Factored-NeuS: Reconstructing Surfaces, Illumination, and Materials of Possibly Glossy Objects

YUE FAN, Chongqing University, China
IVAN SKOROKHODOV, KAUST, Saudi Arabia
OLEG VOYNOV, Skoltech, Russia
SAVVA IGNATYEV, Skoltech, Russia
EVGENY BURNAEV, Skoltech, Russia
PETER WONKA, KAUST, Saudi Arabia
YIQUN WANG, Chongqing University, China

We develop a method that recovers the surface, materials, and illumination of a scene from its posed multi-view images. In contrast to prior work, it does not require any additional data and can handle glossy objects or bright lighting. It is a progressive inverse rendering approach, which consists of three stages. In the first stage, we reconstruct the scene radiance and signed distance function (SDF) with a novel regularization strategy for specular reflections. We propose to explain a pixel color using both surface and volume rendering jointly, which allows for handling complex view-dependent lighting effects for surface reconstruction. In the second stage, we distill light visibility and indirect illumination from the learned SDF and radiance field using learnable mapping functions. Finally, we design a method for estimating the ratio of incoming direct light reflected in a specular manner and use it to reconstruct the materials and direct illumination. Experimental results demonstrate that the proposed method outperforms the current state-of-the-art in recovering surfaces, materials, and lighting without relying on any additional data.

CCS Concepts: • **Computing methodologies** → **Shape modeling**; **Reconstruction**; *Reflectance modeling*.

Additional Key Words and Phrases: Glossy surface, materials, illumination, diffuse, specular

ACM Reference Format:
Yue Fan, Ivan Skorokhodov, Oleg Voynov, Savva Ignatyev, Evgeny Burnaev, Peter Wonka, and Yiqun Wang. 2024. Factored-NeuS: Reconstructing Surfaces, Illumination, and Materials of Possibly Glossy Objects. , (May 2024), 19 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Authors' addresses: Yue Fan, Chongqing University, Chongqing, China, linyao153580@gmail.com; Ivan Skorokhodov, KAUST, Thuwal, Saudi Arabia, iskorokhodov@gmail.com; Oleg Voynov, Skoltech, Moscow, Russia, oyvoinov@gmail.com; Savva Ignatyev, Skoltech, Moscow, Russia, savva.ignatyev@skoltech.ru; Evgeny Burnaev, Skoltech, Russia, burnaevevgeny@gmail.com; Peter Wonka, KAUST, Thuwal, Saudi Arabia, pwonka@gmail.com; Yiqun Wang, Chongqing University, Chongqing, China, yiqun.wang@cqu.edu.cn.

XXXX-XXXX/2024/5-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Reconstructing shape, material, and lighting from multiple views has wide applications in computer vision, virtual reality, augmented reality, and shape analysis. The emergence of neural radiance fields [Mildenhall et al. 2020] provides a framework for high-quality scene reconstruction. Subsequently, many works [Fu et al. 2022; Oechsle et al. 2021; Wang et al. 2021, 2022; Yariv et al. 2021] have incorporated implicit neural surfaces into neural radiance fields, further enhancing the quality of surface reconstruction from multi-views. Recently, several works [Munkberg et al. 2022; Zhang et al. 2021a,b, 2022b] have utilized coordinate-based networks to predict materials and learned parameters to represent illumination, followed by synthesizing image color using physically-based rendering equations to achieve material and lighting reconstruction. However, these methods typically do not fully consider the interdependence between different components, leading to the following issues with glossy surfaces when using real data.

First, surfaces with glossy materials typically result in highlights. The best current methods for reconstructing implicit neural surfaces rarely consider material information and directly reconstruct surfaces. The surface parameters can then be frozen for subsequent material reconstruction. Since neural radiance fields typically model such inconsistent colors as bumpy surfaces as shown in Fig. 1 left, the artifacts from surface reconstruction will affect material reconstruction if surfaces and materials are reconstructed sequentially. Second, a glossy surface can affect the decomposition of the reflected radiance into a diffuse component and a specular component. Typically, the specular component leaks into the diffuse component, resulting in inaccurate modeling as shown in Fig. 1 right. Third, focusing on synthetic data makes it easier to incorporate complex physically-based rendering algorithms, but they may not be robust enough to work on real data.

In this work, we consider the impact of glossy surfaces on surface and material reconstruction. To better handle glossy surfaces, we jointly use surface and volume rendering. Volume rendering does not decompose the reflected radiance, while surface rendering considers the diffuse and specular radiance separately. This approach better regularizes not only the decomposition of reflected light but also the surface reconstruction. To better recover diffuse and specular components, we estimate the ratio of incoming light reflected in a specular manner. By introducing this parameter into a Spherical Gaussian representation of the BRDF, we can better model the

reflection of glossy surfaces and decompose more accurate diffuse albedo information. Furthermore, we propose predicting continuous light visibility for signed distance functions to further enhance the quality of reconstructed materials and illumination. Our experimental results have shown that our factorization of surface, materials, and illumination achieves state-of-the-art performance on both synthetic and real datasets. We improve surface, material, and lighting reconstruction compared to PhySG [Zhang et al. 2021a], NVDiffRec [Munkberg et al. 2022], and IndiSG [Zhang et al. 2022b], the leading published competitors.

We believe that the good results of our approach compared to much recently published and unpublished work in material reconstruction is that we primarily developed our method on real data. The fundamental challenge for working on material and lighting reconstruction is the lack of available ground truth information for real datasets. Our solution to this problem was to work with real data and try to improve surface reconstruction as our main metric by experimenting with different materials and lighting decompositions as a regularizer. While we could not measure the success of the material and lighting reconstruction directly, we could indirectly observe improvements in the surface metrics. By contrast, most recent and concurrent work uses surface reconstruction and real data more as an afterthought. This alternative route is to first focus on developing increasingly complex material and lighting reconstruction on synthetic data. However, we believe that this typically does not translate as well to real data as our approach.

2 RELATED WORK

Neural radiance fields. NeRF [Mildenhall et al. 2020] is a seminal work in 3D reconstruction. Important improvements were proposed by Mip-NeRF [Barron et al. 2021] and Mip-NeRF360 [Barron et al. 2022]. One line of work explores the combination of different data structures with MLPs, such as factored volumes [Chan et al. 2022; Chen et al. 2022; Wang et al. 2023] or voxels [Müller et al. 2022; Reiser et al. 2021; Yu et al. 2021]. Multiple approaches take a step towards extending neural radiance fields to reconstruct material information [Ge et al. 2023; Guo et al. 2022; Verbin et al. 2022; Yang et al. 2022; Yariv et al. 2023].

Implicit neural surfaces. Implicit neural surfaces are typically represented by occupancy functions or signed distance fields (SDFs). Some early works [Chen and Zhang 2019; Mescheder et al. 2019; Park et al. 2019] take point clouds as input and output implicit neural surface representations. Many works have studied how to obtain implicit neural surfaces from images, initially focusing on surface rendering only [Niemeyer et al. 2020; Yariv et al. 2020]. Subsequent methods followed NeRF to employ volume rendering, e.g. UNISURF [Oechsle et al. 2021], VolSDF [Yariv et al. 2021], NeuS [Wang et al. 2021], HF-NeuS [Wang et al. 2022], and Geo-NeuS [Fu et al. 2022].

Joint reconstruction of surface, material, and illumination. Ideally, we would like to jointly reconstruct the 3D geometry, material properties, and lighting conditions of a scene from 2D images. Several methods employ strategies to simplify the problem such as assuming known lighting conditions (NeRV [Srinivasan et al. 2021] and NeRD [Boss et al. 2021a]) or pre-training (ENVIDR [Liang et al. 2023]). PhySG [Zhang et al. 2021a], NeRFactor [Zhang et al. 2021b], and NeROIC [Kuang et al. 2022] use Spherical Gaussians, point light sources, and spherical harmonics, respectively, to decompose unknown lighting from a set of images. Using an illumination integration network, Neural-PIL [Boss et al. 2021b] further reduces the computational cost of lighting integration. IRON [Zhang et al. 2022a] uses SDF-based volume rendering methods to obtain better geometric details in the shape recovery stage. NVDiffrec [Munkberg et al. 2022] explicitly extracts triangle mesh from tetrahedral representation for better material and lighting modeling. IndiSG [Zhang et al. 2022b] uses Spherical Gaussians to represent indirect illumination and achieves good lighting decomposition results. Some concurrent works [Deng et al. 2022; Jin et al. 2023; Wu et al. 2023a; Zhang et al. 2023b,a] continue to improve the efficiency and quality of inverse rendering but do not consider cases with a glossy appearance. NeAI [Zhuang et al. 2023] proposes neural ambient illumination to enhance the rendering quality of glossy appearance. NeRO [Liu et al. 2023] highlights its good performance in reconstructing reflective objects. Despite a lot of recent activity in this area, existing frameworks still struggle to effectively reconstruct reflective or glossy surfaces, lighting, and material information directly from images, especially real-world captured images. Appx Tab. 4 provides a comprehensive overview of recent inverse rendering techniques.

3 METHOD

Our framework has three training stages to gradually decompose the shape, materials, and illumination. The input to our framework is a set of images. In the first stage, we reconstruct the surface from a (possibly glossy) appearance by jointly using surface rendering and surface rendering. After that, we use the reconstructed radiance field to extract direct illumination visibility and indirect illumination in the second stage. In the final stage, we recover the direct illumination map and materials' bidirectional reflectance distribution function (BRDF). We propose to use a BRDF with a learnable specular albedo.

3.1 Stage 1: Surface reconstruction from glossy appearance

Current inverse rendering methods first recover implicit neural surfaces, typically represented as SDFs, from multi-view images to recover shape information, then freeze the parameters of neural surfaces to further recover the material. However, this approach does not consider specular reflections that produce highlights and often models this inconsistent color as bumpy surface geometry as depicted in Fig. 1 left. This incorrect surface reconstruction has a negative impact on subsequent material reconstruction. We propose a neural surface reconstruction method that considers the appearance, diffuse color, and specular color of glossy surfaces at the same time, whose architecture is given in Fig. 2. Our inspiration comes from the following observations. First, according to Geo-NeuS, using SDF point cloud supervision can make the colors of surface points and volume rendering more similar. We abandoned the idea of using additional surface points to supervise SDFs and directly used two different MLPs to predict the surface rendering and volume rendering results and narrow the gap between these two colors using network training. In addition, when modeling glossy surfaces, Ref-NeRF proposes a method of decomposing reflected radiance into

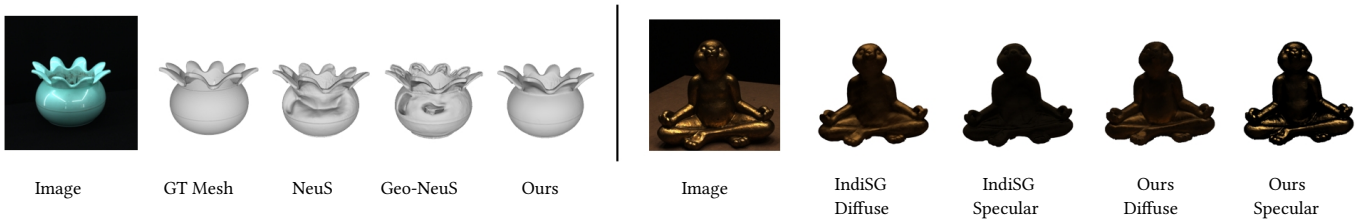

Fig. 1. Left: Geometry visualization for NeuS, Geo-NeuS and our method on the Pot scene from SK3D dataset. Existing surface reconstruction methods struggle to recover the correct geometry of glossy objects due to the complex view-dependent effects they induce. The weak color model of these methods compels to represent such effects through concave geometric deformations rather than proper view-dependent radiance, leading to shape artifacts. In contrast, our method can correctly reconstruct a highly reflective surface due to our joint appearance, diffuse, and specular color training strategy. Right: Visualization of the recovered diffuse color component on the Bunny scene from DTU for IndiSG [Zhang et al. 2022b] and our method. Existing inverse rendering methods overestimate the diffuse material component in the presence of specular highlights. Our regularization strategy allows us to properly disentangle the color into diffuse and specular components.

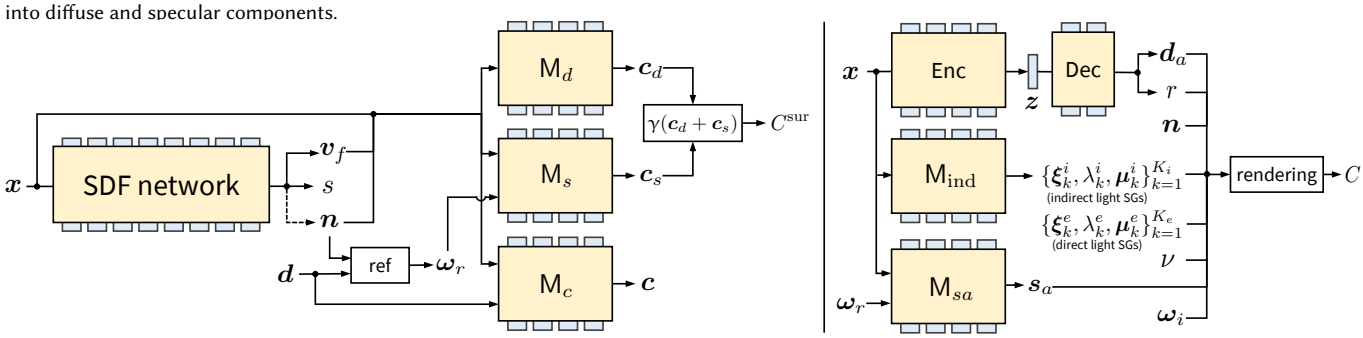

Fig. 2. Overview for Stage 1 (left) and Stage 3 (right) training pipelines (Stage 2 pipeline is omitted due to its simplicity — see Sec 3.2 for details). The first stage (left) trains the SDF network S_θ which outputs a feature vector $\boldsymbol{v}_f \in \mathbb{R}^{256}$, SDF value $s \in \mathbb{R}$, and normal $\boldsymbol{n} \in \mathbb{S}^2$ (as a normalized gradient of s; denoted via the dashed line); diffuse and specular surface color networks M_d and M_s produce their respective colors $\boldsymbol{c}_d, \boldsymbol{c}_s \in \mathbb{R}^3$ via surface rendering, which are then combined through tone mapping $\gamma(\cdot)$ to get the final surface color $C^{\text{sur}} \in \mathbb{R}^3$; volumetric color network M_c produces the volumetrically rendered color $C^{\text{vol}} \in \mathbb{R}^3$. The ref operation denotes computation of the reflection direction $\boldsymbol{\omega}_r \in \mathbb{S}^2$ from normal $\boldsymbol{n}$ and ray direction $\boldsymbol{\omega} \in \mathbb{S}^2$. In the third stage (right), we optimize the material BRDF auto-encoder with the sparsity constraint [Zhang et al. 2022b], our novel specular albedo network M_{sa}, and the indirect illumination network M_{ind}. See Sec 3 for details.

diffuse and specular components, which can better model the glossy appearance. However, this approach is unstable when directly applied to reconstruct implicit surfaces. We propose to simultaneously optimize the radiance from the volumetric rendering and the surface rendering. For surface rendering, we further split the reflected radiance into a diffuse and a specular component. This can achieve an improved surface reconstruction of glossy surfaces.

Shape representation. We model shape as a signed distance function $\mathsf{S}_\theta : \boldsymbol{x} \mapsto (s, \boldsymbol{v}_f)$, which maps a 3D point $\boldsymbol{x} \in \mathbb{R}^3$ to its signed distance value $s \in \mathbb{R}$ and a feature vector $\boldsymbol{v}_f \in \mathbb{R}^{256}$. SDF allows computing a normal $\boldsymbol{n}$ directly by calculating the gradient: $\boldsymbol{n} = \nabla \mathsf{S}_\theta(\boldsymbol{x}) / \|\nabla \mathsf{S}_\theta(\boldsymbol{x})\|$.

Synthesize appearance. Learning implicit neural surfaces from multi-view images often requires synthesizing appearance/color to optimize the underlying surface. The recent use of volume rendering in NeuS [Wang et al. 2021] has been shown to better reconstruct surfaces. According to Eq. 14 in Appx A, the discretization formula for volume rendering is $C^{\text{vol}} = \sum_{i=1}^n T_i \alpha_i \boldsymbol{c}_i = \sum_{i=1}^n w_i \boldsymbol{c}_i$ with n sampled points $\{\boldsymbol{r}(t_i)\}_{i=1}^n$ on the ray. where $\alpha_i = \max(\Phi_s(\mathsf{S}_\theta(\boldsymbol{r}(t_i))) - \Phi_s(\mathsf{S}_\theta(\boldsymbol{r}(t_{i+1})))/\Phi_s(\mathsf{S}_\theta(\boldsymbol{r}(t_i))), 0)$, which is discrete opacity values following NeuS, where $\Phi_s = 1/(1+e^{-x})$ is a sigmoid function and $T_i = \prod_{j=1}^{i-1}(1-\alpha_j)$ is the discrete transparency. Similar to the continuous case, we can also define discrete weights $w_i = T_i \alpha_i$.

To compute color $\boldsymbol{c}_i$ on the point $\boldsymbol{r}(t_i)$, we define a color mapping $\mathsf{M}_c : (\boldsymbol{x}, \boldsymbol{n}, \boldsymbol{d}, \boldsymbol{v}_f) \mapsto \boldsymbol{c}$ from any 3D point $\boldsymbol{x}$ given its feature vector $\boldsymbol{v}_f$, normal $\boldsymbol{n}$ and ray direction $\boldsymbol{d}$.

Synthesize diffuse and specular components. In addition to synthesizing appearance, we also synthesize diffuse and specular components. This idea comes from surface rendering, which better handles surface reflections. From Eq. 16 in Appx A, the radiance L_o of surface point $\boldsymbol{x}$ and outgoing viewing direction $\boldsymbol{\omega}_o$ can be decomposed into two parts: diffuse and specular radiance.

$$L_o(\boldsymbol{x}, \boldsymbol{\omega}_o) = \frac{\boldsymbol{d}_a}{\pi} \int_\Omega L_i(\boldsymbol{x}, \boldsymbol{\omega}_i)(\boldsymbol{\omega}_i \cdot \boldsymbol{n}) d\boldsymbol{\omega}_i \tag{1}$$

$$+ \int_\Omega f_s(\boldsymbol{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o) L_i(\boldsymbol{x}, \boldsymbol{\omega}_i)(\boldsymbol{\omega}_i \cdot \boldsymbol{n}) d\boldsymbol{\omega}_i \tag{2}$$

$$= \mathsf{M}_d(\boldsymbol{x}, \boldsymbol{n}) + \mathsf{M}_s(\boldsymbol{x}, \boldsymbol{\omega}_o, \boldsymbol{n}) \tag{3}$$

We define two neural networks to predict diffuse and specular components separately. We use the term diffuse radiance to refer to the component of the reflected radiance that stems from a diffuse surface reflection. We define a mapping $\mathsf{M}_d : (\boldsymbol{x}, \boldsymbol{n}, \boldsymbol{v}_f) \mapsto \boldsymbol{c}_d$ for diffuse radiance that maps surface points $\boldsymbol{x}$, surface normals $\boldsymbol{n}$, and

feature vectors $\boldsymbol{v}_f$ to diffuse radiance. For simplicity, we assume that the diffuse radiance is not related to the outgoing viewing direction $\boldsymbol{\omega}_o$.

We use the term specular radiance to describe the non-diffuse (view-direction dependent) component of the reflected radiance. Ref-NeRF [Verbin et al. 2022] proposes to model the glossy appearance using the reflection direction instead of the viewing one. However, from Eq. 3, we can observe that specular radiance is also highly dependent on the surface normal, which is particularly important when reconstructing SDF. In contrast to Ref-NeRF, we further condition specular radiance on the surface normal. Therefore, we define specular radiance $\mathsf{M}_s : (\boldsymbol{x}, \boldsymbol{\omega}_r, \boldsymbol{n}, \boldsymbol{v}_f) \mapsto \boldsymbol{c}_s$, which maps surface points $\boldsymbol{x}$, reflection direction $\boldsymbol{\omega}_r$, surface normals $\boldsymbol{n}$, and feature vectors $\boldsymbol{v}_f$ to specular radiance, where $\boldsymbol{\omega}_r = 2(\boldsymbol{\omega}_o \cdot \boldsymbol{n})\boldsymbol{n} - \boldsymbol{\omega}_o$.

Surface rendering focuses the rendering process on the surface, allowing for a better understanding of highlights on the surface compared to volume rendering, but requires calculating surface points. We sample n points on the ray $\{\boldsymbol{r}(t_i) | i = 1, ..., n\}$. We query the sampled points to find the first point $\boldsymbol{r}(t_i')$ whose SDF value is less than zero $\mathsf{S}_\theta(\boldsymbol{r}(t_i')) < 0$. Then the point $\boldsymbol{r}(t_{i'-1})$ sampled before $\boldsymbol{r}(t_i')$ has the SDF value greater than or equal to zero $\mathsf{S}_\theta(\boldsymbol{r}(t_{i'-1})) \geq 0$. To account for the possibility of rays interacting with objects and having multiple intersection points, we select the first point with a negative SDF value to solve this issue.

We use two neural networks to predict the diffuse radiance and specular radiance of two sampling points $\boldsymbol{r}(t_{i'-1})$ and $\boldsymbol{r}(t_i')$. The diffuse radiance of the two points calculated by the diffuse network M_d will be $\boldsymbol{c}_d^{i'-1}$ and $\boldsymbol{c}_d^{i'}$. The specular radiance of the two points calculated by the specular network M_s will be $\boldsymbol{c}_s^{i'-1}$ and $\boldsymbol{c}_s^{i'}$. Therefore, the diffuse radiance and specular radiance of the surface point $\boldsymbol{x}$ can be calculated as follows.

$$\boldsymbol{c}_d = \mathsf{M}_d(\boldsymbol{x}, \boldsymbol{n}) = \frac{w_{i'-1}\boldsymbol{c}_d^{i'-1} + w_i'\boldsymbol{c}_d^{i'}}{w_{i'-1} + w_i'} \tag{4}$$

$$\boldsymbol{c}_s = \mathsf{M}_s(\boldsymbol{x}, \boldsymbol{\omega}_o, \boldsymbol{n}) = \frac{w_{i'-1}\boldsymbol{c}_s^{i'-1} + w_i'\boldsymbol{c}_s^{i'}}{w_{i'-1} + w_i'} \tag{5}$$

The final radiance of the intersection of the ray and the surface is calculated by a tone mapping:

$$C^{\text{sur}} = \gamma(\boldsymbol{c}_d + \boldsymbol{c}_s) \tag{6}$$

where γ is a pre-defined tone mapping function that converts linear color to sRGB [Verbin et al. 2022] while ensuring that the resulting color values are within the valid range of $[0, 1]$.

Training strategies. In our training process, we define three loss functions, namely volume radiance loss $\mathcal{L}_{\text{vol}}$, surface radiance loss $\mathcal{L}_{\text{sur}}$, and regularization loss $\mathcal{L}_{\text{reg}}$. The volume radiance loss $\mathcal{L}_{\text{vol}}$ is measured by calculating the $\mathcal{L}_1$ distance between the ground truth colors C^{gt} and the volume radiances C^{vol} of a subset of rays $\mathcal{R}$. The surface radiance loss $\mathcal{L}_{\text{sur}}$ is measured by calculating the $\mathcal{L}_1$ distance between the ground truth colors C^{gt} and the surface radiances C^{sur}. $\mathcal{L}_{\text{reg}}$ is an Eikonal loss term on the sampled points. We use weights λ_{sur} and λ_{reg} to balance the impact of these three losses. The overall training loss is as follows. See Appx C for details of training strategies.

$$\mathcal{L} = \mathcal{L}_{\text{vol}} + \lambda_{\text{sur}}\mathcal{L}_{\text{sur}} + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}} \tag{7}$$

3.2 Stage 2: Learning direct lighting visibility and indirect illumination

At this stage, we focus on predicting the lighting visibility and indirect illumination of a surface point $\boldsymbol{x}$ under different incoming light direction $\boldsymbol{\omega}_i$ using the SDF in the first stage.

Visibility is an important factor in shadow computation. It calculates the visibility of the current surface point $\boldsymbol{x}$ in the direction of the incoming light $\boldsymbol{\omega}_i$. Path tracing of the SDF is commonly used to obtain a binary visibility (0 or 1) as used in IndiSG [Zhang et al. 2022b], but this kind of visibility is not friendly to network learning. Inspired by NeRFactor [Zhang et al. 2021b], we propose to use an integral representation with the continuous weight function $w(t)$ (from 0 to 1) for the SDF to express light visibility. Specifically, we establish a neural network $\mathsf{M}_v : (\boldsymbol{x}, \boldsymbol{\omega}_i) \mapsto v$, that maps the surface point $\boldsymbol{x}$ and incoming light direction $\boldsymbol{\omega}_i$ to visibility, and the ground truth value of light visibility is obtained by integrating the weights w_i of the SDF of sampling points along the incoming light direction and can be expressed as $v^{gt} = 1 - \sum_{i=1}^{n} w_i$.

Indirect illumination refers to the light that is reflected or emitted from surfaces in a scene and then illuminates other surfaces, rather than directly coming from a light source, which contributes to the realism of rendered images. Following IndiSG [Zhang et al. 2022b], we parameterize indirect illumination $I(\boldsymbol{x}, \boldsymbol{\omega}_i)$ via $K_i = 24$ Spherical Gaussians (SGs). For more details, see Appx D.

3.3 Stage 3: Recovering materials and direct illumination

Reconstructing good materials and lighting from scenes with highlights is a challenging task. Following prior works [Zhang et al. 2021a, 2022b], we use the Disney BRDF model [Burley and Studios 2012] and represent BRDF $f_s(\boldsymbol{\omega}_i \mid \boldsymbol{\xi}_s, \lambda_s, \boldsymbol{\mu}_s)$ via Spherical Gaussians [Zhang et al. 2021a]. Direct (environment) illumination is represented using $K_e = 128$ SGs:

$$E(\boldsymbol{x}, \boldsymbol{\omega}_i) = \sum_{k=1}^{K_e} E_k(\boldsymbol{\omega}_i \mid \boldsymbol{\xi}_k^e, \lambda_k^e, \boldsymbol{\mu}_k^e) \tag{8}$$

and render diffuse radiance and specular radiance of direct illumination in a way similar to Eq. 2.

$$L_d(\boldsymbol{x}) = \frac{\boldsymbol{d}_a}{\pi} \sum_{k=1}^{K_e} (v(\boldsymbol{x}, \boldsymbol{\omega}_i) \otimes E_k(\boldsymbol{\omega}_i)) \cdot (\boldsymbol{\omega}_i \cdot \boldsymbol{n}) \tag{9}$$

$$L_s(\boldsymbol{x}, \boldsymbol{\omega}_o) = \sum_{k=1}^{K_e} (f_s^a \otimes v(\boldsymbol{x}, \boldsymbol{\omega}_i) \otimes E_k(\boldsymbol{\omega}_i)) \cdot (\boldsymbol{\omega}_i \cdot \boldsymbol{n}) \tag{10}$$

where $\boldsymbol{d}_a$ is diffuse albedo.

To reconstruct a more accurate specular reflection effect, we use an additional neural network $\mathsf{M}_{sa} : (\boldsymbol{x}, \boldsymbol{\omega}_r) \mapsto \boldsymbol{s}_a \in [0, 1]$ to predict the specular albedo. The **modified** BRDF f_s^a is as follows:

$$f_s^a = \boldsymbol{s}_a \otimes f_s(\boldsymbol{\omega}_i; \boldsymbol{\xi}, \lambda, \boldsymbol{\mu}) = f_s(\boldsymbol{\omega}_i; \boldsymbol{\xi}, \lambda, \boldsymbol{s}_a\boldsymbol{\mu}) \tag{11}$$

For indirect illumination, the radiance is extracted directly from another surface and does not consider light visibility. The diffuse radiance and specular radiance of indirect illumination are as follows

$$L_d^{\text{ind}}(\boldsymbol{x}) = \frac{\boldsymbol{d}_a}{\pi} \sum_{k=1}^{T} I_k(\boldsymbol{x}, \boldsymbol{\omega}_i) \cdot (\boldsymbol{\omega}_i \cdot \boldsymbol{n}) \tag{12}$$

$$L_s^{\text{ind}}(\boldsymbol{x}, \boldsymbol{\omega}_o) = \sum_{k=1}^{T} (\boldsymbol{s}_a \otimes f_s) \otimes I_k(\boldsymbol{x}, \boldsymbol{\omega}_i) \cdot (\boldsymbol{\omega}_i \cdot \boldsymbol{n}) \tag{13}$$

Our final synthesized appearance is $C = L_d + L_s + L_d^{\text{ind}} + L_s^{\text{ind}}$ and supervised via an $\mathcal{L}_1$ RGB loss.

4 EXPERIMENTS

4.1 Evaluation setup

Datasets. To evaluate the quality of surface reconstruction, we use the DTU [Jensen et al. 2014], SK3D [Voynov et al. 2022], and Shiny [Verbin et al. 2022] datasets. DTU and SK3D are two real-world captured datasets, while Shiny is synthetic. In DTU, each scene is captured by 49 or 64 views at 1600×1200 resolution. For it, we select 4 scenes with high specularities to evaluate our method in terms of surface quality quantitatively and material decomposition qualitatively. In the SK3D dataset, the image resolution is 2368×1952, and 100 views are provided for each scene. Compared to DTU, SK3D contains more reflective objects with complex view-dependent lighting effects. From SK3D, we select 4 glossy surface scenes with high levels of glare. The Shiny dataset has 5 different glossy objects rendered in Blender under conditions similar to the NeRF-synthetic dataset [Mildenhall et al. 2020] (100 training and 200 testing images per scene of the 800×800 resolution).

To evaluate the effectiveness of material and lighting reconstruction, we use the dataset provided by IndiSG [Zhang et al. 2022b], which has self-occlusion and complex materials. Each scene has 100 training images of 800 × 800 resolution. To evaluate the quality of material decomposition, the dataset also provides diffuse albedo, roughness, and masks for testing.

Baselines. Our main competitors are the methods that can also reconstruct all three scene properties: surface geometry, materials, and illumination. We choose NVDiffRec [Munkberg et al. 2022], PhySG [Zhang et al. 2021a], and IndiSG [Zhang et al. 2022b] due to their popularity and availability of the source code. NVDiffRec uses tetrahedral marching to extract triangle meshes and obtains good material decomposition using a triangle-based renderer. PhySG optimizes geometry and material information at the same time using a Spherical Gaussian representation for direct lighting and material. IndiSG first optimizes geometry and then uses a Spherical Gaussian representation for indirect lighting to improve the quality of material reconstruction.

Apart from that, we also compared against more general surface reconstruction methods to provide additional context for our results. For this, we use NeuS [Wang et al. 2021], Geo-NeuS [Fu et al. 2022], and NeRO [Liu et al. 2023]. NeuS is a popular implicit surface reconstruction method that achieves strong results without relying on extra data. Geo-NeuS improves upon NeuS by using additional point cloud supervision, obtained from structure from motion (SfM) [Schönberger and Frahm 2016]. NeRO is capable of reconstructing reflective objects effectively. In addition, we also show a qualitative comparison to Ref-NeRF [Verbin et al. 2022], which considers material decomposition, but due to modeling geometry using density function, it has difficulty extracting smooth geometry.

Evaluation metrics. We use the official evaluation protocol to compute Chamfer distance (lower values are better) for the DTU dataset and also use Chamfer distance for the SK3D dataset. We utilize the PSNR metric (higher values are better), to quantitatively evaluate the quality of rendering, material, and illumination. We follow IndiSG [Zhang et al. 2022b] and employ masks to compute the PSNR metric in the foreground to evaluate the quality of materials and rendering. See Appx B for implementation details.

4.2 Surface reconstruction quality

We first demonstrate quantitative results in terms of Chamfer distance. We provide the numerical results for IndiSG and PhySG for comparison. NVDiffrec is not as good for surface reconstruction as we verify qualitatively in Fig. 6. For completeness, we also compare our method against NeuS, Geo-NeuS, and NeRO. First, we list quantitative results on DTU and SK3D in Tab. 1. As shown in the table, Geo-NeuS achieves better performance on the DTU dataset because the additional sparse 3D points generated by structure from motion (SfM) for supervising the SDF network are accurate. Our approach can also incorporate the components of Geo-NeuS based on extra data, and the surface reconstruction quality will be further improved as shown in Appx F. However, on the SK3D scenes with glossy surfaces, these sparse 3D points cannot be generated accurately by SfM, leading to poor surface reconstruction by Geo-NeuS. In contrast, our approach can reconstruct glossy surfaces on both DTU and SK3D without any explicit geometry information. IndiSG and PhySG share the same surface reconstruction method, but PhySG optimizes it together with the materials, while IndiSG freezes the underlying SDF after its initial optimization. Compared with IndiSG, PhySG cannot optimize geometry and material information well simultaneously on real-world acquired datasets with complex lighting and materials. Our method is the overall best method on SK3D. Most importantly, we demonstrate large improvements over IndiSG and PhySG, our main competitors, on both DTU and SK3D. We further demonstrate the qualitative experimental comparison results in Fig. 5. It can be seen that although Geo-NeuS has the best quantitative evaluation metrics, it loses some of the fine details, such as the small dents on the metal can in DTU 97. By visualizing the results of the SK3D dataset, we can validate that our method can reconstruct glossy surfaces without explicit geometric supervision.

We further conducted additional comparisons of NeRO and NeuS on real-world datasets SK3D and DTU with non-Lambertian surfaces. As shown in Fig. 8 for SK3D dataset, even in scenarios with a simple background and straightforward geometry, NeRO still tends to lose certain details of the dataset and fills in shadowed areas (flower pot model). Chamfer distance metrics results are presented in Tab. 1. For DTU datasets, as shown in Tab. 1 and Appx Fig. 14. NeRO not only struggles to accurately restore detailed information but also fails to address the negative impact of partial highlights on geometry. Moreover, the presence of shadows causes NeRO to mistakenly reconstruct shadowed areas as real objects and fill them in (bricks and skull models). Additional Chamfer distance metrics for the other DTU scenes are presented in Appx Tab. 6.

We show the qualitative results for surface reconstruction compared with NeuS, Ref-NeRF, IndiSG, PhySG, and NVDiffrec on the Shiny dataset, which is a synthetic dataset with glossy surfaces.

Table 1. Quantitative results in terms of Chamfer distance on the scenes with high specularities on DTU [Jensen et al. 2014] and SK3D [Voynov et al. 2022]. Our method achieves state-of-the-art surface reconstruction results on both glossy (this table) and regular (see Appx. Table 6) scenes.

	DTU 63	DTU 97	DTU 110	DTU 122	**Mean**	Pot	Funnel	Snowman	Jug	**Mean**
NeuS [Wang et al. 2021]	1.01	1.21	1.14	0.54	0.98	2.09	3.93	1.40	1.81	2.31
NeRO [Liu et al. 2023]	1.32	1.47	1.14	0.57	1.12	6.03	2.63	1.71	4.23	3.65
Geo-NeuS [Fu et al. 2022]	**0.96**	**0.91**	**0.70**	**0.37**	**0.73**	1.88	2.03	1.64	1.68	1.81
PhySG [Zhang et al. 2021a]	4.16	4.99	3.57	1.42	3.53	14.40	7.39	1.55	7.59	7.73
IndiSG [Zhang et al. 2022b]	1.15	2.07	2.60	0.61	1.61	5.62	4.05	1.74	2.35	3.44
Factored-NeuS (ours)	0.99	1.15	0.89	0.46	0.87	**1.54**	**1.95**	**1.31**	**1.40**	**1.55**

From Fig. 6, we can observe that NeuS is easily affected by highlights, and the geometry reconstructed by Ref-NeRF has strong noise. PhySG is slightly better than IndiSG on the Shiny synthetic dataset with jointly optimizing materials and surfaces, such as toaster and car scenes, but still can not handle bright reflections. NVDiffrec works well on the teapot model but fails on other more challenging glossy surfaces. Our method can produce clean glossy surfaces without being affected by the issues caused by highlights. Overall, our approach demonstrates superior performance in surface reconstruction, especially on glossy surfaces.

4.3 Material reconstruction and rendering quality

In Tab. 2, we evaluate the quantitative results in terms of PSNR metric for material and illumination reconstruction on the IndiSG dataset compared with PhySG, NVDiffrec, and IndiSG. For completeness, we also compare to the case where the specular albedo improvement was not used in Stage 3 (See in Eq. 11 in Section 3.3). Regarding diffuse albedo, although NVDiffrec showed a slight improvement over us in the balloons scene, we achieved a significant improvement over NVDiffrec in the other three scenes. Our method achieved the best results in material reconstruction. Moreover, our method achieves the best results in illumination quality without using the specular albedo improvement. Additionally, our method significantly outperforms other methods in terms of rendering quality and achieves better appearance synthesis results. We present the qualitative results of material reconstruction in Fig. 3, which shows that our method has better detail capture compared to IndiSG and PhySG, such as the text on the balloon. Although NVDiffrec can reconstruct the nails on the backrest, its material decomposition effect is not realistic. The materials reconstructed by our method are closer to ground truth ones. Fig. 12 in Appx. also demonstrates the material decomposition effectiveness of our method on Shiny datasets with glossy surfaces. We show the diffuse albedo and rendering results of NVDiffrec, IndiSG, and our method. The rendering results indicate that our method can restore the original appearance with specular highlights more accurately, such as the reflections on the helmet and toaster compared to the IndiSG and NVDiffrec methods. The material reconstruction results show that our diffuse albedo contains less specular reflection information compared to other methods, indicating our method has a better ability to suppress decomposition ambiguity caused by specular highlights.

In addition to the synthetic datasets with ground truth decomposed materials, we also provide qualitative results on real-world captured datasets such as DTU and SK3D in Fig. 7. From the DTU data, we can observe that our method can separate the specular reflection component from the diffuse reflection component, as seen in the highlights on the apple, can, and golden rabbit. Even when faced with a higher intensity of specular reflection, as demonstrated in the example showcased in SK3D, our method excels at preserving the original color in the diffuse part and accurately separating highlights into the specular part. To offer a more detailed presentation of the reconstruction quality across continuous viewpoints, we include videos of diffuse albedo, indirect illumination, light visibility, and rendering for three different scenes in the supplementary materials. Furthermore, we perform relighting for these three scenes and provide videos to assess the relighting quality.

4.4 Ablation study

Materials and illumination. We conduct an ablation study on the different components we proposed by evaluating their material and lighting performance on a complex scene, the hotdog, as shown in Tab. 3. "SI" refers to surface improvement, which means using networks to jointly synthesize volume color and decomposed surface color. "VI" stands for visibility improvement, which involves continuous light visibility supervision based on the SDF. "SAI" refers to specular albedo improvement, which incorporates specular albedo into the BRDF of Spherical Gaussians. We compare different settings in terms of diffuse albedo, roughness, appearance synthesis, and illumination. We used IndiSG as a reference and find that introducing volume rendering can improve the accuracy of material and lighting reconstruction. When the surface has no defects, further performing the surface improvement will enhance the quality of roughness and rendering but may cause a decrease in lighting reconstruction quality. Making the visibility supervision continuous improves the reconstruction of diffuse albedo, roughness, and lighting, but it also affects rendering quality. Introducing specular albedo can greatly improve roughness and rendering quality but negatively affect lighting reconstruction quality. We further show qualitative results in Fig. 9. It can be observed that after improving the light visibility, the white artifacts at the edges of the plate in diffuse albedo are significantly reduced. Introducing specular albedo also makes the sausage appear smoother and closer to its true color roughness, represented by black. In terms of lighting, when not using specular albedo, the lighting reconstruction achieves the best result, indicating a clearer reconstruction of ambient illumination. In summary, our ablation

Table 2. Quantitative results in terms of PSNR on IndiSG [Zhang et al. 2022b] dataset for IndiSG and our method. "SAI" refers to specular albedo improvement.

	Baloons			Hotdog			Chair			Jugs			**Mean**		
	albedo	illumination	rendering	albedo	illumination	rendering	albedo	illumination	rendering	albedo	illumination	rendering	albedo	illumination	rendering
PhySG	15.91	13.89	27.83	13.95	11.69	25.13	14.86	12.26	28.32	16.84	10.92	28.20	15.39	12.19	27.37
NVDiffrec	**26.88**	14.63	29.90	13.60	22.43	33.68	21.12	15.56	29.16	11.20	10.47	25.30	20.41	13.56	29.51
IndiSG	21.95	25.24	24.40	26.43	21.87	31.77	24.71	**22.17**	24.98	21.44	20.59	24.91	23.63	22.47	26.51
Ours *w/o* IndiLgt	23.13	18.24	29.45	25.62	17.97	35.97	25.22	18.04	34.31	22.87	21.84	26.30	24.21	19.02	31.51
Ours *w/o* SAI	24.09	**25.97**	28.82	30.58	**23.50**	36.05	25.23	22.13	32.64	19.64	20.40	33.56	24.89	**23.00**	32.77
Ours	25.79	21.79	**33.89**	**30.72**	20.23	**36.71**	**26.33**	20.97	**34.58**	**22.94**	**21.84**	**36.48**	**26.28**	21.21	**35.41**

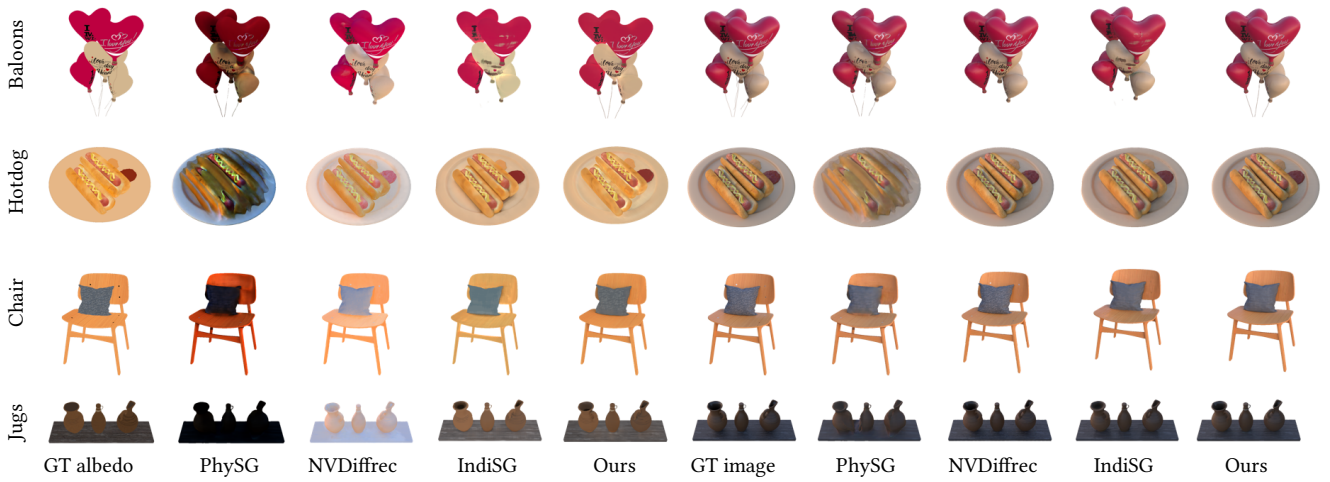

Fig. 3. Qualitative results on IndiSG dataset in terms of albedo reconstruction (left half) and novel view synthesis quality (right half).

Table 3. Ablation study for materials and illumination decomposition in terms of PSNR. "Alb" stands for "diffuse albedo", "Rough" is "roughness", "Rend" is "appearance", and "Illu" is "illumination".

Method	Alb	Rough	Rend	Illu
IndiSG [Zhang et al. 2022b]	26.44	15.97	31.78	21.88
Ours *w/o* SAI *w/o* VI *w/o* SI	29.31	16.98	35.48	23.48
Ours *w/o* SAI *w/o* VI	29.64	17.86	36.36	23.41
Ours *w/o* SAI	30.58	18.83	36.05	**23.50**
Ours	**30.76**	**23.10**	**36.71**	20.24

Fig. 4. Ablation study for different surface reconstruction methods. The numbers in the second line represent Chamfer distance.

study highlights the importance of taking into account various factors when reconstructing materials and illumination from images. By evaluating the performance of different modules, we can better understand their role in improving the reconstruction quality.

In stage 3, if we do not consider indirect illumination during the training process, the predicted results for rendering, material, and lighting will all experience a decline. The results are shown in Fig. 10. The specific PSNR metrics can be found in Tab. 2

Surface reconstruction. To validate our surface reconstruction strategy in Stage 1, we selected the Pot scene from SK3D and ablated the method the following way. "$\mathcal{L}_{vol} + \mathcal{L}^{I}_{sur}$" means that we only use volume rendering and surface rendering MLPs for surface reconstruction, without decomposing material information into diffuse and specular components. "$\mathcal{L}_{vol} + \mathcal{L}^{II}_{vol}$" means we use two volume reconstructions where one of them is split into diffuse and specular components. Just using "$\mathcal{L}^{II}_{vol}$" like Ref-NeRF to split diffuse and specular components fails to reconstruct the correct surface due to inaccurate normal vectors in the reflection direction computation. We provide the quantitative (Chamfer distance) and qualitative results of different frameworks in Fig. 4. It can be seen that synchronizing the volume color and the color on the surface point has a certain effect in suppressing concavities, but still cannot meet the requirements for complex glossy surfaces with strong reflections. Using volume rendering to decompose diffuse and specular components can result in excessive influence from non-surface points, which still causes small concavities. When using our loss "$\mathcal{L}_{vol}(\mathcal{L}^{I}_{vol}) + \mathcal{L}_{sur}(\mathcal{L}^{II}_{sur})$", we can achieve reconstruction results without concavities.

5 CONCLUSIONS

In this work, we propose Factored-NeuS, a novel approach to inverse rendering that reconstructs geometry, material, and lighting from

multiple views. Our first contribution is to simultaneously synthesize the appearance, diffuse radiance, and specular radiance during surface reconstruction, which allows the geometry to be unaffected by glossy highlights. Our second contribution is to train networks to estimate reflectance albedo and learn a visibility function supervised by continuous values based on the SDF, so that our method is capable of better decomposing material and lighting. Experimental results show that our method surpasses the state-of-the-art in both geometry reconstruction quality and material reconstruction quality. A future research direction is how to effectively decompose materials for fine structures, such as nails on the backrest of a chair.

In certain scenarios, our method still faces difficulties. For mesh reconstruction, we can only enhance results on scenes with smooth surfaces and few geometric features. Despite improvements on the glossy parts in the DTU 97 results, the overall Chamfer distance does not significantly decrease. As seen in Fig. 3, the reconstructed albedo of the chair still lacks some detail. The nails on the chair and the textures on the pillow are not accurately captured in the reconstructed geometry. Moreover, we do not foresee any negative societal implications directly linked to our research on surface reconstruction.

In future work, we would like to focus on the reconstruction of dynamic objects and humans. We also would like to include additional data acquisition modalities for improved performance.

REFERENCES

Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. 2021. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5855–5864.

Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. 2022. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5470–5479.

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

Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan Barron, and Hendrik Lensch. 2021b. Neural-pil: Neural pre-integrated lighting for reflectance decomposition. *Advances in Neural Information Processing Systems* 34 (2021), 10691–10704.

Brent Burley and Walt Disney Animation Studios. 2012. Physically-based shading at disney. In *Acm Siggraph*, Vol. 2012. vol. 2012, 1–7.

Eric R Chan, Connor Z Lin, Matthew A Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J Guibas, Jonathan Tremblay, Sameh Khamis, et al. 2022. Efficient geometry-aware 3D generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16123–16133.

Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. 2022. TensoRF: Tensorial Radiance Fields. *arXiv preprint arXiv:2203.09517* (2022).

Zhiqin Chen and Hao Zhang. 2019. Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5939–5948.

Youming Deng, Xueting Li, Sifei Liu, and Ming-Hsuan Yang. 2022. DIP: Differentiable Interreflection-aware Physics-based Inverse Rendering. *arXiv preprint arXiv:2212.04705* (2022).

Qiancheng Fu, Qingshan Xu, Yew Soon Ong, and Wenbing Tao. 2022. Geo-neus: Geometry-consistent neural implicit surfaces learning for multi-view reconstruction. *Advances in Neural Information Processing Systems* 35 (2022), 3403–3416.

Wenhang Ge, Tao Hu, Haoyu Zhao, Shu Liu, and Ying-Cong Chen. 2023. Ref-NeuS: Ambiguity-Reduced Neural Implicit Surface Learning for Multi-View Reconstruction with Reflection. *arXiv preprint arXiv:2303.10840* (2023).

Yuan-Chen Guo, Di Kang, Linchao Bao, Yu He, and Song-Hai Zhang. 2022. Nerfren: Neural radiance fields with reflections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18409–18418.

Rasmus Jensen, Anders Dahl, George Vogiatzis, Engin Tola, and Henrik Aanæs. 2014. Large scale multi-view stereopsis evaluation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 406–413.

Haian Jin, Isabella Liu, Peijia Xu, Xiaoshuai Zhang, Songfang Han, Sai Bi, Xiaowei Zhou, Zexiang Xu, and Hao Su. 2023. TensoIR: Tensorial Inverse Rendering. *arXiv preprint arXiv:2304.12461* (2023).

Zhengfei Kuang, Kyle Olszewski, Menglei Chai, Zeng Huang, Panos Achlioptas, and Sergey Tulyakov. 2022. NeROIC: neural rendering of objects from online image collections. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–12.

Ruofan Liang, Huiting Chen, Chunlin Li, Fan Chen, Selvakumar Panneer, and Nandita Vijaykumar. 2023. ENVIDR: Implicit Differentiable Renderer with Neural Environment Lighting. *arXiv preprint arXiv:2303.13022* (2023).

Yuan Liu, Peng Wang, Cheng Lin, Xiaoxiao Long, Jiepeng Wang, Lingjie Liu, Taku Komura, and Wenping Wang. 2023. NeRO: Neural Geometry and BRDF Reconstruction of Reflective Objects from Multiview Images. *arXiv preprint arXiv:2305.17398* (2023).

Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. 2019. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 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 *European conference on computer vision*. Springer, 405–421.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *arXiv preprint arXiv:2201.05989* (2022).

Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Müller, and Sanja Fidler. 2022. Extracting triangular 3d models, materials, and lighting from images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8280–8290.

Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. 2020. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3504–3515.

Michael Oechsle, Songyou Peng, and Andreas Geiger. 2021. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 165–174.

Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. 2021. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 14335–14345.

Johannes Lutz Schönberger and Jan-Michael Frahm. 2016. Structure-from-Motion Revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

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

Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. 2022. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 5481–5490.

Oleg Voynov, Gleb Bobrovskikh, Pavel Karpyshev, Andrei-Timotei Ardelean, Arseniy Bozhenko, Saveliy Galochkin, Ekaterina Karmanova, Pavel Kopanev, Yaroslav Labutin-Rymsho, Ruslan Rakhimov, et al. 2022. Multi-sensor large-scale dataset for multi-view 3D reconstruction. *arXiv preprint arXiv:2203.06111* (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. *Advances in Neural Information Processing Systems* (2021).

Yiqun Wang, Ivan Skorokhodov, and Peter Wonka. 2022. Improved surface reconstruction using high-frequency details. *Advances in Neural Information Processing Systems* (2022).

Yiqun Wang, Ivan Skorokhodov, and Peter Wonka. 2023. PET-NeuS: Positional Encoding Triplanes for Neural Surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Haoqian Wu, Zhipeng Hu, Lincheng Li, Yongqiang Zhang, Changjie Fan, and Xin Yu. 2023a. NeFII: Inverse Rendering for Reflectance Decomposition with Near-Field Indirect Illumination. *arXiv preprint arXiv:2303.16617* (2023).

Tianhao Wu, Hanxue Liang, Fangcheng Zhong, Gernot Riegler, Shimon Vainer, and Cengiz Oztireli. 2023b. αSurf: Implicit Surface Reconstruction for Semi-Transparent and Thin Objects with Decoupled Geometry and Opacity. *arXiv preprint arXiv:2303.10083* (2023).

Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K Wong. 2022. S ^{3}-NeRF: Neural Reflectance Field from Shading and Shadow under a Single Viewpoint. *Advances in Neural Information Processing Systems* 35 (2022), 1568–1582.

Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. 2021. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems* 34 (2021).

Lior Yariv, Peter Hedman, Christian Reiser, Dor Verbin, Pratul P Srinivasan, Richard Szeliski, Jonathan T Barron, and Ben Mildenhall. 2023. BakedSDF: Meshing Neural SDFs for Real-Time View Synthesis. *arXiv preprint arXiv:2302.14859* (2023).

Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. 2020. Multiview neural surface reconstruction by disentangling geometry and appearance. *Advances in Neural Information Processing Systems* 33 (2020), 2492–2502.

Alex Yu, Sara Fridovich-Keil, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. 2021. Plenoxels: Radiance fields without neural networks. *arXiv preprint arXiv:2112.05131* (2021).

Jingyang Zhang, Yao Yao, Shiwei Li, Jingbo Liu, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. 2023b. NeILF++: Inter-reflectable Light Fields for Geometry and Material Estimation. arXiv:2303.17147 [cs.CV]

Kai Zhang, Fujun Luan, Zhengqi Li, and Noah Snavely. 2022a. Iron: Inverse rendering by optimizing neural sdfs and materials from photometric images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5565–5574.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5453–5462.

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 Transactions on Graphics (TOG)* 40, 6 (2021), 1–18.

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

Youjia Zhang, Teng Xu, Junqing Yu, Yuteng Ye, Junle Wang, Yanqing Jing, Jingyi Yu, and Wei Yang. 2023a. NeMF: Inverse Volume Rendering with Neural Microflake Field. *arXiv preprint arXiv:2304.00782* (2023).

Yiyu Zhuang, Qi Zhang, Xuan Wang, Hao Zhu, Ying Feng, Xiaoyu Li, Ying Shan, and Xun Cao. 2023. NeAI: A Pre-convoluted Representation for Plug-and-Play Neural Ambient Illumination. *arXiv preprint arXiv:2304.08757* (2023).

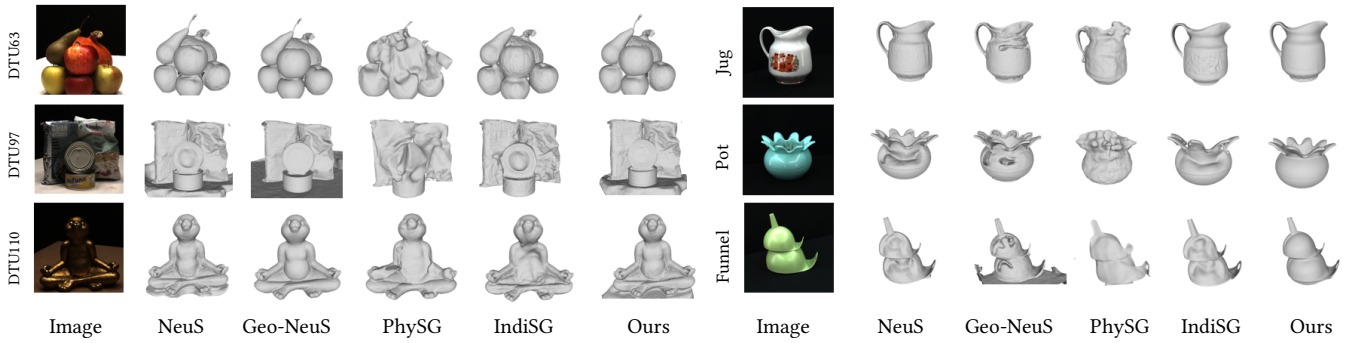

Fig. 5. Qualitative results for DTU (left) and SK3D (right).

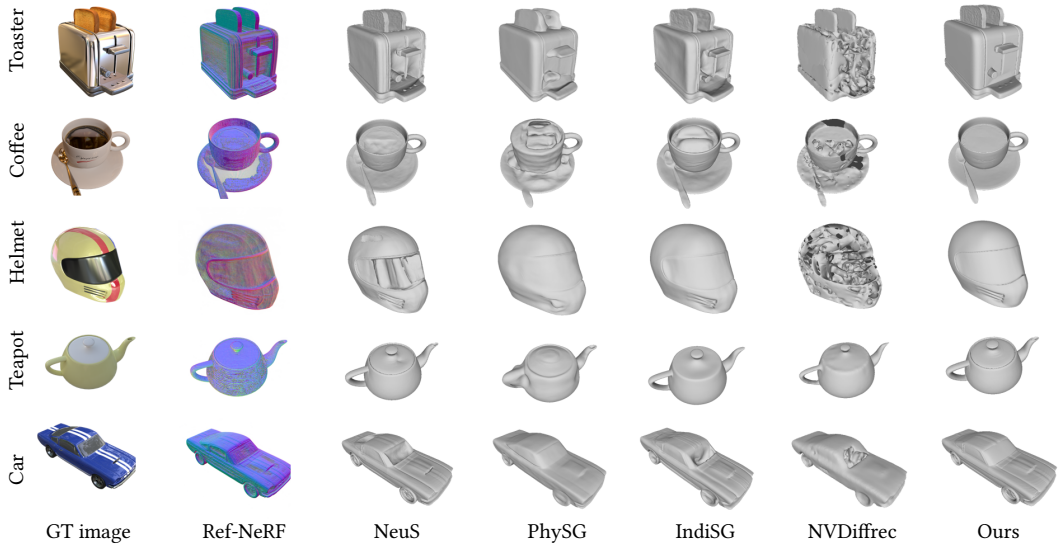

Fig. 6. Qualitative results for the Shiny dataset [Verbin et al. 2022].

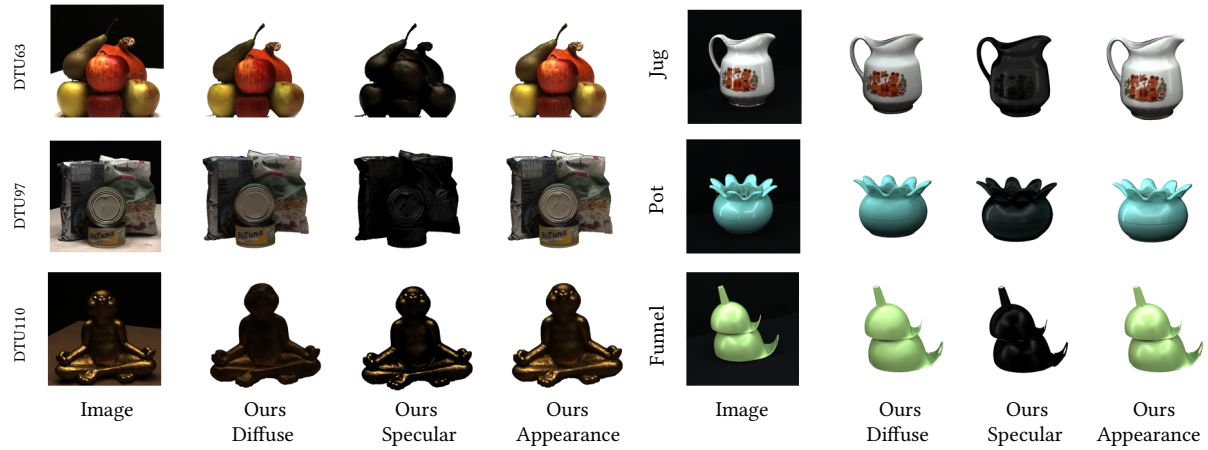

Fig. 7. Qualitative results for the DTU (left) and SK3D (right) datasets.

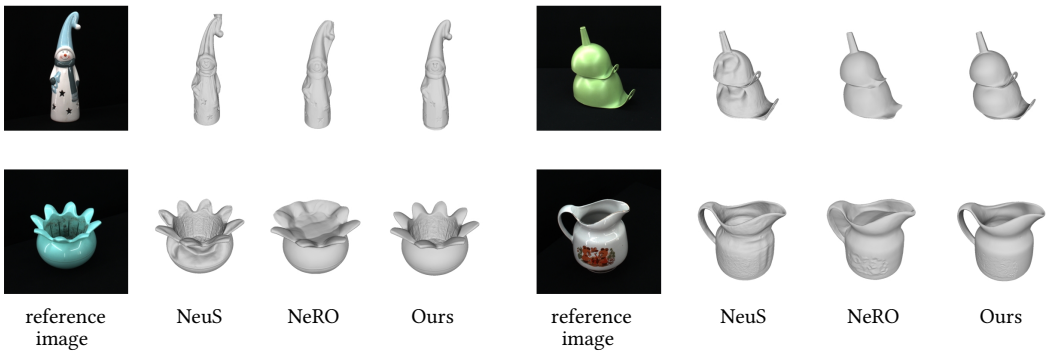

Fig. 8. Comparison with NeRO and NeuS on SK3D dataset.

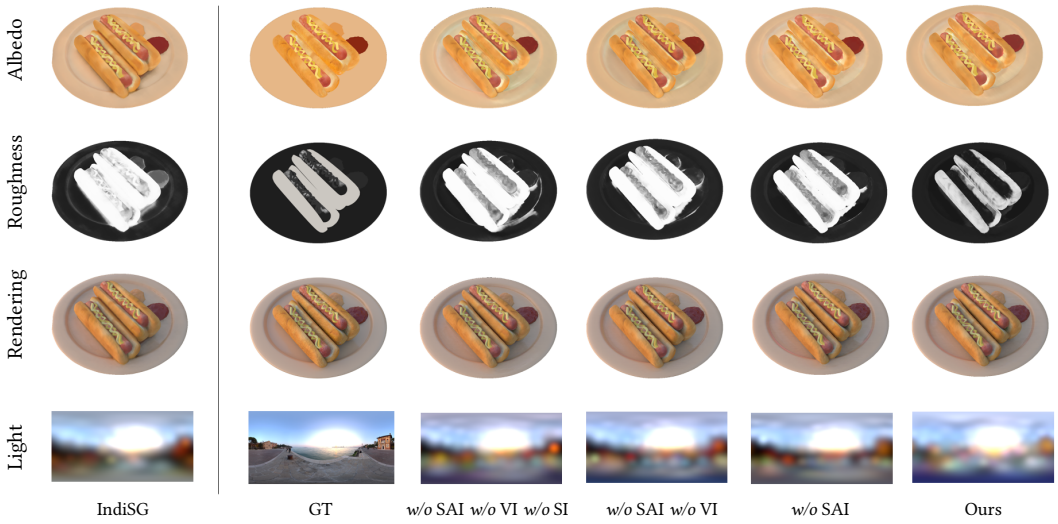

Fig. 9. Ablation study of material and illumination reconstruction.

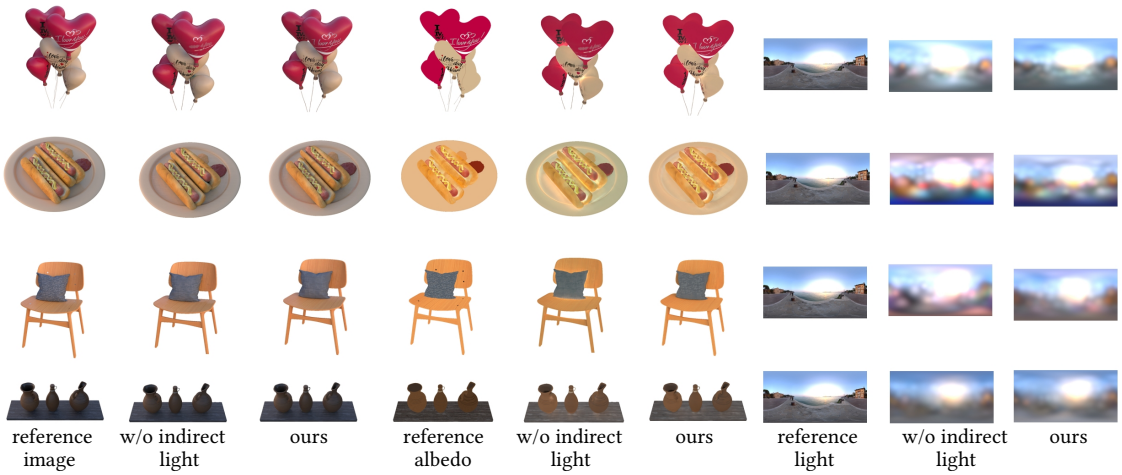

Fig. 10. Ablation study of indirect light.