

FreGS: 3D Gaussian Splatting with Progressive Frequency Regularization

Jiahui Zhang¹ Fangneng Zhan² Muyu Xu¹ Shijian Lu¹ Eric Xing^{3,4}
¹Nanyang Technological University ²Max Planck Institute for Informatics
³Carnegie Mellon University ⁴MBZUAI

jiahui003@e.ntu.edu.sg fzhan@mpi-inf.mpg.de muyu001@e.ntu.edu.sg

shijian.lu@ntu.edu.sg epxing@cs.cmu.edu

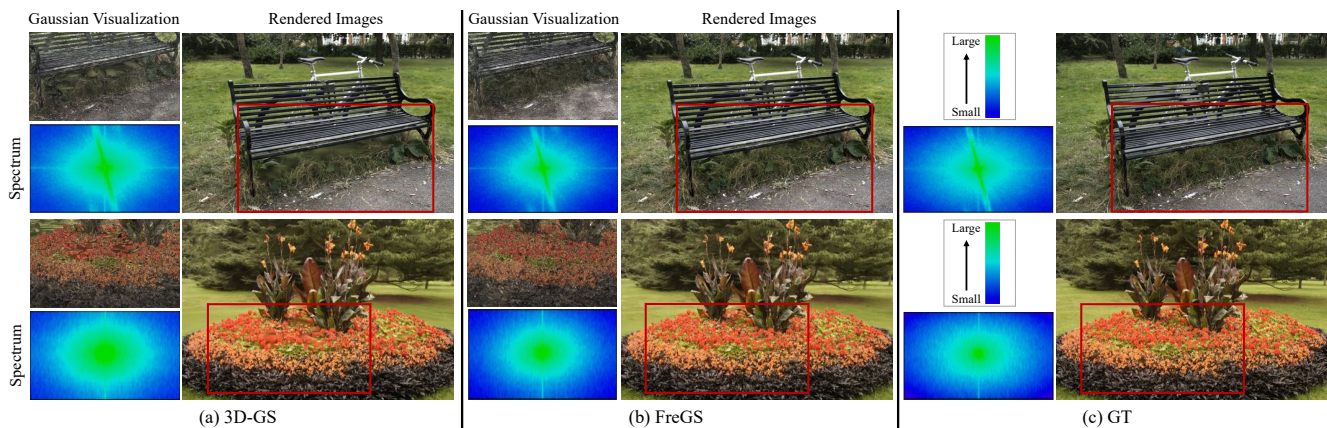


Figure 1. The proposed FreGS mitigates the over-reconstruction of Gaussian densification and renders images with much less blur and artifact as compared with the 3D Gaussian splatting (3D-GS). For the two sample images from Mip-NeRF360 [2], (a) and (b) show the *Rendered Image* and the *Gaussian Visualization* of the highlighted regions, as well as the *Spectra* of over-reconstructed areas in the rendered image by 3D-GS and corresponding areas in FreGS. The *Gaussian Visualization* shows how the learnt rasterized 3D Gaussians compose images (all Gaussians are rasterized with full opacity). The *Spectra* are generated via image Fourier transformation, where the colour changes from blue to green as the spectrum amplitude changes from small to large.

Abstract

3D Gaussian splatting has achieved very impressive performance in real-time novel view synthesis. However, it often suffers from over-reconstruction during Gaussian densification where high-variance image regions are covered by a few large Gaussians only, leading to blur and artifacts in the rendered images. We design a progressive frequency regularization (FreGS) technique to tackle the over-reconstruction issue within the frequency space. Specifically, FreGS performs coarse-to-fine Gaussian densification by exploiting low-to-high frequency components that can be easily extracted with low-pass and high-pass filters in the Fourier space. By minimizing the discrepancy between the frequency spectrum of the rendered image and the corresponding ground truth, it achieves high-quality Gaussian densification and alleviates the over-reconstruction of

Gaussian splatting effectively. Experiments over multiple widely adopted benchmarks (e.g., Mip-NeRF360, Tanks-and-Temples and Deep Blending) show that FreGS achieves superior novel view synthesis and outperforms the state-of-the-art consistently.

1. Introduction

Novel View Synthesis (NVS) has been a pivotal task in the realm of 3D computer vision which holds immense significance in various applications such as virtual reality, image editing, etc. It aims for generating images from arbitrary viewpoints of a scene, often necessitating precise modelling of the scene from multiple scene images. Leveraging implicit scene representation and differentiable volume rendering, NeRF [21] and its extension [1, 2] have recently achieved remarkable progress in novel view synthesis. However, NeRF is inherently plagued by long train-

*Shijian Lu is the corresponding author.

ing and rendering time. Though several NeRF variants [5, 7, 9, 22, 26] speed up the training and rendering greatly, they often sacrifice the quality of rendered images notably, especially while handling high-resolution rendering.

As a compelling alternative to NeRF, 3D Gaussian splatting (3D-GS) [16] has attracted increasing attention by offering superb training and inference speed while maintaining competitive rendering quality. By introducing anisotropic 3D Gaussians together with adaptive density control of Gaussian properties, 3D-GS can learn superb and explicit scene representations for novel view synthesis. It replaces the cumbersome volume rendering in NeRF by efficient splatting, which directly projects 3D Gaussians onto a 2D plane and ensures real-time rendering. However, 3D-GS often suffers from over-reconstruction [16] during Gaussian densification, where high-variance image regions are covered by a few large Gaussians only which leads to clear deficiency in the learnt representations. The over-reconstruction can be clearly observed with blur and artifacts in the rendered 2D images as well as the discrepancy of frequency spectrum of the render images (by 3D-GS) and the corresponding ground truth as illustrated in Fig. 1.

Based on the observation that the over-reconstruction manifests clearly by the discrepancy in frequency spectra, we design FreGS, an innovative 3D Gaussian splatting technique that addresses the over-reconstruction by regularizing the frequency signals in the Fourier space. FreGS introduces a novel frequency annealing technique to achieve progressive frequency regularization. Specifically, FreGS takes a coarse-to-fine Gaussian densification process by annealing the regularization progressively from low-frequency signals to high-frequency signals, based on the rationale that low-frequency and high-frequency signals usually encode large-scale (e.g., global patterns and structures which are easier to model) and small-scale features (e.g., local details which are harder to model), respectively. The progressive regularization strives to minimize the discrepancy of frequency spectra of the rendered image and the corresponding ground truth, which provides faithful guidance in the frequency space and complements the pixel-level L1 loss in the spatial space effectively. Extensive experiments show that FreGS mitigates the over-reconstruction and greatly improves Gaussian densification and novel view synthesis as illustrated in Fig. 1.

The contributions of this work can be summarized in three aspects. *First*, we propose FreGS, an innovative 3D Gaussian splatting framework that addresses the over-reconstruction issue via frequency regularization in the frequency space. To the best of our knowledge, this is the first effort that tackles the over-reconstruction issue of 3D Gaussian splatting from a spectral perspective. *Second*, we design a frequency annealing technique for progressive frequency regularization. The annealing performs

regularization from low-to-high frequency signals progressively, achieving faithful coarse-to-fine Gaussian densification. *Third*, experiments over multiple benchmarks show that FreGS achieves superior novel view synthesis and outperforms the 3D-GS consistently.

2. Related Work

2.1. Neural Rendering for Novel View Synthesis

Novel view synthesis aims to generate new, unseen views of a scene or object from a set of existing images or viewpoints. When deep learning became popular in early days, CNNs were explored for novel view synthesis [12, 27, 32], e.g., they were adopted to predict blending weights for image rendering in [12]. Later, researchers exploit CNNs for volumetric ray-marching [14, 29]. For example, Sitzmann et al. propose Deepvoxels [29] which builds a persistent 3D volumetric scene representation and then achieves rendering via volumetric ray-marching.

Recently, neural radiance field (NeRF) [21] has been widely explored to achieve novel view synthesis via implicit scene representation and differentiable volume rendering. It exploits MLPs to model 3D scene representations from multi-view 2D images and can generate novel views with superb multi-view consistency. A number of NeRF variants have also been designed for handling various challenging conditions, such as dynamic scenes [6, 8, 10, 24, 33], free camera pose [3, 19, 20, 38, 39] and few shot [4, 15, 23, 36], antialiasing [2]. However, novel view synthesis with NeRF often comes at the expense of extremely long training and rendering times. Several studies [5, 7, 9, 13, 22, 26] attempt to reduce the training and rendering times. For example, KiloNeRF [26] speeds up NeRF rendering by utilizing thousands of tiny MLPs instead of one single large MLP. Chen et al. [5] leverage 4D tensor to represent full volume field and factorize the tensor into several compact low-rank tensor components for efficient radiance field reconstruction. Muller et al. propose InstantNGP [22] which introduces multi-resolution hash tables to achieve fast training and real-time rendering. However, most aforementioned work tends to sacrifice the quality of synthesized images especially while handling high-resolution rendering.

As a compelling alternative to NeRF, 3D Gaussian splatting [16] introduces anisotropic 3D Gaussians and efficient differentiable splatting which enables high-quality explicit scene representation while maintaining efficient training and real-time rendering. However, the over-reconstruction of 3D Gaussians during Gaussian densification often introduces blur and artifacts in the rendered images. Our FreGS introduces progressive frequency regularization with frequency annealing to mitigate the over-reconstruction issue, enabling superior Gaussian densification and high-quality novel view synthesis.

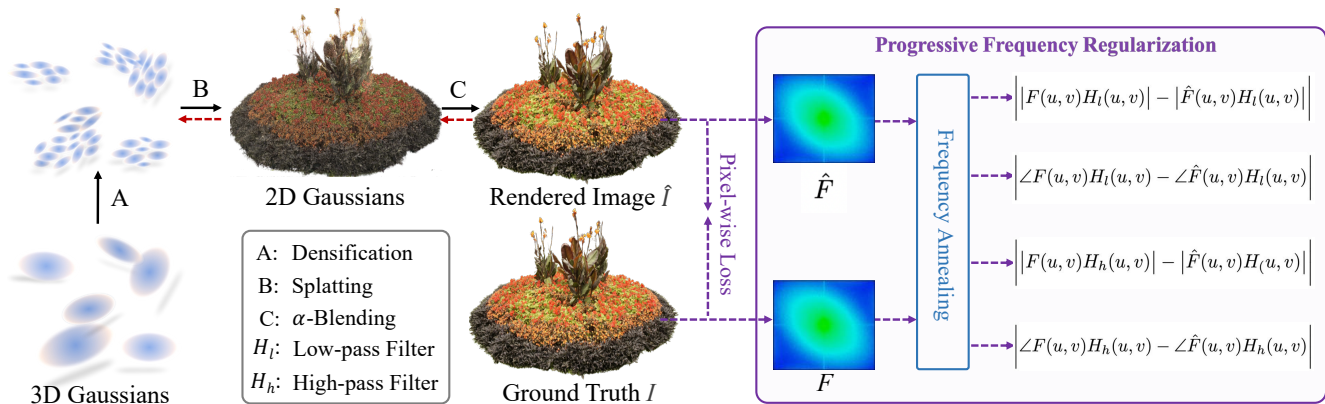


Figure 2. **Overview of the proposed FreGS.** 3D Gaussians are initialized by structure-from-motion. After splatting 3D Gaussians, we can obtain 2D Gaussians and then leverage standard α -blending for rendering. Frequency spectra \hat{F} and F are generated by applying Fourier transform to rendered image \hat{I} and ground truth I , respectively. Frequency regularization is achieved by regularizing discrepancies of amplitude $|F(u, v)|$ and phase $\angle F(u, v)$ in Fourier space. A novel frequency annealing technique is designed to achieve progressive frequency regularization. With low-pass filter H_l and dynamic high-pass filter H_h , low-to-high frequency components are progressively leveraged to perform coarse-to-fine Gaussian densification. Note, the progressive frequency regularization is complementary to the pixel-wise loss between \hat{I} and I . The red dashed line highlights the regularization process for Gaussian densification.

2.2. Frequency in Neural Rendering

NeRF has been widely explored in the frequency space. For example, [21] exploits sinusoidal functions of varying frequencies to encode inputs, overcoming the constraint of neural networks which often struggle to learn high-frequency information from low-dimensional inputs [30, 31, 37]. Several NeRF variants [19, 24, 34–36] also demonstrate the importance of learning in the frequency space under various challenging scenarios. For example, BARF [19] gradually increases frequency for learning NeRF without camera poses. WaveNeRF [35] introduces wavelet frequency decomposition into multi-view stereo for achieving generalizable NeRF. Differently, FreGS boosts 3D Gaussian splatting in the frequency space, demonstrating that progressive frequency regularization with frequency annealing can lead more effective Gaussian densification and advanced novel view synthesis.

3. Proposed Method

We propose FreGS, a novel 3D Gaussian splatting with progressive frequency regularization which is the first to alleviate the over-reconstruction issue of 3D Gaussian splatting from frequency perspective. Fig. 2 shows the overview of FreGS. The original 3D Gaussian splatting [16] (3D-GS), including Gaussian densification, is briefly introduced in Sec. 3.1. In Sec. 3.2, we first reveal the reason for the effectiveness of frequency regularization in addressing the over-reconstruction issue and improving Gaussian densification. Then, we describe amplitude and phase discrepancies employed for frequency regularization within the

Fourier space. To reduce the difficulty of Gaussian densification, we design frequency annealing technique (Sec. 3.3) to achieve progressive frequency regularization, which can gradually exploit low-to-high frequency components to perform coarse-to-fine Gaussian densification.

3.1. Preliminary

3D Gaussian Splatting. 3D-GS models scene representations explicitly with anisotropic 3D Gaussians and achieves real-time rendering by efficient differentiable splatting. Given a sparse point cloud generated by structure-from-motion [11, 28], a set of 3D Gaussians is created, each of which is represented by a covariance matrix Σ , center position p , opacity α and spherical harmonics coefficients representing color c , where the covariance matrix Σ is represented by scaling matrix and rotation matrix for differentiable optimization.

Gaussian densification aims to transform the initial sparse set of Gaussians to a more densely populated set, enhancing its ability to accurately represent the scene. It mainly focuses on two cases. The first is the regions with missing geometric features (corresponding to under-reconstruction) while the other is the large high-variance regions covered by a few large Gaussians only (corresponding to over-reconstruction). Both of these cases result in inadequate representation of regions within scenes. For under-reconstruction, Gaussians are densified by cloning the Gaussians, which increases both the total volume and the number of Gaussians. For over-reconstruction, Gaussian densification is achieved by dividing large Gaussians into multiple smaller Gaussians, which keeps the total vol-

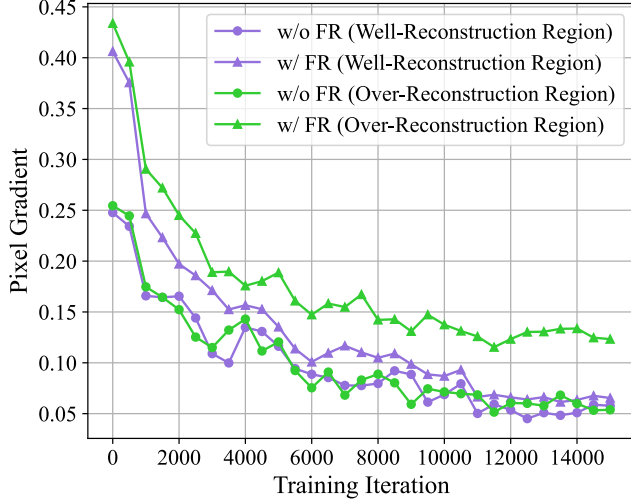


Figure 3. Average pixel gradients within **over-reconstruction regions** and **well-reconstruction regions** in scene ‘Bicycle’. The curve with circle (w/o frequency regularization (FR)) represents the method equivalent to 3D-GS [16], which utilizes pixel-wise L1 loss in the spatial domain only. As the Gaussian densification is terminated after the 15000th iteration as in 3D-GS, we only show comparisons before the 15000th iteration. It can be observed that the frequency regularization can increase the pixel gradient within *over-reconstruction regions* significantly. Thus, compared with L1 loss, the frequency regularization shows superior capability in revealing the over-reconstruction region.

ume but increases the number of Gaussians.

For rendering, 3D Gaussians are projected to a 2D plane by splatting. The rendering can then be achieved via the α -blending. Specifically, the color C of a pixel can be computed by blending N ordered 2D Gaussians that overlap the pixel, which can be formulated by:

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

where the color c_i and α_i are calculated by multiplying the covariance matrix of i -th 2D Gaussian by per-point spherical harmonics coefficients and opacity, respectively.

3.2. Frequency Regularization

In this section, we first explore the reason why 3D-GS leads to over-reconstruction. We compute the average gradient of pixels within the *over-reconstruction regions*, tracking its changes as training progresses. As Fig. 3 shows, with a naive pixel-wise L1 loss, the average gradient could be quite small although the regions are not well reconstructed, which misleads the Gaussian densification. Specifically, the small pixel gradients are back-propagated to 2D splats for this pixel and the corresponding 3D Gaussians. As Gaussian

densification is not applied to Gaussians with small gradients [16], these Gaussians cannot be densified through splitting into smaller Gaussians, leading to over-reconstruction. The consequence of over-reconstruction is an insufficient representation of regions, marked by deficiencies in both overall structure (low-frequency information) and details (high-frequency information). Compared with pixel space, the over-reconstruction region can be better revealed in frequency space by explicitly disentangling different frequency components. Thus, it is intuitive to guide the Gaussian densification by explicitly applying regularization in frequency domain. Fig. 3 shows that the average pixel gradient increases significantly with frequency regularization, demonstrating its effectiveness. We thus conclude that with frequency regularization, Gaussians can be adaptively densified in the over-reconstruction regions. In contrast, L1 loss cannot differentiate over-reconstructed regions from well-reconstructed ones, leading to many redundant Gaussians created in well-reconstructed regions.

Based on the above analysis, we design FreGS which aims to boost 3D Gaussian splatting from frequency perspective. Specifically, it alleviates over-reconstruction and improves Gaussian densification by minimizing the discrepancy between the frequency spectrum of rendered images and corresponding ground truth. Amplitude and phase, as two major elements of frequency, can capture different information of the image. Therefore, we achieve the frequency regularization by regularizing the amplitude and phase discrepancies between rendered images $\hat{I} \in \mathbb{R}^{H \times W \times C}$ and ground truth $I \in \mathbb{R}^{H \times W \times C}$ within Fourier space.

Here, we detail the amplitude and phase discrepancies. We first convert \hat{I} and I to corresponding frequency representations \hat{F} and F by 2D discrete Fourier transform. Take I as an example:

$$F(u, v) = \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} I(x, y) \cdot e^{-i2\pi(u \frac{x}{H} + v \frac{y}{W})}, \quad (2)$$

where (x, y) and (u, v) represent the coordinates in an image and its frequency spectrum, respectively. $I(x, y)$ and $F(u, v)$ denote the pixel value and complex frequency value, respectively. Then, $F(u, v)$ can be expressed in terms of amplitude $|F(u, v)|$ and phase $\angle F(u, v)$ as below:

$$|F(u, v)| = \sqrt{\text{Re}(u, v)^2 + \text{Im}(u, v)^2} \quad (3)$$

$$\angle F(u, v) = \arctan\left(\frac{\text{Im}(u, v)}{\text{Re}(u, v)}\right), \quad (4)$$

where $\text{Im}(u, v)$ and $\text{Re}(u, v)$ represent the imaginary components and the real components of $F(u, v)$.

The amplitude and phase discrepancies (denoted as d_a and d_p) between the rendered image \hat{I} and the ground truth

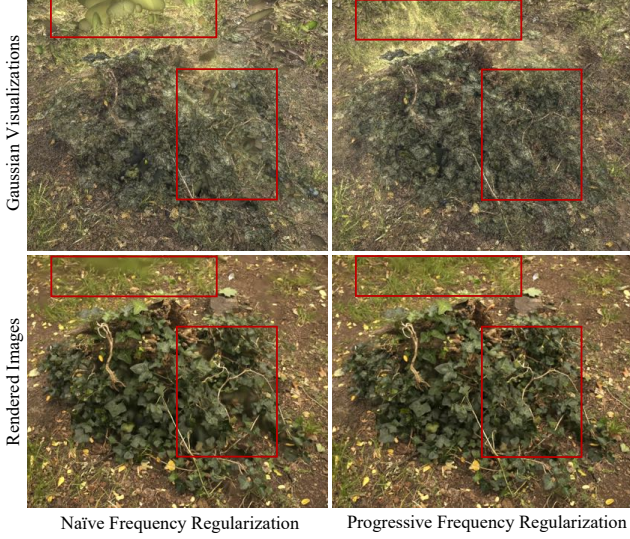


Figure 4. **The comparison of different frequency regularizations.** The naive frequency regularization directly employs amplitude and phase discrepancies without distinguishing between low and high frequency. The proposed progressive frequency regularization introduces frequency annealing technique to achieve low-to-high frequency regularization for coarse-to-fine Gaussian densification. It can be observed that the proposed progressive frequency regularization can achieve finer Gaussian densification and superior novel view synthesis. **Zoom in for best view.**

I can be obtained with the Euclidean metric. Besides, we compute the amplitude and phase of all frequency components to assess the disparities accurately, which are then averaged to derive the final discrepancies as follows:

$$d_a = \frac{1}{\sqrt{HW}} \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} \left| |F(u, v)| - |\hat{F}(u, v)| \right| \quad (5)$$

$$d_p = \frac{1}{\sqrt{HW}} \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} \left| \angle F(u, v) - \angle \hat{F}(u, v) \right|, \quad (6)$$

where $F(u, v)$ and $\hat{F}(u, v)$ denote the complex frequency value of I and \hat{I} , respectively.

3.3. Frequency Annealing

Though naively adopting the amplitude and phase discrepancies (without distinguishing between low and high frequency) as the frequency regularization can mitigate over-reconstruction to some extent, it still suffers from restricted Gaussian densification and significantly biases 3D Gaussian splatting towards undesirable artifacts (as shown in Fig. 4). As low and high frequency relates to large-scale features (e.g., global patterns and structures) and small-scale features (e.g., local details), respectively, we design frequency

annealing technique to perform progressive frequency regularization, which gradually leverages low-to-high frequency to achieve coarse-to-fine Gaussian densification. With frequency annealing technique, superior Gaussian densification can be achieved as shown in Fig. 4.

Specifically, to achieve frequency annealing, we incorporate the low-pass filter H_l and dynamic high-pass filter H_h in Fourier space to extract low and high frequency (denoted as $LF(u, v)$ and $HF(u, v)$), respectively.

$$LF(u, v) = F(u, v)H_l(u, v) \quad (7)$$

$$HF(u, v) = F(u, v)H_h(u, v). \quad (8)$$

The corresponding amplitude and phase discrepancies for low and high frequency can then be formulated as follows:

$$d_{la} = \frac{1}{\sqrt{HW}} \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} \left| |LF(u, v)| - |\hat{L}F(u, v)| \right| \quad (9)$$

$$d_{lp} = \frac{1}{\sqrt{HW}} \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} \left| \angle LF(u, v) - \angle \hat{L}F(u, v) \right| \quad (10)$$

$$d_{ha} = \frac{1}{\sqrt{HW}} \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} \left| |HF(u, v)| - |\hat{H}F(u, v)| \right| \quad (11)$$

$$d_{hp} = \frac{1}{\sqrt{HW}} \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} \left| \angle HF(u, v) - \angle \hat{H}F(u, v) \right| \quad (12)$$

where d_{la} , d_{lp} , d_{ha} and d_{hp} represent low-frequency amplitude discrepancy, low-frequency phase discrepancy, dynamic high-frequency amplitude discrepancy and dynamic high-frequency phase discrepancy, respectively.

For the progressive frequency regularization with the frequency annealing, we initiate it by regularizing the low-frequency discrepancies and then gradually integrate the high-frequency components as the training progresses. The gradual incorporation of high frequency can be achieved with the dynamic high-pass filter H_h , where the frequency band range D_t allowed to pass at the t -th ($t \in [T_0, T]$) iteration can be expressed by:

$$D_0 < D_t < \frac{(t - T_0)(D - D_0)}{T - T_0} + D_0, \quad (13)$$

where D_0 and D denote the maximum range allowed by the low-pass filter and the maximum range of frequency spectrum, respectively. Note that we take the center point $(H/2, W/2)$ as the coordinate origin. t , T_0 and T represent the current iteration, the starting and end iterations of introducing high-frequency components, respectively. Regularization applied to low-to-high frequency results in coarse-to-fine Gaussian densification. The progressive frequency

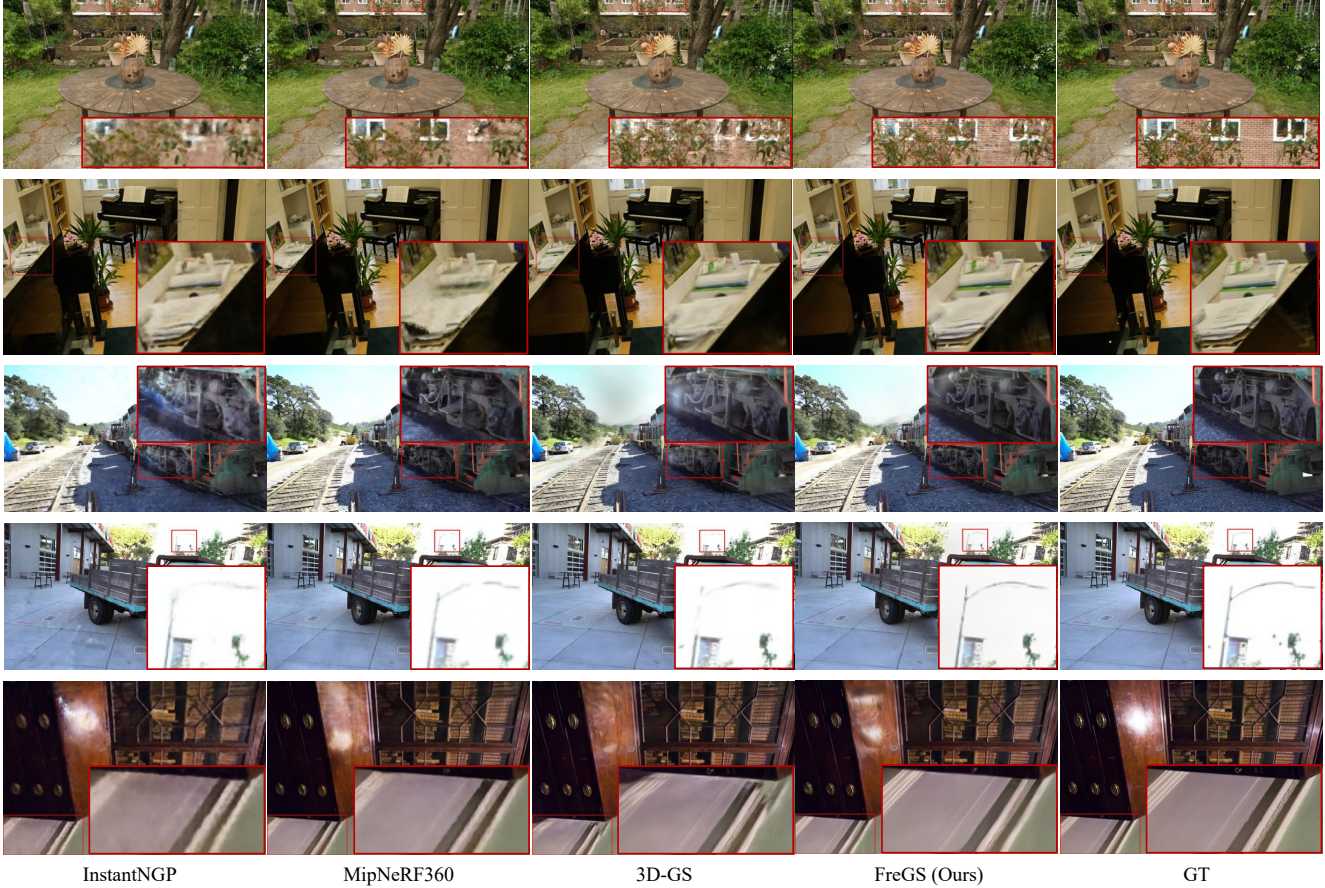


Figure 5. Qualitative comparisons of FreGS with three state-of-the-art methods in novel view synthesis. *Note that for fair comparison as well as trade-off balance between synthesis quality and memory consumption, we train FreGS with similar number of Gaussians as 3D-GS for these datasets (details in Sec.4.2).* The comparisons are conducted over multiple indoor and outdoor scenes including ‘Garden’ and ‘Room’ from Mip-NeRF360, ‘Train’ and ‘Truck’ from Tank&Temple, and ‘DrJohnson’ from Deep Blending. ‘GT’ denotes the ground-truth images. FreGS achieves superior image rendering with much less artifacts but more fine details.

regularization \mathcal{L}_f can be formulated as follows:

$$\mathcal{L}_f = \begin{cases} w_l(d_{la} + d_{lp}), 0 < t \leq T_0 \\ w_l(d_{la} + d_{lp}) + w_h(d_{ha} + d_{hp}), t > T_0, \end{cases} \quad (14)$$

where w_l and w_h represent the training weights for low frequency and high frequency, respectively.

4. Experiments

4.1. Datasets and Implementation Details

Datasets For training and testing, we follow the dataset setting of 3D-GS [16] and conduct experiments on images of a total of 11 real scenes. Specifically, we evaluate FreGS on all nine scenes of Mip-NeRF360 dataset [2] and two scenes from the Tanks&Temples dataset [18]. The selected scenes exhibit diverse styles, ranging from bounded indoor environments to unbounded outdoor ones. To divide the

datasets into training and test sets, we follow 3D-GS and allocate every 8th photo to the test set. The resolution of all involved images is the same as in 3D-GS as well.

Implementation For progressive frequency regularization, we initiate it with low-frequency amplitude and phase discrepancies and then extend the regularization to progressively encompass high-frequency amplitude and phase discrepancies for fine Gaussian densification. Note, the pixel-level L1 loss in the spatial space plus the D-SSIM term is used in the whole training process, which complements the proposed progressive frequency regularization in the frequency space. We stop the Gaussian densification after the 15000th iteration as in 3D-GS [16]. Note that the frequency regularization terminates once Gaussian densification ends. For stable optimization, we start the optimization by working with an image resolution that is four times smaller than

Datasets	Mip-NeRF360			Tanks&Temples			Deep Blending		
Methods	SSIM [↑]	PSNR [↑]	LPIPS [↓]	SSIM [↑]	PSNR [↑]	LPIPS [↓]	SSIM [↑]	PSNR [↑]	LPIPS [↓]
Plenoxels	0.626	23.08	0.463	0.719	21.08	0.379	0.795	23.06	0.510
INGP-Base	0.671	25.30	0.371	0.723	21.72	0.330	0.797	23.62	0.423
INGP-Big	0.699	25.59	0.331	0.745	21.92	0.305	0.817	24.96	0.390
Mip-NeRF360	0.792	27.69	0.237	0.759	22.22	0.257	0.901	29.40	0.245
3D-GS	0.815	27.21	0.214	0.841	23.14	0.183	0.903	29.41	0.243
FreGS(Ours)	0.826	27.85	0.209	0.849	23.96	0.178	0.904	29.93	0.240

Table 1. Quantitative comparisons on the dataset Mip-NeRF360, Tank&Temple and Deep Blending. *Note that for fair comparison as well as trade-off balance between synthesis quality and memory consumption, we train FreGS with similar number of Gaussians as 3D-GS for these datasets (details in Sec.4.2).* All methods are trained with the same training data. INGP-Base and INGP-Big refer to the InstantNGP [22] with a basic configuration and a slightly larger network [22], respectively. Best score, second best score and third best score are in red, orange and yellow respectively.

Datasets	Mip-NeRF360			Tank&Temple		
Methods	PSNR [↑]	SSIM [↑]	LPIPS [↓]	PSNR [↑]	SSIM [↑]	LPIPS [↓]
Base	27.21	0.815	0.214	23.14	0.841	0.183
Base+FR	27.63	0.818	0.213	23.76	0.844	0.181
Base+FR+FA	27.85	0.826	0.209	23.96	0.849	0.178

Table 2. Ablation studies of the proposed FreGS on the datasets Mip-NeRF360 and Tank&Temple. The baseline *Base* adopts pixel-level L1 loss and the D-SSIM term for 3D Gaussian splatting in spatial space. Our *Base+FR* incorporates frequency regularization (FR) to address the over-reconstruction in the frequency space. The *Base+FR+FA* (i.e., FreGS) further introduces our proposed frequency annealing technique (FA) to achieve progressive frequency regularization. *Note that for fair comparison, we train Base+FR+FA with similar number of Gaussians as Base by increasing the gradient threshold. Besides, Base+FR and Base+FR+FA have the same gradient threshold.*

the original images as in 3D-GS. After 500 iterations, we increase the image resolution to the original size by upsampling. We adopt Adam optimizer [17] to train the FreGS and use the Pytorch framework [25] for implementation. For the rasterization, we keep the custom CUDA kernels used in 3D-GS.

4.2. Comparisons with the State-of-the-Art

We compare FreGS with 3D-GS [16] as well as other four NeRF-based methods [2, 7, 22] over various scenes in datasets Mip-NeRF360 and Tank&Temple. **For fair comparison as well as the trade-off balance between memory and performance, we train FreGS with a similar number of Gaussians as 3D-GS for these datasets, which is achieved by introducing progressive frequency regularization while increasing the gradient threshold.** All compared methods are trained with the same training data and hardware. Table 1 shows experimental results over the same test images as described in Section 4.1. We can observe that FreGS outperforms the state-of-the-art 3D-GS consistently

in PSNR, SSIM and LPIPS across all real scenes. The superior performance is largely attributed to our proposed progressive frequency regularization which alleviates the over-reconstruction issue of Gaussians and improves the Gaussian densification effectively. In addition, FreGS surpasses Mip-NeRF360, INGP-Base, INGP-Big, and Plenoxels by significant margins in terms of the image rendering quality. As Fig. 5 shows, FreGS achieves superior novel view synthesis with less artifacts and finer details.

4.3. Ablation Studies

We conduct ablation experiments to examine the proposed frequency regularization that mitigates the over-reconstruction in the frequency space, as well as the frequency annealing that achieves progressive frequency regularization. More details are described in the following two subsections.

Frequency Regularization We first examine how our proposed frequency regularization affects PSNR, SSIM and LPIPS in Gaussian splatting. First, we train a baseline model *Base* that performs Gaussian splatting with a L1 loss in the spatial space. On top of the *Base*, we train a model *Base+FR* that incorporates our proposed frequency regularization in the frequency space. As Table 2 shows, *Base+FR* outperforms the *Base* clearly in PSNR, SSIM and LPIPS, indicating the critical role of frequency regularization in 3D Gaussian splatting and novel view synthesis from a frequency perspective.

Frequency Annealing We then examine how our proposed frequency annealing affects Gaussian splatting. For that, we train a model *Base+FR+FA* (i.e., FreGS) that incorporates the frequency annealing on top of *Base+FR* for progressive frequency regularization. As Table 2 shows, *Base+FR+FA* outperforms *Base+FR* clearly in novel view synthesis, demonstrating the effectiveness of the frequency annealing on progressive frequency regularization.

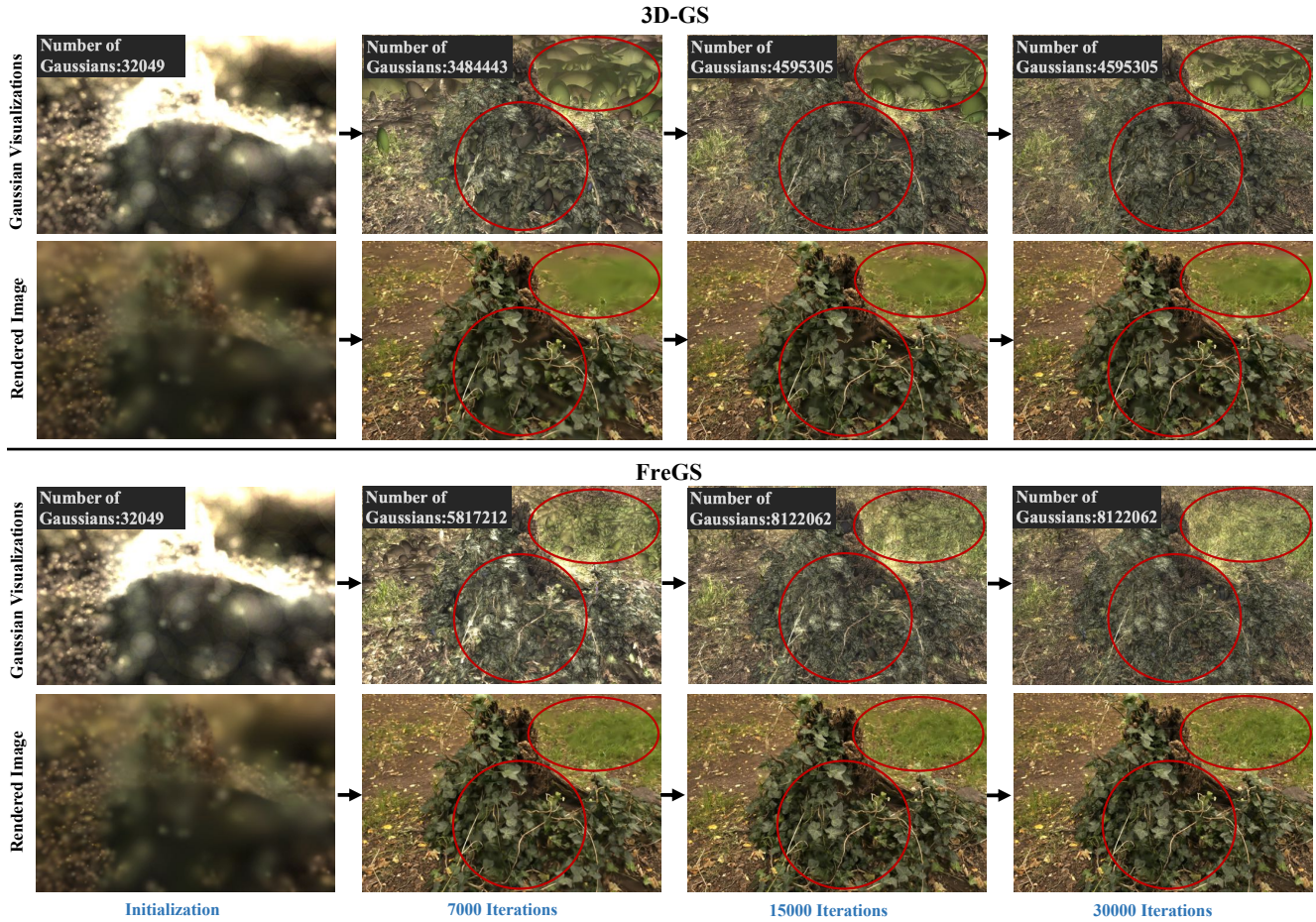


Figure 6. Visualization of Gaussian densification and rendered images along the training process. The proposed FreGS improves the Gaussian densification for over-reconstruction regions progressively and the rendered images also improve progressively with less artifacts and finer details. We visualize both 3D-GS and FreGS at 7000, 15000, and the final 30000 training iterations for comparison. Note that the number of Gaussians does not change any more after the 15,000th iteration due to the termination of Gaussian densification.

4.4. Visualizations

We visualize the Gaussian densification along the training process. As Fig. 6 shows, our FreGS produces more Gaussians and obtains clearly better Gaussian densification for over-reconstruction regions and superior novel view synthesis. Note that the number of Gaussians does not change any more after the 15,000th iteration due to the termination of Gaussian densification. The visualization verifies that the proposed progressive frequency regularization mitigates the over-reconstruction of Gaussians and improves Gaussian densification effectively.

5. Conclusion

This paper presents FreGS, an innovative 3D Gaussian splatting that explores progressive frequency regularization to boost 3D Gaussian splatting from a frequency per-

spective. Specifically, we design a frequency annealing technique for progressive frequency regularization, which performs coarse-to-fine Gaussian densification by progressively leveraging low-to-high frequency components that can be easily extracted with low-pass and high-pass filters in Fourier space. By minimizing the discrepancy between the frequency spectrum of rendered images and the corresponding ground truth, FreGS mitigates the over-reconstruction issue and achieves superior Gaussian densification. Experiments over multiple widely adopted indoor and outdoor scenes show that FreGS achieves superior novel view synthesis and outperforms the state-of-the-art consistently.

6. Acknowledgements

This project is funded by the Ministry of Education Singapore, under the Tier-2 project scheme with a project number MOE-T2EP20220-0003.

References

- [1] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5855–5864, 2021. 1
- [2] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479, 2022. 1, 2, 6, 7
- [3] Wenjing Bian, Zirui Wang, Kejie Li, Jia-Wang Bian, and Victor Adrian Prisacariu. Nope-nerf: Optimising neural radiance field with no pose prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4160–4169, 2023. 2
- [4] Anpei Chen, Zexiang Xu, Fuqiang Zhao, Xiaoshuai Zhang, Fanbo Xiang, Jingyi Yu, and Hao Su. Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14124–14133, 2021. 2
- [5] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *European Conference on Computer Vision*, pages 333–350. Springer, 2022. 2
- [6] Yilun Du, Yinan Zhang, Hong-Xing Yu, Joshua B Tenenbaum, and Jiayun Wu. Neural radiance flow for 4d view synthesis and video processing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14324–14334, 2021. 2
- [7] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5501–5510, 2022. 2, 7
- [8] Chen Gao, Ayush Saraf, Johannes Kopf, and Jia-Bin Huang. Dynamic view synthesis from dynamic monocular video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5712–5721, 2021. 2
- [9] Stephan J Garbin, Marek Kowalski, Matthew Johnson, Jamie Shotton, and Julien Valentin. Fastnerf: High-fidelity neural rendering at 200fps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14346–14355, 2021. 2
- [10] Yudong Guo, Keyu Chen, Sen Liang, Yong-Jin Liu, Hujun Bao, and Juyong Zhang. Ad-nerf: Audio driven neural radiance fields for talking head synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5784–5794, 2021. 2
- [11] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 3
- [12] Peter Hedman, Julien Philip, True Price, Jan-Michael Frahm, George Drettakis, and Gabriel Brostow. Deep blending for free-viewpoint image-based rendering. *ACM Transactions on Graphics (ToG)*, 37(6):1–15, 2018. 2
- [13] Peter Hedman, Pratul P Srinivasan, Ben Mildenhall, Jonathan T Barron, and Paul Debevec. Baking neural radiance fields for real-time view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5875–5884, 2021. 2
- [14] Philipp Henzler, Niloy J Mitra, and Tobias Ritschel. Escaping plato’s cave: 3d shape from adversarial rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9984–9993, 2019. 2
- [15] Ajay Jain, Matthew Tancik, and Pieter Abbeel. Putting nerf on a diet: Semantically consistent few-shot view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5885–5894, 2021. 2
- [16] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (ToG)*, 42(4):1–14, 2023. 2, 3, 4, 6, 7
- [17] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7
- [18] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (ToG)*, 36(4):1–13, 2017. 6
- [19] Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. Barf: Bundle-adjusting neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5741–5751, 2021. 2, 3
- [20] Quan Meng, Anpei Chen, Haimin Luo, Minye Wu, Hao Su, Lan Xu, Xuming He, and Jingyi Yu. Gnerf: Gan-based neural radiance field without posed camera. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6351–6361, 2021. 2
- [21] 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 *European conference on computer vision*, pages 405–421. Springer, 2020. 1, 2, 3
- [22] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *arXiv preprint arXiv:2201.05989*, 2022. 2, 7
- [23] Michael Niemeyer, Jonathan T Barron, Ben Mildenhall, Mehdi SM Sajjadi, Andreas Geiger, and Noha Radwan. Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5480–5490, 2022. 2
- [24] 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 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5865–5874, 2021. 2, 3
- [25] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Ad-*

- vances in neural information processing systems*, 32, 2019. 7
- [26] Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14335–14345, 2021. 2
- [27] Gernot Riegler and Vladlen Koltun. Free view synthesis. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIX 16*, pages 623–640. Springer, 2020. 2
- [28] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. 3
- [29] Vincent Sitzmann, Justus Thies, Felix Heide, Matthias Nießner, Gordon Wetzstein, and Michael Zollhofer. Deepvoxels: Learning persistent 3d feature embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2437–2446, 2019. 2
- [30] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020. 3
- [31] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems*, 33:7537–7547, 2020. 3
- [32] Justus Thies, Michael Zollhöfer, and Matthias Nießner. Deferred neural rendering: Image synthesis using neural textures. *Acm Transactions on Graphics (TOG)*, 38(4):1–12, 2019. 2
- [33] Edgar Tretschk, Ayush Tewari, Vladislav Golyanik, Michael Zollhöfer, Christoph Lassner, and Christian Theobalt. Non-rigid neural radiance fields: Reconstruction and novel view synthesis of a dynamic scene from monocular video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12959–12970, 2021. 2
- [34] Yiqun Wang, Ivan Skorokhodov, and Peter Wonka. Hf-neus: Improved surface reconstruction using high-frequency details. *Advances in Neural Information Processing Systems*, 35:1966–1978, 2022. 3
- [35] Muyu Xu, Fangneng Zhan, Jiahui Zhang, Yingchen Yu, Xiaoqin Zhang, Christian Theobalt, Ling Shao, and Shijian Lu. Wavenerf: Wavelet-based generalizable neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 18195–18204, 2023. 3
- [36] Jiawei Yang, Marco Pavone, and Yue Wang. Freenerf: Improving few-shot neural rendering with free frequency regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8254–8263, 2023. 2, 3
- [37] Fangneng Zhan, Lingjie Liu, Adam Kortylewski, and Christian Theobalt. General neural gauge fields. In *The Eleventh International Conference on Learning Representations*, 2023. 3
- [38] Jiahui Zhang, Fangneng Zhan, Rongliang Wu, Yingchen Yu, Wenqing Zhang, Bai Song, Xiaoqin Zhang, and Shijian Lu. Vmrf: View matching neural radiance fields. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 6579–6587, 2022. 2
- [39] Jiahui Zhang, Fangneng Zhan, Yingchen Yu, Kunhao Liu, Rongliang Wu, Xiaoqin Zhang, Ling Shao, and Shijian Lu. Pose-free neural radiance fields via implicit pose regularization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3534–3543, 2023. 2