RadSplat: Radiance Field-Informed Gaussian Splatting for Robust Real-Time Rendering with 900+ FPS

Michael Niemeyer Fabian Manhardt Marie-Julie Rakotosaona
Michael Oechsle Daniel Duckworth Rama Gosula
Keisuke Tateno John Bates Dominik Kaeser Federico Tombari
Google
m-niemeyer.github.io/radsplat

Abstract

Recent advances in view synthesis and real-time rendering have achieved photorealistic quality at impressive rendering speeds. While radiance field-based methods achieve state-of-the-art quality in challenging scenarios such as in-the-wild captures and large-scale scenes, they often suffer from excessively high compute requirements linked to volumetric rendering. Gaussian Splatting-based methods, on the other hand, rely on rasterization and naturally achieve real-time rendering but suffer from brittle optimization heuristics that underperform on more challenging scenes. In this work, we present RadSplat, a lightweight method for robust real-time rendering of complex scenes. Our main contributions are threefold. First, we use radiance fields as a prior and supervision signal for optimizing point-based scene representations, leading to improved quality and more robust optimization. Next, we develop a novel pruning technique reducing the overall point count while maintaining high quality, leading to smaller and more compact scene representations with faster inference speeds. Finally, we propose a novel test-time filtering approach that further accelerates rendering and allows to scale to larger, house-sized scenes. We find that our method enables state-of-the-art synthesis of complex captures at 900+ FPS.

1. Introduction

Neural fields [5, 36, 47, 78] have emerged as one of the most popular representations for 3D vision due to their simple design, stable optimization, and state-of-the-art performance. After their introduction in the context of 3D reconstruction [5, 36, 47, 78], neural fields have been widely adopted and set new standards in tasks such as view synthesis [1, 2, 37], 3D and 4D reconstruction [30, 44, 48–50, 84], and generative modeling [32, 43, 52, 65, 70].

While neural field methods have achieved unprecedented view synthesis quality even for challenging real-world captures [2, 33, 37], most approaches are limited by the high compute costs of volumetric rendering. In order to achieve real-time frame rates, recent works reduce network complexity [16, 40], cache intermediate outputs [11, 60], or extract 3D meshes [46, 58, 83, 84]. Nevertheless, all methods trade reduced quality and increased storage costs for faster rendering, and are incapable of maintaining state-of-the-art quality in real-time – the goal of this work (see Fig. 1).

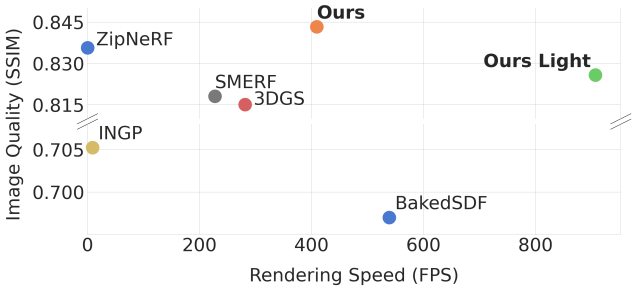

Figure 1. **RadSplat.** By combining benefits of neural fields and point-based representations, we achieve state-of-the-art quality in view synthesis on mip-NeRF 360 [1] while rendering at 900+ frames per second (FPS), indicating a speed up of 3.6× over 3D Gaussian Splatting (3DGS) [26] and 3,000× over Zip-NeRF [2].

Recently, rasterization-based 3D Gaussian Splatting (3DGS) [26] has emerged as a natural alternative to neural fields. The representation admits real-time frame rates with view synthesis quality rivaling the state-of-the-art in neural fields. 3DGS, however, suffers from a challenging optimization landscape and an unbounded model size. The number of Gaussian primitives is not known a priori, and carefully-tuned merging, splitting, and pruning heuristics are required to achieve satisfactory results. The brittleness of these heuristics become particularly evident in large scenes where phenomena such as exposure variation, motion blur, and moving objects are unavoidable (see Fig. 2). An increasing number of primitives further leads to

(a) 3DGS. (b) 3DGS with exposure module.

(c) RadSplat (Ours).

Figure 2. **Robust View Synthesis.** On complex scenes with lighting variations, 3D Gaussian Splatting (3DGS) [26] degrades (2a). When equipped with exposure handling modules [11, 26], results improve but they still contain artifacts and are overly smooth (2b). In contrast, we achieve high quality even for challenging, large-scale captures (2c) by integrating a robust radiance field as prior.

a potentially-unmanageable memory footprint and reduced rendering speed, strongly limiting model quality for larger scenes.

In this work, we present RadSplat, a lightweight method for robust real-time rendering of complex real-world scenes. Our method achieves smaller model sizes and faster rendering than 3DGS while strongly exceeding reconstruction quality. Our key idea is to combine the stable optimization and quality of neural fields to act as a prior and supervision signal for the optimization of point-based scene representations. We further introduce a novel pruning procedure and test-time visibility rendering strategies to significantly reduce memory usage and increase rendering speed without a corresponding loss in quality. In summary, our contributions are as follows:

1. The use of radiance fields as a prior and to handle the complexity of real-world data when optimizing point-based 3DGS representations.
2. A novel pruning strategy that reduces the number of Gaussian primitives by up to 10x whilst improving quality and rendering speed.
3. A novel post-processing step enabling viewpoint-based filtering, further accelerating rendering speed without any reduction in quality.

Our method exhibits state-of-the-art reconstruction quality on both medium and large scenes, with PSNR up to 1.87 dB higher than 3DGS and SSIM exceeding Zip-NeRF, the current state-of-the-art in offline view synthesis (see Fig. 1). At the same time, our method renders up to 907 frames per second, over 3.6× faster than 3DGS [26] and more than 3,000× faster than Zip-NeRF [2].

2. Related Work

Neural Fields. Since their introduction in the context of 3D reconstruction [5, 36, 47, 78], neural fields have become one of the most promising methods for many 3D vision tasks including 3D/4D reconstruction [30, 44, 48–50, 84], 3D generative modeling [4, 32, 43, 52, 65, 70], and view synthesis [1, 2, 37]. Key to their success is among others simplicity, state-of-the-art performance, and robust optimization [33, 73, 78]. In contrast to previous representations such as point- [51, 54, 55], voxel- [35, 53], or mesh-based [22, 72] representations, neural fields do not usually require complex regularization, hand-tuned initialization or optimization control modules as they admit end-to-end optimization and can be queried at arbitrary spatial locations. In the context of view synthesis, Neural Radiance Fields [37] (NeRF) in particular have revolutionized the field by leveraging volumetric rendering, which has proven more robust than prior surface-based rendering approaches [45, 67, 82]. In this work, we employ the robustness and simplicity of neural fields to enable real-time rendering for complex scenes at high quality. More specifically, we use the state-of-the-art radiance field Zip-NeRF [2] to act as a robust prior and source of reliable supervision to train a point-based representation better suited for real-time rendering.

Neural Fields for Real-Time Rendering. NeRF-based models lead to state-of-the-art view synthesis but are typically slow to render so that a variety of works are proposed for speeding up training and inference. While neural fields were initially built on large, compute-heavy multi-layer perceptrons (MLPs) [5, 36, 37, 47], recent works propose the use of voxel representations and interpolation to enable fast training and rendering [16, 40, 59, 87]. Instant NGP [40], for example, demonstrates that a multi-resolution hash grid backbone enables higher quality whilst reducing training time to seconds. However, these works rely on powerful GPUs and often do not achieve real-time rendering for arbitrary scenes. Another line of work aims to represent neural fields as meshes, either as a post-processing step [58, 68] or by direct optimization [6, 23, 61, 71, 74, 75, 84]. These approaches can achieve high frame rates but their quality lacks behind volumetric approaches. More recently, another line of work [11, 17, 24, 60] aims to represent a neural field as a set of easily-cacheable assets such as sparse voxel grids, triplanes, and occupancy grids. These methods retain their high quality but often exhibit large storage requirements, are slower to render on smaller devices, and rely on complex custom rendering implementations [11]. In contrast, we optimize lightweight point-based representations that achieve state-of-the-art quality, are easily compressed, and naturally integrate with graphics software following a rasterization pipeline.

Point-Based Representations. First works propose to render point sets as independent geometry samples [20, 21], which can be implemented efficiently in graphics software [62] and highly parallelized on GPU hardware [27, 64]. To eliminate holes when rendering incomplete surfaces, a line of works explores the "splatting" of points with extents larger than a pixel, e.g. with circular or elliptic shapes [76, 86]. The recent work 3D Gaussian Splatting (3DGS) [26] achieves unprecedented quality and fast training and rendering speed by introducing adaptive density control in combination with efficient rasterization kernels. As a consequence, 3DGS is used in a variety of applications, including 3D human [89] and avatar reconstruction [10, 38, 56], 3D generation [8, 31, 69, 85], SLAM systems [29, 34, 63, 80], 4D reconstruction [77, 79], and open-set segmentation [57, 66]. Further, works are proposed to address aliasing [81, 88] and point densification [7, 15] in the 3DGS representation. Finally, a recent line of works investigate compression for 3DGS [12, 13, 28, 39, 41, 42] and the concurrent work [14] analyzes initialization alternatives. In this work, we combine a NeRF prior for stable optimization with a point-based 3DGS representation for real-time rendering of complex scenes. Compared to prior works, we enable high-quality view synthesis even for complex real-world captures that might contain lighting and exposure variations. Further, we develop pruning and test-time visibility rendering strategies leading to 10× fewer Gaussian primitives at higher quality compared to 3DGS and with inference times of 900+ FPS.

3. Method

Our goal is to develop a lightweight, real-time view synthesis method that is robust even for complex real-world captures. In the following, we discuss the key components for achieving this. First, we optimize radiance fields as a robust prior for complex data (Sec. 3.1). Next, we use the radiance field to first initialize and then to supervise the optimization of point-based 3DGS representations (Sec. 3.2.) We develop a novel pruning technique leading to a significant point count reduction while maintaining high quality (Sec. 3.3). Finally, we cluster input cameras and perform visibility filtering, further accelerating rendering speed to up to 900+ FPS (Sec. 3.4). We show an overview of our method in Fig. 3.

3.1. Neural Radiance Fields as a Robust Prior

Neural Radiance Fields. A radiance field f is a continuous function that maps a 3D point $\mathbf{x} \in \mathbb{R}^3$ and a viewing direction $\mathbf{d} \in \mathbb{S}^2$ to a volume density $\sigma \in \mathbb{R}^+$ and an RGB color value $\mathbf{c} \in \mathbb{R}^3$. Inspired by classical volume rendering [25], a pixel's final color prediction is obtained by approximating the integral via quadrature using sample points:

$$\begin{aligned} \mathbf{c}_{\text{NeRF}} &= \sum_{j=1}^{N_s} \tau_j \alpha_j \mathbf{c}_j \quad \text{with} \\ \tau_j &= \prod_{k=1}^{j-1} (1-\alpha_k), \quad \alpha_j = 1 - e^{-\sigma_j \delta_j} \end{aligned} \tag{1}$$

where τ_j is the transmittance, α_j the alpha value for $\mathbf{x}_j$, and $\delta_j = ||\mathbf{x}_{j+1} - \mathbf{x}_j||_2$ the distance between neighboring sample points. In Neural Radiance Fields [37], f is parameterized as an MLP with ReLU activation f_θ and the network parameters θ are optimized using gradient descent on the reconstruction loss:

$$\mathcal{L}(\theta) = \sum_{\mathbf{r} \in \mathcal{R}_{\text{batch}}} \|\mathbf{c}^{\theta}_{\text{NeRF}}(\mathbf{r}) - \mathbf{c}_{\text{GT}}(\mathbf{r})\|_2^2 \tag{2}$$

where $\mathbf{r} \in \mathcal{R}_{\text{batch}}$ are batches of rays sampled from the set of all pixels / rays $\mathcal{R}$. To further boost training time and quality, Zip-NeRF [2] uses multisampling and an efficient multi-resolution grid backbone [40]. Due to the state-of-the-art performance, we adopt Zip-NeRF as our radiance field prior (see supp. mat. for a comparison to an INGP [40] backbone).

Robust Optimization on Real-World Data. Real-world captures often contain effects such as lighting and exposure variation or motion blur. Crucial for the success of neural fields on such in-the-wild data [33] is the use of Generative Latent Optimization [3] (GLO) embedding vectors or related techniques. More specifically, a per-image latent vector is optimized along with the neural field that enables explaining away these image- and view-dependent effects

$$\mathcal{L}(\theta, \{\mathbf{l}_i\}_{i=1}^N) = \sum_{\mathbf{r}_i \in \mathcal{R}_{\text{batch}}} \|\mathbf{c}^{\theta, l_i}_{\text{NeRF}}(\mathbf{r}_i) - \mathbf{c}_{\text{GT}}(\mathbf{r}_i)\|_2^2 \tag{3}$$

where $\{\mathbf{l}_i\}_{i=1}^N$ indicates the set of GLO vectors and N the number of input images. This allows the model to express appearance changes captured in the input images without introducing wrong geometry such as floating artifacts. At test time, images can be rendered with a constant latent vector (usually the zero vector) to obtain stable and high-quality view synthesis. For all experiments, we follow [2] and optimize a per-image latent vector representing an affine transformation for the bottleneck vector in the Zip-NeRF representation.

3.2. Radiance Field-Informed Gaussian Splatting

Gaussian Splatting. In contrast to neural fields, in 3D Gaussian Splatting [26] an explicit point-based scene representation is optimized. More specifically, the scene is represented as points that are associated with a position $\mathbf{p} \in \mathbb{R}^3$,

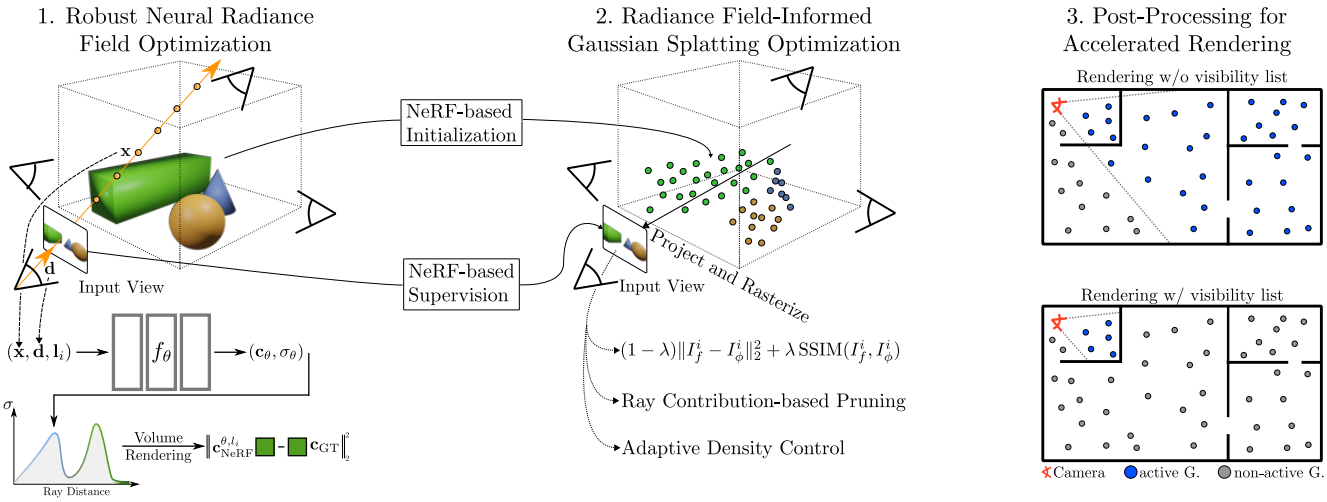

Figure 3. **Overview.** 1. Given posed input images of a scene, we train a robust neural radiance field with GLO embeddings $\mathbf{l}_i$. 2. We use the radiance field prior to initialize and supervise our point-based 3DGS representation that we optimize with a novel pruning technique for more compact, high-quality scenes. 3. We perform viewpoint-based visibility filtering to further accelerate test-time rendering speed.

opacity $o \in [0, 1]$, third-degree spherical harmonics (SH) coefficients $\mathbf{k} \in \mathbb{R}^{16}$, 3D scale $\mathbf{s} \in \mathbb{R}^3$, and the 3D rotation $R \in SO(3)$ represented by 4D quaternions $\mathbf{q} \in \mathbb{R}^4$. Similar to (1), such a representation can be rendered to the image plane for a camera and a list of correctly-sorted points as

$$\mathbf{c}_{\text{GS}} = \sum_{j=1}^{N_p} \mathbf{c}_j \alpha_j \tau_i \quad \text{where} \quad \tau_i = \prod_{i=1}^{j-1} (1 - \alpha_i) \tag{4}$$

where $\mathbf{c}_j$ is the color predicted using the SH coefficients $\mathbf{k}$ and α_j is obtained by evaluating the projected 2D Gaussian with covariance $\Sigma' = JM\Sigma M^T J^T$, multiplied by the per-point opacity o [26], with M being the viewing transformation, J denoting the Jacobian of the affine approximation of the projective transformation [90], and Σ denoting the 3D covariance matrix. To ensure that Σ is a positive semi-definite matrix, it is expressed using the per-point scale matrix $S = \text{diag}(s_1, s_2, s_3)$ and rotation R according to $\Sigma = RSS^T R^T$ [26]. The scene is optimized with a reconstruction loss on the input images and regular densification steps consisting of splitting, merging, and pruning points based on gradient and opacity values.

Radiance Field-based Initialization. A key strength of radiance fields lies in the volume rendering paradigm [37], as opposed to prior surface rendering techniques [45, 67, 82], enabling the ability to initialize, remove, and change density freely in 3D space. In contrast, explicit point-based representations can only provide a gradient signal to already existing geometry prediction due to the rasterization-based approach. The initialization of this representation is hence a crucial property in its optimization process.

We propose to use the radiance field prior for obtaining a suitable initialization. More specifically, for each pixel / ray $\mathbf{r}$ we first define the median depth z_{median} of our NeRF model as the distance to the first sample along the ray with accumulated transmittance $\tau_i > 0.5$. We unproject all pixels / rays into 3D space to obtain our initial point set

$$\mathcal{P}_{\text{init}} = \{\mathbf{p}_i\}_{i \in \mathcal{K}_{\text{rnd}}},\ \mathbf{p}_i = \mathbf{r}_0(i) + \mathbf{d}_{r(i)} \cdot z_{\text{median}}(\mathbf{r}(i)) \tag{5}$$

where $\mathcal{K}_{\text{rnd}}$ are uniform randomly-sampled indices for the list of all rays / pixels, $\mathbf{r}_0(\cdot)$ indicates the ray origin and $\mathbf{d}_{r(\cdot)}$ the normalized ray direction. We found the median depth estimation to perform better than other common choices such as expected depth by being exact sampling point estimates, and we found setting $|\mathcal{K}_{\text{rnd}}|$ to 1 million for all scenes to work well in practice (see Sec. 3.2 of supp. mat. for more details). Further, we initialize

$$\begin{aligned} \mathbf{k}_i &= (\mathbf{k}_i^{1:3}, \mathbf{k}_i^{4:16}),\ \mathbf{k}_i^{1:3} = \mathbf{c}_{\text{NeRF}}(\mathbf{r}(i)),\ \mathbf{k}_i^{4:16} = \mathbf{0}^{1} \\ \mathbf{s}_i &= (s_i, s_i, s_i),\ s_i = \min_{p \in \{p \neq p_i \mid p \in \mathcal{P}_{\text{init}}\}} \|\mathbf{p}_i - \mathbf{p}\|_2 \end{aligned} \tag{6}$$

and set $o_i = 0.1$ and $\mathbf{q}_i$ to the identity rotation. Thus, for each scene we optimize

$$\phi = \{(\mathbf{p}_i, \mathbf{k}_i, \mathbf{s}_i, o_i, \mathbf{q}_i)\}_{i=1}^{N_{\text{init}}} \tag{7}$$

Radiance Field-based Supervision. Radiance fields have been shown to excel even on real-world captures where images contain challenging exposure and lighting variations [2, 33]. We leverage this strength of radiance fields to factor out this complexity and noise of the data to provide a more cleaned up supervision signal than the possibly

[1] We found progressively optimizing $\mathbf{k}^{4:16}$ leads to better results [26].

corrupted input images. More specifically, we render all input images with our NeRF model f_θ and with a constant zero GLO vector

$$\mathcal{I}_f = \{I_f^j\}_{j=1}^N \quad \text{where} \quad I_f^j = \{\mathbf{c}_{\text{NeRF}}^{\theta, l_{\text{zero}}}(\mathbf{r}_j(i))\}_{i=1}^{H \times W} \tag{8}$$

where $\mathbf{l}_{\text{zero}}$ indicates the zero GLO vector, H the height and W the width of the images, and $\mathbf{r}_j(\cdot)$ the rays belonging to the j-th image. We can then use these renderings $\mathcal{I}_f$ to train our point-based representations

$$\mathcal{L}(\phi) = (1-\lambda)\|I_f^i - I_\phi^i\|_2^2 + \lambda\,\text{SSIM}(I_f^i, I_\phi^i) \tag{9}$$

where $i \sim \mathcal{U}$ is drawn from the uniform distribution and we use the default value $\lambda = 0.2$. Another practical benefit of this approach is that we can train from arbitrary camera lens types due to NeRF's flexible ray casting, while the 3D Gaussian Splatting gradient formulation assumes a pinhole camera model and it is unclear how this can be efficiently extended to e.g. fisheye or more complex lens types.

3.3. Ray Contribution-Based Pruning

While 3DGS representations can be efficiently rendered thanks to rasterization, real-time performance still requires a powerful GPU and is not yet achieved on all platforms. The most important property for the rendering performance is the number of points in the scene that need to be rendered.

Importance Score. To obtain a more lightweight representation that can be rendered faster across platforms, we develop a novel pruning technique to reduce the number of Gaussians in the scene whilst maintaining high quality. More specifically, we introduce a pruning step during optimization that removes points that do not contribute significantly to any training view. To this end, we define an importance score by aggregating the ray contribution of Gaussian $\mathbf{p}_i$ along all rays of all input images

$$h(\mathbf{p}_i) = \max_{I_f \in \mathcal{I}_f, r \in I_f} \alpha_i^r \tau_i^r \tag{10}$$

where $\alpha_i^r \tau_i^r$ indicates the ray contribution for the pixel's final color prediction in (4) of Gaussian $\mathbf{p}_i$ along ray $\mathbf{r}$. We find that this formulation leads to improved results compared to concurrent works that investigate similar ideas [12, 28] as we use the exact ray contribution (as opposed to e.g. the opacity) as well as the max operator (instead of e.g. the mean) which is independent of the number of input images, hence more robust to different types of scene coverage [19].

Pruning. We use our importance score during optimization to reduce the overall point count in the scene while maintaining high quality. More specifically, we add a pruning step where we calculate mask values as

$$m_i = m(\mathbf{p}_i) = \mathbb{1}\left(h(\mathbf{p}_i) < t_{\text{prune}}\right),\ t_{\text{prune}} \in [0, 1] \tag{11}$$

and we remove all Gaussians from our scene that have a mask value of one. We apply the pruning step twice over the the course of optimization similar to [12]. The threshold t_{prune} provides a control mechanism over the number of points that are used to represent the scene. In our experiments, we define two values, one value for our default model, and a higher value for a lightweight variant.

3.4. Viewpoint-Based Visibility Filtering

Our pruning technique ensures a compact scene representation with a small overall point count. To scale to larger, more complex scenes such as entire houses or apartments, inspired by classical occlusion culling [9], we introduce a novel viewpoint-based filtering as post-processing step that further speeds up test-time rendering without a quality drop.

Input Camera Clustering. First, we group input cameras together to obtain a meaningful tessellation of the scene space. More specifically, let $(\mathbf{x}_{\text{cam}}^i)_{i=1}^N$ denote the input camera locations for the set of input images $\mathcal{I}$. We run k-means clustering on the input camera locations to obtain k cluster centers $(\mathbf{x}_{\text{cluster}}^i)_{i=1}^k$ and assign the input cameras to the respective cluster centers.

Visibility Filtering. Next, for each cluster center $\mathbf{x}_{\text{cluster}}^j$, we select all assigned input cameras, render the images from these viewpoints, and, similar to (10), calculate an importance score and the respective visibility indicator list

$$\begin{aligned} h_j^{\text{cluster}}(\mathbf{p}_i) &= \max_{I \in \mathcal{I}_c^i, r \in I} \alpha_i^r \tau_i^r, \\ m_j^{\text{cluster}}(\mathbf{p}_i) &= \mathbb{1}\left(h_j^{\text{cluster}}(\mathbf{p}_i) > t_{\text{cluster}}\right) \end{aligned} \tag{12}$$

where $\mathcal{I}_c^i$ is the set of images whose camera positions are assigned to the the cluster center $\mathbf{x}_{\text{cluster}}^i$ and t_{cluster} is a threshold that controls the contribution value of points that should be filtered out (we found setting $t_{\text{cluster}} = 0.001$ to work well in practice). Note that we are not restricted to the input views for calculating these masks. In practice, we hence add random camera samples to $\mathcal{I}_c^i$ to ensure robustness to test views. We calculate the indicator list m_j^{cluster} for each cluster center as a post-processing step after scene optimization.

Visibility List-Based Rendering. To render an arbitrary viewpoint, we first assign its camera center $\mathbf{x}_{\text{cam}}^{\text{test}}$ to the nearest cluster center $\mathbf{x}_{\text{cluster}}^i$. Next, we select the respective indicator list m_i^{cluster}. Finally, we perform default rasterization while only considering the points that are marked as active for the respective cluster. This results in a significant FPS increase up to 45% without any drop in quality.

3.5. Implementation Details

We set the number of initial points N_{init} to 1 million in all experiments. For threshold value t_{prune}, we use 0.01 and

0.25 for our default and lightweight models, respectively, and for the large Zip-NeRF scenes, we use 0.005 and 0.03. We perform pruning after 16 and 24 thousand steps. We follow [26] and use the same densification parameters except for the densification gradient threshold value which we lower to $8.6e^{-5}$ for the Zip-NeRF dataset. We train our radiance fields on 8 V100 GPUs (∼1h) and our 3DGS models on a A100 GPU (∼1h). For the visibility filtering, we use $k = 64$ clusters and we found a small threshold $t_{\text{cluster}} > 0$ to work well in practice and set it to 0.001 for all scenes. For the radiance field training, we follow [2] and use default parameters for all scenes.

4. Experiments

Datasets. We report results on the MipNeRF360 dataset [1], the most common view synthesis benchmark consisting of 9 unbounded indoor and outdoor scenes. We further report results on the Zip-NeRF dataset [2] consisting of 4 large-scale scenes (apartments and houses) with challenging captures that partly contain lighting and exposure variations.

Baselines. On all datasets, we compare against 3DGS [26] as well as MERF [60] and SMERF [11] as the state-of-the-art volumetric approaches that construct efficient voxel and triplane representations together with accelerating structures for empty space skipping. On MipNeRF360, we further compare against mesh-based BakedSDF [84], hash-grid based INGP [40], and point-based approaches Light-Gaussian [12], CompactGaussian [28], and EAGLES [18]. For reference, we always report Zip-NeRF [2], the state-of-the-art offline view synthesis method.

Metrics and Evaluation. We follow common practice and report the view synthesis metrics PSNR, SSIM, and LPIPS. While using techniques such as GLO vectors is essential for high quality on real-world captures (see Sec. 3.1), the evaluation of such models is an open problem such that recent methods [1, 2, 11] train two separate models, one for visualizations, and one (without GLO vectors) purely for the quantitative comparison. In this work, we always train a single model that is robust thanks to the radiance field prior. For evaluation, we simply finetune the trained models on the original training set image data to match potential color shifts and to ensure a fair comparison. Next to measuring quality, we report the rendering speed in frames per second (FPS) on a RTX 3090 GPU and the number of Gaussians in the scenes (only applicable for point-based methods).

4.1. Real-Time View Synthesis

Unbounded Scenes. We observe in Tab. 1a that our method leads to the best quantitative results while achieving faster rendering times than prior state-of-the-art real-time methods such as SMERF [11]. Notably, our model even outperforms the state-of-the-art non-real-time method Zip-NeRF [2] in both SSIM and LPIPS while rendering 1,600× faster. Our lightweight variant ("Ours Light") also exceeds prior works with a mean rendering speed of 907 FPS outpacing even state-of-the-art mesh-based methods such as BakedSDF [84]. Also qualitatively in Fig. 4, we observe that our model achieves the best results. Compared to Zip-NeRF, our method better captures high-frequency textures (e.g., see tablecloth in "Kitchen" scene in Fig. 4) and fine geometric details (e.g., see bicycle spokes in "Bicycle" scene in Fig. 4). Compared to 3DGS, we find that our reconstructions are sharper and more stable while achieving a 2× and 10× overall point count reduction with our default and lightweight variant, respectively.

Large-Scale Scenes. For the Zip-NeRF dataset [2], we observe a similar trend in Tab. 1b. Our default and lightweight variant outperform top-performing real-time SMERF and non-real-time Zip-NeRF in SSIM while rendering significantly faster. Notably, our lightweight variant achieves high quality with a mean SSIM of 0.838 while rendering on average at 748 FPS. In contrast, the state-of-the-

	SSIM↑	PSNR↑	LPIPS↓	FPS↑	#G(M)↓
INGP [40]	0.705	25.68	0.302	9.26	-
BakedSDF [84]	0.697	24.51	0.309	539	-
MERF [60]	0.722	25.24	0.311	171	-
SMERF [11]	0.818	27.99	0.211	228	-
CompactG [28]	0.798	27.08	0.247	128	1.388
LightG [12]	0.799	26.99	0.25	209	1.046
EAGLES [18]	0.809	27.16	0.238	137	1.712
3DGS [26]	0.815	27.20	0.214	251	3.161
Ours Light	0.826	27.56	0.213	907	0.370
Ours	0.843	28.14	0.171	410	1.924
Zip-NeRF [2]	0.836	28.54	0.177	0.25	-

(a) Mip-NeRF360 dataset [1]

	SSIM↑	PSNR↑	LPIPS↓	FPS↑
MERF [60]	0.747	23.49	0.445	318
SMERF [11] ($K = 1$)	0.776	25.44	0.412	356
SMERF [11] ($K = 5$)	0.829	27.28	0.340	221
3DGS [26]	0.809	25.50	0.369	470
Ours Light	0.838	26.11	0.368	748
Ours	0.839	26.17	0.364	630
Zip-NeRF [2]	0.836	27.37	0.305	0.25

(b) Zip-NeRF dataset [2]

Table 1. **Quantitative Comparison.** We compare top-performing real-time rendering approaches and report offline method Zip-NeRF as reference. Our models outperform both NeRF- and GS-based approaches, achieving state-of-the-art view synthesis at higher FPS. Ours Light achieves a 10× reduction of Gaussians (#G) compared to 3DGS while improving quality (1a). Our default model improves even over Zip-NeRF in SSIM and LPIPS while rendering 3,600× faster. On the large-scale scenes in 1b, our models produce the highest SSIM while rendering up to 3.3× faster than top-performing real-time methods such as SMERF.

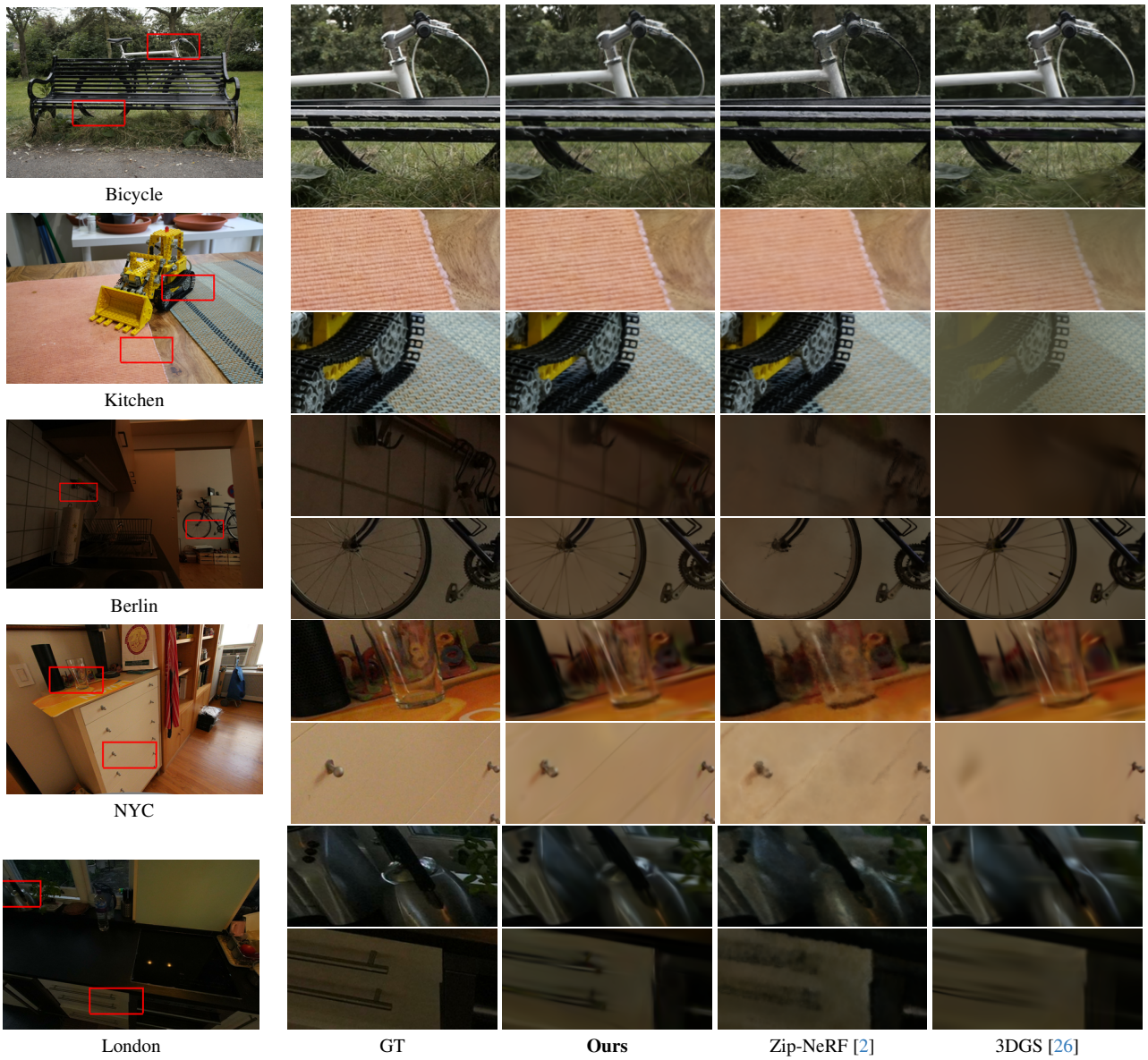

Figure 4. **Qualitative Comparison.** We show results on Bicycle and Kitchen from [1] and on Berlin, NYC, London from [2]. Compared to Zip-NeRF, our method better captures high-frequency textures (e.g., tablecloth in Kitchen) and geometric details (e.g., bicycle spokes in Berlin). Compared to 3DGS, we obtain sharper (e.g., shiny surfaces in London) and more stable results (e.g., color shift in Kitchen).

art real time method for large scenes, i.e. the large variant of SMERF [11] with $5^3 = 125$ submodels, achieves a slightly lower SSIM of 0.829 with a rendering speed of 221 FPS. Also qualitatively in Fig. 4, we observe that our model achieves high visual appeal with sharper and more stable reconstructions. In contrast to 3DGS [26], we find that our method is more robust on challenging captures as shown in Fig. 2 where 3DGS leads to heavily degraded results on the Alameda scene. Note that for 3DGS, results still contain floating artifacts, even when equipped with a per-image module that can handle exposure and lighting variations [11, 26]. In contrast, our method enables high-quality synthesis even for in-the-wild data.

4.2. Ablation Study and Limitations

NeRF-based Initialization. The NeRF-based initialization leads to better quantitative and qualitative results (see Fig. 5). In particular, smaller geometric and texture details might get lost, such as the back of the chair, the books behind the monitor, or the sticky notes on the wall in Fig. 5a.

NeRF-based Supervision. The NeRF-based supervision leads to improved results compared to optimizing the scene

GT **Ours** w/o NeRF Initialization

(a) Qualitative Ablation Study of the NeRF-based Initialization.

GT **Ours** w/o NeRF Supervision

(b) Qualitative Ablation Study of the NeRF-based Supervision.

	SSIM↑	PSNR↑	LPIPS↓	#G (M)↓
Ours	0.839	26.17	0.364	2.022
w/o NeRF Inititialization	0.830	25.71	0.382	1.583
w/o NeRF Supervision	0.835	25.79	0.372	1.849
w/o Pruning	0.839	26.14	0.364	3.049

(c) Quantitative Ablation Study on the Zip-NeRF Dataset.

Figure 5. **Ablation Study.** Without (w/o) the NeRF initialization, geometric and texture details might get lost (5a). Without the NeRF supervision, floating artifacts appear if the views exhibit lighting or exposure changes (5b). W/o pruning, the number of Gaussians is 1.5× larger without any quality improvements (5c).

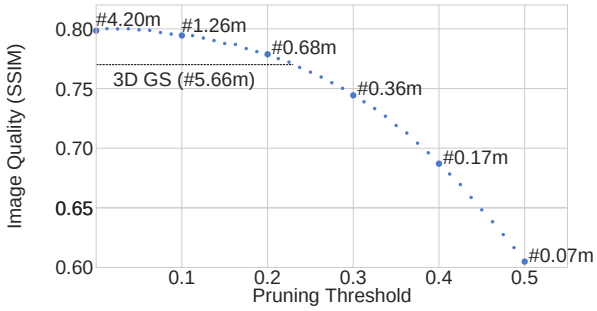

Figure 6. **Pruning.** Pruning thresholds below 0.1 maintain quality while reducing the point count by 4× (shown here on the mip-NeRF 360 Bicycle scene.). We find we can match the 3DGS quality (0.77) with a 10× reduction in Gaussians (5.66m vs. 0.59m).

representation on the input views directly. In particular, for scenes where the input views exhibit exposure or lighting variations, floating artifacts are introduced to model these effects as shown in Fig. 5b. In contrast, our strategy to optimize wrt. the NeRF-based supervision is more stable and leads to better reconstructions for real-world captures.

Pruning. In Tab. 5c, we find that not pruning leads to a small performance drop while exhibiting a significantly larger point count. We hypothesize that a small pruning threshold removes redundancy leading to better generalization. In summary, our pruning technique enables more compact scene representations while maintaining high quality. In Fig. 6 we show that we can match the quality of 3DGS, despite having roughly 10× less Gaussians in the scene.

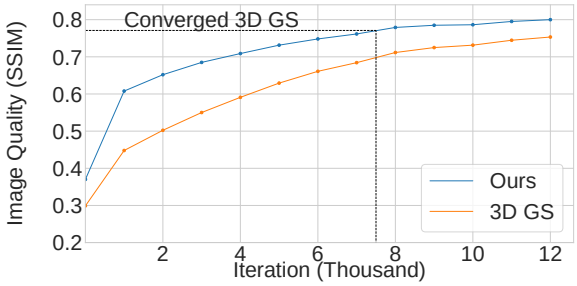

Figure 7. **Quality Progression.** We compare the optimization progression of 3DGS and our default model on the mip-NeRF 360 Bicycle scene. We observe a steeper incline in SSIM and can match the final quality of 3DGS after less than 8k steps.

In Fig. 7, we observe a faster increase in SSIM over the first iterations such that we can match the final 3DGS quality after only 8k iterations.

	mip-NeRF 360 dataset				ZipNeRF dataset			
	SSIM↑	PSNR↑	LPIPS↓	FPS↑	SSIM↑	PSNR↑	LPIPS↓	FPS↑
Ours	0.843	28.14	0.171	410	0.839	26.17	0.364	630
w/o Vis. Fil.	0.843	28.14	0.171	373	0.839	26.17	0.364	435
Ours Light	0.826	27.56	0.213	907	0.838	26.11	0.368	748
w/o Vis. Fil.	0.826	27.56	0.213	887	0.838	26.11	0.368	607

Table 2. **Visibility Filtering.** With this postprocessing, we achieve up to 10% FPS increase on the mip-NeRF 360 scenes and up to 45% improvement in rendering speed when scaling to the larger-scale ZipNeRF scenes while keeping the quality fixed.

Visibility List-Based Rendering. Our visibility list-based rendering enables up to 10% mean FPS speed up on the central object-focused MipNeRF360 scenes and a up to 45% FPS increase on the larger house and apartment-level Zip-NeRF scenes (see Tab. 2). We conclude that this post-processing step is in particular important when scaling to more complex larger-scale scenes.

Limitations. Despite outperforming prior real-time methods, we observe a small gap to ZipNeRF on large-scale scenes which we aim to investigate in the future.

5. Conclusion

We presented RadSplat, a method combining the strengths of neural fields and point-based representations for robust real-time rendering of complex scenes. Using radiance fields as a prior and supervision signal leads to improved results and more stable optimization of point-based 3DGS representations. Our novel pruning leads to more compact scenes with a significantly smaller scene size, whilst improving quality. Finally, our novel test-time filtering further improves rendering speed without a quality drop. Our method achieves state-of-the-art on common benchmarks while rendering up to 3,000× faster than prior works.

References

[1] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *CVPR*, 2022. 1, 2, 6, 7

[2] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Zip-nerf: Anti-aliased grid-based neural radiance fields. In *ICCV*, 2023. 1, 2, 3, 4, 6, 7

[3] Piotr Bojanowski, Armand Joulin, David Lopez-Paz, and Arthur Szlam. Optimizing the latent space of generative networks. In *ICML*, 2018. 3

[4] Eric R. Chan, Connor Z. Lin, Matthew A. Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas Guibas, Jonathan Tremblay, Sameh Khamis, Tero Karras, and Gordon Wetzstein. Efficient geometry-aware 3D generative adversarial networks. In *CVPR*, 2022. 2

[5] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *CVPR*, 2019. 1, 2

[6] Zhiqin Chen, Thomas Funkhouser, Peter Hedman, and Andrea Tagliasacchi. Mobilenerf: Exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures. In *ICCV*, 2023. 2

[7] Kai Cheng, Xiaoxiao Long, Kaizhi Yang, Yao Yao, Wei Yin, Yuexin Ma, Wenping Wang, and Xuejin Chen. Gaussianpro: 3d gaussian splatting with progressive propagation. *arXiv*, 2024. 3

[8] Jaeyoung Chung, Suyoung Lee, Hyeongjin Nam, Jaerin Lee, and Kyoung Mu Lee. Luciddreamer: Domain-free generation of 3d gaussian splatting scenes. *arXiv*, 2023. 3

[9] Satyan R. Coorg and Seth J. Teller. Real-time occlusion culling for models with large occluders. In *SI3D*, 1997. 5

[10] Helisa Dhamo, Yinyu Nie, Arthur Moreau, Jifei Song, Richard Shaw, Yiren Zhou, and Eduardo Pérez-Pellitero. Headgas: Real-time animatable head avatars via 3d gaussian splatting. *arXiv*, 2023. 3

[11] Daniel Duckworth, Peter Hedman, Christian Reiser, Peter Zhizhin, Jean-François Thibert, Mario Lucic, Richard Szeliski, and Jonathan T. Barron. SMERF: streamable memory efficient radiance fields for real-time large-scene exploration. *arXiv*, 2023. 1, 2, 6, 7

[12] Zhiwen Fan, Kevin Wang, Kairun Wen, Zehao Zhu, Dejia Xu, and Zhangyang Wang. Lightgaussian: Unbounded 3d gaussian compression with 15x reduction and 200+ FPS. *arXiv*, 2023. 3, 5, 6

[13] Guangchi Fang and Bing Wang. Mini-splatting: Representing scenes with a constrained number of gaussians. In *ECCV*, 2024. 3

[14] Yalda Foroutan, Daniel Rebain, Kwang Moo Yi, and Andrea Tagliasacchi. Evaluating alternatives to sfm point cloud initialization for gaussian splatting. In *arXiv*, 2024. 3

[15] Linus Franke, Darius Rückert, Laura Fink, Matthias Innmann, and Marc Stamminger. VET: visual error tomography for point cloud completion and high-quality neural rendering. In *SIGGRAPH*, 2023. 3

[16] Fridovich-Keil and Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *CVPR*, 2022. 1, 2

[17] Stephan J Garbin, Marek Kowalski, Matthew Johnson, Jamie Shotton, and Julien Valentin. Fastnerf: High-fidelity neural rendering at 200fps. In *ICCV*, 2021. 2

[18] Sharath Girish, Kamal Gupta, and Abhinav Shrivastava. Eagles: Efficient accelerated 3d gaussians with lightweight encodings. *arXiv*, 2023. 6

[19] Lily Goli, Daniel Rebain, Sara Sabour, Animesh Garg, and Andrea Tagliasacchi. nerf2nerf: Pairwise registration of neural radiance fields. 2023. 5

[20] Markus Gross and Hanspeter Pfister. *Point-based graphics*. Elsevier, 2011. 3

[21] Jeffrey P Grossman and William J Dally. Point sample rendering. In *Rendering Techniques*, 1998. 3

[22] Thibault Groueix, Matthew Fisher, Vladimir G Kim, Bryan C Russell, and Mathieu Aubry. A papier-mâché approach to learning 3d surface generation. In *CVPR*, 2018. 2

[23] Yuan-Chen Guo, Yan-Pei Cao, Chen Wang, Yu He, Ying Shan, and Song-Hai Zhang. Vmesh: Hybrid volume-mesh representation for efficient view synthesis. In *SIGGRAPH Asia*, 2023. 2

[24] Peter Hedman, Pratul P Srinivasan, Ben Mildenhall, Jonathan T Barron, and Paul Debevec. Baking neural radiance fields for real-time view synthesis. In *ICCV*, 2021. 2

[25] James T. Kajiya and Brian Von Herzen. Ray tracing volume densities. In *SIGGRAPH*, 1984. 3

[26] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *SIGGRAPH*, 2023. 1, 2, 3, 4, 6, 7

[27] Samuli Laine and Tero Karras. High-performance software rasterization on gpus. In *SIGGRAPH*, 2011. 3

[28] Joochan Lee, Daniel Rho, Xiangyu Sun, Jong Hwan Ko, and Eunbyung Park. Compact 3d gaussian representation for radiance field. *arXiv*, 2023. 3, 5, 6

[29] Mingrui Li, Shuhong Liu, and Heng Zhou. Sgs-slam: Semantic gaussian splatting for neural dense slam. *arXiv*, 2024. 3

[30] 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 *CVPR*, 2023. 1, 2

[31] Yixun Liang, Xin Yang, Jiantao Lin, Haodong Li, Xiaogang Xu, and Yingcong Chen. Luciddreamer: Towards high-fidelity text-to-3d generation via interval score matching. In *CVPR*, 2024. 3

[32] Chen-Hsuan Lin, Jun Gao, Luming Tang, Towaki Takikawa, Xiaohui Zeng, Xun Huang, Karsten Kreis, Sanja Fidler, Ming-Yu Liu, and Tsung-Yi Lin. Magic3d: High-resolution text-to-3d content creation. In *CVPR*, 2023. 1, 2

[33] Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *CVPR*, 2021. 1, 2, 3, 4

[34] Hidenobu Matsuki, Riku Murai, Paul HJ Kelly, and Andrew J Davison. Gaussian splatting slam. In *CVPR*, 2024. 3

[35] Daniel Maturana and Sebastian Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In *IROS*, 2015. 2

[36] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *CVPR*, 2019. 1, 2

[37] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 1, 2, 3, 4

[38] Arthur Moreau, Jifei Song, Helisa Dhamo, Richard Shaw, Yiren Zhou, and Eduardo Pérez-Pellitero. Human gaussian splatting: Real-time rendering of animatable avatars. In *CVPR*, 2024. 3

[39] Wieland Morgenstern, Florian Barthel, Anna Hilsmann, and Peter Eisert. Compact 3d scene representation via self-organizing gaussian grids. *arXiv*, 2023. 3

[40] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *SIGGRAPH*, 2022. 1, 2, 3, 6

[41] KL Navaneet, Kossar Pourahmadi Meibodi, Soroush Abbasi Koohpayegani, and Hamed Pirsiavash. Compact3d: Compressing gaussian splat radiance field models with vector quantization. *arXiv*, 2023. 3

[42] Simon Niedermayr, Josef Stumpfegger, and Rüdiger Westermann. Compressed 3d gaussian splatting for accelerated novel view synthesis. *arXiv*, 2023. 3

[43] Michael Niemeyer and Andreas Geiger. Giraffe: Representing scenes as compositional generative neural feature fields. In *CVPR*, 2021. 1, 2

[44] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Occupancy flow: 4d reconstruction by learning particle dynamics. In *CVPR*, 2019. 1, 2

[45] Michael Niemeyer, Lars M. Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *CVPR*, 2020. 2, 4

[46] Michael Oechsle, Songyou Peng, and Andreas Geiger. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *ICCV*, 2021. 1

[47] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *CVPR*, 2019. 1, 2

[48] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *ICCV*, 2021. 1, 2

[49] Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M. Seitz. Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields. *SIGGRAPH*, 2021.

[50] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *ECCV*, 2020. 1, 2

[51] Songyou Peng, Chiyu Jiang, Yiyi Liao, Michael Niemeyer, Marc Pollefeys, and Andreas Geiger. Shape as points: A differentiable poisson solver. *NeurIPS*, 2021. 2

[52] Ben Poole, Ajay Jain, Jonathan T Barron, and Ben Mildenhall. Dreamfusion: Text-to-3d using 2d diffusion. In *ICLR*, 2022. 1, 2

[53] Charles R Qi, Hao Su, Matthias Nießner, Angela Dai, Mengyuan Yan, and Leonidas J Guibas. Volumetric and multi-view cnns for object classification on 3d data. In *CVPR*, 2016. 2

[54] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *CVPR*, 2017. 2

[55] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *NeurIPS*, 2017. 2

[56] Shenhan Qian, Tobias Kirschstein, Liam Schoneveld, Davide Davoli, Simon Giebenhain, and Matthias Nießner. Gaussianavatars: Photorealistic head avatars with rigged 3d gaussians. In *CVPR*, 2024. 3

[57] Minghan Qin, Wanhua Li, Jiawei Zhou, Haoqian Wang, and Hanspeter Pfister. Langsplat: 3d language gaussian splatting. In *CVPR*, 2024. 3

[58] Marie-Julie Rakotosaona, Fabian Manhardt, Diego Martin Arroyo, Michael Niemeyer, Abhijit Kundu, and Federico Tombari. Nerfmeshing: Distilling neural radiance fields into geometrically-accurate 3d meshes. In *3DV*, 2024. 1, 2

[59] Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In *ICCV*, 2021. 2

[60] Christian Reiser, Richard Szeliski, Dor Verbin, Pratul P. Srinivasan, Ben Mildenhall, Andreas Geiger, Jonathan T. Barron, and Peter Hedman. MERF: memory-efficient radiance fields for real-time view synthesis in unbounded scenes. *SIGGRAPH*, 2023. 1, 2, 6

[61] Christian Reiser, Stephan Garbin, Pratul P Srinivasan, Dor Verbin, Richard Szeliski, Ben Mildenhall, Jonathan T Barron, Peter Hedman, and Andreas Geiger. Binary opacity grids: Capturing fine geometric detail for mesh-based view synthesis. *arXiv*, 2024. 2

[62] Miguel Sainz and Renato Pajarola. Point-based rendering techniques. *Computers & Graphics*, 2004. 3

[63] Erik Sandström, Keisuke Tateno, Michael Oechsle, Michael Niemeyer, Luc Van Gool, Martin R. Oswald, and Federico Tombari. Splat-slam: Globally optimized rgb-only SLAM with 3d gaussians. *arXiv*, 2024. 3

[64] Markus Schütz, Bernhard Kerbl, and Michael Wimmer. Software rasterization of 2 billion points in real time. *CGIT*, 2022. 3

[65] Katja Schwarz, Yiyi Liao, Michael Niemeyer, and Andreas Geiger. Graf: Generative radiance fields for 3d-aware image synthesis. In *NeurIPS*, 2020. 1, 2

[66] Jin-Chuan Shi, Miao Wang, Hao-Bin Duan, and Shao-Hua Guan. Language embedded 3d gaussians for open-vocabulary scene understanding. In *CVPR*, 2024. 3

[67] Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein. Scene representation networks: Continuous 3d-

structure-aware neural scene representations. In *NeurIPS*, 2019. 2, 4

[68] Jiaxiang Tang, Hang Zhou, Xiaokang Chen, Tianshu Hu, Errui Ding, Jingdong Wang, and Gang Zeng. Delicate textured mesh recovery from nerf via adaptive surface refinement. *arXiv*, 2022. 2

[69] Jiaxiang Tang, Jiawei Ren, Hang Zhou, Ziwei Liu, and Gang Zeng. Dreamgaussian: Generative gaussian splatting for efficient 3d content creation. In *ICLR*, 2024. 3

[70] Christina Tsalicoglou, Fabian Manhardt, Alessio Tonioni, Michael Niemeyer, and Federico Tombari. Textmesh: Generation of realistic 3d meshes from text prompts. *arXiv*, 2023. 1, 2

[71] Haithem Turki, Vasu Agrawal, Samuel Rota Bulò, Lorenzo Porzi, Peter Kontschieder, Deva Ramanan, Michael Zollhöfer, and Christian Richardt. Hybridnerf: Efficient neural rendering via adaptive volumetric surfaces. In *CVPR*, 2024. 2

[72] Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. Pixel2mesh: Generating 3d mesh models from single rgb images. In *ECCV*, 2018. 2

[73] Yuehao Wang, Chaoyi Wang, Bingchen Gong, and Tianfan Xue. Bilateral guided radiance field processing. *SIGGRAPH*, 2024. 2

[74] Zian Wang, Tianchang Shen, Merlin Nimier-David, Nicholas Sharp, Jun Gao, Alexander Keller, Sanja Fidler, Thomas Müller, and Zan Gojcic. Adaptive shells for efficient neural radiance field rendering. In *SIGGRAPH Asia*, 2023. 2

[75] Zian Wang, Tianchang Shen, Merlin Nimier-David, Nicholas Sharp, Jun Gao, Alexander Keller, Sanja Fidler, Thomas Müller, and Zan Gojcic. Adaptive shells for efficient neural radiance field rendering. In *SIGGRAPH*, 2023. 2

[76] Olivia Wiles, Georgia Gkioxari, Richard Szeliski, and Justin Johnson. Synsin: End-to-end view synthesis from a single image. In *CVPR*, 2020. 3

[77] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Wang Xinggang. 4d gaussian splatting for real-time dynamic scene rendering. In *CVPR*, 2024. 3

[78] Yiheng Xie, Towaki Takikawa, Shunsuke Saito, Or Litany, Shiqin Yan, Numair Khan, Federico Tombari, James Tompkin, Vincent Sitzmann, and Srinath Sridhar. Neural fields in visual computing and beyond. In *CFG*, 2022. 1, 2

[79] Zhen Xu, Sida Peng, Haotong Lin, Guangzhao He, Jiaming Sun, Yujun Shen, Hujun Bao, and Xiaowei Zhou. 4k4d: Real-time 4d view synthesis at 4k resolution. In *CVPR*, 2024. 3

[80] Chi Yan, Delin Qu, Dong Wang, Dan Xu, Zhigang Wang, Bin Zhao, and Xuelong Li. Gs-slam: Dense visual slam with 3d gaussian splatting. *arXiv*, 2023. 3

[81] Zhiwen Yan, Weng Fei Low, Yu Chen, and Gim Hee Lee. Multi-scale 3d gaussian splatting for anti-aliased rendering. *arXiv*, 2023. 3

[82] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. In *NeurIPS*, 2020. 2, 4

[83] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. In *NeurIPS*, 2021. 1

[84] Lior Yariv, Peter Hedman, Christian Reiser, Dor Verbin, Pratul P. Srinivasan, Richard Szeliski, Jonathan T. Barron, and Ben Mildenhall. Bakedsdf: Meshing neural sdfs for real-time view synthesis. In *SIGGRAPH*, 2023. 1, 2, 6

[85] Taoran Yi, Jiemin Fang, Junjie Wang, Guanjun Wu, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Qi Tian, and Xinggang Wang. Gaussiandreamer: Fast generation from text to 3d gaussians by bridging 2d and 3d diffusion models. In *CVPR*, 2024. 3

[86] Wang Yifan, Felice Serena, Shihao Wu, Cengiz Öztireli, and Olga Sorkine-Hornung. Differentiable surface splatting for point-based geometry processing. *SIGGRAPH*, 2019. 3

[87] Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. Plenoctrees for real-time rendering of neural radiance fields. In *CVPR*, 2021. 2

[88] Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-splatting: Alias-free 3d gaussian splatting. *arXiv*, 2023. 3

[89] Shunyuan Zheng, Boyao Zhou, Ruizhi Shao, Boning Liu, Shengping Zhang, Liqiang Nie, and Yebin Liu. Gps-gaussian: Generalizable pixel-wise 3d gaussian splatting for real-time human novel view synthesis. In *CVPR*, 2024. 3

[90] Matthias Zwicker, Hanspeter Pfister, Jeroen Van Baar, and Markus Gross. Ewa volume splatting. In *VIS*, 2001. 4