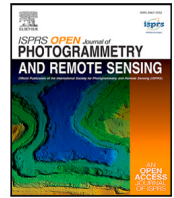




Contents lists available at ScienceDirect

ISPRS Open Journal of Photogrammetry and Remote Sensing

journal homepage: www.journals.elsevier.com/isprs-open-journal-of-photogrammetry-and-remote-sensing

SGS: Shadow-Aware Gaussian Splatting

Ivana Petrovska^{a,*}, Daniel Rebain^b, Lukas Winiwarter^c, Boris Jutzi^a^a Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT), Karlsruhe, 76131, Germany^b University of British Columbia, Vancouver, V6T 1Z4, Canada^c Unit of Geometry and Surveying, Faculty of Engineering Sciences, University of Innsbruck, Innsbruck, 6020, Austria

ARTICLE INFO

Keywords:

Shadow-Aware Gaussian Splatting
3D reconstruction
Novel view synthesis
Shadows
Vegetation
Occlusions

ABSTRACT

Gaussian Splatting algorithms achieve high fidelity under controlled, static conditions with unobstructed viewpoints. However, in unstructured environments hampered by heterogeneous occlusions with sub-optimal lighting conditions or completely dark surroundings, the movement of a co-located camera and light source causes shadow variance as the relative position of objects within the scene produce intricate cast shadows causing the same coordinate of the scene to exhibit different pixel color. This dynamic shifting of shadows leads to transient pixel intensities across the multi-view images, directly violating the photo-consistency assumption inducing the uncertainty for each Gaussian to exhibit varying colors. To address this, we propose SGS: Shadow-Aware Gaussian Splatting, a novel shadow-aware method which explicitly separates intrinsic radiance from extrinsic illumination. Instead of storing separate color sets, each Gaussian is augmented with a single learnable scalar that modulates its intrinsic color features. Furthermore, we introduce a dual-pass rendering pipeline complemented with a bimodal selection loss for an effective disentanglement of shadows from scene geometry. Extensive experiments demonstrate that SGS significantly eliminates volumetric artifacts, while simultaneously increasing overall point coverage complemented with enhanced photometric reconstruction outperforming in all metrics, striking a favorable balance between geometric accuracy, point efficiency and rendering quality.

1. Introduction

Complete and accurate 3D reconstruction from multi-view images is a long standing challenge in 3D Computer Vision and photogrammetry. Differing from traditional Multi-View Stereo (MVS) whose time-consuming cross-view triangulation assumes Lambertian surfaces, radiance-based methods can better capture scene realism and provide a more detailed reconstruction, while being able to approximate different illumination variations. Optimizing an implicit radiance field, Neural Radiance Fields (NeRFs) (Mildenhall et al., 2021) represent a volumetric shape in terms of a spatial density function and view-dependent radiance. Whenever the neural network predicts non-zero density, the gradients encourage updates to the predicted color which is computed on-the-fly by querying the neural network. The volume density is sampled along camera rays and the samples are aggregated to predict the pixel value. However, NeRFs exhibit large learning cost and slow rendering speed due to the reliance on neural networks without explicitly defining the geometry, hindering accurate geometric reconstruction (D. Chen et al., 2023; Jäger and Jutzi, 2023; Wan et al., 2023; Petrovska et al., 2023).

In contrast, an explicit radiance field directly represents the distribution of light in a discrete spatial structure. Gaussian Splatting (GS)

methods, conceived with 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) reconstruct scene geometry as a series of Gaussian ellipsoids initialized on a sparse point cloud. Each Gaussian is associated with 3D position, anisotropic covariance, opacity and spherical harmonic (SH) coefficients describing the color. The parameters are iteratively adjusted through gradient descent until the photometric error between rendered and training images is minimized. To further improve efficiency and quality, the optimization process dynamically adjusts the number and distribution of Gaussians by performing density-based splitting and pruning operations, ensuring a compact and accurate scene representation. During optimization, the parameters are refined through a differentiable rasterization pipeline. The process begins by sorting the Gaussians along each camera ray based on their mean position for subsequent rendering. Each 3D Gaussian is then projected onto the image plane as a 2D Gaussian splat, with its anisotropic properties preserved under perspective projection. However, sorting Gaussians based on their center does not guarantee proper depth ordering and it may result in popping artifacts when rendered from different viewpoints. Moreover, blending the depth of Gaussian centers ignores the shape and orientation of the Gaussian primitives and the position of a 3D point is ill-defined when covered by multiple Gaussians. On top

* Corresponding author.

E-mail address: ivana.petrovska@partner.kit.edu (I. Petrovska).

<https://doi.org/10.1016/j.ophoto.2026.100151>

Received 15 May 2026; Received in revised form 20 August 2026; Accepted 26 August 2026

Available online 21 September 2026

2667-3932/© 2026 The Authors. Published by Elsevier B.V. on behalf of International Society of Photogrammetry and Remote Sensing. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>). Open access funding enabled and organized by Projekt DEAL.

of that, guided solely on photometric loss which prioritizes novel view synthesis over geometric accuracy, semi-transparent floaters in space can be composited to produce the original image without adhering to a single coherent surface. Subsequent research address these challenges focusing on multi-view consistent geometry (Guédon and Lepetit, 2024; Huang et al., 2024; H. Chen et al., 2023; Li et al., 2024; Dai et al., 2024; D. Chen et al., 2024; Song et al., 2024).

While GS algorithms achieve high fidelity under controlled, static conditions with unobstructed viewpoints, such a setup is rarely effective in unstructured environments hampered by heterogeneous occlusions (Petrovska and Jutzi, 2025). These challenges are further compounded in sub-optimal lighting conditions or completely dark surroundings, where the introduction of active light sources is compulsory. Considering that the occlusions are always in front of the target, the movement of a co-located camera and light source induces shadow variance as the relative position of objects within the scene produce wide variety of intricate cast shadows. This causes the same coordinate of the scene to exhibit different pixel color appearing not in shadow in one image and in shadow in the other. The dynamic shifting of shadows leads to transient pixel intensities, directly violating the photo-consistency assumption as the algorithm struggles to match identical points across different images. Since the SH only account for change in the viewing direction under constant light, illumination variations cause the uncertainty for each Gaussian to exhibit varying colors. This prevents accurate representation and decomposition of diffuse and specular components, restricting complex lighting interaction representations (Liu et al., 2025). Adding more SH levels leads to polynomial increase in the parameter count, expanding computational complexity and limiting practical use in real-time applications (Wang et al., 2026).

To explicitly separate intrinsic radiance from extrinsic illumination, we propose Shadow-Aware Gaussian Splatting (SGS). The core of the method is unbaking shadows from scene geometry. Instead of duplicating color parameters, each Gaussian primitive is augmented with a learnable scalar that modulates the basic SH coefficients, enabling the representation of the scene under both full illumination (*Basic*) and in shadow (*Shadow*). This formulation facilitates a novel dual-pass rendering pipeline. During optimization, the renderer produces two competing representations for every viewpoint in *Basic* and *Shadow*. We resolve the ambiguity through a bimodal selection loss, which dynamically generates a pixel-wise mask based on the lowest residual error relative to the ground truth for gradient backpropagation. When a pixel is in shadow in the training data, the loss directs the gradient flow specifically towards *Shadow*, effectively extracting shadows from scene geometry. By preventing shadows from being erroneously baked into the scene color, SGS eliminates the volumetric artifacts and geometric blurring yielding to superior geometric reconstruction and rendering fidelity. Our investigations are based on cast shadows from occlusions on synthetic data to have better control over the occlusion type, coverage and placement, but more importantly the shadows they cast for a comprehensive evaluation and analysis. We focus particularly on vegetation occlusions, as the irregular and complex geometry introduces noise and hinders accurate depth estimation. To investigate the impact of shadows on the geometric reconstruction, we generate images from the same position once under diffuse light (without shadows) and once under direct light (with shadows).

In summary, we propose the following contributions:

- Shadow-aware scene representation by scaling the Spherical Harmonics to separate basic color from illumination-dependent shadows without explicit supervision.
- A dual-pass rendering that effectively disentangles appearance and illumination.
- A bimodal selection loss, which routes gradients according to the minimal photometric residual at each pixel, effectively unbaking shadows from scene geometry, leading to strong geometric and photometric consistency.

- Improved geometric accuracy by significantly reducing floaters, while simultaneously increasing overall completeness and enhanced photometric quality outperforming in all metrics.
- New dataset for the task of 3D reconstruction behind multi-view consistent occlusions under diffuse and direct light where the occlusions cast shadows.

The contribution is organized as follows. In Section 2 an overview of planar-based GS methods, tackling geometric reconstruction behind occlusions and under varying lighting conditions along with current publicly available datasets is given, in Section 3 the fundamentals of our proposed method are introduced, information about the dataset, evaluation metrics along with implementation details is provided in Section 4, in Section 5 the geometric and photometric qualitative and quantitative results are presented, the discussion is provided in Sections 6 and 7 concludes this contribution.

2. Related work

In Section 2.1 we present an overview of GS methods aiming to transform 3D Gaussians into 2D flat surfels for accurate geometric representation, tackling occlusions. Section 2.2 describes techniques to decompose a radiance field to better represent cast shadows due to changing lighting conditions, while Section 2.3 outlines current publicly available datasets for image-based geometric reconstruction hindered by shadows.

2.1. Planar-based Gaussian splatting

Starting from sparse points usually produced during Structure-from-Motion (SfM) (Schonberger and Frahm, 2016), 3DGS (Kerbl et al., 2023) offers efficient scene representation and real-time rendering through explicit 3D Gaussian primitives. Unlike NeRFs (Mildenhall et al., 2021), which rely on volumetric rendering, 3DGS uses a differentiable splatting pipeline to project 3D Gaussians onto the image plane, achieving high-quality novel view synthesis with significantly faster training time. However, 3DGS faces challenges in accurately representing scene geometry due to the volumetric nature of the Gaussian primitives, as the intersection between a pixel ray and 3D Gaussian leads to inconsistent depth from different viewpoints. Inspired by this, SuGaR (Guédon and Lepetit, 2024), NeuSG (H. Chen et al., 2023), Gaussian Surfels (Dai et al., 2024) and GeoGaussian (Li et al., 2024) propose flattening the Gaussians by minimizing the smallest scale to encourage thin Gaussians to conform better with object's surface. 2DGS (Huang et al., 2024) utilizes explicit ray-splat intersection by collapsing the 3D Gaussians into a set of 2D oriented planar Gaussian disks enforced with depth and normal consistency regularization. DeMapGS anchors 2D Gaussian splats to a deformable mesh, which enforces structural and topological consistency. PGSR (D. Chen et al., 2024) and GausSurf (Wang et al., 2024) flatten Gaussians into planes using unbiased depth which directly renders the distance from the camera origin to the Gaussian plane and the corresponding normal map complemented with single-view geometric, multi-view photometric and geometric regularization to further strengthen the surface reconstruction. HDGS (Song et al., 2024) propose aligning the 2D surfels with texture maps and augment them with per-ray depth sorting and pruning for consistency and efficiency. Despite the strong performance in preserving fine details and reducing artifacts, the computational cost is high for moderate geometric accuracy. Incorporating an additional geometric loss based on the planarity derived from the eigenvalue, FeatureGS (Jäger et al., 2025) flattens the volumetric Gaussians and improves the geometric accuracy up to 30% due to the significant reduction of artifacts. GOF (Yu et al., 2024) approximate the surface normals of 3D Gaussians as the normals of intersection planes between the ray and Gaussians. Instead of projecting and evaluating the 3D Gaussians to 2D screen space, the contribution of a Gaussian is calculated with explicit ray-Gaussian

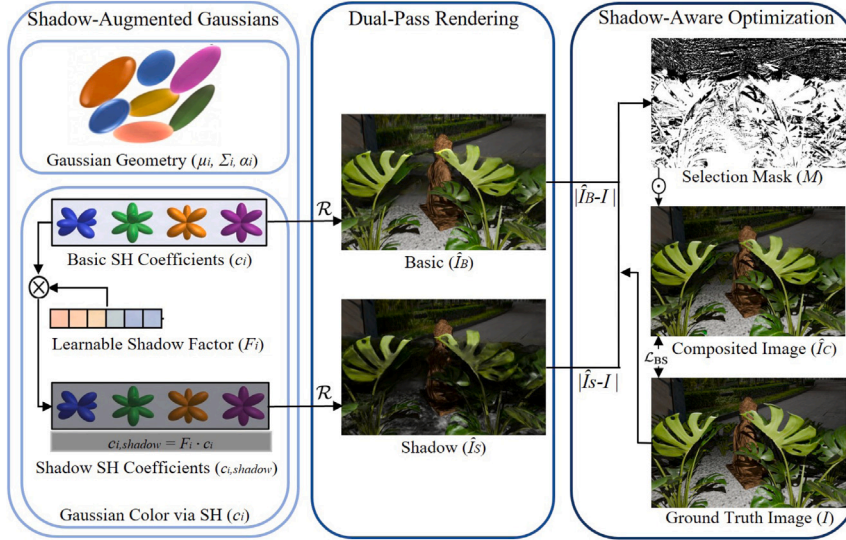


Fig. 1. Overview of the proposed Shadow-Aware Gaussian Splatting (SGS) which explicitly separates intrinsic radiance from extrinsic illumination. The core of the method is unbaking shadows from scene geometry, hence each Gaussian is augmented with a learnable shadow factor (F_i) to account for illumination variations (Section 3.2). To satisfy the multi-view photometric consistency and facilitate learning of the shadow factor, we implement a dual-pass rendering strategy which produces two competing representations from the same geometric primitives, differing only in their color features for every viewpoint, namely under full illumination from the basic Spherical Harmonics (SH) (*Basic*) and in shadow (*Shadow*) (Section 3.3). We resolve the ambiguity and optimize Gaussian geometry and appearance parameters with a shadow-aware Optimization through a bimodal selection loss, which dynamically generates a pixel-wise mask based on the lowest residual error relative to the ground truth for gradient backpropagation (Section 3.4).

intersection where the Gaussian reaches its maximum value along the corresponding ray. 3DGS-to-PC (AG Stuart et al., 2025) increases the overall completeness by sampling points from each Gaussian’s multivariate distribution proportional to its scale and contribution with outlier rejection to ensure geometric consistency.

The above introduced GS variants assume simultaneously captured images with unobstructed line of sight which is very unlikely in real-world setup. Similar to SpotlessSplats (Sabour et al., 2024) which exploit semantic features to effectively identify transient distractors without explicit supervision, DeSplat (Wang et al., 2025) decomposes static and dynamic components. However, if distractors appear static across most views, both methods may mistakenly detect them as part of the static scene. Additionally, a small color difference between distractors and static objects can prevent clear separation potentially leading to loss of detail and reduced image quality. WildGaussians (Kulhanek et al., 2024) and Splatfacto-W (Xu et al., 2024) extend GS to unconstrained environments. However, both methods struggle with high-frequency occlusions that persist across the majority of the training set and fail to distinguish transient distractors from the underlying static geometry, leading to artifacts. Inspired by its predecessor RobustNeRF (Sabour et al., 2023), Robust3DGS (Ungermann et al., 2024) simultaneously optimizes for distractor masks to support the static scene reconstruction by leveraging image residuals in the training process and apply different transformations to obtain local smoothness and distractor contiguity.

2.2. Shadow-aware reconstruction

The first neural radiance-based approach for scene decoupling NeRF-OSR (Rudnev et al., 2022) learns a neural representation of the scene geometry, diffuse albedo and illumination-dependent shadows from a set of images capturing the same scene from different viewpoints at different time. It learns the scene intrinsics and illumination as expressed by the SH coefficients where a dedicated neural component learns shadows. Comparably, OSDG-GS (Feng et al., 2025) addresses the ambiguity of collection of outdoor images with changing and unknown lighting conditions. The images are clustered into several lighting groups by lighting cues distillation for each group to learn

the lighting conditions of the scene to produce realistic cast shadows, similar to NeRF-in-the-Wild (Martin-Brualla et al., 2021). Unlike GS-IR (Liang et al., 2024) which leverages baking-based volumes to store occlusion and accelerate rendering, but struggles to represent indirect lighting, GI-GS (H. Chen et al., 2024) leverages deferred shading to realize efficient path tracing-based occlusion and indirect illumination for both rendering and relighting. However, for complex real-world scenes, it may fail to capture the spatially varying lighting in the environment. SSD-GS (Zheng et al., 2026) separate reflectance into physically interpretable elements to represent soft shadows using a two-stage approach that combines per-ray shadow evaluation with neural refinement. The color of each Gaussian is computed as scalar reflectance intensity for diffuse, specular and subsurface scattering. GaRe (Bai et al., 2025) decomposes global illumination into consistent, physically interpretable shading components, enabling realistic appearance variations through independent interpolation across diverse illumination conditions. Furthermore, it integrates the ray-traced results into Gaussian features, allowing ray-direction adaptive shadow rendering. ShadowGS (Luo et al., 2026) effectively extracts illumination components from appearance attributes using a set of globally shared SH to encode albedo and to represent reflected radiance from nearby surfaces rendered via alpha blending. The method accurately approximates shadow variations across multi-temporal observations from satellite imagery.

2.3. Shadow benchmarks

Several publicly available datasets address the task of reconstruction behind occlusions and cast shadows due to lighting variations. The Phototourism dataset (Snavely et al., 2006) consists of unconstrained internet photo collections of cultural landmarks under different illumination and with distractors. NeRF-OSR (Rudnev et al., 2022) present a new benchmark for outdoor scene relighting containing eight scenes captured from various viewpoints using a DSLR camera at different time of the day. SOBA (Wang et al., 2020), DESOBA (Hong et al., 2022), RdSOBA (Tao et al., 2024) and ADSP (Hsieh and Ding, 2024) contain pairs of shadow and object instances each with individually labeled masks. The predicted shadow and object instances are paired

and matched with the predicted shadow-object associations to generate the final results. Given an image with shadows, the goal is to detect the individual shadow and object instances, as well as to associate the shadows with the objects that cast them. D²NeRF (Wu et al., 2022) introduce a new dataset with various moving shadows in both synthetic and real-world settings. S-EO (Masquil et al., 2025) dataset extends the geometry-aware shadow detection in satellite imagery.

Although the numerous radiance-based methods listed above address geometric reconstruction from outdoor multi-view images captured under unknown lighting conditions, they focus on reconstructing soft shadows and usually from non-overlapping images, thus are more suitable for per-image illumination estimation and instance shadow detection. The evaluation refers to the rendering quality as the goal is photorealistic scene relighting under novel lighting conditions. In contrast, SGS introduces a novel scene rendering and optimization strategy to address the impact of intricate shadows cast from occlusions on the 3D reconstruction and explicitly disentangle shadows from scene geometry. Moreover, our scenes consist of overlapping images with multi-view consistent occlusions present in every image, to evaluate the geometry behind occlusions and impact of their cast shadows as there is no dataset depicting this type of setup.

3. Method

As illustrated in Fig. 1, we introduce SGS, a novel unsupervised Shadow-Aware Gaussian Splatting method that explicitly separates intrinsic radiance from extrinsic illumination effects by a learnable shadow factor which augments the view-dependent SH to account for illumination variations (Section 3.2). Central to our approach is a dual-pass rendering strategy (Section 3.3) combined with a shadow-aware optimization through a bimodal selection loss (Section 3.4). By selecting the minimal residual error relative to the training data, the optimization forces clean separation of shadows from scene geometry, yielding a cleaner gradient for geometric updates. By explicitly handling shadows as a photometric modulation, the positional gradients used for geometric updates are no longer contaminated by lighting residuals. We first present the preliminaries of planar-based GS (Section 3.1) on which SGS is grounded, although it can be incorporated in any GS framework.

3.1. Preliminaries on planar-based GS

A Gaussian primitive is defined by its center $\mu_i \in \mathbb{R}^3$, covariance matrix Σ_i , opacity $\alpha_i \in [0, 1]$ and color features $c_i \in \mathbb{R}^K$ via SH coefficients. The covariance Σ_i can be decomposed into scaling $S_i \in \mathbb{R}^{3 \times 3}$ and rotation matrix $R_i \in \mathbb{R}^{3 \times 3}$: $\Sigma_i = R_i S_i S_i^T R_i^T$. To accurately conform to the scene surface, the 3D Gaussian ellipsoids are flattened into 2D planes along the direction of the minimum scale factor $S_i = \text{diag}(s_1, s_2, s_3)$ for each Gaussian.

Contributions from overlapping Gaussians are accumulated using α -compositing, where the opacity value α determines the influence of each Gaussian on the final rendered pixel. Rendering color $C \in \mathbb{R}^3$ of a pixel is obtained through α -blending.

$$C = \sum_{i \in N} T_i \alpha_i c_i, \quad T_i = \prod_{j=1}^{i-1} (1 - \alpha_j) \quad (1)$$

where $c_i \in \mathbb{R}^3$ denotes the view-dependent color, T_i is the cumulative opacity and N is the number of Gaussians that the ray passes through.

The rendered image $\hat{I}$ is compared against the corresponding ground truth image I using a photometric loss, which drives the refinement of all Gaussian parameters. The $\mathcal{L}_1$ loss ensures that the rendered image is consistent with the ground truth image, while the SSIM term requires the rendered image to have similar structure to the ground truth image with $\lambda_{\text{D-SSIM}} = 0.2$ (Kerbl et al., 2023).

$$\mathcal{L}_{\text{rgb}} = (1 - \lambda_{\text{D-SSIM}}) \mathcal{L}_1 + \lambda_{\text{D-SSIM}} \text{D-SSIM}(\hat{I}, I) \quad (2)$$

3.2. Shadow-augmented Gaussians

Given that the SH coefficients are defined over an angular domain of the viewing direction, illumination changes cause each Gaussian to exhibit varying color, preventing accurate geometric representation. To absorb the photometric variance rather than warping the underlying Gaussian structure, we enhance the view-dependent SH to account for illumination variations without adding more SH levels, thus restricting polynomial increase in the parameter count and computational overhead. Each Gaussian primitive is augmented with a learnable *Shadow Factor* $f_i \in \mathbb{R}$, which is transformed into a normalized visibility coefficient F_i via a sigmoid activation function.

$$F_i = \sigma(f_i) = \frac{1}{1 + e^{-f_i}}, \quad F_i \in [0, 1] \quad (3)$$

Differing from the SH coefficients, which explicitly represent view-dependent appearance variations, the shadow factor acts as a regularized appearance decomposition mechanism rather than an explicit shadow estimator. Consequently, it serves as a latent parameter providing an alternative degree of freedom that allows the optimizer to explain illumination-dependent intensity variations without modifying the underlying geometry.

The color features observed in shadow are calculated as $c_{i,\text{shadow}} = F_i \cdot c_i$ for the Direct Components (DC) and SH higher-order coefficients (rest).

$$c_{i,\text{shadow}}^{\text{DC}} = F_i \cdot c_i^{\text{DC}} \quad (4)$$

$$c_{i,\text{shadow}}^{\text{rest}} = F_i \cdot c_i^{\text{rest}} \quad (5)$$

We initialize the DC features with a neutral constant to prevent bias from the initial point cloud.

$$c_{i,\text{init}}^{\text{DC}} = \frac{0.5}{C_0} \quad (6)$$

where $C_0 \approx 0.28209$ (D. Chen et al., 2024). This forces the algorithm to learn the scene color and shadow distribution entirely from the optimization of the photometric loss without directional dependence. To ensure physically plausible light attenuation, we modify the standard SH-to-RGB conversion which typically includes a constant 0.5 bias to center optimized colors around a neutral gray.

During the densification phase, the shadow representation must remain spatially consistent. When a Gaussian G_p is split into N Gaussians $\{G_1, \dots, G_N\}$, the shadow component is preserved by inheriting Gaussian parameters, such that $f_n = f_p, \forall n \in \{1, \dots, N\}$. Conversely, if a Gaussian is pruned, its associated shadow factor is removed from the optimization.

3.3. Dual-pass rendering

GS methods interpret shadows as physical, dark-colored surface textures which are baked directly into the color of the Gaussians to recreate the shadow's appearance. To satisfy the multi-view photometric consistency and facilitate the learning of the shadow factor, we implement a *Dual-Pass Rendering* strategy. For each training viewpoint, two renderings from the same geometric primitives, differing only in their color feature sets are generated. We define two states of radiance accumulation: *Basic* representing the scene under full illumination and *Shadow* derived by attenuating *Basic* with the shadow factor.

For a set of N Gaussians, the renderer produces two distinct images for each state, $\hat{I}_B$ and $\hat{I}_S$, using the same geometric parameters (position, opacity and scale), but different color coefficients.

$$\hat{I}_B = \mathcal{R}(\{\mu_i, \Sigma_i, \alpha_i, c_i\}, \mathcal{V}) \quad (7)$$

$$\hat{I}_S = \mathcal{R}(\{\mu_i, \Sigma_i, \alpha_i, c_{i,\text{shadow}}\}, \mathcal{V}) \quad (8)$$

where $\mathcal{V}$ denotes the camera pose, $\mathbf{c}_i$ represents the color, encoded through SH coefficients and $\mathbf{c}_{i,\text{shadow}}$ those features multiplied by the learnable shadow factor F_i (as defined in Eqs. (4) and (5)).

Following the point-based alpha-blending, the final color of a pixel in $\hat{C}_S$ is determined by the ordered accumulation of N Gaussians overlapping the pixel. Our modified rendering equation incorporates the Shadow Factor F_i into the primary radiance accumulation.

$$C_S = \sum_{i=1}^N (F_i \cdot \mathbf{c}_i(\mathbf{v})) \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j) \quad (9)$$

where α_i is the opacity of the i th Gaussian and view-dependent color $\mathbf{c}_i(\mathbf{v})$ for a viewing direction $\mathbf{v}$. By making F_i an optimizable parameter, the gradient flow through allows the algorithm to suppress the radiance of specific Gaussians to represent shadows without distorting the underlying geometry.

3.4. Shadow-aware optimization

To resolve the ambiguity between *Basic* and *Shadow* and optimize Gaussian geometry and appearance parameters through gradient back-propagation, we propose a *Bimodal Selection Loss*. Considering that our dual-rendering pipeline generates two competing radiance states for every pixel ($\hat{I}_B$) and ($\hat{I}_S$), the primary challenge lies in the latent assignment of ground truth pixels I to their most plausible state. We resolve this by computing per-pixel residual error maps for both renderings.

$$\epsilon_B = |\hat{I}_B - I| \quad (10)$$

$$\epsilon_S = |\hat{I}_S - I| \quad (11)$$

By minimizing the lower residual at each spatial location, the optimization signal is automatically directed to the state that best explains the observation, effectively extracting shadows from intrinsic color as the $\mathcal{L}_{\text{shadow}}$ loss acts as a gradient routing mechanism.

$$\mathcal{L}_{\text{shadow}} = \frac{1}{HW} \sum_{u,v} \min(\epsilon_B(u, v), \epsilon_S(u, v)) \quad (12)$$

where H and W denote the height and width of the image, and (u, v) index pixel coordinates.

We implicitly define a selection mask M based on these residuals: $M(u, v) = \mathbb{1}(\epsilon_B < \epsilon_S)$. During the backward pass, the gradient $\nabla \mathcal{L}$ is backpropagated only through the parameters of the Gaussians with smaller error with respect to the ground truth image. This effectively assigns pixels to either the *Basic* or *Shadow* state. If a pixel in the ground truth is in shadow, ϵ_S is lower, forcing optimization of the shadow factor F_i . Conversely, the remaining pixels are optimized through the basic SH features. To maintain structural consistency, the composited image Fig. 2 is then defined as.

$$\hat{I}_C = M \odot \hat{I}_B + (1 - M) \odot \hat{I}_S \quad (13)$$

The standard photometric loss (defined in Eq. (2)) forces a single radiance value to account for varying illumination and represent diffuse color and shadow observations across different views, often leading to baked-in shadows in the SH coefficients or blurred geometry. Our proposed loss disentangles appearance thereby preventing shadows from being baked into the intrinsic SH and refines the geometry by allowing F_i to provide additional cues for surface reconstruction. Our Bimodal Selection loss $\mathcal{L}_{BS}$ is a weighted combination of the $\mathcal{L}_{\text{shadow}}$ and the SSIM computed on the composited image $\hat{I}_C$.

$$\mathcal{L}_{BS} = (1 - \lambda) \mathcal{L}_{\text{shadow}} + \lambda_{\text{D-SSIM}} \text{D-SSIM}(\hat{I}_C, I) \quad (14)$$

4. Experiments

We describe the generated dataset with multi-view consistent occlusions under diffuse and direct light where the occlusions cast shadows

in Section 4.1. In Section 4.2 we briefly summarize the geometric and photometric evaluation metrics and in Section 4.3 we elaborate the hyperparameter optimization and implementation details of our method.

4.1. Dataset

To support a comprehensive evaluation, we utilize synthetic data for better control over the occlusions and their cast shadows' placement and coverage. The object whose recovered geometry should be evaluated is a Buddha statue (further on referred to as object) on a rectangular plate. We import the object mesh in Blender¹ and because of the transparent background, we build a textured cube around it to enable proper pose estimation. For the occlusions we use free models from the BlenderKIT² add-on and place them around the object, as the occlusions are always in front. We generate seven scenes with vegetation occlusions consisting of uniform plant type per scene: *Twig* (thin branches of trees without leaves), *Daisy* (flower with small leaves), *Ficus* (plant with delicate leaves), *Lomandra* (bush with narrow, arching leaves), *Orchid* (flower with petals and elongated leaves), *Monstera* (large leaves with splits and holes) and *Birch* (thin-leaved dense tree). To analyze the impact of shadows on the geometric reconstruction, we generate images from the same position once under diffuse light (without shadows) and once with direct light (with shadows), thus acquire two image sets for each scene (Fig. 3). For the first setting, we set a plane which illuminates the scene uniformly. For the latter where the occlusions cast shadows, we mimic digital camera flash and set a light with small offset above the camera. The images are generated in a circular trajectory at a uniform distance from the object with a focal length of 20 mm and f/8 aperture size using the Cycles rendering engine. GPU-acceleration and compositing option are disabled to avoid stochastic deviation and color correction (Schulte et al., 2025). For each scene we generate 125 images with a resolution of 920×614 px each. The images will be publicly available because there is no existing dataset dedicated to this problem setting.

4.2. Evaluation

We evaluate SGS qualitatively and quantitatively in 3D and 2D, namely the geometry on level of point clouds and surface reconstruction from meshes, as well as the generated novel views.

Geometric We quantify the 3D geometric accuracy through cloud-to-mesh distances which compute the closest distance of a point to a triangle in the reference mesh. Furthermore, we assess the surface reconstruction through mesh-to-mesh distances to calculate the distance between vertices of the compared mesh and the nearest triangles of the reference mesh. We crop a bounding box with respect to object's dimensions and for the point clouds additionally filter points within 5 mm which contribute to the completeness (Cpl) to evaluate object's geometry. After aligning the reconstructed point cloud and meshes to the reference mesh we calculate Mean, Standard Deviation (SD) and Root Mean Squared Error (RMSE).

Photometric In order to measure the quality of rendered images, we adopt Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM) and Learned Perceptual Image Patch Similarity (LPIPS) as our evaluation metrics.

4.3. Implementation details

We train with learning rates of 0.0025 for SH, 0.05 for opacity adjustments, 0.005 for scaling operations and 0.001 for rotation

¹ <https://www.blender.org/>.

² <https://www.blenderkit.com/>.

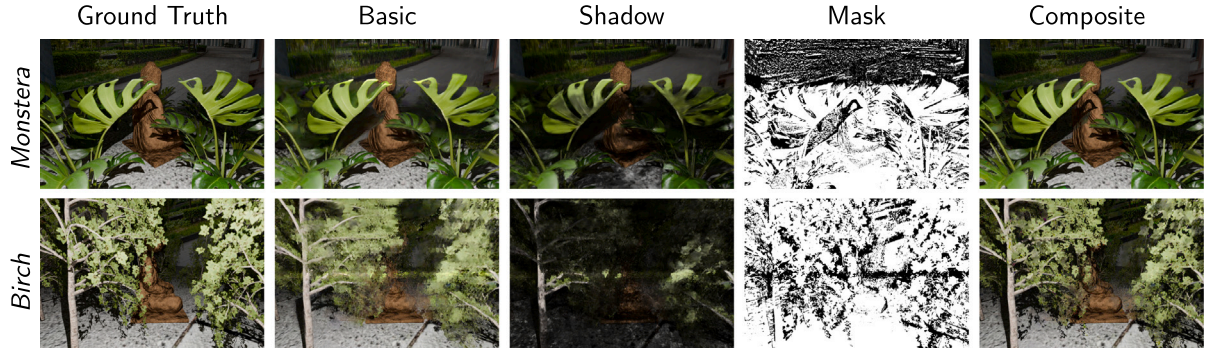


Fig. 2. SGS separates the scene into its underlying base color under full illumination and its interaction with light. By selecting the rendering with the lower photometric error between the two passes, the optimization automatically determines whether each pixel belongs to the illuminated base or shadow region based on the hypothesis that best matches the ground truth image.



Fig. 3. We generate images on a circular trajectory (black circle) around the object from the same position (a) under diffuse light without shadows, thus we set a plane which illuminates the scene uniformly and (b) under direct light where the occlusions cast shadows by setting a light with small offset above the camera.

Table 1

3D and 2D evaluation for different densification thresholds (τ_{grad}) which decide if points should be densified based on the positional gradient for *Orchid*. In combination with the shadow factor, $\tau_{grad} = 0.0001$ achieves best balance between geometry and novel views and is further used in the training. The **first**, **second** and **third** best results are highlighted.

Densify threshold	SD (mm) ↓	RMSE (mm) ↓	Cpl (%) ↑	PSNR ↑	SSIM ↑	LPIPS ↓
$\tau_{grad} = 0.00005$	26.03	26.42	75.71	24.98	0.827	0.234
$\tau_{grad} = 0.0001$	21.43	21.63	62.52	25.71	0.830	0.237
$\tau_{grad} = 0.0002$	24.53	24.81	39.19	25.87	0.831	0.244
$\tau_{grad} = 0.0003$	31.30	31.72	30.65	25.95	0.823	0.258

transformations. In the original SH formulation, the 0.5 constant serves as a numerical bias that centers color optimization around a neutral gray baseline within the normalized $[0, 1]$ RGB space. This offset ensures that at initialization, where SH coefficients are typically zero-centered, the Gaussians possess a visible radiance that facilitates stable gradient flow. This allows the optimizer to efficiently explore the color space as a signed residual, enabling symmetric convergence towards both high and low intensity values. However, this residual-based approach is mathematically incompatible with our multiplicative shadow factor, as the bias prevents the representation of true zero-radiance. We therefore omit this offset from the rasterization kernel and the SH utility functions to ensure purely linear scaling. Following our bimodal selection loss, if a pixel is better represented by the *Basic* state, the *Shadow* state receives zero gradient for that pixel and vice versa. Because gradients only flow through the pixels with smaller error residuals, many Gaussians receive smaller gradient updates during training. Since densification is triggered by the accumulation of large gradients over time, we use a threshold of 0.0001 (Table 1). This leads to a denser point cloud, the increased primitive count is directed towards recovering fine-grained geometric details resulting in a higher-fidelity surface

reconstruction where shadow boundaries are preserved without blurring the underlying geometry.

We estimate the poses on the same scene without occlusions through the incremental SfM (Schonberger and Frahm, 2016) pipeline in COLMAP³ (Schönberger et al., 2016) and use them for all synthetic scenes. We compare SGS with PGSR (D. Chen et al., 2024), on which SGS is grounded, trained with the default parameters. The training iterations for all scenes are set to 15 000 with every 8th frame taken for test. All experiments are conducted on an Nvidia RTX 3090 GPU.

5. Results

We present the results qualitatively in 3D by visualizing the reconstructed point clouds (Figs. 4, 5) along with the cloud-to-mesh errors (Fig. 7) and surface representation through meshes (Fig. 6) as well as in 2D through the rendered novel views (Fig. 8) as well as quantitatively to numerically analyze the accuracy, completeness and rendering fidelity (Tables 2, 5, 3 and 4). To investigate the impact of shadows on the geometric reconstruction we present the results without (Section 5.1) and with shadows (Section 5.2). Note that PGSR* refers to the reconstruction on the scenes captured under diffuse light without shadows.

5.1. Geometric reconstruction without shadows

3D Reconstruction PGSR* reliably approximates the object's geometry behind occlusions maintaining competitive error metrics in scenes with smaller occlusions, like *Twig*, *Daisy* and *Ficus* (Fig. 7). As the

³ <https://github.com/colmap/colmap>.

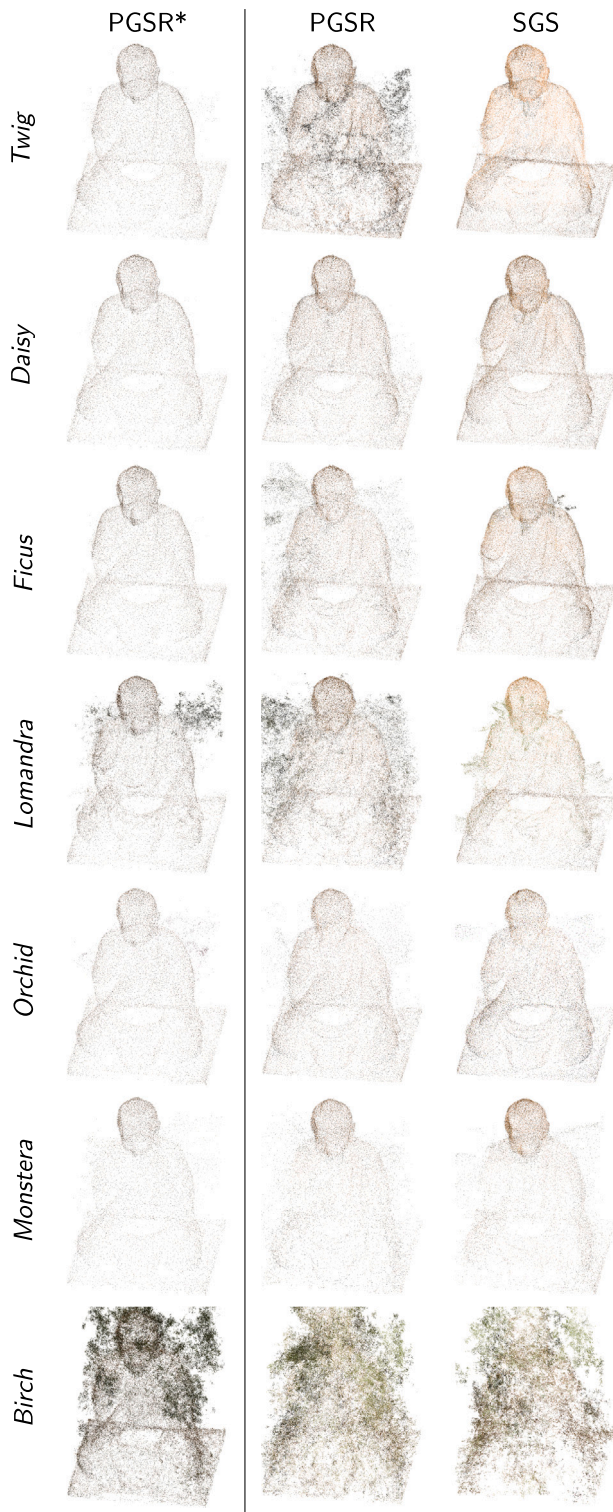


Fig. 4. Point cloud reconstruction cropped around a bounding box with respect to object's dimensions. SGS significantly reduces the floating artifacts caused by the shadows cast from the vegetation occlusions which PGSR bakes as color in scene geometry. Due to the dense foliage with many small leaves, *Birch* imposes strong challenges for the methods without (*) and with shadows.

complexity and coverage of the occlusions increase, the accuracy drops due to the floating artifacts around the object, especially noticeable in *Lomandra* and *Birch* (Fig. 4). Filtering the points within 5 mm with respect to the reference gives better insights of how the object

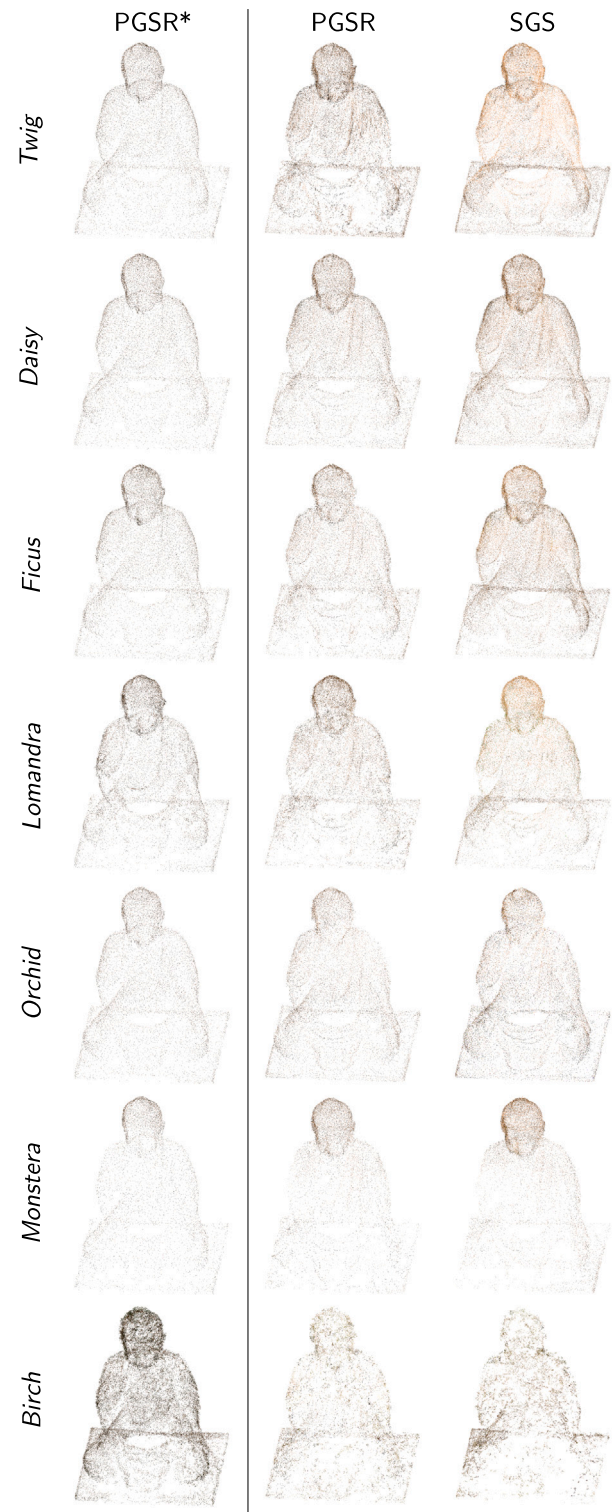


Fig. 5. Point cloud reconstruction filtered within 5 mm. The gaps and sparser areas correspond to the occlusions and cast shadows. SGS is able to reliably approximate the geometry behind shadows increasing overall point coverage even compared to PGSR* without shadows.

geometry is reconstructed, reasonably lower than PGSR with shadows (Fig. 5). The completeness proportionally decreases with more complex and denser occlusions namely, in *Twig* and *Ficus* which consist of thin branches without or small leaves is slightly above 55%, it drops to around 53% in *Daisy* and *Lomandra* caused by the bigger flower

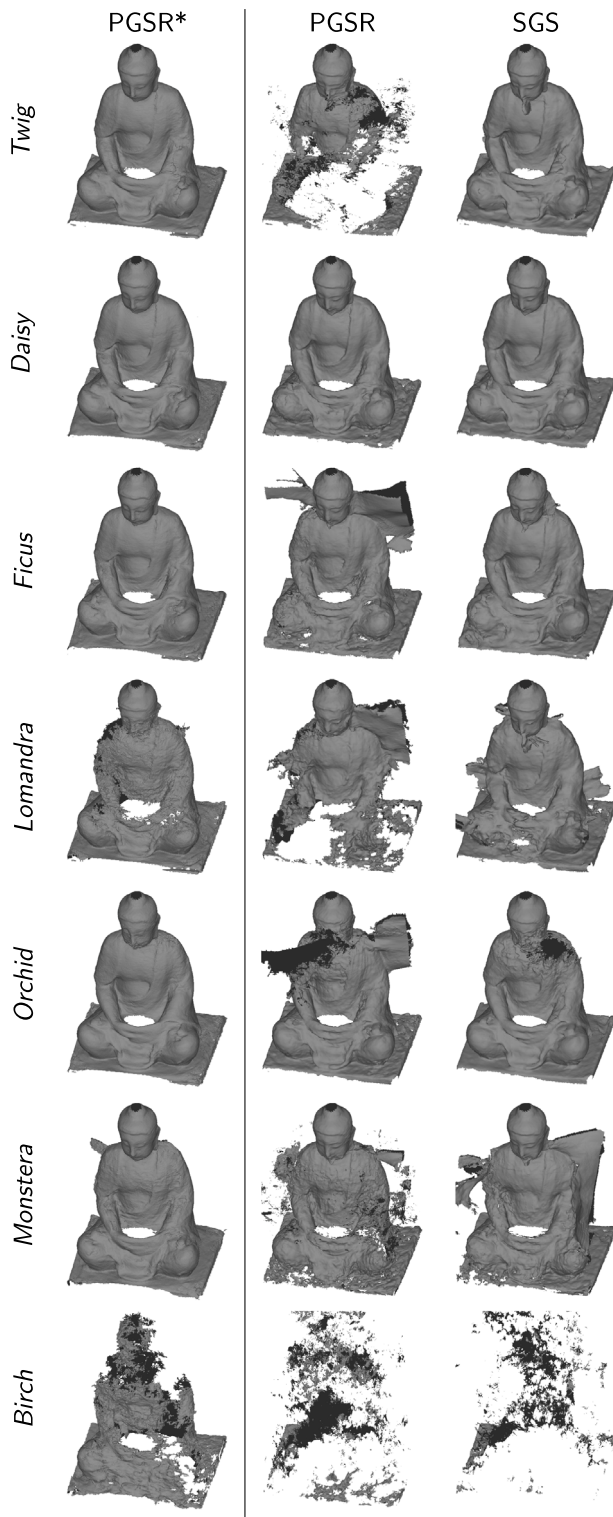


Fig. 6. Mesh reconstruction. SGS achieves better surface representation and watertight meshes, mostly because of the reduced floaters caused by the shadows. Due to the floaters in the point cloud, *Birch* shows poor surface geometry. PGSR* denotes reconstruction on shadows-free images.

blossoms and longer leaves respectively, while challenged by the big petals and leaves covering big object parts drops to 48.69% and 31.30% in *Orchid* and *Monstera* (Tables 2 and 5). The average completeness score of 53.44% with only 17 364 points belonging to the object is

Table 2

Quantitative cloud-to-mesh results after cropping to points within a bounding box defined by the object dimensions. Negative Mean values indicate a directional bias where points are inside the reference mesh, whereas SD and RMSE capture the absolute error magnitude. SGS exhibits 2× better accuracy than PGSR and it is even better than PGSR* without shadows, complemented with highest completeness implying a well-balanced trade-off between precision and surface coverage. The **first** and **second** best results are highlighted.

Scene	Method	Accuracy (mm)			Completeness (%)	
		Mean ↓	SD ↓	RMSE ↓	Npts	Cpl ↑
<i>Twig</i>	PGSR*	0.03	8.86	8.86	16 536	55.40
	PGSR	28.71	49.28	57.04	71 246	67.52
	SGS	0.24	5.72	5.72	54 806	81.06
<i>Daisy</i>	PGSR*	−0.50	6.53	6.54	15 076	53.70
	PGSR	1.00	10.82	10.87	28 338	66.30
	SGS	0.01	5.73	5.73	43 603	78.52
<i>Ficus</i>	PGSR*	−0.43	5.71	5.72	17 028	55.46
	PGSR	22.35	47.61	52.59	32 989	60.07
	SGS	0.94	10.57	10.61	43 276	74.63
<i>Lomandra</i>	PGSR*	26.42	46.01	53.06	35 667	53.57
	PGSR	40.82	56.03	69.32	61 916	55.19
	SGS	5.23	16.70	17.50	40 006	64.74
<i>Orchid</i>	PGSR*	6.85	30.54	31.30	14 006	48.69
	PGSR	8.06	32.33	33.32	21 068	53.84
	SGS	2.88	21.43	21.63	25 228	62.52
<i>Monstera</i>	PGSR*	4.65	23.68	24.13	10 766	31.30
	PGSR	5.94	24.66	25.37	14 996	37.64
	SGS	5.03	25.46	25.95	19 153	41.95
<i>Birch</i>	PGSR*	42.42	55.75	70.05	157 789	75.98
	PGSR	48.10	56.99	74.58	146 818	44.27
	SGS	38.23	49.37	62.44	125 504	36.63

Table 3

Quantitative results of the mesh evaluation along with the training time. As expected, PGSR* without shadows provides more accurate surface representation, however SGS achieves 2× higher accuracy than PGSR for only 4 min of extra time for an explicit separation of shadows from scene geometry. The **first** and **second** best results are highlighted, excluding PGSR*.

Scene	Method	Accuracy (mm)			Time (min) ↓
		Mean ↓	SD ↓	RMSE ↓	
<i>Twig</i>	PGSR*	0.82	3.35	3.45	22
	PGSR	13.29	33.43	35.97	33
	SGS	1.80	4.44	4.79	32
<i>Daisy</i>	PGSR*	0.63	3.32	3.38	19
	PGSR	1.37	3.46	3.72	20
	SGS	1.43	3.48	3.76	26
<i>Ficus</i>	PGSR*	0.59	3.35	3.40	24
	PGSR	13.15	31.22	33.88	26
	SGS	1.56	4.01	4.30	31
<i>Lomandra</i>	PGSR*	0.54	4.03	4.07	31
	PGSR	11.96	28.26	30.68	34
	SGS	5.80	13.79	14.96	32
<i>Orchid</i>	PGSR*	0.62	3.41	3.47	19
	PGSR	19.14	43.17	47.22	19
	SGS	1.79	4.87	5.19	26
<i>Monstera</i>	PGSR*	1.05	5.25	5.35	18
	PGSR	13.58	34.01	36.62	19
	SGS	18.74	40.80	44.89	24
<i>Birch</i>	PGSR*	2.02	6.11	6.44	42
	PGSR	26.99	51.18	57.86	37
	SGS	17.84	36.42	40.55	43

Table 4

2D image metrics of rendered novel views for the vegetation scenes. The large difference in PSNR, SSIM and LPIPS is particularly notable, suggesting that the structural integrity and perceptual quality of SGS are vastly superior, showing robustness. PGSR* denotes reconstruction without shadows. The **first** and **second** best results are highlighted.

Scene	Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
Twig	PGSR*	23.64	0.713	0.357
	PGSR	16.07	0.596	0.444
	SGS	19.65	0.724	0.360
Daisy	PGSR*	22.44	0.759	0.314
	PGSR	18.86	0.691	0.349
	SGS	21.88	0.799	0.261
Ficus	PGSR*	22.59	0.721	0.323
	PGSR	16.80	0.636	0.383
	SGS	20.13	0.770	0.280
Lomandra	PGSR*	22.56	0.707	0.391
	PGSR	16.33	0.549	0.480
	SGS	20.94	0.731	0.336
Orchid	PGSR*	23.62	0.763	0.311
	PGSR	20.79	0.729	0.346
	SGS	25.71	0.830	0.237
Monstera	PGSR*	25.59	0.786	0.351
	PGSR	22.22	0.734	0.367
	SGS	23.92	0.828	0.262
Birch	PGSR*	21.19	0.615	0.474
	PGSR	13.98	0.451	0.538
	SGS	18.65	0.713	0.374

Table 5

Quantitative cloud-to-mesh results filtered within 5 mm. Negative Mean values indicate a directional bias where points are inside the reference mesh, whereas SD and RMSE capture the absolute error magnitude. Compared to PGSR, SGS achieves higher correspondence with the reference, just behind PGSR* without shadows, allocating accurate surface object points thus improving overall completeness. The **first** and **second** best results are highlighted.

Scene	Method	Accuracy (mm)			Completeness (%)	
		Mean $\downarrow$	SD $\downarrow$	RMSE $\downarrow$	Npts	Cpl $\uparrow$
Twig	PGSR*	-0.33	2.34	2.36	12 858	55.40
	PGSR	0.03	2.54	2.54	31 279	67.52
	SGS	-0.14	2.46	2.46	43 362	81.06
Daisy	PGSR*	-0.33	2.27	2.29	12 159	53.70
	PGSR	-0.20	2.44	2.45	20 551	66.30
	SGS	-0.10	2.40	2.40	35 205	78.52
Ficus	PGSR*	-0.26	2.34	2.36	13 819	55.46
	PGSR	-0.08	2.52	2.52	17 296	60.07
	SGS	-0.17	2.42	2.42	32 538	74.63
Lomandra	PGSR*	-0.13	2.49	2.49	16 846	53.57
	PGSR	0.02	2.63	2.62	20 000	55.19
	SGS	-0.29	2.49	2.51	25 985	64.74
Orchid	PGSR*	-0.25	2.43	2.44	9 988	48.69
	PGSR	0.05	2.53	2.53	13 206	53.84
	SGS	0.03	2.42	2.48	18 624	62.52
Monstera	PGSR*	0.10	2.61	2.62	6 650	31.30
	PGSR	-0.48	2.40	2.45	9 308	37.64
	SGS	-0.45	2.30	2.35	13 779	41.95
Birch	PGSR*	0.36	2.57	2.60	49 226	75.98
	PGSR	0.41	2.81	2.84	17 705	44.27
	SGS	0.20	2.76	2.77	22 766	36.63

lowest compared to the reconstructions additionally challenged by cast shadows. It is worth noting that it produces smooth meshes providing

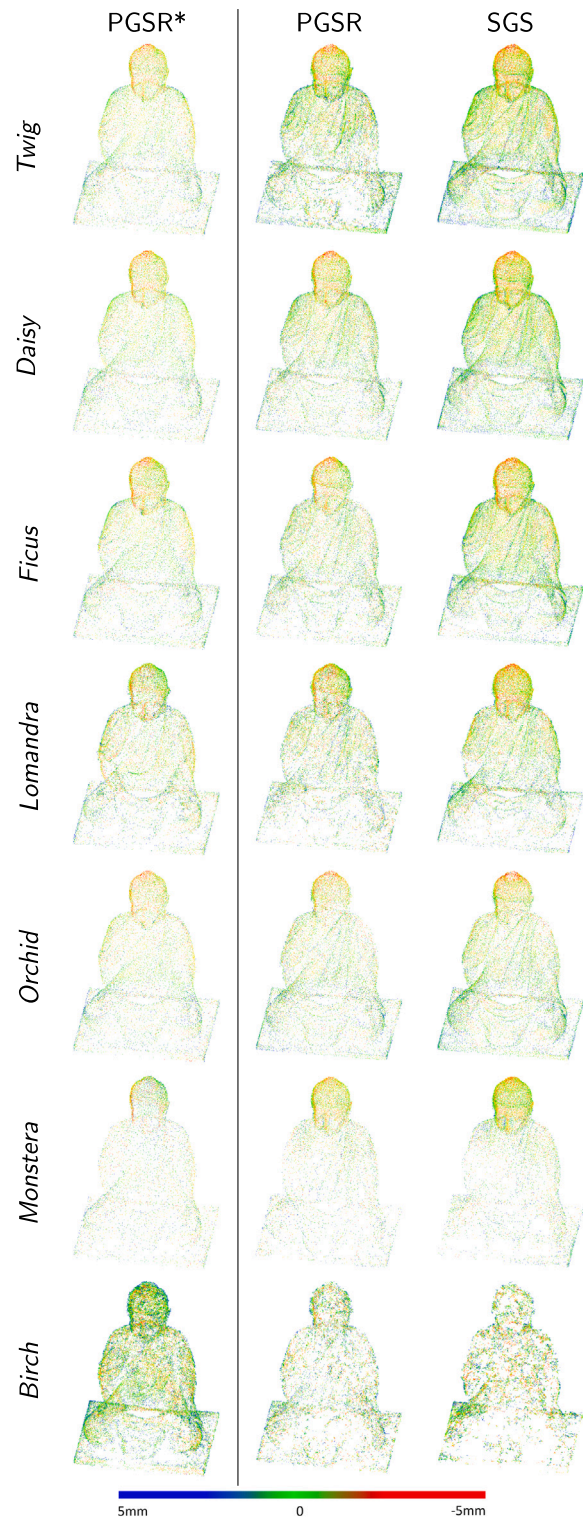


Fig. 7. Cloud-to-mesh distances cropped within 5 mm. Since the reference is a mesh, the distances are signed where positive values indicate points above the mesh and negative when behind. SGS shows better correspondence with the reference than PGSR with less noise and artifacts, competitive with PGSR* without shadows.

reliable surface reconstruction (Table 3). Stemming from the floaters in the point clouds, in *Lomandra* the mesh experiences holes and is poorly reconstructed in *Birch* (Fig. 6).

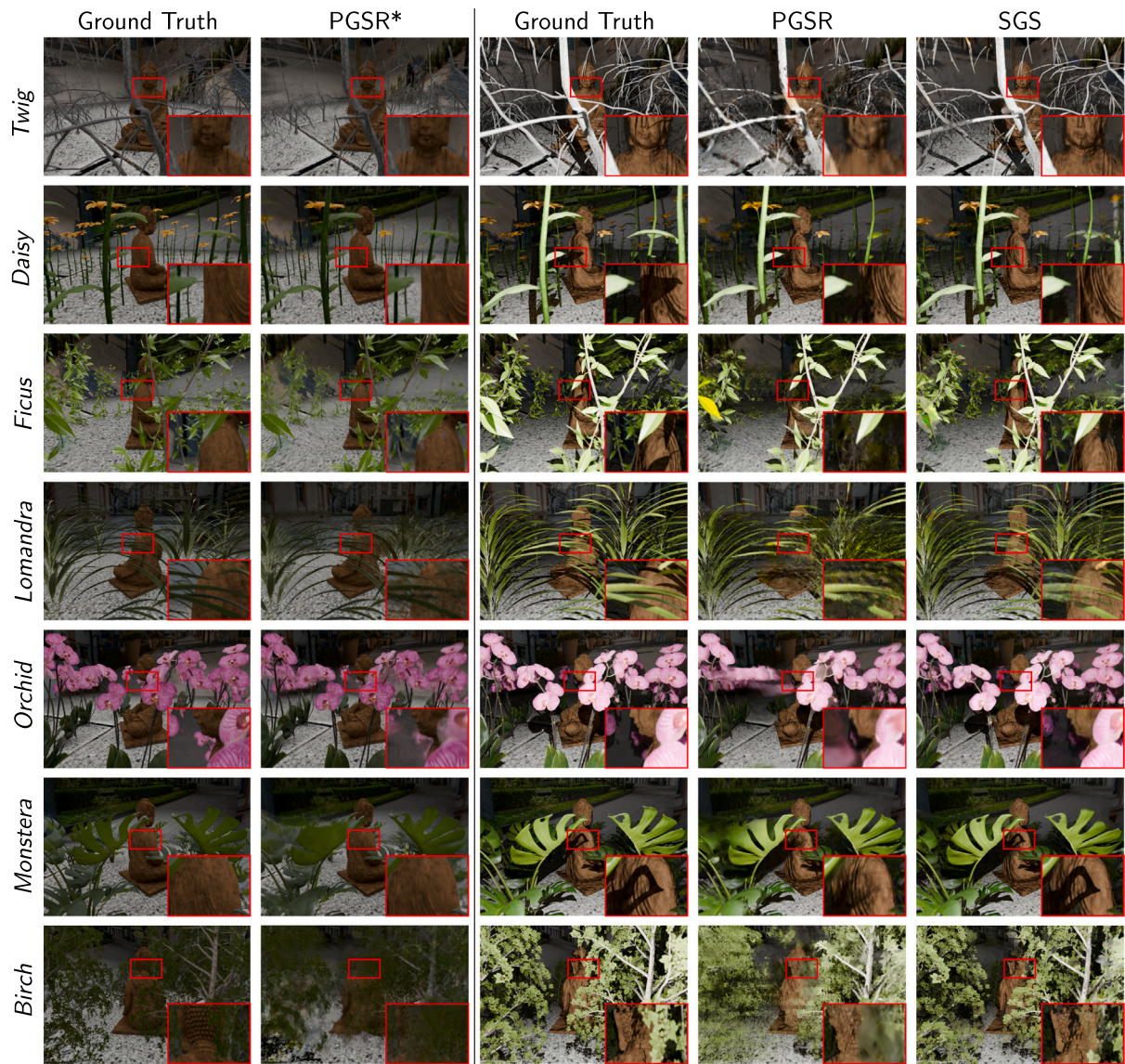


Fig. 8. Novel views. Even without shadows, PGSR* experiences loss of detail and reduced image quality, challenged by the intricate geometry of vegetation occlusions. While PGSR creates big black floaters baking the shadows into the color, SGS faithfully preserves geometric details without blurring the underlying geometry, disentangling shadows from scene geometry.

Novel Views Challenged by the delicate foliage, parts of the vegetation in front of and behind the object fail to be reconstructed or appear smeared in all scenes, leading to less distinct object reconstruction and blurring artifacts observed in *Twig*, *Lomandra* and *Birch* and slightly erroneous highlights visible in *Monstera* (Fig. 8). Due to the band-limited nature of low-degree SH and lack of an explicit occlusion model, PGSR* tries to represent these thin occlusions using the same SH features as the background. If the occlusions are too thin to be consistently seen as a solid surface from different viewpoints, the method treats it as a high-frequency noise and effectively filters it out during densification. Although the RSNR is highest among the methods, the structural and perceptual similarity which better align with human perception are however lower than SGS (Table 4).

5.2. Geometric reconstruction with shadows

3D Reconstruction Under direct illumination, PGSR point cloud reconstruction exhibit a substantial amount of floating artifacts around the object across all scenes (Fig. 4). The RMSE jumps significantly, especially in *Twig*, *Ficus*, *Lomandra* and *Birch* (Table 2). This highlights

a fundamental limitation of GS methods: the entanglement of geometry and illumination as it interprets shadows as physical, dark-colored surface textures. Thus, the algorithm is forced to bake cast shadows directly into the color of the Gaussians. To satisfy the multi-view photometric consistency, GS overfits by physically warping the underlying geometry and hallucinating structural artifacts and planar surfaces to recreate the shadow's appearance. Filtering out the points within 5 mm shows good geometric reconstruction (Fig. 5) preserving object's complex geometry (spikes on the head) and edges (lines on the clothes). Nevertheless, the errors are highest with RMSE values, compared to PGSR* and SGS (Fig. 5). Although the completeness decreases with more shadows, it is worth noting that PGSR generates more object points compared to PGSR* (18 478 vs. 17 364), thus reaches slightly higher completeness of 54.98% compared to 53.44% (Table 5). Cast shadows trigger aggressive Gaussian densification heuristics, leading to a higher point count and a marginal increase in completeness, at the same time sacrificing the accuracy. The holes and surface artifacts in the meshes correspond to the floaters in the point clouds, lowest among the methods.

In contrast, SGS point clouds show realistic representation and can capture the overall object geometry completely while being able to capture complex geometric details, e.g. the spikes on the head with more even point distribution. The geometric accuracy and completeness decline correspondingly to the shadow coverage namely, starting from *Twig* consisting of thin shadows from tree branches which reaches highest completeness with 81.06% to *Birch* representing tree with dense foliage and lowest point coverage of 36.63%. Overall, SGS demonstrates a substantial and consistent improvement over PGSR across all scenes in 3D and 2D metrics with massive improvement in accuracy due to the reduction of floating artifacts by explicitly separating shadows from underlying geometry, while simultaneously increasing the object point coverage. It is also worth noting that it outperforms PGSR* without shadows. For instance, in *Ficus* SGS brings down RMSE from 10.61 mm to 5.72 mm and in *Lomandra* from 69.32 to 17.50 mm for around 4× geometric improvement. SGS is also point efficient. On average, it generates a more compact point cloud and despite having fewer points, sees an increase in completeness over PGSR and it is also better than PGSR* without shadows, allocating accurate surface object points guided by the loss which directs increased primitive density towards recovering fine-grained details. Moreover, the meshes are watertight and smooth for an accurate surface reconstruction (Fig. 6 and Table 3). However *Lomandra* and *Monstera* exhibit surface artifacts along with gaps in the left shoulder of the object in *Orchid* caused by the big elongated shadows occupying large image parts with low pixel variance. Due to the floating artifacts in the point cloud, the object surface is poorly reconstructed in *Birch* scene which imposes biggest challenges for both methods without and with shadows. While SGS introduces a slight computational overhead of 4 min due to the concurrent generation of two radiance states and the execution of the bimodal selection loss, this cost is offset by the significantly higher fidelity of the reconstructed geometry and the successful disentanglement of illumination effects.

Novel Views SH only account for color changes relative to the viewing direction under constant light and are unable to represent additional pixel variations caused by illumination effects. This induces the uncertainty of a Gaussian having different color values along a ray preventing disentanglement of shadows from the underlying geometry. Thus, GS experiences geometric blurring of the vegetation but more importantly the object behind occlusions and cast shadows (*Twig*, *Ficus*, *Lomandra*, *Orchid* and *Birch*) failing to approximate the intricate cast shadows from the vegetation occlusions in *Daisy* and *Monstera* (Fig. 8). Analog to this, the rendering metrics are reasonably better than PGSR*, but far below SGS (Table 4).

For SGS, the novel views support the 3D geometric reconstruction showing strong photometric consistency with the input images. The spatial arrangement of the scene is reliably reconstructed preserving complex vegetation details such as small leaves and thin branches and intricate cast shadows. Object's facial features (*Twig* and *Orchid*), edges (clothes lines in *Daisy* and *Monstera*), homogeneous areas (shoulders visible in *Ficus* and *Lomandra*) and complex geometry (head spikes in *Twig* and *Birch*) are represented with high fidelity. Subsequently, SGS achieves superior rendering metrics in SSIM and LPIPS which best aligns with human perception. When a shadow intersects a surface, SGS absorbs the photometric variance into the Shadow Factor rather than warping the underlying Gaussian structure. This prevents geometric hallucination, maintaining strong balance between image quality and geometric accuracy.

6. Discussion

In Section 6.1 we analyze the geometric and surface reconstruction without and with shadows, followed by Section 6.2 where we discuss the rendered novel views. We provide additional experiments and results on non-transparent occlusions (Section 6.3) to assess the robustness, application and limitations of SGS.

6.1. 3D reconstruction

Although very noisy and with lowest accuracy scores, PGSR* in *Birch* exhibits highest number of object points of 49 226 and highest completeness of 75.98% (Table 5). For the same scene with shadows, PGSR and SGS show higher number of object points 17 705 and 22 766 respectively compared to *Orchid* and *Monstera*. We argue that this is due to the geometry of the leaves which are small with many tiny gaps. As the camera moves, different parts of the object are revealed through different holes in between the leaves. This gives the optimizer more gradient information to densify and fill in the object geometry behind. On the other hand, the petals in *Orchid* and the leaves in *Monstera* are large, flat surfaces. They tend to cover the same parts of the object consistently across multiple frames. If a certain object part is never seen, the Gaussians there will never receive the gradients required to split or clone, leaving that area sparse or empty. Moreover, the number of points does not necessarily imply a higher completeness because they do not conform well with object's geometry and are unevenly distributed on the outer surface (Fig. 5). The visible gaps in the lap and top of the head (Fig. 6) in the meshes which are not reconstructed, arise from the capturing trajectory which is circular around the object instead of hemispheric, as the goal is evaluating the geometry behind cast shadows from occlusions which are always in front of the object.

SGS improves average point coverage behind shadows with a more even point distribution on the object outer surface to 62.86% opposed to 54.98%, however the GS point clouds are still quite sparse. Instead of discrete points, the scene is reconstructed with a limited number of anisotropic Gaussians, which are added only where they improve the rendering loss. Each Gaussian has a spatial covariance, meaning it represents a small volume, not a point. The geometry as point cloud is approximated only through the center of each Gaussian, neglecting geometric properties such as rotation and scale, which are critical for effectively representing areas of the scene.

Notably, SGS creates less floating artifacts around the object (Fig. 4), improves the overall completeness and leads to better renderings than PGSR* without shadows (Fig. 8 and Table 4). While GS methods often rely on shadow-free environments to avoid photometric artifacts, SGS demonstrates that cast shadows can be leveraged as structural cues for geometric refinement. By separating the radiance into intrinsic color and extrinsic shadows, our bimodal selection loss effectively resolves the ambiguity between basic color and illumination variation. This isolation of illumination residuals prevents baking of shadows into the SH, which typically results in volumetric blurring and incomplete geometry. Consequently, the decoupled gradient flow enables a significantly lower densification threshold, facilitating the recovery of high-frequency surface details and improving structural completeness. Our results indicate that treating shadows as an explicit physical component rather than a photometric error, can lead to superior geometric fidelity and novel-view consistency.

6.2. Novel views

We observe localized radiance inconsistencies or discoloration primarily in regions of high geometric complexity, especially visible on the neck of the object in *Twig*, right shoulder between the leaves in *Lomandra* and spikes on the head in *Birch* (Fig. 8). Because our bimodal selection loss performs a discrete selection between the *Basic* and *Shadow* state, marginal differences in the residual error can lead to rapid spatial switching of the selection mask in the absence of a total variation regularizer or spatial smoothing on the shadow factor. Additionally, the removal of the standard 0.5 color bias necessitates absolute radiance optimization; in under-constrained thin structures, the resulting zero-clipping of SH further worsens these visual discontinuities during image composition. While a naive spatial smoothness constraint could mitigate the rapid per-pixel changes, it would inherently blur the naturally discrete boundaries of shadow regions. Instead,

Table 6

Rendering metrics. SGS achieves significantly better novel view synthesis quality compared to PGSR and stronger perceptual fidelity than PGSR* trained on shadow-free images. The **first** and **second** best results are highlighted.

Scene	Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
<i>Fence</i>	PGSR*	26.45	0.805	0.298
	PGSR	19.70	0.689	0.371
	SGS	22.41	0.783	0.293
<i>Urban</i>	PGSR*	27.00	0.842	0.325
	PGSR	19.56	0.743	0.374
	SGS	23.31	0.838	0.274

a more faithful alternative would be to enforce a geometric continuity constraint that penalizes isolated, small-scale shadow regions, operating under the assumption that both the occlusions and the geometry behind are continuous. Also, transitioning from a binary to a softmax selection during the early stages of training would allow both states to contribute and converge to a shared basic color. While the observed photometric artifacts are primarily localized to the radiance field, they introduce high-frequency noise into the gradient flow, which can subtly affect geometric stability. However, we argue that the planar geometric constraints provide sufficient regularization to maintain surface integrity, ensuring that these photometric inconsistencies do not lead to significant geometric degradation.

We also remark that some cast shadows are not rendered, instead the object geometry behind the shadow is reconstructed from the *Basic* state (thin branches on the face in *Twig*, leaf in *Daisy* and *Ficus* and elongated shadows on the upper shoulder in *Lomandra*). This occurs when a Gaussian's Shadow Factor remains static at its fully illuminated state, even when representing a shadow region in the ground truth. If the selection mask erroneously favors the *Basic* state across multiple views during early training, the gradient does not flow to the *Shadow* state. Thus, SGS subsequently bakes the shadow component into the intrinsic SH coefficients resulting in a well-reconstructed geometry that lacks distinct shadow casting.

6.3. Additional non-transparent occlusions

We generate two scenes tackling other non-transparent occlusions, namely *Fence* consisting of different type of metal bars, while *Urban* describes a municipal environment. We use the same poses as for the vegetation scenes. In line with the vegetation occlusions, SGS reduces the number of floaters (Fig. 9) resulting in a smoother mesh (Fig. 10), although some surface artifacts especially visible on the right object side are present. In addition, it achieves higher completeness compared to PGSR and PGSR* with 70.02% and 38.16% for *Fence* and *Urban* respectively (Table 8). However, the *Urban* scene proves challenging because it consists of big occlusions and subsequently shadows with low pixel variance (similar to *Monstera*; Figs. 4, 6, Tables 2 and 3). SGS relies on high-frequency pixel variance across the image sequence to provide the necessary multi-view supervision for the shadow factor. When shadows comprise same image pixels or are spatially static, the optimization of the product $F_i \cdot c_i$ becomes under-constrained, as it cannot distinguish between basic color (surface reflectance) and extrinsic illumination variation. Conversely, when a pixel transitions between *Basic* and *Shadow* state in different viewpoints, the bimodal selection loss effectively disentangles these components by anchoring the intrinsic SH features to the corresponding state. SGS achieves best rendering fidelity preserving geometric details without blurring the underlying geometry with best LPIPS of 0.293 and 0.274 for *Fence* and *Urban* (Table 6), while PGSR creates splotchy black spots and floaters behind shadows (Fig. 11 and Table 7).

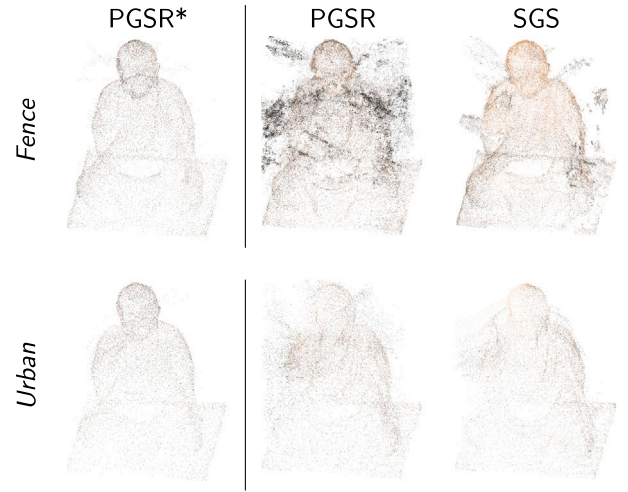


Fig. 9. Point cloud reconstruction for additional non-transparent occlusions. PGSR* without shadows reliably reconstructs the object, while shadows cause floaters in PGSR which SGS successfully reduces.

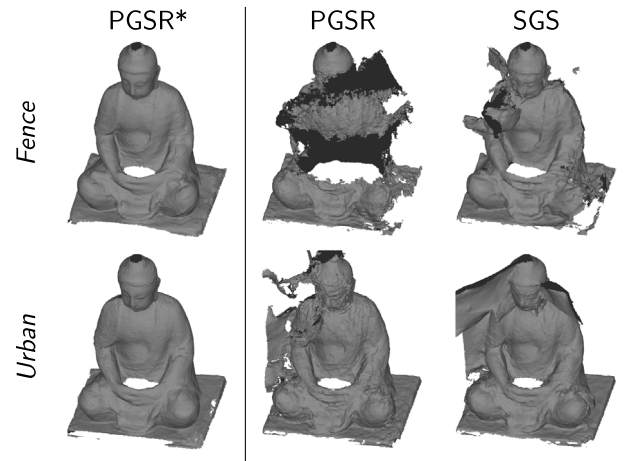


Fig. 10. PGSR* shows smooth mesh reconstruction behind occlusions without shadows. Although SGS better captures surface representation in *Fence*, due to the floaters in the point cloud experiences significant surface artifacts on the right object side, especially in *Urban*.

Table 7

Mesh evaluation. Although SGS shows better correspondence with the reference in *Fence*, challenged by the spatially static shadows with low pixel variance creating floating artifacts, in *Urban* exhibits poor surface representation. The **first** and **second** best results are highlighted.

Scene	Method	Accuracy (mm)			Time (min) $\downarrow$
		Mean $\downarrow$	SD $\downarrow$	RMSE $\downarrow$	
<i>Fence</i>	PGSR*	0.71	3.37	3.44	21
	PGSR	14.30	35.31	38.97	28
	SGS	11.90	29.75	32.04	30
<i>Urban</i>	PGSR*	0.57	3.79	3.83	18
	PGSR	11.44	28.17	30.41	18
	SGS	25.99	50.25	56.57	22

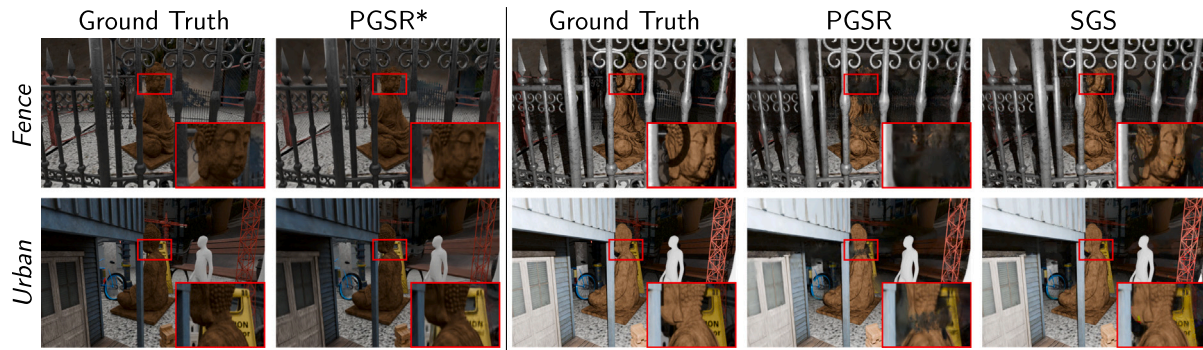


Fig. 11. Novel views. SGS captures finer details and produces more accurate renderings, with higher visual quality than PGSR* producing splotchy black spots and floaters behind shadows. The left side refers to shadow-free renderings.

Table 8

Quantitative cloud-to-mesh evaluation. SGS achieves highest completeness in both scenes, however challenged by the big occlusions and subsequently shadows with low pixel variance in *Urban* sacrifices geometric accuracy. The **first** and **second** best results are highlighted.

Scene	Method	Accuracy (mm)			Completeness (%)	
		Mean ↓	SD ↓	RMSE ↓	Npts	Cpl ↑
Fence	PGSR*	0.45	10.32	10.33	15 209	51.37
	PGSR	58.27	68.05	89.59	71 233	61.58
	SGS	14.36	35.03	37.86	49 665	70.02
Urban	PGSR*	1.38	14.71	14.78	9 903	36.83
	PGSR	15.81	38.04	41.20	16 515	36.41
	SGS	17.16	45.70	48.81	14 361	38.16

7. Conclusion

We propose Shadow-Aware Gaussian Splatting (SGS), a novel GS-based framework that successfully disentangles intrinsic radiance from extrinsic illumination. By leveraging a dual-pass rendering strategy coupled with a bimodal selection loss, SGS effectively unbakes shadows from scene geometry, facilitating high-fidelity geometric reconstruction. We demonstrate that SGS accurately recovers geometry and appearance even under intricate cast shadows, significantly reducing floating artifacts and improving surface representation across challenging datasets. Moreover, it achieves a good balance between geometric reconstruction and rendering quality, it does not depend on priors and can be easily integrated into various GS methods.

Limitations A notable limitation arises in scenes containing large, spatially static shadows with low pixel variance. In such cases, SGS may fail to distinguish between intrinsic surface reflectance and extrinsic illumination because the optimization relies on high-frequency photometric variance across multiple viewpoints to resolve the inherent ambiguity between color and shadow. Without sufficient multi-view supervision, it lacks the necessary cues to disentangle the two components. Consequently, in regions of low pixel variance across views, it may erroneously attribute light attenuation to the intrinsic color coefficients.

Future Work While SGS successfully disentangles shadows, a promising direction for future research is the integration of a physically-based Bidirectional Reflectance Distribution Function (BRDF) (Sun et al., 2018; L. Zhang et al., 2025; Wu et al., 2026; W. Zhang et al., 2025; Zhou et al., 2025). Replacing SHs with an explicit BRDF model would transition the framework from a radiance-based to a reflectance-based representation enabling to perform dynamic relighting, allowing for the insertion of novel light sources while maintaining consistent material properties like specular highlights (Lanvin et al., 2026).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge support by the KIT-Publication Fund of the Karlsruhe Institute of Technology.

References

- AG Stuart, L., Morton, A., Stavness, I., Pound, M.P., 2025. 3DGS-to-PC: 3D Gaussian splatting to dense point clouds. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3730–3739.
- Bai, H., Zhu, J., Jiang, S., Huang, W., Lu, T., Li, Y., Guo, J., Fu, R., Guo, Y., Chen, L., 2025. Gare: Relightable 3d gaussian splatting for outdoor scenes from unconstrained photo collections. arXiv preprint arXiv:2507.20512.
- Chen, H., Li, C., Lee, G.H., 2023. Neusg: Neural implicit surface reconstruction with 3d gaussian splatting guidance. arXiv preprint arXiv:2312.00846.
- Chen, D., Li, H., Ye, W., Wang, Y., Xie, W., Zhai, S., Wang, N., Liu, H., Bao, H., Zhang, G., 2024. Pgsr: Planar-based gaussian splatting for efficient and high-fidelity surface reconstruction. IEEE Trans. Vis. Comput. Graphics 31 (9), 6100–6111.
- Chen, H., Lin, Z., Zhang, J., 2024. Gi-gs: Global illumination decomposition on gaussian splatting for inverse rendering. arXiv preprint arXiv:2410.02619.
- Chen, D., Zhang, P., Feldmann, I., Schreer, O., Eisert, P., 2023. Recovering fine details for neural implicit surface reconstruction. In: 2023 IEEE/CVF Winter Conference on Applications of Computer Vision. WACV, IEEE, pp. 4319–4328.
- Dai, P., Xu, J., Xie, W., Liu, X., Wang, H., Xu, W., 2024. High-quality surface reconstruction using gaussian surfels. In: ACM SIGGRAPH 2024 Conference Papers. pp. 1–11.
- Feng, W., Ye, K., Zhang, Q., Zhang, Q., Li, N., 2025. 2D Gaussian splatting for outdoor scene decomposition and relighting. In: Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence. pp. 990–998.
- Guédon, A., Lepetit, V., 2024. Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5354–5363.
- Hong, Y., Niu, L., Zhang, J., 2022. Shadow generation for composite image in real-world scenes. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 36, pp. 914–922.
- Hsieh, C.-Y., Ding, J.-J., 2024. ADSP: Advanced dataset for shadow processing, enabling visible occluders via synthesizing strategy. In: Proceedings of the Asian Conference on Computer Vision. pp. 1028–1045.
- Huang, B., Yu, Z., Chen, A., Geiger, A., Gao, S., 2024. 2D gaussian splatting for geometrically accurate radiance fields. In: ACM SIGGRAPH 2024 Conference Papers. pp. 1–11.
- Jäger, M., Hillemann, M., Jutzi, B., 2025. FeatureGS: Eigenvalue-feature optimization in 3D Gaussian splatting for geometrically accurate and artifact-reduced reconstruction. ISPRS Open J. Photogramm. Remote. Sens. 100100.
- Jäger, M., Jutzi, B., 2023. 3D density-gradient based edge detection on neural radiance fields (NERFS) for geometric reconstruction. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. 48, 71–78.
- Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., et al., 2023. 3D gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. 42 (4), 139–1.
- Kulhanek, J., Peng, S., Kukulova, Z., Pollefeys, M., Sattler, T., 2024. WildGaussians: 3D Gaussian splatting in the wild. arXiv preprint arXiv:2407.08447.

- Lanvin, A., Hu, J., Lucas, S., Bousseau, A., Drettakis, G., 2026. Intrinsic decomposition and editing of 3D Gaussian splats. *Proc. ACM Comput. Graph. Interact. Tech.* 9 (1), 1–18.
- Li, Y., Lyu, C., Di, Y., Zhai, G., Lee, G.H., Tombari, F., 2024. Geogaussian: Geometry-aware gaussian splatting for scene rendering. *arXiv preprint arXiv:2403.11324*.
- Liang, Z., Zhang, Q., Feng, Y., Shan, Y., Jia, K., 2024. Gs-ir: 3d gaussian splatting for inverse rendering. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 21644–21653.
- Liu, R., Sun, D., Chen, M., Wang, Y., Feng, A., 2025. Deformable beta splatting. In: *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Papers*. pp. 1–11.
- Luo, F., Pan, H., Yang, X., Jiang, B., Liu, F., Huang, T., 2026. Shadowgs: Shadow-Aware 3D Gaussian splatting for satellite imagery. *arXiv preprint arXiv:2601.00939*.
- Martin-Brualla, R., Radwan, N., Sajjadi, M.S., Barron, J.T., Dosovitskiy, A., Duckworth, D., 2021. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 7210–7219.
- Masquil, E., Marí, R., Ehret, T., Meinhardt-Llopis, E., Musé, P., Facciolo, G., 2025. S-EO: A large-scale dataset for Geometry-Aware shadow detection in remote sensing applications. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 2383–2393.
- Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R., 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65 (1), 99–106.
- Petrovska, I., Jäger, M., Haitz, D., Jutzi, B., 2023. Geometric accuracy analysis between neural radiance fields (NeRFs) and terrestrial laser scanning (TLS). *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* 48, 153–159.
- Petrovska, I., Jutzi, B., 2025. Seeing beyond vegetation: A comparative occlusion analysis between Multi-View Stereo, neural radiance fields and Gaussian splatting for 3D reconstruction. *ISPRS Open J. Photogramm. Remote. Sens.* 16, 100089.
- Rudnev, V., Elgharib, M., Smith, W., Liu, L., Golyanik, V., Theobalt, C., 2022. Nerf for outdoor scene relighting. In: *European Conference on Computer Vision*. Springer, pp. 615–631.
- Sabour, S., Goli, L., Kopanas, G., Matthews, M., Lagun, D., Guibas, L., Jacobson, A., Fleet, D.J., Tagliasacchi, A., 2024. SpotlessSplats: Ignoring distractors in 3D Gaussian splatting. *arXiv preprint arXiv:2406.20055*.
- Sabour, S., Vora, S., Duckworth, D., Krasin, I., Fleet, D.J., Tagliasacchi, A., 2023. Robustnerf: Ignoring distractors with robust losses. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 20626–20636.
- Schonberger, J.L., Frahm, J.-M., 2016. Structure-from-motion revisited. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 4104–4113.
- Schönberger, J.L., Zheng, E., Frahm, J.-M., Pollefeys, M., 2016. Pixelwise view selection for unstructured multi-view stereo. In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11–14, 2016, Proceedings, Part III* 14. Springer, pp. 501–518.
- Schulte, F., Brezovsky, M., Günthner, A., Jutzi, B., Mandlbürger, G., Winiwarter, L., 2025. Simulation and validation of underwater scenes for two-media optical 3D reconstruction. *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* 48, 271–278.
- Snively, N., Seitz, S.M., Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. In: *ACM Siggraph 2006 Papers*. pp. 835–846.
- Song, Y., Lin, H., Lei, J., Liu, L., Daniilidis, K., 2024. Hdgs: Textured 2d gaussian splatting for enhanced scene rendering. *arXiv preprint arXiv:2412.01823*.
- Sun, Y., Wu, Y., Wilson, R., 2018. A review of thermal and optical characterisation of complex window systems and their building performance prediction. *Appl. Energy* 222, 729–747.
- Tao, X., Cao, J., Hong, Y., Niu, L., 2024. Shadow generation with decomposed mask prediction and attentive shadow filling. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 38, pp. 5198–5206.
- Ungermann, P., Ettenhofer, A., Niefner, M., Roessle, B., 2024. Robust 3d gaussian splatting for novel view synthesis in presence of distractors. In: *DAGM German Conference on Pattern Recognition*. Springer, pp. 153–167.
- Wan, Z., Richardt, C., Božič, A., Li, C., Rengarajan, V., Nam, S., Xiang, X., Li, T., Zhu, B., Ranjan, R., et al., 2023. Learning neural duplex radiance fields for real-time view synthesis. In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. CVPR, IEEE, pp. 8307–8316.
- Wang, Y., Chen, S., Yi, R., 2026. Sg-splatting: Accelerating 3d gaussian splatting with spherical gaussians. In: *ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing*. ICASSP, IEEE, pp. 8527–8531.
- Wang, T., Hu, X., Wang, Q., Heng, P.-A., Fu, C.-W., 2020. Instance shadow detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 1880–1889.
- Wang, Y., Klasson, M., Turkulainen, M., Wang, S., Kannala, J., Solin, A., 2025. DeSplat: Decomposed Gaussian splatting for distractor-free rendering. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 722–732.
- Wang, J., Liu, Y., Wang, P., Lin, C., Hou, J., Li, X., Komura, T., Wang, W., 2024. Gaussurf: Geometry-guided 3d gaussian splatting for surface reconstruction. *arXiv preprint arXiv:2411.19454*.
- Wu, C., Shi, X., Chu, Z., Huang, Y., 2026. ARS-GS: Anisotropic reflective spherical 3D Gaussian splatting. *J. Imaging* 12 (4), 170.
- Wu, T., Zhong, F., Tagliasacchi, A., Cole, F., Öztireli, C., 2022. D2NeRF: Self-Supervised decoupling of dynamic and static objects from a monocular video. *ArXiv abs/2205.15838*, (2022).
- Xu, C., Kerr, J., Kanazawa, A., 2024. Splatfacto-W: A nerfstudio implementation of Gaussian splatting for unconstrained photo collections. *arXiv preprint arXiv:2407.12306*.
- Yu, Z., Sattler, T., Geiger, A., 2024. Gaussian opacity fields: Efficient adaptive surface reconstruction in unbounded scenes. *ACM Trans. Graph. (ToG)* 43 (6), 1–13.
- Zhang, L., Rupnik, E., Nguyen, T.D., Jacquemoud, S., Klinger, Y., 2025. BRDF-NeRF: Neural radiance fields with optical satellite images and BRDF modelling. *Int. J. Appl. Earth Obs. Geoinf.* 143, 104747.
- Zhang, W., Tang, J., Zhang, W., Fang, Y., Liu, Y.-S., Han, Z., 2025. MaterialRefGS: Reflective gaussian splatting with multi-view consistent material inference. *arXiv preprint arXiv:2510.11387*.
- Zheng, I., Tang, G., Doronin, A., Teal, P., Zhang, F.-L., 2026. SSD-GS: Scattering and shadow decomposition for relightable 3D Gaussian splatting. *arXiv preprint arXiv:2604.13333*.
- Zhou, Y., Zhang, F., Wang, Z., Zhang, L., 2025. RTR-GS: 3D Gaussian splatting for inverse rendering with radiance transfer and reflection. In: *Proceedings of the 33rd ACM International Conference on Multimedia*. pp. 6888–6897.