2D Gaussian Splatting for Outdoor Scene Decomposition and Relighting

Wei Feng, **Kangrui Ye**, **Qi Zhang***, **Qian Zhang** and **Nan Li**
Tianjin University
{wfeng, yekangrui, qizhang118, qianz, linan94}@tju.edu.cn

Abstract

Gaussian splatting techniques have recently revolutionized outdoor scene decomposition and relighting through multi-view images. However, achieving high rendering quality still requires a fixed lighting condition among all input views, which is costly or even impractical to capture in outdoor scenes. In this paper, we propose outdoor scene decomposition and relighting with 2D Gaussian splatting (OSDR-GS), a novel inverse rendering strategy under outdoor changing and unknown lighting conditions. Firstly, we present a lighting-based group learning framework that categorizes input images into multiple lighting groups, to learn the separate lighting from each group individually. Secondly, OSDR-GS introduces a fine-grained outdoor lighting component to represent sun-light and sky-light, respectively, which are also adjusted via the correlative exposure factors adaptively. Finally, we construct a visibility-driven shadow module to characterize the nuanced interplay of light and occlusion realistically, for eliminating the uncertainty of dark pixels on lighting-based group learning. Extensive experiments on multiple challenging outdoor datasets validate the effectiveness of OSDR-GS, which achieves the state-of-the-art performance in changing lighting scene inverse rendering.

1 Introduction

Inverse rendering is a fundamental task in computer vision and graphics. It aims to decompose the 3D scene into intrinsic properties (e.g., material, geometry, and environment lighting) only from multi-view images, then allowing all properties can be manipulated independently.

Recently, advanced novel view synthesis techniques like neural radiance field (NeRF) [Mildenhall *et al.*, 2021] provide a new approach for inverse rendering. Several NeRF-derived researches decompose scenes by multiple multi-layer perceptions (MLP) to learn intrinsic properties separately [Srinivasan *et al.*, 2021; Zhang *et al.*, 2021b; Rudnev *et al.*, 2022]. However, the execution of MLP networks leads to high computation, which significantly suffers from rendering efficiency. Recent 3D Gaussian splatting (3DGS) [Kerbl *et al.*, 2023] has exhibited outstanding efficient performance for novel view synthesis. It utilizes a set of Gaussian primitives (e.g., position and covariance) to represent scene explicitly and employs GPU-accelerated rasterization for real-time rendering. Several GS-derived researches attempt to extend Gaussian primitives with normal, material, or other properties, which play the same role as multi-MLPs of NeRF and fulfill the scene efficient inverse rendering [Liang *et al.*, 2024; Gao *et al.*, 2023; Bi *et al.*, 2024]. However, these methods are required to receive multi-view images with the fixed lighting conditions. Notably, it is costly and even impractical to control a fixed lighting in outdoor scenes (i.e., the changing sun or weather) during the dense view inputs capturing.

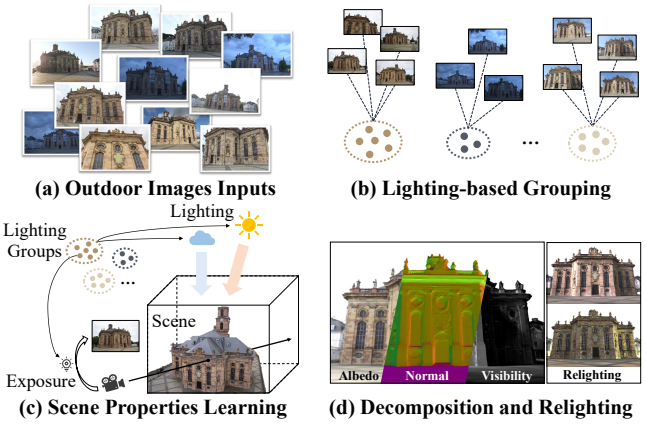

Figure 1: (a) OSDR-GS takes multi-view outdoor images with changing lighting conditions as input. (b) It clusters images into several groups by their lighting cues, and (c) learns scene properties (e.g., geometry, material, and lighting) within lighting groups, (d) enabling the scene decomposition and relighting.

The main challenges of outdoor scene decomposition and relighting with Gaussian splatting are: (1) Images are captured under changing lighting conditions, causing the inconsistency performance of scene's visual appearance; (2) Lighting outdoors is highly complex and changing, leading to the wide variety of cast shadows. The first challenge causes the same coordinate of the scene to exhibit different pixel colors in the same viewpoints. It induces the uncertainty of spheri-

*Corresponding author.

cal harmonic (SH) coefficients for each Gaussian to perform varying colors, which causes inaccurate scene inverse rendering ultimately. The second challenge is the sophisticated mix of multiple light sources in outdoor scene, adding to the complexity of lighting conditions. The relative positions of objects within the scene produce intricate cast shadows, making it more difficult to simulate scene details precisely. Specifically, the mentioned challenges are common and inevitable during the input views capturing in outdoor scenes.

In this work, we propose OSDR-GS, a novel method for **O**utdoor **S**cene **D**ecomposition and **R**elighting leveraging 2D **G**aussian **S**platting. OSDR-GS consists of three key components: *Lighting-based Group Learning*, *Fine-grained Outdoor Lighting Modeling*, and *Visibility-driven Shadow Module*, to decompose the 3D scene into geometry, material, and lighting from multi-view images captured under changing and unknown lighting conditions. For the lighting-based group learning, we cluster images into multiple lighting groups by the elaborately distilled lighting cues, where each image within the same group is considered to have similar lighting and share the same lighting parameters. Specifically, the misgrouped images will be updated regularly according to the presented groups refinement strategy. For outdoor lighting modeling, we adopt a fine-grained parameterized scheme to simulate sun-light and sky-light separately. Then we assign the exposure factors to the input images for compensating the discrepancies adaptively of exposure settings for each capture. For visibility-driven shadow module, we measure the visibility of incident light in the scene to synthesize realistic cast shadows, and formulate an efficient regularization scheme based on the physical of light propagation to optimize more generalized scene shadows.

Extensive experiments on various challenging outdoor scenes, including NeRF-OSR [Rudnev *et al.*, 2022] and Mip-NeRF 360 datasets [Barron *et al.*, 2022] with several state-of-the-art baselines, demonstrate the effectiveness of OSDR-GS. The main contributions are summarized as follows:

- We propose OSDR-GS, a novel scene decomposition and relighting method with Gaussian splatting from multi-view images captured in changing lighting.
- We present a lighting-based group learning framework that clusters input images into multiple lighting groups, and performs group-specific learning to optimize the lighting from each group.
- We introduce a fine-grained lighting scheme with a visibility-based shadow synthesis to measure the visibility of incident light, from arbitrary directions for simulating nuanced interplay of light and shadow.

2 Related Work

3D Scene Representation. In recent years, 3D scene representation technology advanced rapidly [Gao *et al.*, 2022]. Several methods shown impressive performance in high-quality novel view synthesis [Barron *et al.*, 2021; Barron *et al.*, 2022; Chen *et al.*, 2022]. NeRF [Mildenhall *et al.*, 2021] utilizes Multi Layer Perceptron (MLP) to model color and density of sampled 3D coordinates to represent scene implicitly. A number of researchers have endeavored to improve the performance of original NeRF in numerous dimensions. Several approaches work towards improving the quality of image rendering using enhanced neural network architectures [Zhang *et al.*, 2020; Barron *et al.*, 2021]. Some researchers seek to reconstruct scenes without the specific initialized camera pose prior by optimizing poses and MLP simultaneously [Lin *et al.*, 2021; Bian *et al.*, 2023; Zhang *et al.*, 2024]. However, the execution of MLP networks lead to high computation, which significantly suffers from efficiency in rendering [Chen and Wang, 2024]. Recently, 3D Gaussian Splatting [Kerbl *et al.*, 2023] significantly reduce rendering time while maintaining quality comparable to NeRF. It utilize a set of learnable 3D Gaussian primitive to explicitly model scene's color and render images. 3DGS employ a unique tile-based rasterizer tailored for Gaussian splatting, which facilitates real-time rendering for novel view synthesis [Yu *et al.*, 2024; Sun *et al.*, 2024; Xie *et al.*, 2024; Lin *et al.*, 2024]. While the volumetric radiance representation of 3D Gaussians conflicts with the thin nature of surfaces in most scenes. 2D Gaussian Splatting (2DGS) [Huang *et al.*, 2024] propose to utilize a set of plane 2D Gaussian disks to represent surface of scene, which is benefit for high-quality surface reconstruction.

Different from the above methods, we represent the 3D scene by material, visibility and lighting rather than just color, which enables scene decomposition and relighting.

Scene Inverse Rendering. Scene inverse rendering is a fundamental task in computer vision and graphics, which aims to separate scene's geometry, material and lighting from multi-view images. Previous methods [Yu and Smith, 2019; Yu *et al.*, 2020; Xu *et al.*, 2018] utilize several convolution neural networks (CNN) to predict normal, albedo and lighting of an input image. With the rapid development of NeRF, some methods take as input multi-view images captured under known lighting condition and model scene using additional MLPs [Bi *et al.*, 2020; Srinivasan *et al.*, 2021; Xu *et al.*, 2023; Li *et al.*, 2023]. While these methods rely on known lighting, they can not work in most natural scenes. To deal with this, several methods model environmental illumination to decompose and relighting scene under unknown lighting [Zhang *et al.*, 2021b; Zhang *et al.*, 2022; Zhang *et al.*, 2021a; Wu *et al.*, 2023; Boss *et al.*, 2021; Yang *et al.*, 2022]. Other methods attempt to relight outdoor scene by modeling scene properties and learning irrelative environmental light in each views [Rudnev *et al.*, 2022; Sun *et al.*, 2023; Gardner *et al.*, 2024]. Although these NeRF-based methods show infusive performance in outdoor scene inverse rendering, they still suffer from terrible efficient during training and rendering. Recently, 3D Gaussian splatting based method provide novel approach for inverse rendering [Guo *et al.*, 2024; Chen *et al.*, 2024]. GS-IR [Liang *et al.*, 2024] proposes a 3DGS-based inverse rendering framework to deduce the physical attributes of scene. GS^3 [Bi *et al.*, 2024] presents a spatial and angular Gaussian based representation and a triple splatting process for indoor scene relighting. R3DG [Gao *et al.*, 2023] proposes a material and lighting decomposition scheme by introducing additional components for each 3D Gaussian. However, these methods struggle with

outdoor scenes especially those under changing and unknown lighting conditions.

In this work, we propose OSDR-GS that clusters input images into multiple lighting groups to learn lighting within groups, enabling the relighting of outdoor scene.

3 Preliminaries

2D Gaussian Splatting. 2D Gaussian Splatting [Huang *et al.*, 2024] represents the scene by directly modeling the object's surface using 2D planar Gaussians, in contrast to 3DGS [Kerbl *et al.*, 2023], which employs anisotropic 3D Gaussian primitives. The 2D Gaussian disk defined on a local tangent plane is parameterized by its center $\mathbf{p}$, two orthogonal tangent vectors $\mathbf{t}_u$ and $\mathbf{t}_v$, and a scaling vector $\mathbf{S} = (s_u, s_v)$. The normal vector of Gaussian disk is natively represented as the cross product $\mathbf{t}_w = \mathbf{t}_u \times \mathbf{t}_v$. The orientation is arranged into a rotation matrix $\mathbf{R} = [\mathbf{t}_u, \mathbf{t}_v, \mathbf{t}_w]$, and the scaling factors are encoded in a 3×3 diagonal matrix $\mathbf{S} = [s_u, s_v, 0]$. The parameterization of a 2D Gaussian is as follows:

$$P(u, v) = \mathbf{p} + s_u \mathbf{t}_u u + s_v \mathbf{t}_v v = \mathbf{H}(u, v, 1, 1)^\top, \quad (1)$$

where $\mathbf{H} \in 4 \times 4$ is a homogeneous transformation matrix representing the geometry of the 2D Gaussian, uniquely determining a 2D Gaussian in space. Its inhomogeneous form is given by $\hat{\mathbf{H}} = [\mathbf{RS} \quad \mathbf{p}]$, where $\mathbf{RS} = [s_u \mathbf{t}_u \quad s_v \mathbf{t}_v \quad \mathbf{0}]$. The 2D Gaussian value at the point $\mathbf{u} = (u, v)$ in the uv space can be computed by standard 2D Gaussian function $\mathcal{G}(u, v)$.

The center $\mathbf{p}$, scaling (s_u, s_v), and rotation $(\mathbf{t}_u, \mathbf{t}_v)$ are learnable parameters. Similar to 3DGS, each 2D Gaussian primitive has opacity α and viewpoint-dependent color c parameterized by spherical harmonic coefficients. During rendering, it employs volume rendering to integrate alpha-weighted color from near to far:

$$\mathbf{c}(\mathbf{x}) = \sum_{i=1} c_i \alpha_i \mathcal{G}_i(\mathbf{u}(\mathbf{x})) \prod_{j=1}^{i-1} (1 - \alpha_j \mathcal{G}_j(\mathbf{u}(\mathbf{x}))), \quad (2)$$

where $\mathbf{x}$ represents a ray emitted from the camera, $\mathbf{u}(\mathbf{x})$ denotes the uv space coordinates of the intersection point where ray $\mathbf{x}$ passes through the plane of the Gaussian.

Lambertian Reflectance. Lambertian reflectance model is a classic and widely used model for approximating diffuse reflection on surfaces [Kajiya, 1986]. It uses diffuse albedo as the key parameter of the bidirectional reflectance distribution function (BRDF), with the assumption that the reflected radiance is proportional to the cosine of the angle between incident light direction and surface normal. The outgoing light $\mathbf{L}_\mathrm{o}$ at surface point $\mathbf{x}$ is formulated as:

$$\mathbf{L}_\mathrm{o}(\mathbf{x}) = \int_\Omega \frac{\mathbf{a}}{\pi} \mathbf{L}_\mathrm{i}(\mathbf{x}, \omega_i)(\mathbf{n} \cdot \omega_i) \, d\omega_i, \quad (3)$$

where $\mathbf{L}_\mathrm{i}(\mathbf{x}, \omega_i)$ refers to the incident light arriving at surface point $\mathbf{x}$ from the direction ω_i. $\mathbf{a}$ is the albedo at point $\mathbf{x}$, and $(\mathbf{n} \cdot \omega_i)$ is the cosine of the angle between incident light direction ω_i and the surface normal $\mathbf{n}$ at point $\mathbf{x}$.

4 Method

In this section, we introduce OSDR-GS, which is comprised of lighting-based group learning framework (Sec. 4.1), fine-grained outdoor lighting modeling (Sec. 4.2), and visibility-driven shadow module (Sec. 4.3). Eventually, we describe the rendering process and training framework (Sec. 4.4).

4.1 Lighting-based Group Learning

Different lighting conditions can result in varying appearances of multi-view images captured in outdoor scenes. Previous approaches [Rudnev *et al.*, 2022; Sun *et al.*, 2023] assume that these images have totally inconsistent and unrelated lighting conditions, processing each one independently. This will introduce additional lighting uncertainty of the scene since the lighting in different views may be related, which hinders the learning of accurate lighting. In this part, we aim to fully utilize the relationship among multi-view images. As shown in Fig. 2(a), we propose a lighting-based group learning framework that first cluster them into multiple lighting groups with similar lighting conditions, and then assign lighting parameters to each group to learn changing lighting.

Lighting Cues Acquisition. To cluster multi-view images with changing lighting conditions into different lighting groups, we propose an elaborately distilled approach to acquire the lighting cues from each image. The lighting cues is considered to represent lighting condition, and demonstrate a certain level of robustness to changes in viewpoint [Tian *et al.*, 2018]. We start the acquisition by a preprocessing that use pre-trained semantic segmentation model to filter out the background elements (e.g., sky). Next, we convert the image to the HSV (i.e., Hue, Saturation, and Value) color space, and extract the Hue channel. After that, we divide the Hue channel image into k equally spaced bins based on pixels' value, and calculate the frequency of pixels within each bin:

$$h_i = \sum_{p \in \mathcal{P}} \mathcal{I}_B(p), \quad B := \frac{255}{k}(i-1) \leq p < \frac{255}{k}(i), \quad (4)$$

where h_i denotes the frequency of the i_{th} bin ($i = 1, 2, \ldots, k$). $\mathcal{P}$ is the set of all pixel values in Hue channel image. $\mathcal{I}_B(p)$ is an indicator function that equals 1 when the pixel's value p satisfies condition B, and 0 otherwise. Subsequently, we normalize all h and concatenate them into:

$$\mathbf{f}_\mathrm{light} = [h'_1, h'_2, \ldots h'_k], \quad h'_i = \frac{h_i}{\sum_{j=1}^{k} h_j}, \quad (5)$$

where h' is the normalized frequency. Vector $\mathbf{f}_\mathrm{light} \in \mathbb{R}^{1 \times k}$ is the lighting cues of an input image. Since it is a statistical vector, $\mathbf{f}_\mathrm{light}$ demonstrates a certain level of robustness to changes in camera viewpoint, as shown in the left of Fig. 3. This enables it is applicable to images with different views.

Lighting Group Learning. Following the acquisition of lighting cues from each image, we apply the simple K-means clustering algorithm [MacQueen, 1967] using the Bhattacharyya distance [Bhattacharyya, 1943] as the metric to obtain several clusters, which represent lighting groups. Images within the same lighting group have similar lighting conditions, while those in different lighting groups exhibit

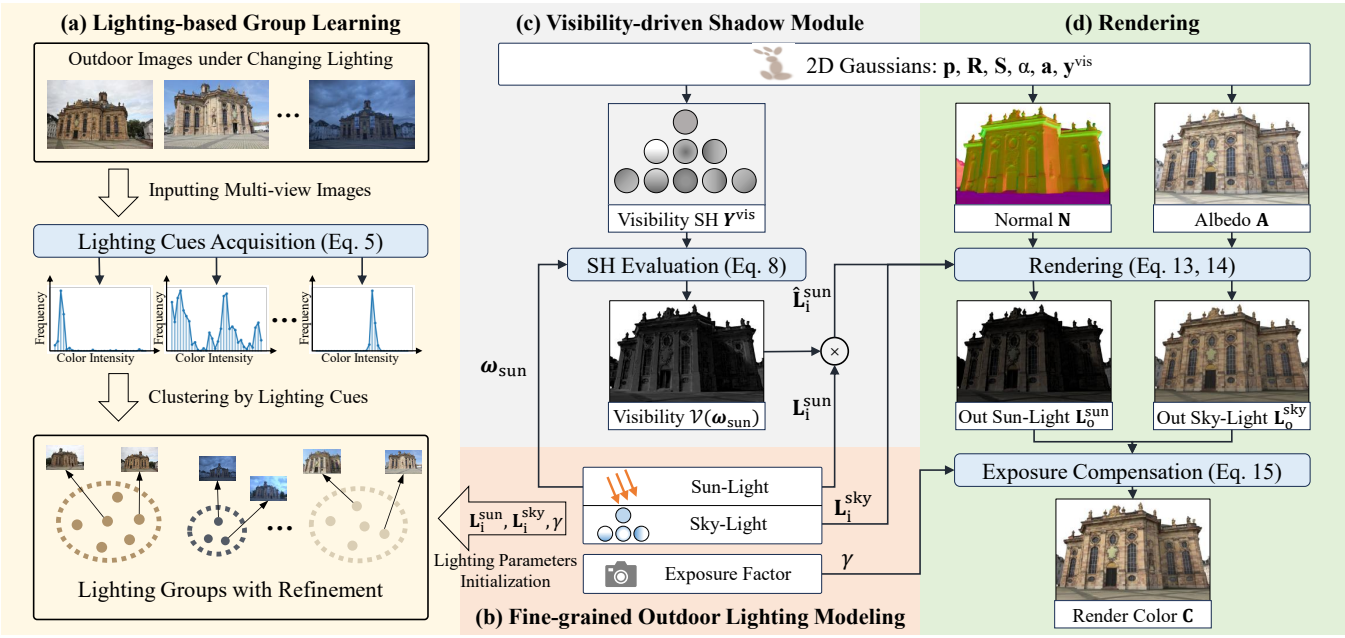

Figure 2: Our **OSDR-GS pipeline** for outdoor scene decomposition and relighting. Given a collection of outdoor images with changing and unknown lighting conditions, (a) we first cluster them into several lighting groups by elaborately distilled lighting cues, and (b) initialize lighting parameters for each group and each image to learn lighting condition of scene. (c) We also measure visibility in the scene for occluding incident light to produce realistic cast shadows. After that, (d) we perform physically based rendering to obtain render image.

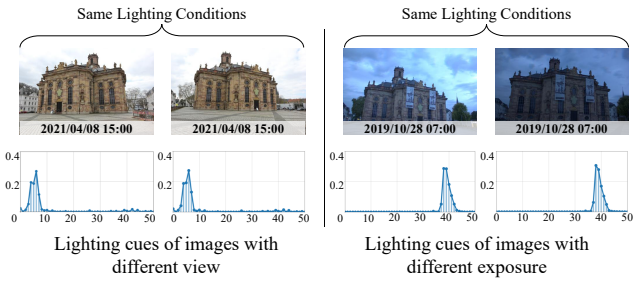

Figure 3: The visualization of lighting cues. We randomly select two sets of images with same lighting conditions from NeRF-OSR dataset. The left pair with different viewpoints exhibit nearly identical lighting cues. The right pair with similar views are different in exposure settings, which also exhibit highly consistent lighting cues. This shows that our lighting cues effectively reflect the lighting conditions in outdoor multi-view images.

significantly different lighting conditions. Next, we assign a set of learnable lighting parameters (detailed in Sec. 4.2) to each lighting group, which means images in the same group share the same lighting parameters. This provides the lighting consistency across different viewpoints. These lighting parameters are jointly optimized with scene's geometry and material through backpropagation via our carefully designed loss function (detailed in Sec. 4.4).

Refining Lighting Groups. Although our lighting grouping strategy is applicable to most situations, there are several misgrouped images inevitably. These misgrouped images will contribute distorted gradients to the current lighting group during optimizing, which disturbs the optimization process and results in inaccurate lighting learning. Inspired by [Feng *et al.*, 2015; Zhang *et al.*, 2018], we assess the deviation between image rendered under particular lighting and the ground truth by the structural similarity item $\mathrm{str}(\mathbf{I}_1, \mathbf{I}_2)$ from structural similarity index (SSIM) [Wang *et al.*, 2004] to solve this problem. We regularly iterate through all lighting groups for each image to find the group with the highest structural similarity $\mathrm{str}(\mathbf{I}_1, \mathbf{I}_2)$ as the optimal lighting group. This achieves the refinement of lighting groups.

4.2 Fine-grained Outdoor Lighting Modeling

Lighting Representation. Outdoor scenes are primarily illuminated by two types of light: direct sun-light and diffuse sky-light. The sun-light comes from the sun itself has strong directional quality. While the sky-light originates from the scattered light in atmosphere and diffuses across the entire sky. Inspired by [Sun *et al.*, 2023], we adopt a fine-grained approach to separately model these two light sources, as shown in Fig. 2(b). For the sun-light, we simply model it as a directional light source, parameterized by a learnable direction $\omega_{\text{sun}} \in \mathbb{R}^3$ and a learnable color-intensity $\mathbf{L}_{\text{i}}^{\text{sun}} \in \mathbb{R}^3$ For sky-light, we parameterize by first-order spherical harmonic (SH). It describes the global lighting distribution on the sphere by a set of basis functions $\mathcal{Y}_{l,m}$ and corresponding learnable coefficients $\mathrm{y}_{l,m}^{\text{sky}} \in \mathbb{R}$, where $l \in \{0, 1\}$, $-l \leq m \leq l$, and $l \in \mathbb{Z}$. The incident sky-light $\mathbf{L}_{\text{i}}^{\text{sky}} \in \mathbb{R}^3$ from direction $\omega_i \in \mathbb{R}^3$ can be described as:

$$\mathbf{L}_{\text{i}}^{\text{sky}}(\omega_i) = \sum_{l=0}^{deg} \sum_{m=-l}^{l} \mathrm{y}_{l,m}^{\text{sky}} \mathcal{Y}_{l,m}(\omega_i), \tag{6}$$

where deg is the degree of SH that equals 1 here. Notably, as we described in Sec. 4.1, we assign lighting parameters to each lighting group, including both sky-light and sun-light.

Adaptive Exposure Compensation. Since the lighting cues is performed in the HSV color space, it is also robust to brightness (typically due to differences in camera exposure settings), as shown in the right of Fig. 3. This causes that within one lighting group, there can be images with the same lighting condition but different brightness. To adaptively compensate each image brightness in the same lighting group, we employ an exposure compensation strategy as follows:

$$\mathbf{C}(\mathbf{x}) = \exp(\gamma)\mathbf{L}_{\mathrm{o}}(\mathbf{x}), \tag{7}$$

where $\mathbf{L}_{\mathrm{o}}(\mathbf{x})$ represents the outgoing light which comes from 3D point $\mathbf{x}$ in scene and arrives at the camera. $\gamma \in \mathbb{R}$ is a learnable exposure factors assigned to each image. $\mathbf{C}(\mathbf{x})$ is the color of pixel that corresponding to point $\mathbf{x}$.

4.3 Visibility-driven Shadow Module

Shadow Modeling. Synthesizing cast shadows is crucial for high-quality outdoor scene rendering, which not only determines the realism of rendering, but also impacts the accuracy of lighting groups refinement. In the real world, shadows are primarily caused by the occlusion of sun-light. To simulate this effect, we introduce a second-order learnable spherical harmonics (SH) coefficients $\mathrm{y}^{\mathrm{vis}}_{l,m} \in \mathbb{R}$ into each Gaussian primitive that measure the visibility of incident light (i.e., sun-light) and block it. The formula is shown as follows:

$$\mathcal{V}(\mathbf{x}, \omega_i) = \sum_{l=0}^{deg} \sum_{m=-l}^{l} Y^{\mathrm{vis}}_{l,m} \mathcal{Y}_{l,m}(\omega_i), \tag{8}$$

where $\mathcal{V}(\mathbf{x}, \omega_i) \in \mathbb{R}$ is the visibility that represents whether the light from direction ω_i is visible at surface point $\mathbf{x}$. It equals 0 if invisible (i.e., the light is occluded) and 1 if visible. deg is the degree of SH that equals 2 here. $Y^{\mathrm{vis}}_{l,m} \in \mathbb{R}^3$ is the spherical harmonic coefficient at surface point $\mathbf{x}$, which obtained by volume rendering using Eq. (2). Then, the occluded incident sun-light $\hat{\mathbf{L}}^{\mathrm{sun}}_{\mathrm{i}}$ can be formulated as:

$$\hat{\mathbf{L}}^{\mathrm{sun}}_{\mathrm{i}} = \mathbf{L}^{\mathrm{sun}}_{\mathrm{i}}\, \mathcal{V}(\mathbf{x}, \omega_{\mathrm{sun}}), \tag{9}$$

where ω_{sun} is the direction of incident sun-light $\mathbf{L}^{\mathrm{sun}}_{\mathrm{i}}$.

Visibility Regularization. In order to improve the generalization of the visibility, we introduce a negative normal loss function based on a physical light propagation: the lower hemisphere of the object surface normal is occluded. Thus we formulate the negative normal $\mathcal{L}_{\mathrm{nn}}$ as below:

$$\mathcal{L}_{\mathrm{nn}} = \frac{1}{p}\sum_{i=1}^{p} \|\mathcal{V}(\mathbf{x}_i, -\mathbf{N}_i + \epsilon)\|_2^2, \tag{10}$$

where p is the number of pixels in an image. $\mathbf{N}_i$ represent the surface normal of 3D point $\mathbf{x}_i$ corresponding to the i-th pixel, which is obtained by volume rendering from 2D Gaussians. $\epsilon \in \mathbb{R}^3$ is a random noise that added in negative normal to represent the direction of negative normal hemisphere.

Practically, the realistic cast shadow will accompany by the binary visibility $\mathcal{V}$. We present a binarization loss function

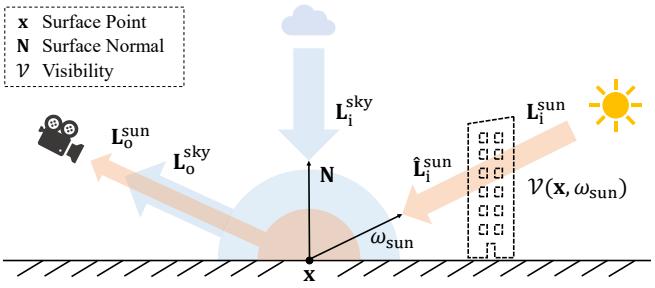

Figure 4: The illustration of rendering process. Our method renders by simulating the propagation of sun-light and sky-light.

$\mathcal{L}_{\mathrm{bin}}$ to enforce a binary visibility pattern, as shown in below:

$$\mathcal{L}_{\mathrm{bin}} = \frac{1}{p}\sum_{i=1}^{p} \min\left(\mathcal{F}(\mathcal{V}(\mathbf{x}_i, \omega_{\mathrm{sun}})), \mathcal{F}(\mathcal{V}(\mathbf{x}_i, \omega_{\mathrm{sun}}) - 1)\right), \tag{11}$$

where $\mathcal{F}(x) = 1 - \exp(-mx^2)$ is an empirical exponential-based function that use to adjust the strength of binarization. The other variables follow the explanation given in Eq. (10).

4.4 Rendering and Training

Physically Based Rendering. In order to decompose the scene, we introduce albedo $\mathbf{a} \in \mathbb{R}^3$ into each Gaussian to render albedo image $\mathbf{A}$ by volume rendering, so that we can employ physically based rendering model to synthesize image. We define the outgoing light $\mathbf{L}_{\mathrm{o}}$ at point $\mathbf{x}$ as follows:

$$\mathbf{L}_{\mathrm{o}}(\mathbf{x}) = \mathbf{L}^{\mathrm{sun}}_{\mathrm{o}}(\mathbf{x}) + \mathbf{L}^{\mathrm{sky}}_{\mathrm{o}}(\mathbf{x}), \tag{12}$$

where $\mathbf{L}^{\mathrm{sun}}_{\mathrm{o}}$ and $\mathbf{L}^{\mathrm{sky}}_{\mathrm{o}}$ represent the outgoing sun-light and outgoing sky-light, respectively. For the outgoing sky-light, we assume that it exclusively comes from the direction of surface normal $\mathbf{N}$, which we formulate it as:

$$\mathbf{L}^{\mathrm{sky}}_{\mathrm{o}}(\mathbf{x}) = \frac{\mathbf{A}}{\pi}\mathbf{L}^{\mathrm{sky}}_{\mathrm{i}}(\omega_i)\ (\mathbf{N} \cdot \omega_i), \tag{13}$$

where $\mathbf{L}^{\mathrm{sky}}_{\mathrm{i}}(\omega_i)$ is the incident sky-light from Eq. (6). ω_i is aligned with normal, which means the cosine item $(\mathbf{N} \cdot \omega_i)$ is equal to 1. For outgoing sun-light $\mathbf{L}^{\mathrm{sun}}_{\mathrm{o}}$, it is formulated as:

$$\mathbf{L}^{\mathrm{sun}}_{\mathrm{o}}(\mathbf{x}) = \frac{\mathbf{A}}{\pi}\hat{\mathbf{L}}^{\mathrm{sun}}_{\mathrm{i}}\ (\mathbf{N} \cdot \omega_{\mathrm{sun}}), \tag{14}$$

where $\hat{\mathbf{L}}^{\mathrm{sun}}_{\mathrm{i}}$ is the occluded sun-light from Eq. (9).

Overall, the aforementioned rendering process is illustrated in Fig. 4. After that, we extend Eq. (7) to adjust the exposure and obtain pixel color $\mathbf{C}$ using outgoing light $\mathbf{L}^{\mathrm{sun}}_{\mathrm{o}}$ and $\mathbf{L}^{\mathrm{sky}}_{\mathrm{o}}$:

$$\mathbf{C}\left(\mathbf{x}\right) = \exp(\gamma)(\mathbf{L}^{\mathrm{sky}}_{\mathrm{o}}(\mathbf{x}) + \mathbf{L}^{\mathrm{sun}}_{\mathrm{o}}(\mathbf{x}))\,. \tag{15}$$

Training Framework. We follow 2DGS [Huang *et al.*, 2024] to mainly employ color loss $\mathcal{L}_{\mathrm{c}}$ and normal loss $\mathcal{L}_{\mathrm{n}}$ to drive the training. To summarize, our loss function is formulated by combining several components, as expressed below:

$$\mathcal{L} = \mathcal{L}_{\mathrm{c}} + \lambda_1\mathcal{L}_{\mathrm{n}} + \lambda_2\mathcal{L}_{\mathrm{nn}} + \lambda_3\mathcal{L}_{\mathrm{bin}}, \tag{16}$$

where $\lambda_1 = 0.05$, $\lambda_2 = 0.01$ and $\lambda_3 = 0.001$ are predefined weighting hyperparameters for each loss terms. By minimizing Eq. (16), our approach returns the 3D scene representation and the optimized lighting parameters.

Method	Ludwigskirche			Staatstheater			Landwehrplatz		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
R3DG [Gao *et al.*, 2023]	18.11	0.600	0.430	15.37	0.632	0.412	15.60	0.650	0.378
NeRF-OSR [Rudnev *et al.*, 2022]	19.19	0.690	0.428	18.91	0.705	0.337	18.70	0.720	0.310
GS-IR [Liang *et al.*, 2024]	19.74	0.689	0.301	18.96	0.733	0.270	18.62	0.757	0.228
OSDR-GS (ours)	**21.27**	**0.723**	**0.296**	**19.51**	**0.759**	**0.263**	**19.62**	**0.795**	**0.216**

Table 1: Relighting Quantitative Comparison on NeRF-OSR Dataset.

Method	Bicycle			Garden			Stump		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
R3DG [Gao *et al.*, 2023]	13.80	0.255	0.735	13.67	0.272	0.755	16.09	0.299	0.701
NeRF-OSR [Rudnev *et al.*, 2022]	16.70	0.303	0.666	16.90	0.355	0.603	17.02	0.315	0.697
GS-IR [Liang *et al.*, 2024]	23.69	0.698	0.267	25.79	0.806	0.158	25.48	0.720	0.255
OSDR-GS (ours)	**24.18**	**0.723**	**0.267**	**25.80**	**0.824**	**0.158**	**25.99**	**0.748**	**0.254**

Table 2: Novel View Synthesis Quantitative Comparison on Mip-NeRF 360 Dataset.

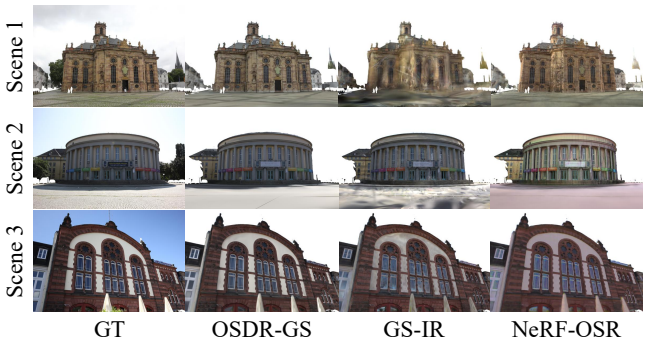

Figure 5: The visualization results of relighting on NeRF-OSR dataset. Scene 1, Scene 2, and Scene 3 represent Ludwigskirche, Staatstheater, and Landwehrplatz, respectively.

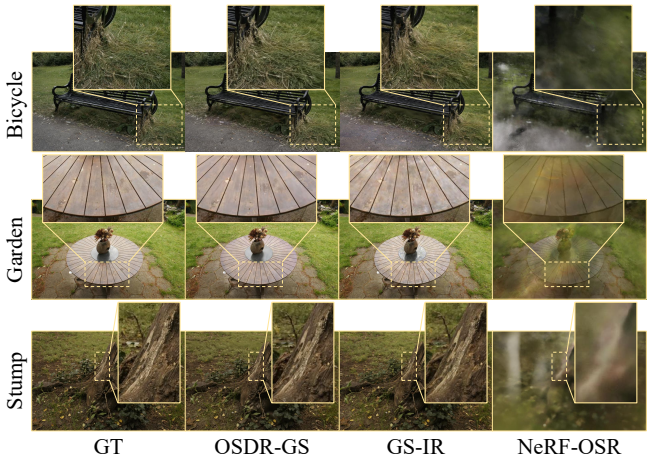

Figure 6: The visualization results of NVS on Mip-NeRF 360 dataset demonstrate that our method synthesizes more realistic colors through fine-grained lighting modeling.

5 Experiments

5.1 Experimental Settings

Datasets. We evaluate our OSDR-GS with SOTA baselines on both NeRF-OSR [Rudnev *et al.*, 2022] and Mip-NeRF 360 [Barron *et al.*, 2022] dataset. ***NeRF-OSR*** dataset consists of 8 outdoor scenes. Each scene includes multi-view images under multiple lighting conditions, with no overlap in lighting between the training and testing sets. ***Mip-NeRF 360*** dataset consists of 5 outdoor scenes with fixed lighting conditions. Following GS-IR [Liang *et al.*, 2024], we use images downsampled by a factor of 4, and pick every eighth image as a test image for dataset splitting.

Baselines and Metrics. We compare our OSDR-GS with three SOTA baselines: NeRF-OSR [Rudnev *et al.*, 2022], GS-IR [Liang *et al.*, 2024], and R3DG [Gao *et al.*, 2023]. 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.

Implementation Details. To adapt the changing lighting conditions for GS-IR and R3DG, we revise them that assign lighting parameters individually to each image rather than to the entire scene. Moreover, due to the different lighting modeling among these methods, ground truth environment maps in NeRF-OSR dataset are not suitable for all of them. Thus, we obtain the necessary lighting parameters of test set by optimizing them exclusively under fixed geometry and material of scene. For each scene, training for 30k iterations on a single NVIDIA RTX 4090 GPU takes approximately 30 minutes. The test-time optimization requires around 5 minutes.

5.2 Comparing with SOTA Methods

In this part, we perform relighting evaluation on NeRF-OSR dataset and novel view sythesis on Mip-NeRF 360 dataset.

Comparison on NeRF-OSR Dataset. We report the relighting performance of our OSDR-GS compared with other baselines. As shown in Tab. 1, our method exhibits outstanding performance in novel view synthesis under novel lighting conditions. Specifically, compared with the best baseline GS-IR [Liang *et al.*, 2024], our method increases by 1.53, 0.55, and 1 on PSNR in all three scenes.

This demonstrates our method achieves more accurate de-

Variants	PSNR ↑	SSIM ↑	LPIPS ↓
w/o LBGL	19.19	0.782	0.234
w/o FGLM	19.12	0.780	0.230
w/o VDSM	18.32	0.778	0.229
w/o RoV	19.37	0.790	0.228
OSDR-GS	**19.62**	**0.795**	**0.216**

Table 3: The ablation study for the proposed components, specifically including lighting-based group learning, fine-grained outdoor lighting modeling, and visibility-driven shadow module.

Grouping	Refinement	PSNR↑	SSIM↑	LPIPS↓
		18.64	0.775	0.234
✓		19.24	0.789	0.221
✓	✓	**19.62**	**0.795**	**0.216**

Table 4: The ablation study for lighting-based group learning.

composition of intrinsic components, resulting in more realistic images rendered under novel lighting condition. We also perform qualitative comparison. As shown in Fig. 5, our method produces more natural relighting effect and more clear image compared with other baselines. Moreover, unlike GS-IR or R3DG that requires multiple stages optimization, our method uses only one-stage optimization process, which significantly reduces the time cost of training.

Comparison on Mip-NeRF 360 Dataset. We report the novel view synthesis (NVS) performance of OSDR-GS on Mip-NeRF 360 dataset. As shown in Tab. 2, we demonstrate results on three outdoor scenes (i.e., Bicycle, Garden, and Stump). Our method achieves slight improvements over GS-IR in terms of PSNR and SSIM, and results comparable performance in LPIPS.

Note that the images in Mip-NeRF 360 dataset are captured under fixed lighting condition. This demonstrates the efficiency of proposed fine-grained lighting modeling and visibility-based shadow module, which are particularly suitable for inverse rendering of outdoor scenes. Moreover, we also conduct qualitative comparison. As shown in Fig. 6, we can observe that our method produces more refined surface details of Garden scene and more realistic colors in Bicycle scene. This indicates the capability of OSDR-GS to decompose scene into precise geometry and lighting.

5.3 Method Analysis

In this part, we conduct elaborate experiments on NeRF-OSR dataset to evaluate the efficacy of our proposed components.

Effect of Proposed Components. We ablate the proposed components of our OSDR-GS, including lighting-based group learning (w/o LBGL), fine-grained lighting modeling (w/o FGLM), visibility-driven shadow module (w/o VDSM), and regularization of visibility (w/o RoV).

As illustrated in Tab. 3, we observe that the last row achieves remarkable performance, which demonstrates that the proposed group learning framework and shadow module are effective for modeling outdoor scene accurately. Moreover, the variant without shadow modeling result in inferior performance. This indicates that our visibility-driven shadow module effectively responds to the changing lighting conditions in the scene and produces realistic shadow effects.

Variants	PSNR↑	SSIM↑	LPIPS↓
w/o VDSM	18.32	0.778	0.229
w/o $\mathcal{L}_{\text{bin}}$	19.44	0.793	0.218
w/o $\mathcal{L}_{\text{nn}}$	19.48	0.792	0.218
OSDR-GS	**19.62**	**0.795**	**0.216**

Table 5: The ablation study for visibility-driven shadow module.

Impact of Lighting-based Group Learning Framework. As summarized in Tab. 4, the first row of the table shows the baseline performance when neither the lighting grouping strategy nor the group refinement approach is applied. The second row indicates the performance improvement when only the lighting grouping strategy is used, and the third row presents the results when both the lighting grouping strategy and the group refinement approach are employed.

These results suggest that both the proposed lighting grouping strategy and the group refinement approach contribute significantly to enhancing the model's performance. Notably, we can observe that variant of only grouping demonstrates results almost as good as the best. This indicates the initialization of lighting groups using lighting cues is quite precise, nearly eliminating the necessity of lighting groups refinement process.

Impact of Visibility-driven Shadow Module. We remove the visibility-driven shadow modeling (w/o VDSM), negative normal regularization (w/o $\mathcal{L}_{\text{nn}}$), and visibility binaryzation regularization (w/o $\mathcal{L}_{\text{bin}}$) from OSDR-GS to evaluate the efficiency of these presented components.

The results are illustrated in Tab. 5. We can observe that the variant of last raw achieves the best results. This demonstrates the outstanding performance of our visibility-driven shadow module in rendering photorealistic images. Both the variant of w/o $\mathcal{L}_{\text{nn}}$ and the variant of w/o $\mathcal{L}_{\text{bin}}$ show slightly worse results, which indicates that both of them play important roles in high-quality visibility modeling.

6 Conclusion

This paper proposes outdoor scene decomposition and relighting with 2D Gaussian splatting (OSDR-GS), a novel inverse rendering strategy under outdoor scenes with changing and unknown lighting conditions. The presented OSDR-GS consists of lighting-based group learning framework, finedgrained outdoor lighting modeling component and visibility-driven shadow module. We conduct extensive experiments to evaluate our method with several state-of-the-art baselines on multiple challenging outdoor datasets. The results demonstrate the effectiveness of OSDR-GS for outdoor decomposition and relighting. Since it is challenging to accurately obtain camera poses in outdoor scene with changing lighting conditions, We plan to consider camera-pose-free decomposition and relighting of outdoor scenes in future research.

Acknowledgments

This work was supported by the National Key R&D Program of China under Grants 2023YFF0906200.

References

[Barron *et al.*, 2021] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5855–5864, 2021.

[Barron *et al.*, 2022] 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.

[Bhattacharyya, 1943] Anil Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distribution. *Bulletin of the Calcutta Mathematical Society*, 35:99–110, 1943.

[Bi *et al.*, 2020] Sai Bi, Zexiang Xu, Pratul Srinivasan, Ben Mildenhall, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. Neural reflectance fields for appearance acquisition. *arXiv preprint arXiv:2008.03824*, 2020.

[Bi *et al.*, 2024] Zoubin Bi, Yixin Zeng, Chong Zeng, Fan Pei, Xiang Feng, Kun Zhou, and Hongzhi Wu. Gs3: Efficient relighting with triple gaussian splatting. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–12, 2024.

[Bian *et al.*, 2023] Wenjing Bian, Zirui Wang, Kejie Li, Jia-Wang Bian, and Victor Adrian Prisacariu. Nope-nerf: Optimising neural radiance field with no pose prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4160–4169, 2023.

[Boss *et al.*, 2021] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T Barron, Ce Liu, and Hendrik Lensch. Nerd: Neural reflectance decomposition from image collections. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12684–12694, 2021.

[Chen and Wang, 2024] Guikun Chen and Wenguan Wang. A survey on 3d gaussian splatting. *arXiv preprint arXiv:2401.03890*, 2024.

[Chen *et al.*, 2022] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *European conference on computer vision*, pages 333–350. Springer, 2022.

[Chen *et al.*, 2024] Danpeng Chen, Hai Li, Weicai Ye, Yifan Wang, Weijian Xie, Shangjin Zhai, Nan Wang, Haomin Liu, Hujun Bao, and Guofeng Zhang. Pgsr: Planar-based gaussian splatting for efficient and high-fidelity surface reconstruction. *arXiv preprint arXiv:2406.06521*, 2024.

[Feng *et al.*, 2015] Wei Feng, Fei-Peng Tian, Qian Zhang, Nan Zhang, Liang Wan, and Jizhou Sun. Fine-grained change detection of misaligned scenes with varied illuminations. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1260–1268, 2015.

[Gao *et al.*, 2022] Kyle Gao, Yina Gao, Hongjie He, Dening Lu, Linlin Xu, and Jonathan Li. Nerf: Neural radiance field in 3d vision, a comprehensive review. *arXiv preprint arXiv:2210.00379*, 2022.

[Gao *et al.*, 2023] Jian Gao, Chun Gu, Youtian Lin, Hao Zhu, Xun Cao, Li Zhang, and Yao Yao. Relightable 3d gaussian: Real-time point cloud relighting with brdf decomposition and ray tracing. *arXiv e-prints*, pages arXiv–2311, 2023.

[Gardner *et al.*, 2024] James AD Gardner, Evgenii Kashin, Bernhard Egger, and William AP Smith. The sky's the limit: Relightable outdoor scenes via a sky-pixel constrained illumination prior and outside-in visibility. In *European Conference on Computer Vision*, pages 126–143. Springer, 2024.

[Guo *et al.*, 2024] Yijia Guo, Yuanxi Bai, Liwen Hu, Ziyi Guo, Mianzhi Liu, Yu Cai, Tiejun Huang, and Lei Ma. Prtgs: Precomputed radiance transfer of gaussian splats for real-time high-quality relighting. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 5112–5120, 2024.

[Huang *et al.*, 2024] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 conference papers*, pages 1–11, 2024.

[Kajiya, 1986] James T Kajiya. The rendering equation. In *Proceedings of the 13th annual conference on Computer graphics and interactive techniques*, pages 143–150, 1986.

[Kerbl *et al.*, 2023] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023.

[Li *et al.*, 2023] Zhong Li, Liangchen Song, Zhang Chen, Xiangyu Du, Lele Chen, Junsong Yuan, and Yi Xu. Relitneulf: Efficient relighting and novel view synthesis via neural 4d light field. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 7007–7016, 2023.

[Liang *et al.*, 2024] Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, and Kui Jia. Gs-ir: 3d gaussian splatting for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21644–21653, 2024.

[Lin *et al.*, 2021] Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. Barf: Bundle-adjusting neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5741–5751, 2021.

[Lin *et al.*, 2024] Jiaqi Lin, Zhihao Li, Xiao Tang, Jianzhuang Liu, Shiyong Liu, Jiayue Liu, Yangdi Lu, Xiaofei Wu, Songcen Xu, Youliang Yan, et al. Vastgaussian: Vast 3d gaussians for large scene reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5166–5175, 2024.

[MacQueen, 1967] James MacQueen. Some methods for classification and analysis of multivariate observations. In

Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, volume 1, pages 281–297. Oakland, CA, USA, 1967.

[Mildenhall *et al.*, 2021] 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.

[Rudnev *et al.*, 2022] Viktor Rudnev, Mohamed Elgharib, William Smith, Lingjie Liu, Vladislav Golyanik, and Christian Theobalt. Nerf for outdoor scene relighting. In *European Conference on Computer Vision*, pages 615–631. Springer, 2022.

[Srinivasan *et al.*, 2021] Pratul P Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T Barron. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7495–7504, 2021.

[Sun *et al.*, 2023] Jia-Mu Sun, Tong Wu, Yong-Liang Yang, Yu-Kun Lai, and Lin Gao. Sol-nerf: Sunlight modeling for outdoor scene decomposition and relighting. In *SIGGRAPH Asia 2023 Conference Papers*, pages 1–11, 2023.

[Sun *et al.*, 2024] Cheng Sun, Jaesung Choe, Charles Loop, Wei-Chiu Ma, and Yu-Chiang Frank Wang. Sparse voxels rasterization: Real-time high-fidelity radiance field rendering. *arXiv preprint arXiv:2412.04459*, 2024.

[Tian *et al.*, 2018] Fei-Peng Tian, Wei Feng, Qian Zhang, Xiaowei Wang, Jizhou Sun, Vincenzo Loia, and Zhi-Qiang Liu. Active camera relocalization from a single reference image without hand-eye calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(12):2791–2806, 2018.

[Wang *et al.*, 2004] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[Wu *et al.*, 2023] Tong Wu, Jia-Mu Sun, Yu-Kun Lai, and Lin Gao. De-nerf: Decoupled neural radiance fields for view-consistent appearance editing and high-frequency environmental relighting. In *ACM SIGGRAPH 2023 conference proceedings*, pages 1–11, 2023.

[Xie *et al.*, 2024] Tianyi Xie, Zeshun Zong, Yuxing Qiu, Xuan Li, Yutao Feng, Yin Yang, and Chenfanfu Jiang. Physgaussian: Physics-integrated 3d gaussians for generative dynamics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4389–4398, 2024.

[Xu *et al.*, 2018] Zexiang Xu, Kalyan Sunkavalli, Sunil Hadap, and Ravi Ramamoorthi. Deep image-based relighting from optimal sparse samples. *ACM Transactions on Graphics (ToG)*, 37(4):1–13, 2018.

[Xu *et al.*, 2023] Yingyan Xu, Gaspard Zoss, Prashanth Chandran, Markus Gross, Derek Bradley, and Paulo Gotardo. Renerf: Relightable neural radiance fields with nearfield lighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 22581–22591, 2023.

[Yang *et al.*, 2022] Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K Wong. Ps-nerf: Neural inverse rendering for multi-view photometric stereo. In *European conference on computer vision*, pages 266–284. Springer, 2022.

[Yu and Smith, 2019] Ye Yu and William AP Smith. Inverserendernet: Learning single image inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3155–3164, 2019.

[Yu *et al.*, 2020] Ye Yu, Abhimitra Meka, Mohamed Elgharib, Hans-Peter Seidel, Christian Theobalt, and William AP Smith. Self-supervised outdoor scene relighting. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16*, pages 84–101. Springer, 2020.

[Yu *et al.*, 2024] 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*, pages 19447–19456, 2024.

[Zhang *et al.*, 2018] Qian Zhang, Wei Feng, Liang Wan, Fei-Peng Tian, and Ping Tan. Active recurrence of lighting condition for fine-grained change detection. In *IJCAI*, pages 4972–4978, 2018.

[Zhang *et al.*, 2020] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020.

[Zhang *et al.*, 2021a] Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5453–5462, 2021.

[Zhang *et al.*, 2021b] Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (ToG)*, 40(6):1–18, 2021.

[Zhang *et al.*, 2022] Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. Modeling indirect illumination for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18643–18652, 2022.

[Zhang *et al.*, 2024] Qi Zhang, Chi Huang, Qian Zhang, Nan Li, and Wei Feng. Learning geometry consistent neural radiance fields from sparse and unposed views. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 8508–8517, 2024.