

LODGE: Level-of-Detail Large-Scale Gaussian Splatting with Efficient Rendering

Jonas Kulhanek^{1,4}, Marie-Julie Rakotosaona¹, Fabian Manhardt¹, Christina Tsalicoglou¹, Michael Niemeyer¹, Torsten Sattler⁵, Songyou Peng², Federico Tombari^{1,3}

¹ Google, ² Google DeepMind, ³ Technical University of Munich,

⁴ Czech Technical University in Prague, Faculty of Electrical Engineering,

⁵ Czech Technical University in Prague, Czech Institute of Informatics, Robotics and Cybernetics

<https://lodge-gs.github.io/>

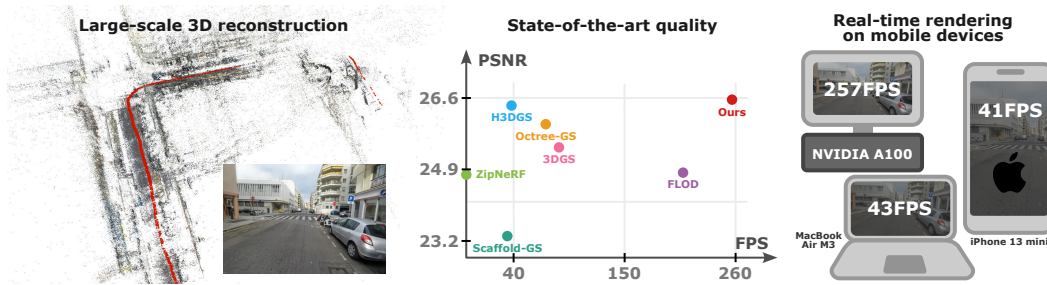


Figure 1: **LODGE**. Applied to large-scale 3D scenes, LODGE achieves outstanding quality while maintaining superior rendering speeds. Furthermore, it enables real-time rendering on mobile devices.

Abstract

In this work, we present a novel level-of-detail (LOD) method for 3D Gaussian Splatting that enables real-time rendering of large-scale scenes on memory-constrained devices. Our approach introduces a hierarchical LOD representation that iteratively selects optimal subsets of Gaussians based on camera distance, thus largely reducing both rendering time and GPU memory usage. We construct each LOD level by applying a depth-aware 3D smoothing filter, followed by importance-based pruning and fine-tuning to maintain visual fidelity. To further reduce memory overhead, we partition the scene into spatial chunks and dynamically load only relevant Gaussians during rendering, employing an opacity-blending mechanism to avoid visual artifacts at chunk boundaries. Our method achieves state-of-the-art performance on both outdoor (Hierarchical 3DGS) and indoor (Zip-NeRF) datasets, delivering high-quality renderings with reduced latency and memory requirements.

1 Introduction

Novel view synthesis is a central area of research in computer vision that enables applications in AR/VR, gaming, interactive maps, and others. The field has recently received a lot of attention with the advent of Neural Radiance Fields (NeRFs) [22] and 3D Gaussian Splatting (3DGS) [10] — the latter pushing the range of applications further as it enables real-time rendering. With the rising popularity of NeRFs and 3DGS, there is an increasing interest in applying such methods to ever larger and more complex scenes [27, 11, 5]. However, standard methods do not scale well to such large-scale environments [11, 18]. The core issue lies in the representation: to capture fine details, a

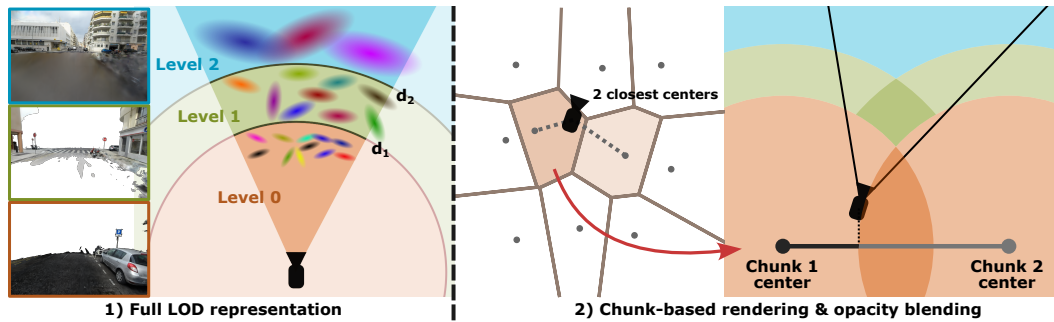


Figure 2: **Method overview.** **Left:** The scene is represented with multiple LODs; ‘active Gaussians’ are selected during training based on camera distance. **Right:** We cluster cameras into chunks, pre-compute ‘active Gaussians’ per chunk, and render the two nearest chunks with ‘opacity blending’.

high number of Gaussians is required. Consequently, even distant regions of the scene are populated with dense Gaussians representing fine-grained geometry, *ie.*, elements that contribute little to the final rendered image. This leads to significant inefficiencies during rendering, as many far-away Gaussians are processed despite having minimal or no visible impact. Furthermore, memory limitations pose an additional challenge: not all Gaussians can fit into GPU memory simultaneously, which is particularly problematic for mobile or low-end devices where memory is severely constrained.

In computer graphics this problem has been extensively studied and — in the context of mesh-based rendering — effectively addressed using level-of-detail (LOD) strategies. These techniques render lower-resolution versions of in-game assets when they are far from the camera and progressively replace them with higher-resolution versions as the camera moves closer. While there are approaches that propose LOD for 3DGS in large-scale scenes [18, 11, 29, 28], they primarily focus on improving rendering speed, without limiting the number of Gaussians loaded in GPU memory, making rendering on smaller devices a challenge. Such methods often require to recompute the subset of Gaussians that need to be rendered at every new frame, adding overhead to the rendering. More importantly, this requires all Gaussians from all different LODs (even more than in 3DGS [10]) to be in GPU memory at all time. Finally, existing LOD approaches [28, 29] require careful parameter tuning for each scene to achieve good quality and performance.

On the other hand, our method is designed to both improve the rendering speed for large-scale scenes and to limit the number of Gaussians needed in memory to enable applications on embedded devices. Similarly to existing LOD based methods, we represent the scene as multiple sets of Gaussians with varying level of detail. However, we propose to define regions in space around a cluster center. Each region activates a fixed set of Gaussians from the precomputed LODs to avoid overhead computation between different frames. Our contributions can be summarized as follows:

- We propose a novel LOD representation for 3DGS which, unlike previous methods [28, 29, 18], does not recompute the list of used Gaussians at each frame and thus can be accelerated and compacted to allow rendering of large-scale scenes even on mobile devices.
- We further design a strategy to automatically select optimal hyperparameters for splitting LODs, whereas most other methods require hyperparameters to be tuned manually for each 3D scene.
- To accelerate rendering further, we split the scene into chunks and pre-compute sets of active Gaussians per chunk.
- Finally, we introduce a novel opacity interpolation scheme to produce visually pleasing rendering and remove any artifacts when transitioning between chunks.

We show that our method outperforms state-of-the-art (SOTA) approaches in terms of quality and rendering speed whilst reducing the number of Gaussians in memory.

2 Related Work

Novel view synthesis has received great attention recently, mainly thanks to the advent of neural radiance fields (NeRFs) [22] and later 3D Gaussian Splatting (3DGS) [10], which offered an on

par alternative with substantially faster rendering. Subsequently, 3DGS was further extended in many ways to handle antialiasing [37, 20, 31], to offer better geometry [38, 8], or to accommodate different camera models [14, 35]. There has been a lot of effort on making 3D Gaussian Splatting representation smaller by compressing attributes of Gaussians [20, 3, 7, 24]. Similarly to our approach, this also reduces the required memory and makes rendering faster. Other approaches modify the rendering procedure to make rendering more efficient [21, 16, 12, 4, 36, 42]. Finally, a lot of interest has been directed towards optimizing scene representation from images captured under different time-of-day, seasons, or exposure levels [13, 25, 15, 39]. We consider these approaches orthogonal to ours as they can also be applied on top of our method. Similarly, some works [21, 17, 25] focused on removing the number of Gaussians by proposing alternative densification and pruning strategies [21, 17], or by pruning Gaussians with little/no contribution to the rendering process [7, 3, 25] – similar to our approach. While these methods make rendering efficient for small-scale scenes, very large scenes are still not tractable to render in real-time without sacrificing quality.

Reconstructing large 3D scenes. When dealing with extremely large scenes that can span several city blocks, another line of works further proposed to split the scene into sub-regions, and reconstructing each one separately. First, as for NeRFs, [32, 33, 40] used per-region MLPs and special rendering procedures to better handle far-away regions. Later, similar techniques were also applied in the context of 3DGS [18, 6, 11, 19]. These methods can be used to train from large image collections efficiently. Our LOD representation can be built from an existing model and can thus be applied on top of these approaches. However, these approaches use a large number of Gaussians in GPU memory and cannot be rendered on low-end devices – a problem addressed by our method.

Level-of-detail 3DGS. The main contribution of our work is a novel level-of-detail (LOD) representation. Nevertheless, we are not the first to propose incorporating LODs into 3DGS [29, 11, 23, 28, 18, 34]. While H3DGS [11] and CLOG [23] build continuous multi-level representations, FLoD [29] and OctreeGS [28] use discrete levels, each being a set of Gaussians (similar to our approach). The authors start from a coarse set of Gaussians, which are then progressively densified to obtain the finer levels. Unlike them, our method builds a LOD structure on top of the standard reconstruction and can be applied to various existing methods while we observed that coarse-to-fine strategy tends to fail on large-scale scenes as the densification fails when the coarse set is too sparse. Finally, all existing LOD methods [29, 11, 28, 23] dynamically select the set of ‘active Gaussians’ (Gaussians used in the rasterization) for each rendered frame. This leads to slower rendering and requires all Gaussians to be loaded in GPU memory. Unlike them, our method splits the 3D scene into sub-regions and for each, it pre-select sets of ‘active Gaussians’, increasing rendering speed and lowering memory usage.

3 Method

Our method aims to make large-scale 3D Gaussian Splatting reconstruction render fast even on mobile devices. To this end, we introduce a novel level-of-detail (LOD) representation and chunk-based caching scheme (see Fig. 2), leading to fast rendering whilst reducing the memory footprint.

Preliminaries: 3D Gaussian Splatting. 3D Gaussian Splatting [10] represents a 3D scene as a set of Gaussian primitives, each having a mean (position) μ , covariance matrix Σ , opacity o , and view-dependent color. The Gaussians are splatted into 2D screen space (obtaining 2D means μ' and covariances Σ' [10, 43]), sorted in z-axis, and composited using alpha blending. The blending weight α for pixel $\mathbf{p}$ of 2D Gaussian with mean μ' , covariance Σ' , and opacity o is then given by the Gaussian function:

$$\alpha = oG(\mathbf{p}), \quad G(\mathbf{p}) = e^{-\frac{1}{2}(\mathbf{p}-\mu')^T(\Sigma')^{-1}(\mathbf{p}-\mu')}. \quad (1)$$

Upon splatting of the 3D Gaussians, the 2D screen is split into tiles of 16 pixels each, with each 2D Gaussian being assigned to all (possibly many) tiles with which it overlaps. Subsequently, the Gaussians for each tile are then sorted and alpha blended in the rasterization process using Eq. (1).

Preliminaries: Importance pruning. As part of our LOD representation, we adapt the importance pruning introduced in RadSplat [25]. During 3DGS training, many Gaussians can become less visible when their opacity decreases or when a Gaussian in front of them becomes opaque. While 3DGS periodically removes Gaussians with low opacity, it does not remove occluded Gaussians (hidden behind other scene geometry). In RadSplat [25], the authors propose to measure each Gaussian’s importance (importance score τ_i) by taking the maximum over the Gaussian’s contribution (rendering weight in alpha blending) to any pixel in all training cameras. By removing all Gaussians with

importance score lower than some threshold, we can effectively prune Gaussians with very little impact on rendering. This makes rendering faster and also reduces memory, which is particularly important when working with low-end devices. We follow [25] and prune twice during training.

Level of Detail (LOD) representation. We propose to represent 3D scenes as multiple sets of Gaussians that correspond to different levels of detail as visualized in Fig. 2. As motivated in the introduction, distant regions of the scene are often represented by a large number of Gaussians that contribute little to the final rendering. Yet, these Gaussians still need to be evaluated by the renderer, causing excessive memory usage and computations. Individual pixels with a large number of visible Gaussians (Gaussians projected into them) can significantly slow down rendering due to how the pixels are processed in 16×16 patches. The main goal is, therefore, to reduce the number of pixels with a large number of visible Gaussians, as shown in Fig. 3.

Therefore, we propose to represent faraway regions with less detailed sets of Gaussians and nearby regions with more detailed sets. To this end, we define multiple LOD levels: $\mathcal{G}^{(l)}$; $L > l > 0$, with $\mathcal{G}^{(0)}$ denoting the most detailed set, obtained by optimizing on the original set of images [10, 11]. We assume each LOD level $\mathcal{G}^{(l)}$ is constructed (as explained in the next section) such that a sufficient rendering quality is achieved when observed from a distance of at least d_l . When rendering the LOD from a given camera pose, we select a subset of Gaussians from each LOD level (see Fig. 2). We call this set the ‘active Gaussians’, and for camera center $\mathbf{c}$, it is defined as

$$\tilde{\mathcal{G}}(\mathbf{c}) = \bigcup_{l=0}^{L-1} \left\{ g_i \in \mathcal{G}^{(l)} : d_l \leq \|\mu_i^{(l)} - \mathbf{c}\|_2 < d_{l+1} \right\}, \quad (2)$$

where $d_0 = 0$, $d_L = \infty$, and $\mu_i^{(l)}$ is the mean of Gaussian g_i from level l . As shown in Fig. 3, LOD rendering reduces long tail of the per-pixel visible Gaussians distribution accelerating the rendering.

Building LOD representation. We aim to construct each LOD level $\mathcal{G}^{(l)}$ such that a sufficient rendering quality is achieved when observed from a distance of at least d_l . To this end, we draw inspiration from the 3D filter proposed by MipSplatting [37]. As described there, while the rendered image is a 2D projection of a continuous 3D scene given by a set of Gaussians, the image itself is instead a grid in which each pixel coordinate defines the sampling from the continuous 3D signal. For a sampling interval of 1 in screen space, the pixel interval in 3D world space at depth d is $T = \frac{d}{f}$, with f being the focal length. According to the Nyquist’s theorem [26, 30, 37], components of the signal can only be reconstructed if they are sampled at intervals smaller than $2T$. Therefore, Gaussians smaller than $2T$ will only result in aliasing artifacts [37] and increase both memory usage and rendering time. To enforce Gaussians of size larger than $2T$, we follow Mip-Splatting [37] and convolve each Gaussian with a smoothing 3D filter. The resulting Gaussian function (for a Gaussian with mean μ and covariance Σ and for depth d) is given by

$$\tilde{G}(\mathbf{x}) = \sqrt{\frac{|\Sigma|}{|\Sigma + \frac{sd}{f} \cdot \mathbf{I}|}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T (\Sigma + \frac{sd}{f} \cdot \mathbf{I})^{-1} (\mathbf{x}-\mu)}, \quad (3)$$

where s is a hyperparameter.

To build the lower-detail representations $\mathcal{G}^{(l)}$ that are viewed at distances larger than d_l , we copy Gaussians from $\mathcal{G}^{(0)}$ and add a smoothing 3D filter (Eq. (3)) for depth d_l . While adding a smoothing 3D filter alone does not directly lead to fewer Gaussians, a significant number of Gaussians will become redundant, reducing their contributions in alpha compositing. Eventually, using the importance score from [25], we iteratively prune unused Gaussians. Note that we always employ a few fine-tuning steps to correct for errors introduced by pruning Gaussians. For the fine-tuning, we use LOD rendering with levels up to the currently optimized LOD level. More details in the *supp. mat.*

Selecting depth thresholds. An important question remains: how should the depth thresholds d_l be selected to maximize the rendering performance. The renderer operates on pixels grouped in 16×16 tiles and the rendering performance is heavily influenced by the number of Gaussians processed inside the same 16×16 -pixels tile. This is because all threads within a tile process the union of all visible Gaussians and must wait for the slowest thread to complete rasterizing. While it is difficult to theoretically analyze the impact of thresholds on the number of processed Gaussians, we can estimate this number (cost) for different thresholds - by rendering a subset of training views - and choose the one that minimizes the average number of Gaussians per tile. In Fig. 4, we show the cost distribution

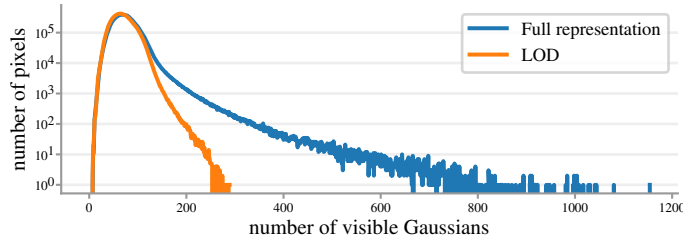


Figure 3: **Visible Gaussians histogram.** For each pixel we compute the number of visible Gaussians and show the histogram for base model ‘Full representation’ and LOD rendering ‘LOD’.

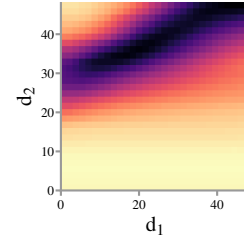


Figure 4: **LOD threshold cost function.** Visualized for 2 depth thresholds. Darker is lower cost.

for two LOD levels at depths d_1 and d_2 . Notice how the value of the minimum is similar for set $\{(x, ax + b) : x \in \mathbb{R}\}$. This enables us to leverage a simple greedy strategy - starting at d_1 , iteratively adding more thresholds. This reduces the complexity of the search problem to a linear one.

Reducing memory with chunk-based rendering. While LOD rendering (described so far) makes rasterization fast by reducing the number of visible Gaussians, all Gaussians still need to be loaded in GPU memory. Moreover, the ‘active Gaussians’ need to be continuously recomputed, leading to overhead computations. This can be prohibitive for small devices where the amount of available memory is limited. To address this, we propose to split the scene into different regions called ‘chunks’ and store a fixed set of ‘active Gaussians’ per chunk. It is important to note that the set of ‘active Gaussians’ associated with a chunk is not merely a subset of all Gaussians located within its spatial boundaries (see Fig. 2). Instead, it represents the entire scene, with regions closer to the chunk center modeled at higher levels of detail. When rendering an image, the rasterizer simply uses the ‘active Gaussians’ pre-computed for the closest chunk.

To define the chunks, we split the scene by performing a K-means clustering on the training camera positions (more details in *supp. mat.*). For each chunk with center $\mathbf{c}$, the ‘active Gaussians’ are defined by Eq. (2) computed at the chunk center (rather than the camera position), with depths d_i offset by the chunk radius (distance to next closest chunk center) to ensure sufficient resolution for all camera positions inside the chunk.

Visibility filtering. Given our LOD chunks, we further accelerate rendering by filtering the set of Gaussians for each LOD chunk. Following RadSplat [25] we additionally perform importance pruning for each LOD chunk, *ie.*, we compute per-Gaussian importance scores and remove all Gaussians with importance score lower than a fixed threshold. However, to further increase robustness, we add additional views by adding random perturbations to existing training views within the LOD chunk. To this end, we employ the original camera positions, but sample random orientations. Note, that if orientations were included alongside positions in defining LOD chunks, visibility filtering could achieve an additional reduction in the number of loaded Gaussians. Nonetheless, this gain would not be significant and would reduce the method’s ability to handle rapid camera rotations and wide fields of view, both of which are common in practical applications (e.g. AR/VR).

Opacity blending for smooth cross-chunk transitions. Having constructed these chunks, a naive approach would be to simply render the ‘active Gaussians’ for the chunk center. While this speeds up rendering and reduces memory, it also introduces sharp changes when rendering a dynamic video of a camera trajectory (see Fig. 7). These artifacts are inherently caused by an abrupt change in the ‘active Gaussians’ while moving, without employing any form of smoothing filter during the transition. To resolve the issue, we propose a smoothing strategy leveraging the two closest chunks (see Fig. 2). When rendering an image, we first seek the two closest chunk centers. We take their respective sets of active Gaussians and combine them. Afterwards, we modulate the opacity of Gaussians, which are not in the intersection of the two sets of active Gaussians, using

$$\hat{\alpha}_i = \alpha_i t, \quad t = \min(1, \max(0, \bar{t})), \quad \bar{t} = \frac{(\mathbf{c} - \mathbf{m}_o)^T (\mathbf{m}_f - \mathbf{m}_o)}{\|\mathbf{m}_o - \mathbf{m}_f\|_2^2}, \quad (4)$$

where $\mathbf{c}$ is the current camera position, $\mathbf{m}_f$ and $\mathbf{m}_o$ are the two closest chunk centers – $\mathbf{m}_f$ being the chunk center to which Gaussian i belongs, and $\mathbf{m}_o$ being the other chunk (not containing Gaussian i). Note that $\bar{t}$ is the normalized length of the projection of $(\mathbf{c} - \mathbf{m}_o)$ onto the line connecting the two chunk centers. We use the length of projection instead of the Euclidean distance to achieve a

	SmallCity					Campus				
	PSNR	SSIM	LPIPS	#G	FPS	PSNR	SSIM	LPIPS	#G	FPS
Zip-NeRF [2]	26.30	0.785	0.368	–	0.09	22.04	0.781	0.416	–	0.20
3DGS [10]	25.42	0.776	0.394	1375K	85.25	24.14	0.785	0.430	1142K	47.38
Mip-Splatting [37]	25.36	0.775	0.394	1445K	75.42	23.96	0.784	0.430	1188K	66.79
Scaffold-GS [20]	23.31	0.753	0.362	1347K	71.38	20.43	0.754	0.436	1335K	46.84
H3DGS [11]	26.42	0.807	0.331	7093K	38.07	24.60	0.798	0.396	6186K	34.32
FLOD [29]	24.82	0.758	0.429	497K	208.41	24.10	0.777	0.453	595K	120.61
OctreeGS [28]	25.98	0.807	0.326	1008K	120.27	25.22	0.800	0.408	642K	119.21
CityGS [18]	25.29	0.772	0.401	2615K	114.07	24.82	0.794	0.419	1881K	121.67
Ours	26.57	0.815	0.325	877K	257.46	24.75	0.803	0.394	1464K	218.96

Table 1: **Hierarchical 3DGS comparison.** We compare baselines on SmallCity and Campus scenes. **Ours** achieves fastest rendering while outperforming others in terms of rendering quality. The first, second, and third values are highlighted.

smooth transition even when the camera doesn’t pass through the chunk center. Given the union of the two sets of active Gaussians with modified opacity, we proceed with the standard rasterization step. However, note that only the union of the two chunks needs to be loaded in memory and at each rendering pass, we only need to update the opacity of the symmetric difference of the two active Gaussian sets.

In practice, reloading of LOD splits can be done in a background process that does not affect the renderer’s runtime. We begin by loading the active Gaussians’ properties (e.g., positions, colors) into GPU memory. As the camera moves, we only need to modulate the opacity of the two loaded chunks until the camera passes the center of the chunk. Then, we remove the Gaussians from the previous chunk, keep the ones from the closest one, and load Gaussians from the second closest chunk. The opacity blending at this point assigns weights close to 1 to all Gaussians from the closest chunk so there are no artifacts during loading the next chunk – even if loading of the next chunk is delayed.

4 Experiments

To validate our method, we conduct experiments on large-scale indoor and outdoor datasets, comparing our method to other SOTA approaches. We further analyze individual components of our method and evaluate rendering speed on various mobile devices. We report standard PSNR, SSIM, and LPIPS (VGG) metrics, but also FPS and the number of Gaussians loaded in GPU memory (#G). The number of loaded Gaussians serves as a proxy for memory usage, offering a fairer comparison across 3DGS implementations (as opposed to peak-memory usage) as it does not depend on the actual implementations but rather on the algorithm itself. For more details on implementation and the evaluation protocol, we kindly refer the reader to the *supp. mat.* All baselines and our method use a single NVIDIA A100 SXM4 40GB GPU for training and evaluation. For mobile experiments, we used two iPhones (13 Mini and 15 Pro), as well as two lower-end laptops without a powerful GPU (MacBook Air M3 and HP Chromebook).

Datasets & baselines. We use two larger-scale datasets to validate our approach, *ie.*, two outdoor scenes from Hierarchical 3DGS dataset [11] and three indoor scenes from Zip-NeRF dataset [2]. Each scene consists of around 1000-2000 images. We compare our approach against the following baselines: Current SOTA on the Zip-NeRF dataset - **Zip-NeRF** [2] - slow to train and without real-time rendering. Traditional 3DGS baselines - **3DGS** [10], **Mip-Splatting** [37], and **Scaffold-GS** [20]. LOD-based methods - **H3DGS** [11] (SOTA on the Hierarchical 3DGS dataset), **FLOD** [29], **Octree-GS** [28], and **CityGS** [18]. Note that H3DGS, Octree-GS, FLOD were trained for 45K, 40K, and 100K iterations, while **ours** was trained for 36K iterations. For FLOD we use 3-4-5 LOD rendering and for H3DGS we employ their default $\tau = 6$. More details can be found in the *supp. mat.*

Evaluation on Hierarchical 3DGS dataset. As shown in Tab. 1, Zip-NeRF struggles on these large scenes, despite requiring much longer training (200K iterations vs. 36K for **ours**). This likely stems from its use of a contiguous scene representation with limited resolution, unlike 3DGS,

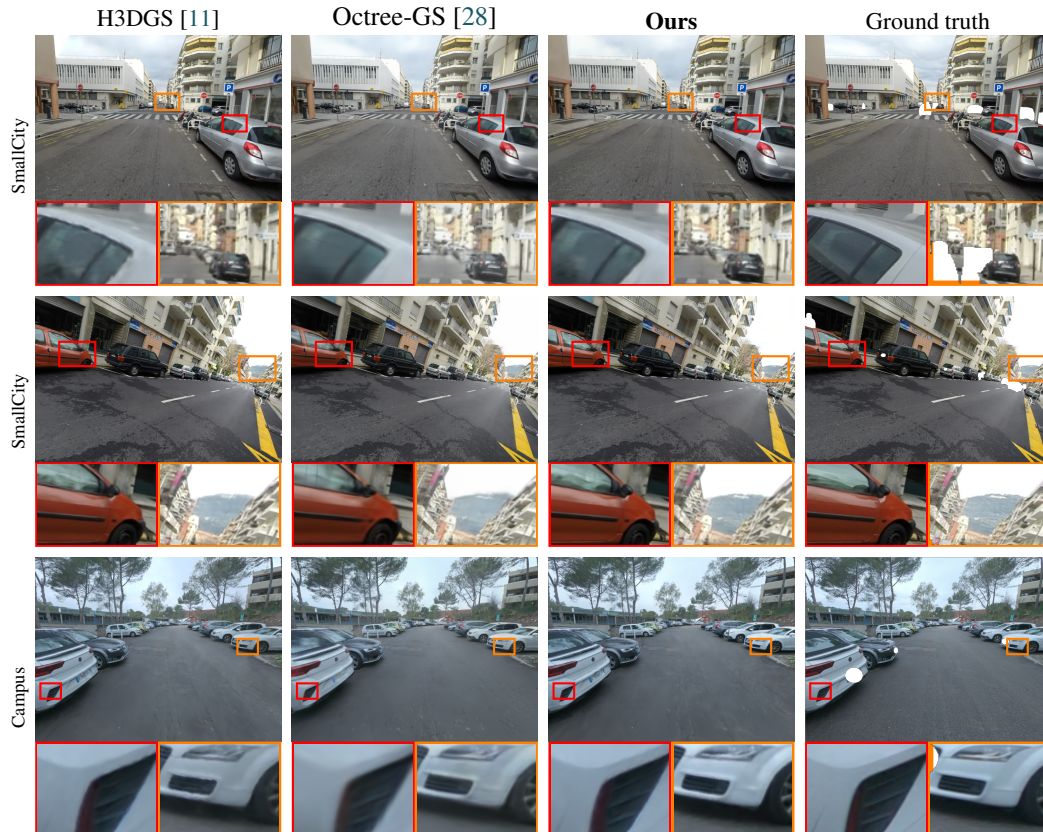


Figure 5: **Hierarchical 3DGS qualitative results.** We compare H3DGS [11] and Octree-GS [28] on SmallCity and Campus scenes. We highlight **close-up region** (left) and **far-away region** (right).

which sparsely encodes only occupied space. Among non-LOD baselines, 3DGS [10] and Mip-Splatting [37] offer competitive quality and rendering speed, whereas Scaffold-GS [20] lags behind, likely due to suboptimal densification in large-scale scenes. H3DGS [11] attains high visual fidelity, but incurs significant overhead from per-frame LOD splitting and the large number of Gaussians, making it slower than other methods. FLOD fails to generate sufficient Gaussians due to their coarse-to-fine training scheme, discarding fine details early and struggling to recover them later, even though it uses three times more training iterations than related methods. Compared to Octree-GS [28] and CityGS [18], **ours** achieves higher overall quality while being twice as fast. Interestingly, both Octree-GS and CityGS excel on Campus in PSNR but underperform in SSIM and LPIPS. We attribute this to our method being more robust to exposure variation but less effective in preserving perceptual similarity under appearance changes. In summary, our method consistently achieves the best results across all scenes whilst maintaining superior rendering efficiency. In qualitative comparisons (Fig. 5), we highlight a **close-up** and **distant** regions to assess both local details and long-range fidelity. Overall, our method delivers sharper reconstructions in both close and distant regions. In contrast, Octree-GS appears blurrier nearby and suffers from color shifts at a distance, while H3DGS often exhibits jagged geometry, potentially due to its reliance on depth supervision.

Zip-NeRF dataset evaluation. As seen in Tab. 2, Zip-NeRF achieves the highest accuracy but requires long training (200K iterations) and does not support real-time rendering. We still include it for reference. Among real-time-capable methods, our approach achieves the best trade-off between quality and speed. CityGS [18] slower than **ours** while also having worse quality. FLOD [29] renders fast but loses fine details due to early pruning in its coarse-to-fine training. H3DGS [11] maintains high fidelity but is significantly slower, requiring over 10M Gaussians and per-frame LOD graph cuts. Octree-GS [28] offers good quality and decent speed, but is slower than ours due to per-frame MLP evaluations. Finally, Scaffold-GS [20] lags behind other non-LOD baselines in both quality and speed. As for qualitative evaluation (Fig. 6), FLOD tends to produce blurrier results and often misses

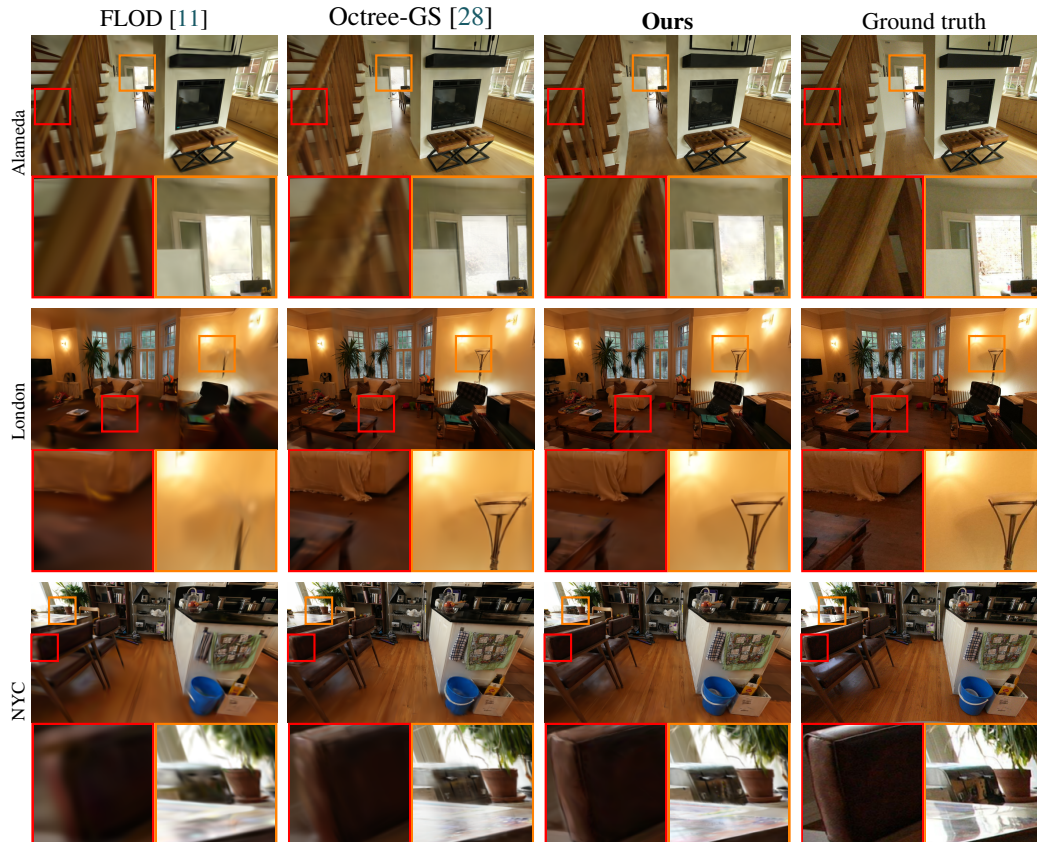


Figure 6: **Zip-NeRF dataset qualitative comparison.** On scenes: Alameda, London, NYC. We highlight **close-up region** (left) and **far-away region** (right).

Method	Alameda			London			NYC		
	PSNR	SSIM	FPS	PSNR	SSIM	FPS	PSNR	SSIM	FPS
<i>Zip-NeRF</i> [2]	22.97	0.738	0.13	26.76	0.822	0.13	28.21	0.845	0.13
3DGS [10]	21.75	0.707	90.12	25.43	0.795	99.19	26.33	0.829	79.69
Mip-Splatting [37]	21.86	0.709	95.52	25.55	0.797	95.56	26.33	0.829	85.24
Scaffold-GS [20]	20.93	0.714	65.61	22.52	0.745	75.71	25.97	0.817	78.74
H3DGS [11]	22.21	0.739	27.82	26.34	0.823	30.49	27.28	0.849	33.11
FLOD [29]	21.35	0.666	276.52	24.38	0.753	195.06	25.01	0.781	260.85
OctreeGS [28]	22.94	0.734	119.83	26.04	0.817	153.06	27.05	0.839	146.29
CityGS [18]	22.43	0.729	174.07	25.87	0.809	190.10	26.55	0.839	144.91
Ours	22.41	0.741	229.99	26.34	0.818	252.58	27.40	0.849	280.22

Table 2: **Zip-NeRF dataset comparison.** On scenes: Alameda, London, NYC. We report PSNR, SSIM, and FPS. The **first**, **second**, and **third** values are highlighted.

distant details (e.g., lamp in 2nd row). Octree-GS struggles with close-up sharpness (e.g., railing in 1st row) compared to our method. For far regions, both methods perform similarly, though our model better preserves high-frequency highlights (e.g., London scene), while Octree-GS slightly reduces color artifacts under exposure variation due to its MLP-based design. In summary, our method can deliver sharper close-up reconstructions and comparable or superior fidelity for far regions, while maintaining faster rendering speed.

In the Hierarchical 3DGS dataset [11], the camera follows a linear trajectory with chunks aligned along it; therefore, only the two nearest chunks are relevant. In Zip-NeRF [2], however, the chunks fully cover the scene and the camera may traverse diagonally, potentially being between three or

	PSNR	Time	#Vis. G	#G
<i>Full representation</i>	26.62	15.17	925k	2639K
LOD (d=10)	26.45	5.61	324K	3465M
LOD (d=10,28)	26.50	4.75	209K	3815K
LOD (d=10,28,47)	26.46	4.17	182K	4016K
LOD (d=10,28,47,63)	26.50	4.07	172K	4145K
LOD (d=15,70)	26.55	5.24	267K	3377K
LOD (d=10,28) + clusters	26.54	3.62	244K	795K
LOD (d=10,28) + clusters + vis. filtering	26.55	3.15	185K	612K
Opacity blending	26.57	3.88	268K	877K

Table 3: **Performance analysis.** We show impact of different number of LOD levels and impact of LOD clusters, visibility filtering, and opacity blending on rendering time and quality (PSNR). For LOD, we vary the number of levels and the depths d . Last three rows use chunk-based rendering.



Figure 7: **Ablation study.** We compare the full representation (without LOD), full LOD, rendering from closest and second closest LOD chunks, and opacity blending.

more chunks – raising the question of whether opacity blending of only two chunks is sufficient. Empirically, no artifacts were observed across Hierarchical 3DGS [11], Zip-NeRF [2], and Mip-NeRF 360 [1]. We hypothesize that: **a)** Close to the training camera distribution, neighboring chunks are well constrained, producing near-identical renderings even without blending. **b)** Far from the distribution, viewpoints lie at chunk boundaries where only the two closest chunks dominate. Thus, two-chunk blending is generally sufficient.

Performance analysis & design validation. We conduct an ablation study on the SmallCity scene to assess the impact of key components of our method on rendering quality and speed. Quantitative results are shown in Tab. 3 (corresponding to SmallCity in Tab. 1), reporting PSNR, render time, number of visible Gaussians, and memory usage. Time for reloading Gaussians at chunk boundaries is excluded, as it is handled asynchronously. Qualitative results are shown in Fig. 7. We start with the base method (‘Full representation’), which includes importance pruning but no LOD. Adding up to four LOD levels significantly improves rendering speed (up to $3\times$ with just one level) at a minimal cost in PSNR and sharpness due to more aggressive pruning. As additional levels yield diminishing returns, we decided to always employ two LODs for all experiments, as a good speed and quality trade-off. Thresholds were selected automatically (Section 3, ‘Selecting depth thresholds’), and we show that when instead setting them manually (15 and 70 m), performance worsens compared to our automatic selection ‘LOD (d=10,28)’.

The last three rows evaluate our chunk-based rendering. Clustering camera positions and adjusting depth thresholds by chunk radius improves quality (‘LOD (d=10,28) + clusters’), but increases Gaussian count due to finer LOD resolutions. Visibility filtering (‘+ vis. filtering’) further reduces both visible and loaded Gaussians, and thus rendering time. However, using the active Gaussians from only the chunk centers introduces visual artifacts at chunk boundaries, noticeable as sharpness discontinuities (Fig. 7, second column from right). To circumvent this issue, we introduce opacity blending between adjacent chunks (last row, corresponding to ‘Ours’ in Tab. 1), which smoothens transitions. While it slightly increases render time and Gaussian count, it runs faster and more memory-efficient than its non-chunk-based LOD counterpart, being particularly beneficial for memory-constrained mobile devices.

	HQ	iPhone 13 Mini		iPhone 15 Pro		MacBook M3		Chromebook	
		FPS	S. time	FPS	S. time	FPS	S. time	FPS	S. time
H3DGS	✓	✗	✗	✗	✗	7	42	2	129
H3DGS $\tau = 6$	✓	✗	✗	✗	✗	13	19	5	97
3DGS	✗	43	8	38	8	38	8	22	15
Full representation	✓	✗	✗	26	18	29	19	17	31
Ours - single cluster	✓	42	6	34	5	41	7	23	10
Ours	✓	41	7	35	5	43	7	22	9

Table 4: **Mobile experiment.** We estimate and compare the rendering speed (FPS) and sorting time in milliseconds (S. Time) on various mobile devices. (✗) means rendering crashed; HQ=high quality.

Rendering on mobile and low-power devices. Finally, we benchmark rendering speed on four devices: iPhone 13 Mini, iPhone 15 Pro, HP Elite Dragonfly Chromebook, and MacBook Air 13 inch, using the web-based 3DGS renderer by Mark Kellogg [9], which performs rendering with asynchronous Gaussian sorting running in parallel with the rasterization. We report both FPS (rasterization only) and sorting time (S. time) in milliseconds in Tab. 4. For H3DGS [11], we show both the full model ($\tau = 0$) and the default ($\tau = 6$) configuration. Vanilla 3DGS renders fast due to its low Gaussian count but at the cost of poor quality (see Tab. 1; PSNR @ 25.42 vs. 26.57 for ours). H3DGS instead fails on iPhones due to memory limitations and cannot render in real time on laptops. Notably, iPhone 13 Mini outperforms the 15 Pro in FPS due to its smaller display. Our method runs efficiently across all devices and achieves the best performance on laptops.

5 Conclusion

We introduced a novel level-of-detail (LOD) approach for 3D Gaussian Splatting that enables real-time rendering of large-scale scenes, even on memory-constrained devices. Our method combines a multi-level LOD representation with chunk-based rendering to avoid per-frame overhead by precomputing active Gaussian sets for spatial regions. We further proposed an automatic threshold selection strategy and a two-cluster opacity blending scheme to ensure smooth transitions between chunks. Extensive experiments on both indoor and outdoor datasets demonstrate that our method outperforms state-of-the-art baselines in both rendering quality and speed. Importantly, our approach is deployable on mobile devices, achieving real-time performance where other methods fail.

Limitations. While our method enables real-time rendering on mobile devices, it assumes that loading Gaussians—and reloading them when crossing chunk boundaries—can be performed efficiently. In practice, this would require optimized web servers and effective compression protocols to stream Gaussians to the device in real time, which we leave as future work.

References

- [1] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *CVPR*, 2022.
- [2] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Zip-nerf: Anti-aliased grid-based neural radiance fields. In *ICCV*, pages 19697–19705, 2023.
- [3] Zhiwen Fan, Kevin Wang, Kairun Wen, Zehao Zhu, Dejia Xu, Zhangyang Wang, et al. LightGaussian: Unbounded 3D Gaussian compression with 15x reduction and 200+ FPS. *NeurIPS*, 2024.
- [4] Guofeng Feng, Siyan Chen, Rong Fu, Zimu Liao, Yi Wang, Tao Liu, Zhilin Pei, Hengjie Li, Xingcheng Zhang, and Bo Dai. FlashGS: Efficient 3D Gaussian Splatting for large-scale and high-resolution rendering. *arXiv*, 2024.
- [5] Tobias Fischer, Jonas Kulhanek, Samuel Rota Buló, Lorenzo Porzi, Marc Pollefeys, and Peter Kotschieder. Dynamic 3D gaussian fields for urban areas. In *NeurIPS*, 2024.

- [6] Yuanyuan Gao, Hao Li, Jiaqi Chen, Zhengyu Zou, Zhihang Zhong, Dingwen Zhang, Xiao Sun, and Junwei Han. CityGS-X: A scalable architecture for efficient and geometrically accurate large-scale scene reconstruction. *arXiv*, 2025.
- [7] Sharath Girish, Kamal Gupta, and Abhinav Shrivastava. EAGLES: Efficient accelerated 3D Gaussians with lightweight encodings. In *ECCV*, 2024.
- [8] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2D gaussian splatting for geometrically accurate radiance fields. *arXiv*, 2024.
- [9] Mark Kellogg. 3d gaussian splatting for three.js. <https://github.com/mkkellogg/GaussianSplats3D>, 2025.
- [10] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM TOG*, 2023.
- [11] Bernhard Kerbl, Andreas Meuleman, Georgios Kopanas, Michael Wimmer, Alexandre Lanvin, and George Drettakis. A hierarchical 3d gaussian representation for real-time rendering of very large datasets. *ACM TOG*, 43(4), July 2024.
- [12] Dmytro Kotovenko, Olga Grebenkova, and Björn Ommer. EDGS: Eliminating densification for efficient convergence of 3DGS. *arXiv*, 2025.
- [13] Jonas Kulhanek, Songyou Peng, Zuzana Kukelova, Marc Pollefeys, and Torsten Sattler. WildGaussians: 3D Gaussian Splatting in the wild. *arXiv*, 2024.
- [14] Zimu Liao, Siyan Chen, Rong Fu, Yi Wang, Zhongling Su, Hao Luo, Li Ma, Linning Xu, Bo Dai, Hengjie Li, et al. Fisheye-GS: Lightweight and extensible gaussian splatting module for fisheye cameras. *arXiv*, 2024.
- [15] Jiaqi Lin, Zhihao Li, Xiao Tang, Jianzhuang Liu, Shiyong Liu, Jiayue Liu, Yangdi Lu, Xiaofei Wu, Songcen Xu, Youliang Yan, and Wenming Yang. Vastgaussian: Vast 3d gaussians for large scene reconstruction. In *CVPR*, 2024.
- [16] Weikai Lin, Yu Feng, and Yuhao Zhu. MetaSapiens: Real-time neural rendering with efficiency-aware pruning and accelerated foveated rendering. In *International Conference on Architectural Support for Programming Languages and Operating Systems*, 2025.
- [17] Wenkai Liu, Tao Guan, Bin Zhu, Luoyuan Xu, Zikai Song, Dan Li, Yuesong Wang, and Wei Yang. EfficientGS: Streamlining Gaussian Splatting for large-scale high-resolution scene representation. *IEEE MultiMedia*, 2025.
- [18] Yang Liu, Chuanchen Luo, Lue Fan, Naiyan Wang, Junran Peng, and Zhaoxiang Zhang. CityGaussian: Real-time high-quality large-scale scene rendering with Gaussians. In *ECCV*, 2024.
- [19] Yang Liu, Chuanchen Luo, Zhongkai Mao, Junran Peng, and Zhaoxiang Zhang. CitygaussianV2: Efficient and geometrically accurate reconstruction for large-scale scenes. *arXiv*, 2024.
- [20] Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, and Bo Dai. Scaffold-GS: Structured 3D gaussians for view-adaptive rendering. *arXiv*, 2023.
- [21] Saswat Subhajyoti Mallick, Rahul Goel, Bernhard Kerbl, Markus Steinberger, Francisco Vicente Carrasco, and Fernando De La Torre. Taming 3DGS: High-quality radiance fields with limited resources. In *ACM TOG*, 2024. doi: 10.1145/3680528.3687694.
- [22] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020.
- [23] N Milef, D Seyb, T Keeler, T Nguyen-Phuoc, A Božič, S Kondguli, and C Marshall. Learning fast 3D gaussian splatting rendering using continuous level of detail. In *CGF*, 2025.
- [24] Simon Niedermayr, Josef Stumpfegger, and Rüdiger Westermann. Compressed 3D Gaussian Splatting for accelerated novel view synthesis. In *CVPR*, 2024.
- [25] Michael Niemeyer, Fabian Manhardt, Marie-Julie Rakotosaona, Michael Oechsle, Daniel Duckworth, Rama Gosula, Keisuke Tateno, John Bates, Dominik Kaeser, and Federico Tombari. RadSplat: Radiance field-informed gaussian splatting for robust real-time rendering with 900+ FPS. *arXiv*, 2024.
- [26] Harry Nyquist. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 1928. doi: 10.1109/T-AIEE.1928.5055024.

- [27] Konstantinos Rematas, Andrew Liu, Pratul P Srinivasan, Jonathan T Barron, Andrea Tagliasacchi, Thomas Funkhouser, and Vittorio Ferrari. Urban radiance fields. In *CVPR*, 2022.
- [28] Kerui Ren, Lihan Jiang, Tao Lu, Mulin Yu, Linning Xu, Zhangkai Ni, and Bo Dai. Octree-GS: Towards consistent real-time rendering with LOD-structured 3D gaussians. *arXiv*, 2024.
- [29] Yunji Seo, Young Sun Choi, Hyun Seung Son, and Youngjung Uh. FLoD: Integrating flexible level of detail into 3D gaussian splatting for customizable rendering. *arXiv*, 2024.
- [30] Claude E Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 1949.
- [31] Jianxiong Shen, Yue Qian, and Xiaohang Zhan. LOD-GS: Achieving levels of detail using scalable Gaussian soup. In *CVPR*, 2025.
- [32] Matthew Tancik, Vincent Casser, Xincheng Yan, Sabeek Pradhan, Ben Mildenhall, Pratul P Srinivasan, Jonathan T Barron, and Henrik Kretzschmar. Block-nerf: Scalable large scene neural view synthesis. In *CVPR*, pages 8248–8258, 2022.
- [33] Haithem Turki, Deva Ramanan, and Mahadev Satyanarayanan. Mega-NeRF: Scalable construction of large-scale nerfs for virtual fly-throughs. In *CVPR*, 2022.
- [34] Zipeng Wang and Dan Xu. PyGS: Large-scale scene representation with pyramidal 3D Gaussian Splatting. *arXiv*, 2024.
- [35] Qi Wu, Janick Martinez Esturo, Ashkan Mirzaei, Nicolas Moenne-Loccoz, and Zan Gojcic. 3DGUT: Enabling distorted cameras and secondary rays in Gaussian Splatting. *arXiv*, 2024.
- [36] Vickie Ye, Ruilong Li, Justin Kerr, Matias Turkulainen, Brent Yi, Zhuoyang Pan, Otto Seiskari, Jianbo Ye, Jeffrey Hu, Matthew Tancik, et al. gsplat: An open-source library for Gaussian splatting. *JMLR*, 2025.
- [37] Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-splatting: Alias-free 3d gaussian splatting. In *CVPR*, 2024.
- [38] Zehao Yu, Torsten Sattler, and Andreas Geiger. Gaussian opacity fields: Efficient high-quality compact surface reconstruction in unbounded scenes. *arXiv*, 2024.
- [39] Dongbin Zhang, Chuming Wang, Weitao Wang, Peihao Li, Minghan Qin, and Haoqian Wang. Gaussian in the wild: 3D Gaussian Splatting for unconstrained image collections. In *ECCV*, 2024.
- [40] Guangyun Zhang, Chaozhong Xue, and Rongting Zhang. SuperNeRF: High-precision 3D reconstruction for large-scale scenes. *IEEE TGRS*, 2024.
- [41] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018.
- [42] Brent Zoomers, Maarten Wijnants, Ivan Molenaers, Joni Vanherck, Jeroen Put, and Nick Michiels. PRoGS: Progressive rendering of Gaussian splats. In *WACV*, 2025.
- [43] Matthias Zwicker, Hanspeter Pfister, Jeroen Van Baar, and Markus Gross. Surface splatting. In *ACM TOG*, 2001.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [Yes]

Justification: The introduction states the claims made, including contributions made in the paper and important assumptions and limitations.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: Limitations discussed in 'Limitations' section in 'Conclusions'.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: The paper does not contain theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [\[Yes\]](#)

Justification: The ‘Method’ and ‘Experiments’ section provide sufficient information to reproduce the main experimental results. Further details (such as exact values of hyperparameters) are given in Supplementary Material.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: Paper code cannot be released at the time of submission due to institutional policies.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: The full details are provided in Supplemental material.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: Due to a restricted computational budget, we weren't able to compute error bars.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: The details are given in the ‘Experiments’ section.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn’t make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: The paper fully conforms to the code of ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: We aren’t aware of any potential misuse of our method. We discuss societal impact in Supplementary Material.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper does not pose a risk of misuse.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The datasets we use were released as part of papers. We include appropriate citations and credit the authors.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The paper does not release any assets alongside the paper submission.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The core method development in this research did involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.

A Supplementary Material

A.1 Video

In the attached video (<https://lodge-gs.github.io/video.html>), we present a qualitative comparison between our method and several existing approaches. First, we compare our method with ZipNeRF [2], Octree-GS [28], H3DGS [11], and 3DGS [10] on the SmallCity and Campus scenes from the Hierarchical 3DGS dataset [11], as well as the NYC and London scenes from the ZipNeRF dataset [2]. ZipNeRF produces less sharp results, particularly for nearby cars and the ground. Octree-GS exhibits over-exposure and reduced detail in close-up cars. H3DGS achieves comparable visual quality to ours but is significantly slower to render. 3DGS, on the other hand, suffers from blurriness and a noticeable loss of detail. Next, we demonstrate how our LOD representation effectively reduces the number of visible Gaussians, thereby accelerating the rendering process. Finally, we show temporal artifacts when using ‘LOD + chunks’ (without opacity blending), and show how opacity blending removes these artifacts.

A.2 Implementational Details

In our method, we first optimize the ‘base’ model – *ie.*, a vanilla 3DGS reconstruction—and then build the LOD representation on top. Given the ‘base’ model, we build the LOD representation (as described in Section ‘LOD structure optimization steps’). We then build the LOD chunks as described in Section 3 of the main paper. In Section ‘Selecting chunk centers’ we give details on how the chunks are selected. To build the ‘base’ 3DGS reconstruction, we combine several improvements from recent literature [37, 11, 25] which we describe in Section ‘Detailed on training the full representation’.

LOD structure optimization steps. In this section we detail the exact steps and parameters used for creating the LOD gaussians. Given a sequence of depth thresholds d_l , we build the LOD sets $\mathcal{G}^{(l)}$ from the finest to the coarsest level. We start from the set $\mathcal{G}^{(0)}$, iteratively adding more levels. At step l , we construct the set $\mathcal{G}^{(l)}$ via applying a 3D smoothing filter to all Gaussians in $\mathcal{G}^{(0)}$. Note that the smoothing filter is only added to the Gaussian function, whereas the parameters remain unchanged. Therefore, the optimization process cannot decrease Gaussians’ sizes below the 3D filter size. Next, we prune all Gaussians in $\mathcal{G}^{(l)}$ using the importance score pruning and then perform 1 000 optimization steps. The optimization steps use the LOD rendering procedure (with LOD levels up to l) described in the previous section, with the modification that depth d_l is replaced by a random number drawn from a uniform distribution $\mathcal{U}(0.7d_l, 1.3d_l)$. This makes the representation more robust towards cameras lying outside of the training trajectory. We use the same loss function (DSSIM+L1) as during the standard optimization, however, we only update parameters of the Gaussians in set G_l . This process of pruning and fine-tuning is repeated in total three times, with importance score thresholds of 0.2γ , 0.6γ , and γ , where γ is a hyperparameter. Therefore, we need overall $N_{\text{levels}} \cdot 3 \cdot 1\,000$ optimization iterations.

Selecting chunk centers. We split the scene into chunks, by performing a K-means clustering on the training camera positions. We set the number of clusters N_{clusters} as follows:

$$N_{\text{clusters}} = \frac{4}{d_1} \max_i \|\mathbf{c}_i - \bar{\mathbf{c}}\|, \quad \bar{\mathbf{c}} = \frac{1}{N_c} \sum_j^{N_c} \mathbf{c}_j, \quad (5)$$

where N_c is the number of training cameras, c_i are camera positions, and d_1 is first LOD depth threshold. For outdoor scenes this roughly corresponds (empirically) to cluster sizes of 5 meters.

Details on training the full representation. In our method, the LOD representation is build from an existing 3DGS representation. To this end, we extended vanilla 3DGS with recent improvements [37, 11, 25] to achieve better quality of the base representation. We employ the original 3DGS renderer augmented with a 2D filter, as proposed in Mip-Splatting [37]. However, for densification, we adopt the modified strategy from H3DGS [11]. Specifically, we replace the original hard threshold on the 2D position gradient norm used for cloning/splitting with the condition:

$$\text{grad_position_2D} \cdot \text{max_radii_2D} \cdot (\text{opacity})^{1/5} \geq \text{threshold}, \quad (6)$$

where max_radii_2D is the largest radius the Gaussian projects into (since the last densification). Moreover, instead of averaging the gradient statistics, we take the maximum. We use a gradient

threshold of 0.015 for the Hierarchical 3DGS [11] dataset and 0.03 for the ZipNeRF [2] dataset. All models were trained for 30 000 iterations. We reset opacity every 3 000 steps until iteration 15 000 and apply densification every 300 steps from step 600 to 15 000. Importance score pruning is performed at steps 8 000, 16 000, and 24 000 with threshold of 0.02 for all scenes, except Campus, where we use lower threshold of 0.01.

A.3 Datasets

To validate our method, we use two larger-scale datasets: two outdoor scenes used in the Hierarchical 3DGS paper [11] and three indoor scenes from the Zip-NeRF paper [2].

Hierarchical 3DGS dataset [11]. The public release of the dataset contains two urban scenes (SmallCity and Campus). To collect the dataset, authors used a bicycle helmet on which they mounted 6 GoPro HERO6 Black cameras (5 for the Campus scene). The SmallCity scene was collected by riding a bicycle at around 6 – 7km/h, while Campus was captured on foot wearing the helmet. Poses were estimated using COLMAP with custom parameters and hierarchical mapper. Additional per-chunk bundle adjustment was performed, see Hierarchical 3DGS [11]. In our experiments, we use the official train/test split provided by the authors. We also use the official segmentation masks provided with the dataset to remove license plates, pedestrians, and moving cars.

Zip-NeRF dataset [2]. The dataset contains four large-scale indoor scenes: Berlin, Alameda, London, and NYC, (1000-2000 photos each) captured using fisheye cameras. We use the provided undistorted images and following Zip-NeRF [2], we use downscale factor of two resulting in downsampled resolutions between 1392×793 and 2000×1140 depending on scene. Note, that we use the provided downscaled JPEG images for training and evaluation. In our experiments, we only use Alameda, London, and NYC, because for some baselines the Berlin scene cannot be fitted into GPU memory. Same as Zip-NeRF, we take each 8th image as testing image (when sorted alphabetically).

A.4 Evaluation Protocol & Baselines

We report standard PSNR, SSIM, and LPIPS metrics. To ensure a fair evaluation and reproducibility, before computing metrics, we round the predictions to uint8 range. For the Hierarchical 3DGS dataset, we mask the prediction and the ground truth image by the provided mask - replacing the pixels inside the mask with black color. For SSIM, we use commonly used default values and average across pixels and color channels. For LPIPS [41], we use the VGG network (unlike some baselines which used AlexNet with lower values). To compute FPS, we compute rendering times for each testing image. We compute each per-image rendering time 7 times and take a median to ensure precise measurement. The resulting FPS is then one over average rendering time for all test images. In the rendering time, we only include the rendering part, not the time required to move the data from GPU memory at the end of the rendering. When reporting the number of Gaussians $\#G$, we report the average number of Gaussian loaded in GPU memory when rendering. For methods which constructs LOD splits on the fly this is the total number of Gaussians in all LOD sets (if applicable). Note, that for Scaffold-GS [20] and Octree-GS [28], some attributes (colors) are not stored explicitly, but are decoded on-the-fly lowering memory requirements. All baselines and our method were trained and evaluated on a single NVIDIA A100 SXM4 40 GB GPU.

Baselines. For the baselines, we include the following: **Zip-NeRF** [2] which is the current state-of-the-art on the Zip-NeRF dataset. Note, that this method was trained for 200K iterations - much longer than other baselines and does not achieve real-time rendering. We only include it for reference. **RadSplat** [25] uses Zip-NeRF for initialization and is, therefore, much more expensive to train than other baselines. It uses the same strategy as our method to remove less important Gaussians during training. The method used Zip-NeRF checkpoint trained for 200K and performed additional 45K training iterations to optimize 3DGS representation. **3DGS** [10] is the basis for most other methods. We use the default parameters, training for 30K iterations. **Mip-Splatting** [37] augments 3DGS with antialiasing 2D and 3D filters. In our method, we also use the same 2D filter and we use the 3D filter when building LOD levels. **Scaffold-GS** [20] reduces memory requirements by decoding attributes of Gaussians on the fly from MLP. Therefore it is important baseline for rendering speed as our method stores all attributes explicitly. **H3DGS** [11] starts from a 3DGS representation optimized at the finest level (same as ours) and builds a representation by iteratively merging Gaussians, however, at render time all Gaussians need to be kept in memory and LOD split needs to be built with every

rendering call. The method is currently SoTA for the Hierarchical 3DGS dataset. The **FLOD** [29], **CityGS** [18], and **Octree-GS** [28] are both recent LOD 3DGS-based approaches. Octree-GS was trained for 40K iterations, while FLOD used 100K iterations for the 5 LOD levels. At rendering, we use the selective 3, 4, 5 rendering split. For all baselines we modified the code to accept masks (masking loss gradient propagation in masked regions). We have performed a limited hyperparameter search to get best performance for the datasets.

A.5 Broader Impact

We are not aware of any potential misuse of our method. On the contrary, we believe it enables positive applications, particularly by making large-scale 3D reconstructions deployable on mobile and resource-constrained devices. This can enhance navigation in complex urban environments, support AR experiences, and improve accessibility for users with visual impairments. Our method may also benefit fields like robotics, autonomous driving, and cultural heritage preservation by providing efficient scene representations.