

SplatSuRe: Selective Super-Resolution for Multi-view Consistent 3D Gaussian Splatting

Pranav Asthana Alex Hanson Allen Tu Tom Goldstein

Matthias Zwicker Amitabh Varshney

University of Maryland, College Park

<https://splatsure.github.io/>



Figure 1. **SplatSuRe** trains a 3D Gaussian Splatting [14] model to produce sharp, high-resolution novel views from low-resolution inputs. By selectively leveraging high-frequency cues already present in low-resolution training views and applying super-resolution only where needed, our method delivers greater detail and multi-view consistency than prior approaches without any additional training.

Abstract

3D Gaussian Splatting (3DGS) enables high-quality novel view synthesis, motivating interest in generating higher-resolution renders than those available during training. A natural strategy is to apply super-resolution (SR) to low-resolution (LR) input views, but independently enhancing each image introduces multi-view inconsistencies, leading to blurry renders. Prior methods attempt to mitigate these inconsistencies through learned neural components, temporally consistent video priors, or joint optimization on LR and SR views, but all uniformly apply SR across every image. In contrast, our key insight is that close-up LR views may contain high-frequency information for regions also captured in more distant views, and that we can use the camera pose relative to scene geometry to inform where to add SR content. Building from this insight, we propose SplatSuRe, a method that selectively applies SR content only in undersampled regions lacking high-frequency supervision, yielding sharper and more consistent results. Across Tanks & Temples, Deep Blending and Mip-NeRF 360, our approach surpasses baselines in both fidelity and perceptual quality. Notably, our gains are most significant in localized foreground regions where higher detail is desired.

1. Introduction

Novel view synthesis aims to render unseen viewpoints given a set of multi-view images. 3D Gaussian Splatting (3DGS) [14] enables real-time, photorealistic novel view synthesis by representing scenes as an explicit set of anisotropic Gaussians optimized through differentiable splatting. While 3DGS excels at efficiency and reconstruction quality, its performance is tightly coupled to the resolution of the training images. Models trained on low-resolution (LR) inputs lack access to high-frequency signals present in high-resolution (HR) views, resulting in blurry textures, over-smoothed surfaces, and aliasing artifacts when rendered at higher test-time resolutions [33].

When restricted to only LR views, a natural strategy is to apply super-resolution (SR) to enhance them before fitting a 3D model [6]. However, single-image SR operates independently on each view and frequently introduces view-dependent hallucinated textures [27]. When these inconsistent SR predictions are used as direct supervision, the resulting 3D optimization receives conflicting gradients across viewpoints, degrading model quality. Prior methods attempt to mitigate these inconsistencies through learned neural components, temporally consistent video priors, or joint optimization on LR and SR views [10, 23, 25, 30, 32]. However, these methods inject SR content *uniformly* across

the image, regardless of whether a region actually benefits from generative detail or is already well-constrained by existing LR observations.

In contrast, the key observation motivating our work is that images of a scene do not sample 3D content uniformly. A low-resolution view captured up close often contains enough high-frequency detail to supervise rendering of more distant views that observe the same region only coarsely. This disparity in multi-view sampling implies that many views already receive high-frequency supervision from closer views, whereas others that do not have any closer views providing sufficient higher-frequency information would benefit from SR guidance. Applying SR indiscriminately therefore introduces unnecessary inconsistencies in well-resolved regions.

Based on this observation, we propose **SplatSuRe**, a selective super-resolution framework for multi-view consistent 3D Gaussian Splatting. Rather than uniformly enhancing all pixels, SplatSuRe identifies 3D regions that lack high-frequency observations and injects SR only where it is beneficial. We first compute a *Gaussian fidelity score* that measures how well each Gaussian is sampled across training views, then render per-view *super-resolution region selection* weight maps that highlight undersampled areas while suppressing SR where LR supervision is already reliable. These maps modulate SR supervision during training, allowing the model to exploit generative detail in under-resolved regions while maintaining consistency elsewhere. Through this geometry-aware selective refinement, SplatSuRe produces sharper reconstructions and improved perceptual quality without introducing additional neural components or modifying the underlying 3DGS pipeline.

In summary, we propose the following contributions:

1. A per-Gaussian fidelity score that quantifies how well it is resolved across views, leveraging LR geometry to estimate available high-frequency information.
2. A per-view spatial map of available frequency information computed using the Gaussian fidelity score that modulates SR supervision during optimization.
3. A selective SR training framework that jointly optimizes a 3DGS model using LR and SR supervision, injecting generative detail only where needed while preserving multi-view consistency and achieving state-of-the-art results across a diverse range of scenes.

2. Related Work

3D Gaussian Splatting (3DGS) [14] enables real-time novel view synthesis by representing scenes as sets of anisotropic Gaussians optimized through differentiable rasterization. While it achieves high rendering efficiency, models trained with low resolution (LR) images suffer from aliasing artifacts when rendered at higher resolutions, since Gaussians optimized at coarse scales become undersampled.

Mip-Splatting [33] mitigates this aliasing by applying scale-adaptive 3D and 2D filtering while preserving radiance energy across resolutions, similar in spirit to Mip-NeRF [1]. While achieving significant improvement over vanilla 3DGS, its rendering quality is still tied to the resolution of training views and blurring can occur at higher resolutions. In contrast, our method uses super-resolution to further inform high-resolution information.

Super-resolution (SR) has long been applied to neural radiance fields [20] to enhance novel view synthesis quality [5, 10, 11, 22, 24, 26, 35]. More recently, SR has been extended to 3D Gaussian Splatting (3DGS). Several variants aim to recover high-resolution detail and improve multi-view consistency [6, 23, 25, 30] through residual feature learning [30], uncertainty modeling [25, 30], per-scene refinement [25] or video super-resolution [23]. Among these, SRGS [6] jointly optimizes Gaussian parameters using both LR ground-truth images and super-resolved views produced by a frozen single-image SR model. However, applying this enhancement uniformly across all regions does not eliminate the effect of inconsistencies introduced by super-resolution. S2Gaussian [25] focuses on sparse view reconstruction and proposes an inconsistency modeling module trained per-scene to reduce inconsistencies in SR images. SuperGaussian [23] applies a video SR network to frames rendered from a low-resolution 3DGS model, using the resulting temporally consistent sequence for retraining the 3D model. While these approaches improve fidelity, they either rely on additional neural components to enforce consistency or lack spatial adaptivity in the use of SR. In contrast, our method utilizes camera pose information relative to scene geometry to determine undersampled regions, and selectively applies super-resolution in those regions to enhance sharpness while maintaining fidelity.

Diffusion methods offer complementary advances in 3D representation and enhancement tasks. Several works jointly optimize diffusion and 3D parameters for view-consistent generation and super-resolution [2, 17, 32, 36]. GaussianSR [32] distills information from 2D super-resolution models as a loss for training a 3DGS model. 3DSR [2] uses a 3DGS model to enforce 3D consistency in the diffusion process, fitting it multiple times through denoising diffusion steps. Diffusion priors have also been applied as post-processing to enhance renders from 3DGS models [18, 28, 29]. While these methods rely on large pretrained diffusion models to predict images that reduce 3D inconsistencies, we leverage explicit geometric relationships between cameras and scene structure to determine where generative detail is needed, which can then be added via any of these complementary methods.

In addition to image and video-based models, SR can also be performed directly in 3D, circumventing multi-view inconsistencies. Geometric point-based networks [3, 4, 19,

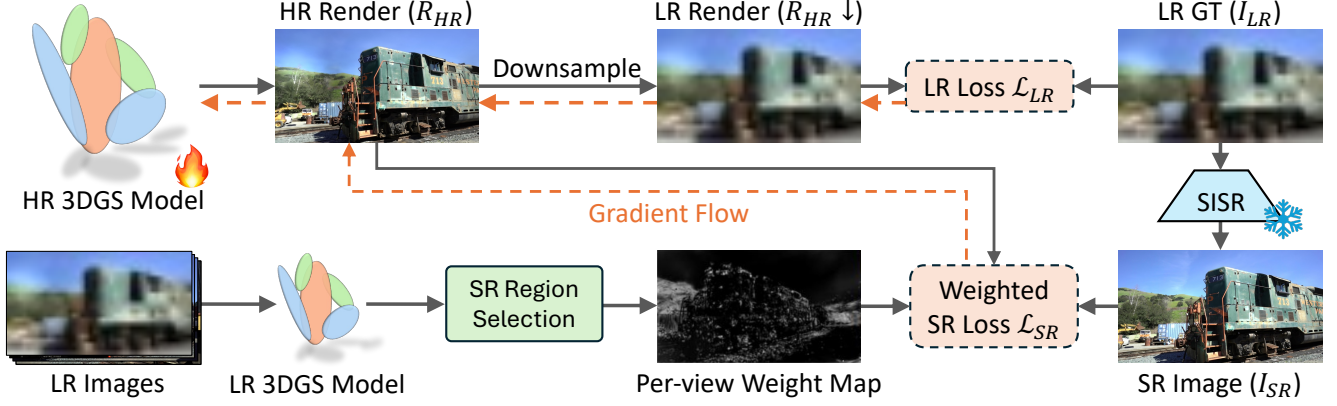


Figure 2. **Overview of our SplatSuRe framework.** A high-resolution (HR) 3D Gaussian Splatting (3DGS) model is trained using low-resolution (LR) and super-resolution (SR) inputs. We first train a 3DGS model on LR inputs to identify undersampled regions and render per-view weight maps that indicate where SR is needed. During training of the HR 3DGS model, the images produced by the frozen single-image super-resolution (SISR) model are spatially weighted by these maps to form the SR loss $\mathcal{L}_{SR}$. A complementary LR loss $\mathcal{L}_{LR}$ compares the downsampled HR render against the original LR ground truth to provide consistent supervision across the entire image.

31] locally upscale point clouds to recover fine-grained geometry. However, these approaches typically only upsample geometry, while texture resolution still relies on 2D methods. As before, our method is orthogonal to these techniques and can be combined with them to further improve reconstruction fidelity.

3. Background

3.1. 3D Gaussian Splatting

3D Gaussian Splatting (3DGS) [14] models a scene as a set of 3D Gaussians, each represented by a mean position $\mu \in \mathcal{R}^3$, per-axis scale $s \in \mathcal{R}^3$, rotation vector $q \in \mathcal{R}^4$, scalar opacity $o \in \mathcal{R}_+$, and view-dependent color c , represented as a base color with spherical harmonic coefficients. Volumetric rendering with alpha blending is used to splat the Gaussians onto the image plane and the resulting color for pixel p is given by:

$$C(p) = \sum_{i=1}^N c^i \alpha^i(p) \prod_{j=1}^{i-1} (1 - \alpha^j(p)), \quad (1)$$

where N is the number of Gaussians intersecting the pixel ray, $\alpha^i(p) = o^i e^{-\frac{1}{2}(p - \mu^i)^T (\Sigma_{2D}^i)^{-1} (p - \mu^i)}$ is the contribution of the i^{th} Gaussian, and c^i is its view-dependent color. The 2D covariance matrix Σ_{2D} is given by the EWA Splatting approximation [37]. Gaussians are sorted in depth order before splatting to ensure that transmittance is computed correctly.

The model is initialized using Structure from Motion (SfM) on the training views to provide camera parameters and produce a sparse point cloud, serving as the initial Gaussian means. During model training, images are randomly selected from the training set, and a weighted sum of $\mathcal{L}_1$ and $\mathcal{L}_{D-SSIM}$ losses is used to optimize the Gaussian

parameters using gradient descent. Gaussians are split and cloned throughout training to increase scene fidelity and ensure sufficient primitives where needed. We retain the standard 3DGS rendering and densification pipeline, modifying only the supervision losses for our method.

3.2. Super-Resolution for Gaussian Splatting

When trained solely on low-resolution (LR) images, 3DGS models lack access to the high-frequency cues present in high-resolution (HR) views, leading to over-smoothed textures and incomplete fine structure. This limitation has motivated the integration of super-resolution (SR) models into the 3DGS training pipeline. SRGS [6] employs a frozen single-image SR model to generate super-resolved views and jointly optimizes Gaussian parameters against both LR ground truth and SR outputs, supplying explicit high-frequency supervision that LR views alone cannot provide. However, SRGS applies SR *uniformly* across the entire image, even in regions that already receive reliable high-frequency supervision from nearby LR views. Since SR is applied independently to each image, the generated details are not necessarily multi-view consistent – injecting them everywhere can introduce geometric or texture inconsistencies, leading to averaging effects in the model that render as blurring.

These observations reveal a fundamental challenge: *super-resolution is not uniformly beneficial across the scene*. As illustrated in Figure 3, some regions already receive sufficient high-frequency supervision from closer LR training views, so adding generated detail introduces unnecessary inconsistencies that harm cross-view coherence. Other regions, particularly those that are distant or sparsely observed, are undersampled and require SR to recover missing details. This motivates a selective, geometry-aware strategy that determines where SR should influence opti-

mization. Instead of treating all pixels equally, we aim to exploit the multi-view sampling pattern of each Gaussian to determine which regions are sufficiently constrained by LR observations and which require additional SR guidance.

4. Method

Super-resolution (SR) mostly benefits regions that lack reliable high-frequency supervision, motivating a geometry-aware mechanism for deciding where SR should influence 3DGS optimization. Our method identifies these undersampled areas and applies SR only where high-frequency detail is missing, improving sharpness while avoiding unnecessary inconsistencies. To achieve this, we compute a Gaussian fidelity score that measures how well each Gaussian is sampled across training views, then render per-view weight maps that highlight undersampled areas while suppressing SR where LR supervision is already reliable. Incorporating these weight maps into a combined LR–SR objective yields our **SplatSuRe** framework, shown in Figure 2, which selectively injects detail where it is beneficial while preserving multi-view consistency elsewhere.

4.1. Gaussian Fidelity Score

Images capturing a scene do not contribute equal amounts of high-frequency information for 3D reconstruction. A low-resolution (LR) view taken at a short distance or with a long focal length can contain more fine detail than a high-resolution (HR) view captured from further away. We leverage this inherent disparity to determine which 3D regions already have adequate high-frequency supervision and which require additional details from super-resolution (SR). Figure 3 illustrates how a nearby LR image can provide the high-frequency detail needed to supervise a distant viewpoint.

To measure per-Gaussian relative sampling frequency, we first train a low-resolution 3DGS model using the LR images. This provides stable scene geometry and allows us to compute each Gaussian’s screen-space radius in every training view. Following 3DGS, the screen-space radius, measured in pixel units, of Gaussian $\mathcal{G}^i$ is:

$$r^i = 3\sqrt{\max(\lambda_1^i, \lambda_2^i)}, \quad (2)$$

where λ_1^i and λ_2^i are the eigenvalues of the Gaussian’s 2D covariance matrix Σ_{2D}^i , calculated as:

$$\lambda_1^i, \lambda_2^i = \frac{1}{2}\text{tr}(\Sigma_{2D}^i) \pm \sqrt{\max\{0.1, \frac{1}{4}\text{tr}^2(\Sigma_{2D}^i) - |\Sigma_{2D}^i|\}}, \quad (3)$$

where $\text{tr}(\Sigma_{2D}^i)$ denotes the trace of the projected covariance matrix. For each Gaussian, we then compute the ratio ρ^i

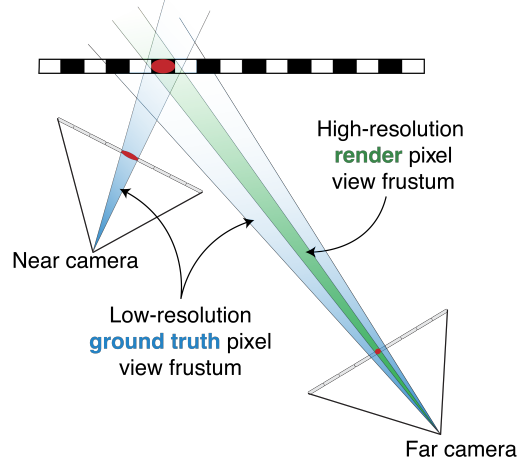


Figure 3. **Disparity in high-frequency ground truth information across different views.** Low-resolution ground truth from near cameras provides high-resolution information for rendering distant views, reducing the need for additional generated detail in those views. Conversely, super-resolution is needed in views where no other camera provides higher-resolution information.

between its maximal and minimal radius across all training views T in which it contributes to the rendering:

$$r_{min}^i = \min_{t \in T} r_t^i, \quad r_{max}^i = \max_{t \in T} r_t^i, \quad (4)$$

$$\rho^i = r_{max}^i / r_{min}^i. \quad (5)$$

A high ratio indicates that the Gaussian is sampled at varied frequencies across views, meaning some views observe it with high fidelity and can supervise the others. Conversely, a ratio close to one indicates uniformly sampling across views, indicating that this region lacks any higher-frequency observations and requires SR to add generated details. This interpretation also holds for view regions where the Gaussian projects near its maximal radius, since no other views provide higher frequency information.

Note that 3DGS dilates each Gaussian by convolving it with a fixed low-pass Gaussian filter to prevent aliasing and ensure a minimal rendering size:

$$\Sigma_{2D}^i = \Sigma_{2D}^i + s\mathbf{I}, \quad (6)$$

where $s=0.3$ is the amount of dilation. Since this dilation artificially inflates the radius, especially for distant Gaussians where the blur dominates, we exclude this dilation when computing radii for the ratio.

We then transform the raw ratio ρ^i into a *Gaussian fidelity score* that maps each Gaussian to a weight in $[0, 1]$, with lower values indicating a greater need for SR. The ratio is first offset by a threshold τ and then mapped into the unit interval using a sigmoid function, giving the per-Gaussian score:

$$\text{score}_{\mathcal{G}^i} = \sigma\left(\frac{\rho^i - \tau}{k}\right), \quad (7)$$

where σ is the sigmoid function and $k=0.05$ controls the smoothness of the transition from 0 to 1. The *ratio threshold* τ is a hyper-parameter that depends on the structure of the scene and the consistency of the SR model across views. Scores for Gaussians visible in fewer than three views are set to zero because these regions are poorly constrained. Higher scores correspond to Gaussians that are already well-captured by LR supervision, while lower scores identify regions where SR should be applied more heavily.

4.2. Super-Resolution Region Selection

Our Gaussian fidelity score provides a scene-level measure of how well a Gaussian is observed across views, but supervision of high-resolution model updates requires pixel-wise weight maps for each training view. For a given training view t , we identify the set of Gaussians whose maximal radius occurs in its rendered view:

$$\mathcal{M}(t) = \{\mathcal{G}^i \mid t = \operatorname{argmax}_{t \in T} r_t^i \forall i \in N\}, \quad (8)$$

where $\mathcal{G}^i$ is the i^{th} Gaussian, r_t^i is its screen-space radius, and N is the total number of Gaussians. These are Gaussians that do not receive higher-frequency information from another view. The weight map for training view t is then rendered as:

$$W'_t = (1 - \operatorname{Render}(\operatorname{score}_{\mathcal{G}})) + \operatorname{Render}(\mathbf{1}_{\mathcal{M}(t)}(\mathcal{G})), \quad (9)$$

where $\operatorname{Render}(\cdot)$ denotes splatting the LR model by alpha-blending the specified per-Gaussian values, $\mathcal{G}$ is the set of all Gaussians, $\operatorname{score}_{\mathcal{G}}$ is the vector of Gaussian fidelity scores, and $\mathbf{1}_{\mathcal{M}(t)}(\mathcal{G})$ is an indicator that is 1 for Gaussians in $\mathcal{M}(t)$ and 0 otherwise. The first term in Equation 9 ensures that undersampled regions with low fidelity scores receive SR, while the second term ensures that areas observed most closely by the current view also receive SR because no alternative view provides higher resolution information. Finally, the weight map is normalized to ensure the magnitude of the SR loss is consistent across views. As illustrated in Figure 4, this per-view weight map is high in regions requiring SR and low in regions already sufficiently supervised by LR views.

4.3. SplatSuRe Training Objective

Our super-resolution region selection method determines how much SR should influence each pixel during optimization. Rather than uniformly applying SR to the entire image during training, we incorporate our weight map W_t to supervise training using two complementary signals: (1) low-resolution ground truth images, and (2) selectively-weighted super-resolved images.

In each training iteration, the model is rendered at the target high resolution. The render R_{HR} is downsampled to



Figure 4. **Super-resolution weight maps.** Bright regions indicate areas where generative detail is required, while dark regions correspond to areas well-sampled by other low-resolution views. Note that high weights are obtained in regions that are either not sampled closely, such as background trees behind the tractor, or where other views do not provide higher resolution information, such as the foreground table in the ballroom.

produce $R_{HR} \downarrow$ and compared with the LR ground truth I_{LR} , yielding the LR loss:

$$\mathcal{L}_{LR} = (1 - \lambda)\mathcal{L}_1(R_{HR} \downarrow, I_{LR}) + \lambda\mathcal{L}_{D-SSIM}(R_{HR} \downarrow, I_{LR}), \quad (10)$$

where $\mathcal{L}_1$ and $\mathcal{L}_{D-SSIM}$ are the losses used in 3DGS. The super-resolved image I_{SR} is compared with the HR render R_{HR} using a spatially weighted loss:

$$\mathcal{L}_{SR} = (1 - \lambda)\mathcal{L}_1^W(R_{HR}, I_{SR}) + \lambda\mathcal{L}_{D-SSIM}^W(R_{HR}, I_{SR}), \quad (11)$$

where each pixel’s contribution is scaled by its weight in W_t . High weights amplify SR supervision in undersampled regions, while low weights suppress it where LR views already provide reliable high-frequency information.

The overall objective combines LR ground truth supervision with selectively-weighted SR guidance:

$$\mathcal{L} = (1 - \gamma)\mathcal{L}_{LR} + \gamma\mathcal{L}_{SR}, \quad (12)$$

where γ controls the relative contribution of each term. This **SplatSuRe** formulation produces a high-resolution 3DGS model that leverages SR only where it improves reconstruction quality, avoiding inconsistencies in regions already well-constrained by LR views.

5. Experiments

We integrate our **SplatSuRe** method into the 3D Gaussian Splatting (3DGS) codebase [18], adding modules for true Gaussian radius computation, weight map computation and rendering, and auxiliary loss terms. Experiments are conducted at $4\times$ super-resolution using StableSR [27] and using a ratio threshold $\tau=1.1$ to generate weight maps. Our choice of ratio threshold and SR model are ablated in Sections 7.1 and 7.2, respectively. We use $\lambda=0.2$ and $\gamma=0.4$ for our losses defined in Equations 10-12.

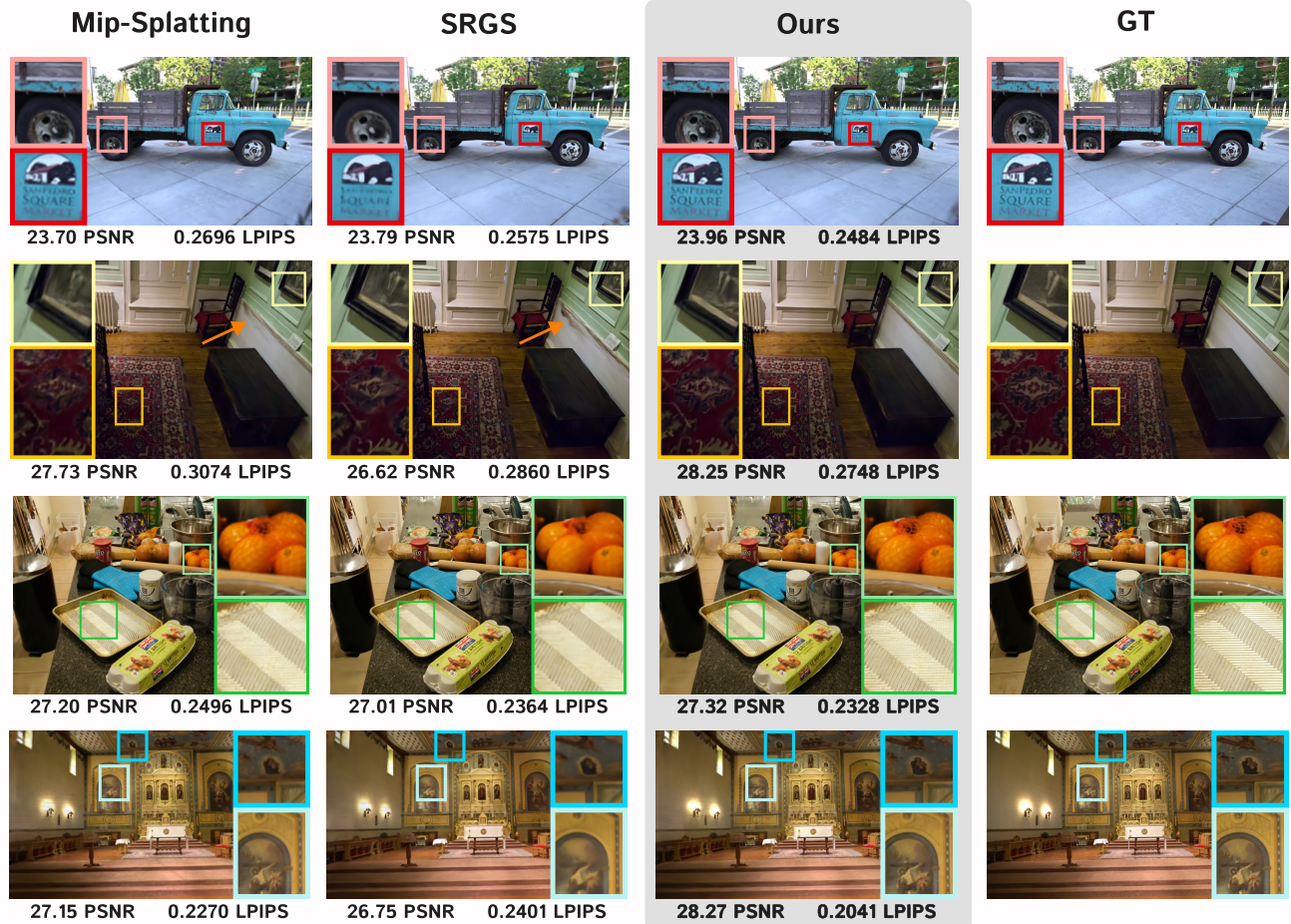


Figure 5. **Qualitative results on Tanks & Temples [15], Deep Blending [8] and Mip-Nerf 360 [1].** Experiments are performed at $4\times$ super-resolution with ratio threshold $\tau=1.1$. Compared to Mip-Splatting [33] and SRGS [6], our method produces sharper, more faithful reconstructions that better align with ground truth while maintaining cross-view consistency. It preserves finer details in text (red box on truck), high-frequency patterns (yellow box on carpet and green box on tray) and recognizable distant objects observed in other views (blue box on church mural). Notably, it reduces Gaussian artifacts (orange arrow) observed in other methods. Additional results in Appendix A.3.

Datasets. We evaluate our method on three real-world datasets that provide a diverse mix of indoor and outdoor environments, scenes with both circular and elongated central objects, and camera paths ranging from smooth trajectories to irregular captures. **Tanks & Temples [15]** includes 21 real-world scenes of varying scales, of which we use 19 – two are excluded because COLMAP fails. Images are downsampled to 240×135 and upsampled $4\times$ to 960×540 . Each scene contains roughly 150-500 images, with every eighth image used for testing and the remainder for training. **Mip-NeRF 360 [1]** consists of nine real-world scenes – five outdoor and four indoor – with about 250–300 images each at approximately $4K \times 3K$ resolution. We downsample these images by $8\times$ ($\sim 500 \times 375$) and upsample by $4\times$ to half the native resolution ($\sim 2K \times 1.5K$), again reserving every eighth image for testing. Finally, we use two scenes from **Deep Blending [8]**, which provides a collection of forward-facing indoor scenes captured under diverse light-

ing conditions. Each scene contains roughly 250 images at $\sim 1K \times 1K$ resolution, which we downsample by $4\times$ for training and evaluate on the original full-resolution renders.

Baselines. We compare SplatSuRe with four representative baselines using their official implementations on our evaluation datasets. **3DGS (LR) [14]** is trained on low-resolution inputs and serves as the primary baseline. **StableSR [27]** applies a single-image SR model to the low-resolution training data and fits a 3DGS model on the super-resolved images. **Mip-Splatting [33]** is trained at low resolution and mitigates aliasing through 3D and 2D multi-scale filtering for high-resolution rendering. Finally, **SRGS [6]** optimizes Gaussian parameters using both low-resolution and super-resolved images from a frozen SR model, without additional pretraining or scene-specific fine-tuning.

Metrics. We evaluate both reconstruction fidelity and perceptual realism using eight complementary metrics. For

Table 1. **Quantitative results on Tanks & Temples [15]**. Experiments are performed at $4\times$ super-resolution using ratio threshold $\tau=1.1$. The **best**, **second best** and **third best** entries are highlighted. Our SplatSuRe method achieves the strongest results across all metrics.

Method	Tanks & Temples [15]							
	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	FID $\downarrow$	CMMD $\downarrow$	DreamSim $\downarrow$	MUSIQ $\uparrow$	NIQE $\downarrow$
3DGS (LR) [14]	0.669	19.41	0.350	71.58	2.013	0.0895	57.776	3.412
StableSR [27]	0.751	22.47	0.300	59.29	1.123	0.0667	56.748	4.945
Mip-Splatting [33]	0.767	23.10	0.303	52.46	1.137	0.0597	46.571	5.043
SRGS [6] + StableSR [27]	0.771	23.32	0.286	49.11	1.048	0.0535	55.209	4.633
Ours + StableSR [27]	0.784	23.81	0.272	37.72	1.040	0.0413	58.332	3.928

Table 2. **Quantitative results on Deep Blending [8] and Mip-NeRF 360 [1]**. Our SplatSuRe method achieves the strongest results across all metrics on Deep Blending and outperforms SRGS [6] on Mip-NeRF 360. Appendix A.1 and A.2 present $8\times$ SR and per-scene results.

Method	Deep Blending [8]					Mip-NeRF 360 [1]				
	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	CMMD $\downarrow$	DreamSim $\downarrow$	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	CMMD $\downarrow$	DreamSim $\downarrow$
3DGS (LR) [14]	0.836	26.72	0.335	0.868	0.0574	0.635	20.67	0.384	0.645	0.0529
StableSR [27]	0.845	27.16	0.325	0.630	0.0505	0.699	24.28	0.351	0.725	0.0387
Mip-Splatting [33]	0.865	28.43	0.327	0.690	0.0398	0.759	26.48	0.292	0.183	0.0132
SRGS [6] + StableSR [27]	0.861	28.23	0.317	0.630	0.0409	0.734	25.92	0.317	0.339	0.0194
Ours + StableSR [27]	0.872	29.01	0.306	0.496	0.0330	0.740	26.34	0.323	0.339	0.0179

reference-based assessment, we report SSIM, PSNR, and LPIPS [34], which measure pixel and feature-level agreement with ground-truth high-resolution images. To capture distributional and semantic perceptual quality, we include FID [9], CMMD [12], and DreamSim [7], which compare the feature distributions or embeddings of generated and real images in pretrained perceptual spaces. Finally, we evaluate no-reference perceptual quality using MUSIQ [13] and NIQE [21], which estimate realism and naturalness without ground-truth supervision. While PSNR and SSIM directly measure pixel fidelity at the native resolution, the perceptual and distributional metrics internally downsample or resize images before feature extraction, reducing sensitivity to high-frequency details. As a result, improvements in fine-scale sharpness – central to super-resolution methods – may be underrepresented by these metrics, motivating our use of a broad set of complementary evaluations.

6. Results

Qualitative Results. Figures 1 and 5 present representative renderings from our method. Our approach produces consistently sharper reconstructions while preserving smoothness in uniformly textured regions. In contrast, Mip-Splatting yields overly blurred results because it is trained only on LR and its 3D anti-aliasing filter excessively blurs all Gaussians. SRGS recovers higher-frequency content in some areas but remains constrained by the underlying SR images – regions where the SR output is consistent appear sharp, whereas areas with view-inconsistent SR predictions become noticeably blurred or distorted.

Quantitative Results. Table 1 presents results on Tanks & Temples [15]. Our method achieves the highest image quality across nearly all metrics, demonstrating both its effectiveness and robustness. The only exception is NIQE, where 3DGS attains a better score. Similarly, SRGS ranks second across most reference and perceptual metrics but lags behind in MUSIQ and NIQE, where 3DGS performs better. This trend arises because the aliased, noisy renderings of low-resolution 3DGS coincidentally resemble the natural image statistics favored by no-reference quality metrics, whereas the smoother and more coherent outputs of Mip-Splatting and SR-based methods appear less “natural” under such measures. In contrast, our approach delivers substantially higher perceptual fidelity and cross-view consistency, producing sharper and more realistic high-resolution reconstructions that align closely with both human perception and ground-truth images.

Table 2 reports results on Deep Blending [8] and Mip-NeRF 360 [1]. Our method achieves the strongest performance across all metrics on Deep Blending. On Mip-NeRF 360, it outperforms SRGS on every metric except LPIPS; however, both SR-based approaches are surpassed by Mip-Splatting. This advantage arises from the characteristics of Mip-NeRF 360, where smooth, circular camera trajectories provide dense multi-view coverage and minimal under-sampling. In such well-sampled settings, Mip-Splatting’s multi-scale anti-aliasing filters effectively preserve radiance energy and suppress aliasing, whereas SR-based methods may introduce slight inconsistencies when enhancing already well-resolved regions. Additionally, the LR images in Mip-NeRF 360 are approximately twice as large as those

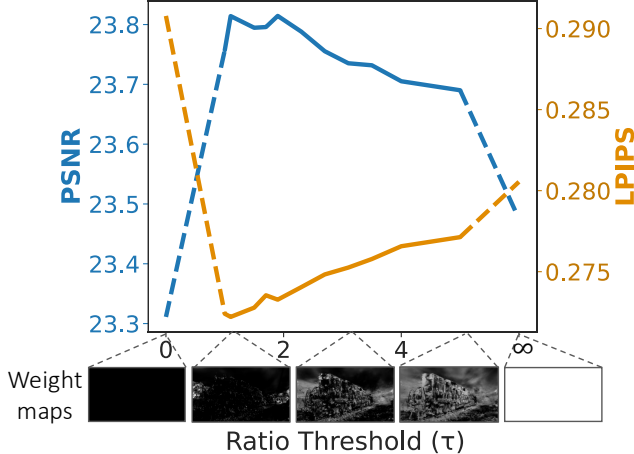


Figure 6. **Effect of ratio threshold on Tanks & Temples [15].** Weight maps, where bright regions indicate higher SR influence, are shown below the corresponding ratio thresholds. $\tau=0$ and $\tau=\infty$ correspond to zero and full use of super-resolution. SR is initially helpful in improving rendering quality, but excessive use worsens results. The effect of ratio threshold on different scenes is analyzed in Appendix A.2.

in the other datasets. Since most high-frequency details are already preserved at this scale, SR offers minimal improvement. Overall, SplatSuRe achieves the highest performance on Tanks & Temples and Deep Blending and outperforms SRGS on Mip-NeRF 360. Results for $8\times$ SR, additional visualizations, and per-scene metrics for Tables 1 and 2 can be found in Appendix A.1, A.3, and A.4.

7. Ablations

7.1. Ablation on Ratio Threshold

Scenes with different geometries and scales exhibit varying distributions of high and low-frequency content across training views. To account for this variability, we examine the effect of the ratio threshold used to determine where SR is applied. As shown in Figure 6, introducing a small amount of SR initially improves both PSNR and LPIPS by recovering fine details in undersampled regions. However, applying SR too aggressively leads to inconsistencies across views, degrading overall performance. This trend is most pronounced in scenes with varied camera-object distances, where sampling density differs significantly across views. Based on this analysis, we select ratio threshold $\tau=1.1$, which provides the best trade-off between sharpness and consistency across scene, in our main experiments. See Appendix A.2 for a detailed analysis on different scenes.

7.2. Ablation on Super-Resolution Model

We evaluate our method using two single-image super-resolution (SISR) models: SwinIR[16] and StableSR[27]. SwinIR is optimized for high fidelity under pixel-based metrics such as PSNR, producing accurate yet often over-

Table 3. **Comparison of SwinIR [16] and StableSR [27] on Tanks & Temples [15].** Experiments are performed at $4\times$ super-resolution using ratio threshold $\tau=1.1$. Our method outperforms SRGS with either model. While SwinIR achieves higher PSNR due to its conservative reconstruction, we choose StableSR for our main experiments for its superior perceptual quality.

Method	SwinIR [16]		StableSR [27]	
	PSNR ↑	CMMD ↓	PSNR ↑	CMMD ↓
SRGS [6]	23.84	1.137	23.32	1.048
Ours	24.06	1.135	23.81	1.043

smoothed reconstructions. In contrast, StableSR incorporates diffusion-based generative priors to synthesize sharper, more detailed, and perceptually realistic images, sometimes at the expense of lower PSNR. As shown in Table 3, our approach improves performance across both models, demonstrating that it is agnostic to the choice of SISR backbone. Notably, our gains are larger with StableSR, as its higher perceptual quality comes at the cost of multi-view inconsistencies. Across both methods, StableSR produces better perceptual quality metrics, motivating our use of StableSR for the main experiments where perceptual realism is prioritized alongside reconstruction fidelity.

8. Limitations and Future Work

Our method reduces multi-view inconsistencies by suppressing super-resolution in regions already well-captured by low-resolution views. However, this conservative strategy may miss cases where SR could still offer useful refinements. For example, many SR models generate stable, visually sharp detail along high-contrast boundaries. Incorporating adaptive mechanisms that selectively apply SR in such regions could further enhance reconstruction quality. Additionally, our framework currently operates at a single upsampling level, modulating only the strength of SR supervision. Extending this to a multi-scale formulation – where smaller upsampling ratios offer more consistent detail than larger ones – may offer finer control over SR integration and lead to additional gains in sharpness.

9. Conclusion

We introduced **SplatSuRe**, a selective super-resolution framework for 3D Gaussian Splatting that applies SR only where high-frequency information is missing. By leveraging scene geometry and camera pose to estimate per-Gaussian sampling fidelity, SplatSuRe identifies undersampled regions and modulates SR supervision through per-view weight maps that highlight where generative detail is truly needed. This geometry-aware strategy injects enhanced detail and sharpness where required while preserving multi-view consistency, achieving state-of-the-art results across a diverse range of scenes.

Acknowledgments

This work was made possible by NSF Grants 21-37229 and 22-35050, DARPA TIAMAT, and the NSF TRAILS Institute (2229885). This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 140D0423C0076. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Additional support was provided by Coefficient Giving.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479, 2022. 2, 6, 7, 11, 12, 13, 16, 17
- [2] Yi-Ting Chen, Ting-Hsuan Liao, Pengsheng Guo, Alexander Schwing, and Jia-Bin Huang. Bridging diffusion models and 3d representations: A 3d consistent super-resolution framework. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13481–13490, 2025. 2
- [3] Chinthaka Dinesh, Gene Cheung, and Ivan V Bajić. 3d point cloud super-resolution via graph total variation on surface normals. In *2019 IEEE international conference on image processing (ICIP)*, pages 4390–4394. IEEE, 2019. 2
- [4] Zheng Fang, Ke Ye, Yaofang Liu, Gongzhe Li, Xianhong Zhao, Jialong Li, Ruxin Wang, Yuchen Zhang, Xiangyang Ji, and Qilin Sun. Egp3d: Edge-guided geometric preserving 3d point cloud super-resolution for rgb-d camera, 2024. 2
- [5] Xiang Feng, Yongbo He, Yubo Wang, Chengkai Wang, Zhenzhong Kuang, Jiajun Ding, Feiwei Qin, Jun Yu, and Jianping Fan. Zs-srt: An efficient zero-shot super-resolution training method for neural radiance fields. *Neurocomputing*, 590:127714, 2024. 2
- [6] Xiang Feng, Yongbo He, Yubo Wang, Yan Yang, Wen Li, Yifei Chen, Zhenzhong Kuang, Jiajun ding, Jianping Fan, and Yu Jun. Srgs: Super-resolution 3d gaussian splatting, 2024. 1, 2, 3, 6, 7, 8, 11, 12, 14, 15, 16, 17
- [7] Stephanie Fu, Netanel Tamir, Shobhita Sundaram, Lucy Chai, Richard Zhang, Tali Dekel, and Phillip Isola. Dreamsim: Learning new dimensions of human visual similarity using synthetic data. In *Advances in Neural Information Processing Systems*, pages 50742–50768, 2023. 7
- [8] Peter Hedman, Julien Philip, True Price, Jan-Michael Frahm, George Drettakis, and Gabriel Brostow. Deep blending for free-viewpoint image-based rendering. *ACM Transactions on Graphics*, 37(6):1–15, 2018. 6, 7, 11, 12, 16, 17
- [9] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017. 7
- [10] Ding-Jiun Huang, Zi-Ting Chou, Yu-Chiang Frank Wang, and Cheng Sun. Assr-nerf: Arbitrary-scale super-resolution on voxel grid for high-quality radiance fields reconstruction, 2024. 1, 2
- [11] Xudong Huang, Wei Li, Jie Hu, Hanting Chen, and Yunhe Wang. Refsr-nerf: Towards high fidelity and super resolution view synthesis. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8244–8253, 2023. 2
- [12] Sadeep Jayasumana, Srikumar Ramalingam, Andreas Veit, Daniel Glasner, Ayan Chakrabarti, and Sanjiv Kumar. Re-thinking fid: Towards a better evaluation metric for image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9307–9315, 2024. 7, 11
- [13] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5148–5157, 2021. 7
- [14] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023. 1, 2, 3, 6, 7, 12, 16, 17
- [15] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics*, 36(4):1–13, 2017. 6, 7, 8, 11, 12, 13, 14, 15, 16, 17
- [16] Jingyun Liang, Jie Zhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 8
- [17] Chenguo Lin, Panwang Pan, Bangbang Yang, Zeming Li, and Yadong Mu. Diffsplat: Repurposing image diffusion models for scalable 3d gaussian splat generation. In *International Conference on Learning Representations (ICLR)*, 2025. 2
- [18] Xi Liu, Chaoyi Zhou, and Siyu Huang. 3dgs-enhancer: Enhancing unbounded 3d gaussian splatting with view-consistent 2d diffusion priors. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. 2, 5
- [19] Yanzhe Liu, Rong Chen, Yushi Li, Yixi Li, and Xuehou Tan. Spu-pmd: Self-supervised point cloud upsampling via progressive mesh deformation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5188–5197, 2024. 2
- [20] 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. *Communications of the ACM*, 65(1):99–106, 2021. 2

- [21] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2013. [7](#)
- [22] Barbara Roessle, Norman Müller, Lorenzo Porzi, Samuel Rota Bulò, Peter Kotschieder, and Matthias Nießner. Ganerf: Leveraging discriminators to optimize neural radiance fields. *ACM Trans. Graph.*, 42(6), 2023. [2](#)
- [23] Yuan Shen, Duygu Ceylan, Paul Guerrero, Zexiang Xu, Niloy J. Mitra, Shenlong Wang, and Anna Fröhstück. Super-gaussian: Repurposing video models for 3d super resolution. In *European Conference on Computer Vision (ECCV)*, 2024. [1](#), [2](#)
- [24] Shrey Vishen, Jatin Sarabu, Saurav Kumar, Chinmay Bharathulwar, Rithwick Lakshmanan, and Vishnu Srinivas. Advancing super-resolution in neural radiance fields via variational diffusion strategies. In *Proceedings of the Winter Conference on Applications of Computer Vision*, pages 336–343, 2025. [2](#)
- [25] Yecong Wan, Mingwen Shao, Yuanshuo Cheng, and Wangmeng Zuo. S2gaussian: Sparse-view super-resolution 3d gaussian splatting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 711–721, 2025. [1](#), [2](#)
- [26] Chen Wang, Xian Wu, Yuan-Chen Guo, Song-Hai Zhang, Yu-Wing Tai, and Shi-Min Hu. Nerf-sr: High-quality neural radiance fields using supersampling. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 6445–6454, 2022. [2](#)
- [27] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin C.K. Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. 2024. [1](#), [5](#), [6](#), [7](#), [8](#), [12](#), [16](#), [17](#)
- [28] Jiaxin Wei, Stefan Leutenegger, and Simon Schaefer. Gs-fix3d: Diffusion-guided repair of novel views in gaussian splatting. 2025. [2](#)
- [29] Jay Zhangjie Wu, Yuxuan Zhang, Haithem Turki, Xuanchi Ren, Jun Gao, Mike Zheng Shou, Sanja Fidler, Zan Gojcic, and Huan Ling. Difix3d+: Improving 3d reconstructions with single-step diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 26024–26035, 2025. [2](#)
- [30] Shiyun Xie, Zhiru Wang, Xu Wang, Yinghao Zhu, Chengwei Pan, and Xiwang Dong. Supergs: Super-resolution 3d gaussian splatting enhanced by variational residual features and uncertainty-augmented learning, 2024. [1](#), [2](#)
- [31] Lequan Yu, Xianzhi Li, Chi-Wing Fu, Daniel Cohen-Or, and Pheng-Ann Heng. Pu-net: Point cloud upsampling network. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. [3](#)
- [32] Xiqian Yu, Hanxin Zhu, Tianyu He, and Zhibo Chen. Gaussiansr: 3d gaussian super-resolution with 2d diffusion priors. *arXiv preprint arXiv:2406.10111*, 2024. [1](#), [2](#)
- [33] Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-splatting: Alias-free 3d gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19447–19456, 2024. [1](#), [2](#), [6](#), [7](#), [11](#), [12](#), [14](#), [15](#), [16](#), [17](#)
- [34] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. [7](#)
- [35] Peng Zheng, Linzhi Huang, Yizhou Yu, Yi Chang, Yilin Wang, and Rui Ma. Supernerf-gan: A universal 3d-consistent super-resolution framework for efficient and enhanced 3d-aware image synthesis, 2025. [2](#)
- [36] Junsheng Zhou, Weiqi Zhang, and Yu-Shen Liu. Diffgs: Functional gaussian splatting diffusion. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. [2](#)
- [37] Matthias Zwicker, Hanspeter Pfister, Jeroen Van Baar, and Markus Gross. Ewa splatting. *IEEE Transactions on Visualization and Computer Graphics*, 8(3):223–238, 2002. [3](#)

A. Appendix

A.1. Results for 8× Super-Resolution

Table 4 reports quantitative results at 8× super-resolution for Tanks & Temples [15]. We downsample the original 1920×1080 images by $16\times$ and upsample by $8\times$ to half the native resolution. Our method achieves the strongest performance across most metrics (SSIM, PSNR, LPIPS, FID and DreamSim). For MUSIQ and NIQE, 3DGS (LR) attains higher scores, consistent with the NIQE behavior discussed in Section 6. Our CMMD score ranks fourth among the compared methods. Prior work [12] shows that CMMD is sensitive to high-frequency distortions introduced by noise in the latent embedding space. We hypothesize that the sharpening effects like mild aliasing or overly crisp edges introduced at large upsampling factors may be interpreted by CMMD as distortions, even though they improve perceptual quality and are reflected positively by other metrics. Since no single metric fully characterizes visual fidelity, we report a broad suite of metrics to provide a more comprehensive assessment of our method.

Table 5 presents the 8× SR results for Deep Blending [8] and Mip-NeRF 360 [1]. For Deep Blending, we downsample the original images by $8\times$ ($\sim 125 \times 125$) and upsample by $8\times$ back to the native resolution ($\sim 1K \times 1K$). For Mip-NeRF 360, we downsample the original images by $16\times$ ($\sim 250 \times 188$) and upsample by $8\times$ to half the native resolution ($\sim 2K \times 1.5K$). SplatSuRe achieves the strongest performance across all metrics on Deep Blending. On Mip-NeRF 360, our method also achieves the strongest results across nearly all metrics; we are slightly outperformed by Mip-Splatting [33] on SSIM, while still outperforming SRGS [6]. In these settings, the LR images – including those in Mip-NeRF 360 – contained limited high-frequency information due to their smaller resolutions, requiring super-resolution to hallucinate substantially more details. These results further demonstrate that our selective SR method produces higher image quality than uniform application, even in settings with highly-generative view-inconsistent SR.

A.2. Additional Ratio Threshold Analysis

Different scenes exhibit distinct behaviors as the ratio threshold and the corresponding amount of super-resolution (SR) information are increased. Most scenes benefit from a moderate amount of SR, but experience a sharp drop in image quality when it is excessively applied. Others plateau in image quality or continue improving, although with diminishing returns. We visualize these trends by plotting PSNR and LPIPS across ratio thresholds for three representative scenes from each category.

Figure 7 illustrates scenes that benefit from an optimal amount of SR. These scenes exhibit a characteristic pattern

in which image quality initially improves with moderate SR, but decreases sharply when it is excessively applied. Our method identifies the most poorly sampled regions and selectively applies SR to them, yielding a substantial initial quality boost. However, applying excessive SR introduces multi-view inconsistencies that rapidly degrades image quality. This behavior appears consistently across most scenes, supporting our hypothesis that selectively applying SR is more beneficial than applying it uniformly.

Figure 8 presents scenes that plateau in image quality or continue to benefit slightly as the amount of SR is increased. Our method identifies the most poorly sampled regions and selectively applies SR to them, producing the sharp early quality gains at lower ratio thresholds. Beyond this point, increasing the amount of SR yields diminishing returns or no improvement. This is particularly observed in scenes where the input images already contain substantial high-frequency detail. In such cases, SR produces simpler sharpening or edge-enhancement effects rather than hallucinating new structure, making uniform application less harmful and sometimes marginally beneficial. This behavior is especially common in outdoor scenes in Mip-NeRF 360 [1], where the downsampled $\sim 500 \times 375$ images remain relatively high-resolution and therefore do not exhibit the multi-view inconsistencies that typically arise when excessive SR is applied.

A.3. Additional Visualizations

Figures 9 and 10 present additional qualitative results on Tanks & Temples [15]. Consistent with the examples in Figures 1 and 5 discussed in Section 6, SplatSuRe produces sharper reconstructions than competing methods while reducing artifacts and preserving smoothness in uniformly textured regions.

A.4. Per-Scene Metrics

Tables 6, 7, 8, 9, 10, 11, 12, and 13 report per-scene SSIM, PSNR, LPIPS, FID, CMMD, DreamSim, MUSIQ, and NIQE results on Tanks & Temples [15], Deep Blending [8], and Mip-NeRF 360 [1] for all methods evaluated in Section 5. Across individual scenes, we observe the same trends as in the averaged results reported in Tables 1 and 2 in Section 6: SplatSuRe achieves the strongest performance on Tanks & Temples and Deep Blending, and outperforms SRGS on Mip-NeRF 360.

Table 4. **Quantitative results on Tanks & Temples [15] at 8× super-resolution.** Experiments are performed using ratio threshold $\tau=1.1$. The best, second best and third best entries are highlighted. Our SplatSuRe method achieves the strongest results on most metrics.

Method	Tanks & Temples [15]							
	SSIM ↑	PSNR ↑	LPIPS ↓	FID ↓	CMMD ↓	DreamSim ↓	MUSIQ ↑	NIQE ↓
3DGS (LR) [14]	0.537	16.40	0.473	172.44	3.795	0.2427	46.870	3.976
StableSR [27]	0.664	21.49	0.406	99.80	1.878	0.1100	38.929	6.293
Mip-Splatting [33]	0.661	21.53	0.428	109.01	1.941	0.1183	32.417	6.645
SRGS [6] + StableSR [27]	0.674	21.97	0.399	92.91	1.891	0.0981	38.322	6.138
Ours + StableSR [27]	0.692	22.48	0.378	77.17	2.007	0.0774	44.428	5.341

Table 5. **Quantitative results on Deep Blending [8] and Mip-NeRF 360 [1] at 8× super-resolution.** Experiments are performed using ratio threshold $\tau=1.1$. Our SplatSuRe method achieves the strongest results on almost all metrics.

Method	Deep Blending [8]					Mip-NeRF 360 [1]				
	SSIM ↑	PSNR ↑	LPIPS ↓	CMMD ↓	DreamSim ↓	SSIM ↑	PSNR ↑	LPIPS ↓	CMMD ↓	DreamSim ↓
3DGS (LR) [14]	0.783	24.50	0.404	2.051	0.1280	0.509	18.17	0.491	2.395	0.1337
StableSR [27]	0.821	26.68	0.383	1.159	0.0705	0.630	24.23	0.445	1.096	0.0546
Mip-Splatting [33]	0.831	27.24	0.391	1.127	0.0689	0.656	24.80	0.420	0.610	0.0243
SRGS [6] + StableSR [27]	0.830	27.35	0.377	1.079	0.0622	0.648	24.88	0.422	0.665	0.0320
Ours + StableSR [27]	0.843	28.03	0.357	0.879	0.0489	0.655	25.08	0.406	0.546	0.0231

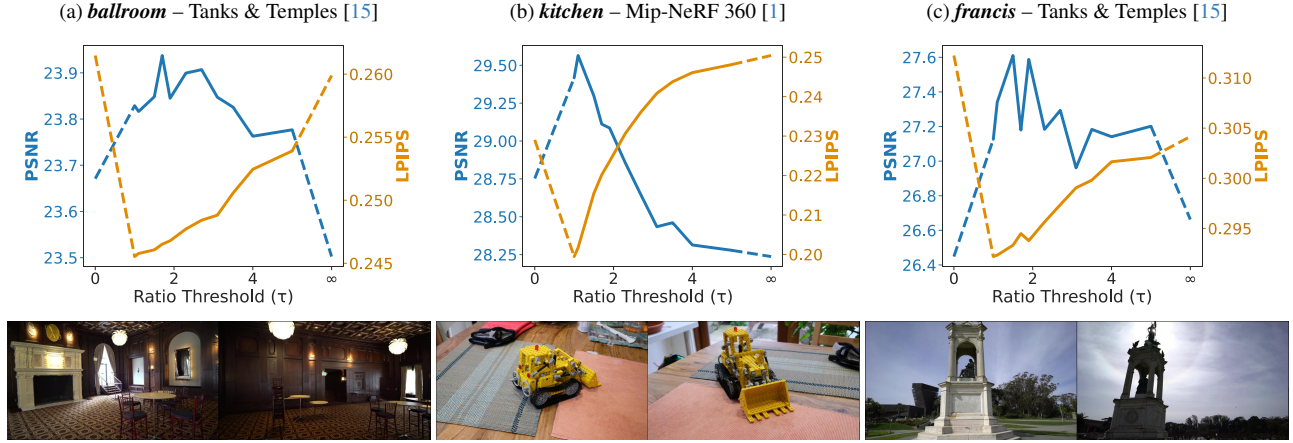


Figure 7. **Representative scenes that benefit from an optimal amount of super-resolution.** Top: Image quality vs. ratio threshold plots. Bottom: ground truth images illustrating scene structure for (a) *ballroom* from Tanks & Temples [15], (b) *kitchen* from Mip-NeRF 360 [1] and (c) *francis* from Tanks & Temples. Applying SR to the most poorly sampled regions yields large gains in image quality, whereas excessive SR introduces multi-view inconsistencies that sharply degrade quality. Most scenes exhibit this behavior, supporting our hypothesis that selectively applying SR is more beneficial than applying it uniformly.

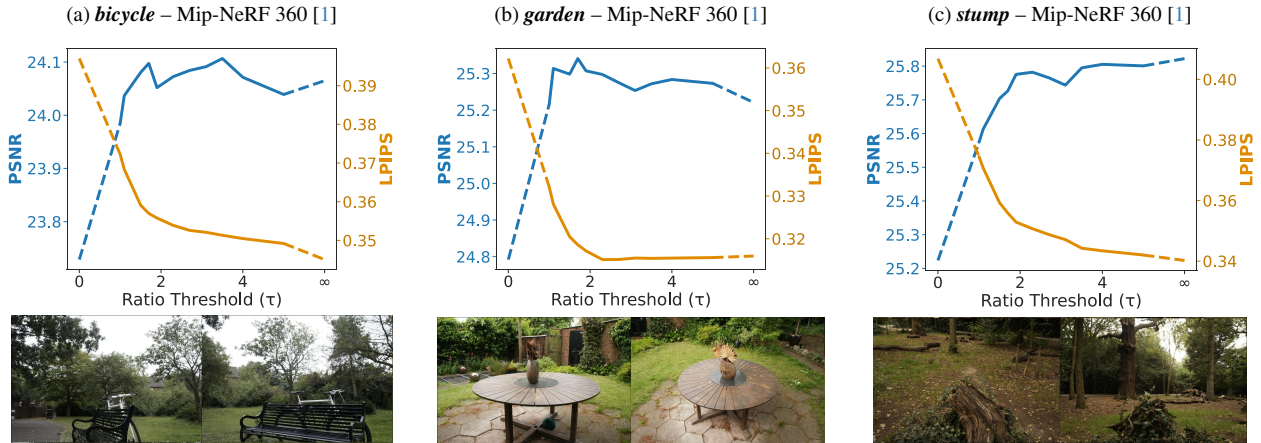


Figure 8. **Representative scenes that plateau in image quality or continue to benefit from increased amounts of super-resolution.** Top: Image quality vs. ratio threshold plots. Bottom: ground truth images illustrating scene structure for (a) *bicycle*, (b) *garden*, and (c) *stump* from Mip-NeRF 360 [1]. Applying SR to the most poorly sampled regions yields large gains in image quality, while further increasing SR provides diminishing returns or no improvement. This is particularly observed in scenes where the input images already contain substantial high-frequency detail. In such cases, SR produces simpler sharpening or edge-enhancement effects rather than hallucinating new structure, making uniform application less harmful and sometimes marginally beneficial.

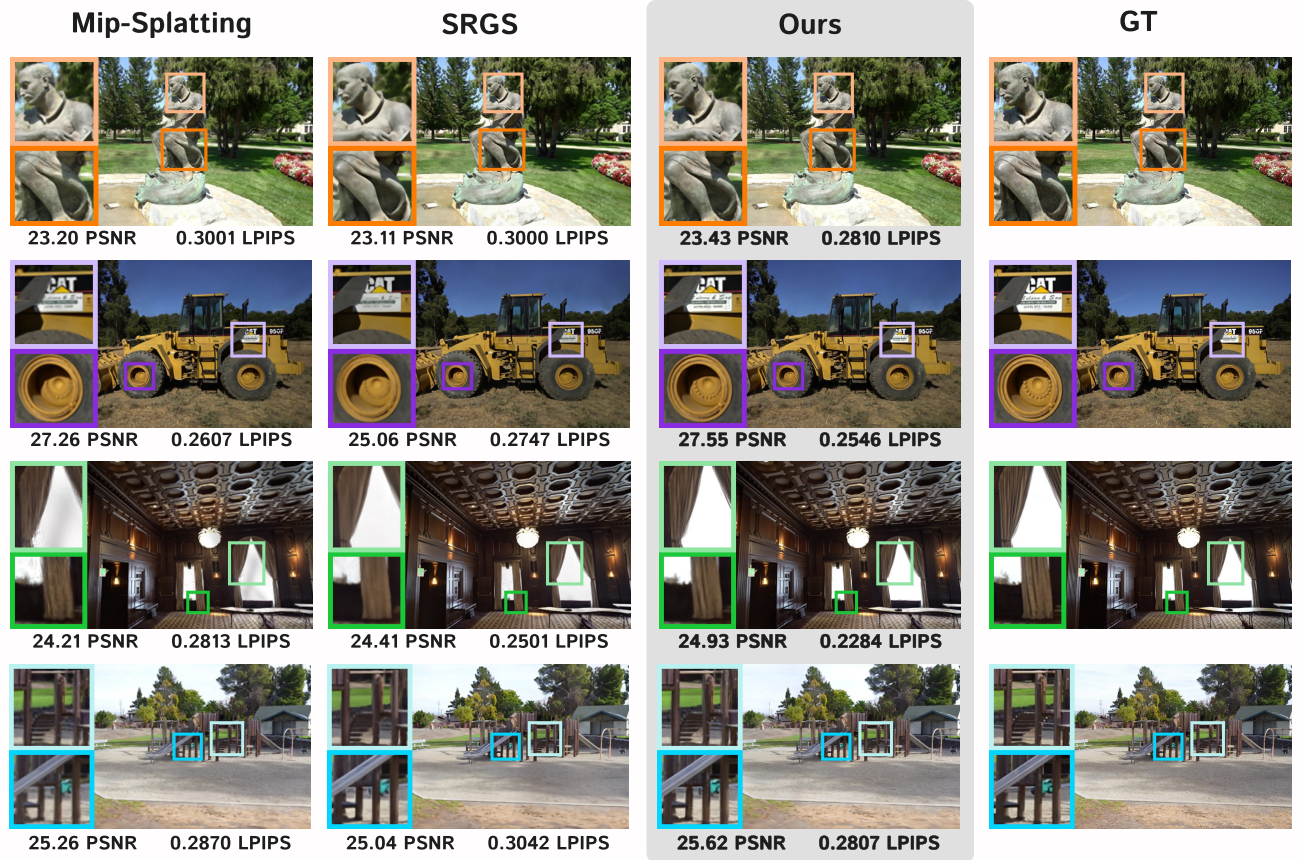


Figure 9. **Additional qualitative results on Tanks & Temples [15]**. Experiments are performed at $4\times$ super-resolution with ratio threshold $\tau=1.1$. Compared to Mip-Splatting [33] and SRGS [6], our method produces sharper, more faithful reconstructions that better align with ground truth while maintaining cross-view consistency. It preserves high-frequency patterns (orange boxes on statue), fine details in text (purple box on vehicle), and reduces artifacts while preserving sharpness (green boxes on curtains, blue boxes in playground).



Figure 10. **Additional qualitative results on Tanks & Temples [15].** Experiments are performed at $4\times$ super-resolution with ratio threshold $\tau=1.1$. Compared to Mip-Splatting [33] and SRGS [6], our method produces sharper, more faithful reconstructions that better align with ground truth while maintaining cross-view consistency. It preserves high-frequency geometry (pink boxes on barn) and retains sharp texture (yellow boxes on lighthouse, red boxes on horse statue, blue boxes on tank).

Table 6. SSIM $\uparrow$ on each scene in Tanks & Temples [15], Deep Blending [8], and Mip-NeRF 360 [1]. Experiments are performed at $4\times$ super-resolution using ratio threshold $\tau=1.1$. The best, second best and third best entries are highlighted.

Method	Tanks & Temples [15]																
	auditorium	ignatius	palace	ballroom	panther	barn	lighthouse	playground	courtroom	m60	temple	caterpillar	family	train	francis	truck	church
3DGS (LR) [14]	0.807	0.527	0.666	0.578	0.752	0.729	0.742	0.628	0.647	0.735	0.715	0.602	0.566	0.648	0.779	0.649	0.675
StableSR [27]	0.849	0.646	0.709	0.702	0.819	0.771	0.777	0.720	0.725	0.810	0.761	0.686	0.753	0.705	0.845	0.759	0.744
Mip-Splatting [33]	0.838	0.681	0.687	0.729	0.847	0.790	0.789	0.746	0.730	0.836	0.762	0.704	0.788	0.733	0.855	0.786	0.753
SRGS [6] + StableSR [27]	0.861	0.671	0.713	0.734	0.846	0.794	0.789	0.742	0.746	0.835	0.773	0.703	0.780	0.729	0.857	0.784	0.761
Ours + StableSR [27]	0.859	0.694	0.717	0.746	0.859	0.808	0.807	0.764	0.752	0.846	0.786	0.714	0.795	0.743	0.866	0.798	0.781

Method	Deep Blending [8]				Mip-NeRF 360 [1]							
	drjohnson	playroom	bicycle	bonsai	counter	flowers	garden	kitchen	room	stump	treehill	
3DGS (LR) [14]	0.828	0.843	0.527	0.776	0.772	0.429	0.559	0.719	0.836	0.581	0.516	
StableSR [27]	0.837	0.854	0.604	0.848	0.833	0.497	0.654	0.751	0.858	0.669	0.571	
Mip-Splatting [33]	0.855	0.874	0.676	0.917	0.872	0.544	0.719	0.870	0.897	0.740	0.596	
SRGS [6] + StableSR [27]	0.851	0.870	0.653	0.896	0.863	0.533	0.692	0.825	0.888	0.671	0.590	
Ours + StableSR [27]	0.867	0.877	0.646	0.906	0.867	0.503	0.690	0.869	0.896	0.706	0.582	

Table 7. PSNR $\uparrow$ on each scene in Tanks & Temples [15], Deep Blending [8], and Mip-NeRF 360 [1].

Method	Tanks & Temples [15]																
	auditorium	ignatius	palace	ballroom	panther	barn	lighthouse	playground	courtroom	m60	temple	caterpillar	family	train	francis	truck	church
3DGS (LR) [14]	22.14	17.13	18.35	18.23	22.49	22.97	20.47	19.85	19.53	21.72	18.99	19.36	15.09	18.80	22.47	17.38	20.00
StableSR [27]	23.62	20.52	19.68	22.20	25.23	25.18	21.65	23.24	21.75	24.72	20.24	22.44	22.54	20.55	25.97	22.86	22.02
Mip-Splatting [33]	23.46	21.22	18.24	23.15	27.11	26.85	21.29	24.13	22.26	26.16	20.31	22.99	24.09	21.33	26.62	23.94	21.99
SRGS [6] + StableSR [27]	24.23	21.26	19.69	23.44	26.98	26.83	21.74	24.12	22.74	26.17	20.49	22.98	23.82	21.28	26.55	23.87	22.54
Ours + StableSR [27]	24.66	21.61	20.24	23.82	27.38	27.33	22.82	25.12	22.99	26.63	21.39	23.34	24.00	21.79	27.34	24.11	23.18

Method	Deep Blending [8]				Mip-NeRF 360 [1]							
	drjohnson	playroom	bicycle	bonsai	counter	flowers	garden	kitchen	room	stump	treehill	
3DGS (LR) [14]	26.29	27.15	18.50	22.75	23.27	17.60	18.64	19.55	25.93	20.34	19.44	
StableSR [27]	26.60	27.71	22.87	26.48	26.44	19.92	23.80	24.80	28.32	24.24	21.64	
Mip-Splatting [33]	27.54	29.33	24.28	30.65	28.28	21.05	25.63	29.48	30.69	26.20	22.08	
SRGS [6] + StableSR [27]	27.38	29.08	24.05	29.81	28.11	20.91	25.20	28.26	30.42	24.58	21.97	
Ours + StableSR [27]	28.46	29.57	24.04	30.54	28.39	20.55	25.31	29.56	31.05	25.61	22.02	

Table 8. LPIPS $\downarrow$ on each scene in Tanks & Temples [15], Deep Blending [8], and Mip-NeRF 360 [1].

Method	Tanks & Temples [15]																
	auditorium	ignatius	palace	ballroom	panther	barn	lighthouse	playground	courtroom	m60	temple	caterpillar	family	train	francis	truck	church
3DGS (LR) [14]	0.301	0.411	0.380	0.337	0.309	0.317	0.303	0.371	0.376	0.302	0.335	0.385	0.408	0.346	0.352	0.345	0.343
StableSR [27]	0.247	0.357	0.364	0.283	0.275	0.295	0.283	0.333	0.294	0.266	0.302	0.340	0.289	0.305	0.316	0.298	0.306
Mip-Splatting [33]	0.275	0.331	0.408	0.288	0.260	0.293	0.296	0.316	0.323	0.250	0.325	0.325	0.281	0.298	0.316	0.284	0.313
SRGS [6] + StableSR [27]	0.242	0.328	0.367	0.266	0.251	0.277	0.276	0.311	0.292	0.242	0.299	0.321	0.271	0.289	0.304	0.274	0.294
Ours + StableSR [27]	0.248	0.310	0.355	0.246	0.228	0.262	0.259	0.284	0.285	0.225	0.290	0.313	0.257	0.281	0.292	0.253	0.272

Method	Deep Blending [8]				Mip-NeRF 360 [1]							
	drjohnson	playroom	bicycle	bonsai	counter	flowers	garden	kitchen	room	stump	treehill	
3DGS (LR) [14]	0.332	0.339	0.424	0.328	0.306	0.477	0.400	0.312	0.293	0.436	0.482	
StableSR [27]	0.315	0.335	0.407	0.256	0.258	0.461	0.356	0.301	0.267	0.394	0.460	
Mip-Splatting [33]	0.334	0.321	0.318	0.212	0.242	0.397	0.292	0.205	0.245	0.314	0.403	
SRGS [6] + StableSR [27]	0.311	0.323	0.349	0.232	0.239	0.420	0.315	0.249	0.250	0.368	0.428	
Ours + StableSR [27]	0.300	0.312	0.368	0.238	0.241	0.454	0.328	0.202	0.253	0.371	0.452	

Table 9. FID $\downarrow$ on each scene in Tanks & Temples [15], Deep Blending [8], and Mip-NeRF 360 [1].

Method	Tanks & Temples [15]																
	auditorium	ignatius	palace	ballroom	panther	barn	lighthouse	playground	courtroom	m60	temple	caterpillar	family	train	francis	truck	church
3DGS (LR) [14]	86.75	85.58	83.25	73.89	40.93	40.48	84.46	82.51	99.19	53.16	81.79	29.93	113.45	43.80	29.64	35.69	58.24
StableSR [27]	68.24	114.03	83.65	49.37	38.50	48.11	72.12	66.81	64.95	40.68	78.22	25.98	68.88	35.61	31.14	32.99	54.69
Mip-Splatting [33]	78.68	52.74	138.29	40.43	24.78	36.49	67.71	59.20	69.94	25.38	82.25	20.50	38.03	30.99	20.36	20.20	48.50
SRGS [6] + StableSR [27]	58.94	79.04	97.00	37.35	26.97	36.74	66.75	54.86	59.58	27.03	75.46	20.10	43.93	29.88	23.64	22.37	45.17
Ours + StableSR [27]	60.48	46.33	70.53	27.26	20.96	27.20	57.93	35.69	54.34	19.79	63.87	16.42	28.28	22.83	16.72	14.82	32.61

Method	Deep Blending [8]				Mip-NeRF 360 [1]							
	drjohnson	playroom	bicycle	bonsai	counter	flowers	garden	kitchen	room	stump	treehill	
3DGS (LR) [14]	60.62	72.04	30.50	72.93	79.04	77.01	47.98	59.89	41.92	89.83	42.08	
StableSR [27]	51.64	67.41	23.55	54.42	64.89	73.04	19.38	49.64	37.09	105.27	40.16	
Mip-Splatting [33]	53.90	50.26	10.21	20.19	36.29	22.76	9.05	12.04	17.57	37.03	20.21	
SRGS [6] + StableSR [27]	47.05	51.10	11.37	26.41	41.87	46.01	10.46	15.43	21.62	42.73	26.52	
Ours + StableSR [27]	43.19	45.09	11.79	22.40	38.03	58.56	10.63	8.47	17.69	50.63	29.83	

Table 10. CMMD ↓ on each scene in Tanks & Temples [15], Deep Blending [8], and Mip-NeRF 360 [1].

Method	Tanks & Temples [15]																
	auditorium	ignatius	palace	ballroom	panther	barn	lighthouse	playground	courtroom	m60	temple	caterpillar	family	train	francis	truck	church
3DGS (LR) [14]	1.770	2.234	2.361	1.602	0.718	1.699	1.540	1.925	1.616	1.202	2.522	1.897	3.038	2.153	2.758	2.140	1.705
StableSR [27]	1.093	1.032	1.491	1.073	0.791	1.068	1.092	1.139	0.987	0.907	1.597	1.464	0.647	1.330	1.401	1.143	1.163
Mip-Splatting [33]	1.246	1.046	1.624	0.787	0.633	1.140	0.994	1.151	1.024	0.579	1.853	1.243	0.863	1.384	1.286	1.080	1.029
SRGS [6] + StableSR [27]	1.066	0.931	1.582	0.787	0.516	0.942	0.978	0.998	0.943	0.523	1.795	1.261	0.557	1.383	1.408	1.140	1.049
Ours + StableSR [27]	0.993	1.121	1.568	0.643	0.351	1.134	0.901	1.025	0.951	0.443	1.699	1.330	0.521	1.529	1.355	1.061	0.944

Method	Deep Blending [8]				Mip-NeRF 360 [1]						
	drjohnson	playroom	bicycle	bonsai	counter	flowers	garden	kitchen	room	stump	treehill
3DGS (LR) [14]	1.003	0.733	0.392	1.011	0.319	1.354	0.242	0.346	0.261	0.926	0.956
StableSR [27]	0.807	0.714	0.621	0.934	0.440	1.409	0.310	0.441	0.578	0.767	1.023
Mip-Splatting [33]	0.828	0.552	0.099	0.346	0.195	0.215	0.036	0.191	0.187	0.125	0.257
SRGS [6] + StableSR [27]	0.715	0.546	0.138	0.417	0.279	0.782	0.081	0.228	0.378	0.261	0.486
Ours + StableSR [27]	0.558	0.434	0.159	0.266	0.139	1.108	0.065	0.119	0.184	0.406	0.610

Table 11. DreamSim ↓ on each scene in Tanks & Temples [15], Deep Blending [8], and Mip-NeRF 360 [1].

Method	Tanks & Temples [15]																
	auditorium	ignatius	palace	ballroom	panther	barn	lighthouse	playground	courtroom	m60	temple	caterpillar	family	train	francis	truck	church
3DGS (LR) [14]	0.0969	0.0884	0.1548	0.0881	0.0399	0.0549	0.1154	0.0806	0.1074	0.0552	0.1156	0.0595	0.1215	0.0747	0.0710	0.0620	0.0795
StableSR [27]	0.0634	0.0703	0.1497	0.0438	0.0443	0.0613	0.0909	0.0759	0.0566	0.0591	0.0955	0.0516	0.0565	0.0652	0.0555	0.0590	0.0558
Mip-Splatting [33]	0.0739	0.0465	0.2459	0.0312	0.0216	0.0392	0.0887	0.0541	0.0592	0.0266	0.1001	0.0392	0.0278	0.0479	0.0341	0.0293	0.0541
SRGS [6] + StableSR [27]	0.0500	0.0511	0.1603	0.0271	0.0264	0.0399	0.0792	0.0564	0.0466	0.0369	0.0936	0.0416	0.0359	0.0504	0.0402	0.0402	0.0459
Ours + StableSR [27]	0.0490	0.0381	0.1272	0.0202	0.0187	0.0314	0.0625	0.0369	0.0422	0.0239	0.0720	0.0315	0.0229	0.0410	0.0309	0.0268	0.0331

Method	Deep Blending [8]				Mip-NeRF 360 [1]						
	drjohnson	playroom	bicycle	bonsai	counter	flowers	garden	kitchen	room	stump	treehill
3DGS (LR) [14]	0.0613	0.0534	0.0624	0.0659	0.0453	0.0562	0.0452	0.0454	0.0322	0.0586	0.0558
StableSR [27]	0.0499	0.0510	0.0348	0.0320	0.0377	0.0598	0.0170	0.0240	0.0345	0.0356	0.0533
Mip-Splatting [33]	0.0451	0.0345	0.0126	0.0122	0.0166	0.0124	0.0070	0.0046	0.0163	0.0117	0.0252
SRGS [6] + StableSR [27]	0.0439	0.0378	0.0150	0.0160	0.0226	0.0331	0.0079	0.0066	0.0212	0.0178	0.0345
Ours + StableSR [27]	0.0332	0.0327	0.0151	0.0123	0.0177	0.0380	0.0071	0.0037	0.0144	0.0168	0.0364

Table 12. MUSIQ ↑ on each scene in Tanks & Temples [15], Deep Blending [8], and Mip-NeRF 360 [1].

Method	Tanks & Temples [15]																
	auditorium	ignatius	palace	ballroom	panther	barn	lighthouse	playground	courtroom	m60	temple	caterpillar	family	train	francis	truck	church
3DGS (LR) [14]	55.443	64.965	51.161	56.309	54.610	60.048	48.785	54.510	55.349	53.631	57.565	59.116	65.987	60.573	59.605	61.675	57.991
StableSR [27]	58.573	58.466	49.326	56.501	57.051	60.822	46.028	53.817	55.209	54.985	54.748	60.520	62.883	60.841	55.264	60.637	55.297
Mip-Splatting [33]	51.773	48.745	42.653	50.651	46.800	49.040	33.568	41.815	45.084	47.621	46.236	51.797	49.461	51.300	40.700	47.208	47.433
SRGS [6] + StableSR [27]	57.276	57.690	48.339	54.321	54.969	58.305	44.228	52.796	54.435	53.662	52.762	58.732	61.687	58.721	54.220	57.781	54.378
Ours + StableSR [27]	58.504	60.078	50.762	59.501	56.915	61.660	47.882	58.036	57.351	55.500	55.978	60.757	65.967	60.214	57.281	62.616	57.574

Method	Deep Blending [8]				Mip-NeRF 360 [1]						
	drjohnson	playroom	bicycle	bonsai	counter	flowers	garden	kitchen	room	stump	treehill
3DGS (LR) [14]	47.916	49.425	54.239	56.108	55.829	58.968	58.961	63.552	50.931	52.085	41.783
StableSR [27]	52.919	53.324	55.767	59.725	57.965	52.194	54.576	59.952	58.135	44.337	38.276
Mip-Splatting [33]	38.800	44.105	38.243	44.656	42.445	35.728	44.342	50.600	41.878	35.866	35.112
SRGS [6] + StableSR [27]	50.849	52.035	54.818	56.237	55.391	53.542	54.318	58.321	55.637	44.039	37.260
Ours + StableSR [27]	50.594	55.444	57.923	58.501	57.752	51.206	56.236	64.952	55.154	49.551	38.014

Table 13. NIQE ↓ on each scene in Tanks & Temples [15], Deep Blending [8], and Mip-NeRF 360 [1].

Method	Tanks & Temples [15]																
	auditorium	ignatius	palace	ballroom	panther	barn	lighthouse	playground	courtroom	m60	temple	caterpillar	family	train	francis	truck	church
3DGS (LR) [14]	4.718	2.628	4.293	2.791	3.884	3.944	4.097	2.744	3.385	3.625	4.081	2.873	2.605	2.994	4.269	2.784	3.169
StableSR [27]	6.531	4.545	5.820	4.335	5.239	5.229	5.415	4.991	4.830	5.022	4.844	4.690	4.214	4.363	5.611	4.361	5.005
Mip-Splatting [33]	5.482	4.693	6.567	4.013	5.614	5.302	5.482	5.086	4.569	4.955	5.039	4.632	5.077	4.449	5.982	5.022	4.653
SRGS [6] + StableSR [27]	6.128	4.187	5.629	3.978	4.904	4.988	5.025	4.537	4.562	4.716	4.573	4.323	4.026	4.137	5.234	4.096	4.644
Ours + StableSR [27]	5.330	3.440	4.720	3.318	3.997	4.657	4.309	3.644	4.029	3.765	4.167	3.340	3.435	3.498	4.559	3.416	3.785

Method	Deep Blending [8]				Mip-NeRF 360 [1]						
	drjohnson	playroom	bicycle	bonsai	counter	flowers	garden	kitchen	room	stump	treehill
3DGS (LR) [14]	4.760	5.329	2.808	3.964	3.535	2.915	2.741	2.986	4.210	3.137	2.744
StableSR [27]	5.231	6.225	4.333	5.452	5.533	4.938	5.516	5.633	6.013	4.919	5.105
Mip-Splatting [33]	6.097	6.426	5.665	6.458	5.730	5.965	6.155	5.600	5.885	5.623	5.462
SRGS [6] + StableSR [27]	5.208	6.188	4.131	5.228	5.214	4.521	5.074	5.546	5.840	4.846	4.381
Ours + StableSR [27]	5.054	5.768	3.378	4.309	3.828	3.825	3.949	3.828	4.531	4.052	3.702