

NeuSG: Neural Implicit Surface Reconstruction with 3D Gaussian Splatting Guidance

Hanlin Chen Chen Li Gim Hee Lee

Department of Computer Science, National University of Singapore
 {hanlin.chen, gimhee.lee}@comp.nus.edu.sg

Abstract

Existing neural implicit surface reconstruction methods have achieved impressive performance in multi-view 3D reconstruction by leveraging explicit geometry priors such as depth maps or point clouds as regularization. However, the reconstruction results still lack fine details because of the over-smoothed depth map or sparse point cloud. In this work, we propose a neural implicit surface reconstruction pipeline with guidance from 3D Gaussian Splatting to recover highly detailed surfaces. The advantage of 3D Gaussian Splatting is that it can generate dense point clouds with detailed structure. Nonetheless, a naive adoption of 3D Gaussian Splatting can fail since the generated points are the centers of 3D Gaussians that do not necessarily lie on the surface. We thus introduce a scale regularizer to pull the centers close to the surface by enforcing the 3D Gaussians to be extremely thin. Moreover, we propose to refine the point cloud from 3D Gaussians Splatting with the normal priors from the surface predicted by neural implicit models instead of using a fixed set of points as guidance. Consequently, the quality of surface reconstruction improves from the guidance of the more accurate 3D Gaussian splatting. By jointly optimizing the 3D Gaussian Splatting and the neural implicit model, our approach benefits from both representations and generates complete surfaces with intricate details. Experiments on Tanks and Temples verify the effectiveness of our proposed method.

1. Introduction

Surface reconstruction from multiple calibrated views is a fundamental task in 3D computer vision. The conventional approach involves estimating a point cloud from images through multi-view stereo (MVS) techniques [4, 7, 21, 39, 40], followed by the extraction of a triangular mesh from the point cloud [17, 26]. More recently, neural implicit surface reconstruction has emerged as a competitive alternative, particularly effective for surfaces with limited texture and

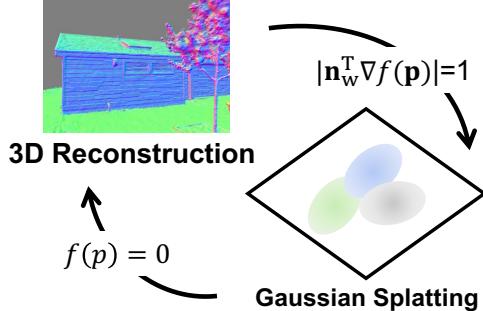


Figure 1. An illustration of joint optimization of implicit surface reconstruction and 3D Gaussian Splatting.

non-Lambertian properties. These methods use multi-layer perceptrons (MLP) or hash encoding [29] to assign geometrical properties such as density [28], occupancy [30], or the signed distance to the nearest surface point [23, 47, 49] to spatial coordinates.

The application of signed distance functions (SDF) in neural surface reconstruction is particularly noteworthy. It involves the use of an SDF-induced density function, allowing for volume rendering to learn an implicit SDF representation. While current neural implicit-based methods using rendering supervision alone yield impressive results for simple scenes, they face challenges with larger scenes, especially those containing extensive textureless areas [9, 57, 60]. Previous works have attempted to address this by integrating structural priors into the optimization process, such as depth prior [31], normal regularization [57], point clouds regularization [9, 60], surface smoothness [32, 62], or semantic regularization [12, 16]. While these methods succeed in generating complete surface reconstructions, they often result in over-smoothed surfaces lacking in fine details. Particularly, works such as those by [9, 60] leverage sparse or dense point clouds from MVS to improve the fidelity of surface reconstruction. Nevertheless, point clouds predicted from MVS are uniformly distributed and sometimes incomplete and hence fails to provide geometry prior for some regions, especially regions with in-

tricate details. Moreover, the generated points inevitably suffer from noisy geometries that result in unreliable priors.

The advantage of the 3D Gaussian Splatting [18] is that it can generate dense point clouds with detailed geometry, which provides sufficient geometric constraints especially for regions with intricate details. However, point cloud generated from the vanilla 3D Gaussian Splatting cannot be directly used as a prior since they are computed as the center of 3D Gaussians that are generally located inside the surface. To this end, we introduce a scale regularizer that enforces the points to be close to the surface. Specifically, we enforce the smallest scaling factor of each 3D Gaussian ellipsoid to be close to zero such that the 3D Gaussian ellipsoid are flattened to a plane. The extremely thin 3D Gaussian are then inherently moved to the surface in order to render the correct color. Similar to existing point cloud-based approaches [9, 60], point clouds generated from 3D Gaussian Splatting are noisy and therefore may provide incorrect prior. To mitigate this problem, we further propose to refine the 3D Gaussians with the normal priors from the surface predicted by a neural implicit model, NeuS [47]. Specifically, the direction of the smallest scaling factor can be regarded as the normal direction since the 3D Gaussian ellipsoid is close to a plane. We regularize the normal direction by aligning it with the normal predicted by NeuS. With this joint optimization, the point cloud generated from 3D Gaussians are refined to provide more reliable prior. As a consequence, our approach is able to benefit from both representations and generate a complete surface with intricate details. An illustration of our proposed mutual optimization is shown in Figure 1.

In summary, our contributions include:

- We propose a novel framework that jointly optimizes NeuS and 3D Gaussian Splatting. This approach uses point clouds generated from 3D Gaussian Splatting to regulate NeuS, and concurrently uses the predicted normals from NeuS to refine 3D Gaussian Splatting for higher quality point clouds.
- We introduce two regularizers to ensure that point clouds from extremely thin 3D Gaussians closely adhere to the surfaces. These include regularizing the smallest scaling of each 3D Gaussian close to zero and aligning the normals of these Gaussians perpendicular to the surfaces.
- We empirically demonstrate the effectiveness of our approach, dubbed NeuSG, highlighting its significant improvements over previous methods in surface reconstruction by experiments.

2. Related Work

2.1. Multi-view surface reconstruction

Surface reconstruction from multiple views is a fundamental aspect of 3D reconstruction within the field of com-

puter vision. Traditional multi-view stereo (MVS) techniques [3, 4, 7, 21, 38–40] utilize either feature matching for depth calculation [3, 38] or voxel-based shape representation [4, 7, 21, 39, 45]. The depth-based approaches focus on creating depth maps that are later combined into a comprehensive point cloud. In contrast, volumetric reconstruction methods estimate space occupancy and color within a voxel grid [4, 7, 25], using color consistency checks to sidestep the need for explicit matching of correspondence. Nonetheless, the finite resolution of the voxel grids limits the precision achievable by these reconstruction methods. Modern learning-based MVS approaches often modify traditional processes, such as feature matching [27, 46, 58], depth integration [36], or depth inference from multiple images [15, 51, 52, 56, 59].

2.2. Neural Rendering and Radiance Fields

The pursuit of novel-view synthesis with volumetric representations began with the Soft3D [34], which was later advanced by integrating deep-learning with volumetric ray-marching to form a continuous, differentiable density field for geometry representation [14, 41]. Although volumetric ray-marching produces high-quality renders, its computational demand is substantial due to the extensive sampling required. The introduction of Neural Radiance Fields (NeRF) [28] with importance sampling and positional encoding marked an improvement in render quality, albeit at the expense of speed due to the sizable Multi-Layer Perceptron (MLP).

Subsequent methods have sought to optimize both quality and speed, often by implementing regularization strategies. To refine rendering quality, some works integrate position encoding or band-limited coordinate networks with neural radiance fields for a pre-filtered scene representation [1, 2, 24]. Different from these methods focusing on encodings and networks, S3IM [50] introduces a multiplex training approach to further improve rendering fidelity. Recent innovations aim to expedite training and rendering by leveraging spatial data structures for feature storage, alternative encodings, and adjusting MLP size [5, 8, 10, 13, 35, 44]. Noteworthy techniques include the usage of hash tables for more efficient encoding and smaller MLPs or even bypassing MLPs entirely [8, 29, 42].

Particularly, InstantNGP [29] employs a hash grid and occupancy grid for swift computation alongside a reduced MLP for depicting density and color, while Plenoxels [8] utilizes a sparse voxel grid to interpolate a continuous density field, allowing it to eliminate neural networks entirely. Both methods incorporate Spherical Harmonics, with the former applying them to represent directional effects and the latter using them to encode color network inputs. Despite their impressive outcomes, these methods may face challenges in accurately representing empty space, influ-

enced by the scene or capture conditions. Moreover, the structured grids chosen for computational acceleration can constrain image quality, and the intensive sampling required for ray-marching can slow down rendering. Addressing these limitations, the unstructured and explicit GPU-optimized 3D Gaussian Splatting [18] presents a solution that achieves enhanced rendering speeds and improved quality without the reliance on neural components. In our paper, we utilize 3D Gaussian Splatting to generate dense point clouds closing the surface to obtain complete and detailed surfaces.

2.3. Neural Surface Reconstruction

In neural representation of 3D surfaces, implicit functions such as occupancy grids [30, 32] and Signed Distance Functions (SDFs) [53] have been preferred over basic volume density fields. To integrate these with neural volume rendering techniques [28], reparameterization methods are developed to convert these representations back to volume density, as shown in previous research [47, 54]. These neural implicit functions have proven more effective in surface prediction and maintaining the quality of view synthesis [53].

Further research has aimed to adapt these methods for real-time applications, but compromising on surface accuracy as reported in recent studies [22, 49]. Simultaneously, additional research has considered extra data to enhance reconstruction quality [6, 9, 57]. For example, NeuralWarp [6] uses co-visibility cues from Structure-from-Motion (SfM) for patch warping, although this may not capture highly variable surfaces well. Other methods [9, 60] apply sparse SfM point clouds to supervise the SDF with their efficacy limited by the point cloud quality, as is the case with traditional techniques [60]. The use of depth and segmentation data has been explored in studies using unrestricted image collections or hash-encoded scene representations [43, 57, 63]. While these can build complete geometry surfaces, they often lack sharpness and detail. Alternatively, some research [48] has introduced a multi-stage optimization to refine surface details using a displacement network to correct shapes from a primary network. At the same time, Neuralangelo [23] employs hash encodings [29] and examines higher-order derivatives to reconstruct surfaces without auxiliary inputs. However, these approaches can result in imperfections such as holes in less textured or observed areas. In comparison, our method utilizes point clouds from Gaussian Splatting to achieve completeness and detail in geometry.

3. Our Method

Our proposed NeuSG efficiently reconstructs complete and detailed surfaces of scenes from multi-view images. Section 3.1 provides an overview of NeuS [47] and 3D Gaussian Splatting [18]. Our proposed scale and normal regu-

larization for 3D Gaussians is detailed in Section 3.2. The joint optimization for both NeuS and 3D Gaussian splatting is described in Section 3.3. Figure 2 shows the illustration of our whole framework.

3.1. Preliminary

3.1.1 Neural Implicit Surfaces by Volume Rendering

NeRF [28] captures a 3D scene using density and color fields for synthesizing novel views through volume rendering. The method faces challenges in defining clear surfaces, often resulting in noisy and unrealistic surfaces when derived from density [47, 54]. Signed Distance Functions (SDFs) offer a common method for representing surfaces implicitly as a zero-level set, $\{\mathbf{x} \in \mathbb{R}^3 \mid f(\mathbf{x}) = 0\}$, where $f(\mathbf{x})$ is the SDF value from an MLP $f(\cdot)$. NeuS [47] replaces the volume density output of NeRF with SDF. A non-learnable but differentiable logistic function is then designed to convert SDF into opacity for volume rendering. Consequently, the intermediate SDF output allows for Eikonal regularization [11] to improve the quality of surface reconstruction. Formally, the opacity α_i for a 3D point \mathbf{x}_i with SDF value $f(\mathbf{x}_i)$ is given by:

$$\alpha_i = \max \left(\frac{\Phi_s(f(\mathbf{x}_i)) - \Phi_s(f(\mathbf{x}_{i+1}))}{\Phi_s(f(\mathbf{x}_i))}, 0 \right), \quad (1)$$

where Φ_s denotes the sigmoid function. Given a posed camera located at \mathbf{o} and a ray direction \mathbf{d} , this opacity is then used to integrate the color radiance along a ray in the volume rendering process. The color of each point \mathbf{c}_i is predicted by an MLP, and the color of a pixel is the Riemann sum of these values:

$$\hat{\mathbf{C}}(\mathbf{o}, \mathbf{d}) = \sum_{i=1}^N w_i \mathbf{c}_i, \text{ where } w_i = T_i \alpha_i, \quad (2)$$

where α_i is the opacity defined in Eq. 1, and T_i is the cumulative transmittance which indicates the fraction of light that reaches the camera. N is the number of sample points along the ray.

3.1.2 3D Gaussian Splatting

As described in [18], 3D Gaussian Splatting is a method for representing 3D scenes with 3D Gaussians. Each Gaussian is defined by a covariance matrix Σ , and a center point \mathbf{p} which is the mean of the Gaussian. The 3D Gaussian distribution can be represented as:

$$G(\mathbf{x}) = \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{p})^\top \Sigma^{-1} (\mathbf{x} - \mathbf{p}) \right\}. \quad (3)$$

For optimization purposes, the covariance matrix Σ is expressed as the product of a scaling matrix \mathbf{S} and a rotation matrix \mathbf{R} :

$$\Sigma = \mathbf{R} \mathbf{S} \mathbf{S}^\top \mathbf{R}^\top. \quad (4)$$

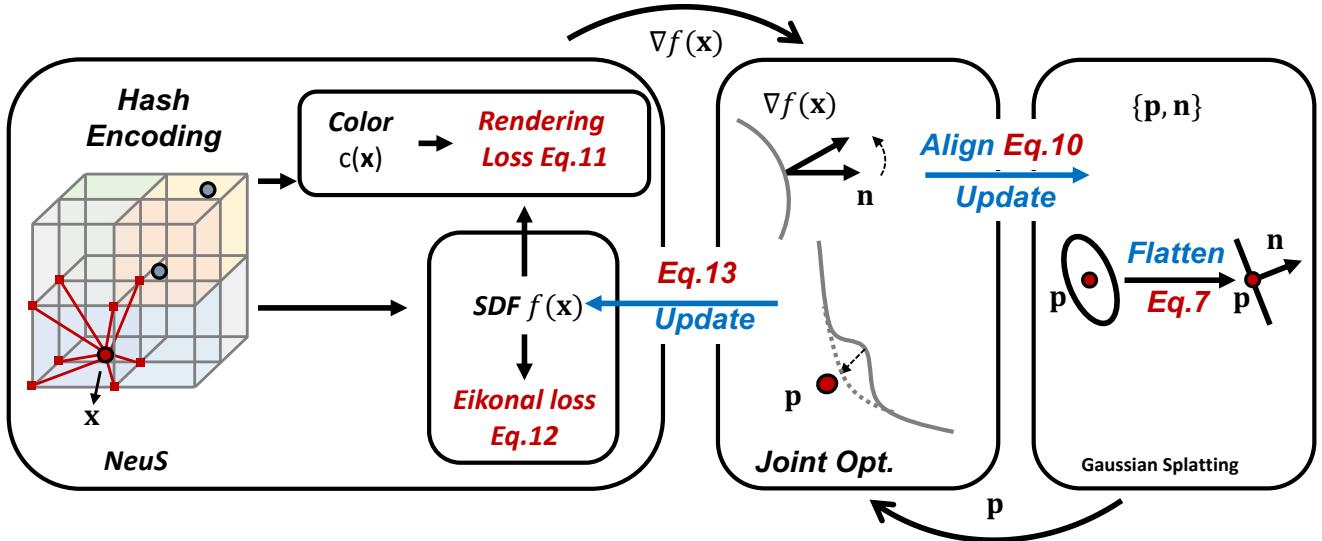


Figure 2. The NeuSG framework includes three principal components: 1) Optimization of neural implicit surface reconstruction. 2) Geometric constraints from point clouds generated from Gaussian Splatting, as formalized in Eq. 13. 3) Refinement of Gaussian Splatting through normal alignment, as detailed in Eq. 10.

\mathbf{S} is a diagonal matrix, stored by a scaling factor s . The rotation matrix \mathbf{R} is represented by a quaternion $\mathbf{r} \in \mathbb{R}^4$.

For novel view rendering, the splatting technique [55] is applied to the Gaussians on camera planes. Using the viewing transform \mathbf{W} and the Jacobian of the affine approximation of the projective transformation \mathbf{J} [64], the transformed covariance matrix Σ' can be determined as:

$$\Sigma' = \mathbf{J}\mathbf{W}\Sigma\mathbf{W}^\top\mathbf{J}^\top. \quad (5)$$

A complete 3D Gaussian is given by its position $\mathbf{p} \in \mathbb{R}^3$, color represented with spherical harmonics coefficients $\mathbf{H} \in \mathbb{R}^k$, opacity $\alpha \in \mathbb{R}$, quaternion $\mathbf{r} \in \mathbb{R}^4$, and scaling factor $s \in \mathbb{R}^3$. For a given pixel, the combined color and opacity from multiple Gaussians are weighted by Eq. 3. The color blending for overlapping points is:

$$\mathbf{C} = \sum_{i \in N} \mathbf{c}_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j), \quad (6)$$

where \mathbf{c}_i, α_i denote the color and density of a point.

3.2. Regularization for 3D Gaussians

We propose the utilization of 3D Gaussian Splatting to recover highly-detailed surfaces. This method is beneficial for creating dense point clouds that exhibit intricate details. However, the 3D Gaussian Splatting is not directly applicable in our context as the centers of the 3D Gaussians may not align with the actual surface. To circumvent this problem, we introduce a scale regularization as described

in Section 3.2.1. Furthermore, given that the points generated from 3D Gaussian Splatting are noisy, we propose to refine the 3D Gaussians using normal priors predicted by NeuS as discussed in Section 3.2.2. The process is shown in Figure 3.

3.2.1 Scale Regularization

As previously noted, conventional 3D Gaussian Splatting generates Gaussian centers $\{\mathbf{p}_i\}$ that are typically positioned inside the surface and thus making them unsuitable for direct surface Regularization. To remedy this, we refine the 3D Gaussian ellipsoids into highly flat shapes. This process encourages the Gaussians to become significantly narrow, thereby pulling their centers closer to the surface. Within the 3D Gaussians, the scaling factor s defines the dimensions of the ellipsoid in each direction. By manipulating the scaling vector, we can alter the shape of the Gaussian. Specifically, we minimize the smallest component of the scaling factor $\mathbf{s} = (s_1, s_2, s_3)^\top \in \mathbb{R}^3$ for each Gaussian towards zero:

$$\mathcal{L}_s = \|\min(s_1, s_2, s_3)\|_1. \quad (7)$$

This process effectively flattens the 3D Gaussian and hence encourages the center points $\{\mathbf{p}_i\}$ to align with the surface.

3.2.2 Normal Regularization

As the Gaussian flattens, the direction of the minimized scaling factor becomes the normal of the thin Gaussian. We

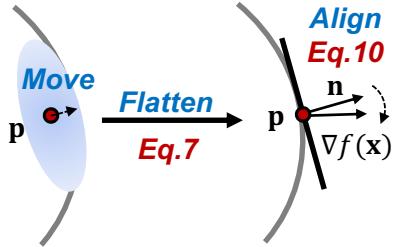


Figure 3. An illustration of 3D Gaussian flattening and normal alignment.

define the normal in the camera coordinate system as:

$$\mathbf{n}_c = \text{OneHot}(\arg \min(s_1, s_2, s_3)) \in \mathbb{R}^3, \quad (8)$$

where $\text{OneHot}(\cdot)$ converts an index to a vector. The normal \mathbf{n}_c is then transformed to the world coordinate system using rotation \mathbf{R} :

$$\mathbf{n}_w = \mathbf{R} \times \mathbf{n}_c. \quad (9)$$

We then align \mathbf{n}_w with the surface normal predicted by NeuS to fine-tune the Gaussians:

$$\mathcal{L}_{\text{align}} = |1 - |\mathbf{n}_w^\top \cdot \nabla f(\mathbf{p}_i)||_1. \quad (10)$$

The absolute value ensures correct orientation regardless of the vector direction. The point cloud derived from 3D Gaussians is refined through the alignment of normals, thereby offering a more dependable prior for surface optimization.

3.3. Joint Optimization

Initially, NeuS is optimized through a composite loss function of rendering loss, Eikonal loss and the constraints imposed by point clouds derived from 3D Gaussians as described in Section 3.3.1. Subsequently, the normals as predicted by NeuS are utilized to further refine the 3D Gaussians, which is detailed in Section 3.3.2.

3.3.1 Implicit Surface Reconstruction

The training of the network is guided by a color loss that measures the discrepancy between the input images \mathbf{C} and the rendered output $\hat{\mathbf{C}}$:

$$\mathcal{L}_{\text{RGB}} = \|\mathbf{C} - \hat{\mathbf{C}}\|_1. \quad (11)$$

To further refine the surface details, an Eikonal regularization term [11] enforces the correct gradient norm of the Signed Distance Function (SDF) in three-dimensional space:

$$\mathcal{L}_{\text{eik}} = \frac{1}{N} \sum_{i=1}^N (\|\nabla f(\mathbf{x}_i)\|_2 - 1)^2, \quad (12)$$

where $\nabla f(\mathbf{x}_i)$ is the gradient of the SDF $f(\mathbf{x}_i)$ evaluated at point \mathbf{x}_i . This is consistent with the NeuS framework for SDF-based volume rendering [47].

The point clouds \mathbf{p}_i contribute to the surface optimization by enforcing the SDF values to be near zero at these points:

$$\mathcal{L}_{\text{pt}} = |f(\mathbf{p}_i)|_1, \quad (13)$$

where $f(\mathbf{p}_i)$ is the predicted SDF value at the point \mathbf{p}_i . The overall loss function combining these elements is:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{RGB}} + \lambda_1 \mathcal{L}_{\text{eik}} + \lambda_2 \mathcal{L}_{\text{pt}}, \quad (14)$$

with λ_1 and λ_2 balancing the individual components.

3.3.2 Gaussian Splatting Refinement

In addition to the color matching loss \mathcal{L}_{RGB} , our approach involves minimizing the scaling factor through \mathcal{L}_s as detailed in Eq. 7, and aligning the normals via $\mathcal{L}_{\text{align}}$ as per Eq. 10. Hence, the combined loss function for Gaussian splatting is:

$$\mathcal{L}_{\text{Gaussian}} = \mathcal{L}_{\text{RGB}} + \lambda_3 \mathcal{L}_s + \lambda_4 \mathcal{L}_{\text{align}}, \quad (15)$$

where λ_3 and λ_4 weight the importance of the scaling and alignment terms, respectively. With this mutual optimization, the point cloud generated from 3D Gaussians will be refined, which in turn provides more reliable prior.

4. Experiment

We evaluate our method on the tasks of 3D surface reconstruction in Section 4.1. Additionally, we validate the effectiveness of the proposed methods in Section 4.2.

Implementation Details. Our approach is implemented in PyTorch [33], utilizing the Adam optimizer [19] with a learning rate of 1×10^{-3} for the task of neural surface reconstruction. The weighting parameters λ_1 , λ_2 , λ_3 , and λ_4 are set to 0.1, 1, 100, and 1, correspondingly. During training, we sample 1,024 rays at each iteration. Our method incorporates hash encoding, a multi-scale optimization strategy, and the numerical gradient techniques as proposed by NeuralAngelo [23]. To enhance efficiency, we utilize 2^{19} hash entries per resolution, which is less than the 2^{22} used in NeuralAngelo, thus reducing both training time and memory usage while maintaining comparable results. The neural surface reconstruction model undergoes optimization over 500k iterations, interspersing the optimization of Gaussian Splatting every 100k iterations, totaling 30k iterations for the latter. In addition to point clouds derived from Gaussian Splatting, we integrate point clouds from Vis-MVSNet [59] for complement. Moreover, we adopt a dual-network architecture akin to NeRF++ [61], employing a NeRF model for

	Barn	Caterpillar	Courthouse	Ignatius	Meetingroom	Truck	Mean	GPU hours
NeuralWarp [6]	0.22	0.18	0.08	0.02	0.08	0.35	0.15	-
COLMAP [37]	0.55	0.01	0.11	0.22	0.19	0.19	0.21	-
Vis-MVSNet [59]	0.49	0.21	0.36	0.25	0.43	0.28	0.34	-
NeuS [47]	0.29	0.29	0.17	0.83	0.24	0.45	0.38	-
NeuS-NGP	0.46	0.32	0.08	0.81	0.08	0.44	0.37	16
Geo-Neus [9]	0.33	0.26	0.12	0.72	0.20	0.45	0.35	-
MonoSDF [57]	0.49	0.31	0.12	0.78	0.23	0.42	0.39	18
RegSDF [60]	-	0.22	-	-	-	0.46	-	-
NAngelo-19 [23]	0.61	0.34	0.13	0.82	0.22	0.45	0.43	15
NAngelo-22 [23]	0.70	0.36	0.28	0.89	0.32	0.48	0.50	128
NeuSG	0.73	0.37	0.22	0.83	0.35	0.46	0.49	16

Table 1. Quantitative assessment on the Tanks and Temples dataset [20] indicates that our

NeuSG achieves the highest quality in surface reconstruction. **Teal** and **Cyan** indicate best and second best results, respectively. Notably, NAngelo-19, representing NeuralAngelo configured with a maximum of 2^{19} hash entries per resolution (same setting as our method), does not match NeuSG performance. NAngelo-22 which extends the capacity of NeuralAngelo to 2^{22} hash entries per resolution, necessitates tremendous increase in training duration but achieves results that are only comparable to our NeuSG. This highlights the efficiency of NeuSG, which attains similar or superior outcomes even with a less resource-intensive setup.

external scene components and a NeuS model for internal geometry. The NeuSG is trained for about 16 hours on a single RTX4090 GPU with 24GB memory.

Datasets. Our experiments mainly focuses on large and complex scenes. We primarily evaluate our methodology on six scenes from the Tanks and Temples dataset [20], encompassing a range of large-scale indoor and outdoor environments. These scenes comprise between 263 to 1,107 images captured with a handheld monocular RGB camera. Ground truth data for evaluation is acquired via LiDAR sensors. Additional results are presented in the supplementary materials.

Evaluation Criteria. For the assessment of surface reconstruction, we report on the Chamfer distance and the F1 score [20]. Image synthesis quality is quantified using the Peak Signal-to-Noise Ratio (PSNR).

4.1. Comparisons to Baselines

We compare with state-of-the-art neural surface reconstruction methods, including NeuS [47], NeuralWarp [6], Geo-Neus [9], MonoSDF [57], RegSDF [60], and NeuralAngelo [23]. In addition, we have trained a variant of NeuS with hashing encoding, referred to as NeuS-NGP. For a fair comparison, all methods utilizing a hash table, such as NeuS-NGP, MonoSDF, NAngelo-19 and our NeuSG are

constrained to a maximum of 2^{19} hash entries per resolution. NAngelo-19 denotes our own training of NeuralAngelo with the official code but limited to 2^{19} hash entries, whereas NAngelo-22 refers to the original implementation with 2^{22} hash entries as reported in their paper. The results for classical multi-view stereo including COLMAP [37] and Vis-MVSNet [59] are also presented for comprehensive evaluation. Qualitative comparisons are depicted in Figure 4 and quantitative results are summarized in Table 1.

We can see that NeuSG outperforms other approaches in terms of the F1 score. When compared to NeuS [47], our method excels in reconstructing surfaces that are not only complete but also exhibit high-fidelity and intricate details (0.49 v.s. 0.38). As depicted in Figure 4, NeuSG shows its ability to recover full surfaces compared with NAngelo-19. Even when compared to NAngelo-22 [23] which utilizes a larger number of hash entries, our NeuSG still attains comparable performance (0.49 vs. 0.5) with significantly reduced computational time (16 hours vs. 128 hours). Furthermore, our method surpasses the geometric constraint-based approaches [9, 57, 60] since they tend to produce over-smoothed results. This is particularly evident in Figure 4, where MonoSDF produce overly smooth surfaces despite using geometric priors, whereas NeuSG preserves fine surface details. Compared with methods that also make use of point clouds for regularization [9, 60], our method achieves better performance (0.49 v.s. 0.35) and

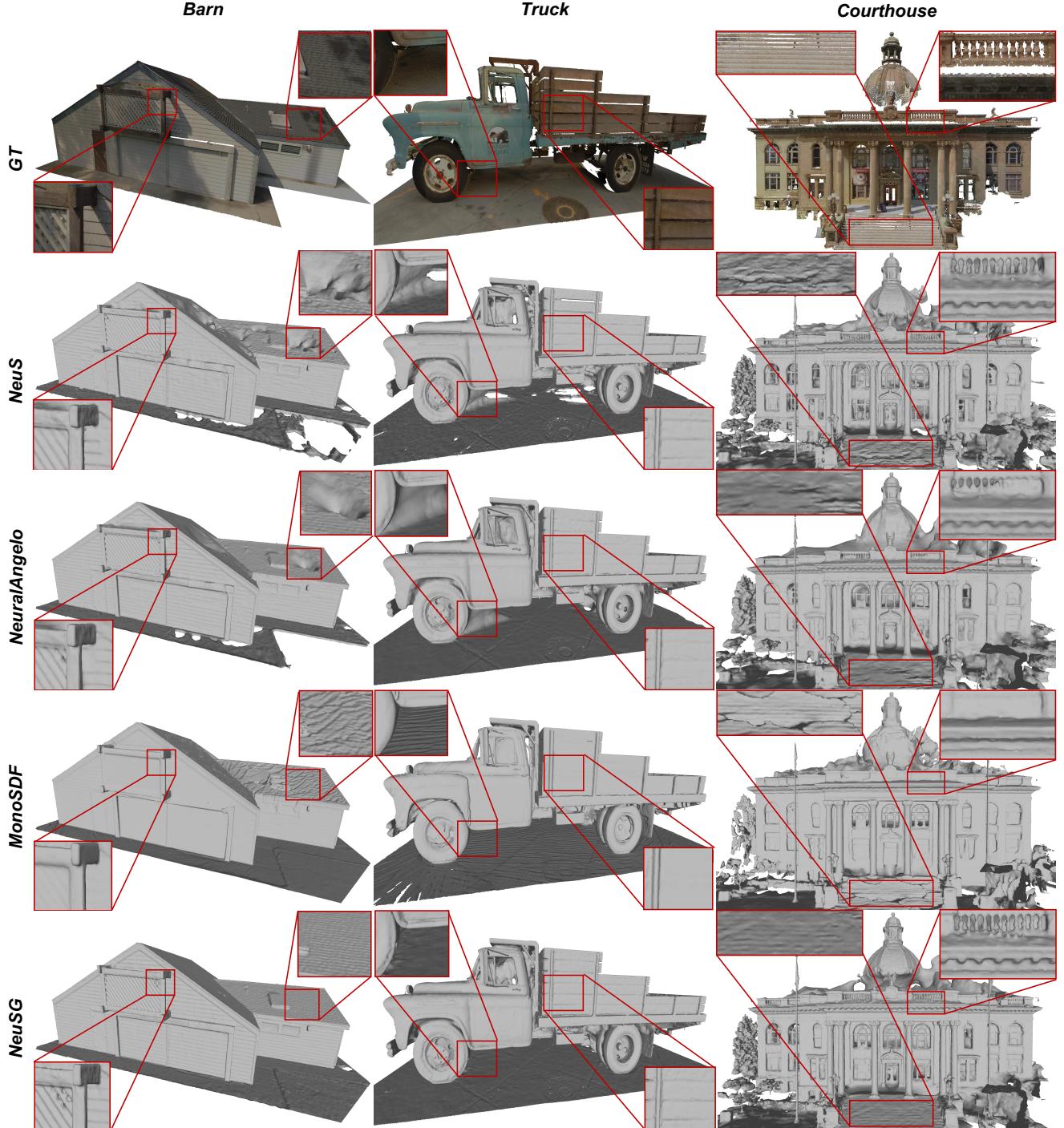


Figure 4. Qualitative comparison on Tanks and Temples dataset [20]. NeuSG excels in achieving both complete and intricately detailed surfaces, in contrast to baseline approaches which often result in surfaces that are either incomplete or marred by noise.

thus demonstrates the effectiveness of joint optimization of NeuS and Gaussian Splatting.

As depicted in Figure 4, our approach successfully achieves complete and detailed surface reconstruction. For

example, our method reconstructs the entire roof structure without any gaps in the ‘Barn’ scene. This is a feat not matched by NeuS or NeuralAngelo, which tend to leave holes in the roof. Although MonoSDF is capable of recon-

structing a complete roof, it falls short in preserving finer details such as the stripe patterns of the wall. Furthermore, our method avoids creating extraneous surface elements. A notable example is the ‘Truck’ scene where our method accurately reconstructs a standalone tire, in contrast to NeuS and NeuralAngelo which incorrectly merge the tire with the ground. Additionally, our method distinctively captures the intricate details of the stairway and handrail in the ‘Courthouse’ scene. This is a level of detail not replicated by the other methods.

4.2. Ablations

4.2.1 Scale and Normal Regularization

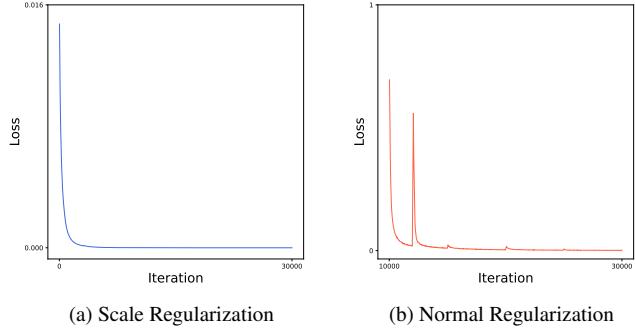


Figure 5. The scale regularizations for Gaussian Splatting. **Left:** Scale regularization. **Right:** Normal regularization.

To validate the transformation of 3D Gaussians into extremely thin structures, we present the scale and normal regularization losses in Figure 5. Figure 5a demonstrates that the scale regularization loss approaches zero (approximately 1×10^{-8}), confirming the successful flattening of 3D Gaussians into extremely thin forms. Figure 5b illustrates that the normal regularization loss also nears zero (about 0.001), indicative of the accurate learning of normals.

4.2.2 Point Clouds Regularization

MVS	Gaussian			F1 Score
	Original	Scale	Normal	
✗	✗	✗	✗	0.61 (+0)
✓	✗	✗	✗	0.63 (+0.02)
✓	✓	✗	✗	0.59 (-0.02)
✗	✗	✓	✗	0.69 (+0.08)
✗	✗	✓	✓	0.73 (+0.12)

Table 2. Ablations of different components.

As depicted in Table 2, we conduct an ablation study

to assess the impact of point cloud regularization in our method. The study commences with a baseline variant without point cloud regularization. Subsequently, we incorporate point clouds from Vis-MVSNet [59] to aid in the optimization process of the network. Following this, point clouds generated via Gaussian Splatting [18] are employed to guide the learning of the Signed Distance Function (SDF). Our findings indicate that while the utilization of point clouds from Vis-MVSNet offers marginal improvements, the integration of point clouds from our proposed Gaussian Splatting method significantly enhances the performance of the model.

We also verify the effectiveness of scale and normal regularization in quality. We can see from the table that without any regularization, only point clouds from original Gaussian Splatting create a negative impact. By adding the scale and normal regularization, the performance is further improved to the best.

5. Limitation

Similar to numerous neural implicit reconstruction methods, our approach is constrained by the number of images available. Dense observations from multiple viewpoints are essential for accurate reconstruction of the object. This reliance on extensive multi-view data can be a significant limitation in scenarios where such comprehensive coverage is not feasible. Furthermore, our method, while effective in detailed surface reconstruction, may struggle in environments with sparse or uneven image distribution, potentially leading to less accurate reconstructions in these areas.

6. Conclusion

In this work, we have introduced a neural implicit surface reconstruction pipeline enhanced by the incorporation of 3D Gaussian Splatting to facilitate the recovery of highly detailed and complete surfaces. The key advantage of 3D Gaussian Splatting lies in its capability to produce dense point clouds with detailed structures. We propose a scale regularization strategy that effectively transforms the 3D Gaussians into exceedingly thin shapes, thereby drawing the points closer to the surface. Furthermore, our method diverges from the conventional approach of using a static set of points as priors. Instead, we refine the 3D Gaussians using normal priors derived from surfaces predicted by neural implicit models. This approach endows 3D Gaussian Splatting with more precise directional guidance, leading to enhanced surface reconstruction. By jointly optimizing both the 3D Gaussian Splatting and the neural implicit model, our proposed method leverages the strengths of each technique, enabling the generation of complete surfaces with a high level of detail.

References

- [1] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5855–5864, 2021. [2](#)
- [2] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479, 2022. [2](#)
- [3] Michael Bleyer, Christoph Rhemann, and Carsten Rother. Patchmatch stereo-stereo matching with slanted support windows. In *Bmvc*, pages 1–11, 2011. [2](#)
- [4] Adrian Broadhurst, Tom W Drummond, and Roberto Cipolla. A probabilistic framework for space carving. In *Proceedings eighth IEEE international conference on computer vision. ICCV 2001*, pages 388–393. IEEE, 2001. [1, 2](#)
- [5] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *European Conference on Computer Vision*, pages 333–350. Springer, 2022. [2](#)
- [6] François Darmon, Bénédicte Bascle, Jean-Clément Devaux, Pascal Monasse, and Mathieu Aubry. Improving neural implicit surfaces geometry with patch warping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6260–6269, 2022. [3, 6](#)
- [7] Jeremy S De Bonet and Paul Viola. Voxels: Probabilistic voxelized volume reconstruction. In *Proceedings of International Conference on Computer Vision (ICCV)*, page 2. Citeseer, 1999. [1, 2](#)
- [8] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5501–5510, 2022. [2](#)
- [9] Qiancheng Fu, Qingshan Xu, Yew Soon Ong, and Wenbing Tao. Geo-neus: Geometry-consistent neural implicit surfaces learning for multi-view reconstruction. *Advances in Neural Information Processing Systems*, 35:3403–3416, 2022. [1, 2, 3, 6](#)
- [10] Stephan J Garbin, Marek Kowalski, Matthew Johnson, Jamie Shotton, and Julien Valentin. Fastnerf: High-fidelity neural rendering at 200fps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14346–14355, 2021. [2](#)
- [11] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. *arXiv preprint arXiv:2002.10099*, 2020. [3, 5](#)
- [12] Haoyu Guo, Sida Peng, Haotong Lin, Qianqian Wang, Guofeng Zhang, Hujun Bao, and Xiaowei Zhou. Neural 3d scene reconstruction with the manhattan-world assumption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5511–5520, 2022. [1](#)
- [13] Peter Hedman, Pratul P Srinivasan, Ben Mildenhall, Jonathan T Barron, and Paul Debevec. Baking neural radiance fields for real-time view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5875–5884, 2021. [2](#)
- [14] Philipp Henzler, Niloy J Mitra, and Tobias Ritschel. Escaping plato’s cave: 3d shape from adversarial rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9984–9993, 2019. [2](#)
- [15] Po-Han Huang, Kevin Matzen, Johannes Kopf, Narendra Ahuja, and Jia-Bin Huang. Deepmvs: Learning multi-view stereopsis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2821–2830, 2018. [2](#)
- [16] Ajay Jain, Matthew Tancik, and Pieter Abbeel. Putting nerf on a diet: Semantically consistent few-shot view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5885–5894, 2021. [1](#)
- [17] Michael Kazhdan and Hugues Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)*, 32(3):1–13, 2013. [1](#)
- [18] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023. [2, 3, 8](#)
- [19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. [5](#)
- [20] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (ToG)*, 36(4):1–13, 2017. [6, 7](#)
- [21] Kiriakos N Kutulakos and Steven M Seitz. A theory of shape by space carving. *International journal of computer vision*, 38:199–218, 2000. [1, 2](#)
- [22] Hai Li, Xingrui Yang, Hongjia Zhai, Yuqian Liu, Hujun Bao, and Guofeng Zhang. Vox-surf: Voxel-based implicit surface representation. *IEEE Transactions on Visualization and Computer Graphics*, 2022. [3](#)
- [23] Zhaoshuo Li, Thomas Müller, Alex Evans, Russell H Taylor, Mathias Unberath, Ming-Yu Liu, and Chen-Hsuan Lin. Neuralangelo: High-fidelity neural surface reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. [1, 3, 5, 6](#)
- [24] David B Lindell, Dave Van Veen, Jeong Joon Park, and Gordon Wetzstein. Bacon: Band-limited coordinate networks for multiscale scene representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16252–16262, 2022. [2](#)
- [25] Shaohui Liu, Yinda Zhang, Songyou Peng, Boxin Shi, Marc Pollefeys, and Zhaopeng Cui. Dist: Rendering deep implicit signed distance function with differentiable sphere tracing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2019–2028, 2020. [2](#)
- [26] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Seminal graphics: pioneering efforts that shaped the field*, pages 347–353. 1998. [1](#)

- [27] Wenjie Luo, Alexander G Schwing, and Raquel Urtasun. Efficient deep learning for stereo matching. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5695–5703, 2016. [2](#)
- [28] Ben Mildenhall, Pratul Srinivasan, Matthew Tancik, Jonathan Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision*, pages 405–421. Springer, 2020. [1, 2, 3](#)
- [29] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4):1–15, 2022. [1, 2, 3](#)
- [30] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3504–3515, 2020. [1, 3](#)
- [31] Michael Niemeyer, Jonathan T Barron, Ben Mildenhall, Mehdi SM Sajjadi, Andreas Geiger, and Noha Radwan. Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5480–5490, 2022. [1](#)
- [32] Michael Oechsle, Songyou Peng, and Andreas Geiger. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5589–5599, 2021. [1, 3](#)
- [33] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017. [5](#)
- [34] Eric Penner and Li Zhang. Soft 3d reconstruction for view synthesis. *ACM Transactions on Graphics (TOG)*, 36(6):1–11, 2017. [2](#)
- [35] Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14335–14345, 2021. [2](#)
- [36] Gernot Riegler, Ali Osman Ulusoy, Horst Bischof, and Andreas Geiger. Octnetfusion: Learning depth fusion from data. In *2017 International Conference on 3D Vision (3DV)*, pages 57–66. IEEE, 2017. [2](#)
- [37] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. [6](#)
- [38] Johannes L Schonberger, Enliang Zheng, Jan-Michael Frahm, and Marc Pollefeys. Pixelwise view selection for unstructured multi-view stereo. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part III 14*, pages 501–518. Springer, 2016. [2](#)
- [39] Steven M Seitz and Charles R Dyer. Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision*, 35:151–173, 1999. [1, 2](#)
- [40] Steven M Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR’06)*, pages 519–528. IEEE, 2006. [1, 2](#)
- [41] Vincent Sitzmann, Justus Thies, Felix Heide, Matthias Nießner, Gordon Wetzstein, and Michael Zollhofer. Deepvoxels: Learning persistent 3d feature embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2437–2446, 2019. [2](#)
- [42] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5459–5469, 2022. [2](#)
- [43] Jiaming Sun, Xi Chen, Qianqian Wang, Zhengqi Li, Hadar Averbuch-Elor, Xiaowei Zhou, and Noah Snavely. Neural 3d reconstruction in the wild. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–9, 2022. [3](#)
- [44] Towaki Takikawa, Joey Litalien, Kangxue Yin, Karsten Kreis, Charles Loop, Derek Nowrouzezahrai, Alec Jacobson, Morgan McGuire, and Sanja Fidler. Neural geometric level of detail: Real-time rendering with implicit 3d shapes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11358–11367, 2021. [2](#)
- [45] Shubham Tulsiani, Tinghui Zhou, Alexei A Efros, and Jitendra Malik. Multi-view supervision for single-view reconstruction via differentiable ray consistency. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2626–2634, 2017. [2](#)
- [46] Benjamin Ummenhofer, Huizhong Zhou, Jonas Uhrig, Niklaus Mayer, Eddy Ilg, Alexey Dosovitskiy, and Thomas Brox. Demon: Depth and motion network for learning monocular stereo. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5038–5047, 2017. [2](#)
- [47] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*, 2021. [1, 2, 3, 5, 6](#)
- [48] Yiqun Wang, Ivan Skorokhodov, and Peter Wonka. Hf-neus: Improved surface reconstruction using high-frequency details. *Advances in Neural Information Processing Systems*, 35:1966–1978, 2022. [3](#)
- [49] Yiming Wang, Qin Han, Marc Habermann, Kostas Daniilidis, Christian Theobalt, and Lingjie Liu. Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3295–3306, 2023. [1, 3](#)
- [50] Zeke Xie, Xindi Yang, Yujie Yang, Qi Sun, Yixiang Jiang, Haoran Wang, Yunfeng Cai, and Mingming Sun. S3im: Stochastic structural similarity and its unreasonable effectiveness for neural fields. In *International Conference on Computer Vision*, 2023. [2](#)

- [51] Yao Yao, Zixin Luo, Shiwei Li, Tian Fang, and Long Quan. Mvsnet: Depth inference for unstructured multi-view stereo. In *Proceedings of the European conference on computer vision (ECCV)*, pages 767–783, 2018. 2
- [52] Yao Yao, Zixin Luo, Shiwei Li, Tianwei Shen, Tian Fang, and Long Quan. Recurrent mvsnet for high-resolution multi-view stereo depth inference. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5525–5534, 2019. 2
- [53] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. *Advances in Neural Information Processing Systems*, 33:2492–2502, 2020. 3
- [54] Lior Yariv, Jitao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems*, 34:4805–4815, 2021. 3
- [55] Wang Yifan, Felice Serena, Shihao Wu, Cengiz Öztireli, and Olga Sorkine-Hornung. Differentiable surface splatting for point-based geometry processing. *ACM Transactions on Graphics (TOG)*, 38(6):1–14, 2019. 4
- [56] Zehao Yu and Shenghua Gao. Fast-mvsnet: Sparse-to-dense multi-view stereo with learned propagation and gauss-newton refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1949–1958, 2020. 2
- [57] Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction. *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. 1, 3, 6
- [58] Sergey Zagoruyko and Nikos Komodakis. Learning to compare image patches via convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4353–4361, 2015. 2
- [59] Jingyang Zhang, Yao Yao, Shiwei Li, Zixin Luo, and Tian Fang. Visibility-aware multi-view stereo network. *British Machine Vision Conference (BMVC)*, 2020. 2, 5, 6, 8
- [60] Jingyang Zhang, Yao Yao, Shiwei Li, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. Critical regularizations for neural surface reconstruction in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6270–6279, 2022. 1, 2, 3, 6
- [61] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020. 5
- [62] Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (ToG)*, 40(6):1–18, 2021. 1
- [63] Fuqiang Zhao, Yuheng Jiang, Kaixin Yao, Jiakai Zhang, Liao Wang, Haizhao Dai, Yuhui Zhong, Yingliang Zhang, Minye Wu, Lan Xu, et al. Human performance modeling and rendering via neural animated mesh. *ACM Transactions on Graphics (TOG)*, 41(6):1–17, 2022. 3
- [64] Matthias Zwicker, Hanspeter Pfister, Jeroen Van Baar, and Markus Gross. Surface splatting. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 371–378, 2001. 4

References

- [1] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5855–5864, 2021. 2
- [2] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479, 2022. 2
- [3] Michael Bleyer, Christoph Rhemann, and Carsten Rother. Patchmatch stereo-stereo matching with slanted support windows. In *Bmvc*, pages 1–11, 2011. 2
- [4] Adrian Broadhurst, Tom W Drummond, and Roberto Cipolla. A probabilistic framework for space carving. In *Proceedings eighth IEEE international conference on computer vision. ICCV 2001*, pages 388–393. IEEE, 2001. 1, 2
- [5] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorrf: Tensorial radiance fields. In *European Conference on Computer Vision*, pages 333–350. Springer, 2022. 2
- [6] François Darmon, Bénédicte Basclé, Jean-Clément Devaux, Pascal Monasse, and Mathieu Aubry. Improving neural implicit surfaces geometry with patch warping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6260–6269, 2022. 3, 6
- [7] Jeremy S De Bonet and Paul Viola. Poxels: Probabilistic voxelized volume reconstruction. In *Proceedings of International Conference on Computer Vision (ICCV)*, page 2. Citeseer, 1999. 1, 2
- [8] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxts: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5501–5510, 2022. 2
- [9] Qiancheng Fu, Qingshan Xu, Yew Soon Ong, and Wenbing Tao. Geo-neus: Geometry-consistent neural implicit surfaces learning for multi-view reconstruction. *Advances in Neural Information Processing Systems*, 35:3403–3416, 2022. 1, 2, 3, 6
- [10] Stephan J Garbin, Marek Kowalski, Matthew Johnson, Jamie Shotton, and Julien Valentin. Fastnerf: High-fidelity neural rendering at 200fps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14346–14355, 2021. 2
- [11] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. *arXiv preprint arXiv:2002.10099*, 2020. 3, 5

- [12] Haoyu Guo, Sida Peng, Haotong Lin, Qianqian Wang, Guofeng Zhang, Hujun Bao, and Xiaowei Zhou. Neural 3d scene reconstruction with the manhattan-world assumption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5511–5520, 2022. 1
- [13] Peter Hedman, Pratul P Srinivasan, Ben Mildenhall, Jonathan T Barron, and Paul Debevec. Baking neural radiance fields for real-time view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5875–5884, 2021. 2
- [14] Philipp Henzler, Niloy J Mitra, and Tobias Ritschel. Escaping plato’s cave: 3d shape from adversarial rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9984–9993, 2019. 2
- [15] Po-Han Huang, Kevin Matzen, Johannes Kopf, Narendra Ahuja, and Jia-Bin Huang. Deepmvs: Learning multi-view stereopsis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2821–2830, 2018. 2
- [16] Ajay Jain, Matthew Tancik, and Pieter Abbeel. Putting nerf on a diet: Semantically consistent few-shot view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5885–5894, 2021. 1
- [17] Michael Kazhdan and Hugues Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)*, 32(3):1–13, 2013. 1
- [18] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023. 2, 3, 8
- [19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5
- [20] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (ToG)*, 36(4):1–13, 2017. 6, 7
- [21] Kiriakos N Kutulakos and Steven M Seitz. A theory of shape by space carving. *International journal of computer vision*, 38:199–218, 2000. 1, 2
- [22] Hai Li, Xingrui Yang, Hongjia Zhai, Yuqian Liu, Hujun Bao, and Guofeng Zhang. Vox-surf: Voxel-based implicit surface representation. *IEEE Transactions on Visualization and Computer Graphics*, 2022. 3
- [23] Zhaoshuo Li, Thomas Müller, Alex Evans, Russell H Taylor, Mathias Unterath, Ming-Yu Liu, and Chen-Hsuan Lin. Neuralangelo: High-fidelity neural surface reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1, 3, 5, 6
- [24] David B Lindell, Dave Van Veen, Jeong Joon Park, and Gordon Wetzstein. Bacon: Band-limited coordinate networks for multiscale scene representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16252–16262, 2022. 2
- [25] Shaohui Liu, Yinda Zhang, Songyou Peng, Boxin Shi, Marc Pollefeys, and Zhaopeng Cui. Dist: Rendering deep implicit signed distance function with differentiable sphere tracing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2019–2028, 2020. 2
- [26] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Seminal graphics: pioneering efforts that shaped the field*, pages 347–353. 1998. 1
- [27] Wenjie Luo, Alexander G Schwing, and Raquel Urtasun. Efficient deep learning for stereo matching. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5695–5703, 2016. 2
- [28] Ben Mildenhall, Pratul Srinivasan, Matthew Tancik, Jonathan Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision*, pages 405–421. Springer, 2020. 1, 2, 3
- [29] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4):1–15, 2022. 1, 2, 3
- [30] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3504–3515, 2020. 1, 3
- [31] Michael Niemeyer, Jonathan T Barron, Ben Mildenhall, Mehdi SM Sajjadi, Andreas Geiger, and Noha Radwan. Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5480–5490, 2022. 1
- [32] Michael Oechsle, Songyou Peng, and Andreas Geiger. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5589–5599, 2021. 1, 3
- [33] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017. 5
- [34] Eric Penner and Li Zhang. Soft 3d reconstruction for view synthesis. *ACM Transactions on Graphics (TOG)*, 36(6):1–11, 2017. 2
- [35] Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14335–14345, 2021. 2
- [36] Gernot Riegler, Ali Osman Ulusoy, Horst Bischof, and Andreas Geiger. Octnetfusion: Learning depth fusion from data. In *2017 International Conference on 3D Vision (3DV)*, pages 57–66. IEEE, 2017. 2
- [37] Johannes L Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. 6
- [38] Johannes L Schönberger, Enliang Zheng, Jan-Michael Frahm, and Marc Pollefeys. Pixelwise view selection for

- unstructured multi-view stereo. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part III* 14, pages 501–518. Springer, 2016. 2
- [39] Steven M Seitz and Charles R Dyer. Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision*, 35:151–173, 1999. 1, 2
- [40] Steven M Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR’06)*, pages 519–528. IEEE, 2006. 1, 2
- [41] Vincent Sitzmann, Justus Thies, Felix Heide, Matthias Nießner, Gordon Wetzstein, and Michael Zollhofer. Deepvoxels: Learning persistent 3d feature embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2437–2446, 2019. 2
- [42] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5459–5469, 2022. 2
- [43] Jiaming Sun, Xi Chen, Qianqian Wang, Zhengqi Li, Hadar Averbuch-Elor, Xiaowei Zhou, and Noah Snavely. Neural 3d reconstruction in the wild. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–9, 2022. 3
- [44] Towaki Takikawa, Joey Litalien, Kangxue Yin, Karsten Kreis, Charles Loop, Derek Nowrouzezahrai, Alec Jacobson, Morgan McGuire, and Sanja Fidler. Neural geometric level of detail: Real-time rendering with implicit 3d shapes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11358–11367, 2021. 2
- [45] Shubham Tulsiani, Tinghui Zhou, Alexei A Efros, and Jitendra Malik. Multi-view supervision for single-view reconstruction via differentiable ray consistency. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2626–2634, 2017. 2
- [46] Benjamin Ummenhofer, Huizhong Zhou, Jonas Uhrig, Niklaus Mayer, Eddy Ilg, Alexey Dosovitskiy, and Thomas Brox. Demon: Depth and motion network for learning monocular stereo. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5038–5047, 2017. 2
- [47] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*, 2021. 1, 2, 3, 5, 6
- [48] Yiqun Wang, Ivan Skorokhodov, and Peter Wonka. Hf-neus: Improved surface reconstruction using high-frequency details. *Advances in Neural Information Processing Systems*, 35:1966–1978, 2022. 3
- [49] Yiming Wang, Qin Han, Marc Habermann, Kostas Daniilidis, Christian Theobalt, and Lingjie Liu. Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3295–3306, 2023. 1, 3
- [50] Zeke Xie, Xindi Yang, Yujie Yang, Qi Sun, Yixiang Jiang, Haoran Wang, Yunfeng Cai, and Mingming Sun. S3im: Stochastic structural similarity and its unreasonable effectiveness for neural fields. In *International Conference on Computer Vision*, 2023. 2
- [51] Yao Yao, Zixin Luo, Shiwei Li, Tian Fang, and Long Quan. Mvsnet: Depth inference for unstructured multi-view stereo. In *Proceedings of the European conference on computer vision (ECCV)*, pages 767–783, 2018. 2
- [52] Yao Yao, Zixin Luo, Shiwei Li, Tianwei Shen, Tian Fang, and Long Quan. Recurrent mvsnet for high-resolution multi-view stereo depth inference. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5525–5534, 2019. 2
- [53] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. *Advances in Neural Information Processing Systems*, 33:2492–2502, 2020. 3
- [54] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems*, 34:4805–4815, 2021. 3
- [55] Wang Yifan, Felice Serena, Shihao Wu, Cengiz Öztureli, and Olga Sorkine-Hornung. Differentiable surface splatting for point-based geometry processing. *ACM Transactions on Graphics (TOG)*, 38(6):1–14, 2019. 4
- [56] Zehao Yu and Shenghua Gao. Fast-mvsnet: Sparse-to-dense multi-view stereo with learned propagation and gauss-newton refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1949–1958, 2020. 2
- [57] Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction. *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. 1, 3, 6
- [58] Sergey Zagoruyko and Nikos Komodakis. Learning to compare image patches via convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4353–4361, 2015. 2
- [59] Jingyang Zhang, Yao Yao, Shiwei Li, Zixin Luo, and Tian Fang. Visibility-aware multi-view stereo network. *British Machine Vision Conference (BMVC)*, 2020. 2, 5, 6, 8
- [60] Jingyang Zhang, Yao Yao, Shiwei Li, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. Critical regularizations for neural surface reconstruction in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6270–6279, 2022. 1, 2, 3, 6
- [61] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020. 5
- [62] Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (ToG)*, 40(6):1–18, 2021. 1

- [63] Fuqiang Zhao, Yuheng Jiang, Kaixin Yao, Jiakai Zhang, Liao Wang, Haizhao Dai, Yuhui Zhong, Yingliang Zhang, Minye Wu, Lan Xu, et al. Human performance modeling and rendering via neural animated mesh. *ACM Transactions on Graphics (TOG)*, 41(6):1–17, 2022. 3
- [64] Matthias Zwicker, Hanspeter Pfister, Jeroen Van Baar, and Markus Gross. Surface splatting. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 371–378, 2001. 4