# Spatial Annealing Smoothing for Efficient Few-shot **Neural Rendering**

Yuru Xiao, Xianming Liu, Member, IEEE, Deming Zhai, Member, IEEE, Kui Jiang, Member, IEEE, Junjun Jiang, Senior Member, IEEE, Xiangyang Ji, Member, IEEE

Abstract—Neural Radiance Fields (NeRF) with hybrid representations have shown impressive capabilities in reconstructing scenes for view synthesis, delivering high efficiency. Nonetheless, their performance significantly drops with sparse view inputs, due to the issue of overfitting. While various regularization strategies have been devised to address these challenges, they often depend on inefficient assumptions or are not compatible with hybrid models. There is a clear need for a method that maintains efficiency and improves resilience to sparse views within a hybrid framework. In this paper, we introduce an accurate and efficient few-shot neural rendering method named Spatial Annealing smoothing regularized NeRF (SANeRF), which is specifically designed for a pre-filtering-driven hybrid representation architecture. We implement an exponential reduction of the sample space size from an initially large value. This methodology is crucial for stabilizing the early stages of the training phase and significantly contributes to the enhancement of the subsequent process of detail refinement. Our extensive experiments reveal that, by adding merely one line of code, SANeRF delivers superior rendering quality and much faster reconstruction speed compared to current few-shot NeRF methods. Notably, SANeRF outperforms FreeNeRF by 0.3 dB in PSNR on the Blender dataset, while achieving 700× faster reconstruction speed. Code is available at https://github.com/pulangk97/SANeRF.

Index Terms—Few-shot neural rendering, Spatial Annealing, Efficient.

## I. INTRODUCTION

TEURAL Radiance Fields (NeRF) have emerged as a groundbreaking approach in the field of computer vision and graphics, enabling highly detailed 3D scene reconstructions from a set of 2D images [1]. The core principle behind NeRF involves modeling the volumetric scene function using a neural network. This function maps a spatial location and a viewing direction to a color and density, which can be used to render photorealistic images from novel viewpoints through volume rendering. However, traditional NeRF approaches require a substantial amount of images from various viewpoints to achieve high-quality reconstructions, limiting their applicability in scenarios where data acquisition is costly or challenging.

The research on few-shot NeRF aims to address this limitation by developing methods that can produce high-fidelity 3D scene reconstructions with a minimal number of input images

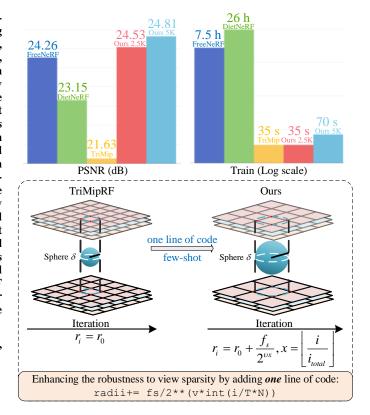


Fig. 1. An overview of our proposed method. We introduce a spatial annealing strategy for a pre-filtering designed efficient NeRF architecture. With the addition of as few as one line of code, our technique substantially improves the rendering quality of the base architecture, outperforming TriMipRF by nearly 3 dB in PSNR in the few-shot setting. Furthermore, our method achieves a reconstruction speed that is 700× faster and delivers superior quality when compared to FreeNeRF.

[2]-[6]. The "few-shot" aspect refers to the ability to work effectively with significantly fewer images than traditional NeRF approaches. This area of research is critical for extending the applicability of NeRF to a wider range of real-world scenarios, such as autonomous driving [7], [8], outdoor environments [9], [10], where capturing a large dataset might be impractical, or for quick 3D modeling in dynamic environments [11], [12].

Numerous methods have emerged to tackle this challenge employing various strategies. Nevertheless, the majority concentrate on alleviating overfitting and refining geometry reconstruction, often neglecting reconstruction efficiency. As shown in Fig. 1, for instance, DietNeRF [2] implements a patchbased semantic consistency regularization approach, necessitating feature extraction from the CLIP [13] to compute the semantic consistency loss. This forward processing, applied to

Y. Xiao, X. Liu, D. Zhai, K. Jiang and J. Jiang are with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China, E-mail: xiaoyuru.30@gmail.com; {csxm, zhaideming, jiangkui, jiangjunjun}@hit.edu.cn.

X. Ji is with the Department of Automation, Tsinghua University, Beijing. 100084, China, E-mail: xyji@tsinghua.edu.cn.

randomly sampled rendered patches, is notably time-intensive. FreeNeRF [5] employs a coarse-to-fine training approach to tackle the few-shot neural rendering problem, seen as a variant of implicit geometry regularization. While FreeNeRF notably improves rendering performance in few-shot scenarios with minimal code adjustments, its training process proves to be time-consuming and labor-intensive, spanning several hours. As a result, there is significant potential and a notable gap in developing a few-shot NeRF approach that not only achieves superior rendering performance but also maintains high efficiency.

To enhance training efficiency alongside high-quality reconstruction, the TriMipRF framework [14] utilizes a tri-plane hybrid representation combined with a pre-filtering strategy for anti-aliasing. Though this methodology significantly accelerates training while achieving high-fidelity rendering, its performance significantly declines due to the overly rapid convergence when input views are sparse as shown in Fig. 1. Directly integrating current regularization methods into the architecture can either compromise training efficiency [2], [6], [15] or be unsuitable for implementation on hybrid representations [5]. Consequently, it is both practical and impressive to develop a straightforward method for TriMipRF-like hybrid representations that enhances robustness to view sparsity without compromising its efficiency.

In this study, we bridge the gap between efficiency and accuracy by introducing a novel few-shot neural rendering technique, specifically designed for the pre-filtering-driven NeRF acceleration architecture. Our results demonstrate that fine-tuning the pre-filtering parameters through a carefully crafted annealing strategy significantly enhances the resilience of the base architecture to sparse view scenarios. Our initial observations drew a parallel between the concept of frequency regularization [5] and the pre-filtering method [16] commonly utilized in anti-aliasing. Though both approaches apply a lowpass filter to frequency positional encoding, the shapes of their masks differ. Leveraging this observation, our method strategically reduces the sample area radius from a generous initial value. This approach intentionally introduces a blurring effect in the early training phases, crucial for establishing the foundational low-frequency geometry. As training progresses, we incrementally refine the details, further elaborating on the initially formed geometry.

In this paper, we introduce a plug-and-play approach that can be seamlessly incorporated into a pre-filtering-driven architecture, as simple as adding a single line of code. Specifically, we integrate spatial annealing smoothing strategy into the popular TriMipRF framework [14] to derive an efficient and accurate few-shot NeRF scheme, dubbed SANeRF, as illustrated in Fig. 1. Extensive experiments on synthetic datasets, Blender [1] and Shiny Blender [17] within a few-shot context, demonstrate the effectiveness and efficiency of our approach. Not only does it markedly exceed the original TriMipRF by approximately 3dB in PSNR, as highlighted in Fig. 1, but it also surpasses the cutting-edge few-shot NeRF technique, FreeNeRF [5], by 0.3dB in PSNR. Additionally, it boasts a training speed 700× faster than FreeNeRF, representing a significant leap forward in both performance and efficiency.

The main contributions of our work are summarized as follows:

- We elucidate the relationship between frequency regularization and the pre-filtering strategy. This insight extends the applicability of frequency regularization-based methods to pre-filtering designed efficient NeRF architectures.
- We introduce a spatial annealing smoothing strategy tailored for an efficient NeRF architecture. By adjusting the sample space in the pre-filtering strategy with just a minimal code addition, we significantly enhance the robustness to view sparsity while maintaining high efficiency.
- We offer extensive experimental results to demonstrate that our method not only attains a superior level of efficiency but also exhibits enhanced performance. Specifically, it surpasses FreeNeRF by 0.3 dB in PSNR on the Blender dataset while achieving 700× faster training speed.

## II. RELATED WORK

Neural Radiance Fields. Neural Radiance Field (NeRF) [1] is a distinguished 3D representation method renowned for its ability to render realistic novel views. It utilizes a Multilayer Perceptron (MLP) network to implicitly associate 3D coordinates with radiance attributes, including density and color. Over the past few years, the research community has developed numerous NeRF extensions, enhancing its application across a variety of domains such as novel view synthesis [16], [18]–[20], surface reconstruction [21]–[24], dynamic scene modeling [12], [25], [26], and 3D content generation [27]-[31]. Although NeRF has pioneered several advancements across different research areas, it encounters significant challenges. The method is particularly known for its slow reconstruction speed and heavy dependence on the density of input views, which limits its efficiency and practicality in broader applications. These challenges highlight the need for ongoing research and development to optimize NeRF's performance and expand its usability in the field of 3D representation and beyond.

Few-shot Neural Rendering. NeRF struggles with accurately fitting the underlying geometry when processing sparse input views. Many approaches incorporate 3D information like sparse point clouds [32], estimated depth [4], [33]–[35], or dense correspondences [36], [37] for enhanced supervision. However, integrating additional algorithms or models to acquire extra 3D data adds complexity and implementation challenges to the overall process. Beyond leveraging 3D information, some methods aim to directly regularize geometry using strategies such as patch-based smooth priors [15], semantic consistency [2], or geometric consistency [6]. These techniques, however, tend to prolong training times because of the added rendering burden on extra sampled patches. Learning-based methods [38]–[45], on the other hand, seek to train a network on extensive multi-view datasets to internalize a geometric prior, yet these approaches require costly pre-training and additional fine-tuning steps.

Recently, FreeNeRF [5] addresses the issue of underconstrained geometry in sparse view settings by progressively increasing the frequency of positional encoding. This technique eliminates the need for additional 3D information and does not increase the computational cost of the base architecture, positioning it as a robust baseline for few-shot NeRF applications. While the coarse-to-fine training approach of FreeNeRF has demonstrated effectiveness in few-shot scenarios, it still necessitates lengthy training periods. Furthermore, its method of applying a mask to frequency positional encoding is not readily adaptable to various NeRF acceleration architectures. Motivated by FreeNeRF's achievements, our goal is to develop a straightforward solution for a NeRF acceleration architecture, aiming to enhance performance in few-shot settings without compromising training efficiency.

**NeRF Acceleration.** NeRF's notable performance is related to its network's high capacity and dense sampling along rays, which, however, result in prolonged training and rendering times. To counter this, various strategies have been developed focusing on hybrid representations, such as low-rank tensors [46], multi-resolution hash grids [47], [48], or tri-plane [14], to expedite training and rendering processes. These approaches avoid querying a large MLP for each sample point, thereby facilitating faster training convergence and reducing computational costs during rendering.

## III. PRELIMINARIES AND MOTIVATION

**Frequency Regularization.** FreeNeRF [5] addresses the fewshot challenge by progressively increasing the frequency of positional encoding throughout the training process, a technique known as frequency regularization. This approach is straightforwardly implemented using a modulated mask as

$$\gamma'(t, T, \mathbf{x}) = \gamma(\mathbf{x}) \odot \mathbf{M}(t, T, L) \tag{1}$$

with

$$\mathbf{M}_i(t,T,L) = \begin{cases} 1 & \text{if } i \leq \frac{t \cdot L}{T} + 3\\ \frac{t \cdot L}{T} - \lfloor \frac{t \cdot L}{T} \rfloor & \text{if } \frac{t \cdot L}{T} + 3 < i \leq \frac{t \cdot L}{T} + 6\\ 0 & \text{if } i > \frac{t \cdot L}{T} + 6 \end{cases},$$

where t and T represent the current and total iteration steps, respectively, while  $\gamma$  and  $\gamma'$  correspond to the initial and masked positional encodings, respectively. M represents the frequency mask, which linearly expands throughout the training process.

**TriMipRF.** TriMipRF [14] introduces a pre-filtering strategy for tri-plane representation. Due to the incompatibility of integrated positional encoding (IPE) introduced by MipNeRF [16] with this hybrid model, TriMipRF performs area-sampling by querying features on the three planes with plane levels correlating with the projected radius of the sampling sphere within the cone. The relevant levels are denoted by  $\lfloor l \rfloor$  and  $\lceil l \rceil$ , where l signifies the query level corresponding to the sphere's radius illustrated as

$$l = \log_2\left(\frac{\tau}{\ddot{n}}\right),\tag{3}$$

where  $\ddot{r}$  represents the base radius associated with the base level  $l_0$ , while  $\tau$  denotes the radius of the sampled sphere, depicted as

$$\tau = \frac{\|\mathbf{x} - \mathbf{o}\|_{2} \cdot f\dot{r}}{\|\mathbf{d}\|_{2} \cdot \sqrt{\left(\sqrt{\|\mathbf{d}\|_{2}^{2} - f^{2}} - \dot{r}\right)^{2} + f^{2}}}.$$
 (4)

The central position of the sphere is denoted by  $\mathbf{x} = \mathbf{o} + t\mathbf{d}$ , where  $\mathbf{o}$  represents the camera center,  $\mathbf{d}$  is the ray's direction, and t is the distance from  $\mathbf{x}$  to  $\mathbf{o}$ . f denotes the focal length. The radius of the disc,  $\dot{r}$ , is calculated as  $\sqrt{\Delta x \cdot \Delta y/\pi}$ , with  $\Delta x$  and  $\Delta y$  representing the pixel's width and height in world coordinates, respectively.

Motivation. Although FreeNeRF [5] is effective in few-shot settings, its application is restricted to implicit representations. Transitioning such methodologies to encompass hybrid representations offers a valuable, albeit challenging, opportunity to boost efficiency. Our objective is to craft a universal form of frequency regularization applicable to both implicit and hybrid representations. It is worth noting that, despite TriMipRF [14] eliminating coordinate positional encoding, it still necessitates spatial area-sampling for its pre-filtering design evolved from MipNeRF [16]. This observation motivates us to explore a dual form of frequency regularization in the spatial domain. This approach is particularly well-suited for hybrid representations that employ an area-sampling strategy. Consequently, we introduce a spatial annealing strategy tailored for the TriMipRF architecture [14]. Through precise adjustments to the size of the sample area, we realize notable performance gains in fewshot contexts, requiring only minor code modifications. We term this method spatial annealing smoothing, which forms the core of our proposed approach, SANeRF.

# IV. METHOD

**Frequency Regularization & Pre-filtering Strategy.** We begin by examining the relationship between frequency regularization (see Eq. 1) and the pre-filtering strategy based on integrated positional encoding. We consider a multivariate Gaussian distribution in 3D space, depicted as

$$G(\mathbf{x}) = e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})},$$
 (5)

where  $\mu$  represents the position, and  $\Sigma$  denotes the covariance matrix. For straightforward comparison, we model the Gaussian as isotropic, characterized by a diagonal  $\Sigma$  with uniform elements  $\sigma_f^2$ , reflecting the sample space size. Following Mip-NeRF [16], we calculate the Gaussian's integrated positional encoding  $\gamma_G$  as

$$\gamma_G(\boldsymbol{\mu}) = \gamma_G'(\boldsymbol{\mu}) \odot \mathbf{M}_l \tag{6}$$

with

$$\mathbf{M}_{l} = \left[e^{-\frac{1}{2}\sigma_{f}^{2}}, e^{-\frac{1}{2}\sigma_{f}^{2}}, ..., e^{-2^{2L-3}\sigma_{f}^{2}}, e^{-2^{2L-3}\sigma_{f}^{2}}\right]$$

$$\gamma'_{G}(\boldsymbol{\mu}) = \left[\sin(\boldsymbol{\mu}), \cos(\boldsymbol{\mu}), ..., \sin(2^{L-1}\boldsymbol{\mu}), \cos(2^{L-1}\boldsymbol{\mu})\right],$$
(7)

where  $\gamma_G'(\mu)$  represents the positional encoding of  $\mu$ , while  $\mathbf{M}_l$  denotes a low-pass mask applied to this encoding. The mask's structure aligns with a Gaussian distribution, featuring

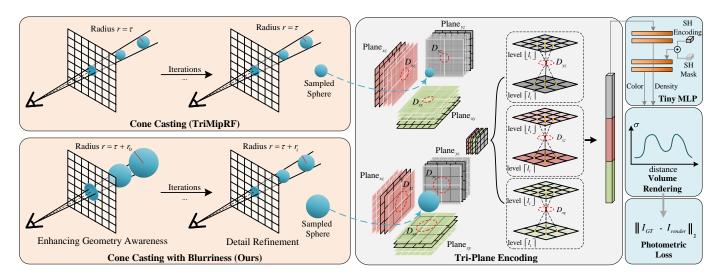


Fig. 2. An overview of the complete framework. We introduce an efficient few-shot NeRF method utilizing TriMipRF. Initially, we set the sample sphere's radius larger than that of the base sphere to optimize global geometry, as depicted in the bottom left corner. During training, we progressively reduce the sphere's radius through exponential decay, thereby refining local details within the predefined global structure.

a covariance  $\sigma_i^2=1/\sigma_f^2.$  Here,  $\sigma_i^2$  regulates the frequency bandwidth.

Upon comparing Eq. 7 with Eq. 1, we observe a notable similarity between the integrated positional encoding and frequency regularization: both apply a low-pass mask to the original positional encoding, with the mask's form being the primary difference. Drawing inspiration from FreeNeRF [5], which linearly expands the frequency mask (effectively exponentially increasing the frequency bandwidth), we adopt an exponential growth model for  $\sigma_i^2$ , defined as  $\sigma_i^2 = 2^x$ , where x is the step increment corresponding to the iteration count. Concurrently, the spatial Gaussian's covariance  $\sigma_f^2$  decreases exponentially, represented as

$$\sigma_f^2 = \frac{1}{2^x}. (8)$$

Consequently, we deduce that the frequency regularization introduced by FreeNeRF [5] can be executed by inversely adjusting the spatial sample space within the pre-filtering strategy, as detailed in Eq. 8. This insight guides the development of our spatial annealing smoothing strategy.

**Spatial Annealing Strategy.** Fig. 2 provides an overview of our method. We have devised a coarse-to-fine training strategy based on the TriMipRF [14]. Initially, we implement blurring in the rendering process by increasing the radius r of the sample sphere, as depicted in the bottom left corner of Fig. 2. The base radius of the sample sphere is defined in Eq. 4. This increment in radius corresponds to a higher level of the three planes, characterized by a lower resolution. At the training's outset, this method assists the optimization process in focusing on the global geometric structures' reconstruction, which is crucial for addressing the overfitting issues prevalent in fewshot scenarios. The sample area's radius is systematically reduced, adhering to an exponential decay as detailed in Eq. 8. This gradual reduction is designed to shift more training focus toward enhancing local geometric and textural details.

The annealing process's specific steps are outlined as

$$r_i = \begin{cases} \tau + \frac{f_s}{2^{vx}}, i < T \\ \tau, i \ge T \end{cases}, x = \left\lfloor \frac{iN_{split}}{T} \right\rfloor$$
 (9)

where  $N_{split}$  is the total number of decrement steps, i is the current iteration count and T stands for the stop point of annealing.  $f_s$  dictates the initial size of the sphere, while  $\vartheta$  controls the speed of the decrement.

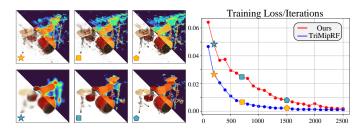


Fig. 3. Comparison results during the training procedure. The training loss curve, displayed on the right, reveals that TriMipRF exhibits premature convergence early in the training, resulting in the underfitting of the geometry depicted on the left. Conversely, our spatial annealing strategy effectively addresses this challenge.

**Rendering.** The queried level of the three planes decreases in response to the exponential reduction in the size of the enlarged sphere. This relationship can be mathematically represented as

$$l_i = log_2(\frac{\tau 2^{\vartheta x} + f_s}{\ddot{r} 2^{\vartheta x}}). \tag{10}$$

We adhere to the TriMipRF framework [14], querying eight feature vectors from feature planes at levels  $\lfloor l_i \rfloor$  and  $\lceil l_i \rceil$ . In the initial training phase, increasing the level leads to the exclusion of higher-resolution features, which are typically linked to premature convergence and overfitting as shown in Fig. 3. This strategy emphasizes the learning of low-frequency geometric structures. As the radius decreases exponentially, the training progressively shifts focus, allowing more iterations

IEEE TRANSACTIONS ON IMAGE PROCESSING

for refining high-frequency details on the pre-established low-frequency geometry. The interpolation of queried features is based on the center position of the projected circle and the sphere level  $l_i$ . The final feature vector  $\mathbf{f}$ , input into the tiny MLP, is the concatenation of the feature vectors  $\{\mathbf{f}_{XY},\mathbf{f}_{XZ},\mathbf{f}_{YZ}\}$  queried from each plane.

## V. EXPERIMENTS

**Datasets and Metrics.** We evaluate our method using the Blender dataset [1], which comprises 8 synthetic scenes, and the shiny Blender dataset [17], featuring 6 synthetic scenes observed from a surround view perspective. Consistent with the DietNeRF [2], we train our model on 8 views identified by the IDs 26, 86, 2, 55, 75, 93, 16, and 73 for both datasets. The evaluation is conducted on 25 images, evenly selected from each dataset's test set. In line with community standards, all images are downsampled by a factor of 2 using an average filter. For quantitative analysis, we report the average scores across all test scenes for PSNR, SSIM, and LPIPS.

To further illustrate our method's efficacy across different base architectures and substantiate the comparison between frequency regularization and spatial annealing strategy, we evaluate our approach using the LLFF dataset [49], employing the MipNeRF architecture [16] in conjunction with the FreeNeRF [5]. The LLFF dataset comprises 8 forward-facing scenes. In alignment with FreeNeRF, we designate every eighth image as a test view and uniformly select from the remaining images to assemble the training views. All input images are downsampled using an average filter, achieving a reduction factor of 8. This approach allows us to assess the robustness and adaptability of our method in handling real-world scenes.

**Degree Truncation of Spherical Harmonic Encoding.** Spherical Harmonic Encodings play a pivotal role in capturing view-dependent appearance in various applications [17]. The real Spherical Harmonic Encoding is denoted by  $\{Real(Y_l^m)\}, l \in \mathbb{N}, m=0,\pm 1,\pm 2,\ldots \pm l.$  These encodings utilize Spherical Harmonic Functions, which are defined as

$$Y_l^m(\lambda, \phi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-|m|)!}{(l+|m|)!}} P_l^m(\cos \lambda) e^{im\phi} . \quad (11)$$

When the input views are sparse, fitting the view-dependent colors becomes challenging. To address this and cover a broader range of unseen view directions, we directly mask the higher levels of the Spherical Harmonic Encoding. This masking process can be represented as

$$\mathbf{M}_{i}(n) = \begin{cases} 1, i \leq \sum_{j=0}^{n-1} 2j + 1\\ 0, i > \sum_{j=0}^{n-1} 2j + 1 \end{cases}$$
 (12)

where n represents the truncated level. The mask zeroes out all Spherical Harmonic components above level n, while retaining all lower-level components.

**Implementation Details.** We implement our methodology based on the TriMipRF codebase [14], utilizing the PyTorch framework [50] with the tiny-cuda-nn extension [51]. Building on TriMipRF's original configurations, we introduce additional

hyperparameters tailored to our spatial annealing strategy. Specifically, we initialize the sphere size  $f_s$  at 0.15, set the decrease rate  $\vartheta$  to 0.2, define the total number of decrement steps  $N_{split}$  as 30, and establish the stop point T at 2K. In the few-shot setting, there is a significant reduction in the number of input rays. Consequently, we limit the maximum training steps for both our method and the baseline TriMipRF to one-tenth of those employed in TriMipRF's full-view setting. We train our model using the AdamW optimizer [52] for 2.5K iterations, with a learning rate of  $2\times 10^{-3}$  and a weight decay of  $1\times 10^{-5}$ . Regarding the degree truncation in Spherical Harmonic Encoding, we consistently set the truncated level n to 2 across all experiments.

5

**Baselines.** We conduct comparative analyses between our method and several baselines: the vanilla NeRF [1], MipNeRF [16], and a range of few-shot NeRF approaches, including FreeNeRF [5], DietNeRF [2], MixNeRF [53], and InfoNeRF [3]. These methods are specifically designed for few-shot settings, aiming to perform effectively without the need for additional 3D data or intricate network architectures.

Furthermore, we extend the TriMipRF framework [14] and the rasterization-based radiance field method, 3DGS [54], to few-shot applications to provide a comprehensive analysis. For our quantitative comparisons, we utilize the benchmark results as reported in the studies on FreeNeRF [5] and MixNeRF [53]. To assess the reconstruction time, we record the rough training durations for all considered baseline methods under their reported configurations, using a single NVIDIA 3090 GPU.

## A. Blender Dataset

Fig. 4 and Tab. I show the qualitative and quantitative comparisons on the Blender dataset [1], respectively. The qualitative analysis reveals that the TriMipRF [14] produces noticeable "floater" artifacts in the "chair" and "hotdog" scenes, as well as distorted geometry in the "mic" scene. In contrast, our approach, which requires only a few additional lines of code, effectively addresses these issues. When compared with the latest few-shot baseline, FreeNeRF [5], our method shows superior performance, particularly in detail-rich areas highlighted in red boxes. Specific examples include the texture in the "chair" scene and the edge of the plate in the "hotdog" scene.

TABLE I QUANTITATIVE RESULTS ON BLENDER WITH 8 INPUT VIEWS.

Method	PSNR↑	SSIM↑	LPIPS↓	Train↓
NeRF [1]	14.93	0.687	0.318	6 h
MipNeRF [16]	18.74	0.830	0.240	9 h
Simplified NeRF [2]	20.09	0.822	0.179	2 h
3DGS [54]	22.20	0.860	0.121	28 s
TriMipRF [14]	21.63	0.818	0.183	35 s
DietNeRF [2]	23.15	0.866	0.109	26 h
DietNeRF, $\mathcal{L}_{\mathbf{MSE}}$ ft [2]	23.59	0.874	0.097	32 h
InfoNeRF [3]	22.01	0.852	0.133	7 h
MixNeRF [53]	23.84	0.878	0.103	9 h
FreeNeRF [5]	24.26	0.883	0.098	7.5 h
Ours $(2.5K \text{ iterations})$	24.53	0.881	0.102	35 s
Ours ( $5K$ iterations)	24.81	0.882	0.101	70 s

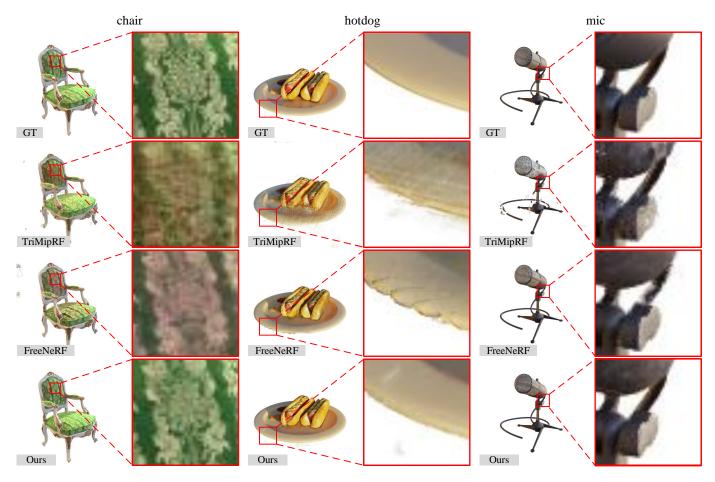


Fig. 4. Qualitative Results on Blender. We present qualitative comparisons of our method with the base architecture TriMipRF and the FreeNeRF few-shot baseline, utilizing 8 input views consistent with the FreeNeRF configuration.

Our method demonstrates significant advancements in quantitative results, outperforming TriMipRF [14] by nearly 3 dB in PSNR with similar 35 seconds reconstruction time. Furthermore, the evaluated training durations demonstrate that our method achieves faster reconstruction speed. Notably, while the FreeNeRF [5] necessitates 7.5 hours for training, and Diet-NeRF [2], with its patch-based semantic regularizer, requires an extensive 26 hours, our approach achieves a remarkable 700× reconstruction speed compared to FreeNeRF. In addition to its efficiency, our method also delivers superior quantitative outcomes, exceeding FreeNeRF by 0.27 dB in PSNR. While these results are impressive, extending the training duration has the potential to further enhance performance. To this end, we increase the training iterations to 5K, while keeping all other settings unchanged. As shown in Tab. I, our method not only surpasses FreeNeRF with a 0.55 dB gain in PSNR but also delivers a substantial 350× acceleration.

## B. Shiny Blender Dataset

We conduct qualitative and quantitative analyses to compare our method with the base architecture using the Shiny Blender dataset [17], as illustrated in Fig. 5 and Tab. II. Qualitatively, our approach significantly improves novel view synthesis over the baseline. For example, TriMipRF [14] displays noticeable

white artifacts near the coffee cup and car, as seen in the first and third rows, respectively. In contrast, our method achieves more accurate geometry and realistic appearances. Quantitatively, our method surpasses TriMipRF by 3.5 dB in PSNR, demonstrating enhanced effectiveness and robustness in various scenes within the Shiny Blender dataset.

TABLE II QUANTITATIVE COMPARISON ON SHINY BLENDER DATASET WITH 8 INPUT VIEWS.

Scene		TriMipRF			Ours	
Scelle	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
ball	13.61	0.843	0.384	16.94	0.898	0.276
car	20.02	0.815	0.158	21.89	0.864	0.086
coffee	19.14	0.799	0.226	26.65	0.913	0.112
helmet	17.74	0.727	0.333	21.08	0.834	0.180
teapot	30.18	0.972	0.051	34.79	0.988	0.018
toaster	15.32	0.624	0.340	15.55	0.721	0.217
AVE	19.34	0.797	0.249	22.82	0.870	0.148

## VI. METHOD ANALYSIS

In this section, we provide a comprehensive analysis of our method. Sec. VI-A delves into the comparison between the frequency regularization approach and our spatial annealing strategy, as introduced in Sec. IV. While our method draws

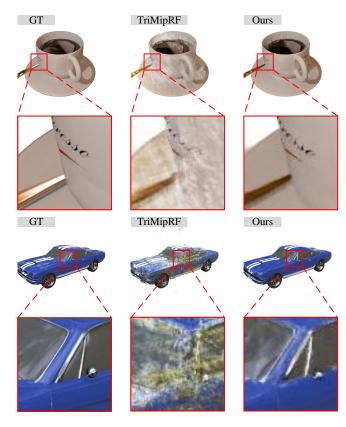


Fig. 5. Comparison with TriMipRF on Shiny Blender dataset. TriMipRF exhibits noticeable white artifacts. In contrast, our method reconstructs scenes with significantly enhanced rendering quality.

inspiration from frequency regularization, it exhibits enhanced generality and adaptability, rendering it more effective for hybrid representation architectures. In Sec. VI-B, we evaluate our method's robustness against view sparsity through thorough experimentation. Sec. VI-C then offers an exhaustive ablation study, further elucidating the strengths and nuances of our approach.

# A. Frequency Regularization vs Spatial Annealing

In Sec. IV, we provide a theoretical analysis that clarifies the relationship between the pre-filtering strategy and frequency regularization based on integrated positional encoding. To demonstrate the strategy's validity and to evaluate our method's performance across various base architectures, we apply it using the JAX MipNeRF framework [16]. In this setup, we adjust the initial sphere size to 0.2, establish the decrease rate  $\vartheta$  at 1, and set the total number of decrease steps  $N_{split}$  to 33.

Fig. 6 and Tab. III present qualitative and quantitative comparisons of our method with FreeNeRF [5], as implemented using the JAX MipNeRF framework on the LLFF dataset [49]. Quantitatively, our method matches or slightly outperforms FreeNeRF, showing up to a 0.1 dB increase in PSNR across the 8 forward-facing scenes. Qualitatively, our method often provides improved novel view synthesis and depth map rendering. For instance, whereas FreeNeRF struggles with accurately reconstructing the leaf edges in the first row and the ceiling's

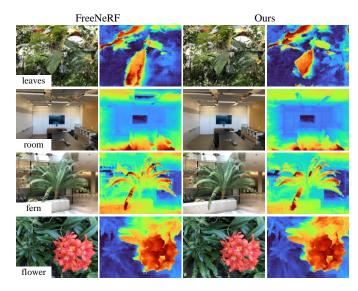


Fig. 6. Qualitative comparison with FreeNeRF on LLFF with 3 input views. We implement our method utilizing the JAX MipNeRF architecture to thoroughly evaluate our method's effectiveness across different base architectures.

TABLE III

QUANTITATIVE COMPARISON ON LLFF WITH 3 INPUT VIEWS.

C		FreeNeRF		Ours			
Scene	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	
fern	20.79	0.647	0.297	20.80	0.648	0.288	
flower	20.68	0.612	0.297	20.90	0.637	0.263	
fortress	23.42	0.594	0.316	23.36	0.555	0.317	
horns	17.93	0.581	0.362	18.28	0.568	0.354	
leaves	16.25	0.504	0.368	16.48	0.551	0.312	
orchids	15.61	0.435	0.365	15.32	0.432	0.338	
room	21.71	0.815	0.213	22.00	0.821	0.204	
trex	20.26	0.708	0.250	20.24	0.720	0.232	
AVE	19.58	0.612	0.308	19.67	0.616	0.288	

middle section in the second row, our method maintains geometric consistency and offers more detailed depth maps. This advantage likely stems from our method's greater adaptability. Unlike frequency regularization, constrained by a fixed frequency range, our method dynamically adjusts the sample space to a wider range, thereby enhancing global geometry reconstruction. This aligns with FreeNeRF's recommendation that a larger sample sphere combined with a lower positional encoding frequency can significantly boost global geometry accuracy.

# B. Robustness to View Sparsity

We validate our method alongside the TriMipRF [14] using varying numbers of input views to assess our method's resilience to view sparsity, as shown in Fig. 7. We use the first n images from the Blender's training set as input, where n is varied among 8, 20, 40, 60, 80, and 100, with the dataset containing a total of 100 images. We fix the total iteration steps at 10,000 and the spatial annealing strategy's endpoint at 2,000, maintaining consistency with the settings detailed in Sec. V. The bottom curve of Fig. 7 demonstrates that our method consistently outperforms TriMipRF under different input view conditions. Notably, it shows marked improvements in scenarios with sparse input views, achieving a 2.50 dB and

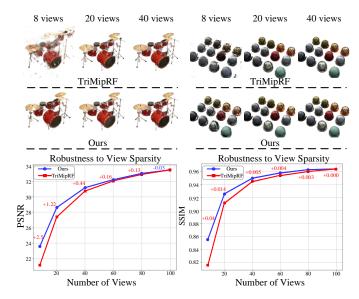


Fig. 7. We conduct a validation to assess the robustness of our method to view sparsity, implementing it alongside the base TriMipRF architecture with various input view configurations. The top part of the figure displays qualitative results, while the bottom part presents the quantitative curves.

1.22 dB higher PSNR with 8 and 20 input views, respectively. While the enhancement diminishes as the number of views increases, our approach still outperforms TriMipRF by 0.13 dB with 80 input views. This indicates a significant boost in robustness to view sparsity. our method exhibits a negligible quality loss (under 0.05 dB) even with the full set of training views. The qualitative analysis at the top of Fig. 7 reveals fewer artifacts and geometric distortions, especially noticeable with 8 input views, further confirming our method's efficacy in enhancing robustness against view sparsity.

# C. Ablation Study

While our method introduces two additional hyperparameters: decreasing speed  $\vartheta$  and initial sphere size  $f_s$ , both are crucial for modulating early training stage blurriness. We observe that enhancements over the base architecture are consistently robust, particularly with a larger  $f_s$  and smaller  $\vartheta$ , which favor global geometry reconstruction. Our robustness validation, presented in Tab. IV, involved varying  $f_s$  and  $\vartheta$  from 0.05 to 0.35 in increments of 0.05, while maintaining other parameters as detailed in Sec. V. Results in normal black indicate improved PSNR metrics over TriMipRF [14], whereas blue signifies reductions. Despite some performance drops at high  $\vartheta$  and low  $f_s$ , leading to suboptimal geometry, our method predominantly enhances metrics with higher  $f_s$  and lower  $\vartheta$ . These findings affirm our method's practicality and robustness concerning hyperparameter variations.

To thoroughly assess our proposed method's efficacy, we conduct an ablation study using the Blender dataset [1] and the Shiny Blender dataset [17], each with 8 input views, as detailed in Tab. V. We evaluate the contributions of the newly introduced spatial annealing strategy (Sa) and the spherical harmonic mask (SH Mask) separately. The findings reveal that the spatial annealing strategy significantly enhances perfor-

TABLE IV

ABLATION STUDY ON THE IMPACT OF  $f_s$  and  $\vartheta$ . We assess our method's PSNR metrics across a range of hyperparameter combinations.

					θ			
		0.05	0.1	0.15	0.2	0.25	0.3	0.35
	0.05	22.59	22.19	22.04	20.65	19.63	19.76	19.48
	0.1	22.91	23.28	23.60	22.78	22.47	21.86	21.16
	0.15	23.31	23.73	24.33	23.31	23.30	23.38	22.31
$f_s$	0.2	23.23	23.62	24.53	23.92	23.79	23.63	22.84
	0.25	23.23	23.43	24.25	23.90	23.77	23.64	22.86
	0.3	23.41	23.24	24.35	24.34	24.04	23.34	22.73
	0.35	23.35	23.69	24.57	24.42	23.99	23.76	23.05

mance in a few-shot setting, achieving a PSNR increase of 1.2 dB on the Blender dataset and 2.3 dB on the Shiny Blender dataset with Sa alone. These results underscore the substantial improvement afforded by our method.

TABLE V ABLATION STUDY.

Blender								
	Sa	SH Mask	PSNR↑	SSIM↑	LPIPS↓	Train		
TriMipRF	×	×	21.63	0.818	0.183	35 s		
	<b>√</b>	×	22.83	0.856	0.120			
Ours	×	✓	22.27	0.825	0.173	35 s		
	✓	✓	24.53	0.881	0.102			
Shiny Blender								
TriMipRF	X	×	19.34	0.797	0.249	35 s		
	<b>√</b>	×	21.69	0.859	0.152			
Ours	×	✓	20.51	0.805	0.229	35 s		
	✓	✓	22.82	0.870	0.148			

## VII. CONCLUSION

In this study, we present a novel spatial annealing smoothing strategy tailored for a hybrid representation architecture equipped with a pre-filtering strategy. This approach adaptively modifies the spatial sampling size using a meticulously designed annealing process, enhancing geometric reconstruction and detail refinement. Remarkably, our approach requires only minimal modifications to the base architecture's code to achieve state-of-the-art performance in few-shot scenarios while preserving efficiency. We acknowledge that our spatial annealing strategy possesses the potential to enhance the training stability of diverse architectures crafted with pre-filtering.

#### REFERENCES

- B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.
- [2] A. Jain, M. Tancik, and P. Abbeel, "Putting nerf on a diet: Semantically consistent few-shot view synthesis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5885–5894.
- [3] M. Kim, S. Seo, and B. Han, "Infonerf: Ray entropy minimization for few-shot neural volume rendering," in *Proceedings of the IEEE/CVF* Conference on Computer Vision and Pattern Recognition, 2022, pp. 12912–12921.
- [4] G. Wang, Z. Chen, C. C. Loy, and Z. Liu, "Sparsenerf: Distilling depth ranking for few-shot novel view synthesis," in *Proceedings of* the IEEE/CVF International Conference on Computer Vision, 2023, pp. 9065–9076.

- [5] J. Yang, M. Pavone, and Y. Wang, "Freenerf: Improving few-shot neural rendering with free frequency regularization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 8254–8263.
- [6] M.-S. Kwak, J. Song, and S. Kim, "Geconerf: few-shot neural radiance fields via geometric consistency," in *Proceedings of the 40th Interna*tional Conference on Machine Learning, 2023, pp. 18023–18036.
- [7] Z. Xie, J. Zhang, W. Li, F. Zhang, and L. Zhang, "S-nerf: Neural radiance fields for street views," in *International Conference on Learning Representations (ICLR)*, 2023.
- [8] Y. Chen, F. Rong, S. Duggal, S. Wang, X. Yan, S. Manivasagam, S. Xue, E. Yumer, and R. Urtasun, "Geosim: Realistic video simulation via geometry-aware composition for self-driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 7230–7240.
- [9] C. Wang, J. Sun, L. Liu, C. Wu, Z. Shen, D. Wu, Y. Dai, and L. Zhang, "Digging into depth priors for outdoor neural radiance fields," in Proceedings of the 31st ACM International Conference on Multimedia, 2023, pp. 1221–1230.
- [10] K. Rematas, A. Liu, P. P. Srinivasan, J. T. Barron, A. Tagliasacchi, T. Funkhouser, and V. Ferrari, "Urban radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 12932–12942.
- [11] W. Xian, J.-B. Huang, J. Kopf, and C. Kim, "Space-time neural irradiance fields for free-viewpoint video," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9421–9431.
- [12] K. Park, U. Sinha, J. T. Barron, S. Bouaziz, D. B. Goldman, S. M. Seitz, and R. Martin-Brualla, "Nerfies: Deformable neural radiance fields," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 5865–5874.
- [13] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark et al., "Learning transferable visual models from natural language supervision," in *International* conference on machine learning. PMLR, 2021, pp. 8748–8763.
- [14] W. Hu, Y. Wang, L. Ma, B. Yang, L. Gao, X. Liu, and Y. Ma, "Tri-miprf: Tri-mip representation for efficient anti-aliasing neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 19774–19783.
- [15] M. Niemeyer, J. T. Barron, B. Mildenhall, M. S. Sajjadi, A. Geiger, and N. 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*, 2022, pp. 5480–5490.
- [16] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan, "Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5855–5864.
- [17] D. Verbin, P. Hedman, B. Mildenhall, T. Zickler, J. T. Barron, and P. P. Srinivasan, "Ref-nerf: Structured view-dependent appearance for neural radiance fields," in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022, pp. 5481–5490.
- [18] R. Martin-Brualla, N. Radwan, M. S. Sajjadi, J. T. Barron, A. Dosovitskiy, and D. Duckworth, "Nerf in the wild: Neural radiance fields for unconstrained photo collections," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7210–7219.
- [19] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Mip-nerf 360: Unbounded anti-aliased neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5470–5479.
- [20] B. Mildenhall, P. Hedman, R. Martin-Brualla, P. P. Srinivasan, and J. T. Barron, "Nerf in the dark: High dynamic range view synthesis from noisy raw images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 16190–16199.
- [21] M. Oechsle, S. Peng, and A. Geiger, "Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5589–5599.
- [22] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang, "Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction," *Advances in Neural Information Processing Systems*, vol. 34, pp. 27 171–27 183, 2021.
- [23] Z. Li, T. Müller, A. Evans, R. H. Taylor, M. Unberath, M.-Y. Liu, and C.-H. Lin, "Neuralangelo: High-fidelity neural surface reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 8456–8465.

- [24] Y. Wang, I. Skorokhodov, and P. Wonka, "Pet-neus: Positional encoding tri-planes for neural surfaces," in *Proceedings of the IEEE/CVF Confer*ence on Computer Vision and Pattern Recognition, 2023, pp. 12598– 12607.
- [25] J. Fang, T. Yi, X. Wang, L. Xie, X. Zhang, W. Liu, M. Nießner, and Q. Tian, "Fast dynamic radiance fields with time-aware neural voxels," in SIGGRAPH Asia 2022 Conference Papers, 2022, pp. 1–9.
- [26] Y.-L. Liu, C. Gao, A. Meuleman, H.-Y. Tseng, A. Saraf, C. Kim, Y.-Y. Chuang, J. Kopf, and J.-B. Huang, "Robust dynamic radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 13–23.
- [27] J. Gu, A. Trevithick, K.-E. Lin, J. M. Susskind, C. Theobalt, L. Liu, and R. Ramamoorthi, "Nerfdiff: Single-image view synthesis with nerf-guided distillation from 3d-aware diffusion," in *International Conference on Machine Learning*. PMLR, 2023, pp. 11808–11826.
- [28] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall, "Dreamfusion: Text-to-3d using 2d diffusion," in *The Eleventh International Conference on Learning Representations*, 2022.
- [29] J. R. Shue, E. R. Chan, R. Po, Z. Ankner, J. Wu, and G. Wetzstein, "3d neural field generation using triplane diffusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 20875–20886.
- [30] H. Chen, J. Gu, A. Chen, W. Tian, Z. Tu, L. Liu, and H. Su, "Single-stage diffusion nerf: A unified approach to 3d generation and reconstruction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 2416–2425.
- [31] Y. Hong, K. Zhang, J. Gu, S. Bi, Y. Zhou, D. Liu, F. Liu, K. Sunkavalli, T. Bui, and H. Tan, "Lrm: Large reconstruction model for single image to 3d," arXiv preprint arXiv:2311.04400, 2023.
- [32] K. Deng, A. Liu, J.-Y. Zhu, and D. Ramanan, "Depth-supervised nerf: Fewer views and faster training for free," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 12882–12891.
- [33] A. Cao, C. Rockwell, and J. Johnson, "Fwd: Real-time novel view synthesis with forward warping and depth," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 15713–15724.
- [34] Y. Wei, S. Liu, Y. Rao, W. Zhao, J. Lu, and J. Zhou, "Nerfingmvs: Guided optimization of neural radiance fields for indoor multi-view stereo," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5610–5619.
- [35] B. Roessle, J. T. Barron, B. Mildenhall, P. P. Srinivasan, and M. Nießner, "Dense depth priors for neural radiance fields from sparse input views," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 12892–12901.
- [36] P. Truong, M.-J. Rakotosaona, F. Manhardt, and F. Tombari, "Sparf: Neural radiance fields from sparse and noisy poses," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 4190–4200.
- [37] Y. Lao, X. Xu, X. Liu, H. Zhao et al., "Corresnerf: Image correspondence priors for neural radiance fields," Advances in Neural Information Processing Systems, vol. 36, 2024.
- [38] K.-E. Lin, Y.-C. Lin, W.-S. Lai, T.-Y. Lin, Y.-C. Shih, and R. Ramamoorthi, "Vision transformer for nerf-based view synthesis from a single input image," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 806–815.
- [39] A. Yu, V. Ye, M. Tancik, and A. Kanazawa, "pixelnerf: Neural radiance fields from one or few images," in *Proceedings of the IEEE/CVF* Conference on Computer Vision and Pattern Recognition, 2021, pp. 4578–4587.
- [40] A. Chen, Z. Xu, F. Zhao, X. Zhang, F. Xiang, J. Yu, and H. Su, "Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 14124–14133.
- [41] Q. Wang, Z. Wang, K. Genova, P. P. Srinivasan, H. Zhou, J. T. Barron, R. Martin-Brualla, N. Snavely, and T. Funkhouser, "Ibrnet: Learning multi-view image-based rendering," in *Proceedings of the IEEE/CVF* Conference on Computer Vision and Pattern Recognition, 2021, pp. 4690–4699
- [42] X. Long, C. Lin, P. Wang, T. Komura, and W. Wang, "Sparseneus: Fast generalizable neural surface reconstruction from sparse views," in European Conference on Computer Vision. Springer, 2022, pp. 210– 227.
- [43] Y. Liu, S. Peng, L. Liu, Q. Wang, P. Wang, C. Theobalt, X. Zhou, and W. Wang, "Neural rays for occlusion-aware image-based rendering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 7824–7833.

- [44] J. Chibane, A. Bansal, V. Lazova, and G. Pons-Moll, "Stereo radiance fields (srf): Learning view synthesis for sparse views of novel scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7911–7920.
- [45] A. Trevithick and B. Yang, "Grf: Learning a general radiance field for 3d representation and rendering," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15182–15192.
- [46] A. Chen, Z. Xu, A. Geiger, J. Yu, and H. Su, "Tensorf: Tensorial radiance fields," in *European Conference on Computer Vision*. Springer, 2022, pp. 333–350.
- [47] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM transactions on graphics (TOG)*, vol. 41, no. 4, pp. 1–15, 2022.
- [48] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Zip-nerf: Anti-aliased grid-based neural radiance fields," in *Proceedings* of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 19697–19705.
- [49] B. Mildenhall, P. P. Srinivasan, R. Ortiz-Cayon, N. K. Kalantari, R. Ramamoorthi, R. Ng, and A. Kar, "Local light field fusion: Practical view synthesis with prescriptive sampling guidelines," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, pp. 1–14, 2019.
- [50] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga et al., "Pytorch: An imperative style, high-performance deep learning library," Advances in neural information processing systems, vol. 32, 2019.
- [51] T. Müller, "tiny-cuda-nn, 4 2021," URL: https://github. com/NVlabs/tinycuda-nn, vol. 4.
- [52] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," arXiv preprint arXiv:1711.05101, 2017.
- [53] S. Seo, D. Han, Y. Chang, and N. Kwak, "Mixnerf: Modeling a ray with mixture density for novel view synthesis from sparse inputs," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 20659–20668.
- [54] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," ACM Transactions on Graphics, vol. 42, no. 4, pp. 1–14, 2023.