

3DGSR: Implicit Surface Reconstruction with 3D Gaussian Splatting

XIAOYANG LYU, The University of Hong Kong, Hong Kong
YANG-TIAN SUN, The University of Hong Kong, Hong Kong
YI-HUA HUANG, The University of Hong Kong, Hong Kong
XIUZHE WU, The University of Hong Kong, Hong Kong
ZIYI YANG, The University of Hong Kong, Hong Kong
YILUN CHEN, Shanghai AI Lab, China
JIANGMIAO PANG, Shanghai AI Lab, China
XIAOJUAN QI*, The University of Hong Kong, Hong Kong

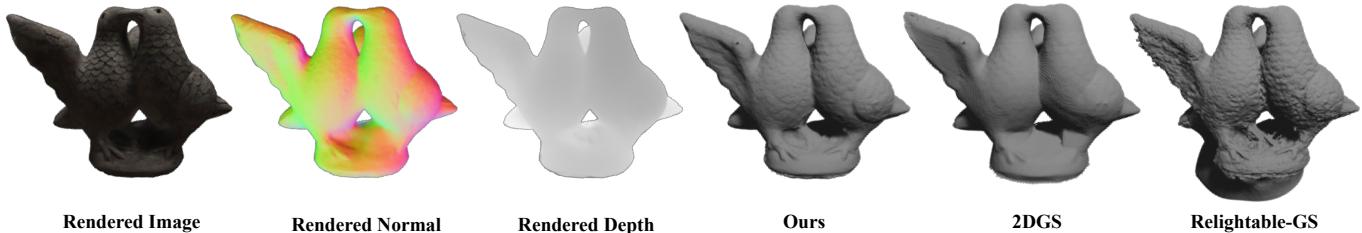


Fig. 1. Our method, called 3DGSR, achieves accurate 3D surface reconstruction with rich details while maintaining the efficiency and high-quality rendering of 3DGS. The left part shows the capability of our method to achieve high-quality reconstruction and rendering results simultaneously. 2DGs [Huang et al. 2024] is the state-of-the-art Gaussian-based reconstruction method.

In this paper, we present an implicit surface reconstruction method with 3D Gaussian Splatting (3DGS), namely 3DGSR, that allows for accurate 3D reconstruction with intricate details while inheriting the high efficiency and rendering quality of 3DGS. The key insight is to incorporate an implicit signed distance field (SDF) within 3D Gaussians for surface modeling, and to enable the alignment and joint optimization of both SDF and 3D Gaussians. To achieve this, we design coupling strategies that align and associate the SDF with 3D Gaussians, allowing for unified optimization and enforcing surface constraints on the 3D Gaussians. With alignment, optimizing the 3D Gaussians provides supervisory signals for SDF learning, enabling the reconstruction of intricate details. However, this only offers sparse supervisory signals to the SDF at locations occupied by Gaussians, which is insufficient for learning a continuous SDF. Then, to address this limitation, we incorporate volumetric rendering and align the rendered geometric attributes (depth, normal) with that derived from 3DGS. In sum, these two designs

allow SDF and 3DGS to be aligned, jointly optimized, and mutually boosted. Our extensive experimental results demonstrate that our 3DGSR enables high-quality 3D surface reconstruction while preserving the efficiency and rendering quality of 3DGS. Besides, our method competes favorably with leading surface reconstruction techniques while offering a more efficient learning process and much better rendering qualities.

CCS Concepts: • Computing methodologies → Mesh geometry models.

Additional Key Words and Phrases: Gaussian Splatting, Implicit Function, Signed Distance Function, Volumetric Rendering

ACM Reference Format:

Xiaoyang Lyu, Yang-Tian Sun, Yi-Hua Huang, Xiuzhe Wu, Ziyi Yang, Yilun Chen, Jiangmiao Pang, and Xiaojuan Qi. 2024. 3DGSR: Implicit Surface Reconstruction with 3D Gaussian Splatting. *ACM Trans. Graph.* 43, 6, Article 198 (December 2024), 12 pages. <https://doi.org/10.1145/3687952>

*Corresponding author

Authors' addresses: Xiaoyang Lyu, The University of Hong Kong, Hong Kong, shawlyu@connect.hku.hk; Yang-Tian Sun, The University of Hong Kong, Hong Kong, sunyt98@connect.hku.hk; Yi-Hua Huang, The University of Hong Kong, Hong Kong, huangyihua16@mails.ucas.ac.cn; Xiuzhe Wu, The University of Hong Kong, Hong Kong, xzwu@eee.hku.hk; Ziyi Yang, The University of Hong Kong, Hong Kong, 14ziyiyang@gmail.com; Yilun Chen, Shanghai AI Lab, Shanghai, China, chenyilun95@gmail.com; Jiangmiao Pang, Shanghai AI Lab, Shanghai, China, pangjiangmiao@gmail.com; Xiaojuan Qi, The University of Hong Kong, Hong Kong, xjq@eee.hku.hk.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM. 0730-0301/2024/12-ART198 \$15.00 <https://doi.org/10.1145/3687952>

1 INTRODUCTION

3D Gaussian Splatting (3DGS) [Kerbl et al. 2023] has emerged as a new state-of-the-art approach for high-quality novel view synthesis. This method represents the geometry and appearance of a 3D scene as a set of Gaussians, which are then optimized from posed multi-view images. Thanks to its Gaussian rasterization pipeline, 3DGS achieves real-time efficiency, even when rendering high-resolution outputs. Despite its impressive rendering quality and efficiency, 3DGS generates only noisy, incomplete point clouds for 3D geometry and struggles to accurately reconstruct the 3D surface of a scene [Tang et al. 2023a]. These, however, are highly desired in various geometry-related applications, such as 3D reconstruction, geometry editing, animation, and relighting, among others.

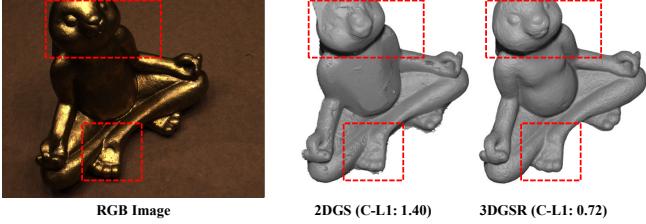


Fig. 2. Oversmooth reconstruction results from 2DGs [Huang et al. 2024] lead to poorer performance, reflected by a higher chamfer- \mathcal{L}_1 error (1.40) compared to 3DGSR (0.72).

This motivates us to investigate how to enable 3DGS for high-quality surface reconstruction while preserving its rendering capabilities and speed. The challenges faced by 3DGS in accurately representing a 3D surface stem from (1) its unstructured point-based geometry representation (i.e., the center of Gaussian), making it difficult to extract 3D surfaces through post-processing, such as Poisson surface reconstruction [Kazhdan and Hoppe 2013]; and (2) its elliptical shape, which is not naturally aligned with surface representations without proper constraints. Furthermore, the optimization process does not consider the geometric constraints among the points that are indeed constrained by the underlying geometry (i.e., surfaces), leading to noisy points. The limited number of views utilized for learning further exacerbates this issue.

To this end, some recent works have explored the regularization of 3DGS [Guédon and Lepetit 2023] or the use of 2DGs [Huang et al. 2024] or Gaussian surfels [Dai et al. 2024] for learning. Subsequently, a post-processing step, employing depth fusion [Huang et al. 2024] or Poisson reconstruction [Dai et al. 2024], is required to obtain a 3D surface. Despite achieving promising results, these approaches tend to sacrifice detailed structures, resulting in smoother surfaces (Fig. 2) and suffering from limitations inherent to the reconstruction method. Moreover, these formulations essentially decouple surface reconstruction from GS learning, which not only prevents the surface from regularizing the learning of GS but also leads to alignment issues and the accumulation of errors.

In this paper, we introduce 3DGSR, a simple yet effective method that integrates a lightweight neural implicit signed distance field (SDF) within Gaussians for surface modeling, which is optimized and jointly aligned with 3D Gaussians. We start by examining the alignment of an SDF field and 3DGS, enabling unified optimization and synchronization of the surface and appearance. Inspired by the intuition that Gaussians near the surface should possess a high opacity value, we explore a *tight coupling* strategy that uses SDF to directly control Gaussian opacity with a differentiable SDF-to-opacity transformation function. Although this improves reconstruction quality, we find that surface reconstruction can be disrupted by high-frequency textures, particularly in low-frequency flat surface areas, leading to noisy reconstruction (Fig. 3). This issue can also be observed in the concurrent work GOF [Yu et al. 2024b] (Fig. 3), which directly leverages Gaussian opacity and volume rendering for surface extraction. Therefore, instead of tight coupling SDF and 3D Gaussians, we further investigate a *loose coupling* strategy that only utilizes SDF to constrain the distribution and orientations of

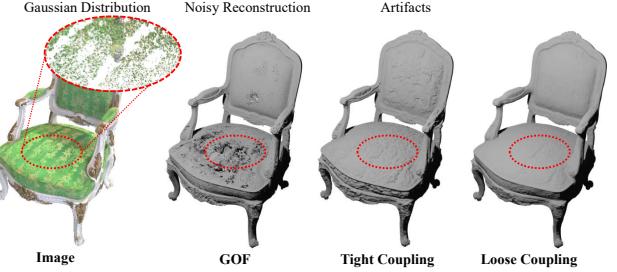


Fig. 3. Imperfect reconstruction from the non-uniform distributions of Gaussians. Directly defining the surface from the Gaussian points may lead the noisy reconstruction or artifacts.

Gaussians. This approach forces the Gaussians to be near the surface derived from SDF and aligns their orientations (the direction of the shortest axis) with the normal of the surface. By doing so, we facilitate the alignment of surface and 3D Gaussians while providing flexibility for the SDF to model surface geometry and the 3D Gaussians to capture appearance details, yielding superior results in both rendering and reconstruction (Fig. 1).

By employing the coupling strategies, we enable the optimization of Gaussians to supply supervisory signals for the SDF. However, these signals are only provided at locations occupied by the Gaussians, resulting in sparse supervision for the SDF. This sparsity proves insufficient for optimizing a continuous SDF and may lead to the creation of redundant structures. To address this issue, we integrate volumetric rendering techniques [Mildenhall et al. 2021; Wang et al. 2021] and align the rendered depth (or normal) with that obtained from the 3D Gaussians through a consistency loss. This approach effectively regularizes the regions not covered by the 3D Gaussians. It is worth noting that during each learning iteration, we only sample a few rays for rendering depth and normal using a single-layer MLP. This not only avoids introducing significant computational costs for training but also enables fast learning of SDF from the geometric cues provided by 3DGS. In turn, the improved SDF helps to regularize the 3D Gaussians effectively.

Extensive experiments conducted on multiple datasets demonstrate that our method can produce high-quality reconstruction results while maintaining the efficiency and rendering qualities of 3DGS. When compared to neural implicit reconstruction methods [Tang et al. 2023b; Wang et al. 2021; Yariv et al. 2021], our approach performs on par or even surpasses them while being significantly faster and achieving superior novel view synthesis results. In comparison to GS-based methods [Huang et al. 2024; Yu et al. 2024b], our approach outperforms them in both surface reconstruction and novel view synthesis.

2 RELATED WORK

Our method builds upon prior works in 3D reconstruction with multi-view images to enable fast surface reconstruction and novel view synthesis. We discuss close related prior works in neural implicit representation and 3D Gaussian splatting (3DGS).

Neural Implicit Representations. In recent years, the fields implicitly predicted by neural networks have gained significant attention

in both novel view synthesis [Lombardi et al. 2019; Mildenhall et al. 2021; Sitzmann et al. 2019] and 3D reconstruction [Jiang et al. 2020; Kellnhofer et al. 2021; Liu et al. 2020b; Niemeyer et al. 2020; Yariv et al. 2020; Yu et al. 2024a]. Neural Radiance Field (NeRF) [Mildenhall et al. 2021] has demonstrated excellent performance in high-quality novel view synthesis by modeling scenes with a density field and a view-dependent color field. However, NeRF’s density field cannot extract high-quality surfaces. To achieve accurate surface reconstruction, NeuS [Wang et al. 2021] represents the scene with a signed distance function (SDF) and derives the alpha value from the precondition that the zero-level surface contributes the most to the volume rendering. Following NeuS, Darmon *et al.* [Darmon et al. 2022] improves the reconstruction quality with an additional warping loss that ensures photometric consistency. Neuralangelo [Li et al. 2023b] introduces multi-resolution hash grids [Müller et al. 2022] and regularizes the model with numerical gradients and coarse-to-fine training. However, these methods often require extensive training time, which can limit their practical applicability.

3D Gaussian Splatting. Most recently, 3DGS [Kerbl et al. 2023] has demonstrated remarkable performance in terms of high-level rendering quality and real-time rendering. Subsequent works have extended 3DGS to handle 3D/4D generation [Ren et al. 2023; Tang et al. 2023a; Yang et al. 2023a; Yi et al. 2023; Zielonka et al. 2023], dynamic scenes [Das et al. 2023; Huang et al. 2023; Li et al. 2023a; Luiten et al. 2023; Yang et al. 2023b, 2024], relighting [Gao et al. 2023], and animation [Jung et al. 2023; Ye et al. 2023]. Some concurrent works [Guédon and Lepetit 2023; Huang et al. 2024; Yu et al. 2024b] aim to reconstruct the 3D surface from multi-view images using 3DGS. SuGaR [Guédon and Lepetit 2023] enforces Gaussians to have limited overlap with others during the optimization of 3DGS. After that, an initial mesh can be extracted by performing Poisson reconstruction [Kazhdan and Hoppe 2013] on Gaussians. The mesh is then bound with flat Gaussians distributed on its faces and optimized to obtain refined results. 2DGS [Huang et al. 2024] represents the points as 2D surfels and leverages TSDF fusion to obtain the final mesh. GOF [Yu et al. 2024b] leverages ray-tracing-based volume rendering of 3D Gaussians by identifying its level. Different from the SuGaR and 2DGS, they extract the mesh from a trained tetrahedral grid with marching tetrahedra. Though these methods achieve promising reconstruction results, their surface still misses fine-grained geometry details and high-quality rendering results. Here, we derive 3DGSR combined with explicit Gaussian rasterization and implicit volumetric rendering to get detailed surface reconstruction and high-quality rendering results.

3 PRELIMINARIES

3D Gaussian Splatting (3DGS). 3DGS [Kerbl et al. 2023] employs 3D Gaussian points, characterized by a central point μ and a covariance matrix Σ , to effectively render images from given viewpoints:

$$G(x) = \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right), \quad (1)$$

where x is the point position in world space.

To facilitate optimization, Σ is factorized into the product of a scaling matrix S , represented by scale factors s , and a rotation matrix

R encoded by a quaternion $r \in \mathbb{R}^4$:

$$\Sigma = RSS^T R^T, \quad (2)$$

For rendering novel views, the 3D Gaussians are projected onto a 2D image plane according to elliptical weighted average (EWA) [Zwicker et al. 2001]:

$$\Sigma' = JW\Sigma W^T J^T, \quad (3)$$

where W is the viewing transformation matrix, and J is the Jacobian derived from projective transformations. The final color $C(u)$ at pixel u is rendered by combining N ordered Gaussian splats using point-based α -blending:

$$C(u) = \sum_{i \in N} T_i \alpha_i c_i, \quad T_i = \prod_{j=1}^{i-1} (1 - \alpha_j). \quad (4)$$

Here, c_i is determined by the spherical harmonic (SH) function and the viewing direction, while α_i is computed from the 2D covariance Σ' , modulated by the opacity σ of each Gaussian [Kerbl et al. 2023]. In addition to rendering color c_i , Eq. (4) is also employed to render normal and depth as detailed below.

Normal and Depth Estimation from Gaussian. Typically, each 3D Gaussian can be treated as a point by using the centers of 3D Gaussians. Then, we can leverage the estimated depth and render them into a depth map using point-based alpha blending [Gao et al. 2023] as

$$\mathcal{D} = \sum_{i \in N} T_i \alpha_i d_i, \quad (5)$$

where d_i is the depth of i th point. We then project the rendering depth into 3D space and get the pseudo normal from the local plane.

4 METHOD

Given posed multiview images $\{\mathcal{I}_k\}$ and a point cloud obtained from the Structure from Motion algorithm [Schönberger and Frahm 2016], 3DGSR integrates an implicit signed distance field (SDF) within 3D Gaussians (Fig. 4) to reconstruct high-quality surfaces while preserving the rendering quality and efficiency of 3DGS. To achieve this, 3DGSR incorporates two key designs: **Alignment of Surface and 3D Gaussians** (Sec. 4.1) We investigate coupling strategies to synchronize surface and 3D Gaussians, enabling their alignment and joint optimization. The *tight coupling* strategy utilizes a differentiable SDF-to-opacity transformation function to connect the SDF field and the opacity of 3D Gaussians; and the *loose coupling strategy* uses surface derived from SDF to constrain the distribution of Gaussians, make them near the surface and align the orientation of 3D Gaussians with surface normals.

By aligning 3D Gaussian and the surface, we allow unified optimization of 3D Gaussians and SDF. However, the supervisory signal only exists at locations occupied by 3D Gaussians and is thus not sufficient for supervising the dense SDF field. Thus, we propose the second design: **Regularization with Volumetric Rendering** (Sec. 4.2). We employ volumetric rendering to render depth and normal by querying points along the cast ray from the neural implicit SDF and align them with that derived from 3D GS with a consistency regularization loss. This provides dense supervisory signals for learning the SDF. The two designs work together to enforce surface constraints on 3D Gaussians, which in turn allows 3D Gaussians to assist in

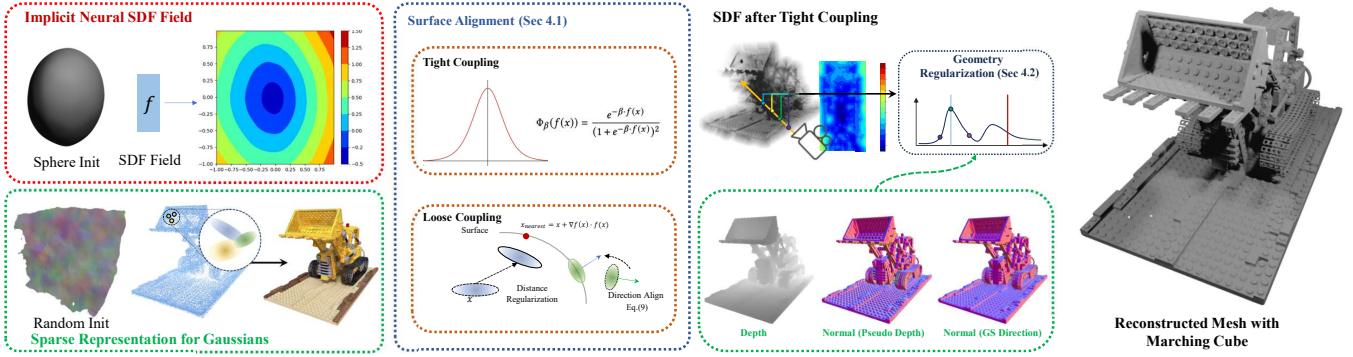


Fig. 4. Pipeline of our proposed approach for implicit surface reconstruction. We model the surface with an implicit SDF field, with which the SDF value of each 3D Gaussian can be predicted. We propose two different coupling strategies to make the distribution of Gaussians align with the implicit SDF field. The geometry attributes of 3D Gaussians serve as a regularization for the SDF field, while the rendered image is supervised by the captured image.

optimizing the SDF, resulting in high-quality surfaces with aligned 3D Gaussians.

Compared to previous neural implicit surface reconstruction methods [Wang et al. 2021; Yariv et al. 2021], our approach utilizes 3DGS for efficient and high-quality rendering while avoiding the dependence on time-consuming color prediction branches and volumetric rendering loss for optimization. Instead, it can be supervised by 3DGS and directly learn the SDF from optimized geometric clues (e.g., depth and normal) derived from 3DGS, making the training process more than 20 times faster (12 hours vs. 30 minutes). In comparison to Gaussian-based reconstruction methods [Huang et al. 2024], our approach unifies the optimization of 3DGS and an implicit surface, aligning them in an end-to-end manner. This eliminates the need for post-processing steps such as depth fusion or Poisson reconstruction and circumvents their limitations and error accumulations, leading to improved reconstruction quality and better alignment.

4.1 Alignment of Surfaces and Gaussians

The vanilla 3DGS method optimizes millions of Gaussians to accurately fit the training views, enabling photorealistic synthesis of novel views. However, its main emphasis is on fitting the target views, which does not guarantee high-quality 3D geometries. Moreover, the 3D Gaussians should be constrained by the underlying geometry (i.e., surfaces) of the 3D scene, which is absent in existing methods.

Here, we introduce neural implicit SDF, denoted as $f : \mathbb{R}^3 \rightarrow \mathbb{R}^1$ for surface modeling, which maps each 3D position $x \in \mathbb{R}^3$ to its signed distance from the surface. This is implemented using multi-resolution hash grids [Müller et al. 2022] in conjunction with a single-layer MLP. The surface S of a scene can then be derived by using the zero-level set of its SDF, defined by:

$$S = \{x \in \mathbb{R}^3 \mid f(x) = 0\}. \quad (6)$$

Given the SDF f , the next question is to connect 3DGS with the implicit SDF to enable unified surface and appearance modeling and

optimization. Here, we explore two coupling strategies: tight coupling with a differentiable SDF-to-opacity transformation function and loose coupling with distance and orientation regularization.

Tight Coupling Strategy. Intuitively, points near the surface will have large opacity values and contribute more to rendering images. Thus, we introduce a differentiable SDF-to-opacity transformation function to tightly couple the SDF field and 3DGS. Consequently, we select a bell-shaped function $\Phi_\beta(f(x))$ controlled by a hyperparameter β , expressed as:

$$\Phi_\beta(f(x)) = \frac{e^{-\beta \cdot f(x)}}{(1 + e^{-\beta \cdot f(x)})^2}, \quad (7)$$

where β is a learnable parameter controlling the shape of the function. This function transforms SDF values at a query location into the opacity of the corresponding Gaussian at that location. By linking the SDF with 3D Gaussians using a differentiable function, the SDF field can be updated through the optimization of 3D Gaussians by minimizing the photometric loss:

$$\mathcal{L}_c = \lambda \mathcal{L}_{\text{D-SSIM}} + (1 - \lambda) \mathcal{L}_1, \quad (8)$$

This loss function encourages the Gaussian at a location x that contributes more to rendering a pixel to have a higher opacity value $\Phi_\beta(f(x))$. This, in turn, encourages the SDF $f(x)$ to be closer to 0 and, therefore, nearer to the surface.

Loose Coupling. Although the tight coupling strategy can align Gaussian points with the implicit surface and enable unified optimization for enhanced reconstructions, it may slightly compromise rendering quality and produce unsatisfactory reconstruction results with visible artifacts (See Fig. 3). Upon investigation, we find that the surface and color appearances indeed exhibit different levels of detail and frequencies, which could be a contributing factor.

For example, as shown in Fig. 3, a chair with flat low-frequency surface areas may require numerous non-uniformly distributed Gaussian points to represent its high-frequency textures. In areas with dense Gaussian points, the tight coupling strategy enforces the SDF values to be closer to zero, potentially leading to noisy surface structures with unnecessary details (see Fig. 10) if the Gaussian

locations are not accurate. This suggests that a tight alignment might not be the optimal solution.

Thus, we introduce a loose coupling strategy as an alternative to the tight coupling approach. The central concept involves using the surface derived from the SDF to regularize the distribution of 3D Gaussians. This strategy ensures that 3D Gaussians are positioned near the surface and their orientations are aligned with the surface normals. By aligning the surface normals instead of directly constraining the opacity, we can prevent interference with the supervision needed for high-frequency appearances.

Specifically, the direction of a 3D Gaussian at location x_g is determined by the direction of its shortest axis of the covariance matrix, represented by n_g . The normal at x_g in the SDF field is derived using $\nabla f(x_g)$. Note we normalize $\nabla f(x_g)$ to a unit vector. We then align the direction of the 3D Gaussian and the surface normal by minimizing the direction-aligned loss, which is defined as follows:

$$\mathcal{L}_a = \|1 - n_g \nabla f(x_g)\|. \quad (9)$$

Additionally, we constrain the 3D Gaussians to be near the surface. To achieve this, we query the SDF value of each Gaussian point x_g and calculate its nearest surface location using the gradient of the SDF field as follows:

$$x_{\text{nearest}} = x_g + f(x_g) \cdot \nabla f(x_g), \quad (10)$$

where $x_{\text{nearest}} \in R$ represent the nearest 3D points of point x_g on the surface from SDF. Then, we can use \mathcal{L}_1 loss to minimize their distance. As shown in Fig. 3, despite its simplicity, the loose coupling strategy yields detailed and accurate surface reconstruction results where the 3D Gaussians near the surface with their orientations aligned with surface normals. Moreover, this approach also helps reduce the number of Gaussians by preventing the generation of Gaussian points that are far away from the surface. Our experiments demonstrate that we can decrease the number of Gaussian points by 20% without compromising, or even enhancing, the rendering quality. If not explicitly specified, we use the loose coupling strategy for alignment by default.

4.2 Regularization with Volumetric Rendering

By integrating the SDF field with 3D Gaussians and jointly training them, as discussed in Section 4.1, we can achieve high-fidelity rendering results with an optimized SDF field. However, due to the sparse and discrete nature of 3D Gaussians, the supervision signals mentioned earlier are only available for locations occupied by 3D Gaussians, primarily near the surface. This limited availability is insufficient to regularize the continuous SDF, leading to artifacts in locations not covered by 3D Gaussians, such as floating artifacts in unoccupied free spaces, as shown in Fig. 4.

Volumetric rendering employs ray-casting to render a pixel, sampling points along a ray that includes both occupied and free points. Thus, we propose incorporating volumetric rendering to optimize the dense SDF field. Specifically, we use volumetric rendering to render depth and normal, which can be directly derived from the SDF field, and align them with those derived from 3DGS for optimization. There are two main reasons for using rendered depth and normal instead of color. First, rendering color would require an additional color prediction MLP network, significantly increasing

the learning time. Second, and more importantly, learning the SDF directly from geometric clues (e.g., depth and normal) can alleviate the burden of appearance modeling, accelerate the learning process, and enhance the quality.

More specifically, for each pixel, the ray cast from it is denoted as $\{\mathbf{p}(t) = r_o + t \cdot r_d \mid t > 0\}$, where r_o is the camera center and r_d is the unit direction vector of the ray. We accumulate the depth $\hat{D}(r)$ and normal $\hat{\mathbf{N}}(r)$ along the cast ray [Wang et al. 2021]:

$$\tilde{D}(r) = \sum_{i=1}^M T_i^r \alpha_i t_i^r, \quad \tilde{\mathbf{N}}(r) = \sum_{i=1}^M T_i^r \alpha_i \tilde{\mathbf{n}}_i^r, \quad (11)$$

where $\tilde{\mathbf{n}}$ is the normal computed by the gradient of SDF field ∇f , T_i^r and α_i represent the transmittance and alpha value (a.k.a opacity) of the sample point, and their values can be computed by

$$T_i^r = \prod_{j=1}^{i-1} (1 - \alpha_j), \quad \alpha_i = \max \left(\frac{\phi_s(f(x_i)) - \phi_s(f(x_{i+1}))}{\phi_s(f(x_i))}, 0 \right), \quad (12)$$

where ϕ_s is the Sigmoid function.

After that, we utilize the depth \mathcal{D} and normal \mathcal{N} rendered by Gaussian Splatting to supervise the volumetric rendering results \mathbf{D} and \mathbf{N} by optimizing the following objectives:

$$\mathcal{L}_{vd} = \sum_{r \in \mathcal{R}} \|\mathcal{D}(r) - \tilde{D}(r)\|_2, \quad (13)$$

and

$$\mathcal{L}_{vn} = \sum_{r \in \mathcal{R}} \|\mathcal{N}(r) - \tilde{\mathbf{N}}(r)\|_1 + \|1 - \mathcal{N}(r) \tilde{\mathbf{N}}(r)\|_1, \quad (14)$$

where \mathcal{R} is the union of all training rays. These constraints regularize all the sampling points along the training rays, resulting in dense supervision for the implicit field and eliminating floating artifacts. As shown in Fig. 4, there exist noisy SDF distributions (negative value) floating over the free space between the camera and the object, which is then removed by applying the regularization with volumetric rendering. Accurate geometry can be obtained through our careful training strategy and model design.

4.3 Training Objectives and Implementation Details

To train 3DGRS, our objective is to get high-quality rendering and reconstruction results. Thus, the final loss function is defined by

$$\mathcal{L} = \mathcal{L}_c + \alpha_1 \mathcal{L}_{vd} + \alpha_2 \mathcal{L}_{vn} + \alpha_3 \mathcal{L}_a + \alpha_4 \mathcal{L}_{eik}$$

where \mathcal{L}_{eik} is the Eikonal equation [Gropp et al. 2020] to penalize the deviation of the magnitude of grad f from unit length:

$$\mathcal{L}_{eik} = \mathbb{E}_X [(\|\nabla f - 1\|)^2]. \quad (15)$$

In our experiments, we set $\alpha_1 = 1.0, \alpha_2 = 0.1, \alpha_3 = 0.0001, \alpha_4 = 0.1$.

We implement the 3DGSR with CUDA and tinyCUDANN [Müller 2021], building upon the framework of 3DGS. We customize the normal and depth maps CUDA rendering kernel. During training, we increase the number of 3D Gaussian primitives following the adaptive control strategy in 3DGS. During the volumetric rendering stage, we sample $N = 1024$ rays per step.

For the synthetic data, we initialize the implicit function as a sphere, and the position of Gaussian is randomly initialized. For the real-world scenes, which frequently exhibit intricate geometry and noisy poses, we adopt the SFM points as the initialized Gaussian

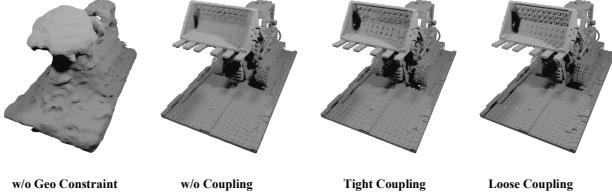


Fig. 5. Qualitative comparisons of different coupling methods and the influence of geometric constraint.

position, and the network is still initialized as a sphere. Furthermore, we also apply the sparse SFM points to ensure the consistency between the rendered depth D and projected depth map D_{sfm} with \mathcal{L}_2 loss. Specifically, when optimizing the pose using SFM algorithms like COLMAP, we can also obtain sparse 2D key points along with their corresponding 3D point locations. Given an image and its camera pose, we project visible 3D points into the camera space and use the z-value as the depth D_{sfm} for the corresponding 2D key points. We then apply an \mathcal{L}_2 loss to enforce consistency between the rendered depth D and the projected sparse depth D_{sfm} based on the measurements. Our models are trained for 30000 iterations, and all the experiments are conducted on a single NVIDIA RTX 3090 GPU.

5 EXPERIMENTS

5.1 Experimental Settings

Datasets. Our evaluation encompasses three datasets: the NeRF synthetic dataset [Mildenhall et al. 2021], the DTU real-captured dataset [Jensen et al. 2014], and the Tanks & Temples (TNT) large-scale dataset [Knapitsch et al. 2017]. The NeRF synthetic dataset consists of eight objects, each rendered via path tracing to generate 100 views. The DTU and TNT datasets are processed to obtain camera poses in accordance with [Gao et al. 2023], while the images and masks are processed following the methodology outlined in [Liu et al. 2020a]. The image resolutions for the DTU and TNT datasets are 800×600 and 1920×1080 , respectively.

Comparisons. We benchmark our method on two tasks: surface reconstruction and novel view synthesis. Our approach is compared against several state-of-the-art **1) Neural surface reconstruction** methods, including NeuS [Wang et al. 2021], NeuSG [Chen et al. 2023], RelightableGaussian [Gao et al. 2023], NeRF2Mesh [Tang et al. 2023b], Neuralangelo [Li et al. 2023b], NeUS2 [Wang et al. 2023], BakedSDF [Yariv et al. 2023], VolSDF [Yariv et al. 2021], and **2) Gaussian-based reconstruction methods** such as SuGaR [Guédon and Lepetit 2023], 2DGS [Huang et al. 2024], and GOF [Yu et al. 2024b]. It is worth noting that GOF [Yu et al. 2024b] is a concurrent work. Additionally, we also compare the novel view synthesis capabilities of our method with works that focus solely on **novel view synthesis** tasks, including NeRF [Mildenhall et al. 2021], Mip-NeRF [Barron et al. 2021], Ins-NGP [Müller et al. 2022], NeUS2 [Wang et al. 2023], and 3DGS [Kerbl et al. 2023]. The quality of image synthesis is quantified using PSNR. For surface reconstruction accuracy, we employ Chamfer distance and the F1 score [Knapitsch et al. 2017].

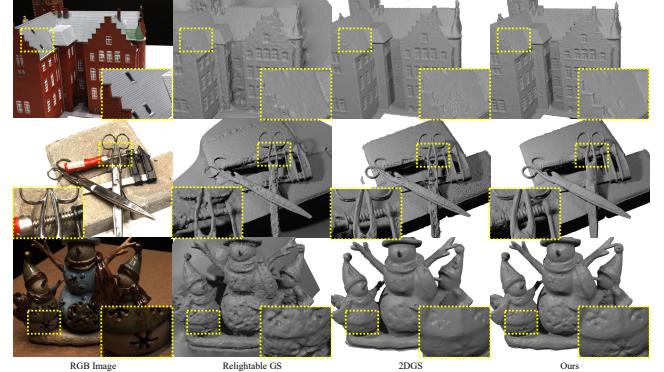


Fig. 6. Qualitative comparisons of reconstructed results between Relightable 3D Gaussian [Gao et al. 2023], 2DGS [Huang et al. 2024] and our method on the DTU dataset. Our method yields the highest-quality surface reconstruction preserving fine details and intricate structures.

5.2 Main Results and Comparisons

Quantitative analysis. Tables 1, 2 and 3 present the quantitative results for NeRF-synthetic and DTU datasets, respectively. On the NeRF-synthetic dataset (Table 1 and Table 3), our approach achieves the best performance in both 3D surface reconstruction and novel view synthesis. Notably, our approach even outperforms the leading neural surface reconstruction methods for reconstruction evaluation while being much faster in training and yielding much better novel view synthesis results. Besides, our approach also delivers higher novel-view synthesis results than 3DGS [Kerbl et al. 2023] and 2DGS [Huang et al. 2024] in terms of PSNR, while achieving performance comparable to 3DGS on SSIM and LPIPS. On the DTU dataset (Table 2), our method also exhibits the best performance in terms of average surface reconstruction quality when compared with Gaussian-based reconstruction methods. It also performs favorably when compared with leading neural implicit reconstruction methods, which take much longer time to train and often yield inferior rendering qualities. We find that existing methods that achieve better reconstruction qualities tend to sacrifice the rendering qualities with lower performance than the NeRF baseline. This might be attributed to the different levels of detail required by surface reconstruction and novel view synthesis. Overall, our 3DGSR achieves excellent performance in both surface reconstruction and novel view synthesis.

Qualitative analysis. We provide qualitative comparisons of the reconstruction results in Figures 6 and 8. The state-of-the-art NeRF-based 3D reconstruction method, NeuS [Wang et al. 2021], fails to preserve fine details such as intricate Mic textures, Lego tracks, and stripes on the Hotdog disk. The 2DGS method [Huang et al. 2024], which utilizes 2D Gaussians and relies on an additional post-processing depth fusion step to obtain 3D surfaces, tends to produce over-smoothed reconstruction results, as evidenced by the smoothed Lego track and house. GOF [Yu et al. 2024b] employs Gaussian opacities and volumetric rendering to extract meshes directly from 3DGS, unifying geometry and appearance modeling for 3DGS. However, due to the inherent differences in the level of detail for geometry

		Chair	Drums	Ficus	Hotdog	Lego	Materials	Mic	Ship	Avg
PSNR	NeRF [Mildenhall et al. 2021]	34.17	25.08	30.39	36.82	33.31	30.03	34.78	29.30	31.74
	Ins-NGP [Müller et al. 2022]	35.00	26.02	33.51	37.40	36.39	29.78	36.22	31.10	33.18
	Mip-NeRF [Barron et al. 2021]	35.14	25.48	33.29	37.48	35.70	30.71	36.51	30.41	33.09
	3D-GS [Kerbl et al. 2023]	35.36	26.15	34.87	37.72	35.78	30.00	35.36	30.80	33.32
	NeuS [Wang et al. 2021]	31.22	24.85	27.38	36.04	34.06	29.59	31.56	26.94	30.20
	NeuS2 [Wang et al. 2023]	31.55	24.65	29.61	34.84	31.63	27.68	34.23	28.92	30.39
	NeRO [Liu et al. 2023]	28.74	24.88	28.38	32.13	25.66	24.85	28.64	26.55	27.48
	BakedSDF [Yariv et al. 2023]	31.65	20.71	26.33	36.38	32.69	30.48	31.52	27.55	29.66
	NeRF2Mesh [Tang et al. 2023b]	34.25	25.04	30.08	35.70	34.90	26.26	32.63	29.47	30.88
	2DGS [Huang et al. 2024]	35.05	26.05	35.57	37.36	35.10	29.74	35.09	30.60	33.07
Ours (Tight Coupling)	Ours (Tight Coupling)	34.85	26.08	35.17	36.88	34.90	30.03	36.44	31.48	33.23
	Ours (Loose Coupling)	35.69	26.35	35.65	37.98	36.19	30.31	37.12	31.65	33.86
$C\text{-}L_1$	VolSDF [Yariv et al. 2021]	1.18	4.03	3.01	3.22	2.26	1.57	1.13	6.42	2.86
	NeuS [Wang et al. 2021]	3.99	1.27	0.94	2.12	2.56	1.39	1.00	5.38	2.33
	NeRO [Liu et al. 2023]	1.27	1.97	1.22	1.88	1.90	1.33	0.87	4.95	1.92
	BakedSDF [Yariv et al. 2023]	1.83	1.43	1.09	1.68	1.13	1.36	0.84	3.88	1.66
	NeRF2Mesh [Tang et al. 2023b]	1.62	1.11	0.65	2.73	1.93	1.42	0.78	2.20	1.55
	RelightableGaussian [Gao et al. 2023]	3.65	2.34	1.26	3.11	1.63	1.35	1.76	3.35	2.31
	Ours (Tight Coupling)	1.01	0.95	0.69	1.92	1.35	1.35	1.15	3.35	1.50
Ours (Loose Coupling)	Ours (Loose Coupling)	0.99	0.93	0.68	1.90	1.32	1.33	1.15	2.64	1.37

Table 1. We assess the quality of synthesized images and the accuracy of surface reconstruction, with each cell colored to indicate the best and second best. Our method is compared against various state-of-the-art (SOTA) approaches in tasks of novel view synthesis and surface reconstruction. It outperforms all competitors in both tasks, achieving the highest PSNR and the lowest Chamfer- L_1 ($C\text{-}L_1$) distance.

Scan ID	24	37	40	55	63	65	69	83	97	105	106	110	114	118	122	Mean
NeRF [Mildenhall et al. 2021]	1.90	1.60	1.85	0.58	2.28	1.27	1.47	1.67	2.05	1.07	0.88	2.53	1.06	1.15	0.96	1.49
NeuS [Wang et al. 2021]	1.00	1.37	0.93	0.43	1.10	0.65	0.57	1.48	1.09	0.83	0.52	1.20	0.35	0.49	0.54	0.84
VolSDF [Yariv et al. 2021]	1.14	1.26	0.81	0.49	1.25	0.70	0.72	1.29	1.18	0.70	0.66	1.08	0.42	0.61	0.55	0.86
NeuS2 [Wang et al. 2023]	0.56	0.76	0.49	0.37	0.92	0.71	0.76	1.22	1.08	0.63	0.59	0.89	0.40	0.48	0.55	0.70
Neuralangelo [Li et al. 2023b]	0.37	0.72	0.35	0.35	0.87	0.54	0.53	1.29	0.97	0.73	0.47	0.74	0.32	0.41	0.43	0.61
ReGS [Gao et al. 2023]	1.21	1.00	1.10	0.87	1.03	1.51	1.51	1.06	1.63	0.97	1.36	1.21	0.94	1.38	1.26	1.20
SuGaR [Guédon and Lepetit 2023]	1.47	1.33	1.13	0.61	2.25	1.71	1.15	1.63	1.62	1.07	0.79	2.45	0.98	0.88	0.79	1.33
2DGS [Huang et al. 2024]	0.48	0.91	0.39	0.39	1.01	0.83	0.81	1.36	1.27	0.76	0.70	1.40	0.40	0.76	0.52	0.80
GOF [Yu et al. 2024b] (concurrent)	0.50	0.82	0.37	0.37	1.12	0.74	0.73	1.18	1.29	0.68	0.77	0.90	0.42	0.66	0.49	0.74
Ours	0.44	0.96	0.40	0.36	1.02	0.80	0.64	1.20	1.08	0.97	0.54	0.72	0.37	0.52	0.42	0.70

Table 2. Quantitative assessment on the DTU dataset with each cell colored to indicate the best second best and third best. Our method achieves the highest average quality of surface reconstruction and the lowest Chamfer- L_1 distance with real-world data.

Method	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
NeRF [Barron et al. 2023]	31.74	0.947	0.056
NeuS2 [Wang et al. 2023]	30.39	0.944	0.064
3D-GS [Kerbl et al. 2023]	33.32	0.970	0.031
2DGS [Huang et al. 2024]	33.07	0.967	0.033
Ours	33.86	0.970	0.031

Table 3. More quantitative assessment on the NeRF synthetic dataset.

and appearance modeling, their results tend to produce noisy and non-existent structures needed for modeling high-frequency appearances, as seen in the noisy balls and Lego tracks in Figure 8. In

contrast, our approach is capable of producing high-quality reconstructions, effectively preserving the finer details and maintaining the overall structure of the objects. More qualitative results and comparisons are included in Figure 9 and 10.

Speed comparison. We compared the training and inference speed of our method with Neuralangelo [Li et al. 2023b] and 2DGS [Huang et al. 2024], as shown in Table 4. When training on the DTU dataset, neural surface reconstruction methods like Neuralangelo [Li et al. 2023b] require over 30 hours to produce the final results, while Gaussian-splatting-based methods like 2DGS [Huang et al. 2024] take only 18.8 minutes. Our method, 3DGSR, takes 30 minutes—slightly longer than 2DGS [Huang et al. 2024], but significantly faster than Neuralangelo [Li et al. 2023b]. This discrepancy arises

Method	Training Speed	Inference
Neuralangelo [Li et al. 2023b]	> 30 h	< 1 FPS
2DGS [Huang et al. 2024]	18.8 min	175 FPS
Ours	30 min	175 FPS

Table 4. Training and inference speed comparison on DTU dataset.

Method	C- \mathcal{L}_1	PSNR	LPIPS	SSIM
static β (<i>Tight coupling</i>)	1.95	33.02	0.0385	0.966
Poisson (<i>Tight coupling</i>)	2.03	-	-	-
Ours (<i>Tight coupling</i>)	1.50	33.23	0.0340	0.968
Ours (<i>Loose coupling</i>)	1.37	33.86	0.0330	0.971
w/o depth,normal	3.63	32.83	0.0380	0.963
w/o normal	1.74	33.18	0.0383	0.967

Table 5. Ablation study of the surface and GS alignment on the NeRF Synthetic dataset and the volumetric rendering depth (normal) loss, where we evaluate the effect of each proposed component. C- \mathcal{L}_1 is an abbreviation for the Chamfer- \mathcal{L}_1 distance.

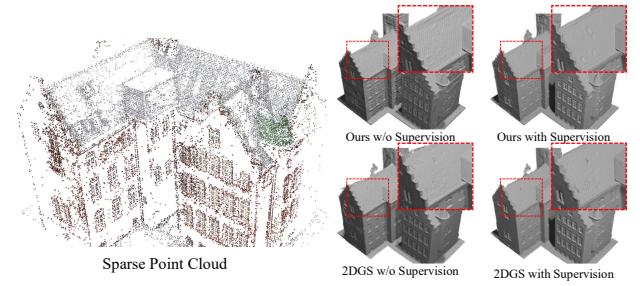
because 3DGSR incorporates geometric constraints that effectively optimize the implicit representation, though querying the rays during training incurs additional time costs. However, when rendering an 800x600 image on the DTU dataset, both 2DGS [Huang et al. 2024] and our method achieve a frame rate of 175 fps. This is because surface regularization is applied solely during the training phase, allowing for the simultaneous learning of high-quality surfaces and Gaussians. Once the optimization is complete, the surface can be extracted offline and reused for future applications.

5.3 Ablation Studies

We conduct ablation studies on the NeRF Synthetic dataset in Table 5 to evaluate the effectiveness of each module in our method. Four different configurations are investigated to train our model. We also show the qualitative results in Fig. 10.

Volumetric Constraints. The volumetric constraints have a minimal impact on novel view synthesis (PSNR: 32.83-33.86) but significantly influence the reconstruction results (Chamfer- \mathcal{L}_1 : 3.63-1.37). This is because the volumetric rendering for depth and normal supervision can provide dense supervisory signals for learning the continuous SDF. Without these constraints, numerous noisy structures would be generated in areas not occupied by Gaussians, negatively affecting the overall reconstruction quality. As shown in Fig. 10, if we directly use the tight coupling strategy without any geometry constraint, we can only get the coarse result.

Reconstruction Strategy. Compared with the marching cubes on the SDF field for surface extraction, we also employ Poisson reconstruction to extract surface from Gaussians with normals derived from the implicit SDF, which yields a much lower performance. This demonstrates the advantage of using learned implicit SDF to represent the surface that is continuous, more tolerant to noisy supervisions, and allows for easy extraction of high-fidelity surfaces. Notably, the surface is modeled as a hash grid and a single-layer

Fig. 7. Sparse point cloud and the influence of projected depth supervision D_{sf} in 2DGS and 3DGSR.

Method	D_{sf}	Chamfer- \mathcal{L}_1
2DGS [Huang et al. 2024]	Yes	0.85
2DGS [Huang et al. 2024]	No	0.80
Ours	Yes	0.70
Ours	No	0.81

Table 6. Ablation study of sparse depth supervision. Yes indicates the use of D_{sf} as a supervision signal, while No indicates that D_{sf} is not used for supervision.

MLP, which is lightweight and does not introduce many burdens for computation.

Coupling Strategy. We experiment with the tight coupling and loose coupling strategies proposed by our approach. For the tight coupling strategy, the learning of the parameter β impacts our performance. In contrast, the loose coupling strategy leads to improved results in both reconstruction and novel view synthesis, demonstrating its effectiveness in balancing the different aspects of the modeling process.

Projected Sparse Depth D_{sf} . To ensure a fair comparison, we conducted experiments to analyze the influence of the supervision by projected sparse depth D_{sf} on both 3DGSR and 2DGS [Huang et al. 2024], presenting the quantitative results in Table 6 and qualitative results in Fig. 7. As shown in Table 6, we observed that adding sparse depth cues negatively affects the performance of 2DGS [Huang et al. 2024], causing a worsening in Chamfer- \mathcal{L}_1 from 0.80 to 0.85. This performance drop may stem from the fact that 2DGS [Huang et al. 2024] relies on depth fusion for reconstruction, as opposed to our approach, which jointly optimizes and aligns the surface with the 3D Gaussians. In contrast, the addition of sparse depth supervision proves beneficial in our method, resulting in improved quantitative results (Chamfer- \mathcal{L}_1 from 0.81 to 0.70). These cues assist SDF optimization in regions where image supervision is sparse and camera poses are inaccurate, though they are less impactful in areas where sufficient observations are available. When trained using only image signals, our method can be comparable with 2DGS [Huang et al. 2024]. Fig. 7 further demonstrates that 2DGS [Huang et al. 2024] shows only minor improvements with sparse depth supervision, whereas our method achieves more detailed reconstructions even in the absence of such supervision. In conclusion, these depth cues are

valuable for improving performance in cases of inaccurate camera poses and ill-posed sparse view data, especially in real-world scenes. Our approach can effectively utilize these free cues to enhance reconstruction quality.

6 CONCLUSION

In this paper, we introduce 3DGSR, an efficient method for high-quality surface reconstruction using 3DGS. Our approach is based on two key components: 1) the integration of neural implicit SDF and its alignment with 3DGS, and 2) the utilization of volumetric rendering and the SDF-and-Gaussian geometry consistency regularization for SDF optimization. The first component allows for the joint optimization of SDF and 3D Gaussians, with the optimization of Gaussians providing supervision signals for learning the SDF. The second component supplies additional supervision signals for refining the SDF in areas not occupied by Gaussians, using a consistency loss that aligns depth (normal) from SDF with that derived from Gaussians. Our extensive experiments showcase the effectiveness of 3DGSR in reconstructing high-quality surfaces outperforming those obtained from state-of-the-art reconstruction pipelines, without compromising the rendering capabilities and efficiency of 3D Gaussians. We hope our approach could inspire more future work. Despite its promising results, our approach still has limited capability to handle unbounded scenes, due to the use of a hash grid and implicit SDF, and also has worse results on real datasets with noisy pose estimation results. Furthermore, it may encounter difficulties in reconstructing transparent objects. These limitations highlight potential areas for further improvement and development in our method.

7 ACKNOWLEDGEMENT

This work has been supported by the Hong Kong Research Grant Council - Early Career Scheme (Grant No. 27209621), General Research Fund Scheme (Grant No. 17202422, 17212923), Theme-based Research (Grant No. T45-701/22-R) and RGC Matching Fund Scheme (RMGS). Part of the described research work is conducted in the JC STEM Lab of Robotics for Soft Materials funded by The Hong Kong Jockey Club Charities Trust.

REFERENCES

- Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. 2021. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5855–5864.
- Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. 2023. Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Fields. *ICCV* (2023).
- Hanlin Chen, Chen Li, and Gim Hee Lee. 2023. NeuSG: Neural Implicit Surface Reconstruction with 3D Gaussian Splatting Guidance. *arXiv preprint arXiv:2312.00846* (2023).
- Pinxuan Dai, Jiamin Xu, Wenxiang Xie, Xinguo Liu, Huamin Wang, and Weiwei Xu. 2024. High-quality Surface Reconstruction using Gaussian Surfels. In *SIGGRAPH 2024 Conference Papers*. Association for Computing Machinery. <https://doi.org/10.1145/3641519.3657441>
- François Darmon, Bénédicte Basclé, Jean-Clément Devaux, Pascal Monasse, and Mathieu Aubry. 2022. Improving neural implicit surfaces geometry with patch warping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6260–6269.
- Devikalyan Das, Christopher Wewer, Raza Yunus, Eddy Ilg, and Jan Eric Lenssen. 2023. Neural Parametric Gaussians for Monocular Non-Rigid Object Reconstruction. *arXiv preprint arXiv:2312.01196* (2023).
- Jian Gao, Chun Gu, Youtian Lin, Hao Zhu, Xun Cao, Li Zhang, and Yao Yao. 2023. Relightable 3D Gaussian: Real-time Point Cloud Relighting with BRDF Decomposition and Ray Tracing. *arXiv preprint arXiv:2311.16043* (2023).
- Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. 2020. Implicit Geometric Regularization for Learning Shapes. In *Proceedings of Machine Learning and Systems* 2020. 3569–3579.
- Antoine Guédon and Vincent Lepetit. 2023. SuGaR: Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High-Quality Mesh Rendering. *arXiv preprint arXiv:2311.12775* (2023).
- Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2024. 2D Gaussian Splatting for Geometrically Accurate Radiance Fields. *SIGGRAPH* (2024).
- Yi-Hua Huang, Yang-Tian Sun, Ziyi Yang, Xiaoyang Lyu, Yan-Pei Cao, and Xiaojuan Qi. 2023. SC-GS: Sparse-Controlled Gaussian Splatting for Editable Dynamic Scenes. *arXiv preprint arXiv:2312.14937* (2023).
- Rasmus Jensen, Anders Dahl, George Vogiatzis, Engin Tola, and Henrik Aanaes. 2014. Large scale multi-view stereopsis evaluation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 406–413.
- Yue Jiang, Dantong Ji, Zhizhong Han, and Matthias Zwicker. 2020. Sdfdiff: Differentiable rendering of signed distance fields for 3d shape optimization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1251–1261.
- HyunJin Jung, Nikolas Brasch, Jifei Song, Eduardo Perez-Pellitero, Yiren Zhou, Zhihao Li, Nassir Navab, and Benjamin Busam. 2023. Deformable 3D Gaussian Splatting for Animatable Human Avatars. *arXiv preprint arXiv:2312.15059* (2023).
- Michael Kazhdan and Hugues Hoppe. 2013. Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)* 32, 3 (2013), 1–13.
- Petr Kellnhofer, Lars C Jebe, Andrew Jones, Ryan Spicer, Kari Pulli, and Gordon Wetzstein. 2021. Neural lumigraph rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4287–4297.
- Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics* 42, 4 (July 2023). <https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/>
- Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. 2017. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (ToG)* 36, 4 (2017), 1–13.
- Zhan Li, Zhang Chen, Zhong Li, and Yi Xu. 2023a. Spacetime Gaussian Feature Splatting for Real-Time Dynamic View Synthesis. *arXiv preprint arXiv:2312.16812* (2023).
- Zhaohui Li, Thomas Müller, Alex Evans, Russell H Taylor, Mathias Unberath, Ming-Yu Liu, and Chen-Hsuan Lin. 2023b. Neuralangelo: High-Fidelity Neural Surface Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8456–8465.
- Lingjie Liu, Jiatao Gu, Kyaw Zaw Lin, Tat-Seng Chua, and Christian Theobalt. 2020a. Neural sparse voxel fields. *Advances in Neural Information Processing Systems* 33 (2020), 15651–15663.
- Shaohue Liu, Yinda Zhang, Songyu Peng, Boxin Shi, Marc Pollefeys, and Zhaopeng Cui. 2020b. Dist: Rendering deep implicit signed distance function with differentiable sphere tracing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019–2028.
- Yuan Liu, Peng Wang, Cheng Lin, Xiaoxiao Long, Jiepeng Wang, Lingjie Liu, Taku Komura, and Wenping Wang. 2023. NeRO: Neural Geometry and BRDF Reconstruction of Reflective Objects from Multiview Images.
- Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. 2019. Neural volumes: learning dynamic renderable volumes from images. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–14.
- Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. 2023. Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. *arXiv preprint arXiv:2308.09713* (2023).
- Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 1 (2021), 99–106.
- Thomas Müller. 2021. *tiny-cuda-nn*. <https://github.com/NVLabs/tiny-cuda-nn>
- Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)* 41, 4 (2022), 1–15.
- Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. 2020. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3504–3515.
- Jiawei Ren, Liang Pan, Jiaxiang Tang, Chi Zhang, Ang Cao, Gang Zeng, and Ziwei Liu. 2023. DreamGaussian4D: Generative 4D Gaussian Splatting. *arXiv preprint arXiv:2312.17142* (2023).
- Johannes Lutz Schönberger and Jan-Michael Frahm. 2016. Structure-from-Motion Revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein. 2019. Scene representation networks: Continuous 3d-structure-aware neural scene representations. *Advances in Neural Information Processing Systems* 32 (2019).

- Jiaxiang Tang, Jiawei Ren, Hang Zhou, Ziwei Liu, and Gang Zeng. 2023a. DreamGaussian: Generative Gaussian Splatting for Efficient 3D Content Creation. *arXiv preprint arXiv:2309.16653* (2023).
- Jiaxiang Tang, Hang Zhou, Xiaokang Chen, Tianshu Hu, Errui Ding, Jingdong Wang, and Gang Zeng. 2023b. Delicate Textured Mesh Recovery from NeRF via Adaptive Surface Refinement. In *International Conference on Computer Vision (ICCV)*.
- Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. 2021. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689* (2021).
- Yiming Wang, Qin Han, Marc Habermann, Kostas Daniilidis, Christian Theobalt, and Lingjie Liu. 2023. Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3295–3306.
- Xiaofeng Yang, Yiwen Chen, Cheng Chen, Chi Zhang, Yi Xu, Xulei Yang, Fayao Liu, and Guosheng Lin. 2023a. Learn to Optimize Denoising Scores for 3D Generation: A Unified and Improved Diffusion Prior on NeRF and 3D Gaussian Splatting. *arXiv preprint arXiv:2312.04820* (2023).
- Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. 2023b. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. *arXiv preprint arXiv:2309.13101* (2023).
- Zeyu Yang, Hongye Yang, Zijie Pan, Xiatian Zhu, and Li Zhang. 2024. Real-time Photorealistic Dynamic Scene Representation and Rendering with 4D Gaussian Splatting. In *International Conference on Learning Representations (ICLR)*.
- Lior Yaniv, Jitao Gu, Yoni Kasten, and Yaron Lipman. 2021. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems* 34 (2021), 4805–4815.
- Lior Yaniv, Peter Hedman, Christian Reiser, Dor Verbin, Pratul P. Srinivasan, Richard Szeliski, Jonathan T. Barron, and Ben Mildenhall. 2023. BakedSDF: Meshing Neural SDFs for Real-Time View Synthesis. *arXiv* (2023).
- Lior Yaniv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. 2020. Multiview neural surface reconstruction by disentangling geometry and appearance. *Advances in Neural Information Processing Systems* 33 (2020), 2492–2502.
- Keyang Ye, Tianjia Shao, and Kun Zhou. 2023. Animatable 3d gaussians for high-fidelity synthesis of human motions. *arXiv preprint arXiv:2311.13404* (2023).
- Taoran Yi, Jiemin Fang, Guanjun Wu, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Qi Tian, and Xinggang Wang. 2023. Gaussiandreamer: Fast generation from text to 3d gaussian splatting with point cloud priors. *arXiv preprint arXiv:2310.08529* (2023).
- Mulin Yu, Tao Lu, Liming Xu, Lihan Jiang, Yuanbo Xiangli, and Bo Dai. 2024a. GSDF: 3DGs Meets SDF for Improved Rendering and Reconstruction. *arXiv:2403.16964* [cs.CV] <https://arxiv.org/abs/2403.16964>
- Zehao Yu, Torsten Sattler, and Andreas Geiger. 2024b. Gaussian Opacity Fields: Efficient High-quality Compact Surface Reconstruction in Unbounded Scenes. *arXiv:2404.10772* (2024).
- Wojciech Zienlonka, Timur Bagautdinov, Shunsuke Saito, Michael Zollhöfer, Justus Thies, and Javier Romero. 2023. Drivable 3d gaussian avatars. *arXiv preprint arXiv:2311.08581* (2023).
- M. Zwicker, H. Pfister, J. van Baar, and M. Gross. 2001. Surface Splatting. In *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*. 371–378.

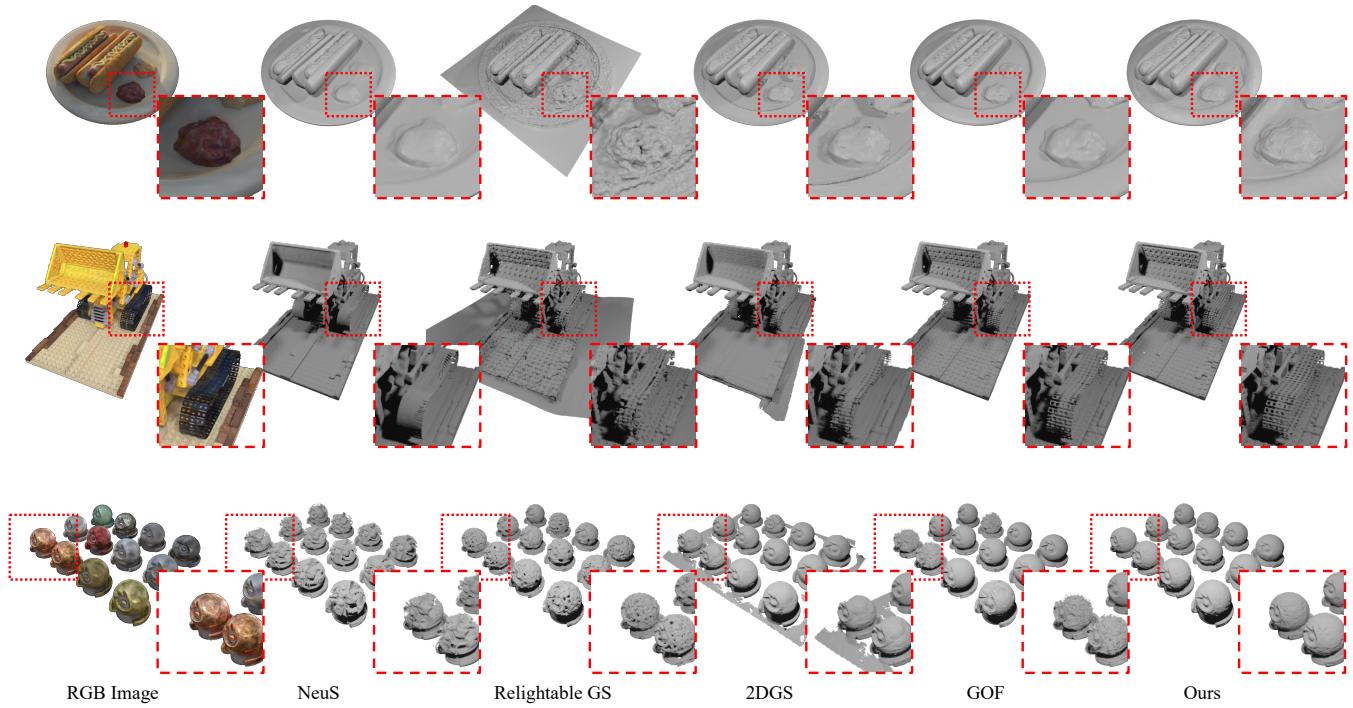


Fig. 8. Qualitative comparison on NeRF-synthetic.

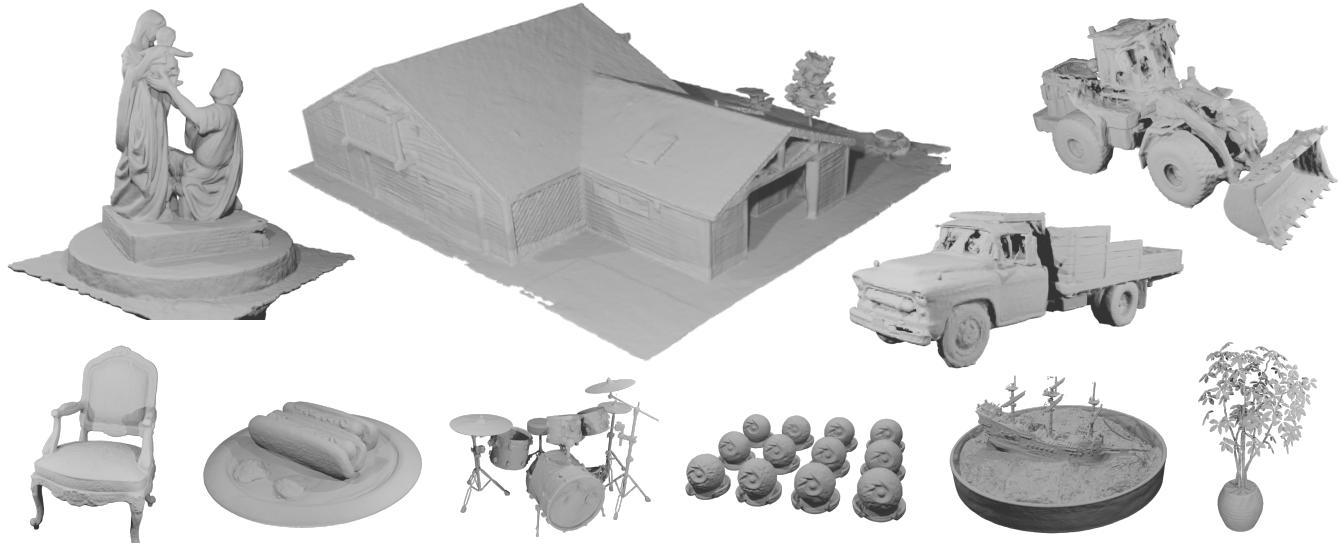


Fig. 9. Qualitative results of our 3DGSR on the NeRF synthetic dataset [Mildenhall et al. 2021] and TNT dataset [Knapitsch et al. 2017].

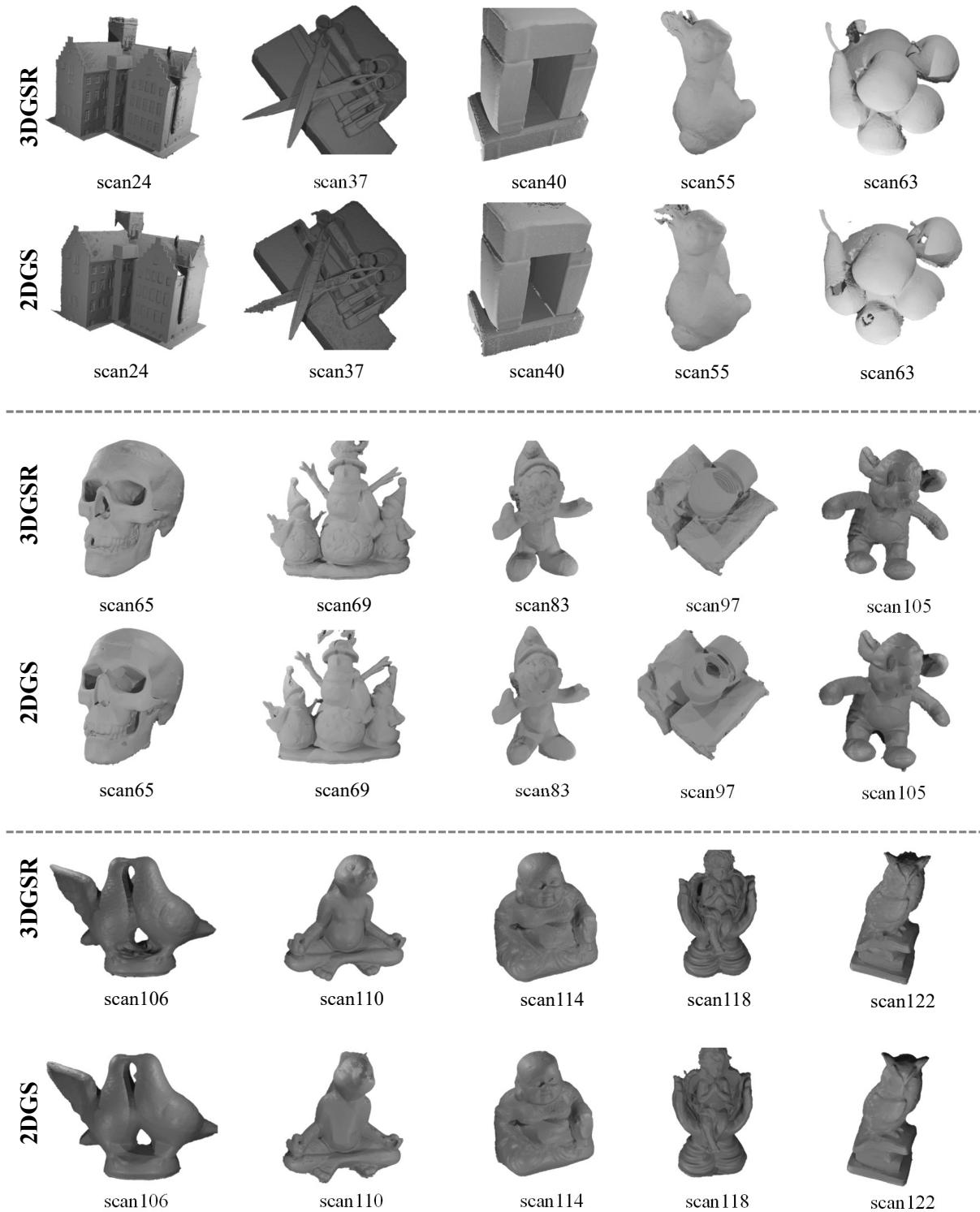


Fig. 10. Comparison of surface reconstruction using our 3DGSR and 2DGS [Huang et al. 2024].