm shift, enabling the creation of highly detailed and realistic models Khairina et al. (2023), Fontanella et al. (2020). Neural Radiance Fields (NeRF), in particular, represent a significant advancement in this field Mildenhall et al. (2021). NeRF synthesizes photorealistic images using a neural network to model the volumetric function of a scene, capturing complex light and shadow relationships. This technique surpasses traditional methods in handling occlusions and complex textures, generating continuous, detailed 3D representations even from limited image sets. NeRF is particularly beneficial for documenting fragile or inaccessible cultural heritage sites non-invasively Tavakoli et al. (2023).

Despite these advances, a comprehensive and comparative evaluation of novel 3D reconstruction approaches in DCH domain remains a gap in the existing literature. Such an evaluation is essential to understand the relative strengths and weaknesses of each method and to determine their most appropriate application contexts. More specifically, the CH sector faces the challenges of defining low cost strategies for its documentation, while keeping details and precision as faithful as possible with reality. The recent literature uncovered new potential of Neural Rendering in this field, Croce et al. (2024b) though their performances in terms of reliability are partially uncovered. This paper aims to fill this gap by presenting a detailed comparative study of different Neural Rendering (NR) techniques: NeRF, Signed Distance Function (SDF) and 3D Gaussian Splatting Kerbl et al. (2023). We explore the state-of-the-art for each technology, with a strong focus on quality, training time and resource consumption, reflecting the latest research developments. The aim is to provide a comprehensive guide that compares the quantitative metrics and training characteristics of the networks that support these technologies.

We developed a complete pipeline that starts from dataset acquisition, to camera pose estimation, to NR networks training, metrics evaluation and 3D mesh extraction. This comparative pipeline is intended to help practitioners make informed decisions about the most appropriate tools for their specific datasets and usage conditions. The outcome of this research presents a detailed analysis of the results, conditions of use and recommended applications for each method, distinguishing between amateur and professional applications, including those aimed at accurate surface reconstruction for both digital interaction and in-depth analytical studies of reconstructed morphologies. Furthermore, we have carefully collected datasets from heritage sites in the Marche region, located in the center of Italy. The motivation for this experiment lies in the broad territorial scope of the project itself. In fact, the contribution of the research group to the inner areas of the region is to promote the heritage, both from a landscape and architectural point

* Corresponding author. Via Brecce Bianche, 12, 60131, Ancona, Italy.
E-mail addresses: e.balloni@pm.univpm.it (E. Balloni), S1104315@studenti.univpm.it (D. Ceka), r.pierdicca@staff.univpm.it (R. Pierdicca), marina.paolanti@unimc.it (M. Paolanti), a.mancini@univpm.it (A. Mancini), p.zingaretti@univpm.it (P. Zingaretti).

https://doi.org/10.1016/j.daach.2024.e00371
Received 16 April 2024; Received in revised form 5 August 2024; Accepted 28 August 2024
Available online 31 August 2024
2212-0548/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

of view, for the valorisation and promotion of tourism.

The main contributions of this paper can be summarized as follows:

- An in-depth comparative study of three state-of-the-art 3D reconstruction methods: NeRF, SDF and 3D Gaussian Splatting, providing a comprehensive view of their capabilities and limitations in the context of DCH.
- A detailed analysis of the resource efficiency of each method, including training time and computational resources required. This analysis is critical to understanding the practicality of using these methods in different scenarios, especially where resources may be limited.
- Based on the comparative analysis, the paper proposes guidelines to help practitioners select the most appropriate 3D reconstruction tool for their specific needs.
- Analysis of potential applications of each method in real-world scenarios, providing insights into their suitability for different tasks.

This paper is organized as follows: Section 2 provides an overview of relevant literature and systems. Section 3 details the technical framework used in our study. Section 4 presents the experimental results. Finally, Section 5 summarizes the findings and suggests future research directions.

2. Related works

The challenge of extracting three-dimensional (3D) information from images is a well-known challenge in the field of computer vision. Historically, this task has been approached through conventional geometric methods. Over the years, these methods have relied on techniques such as photometric consistency and precisely designed features to extract depth information from visual data. While these methods have provided significant advances and insights, they often fail to capture the intricate details and complexities of real-world scenes, particularly in scenarios with complex geometries and lighting conditions Strecha et al. (2006); Goesele et al. (2007); Remondino et al. (2008); Hirschmuller (2008); Barnes et al. (2009); Furukawa and Ponce (2010); Jancosek and Pajdla (2011); Bleyer et al. (2011); Schönberger et al. (2016). Recent years have seen a paradigm shift in 3D reconstruction methods, with the emergence of novel approaches driven by AI techniques. In this section, we review relevant works and systems in the field of NR for CH, contextualising our research within the broader landscape of advances and identifying potential areas for innovation and improvement.

2.1. Neural rendering methods

In the study of 3D scene representation and rendering, NeRF stands out as a prominent approach (Mildenhall et al., 2021). NeRF represents scenes using neural networks that capture complex spatial information and viewpoints through a robust framework. Unlike traditional methods, NeRF can represent scenes using a deep, fully connected neural network without convolutional layers, enabling the synthesis of photo-realistic images. The network takes as input a single continuous 5D coordinate set representing spatial locations and viewing directions, and outputs volume density and view-dependent emitted radiance for each location and direction. Many advances have been made in NeRF techniques, including Mip-NeRF Barron et al. (2021) and Mip-NeRF 360 Barron et al. (2022), which address challenges such as anti-aliasing and handling scenes with 360-degree camera rotation. Another notable improvement is Instant-NGP Müller et al. (2022), which uses multi-resolution hash grids to greatly improve rendering speed and lower memory requirements.

The SDF approach is another powerful technique for 3D reconstruction from visual data. Unlike methods such as NeRF, SDF methods directly represent the scene geometry using a scalar field that encodes the signed distance from each point in space to the surface of the object.

Table 1
Comparison of NeRF, SDF and 3D Gaussian Splatting methods based on usage scenarios and key features.

Method	Key Features	Application scenarios
NeRF	• High speed Li et al. (2023a) • Good visual rendering of scenes Lazova et al. (2023) • Limited mesh detail Yuan et al. (2022)	• Fast exploration of 3D scenes with non-critical mesh details Ge et al. (2024)
SDF	• High-level mesh Qiao et al. (2023) • Precise details Zhu et al. (2023) • Slow computation Wu et al. (2023b)	• Reconstruction of high-fidelity meshes for complex scenes Zhu et al. (2023) • Monument degradation studies Hasselgren et al. (2022) • 3D models for video games Yunus et al. (2024)
3D Gaussian Splatting	• Moderate computation speed Lee et al. (2024) • Good mesh quality Waczyńska et al. (2024) • Suitable for a wide range of applications Tosi et al. (2024)	• Compromise between computation speed and mesh quality Wu et al. (2023a)

One of the key advantages of SDF-based approaches is their ability to provide a compact and efficient representation of complex geometric structures. By storing the signed distance values at discrete grid points or using implicit function representations, SDF methods can capture intricate surface details while maintaining a relatively small memory footprint Park et al. (2019). SDF methods offer several advantages, including fast inference times and the ability to handle large scenes with high fidelity Yariv et al. (2021). In addition, SDF representations can be easily combined with traditional graphics techniques such as ray tracing and rasterisation, allowing seamless integration into existing rendering pipelines Wang et al. (2021). A notable example of SDF-based reconstruction is the work of NeuS, which introduced a neural network architecture capable of efficiently learning SDF representations Wang et al. (2021).

3D Gaussian Splatting Kerbl et al. (2023) is a novel approach to rendering scenes in three dimensions. Using anisotropic 3D Gaussians as the basic building blocks, it provides a highly flexible and expressive representation of radiation fields. By optimising properties such as position, opacity and anisotropic covariance through iterative processes, Gaussian splatting achieves a compact and accurate scene representation. Rendering uses fast GPU sorting algorithms inspired by tile-based rasterisation techniques. This method ensures visibility-aware rendering, allows for anisotropic splatting, and enables fast and accurate backward propagation for high-quality novel view synthesis. Through its innovative implementation, Gaussian splatting significantly advances the state of the art in real-time rendering, offering unparalleled visual fidelity and efficiency in 3D scene rendering and synthesis. Reconstructing meshes from Gaussian splatting methods presents challenges primarily due to the inherent nature of the representation. Unlike traditional mesh-based representations, Gaussian splatting relies on anisotropic 3D Gaussians that are volumetric in nature and lack explicit surface definitions. These Gaussian distributions represent radiation fields in a continuous manner, making it difficult to directly extract discrete surface geometry. In addition, the optimization process involved in Gaussian splatting aims to produce a compact and accurate scene representation, which may prioritise the capture of radiance properties over detailed surface geometry. The first 3D Gaussian Splatting application, allows fast reconstruction of a 3D space, but lacks mesh reconstruction: in fact, the disordered nature of millions of tiny 3D Gaussians resulting from the process optimization poses a significant problem in mesh generation. To overcome this challenge, SuGaR Guédon and Lepetit (2023) introduces a key regularization term, which encourages the 3D Gaussians to align with the scene surface, allowing

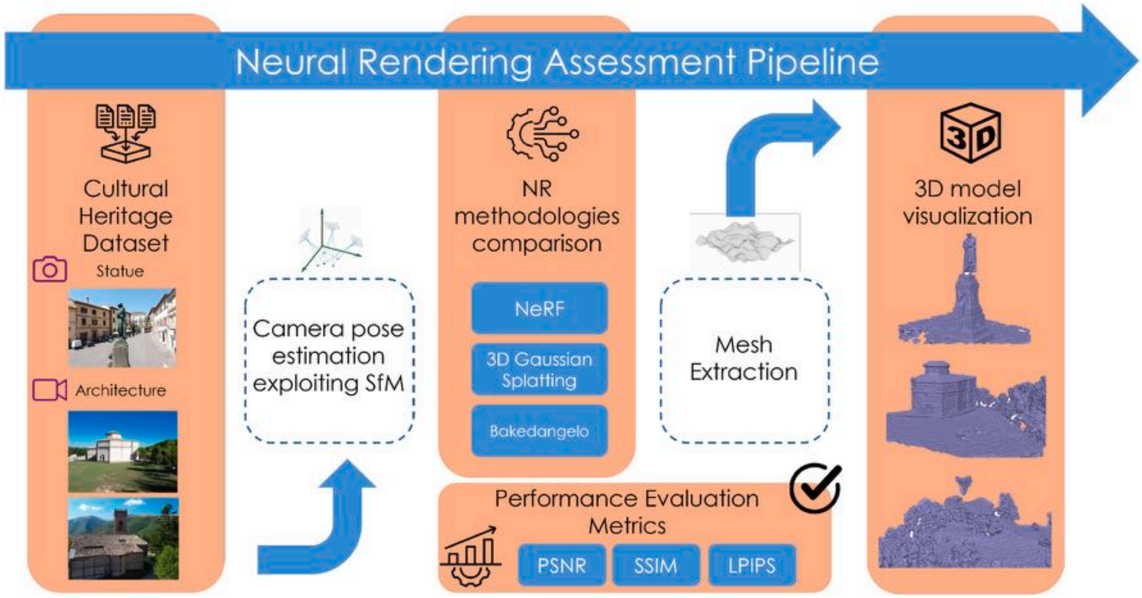


Fig. 1. Workflow of the key components of the methodology, including model selection, evaluation metrics, and dataset collection.

the Gaussians to be optimised towards a more coherent and surface-oriented disposition, facilitating mesh extraction.

Table 1 summarizes the main features of the approaches considered in this paper. NeRF is highlighted for its high speed and visual rendering capabilities but noted for its limited mesh detail, making it suitable for non-critical mesh detail scenarios. SDF stands out for its ability to create high-fidelity meshes with precise details, but with slower computation speeds, making it ideal for applications requiring detailed reconstructions, such as monument degradation studies and video game models. 3D Gaussian Splatting is presented as a middle-ground solution, balancing computation speed and mesh quality, and is applicable in a broader range of scenarios.

2.2. NR and photogrammetry in CH

Significant research efforts in the field of CH have been devoted to comparing different methods for 3D reconstruction. One notable comparison involves contrasting NeRFs with photogrammetry in terms of their operational procedures and output representations. This comparison highlights the strengths and limitations of each approach, particularly in the representation of complex objects and scenes. In Croce et al. (2023), after a preliminary critical review of the scientific and technical literature on NeRFs, authors have highlighted possible applications of the latter in the field of CH, for image-based reconstruction of 3D models of real multi-scale objects, even in combination with more established photogrammetric techniques. In this work, NeRFs have been shown to have distinct advantages in representing objects that are challenging for traditional photogrammetric methods. These include objects with metallic, translucent or transparent surfaces, as well as those with homogeneous textures or intricate details. In addition, NeRFs excel in scenarios involving occlusions, vegetation and elements with exceptionally fine detail, where photogrammetry can have difficulties in capturing accurate representations.

Another contribution comes from a recent study that presented an innovative approach to 3D documentation of ancient statues Balloni et al. (2023). This method integrated terrestrial laser scanning (TLS) acquisition with a multi-view stereo (MVS) pipeline using images captured by a DJI Mavic 2 drone. The study also aimed to compare the accuracy and final output of two different methods: Deep Points (DP) and Neural Radiance Fields (NeRF), using the TLS acquisition as validation ground truth. The investigation began with a TLS acquisition of the ancient statue using a Faro Focus 2 scanner. Subsequent steps include the application of an MVS pipeline using 2D images captured by a DJI Mavic 2 drone, and the training of the NeRF network with the same images after a 90% volume reduction. The main contribution of this research was to improve the understanding of this methodological approach while comparing the accuracy and final results of DP and NeRF techniques. In particular, NeRF outperforms DP in terms of accuracy and realism.

In line with the aforementioned investigations, a study by Mazzacca et al. Mazzacca et al. (2023) contributes to the ongoing research on NeRF in the field of CH preservation. The research investigates the efficacy of NeRF techniques applied to a variety of CH datasets, ranging from smartphone-captured videos to SLR camera images. In particular, the study evaluated several NeRF methods, with a focus on Instant-NGP and Nerfacto, which emerge as the leading methods, outperforming alternative methods in terms of performance and accuracy. Furthermore, a comprehensive analysis, including both qualitative and quantitative assessments across different datasets, highlights the robustness of NeRF approaches, particularly in scenarios characterised by uniform textures, specular surfaces and limited data availability.

Building on the aforementioned research efforts, the work of Murtiyoso et al. Murtiyoso and Grussenmeyer (2023) represents another significant contribution in this field. The study used the Nerfacto architecture to reconstruct two instances of CH objects and then subjected them to a comparative analysis with traditional Multi-View Stereo (MVS) photogrammetric techniques. Despite the initial hypothesis that NeRF might not achieve the same level of accuracy and density as MVS results, the study revealed NeRF's remarkable potential, particularly in terms of processing speed. While MVS techniques may excel in achieving higher accuracy and density, NeRF showed promising capabilities with faster processing times, highlighting its viability as a complementary tool in the field of CH preservation and documentation.

The study by Croce et al. Croce et al. (2024a) aims to extend the discourse by assessing the intrinsic advantages or limitations of NeRFs compared to photogrammetry, while also exploring the potential benefits of integrating these two methods for digital 3D reconstruction of CH objects. Using the same set of input images with known camera positions, the study attempts to provide a comprehensive comparison between NeRF and photogrammetry, evaluating various aspects such as quality and consistency of results, handling of challenging scenes (e.g. objects with reflective, metallic or translucent surfaces), realism of

Fig. 2. Some frames taken from the video of the scene set at the Macereto sanctuary.

renderings, processing time, and the impact of image resolution and number on the accuracy and fidelity of the 3D reconstruction. Concerning SDF and 3D Gaussian Splatting methods, to the authors' knowledge, no research has yet been developed on the topic of comparison between photogrammetry and SDF or 3D Gaussian Splatting in the context of CH.

In light of this comprehensive state of the art, our work aims to further contribute to the advancement of 3D reconstruction methods for DCH preservation. Building on the insights gained from previous research efforts, our study aims to explore the comparative analysis of NeRF alongside other novel techniques such as SDF and 3D Gaussian Splatting. By systematically evaluating and comparing these methods on various metrics such as quality, resource efficiency and real-world applicability, our goal is to provide practitioners with useful guidelines for selecting the most appropriate tool for their specific needs and constraints. In addition, our research highlights the importance of region-specific data collection efforts, as demonstrated by our datasets collected from heritage sites in the Marche region of Italy. This combined effort is made to improve our understanding of the performance and applicability of each method and also to encourage interdisciplinary collaboration and progress in the field of DCH conservation. Ultimately, our work aims to contribute to the ongoing efforts to document, visualize and preserve our CH for future generations.

3. Materials and methods

In this section, we present our approach to evaluating NR models, including the selection of models, evaluation metrics, and datasets. We used a selection of NR models, chosen based on their relevance and state-of-the-art capabilities in the field of DCH preservation. To comprehensively assess the performance of these models and compare them, we used a set of evaluation metrics that included both objective measures, such as Peak Signal-to-Noise Ration (PSNR) Hore and Ziou (2010) and Structural Similarity Index Measure (SSIM) Zhang et al. (2018), and perceptual metrics, such as Learned Perceptual Image Patch Similarity (LPIPS) Zhang et al. (2018). These metrics aimed to provide a comprehensive evaluation of the quality and fidelity of the reconstructed 3D models. Our evaluation was carried out using datasets collected from heritage sites in the Marche region of Italy, known for their rich historical and architectural significance, consisting of high-resolution images and videos captured using different imaging devices. By using datasets from different heritage sites, we aimed to evaluate the robustness and generalisability of the NeRF models under different environmental conditions and cultural contexts. Fig. 1 illustrates our approach to evaluating NeRF models for 3D reconstruction in digital heritage preservation, providing a visual overview of the steps involved in our methodology, including model selection, evaluation metrics, and dataset collection. The next subsections describe this approach in more detail.

3.1. NR reconstruction methods

Selecting the correct NR methods is crucial, particularly in the context of CH, where intricate details and materials are widely present. In this work, we leveraged the Nerfacto NeRF model from the Nerfstudio framework Tancik et al. (2023). Nerfacto is renowned for its efficient balance of speed and optimization, making it an excellent choice for applications requiring the swift and precise reconstruction of 3D scenes. This framework integrates different NeRF methodologies to encode volumetric scene representations, facilitating high-quality rendering with quicker convergence.

The SDF method employed in this study is Bakedangelo from the SDFStudio framework Yu et al. (2022). Bakedangelo merges the foundational principles of BakedSDF Yariv et al. (2023) with the advanced numerical gradients and progressive training strategies of Neuralangelo Li et al. (2023b). This combination allows for precise and detailed modeling of 3D surfaces, harnessing the advantages of both techniques to achieve superior surface accuracy and detail retention.

For the 3D Gaussian Splatting method, we selected SuGaR Guédon and Lepetit (2023). SuGaR introduces a significant regularization term that encourages the 3D Gaussians to conform to the scene surface. This regularization promotes a more coherent and surface-aligned arrangement of the Gaussians, enhancing the optimization process and facilitating mesh extraction. This approach is particularly advantageous for attaining accurate surface representations in complex scenes.

Concerning the mesh extraction phase, the process is the same for both Nerfacto and Bakedangelo: initially, a point cloud is derived by creating depth images for each frame and subsequently reconstructing

Fig. 3. Some images from the video of the scene at Magalotti Castle.

Fig. 4. Some images from the photo collection of the scene set in Piazza A. Gentili in San Ginesio.

the 3D scene through the back-projection of points from these depth maps into the global coordinate system. Subsequently, employing the Poisson Surface Reconstruction method Kazhdan et al. (2006), the point cloud is transformed into a 3D mesh. SuGaR also leverages Poisson Surface Reconstruction, but combines it with different optimizations and refinements to extract a mesh from the generated Gaussians.

All the experiment were performed with a single RTX A6000 GPU (48 GB VRAM) on an Ubuntu 22.04 operating system.

3.2. Marche cultural heritage datasets

For this study, we collected three different datasets from sites of significant CH in Marche region in the center of Italy. These include:

- *The Sanctuary of Macereto*: Located in Visso, on the western slopes of the Sibillini Mountains, this religious complex serves as one of the datasets. We captured 2160p@30fps video footage from which only the key frames were extracted, resulting in a total of 300 frames for analysis (Fig. 2).
- *Magalotti Castle*: Located in the municipality of Fiastra, this historic site is the second dataset. Similar to the first, 2160p@30fps video footage was used and 725 key frames were extracted for further study (Fig. 3).
- *A. Gentili Statue in San Ginesio*: For this urban environment, the dataset consisted of 224 photos of the statue of Alberico Gentili, located in San Ginesio (Fig. 4).

The 3 datasets have been chosen due to their different features. In particular, the San Ginesio dataset consists of images that focus on a single object, namely a statue, that is positioned centrally within the images. This singular focus facilitates precise localisation and characterisation of the object within the image. Conversely, the Macereto and Magalotti datasets present environments with architectural structures and surrounding panoramas. However, while Macereto provides mostly comprehensive views of the buildings that are conducive to 3D reconstruction, Magalotti presents a more challenging scenario due to sub-optimal viewing angles, requiring advanced methods for accurate reconstruction. These datasets diversification allows for a comprehensive exploration of different approaches to 3D reconstruction techniques.

The datasets have been processed for two different types of train validation splits, tailored to the requirements of the modelling approaches used, namely Nerfstudio and SDFStudio. Nerfstudio uses a 90-10 split to divide the dataset into training and validation sets, ensuring a clear separation between the two. On the other hand, SDFStudio uses a unique strategy where all training images are also used for validation. However, it imposes constraints by using only a percentage of these images for validation purposes, rather than the entire collection. This methodological divergence in the handling of the datasets highlights the differences in the approaches of Nerfstudio and SDFStudio to dataset usage and validation.

3.3. Performance evaluation metrics

To evaluate the results obtained from each experiment, various metrics are required to express the overall performances. The most common metric used in image comparison is the PSNR Hore and Ziou (2010), a widely used metric to assess the quality of image reproduction after compression or processing. It measures the ratio between the peak signal and the noise, giving an indication of the fidelity of the reproduction compared to the original. This technique uses the calculation of the Mean Squared Error (MSE) for each point of the difference between the two input images, with the value reported on a logarithmic scale.

$$\mathrm{MSE} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}[I(i,j) - K(i,j)]^2 \tag{1}$$

$$\mathrm{PSNR} = 10\cdot\log_{10}\left(\frac{\mathrm{Peak\ Signal}^2}{\mathrm{MSE}}\right) \tag{2}$$

The MSE formula calculates the squared difference between each pair of pixels and averages them, providing an overall measure of the discrepancy between the original and reconstructed images. A lower MSE indicates less discrepancy and, theoretically, better fidelity of the reconstructed image compared to the original. In addition, the ”signal peak” in the PSNR formula refers to the maximum value that a pixel in the image can represent. This value depends on the bit depth of the image, so in the case of an RGB image, the signal peak will be 255^3 (3 channels).

If PSNR makes a pixel-level comparison, the SSIM Zhang et al. (2018) is a metric used to evaluate the quality of an image, taking into account aspects such as luminance, contrast and structure. Unlike PSNR, SSIM attempts to model human visual perception more accurately. The basic formula for SSIM is complex, but can be broken down into three main components: luminance l, contrast c and structure s.

$$\mathrm{SSIM}(x,y) = [l(x,y)]^{\alpha}\cdot[c(x,y)]^{\beta}\cdot[s(x,y)]^{\gamma} \tag{3}$$

Where:

- x and y are the two images being compared.
- $l(x, y)$ represents the similarity in luminance.
- $c(x, y)$ represents the similarity in contrast.
- $s(x, y)$ represents the similarity in structure.
- The parameters α, β, γ are constants that regulate the relative importance of each component.

The last metric presented is based on machine learning models that attempt to model human visual perception (HVS). LPIPS is a similarity metric for assessing the perceptual quality of images, taking into account local and perceptual features. Instead of focusing solely on luminance, contrast and structure, LPIPS uses deep neural networks to extract high-level features from the image to capture more complex aspects of human perception. Unlike the previous metrics, there is no standard formula, but the LPIPS calculation process involves the use of pre-trained neural models. LPIPS is particularly useful in applications where the perceptual quality of the image is critical, such as image generation, deep learning for computer vision, and scenarios where traditional metrics may not accurately reflect human perception.

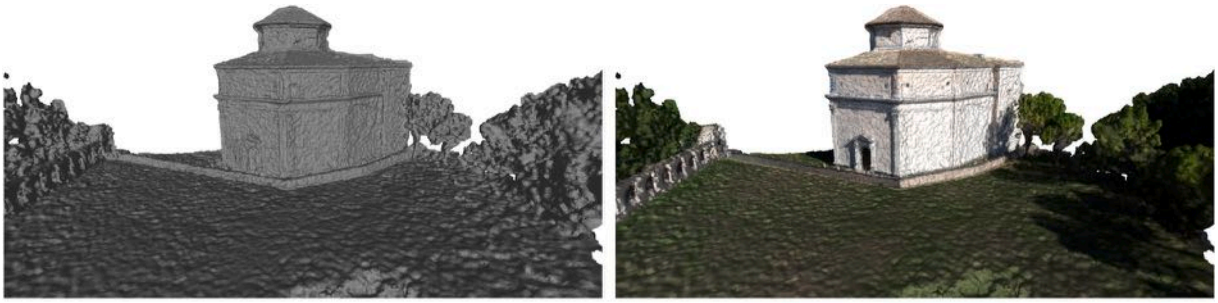
Table 2
Comparison of results between different datasets using the Nerfacto, Bakedangelo, and Gaussian Splatting methods in terms of PSNR, SSIM, and LPIPS. Values are truncated to the second decimal.

Method	Macereto			San Ginesio		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Nerfacto	21.38	0.65	0.21	17.61	0.59	0.42
Bakedangelo	27.42	**0.88**	**0.09**	19.43	**0.77**	**0.16**
Gaussian Splatting	**27.70**	0.84	0.26	**20.16**	0.76	0.31

Method	Magalotti			Training Time
	PSNR ↑	SSIM ↑	LPIPS ↓	
Nerfacto	26.72	0.81	0.13	**~ 30 min**
Bakedangelo	**28.37**	**0.89**	**0.06**	~ 24 h
Gaussian Splatting	24.17	0.71	0.33	~ 1 h

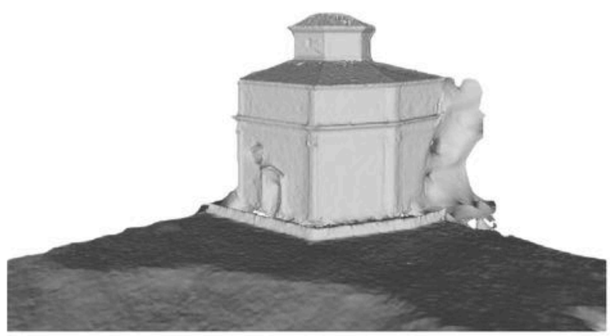
4. Results and discussions

In this section, we provide a comprehensive overview of the results of our experiments and analyses. In Table 2 we present the results and observations obtained by applying different 3D reconstruction methods to the CH datasets described above, in terms of quantitative metrics and training times. Results also include visual representations and qualitative assessments of the reconstructed meshes generated by the different networks (Figs. 5–7). Through detailed comparisons and evaluations, we aim to clarify the effectiveness, limitations and potential applications of each approach in the context of CH preservation and documentation.



(a) Nerfacto mesh

(b) 3DGS mesh



(c) Bakedangelo mesh

Fig. 5. Comparison between the three outputs on the Macereto Dataset. Note: Bakedangelo does not provide texture extraction.

4.1. Results and discussion on the Macereto dataset

The quantitative results of the analysis of the Macereto dataset are reported in Table 2. The evaluation focuses on assessing the effectiveness of different 3D reconstruction approaches, including Nerfacto, Bakedangelo and Gaussian Splatting. The qualitative results in Fig. 5 highlight the differences in the reconstructed meshes produced by these methods. In particular, the mesh resulting from the Nerfacto approach has a coarse representation characterised by clusters and jagged surfaces, particularly noticeable on the walls and roof of the sanctuary. Despite these shortcomings, the model effectively captures the entire scene. Conversely, the Bakedangelo mesh has more refined surfaces with fewer clusters, showing detailed features such as roof tiles and bricks. However, both Nerfacto and Bakedangelo have limitations in accurately modelling vegetation, although this is considered secondary, as the primary focus is on the building. Finally, the Gaussian Splatting approach produces meshes comparable to those generated by Nerfacto, but with improved scene representation; in particular, the inclusion of the cloister of the sanctuary in the reconstructed mesh demonstrates the enhanced capabilities of this method. In addition, for the Bakedangelo model, the increase in vertex count further improves mesh complexity and quality.

Evaluation metrics show significant improvements for the Macereto dataset trained on Bakedangelo. These include notable increases in SSIM (0.88) accompanied by a reduction in LPIPS value (0.09). Furthermore, PSNR values of Bakedangelo and Gaussian Splatting are comparable (27.42 and 27.70, respectively). While Gaussian Splatting outperforms Nerfacto for the PSNR and SSIM metrics, there is a slight degradation in performance for LPIPS. These results are mainly due to the nature of the different models, with Bakedangelo being the most computationally demanding and slow approach, but giving the best results.

4.2. Results and discussion on the San Ginesio dataset

The evaluation of the San Ginesio dataset reflects the observations made in the Macereto experiments, showing similar advantages and limitations of the different reconstruction approaches, as reported in Table 2. As shown in Fig. 6, the Bakedangelo method stands out for its superior reconstruction quality, although it requires longer training times compared to other approaches.

On the San Ginesio dataset, Bakedangelo shows significant improvements in the evaluation metrics. In particular, it achieved the highest SSIM (0.77) and the lowest LPIPS values (0.16). Despite a modest increase in vertex count, Bakedangelo utilises vertices more

(a) Nerfacto mesh

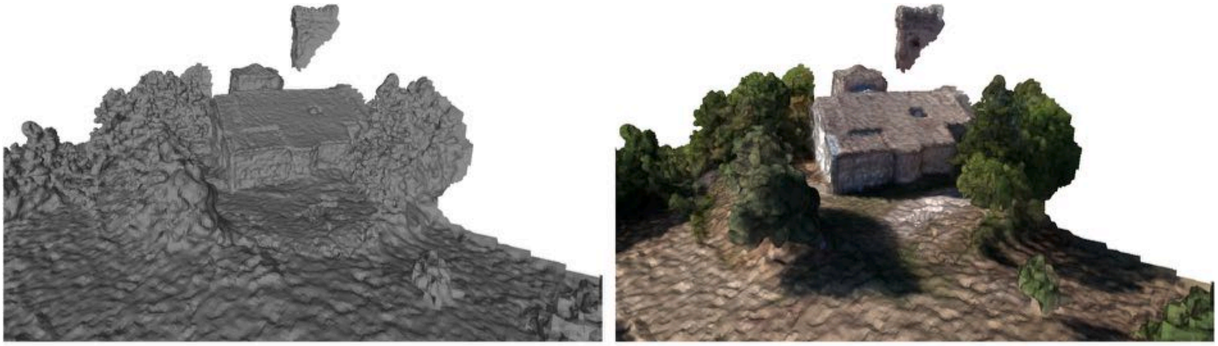
(b) 3DGS mesh (c) Bakedangelo mesh

Fig. 6. Comparison between the three outputs on the San Ginesio dataset. Note: Bakedangelo does not provide texture extraction.

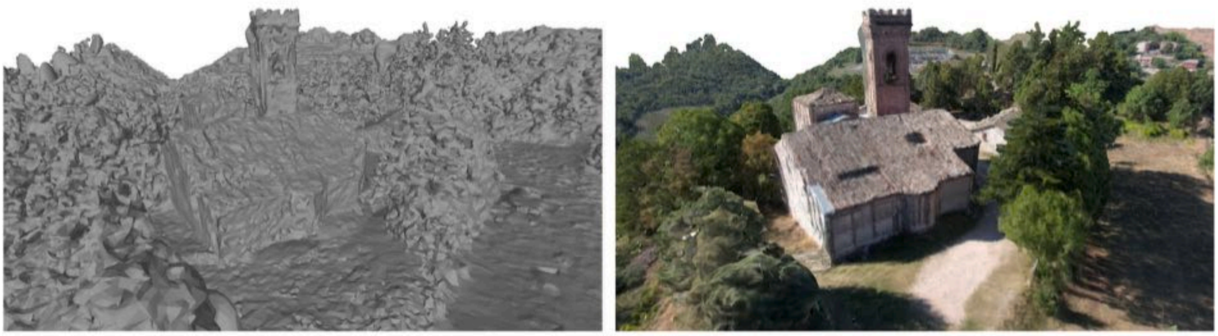
efficiently, resulting in a superior mesh quality compared to its Nerfacto counterpart. Similarly, Gaussian splatting shows significant performance gains over Nerfacto in all the metrics, highlighting the improved fidelity and structural similarity of the rendered images. Also, it performed the best in terms of PSNR (20.16), even exceeding Bakedangelo PSNR value (19.43). This could be attributed to the different environment present in the datasets. Furthermore, Nerfacto presents a marginal degradation in LPIPS performance (0.42), suggesting a slight increase in error perception compared to Gaussian Splatting. Overall, the results highlight the effectiveness of Bakedangelo in producing high quality reconstructions, albeit with longer training times. Meanwhile, Gaussian Splatting offers competitive performance gains, particularly in terms of image fidelity and structural similarity, despite minor drawbacks in error perception.

4.3. Results and discussion on the Magalotti dataset

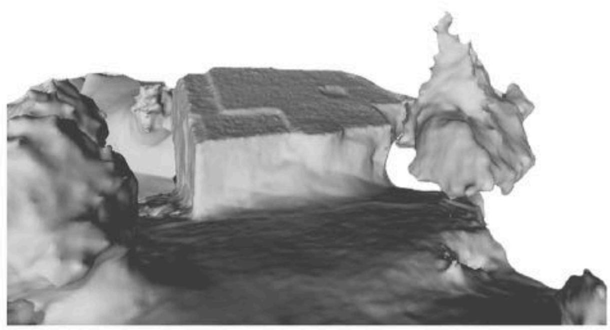
The experiments conducted on the Magalotti dataset produced poor results compared to the other settings, as it can be seen in Fig. 7. This is mainly due to inherent limitations in the dataset itself. The video capture angles were sparse and repetitive, resulting in a lack of consistency between the model and the original scene. As noted above, traditional reconstruction models struggled to adequately represent the view angles due to these limitations. Despite this, the SuGaR model for 3D Gaussian Splatting was able to create a complete representation of the scene (Fig. 5b), overcoming the limitations imposed by the viewing angle of



(a) Nerfacto mesh



(b) 3DGS mesh



(c) Bakedangelo mesh

Fig. 7. Comparison between the three outputs on the Magalotti Dataset. Note: Bakedangelo does not provide texture extraction.

the dataset. SuGaR successfully reconstructed not only the front part of the scene, but also areas captured from few shot angles, making it the preferred choice for scenarios with limited shot representations.

In terms of evaluation metrics, as reported in Table 2, Bakedangelo showed significant improvements from all the other models, with increases in PSNR (28.37) and SSIM (0.89) accompanied by decreases in LPIPS value (0.06). On the other hand, Gaussian Splatting showed a PSNR of 24.71, a SSIM of 0.71 and a LPIPS of 0.33. The decrease in SSIM suggests less structural coherence compared to other methods, while the increased LPIPS value indicates a higher perceived error. Compared to Nerfacto (PSNR 26.72, SSIM 0.81, LPIPS 0.13), Gaussian Splatting showed inferior performance across all metrics, with decreases in PSNR, SSIM and increases in LPIPS values.

These results highlight the challenges posed by limited view angle representation in traditional reconstruction models, and the potential of novel approaches such as SuGaR to overcome such limitations.

4.4. Results comparison

After a thorough analysis of the data, a number of key findings emerge, providing valuable guidance for selecting the most appropriate method based on specific application requirements. In terms of application scenarios, the NeRF method proves to be particularly adept at rapidly exploring 3D scenes (~30 min for a complete training). However, it has limitations in capturing critical mesh details. Its high execution speed makes it well suited to contexts where rapid visualisation of the scene is crucial, prioritising visual rendering over intricate mesh details. On the other hand, the SDF approach excels at reconstructing high fidelity meshes, especially for complex scenes. This makes it ideal for applications that require precise detail, such as monument degradation studies or video game use. It is important to note, however, that achieving such precision comes at the cost of longer training times (~24 h for a complete training). 3D Gaussian Splatting, on the other hand, is a versatile middle ground between speed and mesh quality. Its moderate execution speed (~1 h for a complete training) makes it suitable for a wide range of applications, while its ability to produce meshes of good quality makes it an interesting option for scenarios requiring a balance between detail and computational efficiency.

(a) Macereto in LumaAI (b) San Ginesio in LumaAI (c) Magalotti in LumaAI

Fig. 8. Examples taken from the visualisation with LumaAI.

4.5. Results visualisation

To further explore and visualize the results obtained from the Nerf model, we used Luma AI,[1] a powerful tool capable of rendering high-quality video output from the synthesised 3D scenes. Some examples are reported in Fig. 8. The visualisation process enables the creation of immersive and photorealistic video renderings. Leveraging the advanced capabilities of Luma AI, we were able to showcase the intricate spatial details and viewpoints captured by the NeRf model in a visually compelling way. The resulting videos,[2] ,[3] [4] allowed viewers to dynamically explore the reconstructed 3D scenes and gain a full understanding of the model's performance and fidelity. Through this visualisation approach, we effectively communicated the results of our research, highlighting the strengths and limitations of the NeRF model in reconstructing complex real-world environments.

5. Conclusions and future works

In this research, we investigated innovative approaches to reconstruct three-dimensional scenes from two-dimensional images. We evaluated three state-of-the-art models - Nerfacto, Bakedangelo and 3D Gaussian Splatting - and their capability of creating accurate meshes, comparing their results using quantitative metrics such as PSNR, SSIM and LPIPS, in addition to visual evaluations. Our goal was to provide actionable insights for those entering the field of 3D scene reconstruction, enabling them to choose the most appropriate approach for their specific needs. Results showed how the Nerfacto model excels at rendering 3D scenes for exploration via virtual camera positioning, although with limited mesh detail, making it suitable for scenarios where mesh complexity is less critical. The Bakedangelo network introduces significant innovations in high-fidelity mesh reconstruction, ideal for detailed scenes, but with longer training times. In between these two models, 3D Gaussian Splatting emerges as a rival to NeRF in terms of computational speed without significantly compromising mesh quality, offering promising prospects for future development.

Future research could explore further advances in NeRF, SDF and 3D Gaussian splatting networks, as each technology has different implementations, each with its own strengths and weaknesses. This work serves as a starting point for such advances, highlighting the potential for integration with additional analyses and the incorporation of novel approaches such as NeuS2 Wang et al. (2023). These techniques aim to drastically reduce the time required to produce high-resolution meshes to minutes, without compromising detail. However, these technologies are still in the early stages of research and, due to the dynamic nature of the field, require continuous updates. Furthermore, while Nerfacto combines multiple approaches, exploring potential synergies and areas for improvement between NeRF and Gaussian splatting could provide invaluable insights Chen et al. (2024). Similarly, exploring the integration of SDF approaches within NeRF methodologies could pave the way for approaches that harness the strengths of both to achieve greater photorealism and faithful projection of real-world 3D spaces within the digital domain. Finally, concerning the datasets used in this study, it would be interesting to evaluate model performance with enhanced pose quality and quantity, addressing limitations observed in datasets such as Magalotti, where the lack of poses resulted in low-quality 3D models. Incorporating these enhancements aims to address the identified shortcomings in each model, thereby increasing the detail and fidelity of the reconstruction.

CRediT authorship contribution statement

Emanuele Balloni: Writing – review & editing, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Formal analysis, Conceptualization. **David Ceka:** Writing – original draft, Investigation. **Roberto Pierdicca:** Writing – review & editing, Supervision. **Marina Paolanti:** Writing – review & editing, Supervision. **Adriano Mancini:** Supervision, Funding acquisition. **Primo Zingaretti:** Supervision, Funding acquisition.

Declaration of competing interest

None.

Acknowledgement

This work is funded by PESARO CTE SQUARE; CUP D74J22000930008, FSC MISE 2014–2020.

References

Abergel, V., Manuel, A., Pamart, A., Cao, I., De Luca, L., 2023. Aïoli: a reality-based 3d annotation cloud platform for the collaborative documentation of cultural heritage artefacts. Digital Applications in Archaeology and Cultural Heritage 30, e00285.

Balloni, E., Gorgoglione, L., Paolanti, M., Mancini, A., Pierdicca, R., et al., 2023. Few shot photogrammetry: a comparison between nerf and mvs-sfm for the documentation of cultural heritage. Int. Arch. Photogram. Rem. Sens. Spatial Inf. Sci. 48.

Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B., 2009. Patchmatch: a randomized correspondence algorithm for structural image editing. ACM Trans. Graph. 28, 24.

Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P., 2021. Mip-nerf: a multiscale representation for anti-aliasing neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 5855–5864.

Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P., 2022. Mip-nerf 360: unbounded anti-aliased neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5470–5479.

[1] https://lumalabs.ai/.

[2] https://lumalabs.ai/capture/813208f2-73c9-4123-8db9-596044129dfa.

[3] https://lumalabs.ai/capture/0b35f7b6-6ef8-4adc-9e9c-e3c53e2b85f4.

[4] https://lumalabs.ai/capture/361d12c1-3905-4eaf-9c6b-cf30a49b1a1e.

Bleyer, M., Rhemann, C., Rother, C., 2011. Patchmatch stereo-stereo matching with slanted support windows. In: Proceedings of the British Machine Vision Conference (BMVC).
Chen, Y., He, T., Huang, D., Ye, W., Chen, S., Tang, J., Chen, X., Cai, Z., Yang, L., Yu, G., et al., 2024. Meshanything: Artist-Created Mesh Generation with Autoregressive Transformers arXiv preprint arXiv:2406.10163.
Cianci, M.G., Colaceci, S., 2023. Laser scanner and uav for the 2d and 3d reconstructions of cultural heritage. SCIRES-IT-SCIentific RESearch and Information Technology 12, 43–54.
Cotella, V.A., 2023. From 3d point clouds to hbim: application of artificial intelligence in cultural heritage. Autom. ConStruct. 152, 104936.
Croce, V., Billi, D., Caroti, G., Piemonte, A., De Luca, L., Véron, P., 2024a. Comparative assessment of neural radiance fields and photogrammetry in digital heritage: impact of varying image conditions on 3d reconstruction. Rem. Sens. 16, 301.
Croce, V., Caroti, G., De Luca, L., Piemonte, A., Véron, P., 2023. Neural radiance fields (nerf): review and potential applications to digital cultural heritage. Int. Arch. Photogram. Rem. Sens. Spatial Inf. Sci. 48, 453–460.
Croce, V., Forleo, G., Billi, D., Bevilacqua, M., Piemonte, A., Caroti, G., 2024b. Neural radiance fields (nerf) for multi-scale 3d modeling of cultural heritage artifacts. Int. Arch. Photogram. Rem. Sens. Spatial Inf. Sci. 48, 165–171.
Fontanella, F., Colace, F., Molinara, M., Di Freca, A.S., Stanco, F., 2020. Pattern Recognition and Artificial Intelligence Techniques for Cultural Heritage.
Furukawa, Y., Ponce, J., 2010. Accurate, dense and robust multiview stereopsis. IEEE Trans. Pattern Anal. Mach. Intell. 32, 1362–1376.
Ge, Y., Guo, B., Zha, P., Jiang, S., Jiang, Z., Li, D., 2024. 3d reconstruction of ancient buildings using uav images and neural radiation field with depth supervision. Rem. Sens. 16, 473.
Goesele, M., Snavely, N., Curless, B., Hoppe, H., Seitz, S.M., 2007. Multi-view stereo for community photo collections. In: Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV). IEEE, pp. 1–8.
Guédon, A., Lepetit, V., 2023. Sugar: Surface-Aligned Gaussian Splatting for Efficient 3d Mesh Reconstruction and High-Quality Mesh Rendering arXiv preprint arXiv: 2311.12775.
Hasselgren, J., Hofmann, N., Munkberg, J., 2022. Shape, light, and material decomposition from images using Monte Carlo rendering and denoising. Adv. Neural Inf. Process. Syst. 35, 22856–22869.
Hirschmuller, H., 2008. Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. 30, 328–341. https://doi.org/10.1109/TPAMI.2007.1166.
Hore, A., Ziou, D., 2010. Image quality metrics: psnr vs. ssim. In: 2010 20th International Conference on Pattern Recognition. IEEE, pp. 2366–2369.
Jancosek, M., Pajdla, T., 2011. Multi-view reconstruction preserving weakly-supported surfaces. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 3121–3128.
Kazhdan, M., Bolitho, M., Hoppe, H., 2006. Poisson surface reconstruction. In: Proceedings of the Fourth Eurographics Symposium on Geometry Processing.
Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., 2023. 3d Gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. 42, 1–14.
Khairina, N., Barus, R.K.I., Ula, M., Sahputra, I., et al., 2023. Preserving cultural heritage through ai: developing lenet architecture for wayang image classification. Int. J. Adv. Comput. Sci. Appl. 14.
Lazova, V., Guzov, V., Olszewski, K., Tulyakov, S., Pons-Moll, G., 2023. Control-nerf: editable feature volumes for scene rendering and manipulation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 4340–4350.
Lee, J., Lee, S., Lee, J., Park, J., Sim, J., 2024. Gscore: Efficient Radiance Field Rendering via Architectural Support for 3d Gaussian Splatting.
Li, S., Li, H., Wang, Y., Liao, Y., Yu, L., 2023a. Steernerf: accelerating nerf rendering via smooth viewpoint trajectory. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20701–20711.
Li, Z., Müller, T., Evans, A., Taylor, R.H., Unberath, M., Liu, M.Y., Lin, C.H., 2023b. Neuralangelo: high-fidelity neural surface reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8456–8465.
Mazzacca, G., Karami, A., Rigon, S., Farella, E., Trybala, P., Remondino, F., et al., 2023. Nerf for heritage 3d reconstruction. Int. Arch. Photogram. Rem. Sens. Spatial Inf. Sci. 48, 1051–1058.
Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R., 2021. Nerf: representing scenes as neural radiance fields for view synthesis. Commun. ACM 65, 99–106.
Müller, T., Evans, A., Schied, C., Keller, A., 2022. Instant neural graphics primitives with a multiresolution hash encoding. ACM Trans. Graph. 41 (102), 1–102:15. https://doi.org/10.1145/3528223.3530127 doi:10.1145/3528223.3530127.
Murtiyoso, A., Grussenmeyer, P., 2023. Initial assessment on the use of state-of-the-art nerf neural network 3d reconstruction for heritage documentation. The International Archives of the Photogrammetry. Remote Sensing and Spatial Information Sciences 48, 1113–1118.
Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S., 2019. Deepsdf: learning continuous signed distance functions for shape representation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 165–174.
Qiao, Y.L., Gao, A., Xu, Y., Feng, Y., Huang, J.B., Lin, M.C., 2023. Dynamic mesh-aware radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 385–396.
Remondino, F., El-Hakim, S., Gruen, A., Zhang, L., 2008. Turning images into 3d models-development and performance analysis of image matching for detailed surface reconstruction of heritage objects. IEEE Signal Process. Mag. 25, 55–65.
Schönberger, J., Zheng, E., Pollefeys, M., Frahm, J.M., 2016. Pixelwise View Selection for Unstructured Multi-View Stereo. https://doi.org/10.1007/978-3-319-46487-9_31.
Strecha, C., Fransens, R., Van Gool, L., 2006. Combined depth and outlier estimation in multi-view stereo. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR'06, pp. 2394–2401. https://doi.org/10.1109/CVPR.2006.78.
Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Wang, T., Kristoffersen, A., Austin, J., Salahi, K., Ahuja, A., et al., 2023. Nerfstudio: a modular framework for neural radiance field development. In: ACM SIGGRAPH 2023 Conference Proceedings, pp. 1–12.
Tavakoli, S., Nicu, I.C., Frauenfelder, R., Gilbert, G., 2023. First geophysical investigations to study a fragile pomor cultural heritage site at russekeila–kapp linné), svalbard. J. Cult. Herit. 63, 187–193.
Tosi, F., Zhang, Y., Gong, Z., Sandström, E., Mattoccia, S., Oswald, M.R., Poggi, M., 2024. How Nerfs and 3d Gaussian Splatting Are Reshaping Slam: a Survey arXiv preprint arXiv:2402.13255.
Waczyńska, J., Borycki, P., Tadeja, S., Tabor, J., Spurek, P., 2024. Games: Mesh-Based Adapting and Modification of Gaussian Splatting arXiv preprint arXiv:2402.01459.
Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W., 2021. Neus: Learning Neural Implicit Surfaces by Volume Rendering for Multi-View Reconstruction. NeurIPS.
Wang, Y., Han, Q., Habermann, M., Daniilidis, K., Theobalt, C., Liu, L., 2023. Neus2: fast learning of neural implicit surfaces for multi-view reconstruction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3295–3306.
Wu, G., Yi, T., Fang, J., Xie, L., Zhang, X., Wei, W., Liu, W., Tian, Q., Wang, X., 2023a. 4d Gaussian Splatting for Real-Time Dynamic Scene Rendering arXiv preprint arXiv: 2310.08528.
Wu, Q., Wang, K., Li, K., Zheng, J., Cai, J., 2023b. Objectsdf++: improved object-compositional neural implicit surfaces. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 21764–21774.
Yariv, L., Hedman, P., Reiser, C., Verbin, D., Srinivasan, P.P., Szeliski, R., Barron, J.T., Mildenhall, B., 2023. Bakedsdf: meshing neural sdfs for real-time view synthesis. In: ACM SIGGRAPH 2023 Conference Proceedings, pp. 1–9.
Yariv, L., Kasten, Y., Moran, D., Galun, M., Atzmon, M., Ronen, B., Lipman, Y., 2021. Volume rendering of neural implicit surfaces. In: Advances in Neural Information Processing Systems.
Yu, Z., Chen, A., Antic, B., Peng, S.P., Bhattacharyya, A., Niemeyer, M., Tang, S., Sattler, T., Geiger, A., 2022. Sdfstudio: A Unified Framework for Surface Reconstruction.
Yuan, Y.J., Sun, Y.T., Lai, Y.K., Ma, Y., Jia, R., Gao, L., 2022. Nerf-editing: geometry editing of neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18353–18364.
Yunus, R., Lenssen, J.E., Niemeyer, M., Liao, Y., Rupprecht, C., Theobalt, C., Pons-Moll, G., Huang, J.B., Golyanik, V., Ilg, E., 2024. Recent Trends in 3d Reconstruction of General Non-rigid Scenes arXiv preprint arXiv:2403.15064.
Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O., 2018. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 586–595.
Zhu, J., Huo, Y., Ye, Q., Luan, F., Li, J., Xi, D., Wang, L., Tang, R., Hua, W., Bao, H., et al., 2023. I2-sdf: intrinsic indoor scene reconstruction and editing via raytracing in neural sdfs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12489–12498.