

Compact 3D Scene Representation via Self-Organizing Gaussian Grids

Wieland Morgenstern¹ Florian Barthel^{1,2} Anna Hilsmann¹ Peter Eisert^{1,2}

¹ Fraunhofer Heinrich Hertz Institute, HHI

² Humboldt University of Berlin

{first}.{last}@hhi.fraunhofer.de

Abstract

3D Gaussian Splatting has recently emerged as a highly promising technique for modeling of static 3D scenes. In contrast to Neural Radiance Fields, it utilizes efficient rasterization allowing for very fast rendering at high-quality. However, the storage size is significantly higher, which hinders practical deployment, e.g. on resource constrained devices. In this paper, we introduce a compact scene representation organizing the parameters of 3D Gaussian Splatting (3DGS) into a 2D grid with local homogeneity, ensuring a drastic reduction in storage requirements without compromising visual quality during rendering. Central to our idea is the explicit exploitation of perceptual redundancies present in natural scenes. In essence, the inherent nature of a scene allows for numerous permutations of Gaussian parameters to equivalently represent it. To this end, we propose a novel highly parallel algorithm that regularly arranges the high-dimensional Gaussian parameters into a 2D grid while preserving their neighborhood structure. During training, we further enforce local smoothness between the sorted parameters in the grid. The uncompressed Gaussians use the same structure as 3DGS, ensuring a seamless integration with established renderers. Our method achieves a reduction factor of 8x to 26x in size for complex scenes with no increase in training time, marking a substantial leap forward in the domain of 3D scene distribution and consumption. Additional information can be found on our project page:

fraunhoferhhi.github.io/Self-Organizing-Gaussians/

1. Introduction

3D reconstruction and rendering of real-world scenes has been a cornerstone of computer vision since its inception. Neural Radiance Fields (NeRFs) [24] and 3D Gaussian Splatting (3DGS) [14] have recently revolutionized 3D representation and novel-view synthesis. Assessing the efficacy of 3D scene reconstruction methods and representa-

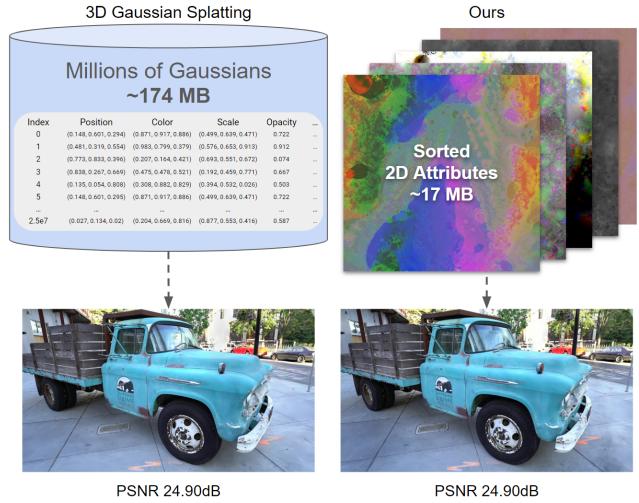


Figure 1. Our 3DGS training method allows for high compression of 3D Gaussian Splatting attributes while maintaining high rendering quality. By sorting the Gaussian features periodically into a 2D grid and applying a smoothness loss during scene generation, we are able to significantly reduce the memory footprint using state-of-the-art image compression methods.

tions typically involves three important metrics: rendering quality, rendering speed, and storage size. While rendering quality is of general importance, rendering speed and storage size play a crucial role especially for real time applications, portability on devices with limited performance and memory or, for fast web applications.

NeRF-based rendering methods achieve high rendering quality [1]. However, they are costly to train and synthesis speed is slow, as each pixel value is computed by aggregating multiple results from a forward pass through a neural network. Recent methods have enabled considerably faster training and inference [7, 27, 34], often, however, at the cost of decreased quality and/or increased disk memory for each model. A recent alternative to NeRF is 3D Gaussian Splatting (3DGS), explicitly representing the scene with localized 3D Gaussian splats [14]. In contrast to

NeRFs, 3DGS employs an efficient rasterization algorithm allowing for very fast rendering at high quality. In addition, the representation enables the modeling of consistent object motion in dynamic scenes [22]. On the downside, however, the storage size of 3DGS models is significantly higher than with NeRFs, because 3DGS stores millions of parameters in a large unorganized list. Such a data structure can take up to several hundred megabytes depending on the scene, which hinders practical deployment, e.g. on resource constrained devices.

To address this limitation, we propose a new compact representation based on 3DGS, allowing for efficient encoding and storage. We exploit the fact that numerous permutations of Gaussian parameters can equivalently represent it with the same visual quality by arranging the millions of multidimensional list entries of 3DGS into a structured 2D grid with increased smoothness that can efficiently be encoded. In essence, we integrate a smoothness loss into the 3DGS scene optimization process to guide the ambiguous arrangements of the splats towards a visually equivalent but better compressible parameter set. We finally quantize the parameters and encode them by efficient 2D image compression methods, reducing the overall memory footprint by a factor of 8x to 26x, without sacrificing rendering quality.

With our method, we enable novel-view synthesis at high quality, high rendering speed, and with a small memory footprint, bringing high quality 3D scene reconstruction and rendering one step closer towards applications on small devices with limited memory or fast web applications.

Our main contributions are:

1. We propose a new compact scene representation and training concept for 3DGS, structuring the high-dimensional features in a smooth 2D grid, which can be efficiently encoded using state-of-the-art compression methods.
2. We introduce an efficient 2D sorting algorithm that sorts millions of 3DGS parameters in parallel on the GPU.
3. We provide a simple to use interface for compressing and decompressing the resulting 3D scenes. The decompressed reconstructions share the structure of 3DGS, allowing integration into established renderers.
4. We efficiently reduce the storage size by a factor of 8x to 26x while maintaining high visual quality

2. State of the Art

We first overview recent advances in 3D reconstruction and representation, before we take a closer look at high dimensional data structuring and sorting.

2.1. 3D Scene Representation

Over the past years, 3D scene representation has seen rapid improvements, especially with the advent of deep learning.

The general goal is to allow novel-view synthesis of 3D objects captured by a number of 2D images. In the following, we will give a brief overview of the most relevant technologies and highlight their key differences. For a more comprehensive review in this field, please refer to Tewari *et al.* [36].

Early novel-view synthesis approaches were based on pure image interpolation as for the lightfields [10, 19], or reconstructed 3D geometry by means of regular voxel grids [29], point clouds, or meshes. Structure-from-Motion (SfM) and Multi-View-Stereo (MVS) were proposed to estimate such representations from a set of input images [5, 9, 31]. These methods were the basis for view-synthesis approaches guided by explicit representations of the geometry [3, 17, 28]. Also compression of the large 3D point cloud datasets has been addressed [11, 26], but is typically restricted to simple colored points without additional attributes.

In recent years, implicit representations have been proposed to represent geometry [30]. Especially, Neural Radiance Fields (NeRFs) have revolutionized novel-view synthesis [24] by representing a scene volume as a multi-layer perceptron (MLP). This representation allows the visualization of thin structures, semi-transparent surfaces and highly realistic light reflections. In general, NeRFs show an outstanding rendering quality, however, they are slow during training and inference.

The success of NeRF has resulted in numerous follow-up methods improving rendering quality, e.g. for large scenes [23], or speed [1, 7, 25, 35]. While Mip-NeRF360 [1] as one of the state-of-the-art approaches achieves an outstanding quality, it is still very costly to train and render. Methods focusing on faster training and/or rendering exploit spatial data structures and encodings in order to accelerate computation, representing the scene as a grid [7], tensors [4], latent features [13, 21, 34], hash grids [25] or using vector quantizations [35]. Among these, InstantNGP is a widely recognized example, using a smaller MLP to represent density and appearance though the use of a hash table and an occupancy grid [25]. Plenoxels estimate explicit spherical harmonics in addition to color and density to represent view dependent appearance effects [7]. In order to improve the rendering, the values are interpolated from a sparse 3D grid of features using trilinear interpolation, resulting in considerably higher rendering speed at the cost of lower quality. In addition, the memory size of Plenoxels is high, given that the number of elements in the 3D grid grows by $\mathcal{O}(n^3)$.

Other approaches aiming at real-time rendering involved precomputing and storing NeRF's view-dependent colors and opacities in volumetric data structures [4, 8, 12, 18, 37, 38] or partitioning the scene into voxels, each represented by separate small MLPs. However, these representations suffer from quality degradation and impose substan-

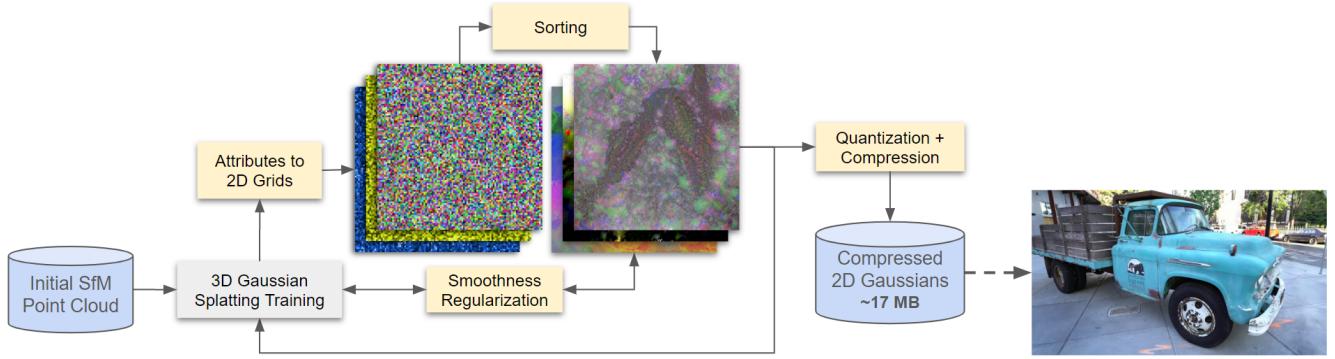


Figure 2. An overview of our novel 3DGS training method. During training, we arrange all high dimensional attributes into multiple 2D grids. Those grids are sorted and a smoothness regularization is applied. This creates redundancy which help to compress the 2D grids into small files using off-the-shelf compression methods.

tial graphics memory demands, restricting their application to objects rather than entire scenes. Several methods have been proposed to reduce the required disk space. Among these, Memory-efficient NeRF (MERF) [18] and TensoRF [4], share a similar approach, storing the 3D scene representation inside lightweight 2D features. Other methods to compress NeRF reconstructions use incremental pruning [6], vector quantization [18, 39] or compressed codebook representations [35]. However, these approaches still require costly network inference passes for each query point in the scene and are therefore not capable of real-time rendering of large scenes.

Our approach is inspired by a recently proposed approach for 3D scene representation by Kerbl *et al.* [14] combining the advantages of explicit representations (fast training and rendering) with those of implicit representation (stochastic sampling) by representing the scene by millions of 3D Gaussians, called 3D Gaussian Splatting (3DGS). The parameters (3D coordinates, color, size, density, orientation, and spherical harmonics) of these Gaussians are optimized using an efficient differentiable rasterization algorithm starting from a set of Structure-from-Motion points to best represent a given 3D scene. The main contribution of this approach is a representation that is continuous and differentiable while also allowing fast rendering: 3DGS is currently the state-of-the-art approach in terms of quality, rendering time, and performance. However, this comes at the cost of significantly higher memory requirements, as for each scene, millions of high dimensional Gaussians are stored in an unstructured list.

2.2. Mapping Higher Dimensional Data to 2D Grids

The ambiguity of ordering of the high dimensional point clouds in approaches like 3DGS [14] enables optimization of point indices for more efficient representation and storage. Grid-based sorting techniques like Self-Organizing Map [15, 16], Self-Sorting Map [32, 33], or Fast Linear

Assignment Sorting [2], allow mapping high-dimensional data points onto a 2D grid, organized by similarity, thereby facilitating an efficient compression of the data.

A self-organizing map (SOM) [15, 16] is a neural network trained using unsupervised learning to produce a low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples. Initialized with random vectors, the map is refined through iterative adjustments to represent the input data in a topologically organized manner until convergence. Due to the sequential nature of the SOM, the last input vectors can only be assigned to the few remaining unassigned map positions, resulting in isolated and poorly positioned vectors within the map. Self-Sorting Maps [32, 33] avoid this problem by swapping assigned positions of four input vectors at a time. Due to the factorial number of permutations, adding more candidates would be computationally too complex. Linear Assignment Sorting (LAS) [2] outperforms other sorting methods in terms of speed and sorting quality for non-parallel computations, by swapping all (or several (Fast LAS)) vectors simultaneously using a continuously filtered map.

However, these grid-sorting algorithms, effective for handling thousands of items, prove inadequate for our specific application, where we seek to structure millions of Gaussians spread across various dimensions. To overcome this, we adapt and expand upon existing sorting methodologies, enabling continuous sorting of Gaussian data during training without inflating the training duration.

3. Method

Our goal is a new representation and training scheme for 3D Gaussian Splatting with drastically reduced storage requirements while maintaining visual quality during rendering. The idea of our method is to map originally unstructured 3D Gaussians to a 2D grid in such a way that local spatial

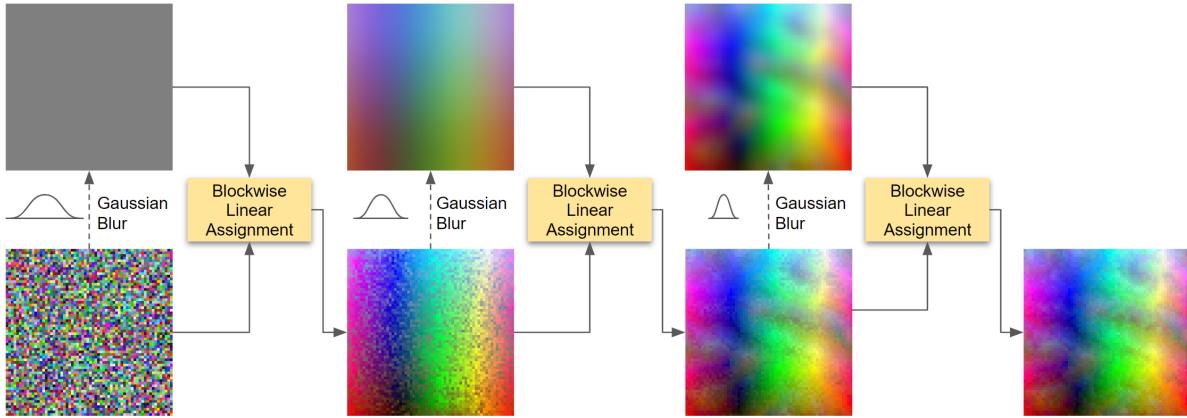


Figure 3. We sort the Gaussians into a 2D grid with an iterative approach. Gaussians are assigned to the position matching closest to a smoothed version of the grid. With increasing iterations, we decrease the kernel size and sigma of the 2D gaussian filter.

relationships are preserved. By mapping all Gaussian parameters (position, color, etc.) to the same 2D position, we effectively create multiple data layers with the same layout. In this 2-dimensional layout, neighboring Gaussians will have similar attribute values, facilitating compression and enabling the storage of the same scene using reduced memory.

Figure 2 illustrates an outline of our method. Our training introduces two new components to the 3DGGS framework: a highly-parallel 2D sorting step (Section 3.1), and an enhanced scene optimization with a local neighborhood loss on the 2D grid (Section 3.2). The sorting step initializes the optimization process by explicitly organizing the values, establishing an initial smooth local neighborhood without requiring a global optimization. Once sorted, our enhanced training incorporates a neighborhood loss, further enforcing local smoothness on the 2D grid, while the differentiable renderer of the 3D Gaussian splatting optimizes the Gaussians attributes to accurately represent the scene. Similar to the original 3DGGS approach, Gaussians are pruned from the scene in regular intervals, and new Gaussians are added in under-reconstructed parts of the scene. After each densification step, we re-sort all Gaussians into the 2D grid. This step re-establishes local neighborhoods after pruning and densification may have disrupted them.

To store the Gaussian attributes, we leverage off-the-shelf compression methods, which can be applied effectively to the well-organized data. Inspired by [1], we map the position of the Gaussians to a logarithmic scale. Thus, Gaussians located in the center of the scene have more precise coordinates.

3.1. Sorting High Dimensional Gaussians into 2D Grids

Traditional grid-sorting algorithms are suitable for thousands of items, yet fall short for our application, where even

small scenes contain millions of Gaussians across multiple dimensions. To address this, we combine and enhance concepts from existing sorting methods discussed in Section 2.2. Our aim is to handle the increased data volume and complexity efficiently, while ensuring that regular sorting of Gaussian data during training does not negatively affect training time.

Our strategy involves an approximation approach to the highly complex assignment problem, inspired by the principles of the Fast Linear Assignment Sorting (FLAS) algorithm [2]. We aim to find an optimal balance between computational efficiency and maintaining the accuracy of the spatial relationships among the Gaussian attributes. The parallel assignment process is visualized with random RGB colors in Figure 3: We initialize our grids by mapping the Gaussians to random positions, even if they are already sorted, to avoid getting stuck in local minima. Subsequently, we perform a low-pass filtering operation on the grid to construct an idealized target grid, and re-assign all elements to their best-matching positions regarding the smooth target. This process is repeated while decreasing the filter size, gradually sorting the grid.

For each re-assignment, the grid is divided into multiple blocks, which are processed independently. The block size β is set as $\beta = \phi + 1$, where ϕ is the radius of the Gaussian blur applied to compute the target. We do not create blocks smaller than $\beta = 16$, as tiny blocks inhibit efficient parallelization. To be able to sort across block borders and to cover all borders of the grid when not aligned at multiples of the current block size, we shift all blocks by a random Δ_y and Δ_x before starting re-assignment.

After each block-wise linear assignment, we measure the mean L2 distance between sorted grid and target grid and compare it to its previous value. If it has improved by less than 0.01%, we start with a new random block position. When the assignment with the new block positions does not

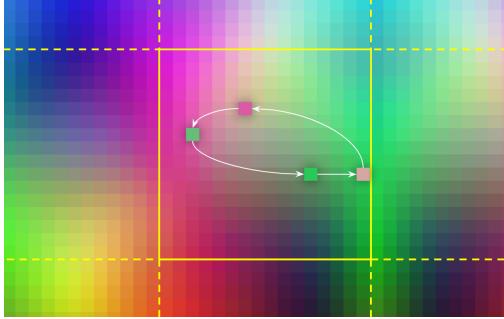


Figure 4. All values within each sub-block of the 2D grid are re-assigned in the same step. The elements from the block are split into randomly assigned, non-overlapping groups of 4. Out of the 24 possible permutations for each 4 elements, the positions that have the lowest total distance to the smoothed grid target are chosen.

improve the distance further (using the same threshold), the radius is reduced by applying $\phi = 0.95\phi$. We start with a radius $\phi = \frac{w}{2} - 1$, where w is the total grid width, and we finish the work when the radius value would fall below 1, and the sorting has converged.

Blockwise Linear Assignment: As traditional sorting algorithms do not map well to GPU execution, we perform the assignment by swapping four elements at a time, as done in Self-Organizing Structured Maps [33]. The groups of four are chosen from a random permutation of all elements from a block, using each element only once. The random grouping is important for pairing up elements with possible exchange targets from different areas within the block. All blocks are processed in parallel, and use the same permutation for the grouping. Finally, the new assignment for the groups of four employs a brute-force approach, iterating through the cost for all 24 possible permutations to find the optimal arrangement of elements. A single re-assignment is visualized in Figure 4.

Organizing Gaussians into a 2D Grid: The presented algorithm can be implemented highly parallel on the GPU, and will organize millions of Gaussians in a few seconds, using a few thousand iterations of the blockwise linear assignment. In our application, we normalize all attributes of the Gaussians (coordinates, scale, etc.) independently to the range $[0, 1]$ and then scale them with individual weights before calculating the distance of the possible re-assessments. We use the activated values for all attributes (e.g. exponentially activated scale), as they would be seen by the renderer. At the start of the training, we sort once, and continue sorting with each densification step. As the number of Gaussians grows over time, so will the size of the grid. We find the largest square grid size that can be completely filled with Gaussians for simplicity, and prune the Gaussians with the lowest opacity values that will not fit.

3.2. Smoothness Regularization

2D image compression methods work best if the content exhibits little noise and shows smooth structures that can be accurately predicted from neighboring values or represented by only a few transform coefficients. To further enforce such features in our 2D attribute grids, we employ a smoothness regularization on the sorted Gaussians during the training for scene optimization. Here, we exploit the fact that different Gaussian configurations can lead to similar visual quality while varying in local 2D grid smoothness.

For the regularization, we first create a blurred and thus smoother version of the 2D grid by applying a 2D Gaussian filter on all attribute channels. Deviations of the current Gaussian parameters at an original 2D grid position from those of the smoothed grid form the additional smoothness loss for training. For comparison, we select a Huber loss, as it is less sensitive to outliers than MSE. This has shown to improve the compression rate. Formally, our smoothness regularization term can be expressed as follows:

$$R_s(I) = \begin{cases} \frac{1}{2}(I - I_b)^2, & \text{for } |I - I_b| \leq 1 \\ |I - I_b| - \frac{1}{2}, & \text{otherwise.} \end{cases} \quad (1)$$

Here, I and I_b denote the original and blurred 2D attribute grid, respectively. With this formulation, the smoothness regularization penalizes all pixels on the 2D grid that have very different values compared to their local neighborhood, thus creating smoother 2D features. We add the loss of the smoothness regularization to the loss of the 3D Gaussian Splatting algorithm using a weight λ . As we propagate the gradient of the smoothness regularization through the Gaussian renderer during training, we specifically enforce the 3D Gaussian Splatting optimization process to prefer Gaussians that improve the quality of the rendering while also considering the respective local smoothness. This differentiates our method from compression methods that focus on reducing existing data in a pure post-processing step. Instead, we manipulate the data during creation to achieve high compression and high visual quality afterwards.

4. Evaluation

4.1. Implementation Details

Our experiments are based off the official 3DGS implementation on GitHub¹. The parallel sorting is implemented in PyTorch and also uses the CUDA backend. For our comparisons to prior methods, we select three real-world 3D scene reconstruction datasets: Mip-NeRF360, Tanks&Temples and Deep Blending and a synthetic dataset: Synthetic-NeRF.

¹<https://github.com/graphdeco-inria/gaussian-splatting>

Dataset Method/Metric	Mip-NeRF360			Tanks&Temples			Deep Blending			Synthetic-NeRF		
	PSNR↑	LPIPS↓	Size↓	PSNR↑	LPIPS↓	Size↓	PSNR↑	LPIPS↓	Size↓	PSNR↑	LPIPS↓	Size
Plenoxels †	23.08	0.463	2100	21.08	0.379	2300	23.06	0.510	2700	31.76	-	-
M-NeRF360 †	27.69	0.237	8.6	22.22	0.257	8.6	29.40	0.245	8.6	33.09	-	-
INGP-Base †	25.30	0.371	13	21.72	0.330	13	23.62	0.423	13	33.18	-	-
INGP-Big †	25.59	0.331	48	21.92	0.305	48	24.96	0.390	48	-	-	-
VQ-TensoRF †	-	-	-	28.20	-	3.3	-	-	-	32.86	-	3.6
3DGS	27.55	0.222	785	25.54	0.201	454	30.07	0.248	699	33.88	0.031	71.6
3DGS w/o SH	26.94	0.234	212	25.10	0.210	122	30.24	0.249	191	32.06	0.039	19.4
Ours-Q	26.01	0.259	23.9	24.52	0.237	11.9	28.92	0.276	8.4	31.05	0.047	2.2
Ours-S	25.83	0.273	18.2	24.59	0.247	9.9	29.08	0.280	7.2	30.83	0.051	1.7

Table 1. A comparison of our method to the default 3DGS with and without spherical harmonics (SH) and prior NeRF-based renderer. Compared to 3DGS without spherical harmonics, a size reduction between 8.8x and 26.5x is reached, depending on the dataset. All sizes are in MB. Results with a † are directly copied from [14, 20]. Results for SSIM are included in the supplementary material.

During optimization, we have found two sets of parameters that showed promising results in terms of rendering quality and size. In the following we will refer to those configurations as Q and S . Q denotes the configuration that mainly focuses on rendering quality, while S rather focuses on size. A more detailed description of both parameter configurations is given in the supplementary material of this paper. For the compression of our 2D attribute grids, we use a combination of EXR and JPEG-XL. A more comprehensive study on different compression methods in combination with different quantization levels is also given in the supplementary material. For all our experiments, we disable the spherical harmonics from 3DGS.

4.2. Results

Quantitative results: In table 1 we compare our method to the default 3D Gaussian Splatting algorithm and prior NeRF-based 3D reconstruction methods, reporting the standard PSNR, L-PIPS, and SSIM metrics. We demonstrate that we reduce the average memory size by a factor of 8x to 26x compared to 3DGS [14] without spherical harmonics, without sacrificing visual quality. The highest reduction is observed on the Deep Blending dataset. For some datasets, compared to Mip-NeRF360 [1] and VQ-TensoRF [20], we achieve a slightly higher PSNR and slightly lower L-PIPS, but at the same time allow for real time rendering and fast training, given the efficient rasterization algorithm of 3DGS [14]. Additionally, while Mip-NeRF360 trains one scene for several hours, our method only uses 10 to 30 minutes. Most notably, we achieve the same training time as 3DGS, even though our method brings some computational overhead compared to the default 3DGS, arising from the periodic sorting of Gaussians and the necessity to blur grids for computing the neighbor loss. This is due to our parameter configuration enforcing the optimization process to create a considerably smaller number of Gaussians. For example, for the *Truck* dataset, our method creates 1.55M Gaussians, whereas the default 3DGS creates 2.58M.

For the same reason of the reduced number of Gaussians, our scenes render much faster in the 3DGS viewer than the original models. The vanilla *Truck* with 2.58M Gaussians renders at 385 fps with the default settings on the same GPU, while our scene with 1.55M Gaussians renders at 515 fps with nearly the same visual quality.

Qualitative results: The results of the metrics suggest that the rendering quality of our method is very similar to the vanilla 3DGS. This is also demonstrated in figure 5, where we directly compare the renderings for the truck scene side by side. Here, the training with activated spherical harmonics shows the best rendering results, especially when inspecting transparent materials such as the windshield of the truck. Besides that, all other renderings that do not use spherical harmonics show very similar quality. This underlines that our method is able to maintain a high rendering quality, while only using a fraction of the memory.

Figure 6 displays examples of the sorted 2D grids during the training process. Specifically, we show the 2D RGB color grids for the counter, the flowers and the kitchen dataset. We observe that the colors of the scenes organize themselves into clusters of similar colors. The grids, however, do not appear to be more organized with more iterations. This has two reasons. Firstly, the number of Gaussians grows substantially during the training, posing an increasingly more difficult sorting task. And secondly, as our method has to sort all attributes at once using the same permutation, isolated features, such as the color can not be organized perfectly.

4.3. Smoothness Regularization

The magnitude of the smoothness regularization λ decides the size of the gradient for local smoothness in the 2D attribute grids in relation to the size of the gradient of the Gaussian training algorithm. To find a good balance between both optimization targets, we tested a series of parameters for λ . The results for this study is displayed in table 2. As expected, lower magnitudes for λ result in higher

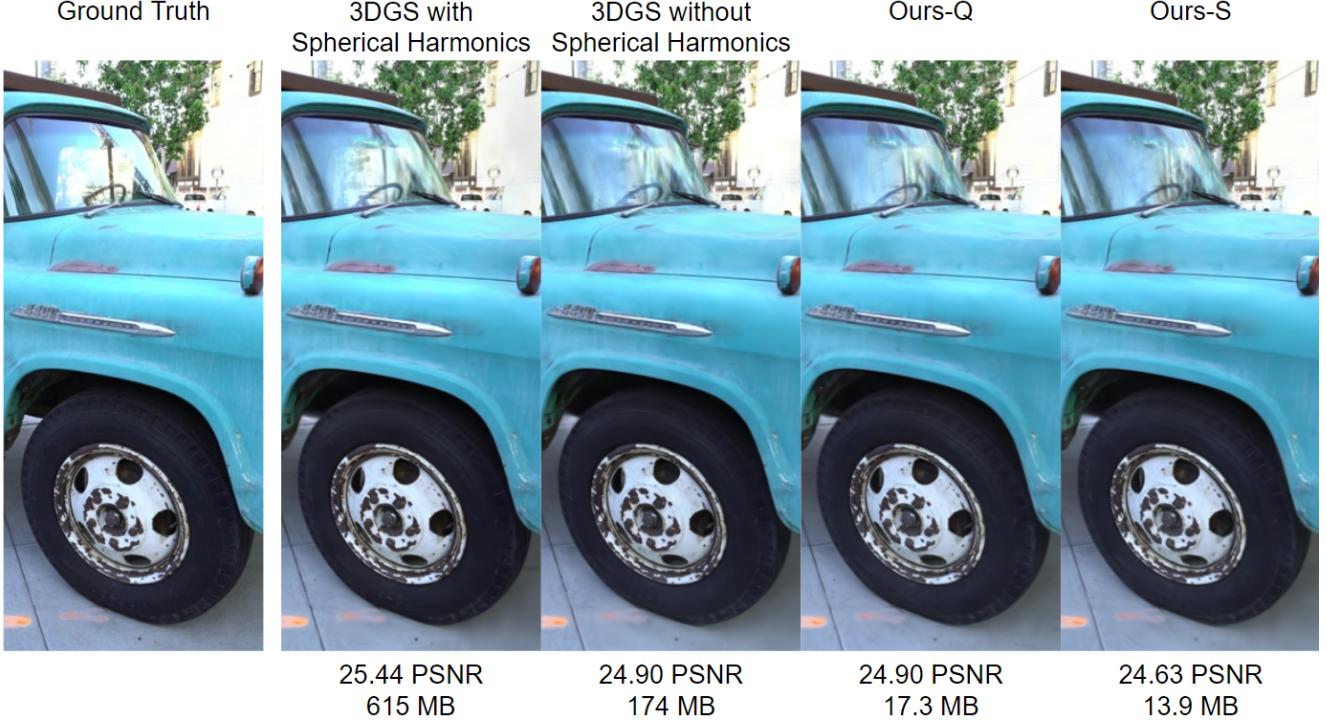


Figure 5. A qualitative comparison on the truck dataset between our method and 3DGS with and without spherical harmonics. Ours-Q denotes the configuration that focuses on rendering quality, while Ours-S focuses on disk size.

rendering quality, but also larger storage size.

Neighbor Loss λ	PSNR	Size in MB
0.01	24.39	18.02
0.05	24.48	18.35
0.1	24.18	17.50
0.5	24.40	14.62
1.0	24.24	12.43
1.5	24.02	10.73

Table 2. Testing different magnitudes for the neighbor loss λ .

To demonstrate the importance of the smoothing regularization, we only apply our sorting method to a normally trained 3DGS model and compress the result. This is demonstrated in table 3. Here, 3DGS + sort denotes the experiment without smoothness regularization. With a size of only 36 MB compared to 174 MB, it is substantially smaller than before. However, the PSNR is also reduced. This is because the compression method relies on quantization and smooth structures, which the sorting algorithm only partially achieves. Both our experiments, on the other hand, both perform better in terms of PSNR and size, underlining the significance of the smoothing regularization.

Method	PSNR	Size in MB
3DGS	24.89	174.0
3DGS + sort	23.90	36.4
Ours-Q	24.90	17.3
Ours-S	24.63	13.9

Table 3. Evaluating the PSNR and size of a 3DGS model that was not trained with the smoothness regularization. Instead we only apply our sorting method after training.

4.4. Features to 2D Grid

To convert the millions of Gaussians into a 2D grid, we reshape the features to a square grid. This grid has a side length of the rounded square root of the number of Gaussians. This means that we potentially lose a row and a column of Gaussians at most during the reshape process. To ensure that we do not throw away any Gaussians that might bring meaningful contributions to the scene, we sort all Gaussians by opacity and only discard Gaussians with the lowest opacity. In figure 7, we show that the PSNR of a rendered scene is not affected when removing up to 30% of Gaussians that have a low opacity. In contrast, removing Gaussians with small scaling, affects the quality of the scene considerably more.

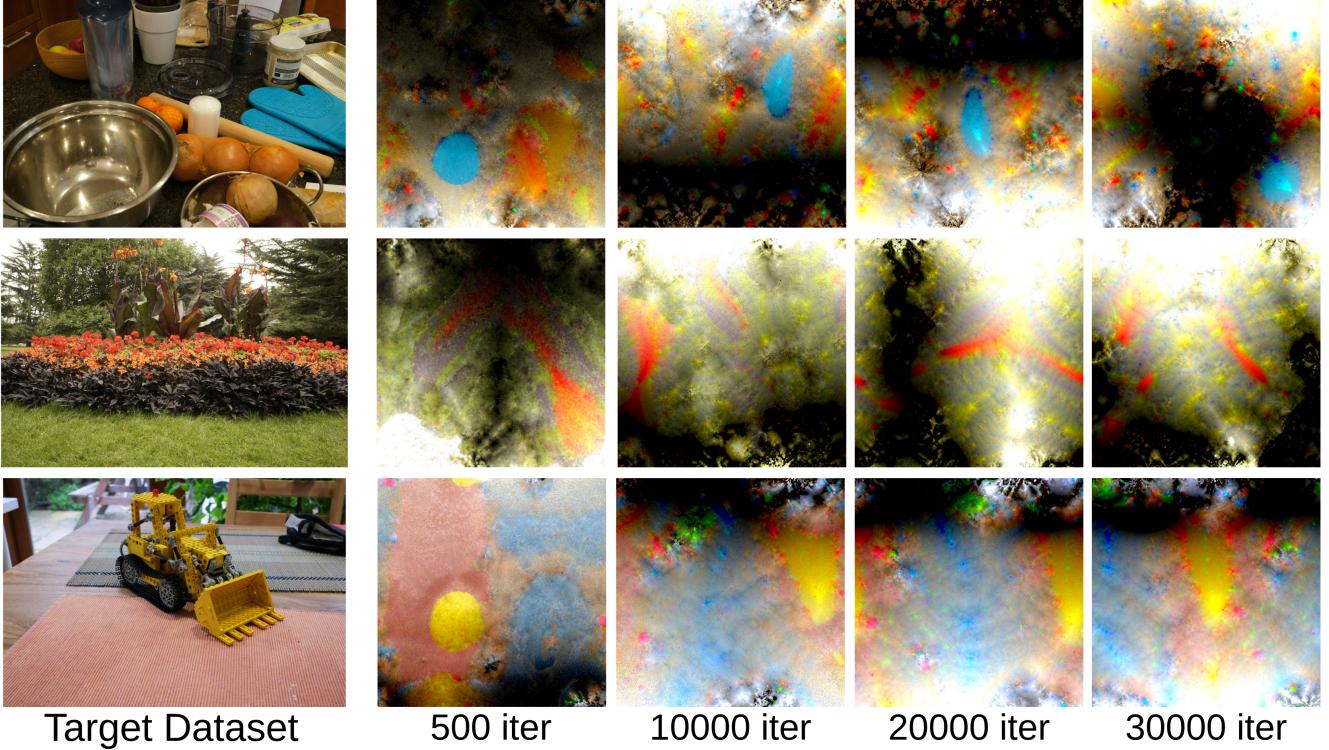


Figure 6. Resulting 2D attribute grids for the colors for the respective scene during training. Until iteration 15000, the number of Gaussians grow from about 100k-200k up to 1.5-2 million.

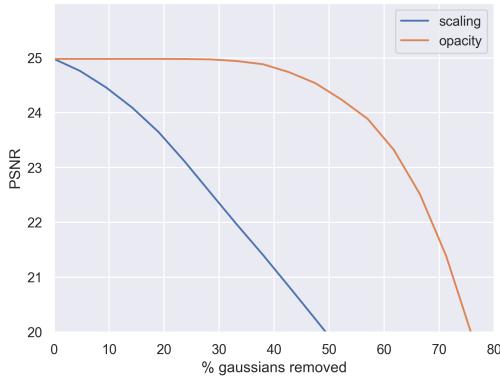


Figure 7. Removing the first % of Gaussians, ordered by scaling and opacity from the lowest to the highest.

4.5. Sorting Performance

To demonstrate the effectiveness of our parallel sorting algorithm, we compare it to the state of the art in quality, FLAS [2]. An implementation of Self-Sorting Maps [33] was not available to us. We measure runtime and, to estimate the quality of the sort, the variance of the absolute differences (VAD) between all neighboring values, over all channels.

Sorting a random 512x512x3 grid on an *AMD Ryzen*

Threadripper PRO 5955WX CPU with FLAS takes 131s and achieves a VAD of 3.53. Our algorithm running on an *Nvidia RTX 4090* finishes in 5.7 seconds after 8015 re-orders with a VAD of 4.02 (the shuffled data has a VAD of 3607.45).

5. Conclusion

We have presented a novel optimized representation of 3D Gaussian Splatting, targeting efficient compression and storage. This is achieved by a novel, highly parallelized sorting strategy that smoothly arranges the high dimensional, initially unordered Gaussian parameter sets in a 2D grid, which can be encoded by standard image coding techniques like jpeg-xl. In addition, an optimized training scheme for the 3D Gaussians with a new smoothness loss leads to splat configurations with even higher smoothness in the 2D grid, while still preserving the original accuracy of the 3D scene. In spite of the proposed extensions, training time only negligibly changes compared to the original 3DGS approach. With the proposed solution, we can reduce the data down to 4% of the original size at the same visual quality, enabling the efficient handling of large 3DGS scenes in practical applications.

In future work, we target a further increase in compression efficiency. Besides an adapted weighting of the indi-

vidual parameters of the smoothness loss, better and less correlated representations for rotations, shape and spherical harmonics will be investigated. The main focus, however, will be on the extension to 4D scenes with consideration of temporal dependencies, supporting the strength of 3D Gaussians in the representation of dynamic scenes.

Acknowledgements

This work has partly been funded, by the German Federal Ministry for Economic Affairs and Climate Action (To-HyVe, grant no. 01MT22002A) and the German Research Foundation (3DIL, grant no. 502864329).

References

- [1] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5460–5469, 2022. [1](#), [2](#), [4](#), [6](#)
- [2] K. U. Barthel, N. Hezel, K. Jung, and K. Schall. Improved evaluation and generation of grid layouts using distance preservation quality and linear assignment sorting. *Computer Graphics Forum*, 42(1):261–276, 2023. [3](#), [4](#), [8](#)
- [3] G. Chaurasia, S. Duchene, O. Sorkine-Hornung, and G. Drettakis. Depth synthesis and local warps for plausible image-based navigation. *ACM Trans. on Graphics*, 32(3), 2013. [2](#)
- [4] A. Chen, Z. Xu, A. Geiger, J. Yu, and H. Su. TensoRF: Tensorial radiance fields. In *Proc. European Conference on Computer Vision (ECCV)*, 2022. [2](#), [3](#)
- [5] F. Dellaert, S.M. Seitz, C.E. Thorpe, and S. Thrun. Structure from motion without correspondence. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 557–564, 2000. [2](#)
- [6] Chenxi Lola Deng and Enzo Tartaglione. Compressing explicit voxel grid representations: Fast NeRFs become also small. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1236–1245, 2023. [3](#)
- [7] S. Fridovich-Keil, A. Yu, M. Tancik, Q. Chen, B. Recht, and A. Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5491–5500, 2022. [1](#), [2](#)
- [8] S. J. Garbin, M. Kowalski, M. Johnson, J. Shotton, and J. Valentin. FastNeRF: High-fidelity neural rendering at 200fps. In *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14326–14335, 2021. [2](#)
- [9] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. M. Seitz. Multi-view stereo for community photo collections. In *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1–8, 2007. [2](#)
- [10] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Proc. of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, page 43–54, 1996. [2](#)
- [11] D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki, and A. Tabatabai. An overview of ongoing point cloud compression standardization activities: video-based (v-pcc) and geometry-based (g-pcc). *APSIPA Trans. on Signal and Information Processing*, 9, 2020. [2](#)
- [12] P. Hedman, P. P. Srinivasan, B. Mildenhall, J. T. Barron, and P. Debevec. Baking neural radiance fields for real-time view synthesis. In *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5855–5864, 2021. [2](#)
- [13] A. Karnewar, T. Ritschel, O. Wang, and N. Mitra. ReLU Fields: The little non-linearity that could. *ACM Trans. on Graphics*, 2022. [2](#)
- [14] B. Kerbl, G. Kopanas, T. Leimkuehler, and G. Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. on Graphics*, 42(4), 2023. [1](#), [3](#), [6](#), [2](#)
- [15] Teuvo Kohonen. Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, 43:59–69, 1982. [3](#)
- [16] Teuvo Kohonen. Essentials of the self-organizing map. *Neural Networks*, 37:52–65, 2013. [3](#)
- [17] G. Kopanas, J. Philip, T. Leimkühler, and G. Drettakis. Point-based neural rendering with per-view optimization. *Computer Graphics Forum (Proc. of the Eurographics Symposium on Rendering)*, 40(4), 2021. [2](#)
- [18] Yongjae Lee, Li Yang, and Deliang Fan. Mf-nerf: Memory efficient nerf with mixed-feature hash table. *ArXiv*, abs/2304.12587, 2023. [2](#), [3](#)
- [19] M. Levoy and P. Hanrahan. Light field rendering. In *Proc. of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, page 31–42, 1996. [2](#)
- [20] L. Li, Z. Shen, Z. Wang, L. Shen, and L. Bo. Compressing volumetric radiance fields to 1 mb. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4222–4231, Los Alamitos, CA, USA, 2023. IEEE Computer Society. [6](#), [2](#)
- [21] L. Liu, J. Gu, K. Z. Lin, T.-S. Chua, and C. Theobalt. Neural sparse voxel fields. *NeurIPS*, 2020. [2](#)
- [22] Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. In *Proc. International Conference on 3D Vision (3DV)*, 2024. [2](#)
- [23] R. Martin-Brualla, N. Radwan, M. S. M. Sajjadi, J. T. Barron, A. Dosovitskiy, and D. Duckworth. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. [2](#)
- [24] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *Proc. European Conference on Computer Vision (ECCV)*, 2020. [1](#), [2](#)
- [25] T. Müller, A. Evans, C. Schied, and A. Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. on Graphics*, 41(4):102:1–102:15, 2022. [2](#)
- [26] M. Quach, J. Pang, D. Tian, G. Valenzise, and F. Dufaux. Survey on deep learning-based point cloud compression. *Frontiers in Signal Processing*, 2022. [2](#)

- [27] C. Reiser, R. Szeliski, D. Verbin, P. Srinivasan, B. Mildenhall, A. Geiger, J. Barron, and P. Hedman. MERF: Memory-efficient radiance fields for real-time view synthesis in unbounded scenes. *ACM Trans. on Graphics*, 42(4), 2023. 1
- [28] D. Rückert, L. Franke, and M. Stamminger. ADOP: Approximate differentiable one-pixel point rendering. *ACM Trans. on Graphics*, 41(4), 2022. 2
- [29] S. Seitz and C. Dyer. Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision*, 35(2), 1999. 2
- [30] V. Sitzmann, J. Thies, F. Heide, M. Nießner, G. Wetzstein, and M. Zollhöfer. Deepvoxels: Learning persistent 3d feature embeddings. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2
- [31] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3d. *ACM Trans. on Graphics*, 25(3):835–846, 2006. 2
- [32] Grant Strong and Minglun Gong. Self-sorting map: An efficient algorithm for presenting multimedia data in structured layouts. *IEEE Trans. Multim.*, 16(4):1045–1058, 2014. 3
- [33] Grant Strong, Rune Jensen, Minglun Gong, and Anne C. Elster. Organizing visual data in structured layout by maximizing similarity-proximity correlation. In *ISVC* (2), pages 703–713. Springer, 2013. 3, 5, 8
- [34] C. Sun, M. Sun, and H. Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1, 2
- [35] T. Takikawa, A. Evans, J. Tremblay, T. Müller, M. McGuire, A. Jacobson, and S. Fidler. Variable bitrate neural fields. *ACM Trans. on Graphics*, 2022. 2, 3
- [36] A. Tewari, J. Thies, B. Mildenhall, P. Srinivasan, E. Tretschk, W. Yifan, C. Lassner, V. Sitzmann, R. Martin-Brualla, S. Lombardi, T. Simon, C. Theobalt, M. Nießner, J. T. Barron, G. Wetzstein, M. Zollhöfer, and V. Golyanik. Advances in Neural Rendering. *Computer Graphics Forum (EG STAR 2022)*, 2022. 2
- [37] A. Yu, R. Li, M. Tancik, H. Li, R. Ng, and A. Kanazawa. Plenoctrees for real-time rendering of neural radiance fields. In *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5732–5741, 2021. 2
- [38] J. Zhang, J. Huang, B. Cai, H. Fu, M. Gong, C. Wang, J. Wang, H. Luo, R. Jia, B. Zhao, and X. Tang. Digging into radiance grid for real-time view synthesis with detail preservation. In *Proc. European Conference on Computer Vision (ECCV)*, page 724–740, 2022. 2
- [39] Hongliang Zhong, Jingbo Zhang, and Jing Liao. VQ-NeRF: Neural reflectance decomposition and editing with vector quantization, 2023. 3

Compact 3D Scene Representation via Self-Organizing Gaussian Grids

Supplementary Material

6. Results for SSIM

In addition to the results from table 1, we also measure the structural similarity index measure (SSIM) in table 4. Those results correlate with the PSNR results from table 1.

7. Parameter Selection

The 3D Gaussian Splatting (3DGS) algorithm comes with many different training parameters. In the following we will give a brief overview of them and highlight the parameters that we changed for our new 3DGS training algorithm.

3DGS Parameters: To minimize the sorting time during training, we make slight changes to the default parameters of 3DGS. I.e. we reduce the number of Gaussians that are created during optimization. To achieve this, we modify the following five parameters:

- *Densification interval* $\in \mathbb{N}$: Determines how often the densification process is taking place. A low value corresponds to frequent densification, which results in a large amount of gaussians.
- *Densify grad threshold* $\in \mathbb{R}$: All Gaussians with a larger accumulated gradient for the xyz position, are split and cloned. A low threshold results in more Gaussians generated during training.
- *Densify min opacity* $\in \mathbb{R}$: A threshold that filters all Gaussians with smaller opacity during densification. A high value leads to fewer Gaussians.
- *Opacity reset interval* $\in \mathbb{N}$: In every x steps, the opacity of all Gaussians is set to 0.01.
- *Percent dense* $\in \mathbb{R}$: During the densification process, large Gaussians are split into two smaller copies with each 80% of the original size and small Gaussians are cloned identically. The percent dense parameter specifies the threshold at which a gaussian is classified as small or large. The value, which set is between 0 and 1, is multiplied by the extend of the scene and used as comparison.

Smoothness Reg. Parameters: The smoothness regularization first blurs the 2D attribute grids and then computes a error term with a Huber loss. This gives us several new parameters to tune:

- *Kernel size* $\in \mathbb{N}$: The size of the blur filter. A higher value leads to smoother target images.
- *Sigma* $\in \mathbb{R}$: The standard deviation of the gaussian blur.
- *Overall multiplier* $\lambda \in \mathbb{R}$: A multiplier that scales the loss of the Smoothness Regularization before adding it to the loss of 3DGS.
- *Separate multiplier* $\in \mathbb{R}$: A loss scalar for each of the five (position, color, opacity, scaling and rotation) 2D attribute

grids.

Sorting Parameters: The sorting algorithm that takes place after every densification can be adjusted with separate multipliers for each of the five 3DGS attributes (position, color, opacity, scaling and rotation).

In table 5 we give a detailed description of our parameter selection for our two different configurations.

8. Sorting performance

In this section, we provide additional measurements of the performance of our novel sorting algorithm.

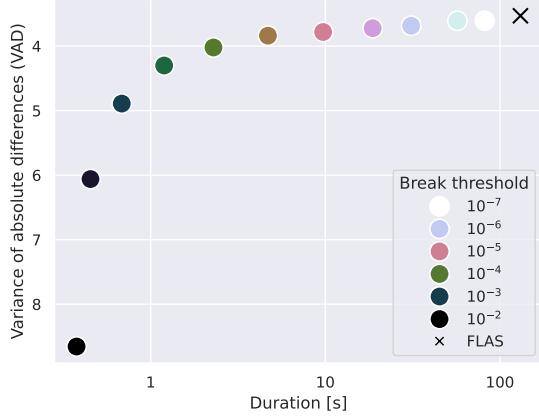


Figure 8. Behavior of the sorting algorithm under different break thresholds. Decreasing the relative L2 threshold trades off additional sorting quality for longer runtime. These values were measured with an *NVidia RTX 4090* on a random 512x512x3 grid. In all our training experiments, we have fixed this parameter to 10^{-4} .

The parameter that has the largest influence on the grid sorting performance is the threshold, which decides whether the relative reduction in L2 distance between two iterations should stop sorting with the current configuration, and potentially continue with the next permutation or next lower level. Decreasing it leads to a higher sorting quality, but the increase in iterations takes additional runtime. In Figure 8 we have plotted the sorting quality over the total runtime.

In Figure 9, we are showing the runtime performance of the sorting algorithm depending on the input grid size. With the chosen parameters, sorting will usually take less than 10 seconds during training. The highest number of Gaussians of any of our models over all datasets is the *Garden* scene with 4.37M Gaussians, which requires a grid of a side length of 2091. Thus even the largest of scenes in the used

Dataset Method/Metric	Mip-NeRF360		Tanks&Temples		Deep Blending		Synthetic-NeRF	
	SSIM↑	Size (MB)	SSIM↑	Size (MB)	SSIM↑	Size (MB)	SSIM↑	Size (MB)
Plenoxels †	0.626	2100	0.719	2300	0.795	2700	-	-
M-NeRF360 †	0.792	8.6	0.759	8.6	0.901	8.6	-	-
INGP-Base †	0.671	13	0.723	13	0.797	13	-	-
INGP-Big †	0.699	48	0.745	48	0.817	48	-	-
VQ-TensoRF †	-	-	0.913	3.3	-	-	0.960	3.6
3DGS	0.814	607.3	0.866	446.6	0.907	452.8	0.970	55.4
3DGS w/o SH	0.803	54.3	0.859	39.5	0.908	39.9	0.962	5.1
Ours-Q	0.772	23.9	0.838	11.9	0.891	8.4	0.955	2.2
Ours-S	0.763	18.2	0.831	9.9	0.892	7.2	0.953	1.7

Table 4. A comparison over structural similarity index measure (SSIM) between our method, the default 3DGS model and prior NeRF-based methods. Results with a † are directly copied from [14, 20]

Parameter / Method	3DGS	Ours-S	Ours-Q
3DGS			
Densification interval	100	1000	1000
Densify grad threshold	0.0002	0.00009	0.00007
Densify min opacity	0.005	0.1	0.1
Opacity reset interval	3000	∞	∞
Percent dense	0.01	1.0	0.1
Smoothness Reg.			
Kernel size	-	5	5
Sigma	-	4	3
Overall multiplier λ	-	1.0	1.0
Position weight	-	0.0	0.0
Color weight	-	0.0	0.0
Opacity weight	-	0.33	0.09
Scaling weight	-	0.0	0.0
Rotation weight	-	0.67	0.91
Sorting			
Position weight	-	1.0	1.0
Color weight	-	1.0	1.0
Opacity weight	-	0.0	0.0
Scaling weight	-	1.0	1.0
Rotation weight	-	0.0	0.0

Table 5. A detailed list of our training parameters compared to the default 3DGS training. An opacity reset interval of ∞ denotes that we deactivate the opacity reset.

datasets can be sorted in well below a minute.

9. Quantization & Compression

Once training has finished, we can use off-the-shelf image compression methods to store the data on disk.

We store the un-activated Gaussian parameters, retaining the activation methods from 3DGS: sigmoid actiation for opacity and exponential activation for the scale. As mentioned in Section 3, we introduce an additional exponential activation for the coordinates x used during training, allow-

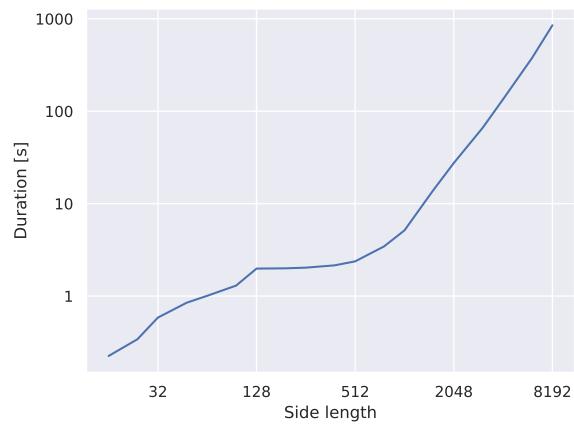


Figure 9. Runtime of the sorting algorithm over the side length of the 2D grid. Here measured using 3 layers for the grids, while there may be up to 14 when sorting Gaussians.

ing for more relative precision in coordinates in the center of the scene:

$$x = \text{sign}(x_{\log}) \times (\exp(|x_{\log}|) - 1) \quad (2)$$

We clip the RGB values (DC values of the spherical harmonics) to the range $[-2, 4]$ and normalize it to the range $[0, 1]$. Opacity values are clipped to the range $[-6, 12]$ and similarly normalized. These ranges are chosen to cover the 1st to 99th percentile of these values across all models from the 360 dataset.

We quantize the different attributes by rounding them to the closest value of a linear range of q total values, with $q_{\text{coords}} = 2^{13}$, $q_{\text{scale}} = q_{\text{rotation}} = 2^6$ and $q_{\text{opacity}} = 2^5$. Clipping, normalization and quantization are done post-training in our method. In future work, it would be interesting to investigate applying them during training for potentially further improving the results.

Name	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	Size (MB) \downarrow
PLY	25.41	0.880	0.152	104.89
NPZ	25.41	0.880	0.152	76.10
JXL ll	25.41	0.880	0.152	66.05
PNG 16	24.08	0.873	0.157	37.00
EXR	25.36	0.878	0.154	30.56
EXR+JXL q	24.89	0.862	0.170	17.25

Table 6. The truck scene, trained with our method, compressed in different formats. Marked in bold is the compression method used in our experiments for all datasets.

We compress the RGB grid with JPEG XL, as an 8-Bit image with a quality level of 90, and store all other attributes as 32-Bit OpenEXR images with zip compression.

Table 6 contrasts different compression methods on the Truck scene. Described in the text above was the *EXR+JXL q* method, which we used as the default method for all data sets in the comparisons in the paper.

Notably, the PLY file when trained with our method is of smaller size and of higher quality than training with the default 3DGS parameters (which yields 24.90 PSNR at 174 MB). *NPZ* is using NumPy’s `save_compressed` method to directly store the tensors. *JXL ll* stores all tensors as lossless JPEG XL images. *PNG 16* truncates the value below the 1st and above the 99th percentile (over all 360 datasets), then normalizes the data to the [0, 1] range and stores them as 32-Bit PNGs. *EXR* stores the values as 32-Bit OpenEXR files with zip compression. Even though we are not using spherical harmonics, this is providing close to the same quality as 3DGS with spherical harmonics (25.44 PSNR at 615 MB), at a compression rate of over 20x.

In our work, we have demonstrated the feasibility of organizing Gaussians into 2D grids for a compact representation. We have not used spherical harmonics (SH), as they have a relatively small effect on scene quality, while taking three quarters of all space in vanilla 3DGS (45 of 59 attributes in 3DGS are spherical harmonics, with the default third degree SH). We believe that a better space-for-quality trade-off is required for efficient representation of view-dependent effects. When this is identified in future work, we think it will be worthwhile to perform additional fine-tuning of the quantization and compression parameters.