

SOL-NeRF: Sunlight Modeling for Outdoor Scene Decomposition and Relighting

Jia-Mu Sun

Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, CAS and University of Chinese Academy of Sciences
China
sunjiamu21s@ict.ac.cn

Yong-Liang Yang

Department of Computer Science
University of Bath
United Kingdom
y.yang@cs.bath.ac.uk

Tong Wu

Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, CAS and University of Chinese Academy of Sciences
China
wutong19s@ict.ac.cn

Yu-Kun Lai

School of Computer Science and Informatics
Cardiff University
United Kingdom
LaiY4@cardiff.ac.uk

Lin Gao*

Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, CAS and University of Chinese Academy of Sciences
China
gaolin@ict.ac.cn

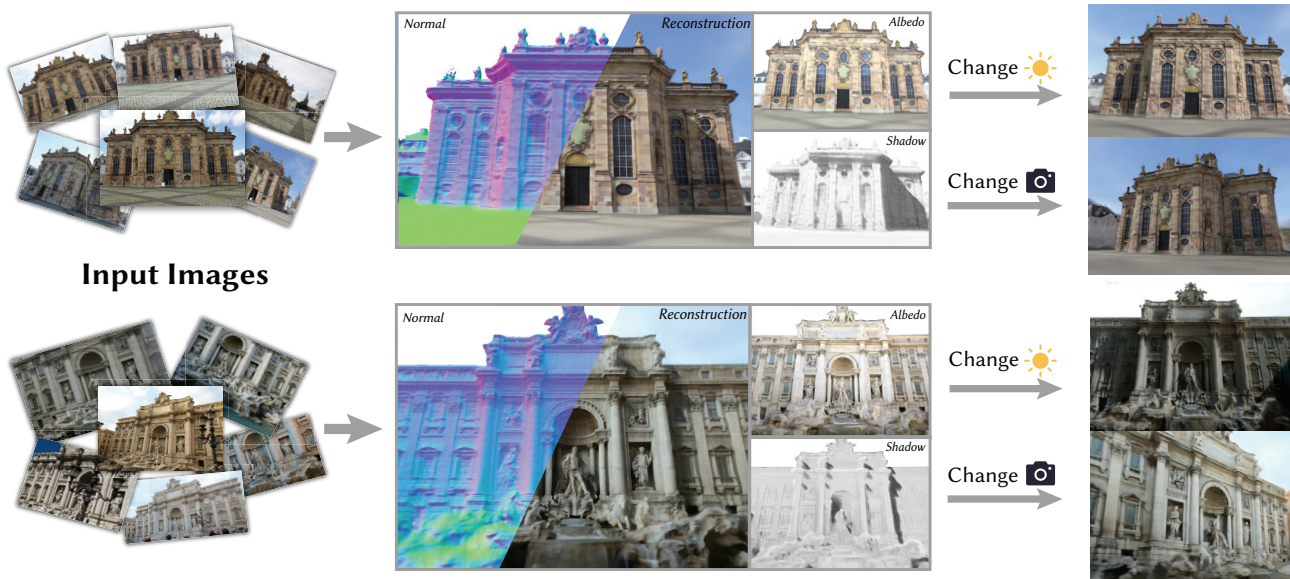


Figure 1: Given a set of input images of an outdoor scene, our SOL-NeRF pipeline decomposes them into geometry and material properties, which enables rendering the input scene from a novel viewpoint and relighting it with a different illumination.

*Corresponding author is Lin Gao (gaolin@ict.ac.cn).

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SA Conference Papers '23, December 12–15, 2023, Sydney, NSW, Australia
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0315-7/23/12.
<https://doi.org/10.1145/3610548.3618143>

ABSTRACT

Outdoor scenes often involve large-scale geometry and complex unknown lighting conditions, making it difficult to decompose them into geometry, reflectance and illumination. Recently researchers made attempts to decompose outdoor scenes using Neural Radiance Fields (NeRF) and learning-based lighting and shadow representations. However, diverse lighting conditions and shadows in outdoor scenes are challenging for learning-based models. Moreover, existing methods may produce rough geometry and normal reconstruction and introduce notable shading artifacts when the scene is rendered under a novel illumination. To solve the above problems, we propose SOL-NeRF to decompose outdoor scenes with the help of a hybrid lighting representation and a signed distance field geometry reconstruction. We use a single Spherical Gaussian (SG) lobe to approximate the sun lighting, and a first-order Spherical Harmonic (SH) mixture to resemble the sky lighting. This hybrid representation is specifically designed for outdoor settings, and compactly models the outdoor lighting, ensuring robustness and efficiency. The shadow of the direct sun lighting can be obtained by casting the ray against the mesh extracted from the signed distance field, and the remaining shadow can be approximated by Ambient Occlusion (AO). Additionally, sun lighting color prior and a relaxed Manhattan-world assumption can be further applied to boost decomposition and relighting performance. When changing the lighting condition, our method can produce consistent relighting results with correct shadow effects. Experiments conducted on our hybrid lighting scheme and the entire decomposition pipeline show that our method achieves better reconstruction, decomposition, and relighting performance compared to previous methods both quantitatively and qualitatively.

CCS CONCEPTS

• **Computing methodologies** → *Image-based rendering*.

KEYWORDS

neural radiance fields, inverse rendering, outdoor scene reconstruction

ACM Reference Format:

Jia-Mu Sun, Tong Wu, Yong-Liang Yang, Yu-Kun Lai, and Lin Gao. 2023. SOL-NeRF: Sunlight Modeling for Outdoor Scene Decomposition and Relighting. In *SIGGRAPH Asia 2023 Conference Papers (SA Conference Papers '23)*, December 12–15, 2023, Sydney, NSW, Australia. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3610548.3618143>

1 INTRODUCTION

Neural Radiance Fields (NeRF) [Mildenhall et al. 2020] that are learnable in the differentiable rendering process have become increasingly important for inverse rendering tasks. Several pioneering works [Oechsle et al. 2021; Wang et al. 2021; Yariv et al. 2021] learn to reconstruct the geometry of the input scene with volume rendering based on NeRF and implicit surface representations. While achieving good-quality geometry and appearance on relatively small scenes under a fixed lighting condition, their performance drops on larger scenes with varying illuminations, and none of them

can relight the input scene. This is because they learn the appearance using a single MLP (Multi-layer Perceptron) that entangles material, lighting and shadow.

To reconstruct outdoor scenes with varying lighting conditions, several methods have been proposed. NeRF-W [Martin-Brualla et al. 2021] splits the original NeRF network into two: a static one and a transient one, where the latter is controlled by a per-frame appearance embedding. This strategy successfully peels off the lighting effect. However, it is still an implicit illumination representation, rather than an explicit decomposition of material and lighting. Therefore, one can only relight the scene by interpolating the latent codes representing two lighting conditions. NeuralRecon-W [Sun et al. 2022] also uses latent embeddings to model different appearances across images while focusing on geometry reconstruction. By utilizing a surface-based sampling strategy, it can efficiently constrain the samples to be around the surface, yielding satisfactory reconstruction results. Nonetheless, NeuralRecon-W is incapable of relighting also due to its implicit lighting representation.

Unlike the above methods, NeRF-OSR [Rudnev et al. 2022] explicitly decomposes the lighting and material of outdoor scenes. It extends the original NeRF by modeling the lighting with a second-order Spherical Harmonics (SH) mixture. The material is predicted using an albedo network, while shadows are estimated by a shadow network that takes the SH coefficients as input. The SH lighting for an image is predicted by a learnable embedding for each image. NeRF-OSR shows promising decomposition and relighting results, but the learned shadow network can easily overfit the input images. This is because it merely outputs a scalar value (between 0 and 1) that suppresses the non-shadowed results in the image space, rather than explicitly computing shadow effects from lighting and geometry. Very recently, FEGR [Wang et al. 2023] proposed to use hash grid encoding [Müller et al. 2022] to model scene properties while utilizing explicit ray tracing to obtain high-order effects such as shadows. It models lighting with a lighting prediction network that predicts the lighting intensity from a specific direction. However, the lighting network may predict an inaccurate illumination as it learns a single lighting model which mixes high-intensity sunlight and low-intensity sky light. As a result, the learned sunlight may have an irregular shape and the intensity distribution of the sky can be less accurate.

Aiming at high-quality scene decomposition and relighting, we propose a novel method called SOL-NeRF (Sunlight Modeling for Outdoor Scene Decomposition and ReLighting) with a hybrid lighting representation enhanced by geometry and lighting priors (see Fig. 1). To address the complication of outdoor lighting with varying intensity distributions, we use a hybrid representation based on Spherical Gaussian (SG) and Spherical Harmonics (SH) to model sunlight and sky light respectively. The proposed hybrid representation also allows more accurate and reliable estimation of shadow effects, including explicit ray-tracing using SG, and learned Ambient Occlusion (AO) using SH. To enable quality and efficient geometry reconstruction, we employ the octree-accelerated sampling strategy [Sun et al. 2022], and transform signed distance values to opacity values for volume rendering using the formulation in [Wang et al. 2021]. The learned Signed Distance Field (SDF) also provides necessary inside-outside information for the shadow calculation. Moreover, we utilize essential priors to resolve

ambiguities and enhance the performance of scene decomposition. More specifically, we analytically link the color of sunlight with the elevation of the sun in the sky dome, making the learned lighting more accurate. A relaxed Manhattan world assumption adapted from [Furukawa et al. 2009] is used to improve the surface normal distribution, thus improving the reconstructed scene geometry.

Overall, our contribution can be summarized as follows:

- We propose SOL-NeRF, a pipeline that decomposes outdoor scenes into geometry, reflectance and lighting. It can efficiently extract high-quality geometry and reflectance from a set of images of outdoor scenes under varying illuminations.
- We introduce a novel hybrid lighting representation for skylights in the daytime composed of both Spherical Gaussian and Spherical Harmonics. This representation can approximate shadows with a low computation overhead, which is essential for inverse rendering applications.
- Experiments show that the proposed lighting representation can efficiently model various lighting conditions and built upon this lighting representation, SOL-NeRF can reconstruct and relight outdoor scenes more faithfully compared with existing methods.

2 RELATED WORK

2.1 Neural Inverse Rendering.

Inverse rendering [Ramamoorthi and Hanrahan 2001b; Sato et al. 1997] is a fundamental problem in computer graphics. It aims to decompose the geometry, material, and lighting from observed images. Early works [Gao et al. 2020; Yu et al. 2020] often involve intrinsic images or geometry proxies to decompose a scene. Recently, with the emergence of the implicit representation and neural radiance fields (NeRF), a number of works proposed to learn the underlying geometry, material and lighting from given observations. NeRD [Boss et al. 2021a] is the pioneering work that decomposes material and geometry with given lighting conditions. It takes direct lighting and one-bounce indirect lighting into consideration and models the visibility with a neural network. Later, PhySG [Zhang et al. 2021a] works under an unknown lighting condition and uses signed distance fields (SDF) and Spherical Gaussian (SG) functions as geometry and lighting representation respectively but assumes the whole scene shares the same roughness and specular components. Based on PhySG, InvRender [Zhang et al. 2022] introduces another set of SG functions to model indirect illumination. Boss et al. [2021a] also use SG as its representation and adopt a two-stage training strategy by first optimizing a sampling network and then a decomposition network. To better predict material, Neural-PIL [Boss et al. 2021b] builds a smooth material autoencoder to predict the material parameters. NeRFactor [Zhang et al. 2021b] also trains a learned BRDF autoencoder and further separates shadows from it with a visibility network. NeROIC [Kuang et al. 2022] separates the static appearance and the transient appearance with two branches similar to NeRF-W [Martin-Brualla et al. 2021] and uses Spherical Harmonic (SH) functions to approximate lighting. To model high-frequency lighting beyond the capability of SG and SH, NvDiffRec [Munkberg et al. 2022] and NvDiffRecMC [Hasselgren et al. 2022] propose to model it with a large environment

image and use DMTet [Shen et al. 2021] as its geometry representation. These methods mainly focus on decomposing different components from a scene with a few objects under synthetic or real illumination, and only a small number of methods target *outdoor* scenes. NeRF-OSR [Rudnev et al. 2022] is the first work that attempts decomposition and relighting on outdoor scenes with NeRF. NeuLighting [Li et al. 2022a] directly predicts the scene properties from latent codes extracted from the input images. FEGR [Wang et al. 2023] further improves the decomposition quality by calculating ray intersections with the explicit surface extracted from an SDF network.

2.2 Scene-Level Reconstruction with Neural Rendering.

With the popularity of implicit representations [Chen and Zhang 2019; Mescheder et al. 2019; Park et al. 2019] and neural surface rendering [Thies et al. 2019; Yariv et al. 2020], a few methods [Guo et al. 2022; Sun et al. 2021] start to work on scene-level reconstruction by optimizing the underlying implicit geometric representation and the appearance network. Later, neural volumetric rendering becomes mainstream and several works build the relationship between implicit representations and volume rendering. NeuS [Wang et al. 2021] and VOLSDF [Yariv et al. 2021] transform signed distance values to density values in volume rendering. UNISURF [Oechsle et al. 2021] replaces alpha values in volume rendering with occupancy values. MonoSDF [Yu et al. 2022] and NeuRIS [Wang et al. 2022] utilize the normal prior learned from a 2D normal estimation network to guide the geometry reconstruction. Apart from these methods focusing on objects and indoor scenes, NeuralReconW [Sun et al. 2022] works on outdoor scenes and improves the sampling strategy with an octree with the point cloud produced by Structure-from-Motion (SfM) methods [Schönberger and Frahm 2016; Schönberger et al. 2016]. FEGR [Wang et al. 2023] also utilizes NeuS’s formulation and accelerates the reconstruction process with a multi-resolution hash table [Müller et al. 2022].

2.3 Outdoor Lighting Estimation.

For outdoor lighting estimation, there are two commonly used settings: single-view lighting estimation and multi-view lighting estimation. Given a single view observation, previous methods [Hold-Geoffroy et al. 2019, 2017; Song and Funkhouser 2019; Zhang et al. 2019; Zhu et al. 2021] propose feed-forward networks to predict the sun’s location and the sky intensity. However, single-view lighting estimation is extremely ill-posed, which makes it hard to predict an accurate sun location and sky intensity. Under the multi-view setting, Duchene et al. [2015] use a progressive method to predict both the scene material and lighting, while Philip et al. [2019] directly use a multi-view dataset and target lighting to relighting outdoor scenes. NeRF-W [Martin-Brualla et al. 2021] models the lighting implicitly with a learnable feature vector for each input image. NeRF-OSR [Rudnev et al. 2022] explicitly approximates the outdoor lighting with SH functions. FEGR [Wang et al. 2023] instead models the lighting intensity with a neural network from a specific direction. However, this lighting formulation mixes the high-intensity sunlight and the low-intensity sky light, and thus may predict the sun with an irregular shape and the sky with an

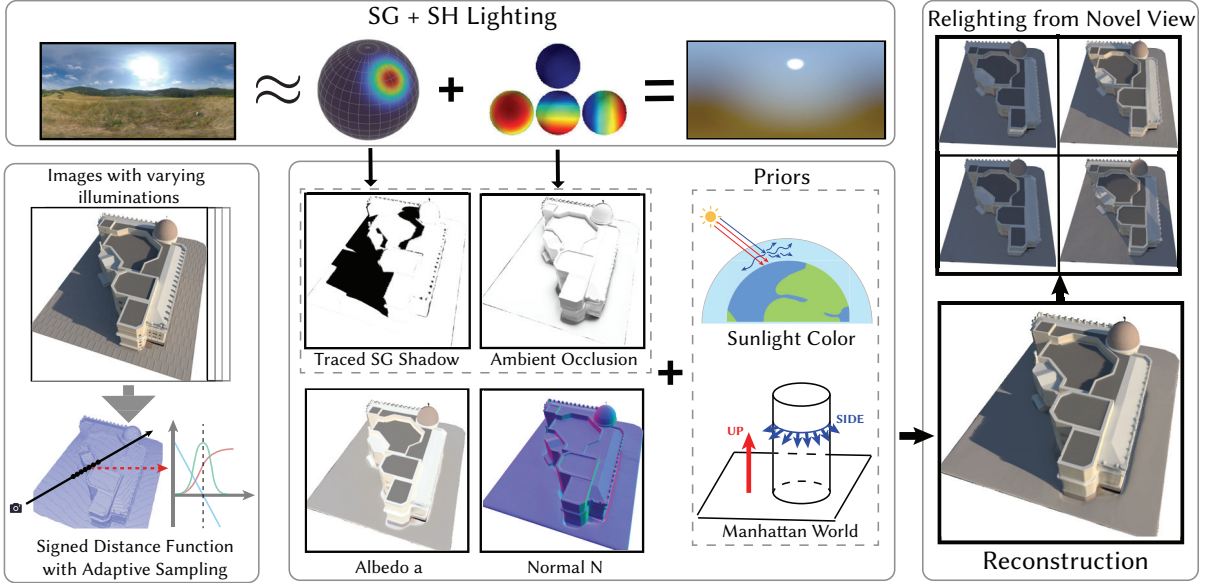


Figure 2: The overview of SOL-NeRF pipeline. Given a set of images under multiple different lighting conditions, we model the scene’s geometry with a signed distance field (SDF) and apply an adaptive sampling strategy in the neural volume rendering process. To decompose geometry, material, shadow and lighting, we predict the diffuse albedo a with an MLP network and the normal is derived by the gradient of the SDF network. Our lighting is composed of a Spherical Gaussian (SG) function and the first-order Spherical Harmonic (SH) functions. The SG function is responsible for high-intensity lighting like the sun while the SH functions are designed to represent relatively low-intensity lighting like the sky light. We consider both the shadow cast by the directional SG light and the ambient occlusion. To enhance decomposition quality, we introduce priors for the sunlight color and geometry. Overall, SOL-NeRF enables realistic reconstruction and relighting the input scene under a novel lighting condition.

inaccurate intensity distribution. Different from NeRF-OSR and FEGR, we separately model the high-intensity sunlight and the low-intensity sky light with an SG function and a first-order SH respectively to reduce the learning difficulty.

3 METHOD

We propose SOL-NeRF, an implicit scene representation based method that decomposes scene properties from multi-view images. Our method is briefly described with illustrations in Fig. 2. The rest of the section elaborates on the key components and is organized as follows. Firstly, we introduce some of the preliminaries, including the formulation of neural rendering and NeuS [Wang et al. 2021], the rendering equation, and the sampling strategy adopted from NeuralRecon-W [Sun et al. 2022] (Sec. 3.1). Secondly, we elaborate on our novel hybrid lighting representation and present the rendering method accordingly (Sec. 3.2). Thirdly, we discuss how to cast shadows under our hybrid lighting representation (Sec. 3.3). Finally, we introduce additional priors that improve the decomposition and relighting performances of our method (Sec. 3.4).

3.1 Preliminaries

3.1.1 Neural Rendering based on an Implicit Representation. NeRF uses a density field σ to model geometry. At rendering time, the renderer generates a ray per pixel and samples with distances t_i along

the ray. The final color of the pixel is obtained through discrete integration $C = \sum_i^N w_i c_i$, where N is the number of sampled points along a ray and $w_i = \exp(\sum_{j<i} \sigma_i(t_{j+1} - t_j))(1 - \exp(-\sigma_i(t_{j+1} - t_j)))$. However, the density field cannot accurately express the surface geometry. To tackle this problem, NeuS [Wang et al. 2021] models geometry with a signed distance field (SDF) function f_{sdf} and transforms the signed distance value to the density value with $\sigma(t) = \max\left(-\frac{d\Phi_s}{dt}(f_{sdf}(x(t))), 0\right)$, where $\Phi_s(x) = (1 + e^{-sx})^{-1}$ and s is a trainable deviation parameter.

3.1.2 The rendering equation. Rather than directly using the output of the color network as in NeRF, we perform shading based on the rendering equation [Kajiya 1986] for every sample point \mathbf{x}_i . We omit the specular lobe of the BRDF, and only preserve the diffuse one with direct lighting to calculate the color of the sample point:

$$c_i = \frac{1}{\pi} \mathbf{a}_i \int_S V(\mathbf{x}_i, \omega_j) L_{in}(\mathbf{x}_i, \omega_j) \max(\hat{\mathbf{N}}_i \cdot \omega_j, 0) d\omega_j \quad (1)$$

where \mathbf{a}_i and $\hat{\mathbf{N}}_i$ denote albedo and surface normal for point \mathbf{x}_i , $V(\mathbf{x}_i, \omega_j)$ denotes visibility at \mathbf{x}_i from direction ω_j , $L_{in}(\mathbf{x}_i, \omega_j)$ denotes incoming radiance. L_{in} is approximated by the hybrid lighting of SG and SH (Sec. 3.2).

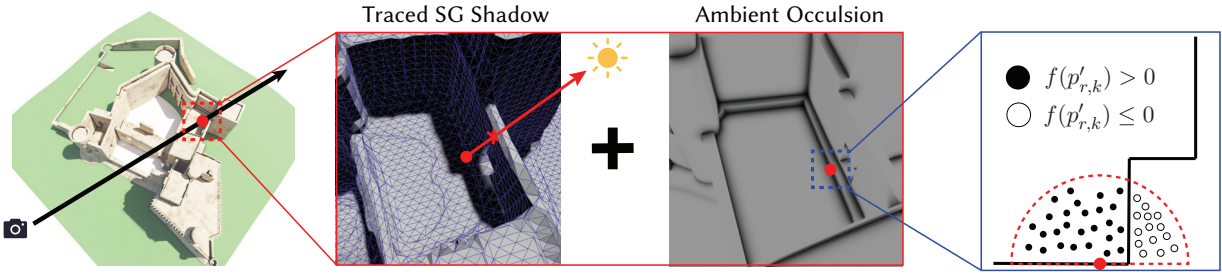


Figure 3: The calculation of shadow under the proposed hybrid lighting. The final shadow consists of both the SG shadow and the shadow caused by ambient occlusion. Specifically, we calculate the SG shadow by casting the SG lighting direction to the extracted mesh. The ambient occlusion value of a surface point p_r is determined by the ratio of points outside the surface and the total sample points $\{p'_{r,k}\}$ in a hemisphere centered at p_r . Note that ambient occlusion is calculated by directly querying the SDF, while traced shadow is based on extracted mesh.

3.1.3 Sampling Strategy. As an outdoor scene is often at a large scale, there are usually plenty of empty spaces without any geometric object in the scene, making the standard importance sampling-based rendering method struggle to infer the real geometry. Hence, we adopt the sampling strategy proposed by NeuralRecon-W [Sun et al. 2022], which builds an octree based on the point cloud generated by SfM, and stores the SDF values into the octree. The SDF values can efficiently guide the sampling process, boosting both the training speed and geometry quality.

3.2 Hybrid Lighting Representation

Our method is dedicated to outdoor scenes, which are mostly captured during cloudy or sunny daytime. Hence we propose a lighting representation specially optimized to model daylight, which can also handle scenes without direct sunlight as a degenerate case. The upper part of Fig. 2 illustrates our lighting formulation. The basic idea is to separate the daylight into two components: 1) *direct sunlight*, which is a highly concentrated and bright ‘local’ light source. 2) *sky light*, which is a more uniformly distributed ‘global’ light source on the sky dome. Actually, both components are effects of sunshine. The sky light is simply the scattered sunlight, the color of which can be calculated by Reyleigh and Mie Scattering [Tyndall 1869].

However, directly calculating the whole sky and its lighting effect at every sample point using Reyleigh and Mie Scattering equations is expensive. Instead, we model the two components based on two lighting representations: 1) a Spherical Gaussian (SG) illumination model to represent direct sunlight, and 2) a Spherical Harmonics (SH) illumination model to represent sky light.

3.2.1 The Formulation and Rendering of SG and SH. Spherical Gaussians are Gaussian functions defined on spheres: $SG(\mathbf{v}; \boldsymbol{\mu}, \lambda, \mathbf{l}) = \mathbf{l}e^{\lambda(\boldsymbol{\mu} \cdot \mathbf{v} - 1)}$, where \mathbf{v} is the function input representing the light direction, $\boldsymbol{\mu}$ and λ control the center and deviation, and \mathbf{l} controls the peak value. We use 7 learnable parameters (3 for peak intensity \mathbf{l} , 3 for direction $\boldsymbol{\mu}$, and 1 for deviation λ) to represent the SG lobe and employ the same rendering equation as [Wang et al. 2009]. As for Spherical Harmonics [Ramamoorthi and Hanrahan 2001a], it uses a set of basis functions $Y_l^m(\mathbf{v})$, where $-l \leq m \leq l$ to

decompose a global lighting distribution on a sphere. Here we use the first-order SH mixture ($l = 0, 1$), which reduces the parameters needed to represent the sky light, and is beneficial to learning the near-uniform sky light. The first-order SH has 4 lobes, meaning we need 12 parameters for SH.

3.2.2 Motivation of the Hybrid Representation. The separation modeling of sunlight and sky light using SG and SH has several advantages: 1) The rendering with SG and SH illumination is memory-efficient and fast because only 7 SG and 12 SH parameters are required and the shading calculation is close-form. 2) It separates the high-intensity sunlight and the relatively low-intensity sky light to reduce the learning difficulty of the complex outdoor illumination. 3) Casting shadows under this representation is efficient since only one ray-mesh intersection test is needed (see Sec. 3.3). The explicit shadow calculation can produce more realistic relighting results compared with the implicitly learned shadow in NeRF-OSR [Rudnev et al. 2022] that may overfit input images (see Sec. 4).

3.3 Shadow Calculation

As a direct effect of the interaction between light and geometry, shadow is largely influenced by the lighting. We calculate the shadow according to the SG and SH components of the lighting, respectively. We show the shadow calculation process in Fig. 3.

3.3.1 Ray Tracing for SG Shadow. The proposal of SG illumination is to simulate direct sunlight. The sunlight is nearly a directional light source given its highly concentrated nature. Consequently, the shadow cast by the sunlight can be simply obtained by tracing the light ray reversely (from the shading point along the reversed SG direction). To accelerate the ray tracing, we construct an octree based on the reconstructed geometry by f_{sdf} .

3.3.2 Learned Ambient Occlusion for SH Shadow. The SH illumination is utilized to approximate the near-uniform sky light. If we further assume the sky light is uniformly distributed, what we get is exactly the so-called ‘ambient light’ in the real-time rendering context. Thus, we adapt the Screen Space Ambient Occlusion (SSAO) [Mittring 2007; Shanmugam and Arikan 2007] techniques

to our SDF-based geometry representation. For every ray, we get the surface point $\mathbf{p}_r = \mathbf{o}_r + t_r \mathbf{d}_r$ with the corresponding normal \mathbf{N}_r , where \mathbf{o}_r and \mathbf{d}_r denote the origin and direction of the r -th ray, and $t_r = \sum_{i=1}^N w_i t_i$ denotes the depth estimated by volume rendering. We sample M points $\mathbf{p}'_{r,1}, \mathbf{p}'_{r,2}, \dots, \mathbf{p}'_{r,M}$ within the hemisphere centered at \mathbf{p}_r and with upper direction \mathbf{N}_r . The SDF value of $\mathbf{p}'_{r,k}$ is queried, and the final ambient occlusion shadow is calculated by:

$$AO(\mathbf{p}_r) = \frac{1}{M} \sum_{k=1}^M \mathbb{I}(f_{sdf}(\mathbf{p}'_{r,k}) > 0), \quad (2)$$

where $\mathbb{I}(\ast)$ is an indicator function, and $f_{sdf}(\ast)$ is the learned SDF function. However, the sampled ambient occlusion $AO(\cdot)$ will cause noises since it is an approximate function based on random sampling. So we further define an MLP f_{ao} as a smooth, learnable model to better estimate ambient occlusion, which is optimized during end-to-end training with data loss L_{ao} and regularization loss L_{aor} :

$$L_{ao} = \left| \sum_{i=1}^N w_i f_{ao}(\mathbf{x}_i) - AO(\mathbf{p}_r) \right| \quad L_{aor} = \left| \sum_{i=1}^N w_i f_{ao}(\mathbf{x}_i) - 1 \right| \quad (3)$$

Experiments in Fig. 11 show the learned f_{ao} is smoother than the original sampled $AO(\cdot)$ and can boost reconstruction results.

Now we have introduced both the lighting representation and the shadow calculation. Eq. 1 can be written as follows:

$$\begin{aligned} \mathbf{c}_i = & \frac{1}{\pi} \mathbf{a}_i \int_{\mathcal{S}} V_t(\mathbf{x}_i, \boldsymbol{\omega}) SG(\boldsymbol{\omega}; \mu, \lambda, \mathbf{l}) \max(\hat{\mathbf{N}}_i \cdot \boldsymbol{\omega}, 0) d\boldsymbol{\omega} \\ & + \frac{1}{\pi} \mathbf{a}_i f_{ao}(\mathbf{x}_i) \int_{\mathcal{S}} \sum_{l,m} y_l^m Y_l^m(\boldsymbol{\omega}) \max(\hat{\mathbf{N}}_i \cdot \boldsymbol{\omega}, 0) d\boldsymbol{\omega}, \end{aligned} \quad (4)$$

where $V_t(\mathbf{x}_i, \boldsymbol{\omega})$ is the intersection test result between the reconstructed mesh and the ray $\mathbf{x}_i + t\boldsymbol{\omega}$. It equals 0 when the ray-mesh intersection exists. y_l^m is the mixture weight of spherical harmonics basis functions and \mathbf{l} is the color of the sun.

3.4 Priors

In addition to the proposed lighting representation, we also exploit two priors to further improve the performance of SOL-NeRF, including the sunlight color prior and relaxed Manhattan world geometry prior.

3.4.1 Sunlight Color Prior. The proposed hybrid lighting representation is under-constrained regarding the sunlight color. In the real world, the color of the sunlight is closely related to sun elevation angle and altitude (see Fig. 4). Various models have been proposed to model the sunlight and sky color with respect to these parameters [Hosek and Wilkie 2012; Nishita et al. 1993; Preetham et al. 1999]. We adopt the model proposed in [Nishita et al. 1993], since it is purely based on physical calculations. However, if we calculate the color at rendering time, the computation overhead makes training impractical. We choose to fit the R,G,B channels of the light with an analytical function $\mathbf{a}_{sun} = f_{sun}(\theta)$, where f_{sun} is a piecewise polynomial function and θ is the elevation angle of the sun. However, this model only considers the clear sky with a certain density of air and water droplets. In practice, there are many factors (e.g., location-dependent air and water droplets density, cloud and mist, and errors of the imaging system) that make this model

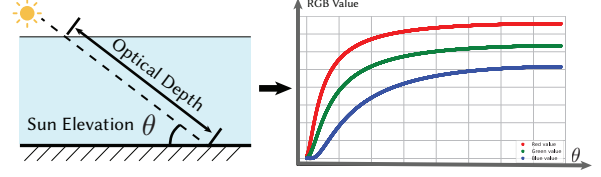


Figure 4: The sunlight color is related to the sun elevation θ . According to the learned sun elevation θ , atmosphere thickness, and air density, the optical depth is calculated. Then the color of the sunlight that reaches the surface is obtained using wavelength-dependent Rayleigh and Mie Scattering laws [Nishita et al. 1993; Tyndall 1869]. Using this method we can associate every θ with a color. This function is approximated using a polynomial that can be evaluated efficiently in training time.

inaccurate. To make up for this, we modify the final color of the sun SG to $\mathbf{l} = \mathbf{c} \odot \mathbf{a}_{sun} + \mathbf{b}$, where \mathbf{c}, \mathbf{b} are learnable parameters, and are initialized to $\mathbf{1}$ and $\mathbf{0}$ respectively. For more details of the function f_{sun} , please refer to the supplementary material.

3.4.2 Relaxed Manhattan World Prior. Man-made buildings often have surfaces aligned with three orthogonal directions. This assumption is called the Manhattan world assumption [Furukawa et al. 2009] and is widely used in 3D reconstruction methods. Since the target of our outdoor scene is often buildings, we consider applying this assumption to our pipeline.

However, many non-Manhattan-style buildings do not fully obey this assumption. So we instead apply a relaxed version of the assumption: we assume the scene has one upright direction (ground normal), and most surface normals are either parallel or perpendicular to this direction. This is natural as a building usually sits on the ground, and its facade normal is naturally perpendicular to the ground normal. We formulate this prior as a loss L_{man} :

$$L_{man}(\hat{\mathbf{N}}) = \min(|\hat{\mathbf{N}} \cdot \mathbf{U}|, |\hat{\mathbf{N}} \times \mathbf{U}|_2), \quad (5)$$

where $\hat{\mathbf{N}}$ and \mathbf{U} are normalized vectors of normal and upright direction.

After introducing all the components, we formulate the overall loss function of our pipeline as follows:

$$L = L_{color} + \lambda_1 L_{reg} + \lambda_2 L_{mask} + \lambda_3 L_{ao} + \lambda_4 L_{aor} + \lambda_5 L_{man} \quad (6)$$

where the color loss L_{color} minimizes the difference between rendered color and the ground truth pixel color and is defined as:

$$L_{color} = \frac{1}{R} \sum_{r=1}^R \|C(r) - C^t(r)\|_1 \quad (7)$$

where R is the number of sampled rays in a training batch. $C(r)$ and $C^t(r)$ are the rendered pixel color and ground truth pixel color for the r th ray. $L_{reg} = \|\nabla f_{sdf}\|_2 - 1$ is the eikonal loss to allow smooth surface reconstruction. The foreground mask loss L_{mask} is the same as that in NeuS [Wang et al. 2021] and NeuralReconW [Sun et al. 2022]. For detailed hyperparameter settings, please refer to the supplementary material.

Table 1: Quantitative comparison of reconstruction results using SSIM, PSNR and MSE metrics on the synthetic dataset.

Methods	Rendered			Albedo			Normal
	PSNR \uparrow	SSIM \uparrow	MSE \downarrow	PSNR \uparrow	SSIM \uparrow	MSE \downarrow	MAE($^\circ$) \downarrow
NeRF-OSR	22.748	0.808	0.00903	20.157	0.825	0.0142	26.211
Ours	23.437	0.863	0.00610	24.891	0.827	0.00579	16.334

4 RESULTS AND EVALUATIONS

4.1 Datasets and Metrics

We evaluate on two datasets: Sites 1, 2, 3 of the *OSR* dataset from NeRF-OSR [Rudnev et al. 2022], and a synthetic dataset (denoted *synthetic*) with three scenes (Syn 1, 2, 3) collected from *Blendswap.com*. We evaluate the novel view synthesis results on the synthesis dataset and relighting results on both datasets. The ablation studies are conducted on the synthetic dataset. For the scene decomposition (including novel view synthesis) and relighting quality, we compare the PSNR, SSIM [Wang et al. 2004], and MSE (Mean Squared Error) to evaluate the similarity of the novel view rendering with the ground truth image and albedo (the ground truth albedo is only available in our synthetic dataset). For the geometry, we compare MAE (Mean Absolute Error) between the rendered and ground truth normals. We additionally evaluate the proposed hybrid lighting representation with eight novel HDR (High Dynamic Range) environment maps collected from the Blender software and *hdrmaps.com*.

4.2 Scene Decomposition

In Fig. 6 we include decomposition results for one scene (Site 3) from the *OSR* dataset and two scenes (Syn 1, 3) from the *synthetic* dataset. Our method can recover a better geometry as indicated by the normal quality. On the synthetic scenes, our decomposed albedos are less shadow-contaminated and more accurate in terms of color, while the ones produced by NeRF-OSR have shadow leftovers and appear reddish. These can be seen more clearly by the quantitative evaluation shown in Table 1. Our method has a similar performance to NeRF-OSR for novel view synthesis as a compound task under known lighting, while it outperforms NeRF-OSR in the decomposition results of albedo and normal, benefiting also the scene relighting as presented below.

4.3 Scene Relighting

In Fig. 9, we show relighting results for two scenes (Sites 1, 2) of the *OSR* dataset and one scene (Syn 2) of the synthetic dataset. For each scene, the ground truth image and two relighting results of that view are displayed. The calculated or predicted shadow is shown along with the relighting results. From Fig. 9 we can see the predicted shadow of NeRF-OSR does not effectively reflect the change of the lighting condition, but look more like a ‘monochrome’ version of the albedo. The reason behind this is the shadow prediction network only sees 200 to 300 lighting conditions, causing it to overfit these lighting settings. And the inaccurate and noisy normals produced by NeRF-OSR also negatively affect the relighting results, making the geometry surface look bumpy. Our pipeline solves these problems by utilizing the hybrid SG and SH lighting and its shadow

Table 2: Quantitative comparison of relighting results on real NeRF-OSR dataset and the synthetic dataset using SSIM, PSNR and LPIPS metrics. Results are averaged over ten different viewpoints with five different environment maps.

Methods	OSR			Synthetic		
	PSNR \uparrow	SSIM \uparrow	MSE \downarrow	PSNR \uparrow	SSIM \uparrow	MSE \downarrow
NeRF-OSR	17.314	0.576	0.025	21.776	0.689	0.017
Ours	19.001	0.683	0.0189	25.278	0.854	0.005

calculation techniques described in Sec. 3. The measures reported in Table 2 show the performance gain of our method over NeRF-OSR.

4.4 Hybrid Lighting Evaluation

Since the SG and SH hybrid lighting representation is newly proposed, we need to evaluate how well it can approximate outdoor lighting. For the lighting evaluation experiment, we collected six equirectangular environment maps, representing different conditions of sky lighting. For each environment map, we optimize our hybrid lighting representation to minimize the error between the ground truth map and the rendered map of our lighting. We then put the ground truth lighting and our optimized lighting into Blender [Community 2023], using the Cycles engine to get the rendered results of a diffuse scene under these two lighting settings. The SH lighting is directly rendered into an environment map, and the SG lighting is implemented using an explicit ‘Sun’ light source. In average, the PSNR between the rendered results of our method and ground truth reaches **32.829**. We show the results of this experiment in Fig. 10. It can be seen that the SG and SH hybrid lighting can effectively approximate various environment maps with minor errors in rendering results both visually and quantitatively.

4.5 Ablation Studies

In this subsection, we ablate several important design choices in our framework.

4.5.1 Sunlight Prior. The sunlight prior connects the elevation and color of the sun, which reduces the degree of freedom of the learnable lighting and its learning difficulty. We show the ablation results in the first two rows of Fig. 5 and Table 3. The absence of this prior can cause inaccurate albedo decomposition since we do not limit the light color. This effect is observed in other methods that model the sunlight without any constraints, like NeRF-OSR [Rudnev et al. 2022] and FEGR [Wang et al. 2023]. Moreover, the training process can be sensitive to the initial values of SG and SH parameters and may fail to reconstruct the input scene (the second row of Fig. 5).

4.5.2 Relaxed Manhattan-World Assumption. The relaxed Manhattan-world assumption regularizes the learned geometry and helps to reduce the bumpy surface artifact caused by the positional encoding. Qualitative comparisons between without and with the Manhattan-world assumption are shown in the third row of Fig. 5. It can be seen that the reconstructed normal is smoother when the Manhattan-world assumption is applied while the learned normal may be bumpy without this regularization. We also report the MAE between reconstructed normals and the ground truth normals

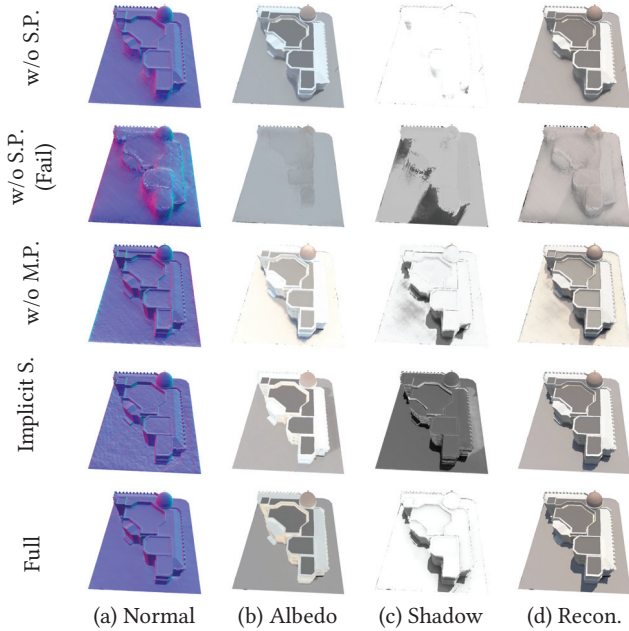


Figure 5: Qualitative comparison of decomposition results between the full pipeline (Full) and the baselines including without sunlight prior (w/o S.P.), without Manhattan-world prior (w/o M.P.), and predicting shadow with an MLP network implicitly (Implicit S.).

Table 3: Quantitative comparison of reconstruction results between the baseline (w/o S.P.) without sunlight color prior and our method (w/ S.P.) using SSIM, PSNR, and MSE metrics on Syn 3 scene.

Methods	Rendered			Albedo			Normal
	PSNR \uparrow	SSIM \uparrow	MSE \downarrow	PSNR \uparrow	SSIM \uparrow	MSE \downarrow	MAE($^\circ$) \downarrow
w/o S.P.	24.241	0.879	0.00377	22.041	0.904	0.00625	19.364
w/ S.P.	25.532	0.887	0.00279	27.905	0.941	0.00162	16.205

in Table 4, which indicates this assumption can enhance normal reconstruction.

4.5.3 Shadow Calculation Strategy. As mentioned in Sec. 3.3, the shadow is approximated by explicitly casting rays to the mesh extracted by the SDF network and the ambient occlusion. Another possible shadow calculation strategy is to predict it from lighting conditions similar to NeRF-OSR. Specifically, we set up another MLP network to predict shadows from the SG and SH parameters in our framework. We compare this strategy (Implicit S) with our method (Explicit S) on the scene decomposition task in the fourth row of Fig. 5 and the relighting task in Fig. 8. It can be seen that the implicit shadow strategy fails to correctly decompose shadow and render the input scene under novel lighting conditions.

4.5.4 Ambient Occlusion Calculation Strategy. As mentioned in Sec. 3.3, we use an MLP network f_{ao} to predict ambient occlusion

Table 4: Quantitative comparison of reconstructed normals by our method (w/ M.P.) and the baseline (w/o M.P.) without the Manhattan-world assumption using the MAE metric on Syn 3 scene.

	w/o M.P.	w/ M.P.
Normal MAE ($^\circ$) \downarrow	19.564	16.205

and encourage it to be consistent with the sampled ambient occlusion in Eqn. 2. Here we ablate these two strategies of ambient occlusion calculation in Fig. 11. It shows that directly calculating ambient occlusion can cause noises in rendered results since it is computationally expensive to sample a lot of points around a single sample point while the predicted ambient occlusion by our network is smooth and noise-free.

5 CONCLUSION

In this paper, we propose SOL-NeRF for outdoor scene decomposition and relighting based on neural radiance fields. Specifically, we use a signed distance field as our geometry representation. Instead of modeling the sky as a whole like previous methods, we separate the learning of the sunlight and the sky light and introduce a hybrid sunlight representation where the sun is modeled as a Spherical Gaussian function and the sky is approximated by first-order Spherical Harmonics. Based on the proposed geometry and lighting representation, the shadow can be approximated by casting rays to the mesh extracted from the signed distance field and learned ambient occlusion. Our method also benefits from the relaxed Manhattan-world geometry prior and sunlight color prior. These priors reduce the learning difficulty of the ill-posed inverse rendering problem and stabilize the training process. Nevertheless, our method still has the following limitations: Firstly, our method works better on diffuse scenes since we do not take specular reflection into account. Decomposition results can be less faithful when specular reflection exists as shown in Fig. 7. Secondly, our lighting representation is based on the environment map that does not handle emitters like streetlights. For future directions, we would like to decompose more complicated lighting effects like specular reflection and emitters from the input scene similar to inverse rendering methods like [Li et al. 2022b; Wu et al. 2023; Zhu et al. 2023].

ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (No. 62061136007), the Beijing Municipal Natural Science Foundation for Distinguished Young Scholars (No. JQ21013), the Royal Society Newton Advanced Fellowship (No. NAF\R2\192151), RCUK grant CAMERA (EP/M023281/1, EP/T022523/1), and a gift from Adobe.

REFERENCES

- Mark Boss, Raphael Braun, Varun Jampani, Jonathan T Barron, Ce Liu, and Hendrik Lensch. 2021a. NeRD: Neural reflectance decomposition from image collections. In *ICCV*. 12684–12694.
- Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan T. Barron, and Hendrik P.A. Lensch. 2021b. Neural-PIL: Neural Pre-Integrated Lighting for Reflectance Decomposition. In *Advances in Neural Information Processing Systems*.
- Zhiqin Chen and Hao Zhang. 2019. Learning implicit fields for generative shape modeling. In *CVPR*. 5939–5948.

- Blender Online Community. 2023. *Blender - a 3D modelling and rendering package*. Blender Foundation.
- Sylvain Duchêne, Clement Riant, Gaurav Chaurasia, Jorge Lopez Moreno, Pierre-Yves Laffont, Stefan Popov, Adrien Bousseau, and George Drettakis. 2015. Multiview Intrinsic Images of Outdoors Scenes with an Application to Relighting. *ACM Trans. Graph.* 34, 5, Article 164 (2015), 16 pages.
- Yasutaka Furukawa, Brian Curless, Steven M Seitz, and Richard Szeliski. 2009. Manhattan-world stereo. In *CVPR*. 1422–1429.
- Duan Gao, Guojun Chen, Yue Dong, Pieter Peers, Kun Xu, and Xin Tong. 2020. Deferred Neural Lighting: Free-Viewpoint Relighting from Unstructured Photographs. *ACM Trans. Graph.* 39, 6, Article 258 (2020), 15 pages.
- Haoyu Guo, Sida Peng, Haotong Lin, Qianqian Wang, Guofeng Zhang, Hujun Bao, and Xiaowei Zhou. 2022. Neural 3D Scene Reconstruction with the Manhattan-world Assumption. In *CVPR*. 5511–5520.
- Jon Hasselgren, Nikolai Hofmann, and Jacob Munkberg. 2022. Shape, Light, and Material Decomposition from Images using Monte Carlo Rendering and Denoising. In *Advances in Neural Information Processing Systems*.
- Yannick Hold-Geoffroy, Akshaya Athawale, and Jean-François Lalonde. 2019. Deep Sky Modeling For Single Image Outdoor Lighting Estimation. In *CVPR*.
- Yannick Hold-Geoffroy, Kalyan Sunkavalli, Sunil Hadap, Emiliano Gambaretto, and Jean-François Lalonde. 2017. Deep Outdoor Illumination Estimation. In *CVPR*.
- Lukas Hosek and Alexander Wilkie. 2012. An Analytic Model for Full Spectral Sky-Dome Radiance. *ACM Trans. Graph.* 31, 4, Article 95 (2012), 9 pages.
- James T. Kajiya. 1986. The Rendering Equation (*SIGGRAPH '86*). 143–150.
- Zhengfei Kuang, Kyle Olszewski, Menglei Chai, Zeng Huang, Panos Achlioptas, and Sergey Tulyakov. 2022. NeROIC: Neural Rendering of Objects from Online Image Collections. *ACM Trans. Graph.* 41, 4, Article 56 (2022), 12 pages.
- Quewei Li, Jie Guo, Yang Fei, Feichao Li, and Yanwen Guo. 2022a. NeuLighting: Neural Lighting for Free Viewpoint Outdoor Scene Relighting with Unconstrained Photo Collections. In *SIGGRAPH Asia 2022 Conference Papers*. Article 13, 9 pages.
- Zhengqin Li, Jia Shi, Sai Bi, Rui Zhu, Kalyan Sunkavalli, Milos Hasan, Zexiang Xu, Ravi Ramamoorthi, and Manmohan Chandraker. 2022b. Physically-Based Editing of Indoor Scene Lighting from a Single Image. In *ECCV*, Vol. 13666. 555–572.
- Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. 2021. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In *CVPR*.
- Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. 2019. Occupancy networks: Learning 3D reconstruction in function space. In *CVPR*. 4460–4470.
- Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing scenes as neural radiance fields for view synthesis. In *ECCV*. 405–421.
- Martin Mittring. 2007. Finding next Gen: CryEngine 2 (*SIGGRAPH '07*). 97–121.
- Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Trans. Graph.* 41, 4 (2022), 102:1–102:15.
- Jacob Munkberg, Wenzheng Chen, Jon Hasselgren, Alex Evans, Tianchang Shen, Thomas Müller, Jun Gao, and Sanja Fidler. 2022. Extracting Triangular 3D Models, Materials, and Lighting From Images. In *CVPR*. 8270–8280.
- Tomoyuki Nishita, Takao Sirai, Katsumi Tadamura, and Eihachiro Nakamae. 1993. Display of the Earth Taking into Account Atmospheric Scattering (*SIGGRAPH '93*). 175–182.
- Michael Oechsle, Songyou Peng, and Andreas Geiger. 2021. UNISURF: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *ICCV*. 5589–5599.
- Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. 2019. DeepSDF: Learning continuous signed distance functions for shape representation. In *CVPR*. 165–174.
- Julien Philip, Michaël Gharbi, Tinghui Zhou, Alexei A. Efros, and George Drettakis. 2019. Multi-view relighting using a geometry-aware network. *ACM Trans. Graph.* 38, 4 (2019), 78:1–78:14.
- A. J. Preetham, Peter Shirley, and Brian Smits. 1999. A Practical Analytic Model for Daylight (*SIGGRAPH '99*). 91–100.
- Ravi Ramamoorthi and Pat Hanrahan. 2001a. An efficient representation for irradiance environment maps. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. 497–500.
- Ravi Ramamoorthi and Pat Hanrahan. 2001b. A signal-processing framework for inverse rendering. In *SIGGRAPH*. 117–128.
- Viktor Rudnev, Mohamed Elgharib, William Smith, Lingjie Liu, Vladislav Golyanik, and Christian Theobalt. 2022. NeRF for Outdoor Scene Relighting. In *ECCV*.
- Yoichi Sato, Mark D. Wheeler, and Katsushi Ikeuchi. 1997. Object shape and reflectance modeling from observation. In *SIGGRAPH*. 379–387.
- Johannes Lutz Schönberger and Jan-Michael Frahm. 2016. Structure-from-Motion Revisited. In *CVPR*.
- Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. 2016. Pixelwise View Selection for Unstructured Multi-View Stereo. In *ECCV*.
- Perumal Shanmugam and Okan Arıkan. 2007. Hardware Accelerated Ambient Occlusion Techniques on GPUs (*3D '07*). 73–80.
- Tianchang Shen, Jun Gao, Kangxue Yin, Ming-Yu Liu, and Sanja Fidler. 2021. Deep marching tetrahedra: a hybrid representation for high-resolution 3D shape synthesis. *Advances in Neural Information Processing Systems* 34 (2021), 6087–6101.
- Shuran Song and Thomas A. Funkhouser. 2019. Neural Illumination: Lighting Prediction for Indoor Environments. In *CVPR*. 6918–6926.
- Jiaming Sun, Xi Chen, Qianqian Wang, Zhengqi Li, Hadar Averbuch-Elor, Xiaowei Zhou, and Noah Snavely. 2022. Neural 3D Reconstruction in the Wild. In *SIGGRAPH Conference Proceedings*.
- Jiaming Sun, Yiming Xie, Linghao Chen, Xiaowei Zhou, and Hujun Bao. 2021. NeuralRecon: Real-Time Coherent 3D Reconstruction from Monocular Video. *CVPR* (2021).
- Justus Thies, Michael Zollhöfer, and Matthias Nießner. 2019. Deferred neural rendering: image synthesis using neural textures. *ACM Trans. Graph.* 38, 4 (2019), 66:1–66:12.
- John Tyndall. 1869. IV. On the blue colour of the sky, the polarization of skylight, and on the polarization of light by cloudy matter generally. *Proceedings of the Royal Society of London* 17 (1869), 223–233.
- Jiaping Wang, Peiran Ren, Minmin Gong, John Snyder, and Baining Guo. 2009. All-Frequency Rendering of Dynamic, Spatially-Varying Reflectance. *ACM Trans. Graph.* 28, 5 (2009), 1–10.
- Jiepeng Wang, Peng Wang, Xiaoxiao Long, Christian Theobalt, Taku Komura, Lingjie Liu, and Wenping Wang. 2022. NeuRIS: Neural reconstruction of indoor scenes using normal priors. *arXiv preprint arXiv:2206.13597* (2022).
- Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. 2021. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. In *Advances in Neural Information Processing Systems*, Vol. 34.
- Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. 2004. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing* 13, 4 (2004), 600–612.
- Zian Wang, Tianchang Shen, Jun Gao, Shengyu Huang, Jacob Munkberg, Jon Hasselgren, Zan Gojcic, Wenzheng Chen, and Sanja Fidler. 2023. Neural Fields meet Explicit Geometric Representations for Inverse Rendering of Urban Scenes. In *CVPR*.
- Tong Wu, Jia-Mu Sun, Yu-Kun Lai, and Lin Gao. 2023. DE-NeRF: DEcoupled Neural Radiance Fields for View-Consistent Appearance Editing and High-Frequency Environmental Relighting. In *SIGGRAPH 2023*. ACM, 74:1–74:11.
- Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. 2021. Volume rendering of neural implicit surfaces. In *Advances in Neural Information Processing Systems*.
- Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Ronen Basri, and Yaron Lipman. 2020. Multiview Neural Surface Reconstruction by Disentangling Geometry and Appearance. In *Advances in Neural Information Processing Systems*.
- Ye Yu, Abhimitra Meka, Mohamed Elgharib, Hans-Peter Seidel, Christian Theobalt, and William A. P. Smith. 2020. Self-supervised Outdoor Scene Relighting. In *ECCV 2020*, Vol. 12367. 84–101.
- Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. 2022. MonoSDF: Exploring Monocular Geometric Cues for Neural Implicit Surface Reconstruction. *arXiv preprint arXiv:2206.00665* (2022).
- Jinsong Zhang, Kalyan Sunkavalli, Yannick Hold-Geoffroy, Sunil Hadap, Jonathan Eisenmann, and Jean-François Lalonde. 2019. All-Weather Deep Outdoor Lighting Estimation. In *CVPR*.
- Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. 2021a. PhysG: Inverse Rendering with Spherical Gaussians for Physics-based Material Editing and Relighting. In *CVPR*.
- Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. 2021b. NeRFactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Trans. Graph.* 40, 6 (2021), 1–18.
- Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. 2022. Modeling Indirect Illumination for Inverse Rendering. In *CVPR*.
- Jingsen Zhu, Yuchi Huo, Qi Ye, Fujun Luan, Jifan Li, Dianbing Xi, Lisha Wang, Rui Tang, Wei Hua, Hujun Bao, and Rui Wang. 2023. I²-SDF: Intrinsic Indoor Scene Reconstruction and Editing via Raytracing in Neural SDFs. In *CVPR*.
- Yongjie Zhu, Yinda Zhang, Si Li, and Boxin Shi. 2021. Spatially-Varying Outdoor Lighting Estimation From Intrinsic. In *CVPR*. 12834–12842.

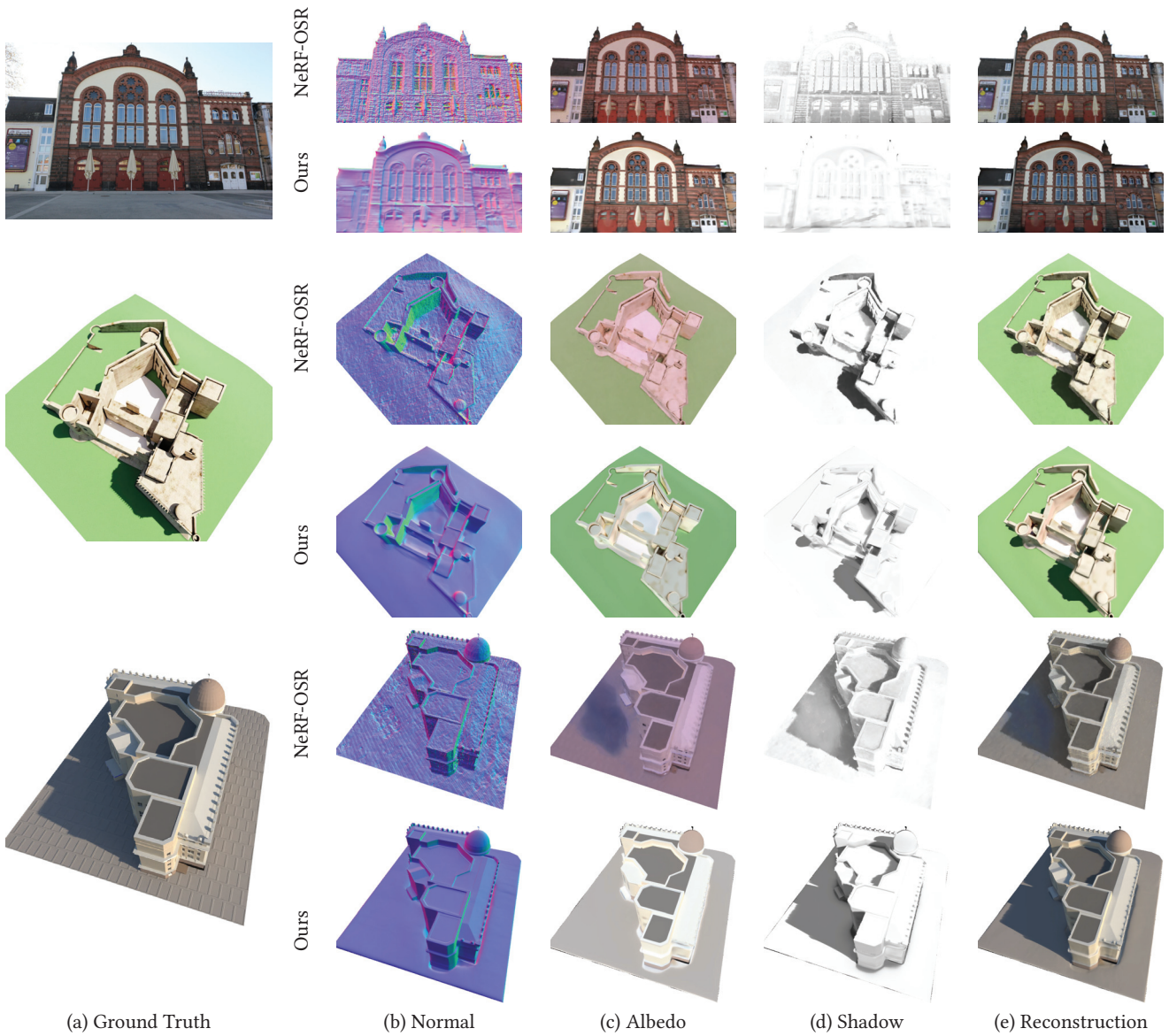


Figure 6: Decomposition results of NeRF-OSR and our method. For each scene, we show different decomposed components (normal, albedo and shadow) and the reconstructed image.

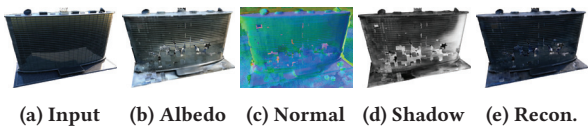


Figure 7: Failure case: our method is incapable of decomposing the scene correctly when strong specular reflections exist.

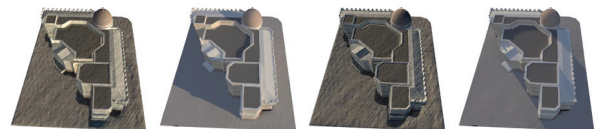


Figure 8: Qualitative comparisons of relighting results between our explicit shadow calculation (Explicit S.) and the baseline (Implicit S.) using an implicit MLP network to predict shadows with two novel lighting conditions.

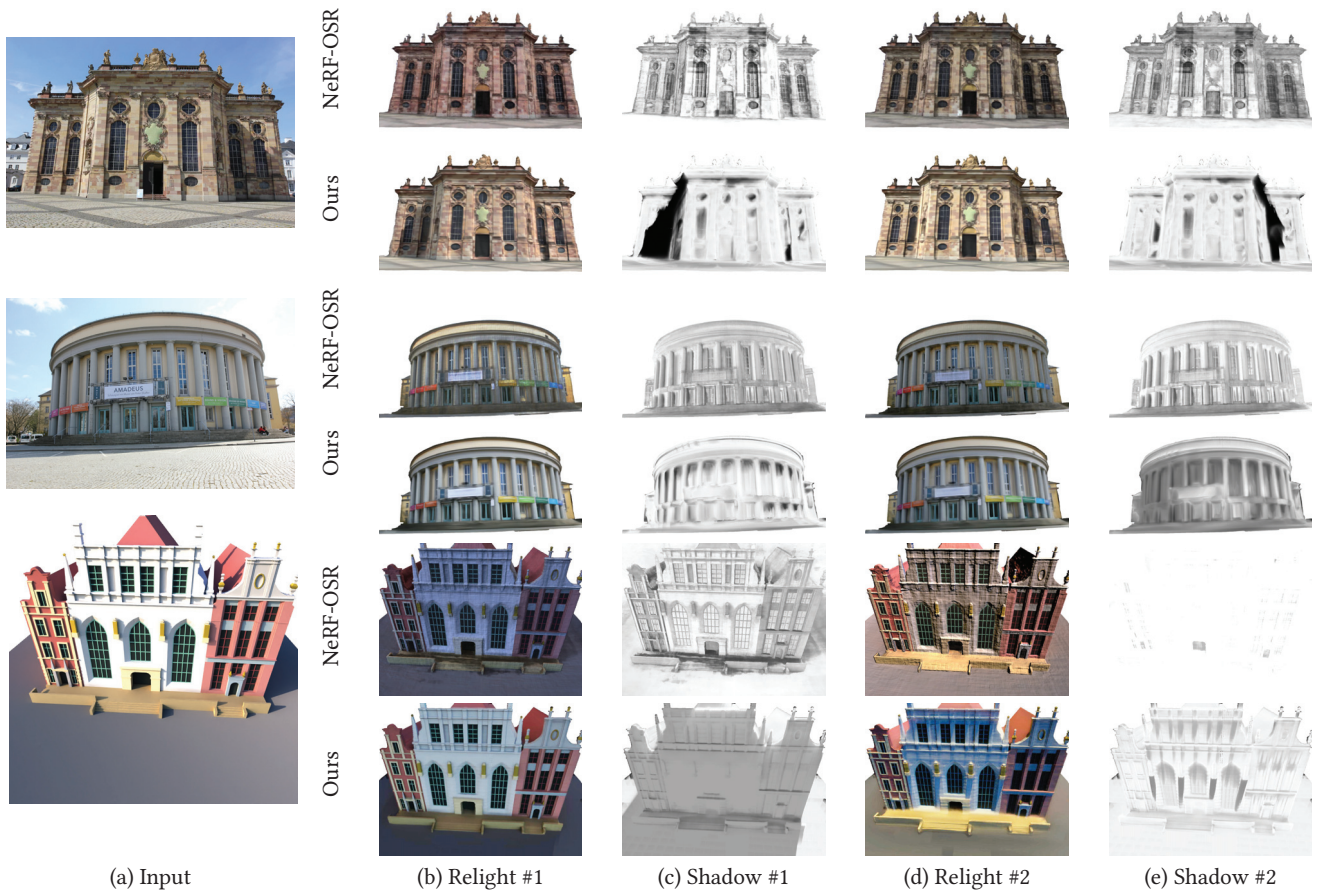


Figure 9: Relighting results of NeRF-OSR and our method. For each input scene, we relight it with two different lighting conditions and show rendered images and shadows.

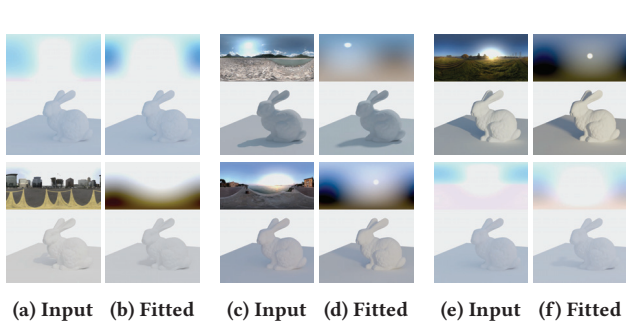


Figure 10: We use our hybrid lighting representation to fit six environment maps. In each pair, we compare the rendered images under the ground truth environment map and the fitted environment map. The average PSNR between the images pairs is 32.829. The per-image PSNRs are reported as follows: 49.316, 37.336, 32.507, 32.467, 32.077, and 29.758. When calculating average, the first sample(49.316) is regarded as an outlier and omitted.

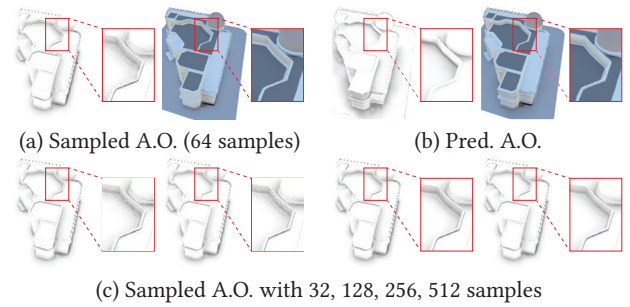


Figure 11: Qualitative comparison of Ambient Occlusion and reconstruction results between the explicit ambient occlusion calculation baseline (Sampled A.O.) and our method that uses an MLP network to predict the A.O. (Pred. A.O.). We additionally include sampled A.O. with different numbers of samples for reference. Please zoom in for more details.