LumiGauss: Relightable Gaussian Splatting in the Wild

Joanna Kaleta[1,2*] **Kacper Kania**[1] **Tomasz Trzciński**[1,4,5] **Marek Kowalski**[3]

[1]Warsaw University of Technology [2]Sano Centre for Computational Medicine [3]Microsoft [4]IDEAS NCBR [5]Tooploox

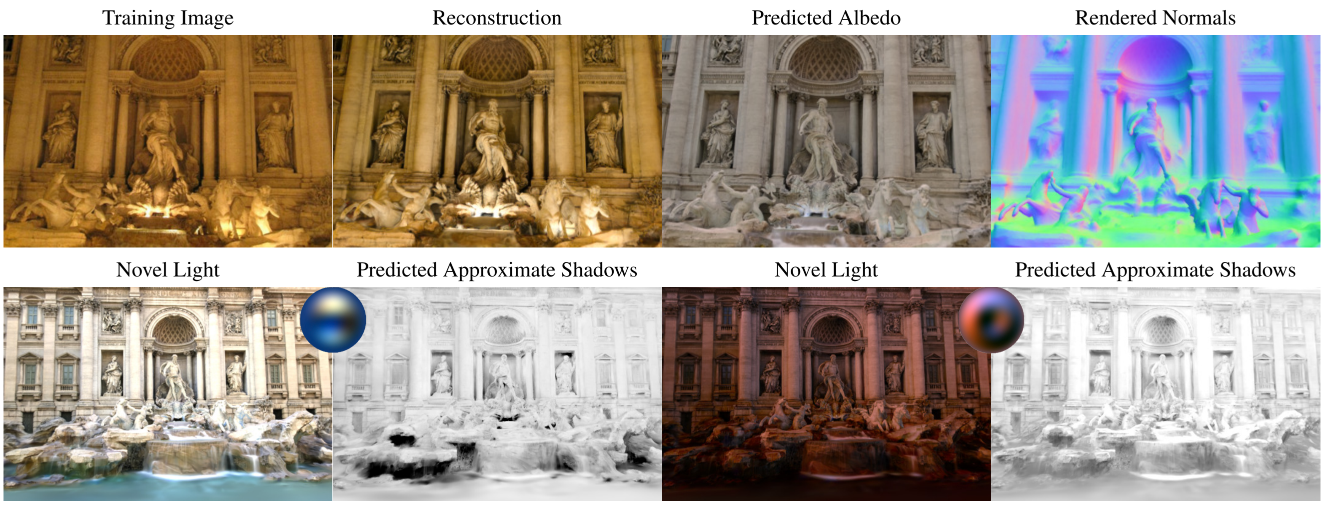

Figure 1. **Teaser** – LumiGauss reconstructs environment maps and object surfaces from *in-the-wild* images. Our model decouples the scene color and its normals (*second and fourth column in the top row*). At inference, it can synthesize novel views (*bottom row*) and realistic lighting (*first and third columns in the bottom*) with high-fidelity shadows (*second and fourth columns in the bottom*).

Abstract

Decoupling lighting from geometry using unconstrained photo collections is notoriously challenging. Solving it would benefit many users as creating complex 3D assets takes days of manual labor. Many previous works have attempted to address this issue, often at the expense of output fidelity, which questions the practicality of such methods. We introduce LumiGauss - a technique that tackles 3D reconstruction of scenes and environmental lighting through 2D Gaussian Splatting. Our approach yields high-quality scene reconstructions and enables realistic lighting synthesis under novel environment maps. We also propose a method for enhancing the quality of shadows, common in outdoor scenes, by exploiting spherical harmonics properties. Our approach facilitates seamless integration with game engines and enables the use of fast precomputed radiance transfer. We validate our method on the NeRF-OSR dataset, demonstrating superior performance over baseline methods. Moreover, LumiGauss can synthesize realistic images for unseen environment maps. Our code: https://github.com/joaxkal/lumigauss.

*Corresponding authors: joanna.kaleta.dokt@pw.edu.pl

1. Introduction

The colors emitted by objects are a combination of a spectrum of the light hitting the object and the material properties of that object. The light hitting the object's surface is a sum of the light scattered in the medium and bounced from neighboring objects [35]. In computer graphics, we often simplify this effect and decouple it into two entities: an intrinsic object's color or *albedo* and an omnidirectional texture representing the illumination [22]—*environment map*. Acquiring those assets enables the designing of realistic scenes in games or movies.

In many scenarios, creating realistic albedo textures and environment maps requires skilled technicians and artists to be involved in the process. To democratize it, the previous approaches [7, 23, 34] tried to use photographs taken with commodity cameras and *invert* the capturing process to recover albedo and an environment map. Given the abundance of casual, in-the-wild photographs available on the Internet, solving that issue is of high importance.

Recent advancements in reconstruction in-the-wild include NeRF-in-the-Wild [18] (NeRF-W). NeRF-W leverages neural radiance fields [19] which reconstruct a scene given its photos with calibrated cameras. NeRF-W can further work in realistic scenarios where the pictures come from the *in-the-wild* collections—the images in such may differ in the lighting conditions or scene content. However, NeRF-W and its follow-up works, HA-NeRF [3] and CR-NeRF [39], cannot decouple the object's albedo and the environment map, making it difficult to use in practice. NeRF-OSR [23] approaches that problem, but its shading model requires neural network execution at runtime, making integration with graphics engines difficult, and the reconstruction quality leaves space for improvement.

3D Gaussian Splatting [13] (3DGS) solves one of the main bottlenecks of NeRF - the training speed and output fidelity. In contrast to NeRFs, 3DGS models the scene as a composition of 3D Gaussians attributed with colors and opacity which are rasterized, or *splatted*, to render the output image. Recovering an object's surface from them requires specialized training techniques [9]. On the other hand, 2DGS [12] proposes reformulating 3D Gaussians as their 2D alternative where one of the axes is collapsed. The final scene representation ends up being composed of 2D *surfels* which provide a flat surface crucial for our relighting approach.

In this work, we propose LumiGauss, a method that uses 2DGS [12] to perform inverse graphics on images taken in the wild. In contrast to past approaches, our method is imbued with fast training and inference speed while maintaining high-quality renderings and being easy to integrate with graphics engines. In our method, the light is modeled as a combination of an environment map and a radiance transfer function that represents which parts of the environment map illuminate a given surfel—both are modeled by spherical harmonics [22]. This approach allows for modeling shadows, which is our main goal, but also has the potential to represent light reflected off of other objects. The output from LumiGauss enables both novel view synthesis and relighting using environment maps beyond those available during training. Leveraging the possibilities offered by the precomputed radiance transfer, our representation integrates seamlessly into game engines, enabling fast and efficient relighting.

Our contributions:

- We repurpose 2D Gaussian Splatting for an inverse graphics pipeline in an in-the-wild setting. With our approach, we recover high-quality albedo and environment maps.
- To enable shadows we learn the radiance transfer function for each 2D splat and represent it using spherical harmonics.
- Finally, we demonstrate that our reconstructed environment maps can be effectively used to relight arbitrary objects within graphic engines.

2. Related Works

Relighting. Relighting outdoor scenes is a key challenge in computer graphics and VR/AR. Early works [1, 5, 10, 14, 30, 31, 36] used training-free methods like statistical inference. Deep learning approaches, such as Yu et al. [40] with a neural renderer, and Philip et al. [20] with proxy geometry, face limitations in reconstruction quality and viewpoint flexibility.

NeRF-based methods [19] enabled simultaneous viewpoint and lighting changes. However, methods like [27, 41, 44] handle a single illumination only or specific illumination setup during. Others are object-specific, such as for faces [29]. Many unconstrained photo collection methods focus on appearance, not lighting, complicating integration with other graphical components [3, 15, 18, 39].

NeRF-based approaches, such as [21] and [11], focus on inverse rendering for outdoor scenes, particularly in applications like autonomous driving. However, these methods are designed for single video sequences rather than unstructured photo collections. Rudnev et al. [23] proposed a method for relighting landmarks from unconstrained photo collections, using NeRF with external lighting extraction. Similarly, [16] compresses the per-image illumination into a disentangled latent vector. Wang et al. [34] target static scenes and works with unconstrained photo collections but rely on costly mesh extraction. Some methods incorporate additional priors, environmental assumptions, or regularizations [28, 38]. Gardner et al. [7] leverage externally trained models to provide environmental lighting priors. Despite their potential, these methods cannot be used in real-time applications due to NeRF's slow training and rendering times.

In contrast, the TensoRF-based approach by Chang et al. [2] aligns time information and sun direction with images for relighting, eliminating the need for external lighting models. While this method is faster than NeRF, it still lacks seamless integration with graphics engines and is unsuitable for synthetic light integration.

Notable Gaussian Splatting works designed for unconstrained photo collections [4, 13, 32, 37, 42] focused on appearance editing, not seamless graphical component integration. Relightable Gaussian approaches, like [6, 17, 25], tackle material decomposition but are not adapted to handle varying lighting conditions of *in-the-wild* training setup. Radiance transfer properties, employed in a similar way to LumiGauss, are utilized in [24, 43]. However, these methods rely on a burdensome dataset setup, restricting their applicability to specific use cases.

Gaussian Splatting. Kerbl et al. [13] introduced a notion of using learnable 3D Gaussian primitives from point clouds. Those Gaussians are parametrized with 3D covariance matrix Σ_k and their location $\mathbf{t}_k$:

$$\mathcal{G}(\mathbf{t}) = \exp(\frac{1}{2}(\mathbf{t}-\mathbf{t}_k)^\top \Sigma_k^{-1}(\mathbf{t}-\mathbf{t}_k)), \quad (1)$$

where the covariance matrix is factorized into a scaling diagonal matrix $\mathbf{s}_k$ and a rotation matrix $\mathbf{R}_k$ as $\Sigma_k{=}\mathbf{R}_k\mathbf{s}_k\mathbf{s}_k^\top\mathbf{R}_k^\top$. An image is rendered with a splatting operator $\mathcal{S}(\cdot)$ which projects Gaussians into the camera coordinates with a world-to-camera matrix and then to image plane with a local affine transformation [45]:

$$\mathcal{S}(\mathcal{C}_c \mid \mathcal{G}) = \sum_{k=1}^{K} \mathbf{c}_k o_k \mathcal{G}_k \prod_{j=1}^{k-1}(1-o_j\mathcal{G}_k). \quad (2)$$

The operator produces an RGB image, given a calibrated camera matrix $\mathcal{C}_c$ and their additional Gaussians' attributes: their colors $\mathbf{c}$ and opacities o. Attributes are learned using a stochastic gradient descent.

Huan et al. [12] argues that 3DGS although producing high-quality images, the implicit surface representation is noisy, limiting its applicability in relighting scenarios. They propose using 2D Gaussians instead to create smooth, coherent meshes thanks to their exact 2D surfel projection. We leverage that representation in our LumiGauss—a relightable model that decouples albedo, environment light and shadows thanks to our proposed physical constraints.

3. Method

3.1. Preliminaries on Radiance Transfer

The rendering equation, in its simplified form [8], is an integral function that represents light $L(\mathbf{x}, \boldsymbol{\omega}_o)$ exiting point $\mathbf{x}$ along the vector $\boldsymbol{\omega}_o$:

$$L(\mathbf{x}, \boldsymbol{\omega}_o) = \int_s f_r(x, \boldsymbol{\omega}_o, \boldsymbol{\omega}_i) L_i(\mathbf{x}, \boldsymbol{\omega}_i) D(\mathbf{x}, \boldsymbol{\omega}_i) d\boldsymbol{\omega}_i \quad (3)$$

where $f_r(\cdot)$ is a BRDF function, $L_i(\cdot)$ an incoming light along the vector $\boldsymbol{\omega}_i$, and $D(\cdot)$ is a radiance transfer function. Intuitively, $f_r(\cdot)$ represents the surface material, $L_i(\cdot)$ represents the intensity and color of the illumination, and $D(\cdot)$ is a term that takes into account shadows or light reflections from other surfaces. Depending on the formulation of those functions, the rendering equation can range from a straightforward and inaccurate light model to a highly complex and accurate one.

Unshadowed model. One example of a reflection model that can be represented with Eq. (3) is the diffuse surface reflection model, also known as *dot product lighting*. A diffuse BRDF reflects light uniformly, making the lighting view-independent and simplifying the BRDF as follows:

$$L_D(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\pi} \int_s L_i(\mathbf{x}, \boldsymbol{\omega}_i) \max(\mathbf{n}(\mathbf{x}) \cdot \boldsymbol{\omega}_i, 0) d\boldsymbol{\omega}_i \quad (4)$$

where $\rho(\cdot)$ is the surface albedo, $\mathbf{n}(\mathbf{x})$ a surface normal at the point x. Shadows are neglected.

The incoming light $L_i(\mathbf{x}, \boldsymbol{\omega}_i)$ can be represented in several ways. In this work, we assume that the scene is illuminated with an **omnidirectional environment map** that is parametrized using spherical harmonics (SH) of degree n with $(n{+}1)^2$ coefficients. Because the environment map is positioned infinitely far from the scene, the light is position-independent, and thus, the rendering equation is further simplified:

$$L_U(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\pi} \int_s L_i(\boldsymbol{\omega}_i) \max(\mathbf{n}(\mathbf{x}) \cdot \boldsymbol{\omega}_i, 0) d\boldsymbol{\omega}_i \quad (5)$$

With illumination parametrized with SH, we can evaluate the integral in Eq. (5) using a closed-form solution from Eq. (12) in [22]. From this point onward, we refer to rendering with Eq. (5) as *unshadowed*.

Shadowed model. In addition to the *unshadowed* lighting model, we propose a *shadowed* model, where $D(\mathbf{x}, \boldsymbol{\omega}_i)$ is parameterized using spherical harmonics (SH) and learned from training data. In $D(\mathbf{x}, \boldsymbol{\omega}_i)$, SH represents a spherical signal that quantifies the light arriving from each direction of the environment map to an associated point in space. The *shadowed* model is derived by replacing the dot product term in Eq. (5):

$$L_S(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\pi} \int_s L_i(\boldsymbol{\omega}_i) D(\mathbf{x}, \boldsymbol{\omega}_i) d\boldsymbol{\omega}_i. \quad (6)$$

In addition to modeling shadows, this approach also has the potential to model the interreflection of light between objects in the scene.

Using SH of the same degree for both the environment map and transfer function allows efficient evaluation of the rendering equation Eq. (6). A key SH property simplifies the integral of two SH-based functions to a dot product of their coefficients, thanks to SH orthogonality. With this property Eq. (6) can be re-written as:

$$L_S(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\pi} \mathbf{l} \cdot \mathbf{d}, \quad (7)$$

where $\mathbf{l} \in R^{(n+1)^2}$ are the SH coefficients of $L_i(\boldsymbol{\omega}_i)$ and $\mathbf{d} \in R^{(n+1)^2}$ are the SH coefficients of $D(\mathbf{x}, \boldsymbol{\omega}_i)$. Please see [8, 26] for derivation. This property is commonly used in real-time rendering where the radiance transfer function is pre-computed and only Eq. (7) is evaluated at runtime.

3.2. LumiGauss

LumiGauss creates a 3D representation of a relightable model using 2D Gaussians [12] from $c \leq C$ images taken *in-the-wild* $\{\mathcal{I}_c\}_{c=1}^C$ with associated calibrated cameras $\{\mathcal{C}_c\}_{c=1}^C$. Our goal is to find Gaussian parameters $\mathcal{G}{=}\{\mathbf{t}_k, \mathbf{R}_k, \mathbf{s}_k, o_k, \boldsymbol{\rho}_k, \mathbf{d}_k\}_{k=1}^K$ that after the rasterization [13] recreate those images. We optimize Gaussians by

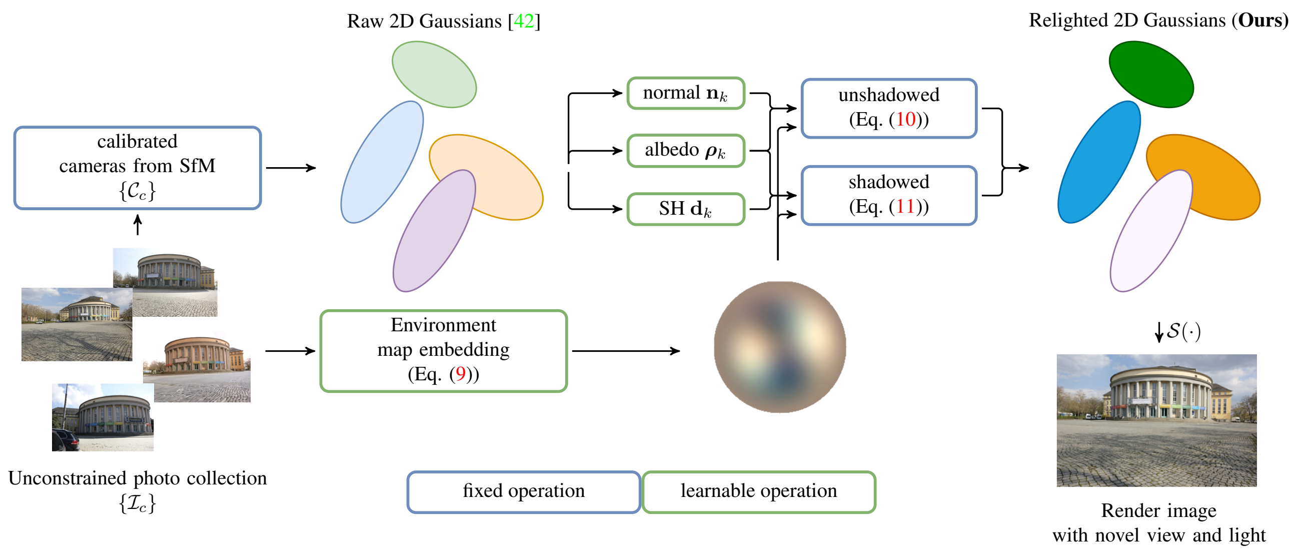

Figure 2. **Pipeline –** LumiGauss learns the relightable 2D Gaussian [42] representation from unconstrained photo collection with variable camera parameters and lighting conditions. Each of k Gaussians holds: a normal $\mathbf{n}_k$, albedo $\boldsymbol{\rho}_k$, and learnable transfer function $\mathbf{d}_k$. Our contributed method composes the Gaussians in two modes—*shadowed* and *unshadowed*. The *shadowed* model reconstructs additional shadows (see Fig. 1) on top of the unshadowed model thanks to our proposed use of a radiance transfer function. The Gaussians are splatted [13, 42] to render the output image in a novel view and light.

minimizing the objective:

$$\arg\min_{\mathcal{G},\mathcal{E},\boldsymbol{\theta}} \mathbb{E}_{\mathcal{C}_c\sim\{\mathcal{C}_c\}} \underbrace{\ell_{\text{rgb}}(\mathcal{S}(\mathcal{C}_c \mid \mathcal{G},\mathcal{E},\boldsymbol{\theta}),\mathcal{I}_c)}_{\text{Sec. 3.4}} + \underbrace{\mathcal{R}(\mathcal{G})}_{\text{Sec. 3.3}}, \tag{8}$$

where $\mathcal{E}=\{\mathbf{e}_c\}_{c=1}^{C}$ is a set of scene-dependent, learnable environment embeddings, ℓ_{rgb} is a photometric objective that compares the rendered image from an operator $\mathcal{S}(\cdot)$ (Eq. (2)), and $\mathcal{R}$ are additional regularization terms. In contrast to 2DGS [12], for each Gaussian we model the base color $\boldsymbol{\rho}$ as diffuse[1], and introduce SH coefficients for the transfer function $\mathbf{d}$[2]. 2DGS provides smooth normals that make relighting possible.

In what follows, we drop the dependence of functional forms on the positions $\mathbf{x}$ we introduced in Sec. 3.1 for brevity.

Relighting. To handle the diverse lighting conditions in *in-the-wild* images, we associate each training image with a learnable latent code $\mathbf{e}_c$ that encodes its lighting conditions. Using this embedding, we predict the environment map coefficients via an MLP:

$$\mathbf{l}_c = \text{MLP}(\mathbf{e}_c|\boldsymbol{\theta}), \tag{9}$$

where $\mathbf{l}_c \in \mathbb{R}^{3\times(n+1)^2}$ represents the SH coefficients of the environment map, and n=2 is the SH degree. As shown in [22], second-order SH is sufficient to approximate environment lighting in many scenarios.

[1] As per Sec. 3.1, view-dependent effects are not modeled in diffuse reflections.

[2] These coefficients correspond to a single channel in practice.

The predicted illumination is used in the rendering process in one of two ways: *unshadowed* and *shadowed*. Those two approaches correspond to Eq. (5) and Eq. (7) respectively, and are described below.

Unshadowed model. For the unshadowed scenario, we follow Eq. (5), which integrates light over the hemisphere in the direction of the surface normal. The color $\mathbf{c}_k$, *radiance*, for each Gaussian $\mathcal{G}_k$ given its normal $\mathbf{n}_k$ and the illumination parameters $\mathbf{l}_c$ equates to:

$$\mathbf{c}_k = \boldsymbol{\rho}_k \odot \underbrace{\mathbf{n}_k^t M(\mathbf{l}_k)\mathbf{n}_k}_{\text{unshadowed irradiance}}, \tag{10}$$

where M is a 4×4 matrix derived from the SH parameters of the environment map. It is the closed form solution of the integral in Eq. (5), please see Eq. (12) in [22] for details.

This simple yet effective model already imbues the model with relighting capabilities. However, as described in Fig. 3 it does not capture shadows correctly, limiting the output's fidelity.

Shadowed model. To effectively capture shadows in the model, we redefine the output color of a Gaussian as $\tilde{\mathbf{c}}_k$, a function of learnable radiance transfer function D_k parametrized by spherical harmonics $\mathbf{d}_k \in \mathbb{R}^{(n+1)^2}$, light $\mathbf{l}_c$ and albedo $\boldsymbol{\rho}_k$. Using a learned radiance transfer function (instead of fixing it to capture light from the hemisphere above the normal as we do in *unshadowed*) allows for creating shadows, as described in Sec. 3.1. Overall, following

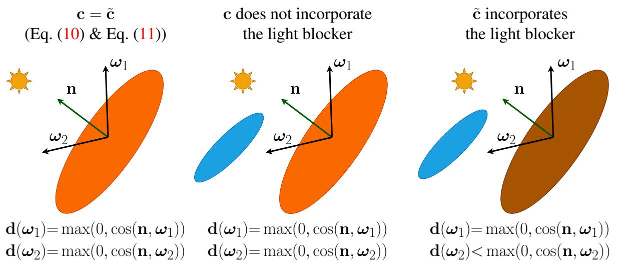

Figure 3. Unshadowed $\mathbf{c}$ (Eq. (10)) and shadowed $\tilde{\mathbf{c}}$ (Eq. (11)) may give the same output color if a Gaussian is fully exposed to the environment light. In the case of any occluder, $\mathbf{c}$ does not handle, and the color does not change. However, our proposed $\tilde{\mathbf{c}}$ properly reacts to the occluder and makes the output color darker.

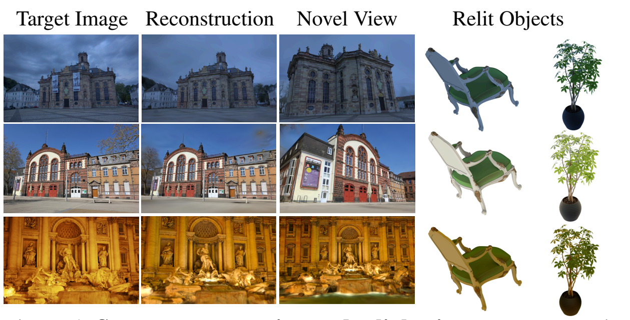

Figure 4. **Scene reconstruction and relighting –** Reconstruction and relighting capabilities of LumiGauss. LumiGauss reproduces sharp and clean landmarks, and the learned environment lighting enables accurate scene relighting. We use learned environment maps to relight the scene from novel viewpoints and then relight arbitrary objects within a graphics engine.

Eq. (7), the output shadowed color or *radiance* reduces to:

$$\tilde{\mathbf{c}}_k = \boldsymbol{\rho}_k \odot \underbrace{\sum_{i=1}^{(n+1)^2} \mathbf{l}_c^i \cdot \mathbf{d}_k^i}_{\text{shadowed irradiance}}, \tag{11}$$

As we show later in the experiments, the addition of shadows leads to more accurate relighting. Additionally, it does not require learnable MLP to reconstruct shadows at the inference stage, differentiating it from NeRF-OSR [23] and making our approach applicable to rendering engines directly.

3.3. Physical constraints

The regularizations proposed in 2DGS [12] keep the Gaussians close to the surface and smooth locally, which is crucial in our relighting scenario. Aside from them, we propose new loss terms based on the physical light properties that restrict the optimization from achieving degenerate, *non-relightable* cases. We restrict radiance transfer D_k function to remain within the range of 0 to 1, where 0 indicates complete shadowing and 1 signifies full exposure to lighting:

$$\ell_{0\text{-}1} = \mathbb{E}_k \mathbb{E}_{\boldsymbol{\omega}_i} [\| \max(D_k(\boldsymbol{\omega}_i), 1) - 1\|_2^2 + \| \min(D_k(\boldsymbol{\omega}_i), 0)\|_2^2], \tag{12}$$

and allow the environment light to remain in the $\mathbb{R}_+$ domain:

$$\ell_+ = \mathbb{E}_k \mathbb{E}_{\boldsymbol{\omega}_i} \| \min(L_c(\boldsymbol{\omega}_i), 0)\|_2^2, \tag{13}$$

which allows the environment light to brighten the scene arbitrarily.

The shadowed radiance transfer should remain close to the unshadowed version. If not, the shadowed version might include light from any direction, resulting in degenerate solutions and incorrect relighting. We visualize the shadowed and unshadowed transfer functions in Fig. 3. To address this issue, we propose the following loss function:

$$\ell_{\LEFTcircle\leftrightarrow\circ} = \mathbb{E}_k \mathbb{E}_{\boldsymbol{\omega}_i} \| \max(\mathbf{n}_k \cdot \boldsymbol{\omega}_i, 0) - D_k(\boldsymbol{\omega}_i)\|_2^2, \tag{14}$$

The applied transfer function inherently accounts for shadows and interreflections. To focus specifically on modeling shadows and restrict the use of Eq. (11) for other cases, we impose a loss function ensuring that shadowed radiance should not be brighter than unshadowed one:

$$\ell_{\LEFTcircle} = \mathbb{E}_k \mathbb{E}_{\boldsymbol{\omega}_i} \| \max(D_k(\boldsymbol{\omega}_i) - \max(\mathbf{n}_k \cdot \boldsymbol{\omega}_i, 0), 0)\|_2^2, \tag{15}$$

Those losses are weighted with scalars $\{\lambda_{1,\ldots,4}\}$ fixed across experiments and contribute to our regularization term:

$$\mathcal{R}(\mathcal{G}) = \lambda_1 \ell_{0\text{-}1} + \lambda_2 \ell_+ + \lambda_3 \ell_{\LEFTcircle\leftrightarrow\circ} + \lambda_4 \ell_{\LEFTcircle} \tag{16}$$

Calculating it exactly requires us to compute the expectation over the hemisphere $\mathbb{S}^2$. Instead, we approximate the expectations over directions $\boldsymbol{\omega}_i$ with a Monte Carlo estimator by randomly sampling the SH lobe with N samples at each training step.

3.4. Reconstruction

We render images using the splatting algorithm $\mathcal{S}(\cdot)$ proposed in 2DGS [12]. We compare the rendered images with ground-truth $\{\mathcal{I}_c\}$ taken with $\{\mathcal{C}_c\}$ cameras. Our method builds on 2DGS [12] and therefore our reconstruction loss ℓ_{rgb} follows the following term:

$$\ell_{\text{rgb}} = \lambda_{\text{rec}}(\LEFTcircle)\ell_{\text{rec}}(\LEFTcircle) + \lambda_{\text{rec}}(\circ)\ell_{\text{rec}}(\circ), \tag{17}$$

$$\ell_{\text{rec}}(\{\LEFTcircle, \circ\}) = \ell_1(\{\LEFTcircle, \circ\}) + \lambda \ell_{\text{D-SSIM}}(\{\LEFTcircle, \circ\}), \tag{18}$$

where the ℓ_1 is the L_1 loss comparing either the image rendered from our shadowed or unshadowed models and $\ell_{\text{D-SSIM}}$ is a differentiable D-SSIM [33] further improving the quality. We use λ=0.2 throughout all the experiments. Our proposed $\ell_{\text{rec}}(\circ)$ resembles a pretraining stage.

Table 1. **Quantitative results –** Comparison between our LumiGauss and selected baselines for two different . We report the reconstruction quality regarding PSNR, MSE, MAE, SSIM on full and 4x downsampled image resolutions. *u/s* denotes using upsampled, *d/s* downsampled images for the evaluation, and the last delimited area presents the ablation study on downsampled data. We denote NeRF-OSR [23] results reproduced by FEGR [34] with *. We use † to further annotate our approach where we remove loss terms $\ell_{◐\leftrightarrow○}$, $\ell_{\text{rec}}(○)$ from the second training stage. In ‡, we omit the first training stage. Compared to the baselines, LumiGauss achieves reconstructions of high fidelity. It reliably produces smooth surfaces and sharp edges, reflected in its high SSIM values. Additionally, our proposed components either enhance reconstruction or preserve physical accuracy without negatively impacting the results. Please note, that NeuSky [7] is a concurrent work, published prior to the WACV's deadline at ECCV 2024.

Method	Landwehrplatz				Ludwigskirche				Staatstheater			
	PSNR ↑	MSE ↓	MAE ↓	SSIM ↑	PSNR ↑	MSE ↓	MAE ↓	SSIM ↑	PSNR ↑	MSE ↓	MAE ↓	SSIM ↑
Yu et al. $_{u/s}$ [40]	15.17	0.033	0.133	0.376	17.87	0.017	0.097	0.378	15.28	0.032	0.138	0.385
Philip et al. [20]	12.28	0.062	0.179	0.319	16.63	0.023	0.113	0.367	12.34	0.065	0.200	0.272
NeRF-OSR [23]	16.65	0.024	0.114	0.501	18.72	0.014	0.090	0.468	15.43	0.029	0.133	0.517
NeRF-OSR* [23]	15.66	0.029	-	-	19.34	0.012	-	-	16.35	0.027	-	-
SR-TensoRF [2]	16.74	0.024	0.093	0.653	17.30	0.021	0.096	0.542	15.43	0.030	0.111	0.632
FEGR [34]	17.57	0.018	-	-	21.53	0.007	-	-	17.00	0.023	-	-
SOL-NeRF [28]	17.58	0.028	-	0.618	21.23	0.008	-	0.749	18.18	0.019	-	0.680
NeuSky [7]	18.31	0.016	-	-	22.50	0.005	-	-	16.66	0.023	-	-
Ours	18.01	0.017	0.096	0.778	19.59	0.012	0.085	0.700	17.02	0.021	0.107	0.729
Yu et al. [40]	15.84	0.028	0.123	0.392	18.71	0.014	0.088	0.400	15.43	0.031	0.136	0.363
Philip et al. $_{d/s}$ [20]	12.85	0.054	0.169	0.164	17.37	0.019	0.105	0.429	11.85	0.070	0.210	0.184
NeRF-OSR $_{d/s}$ [23]	17.38	0.021	0.106	0.576	19.86	0.011	0.080	0.626	15.83	0.026	0.128	0.556
Ours $_{d/s}$	18.40	0.016	0.094	0.746	20.13	0.011	0.080	0.727	17.24	0.020	0.105	0.715
Ours †	15.03	0.034	0.139	0.58	19.34	0.015	0.094	0.693	16.09	0.028	0.124	0.665
Ours ‡	17.59	0.019	0.100	0.733	19.05	0.016	0.097	0.680	16.83	0.022	0.110	0.694
Ours $\setminus \ell_{\text{0-1}}$	18.30	0.016	0.095	0.744	20.15	0.010	0.080	0.734	17.25	0.020	0.105	0.712
Ours $\setminus \ell_{+}$	17.35	0.020	0.104	0.728	20.17	0.012	0.081	0.729	17.10	0.020	0.106	0.703

Table 2. **Performance comparison –** Training time and inference speed comparison between the baselines and our LumiGauss.

Method	Training time	FPS
NeRF-OSR [23]	31h	0.003
NeuSky [7]	14h	0.004
Ours	1h 20min	20.7

As the more complex shadowed model lands in local minima if trained from scratch, we initiate the training with $\lambda_{\text{rec}}(◐)$=0.0 and $\lambda_{\text{rec}}(○)$=1.0. Once the simpler model converges, we switch $\lambda_{\text{rec}}(◐)$=1.0 and $\lambda_{\text{rec}}(○)$ to a small value so as not to deteriorate the quality of the model. In short, the shadowed model explains the parts of an image with shadows, which the unshadowed could not with its simpler lighting model.

4. Experiments

4.1. Datasets and baselines

To evaluate our approach, we followed the protocol from NeRF-OSR [23] using ground truth environment maps. We use the official data split for Staatstheatert, Landwehrplatz, and Ludwigskirche. We use segmentation masks for test images provided in the OSR dataset and calculate MSE, MAE, SSIM and PSNR on masked regions only. We compare LumiGauss against several NeRF-based baselines[3] and TensoRF baseline. We provide the implementation details in `Supplementary`.

4.2. Scene reconstruction and relightning

We present the qualitative results in Tab. 1 and quantitative in Figs. 4 and 5. As Yu et al. [40] evaluates their model on downsampled images, we show the metric values on downsampled (*d/s*), and upsampled (*u/s*) to identify the quality differences. As we can see, LumiGauss performs better or on par with the baselines. NeuSky [7] is a concurrent work which models the environment maps and the sky using a prior, pretrained model.

As our backbone, 2DGS [12] incorportates priors to produce sharp edges and smooth surfaces, our model inherently performs better as expressed by SSIM. Please also see the zoom-ins in Fig. 5. Those shape reconstruction qualities allow us to relight the scene with high fidelity. We demonstrate that in Fig. 7 where one can see that our method effectively relights landmarks under various lighting conditions. We finally visualize the rendered shadows produced thanks to our proposed physical constraints at training time. Since

[3]We include the concurrent NeuSky [7] which has been published officially after the WACV deadline.

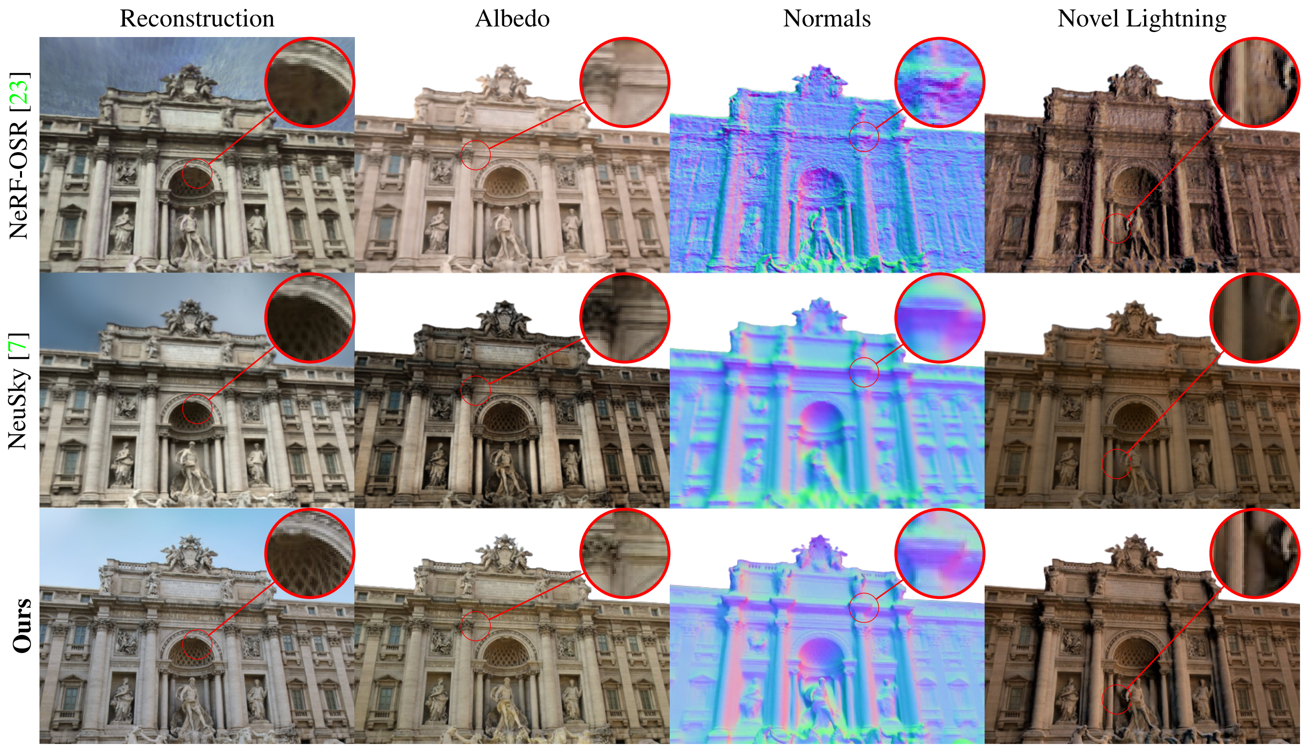

Figure 5. **Qualitative comparison of albedo, normals, and relighting under similar lighting conditions on Trevi Fountain.** LumiGauss produces albedo with fewer baked-in shadows, sharp normals, smooth surfaces, and more accurate novel lighting compared to the baselines. Results for NeuSky originally reported in [7]. Please, zoom in for details.

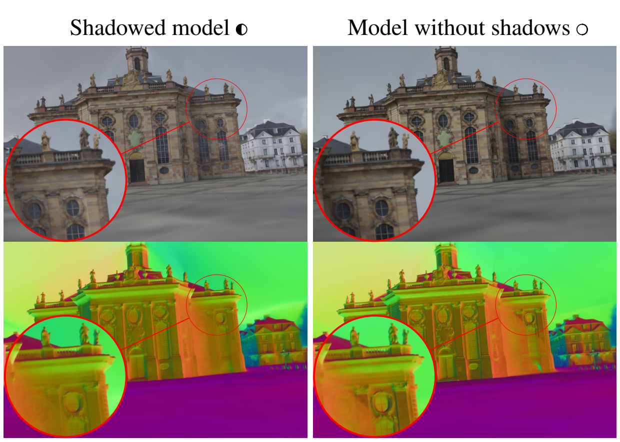

Figure 6. **Effects of shadowed training –** We show the comparison of **albedo** between the shadowed (*left*) and unshadowed (*right*) models. The albedo in the shadowed training is brighter with fewer shadows. The shadowed model recovers more accurate normals.

LumiGauss does not predict shadows explicitly, we visualize them as grayscaled difference of output irradiances between the *unshadowed* (Eq. (10)) and *shadowed* (Eq. (11)) to approximate shadow effects:

$$\max(\mathfrak{g}(\mathbf{c}_k \oslash \boldsymbol{\rho}_k - \tilde{\mathbf{c}}_k \oslash \boldsymbol{\rho}_k), 0), \tag{19}$$

where $\oslash$ is an element-wise division and $\mathfrak{g}(\cdot)$ converts from the RGB space to the grayscale space.

We also display the illumination in Fig. 7 to differentiate between shadows and dark illumination from the environment map. Additional detailed results on scene reconstruction, relightning and more comparisons to other works are included in `Supplementary`.

4.3. Ablations

We prioritize enhancing relighting capabilities over accurate appearance recreation during the optimization process, contrasting with recent Gaussian splatting methods that target novel view synthesis based on unconstrained photo collections [4, 37, 42]. Consequently, our ablation study primarily focuses on the degradation of relighting capabilities when removing any of the proposed components. We compare shadowed and unshadowed modeling and investigate the contributions of each loss term. We present the results in Tab. 1.

Gaussians can optimize to shadowed surfaces and represent shadows as normals and albedo colors (effect known as albedo/illumination ambiguity). Therefore, gains from separating shadows from lightning are not visible in metrics computed on a limited data subset. We noticed that adding a shadowed version can help restore proper albedo

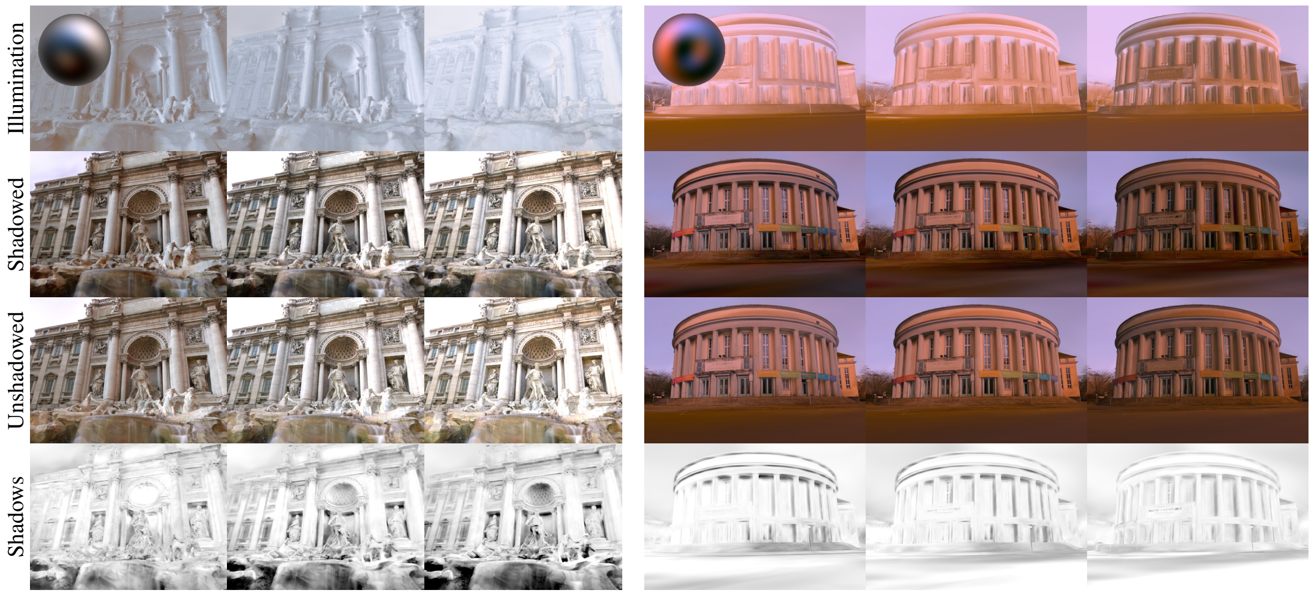

Figure 7. **Environment map rotation –** The top row shows the illumination entering the scene. The second and third rows display the shadowed and unshadowed renderings, respectively. The last row represents the approximate predicted shadows. Please zoom in for details.

and normal vectors of surfaces that during the training were distorted or had low brightness (see Fig. 6).

4.4. Performance comparison

We compare LumiGauss' efficiency with two NeRF baselines. Our method achieves plausible relightning results while being orders of magnitudes faster both in terms of training and inference as shown in Tab. 2.

4.5. Limitations

We identify the following limitations of our approach. Notably, surface albedo and normals may attempt to simulate shadows in scenarios with hard and frequent shadows. This can pose challenges for shadow training, especially when shadows are visible in several training images, potentially hindering the accurate representation of surface normals. Incorporating priors for environment light and shadowing could further enhance disentanglement and light transport modeling as presented in the concurrent NeuSky [7]. While we assume diffuse albedo, valid for most outdoor cases, shadows can appear unnaturally on reflective surfaces such as windows. Separate background optimization could enhance the synthesis of scenes with extensive sky areas. Finally, our shadow modeling baked-in the spherical harmonics representations is non trivial to extend to dynamic applications, such as autonomous driving.

5. Conclusions

We present LumiGauss—the method capable of decoupling environment lighting and albedo of objects from images *in-the-wild*. To this end, we apply 2DGS [12] to reconstruct the object's surface accurately and then use our proposed training components that correctly disentangle light properties from the rendered colors. As we show in the experiments, our approach achieves better reconstruction results than the baselines. We also present that one of our contributions—modeling shadows via leveraging Spherical Harmonics properties—provides shadows of high fidelity that react appropriately to changing environment light. LumiGauss is a novel approach in the direction of inverting the rendering process from images *in-the-wild*, reconstructing high-quality scene properties without sacrificing the fidelity of the output.

Acknowledgments

This research was supported by Microsoft Research through the EMEA PhD Scholarship Programme, National Science Centre, Poland (grant no. 2022/47/O/ST6/01407), the European Union's Horizon 2020 research and innovation program (grant agreement no. 857533, Sano), the International Research Agendas program of the Foundation for Polish Science, co-financed by the European Regional Development Fund, and the Horizon Europe Programme (HORIZON-CL4-2022-HUMAN-02) under the project "ELIAS: European Lighthouse of AI for Sustainability" (GA no. 101120237). We thank NVIDIA Corporation for providing access to GPUs via NVIDIA's Academic Hardware Grants Program.

References

[1] Jonathan T Barron and Jitendra Malik. Shape, illumination, and reflectance from shading. IEEE transactions on pattern analysis and machine intelligence, 37(8):1670–1687, 2014. 2

[2] Yeonjin Chang, Yearim Kim, Seunghyeon Seo, Jung Yi, and Nojun Kwak. Fast sun-aligned outdoor scene relighting based on tensorf. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 3626–3636, 2024. 2, 6

[3] Xingyu Chen, Qi Zhang, Xiaoyu Li, Yue Chen, Ying Feng, Xuan Wang, and Jue Wang. Hallucinated neural radiance fields in the wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 12943–12952, 2022. 2

[4] Hiba Dahmani, Moussab Bennehar, Nathan Piasco, Luis Roldao, and Dzmitry Tsishkou. Swag: Splatting in the wild images with appearance-conditioned gaussians. In European Conference on Computer Vision, pages 325–340. Springer, 2025. 2, 7

[5] Sylvain Duchêne, Clement Riant, Gaurav Chaurasia, Jorge Lopez-Moreno, Pierre-Yves Laffont, Stefan Popov, Adrien Bousseau, and George Drettakis. Multi-view intrinsic images of outdoors scenes with an application to relighting. ACM Transactions on Graphics, page 16, 2015. 2

[6] 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 preprint arXiv:2311.16043, 2023. 2

[7] 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, 2025. 1, 2, 6, 7, 8

[8] Robin Green. Spherical harmonic lighting: The gritty details. In Archives of the game developers conference, volume 56, page 4, 2003. 3

[9] Antoine Guédon and Vincent Lepetit. 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, pages 5354–5363, 2024. 2

[10] Tom Haber, Christian Fuchs, Philippe Bekaer, Hans-Peter Seidel, Michael Goesele, and Hendrik PA Lensch. Relighting objects from image collections. In 2009 IEEE Conference on Computer Vision and Pattern Recognition, pages 627–634. IEEE, 2009. 2

[11] Lei He, Leheng Li, Wenchao Sun, Zeyu Han, Yichen Liu, Sifa Zheng, Jianqiang Wang, and Keqiang Li. Neural radiance field in autonomous driving: A survey. arXiv preprint arXiv:2404.13816, 2024. 2

[12] 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. 2, 3, 4, 5, 6, 8

[13] 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. 2, 3, 4

[14] Jean-François Lalonde, Alexei A Efros, and Srinivasa G Narasimhan. Webcam clip art: Appearance and illuminant transfer from time-lapse sequences. ACM Transactions on Graphics (TOG), 28(5):1–10, 2009. 2

[15] Peihao Li, Shaohui Wang, Chen Yang, Bingbing Liu, Weichao Qiu, and Haoqian Wang. Nerf-ms: Neural radiance fields with multi-sequence. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 18591–18600, 2023. 2

[16] Quewei Li, Jie Guo, Yang Fei, Feichao Li, and Yanwen Guo. Neulighting: Neural lighting for free viewpoint outdoor scene relighting with unconstrained photo collections. In SIGGRAPH Asia 2022 Conference Papers, pages 1–9, 2022. 2

[17] 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. 2

[18] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 7210–7219, 2021. 1, 2

[19] 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. 1, 2

[20] Julien Philip, Michaël Gharbi, Tinghui Zhou, Alexei A Efros, and George Drettakis. Multi-view relighting using a geometry-aware network. ACM Trans. Graph., 38(4):78–1, 2019. 2, 6

[21] Ava Pun, Gary Sun, Jingkang Wang, Yun Chen, Ze Yang, Sivabalan Manivasagam, Wei-Chiu Ma, and Raquel Urtasun. Neural lighting simulation for urban scenes. In Thirty-seventh Conference on Neural Information Processing Systems, 2023. 2

[22] Ravi Ramamoorthi and Pat Hanrahan. An efficient representation for irradiance environment maps. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pages 497–500, 2001. 1, 2, 3, 4

[23] 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. 1, 2, 5, 6, 7

[24] Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, and Giljoo Nam. Relightable gaussian codec avatars. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 130–141, 2024. 2

[25] Yahao Shi, Yanmin Wu, Chenming Wu, Xing Liu, Chen Zhao, Haocheng Feng, Jingtuo Liu, Liangjun Zhang, Jian Zhang, Bin Zhou, et al. Gir: 3d gaussian inverse ren-

dering for relightable scene factorization. arXiv preprint arXiv:2312.05133, 2023. 2

[26] Marcos Paulo Berteli Slomp, Manuel Menezes de Oliveira Neto, and Diego Inácio Patrício. A gentle introduction to precomputed radiance transfer. Revista de informática teórica e aplicada. Porto Alegre. Vol. 13, n. 2 (2006), p. 131-160, 2006. 3

[27] 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. 2

[28] 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, SA '23, New York, NY, USA, 2023. Association for Computing Machinery. 2, 6

[29] Tiancheng Sun, Kai-En Lin, Sai Bi, Zexiang Xu, and Ravi Ramamoorthi. Nelf: Neural light-transport field for portrait view synthesis and relighting. In Eurographics Symposium on Rendering, 2021. 2

[30] Kalyan Sunkavalli, Wojciech Matusik, Hanspeter Pfister, and Szymon Rusinkiewicz. Factored time-lapse video. In ACM SIGGRAPH 2007 papers, pages 101–es. 2007. 2

[31] Alejandro Troccoli and Peter K Allen. Relighting acquired models of outdoor scenes. In Fifth International Conference on 3-D Digital Imaging and Modeling (3DIM'05), pages 245–252. IEEE, 2005. 2

[32] Yuze Wang, Junyi Wang, and Yue Qi. We-gs: An in-the-wild efficient 3d gaussian representation for unconstrained photo collections. arXiv preprint arXiv:2406.02407, 2024. 2

[33] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4):600–612, 2004. 5

[34] Zian Wang, Tianchang Shen, Jun Gao, Shengyu Huang, Jacob Munkberg, Jon Hasselgren, Zan Gojcic, Wenzheng Chen, and Sanja Fidler. Neural fields meet explicit geometric representations for inverse rendering of urban scenes. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2023. 1, 2, 6

[35] Turner Whitted. An improved illumination model for shaded display. In Proceedings of the 6th annual conference on Computer graphics and interactive techniques, page 14, 1979. 1

[36] Guanyu Xing, Xuehong Zhou, Qunsheng Peng, Yanli Liu, and Xueying Qin. Lighting simulation of augmented outdoor scene based on a legacy photograph. In Computer Graphics Forum, volume 32, pages 101–110. Wiley Online Library, 2013. 2

[37] Jiacong Xu, Yiqun Mei, and Vishal M Patel. Wild-gs: Real-time novel view synthesis from unconstrained photo collections. arXiv preprint arXiv:2406.10373, 2024. 2, 7

[38] Siqi Yang, Xuanning Cui, Yongjie Zhu, Jiajun Tang, Si Li, Zhaofei Yu, and Boxin Shi. Complementary intrinsics from neural radiance fields and cnns for outdoor scene relighting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 16600–16609, 2023. 2

[39] Yifan Yang, Shuhai Zhang, Zixiong Huang, Yubing Zhang, and Mingkui Tan. Cross-ray neural radiance fields for novel-view synthesis from unconstrained image collections. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 15901–15911, 2023. 2

[40] 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. 2, 6

[41] Chong Zeng, Guojun Chen, Yue Dong, Pieter Peers, Hongzhi Wu, and Xin Tong. Relighting neural radiance fields with shadow and highlight hints. In ACM SIGGRAPH 2023 Conference Proceedings, 2023. 2

[42] Dongbin Zhang, Chuming Wang, Weitao Wang, Peihao Li, Minghan Qin, and Haoqian Wang. Gaussian in the wild: 3d gaussian splatting for unconstrained image collections. In European Conference on Computer Vision, pages 341–359. Springer, 2025. 2, 4, 7

[43] Libo Zhang, Yuxuan Han, Wenbin Lin, Jingwang Ling, and Feng Xu. Prtgaussian: Efficient relighting using 3d gaussians with precomputed radiance transfer, 2024. 2

[44] 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. 2

[45] Matthias Zwicker, Hanspeter Pfister, Jeroen Van Baar, and Markus Gross. Ewa volume splatting. In Proceedings Visualization, 2001. VIS'01., pages 29–538. IEEE, 2001. 3