

3DOF+Quantization: 3DGS quantization for large scenes with limited Degrees of Freedom

Matthieu GENDRIN
Stéphane PATEUX
Théo LADUNE
Orange Innovation

July 12, 2024

Abstract

3DGS [Kerbl et al., 2023] is a major breakthrough in 3D scene reconstruction. With a number of views of a given object or scene, the algorithm trains a model composed of 3D gaussians, which enables the production of novel views from arbitrary points of view. This freedom of movement is referred to as 6DoF for 6 degrees of freedom: a view is produced for any position (3 degrees), orientation of camera (3 other degrees). On large scenes, though, the input views are acquired from a limited zone in space, and the reconstruction is valuable for novel views from the same zone, even if the scene itself is almost unlimited in size. We refer to this particular case as 3DoF+, meaning that the 3 degrees of freedom of camera position are limited to small offsets around the central position. Considering the problem of coordinate quantization, the impact of position error on the projection error in pixels is studied. It is shown that the projection error is proportional to the squared inverse distance of the point being projected. Consequently, a new quantization scheme based on spherical coordinates is proposed. Rate-distortion performance of the proposed method are illustrated on the well-known Garden scene.

1 Introduction

3DGS [Kerbl et al., 2023] has opened new possibilities in terms of novel view synthesis of 3D scenes. The quality and training performance are such that a major part of the 3D research community has switched to this model. With this success comes the need to compress such models, which is achieved in several ways in the literature. Papantonakis *et al.* [Papantonakis et al., 2024] proposes to optimize the number of gaussians and the color coefficients, and to use a codebook-based quantization method. Scaffold-GS ([Lu et al., 2024]) introduces a structured description of the model to obtain additional compression performance. HAC ([Chen et al., 2024]) builds upon Scaffold-GS adding

entropy minimization for further rate savings. These previous work provides compelling rate-distortion, without any hypothesis on the degrees of freedom of the camera. As a complement, this paper analyses how the 3DoF+ hypothesis can be leveraged to perform a more accurate bit allocation to the spatial coordinates, giving more precision to the gaussians near the cameras, to the expense of gaussians located further away. Note that this work is complementary to the existing methods.

2 Preliminaries

3D Gaussian Splatting (3DGS) [Kerbl et al., 2023] models a 3D scene with 3D gaussians, and renders viewpoints through a differentiable splatting and tile-based rasterization. Each Gaussian is defined by a 3D covariance matrix $\Sigma \in \mathbb{R}^{3 \times 3}$ and location (mean) $\mu \in \mathbb{R}^3$, where $\mathbf{x} \in \mathbb{R}^3$ is a random 3D point, and Σ is defined by a diagonal matrix $\mathbf{S} \in \mathbb{R}^{3 \times 3}$ representing scaling and rotation matrix $\mathbf{R}_g \in \mathbb{R}^{3 \times 3}$ to guarantee its positive semi-definite characteristics, such that $\Sigma = \mathbf{R}_g \mathbf{S} \mathbf{S}^\top \mathbf{R}_g^\top$.

$$G(\mathbf{x}) = \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu)\right) \quad (1)$$

To render an image from a random viewpoint, 3D Gaussians are first splatted to 2D, and render the pixel value $\mathbf{C} \in \mathbb{R}^3$ using α -composed blending, where $\alpha \in \mathbb{R}$ measures the contribution to this pixel of each Gaussian after 2D projection, $\mathbf{c} \in \mathbb{R}^3$ is view-dependent color modeled by Spherical Harmonic (SH) coefficients, and N is the number of sorted Gaussians contributing to the rendering. The 3DGS rendering is illustrated in figure 1.

$$\mathbf{C} = \sum_{n \in N} c_n \alpha_n \prod_{j=1}^{n-1} (1 - \alpha_j) \quad (2)$$

Note that the gaussian calculation exposed in eq. (1) is actually done in 2D after projection of the gaussians center and covariance matrix on the screen image. The projection of the center is done with the classical formulation given in eq. (3).

3 Position dependent quantization methodology

Perspective projection For the sake of simplicity, let us first consider a toy system with a single camera. Any point with coordinates (x, y, z) in the camera referential is projected to the image plan as (u, v) :

$$\begin{aligned} u &= f \frac{x}{z} \\ v &= f \frac{y}{z} \end{aligned} \quad (3)$$



Figure 1: Gaussian render example. The rendering has been modified on the left part to highlight the gaussians, while the right part is the standard 3DGS rendering.

Under a high-rate hypothesis, quantizing the point coordinates (x, y, z) with a scalar, unitary quantizer can be modeled as adding an independent noise δ to the coordinates.

1. the impact on (u, v) of the noise δ on x, y is proportional to $\frac{1}{z}$
2. the impact on (u, v) of the noise δ on z is proportional to $\frac{1}{z^2}$

Of course, we don't want to encode the model for one precise camera, but This gives the intuition of how the 3DoF+ assumption can be leveraged: the z coordinate of the local referential is to be quantized differently than the (x, y) . For example, we could use larger quantization steps for x, y, z when z is large. And even larger step for z than for x, y .¹ To generalize from this toy example, we note that z is approximately the distance between the camera and the gaussian, and that the cameras are all in the same area in 3DoF+. As such, the reasoning holds for all cameras.

Figure 3 illustrates a typical 3DoF+ scene, with the camera positions in a small spatial zone compared to the distance to the closest points.

¹Note that this encourages to use spherical coordinates here for performing positional quantization. Simple scalar quantization on spherical coordinates would then lead to well adapted vector quantization of positions.

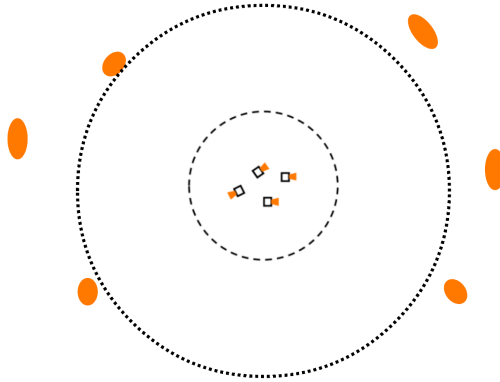


Figure 2: 3DoF+ scene.

Inner circle, radius R_i , is the limit of the possible camera poses.

Outer circle, radius R , defines the minimum gaussian distance to the center.

Spherical coordinates To make the distance between a gaussian and a camera more explicit, the gaussian coordinates are now expressed in spherical coordinates (ρ, θ, ϕ) . The origin is set in the area of the cameras centers. With this referential and the 3DoF+ assumption, the ρ coordinate of any point approximates the distance between this point and any camera.

Now on, we'll consider a point of spherical coordinates $\rho \mathbf{d}$ is the unit direction vector defined by:

$$\mathbf{d} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)^T \quad (4)$$

Considering that a camera can point to any direction, we will work on the 360° projection. This projection associates each point with its projection on the sphere of radius f . We will now work in a referential with the camera as center, keeping the orientation of the world referential. This referential will be referred to as the local referential.

$$\mathbf{P} = \mathbf{P}_0 + \rho \mathbf{d} \quad (5)$$

Where \mathbf{P}_0 is the world origin in the local referential.

The projection \mathbf{p} of \mathbf{P} on the sphere is:

$$\mathbf{p} = \frac{\mathbf{P}}{\|\mathbf{P}\|} \quad (6)$$

And the derivation of this projection gives (cf details in appendix):

$$\begin{aligned}
 \frac{\partial \mathbf{p}}{\partial \theta} &= f \frac{\rho}{\|\mathbf{P}\|} (\cos \theta \cos \phi, \cos \theta \sin \phi, -\sin \theta)^T + O(\epsilon) \\
 \frac{\partial \mathbf{p}}{\partial \phi} &= f \frac{\rho \sin \theta}{\|\mathbf{P}\|} (-\sin \phi, \cos \phi, 0)^T + O(\epsilon) \\
 \frac{\partial \mathbf{p}}{\partial \rho} &= -f \frac{1}{\rho^2} ((\mathbf{P}_0^T \mathbf{d}) \mathbf{d} - \mathbf{P}_0) + o(\epsilon^2) \\
 \epsilon &= \frac{\|\mathbf{P}_0\|}{\|\mathbf{P}\|}
 \end{aligned} \tag{7}$$

This means that uniform quantization is relevant for θ and ϕ , since their impact is bounded by finite values close to 1. The impact of ρ quantization on the other hand depends on the value of ρ itself. Thus quantizing the coordinate uniformly would be suboptimal.

It is proposed to parameterize ρ as: $\rho = \frac{1}{t}$, yielding:

$$\frac{\partial \mathbf{p}}{\partial t} = f((\mathbf{P}_0^T \mathbf{d}) \mathbf{d} - \mathbf{P}_0) + o(\epsilon^2) \tag{8}$$

Which is a good candidate for uniform quantization, since it does not depend on the position of \mathbf{P} .

Center vs periphery Since the calculations assume the gaussians are far away, compared to the distance between the cameras. In most cases, this hypothesis is not verified for all gaussians, and quantization scheme can not be used for the whole scene. A center zone is thus defined, and uses a uniform quantization instead of the proposed model. Outside of the center zone lies the peripheral zone where the proposed schema is used.

In short:

1. In the center, x, y, z are quantized uniformly
2. In the periphery, $\theta, \phi, 1/\rho$ are quantized uniformly

The center is defined as the points matching: $\rho < R$ with R being roughly twice the distance of the cameras from the center of the scene.

²

4 Experiments

Test conditions We tested the proposed quantization a reconstruction of the Garden scene, from [Barron et al., 2022], after 30k iterations on the original

²The simple fact to divide the scene in center vs periphery is already a way to quantize the coordinates of the central part with a finer step, without changing the quantization step of the rest of the scene. This optimization can be used on x, y, z coordinate, it is tested in the ablation study below.

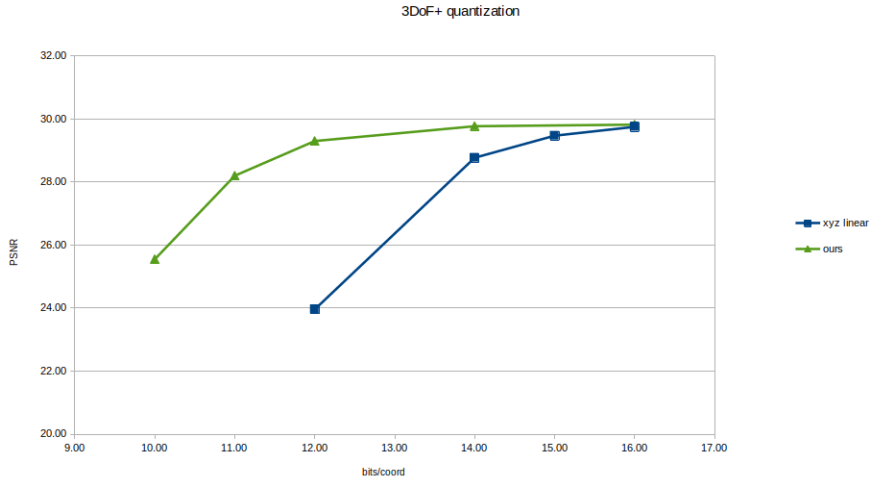


Figure 3: PSNR versus bits/coord

code of 3DGS. The PSNR is evaluated on the training views, with the configurations:

1. uniform: each x,y,z coordinate is quantized independently, the step depending on the extent of the scene
2. ours: we use $R = 1.5$ times the radius of the training cameras positions (bigger values of R did not improve the quality of the novel views)

Table 1: Results on garden scene, mip-nerf dataset

bits / coord	uniform	ours
16	29.76	29.82
14	28.77	29.77
12	23.96	29.30

The measures listed in table 1 show a clear improvement in terms of PSNR when lowering the number of bits per coordinate.

Discussions This document proposes an analysis of the impact of quantization noise in terms of projection on the screen plan. Another impact of quantization noise comes from the use of the position to define in which order the gaussians are drawn. The analysis of this aspect is left for future work. The split of the model points in center vs periphery is a new information, which should be added to the information to be coded. One may argue that the order of the points in the file is an easy way to encode this information. If the center

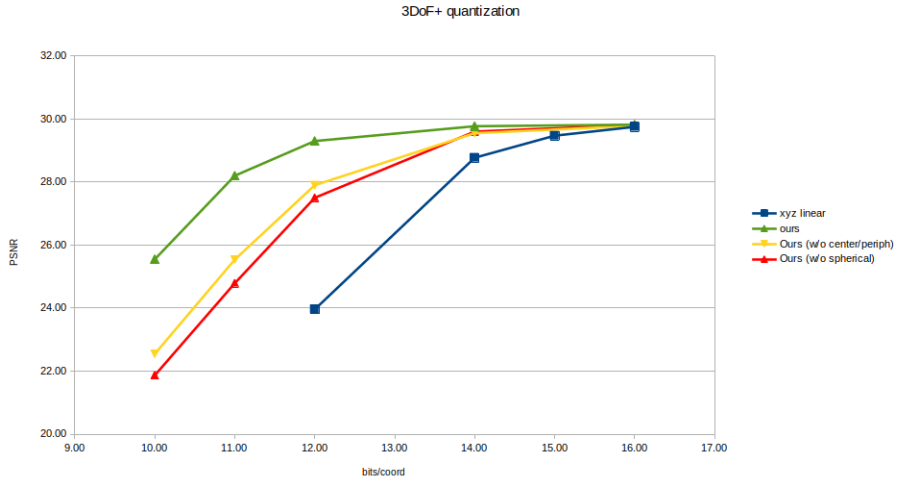


Figure 4: Ablations.

points are transferred first, only the index of the first periphery point has to be provided to differentiate the two populations. A more basic way to transfer this information would be to add one bit per gaussian, which costs 0.33 bit per coordinate. With this extra cost, the proposed solution keeps better than uniform quantization.

5 Ablation study

The proposed quantization scheme includes the use of spherical coordinates and the split of the scene in two parts: center vs periphery. Spherical coordinates, including the inversion of ρ , bring some value by giving more precision to points close to the center. The added value of this part is illustrated by "w/o center/periphery", where $1/\rho, \theta, \phi$ are quantize independently of the gaussian position. The split in center vs periphery is another way to enable finer precision on the center, without sacrificing the precision of the periphery. This part is illustrated by "w/o spherical", where x, y, z are quantized, with two sets of bounds: $[-R, R]$ for the center gaussians, and the extent of the whole scene for the periphery. The figure 4 shows the PSNR reached at different bits/coord values.

6 Conclusions

This article proposes a simple parameterization of spatial coordinates, which minimizes the projection error due to the positions quantization, compared to a standard uniform quantization. This straightforward technique does not in-

terfere with 3DGS training and is compatible with many other compression algorithms. Though exposed in a simple 3DoF+ context, this technique can be adapted to many 3D scenes to avoid storing and transmitting more information than needed for zones that will be viewed far away at render time.

Appendix A Partial derivations

This section details the partial derivatives exposed in the document.
Unit vector of the spherical coordinates:

$$\begin{aligned}
 \mathbf{d} &= (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)^T \\
 \frac{\partial \mathbf{d}}{\partial \theta} &= (\cos \theta \cos \phi, \cos \theta \sin \phi, -\sin \theta)^T \\
 \mathbf{d}^T \frac{\partial \mathbf{d}}{\partial \theta} &= 0 \\
 \frac{\partial \mathbf{d}}{\partial \phi} &= (-\sin \theta \sin \phi, \sin \theta \cos \phi, 0)^T \\
 \mathbf{d}^T \frac{\partial \mathbf{d}}{\partial \phi} &= 0
 \end{aligned} \tag{9}$$

Point in the camera referential:

$$\begin{aligned}
 \mathbf{P} &= \mathbf{P}_0 + \rho \mathbf{d} \\
 \frac{\partial \mathbf{P}}{\partial \rho} &= \mathbf{d} \\
 \frac{\partial \mathbf{P}}{\partial \mathbf{d}} &= \rho \mathbf{I}
 \end{aligned} \tag{10}$$

With \mathbf{I} the 3×3 identity matrix.

Projection of the point on the unit sphere:

$$\begin{aligned}
 \|\mathbf{P}\| &= \sqrt{p_x^2 + p_y^2 + p_z^2} \\
 \frac{\partial \|\mathbf{P}\|}{\partial \mathbf{P}} &= (p_x / \sqrt{p_x^2 + p_y^2 + p_z^2}, p_y / \sqrt{p_x^2 + p_y^2 + p_z^2}, p_z / \sqrt{p_x^2 + p_y^2 + p_z^2}) \\
 &= \frac{1}{\|\mathbf{P}\|} \mathbf{P}^T \\
 \mathbf{p} &= \frac{1}{\|\mathbf{P}\|} \mathbf{P} \\
 \frac{\partial \mathbf{p}}{\partial \mathbf{P}} &= \frac{1}{\|\mathbf{P}\|} \mathbf{I} + \mathbf{P} \left(\frac{-1}{\|\mathbf{P}\|^2} \frac{\partial \|\mathbf{P}\|}{\partial \mathbf{P}} \right) \\
 &= \frac{1}{\|\mathbf{P}\|} \mathbf{I} + \mathbf{P} \left(\frac{-1}{\|\mathbf{P}\|^2} \frac{1}{\|\mathbf{P}\|} \mathbf{P}^T \right) \\
 &= \frac{1}{\|\mathbf{P}\|} \mathbf{I} - \frac{1}{\|\mathbf{P}\|^3} \mathbf{P} \mathbf{P}^T
 \end{aligned} \tag{11}$$

Derivation of the projection by angles:

$$\begin{aligned}
\frac{\partial \mathbf{p}}{\partial \theta} &= \left(\frac{1}{\|\mathbf{P}\|} \mathbf{I} - \frac{1}{\|\mathbf{P}\|^3} \mathbf{P} \mathbf{P}^T \right) \frac{\partial \mathbf{P}}{\partial \mathbf{d}} \frac{\partial \mathbf{d}}{\partial \theta} \\
&= \left(\frac{1}{\|\mathbf{P}\|} \mathbf{I} - \frac{1}{\|\mathbf{P}\|^3} \mathbf{P} \mathbf{P}^T \right) \rho \frac{\partial \mathbf{d}}{\partial \theta} \\
&= \frac{1}{\|\mathbf{P}\|} \rho \frac{\partial \mathbf{d}}{\partial \theta} - \frac{1}{\|\mathbf{P}\|^3} \mathbf{P} \mathbf{P}^T \left(\rho \frac{\partial \mathbf{d}}{\partial \theta} \right) \\
&= \frac{\rho}{\|\mathbf{P}\|} \left(\frac{\partial \mathbf{d}}{\partial \theta} - \frac{1}{\|\mathbf{P}\|} (\mathbf{P}_0^T \frac{\partial \mathbf{d}}{\partial \theta}) \mathbf{p} \right) \\
&= \frac{\rho}{\|\mathbf{P}\|} \frac{\partial \mathbf{d}}{\partial \theta} + O(\epsilon) \\
&= \frac{\rho}{\|\mathbf{P}\|} (\cos \theta \cos \phi, \cos \theta \sin \phi, -\sin \theta) + O(\epsilon) \\
\frac{\partial \mathbf{p}}{\partial \phi} &= \left(\frac{1}{\|\mathbf{P}\|} \mathbf{I} - \frac{1}{\|\mathbf{P}\|^3} \mathbf{P} \mathbf{P}^T \right) \frac{\partial \mathbf{P}}{\partial \mathbf{d}} \frac{\partial \mathbf{d}}{\partial \phi} \\
&= \frac{1}{\|\mathbf{P}\|} \rho \frac{\partial \mathbf{d}}{\partial \phi} - \frac{1}{\|\mathbf{P}\|^3} \mathbf{P} \mathbf{P}^T \left(\rho \frac{\partial \mathbf{d}}{\partial \phi} \right) \\
&= \frac{\rho}{\|\mathbf{P}\|} \left(\frac{\partial \mathbf{d}}{\partial \phi} - \frac{1}{\|\mathbf{P}\|} (\mathbf{P}_0^T \frac{\partial \mathbf{d}}{\partial \phi}) \mathbf{p} \right) \\
&= \frac{\rho}{\|\mathbf{P}\|} \frac{\partial \mathbf{d}}{\partial \phi} + O(\epsilon) \\
&= \frac{\rho \sin \theta}{\|\mathbf{P}\|} (-\sin \phi, \cos \phi, 0) + O(\epsilon)
\end{aligned} \tag{12}$$

Derivation of the projection by ρ :

$$\begin{aligned}
\frac{\partial \mathbf{p}}{\partial \rho} &= \left(\frac{1}{\|\mathbf{P}\|} \mathbf{I} - \frac{1}{\|\mathbf{P}\|^3} \mathbf{P} \mathbf{P}^T \right) \frac{\partial \mathbf{P}}{\partial \rho} \\
&= \frac{1}{\|\mathbf{P}\|} \mathbf{d} - \frac{1}{\|\mathbf{P}\|^3} \mathbf{P} \mathbf{P}^T \mathbf{d} \\
&= \frac{1}{\|\mathbf{P}\|} \mathbf{d} - \frac{1}{\|\mathbf{P}\|^3} (\mathbf{P}_0 + \rho \mathbf{d})(\rho + \mathbf{P}_0^T \mathbf{d}) \\
&= \frac{1}{\|\mathbf{P}\|^3} (\|\mathbf{P}\|^2 \mathbf{d} - \rho \mathbf{P}_0 - (\mathbf{P}_0^T \mathbf{d}) \mathbf{P}_0 - \rho^2 \mathbf{d} - \rho (\mathbf{P}_0^T \mathbf{d}) \mathbf{d}) \\
&= \frac{1}{\|\mathbf{P}\|^3} ((\rho^2 + 2\rho \mathbf{P}_0^T \mathbf{d} + \|\mathbf{P}_0\|^2) \mathbf{d} - \rho \mathbf{P}_0 - (\mathbf{P}_0^T \mathbf{d}) \mathbf{P}_0 - \rho^2 \mathbf{d} - \rho (\mathbf{P}_0^T \mathbf{d}) \mathbf{d}) \\
&= \frac{1}{\|\mathbf{P}\|^3} ((\rho \mathbf{P}_0^T \mathbf{d} + \|\mathbf{P}_0\|^2) \mathbf{d} - (\rho + \mathbf{P}_0^T \mathbf{d}) \mathbf{P}_0) \\
&= \frac{\rho}{\|\mathbf{P}\| \|\mathbf{P}\|^2} ((\mathbf{P}_0^T \mathbf{d}) \mathbf{d} - \mathbf{P}_0) + \frac{1}{\|\mathbf{P}\|} \left(\frac{\|\mathbf{P}_0\|^2}{\|\mathbf{P}\|^2} \mathbf{d} - \frac{\mathbf{P}_0^T \mathbf{d}}{\|\mathbf{P}\|} \frac{\mathbf{P}_0}{\|\mathbf{P}\|} \right) \\
&= \frac{\rho}{\|\mathbf{P}\| \|\mathbf{P}\|^2} ((\mathbf{P}_0^T \mathbf{d}) \mathbf{d} - \mathbf{P}_0) + O(\epsilon^2) \\
&= \frac{1}{\|\mathbf{P}\|^2} ((\mathbf{P}_0^T \mathbf{d}) \mathbf{d} - \mathbf{P}_0) + O(\epsilon^2) \tag{13}
\end{aligned}$$

Please note that in these equations, \mathbf{p} is the projection on the unit sphere, as in the paper it's the projection on the sphere of radius f .

References

- [Barron et al., 2022] Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., and Hedman, P. (2022). Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5470–5479.
- [Chen et al., 2024] Chen, Y., Wu, Q., Cai, J., Harandi, M., and Lin, W. (2024). Hac: Hash-grid assisted context for 3d gaussian splatting compression. *arXiv preprint arXiv:2403.14530*.
- [Kerbl et al., 2023] Kerbl, B., Kopanas, G., Leimkühler, T., and Drettakis, G. (2023). 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1.
- [Lu et al., 2024] Lu, T., Yu, M., Xu, L., Xiangli, Y., Wang, L., Lin, D., and Dai, B. (2024). Scaffold-gs: Structured 3d gaussians for view-adaptive rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20654–20664.
- [Papantonakis et al., 2024] Papantonakis, P., Kopanas, G., Kerbl, B., Lanvin, A., and Drettakis, G. (2024). Reducing the memory footprint of 3d gaussian

splatting. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 7(1):1–17.