A Review of Neural Radiance Fields and 3D Gaussian Splatting for 3D Reconstruction

Ruiyang Chen

Nanjing University of Posts and Telecommunications, Nanjing, China
*Corresponding author: p22000719@njupt.edu.cn

Abstract:

Within the disciplines of computer vision and deep learning, the ability to construct 3D models from data has become a fundamental capability. A recent wave of progress has significantly advanced the two leading methods for scene representation:Neural Radiance Fields for implicit modeling and 3D Gaussian Splatting for explicit construction. This review aims to build a systematic cognitive framework of 3D reconstruction technology for readers by comparing the performance of these two technologies and their variants in static and dynamic scenes. This review selects representative evaluation parameters such as Peak Signal-to-Noise Ratio. By collecting experimental data from public datasets, it compares, organizes and summarizes the high-quality rendering ability of neural radiance field technology for geometric details and the high-speed real-time rendering ability of 3D Gaussian splatting. Looking ahead, this paper proposes key development directions such as the integration of expression paradigms and overcoming the slow speed of implicit expression, hoping to provide a structural knowledge framework and research inspiration for the professional field.

Keywords: Neural Radiance Fields, 3D Gaussian Splatting method, Representation Paradigm, Frames Per Second

1. Introduction

Recently, cutting - edge technologies such as virtual reality, autonomous driving, and the metaverse have witnessed rapid development. The 3D reconstruction technology, which reconstructs objects in the 3D world from 2D data, has become the core driving force behind these technologies. 3D reconstruction technology has undergone a staged advancement, transitioning from conventional techniques like Structure from Motion that are based on geometric principles and manual features, to the age of deep learning propelled by data[1]. To systematically sort out the development context of 3D reconstruction in the era of deep learning, this review takes the "representation paradigm“ as the core logical framework for discussion, and conducts in - depth research and analysis on the performance of their respective representative technologies in different scenarios.

The Neural Radiance Field technique uses a multi-layer perceptron to create an implicit scene representation. The final image is then synthesized by volume rendering the network‘s predicted color and density along camera rays [2].This principle enables it to generate high - quality novel view images, and it is mainly applied in high - fidelity novel view synthesis, implicit 3D reconstruction, as well as in virtual reality and film production. Different from NeRF‘s implicit representation, Three-dimensional Gaussian splashing is a representative of display expression paradigms, describing the scene through millions of 3D Gaussian points with well - defined attributes. During the rendering phase, Gaussian points are projected onto the image plane and subsequently composited. The 3DGS technique demonstrates significant promise for real-time applications, facilitating a substantial increase in frame rate while preserving high visual quality. Its performance makes the technology ideal for interactive uses like VR/AR platforms, game engines, and web-based 3D viewers.

The primary contributions of this review are outlined as follows: To begin, this paper presents a structured framework for comprehending the technology of three-dimensional cognitive reconstruction.Second, this review uses different benchmark datasets and evaluation indicators to summarize and compare the performance of different methods in different scenarios. Finally, the future development of this technology field is prospected[3].

2. Three-dimensional reconstruction methods and the development of their variants

2.1 Three-dimensional reconstruction technology based on implicit representation — Neural Radiance Fields

In 2020, Ben’s team pioneered the foundational Neural Radiance Field (NeRF) to generate realistic new views from a limited set of images by conceptualizing scenes as continuous volumetric fields [4]. To model a stationary environment, NeRF relies on a Multi-Layer Perceptron to approximate a continuous 5D function. This function takes a 3D coordinate and a viewing angle as inputs to generate the scene’s volumetric density and direction-dependent color at that point. To render an image, rays are cast from the camera’s viewpoint. The rendering process for each ray involves sampling points along its trajectory. The network then processes the 5D coordinates of these points to yield a color and density for each. Applying standard volume rendering principles, these individual outputs are integrated along the ray’s path. The ultimate color of a pixel is determined by the integral formula that follows:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t),\mathbf{d})dt$$

The term $T(t)$ signifies the transmittance, which is the accumulated probability of the ray traveling from the near bound t_n to the point t without being occluded. Its value is computed with the subsequent formula:

$$T(t) = \exp\left(-\int_{t_n}^{t} \sigma(\mathbf{r}(s))ds\right)$$

To enable the model to capture fine, high-frequency details, a crucial step is the application of positional encoding, which transforms the raw input coordinates. The standard approach for this encoding involves mapping inputs from a low-dimensional space into a higher-dimensional one. As an illustration, a single component of a coordinate can be represented as follows:

$$\gamma(p) = \left(\ldots, \sin(2^k \pi p), \cos(2^k \pi p), \ldots\right)_{k=0}^{L-1}$$

NeRF‘s novel method for view synthesis uses a simple MLP to create detailed, implicit 3D models, surpassing prior techniques. Despite its superior results, the approach is hindered by extremely slow training and rendering speeds and significant memory requirements. Figure 1 illustrates this fundamental concept, referencing the official developers‘ work.

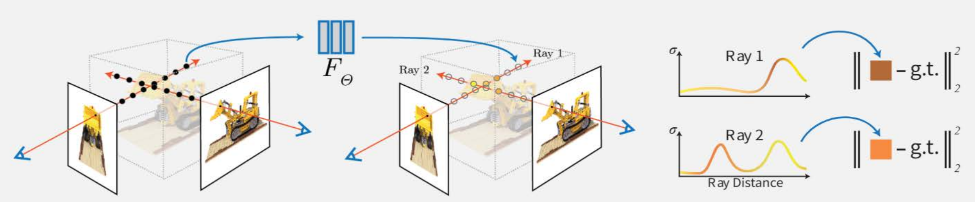
F_Θ
Ray 1
Ray 2
Ray 1
Ray 2
σ
σ
$\|\blacksquare - \text{g.t.}\|_2^2$
$\|\blacksquare - \text{g.t.}\|_2^2$
Ray Distance

Fig 1. Schematic diagram of the basic principle of NeRF technology

2.1.1 NeRF 3D reconstruction of static scenes

To address the slow processing of static environments in conventional NeRF, Alex Yu and colleagues introduced PlenOctrees [5]. This method utilizes octree precomputation to enable real-time rendering, effectively resolving the performance bottleneck of earlier approaches.PlenOctrees shows that significant acceleration can be achieved by "baking" implicit NeRF into an explicit hierarchical data structure.

Speed is crucial, but the ultimate goal is photo-realistic rendering. To improve the rendering quality and anti-aliasing of static scenes, Mip-NeRF was developed by Jonathan T. Barron et al. in 2021 [6]. This technology solves the problem of jagged artifacts in grid-based NeRF acceleration methods (such as Instant-NGP).

During the development of NeRF, the memory required to store large or detailed scene models has become a significant bottleneck. In 2022, Thomas Müller et al. used a multi-resolution structure to eliminate hash collisions as much as possible, and the entire system was implemented with highly optimized and fully fused CUDA cores to minimize memory bandwidth and computational operations [7]. This enables high-quality NeRF training to be completed within seconds and rendering within tens of milliseconds.

2.1.2 NeRF 3D reconstruction of dynamic scenes

However, considering the need to more accurately capture common non-rigid motions and even topological changes in the real world, researchers have proposed a series of more complex dynamic NeRF models. In 2021, Tretschk, E. and his team proposed NR - NeRF [8], which solved the technical problem of reconstructing and synthesizing new views of scenes with general non - rigid deformations from monocular videos and improved the constraints on the rigid regions of the scene.

In 2021, Park, K. et al. proposed HyperNeRF, which elevated NeRF to a higher - dimensional „hyperspace"[9]. Within this hyperspace, the 5D light representation of each input image is treated as a distinct cross-section. This framework accommodates topological changes in shape by simply moving the cross-section. Warping this slice also enables the generation of a more detailed template.

Regarding the 3D reconstruction of the human form, approaches such as Animatable NeRF yield positional inaccuracies in points and generate visual artifacts in details when managing significant non-rigid bodily transformations. In 2023, Xie introduced Deform2NeRF, a model that advances the Animatable NeRF framework [10]. It integrates a network dedicated to modeling non-rigid deformations with a module that fuses features from both 2D and 3D domains. The deformation component is specifically tasked with rectifying point location errors that arise from substantial non-rigid motion.Concurrently, the feature fusion module employs a cross-attention mechanism to pull features from various image perspectives and combine them with 3D data, thereby diminishing artifacts and ameliorating view inconsistency.

2.2 Three-dimensional reconstruction technology based on explicit representation — Three-dimensional Gaussian splatting

As a solution to the performance limitations of Neural Radiance Fields, Kerbl and his team introduced an approach in 2023 called 3D Gaussian Splatting [11]. This technique fundamentally changes the scene model, moving from a continuous implicit field to a collection of discrete, explicit 3D Gaussians. This shift enables photorealistic rendering at real-time frame rates. The core idea behind this technique is the parameterization of each individual Gaussian. Every Gaussian is made of its 3D location μ ,its shape and orientation via a covariance matrix Σ , its transparency level α , and its appearance, which is represented by spherical harmonics to handle view-dependent color effects. During rendering, a differentiable rasterizer projects these Gaussians onto the 2D screen and performs alpha blending after sorting them by depth. The final color value for any given pixel, denoted as C, is derived from the summation formula that follows:

$$C = \sum_{i=1}^{N} c_i \alpha' i \prod j = 1^{i-1} (1 - \alpha'_j)$$

The differentiable property of this formula is the cornerstone of optimization, enabling the system to directly adjust all Gaussian parameters through gradient descent. Combined with the strategy of adaptively adding and removing Gaussian functions, it can efficiently reproduce fine scene details. In order to establish a more systematic cognitive framework for three-dimensional reconstruction, this review refers to the official website of 3DGS technology developers and provides readers with a block diagram of the 3DGS technology principle, as shown in Figure 2.

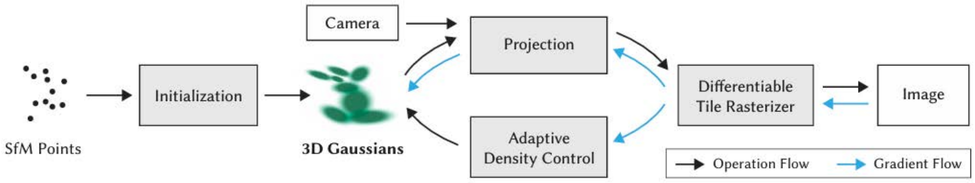

Fig 2. Linear block diagram of the principle of 3DGS technology

2.2.1 3DGS three-dimensional reconstruction of static scenes

For very large-scale static scenes or higher resolutions, CityGS developed by Liu in 2024 adopts A method of breaking down problems and training them individually[12]. Guided by global priors to achieve seamless fusion, it further improves the rendering speed. Beyond the pursuit of rendering speed, in 2023, Chen et al. used a two-pass deferred shading method to improve the visual fidelity of 3DGS for scenes such as specular reflection or complex lighting in static scenes [13], further enhancing the general applicability of 3DGS.

The large memory and storage requirements of 3DGS due to a large number of Gaussian functions and their properties have always been a core issue in the development of 3DGS technology. Lee and his team mainly adopted two strategies to achieve a compact 3D Gaussian splatting technology[14]. One is the learnable mask strategy, and the other is to use a grid-based neural field representation and learn the codebook of geometric attributes through vector quantization to compress Gaussian attributes. This achieves significant compression while maintaining quality, making 3DGS more practical for applications with limited storage or for network transmission.

2.2.2 3DGS 3D reconstruction of dynamic scenes

To address the dual challenges of achieving interactive rendering for dynamic scenes and maintaining efficiency in training and storage, researchers sought to overcome the per-frame modeling limitations inherent in 3DGS. This led to a swift expansion of research into the temporal domain, culminating in the development of 4DGS and a family of related variants.

In 2024, Wu‘s team extended 3D Gaussians to 4D Gaussian primitives, directly modeling the spatio - temporal volume [15]. Each 4D Gaussian has explicit spatio - temporal geometry and appearance features. By learning these 4D Gaussians to fit the underlying spatio - temporal volume, relevant information in space and time can be captured. The initial 4DGS method, although faster, usually involves more parameters, posing new challenges in terms of storage and training data requirements.

To achieve rapid reconstruction and real - time rendering of dynamic scenes, especially to explicitly model the attributes of each Gaussian point under monocular and multi - view video inputs, Lin proposed a dual - domain deformation model to explicitly model the deformation of each Gaussian point‘s attributes [16]. The compact dynamic representation of this method reduces the computational cost of the deformation model and introduces an adaptive timestamp scaling technique to avoid overfitting to frames with drastic motion.

In addition to basic motion and deformation, like NeRF, 3DGS technology is also constantly exploring how to represent more complex dynamic phenomena, such as changes in the topological structure of objects and applications in vast large - scale scenes. To overcome the challenge of precisely reconstructing and tracking dynamic surfaces with 3D Gaussians amidst complex topological changes, Zheng and his team proposed GauSTAR in 2025 [17]. The representation of dynamic objects is achieved in this method by linking Gaussian functions directly to corresponding mesh patches. To achieve surfaces with consistent topology, GauSTAR preserves the mesh structure and follows its movement using Gaussians. GauSTAR accommodates topological alterations by adaptively disassociating Gaussian distributions from the mesh in the affected regions. This unbinding strategy allows for the subsequent creation of new surface geometry and precise registration, all of which are based on the resulting optimized Gaussians.

3. Performance experiment of 3D reconstruction methods

3.1 Dataset

The development and evaluation of 3D reconstruction technology largely depend on high - quality and diverse datasets. Early object - centered datasets under controlled conditions (such as DTU) helped establish the baseline capabilities of NeRF and 3DGS. Subsequently, more complex and realistic real - world datasets (such as Tanks and Temples, LLFF, KITTI - 360, Waymo) have driven these technologies to address challenges such as scalability, robustness to imperfect data, dynamic elements, and lighting changes. To facilitate a more specific performance comparison of methods in static and dynamic scenarios, this review selects representative datasets. A subsequent analysis is then conducted to compile and assess how each approach performs on these selected datasets.

Local Light Field Fusion Dataset:

The LLFF dataset provides a key evaluation platform for NeRF and 3DGS models focused on synthesizing novel views. It is composed of forward-facing, real-world scenes and is specifically used to assess model performance under less-than-ideal, handheld capture conditions. It contains 24 real forward - facing scenes, captured by hand - held mobile phones (image resolution: 1008x756), and the poses are estimated using COLMAP.

Mip - NeRF 360 Dataset:

The Mip-NeRF 360 dataset acts as a benchmark for NeRF and 3DGS applications in synthesizing novel views for unbounded scenes. It was developed to handle „inside-out" capture scenarios, where cameras point in various directions and content exists at arbitrary distances, a challenge for conventional NeRFs. This dataset includes nine complex indoor and outdoor scenes, all featuring 360-degree photography around a central point with detailed backgrounds. To evaluate and improve how models handle expansive, intricate real-world environments, the dataset incorporates several key techniques. These include non-linear parameterization for the scene, online knowledge distillation, and a regularization strategy based on distortion.

Waymo Open Dataset:

As a large-scale, high-quality, multi-modal sensor resource, the Waymo Open Dataset is extensively utilized within the autonomous driving field. Collected by Waymo's self-driving vehicles across varied urban and suburban settings, it encompasses a range of lighting and weather conditions. The dataset is structured into three parts—perception, motion, and end-to-end driving,and delivers high-resolution, synchronized LiDAR and camera data, sensor calibrations, and vehicle pose details. It includes fine-grained object annotations for vehicles, pedestrians, cyclists, and traffic signs, offering robust data support for evaluating and training autonomous systems on tasks like 3D object detection, motion prediction, semantic segmentation, and behavior forecasting..

KITTI / KITTI - 360 Dataset:

It is also a dataset widely used in the field of autonomous driving, providing outdoor driving scene sequences containing dynamic elements such as vehicles and pedestrians. KITTI - 360 further provides more comprehensive 360 - degree sensor data for evaluating the reconstruction, dynamic scene processing, and semantic understanding capabilities in large - scale outdoor environments.

3.2 Reference indicators

To assess the comparative advantages of several methods, this study applies key quantitative indicators. These are used to numerically illustrate the results achieved by each approach across multiple data collections. The following key indicators were selected to guide this evaluation:

Peak Signal-to-Noise Ratio: It is an established method for assessing the quality of a generated image by measuring its deviation from a perfect reference. The underlying calculation determines the average of the squared intensity differences, pixel by pixel, between the synthesized output and its corresponding "ground truth" target. The final value, expressed in decibels, is often used as a proxy for quality, where higher numbers suggest a better result.

Structural Similarity Index : As a metric designed to better align with human visual assessment,it offers an alternative to traditional methods like PSNR [18]. Its core methodology is a composite analysis based on three key image attributes: overall illumination, dynamic range, and the inter-dependencies between pixels. The output is an index scaled from 0 to 1, where a higher value signifies a closer perceptual match between the two images.

Learned Perceptual Image Patch Similarity: To capture human-like judgments of image similarity, LPIPS metric was developed. This approach leverages the internal representations of neural networks that have been extensively trained on vast image datasets. By comparing these deep features, it offers a measure of perceptual distance.

Frames Per Second: It is a standard indicator used to measure rendering performance and speed. FPS represents the number of image frames that a graphics system or 3D model can generate and output per second. This value is the reciprocal of the time required to render a single frame. In the fields of 3D reconstruction and real-time graphics, FPS is the core standard for evaluating whether a method can achieve real-time interactive applications (such as games and virtual reality). A higher FPS value means smoother and more immediate visual feedback.

3.3 Performance comparison of static scenes

This review assesses how various 3D reconstruction techniques perform in non-dynamic environments by examining key metrics: PSNR, SSIM, LPIPS, and FPS. The comparison draws upon data from multiple NeRF and 3DGS-based models tested on the Mip-NeRF dataset, with a summary of these findings presented in Table 1.

Table 1.Performance metrics of NeRF and 3DGS methods on Mip-NeRF 360 datasets

Methods	PSNR↑	SSIM↑	LPIPS↓	FPS↑
BakedSDF[19]	24.51	0.697	0.309	539
3DGS [3]	27.20	0.815	0.214	251
Zip-NeRF[6]	28.54	0.836	0.177	0.25
K-Buffers[20]	29.19	0.859	0.126	N/A

Table 1 offers a comparative analysis of different techniques, with all results generated using the Mip-NeRF

360 benchmark. This benchmark is specifically tailored to evaluate the generation of new viewpoints within large-scale, stationary environments. From the compiled data, K-Buffers emerges as the leading method across every quality metric, demonstrating exceptional results that PSNR is 29.19, SSIM is 0.859, and LPIPS is 0.126. These scores position it as a top-tier solution for reconstructing static scenes with high fidelity.Zip - NeRF also demonstrates top - notch rendering quality, but its rendering speed is only 0.25 FPS, indicating that it focuses more on offline high - precision synthesis. When balancing rendering speed against visual quality, approaches based on 3D Gaussian Splatting exhibit a notable advantage. The original 3DGS achieves an extremely high rendering speed (251 FPS) while maintaining a highly competitive rendering quality (PSNR > 27). In sharp contrast, BakedSDF sacrifices some image quality to achieve the highest rendering speed of 539 FPS. Overall, for static scenes, although K - Buffers and Zip - NeRF reach the peak in terms of quality, 3DGS and its variants (such as SMERF) show great advantages and potential in achieving real - time, high - quality rendering.

3.4 Performance comparison of dynamic scenarios

For the evaluation of techniques in dynamic contexts, this review leverages performance data from the KITTI/KITTI 360 benchmark. Table 2 summarizes the quantitative findings from this comparative study.

Table 2. Performance metrics of NeRF and 3DGS methods on KITTI/KITTI-360 datasets

Methods	PSNR↑	SSIM↑	LPIPS↓	FPS↑
3DGS[3]	19.54	0.776	0.224	125
SplatFlow[21]	28.32	0.932	0.089	44
MVSNeRF [22]	18.44	0.638	0.317	0.025
EVolSplat [23]	23.26	0.797	0.179	83.81

Table 2 presents a performance assessment for several methods tested on the large-scale, dynamic autonomous driving scenarios provided by KITTI/KITTI-360.Due to the complexity of the scenarios, the overall metric scores are generally lower than those in static scenarios. Among these methods, the Gaussian Splatting variant specifically designed for dynamic scenarios performs the best. SplatFlow achieved the best results in the perceptual metrics LPIPS (0.089) and SSIM (0.932), which demonstrates the effectiveness of explicitly modeling scene dynamics.It is worth noting that although the unmodified 3DGS achieves the fastest rendering speed of 125 FPS, its rendering quality (PSNR 19.54) drops significantly compared to other dynamic methods, highlighting the limitations of the basic model in handling dynamic elements. On the other hand, NeRF-based method such as MVSNeRF has extremely low rendering speeds and are not feasible for real-time applications. Operating at 83.81 FPS, EVolSplat strikes an effective equilibrium between the quality of the final image and the rate at which it is rendered. In conclusion, for the challenge of reconstructing scenes with motion, the most advanced solutions currently available are based on the Gaussian Splatting framework. The leading edge in this research area is represented by approaches like SplatFlow, which integrate either temporal data or optical flow estimation..

4. Conclusion

This review systematically reviews the domain of 3D reconstruction, focusing on its two most influential contemporary methods: the implicit modeling paradigm of neural radiance fields and the explicit representation offered by 3D Gaussian splatting.A central finding of this work is that the paradigm shift from implicit neural representations to explicit Gaussian-based methods constitutes a major breakthrough, enabling photorealistic rendering at interactive speeds. When evaluated in both static and dynamic contexts, 3DGS and its derivatives demonstrate clear superiority in terms of visual fidelity and processing speed. For modeling complex dynamic scenes, 4DGS derivative methods that incorporate temporal information or motion priors represent the current state-of-the-art. Nevertheless, NeRF maintains its strength in rendering with ultra-high geometric detail. A key challenge for implicit 3D reconstruction is to continuously improve NeRF's rendering speed to achieve real-time performance. Looking forward, this review believes that the development of efficient hybrid models is the future direction of work. High-fidelity 3D reconstruction coupled with rapid rendering is attainable by merging the continuous nature of implicit representations with the performance benefits of explicit structures. Secondly, the real - time interaction ability of 3D reconstruction technology can be enhanced to further expand the technology from basic reconstruction to content creation and other aspects.

References

[1] Wang P, Liu Y, Liu Z, et al. NeRF-based 3D human-body reconstruction: a survey. arXiv preprint arXiv:2305.16823, 2023.

[2] Mildenhall B, Srinivasan P P, Tancik M, et al. NeRF: representing scenes as neural radiance fields for view synthesis. European Conference on Computer Vision (ECCV), 2020: 405-421.

[3] Kerbl B, Kopanas G, Leimkühler T, Drettakis G. 3D gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics, 2023, 42(4): 1-14.

[4] Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R, Ng R. NeRF: representing scenes as neural radiance fields for view synthesis. European Conference on Computer Vision (ECCV), 2020: 405-421.

[5] Yu A, Li R M, Tancik M, et al. PlenOctrees for real-time rendering of neural radiance fields. IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 5752-5761.

[6] Barron J T, Mildenhall B, Verbin D, et al. Zip-NeRF: anti-aliased grid-based neural radiance fields. IEEE/CVF International Conference on Computer Vision (ICCV), 2023: 19755-19764.

[7] Müller T, Evans A, Schied C, Keller A. Instant neural graphics primitives with a multiresolution hash encoding. ACM Transactions on Graphics, 2022, 41(4): 102.

[8] Tretschk E, Tewari A, Golyanik V, et al. Non-rigid neural radiance fields: reconstruction and novel view synthesis of a dynamic scene from monocular video. IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 12959-12970.

[9] Park K, Sinha U, Hedman P, et al. HyperNeRF: a higher-dimensional representation for topologically varying neural radiance fields. ACM Transactions on Graphics, 2021, 40(6): 238.

[10] Xie X, Guo W, Li J, Liu J, Xu J. Deform2NeRF: non-rigid deformation and 2D-3D feature fusion with cross-attention for dynamic human reconstruction. Electronics, 2023, 12(21): 4382

[11] Kerbl B, Kopanas G, Leimkühler T, Drettakis G. 3D gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics, 2023, 42(4): 139.

[12] Liu Y, Guan H, Luo C, et al. CityGaussian: real-time high-quality large-scale scene rendering with gaussians. European Conference on Computer Vision (ECCV), 2024.

[13] Chen Y, Fan L, Wang Y, Liu X, Wang J. Split-SSG: splitting screen-space-accurate 3D gaussian splatting for high-fidelity real-time rendering. arXiv preprint arXiv:2312.01317, 2023

[14] Lee J C, Rho D, Sun X, Ko J H, Park E. Compact 3D gaussian representation for radiance field. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024.

[15] Wu G, et al. 4D gaussian splatting for real-time dynamic scene rendering. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 20310-20320.

[16] Lin Y, Dai Z, Zhu S, Yao Y. Gaussian-Flow: 4D reconstruction with dynamic 3D gaussian particle. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024

[17] Zheng C, Xue L, Zarate J, Song J. GauSTAR: gaussian surface tracking and reconstruction. arXiv preprint arXiv:2501.10283, 2025.

[18] Wang Z, Bovik A C, Sheikh H R, Simoncelli E P. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

[19] Yariv L, Hedman P, Reiser C, et al. BakedSDF: Meshing Neural SDFs for Real-Time View Synthesis. ACM Transactions on Graphics (TOG), 2023, 42(4): 1-16.

[20] Ren Z, Chen W, Zhang J, et al. K-Buffers: A Plug-in Method for Enhancing Neural Fields with Multiple Buffers. arXiv preprint arXiv:2405.18318, 2024.

[21] Guédon A, Lepetit V. SplatFlow: Splatting Gaussian Kernels for Dynamic Scenes from Monocular Videos. arXiv preprint arXiv:2311.12337, 2023.

[22] Chen A, Xu Z, Zhao F, et al. MVSNeRF: Fast Generalizable Radiance Field Reconstruction from Multi-View Stereo. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 14124-14133.

[23] Miao Z, Yang S, Yu K, et al. EVO-Splat: Evolving 3D Gaussian Splatting for Real-time Dynamic Scene Reconstruction. arXiv preprint arXiv:2403.18946, 2024.