

# VRSplat: Fast and Robust Gaussian Splatting for Virtual Reality

XUECHANG TU, Peking University, China

LUKAS RADL, Graz University of Technology, Austria

MICHAEL STEINER, Graz University of Technology, Austria

MARKUS STEINBERGER, Graz University of Technology, Austria and Huawei Technologies, Austria

BERNHARD KERBL, Carnegie Mellon University, USA

FERNANDO DE LA TORRE, Carnegie Mellon University, USA

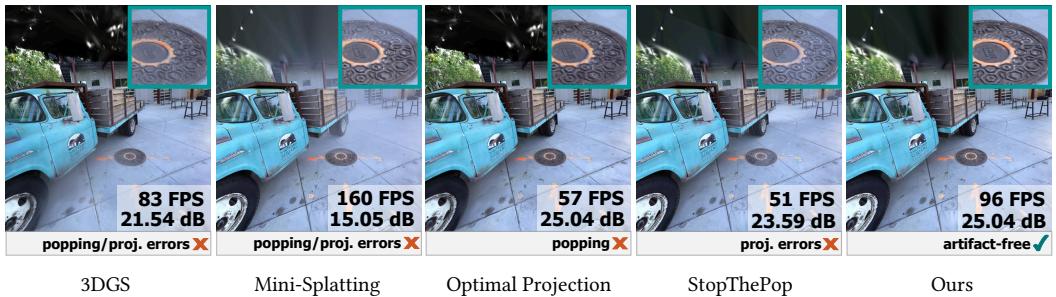


Fig. 1. For a large field-of-view as commonly used in VR, 3DGS’s projection error produces elongated or large Gaussians that can cause cloudy appearance, leading to significantly worse PSNR values. Optimal Projection faithfully projects these Gaussians, but incurs a performance cost and still suffers from immersion-breaking popping artifacts. Mini-Splatting can drastically reduce model size and rendering times, but resolving remaining popping artifacts (StopThePop) introduces significant overhead. Our full method elegantly combines these techniques with an efficient foveated rendering routine and provides an immersive and artifact-free VR experience with framerates above 72FPS rendered on an NVIDIA RTX 4090 and displayed on a Meta Quest 3.

3D Gaussian Splatting (3DGS) has rapidly become a leading technique for novel-view synthesis, providing exceptional performance through efficient software-based GPU rasterization. Its versatility enables real-time applications, including on mobile and lower-powered devices. However, 3DGS faces key challenges in virtual reality (VR): (1) temporal artifacts, such as popping during head movements, (2) projection-based distortions that result in disturbing and view-inconsistent floaters, and (3) reduced framerates when rendering large numbers of Gaussians, falling below the critical threshold for VR. Compared to desktop environments, these issues are drastically amplified by large field-of-view, constant head movements, and high resolution of head-mounted displays (HMDs). In this work, we introduce VRSplat: we combine and extend several recent advancements in 3DGS to address challenges of VR holistically. We show how the ideas of Mini-Splatting, StopThePop, and Optimal Projection can complement each other, by modifying the individual techniques and core 3DGS rasterizer. Additionally, we propose an efficient foveated rasterizer that handles focus and

---

Authors’ addresses: [Xuechang Tu](#), Peking University, China, cmxrynp@gmail.com; [Lukas Radl](#), lukas.radl@tugraz.at, Graz University of Technology, Austria; [Michael Steiner](#), michael.steiner@tugraz.at, Graz University of Technology, Austria; [Markus Steinberger](#), Graz University of Technology, Austria and Huawei Technologies, Austria, steinberger@tugraz.at; [Bernhard Kerbl](#), kerbl@cg.tuwien.ac.at, Carnegie Mellon University, USA; [Fernando de la Torre](#), ftorre@andrew.cmu.edu, Carnegie Mellon University, USA.

---

© 2025 Copyright held by the owner/author(s).

This is the author’s version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, <https://doi.org/10.1145/3728311>.

peripheral areas in a single GPU launch, avoiding redundant computations and improving GPU utilization. Our method also incorporates a fine-tuning step that optimizes Gaussian parameters based on StopThePop depth evaluations and Optimal Projection. We validate our method through a controlled user study with 25 participants, showing a strong preference for VRSplat over other configurations of Mini-Splatting. VRSplat is the first, systematically evaluated 3DGS approach capable of supporting modern VR applications, achieving 72+ FPS while eliminating popping and stereo-disrupting floaters.

CCS Concepts: • Computing methodologies → Virtual reality; Rasterization; Machine learning.

#### ACM Reference Format:

Xuechang Tu, Lukas Radl, Michael Steiner, Markus Steinberger, Bernhard Kerbl, and Fernando de la Torre. 2025. VRSplat: Fast and Robust Gaussian Splatting for Virtual Reality. *Proc. ACM Comput. Graph. Interact. Tech.* 8, 1, Article 1 (May 2025), 22 pages. <https://doi.org/10.1145/3728311>

## 1 INTRODUCTION

The field of novel-view synthesis has experienced remarkable advancements in recent years, driven by the introduction of Neural Radiance Fields (NeRFs) [Mildenhall et al. 2020] and the wide array of subsequent developments. More recently, 3D Gaussian Splatting (3DGS) [Kerbl et al. 2023] has introduced a paradigm shift by representing scenes with explicit 3D Gaussian point clouds, rather than implicit volumetric fields. 3DGS offers not only high visual quality and low training times, but also enables efficient rendering through GPU-based software rasterization or the traditional graphics pipeline.

Capturing and synthesizing novel views of real-world scenes is crucial in a variety of applications, such as e-commerce, visual effects in film production, and immersive experiences. Virtual reality (VR) has emerged as a particular focus, with its stringent demands for high framerates and interactivity. The impressive performance of 3DGS rasterization makes it well-suited for rendering on low-powered head-mounted displays (HMDs), but several limitations in the original 3DGS implementation can disrupt the immersive experience.

Firstly, the reliance on global primitive sorting during rasterization introduces popping artifacts when the viewpoint changes [Radl et al. 2024b], which can be particularly jarring in VR due to the constant micro-movements of the head in HMDs. Secondly, the projection of 3D Gaussians into 2D splats is based on a local affine approximation [Zwicker et al. 2001], leading to projection errors that worsen with increasing distance from the image plane center [Huang et al. 2024]. These errors are further amplified in VR due to the wide field-of-view (FOV), causing Gaussians to become distorted or cloud-like, which detracts from the visual experience. Lastly, the heuristic-driven densification in 3DGS reconstructions often results in excessively large point clouds, making it challenging to maintain interactive framerates since the runtime is directly tied to the number of primitives.

To address these limitations, we introduce VRSplat, a robust and efficient solution for rendering Gaussian splats in VR. Our method combines recent advancements in 3DGS in a meaningful way and introduces novel, targeted high-performance optimizations to ensure view-consistent, artifact-free rendering while maintaining the required framerates for VR. While VRSplat can be used with *any* scene reconstruction method that generates a compact set of 3D Gaussians, we choose *Mini-Splatting* [Fang and Wang 2024] due to its reduced primitive count. We mitigate popping artifacts by employing hierarchical rasterization as seen in StopThePop [Radl et al. 2024b], and apply Optimal Projection [Huang et al. 2024] to eliminate projection artifacts. However, both methods come with a significant performance cost. To counter this, we propose a single-pass foveated rendering technique, which boosts performance and ensures that the recommended frame rate of  $\geq 72$  FPS at native HMD resolutions is consistently achieved. In addition to a novel system design for real-time artifact-free rendering, VRSplat makes the following technical contributions:

- A fine-tuning approach that takes a 3DGS scene trained with any (compact) optimization strategy as input—for example Mini-Splatting which reduces the number of Gaussians by up to 10×—and optimizes it for both StopThePop rendering and Optimal Projection. This approach not only removes large floating artifacts and popping but also significantly improves overall reconstruction quality.
- The integration of StopThePop rendering and Optimal Projection within the core 3DGS CUDA renderer. This combination enables correct and efficient tile-based culling by accounting for tight bounds from Optimal Projection and providing efficient hierarchical sorting and culling via StopThePop.
- An optimized Gaussian Splatting rasterizer that supports rendering tiles at varying resolutions, enabling effective foveated rendering in a single pass. This approach eliminates redundant computations, minimizes kernel launch overhead, reduces memory access during preprocessing and sorting, and improves GPU utilization by handling the entire workload in one go.

Finally, this paper contributes a formal validation of the enhanced VR experience provided by VRSplat, via an extensive user study; its evaluation shows overwhelming preference for VRSplat compared to baseline, highlighting our method’s ability to deliver an immersive VR experiences and meet the performance requirements of smooth VR applications. We additionally demonstrate significantly improved image quality metrics when adapting the evaluation setup to the requirements of VR rendering (*i.e.* high-resolution, large FOV). Fig. 1 illustrates the limitations of the individual techniques we employ, and how, when combined with our single-pass foveated rendering solution, they form a view-consistent VR renderer capable of rendering large 3D scenes.

## 2 RELATED WORK

This section revisits recent work on radiance fields, their applications in VR, and approaches toward interactive rendering on HMDs.

*Radiance Fields.* Novel-view synthesis is a long-standing problem in visual computing. Neural Radiance Fields (NeRFs) [Mildenhall et al. 2020] have demonstrated impressive results by combining classical volumetric rendering along view rays combined with large MLPs. While NeRFs exhibit astonishing visual quality, they suffer from long training and rendering times due to expensive raymarching for each pixel. Recent advances have focused on accelerating rendering and training [Fridovich-Keil et al. 2022; Müller et al. 2022; Steiner et al. 2024], extended NeRFs to unbounded scenes [Barron et al. 2022, 2023] or applied them for various downstream applications, *e.g.* 3D Stylization [Nguyen-Phuoc et al. 2022; Radl et al. 2024a; Zhang et al. 2023].

3D Gaussian Splatting (3DGS) [Kerbl et al. 2023] thus introduces 3D Gaussians as an alternative scene representation, leveraging efficient software rasterization. Similarly to NeRF, this resulted in another wave of follow-up work, exploring artifact-free rendering [Huang et al. 2024; Radl et al. 2024b; Yu et al. 2024], reducing the number of primitives [Fan et al. 2024; Fang and Wang 2024; Kheradmand et al. 2024; Mallick et al. 2024] and rendering of very large datasets [Kerbl et al. 2024; Lin et al. 2024b], amongst many other works.

*Radiance Fields for Virtual Reality.* VR is a natural use case for radiance fields, as it provides more immersive experiences when exploring scenes. FoV-NeRF [Deng et al. 2022] utilizes a gaze-contingent neural representation and foveated rendering to achieve near real-time rendering of NeRF in VR. VR-NeRF [Xu et al. 2023] instead distributes computation across multiple GPUs and leverages an occupancy grid to improve performance. However, their trade-off between quality and speed is far from satisfactory.



Fig. 2. Normal FOV (top) and VR (bottom) views of Mini-Splatting [Fang and Wang 2024] and ours. Increasing the field-of-view reveals shortcomings of current methods in VR applications: large Gaussians when projected with the local affine approximation can cause cloud-like artifacts and elongated Gaussians (*i.e.*, “god rays”). VRSplat eliminates the artifacts while meeting the performance demands of smooth VR experiences.

VR-GS [Jiang et al. 2024] builds an interactive VR setting on top of the original 3DGS renderer that offers intuitive interaction with objects using simulation. Our approach could be used as a drop-in replacement to offer higher-quality rendering. Concurrent work to ours RTGS [Lin et al. 2024a] also explores rendering of Gaussians on mobile devices and proposes a foveated rendering solution to achieve immersive framerates. Our own method, VRSplat, substantiates the ideas presented in a recent poster abstract [Tu et al. 2024]. The contributions of RTGS are complementary to ours: they propose a training strategy that leads to scene reconstructions that should increase rendering speeds by reducing the number of overlapping Gaussians. Their second contribution is a trained, level-of-detail-based foveated rendering strategy that selects a lower number of Gaussians for lower-resolution rendering. In contrast to RTGS, our fine-tuning procedure is significantly more lightweight due to a single model being used, rendered at different resolution. Our approach could be combined with RTGS by fine-tuning their output to our model to reduce artifacts and render foveated images in a single pass to further increase VR framerates.

*Accelerating Virtual Reality Rendering.* The characteristics of the human visual system have long been exploited to accelerate rendering. For instance, due to reduced visual acuity in the periphery, a plethora of works [Chakravarthula et al. 2021; Guenter et al. 2012; Kaplanyan et al. 2019; Konrad et al. 2020; Krajancich et al. 2023; Patney et al. 2016; Rolff et al. 2023a,b; Sun et al. 2017] reduce rendering quality in these regions, thereby increasing rendering speed. Recent work has also leveraged procedural noise generation in foveated rendering, effectively simulating the presence of high-frequency content outside the fovea in a post-processing step [Tariq et al. 2022]. Our single-pass foveated rendering solution, although related to the aforementioned works, specifically operates on 3DGS’s tile-based rasterization to extract the maximum performance.

### 3 3D GAUSSIAN SPLATTING

3DGS [Kerbl et al. 2023] represents a scene using a mixture of  $N$  3D Gaussians, each given by:

$$G(\mathbf{x}) = e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})}, \quad (1)$$

with the Gaussian 3D mean  $\boldsymbol{\mu} \in \mathbb{R}^3$ , and the 3D covariance matrix  $\Sigma = RSS^T R^T$  being constructed from a rotation matrix  $R \in \mathbb{R}^{3 \times 3}$  and a diagonal scaling matrix  $S \in \mathbb{R}^{3 \times 3}$ . For a pixel  $\mathbf{p}$ , the resulting color is given by the volume rendering equation:

$$C(\mathbf{p}) = \sum_{i=0}^N \mathbf{c}_i \alpha_i \prod_{j=0}^{i-1} (1 - \alpha_j), \quad (2)$$

where  $\alpha_i = \sigma_i G_{2D}(\mathbf{p})$  is the alpha value of the projected 2D splat, dependent on the value of the 2D Gaussian  $G_{2D}$  and the opacity value  $\sigma_i$ , and  $\mathbf{c}_i$  is the Gaussian's view-dependent color.

Rather than considering a fully-sorted list of all primitives for each pixel, 3DGS performs tile-based rasterization; based on the extent of each Gaussian on the image plane, it is assigned to tiles each covering  $16 \times 16$  pixels. After instantiation of all Gaussian/tile combinations, this arrangement is sorted and only potentially contributing Gaussians are considered during rendering, simplifying and accelerating the pipeline.

The projection of each 3D Gaussian onto the image plane is demonstrated by Zwicker et al. [2001]: the covariance matrix of the 2D Gaussian  $\Sigma_{2D}$  is obtained by dropping the last row/column of

$$\Sigma' = JW\Sigma W^T J^T, \quad (3)$$

with  $W$  the camera matrix and  $J$  the local affine approximation of the projective transform. As noted in [Zwicker et al. 2001], this is only correct for the center of the image plane—for Gaussians at extreme positions (*i.e.* on the boundary of the image plane), the approximation error grows larger and causes elongated Gaussians or cloud-like artifacts (see Fig. 2). Optimal Projection [Huang et al. 2024] describes an alternative approach to remedy this issue: Namely, they project Gaussians onto distinct tangent planes on the unit sphere around the camera  $\mathbf{o}$ , in turn removing projection-based artifacts with a small performance overhead.

The sort order in Eqn. (2) is given by the  $z$ -coordinate of  $\boldsymbol{\mu}$  in view-space. Clearly, this results in changes of ordering during view rotation, causing popping artifacts [Radl et al. 2024b] — an artifact also common in other splatting approaches [Mueller and Crawfis 1998] (cf. the supplemental video). While current VR solutions [Cier et al. 2024] thus prefer sorting by the distance to the camera, *i.e.*  $\|\mathbf{o} - \boldsymbol{\mu}\|_2$ , this now results in popping during translation. The only viable solution to eliminate all popping artifacts is perform a full sort of primitives considered for each pixel; despite tile-based rasterization, a full sort is computationally prohibitive if interactive framerates are required.

As the performance of each stage in 3DGS's rendering pipeline is directly tied to primitive counts, reducing the number of Gaussians can drastically improve performance. While 3DGS's model sizes are tolerable for desktop viewing, the high resolution and demanding framerate requirements of HMDs necessitate smaller Gaussian point clouds. The underlying root cause of the large primitive counts is 3DGS's heuristics-driven densification [Kerbl et al. 2023], which leads to a large number of redundant primitives in over-reconstructed regions. A plethora of recent works investigate model size reduction whilst retaining high image quality metrics, either through importance-based pruning [Fan et al. 2024; Niemeyer et al. 2024], score-based densification [Mallick et al. 2024], oversampling and subsequent simplification [Fang and Wang 2024] or sampling-based densification [Kheradmand et al. 2024].

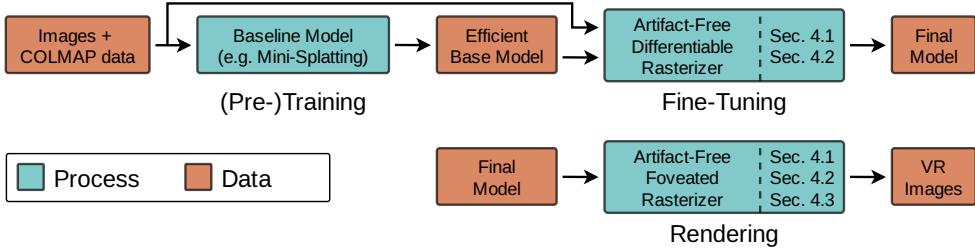


Fig. 3. An overview of our pipeline for high-fidelity VR rendering: Starting from an efficient baseline, such as Mini-Splatting [Fang and Wang 2024], we fine-tune the Gaussians using our artifact-free differentiable rasterizer to arrive at a 3DGS model capable of VR rendering without artifacts. We thus remain largely baseline-agnostic, and the overhead for our fine-tuning is only a small fraction of the initial baseline reconstruction.

## 4 METHOD

This section introduces our full model for immersive VR experiences on consumer-grade HMDs. We synergize recent advances in 3DGS to arrive at a full model capable of view-consistent rendering of Gaussians. To address the implied performance penalty, we introduce a novel single-pass foveated rendering solution capable of achieving the required rendering times for immersive VR. We provide an overview of our complete pipeline in Fig. 3.

We choose Mini-Splatting [Fang and Wang 2024] as our model backbone, as their more uniform distribution of Gaussians also improves rendering times by balancing per-tile workloads. Subsequently, we propose a fine-tuning scheme to incorporate StopThePop, Optimal Projection and Mini-Splatting into a single, high-quality representation—which we then accelerate by exploiting foveated rendering.

### 4.1 Eliminating Popping Artifacts

Due to Mini-Splatting representing scenes with a significantly reduced primitive count, Gaussians are naturally larger, and therefore more prone to popping. Radl and Steiner *et al.* [Radl et al. 2024b] demonstrate that hierarchical per-pixel resorting of Gaussian splats according to depth along view rays is able to overcome popping effectively. Hence, we opt to use their renderer to achieve view-consistent rendering.

3DGS employs its differentiable renderer to optimize Gaussian parameters—including position, rotation, scale, opacity, and spherical harmonics coefficients—via backpropagation. Sort order plays an important role during optimization, *i.e.* models need to be rendered with the same sort order as was used during training. Consequently, naïvely applying StopThePop to a Mini-Splatting [Fang and Wang 2024] model—which uses a global sort order during training—reduces image quality. Therefore, we opt for fine-tuning our small Mini-Splatting models with StopThePop rasterization (without densification, thus preserving the low primitive count), which proves to be sufficient to achieve equal or higher scores on standard image quality metrics, depending on the dataset. This strategy has the additional advantage that it makes our method largely agnostic to the pre-trained model and densification objective, as long as they yield coherent 3D Gaussian representations.

### 4.2 Minimizing Projection Error

The projection-based distortions due to 3DGS’s local affine approximation are especially disturbing in VR, where the large FOV and constant movement of the user’s head causes cloud-like artifacts to rotate and block the views. Optimal Projection for 3DGS [Huang et al. 2024] conducted a detailed

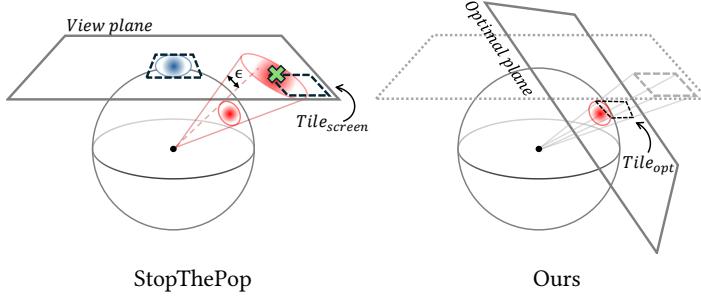


Fig. 4. Left: StopThePop computes the maximum contribution within a tile directly in screen space. Right: As we employ Optimal Projection [Huang et al. 2024], we find the maximum by reasoning about the maximum on the per-Gaussian optimal plane, leveraging the fact that tiles form general quadrilaterals. Subsequently, projecting this maximum back to screen space yields the maximum contribution within a tile [Tu et al. 2024].

analysis of the error introduced in the projection step and proposed a solution: Namely, each Gaussian is projected to the tangent plane of the unit sphere at  $\mathbf{o}$  that is perpendicular to the line connecting  $\mathbf{o}$  and  $\boldsymbol{\mu}$ , where  $\mathbf{o}$  denotes the camera position.

However, Optimal Projection is not directly compatible with tile-based culling, as proposed in StopThePop [Radl et al. 2024b]. Tile-based culling reduces the number of Gaussian/tile combinations by: (1) computing the maximum over  $G_{2D}$  for each Gaussian/tile combination and (2) pruning the Gaussian from the specific tile if  $\alpha < \frac{1}{255}$ . Radl and Steiner *et al.* also show that the maximum over  $G_{2D}$  can be computed by maximizing the Gaussian’s contribution over the tile boundaries closest to the 2D mean [Radl et al. 2024b]. Nevertheless, note that this computation is done fully in screen-space with axis-aligned tiles, thus requiring an adapted algorithm to work with Optimal Projection.

When the 2D Gaussian’s mean  $\boldsymbol{\mu}_{2D}$  is inside the tile, it follows that the point of maximum contribution  $\hat{\mathbf{x}} = \boldsymbol{\mu}_{2D}$ , otherwise  $\hat{\mathbf{x}}$  must lie on one of the edges that are reachable from  $\boldsymbol{\mu}_{2D}$ . For Optimal Projection, tile-based culling can no longer be performed on the image plane, but needs to be evaluated on the optimal plane of each Gaussian. Projecting these screen-space tiles yields again convex quadrilaterals  $Tile_{opt}$ . Consequently, at most two neighboring edges may be reachable from  $\boldsymbol{\mu}_{2D}$  because projection preserves their geometrical relationship in the original view plane. For each candidate edge  $\mathbf{p} + t \cdot \mathbf{d}$ , by taking the derivative of density with respect to  $t$  and letting it equal zero, the maximum density point along the line can be calculated as

$$\hat{\mathbf{x}} = \mathbf{p} + \frac{\mathbf{d}^\top \Sigma_{2D}^{-1} (\boldsymbol{\mu}_{2D} - \mathbf{p})}{\mathbf{d}^\top \Sigma_{2D}^{-1} \mathbf{d}} \cdot \mathbf{d}, \quad (4)$$

where  $\Sigma_{2D}$  is the 2D covariance matrix of the splat. Evaluating the  $\alpha$  value at  $\hat{\mathbf{x}}$  allows us to determine the culling criteria. Projecting  $\hat{\mathbf{x}}$  back onto the image plane yields the maximum point inside the screen-space tile, which StopThePop requires for its per-tile depth evaluation [Radl et al. 2024b]. See Fig. 4 for a visualization of StopThePop’s culling strategy compared to our approach.

### 4.3 Single-Pass Foveated Rendering

To remedy the implied performance penalty of StopThePop rasterization and Optimal Projection, we leverage the reduced visual acuity in the periphery by employing foveated rendering. While multi-pass solutions (rendering the center and peripheral regions separately) require only minor changes to the pipeline and can be effective when combined with proper culling [Tu et al. 2024],

they necessitate processing the same Gaussians multiple times. Hence, we propose a targeted single-pass solution that ties in elegantly with the tile-based 3DGS rasterizer.

*Single-Pass Foveated Rendering.* In the following, we differentiate only between a high and a low-resolution region by launching differently-sized tiles. However, all our strategies can directly be applied to further resolution reductions as well. As visualized in Fig. 5a, we realize the split between high and low resolution regions by using  $16 \times 16$  pixel tiles for the center, and  $32 \times 32$  tiles for the periphery. In practice, we always first partition the image into  $32 \times 32$  pixel tiles, which are split into four separate subtiles of size  $16 \times 16$  in the center region. This mapping can be efficiently computed at the start of each render pass and enables us to compute the exact number of blocks to launch during the rendering stage. Consumer-grade HMDs, like the Meta Quest 3, often do not provide eye tracking, thus we can precompute this mapping once for each eye and re-use it across frames.

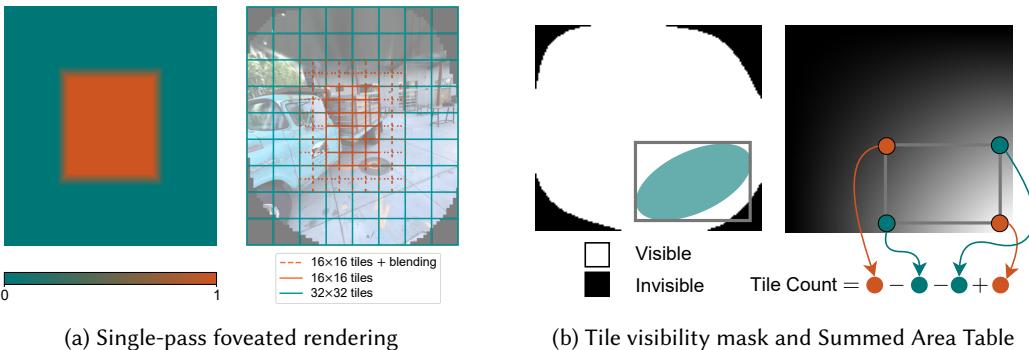


Fig. 5. We design a single-pass foveated rendering routine with visibility culling. (a) Depiction of our chosen tile configurations for our single-pass foveated rendering: we split the center region into tiles of size  $16 \times 16$ , and use  $32 \times 32$  tiles in the periphery; hybrid tiles in-between (dotted) are treated as  $16 \times 16$  tiles but the values of groups of  $2 \times 2$  pixels are averaged and blended with the individual pixel values, based on the continuous blending mask (left). (b) We sample the visibility mask from OpenXR to first compute a mask of all visible tiles, from which we construct its summed-area table. We then compute the exact number of visible tiles covered by the rectangular extent of a 2D Gaussian splat from the summed-area values of its corner points.

During tile-based rasterization, we assign Gaussians to  $32 \times 32$  tiles. This results in fewer Gaussian/tile combinations overall, but additional workload during the rendering stage as subtiles in the center region need to load more Gaussians. However, we utilize StopThePop’s hierarchical culling [Radl et al. 2024b] to mitigate this issue, allowing us to efficiently cull unnecessary Gaussians early for the  $16 \times 16$  subtiles. During the render stage, we launch blocks with 256 threads (same as StopThePop and 3DGS) for both  $16 \times 16$  and  $32 \times 32$  tiles. However, each thread in the larger tiles handles groups of  $2 \times 2$  pixels and treats them as a single pixel, effectively halving the resolution.

Finally, we blend the high-resolution region with the low-resolution region to avoid noticeable discontinuities. For the transitional subtiles in the high resolution region (visualized as dotted tiles in Fig. 5a), we approximate the color of the low resolution tile by averaging the values of  $2 \times 2$  pixels and blend it with the individual pixel values, based on the continuous blending mask. For the low-resolution region, we perform nearest neighbor upsampling and blurring with a  $3 \times 3$  Gaussian kernel, which proves effective in suppressing aliasing artifacts.

*Visibility Culling.* We propose to further accelerate rendering by culling tiles not visible in the HMD. We process the visibility mask available for the HMD (e.g. through OpenXR) to create a bitfield with one bit for each tile, indicating whether or not any pixels of the tile are visible. Subsequently, we compute a summed-area table from this bitfield, and instantiate Gaussian/tile combinations in two stages: First, we calculate the exact number of visible tiles each Gaussian may touch using the summed-area table (see Fig. 5b); We then use this information to allocate the global sorting buffer and to compute the range of each Gaussian’s instances inside this buffer.

This leads to a reduction in sort entries, as Gaussian/tile combinations outside the visible area of the image are not instantiated. In our single-pass foveated rendering approach, we can reduce the amount of tiles further by removing the “invisible” tiles entirely from the precomputed mapping, resulting in a reduction of  $\sim 11\%$  in the number of tiles and  $\sim 6\%$  in the total number of Gaussian/tile combinations.

## 5 RESULTS AND EVALUATION

We use two variants of Mini-Splatting [Fang and Wang 2024] as the base for our model: the original Mini-Splatting ( $z$ ), where the sort order is given by the view-space  $z$ -coordinate as well as Mini-Splatting (Dist), which we modified to order Gaussians according to  $\|\mu - o\|_2$ . Ours denotes our full method, using a Mini-Splatting backbone fine-tuned for 5K iterations with StopThePop and Optimal Projection. For both Mini-Splatting and StopThePop, we use the default training parameters. For fine-tuning, we set the total number of iterations to 35K and start from 30K. We use tile sizes of  $16 \times 16$  for center tiles, and  $32 \times 32$  for peripheral tiles. The center region is half the rendered image size, with 10% padding of linearly increasing blending weights.

We evaluate all methods quantitatively and qualitatively in a controlled user study, using three well-established datasets: the Mip-NeRF 360 dataset [Barron et al. 2022], Tanks & Temples [Knapitsch et al. 2017] and Deep Blending [Hedman et al. 2018]. We perform our performance evaluation and user study using the SIBR framework [Bonopera et al. 2020] with OpenXR support and our custom software rasterizer, where we used a Meta Quest 3 tethered to a desktop equipped with an NVIDIA RTX 4090 and render at native resolution ( $2064 \times 2272$  pixels). In addition to our method, we also ablated an optimized two-pass foveated renderer to demonstrate the performance improvements of our single-pass approach. We provide implementation details for this alternative renderer in Appendix A.

### 5.1 User Study

To enable a well-founded qualitative evaluation of our method, we performed a controlled user study with 25 participants for 3DGS-based VR scene viewing using a Meta Quest 3. We use Mini-Splatting [Fang and Wang 2024] for our comparison, as 3DGS [Kerbl et al. 2023] fails to reach the desired framerate target of  $\geq 72$ FPS for some scenes due to its excessive primitive count.

We performed two pairwise comparisons:

- Ours vs Mini-Splatting ( $z$ )
- Ours vs Mini-Splatting (Dist)

Participants were instructed to focus on three distinct criteria: *artifacts*, *quality*, and *preference*, leading to the following hypotheses:

- (1) Our approach produces fewer *artifacts* than other splatting approaches as it avoids popping, elongated Gaussians and cloud-like artifacts.
- (2) Using foveated rendering in our approach does not reduce *quality* compared to other approaches.

- (3) Our approach is fast and robust while avoiding artifacts and thus leads to a *preferable* VR experience compared to other approaches.

To test these hypotheses, we used a pairwise comparison design [Kiran Adhikarla et al. 2017; Mantiuk et al. 2012], where participants experienced a scene rendered with two different methods and then rated them relative to one another. We selected two scenes from the Mip-NeRF 360 dataset [Barron et al. 2022] (*Bonsai* and *Bicycle*) and one scene each from Deep Blending [Hedman et al. 2018] and Tanks & Temples [Knapitsch et al. 2017] (*Dr Johnson* and *Truck*). This selection provided a balanced mix of indoor and outdoor scenes. Each participant viewed two scenes for the Ours vs. Mini-Splatting ( $z$ ) comparison and two for Ours vs. Mini-Splatting (Dist), with the order of scenes and methods randomized to prevent learning effects.

We recruited 25 participants from a local university. After a brief pre-questionnaire, participants rated each pairwise comparison based on general preference, artifacts, and visual quality using a 5-point Likert scale (from “Strongly Prefer Method 1” to “Strongly Prefer Method 2”), with an option for “No Preference.” Participants were instructed to consider *quality* while standing still and focusing on the central object of the current scene, and to move around frequently to assess potential *artifacts*. Each participant could freely explore the scene but could not switch back and forth between conditions. While there was no strict time limit, we encouraged them to transition to the next condition after approximately 90 seconds. Responses were converted into scores  $s \in \{-2, -1, 0, +1, +2\}$ , where  $-2$  indicated a strong preference for Mini-Splatting and  $+2$  indicated a strong preference for our method.

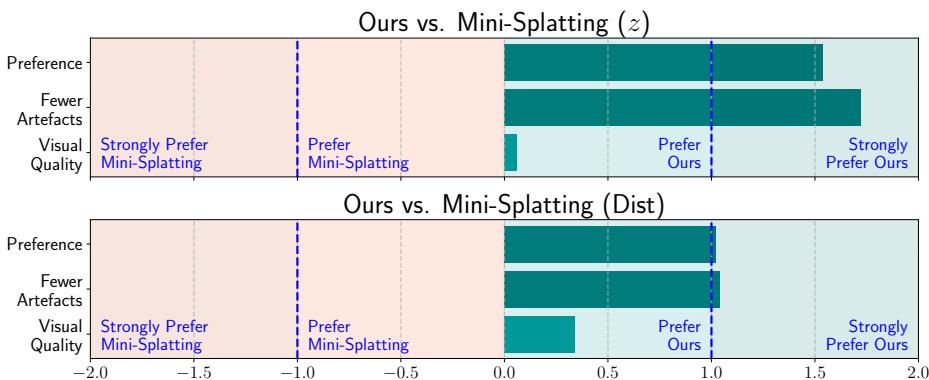


Fig. 6. User Study results: our method is preferred by users across all tested modalities including quality, artifacts and preference. Due to the prevalence of popping artifacts in Mini-Splatting ( $z$ ), users strongly prefer our method. In Mini-Splatting (Dist), projection errors cause the preference for our method.

To analyze the results, we conducted Wilcoxon T-tests on the pairwise results. The average results are shown in Fig. 6. For the Ours vs. Mini-Splatting ( $z$ ) comparison, both *preference* ( $t = 18.0$ ,  $p < .001$ ) and *artifacts* ( $t = 0.0$ ,  $p < .001$ ) were statistically significant, while *quality* ( $t = 209.5$ ,  $p = .62$ ) was not. For the Ours vs. Mini-Splatting (Dist) comparison, all measures were statistically significant: *preference* ( $t = 61.0$ ,  $p < .001$ ), *artifacts* ( $t = 15.0$ ,  $p < .001$ ) and *quality* ( $t = 121.0$ ,  $p < .03$ ).

Clearly, our method is preferred by a large margin in terms of general preference and artifacts. Mini-Splatting ( $z$ ) was rated poorly due to immersion-breaking popping artifacts and errors due to the projection. In contrast to quantitative evaluation on still images, Mini-Splatting obtains *better* qualitative ratings in VR when sorting by the Euclidean distance instead. However, due to projection

errors and occasional popping artifacts, Ours is still preferred by a significant margin. Please refer to the supplemental video for a direct visual comparison and illustration of artifacts. In terms of visual quality, our method is preferred by a small margin despite our foveated rendering solution—this shows that with our blending CUDA kernel, we can effectively reduce resolution in the peripheral region without concerns about visual quality. Although we did not directly compare Mini-Splatting (z) and Mini-Splatting (Dist), the comparison with our method suggests that sorting by Euclidean distance reduces popping artifacts to some extent, slightly improving the VR experience, while the same projection-based artifacts remain disturbing..

In summary, all of our hypotheses were supported, as confirmed by participant feedback. Popping and non-uniform artifacts are significant issues for Mini-Splatting in VR, which our method avoids entirely through the combination of StopThePop and Optimal Projection. Participants commented: “The lighting changes abruptly when I move!” for Mini-Splatting and “This is so much more stable” for our approach. Our combination of strategies, including foveated rendering, does not reduce quality and even outperforms Mini-Splatting (Dist) in terms of user-perceived quality. One participant noted for Mini-Splatting: “When I stand still there are still some unnatural God rays here. And they are even different for the left and right eye.” For our method: “This is clearly better, with no weird artifacts, and it stays consistent when I move my head. Both methods seem equally fast.” While we did not specifically compare our method to 3DGS, previous work suggests that the same preference trends prevail [Tu et al. 2024]. For more details about the user study and per-scene results, see Appendix B.

## 5.2 Image Metrics

We compare our final method against 3DGS [Kerbl et al. 2023], Mini-Splatting [Fang and Wang 2024], and StopThePop [Radl et al. 2024b] in terms of visual quality. We use PSNR, SSIM, LPIPS [Zhang et al. 2018] and present our results in Tab. 1.

Similar to Huang et al. [2024], we follow an adapted protocol, which highlights our robustness to projection errors for large FOVs, as used in VR. Namely, we decrease focal length and increase the render resolution by a factor of 3 $\times$ , which produces elongated and cloud-like Gaussians with standard 3DGS rasterization. We then use a center crop, allowing for a pixel-perfect cutout that can be compared against the test images. Without projection errors, this large FOV cutout should exactly match the low FOV render. As can be seen in Tab. 1, our approach consistently beats related work in this scenario. We also show some qualitative comparisons in Fig. 7.

We also provide image metrics following the standard setup Appendix F. When comparing the numbers across tables, our artifact-free rasterizer is able to retain the image quality, whereas competing methods suffer from severely reduced image metrics. Overall, this evaluation setup demonstrates our improved image quality when considering the requirements for VR rendering.

## 5.3 Ablation Studies

To validate our design choice of fine-tuning with StopThePop [Radl et al. 2024b] and Optimal Projection [Huang et al. 2024] rasterization for 5K iterations, we run a small-scale ablation study. We evaluate Mini-Splatting (z) [Fang and Wang 2024] after {1K, 2.5K, 5K, 10K} iterations of fine-tuning and plot the PSNR in Fig. 8. We include the PSNR score of Mini-Splatting as well as Mini-Splatting with the new rasterization (no fine-tuning) as baselines.

Mini-Splatting with the new renderer initially decreases image quality as the ordering of Gaussians diverges from the learned ordering, causing vastly different per-pixel RGB colors and a significant drop in image quality ( $\sim 1.5$  dB). However, image quality recovers quickly and already approaches the results of Mini-Splatting after 5K fine-tuning steps. Since 5K additional training iterations only result in a minor performance improvement (27.04 vs. 27.09 dB), we choose 5K

Table 1. Standard image quality metrics for our large FOV evaluation protocol. Our proposed method does not suffer from projection artifacts and, therefore, outperforms all other methods on a variety of datasets.

Dataset	Mip-NeRF 360			Tanks & Temples			Deep Blending		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
StopThePop	27.040	0.812	0.213	20.241	0.809	0.192	27.553	0.889	0.243
3DGS	26.820	0.804	0.219	17.112	0.741	0.229	26.192	0.875	0.247
Mini-Splatting (z)	24.765	0.786	0.229	15.355	0.687	0.284	24.347	0.871	0.262
Mini-Splatting (Dist)	25.373	0.794	0.225	17.018	0.754	0.239	26.133	0.880	0.255
Ours (z)	27.293	0.823	0.212	22.901	0.836	0.197	30.333	0.912	0.242
Ours (Dist)	27.262	0.822	0.213	22.840	0.835	0.198	30.337	0.912	0.242

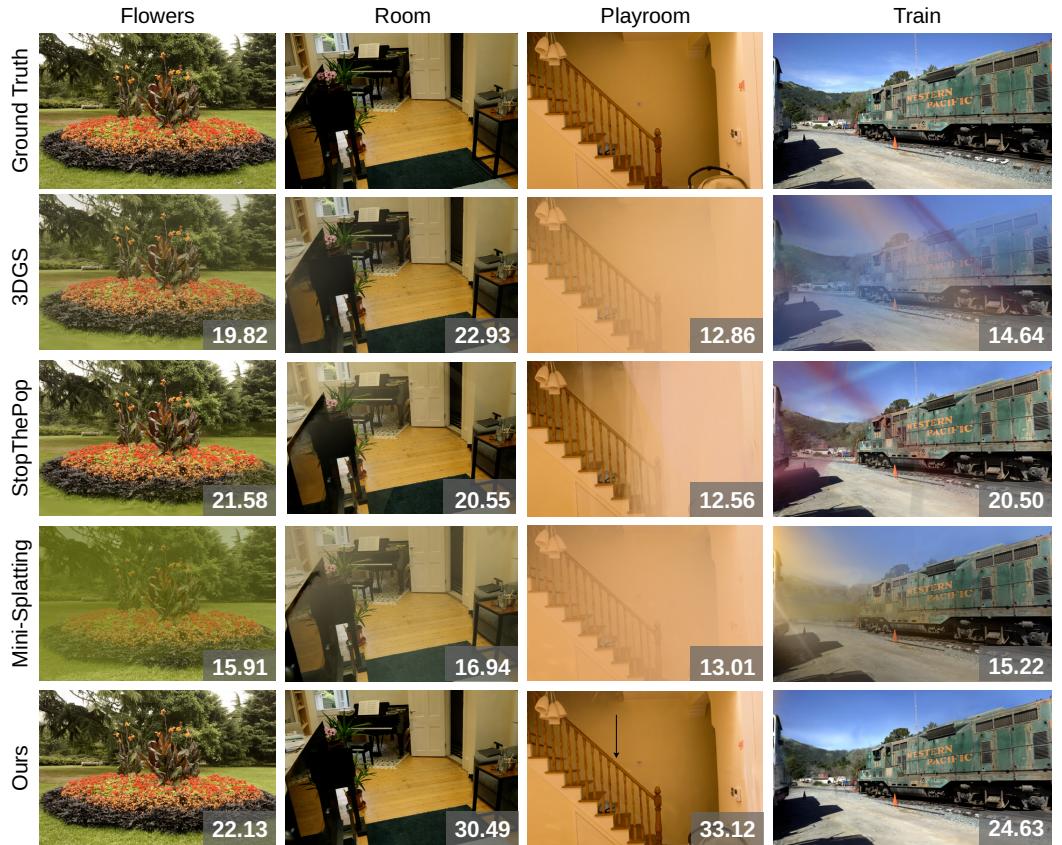


Fig. 7. Images rendered using our large FOV evaluation protocol with inset PSNR values. Projection errors in other methods lead to disturbing artifacts, as also observed in a VR setting.

fine-tuning iterations in the interest of fast optimization. See Appendix C for detailed results and a discussion. We also provide a qualitative evaluation of our method’s components in Fig. 9.

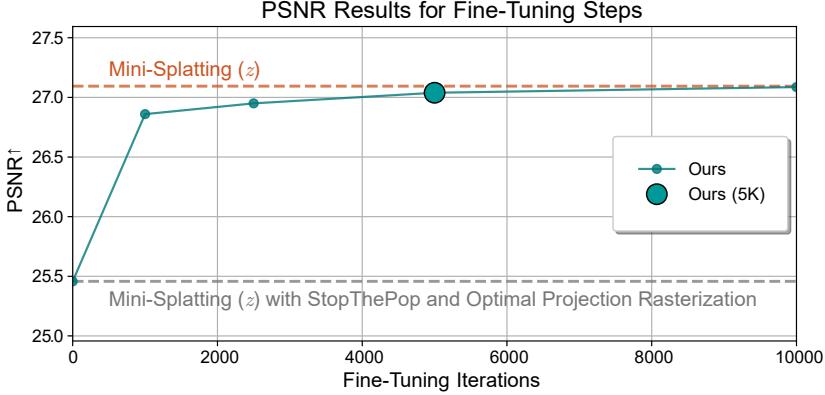


Fig. 8. Ablation study on the number of fine-tuning iterations. We show PSNR averaged over all tested scenes. Whereas the initial quality is significantly reduced with the new rasterization, metrics recover quickly and approach Mini-Splatting ( $z$ ) [Fang and Wang 2024] with 5K fine-tuning iterations.



Fig. 9. Qualitative evaluation of our method's components. When rendering a Mini-Splatting [Fang and Wang 2024] model directly with StopThePop, the changes in sort-order lead to artifacts for both normal and large FOV. Our fine-tuning step allows the model to learn this updated sort order. Finally, Optimal Projection removes artifacts due to projection errors, when rendering with a larger FOV

#### 5.4 Performance Timings

We provide additional render timings to quantify the performance impact of the different techniques in our pipeline. The images were rendered in the resolution as requested by the Meta Quest 3, *i.e.* 2064×2272, and timings were doubled to recreate the stereoscopic setting. To obtain representative timings, we interpolate a camera path between all provided training and test set cameras for each scene and average our measurements over 4 runs. We use CUDA 12.6 with an NVIDIA RTX 4090 for all presented measurements.

Table 2. Average performance timings (in ms) per dataset with our VR evaluation setup (two images with 2064×2272 resolution). Our average render times are below the 14ms threshold, and significantly outperform the simple two-pass foveated rendering routine.

Timings in ms	Deep Blending	Mip-NeRF 360		Tanks & Temples
		Indoor	Outdoor	
3DGS	8.65	8.06	10.35	9.94
Mini-Splatting	4.51	6.24	4.98	4.76
Mini-Splatting + StopThePop	10.55	14.48	11.11	10.95
Ours (w/o fov.)	14.54	20.48	15.91	15.42
Ours (two-pass fov.)	10.35	15.01	12.49	11.18
Ours	8.14	12.36	10.47	9.22

The general performance target for a smooth VR experience on the Meta Quest 3 is 72 FPS (~14ms). As we show in Tab. 2, 3DGS is able to reach this target despite its large model sizes. Mini-Splatting’s considerably smaller models can lead to large performance gains of up to 2.2×. Adding StopThePop on top of Mini-Splatting results in a large increase in rendering times due to the additional sorting overhead. Note that all these methods suffer either from popping and/or projection error artifacts.

Ours without any foveated rendering (w/o fov.) is slower than Mini-Splatting + StopThePop due to the additional overhead from Optimal Projection. A simple two-pass foveated rendering approach accelerates our method, however, there are still inefficiencies due to a duplicate rendering of the blended region (transitional band) and redundant processing of Gaussians in the early, per-Gaussian stages of the rasterizer.

Ours allows us to fully utilize synergies, by fusing kernel calls and rendering the blended region only once. Additionally, as we assign Gaussians to large 32×32 tiles, the workload during Gaussian/tile instantiation is significantly reduced, while the implied rendering overhead is largely mitigated by StopThePop’s hierarchical culling. Overall, Ours reaches framerates comparable to 3DGS and clearly stays below the 14ms render budget for most scenes and viewpoints, while providing an artifact-free VR experience.

The image resolution of 2064 × 2272 results in a total of 4615 tiles (of size 32 × 32), which are further divided into 1085 high-resolution tiles, 2921 low-resolution tiles, 102 hybrid tiles (where blending is performed), and 507 invisible tiles.

We investigate per-scene performance timings in more detail in Tab. 3, where we see that performance gains between 3DGS and Mini-Splatting correlate heavily with model size (number of Gaussians). For example, *Truck* sees a massive decrease in model size by 12× and performance improvements of 2.11×, while *Counter*’s model size only decreases by 3× and performance improves by 1.15×. In scenes, where primitive counts cannot be sufficiently decreased (e.g. *Bonsai*), Ours is unable to beat 3DGS’s performance, mostly due to the per-pixel sorting overhead. Note that performance is also influenced by factors other than model size, such as Gaussian extent, opacity values, and distribution of Gaussians in the scene.

## 6 CONCLUSION, LIMITATIONS AND FUTURE WORK

In this work, we identified the limitations of 3D Gaussian Splatting in delivering high-quality virtual reality experiences. By reviewing recent advances that address some of these challenges, we developed an elegant solution that integrates multiple techniques to enable immersive, artifact-free VR rendering. In addition, we presented a novel single-pass foveated rendering solution building

Table 3. Average performance timings (in ms) per scene for all datasets with our VR evaluation setup (two images with 2064×2272 resolution). Additionally, we report the number of Gaussians  $N$  (in millions) for 3DGS and Mini-Splatting/Ours. Our full method outperforms the two-pass foveated solution and stays within the predefined render budget.

Scene	Bicycle	Flowers	Garden	Stump	Treehill	Bonsai	Counter	Kitchen	Room	DrJ	Playroom	Train	Truck
$N$ (3DGS)	5.95M	3.60M	5.49M	4.84M	3.85M	1.25M	1.20M	1.81M	1.55M	3.28M	2.33M	1.05M	2.56M
N (Mini-Splatting/Ours)	0.52M	0.56M	0.56M	1.00M	0.55M	0.36M	0.41M	0.43M	0.43M	0.37M	0.32M	0.19M	0.21M
3DGS	13.97	8.53	11.91	8.45	8.90	7.32	8.22	9.46	7.24	9.39	7.91	9.81	10.06
Mini-Splatting (MS)	4.91	4.99	5.06	4.99	4.97	6.26	7.11	6.31	5.26	4.69	4.34	4.76	4.76
MS + StopThePop	11.12	10.37	12.18	10.91	10.97	14.39	16.72	14.89	11.91	10.55	10.54	10.73	11.18
Ours (w/o fov.)	15.92	14.91	17.84	15.17	15.71	20.26	23.65	21.22	16.80	14.22	14.86	15.01	15.83
Ours (two-pass fov.)	12.39	11.83	14.24	11.70	12.28	15.24	16.87	15.56	12.36	10.15	10.55	10.74	11.63
Ours	10.14	9.89	12.19	9.80	10.35	12.60	14.11	12.97	9.76	7.88	8.39	8.89	9.55

on the hierarchical StopThePop rasterizer. Our implementation consistently achieves real-time framerates across all tested scenes on a Meta Quest 3 HMD. Furthermore, our formal user study validates these results, demonstrating a clear preference for our method across a diverse set of scenes.

Although our approach successfully mitigates the most prominent artifacts, approximate hierarchical depth sorting may still cause flickering in regions with complex geometric relationships. This highlights the need for a robust level-of-detail (LOD) scheme or fully accurate volume rendering for 3D Gaussians. While ray-tracing 3D Gaussians is a potential solution, recent studies [Blanc et al. 2024; Moenne-Loccoz et al. 2024] indicate that it currently remains too computationally expensive for high-quality VR experiences. There are also areas for optimization in our VR rendering pipeline, such as monocular rendering and stereoscopic reuse of distant content, and the fusion of stereo rendering passes to maximize synergy effects.

The code for fine-tuning our models and the VR viewer is publicly available at <https://github.com/Cekavis/VRSplat>.

## ACKNOWLEDGMENTS

LR, MS, & MS: This research was funded in whole or in part by the Austrian Science Fund (FWF) [10.55776/I6663]. For open access purposes, the author has applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission.

## REFERENCES

- Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. 2022. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. 2023. Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Fields. In *IEEE/CVF International Conference on Computer Vision*.
- Hugo Blanc, Jean-Emmanuel Deschaud, and Alexis Paljic. 2024. RayGauss: Volumetric Gaussian-Based Ray Casting for Photorealistic Novel View Synthesis. arXiv:2408.03356 [cs.CV] <https://arxiv.org/abs/2408.03356>
- Sebastien Bonopera, Jerome Esnault, Siddhant Prakash, Simon Rodriguez, Theo Thonat, Mehdi Benadel, Gaurav Chaurasia, Julien Philip, and George Drettakis. 2020. sibr: A System for Image Based Rendering. [https://gitlab.inria.fr/sibr/sibr\\_core](https://gitlab.inria.fr/sibr/sibr_core)
- Praneeth Chakravarthula, Zhan Zhang, Okan Tursun, Piotr Didyk, Qi Sun, and Henry Fuchs. 2021. Gaze-contingent retinal speckle suppression for perceptually-matched foveated holographic displays. *IEEE Transactions on Visualization and Computer Graphics* 27, 11 (2021), 4194–4203.
- Sean Cier, Xijie Yang, Anton Maridi, and Kem Chen. 2024. MetalSplatter. <https://github.com/scier/MetalSplatter>
- Nianchen Deng, Zhenyi He, Jiannan Ye, Budmonde Duinkharjav, Praneeth Chakravarthula, Xubo Yang, and Qi Sun. 2022. FoV-NeRF: Foveated Neural Radiance Fields for Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (2022), 3854–3864.

- Zhiwen Fan, Kevin Wang, Kairun Wen, Zehao Zhu, Dejia Xu, and Zhangyang Wang. 2024. LightGaussian: Unbounded 3D Gaussian Compression with 15x Reduction and 200+ FPS. In *Advances in Neural Information Processing Systems*.
- Guangchi Fang and Bing Wang. 2024. Mini-Splatting: Representing Scenes with a Constrained Number of Gaussians. In *European Conference on Computer Vision*.
- Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. 2022. Plenoxels: Radiance Fields without Neural Networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. 2012. Foveated 3D Graphics. *ACM TOG* 31, 6, Article 164 (2012), 10 pages.
- Peter Hedman, Julien Philip, True Price, Jan-Michael Frahm, George Drettakis, and Gabriel Brostow. 2018. Deep Blending for Free-viewpoint Image-based Rendering. *ACM TOG* 37, 6, Article 257 (2018), 15 pages.
- Letian Huang, Jiayang Bai, Jie Guo, Yuanqi Li, and Yanwen Guo. 2024. On the Error Analysis of 3D Gaussian Splatting and an Optimal Projection Strategy. In *European Conference on Computer Vision*.
- Ying Jiang, Chang Yu, Tianyi Xie, Xuan Li, Yutao Feng, Huamin Wang, Minchen Li, Henry Lau, Feng Gao, Yin Yang, and Chenfanfu Jiang. 2024. VR-GS: A Physical Dynamics-Aware Interactive Gaussian Splatting System in Virtual Reality. In *SIGGRAPH Conference Papers*. Article 78, 1 pages.
- Anton S Kaplanyan, Anton Sochenov, Thomas Leimkühler, Mikhail Okunev, Todd Goodall, and Gizem Rufo. 2019. DeepFovea: Neural Reconstruction for Foveated Rendering and Video Compression using Learned Statistics of Natural Videos. *ACM TOG* 38, 6 (2019), 1–13.
- Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM TOG* 42, 4 (2023).
- Bernhard Kerbl, Andreas Meuleman, Georgios Kopanas, Michael Wimmer, Alexandre Lanyin, and George Drettakis. 2024. A Hierarchical 3D Gaussian Representation for Real-Time Rendering of Very Large Datasets. *ACM TOG* 43, 4 (2024).
- Shakiba Kheradmand, Daniel Rebain, Gopal Sharma, Weiwei Sun, Yang-Che Tseng, Hossam Isack, Abhishek Kar, Andrea Tagliasacchi, and Kwang Moo Yi. 2024. 3D Gaussian Splatting as Markov Chain Monte Carlo. In *Advances in Neural Information Processing Systems*.
- Vamsi Kiran Adhikarla, Marek Vinkler, Denis Sumin, Rafal K Mantiuk, Karol Myszkowski, Hans-Peter Seidel, and Piotr Didyk. 2017. Towards a quality metric for dense light fields. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Arno Knapsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. 2017. Tanks and Temples: Benchmarking Large-Scale Scene Reconstruction. *ACM TOG* 36, 4, Article 78 (2017), 13 pages.
- Robert Konrad, Anastasios Angelopoulos, and Gordon Wetzstein. 2020. Gaze-Contingent Ocular Parallax Rendering for Virtual Reality. *ACM TOG* 39, 2 (2020), 1–12.
- Brooke Krajancich, Petr Kellnhofer, and Gordon Wetzstein. 2023. Towards Attention-aware Foveated Rendering. *ACM TOG* 42, 4 (2023), 1–10.
- Jiaqi Lin, Zihao Li, Xiao Tang, Jianzhuang Liu, Shiyong Liu, Jiayue Liu, Yangdi Lu, Xiaofei Wu, Songcen Xu, Youliang Yan, and Wenming Yang. 2024b. VastGaussian: Vast 3D Gaussians for Large Scene Reconstruction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Weikai Lin, Yu Feng, and Yuhao Zhu. 2024a. RTGS: Enabling Real-Time Gaussian Splatting on Mobile Devices Using Efficiency-Guided Pruning and Foveated Rendering. *arXiv preprint 2407.00435* (2024). <https://arxiv.org/abs/2407.00435>
- Saswat Mallick, Rahul Goel, Bernhard Kerbl, Francisco Vicente Carrasco, Markus Steinberger, and Fernando De La Torre. 2024. Taming 3DGs: High-Quality Radiance Fields with Limited Resources. In *SIGGRAPH Asia Conference Papers*.
- Rafal K Mantiuk, Anna Tomaszewska, and Radoslaw Mantiuk. 2012. Comparison of four subjective methods for image quality assessment. *Computer Graphics Forum* 31, 8 (2012), 2478–2491.
- Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *European Conference on Computer Vision*.
- Nicolas Moenne-Loccoz, Ashkan Mirzaei, Riccardo Perel, or de Lutio, Janick Martinez Esturo, Gavriel State, Sanja Fidler, Nicholas Sharo, and Zan Gojcic. 2024. 3D Gaussian Ray Tracing: Fast Tracing of Particle Scenes. *ACM TOG* 43, 6 (2024).
- Klaus Mueller and Roger Crawfis. 1998. Eliminating Popping Artifacts in Sheet Buffer-Based Splatting. In *IEEE Visualization*.
- Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM TOG* 41, 4, Article 102 (2022), 15 pages.
- Thu Nguyen-Phuoc, Feng Liu, and Lei Xiao. 2022. SNeRF: Stylized Neural Implicit Representations for 3D Scenes. *ACM TOG* 41, 4 (2022), 142:1–142:11.
- Michael Niemeyer, Fabian Manhardt, Marie-Julie Rakotosaona, Michael Oechsle, Daniel Duckworth, Rama Gosula, Keisuke Tateno, John Bates, Dominik Kaeser, and Federico Tombari. 2024. RadSplat: Radiance Field-Informed Gaussian Splatting for Robust Real-Time Rendering with 900+ FPS. *arXiv preprint 2403.13806* (2024). <https://arxiv.org/abs/2403.13806>
- Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards Foveated Rendering for Gaze-Tracked Virtual Reality. *ACM TOG* 35, 6, Article 179 (2016), 12 pages.

- Lukas Radl, Michael Steiner, Andreas Kurz, and Markus Steinberger. 2024a. LAENeRF: Local Appearance Editing for Neural Radiance Fields. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Lukas Radl, Michael Steiner, Mathias Parger, Alexander Weinrauch, Bernhard Kerbl, and Markus Steinberger. 2024b. StopThePop: Sorted Gaussian Splatting for View-Consistent Real-time Rendering. *ACM TOG* 43, 4, Article 64 (2024).
- Tim Rolff, Ke Li, Julia Hertel, Susanne Schmidt, Simone Frintrop, and Frank Steinicke. 2023a. Interactive VRS-NeRF: Lightning fast Neural Radiance Field Rendering for Virtual Reality. In *Proceedings of the ACM Symposium on Spatial User Interaction*.
- Tim Rolff, Susanne Schmidt, Ke Li, Frank Steinicke, and Simone Frintrop. 2023b. VRS-NeRF: Accelerating Neural Radiance Field Rendering with Variable Rate Shading. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*. IEEE.
- Michael Steiner, Thomas Köhler, Lukas Radl, and Markus Steinberger. 2024. Frustum Volume Caching for Accelerated NeRF Rendering. *Proceedings of the ACM on Computer Graphics and Interactive Technologies* 7, 3, Article 39 (2024), 22 pages.
- Qi Sun, Fu-Chung Huang, Joohwan Kim, Li-Yi Wei, David Luebke, and Arie Kaufman. 2017. Perceptually-Guided Foveation for Light Field Displays. *ACM TOG* 36, 6 (2017), 1–13.
- Taimoor Tariq, Cara Tursun, and Piotr Didyk. 2022. Noise-based Enhancement for Foveated Rendering. *ACM TOG* 41, 4, Article 143 (2022), 14 pages.
- Xuechang Tu, Bernhard Kerbl, and Fernando de la Torre. 2024. Fast and Robust 3D Gaussian Splatting for Virtual Reality. In *SIGGRAPH Asia Posters*.
- Linning Xu, Vasu Agrawal, William Laney, Tony Garcia, Aayush Bansal, Changil Kim, Samuel Rota Bulò, Lorenzo Porzi, Peter Kortschieder, Aljaž Božič, Dahua Lin, Michael Zollhöfer, and Christian Richardt. 2023. VR-NeRF: High-Fidelity Virtualized Walkable Spaces. In *SIGGRAPH Asia Conference Papers*.
- Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. 2024. Mip-Splatting: Alias-free 3D Gaussian Splatting. *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024).
- Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Yuechen Zhang, Zexin He, Jinbo Xing, Xufeng Yao, and Jiaya Jia. 2023. Ref-NPR: Reference-Based Non-Photorealistic Radiance Fields for Controllable Scene Stylization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Mathias Zwicker, Hanspeter Pfister, Jeroen van Baar, and Markus Gross. 2001. EWA Volume Splatting. In *IEEE Visualization*.

## A TWO-PASS FOVEATED RENDERING

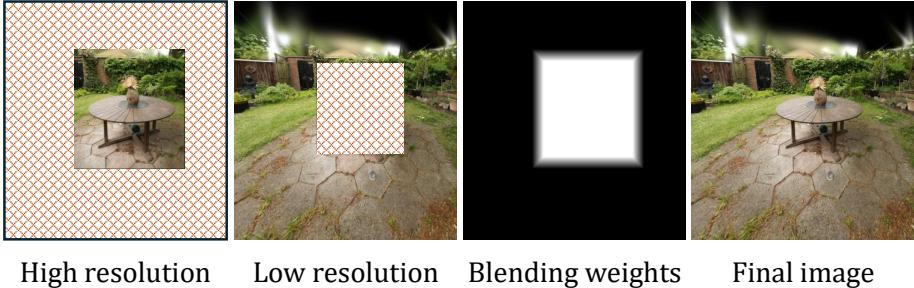


Fig. 10. Depiction of our simple two-pass foveated rendering solution: we render the periphery at half resolution and the center at full resolution, performing blending at the boundary of the visibility mask [Tu et al. 2024].

In addition to our proposed single-pass foveated renderer, we provide a simple, but optimized two-pass foveated rendering approach for comparison. The first pass renders the center region at full resolution, using a tight projection frustum to maximize culling efficiency. The second pass only renders peripheral content at a lower resolution, also considering the visibility mask as provided by OpenXR (see Sec. 4.3).

To guarantee a smooth transition, we render content inside a narrow transition band twice, bilinearly upsample the low resolution content using NVIDIA Performance Primitives, and blend the two render pass images based on the continuous blending mask (see Fig. 10). The size of the center and transitional region for this setup is identical to Ours. This provides a strong baseline for our optimized single-pass foveated rendering scheme.

## B USER STUDY DETAILS

In this section, we detail complimentary demographic information about our set of participants and provide detailed, per-scene user study results.

*Demographic Survey.* Our set of participants for the user study was comprised of 25 participants, aged 23-57. All but one participant had normal or corrected vision: this participant exhibited color blindness. Note that 5 participants did not use their visual aid during the user study.

The majority of the tested population were familiar with computer graphics (3-5 on a 5 point Likert scale): the remaining 4 participants ranked their knowledge of computer graphics as “very poor” or “poor”. 9 participants had never used a VR headset before, the remainder described their usage as “Yearly” up to “Daily”. While 6 participants reported mild motion sickness during the user study, 20/25 ranked the overall experience as “Very satisfying” to “Completely satisfying”. Each user study lasted between 20 and 30 minutes.

*Per-Scene results.* The prevalence of various artifacts within 3DGS is scene-dependent: therefore, we provide per-scene user study results in Fig. 11. As can clearly be seen, there is no clear outlier where Mini-Splatting outperforms our method to a large degree. However, we find results for *Bonsai* vs. Mini-Splatting (Dist) are the least significant out all of tested scenarios. This stems from the fact that artifacts due to projection are less prevalent in this scene (see the supplemental video) and that popping artifacts are significantly reduced when sorting with Euclidean distance.

## C DETAILS ON THE FINE-TUNING ABLATION

In Tab. 4, we provide detailed results for the ablation study presented in Fig. 8. As can be seen, the scores for other image quality metrics generally align with the PSNR results.

For SSIM and LPIPS, the top-3 methods (highlighted) do not include Mini-Splatting; however, StopThePop generally performs very well for these metrics, as can also be seen in Tab. 6.

Table 4. Detailed results for our fine-tuning ablation study: we provide image metrics, averaged over all tested scenes. Clearly, StopThePop fine-tuning is able to recover image quality quickly despite the modified sort order.

Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	FILIP $\downarrow$
Ours (no fine-tuning)	25.458	0.795	0.254	0.184
Ours 1K	26.860	0.832	0.221	0.153
Ours 2.5K	26.950	0.835	0.219	0.151
Ours 5K	27.039	0.836	0.217	0.150
Ours 10K	27.087	0.837	0.215	0.149
Mini-Splatting	27.094	0.835	0.222	0.146

## D LARGE FOV EVALUATION PROTOCOL

As previously discussed, we evaluate the robustness of our approach to projection artifacts in a VR setting, by employing a large FOV evaluation. Huang et al. [2024] evaluate this by changing focal



Fig. 11. Per-Scene user study results: Depending on the scene, the prevalence of errors due to projection or popping artifacts can vary significantly. Despite this, our method still performs favorably for each configuration.

lengths to 0.2 and 0.3 times the original, cut out a patch that corresponds to the original focal length, and resizing the ground truth image for comparison. We instead chose to decrease the original focal length by a factor of 3 and increase render resolution by 3 (see Fig. 12). This has several advantages: (1) the center crop has the exact same resolution as the original render; (2) we receive a pixel-perfect cutout; (3) the dilation of 2D Gaussians (making them at least 1 pixel wide) does not change. This also prevents any unwanted influence on image metrics due to interpolation during resizing or cropping; Hence, for a method without any projection errors the cutout should exactly

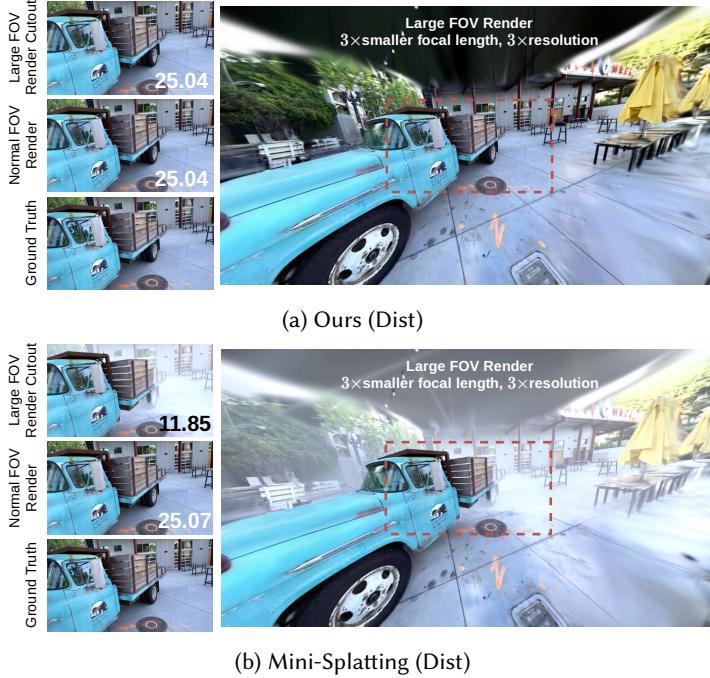


Fig. 12. Visualization of our large FOV evaluation setup, which more closely resembles the VR scenario and allows for comparison against ground truth test images (see inset PSNR values) via a pixel-perfect cutout. Ours does not suffer from projection errors, while Mini-Splatting exhibits very strong artifacts.

match the image rendered with the original focal length. In Fig. 12, the inset PSNR metrics show that this is the case for our method, but not for Mini-Splatting (as well as 3DGS and StopThePop), where projection errors lead to disturbing artifacts and much worse image metrics.

## E PER-SCENE LARGE FOV IMAGE METRICS

In Tab. 5, we provide accompanying per-scene results for the averaged results of Tab. 1.

## F STANDARD IMAGE METRICS

We provide per-scene results for our standard image metrics evaluation (original focal length) in Tab. 7, as well as results averaged over all datasets in Tab. 6.

Table 5. Detailed per-scene image metrics for evaluation with large FOV for all tested scenes.

Dataset Scene	Mip-NeRF 360 Outdoor					Mip-NeRF 360 Indoor				Deep Blending		Tanks & Temples		
	Bicycle	Flowers	Garden	Stump	Treehill	Bonsai	Counter	Kitchen	Room	DrJ	Playroom	Train	Truck	
PSNR $\uparrow$														
3DGS	25.08	20.09	27.10	26.34	20.98	31.78	28.92	30.73	30.37	26.92	25.46	15.92	18.30	
StopThePop	25.21	21.29	27.01	26.70	22.22	31.80	28.57	31.03	29.53	27.52	27.58	19.53	20.95	
Mini-Splatting (z)	23.94	17.64	25.29	27.11	20.76	29.34	27.30	28.77	22.74	23.46	25.24	14.63	16.08	
Mini-Splatting (Dist)	24.16	19.10	25.49	27.06	20.80	29.85	28.32	29.89	23.69	25.31	26.95	16.74	17.29	
Ours (z)	25.28	21.57	26.77	27.19	22.76	31.40	28.65	31.18	30.85	29.81	30.85	21.20	24.61	
Ours (Dist)	25.27	21.67	26.76	27.23	22.79	31.46	28.62	30.74	30.83	29.85	30.82	21.18	24.50	
SSIM $\uparrow$														
3DGS	0.759	0.571	0.864	0.768	0.598	0.940	0.907	0.923	0.902	0.883	0.866	0.692	0.790	
StopThePop	0.766	0.599	0.863	0.775	0.631	0.940	0.904	0.925	0.903	0.889	0.890	0.776	0.842	
Mini-Splatting (z)	0.748	0.526	0.838	0.804	0.615	0.931	0.897	0.915	0.798	0.864	0.878	0.651	0.722	
Mini-Splatting (Dist)	0.742	0.564	0.835	0.804	0.623	0.931	0.903	0.921	0.824	0.872	0.888	0.723	0.785	
Ours (z)	0.775	0.628	0.845	0.808	0.654	0.939	0.906	0.927	0.921	0.908	0.915	0.798	0.874	
Ours (Dist)	0.774	0.628	0.843	0.807	0.653	0.939	0.906	0.926	0.921	0.909	0.915	0.798	0.872	
LPIPS $\downarrow$														
3DGS	0.213	0.353	0.107	0.217	0.338	0.200	0.198	0.126	0.217	0.247	0.246	0.264	0.194	
StopThePop	0.205	0.335	0.106	0.209	0.319	0.200	0.197	0.125	0.216	0.243	0.244	0.218	0.166	
Mini-Splatting (z)	0.231	0.396	0.150	0.198	0.325	0.198	0.198	0.130	0.237	0.274	0.249	0.309	0.258	
Mini-Splatting (Dist)	0.235	0.364	0.154	0.196	0.327	0.197	0.196	0.130	0.227	0.268	0.242	0.277	0.202	
Ours (z)	0.218	0.322	0.148	0.195	0.308	0.196	0.193	0.126	0.206	0.246	0.239	0.238	0.157	
Ours (Dist)	0.220	0.322	0.150	0.194	0.310	0.195	0.193	0.127	0.205	0.245	0.239	0.239	0.158	

Table 6. Standard image quality metrics (without changes in focal length): our proposed method performs well on a variety of datasets. Additionally, our image quality metrics rival those of Mini-Splatting [Fang and Wang 2024], even though only 5K fine-tuning iterations were used

Dataset	Mip-NeRF 360			Tanks & Temples			Deep Blending		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
StopThePop	27.304	0.815	0.211	23.226	0.846	0.171	29.929	0.908	0.231
3DGS	27.443	0.814	0.215	23.734	0.847	0.175	29.510	0.902	0.237
Mini-Splatting (z)	27.326	0.821	0.215	23.308	0.835	0.200	29.994	0.907	0.243
Mini-Splatting (Dist)	27.227	0.821	0.214	23.084	0.834	0.198	30.148	0.909	0.239
Ours (z)	27.294	0.823	0.212	22.901	0.836	0.197	30.333	0.912	0.242
Ours (Dist)	27.263	0.822	0.213	22.840	0.835	0.198	30.337	0.912	0.242

Table 7. Detailed per-scene image metrics for all tested scenes (original focal length).

Dataset Scene	Mip-NeRF 360 Outdoor					Mip-NeRF 360 Indoor				Deep Blending		Tanks & Temples		
	Bicycle	Flowers	Garden	Stump	Treehill	Bonsai	Counter	Kitchen	Room	DrJ	Playroom	Train	Truck	
PSNR $\uparrow$														
3DGS	25.19	21.53	27.30	26.62	22.46	32.11	28.97	31.33	31.48	29.05	29.97	22.05	25.41	
StopThePop	25.22	21.54	27.23	26.70	22.44	31.98	28.60	31.18	30.84	29.45	30.40	21.49	24.96	
Mini-Splatting (z)	25.21	21.46	26.89	27.21	22.65	31.48	28.55	31.11	31.38	29.47	30.52	21.56	25.06	
Mini-Splatting (Dist)	25.22	21.61	26.91	27.22	22.69	31.35	28.52	30.68	30.84	29.64	30.66	21.33	24.84	
Ours (z)	25.28	21.57	26.77	27.19	22.76	31.40	28.65	31.18	30.84	29.81	30.85	21.20	24.61	
Ours (Dist)	25.27	21.67	26.76	27.23	22.79	31.45	28.62	30.74	30.83	29.85	30.82	21.18	24.50	
SSIM $\uparrow$														
3DGS	0.764	0.605	0.864	0.772	0.633	0.941	0.907	0.926	0.919	0.900	0.905	0.814	0.880	
StopThePop	0.768	0.605	0.864	0.776	0.635	0.941	0.905	0.926	0.918	0.906	0.910	0.810	0.882	
Mini-Splatting (z)	0.772	0.625	0.847	0.805	0.653	0.939	0.904	0.924	0.921	0.903	0.910	0.798	0.873	
Mini-Splatting (Dist)	0.773	0.626	0.847	0.805	0.653	0.939	0.904	0.923	0.920	0.905	0.912	0.796	0.873	
Ours (z)	0.775	0.628	0.845	0.808	0.654	0.939	0.906	0.927	0.921	0.908	0.915	0.798	0.874	
Ours (Dist)	0.774	0.628	0.843	0.807	0.653	0.939	0.906	0.926	0.921	0.909	0.915	0.798	0.872	
LPIPS $\downarrow$														
3DGS	0.210	0.335	0.107	0.214	0.326	0.200	0.198	0.125	0.216	0.240	0.234	0.205	0.144	
StopThePop	0.204	0.332	0.106	0.208	0.316	0.199	0.197	0.125	0.214	0.231	0.231	0.202	0.140	
Mini-Splatting (z)	0.222	0.326	0.147	0.197	0.313	0.195	0.195	0.128	0.208	0.252	0.235	0.243	0.157	
Mini-Splatting (Dist)	0.222	0.324	0.147	0.195	0.313	0.194	0.195	0.129	0.206	0.248	0.230	0.241	0.155	
Ours (z)	0.218	0.322	0.148	0.195	0.308	0.196	0.193	0.126	0.206	0.246	0.239	0.238	0.157	
Ours (Dist)	0.220	0.322	0.150	0.194	0.310	0.195	0.193	0.127	0.205	0.245	0.239	0.239	0.158	