NU-NeRF: Neural Reconstruction of Nested Transparent Objects with Uncontrolled Capture Environment

JIA-MU SUN, Institute of Computing Technology, CAS and KIRI Innovations, China
TONG WU, Institute of Computing Technology, CAS and University of Chinese Academy of Sciences, China
LING-QI YAN, Department of Computer Science, University of California, Santa Barbara, United States
LIN GAO[*], Institute of Computing Technology, CAS and University of Chinese Academy of Sciences , China

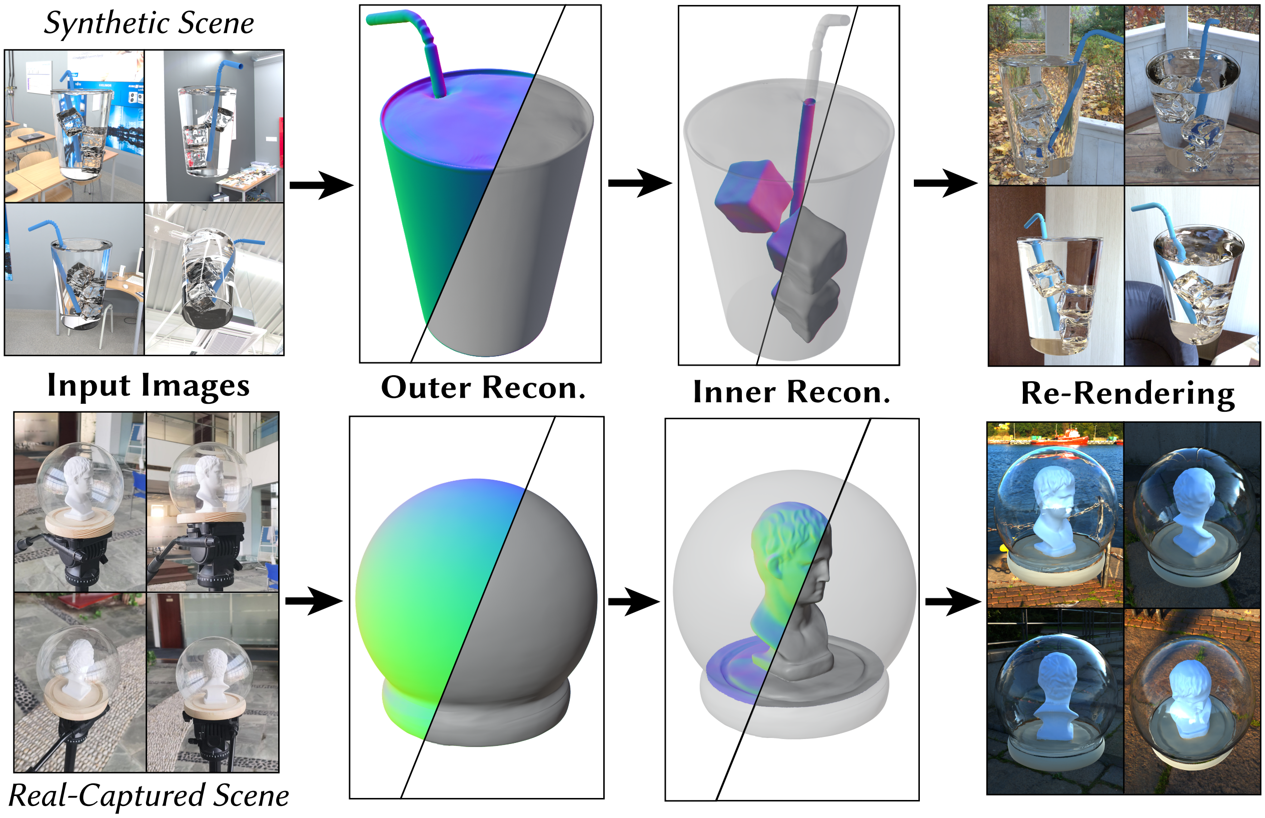


Fig. 1. Given a set of input images of a nested transparent object, our NU-NeRF pipeline can conduct high-quality reconstruction of both the outer and inner surfaces in a two-stage manner. The reconstruction results can be used for realistic re-rendering.

The geometry reconstruction of transparent objects is a challenging problem due to the highly noncontinuous and rapidly changing surface color caused by refraction. Existing methods rely on special capture devices, dedicated backgrounds, or ground-truth object masks to provide more priors and reduce the ambiguity of the problem. However, it is hard to apply methods with these special requirements to real-life reconstruction tasks, like scenes captured in the wild using mobile devices. Moreover, these methods can only cope with solid and homogeneous materials, greatly limiting the scope of the application. To solve the problems above, we propose NU-NeRF to reconstruct nested transparent objects without requiring a dedicated capture environment or additional input. NU-NeRF is built upon a neural signed distance field formulation and leverages neural rendering techniques. It consists of two main stages. In Stage I, the surface color is separated into reflection and refraction. The reflection is decomposed using physically based material and rendering. The refraction is modeled using a single MLP given the refraction and view directions, which is a simple yet effective solution

[*] Corresponding author: Lin Gao (gaolin@ict.ac.cn).

Authors' addresses: Jia-Mu Sun, Tong Wu, and Lin Gao are with Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, Chinese Academy of Sciences. Jia-Mu Sun is also with KIRI Innovations. Tong Wu and Lin Gao are also with University of Chinese Academy of Sciences. Ling-Qi Yan is with Department of Computer Science, University of California, Santa Barbara.
Authors' e-mails: sunjiamu21s@ict.ac.cn, wutong19s@ict.ac.cn, lingqi@cs.ucsb.edu, gaolin@ict.ac.cn.

0730-0301/2024/12-ART262 $15.00
https://doi.org/10.1145/3687757

of refraction modeling. This step produces high-fidelity geometry of the outer surface. In stage II, we use explicit ray tracing on the reconstructed outer surface for accurate light transport simulation. The surface reconstruction is executed again inside the outer geometry to obtain any inner surface geometry. In this process, a novel transparent interface formulation is used to cope with different types of transparent surfaces. Experiments conducted on synthetic scenes and real captured scenes show that NU-NeRF is capable of producing better reconstruction results than previous methods and achieves accurate nested surface reconstruction under an uncontrolled capture environment.

CCS Concepts: • **Computing methodologies** → *Image-based rendering*.

Additional Key Words and Phrases: neural radiance fields, transparent object, reconstruction

1 INTRODUCTION

Transparency is a common phenomenon that can be observed in everyday materials like water or glass, thus the reconstruction of transparent objects is required by numerous downstream applications. However, it remains difficult for computer algorithms to conduct reconstruction in the presence of transparency due to the highly complex light paths caused by refraction. The light rays can be greatly bent in this process, introducing inherent ambiguity. Thus, it remains an ill-posed problem to reconstruct the object's geometry.

To solve this ill-posed problem, a few methods [Huynh et al. 2010; Trifonov et al. 2006; Wetzstein et al. 2011] utilize special capture devices like polarisation cameras to obtain additional information apart from observed color. Later, works attempt to relax the requirements for capture devices but still need to capture the scene under a *controlled* environment like specially designed background patterns [Li et al. 2023a; Lyu et al. 2020; Wu et al. 2018; Xu et al. 2022] and opaque plane placed underneath the object of interest [Gao et al. 2023] to provide information about the intersection location between the refracted ray and the reference background or plane. Others do not require special capture devices or environments but require object masks [Chen et al. 2023; Li et al. 2020] or ground truth lighting conditions [Wang et al. 2023] as additional inputs. Moreover, vision-based methods like Li et al. [2020] needs extensive training dataset, and the domain gap between training scenes and actual scenes cannot be easily solved.

Despite the efforts made on transparent object reconstruction, *nested* transparent objects attract less attention. ReNeuS [Tong et al. 2023] reconstructs the opaque object inside a refractive interface, where the ground truth geometry and IoR of the outer surface are given. In addition, they need the object to be captured in a homogeneous lighting condition, which is a strong requirement. Bemana et al. [2022] leverage neural radiance fields to learn a volume with varying IoR (Index of Refraction), and calculate the light path numerically. To cope with nested objects, they use another radiance field to fit the inner surface. This method can support varying IoR and non-homogeneous materials and produces good novel view synthesis results, but it does not serve the purpose of geometry reconstruction. Therefore, the controlled capture environment (special background, homogeneous lighting, and opaque plane...), additional inputs (masks, lighting, and outer geometry/IoR), and the assumption on inputs (solid objects, homogeneous material) prohibit their practical use in the geometry reconstruction of nested transparent objects under an uncontrolled capture setting.

Aiming at overcoming both drawbacks, we propose NU-NeRF, a geometry reconstruction method for *nested* transparent objects under an *uncontrolled* environment. NU-NeRF leverages SDF-based neural implicit representation [Mildenhall et al. 2020; Wang et al. 2021] for high-fidelity surface reconstruction. Exemplar results of NU-NeRF are shown in Fig. 1. Our key observation is that although MLP is not perfectly accurate when modeling refracted lights, it can compensate for the remaining residual between the observed color and the reflection component, producing high-quality reconstruction results. In contrast, previous methods only use explicit ray tracing to obtain the refraction component, thus requiring additional knowledge such as object masks or special capture environments. NU-NeRF consists of two main stages: 1) Outer surface reconstruction. NU-NeRF utilizes an outer NeRF to model the surrounding environment and constrains the reflection on the object surface according to it. The refraction of the surface is modeled with a direction-dependent MLP, with the Fresnel effect considered. To cope with an uncontrolled environment, no prior (e.g. mask, background, and lighting) is used in this process. 2) Ray-traced inner surface reconstruction. During this stage, a ray-traced reconstruction technique is employed to obtain the surfaces in an outer-to-inner manner. The ray-tracing procedure ensures the accurate simulation of rays, and the surface reconstruction formulation is repeatedly used in the interior of the objects. We observe that our method can handle the reconstruction of complex surfaces while supporting various combinations of outer and inner surfaces with different types of materials. To sum up, our technical contributions include:

- We propose NU-NeRF, a pipeline that reconstructs nested transparent objects in an uncontrolled environment. NU-NeRF can reconstruct inner transparent or opaque surfaces in the transparent surface, greatly extending the applicability.
- We introduce a formulation of surface color components to support the reconstruction of transparent objects without prior correspondence knowledge.
- We propose a ray-traced iterative reconstruction strategy with a novel interface formulation corresponding to it that can be executed along with the surface reconstruction to cope with nested surfaces using ray-tracing.
- Experiments conducted on synthetic and real scenes demonstrate the proposed method can achieve better results than the baseline that requires object masks.

2 RELATED WORK

2.1 Neural Implicit Representations

Traditionally, explicit representations like voxel, mesh or point clouds are used in various applications like geometry processing and rendering [Xiao et al. 2020]. Recently, neural implicit representations like Neural Radiance Fields(NeRF) [Barron et al. 2021; Mildenhall et al. 2020] have gained more popularity, since only a few posed images are needed to obtain the geometry. The implicit representation is also flexible enough for applications like deformation [Park et al. 2021a,b; Pumarola et al. 2021; Tretschk et al. 2021] or generation [Chan et al. 2022; Niemeyer and Geiger 2021; Schwarz et al. 2020]. However, the density-based representation adopted by vanilla NeRF makes it hard to perform surface reconstruction. This also hinders further utilization of neural implicit representations like inverse rendering, since they heavily depend on accurate normal and surface estimates. Some works [Boss et al. 2021; Verbin et al. 2022; Zhang et al. 2021] try to implement decomposition on NeRF, they either are not physically-based or suffer suboptimal rendering quality due to the inaccurate surface and normal estimation. To solve this problem, VolSDF [Yariv et al. 2021] and NeuS [Wang et al. 2021] proposed to replace the density field with a Signed Distance Field (SDF), which enables high-fidelity surface reconstruction based on neural implicit representations while preserving the rendering quality of NeRF. Later, more works improved the quality of geometry reconstruction [Li et al. 2023b; Sun et al. 2022], which further enables more accurate inverse rendering and editing [Munkberg et al. 2022; Wu et al. 2023]. Recently, the focus has moved to reconstructing scenes with challenging visual effects. A relevant work, NeRO [Liu et al. 2023], is dedicated to reconstructing geometry and surface BRDF in the presence of strong reflections and achieved SOTA results. We further propose NU-NeRF to reconstruct the scene with refraction, which is a more challenging effect, in addition to reflection.

2.2 Transparent Object Reconstruction

Reconstructing transparent objects is a classical problem that has been extensively explored [Ihrke et al. 2010]. The multiple refractions of light rays can cause ambiguity and singularity, preventing accurate reconstruction. Previous works on transparent object reconstruction are generally about finding correspondences, thus requiring special capture setups. These setups include light fields probes [Wetzstein et al. 2011], tomography[Trifonov et al. 2006], and polarization capture devices [Huynh et al. 2010; Miyazaki and Ikeuchi 2005; Shao et al. 2022]. To eliminate the need for special capture devices, a series of works [Morris and Kutulakos 2011; Qian et al. 2016; Wu et al. 2018] proposed to find correspondence using specially designed background patterns. Given the ground truth locations of background intersections of the refracted rays and the screen, the normal and shape of the object can be optimized. DRT [Lyu et al. 2020] further improves this pipeline, leveraging differential ray tracing, coarse-to-fine optimization, and mask constraints. NeTO [Li et al. 2023a] refines the shape reconstruction quality by taking multiple refractions into consideration. Xu et al. [2022] propose to use a neural and explicit mesh hybrid pattern for reconstruction of transparent objects. Lin et al. [2023] utilize sinusoidal patterns and binary patterns as the background to conduct the reconstruction of more complex objects (with colored fluids and diffuse material). Apart from the background pattern, object masks or environment maps are also used as priors [Li et al. 2020], these priors can be used to guide the prediction of the refracted ray direction by neural networks and improve reconstruction results [Wang et al. 2023]. There are also methods dedicated to solving a slightly different problem: reconstructing objects behind transparent surfaces like water surface and the surface itself [Zhan et al. 2023].

Recently, with the ongoing trend of neural implicit representations, some neural radiance field-based works are proposed, achieving transparent object reconstruction without special capture devices, specially designed backgrounds, masks, or ground truth environment maps. Bemana et al. [2022] proposed conduct novel view synthesis by learning an IoR field along with the radiance and density, and solved the bent light path along the field. Recently, Deng et al. [2024] utilize a deformation network to predict the ray path and obtain faithful novel view synthesis results, but they are not capable of reconstructing geometry because they use density field instead of SDF as the representation. Gao et al. [2023] propose a two-stage method to first predict the multi-view silhouettes of the object and then the exact shape of a refractive object placed on an opaque plane. The plane is required by the method, since it is used for both geometry and appearance prior. NeRRF [Chen et al. 2023] uses a given input mask to obtain the object shape and then adopts explicit ray tracing to obtain the radiance estimate.

Our proposed NU-NeRF, on the other hand, eliminates all the capture setup (special capture devices and specially designed background) and additional input requirements (masks and ground truth lighting) used by the previous methods. Moreover, these mentioned methods are only capable of reconstructing “solid and homogeneous” objects made with materials of constant IoR (Index of Refraction), with no geometry inside. In contrast, our method can take input images of non-solid objects with varying IoR such as a plastic bottle half filled with water. For this type of objects, NU-NeRF can reconstruct the outer surface as well as any inner surfaces like the surface of the water. In the following sections, we cover the two components of NU-NeRF: Outer Surface Reconstruction (Sec. 3) and Ray-traced Inner Reconstruction (Sec. 4). The overview of the pipeline is shown in Fig. 2.

3 STAGE I: OUTER SURFACE RECONSTRUCTION

In this section, we first go through the surface reconstruction method, which is the building block used in both stages of the iterative geometry reconstruction strategy. Firstly, we introduce some preliminaries of neural rendering and surface reconstruction (Sec. 3.1). Secondly, we elaborate on the surface transmission formulation in the reconstruction process, which describes the rendering process of a shading point(Sec. 3.2). Thirdly, the optimization losses are described (Sec. 3.3).

3.1 Preliminaries

Neural Rendering and Implicit Representation. Neural implicit representations like NeRF [Mildenhall et al. 2020] generally adopt volume rendering techniques, sampling discrete points $\mathbf{p}_i$ along a ray and aggregating the colors $\mathbf{c}_i$ of these points using calculated weights w_i to obtain the final color $C = \sum_{i=0}^{N} w_i \mathbf{c}_i$. NeRF predicts density values σ_i along the ray, and obtains the weights assuming the scene consists of emissive volume: $w_i = \exp(\sum_{j<i} \sigma_j \Delta_j)(1 - \exp(-\sigma_i \Delta_i))$, where Δ_i is the distance between two neighboring sampled points. To improve the surface reconstruction quality, NeuS [Wang et al. 2021] proposes to predict the signed distance values s instead of density values. The SDF value

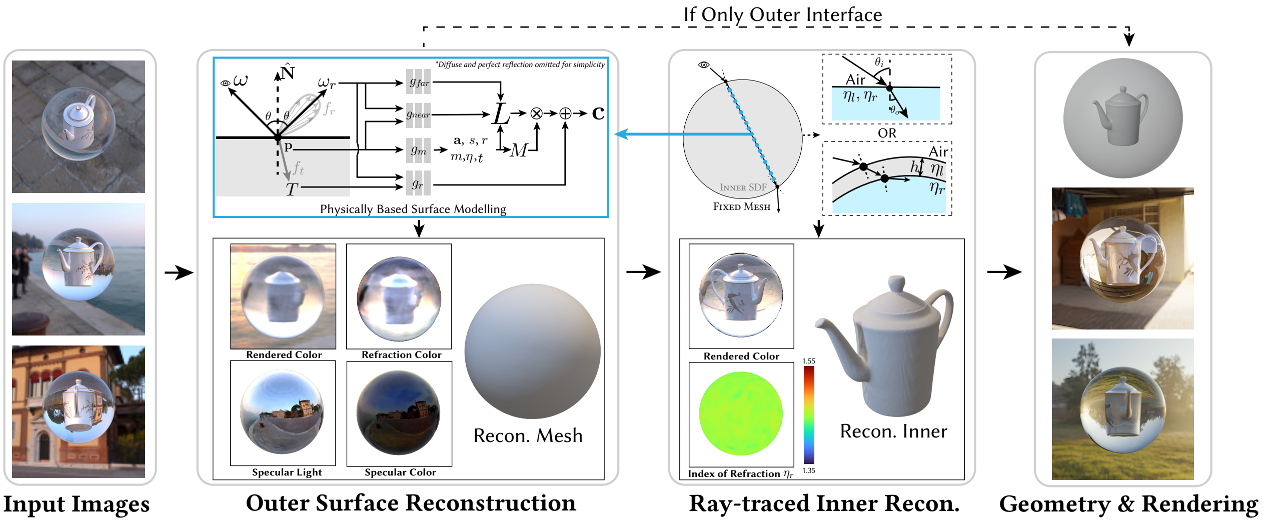


Fig. 2. **The overview of NU-NeRF pipeline.** Given a set of images of a nested transparent object, the reconstruction pipeline of NU-NeRF is separated into two stages. In the first outer interface reconstruction stage, neural rendering techniques are adopted. For each sample point, the split sum approximation is used to calculate the physically based reflection. Additionally, an MLP is used to predict the refracted light. Despite the blurry result predicted by the refraction MLP, it is vital for high-fidelity reconstruction of the outer geometry. In the second ray-traced inner surface reconstruction stage, the outer interface is modeled using two IoRs and an optional thickness. For each refracted ray, another neural rendering process is executed within the surface to obtain the inner geometry. Note that the surface formulation is used again in the second stage (marked by light blue). Finally, the outer and inner geometry can be merged together for downstream applications.

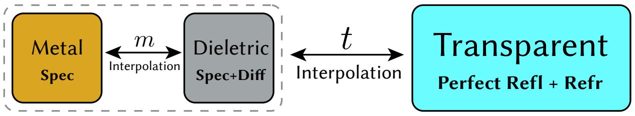


Fig. 3. **The definition of surface material.** To cope with the transparent refraction, we additionally introduce a “transparent” type of material and use a parameter t to interpolate between “transparent” and the regular “metal and dielectric” material used in previous literature.

can be transferred to density value with $\sigma_i = \max\left(-\frac{\frac{d\Phi_s}{dt}[s(t)]}{\Phi_s[s(t)]}, 0\right)$, where t is distance between camera origin and the sample point, $\Phi_s(x) = 1/(1 + e^{-zx})$, z is a trainable parameter. In this work, we adopt the basic formulation of NeuS, using SDF as the geometry representation. We assume the object resides in the unit sphere, and model the outer background using a NeRF $g_{bkgr}(\mathbf{p}, \boldsymbol{\omega})$, where $-\boldsymbol{\omega}$ is the ray direction.

The rendering equation. Aiming at the physically correct rendering of the surface geometry, we perform shading following the rendering equation [Kajiya 1986] considering transmission rather than mere reflection:

$$\mathbf{c} = \int_S f(\mathbf{p}, \omega_{in}, \omega) L_{in}(\mathbf{p}, \omega_{in}) |\mathbf{N} \cdot \omega_{in}| \, d\omega_{in} \tag{1}$$

where $\mathbf{c}$ is the final color, L_{in} is the incoming radiance, $\mathbf{N}$ is the surface normal, $\boldsymbol{\omega}$ is the reversed direction of the ray in the volume rendering process. f is the Bidirectional Scattering Distribution Function (BSDF) of the surface point $\mathbf{p}$.

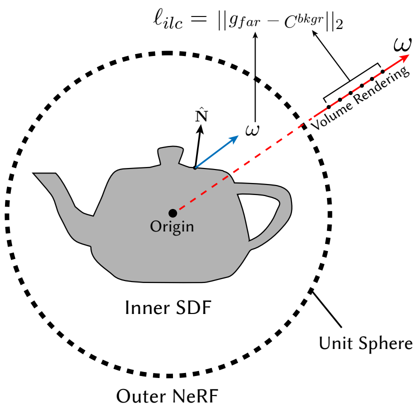
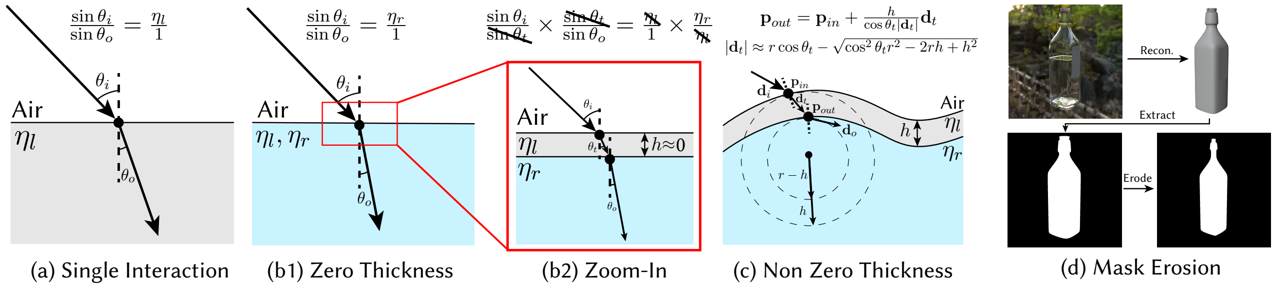


Fig. 4. **The proposed ℓ_{ilc}.** To address the overfitting problem of the refraction predictor g_r, ℓ_{ilc} is proposed. ℓ_{ilc} encourages the far light g_{far} to be the same as the background color $C^{bkgr} = \sum_i w_i^{bkgr} \mathbf{c}_i^{bkgr}$ obtained by background NeRF g_{bkgr}. Since the final color is the sum of reflection color and refraction color, ℓ_{ilc} prevents g_r from overfitting the color and helps to reconstruct more details of the geometry.

3.2 Formulation of Surface Reflection and Transmission

Material. As shown in Fig. 3, to enable physically based rendering of reflection and tramsmission, we parameterize the surface material as base color $\mathbf{a}$, roughness r, metallic m, Index of Refraction(IoR) η, and a transparent parameter t. All these parameters are predicted by an MLP (Multi-Layer Perceptron) $g_m(\mathbf{p})$, the input of which is the positions of the sample points.


$\frac{\sin\theta_i}{\sin\theta_o} = \frac{\eta_l}{1}$
θ_i
Air
η_l
θ_o
$\frac{\sin\theta_i}{\sin\theta_o} = \frac{\eta_r}{1}$
Air
η_l, η_r
$\frac{\sin\theta_i}{\cancel{\sin\theta_t}} \times \frac{\cancel{\sin\theta_t}}{\sin\theta_o} = \frac{\cancel{\eta_l}}{1} \times \frac{\eta_r}{\cancel{\eta_l}}$
Air
η_l
η_r
θ_t
$h\approx 0$
$\mathbf{p}_{out} = \mathbf{p}_{in} + \frac{h}{\cos\theta_t|\mathbf{d}_t|}\mathbf{d}_t$
$|\mathbf{d}_t| \approx r\cos\theta_t - \sqrt{\cos^2\theta_t r^2 - 2rh + h^2}$
$\mathbf{d}_i$
$\mathbf{p}_{in}$
$\mathbf{d}_t$
$\mathbf{p}_{out}$
$\mathbf{d}_o$
h
$r-h$
Air
η_l
η_r
Recon.
Extract
Erode

(a) Single Interaction (b1) Zero Thickness (b2) Zoom-In (c) Non Zero Thickness (d) Mask Erosion

Fig. 5. **Zero and non-zero thickness formulations.** When reconstructing non-solid objects like containers, two IoRs η_r, η_l are used to model the interface, corresponding to the IoR of the container itself and the inner substance. For very thin interface which can be regarded as zero-thick, we assume its two faces are parallel, thus the refraction direction only depends on the inner IoR η_r. When the thickness can not be ignored, we introduce another parameter h to model the thickness. We additionally use spheres to approximate the local area of the incident position. The normals and outgoing directions at the intersection points can be calculated using analytical calculations. For this type of surface, we utilize an eroded mask to ignore the pixels at the edge of the geometry since the light in that area will undergo complex total internal refraction.

Overall Formulation of BSDF. In the scope of this paper, we only consider the transmission component of a nearly perfect smooth surface. Thus, we explicitly categorize the surface into "transmissive" and "reflective" types and use an additional material parameter t to interpolate between them. This is similar to the "metal" and "dielectric" interpolation used in the variants of the Disney BRDF [Burley 2012], and both interpolations are used in this paper. The material formulation is illustrated in Fig. 3.

The two types of materials are separately represented by two BSDFs: f_r (reflective material) and f_t (transmissive material). We define f_r as the sum of Lambertian diffuse reflectance and Cook-Torrance specular reflectance [Cook and Torrance 1982], as in the case of previous methods aiming to perform inverse rendering [Liu et al. 2023; Zhang et al. 2022]. f_t defines the sum of the reflection and refraction color of a perfectly smooth transparent surface. Because f_r and f_t represent two different types of material that both can reflect light, the reflection component appears in both f_r and f_t, but they are calculated differently. Note that the Fresnel coefficient F_p is calculated using the normal N, in contrast with F in the Cook-Torrance reflectance that uses the half vector. An additional transparent value t is introduced to control how transparent the material is.

$$f_r = \frac{(1-m)\mathbf{a}}{\pi} + \frac{DFG}{4(\boldsymbol{\omega}_{in}\cdot\mathbf{N})(\boldsymbol{\omega}\cdot\mathbf{N})} \tag{2}$$

$$f_t = \frac{(1-F_p)\delta(\boldsymbol{\omega}, \mathrm{Refr}(\boldsymbol{\omega}_{in}, \eta, \mathbf{N})) + F_p\delta(\boldsymbol{\omega}, \mathrm{Refl}(\boldsymbol{\omega}_{in}, \mathbf{N}))}{|\mathbf{N}\cdot\boldsymbol{\omega}_{in}|} \tag{3}$$

$$f = (1-t)f_r + tf_t \tag{4}$$

where $\mathbf{a}$ is the diffuse albedo, D, F, G are the classic microfacet distribution, Fresnel and geometry components. They depend on directions $\boldsymbol{\omega}, \boldsymbol{\omega}_{in}$, surface roughness r, and metallic m. In Eqn. 2, we omit the specific formulae of them for simplicity. δ is the Dirac delta function, η is the ratio of the IoRs of the inner and outer material, and Refr is the refraction direction according to Snell's law, Refl is the perfect reflection direction. The detailed formulation of Eqns. 1,2,3,4 to the neural rendering will be covered in the following subsections.

Reflection. For reflection rendering, we adopt the basic formulation of NeRO [Liu et al. 2023] and use the split sum [Karis 2013] approximation. The split sum is a technique dedicated to calculating the rendering equation (Eqn. 1) efficiently. It replaces the integral of specular reflection with the multiplication of two integrals

$$\int_{\mathcal{H}} \frac{DFG}{4(\boldsymbol{\omega}_{in}\cdot\mathbf{N})}\,\mathrm{d}\boldsymbol{\omega}_{in} \approx \underbrace{\int_{\mathcal{H}} L_{in}\,\mathrm{d}\boldsymbol{\omega}}_{L} \cdot \underbrace{\int_{\mathcal{H}} \frac{DFG}{(4\boldsymbol{\omega}\cdot\mathbf{N})}\,\mathrm{d}\boldsymbol{\omega}_{in}}_{M} \tag{5}$$

where $\mathcal{H}$ is the hemisphere above the surface, defined by the normal vector $\mathbf{N}$. In Eqn. 5, L depends on the incoming light, and the incoming light is pre-filtered according to different roughness by convolving the light with GGX distribution in the original work by Karis [2013] and also in some concurrent works of it [McAuley et al. 2012, 2013]. However, in our setting, the light is unknown in the training stage, thus the pre-filtering technique in the original split sum method is not available. Thus we introduce the Integrated Directional Encoding (IDE) techniques [Verbin et al. 2022] to achieve "filtering" inside of the MLP. To model the light, two MLPs g_{far} and g_{near} are fitted, corresponding to the incident light from infinite far away and the indirect inter-reflected light by the geometry itself. An additional interpolation factor s (also referred to as 'occlusion value') is predicted for every sample point by the material network g_m following NeRO [Liu et al. 2023]. According to the analysis given by NeRO [Liu et al. 2023], L can be approximated by:

$$L \approx (1-s)g_{far}(\mathrm{IDE}(\boldsymbol{\omega}_i, r)) + sg_{near}(\mathrm{IDE}(\boldsymbol{\omega}_i, r), \mathbf{p}) \tag{6}$$

where IDE is the Integrated Directional Encoding [Verbin et al. 2022], which transforms the integral of the lighting into the "integral" of the direction and significantly reduces the number of queries of the MLPs. As a special case, diffuse color can be approximated using $\mathbf{c}_d = \frac{\mathbf{a}}{\pi} g_{far}(\mathrm{IDE}(\boldsymbol{\omega}_i, 1))$.

On the other hand, M depends only on the surface material $\mathbf{a}, r, m, \eta, t$. The Fresnel Reflectance F can be approximated using Schlick Approximation $F \approx F_{sch} = F_0 + (1-F_0)(1-\boldsymbol{\omega}\cdot\mathbf{H})^5$ [Schlick 1994], where F_0 is the reflectance when the incident angle equals 0, and $\mathbf{H}$ is the half vector $\mathbf{H} = \frac{(\omega+\omega_{in})}{||(\omega+\omega_{in})||}$. In our setting, $F_0 =$

$((1-m)\cdot a_s + m\cdot \mathbf{a})$, where a_s is the specular color of dielectric material which can be calculated using $a_s = (\frac{\eta-1}{\eta+1})^2$. If we substitute F in M with F_{sch}, M can be separated into the following form:

$$M = F_0 \int_{\mathcal{H}} \frac{DG(1-(1-\boldsymbol{\omega}\cdot\mathbf{H})^5)}{(4\boldsymbol{\omega}\cdot\mathbf{N})}\,\mathrm{d}\boldsymbol{\omega}_{in} + \int_{\mathcal{H}} \frac{DG(1-\boldsymbol{\omega}\cdot\mathbf{H})^5}{(4\boldsymbol{\omega}\cdot\mathbf{N})}\,\mathrm{d}\boldsymbol{\omega}_{in} \tag{7}$$

Both the integrals in Eqn. 7 can be pre-calculated and stored in 2-D lookup textures since they depend on two scalar parameters: roughness r and $\boldsymbol{\omega}\cdot\mathbf{N}$. Thus, at rendering time, only two texture queries are needed to evaluate M. We rename these two integrals into F_1, F_2, then M can be calculated by:

$$M = ((1-m)\cdot a_s + m\cdot\mathbf{a})\cdot F_1 + F_2 \tag{8}$$

Transmission. For the transmission component, it is difficult to directly model it using explicit ray tracing and light transport laws due to the absence of surface geometry in the training stage. Instead, we choose to use an MLP $g_r(\mathbf{p}, \boldsymbol{\omega}, \mathrm{Refr}(\boldsymbol{\omega}, \eta, \mathbf{N}))$ to directly model the refraction. Note that both the original ray direction $\boldsymbol{\omega}$ and the refracted direction are input to g_r. We now write the full formulation for surface reconstruction:

$$\begin{aligned}\mathbf{c} &= (1-t)(\mathbf{c}_d + L\cdot M)\\ &+ t(1-F_{sch,p})(g_r(\mathbf{p}, \boldsymbol{\omega}, \mathrm{Refr}(\boldsymbol{\omega}, \eta, \mathbf{N}))\\ &+ tF_{sch,p}g_l(\mathbf{p}, \mathrm{Refl}(\boldsymbol{\omega}, \mathbf{N}))\end{aligned}$$

where $F_{sch,p}$ is the schlick approximation with the normal vector input $F_{sch,p} = F_0 + (1-F_0)(1-\boldsymbol{\omega}\cdot\mathbf{N})^5$. To g_l is the predicted incident light with roughness set to 0: $g_l(\mathbf{p}, \boldsymbol{\omega}) = (1-s)g_{far}(\mathrm{IDE}(\boldsymbol{\omega}_i, 0)) + sg_{near}(\mathrm{IDE}(\boldsymbol{\omega}_i, 0), \mathbf{p})$. According to our observation, g_r cannot predict accurate refraction color due to the highly non-continuous and rapidly changing nature of refraction, resulting in blurred rendered color (Please refer to Fig. 2). However, g_r can greatly compensate for the reflection color by providing an average among the position and direction, allowing high-fidelity reconstruction. If g_r is not applied, the geometry will suffer from significant degeneration, or the method will fail to reconstruct any meaningful geometry (Please refer to Sec. 5.5).

3.3 Optimization Losses

In this section, we go through the optimization losses used in the surface reconstruction process. Firstly, we adopt all the losses from the Stage I reconstruction in NeRO [Liu et al. 2023], including the losses $\ell_{render}, \ell_{eikonal}, \ell_{occ}, \ell_{stable}$. ℓ_{render} is the Charbonier loss [Charbonnier et al. 1994] between the rendered color and the input image pixel color. $\ell_{eikonal}$ is the eikonal loss that regularizes the gradients of the SDF to 1, as applied in NeuS [Wang et al. 2021]. ℓ_{occ} is the occlusion loss encouraging the occulsion value s to be the same as the ray-traced ground truth s_{march}: $\ell_{occ} = ||s_{march} - s||_1$. ℓ_{stable} is a stabilization loss that prevents the zero level set of the SDF from overly expanding or shrinking applied at the first 1,000 steps of training. For more detail about this loss, please refer to NeRO [Liu et al. 2023].

However, we observed this version suffers from a suboptimal reconstruction of geometry details (Please refer to Sec. 5.5, w/o ℓ_{ilc} ablation). This is caused by the introduction of refraction predictor g_r. g_r tends to overfit the reflection component, causing the actual reflection predictors g_{far} to produce inaccurate light. Since g_r is hard to constrain, we choose to add a regularization on g_{far}. Because the final rendered color is the sum of refraction and reflection, this regularization is effective even if it is not directly applied on g_r. It is called **incident light correspondence loss** ℓ_{ilc}, which encourages the incident light prediction g_{far} to be the same as the background NeRF:

$$\ell_{ilc} = ||g_{far}(\mathrm{IDE}(\boldsymbol{\omega}, 0)) - \sum_i w_i^{bkgr}\mathbf{c}_i^{bkgr}||_2 \tag{9}$$

where $w_i^{bkgr}, \mathbf{c}_i^{bkgr}$ are the calculated weights and colors from the background NeRF g_{bkgr} along the ray. We show the formulation and purpose of ℓ_{ilc} in Fig. 4.

We write the total loss in the surface reconstruction phase as follows:

$$\mathcal{L} = \ell_{render} + \mathbb{I}(\mathrm{step} < 1000)\ell_{stable} + \sum_k \lambda_k\ell_k \quad (k\in\{occ,eikonal,ilc\}) \tag{10}$$

where $\mathbb{I}$ is an indicator function, step is the training step, and $\lambda_{occ}, \lambda_{eikonal}, \lambda_{ilc}$ are hyperparameters, controlling the multiplier of the corresponding losses.

4 STAGE II: RAY-TRACED ITERATIVE RECONSTRUCTION

In this section, we describe the ray-traced iterative reconstruction. The strategy can enable NU-NeRF to reconstruct the surface inside the outer transparent surface. We call the outer transparent surface "interface" from now on. We first go over an ideal case where the interface has zero thickness (Sec. 4.1) and then extend the case to non-zero thickness interfaces (Sec. 4.2). Finally, we combine the interface ray-tracing and the interface reconstruction described in Sec. 3 and form a complete iterative strategy (Sec. 4.3). In Fig. 5, we show our formulation of zero and non-zero thickness interface formulations.

4.1 Zero Thickness Interface

In objects made of solid transparent material (e.g. glass), the light goes through the interface and is both reflected and refracted. The incident energy is distributed according to the Fresnel equations (See Fig. 5(a)). In this case, the light only undergoes one interaction with the interface. However, in real life, it is common for light to undergo two interactions, like on the surface of transparent containers. We will first describe a simple case: the interface is very thin. Thus, the light is refracted into it and immediately shoots out. Since two consequent interactions occur in this process, the Fresnel term needs to be applied twice. As the material is super thin, two interactions can be considered to happen in the same position, and the surface normals at both points are the same (See Fig. 5(b)). We then formulate this type of interface with **two separate IoRs** η_l, η_r. η_l is the IoR of the interface material and is used to calculate the specular color a_s. η_r is the IoR of the material of the substance inside, used for calculating the refraction direction. For example, for a plastic bottle with no water inside, $\eta_l = 1.5, \eta_r = 1.0$. As a special case, the solid objects with a single interaction are modeled using two identical IoRs to simplify implementation.

4.2 Non-zero Thickness Interface

The extension from zero thickness interface to non-zero thickness interface is natural. We additionally introduce a parameter h to model the thickness of the interface. Since the thickness cannot be ignored, the incident and outgoing positions of the light are not the same, and the normals at these points are different. To capture this effect accurately, we model the local area of the incident point using a sphere, the radius of which is calculated using the Gaussian curvature K of the incident point, which can be calculated on meshes using numerical methods [Meyer et al. 2002]. The radius is obtained using $r = 1/\sqrt{K}$, assuming two principal curvatures κ_1, κ_2 are the same (See Fig. 5(c)).

However, there is another concern about the non-zero thickness interface. When the incident position is near the edge of the geometry, the light will repeatedly undergo total internal reflections inside the interface. In this case, the light does not go into the inner of the interface. To cope with this case, we render a mask using the geometry obtained in the first stage (Sec. 3), and apply an erosion filter on the mask, eliminating the samples near the edge of the geometry (See Fig. 5(d)). The kernel size e_s is treated as an adjustable hyperparameter.

4.3 "Onion"-like Iterative Strategy

We now combine the interface reconstruction and the ray-traced reconstruction. The overall pipeline is like "peeling an onion", repeating the same procedure for the outer and inner surfaces:

(1) Given the input images and the corresponding poses of the object, apply the interface reconstruction (Sec. 3). In this step, the outer geometry (can be transparent or opaque) is reconstructed, the materials of the geometry and background NeRF are also learned.

(2) The geometry of the first step is fixed and transformed to an interface with two IoRs and an optional thickness defined on each surface point (Secs. 4.1, 4.2). The rays are traced and refracted into the interface. Another interface reconstruction process is performed within the interface to obtain the inner geometry. All the networks from Stage I, excluding the SDF network, are directly loaded and learned jointly with a low learning rate. This is to conduct a refinement of Stage I networks via the accurate ray tracing of Stage II. In the second interface reconstruction process, the outer NeRF and ℓ_{ilc} are removed since the region of interest is completely fixed.

In theory, (2) can be repeated to cope with geometry with more than two layers. However, in real life, geometries with more than one nested transparent interface are rare. Considering the simplicity of the pipeline, we only consider two-layer geometry in this paper.

5 RESULTS AND EVALUATIONS

5.1 Experiment Settings.

Datasets and Evaluation Metrics We evaluate our method and baselines on two types of datasets:

(1) Synthetic dataset. We collected three types of objects from the public repository: **Solid Objects** like those evaluated in previous works contain datasets *Pig* and *Monkey*. **Transparent + Transparent Objects**, in which both outer and inner geometries are transparent. These contain *PlasticWater* and *Glasswater*. **Transparent + Opaque Objects**, in which the inner geometry is opaque. This contains *Spherepot*. **Complex Combined Objects**, in which the inner geometry contains both opaque and transparent surfaces. This contains *GlassIce*. Each dataset contains 250 images. We render the ground truth object masks along with the images for baseline methods requiring the mask.

(2) Real dataset. We take three datasets from Bemana et al. [2022]: *Ball, Glass, WineGlass*, and collect 2 datasets from the internet: *Lamp* and *PlasticBottle*. These datasets do not contain ground truth shapes. We additionally captured 3 datasets by ourselves: *BallStatue, RealBottle, RealBottle2*. The input images are taken from a 1-minute video clip that is captured using a cellphone. It is important to notice there are various artifacts like defocusing and contre-jour in about 10% of the images (See Fig. 6 for an example). To obtain the ground truth shapes, we paint the objects with AESUB Blue Scanning Spray and scan them using Revopoint POP 3 scanner. Quantitative experiments are only conducted on the self-captured datasets. The object masks for the use of baseline methods are annotated using off-the-shelf methods [Contributors 2020] and manually adjusted for better accuracy.

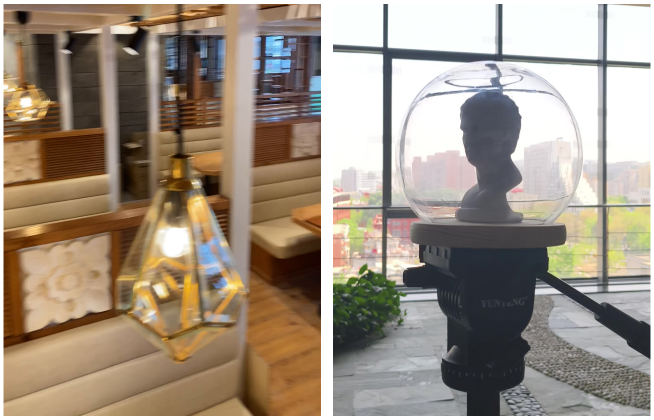

(a) Defocus Artifact (b) Contre-Jour Artifact

Fig. 6. **Examples of artifacts in the real-captured dataset.** The defocus artifact comes from the rapid movement of the capture devices, making the automatic focusing system fail. The contre-jour artifact is generated by the strong light in the background and the low dynamic range imaging system, making the foreground object appear overly dark.

For the quantitative evaluation of the reconstructed geometry, we compare the Chamfer Distance (CD) [Barrow et al. 1977] and Earth Mover's Distance(EMD) [Rubner et al. 2000] between the reconstructed geometry and the ground truth mesh. Both distances are calculated with 50,000 randomly sampled points on the meshes. Note that for nested objects, CDs of both the outer surface and the inner surface are calculated. For the baseline methods or ablated versions that are not able to reconstruct the inner surface, only the

Table 1. Quantitative comparison of reconstruction results using the chamfer distance metric $(\times 10^{-4})$ and Earth Mover's Distance metric $(\times 10^{-3})$ on the synthetic dataset.

Scenes	Ours (Outer)		NeMTO [Wang et al. 2023]		Li et al. [2020]		Ours (Inner)	
	CD ↓	EMD ↓	CD ↓	EMD ↓	CD ↓	EMD ↓	CD ↓	EMD ↓
Pig	**0.73**	6.31	1.16	**4.25**	5.57	22.6	N/A	N/A
Monkey	**1.02**	**4.80**	1.43	11.6	8.56	76.3	N/A	N/A
SpherePot	**0.99**	**3.75**	2.41	8.27	5.50	8.38	1.32	3.27
PlasticWater	**1.67**	**55.0**	3.79	62.3	21.8	91.3	1.64	0.92
GlassWater	**0.61**	**14.8**	4.09	15.6	5.89	17.4	1.11	1.52
GlassIce	**1.62**	**28.6**	68.1	28.8	48.3	32.0	1.48	4.42

distance of the outer surface is calculated. The ablation studies are only conducted on the synthetic dataset.

Baseline. Since our method focuses on the geometry reconstruction of nested transparent objects with an uncontrolled capture setup, there is no State-of-the-art method with the exact same setting as ours (Please refer to the supplementary material for settings of different methods). Recent methods require either dedicated background [Li et al. 2023a; Lyu et al. 2020; Xu et al. 2022] or a certain environment [Gao et al. 2023]. On the other hand, Li et al. [2020] and NeMTO [Wang et al. 2023] only require object masks and the environment map, both of which can be estimated given only the input images without re-capturing the images, which is closest to our settings. Therefore, we set Li et al. [2020] and NeMTO [Wang et al. 2023] as baselines for both qualitative and quantitative comparisons. Recent Gao et al. [2023] and ReNeuS [Tong et al. 2023] also explore transparent object reconstruction but have more requirements on the capture environment including a large enough opaque plane, ground truth outer surface, and homogeneous lighting. Therefore, we select a few cases that meet the capture requirements for these methods for qualitative comparisons. For detailed settings for different baselines, please refer to the supplementary material.

5.2 Results on Synthetic Datasets

The qualitative and quantitative reconstruction results of our NU-NeRF, NEMTO [Wang et al. 2023], and Li et al. [2020] on synthetic datasets are shown in Fig. 7 and Table 1. The results on *Pig* and *Monkey* cases show that our method can reconstruct solid transparent shapes with more geometric details compared to previous methods that require additional input despite some sharp areas being "smoothed", like the eyes of the *Pig* and *Monkey* datasets. This is because strong total internal reflection appears in that area. This effect can also be observed in other SOTA (State-Of-The-Art) methods [Li et al. 2020; Wang et al. 2023]. For more complicated cases with nested surfaces like *SpherePot*, *PlasticWater*, and *GlassWater*, our method is not only capable of reconstructing the detailed outer geometry (e.g. the folds on the *PlasticWater* case) but also can reconstruct the inner surface. On the contrary, NEMTO [Wang et al. 2023] and Li et al. [2020] even fail to recover the detailed geometry of the outer surfaces, for example, the *PlasticWater* case. This is because NeMTO's ray bending network fails to generalize to such cases and predicts wrong light paths and Li et al.'s normal and point cloud prediction networks only estimate the normal and point cloud of the outer surface. The *GlassIce* case is even more challenging, with the inner objects containing both opaque straw and transparent ice and the straw having parts in both inner and outer sections in the scene. Our method can reconstruct all the inner geometry faithfully, and the inner straw and the outer straw are aligned with each other. Re-rendering of the scene in Fig. 1 shows the reconstruction result is accurate. Such a case is obviously beyond the reconstruction capability of the baseline methods [Li et al. 2020; Wang et al. 2023] since they only focus on the outer surface reconstruction and rely on inaccurate light path, normal, and point cloud estimation. As a result, their reconstruction results of the outer surface tend to be smoother than ours and leave out details.

Table 2. Quantitative comparison of reconstruction results using the chamfer distance metric $(\times 10^{-4})$ and Earth Mover's Distance metric $(\times 10^{-3})$ on the real dataset with ground truth geometry.

Scenes	Ours (Outer)		NeMTO [Wang et al. 2023]		Li et al. [2020]		Ours (Inner)	
	CD ↓	EMD ↓	CD ↓	EMD ↓	CD ↓	EMD ↓	CD ↓	EMD ↓
BallStatue	**3.02**	**25.5**	24.4	59.3	13.1	89.3	0.94	8.66
Bottle	**4.57**	**6.78**	48.0	50.1	37.1	48.2	N/A	N/A
Bottle2	**3.34**	**27.1**	15.7	74.2	21.7	126.9	N/A	N/A

5.3 Results on Real-Captured Datasets

The reconstruction results of our NU-NeRF, NEMTO [Wang et al. 2023], and Li et al. [2020] on real-captured datasets are shown in Fig. 8 and Fig. 9. Fig. 8 includes five real scenes with no ground truth geometry, three of which (*Ball*, *Glass*, and *Plastic*) contain objects without geometry located inside, and the other two contain transparent outer surfaces with objects inside. Fig. 9 includes three scenes captured with a scanner, where the ground truth inner and outer geometry can be obtained for both qualitative and quantitative comparisons. It can be observed that our method can reconstruct outer geometry without object masks in complex real-captured scenes in Fig. 8. Although compared to the results on synthetic data, the lack of images (about 100 images every scene, significantly fewer than 250 in synthetic scenes) and inaccurate camera poses negatively affect some regions of the reconstructed results, e.g. the bottom area of *Glass* case. In the *WineGlass* case, our NU-NeRF can reconstruct the inner surface although there are inaccuracies in the outer geometry. The shape of the pencil in both outer and inner sections is aligned, thanks to the ray-tracing technique used in Stage II. In the *Lamp* case, a light bulb with strong emitted color is included, and our method can robustly reconstruct both the outer and inner geometry. However, both baseline methods fail to reconstruct the inner geometry and perform worse in outer geometry reconstruction. In Fig. 9, the *BallStatue* case contains both transparent outer surface and opaque inner geometry. Additionally, there are reflections on the outer surface. Our method deals with the complex visual effects well without losing too much geometry detail as reflected in the quality of inner and outer surfaces. Both the baseline methods [Li et al. 2020; Wang et al. 2023] can produce plausible results on solid objects but fails to produce faithful results when there are nested surfaces and both transparent and opaque materials. NeMTO [Wang

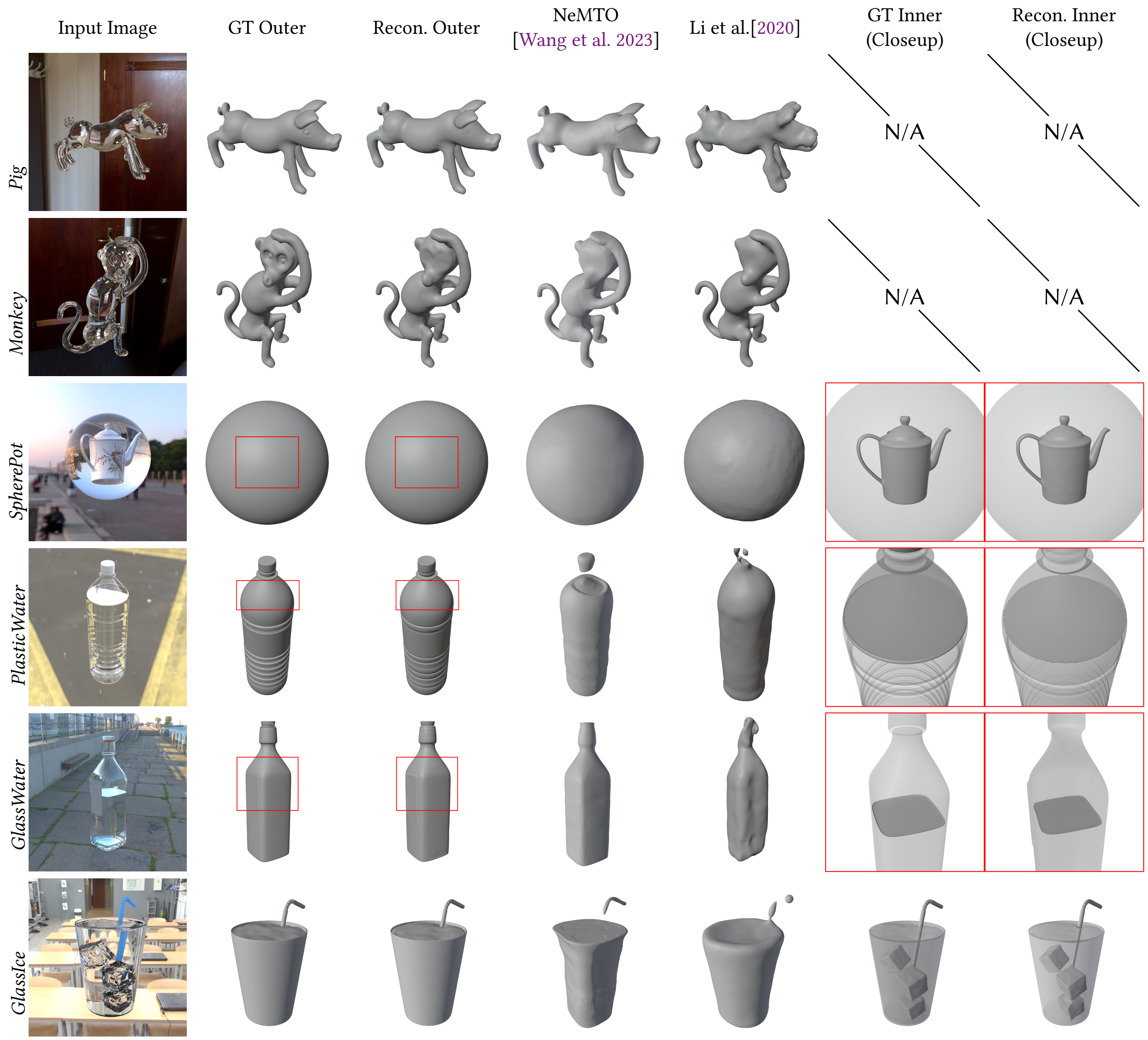


Fig. 7. **Reconstruction and Rendering results on synthetic scenes.** For each scene, we show the input image, GT inner/outer shapes, reconstructed inner/outer shapes and reconstruction results of NeMTO [Wang et al. 2023] and Li et al. [2020].

et al. 2023] struggles when the background is not "infinitely far" or the input lighting estimation is not accurate. Li et al. [2020] tend to "over smooth" the geometry because it includes a geometry smoothing step. The quantitative comparisons between our method and baseline methods are shown in Table 2. The better geometry reconstruction quality also reflects on the metrics and our method comes to the top.

5.4 Comparison with Baselines Requiring Controlled Capture Environment.

In this section, we choose to compare the proposed NU-NeRF pipeline with other methods that target transparent object reconstruction but requires a controlled capture environment. Namely, we choose two baselines: Gao et al. [2023] and ReNeuS [Tong et al. 2023]. Gao et al. [2023] aim to reconstruct transparent objects with no additional inputs other than images, but they need the object to be placed on a sufficiently large opaque plane, and cannot support nested objects. ReNeuS [Tong et al. 2023], on the other hand, is only dedicated to reconstructing opaque objects inside transparent objects with known outer surface geometry, captured in a homogeneous light.

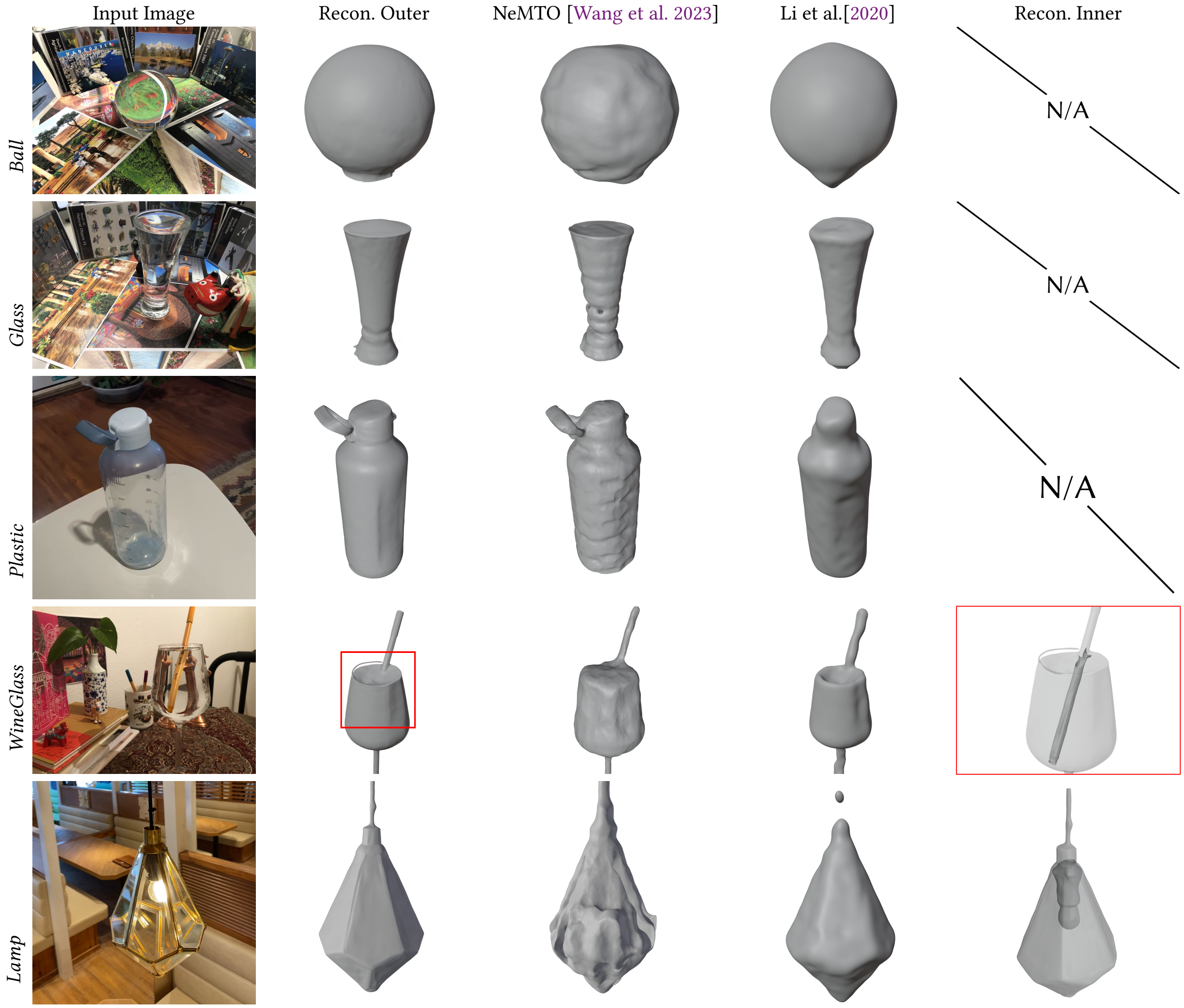


Fig. 8. **Reconstruction and Rendering results on real scenes.** For each scene, we show the input image, reconstructed inner/outer shapes, and reconstruction results of NeMTO [Wang et al. 2023] and Li et al. [2020].

Gao et al. [2023]. We compare our method with Gao et al. [2023] on two different scenes: the first is *PigPlane*, which simply adds a plane needed by Gao et al. [2023] into the *Pig* dataset. The second is *Ball* dataset from Bemana et al. [2022], which places the object on a relatively small plane. Since the code of their method is not released publicly at the time we do our experiments, we choose to implement the method ourselves based on the code of NeuS [Wang et al. 2021]. As shown in the first row of Fig. 10, Gao et al. [2023] can reconstruct reasonable geometry when the plane is large enough but may miss geometry details (the feet) due to its less accurate silhouette estimation. When the plane is not sufficiently large as shown in the second row, the inaccurate plane parameter estimation causes much worse results since rendered colors are determined by intersecting the bent rays and the estimated plane.

ReNeuS [Tong et al. 2023]. We compare our method with ReNeuS on two different scenes: the first is *GlassIce*, and the second is *BallStatue*. When running the method, we provide the ground truth shape of the outer geometry. Again, we choose to implement the method ourselves since the code is also unavailable. As shown in Fig. 11, ReNeuS fails to deal with more complex visual effects like the transparent inner geometry and the reflective outer geometry, which causes incorrect geometry reconstruction like the collapsed surface on the ice and the top of the statue. On the contrary, we model more complex lighting interactions including reflection and refraction, leading to more faithful geometry reconstruction.

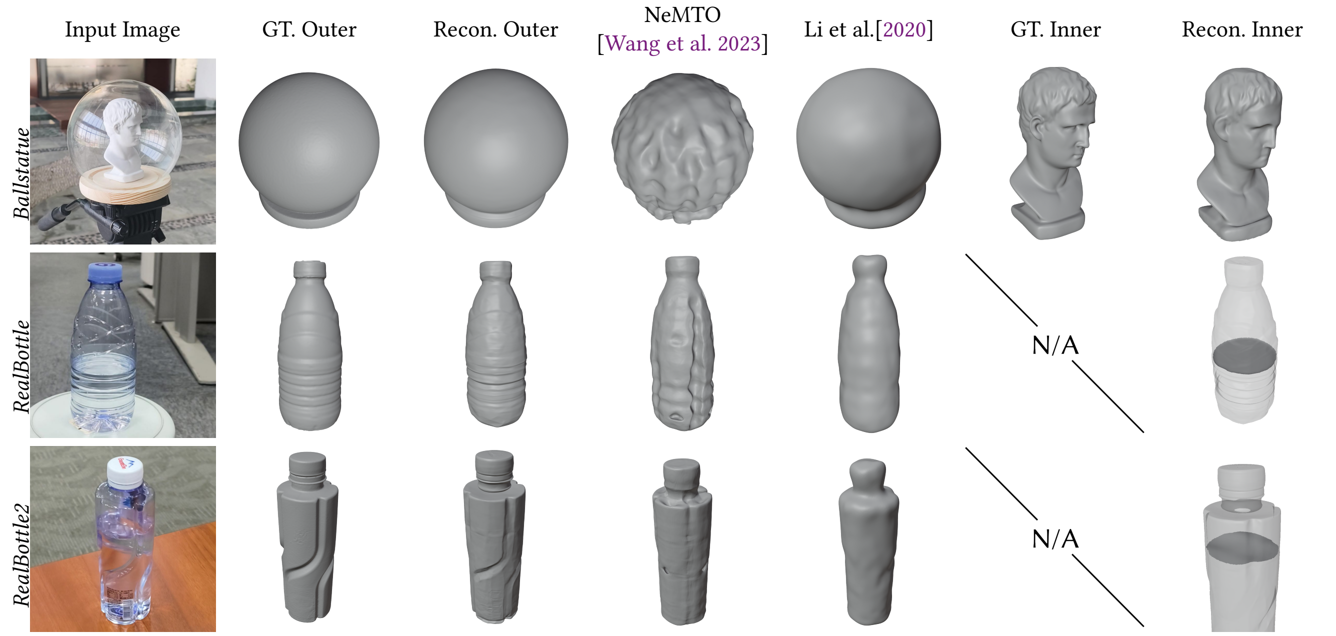


Fig. 9. **Reconstruction and Rendering results on real scenes with ground truth geometry captured by ourselves.** For each scene, we show the input image, GT inner/outer shapes, reconstructed inner/outer shapes, and reconstruction results of NeMTO [Wang et al. 2023] and Li et al. [2020].

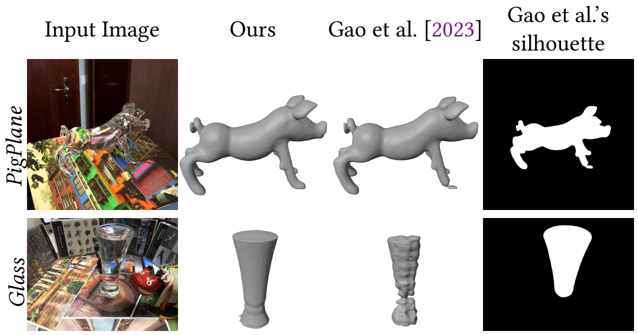


Fig. 10. **Qualitative comparison of our method and Gao et al. [2023].**

Table 3. Ablation studies of reconstruction results of outer interface using the chamfer distance metric ($\times 10^{-4}$) and Earth Mover's Distance metric ($\times 10^{-3}$) on the synthetic dataset. "Fail" means this ablated version produces no geometry, and is unable to calculate the corresponding metric.

Scenes	Full		w/o g_r		Only g_r		w/o ℓ_{ilc}	
	CD ↓	EMD ↓	CD ↓	EMD ↓	CD ↓	EMD ↓	CD ↓	EMD ↓
Pig	**0.73**	**6.31**	1727.81	362.80	1.25	12.3	1.16	8.49
Monkey	**1.02**	**4.80**	59.23	68.14	Fail	Fail	1.38	24.55
SpherePot	**0.99**	**3.75**	Fail	Fail	1969.9	285.9	2.31	4.98
PlasticWater	**1.67**	**55.01**	80.52	70.30	4.75	102.3	1.83	64.7
GlassWater	**0.61**	**14.82**	Fail	Fail	1.36	18.6	0.74	15.53

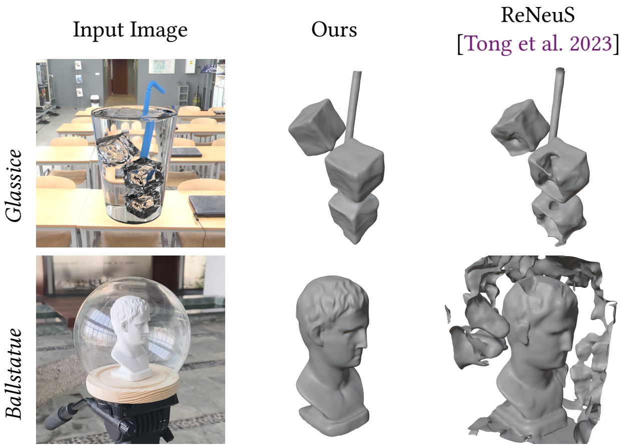


Fig. 11. **Qualitative comparison of our method and ReNeuS [Tong et al. 2023].**

Table 4. Ablation studies of reconstruction results of inner interface using the chamfer distance metric ($\times 10^{-4}$) and Earth Mover's Distance metric ($\times 10^{-3}$) on the synthetic dataset.

Scenes	Full		w/o two IORs		w/o non-zero	
	CD ↓	EMD ↓	CD ↓	EMD ↓	CD ↓	EMD ↓
PlasticWater	**1.64**	**0.92**	45.35	10.26	N/A	N/A
GlassWater	**1.11**	**1.52**	984.21	114.92	941.11	110.97

5.5 Ablation Studies

NU-NeRF consists of two stages, each stage contains multiple design choices. To test the effectiveness of its design, we remove some of

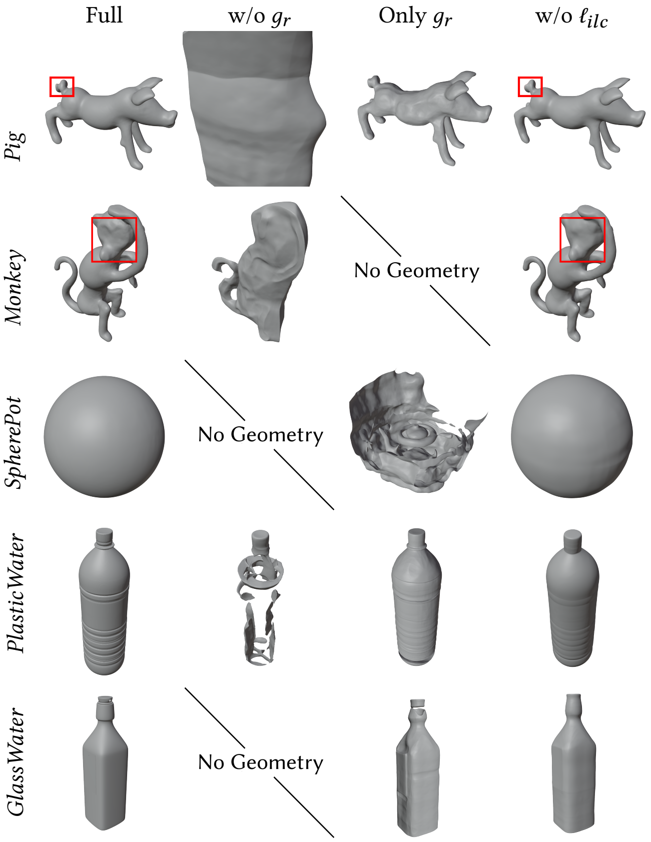


Fig. 12. Qualitative comparison of reconstruction results of outer surface on synthetic datasets between the full pipeline and two baselines: without g_r, without ℓ_{ilc}. For *Pig* and *Monkey* datasets, please zoom in at the highlighted area to see the better details learned by our method than the without ℓ_{ilc} ablation.

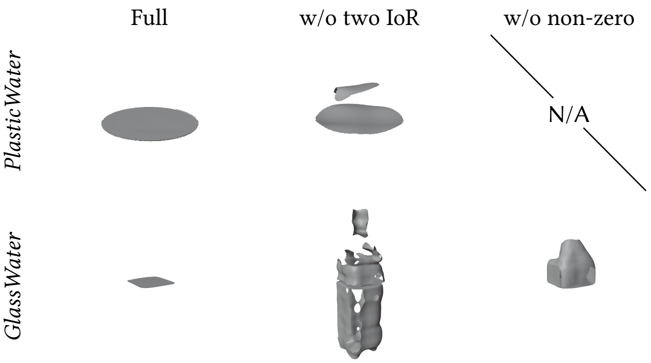


Fig. 13. Qualitative comparison of reconstruction results of inner surface on synthetic datasets between the full pipeline and two ablations: without two IoR formulation, and without non-zero thickness formulation.

the components of the full pipeline and compare these versions against the full version, these ablations include:

- **Without g_r.** We start with removing the g_r network. g_r serves the purpose of modeling refraction light, so this version is NeRO [Liu et al. 2023] plus the loss ℓ_{ilc}.
- **Only g_r.** g_r itself is a NeRF color network. To validate that our method does not solely rely on g_r to learn both the reflection and refraction colors, we remove the reflection part g_l and only use g_r as an ablation.
- **Without ℓ_{ilc}.** ℓ_{ilc} is added into our pipeline to increase the consistency between the learned lighting g_{inf} and the outer NeRF.
- **Without two refraction indices η_r, η_l (i.e. solid assumption).** We model the interface and the inner substance with two IoRs to deal with nested surfaces. We remove the two-IoR formulation and use a single-layer interface with spatially varying IoR to model the input scene. This is evaluated on the *PlasticWater* and *GlassWater* cases.
- **Without non-zero thickness formulation.** In this ablation, we remove the non-zero thickness formulation, with only the zero thickness interface applied. This version is evaluated on the *GlassWater* case.

The ablation studies results are shown in Fig. 12, Fig. 13, Table 3, and Table 4. It can be observed in Fig. 12 that the method can not produce meaningful results without g_r since the refraction color is not modeled. The results produced by only g_r in Fig. 12 are also worse since the reflection part is ignored. And if ℓ_{ilc} is not applied, the geometry details will be lost as shown in Fig. 12. In addition, as illustrated in Fig.13, if the two-IoR formulation is not introduced, the network will learn an "average" version of the IoRs of the interface and inner material, and the inner SDF will produce superfluous geometry to compensate for the inaccurate refraction. Finally, if no non-zero thickness formulation is used, superfluous geometry will appear at the top area of *Glasswater* case. This is because the thickness of the interface causes the light to bend in this area. If the zero-thickness assumption is used, the network fails to predict this type of bending and distortion, which causes the compensation of inner SDF.

6 CONCLUSION

In this paper, we propose NU-NeRF for the geometry reconstruction of nested transparent objects under an uncontrolled capture environment that overcomes the drawbacks of current transparent object reconstruction methods including only applying to solid objects and having extra requirements for inputs and the capture environment. To eliminate the need for any capture environment and additional inputs, we incorporate the neural implicit representation and use the Signed Distance Field to enable surface reconstruction. We model the interface using physically correct BSDF defined with Cook-Torrance reflectance and transmission. We additionally leverage the split sum approximation to make efficient rendering plausible. To model the refraction, we introduce a simple yet effective single MLP into the pipeline to predict the refraction color. A novel incident light consistency loss is added to improve the reconstruction fidelity of the outer surface. Furthermore, our method enables nested object reconstruction by using ray-traced iterative reconstruction. Learnable IoRs on the outer surface and the inner substance are tuned and ray tracing-based rendering with light path explicitly modeled is introduced to enable geometry reconstruction of the inner surface. We evaluate our method on both synthetic and real-captured datasets, where our method outperforms current methods targeting the geometry reconstruction problem of transparent objects. Nevertheless, our method still has the following limitations: Firstly, our method does not model complex optics effects like total internal reflection. Secondly, although theoretically plausible, our method cannot handle more than two layers of surfaces now. For future directions, we would like to improve the quality of reconstruction by taking the actual light transport into account in both the first and second stage and extend the scope to more complex geometry with three layers or more.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (No.62322210), the Innovation Funding of ICT, CAS (No.E461020), Beijing Municipal Natural Science Foundation for Distinguished Young Scholars (No.JQ21013), and Beijing Municipal Science and Technology Commission (No.Z231100005923031). Ling-Qi Yan is also supported by gift funds from Adobe, Intel, Lintex, Meta and XVerse. We also thank KIRI Innovations for their support in data capture.

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 *ICCV*. 5835–5844.

Harry G. Barrow, Jay M. Tenenbaum, Robert C. Bolles, and Helen C. Wolf. 1977. Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Matching. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence. Cambridge, MA, USA, August 22-25, 1977*, Raj Reddy (Ed.). William Kaufmann, 659–663.

Mojtaba Bemana, Karol Myszkowski, Jeppe Revall Frisvad, Hans-Peter Seidel, and Tobias Ritschel. 2022. Eikonal Fields for Refractive Novel-View Synthesis. In *SIGGRAPH '22: Special Interest Group on Computer Graphics and Interactive Techniques Conference, Vancouver, BC, Canada, August 7 - 11, 2022*, Munkhtsetseg Nandigjav, Niloy J. Mitra, and Aaron Hertzmann (Eds.). ACM, 39:1–39:9.

Mark Boss, Raphael Braun, Varun Jampani, Jonathan T Barron, Ce Liu, and Hendrik Lensch. 2021. NeRD: Neural reflectance decomposition from image collections. In *ICCV*. 12684–12694.

Brent Burley. 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, Tero Karras, and Gordon Wetzstein. 2022. Efficient Geometry-aware 3D Generative Adversarial Networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition,*. IEEE, 16102–16112.

Pierre Charbonnier, Laure Blanc-Feraud, Gilles Aubert, and Michel Barlaud. 1994. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proceedings of 1st international conference on image processing*, Vol. 2. IEEE, 168–172.

Xiaoxue Chen, Junchen Liu, Hao Zhao, Guyue Zhou, and Ya-Qin Zhang. 2023. NeRRF: 3D Reconstruction and View Synthesis for Transparent and Specular Objects with Neural Refractive-Reflective Fields. *CoRR* abs/2309.13039 (2023).

MMSegmentation Contributors. 2020. MMSegmentation: OpenMMLab Semantic Segmentation Toolbox and Benchmark. https://github.com/open-mmlab/mmsegmentation.

Robert L. Cook and Kenneth E. Torrance. 1982. A Reflectance Model for Computer Graphics. *ACM Trans. Graph.* 1, 1 (1982), 7–24.

Weijian Deng, Dylan Campbell, Chunyi Sun, Shubham Kanitkar, Matthew Shaffer, and Stephen Gould. 2024. Ray Deformation Networks for Novel View Synthesis of Refractive Objects. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 3118–3128.

Fangzhou Gao, Lianghao Zhang, Li Wang, Jiamin Cheng, and Jiawan Zhang. 2023. Transparent Object Reconstruction via Implicit Differentiable Refraction Rendering. In *SIGGRAPH Asia 2023 Conference Papers, SA 2023, Sydney, NSW, Australia, December 12-15, 2023*, June Kim, Ming C. Lin, and Bernd Bickel (Eds.). ACM, 57:1–57:11.

Cong Phuoc Huynh, Antonio Robles-Kelly, and Edwin R. Hancock. 2010. Shape and refractive index recovery from single-view polarisation images. In *The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 June 2010*. IEEE Computer Society, 1229–1236.

Ivo Ihrke, Kiriakos N. Kutulakos, Hendrik P. A. Lensch, Marcus A. Magnor, and Wolfgang Heidrich. 2010. Transparent and Specular Object Reconstruction. *Comput. Graph. Forum* 29, 8, 2400–2426.

James T. Kajiya. 1986. The Rendering Equation *(SIGGRAPH '86)*. 143–150.

Brian Karis. 2013. Real shading in unreal engine 4. *Proc. Physically Based Shading Theory Practice* 4, 3 (2013), 1.

Zongcheng Li, Xiaoxiao Long, Yusen Wang, Tuo Cao, Wenping Wang, Fei Luo, and Chunxia Xiao. 2023a. NeTO: Neural Reconstruction of Transparent Objects with Self-Occlusion Aware Refraction-Tracing. *CoRR* abs/2303.11219 (2023).

Zhaoshuo Li, Thomas Müller, Alex Evans, Russell H Taylor, Mathias Unberath, Ming-Yu Liu, and Chen-Hsuan Lin. 2023b. Neuralangelo: High-Fidelity Neural Surface Reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhengqin Li, Yu-Ying Yeh, and Manmohan Chandraker. 2020. Through the Looking Glass: Neural 3D Reconstruction of Transparent Shapes. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*. Computer Vision Foundation / IEEE, 1259–1268.

Arvin Lin, Yiming Lin, and Abhijeet Ghosh. 2023. Practical Acquisition of Shape and Plausible Appearance of Reflective and Translucent Objects. In *Computer Graphics Forum*, Vol. 42. Wiley Online Library, e14889.

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. In *SIGGRAPH*.

Jiahui Lyu, Bojian Wu, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. 2020. Differentiable refraction-tracing for mesh reconstruction of transparent objects. *ACM Trans. Graph.* 39, 6 (2020), 195:1–195:13.

Stephen McAuley, Stephen Hill, Naty Hoffman, Yoshiharu Gotanda, Brian Smits, Brent Burley, and Adam Martinez. 2012. Practical physically-based shading in film and game production. In *ACM SIGGRAPH 2012 Courses* (Los Angeles, California) *(SIGGRAPH '12)*. Association for Computing Machinery, New York, NY, USA, Article 10, 7 pages. https://doi.org/10.1145/2343483.2343493

Stephen McAuley, Stephen Hill, Adam Martinez, Ryusuke Villemin, Matt Pettineo, Dimitar Lazarov, David Neubelt, Brian Karis, Christophe Hery, Naty Hoffman, and Hakan Zap Andersson. 2013. Physically based shading in theory and practice. In *ACM SIGGRAPH 2013 Courses* (Anaheim, California) *(SIGGRAPH '13)*. Association for Computing Machinery, New York, NY, USA, Article 22, 8 pages. https://doi.org/10.1145/2504435.2504457

Mark Meyer, Mathieu Desbrun, Peter Schröder, and Alan H. Barr. 2002. Discrete Differential-Geometry Operators for Triangulated 2-Manifolds. In *Third International Workshop "Visualization and Mathematics", VisMath 2002, Berlin, Germany, May 22-25, 2002 (Mathematics and Visualization)*, Hans-Christian Hege and Konrad Polthier (Eds.). Springer, 35–57.

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.

Daisuke Miyazaki and Katsushi Ikeuchi. 2005. Inverse Polarization Raytracing: Estimating Surface Shapes of Transparent Objects. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20-26 June 2005, San Diego, CA, USA*. IEEE Computer Society, 910–917.

Nigel J. W. Morris and Kiriakos N. Kutulakos. 2011. Dynamic Refraction Stereo. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 8 (2011), 1518–1531.

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.

Michael Niemeyer and Andreas Geiger. 2021. GIRAFFE: Representing Scenes As Compositional Generative Neural Feature Fields. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11453–11464.

Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. 2021a. Nerfies: Deformable neural radiance fields. In *IEEE/CVF International Conference on Computer Vision*. 5865–5874.

Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M Seitz. 2021b. HyperNeRF: a higher-dimensional representation for topologically varying neural radiance fields. *ACM Transactions on Graphics* 40, 6 (2021), 1–12.

Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. 2021. D-NeRF: Neural radiance fields for dynamic scenes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10318–10327.

Yiming Qian, Minglun Gong, and Yee-Hong Yang. 2016. 3D Reconstruction of Transparent Objects with Position-Normal Consistency. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*. IEEE Computer Society, 4369–4377.

Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. 2000. The Earth Mover's Distance as a Metric for Image Retrieval. *Int. J. Comput. Vis.* 40, 2 (2000), 99–121.

Christophe Schlick. 1994. An Inexpensive BRDF Model for Physically-Based Rendering. *Comput. Graph. Forum* 13, 3 (1994), 233–246.

Katja Schwarz, Yiyi Liao, Michael Niemeyer, and Andreas Geiger. 2020. GRAF: Generative Radiance Fields for 3D-Aware Image Synthesis. In *Advances in Neural Information Processing Systems.*

Mingqi Shao, Chongkun Xia, Dongxu Duan, and Xueqian Wang. 2022. Polarimetric Inverse Rendering for Transparent Shapes Reconstruction. *CoRR* abs/2208.11836 (2022).

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.*

Jinguang Tong, Sundaram Muthu, Fahira Afzal Maken, Chuong Nguyen, and Hongdong Li. 2023. Seeing Through the Glass: Neural 3D Reconstruction of Object Inside a Transparent Container. In *CVPR*. IEEE, 12555–12564.

Edgar Tretschk, Ayush Tewari, Vladislav Golyanik, Michael Zollhofer, Christoph Lassner, and Christian Theobalt. 2021. Non-Rigid Neural Radiance Fields: Reconstruction and Novel View Synthesis of a Dynamic Scene From Monocular Video. In *IEEE/CVF International Conference on Computer Vision.* 12959–12970.

Borislav Trifonov, Derek Bradley, and Wolfgang Heidrich. 2006. Tomographic Reconstruction of Transparent Objects. In *Proceedings of the Eurographics Symposium on Rendering Techniques, Nicosia, Cyprus, 2006*, Tomas Akenine-Möller and Wolfgang Heidrich (Eds.). Eurographics Association, 51–60.

Dor Verbin, Peter Hedman, Ben Mildenhall, Todd E. Zickler, Jonathan T. Barron, and Pratul P. Srinivasan. 2022. Ref-NeRF: Structured View-Dependent Appearance for Neural Radiance Fields. In *CVPR*. 5481–5490.

Dongqing Wang, Tong Zhang, and Sabine Süsstrunk. 2023. NEMTO: Neural Environment Matting for Novel View and Relighting Synthesis of Transparent Objects. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).*

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.

Gordon Wetzstein, David Roodnick, Wolfgang Heidrich, and Ramesh Raskar. 2011. Refractive shape from light field distortion. In *IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November 6-13, 2011*, Dimitris N. Metaxas, Long Quan, Alberto Sanfeliu, and Luc Van Gool (Eds.). IEEE Computer Society, 1180–1186.

Bojian Wu, Yang Zhou, Yiming Qian, Minglun Gong, and Hui Huang. 2018. Full 3D reconstruction of transparent objects. *ACM Trans. Graph.* 37, 4 (2018), 103.

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.

Yun-Peng Xiao, Yu-Kun Lai, Fang-Lue Zhang, Chunpeng Li, and Lin Gao. 2020. A survey on deep geometry learning: From a representation perspective. *Comput. Vis. Media* 6, 2 (2020), 113–133.

Jiamin Xu, Zihan Zhu, Hujun Bao, and Weiwei Xu. 2022. A Hybrid Mesh-neural Representation for 3D Transparent Object Reconstruction. *CoRR* abs/2203.12613 (2022).

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

Yifan Zhan, Shohei Nobuhara, Ko Nishino, and Yinqiang Zheng. 2023. Nerfrac: Neural radiance fields through refractive surface. In *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 18402–18412.

Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. 2021. 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.*