

SUND_{AE}: Spectrally Pruned Gaussian Fields with Neural Compensation

Runyi Yang
AIR, Tsinghua University
Imperial College London
runyi.yang23@imperial.ac.uk

Zhenxin Zhu
AIR, Tsinghua University
Beihang University

Zhou Jiang
AIR, Tsinghua University
Beijing Institute of Technology

Baijun Ye
AIR, Tsinghua University
Beijing Institute of Technology

Xiaoxue Chen
AIR, Tsinghua University

Yifei Zhang
AIR, Tsinghua University
UCAS

Yuantao Chen
AIR, Tsinghua University
CUHK (SZ)

Jian Zhao[†]
EVOL Lab, Institute of AI (TeleAI),
China Telecom

Hao Zhao[†]
AIR, Tsinghua University
zhaohao@air.tsinghua.edu.cn

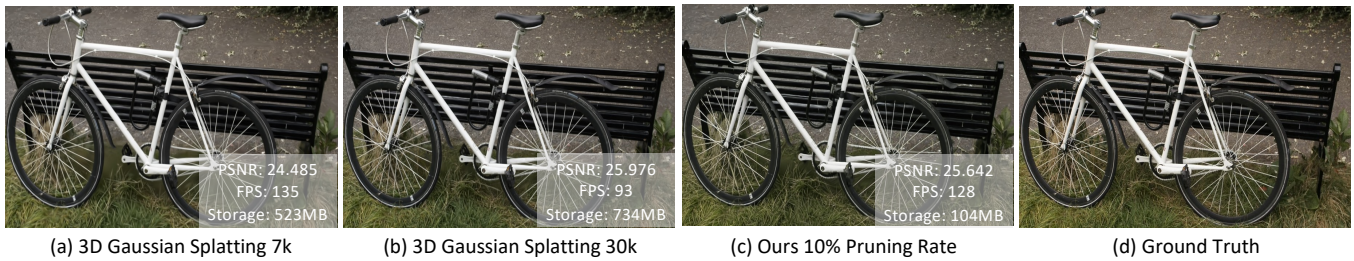


Figure 1: (a) 3D Gaussian splatting (3DGS) results trained for 7K iterations. (b) 3DGS results trained for 30K iterations, in which more Gaussian primitives are allocated so quality gets higher, speed gets slower, and storage gets larger compared to (a). (c) SUND_{AE} results by pruning 90% primitives upon (a), being much smaller in storage, more accurate and a little bit slower than (a). Note that the storage usage is not 10% of (a) because an additional neural compensation head is used.

ABSTRACT

Recently, 3D Gaussian Splatting, as a novel 3D representation, has garnered attention for its fast rendering speed and high rendering quality. However, this comes with high memory consumption, e.g., a well-trained Gaussian field may utilize three million Gaussian primitives and over 700 MB of memory. We credit this high memory footprint to the lack of consideration for the **relationship** between primitives. In this paper, we propose a memory-efficient Gaussian field named SUND_{AE} with spectral pruning and neural compensation. On one hand, we construct a graph on the set of Gaussian primitives to model their relationship and design a spectral down-sampling module to prune out primitives while preserving desired signals. On the other hand, to compensate for the quality loss of pruning Gaussians, we exploit a lightweight neural network head to mix splatted features, which effectively compensates for quality losses while capturing the relationship between primitives in its weights. We demonstrate the performance of SUND_{AE} with extensive results. For example, SUND_{AE} can achieve 26.80 PSNR at 145 FPS using 104 MB memory while the vanilla Gaussian splatting

algorithm achieves 25.60 PSNR at 160 FPS using 523 MB memory, on the Mip-NeRF360 dataset. Codes are publicly available at <https://runyiyang.github.io/projects/SUNDAE/>.

CCS CONCEPTS

• **Computing methodologies** → **Computer vision; Computer graphics.**

KEYWORDS

3D Gaussian Splatting, Graph Signal Processing, Neural Rendering

1 INTRODUCTION

Representing 3D scenes has been a longstanding problem in computer vision and graphics, serving as the foundation for various VR/AR [59][43] and robotics [61][72][71] applications. With the rise of the neural radiance field (NeRF) [38], a series of methods have emerged to enhance the quality and efficiency of NeRF [9, 14, 19, 40, 52, 62, 66]. Recently, 3D Gaussian splatting (3DGS) [29][64][51] has been proposed as a novel 3D scene representation, utilizing a set of 3D positions, opacity, anisotropic covariance, and spherical harmonic (SH) coefficients to represent a 3D scene. Compared to neural rendering methods, this technique demonstrates

[†]Corresponding Author

notable advantages in rendering speed, rendering quality, and training time, making it widely used for applications such as 3D editing [13, 53] and digital twins [28, 31].

Although 3D Gaussian splatting (3DGS) offers several advantages over other implicit 3D representations, training a 3DGS model faces a large storage challenge (i.e., checkpoint storage on ROM), as depicted in Fig. 1 (a) and (b). This issue arises from the training process to gradually populate empty areas with Gaussian primitives, so that the rendering results can better fit input images. Compared with recent neural rendering methods [5, 40], 3DGS requires a much larger memory cost for the same scene, which limits the application of 3DGS on mobile platforms and edge computing.

In this paper, our goal is to address the high memory cost of 3D Gaussian splatting. We introduce a memory-efficient method, which achieves low storage usage while maintaining high rendering speed and good quality. As illustrated in Fig. 1 (c), our approach achieves photorealistic rendering quality with much lower storage.

As previously mentioned, training a vanilla 3DGS leads to a large number of Gaussian primitives, some of which are redundant. We credit the redundancy of Gaussian primitives to the fact that these primitives are independent of each other in the 3DGS formulation. In response, we aim to improve the modeling of the **relationship** between Gaussians to reduce primitive redundancy. We achieve this through two complementary techniques: spectrally pruning the primitive graph and incorporating a neural compensation head. Consequently, our method is referred to as spectrally pruned Gaussian fields with neural compensation, abbreviated as **SUNDAE**.

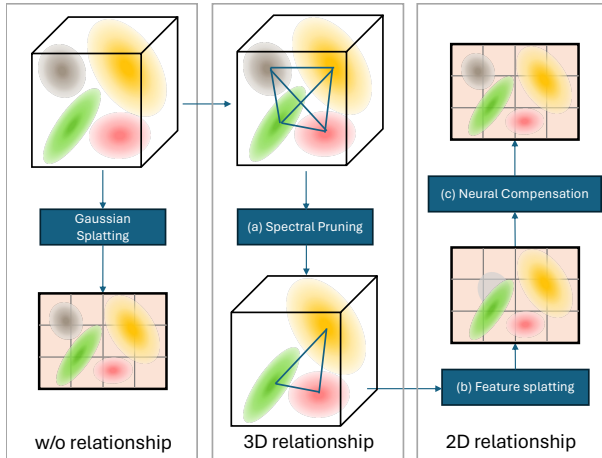


Figure 2: The left panel shows vanilla 3D Gaussian splatting, which requires a large amount of storage as it does not capture the relationship between primitives. The middle panel shows our spectral pruning technique that is based upon the relationship between 3D Gaussians. The right panel shows that the neural compensation head exploits the relationship between 2D feature splatting results to improve rendering.

Spectral Graph Pruning: Gaussian fields utilize a collection of Gaussian primitives as the representation of the scene. As these primitives are irregularly distributed in 3D space, we propose a graph-based data structure, rather than regular structures like grids,

to capture the **relationship** between these primitives. This is conceptually illustrated in the middle panel of Fig. 2. Specifically, We introduce the graph signal processing theory [11] to derive an optimal stochastic sampling strategy that preserves band-limited information on this graph. By controlling the spectrum’s band, we achieve flexible control over pruning ratios. For example, in Fig.1 (c), 90% of Gaussian primitives from Fig.1 (a) are pruned out.

Neural Compensation: After spectral pruning, there is an inevitable decrease in rendering quality. To address this, we employ a neural compensation head to compensate for this quality loss, as conceptually illustrated in the right panel of Fig. 2. We transition from the Gaussian splatting paradigm to a feature splatting paradigm by attaching feature vectors to Gaussian primitives. Subsequently, a lightweight neural network is introduced to predict RGB values on the splatted feature images, thereby integrating information from different primitives. This allows the **relationship** between primitives to be captured in the weights of the compensation network, indirectly in the 2D domain of splatted features.

In summary, these two techniques operate in a complementary manner to tackle the absence of primitive **relationship** modeling in 3DGS. The spectral graph pruning technique removes less important primitives on the primitive graph, while the neural compensation technique integrates information from the remaining primitives. Through comprehensive qualitative and quantitative benchmarking on three datasets (Mip-NeRF360, Tanks&Temples, and Deep Blending), we demonstrate that SUNDAE effectively reduces the size of Gaussian fields while preserving good quality and fast rendering speed. Our contributions can be summarized as:

- A newly proposed primitive pruning framework for Gaussian fields based upon the spectrum of primitive graphs;
- A novel feature splatting and mixing module to compensate for the performance drop caused by the pruning;
- State-of-the-art results, in terms of both quality and speed, on various benchmarks with low memory footprint.

2 RELATED WORKS

Conventional 3D Scene Representations. Various 3D representations have been proposed for 3D reconstruction. Point cloud based representation [35, 39, 63] is extensively employed due to its simplicity and effectiveness in depicting 3D scenes. VoxelMap[65] introduced a novel and efficient probabilistic adaptive voxel mapping method within the parameterized plane feature. While most of the voxel-based representation is more suitable for the downstream planning task to avoid the collision. Moreover, mesh-based method [8, 36] offer distinct advantages in representing complex geometries. ElasticFusion [60] builds on the surfel-based method to reconstruct the dense map with color. This representation is more continuous and realistic. They exploit relationship within the 3D representation in various ways, motivating our study.

Efficient Neural Rendering. Vanilla NeRF [38] involves extensive neural network calculations at 3D positions on a pixel-by-pixel basis, which hinders real-time rendering. Recent approaches [9, 14, 19, 40, 52, 62, 66] have focused on improving the efficiency of Neural radiance fields, and the main perspectives include, for example, fast training, fast rendering and low memory footprint. Instant-NGP [40] uses a multiresolution hash table to store learnable



features and additionally uses a small neural network for decoding, enabling fast training. Other methods for accelerated training either use dense voxel grids, such as DVGO [52] which uses a density voxel grid and a feature voxel grid for explicit and discretized volume representation, or sparse voxel grids, such as Plenoxels [19] which uses a sparse 3D grid with spherical harmonics for scene representation. While these approaches significantly accelerate training, they do not render in real-time and necessitate a large consumer GPU for rendering. 3D Gaussian splatting [29] represents the scene using 3D Gaussians and utilizes a fast, differentiable rendering approach to achieve both high-quality novel view synthesis and real-time rendering. Our SUNDAE focuses on integrating two novel techniques to leverage the primitive relationship in Gaussian fields and improve storage efficiency.

Primitive-based Neural Rendering. Different from the Vanilla NeRF that utilizes a fully implicit network for scene representation, or grid-based neural radiance fields [19, 40] that evenly distribute the model parameters into each voxel grid vertex, point-based neural radiance fields [20, 23, 24, 26, 29, 33, 37, 46, 49, 62, 68, 70] introduce explicit graphic primitives into the construction of the neural implicit fields that allows adaptive control on the model expressiveness within limited memory consumption. Specifically, PointNeRF [62] associates neural features with points and performs k-NN interpolation upon arbitrary point query, allowing the network to capture finer scene details by increasing local neural point density. Different from raytracing-based methods, ADOP [46] and 3DGS [29] use rasterization techniques to render novel viewpoints.

Graph Signal Processing. Graph Signal Processing aims to develop tools to process data with an irregular structure, i.e., on a graph domain. The first and most important area in graph signal processing is to design graph representations [42]. Groundbreaking contributions [16] and [15] presented initial samples of designs founded on the vertex and spectral domain properties, respectively. However, both frequency and time domains alone have their drawbacks [42]. Recent research [3, 18, 32, 41, 54, 55, 67] are gradually moving their focus on the development of critically sampled filterbanks that have a vertex-localized implementation as well as a spectral interpretation. Another problem in graph signal processing is sampling graph signals. The key idea is to define a class of graph signals, i.e., bandlimited, and then define conditions to reconstruct a signal in that class. The concepts are first presented in [44]. A sufficient and necessary condition for unique recovery is defined in [1] and is soon generalized to other types of graphs and signals [11, 47, 48]. However, sampling signals on large graphs is a great challenge due to its complexity. Some techniques require computing the first K basis vectors [11] and other work [2] utilizes spectral proxies instead of exact graph frequencies to reduce complexity. Random strategy has also been proposed in [45] which leads to significantly lower complexity but less comparable performance.

Graph signal processing has been used in a wide variety of applications, including sensor networks [12, 17, 21, 57], biology networks [4, 6, 34], data science [7, 27, 58] and image&pointcloud processing [10, 25, 56, 69]. To the best of our knowledge, we are the first to introduce graph signal processing to the pruning of primitive-based neural rendering methods.

3 METHODS

Given a set of images $I_i \in \mathcal{I}$, along with the corresponding camera calibration parameters C_i , 3DGS [29] reconstructs the scene by representing it with Gaussian primitives yet with a high storage usage. We credit this weakness to the fact that the relationship between Gaussians are not exploited in 3DGS. We introduce SUNDAE, a memory-efficient Gaussian field, featuring a graph-based pruning approach and a neural compensation module. The overall framework is depicted in Fig. 3. We begin by fitting 3D Gaussians as a “warm-up” process in Section 3.1. To reduce memory usage, we introduce a one-step graph signal processing approach in Section 3.2. This involves constructing a Gaussian primitive graph to model the relationship between primitives and utilizing a band-limit graph filter to prune redundant Gaussian primitives. To mitigate quality loss caused by pruning, we adopt a neural compensation module in Section 3.3, which restores the RGB image \hat{I}_i from the neural image F_i using a lightweight neural network. Finally, we introduced a continuous pruning as an alternative strategy in Sec. 3.4.

3.1 3D Gaussian Splatting Warm Up

We utilize the vanilla 3D Gaussian Splatting [29] as the initial step for generating a dense representation of a scene using Gaussian primitives. This step starts by using a sparse point cloud to initialize Gaussian centers. Then, an efficient densification strategy is exploited to increase the number of these primitives. Additionally, the rasterization process involves splatting the 3D Gaussian primitives onto the 2D plane to reconstruct input images.

3.1.1 Gaussian Primitive Initialization. The point cloud P_c obtained through Structure-from-Motion (SfM) [50] serves as the initial input for representing the 3D scene using Gaussian primitives P . We turn the 3D coordinates $x \in P_c$ into Gaussian primitives $p \in P$, as described by the following equation:

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

where the Σ is defined as 3D covariance matrix in the world space. To ensure the positive semi-definiteness and to uphold the physical interpretation of the covariance matrix, an ellipsoid configuration is used to represent the 3D Gaussian covariance. The decomposition of Σ is achieved using a scaling matrix S and a rotation matrix R , as expressed in the equation:

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

This representation of anisotropic covariance is particularly advantageous for optimization processes. The generated Gaussian primitives are characterized by positions x , opacity α , and a covariance matrix Σ . Additionally, the chromatic attributes of Gaussian primitives are encoded by spherical harmonic (SH) coefficients.

3.1.2 Gaussian Primitive Densification. During the training process, all parameters of Gaussian primitives are optimized, and a densification strategy is integrated to improve representation power. Initially, Gaussian primitives exhibit a sparsity level similar to that of the point cloud generated via SfM, which is not enough for representing detailed parts of the scene, such as grass and trees in outdoor scenes. So Gaussian primitives with extensive covariance, which tend to oversimplify the geometric intricacies of detailed

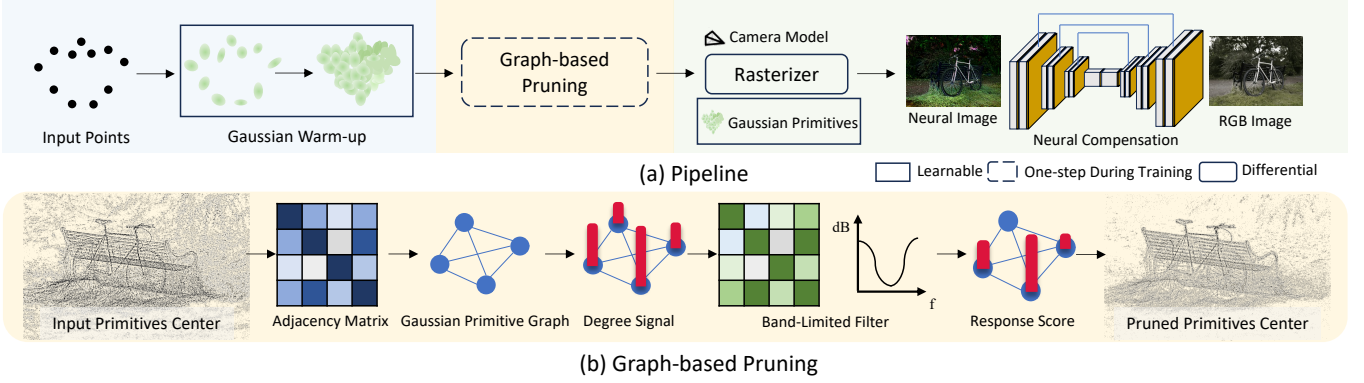


Figure 3: (a) Pipeline: Our proposed method warms up a 3D Gaussian field firstly, followed by a Graph-based pruning strategy to down-sample the Gaussian primitives, and a convolutional neural network to compensate the losses caused by pruning. **(b) Graph-based Pruning:** A graph based on the spatial relationship between the Gaussian primitives, is utilized for pruning post warm-up. Employing a band-limited graph filter, this process facilitates the extraction of fine details from high-frequency components, alongside capturing general features from low-frequency parts, thereby enabling a comprehensive and efficient representation of the entire scene.

scene segments, are subdivided into smaller Gaussians. Meanwhile, those with minimal covariance, indicating under-representation, are duplicated to enhance coverage. We adopt the densification parameters from [29] to increase the density of Gaussians and achieve a high-quality warm-up for the subsequent pruning process.

3.2 Spectral Graph Pruning

After warm-up, a dense representation using Gaussian primitives incurs significant storage consumption, for example, approximately 1.33GB for the Bicycle scene in MipNeRF360 [5] dataset. We credit this inefficiency to the redundancy in primitives. However, determining which primitives are redundant is challenging without establishing **relationships** between them. Therefore, we introduce the graph signal processing theory and construct a graph based on Gaussian primitives to efficiently prune redundant primitives.

3.2.1 Graph Signal Processing Preliminaries. Graph shift. We denote a weighted graph by $\mathcal{G} = (\mathcal{V}, A)$, where \mathcal{V} is the set of nodes $\{v_0, v_1, \dots, v_{N-1}\}$, $N = |\mathcal{V}|$ and $A \in \mathbb{C}^{N \times N}$ is the graph shift, or the weighted adjacency matrix. Graph shift reflects the connection between the nodes using edge weight which is a quantitative description of primitive relationship. When a graph shift acts on a graph signal, it can represent the diffusion of the graph signal. In this work, we build the graph shift to model the spatial relationship among Gaussian primitives. A graph shift is usually normalized for proper scaling, ensuring $\|A\| = 1$.

Graph signal. Given a graph $G = (\mathcal{V}, A)$, a graph signal on this graph can be seen as a map assigning each node v_i with a value $x_i \in \mathbb{C}$. If the order of the nodes is fixed, then the graph signal is defined as a N dimensional vector $x = (x_1, x_2, \dots, x_N)$. In this work, the input graph node is Gaussian primitives central position $x \in \mathbb{R}^3$, the graph signal is the Euclidean distance between primitives.

Graph Fourier Transform. A Fourier transform corresponds to the expansion of a signal using a set of bases. When performing graph Fourier transform, the bases are the eigenbasis of the graph

shift. For simplicity, assume that A has a complete eigenbasis and the spectral decomposition of A is $A = V\Lambda V^{-1}$ [47], where the eigenbasis of A form the columns of matrix V , and $\Lambda \in \mathbb{C}^{N \times N}$ is the diagonal matrix of eigenvalues.

3.2.2 Graph Construction. Given a set of Gaussian Primitives P , we want to construct a nearest neighbor graph. The adjacent matrix W of the graph is defined as:

$$W_{ij} = \begin{cases} \exp(-\frac{\|x_i - x_j\|_2^2}{2 * \sigma^2}), & \|x_i - x_j\|_2^2 < \tau \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where x_i and x_j are central points in P , τ is a hyperparameter, chosen as ten times of the minimum nearest neighborhood distance between primitives experimentally, and σ is the variance of the distance matrix. Equation 3 demonstrates that when the Euclidean distance of two Gaussian primitives is smaller than a threshold τ , the two primitives are connected by the graph edge, whose weight corresponds to geometric information between two primitives in Gaussian fields. A weighted degree matrix D is a diagonal matrix $D_{i,i} = \sum_j W_{i,j}$, reflecting the density around the i th primitive.

3.2.3 Graph Filtering and Sampling. We propose a band-limited graph filter, combined with a high-frequency filter and low-frequency filter (conceptually shown in Fig. 3), to catch the detailed information and general information of the scene. Specifically, the input graph signal x represents the central of Gaussian primitives.

A simple design of a high-pass filter is a Haar-like one:

$$\mathcal{H}_H = I - A = V(I - \Lambda)V^{-1}, \quad (4)$$

where A is the graph shift and V and Λ are the corresponding eigenvectors and eigenvalues in diagonal form. Denote that all eigenvalues are $\lambda_i, i \in 0, 1, \dots, N-1$. We order λ_i in descending order, thus we have $1 - \lambda_i \leq 1 - \lambda_{i+1}$ and $\lambda_0 = 1$. This indicates low-frequency response attenuates and high-frequency response amplifies [10]. Similarly, a Haar-like low-pass graph filter is:



$$\mathcal{H}_L = I + \frac{A}{\lambda_0} = V(I + \Lambda/\lambda_0)V^{-1}. \quad (5)$$

Then we have the response of the input signal x corresponding to filters and the response magnitude could be written as:

$$\pi_i = \|f_i\|_2^2. \quad (6)$$

Controlling bandwidth with γ . In implementation, we prune the abundant primitives according to the response magnitude of the high-pass filter. We sample total $k\%$ of all primitives, among where γ high-frequency primitives and $(1 - \gamma)$ low-frequency primitives by querying the top γ highest magnitude and the top $(1 - \gamma)\%$ lowest magnitude respectively. The value of γ is ablated in Section 4.4 and we used $\gamma = 0.5$ in main experiments to maintain consistency.

3.3 Neural Compensation

Although our graph-based pruning effectively removes unnecessary primitives while retaining important ones, there is inevitably a decrease in rendering quality for large pruning ratio. To address this, we employ a neural compensation network to model the relationship between primitives in the 2D domain.

To allow neural compensation after rasterization, we need to render the 3D Gaussian primitives into neural images in a differentiable manner. Specifically, we leverage the differentiable 3D Gaussian renderer from 3DGS [29] and switch from RGB rendering to feature rendering. The center of the Gaussian primitive is projected using the standard point rendering method:

$$x_{\text{img}} = K_c((T_c x)/(T_c x)_z), \quad (7)$$

where K_c and T_c are the intrinsic and extrinsic parameters of camera C , and x_{img} indicates the pixel coordinates in neural image. The covariance Σ_f in neural image space could be formulated as:

$$\Sigma_f = J T_c \Sigma T_c^T J^T, \quad (8)$$

where the J is the approximated Jacobian of the projective transformation. The neural image is computed by:

$$f = \sum_i c_i \beta_i \prod_{j=1}^{i-1} (1 - \beta_j), \quad (9)$$

where c_i is a feature vector instead of SHs in the original 3DGS [29] and β is the result of neural image covariance Σ_f multiplied with the opacity α of the Gaussian primitive.

In this manner, instead of straightforwardly rendering the RGB image like 3DGS, we obtain a neural image through the differentiable rasterizer for 3D Gaussians, which projects feature of 3D Gaussian primitives to 2D neural image F . Then, we utilize a light-weight neural network Φ to compensate for the quality drop post spectral pruning. This network Φ consists of a four-layer fully convolutional U-Net with skip-connections, which aggregates information from different primitives. Downsampling is performed using average pooling, and the images are upsampled using bilinear interpolation. The network takes the rasterized neural images as input and outputs the RGB images.

$$\hat{I} = \Phi(F) \quad (10)$$

The overall optimization is based on the difference between the rendered images and ground truth images in the dataset. The

compensation network and 3D Gaussian primitives are optimized simultaneously during training. The loss function is a combination of \mathcal{L}_1 and a D-SSIM loss:

$$\begin{aligned} \mathcal{L} &= \lambda_1 \mathcal{L}_1 + \lambda_2 \mathcal{L}_{\text{SSIM}} \\ &= \lambda_1 |\hat{I} - I| + \lambda_2 (1 - \text{SSIM}(\hat{I}, I)) \end{aligned} \quad (11)$$

3.4 Continuous Pruning as a Strategy

In addition to the training-then-pruning strategy described in Sec. 3.2, we further explore a strategy that integrates continuous pruning into training. Unlike training-then-pruning, which prunes out primitives from a fully densified Gaussian field, continuous pruning involves periodically removing a specific number or percentage of primitives at pre-defined intervals throughout the training process. This approach aims to consistently control the maximum number of primitives while training the 3D Gaussian field, thereby lowering **peak memory** requirements during training and allowing training on GPU devices with low GPU memory.

Empirically, the advantage of lower peak memory comes at the cost of weaker control of final memory footprint. For instance, if we prune 20% of the primitives every 2000 iterations, the final converged state of the 3D Gaussian field might deviate from the expected 20% reduction. Additionally, this variance can differ across different scenes, complicating the predictability and consistency of the pruning effects. Therefore, we treat the continuous pruning strategy as an alternative when needed.

4 EXPERIMENTS

4.1 Implementation Details

We implement SUNDAE in Python using the PyTorch framework. The differential rasterizer component is based on the CUDA code-base, as delineated in 3DGS [29]. Additionally, for the construction of graphs and the pruning of primitives, we employed a C++ implementation, accelerated with the Numba library for computational efficiency. All training and testing processes were conducted on a single NVIDIA RTX 3090.

Our algorithm was rigorously tested across three publicly available datasets. Specifically, we evaluated its performance on all seven scenes of the MipNeRF360 dataset, as presented in [5]. Moreover, SUNDAE was tested on two scenes from the Tanks & Temples [30] and two scenes from the Deep Blending [22], adhering to the benchmark criteria established in [29]. The selected scenes encompass a diverse range of capture styles, including both bounded indoor settings and large unbounded outdoor environments.

4.2 Quantitative Results

Evaluation Metrics. In our evaluation, we employ the most widely recognized metrics in the field, PSNR, SSIM, and LPIPS. Additionally, the Frames Per Second (FPS) performance of SUNDAE was assessed by the average rendering time computed across all scenes. Memory was evaluated by calculating the average size of checkpoints in each of the scenes, which is the sum of 3D Gaussian fields and neural compensation module. We adopt the train-test split for the MipNeRF360 dataset suggested by [5], using every 8th photo for the test, and for the rest two datasets, we followed the setting of [29]. The tested results are shown in Tab. 1.

Dataset Method/Metric	Mip-NeRF360					Tanks&Temples					Deep Blending				
	PSNR↑	SSIM↑	LPIPS↓	FPS	Mem	PSNR↑	SSIM↑	LPIPS↓	FPS	Mem	PSNR↑	SSIM↑	LPIPS↓	FPS	Mem
Plenoxels	23.08	0.626	0.463	6.79	2.1GB	21.08	0.719	0.379	13.0	2.3GB	23.06	0.795	0.510	11.2	2.7GB
INGP-Base	25.30	0.671	0.371	11.7	13MB	21.72	0.723	0.330	17.1	13MB	23.62	0.797	0.423	3.26	13MB
INGP-Big	25.59	0.699	0.331	9.43	48MB	21.92	0.745	0.305	14.4	48MB	24.96	0.817	0.390	2.79	48MB
M-NeRF360	27.69	0.792	0.237	0.06	8.6MB	22.22	0.759	0.257	0.14	8.6MB	29.40	0.901	0.245	0.09	8.6MB
3DGS-7K	25.60	0.770	0.279	160	523MB	21.20	0.767	0.280	197	270MB	27.78	0.875	0.317	172	386MB
3DGS-30K	27.21	0.815	0.214	134	734MB	23.14	0.841	0.183	154	411MB	29.41	0.903	0.243	137	676MB
Ours-1%	24.70	0.716	0.375	171	38MB	20.49	0.703	0.375	127	33MB	26.57	0.861	0.355	165	36MB
Ours-10%	26.80	0.805	0.264	145	104MB	22.50	0.787	0.282	122	64MB	28.65	0.892	0.287	163	86MB
Ours-30%	27.24	0.826	0.228	109	279MB	23.46	0.817	0.242	116	148MB	29.40	0.899	0.248	156	203MB
Ours-50%	27.31	0.827	0.213	88	393MB	23.70	0.830	0.219	92	228MB	28.86	0.900	0.242	155	312MB

Table 1: Quantitative evaluation of our method with different downsampling rate compared to previous work over three datasets.

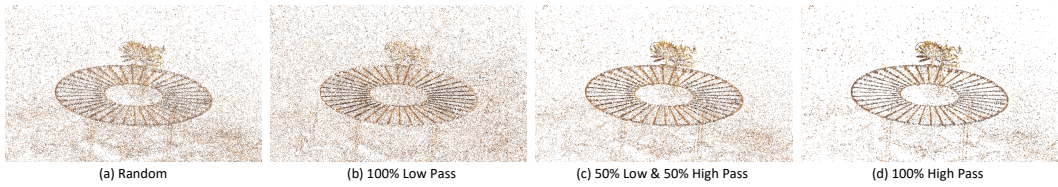


Figure 4: The pruned Gaussian primitive centers with different pruning strategy.

Tradeoff of rendering quality, rendering time and storage. Considering the MipNeRF360 dataset, Mip-NeRF360 method achieved the highest image fidelity with a PSNR of 27.69 and SSIM of 0.792, but had a low FPS of 0.06. Conversely, the 3DGS-7K and 3DGS-30K methods showed high FPS rates of 160 and 134, respectively, though with significant memory requirements (523MB for 7K and 734MB for 30K). Our method, particularly at 30% and 50% sampling rates, struck a balance between quality and efficiency, ranking within the top three in performance metrics while maintaining rapid rendering speeds and manageable memory usage (88 FPS at 393MB for 50%). The SUNDAAE variant (10% and 1%) also displayed remarkable efficiency, achieving a PSNR of 26.80 at 145 FPS with only 104 MB memory, and 24.70 PSNR at 171 FPS using 38 MB. This efficiency suggests that our primitive graph and neural compensation module adeptly model the relationships among primitives in 3D and 2D domains, and our pruning strategy effectively retains the primary information of the scenes. These results indicate that SUNDAAE can represent scenes in a more compact manner.

As presented in Table 1, our method also shows start-of-the-art performance in other datasets by using only around 50% or even 30% of memory. At a very low sampling rate of 1%, our method remains competitive, closely aligning with the performance of established approaches such as Instant-NGP [40] and Plenoxels [19], with minimal compromise in quality. This performance balance highlights the robustness of our spectral pruning and neural compensation techniques in managing Gaussian primitive relationships, thus there’s not much decline of pruning abundant primitives.

4.3 Qualitative Results

The qualitative results can be seen in Fig. 5. We compare the qualitative results of our SUNDAAE of 1% and 10% sampling rate with 3DGS [29] and InstantNGP [40]. The qualitative results show that SUNDAAE could achieve a similar quality of novel view synthesis using only 10% or even 1% memory consumption. The graph could

successfully build the relationship of primitives and the neural compensation head effectively maintains rendering quality. Interestingly, we could see from the 4-th row and the last row of Fig. 5, that the spectral pruning could remove the **floaters** near the camera.

Band visualization. As seen in Fig. 4, we visualize the Gaussian primitive centers post-pruning. Results show that the low-pass part could capture the background or smoothed area more, while the high-pass filter better captures the high-frequency details. The combined filter is proper to capture the high-frequency detailed object-level information as well as reserve the background points.

4.4 Ablation Study

Band-limited ratio of Graph-based pruning. The band-limited filter’s ratio is represented by γ . Specifically, we sample n primitives during graph-based pruning, comprising $n \times \gamma$ high-pass primitives and $n \times (1 - \gamma)$ low-pass primitives, as detailed in Sec. 3.2.3. As demonstrated in Figure 7, we retain 1% of primitives and vary the filter’s ratio γ . The results indicate that γ has a significant impact on the rendering quality. Notably, a γ value of 50% delivers the most favorable outcomes, while a disproportionate emphasis on either low-frequency or high-frequency signals leads to a deterioration in quality. This highlights the advantage of spectral pruning method as it naturally preserves important high-frequency details and low-frequency background using a 50% ratio.

The compensation performance of the network. To verify the effectiveness of our neural compensation module, we conducted experiments with and without neural compensation at different sampling rates. As indicated in Table 3, across all sampling rates, employing neural compensation leads to improved performance compared to not using it. This is further supported by the visualization results depicted in Fig. 6. These findings demonstrate the

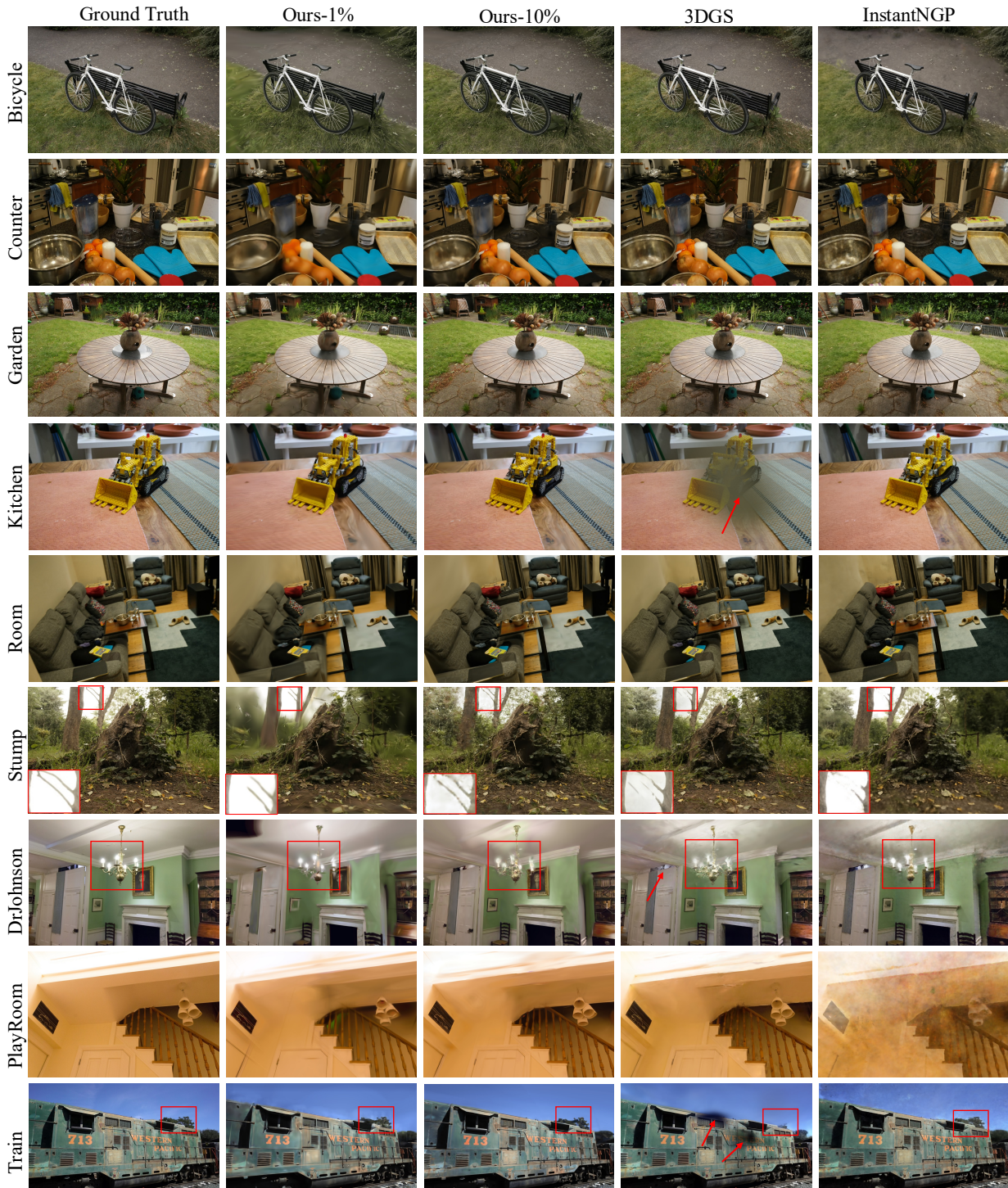


Figure 5: Qualitative results of our novel view synthesis. The scenes are, from the top down: Bicycle, Counter, Garden, Kitchen, Room and Stump from the Mip-NeRF360 dataset; DrJohnson, Playroom from the Deep Blending dataset and Train from Tanks&Temples. Non-obvious differences in quality highlighted by arrows/insets.

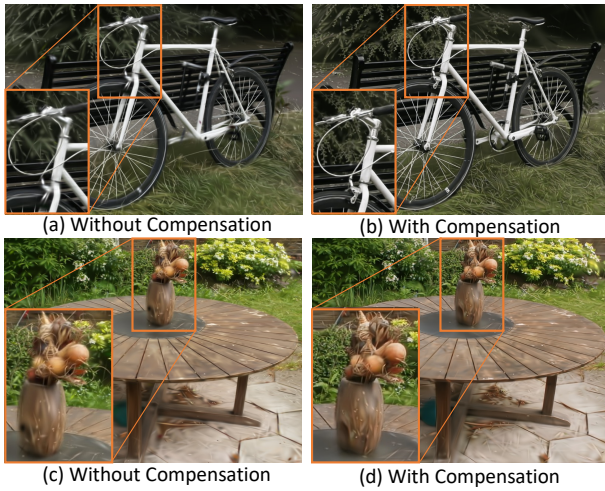


Figure 6: Visualization with and without neural compensation.

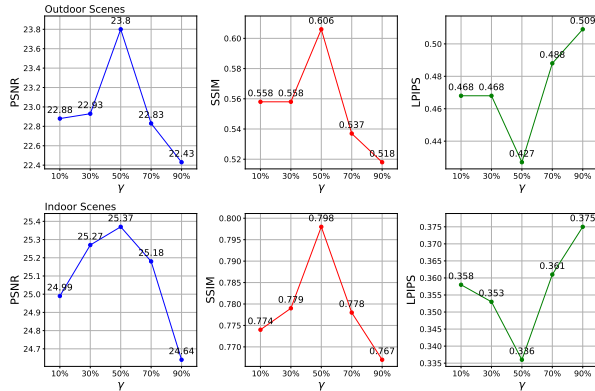


Figure 7: Ablations experiment on the ratio γ of the band-limited filter of graph based pruning.

compensatory capability of this module in mitigating the performance drops caused by spectral pruning. It is also evidenced that the relationship between primitives are well modeled.

Dataset Methods	Bicycle			Garden		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
1% + neural comp	23.23	0.579	0.449	24.42	0.651	0.399
1% w/o. neural comp	22.79	0.542	0.466	24.01	0.614	0.422
10% + neural comp	24.36	0.692	0.321	26.78	0.811	0.218
10% w/o. neural comp	23.93	0.654	0.344	26.18	0.778	0.251
30% + neural comp	24.40	0.714	0.271	27.36	0.843	0.150
30% w/o. neural comp	24.05	0.710	0.275	26.91	0.826	0.175

Table 2: Ablations of neural compensation.

Neural Compensation Module Size. Tab. 3 below shows that increasing the network size does not necessarily enhance rendering quality, aligning with findings from ADOP, indicating a similar trend. We adopt 4-layer UNet of 30MB as the default setting to best balance the quality and the memory.

Sample More Points. As seen in Tab. 1, preserving 50% of primitives outperformed the original 3DGS in rendering quality.

Dataset Methods	Bicycle			Garden		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
10% w/o. neural comp	23.93	0.654	0.344	26.18	0.778	0.251
10% + small NN (7.2MB)	24.01	0.678	0.315	26.83	0.820	0.207
10% + regular NN (30MB)	24.36	0.692	0.321	26.98	0.820	0.202
10% + large NN (120MB)	24.13	0.683	0.320	26.86	0.820	0.201

Table 3: Ablations of neural compensation module size.

We additional test keeping 80% and keeping all primitives to test how sample rates affect the final results as seen in Tab. 4. Further, keeping 80% of primitives improves rendering quality, as indicated by LPIPS, but showed minimal visual enhancement per PSNR and SSIM. Keeping all primitives (and training more epochs) could not further improve the quality, which also shows the importance of modelling the relationship of primitives. Without efficient modelling the relationship, **more primitives makes the model hard to convergence and abundant primitives are contributing negatively to the scene representation.** In addition, our goal is to balance rendering quality with storage efficiency; however, increasing storage to 620 MB for 80% primitives yields only slight quality improvements, diminishing storage efficiency.

Methods	PSNR	SSIM	LPIPS
3DGS	27.21	0.815	0.214
Ours-80%	27.73	0.835	0.185
Ours-100%	27.13	0.821	0.193

Table 4: Ablations of sampling rate on the dataset MipN-eRF360.

Continuous Pruning. In Sec. 3.4, we propose a continuous pruning strategy. We tested it on Bicycle and Counter scenes in MipNerf360 dataset according to different pruning interval iterations and pruning rate. As seen in Tab. 5, points are the number of primitives after training, and ratio is the rough ratio of number of primitives and the original 3DGS after training. Results show that this strategy could reduce the peak memory, but it is hard to control the final memory (reflected by points and ratio). And it would cause a quality loss for its parameter sensitivity. So we justify our training-then-pruning strategy but still provide the continuous pruning strategy as an alternative in our open-source toolbox.

Table 5: Evaluation for different pruning strategies on MipN-eRF360 dataset.

Pruning Strategy	Scene	PSNR	SSIM	LPIPS	Points	Ratio	Peak Memory
Prune every 2k [70%]	Bicycle	24.50	0.730	0.230	531w	70%	7032MB
Prune every 3k [70%]	Bicycle	24.47	0.730	0.224	673w	80%	8301MB
Training-then-pruning [50%]	Bicycle	24.46	0.719	0.251	421w	50%	18036MB
Prune every 2k [50%]	Bicycle	24.35	0.723	0.243	412w	50%	6386MB
Prune every 2k [30%]	Bicycle	24.17	0.699	0.264	272w	30%	6338MB
Prune every 2k [10%]	Bicycle	23.52	0.622	0.363	156w	20%	5056MB
Training-then-pruning [50%]	Counter	28.10	0.882	0.192	63.4w	50%	3730MB
Prune every 2k [30%]	Counter	27.56	0.863	0.160	62.2w	50%	1978MB
Prune every 2k [10%]	Counter	27.00	0.838	0.194	42.1w	30%	1805MB
Prune every 2k [1%]	Counter	24.54	0.777	0.194	15.7w	10%	1542MB

4.5 Efficiency Evaluation

For details on training time, CUDA memory, rendering FPS, and ROM storage, refer to Tab. 6. Notably, ‘Ours-50%’ achieves state-of-the-art rendering quality with an acceptable training duration of



1.41 hours, while achieving real-time rendering, and significantly lowering both CUDA memory usage during training and ROM storage.

Methods	Training Time	FPS	CUDA Mem	ROM Storage
3DGS	37.98min	134	10.53 GB	1.33 GB
Ours-1%	51.99min	171	2451 MB	43.6 MB
Ours-10%	1.10h	145	3169 MB	141 MB
Ours-30%	1.26h	109	4790 MB	447 MB
Ours-50%	1.41h	88	6426 MB	710 MB

Table 6: Evaluation metrics on the bicycle scene of MipNerf360 dataset.

5 CONCLUSION

In this paper, we present a novel method of spectrally pruned Gaussian fields (SUNDAE) with neural compensation, that robustly and efficiently models the relationship between Gaussian primitives by introducing the graph signal processing framework and mix the information from different primitives to compensate the information loss caused by pruning. We use the spatial information between Gaussian primitives to construct a graph to model the relationship and spectrally prune the less important ones. A lightweight neural network is utilized to compensate for the inevitable rendering quality loss post-pruning. Experimental results show that SUNDAE well maintains the efficiency of 3DGS while being much more smaller on a wide spectrum.

REFERENCES

- [1] Aamir Anis, Akshay Gadge, and Antonio Ortega. 2014. Towards a sampling theorem for signals on arbitrary graphs. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 3864–3868.
- [2] Aamir Anis, Akshay Gadge, and Antonio Ortega. 2016. Efficient sampling set selection for bandlimited graph signals using graph spectral proxies. *IEEE Transactions on Signal Processing* 64, 14 (2016), 3775–3789.
- [3] Aamir Anis and Antonio Ortega. 2017. Critical sampling for wavelet filterbanks on arbitrary graphs. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 3889–3893.
- [4] Selen Atasoy, Isaac Donnelly, and Joel Pearson. 2016. Human brain networks function in connectome-specific harmonic waves. *Nature communications* 7, 1 (2016), 10340.
- [5] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. 2022. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. *CVPR* (2022).
- [6] Hamid Behjat, Nora Leonardi, Leif Sörnmo, and Dimitri Van De Ville. 2015. Anatomically-adapted graph wavelets for improved group-level fMRI activation mapping. *NeuroImage* 123 (2015), 185–199.
- [7] Kirell Benzi, Vassilis Kalofolias, Xavier Bresson, and Pierre Vandergheynst. 2016. Song recommendation with non-negative matrix factorization and graph total variation. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Ieee, 2439–2443.
- [8] Fausto Bernardini, Joshua Mittleman, Holly Rushmeier, Cláudio Silva, and Gabriel Taubin. 1999. The ball-pivoting algorithm for surface reconstruction. *IEEE transactions on visualization and computer graphics* 5, 4 (1999), 349–359.
- [9] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. 2022. Tensorf: Tensorial radiance fields. In *European Conference on Computer Vision*. Springer, 333–350.
- [10] Siheng Chen, Dong Tian, Chen Feng, Anthony Vetro, and Jelena Kovačević. 2017. Fast resampling of three-dimensional point clouds via graphs. *IEEE Transactions on Signal Processing* 66, 3 (2017), 666–681.
- [11] Siheng Chen, Rohan Varma, Aliaksei Sandryhaila, and Jelena Kovačević. 2015. Discrete Signal Processing on Graphs: Sampling Theory. *IEEE Transactions on Signal Processing* 63, 24 (2015), 6510–6523. <https://doi.org/10.1109/TSP.2015.2469645>
- [12] Siheng Chen, Yaoqing Yang, José Moura, Jelena Kovačević, et al. 2016. Localization, decomposition, and dictionary learning of piecewise-constant signals on graphs. *arXiv preprint arXiv:1607.01100* (2016).
- [13] Yiwen Chen, Zilong Chen, Chi Zhang, Feng Wang, Xiaofeng Yang, Yikai Wang, Zhongang Cai, Lei Yang, Huaping Liu, and Guosheng Lin. 2023. Gaussiandeditor: Swift and controllable 3d editing with gaussian splatting. *arXiv preprint arXiv:2311.14521* (2023).
- [14] Zhiqin Chen, Thomas Funkhouser, Peter Hedman, and Andrea Tagliasacchi. 2023. Mobilenerf: Exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16569–16578.
- [15] Ronald R Coifman and Mauro Maggioni. 2006. Diffusion wavelets. *Applied and computational harmonic analysis* 21, 1 (2006), 53–94.
- [16] Mark Crovella and Eric Kolaczyk. 2003. Graph wavelets for spatial traffic analysis. In *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No. 03CH37428)*, Vol. 3. IEEE, 1848–1857.
- [17] Xiaowen Dong, Antonio Ortega, Pascal Frossard, and Pierre Vandergheynst. 2013. Inference of mobility patterns via spectral graph wavelets. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 3118–3122.
- [18] Venkatesan N Ekambaram, Giulia C Fanti, Babak Ayazifar, and Kannan Ramchandran. 2015. Spline-like wavelet filterbanks for multiresolution analysis of graph-structured data. *IEEE Transactions on Signal and Information Processing over Networks* 1, 4 (2015), 268–278.
- [19] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. 2022. Plenoxels: Radiance Fields without Neural Networks. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Canada, 5491–5500. ISSN: 2575-7075.
- [20] Quankai Gao, Qiangeng Xu, Hao Su, Ulrich Neumann, and Zexiang Xu. 2023. Strivec: Sparse Tri-Vector Radiance Fields. 17569–17579. https://openaccess.thecvf.com/content/ICCV2023/html/Gao_Strivec_Sparse_Tri-Vector_Radiance_Fields_ICCV_2023_paper.html
- [21] Kanghang He, Lina Stankovic, Jing Liao, and Vladimir Stankovic. 2016. Non-intrusive load disaggregation using graph signal processing. *IEEE Transactions on Smart Grid* 9, 3 (2016), 1739–1747.
- [22] Peter Hedman, Julien Philip, True Price, Jan-Michael Frahm, George Drettakis, and Gabriel Brostow. 2018. Deep Blending for Free-viewpoint Image-based Rendering. 37, 6 (2018), 257:1–257:15.
- [23] Tao Hu, Xiaogang Xu, Ruihang Chu, and Jiaya Jia. 2023. TriVol: Point Cloud Rendering via Triple Volumes. [arXiv:2303.16485](https://arxiv.org/abs/2303.16485) [cs].
- [24] Tao Hu, Xiaogang Xu, Shu Liu, and Jiaya Jia. 2023. Point2Pix: Photo-Realistic Point Cloud Rendering via Neural Radiance Fields. <http://arxiv.org/abs/2303.16482> [arXiv:2303.16482](https://arxiv.org/abs/2303.16482) [cs].
- [25] Wei Hu, Gene Cheung, Antonio Ortega, and Oscar C Au. 2014. Multiresolution graph fourier transform for compression of piecewise smooth images. *IEEE Transactions on Image Processing* 24, 1 (2014), 419–433.
- [26] Jiahui Huang, Zan Gojcic, Matan Atzmon, Or Litany, Sanja Fidler, and Francis Williams. 2023. Neural Kernel Surface Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4369–4379. https://openaccess.thecvf.com/content/CVPR2023/html/Huang_Neural_Kernel_Surface_Reconstruction_CVPR_2023_paper.html
- [27] Weiyu Huang, Antonio G Marques, and Alejandro Ribeiro. 2017. Collaborative filtering via graph signal processing. In *2017 25th European signal processing conference (EUSIPCO)*. IEEE, 1094–1098.
- [28] HyunJun Jung, Nikolas Brasch, Jifei Song, Eduardo Perez-Pellitero, Yiren Zhou, Zhihao Li, Nassir Navab, and Benjamin Busam. 2023. Deformable 3d gaussian splatting for animatable human avatars. *arXiv preprint arXiv:2312.15059* (2023).
- [29] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (ToG)* 42, 4 (2023), 1–14.
- [30] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. 2017. Tanks and Temples: Benchmarking Large-Scale Scene Reconstruction. *ACM Transactions on Graphics* 36, 4 (2017).
- [31] Muhammed Kocabas, Jen-Hao Rick Chang, James Gabriel, Oncel Tuzel, and Anurag Ranjan. 2023. Hugs: Human gaussian splats. *arXiv preprint arXiv:2311.17910* (2023).
- [32] Madeleine S Kotzagiannidis and Pier Luigi Dragotti. 2019. Splines and wavelets on circular graphs. *Applied and Computational Harmonic Analysis* 47, 2 (2019), 481–515.
- [33] Jonas Kulhanek and Torsten Sattler. 2023. Tetra-NeRF: Representing Neural Radiance Fields Using Tetrahedra. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, Canada, 18458–18469. https://openaccess.thecvf.com/content/ICCV2023/html/Kulhanek_Tetra-NeRF_Representing_Neural_Radiance_Fields_Using_Tetrahedra_ICCV_2023_paper.html
- [34] Nora Leonardi and Dimitri Van De Ville. 2013. Tight wavelet frames on multislice graphs. *IEEE Transactions on Signal Processing* 61, 13 (2013), 3357–3367.
- [35] Jiarong Lin and Fu Zhang. 2022. R 3 LIVE: A Robust, Real-time, RGB-colored, LiDAR-Inertial-Visual tightly-coupled state Estimation and mapping package. In

- 2022 International Conference on Robotics and Automation (ICRA). IEEE, 10672–10678.
- [36] William E Lorensen and Harvey E Cline. 1998. Marching cubes: A high resolution 3D surface construction algorithm. In *Seminal graphics: pioneering efforts that shaped the field*. 347–353.
- [37] Fan Lu, Yan Xu, Guang Chen, Hongsheng Li, Kwan-Yee Lin, and Changjun Jiang. 2023. Urban Radiance Field Representation with Deformable Neural Mesh Primitives. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, Canada, 465–476. https://openaccess.thecvf.com/content/ICCV2023/html/Lu_Urban_Radiance_Field_Representation_with_Deformable_Neural_Mesh_Primitives_ICCV_2023_paper.html
- [38] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *Computer Vision – ECCV 2020 (Lecture Notes in Computer Science)*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer International Publishing, Cham, 405–421. <https://arxiv.org/abs/2003.08934>
- [39] Raúl Mur-Artal and Juan D. Tardós. 2017. ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras. *IEEE Transactions on Robotics* 33, 5 (2017), 1255–1262. <https://doi.org/10.1109/TRO.2017.2705103>
- [40] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics* 41, 4 (July 2022), 1–15. <https://dl.acm.org/doi/10.1145/3528223.3530127>
- [41] Sunil K Narang and Antonio Ortega. 2012. Perfect reconstruction two-channel wavelet filter banks for graph structured data. *IEEE Transactions on Signal Processing* 60, 6 (2012), 2786–2799.
- [42] Antonio Ortega, Pascal Frossard, Jelena Kovačević, José MF Moura, and Pierre Vanderghenst. 2018. Graph signal processing: Overview, challenges, and applications. *Proc. IEEE* 106, 2 (2018), 808–828.
- [43] Ziqiao Peng, Wentao Hu, Yue Shi, Xiangyu Zhu, Xiaomei Zhang, Hao Zhao, Jun He, Hongyan Liu, and Zhaoxin Fan. 2023. SyncTalk: The Devil is in the Synchronization for Talking Head Synthesis. *arXiv preprint arXiv:2311.17590* (2023).
- [44] Isaac Pesenson. 2008. Sampling in Paley-Wiener spaces on combinatorial graphs. *Trans. Amer. Math. Soc.* 360, 10 (2008), 5603–5627.
- [45] Gilles Puy, Nicolas Tremblay, Rémi Gribonval, and Pierre Vanderghenst. 2018. Random sampling of bandlimited signals on graphs. *Applied and Computational Harmonic Analysis* 44, 2 (2018), 446–475.
- [46] Darius Rückert, Linus Franke, and Marc Stamminger. 2022. ADOP: approximate differentiable one-pixel point rendering. *ACM Transactions on Graphics* 41, 4 (July 2022), 1–14. <https://doi.org/10/gr8wfr>
- [47] Aliaksei Sandryhaila and Jose MF Moura. 2014. Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure. *IEEE signal processing magazine* 31, 5 (2014), 80–90.
- [48] Aliaksei Sandryhaila and Jose MF Moura. 2014. Discrete signal processing on graphs: Frequency analysis. *IEEE Transactions on Signal Processing* 62, 12 (2014), 3042–3054.
- [49] Erik Sandström, Yue Li, Luc Van Gool, and Martin R. Oswald. 2023. Point-SLAM: Dense Neural Point Cloud-based SLAM. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 18433–18444. https://openaccess.thecvf.com/content/ICCV2023/html/Sandstrom_Point-SLAM_Dense_Neural_Point_Cloud-based_SLAM_ICCV_2023_paper.html
- [50] Johannes L Schonberger and Jan-Michael Frahm. 2016. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4104–4113.
- [51] Xiaowei Song, Jv Zheng, Shiran Yuan, Huan-ang Gao, Jingwei Zhao, Xiang He, Weihao Gu, and Hao Zhao. 2024. SA-GS: Scale-Adaptive Gaussian Splatting for Training-Free Anti-Aliasing. *arXiv preprint arXiv:2403.19615* (2024).
- [52] Cheng Sun, Min Sun, and Hwann-Tzong Chen. 2022. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5459–5469.
- [53] Jiaxiang Tang, Jiawei Ren, Hang Zhou, Ziwai Liu, and Gang Zeng. 2023. Dream-gaussian: Generative gaussian splatting for efficient 3d content creation. *arXiv preprint arXiv:2309.16653* (2023).
- [54] David BH Tay and Jingxin Zhang. 2015. Techniques for constructing biorthogonal bipartite graph filter banks. *IEEE Transactions on Signal Processing* 63, 21 (2015), 5772–5783.
- [55] Oguzhan Teke and Palghat P Vaidyanathan. 2016. Extending classical multirate signal processing theory to graphs—Part I: Fundamentals. *IEEE Transactions on Signal Processing* 65, 2 (2016), 409–422.
- [56] Dong Tian, Hassan Mansour, Andrew Knyazev, and Anthony Vetro. 2014. Chebyshev and conjugate gradient filters for graph image denoising. In *2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. IEEE, 1–6.
- [57] Paola Valdivia, Fabio Dias, Fabiano Petronetto, Cláudio T Silva, and Luis Gustavo Nonato. 2015. Wavelet-based visualization of time-varying data on graphs. In *2015 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 1–8.
- [58] Michal Valko, Rémi Munos, Branislav Kveton, and Tomáš Kocák. 2014. Spectral bandits for smooth graph functions. In *International Conference on Machine Learning*. PMLR, 46–54.
- [59] Yuxi Wei, Zi Wang, Yifan Lu, Chenxin Xu, Changxing Liu, Hao Zhao, Siheng Chen, and Yanfeng Wang. 2024. Editable Scene Simulation for Autonomous Driving via Collaborative LLM-Agents. *arXiv preprint arXiv:2402.05746* (2024).
- [60] Thomas Whelan, Stefan Leutenegger, Renato Salas-Moreno, Ben Glocker, and Andrew Davison. 2015. ElasticFusion: Dense SLAM without a pose graph. *Robotics: Science and Systems*.
- [61] Zirui Wu, Tianyu Liu, Liyi Luo, Zhide Zhong, Jianteng Chen, Hongmin Xiao, Chao Hou, Haozhe Lou, Yuantao Chen, Runyi Yang, et al. 2023. Mars: An instance-aware, modular and realistic simulator for autonomous driving. In *CAAI International Conference on Artificial Intelligence*. Springer, 3–15.
- [62] Qiangeng Xu, Zexiang Xu, Julien Philip, Sai Bi, Zhixin Shu, Kalyan Sunkavalli, and Ulrich Neumann. 2022. Point-NeRF: Point-based Neural Radiance Fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 5428–5438. <https://doi.org/10/gq9msm>
- [63] Wei Xu, Yixi Cai, Dongjiao He, Jiarong Lin, and Fu Zhang. 2022. Fast-lio2: Fast direct lidar-inertial odometry. *IEEE Transactions on Robotics* 38, 4 (2022), 2053–2073.
- [64] Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. 2023. Mip-splatting: Alias-free 3d gaussian splatting. *arXiv preprint arXiv:2311.16493* (2023).
- [65] Chongjian Yuan, Wei Xu, Xiyuan Liu, Xiaoping Hong, and Fu Zhang. 2022. Efficient and probabilistic adaptive voxel mapping for accurate online lidar odometry. *IEEE Robotics and Automation Letters* 7, 3 (2022), 8518–8525.
- [66] Shiran Yuan and Hao Zhao. 2023. SlimmerRF: Slimmable Radiance Fields. *arXiv preprint arXiv:2312.10034* (2023).
- [67] Jin Zeng, Gene Cheung, and Antonio Ortega. 2017. Bipartite approximation for graph wavelet signal decomposition. *IEEE Transactions on Signal Processing* 65, 20 (2017), 5466–5480.
- [68] Chen Zhang, Wanjuan Su, and Wenbing Tao. 2023. Point-NeuS: Point-Guided Neural Implicit Surface Reconstruction by Volume Rendering. <http://arxiv.org/abs/2310.07997> arXiv:2310.07997 [cs].
- [69] Yifei Zhang, Hao Zhao, Hongyang Li, and Siheng Chen. 2024. FastMAC: Stochastic Spectral Sampling of Correspondence Graph. *arXiv preprint arXiv:2403.08770* (2024).
- [70] Yufeng Zheng, Wang Yifan, Gordon Wetzstein, Michael J. Black, and Otmar Hilliges. 2023. PointAvatar: Deformable Point-based Head Avatars from Videos. <http://arxiv.org/abs/2212.08377> arXiv:2212.08377 [cs].
- [71] Qiang Zhou, Weize Li, Lihan Jiang, Guoliang Wang, Guyue Zhou, Shanghang Zhang, and Hao Zhao. 2024. Pad: A dataset and benchmark for pose-agnostic anomaly detection. *Advances in Neural Information Processing Systems* 36 (2024).
- [72] Zhenxin Zhu, Yuantao Chen, Zirui Wu, Chao Hou, Yongliang Shi, Chuxuan Li, Pengfei Li, Hao Zhao, and Guyue Zhou. 2023. Latitude: Robotic global localization with truncated dynamic low-pass filter in city-scale nerf. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 8326–8332.