

PBIR-NIE: Glossy Object Capture under Non-Distant Lighting

GUANGYAN CAI, University of California, Irvine, USA and Adobe Research, USA

FUJUN LUAN, Adobe Research, USA

MILOŠ HAŠAN, Adobe Research, USA

KAI ZHANG, Adobe Research, USA

SAI BI, Adobe Research, USA

ZEXIANG XU, Adobe Research, USA

ILIJAN GEORGIEV, Adobe Research, UK

SHUANG ZHAO, University of California, Irvine, USA



Fig. 1. We propose PBIR-NIE, a physics-based inverse rendering pipeline that optimizes an object’s shape, glossy surface reflectance, and non-distant lighting representation. Our method faithfully recovers the shiny and specular appearance, produces relighting results with high fidelity, and accurately captures geometric details from a rough visual hull initialization.

Glossy objects present a significant challenge for 3D reconstruction from multi-view input images under natural lighting. In this paper, we introduce PBIR-NIE, an inverse rendering framework designed to holistically capture the geometry, material attributes, and surrounding illumination of such objects. We propose a novel parallax-aware non-distant environment map as a lightweight and efficient lighting representation, accurately modeling the near-field background of the scene, which is commonly encountered in real-world capture setups. This feature allows our framework to accommodate complex parallax effects beyond the capabilities of standard infinite-distance environment maps. Our method optimizes an underlying signed distance field (SDF) through physics-based differentiable rendering, seamlessly connecting surface gradients between a triangle mesh and the SDF via neural implicit evolution (NIE). To address the intricacies of highly glossy BRDFs in differentiable rendering, we integrate the antithetic sampling algorithm to mitigate variance in the Monte Carlo gradient estimator. Consequently, our framework exhibits robust capabilities in handling glossy object reconstruction, showcasing superior quality in geometry, relighting, and material estimation.

1 INTRODUCTION

The joint reconstruction of an object’s surface geometry, material reflectance, and its surrounding illumination, commonly referred to as inverse rendering, stands as a foundational task in computer vision and graphics for 3D content creation, with applications across various fields: film making, game production, product design, and AR/VR. Despite its broad applicability, the reconstruction of non-diffuse objects remains particularly challenging due to the presence of specular reflections. These reflective surfaces introduce inconsistencies in color when captured from multiple viewpoints, creating difficulties for multi-view reconstruction methods (such as NeRF [Mildenhall et al. 2020] and 3D Gaussian Splatting [Kerbl et al. 2023]) that heavily rely on view consistency and feature matching.

Recovering the appearance of glossy objects presents challenges due to the high-frequency reflections of complex surrounding environments. Previous inverse rendering methods often rely on an *infinitely distant* 2D environment map [Munkberg et al. 2022], either in a discretized form or parameterized using mixtures of spherical Gaussians [Zhang et al. 2021b; Jin et al. 2023] for representing the

background. However, the infinite-distance assumption is easily violated in real-world capture setups. One solution for capturing complex background parallax effects is to represent the lighting as a neural radiance field (NeRF) [Mildenhall et al. 2020], as demonstrated by prior works [Ling et al. 2024; Wang et al. 2024; Zhuang et al. 2023]. Utilizing NeRF as a background can however be computationally expensive, and the absence of an efficient importance sampling method further complicates its practical use in inverse rendering applications. To address this, we propose a lightweight background representation, ENVMAP++, designed to efficiently model both near-field and far-field background illumination with a parallax-aware environment map. This representation strikes a balance between quality and performance, as demonstrated in our glossy object reconstruction task.

Prior inverse rendering works [Luan et al. 2021; Cai et al. 2022; Sun et al. 2023] rely on good initialization of the surface topology (e.g. via neural rendering methods such as NeuS [Wang et al. 2021]), and further refine the geometric details with a differentiable renderer. On the other hand, neural implicit evolution (NIE) [Mehta et al. 2022] formulates a level-set evolution method for parametrically defined implicit surfaces that does not require mesh extraction to be differentiable. We show that this method can be combined with a physics-based differentiable renderer and allows topology change during optimization. We integrate NIE into our framework and optimize an underlying neural signed distance field (SDF). This integration allows our method to handle objects with complex topology, including thin structures, holes, and other subtle geometric features; thus, it is robust to poor initialization. Our framework, dubbed PBIR-NIE, not only captures the geometry, material properties, and surrounding illumination of glossy objects but also handles topology changes seamlessly.

Lastly, when dealing with highly specular BRDFs (i.e. with very low roughness) in physics-based differentiable rendering, the naively estimated geometric gradients can be very noisy. Zhang et al. [2021a] demonstrate that antithetic sampling can significantly reduce the variance in such cases. However, despite enabling faster convergence, this method can be tedious to implement, especially for indirect illumination where the number of light paths grows exponentially. We propose a simple yet efficient modified variant of antithetic sampling, incorporating Russian roulette and roughness regularization beyond secondary bounces of each light path. This variant is easy to implement and we show it is robust for glossy object reconstruction.

In summary, our contributions include:

- ENVMAP++, a lightweight non-distant lighting representation that efficiently models both the near- and far-field background illumination with a parallax-aware environment map, overcoming limitations of infinitely distant and NeRF emitters.
- Integration of neural implicit evolution into PBIR, allowing for topological changes and enhancing the framework’s ability to handle complex object geometries.
- An efficient antithetic sampling variant that improves the handling of highly glossy BRDFs in differentiable rendering with gradient variance reduction for fast and accurate reconstruction.

- An end-to-end pipeline achieving state-of-the-art geometry, material, and lighting estimation for glossy-object reconstruction, enabling realistic view synthesis and relighting.

2 RELATED WORK

2.1 Neural Surface Reconstruction

Neural rendering, particularly neural implicit representation, has seen significant advancements in 3D reconstruction in recent years. Neural radiance fields (NeRF) [Mildenhall et al. 2020] and its variants [Müller et al. 2022; Chen et al. 2022] utilize volume rendering on scene representations with neural density fields and view-dependent color fields, resulting in impressive photorealism in novel view synthesis. However, geometry extracted from volumetric density fields often exhibits flawed surfaces.

Alternatively, representing the underlying geometry with a signed distance field (SDF) has shown promise for improved surface reconstruction in recent neural SDF-based approaches such as NeuS [Wang et al. 2021], VolSDF [Yariv et al. 2021], and PermutoSDF [Rosu and Behnke 2023a]. Unfortunately, when capturing glossy objects with shiny and metallic materials (such as a soda can, a stainless kettle, or a polished silver spoon), both density-based and SDF-based approaches struggle to faithfully reconstruct the geometry.

Cai et al. [2022] leverage MeshSDF [Remelli et al. 2020], a differentiable version of Marching Cubes, to implicitly optimize an SDF by rendering the extracted mesh. However, this method is computationally expensive. Conversely, Vicini et al. [2022] and Bangaru et al. [2022] reparameterize the discontinuities in direct SDF rendering to avoid meshes entirely, but this approach complicates extending the method to multiple bounces and implementing variance reduction techniques.

2.2 Glossy Surface Reconstruction

Recently, glossy surface reconstruction has received increased attention in the neural and inverse rendering community. Ref-NeRF [Verbin et al. 2022] introduced integrated directional encoding, replacing NeRF’s view-dependent color field with a representation of reflected radiance based on surface normals, which improved the reconstructed surface quality. This approach was extended to SDF-based frameworks in Ref-NeuS [Ge et al. 2023]. SpecNeRF [Ma et al. 2023] proposed 3D Gaussian-based encoding to enhance NeRF’s reflection modeling capabilities. Neural directional encoding [Wu et al. 2024] transfers feature-grid-based spatial encoding into the angular domain and considers near-field specular interreflections with cone tracing, further improving the modeling of complex reflections. Neural plenoptic function [Wang et al. 2024] proposed to represent global illumination via a 5D representation based on NeRFs and raytracing. Our most relevant baseline is NeRO [Liu et al. 2023], which introduced a two-stage, NeuS-based pipeline that explicitly incorporates the rendering equation into the neural reconstruction framework, demonstrating superior geometry quality on reflective objects. However, their pipeline relies on approximation to ensure the computation is tractable, for instance, using neural networks to predict occlusion and indirect illumination. While our work shares a similar goal, we primarily approach the problem by solving the rendering equation without approximation.

2.3 Material and Lighting Estimation

Beyond surface geometry, inverse rendering [Marschner 1998; Ramamoorthi and Hanrahan 2001] typically also involves estimating the material properties (e.g., SVBRDF) and, in some cases, the surrounding illumination (depending on the capture setup). Traditional data-driven acquisition methods [Xia et al. 2016; Dong et al. 2014, 2010; Aittala et al. 2013; Nam et al. 2018, 2016; Zhou et al. 2016] often assume sparsity in spatially-varying surface reflectance and frame the capture as a complex non-linear optimization problem. Similarly, recent neural reconstruction methods address this analysis-by-synthesis problem through differentiable rendering for intrinsic decomposition. They employ various differentiable rendering techniques, including neural renderers (PhysSG [Zhang et al. 2021b], NeRFactor [Zhang et al. 2021c], TensorIR [Jin et al. 2023] and others [Boss et al. 2021a,b; Zhang et al. 2022a,b; Kuang et al. 2022; Srinivasan et al. 2021]), fast differentiable rasterizers [Munkberg et al. 2022] or differentiable Monte Carlo raytracers [Luan et al. 2021; Cai et al. 2022; Sun et al. 2023; Hasselgren et al. 2022]. Our approach operates within the differentiable path tracing framework, parameterizing spatially-varying surface reflectance with an analytic microfacet BRDF model using the GGX distribution [Walter et al. 2007a]. However, unlike previous methods that typically represent lighting with a 2D distant environment map, we also consider near-field background illumination, crucial for glossy object reconstruction. Similar to Ling et al. [2024] and NeRO [Liu et al. 2023] that represent the background illumination with non-distant environment emitters (either a NeRF or two direct-indirect separable lighting MLPs), we introduce a lightweight, parallax-aware environment map representation. This approach demonstrates robustness to near-field and far-field lighting conditions while remaining efficient in inverse rendering optimization.

3 PRELIMINARIES

In this section, we briefly revisit mathematical and algorithmic preliminaries related to physics-based differentiable rendering [Zhao et al. 2020], and discuss the advantages and disadvantages of the differentiable renderer we use and how it affects our pipeline.

Given a virtual object described by parameters ξ , we render the image with Monte Carlo rendering based on the path integral formulation introduced by Veach [1997]:

$$I = \int_{\Omega} f(\bar{x}) d\mu(\bar{x}), \quad (1)$$

where $\Omega := \cup_{N \geq 1} \mathcal{M}^{N+1}$ is the path space consisting of light transport paths $\bar{x} = (x_0, x_1, \dots, x_N)$ with \mathcal{M} being the union of all object surfaces, f is the measurement contribution function, and μ is the corresponding area-product measure.

Computing image gradients involves differentiating pixel intensities in Eq. (1) with respect to ξ , which is not a trivial process due to the existence of discontinuities in the integrand. Zhang et al. [2020a] and Bangaru et al. [2020] presented two different paradigms for tackling this problem: one directly track the discontinuities and one eliminates the discontinuities via reparameterization. Both can accurately differentiate Eq. (1).

Given estimation of the gradient $\frac{dI}{d\xi}$, we are able to reconstruct scene parameters from images using an optimization approach: render images using the initial scene parameters, compute the loss with regard to the target images, obtain the gradient of the loss with respect to the scene parameters, update the scene parameters using gradient descent and repeat until convergence.

4 OUR METHOD

Our **PBIR-NIE** pipeline reconstructs the shape, material, and background lighting of opaque objects from multi-view images with known camera poses. While it can handle diffuse objects, it is specifically designed to perform well on glossy ones, which are more sensitive to light reconstruction quality due to reflected background details. Previous methods often use an environment map to represent background illumination, assuming light originates from infinitely far away. However, this results in blurry reconstructions under indoor lighting, often compensated by increased roughness. To address this, we introduce a new model **ENVMAP++** (Sec. 4.1), a lightweight non-distant lighting representation that is suitable for both near-field and far-field lighting conditions. Additionally, optimizing the shape and material of glossy objects requires special considerations; for shape optimization, we employ neural implicit evolution with careful initialization to ensure robust geometry handling (Sec. 4.2); we introduce a modified version of antithetic sampling [Zhang et al. 2021a] for variance reduction in the Monte Carlo gradient estimation (Sec. 4.3).

The differentiable renderer is a crucial component of our pipeline. We have selected Mitsuba 3 [Jakob et al. 2022b] due to its implementation of numerous state-of-the-art techniques pertinent to our application. However, one limitation constrains our choice of scene primitives: although neural networks serve as effective representations for primitives, evaluating them within the rendering loop is suboptimal. Mitsuba 3 supports two evaluation modes—megakernel and wavefront—enabled by its auto-differentiation engine, Dr. Jit [Jakob et al. 2022a]. While the megakernel mode is significantly more efficient than the wavefront mode for Monte Carlo rendering, it currently does not support the evaluation of neural networks within the kernel. Conversely, the wavefront mode allows for such evaluations but incurs substantial performance and memory costs. To facilitate practical applications, we opt for the megakernel mode, trading some flexibility for efficiency by evaluating all neural networks prior to the rendering step. This trade-off plays an important role in our pipeline design.

4.1 ENVMAP++: Non-Distant Environment Map

We propose representing the background using a deformed spherical emitter, optimizing its vertices to simulate the parallax effect. Assume (without loss of generality) the object and cameras are bounded by a unit sphere \mathcal{S} . The emitter consists of a mesh \mathcal{M}_{env} deformed from the starting sphere S . Each vertex x^M of \mathcal{M}_{env} is represented as $x - n(x)d(x)$, where x and $n(x)$ are the corresponding vertex and vertex normal on \mathcal{S} , and $d(x) > 0$ is the displacement amount to be optimized. Furthermore, $L_e(x)$ is the optimized uniformly emitted radiance of each vertex. The normals point inside the sphere to ensure visibility. To handle the wide range of d , we

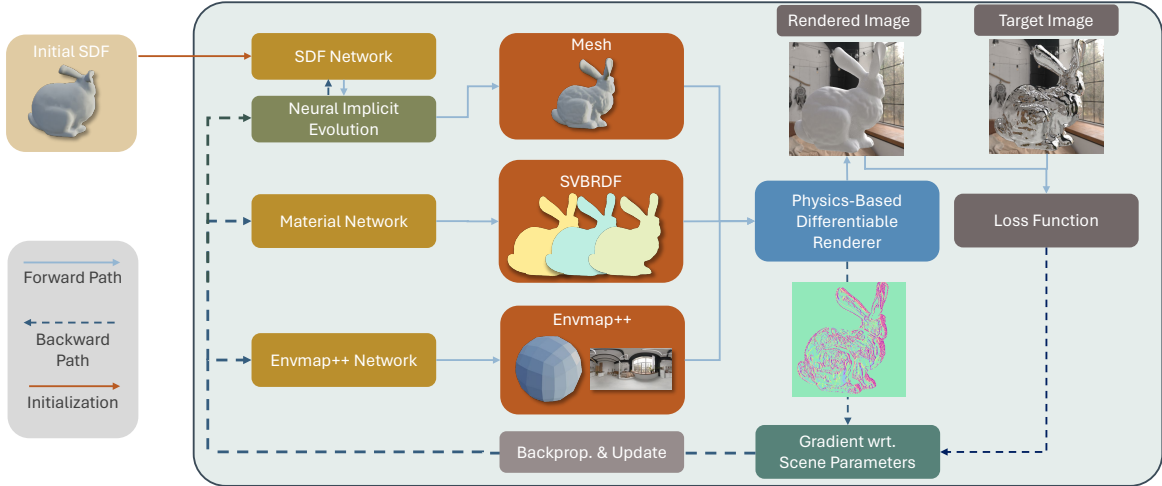


Fig. 2. **Overview of our PBIR-NIE pipeline.** Our pipeline takes a set of multi-view images capturing a glossy object and an initial shape as input. It then reconstructs the scene’s geometry, material properties, and lighting using a physics-based inverse rendering (PBIR) approach. The iterative refinement process includes: 1) **Forward Pass:** Rendering an image by employing physics-based differentiable rendering. This involves using an explicit mesh extracted with a non-differentiable Marching Cubes algorithm to represent the neural implicit surface for shape, and material networks for surface properties, while leveraging information from input training views. Additionally, ENVMAP++ is utilized for enhanced lighting representation, replacing the standard infinite-distance environment map to handle non-distant background illumination. 2) **Backward Pass:** Comparing the rendered image to the ground truth and computing gradients with respect to scene parameters. We use neural implicit evolution (NIE) [Mehta et al. 2022] to facilitate the backpropagation of gradients from the extracted mesh to the neural implicit surface, bypassing the non-differentiable extraction step. 3) **Update:** Adjusting scene parameters (geometry, material, lighting) via backpropagation to minimize the difference between the rendered and ground truth image.

adopt the *inverted sphere parameterization* from NeRF++ [Zhang et al. 2020b] and optimize $r(x) \in (0, 1)$ such that $d(x) = \frac{1}{r(x)} - 1$. This adaptation leads to our light representation termed ENVMAP++.

More generally, for any bounding sphere, \mathcal{M}_{env} can be represented as:

$$\mathcal{M}_{\text{env}} = \left\{ s \left(x - n(x) \left(\frac{1}{r(x)} - 1 \right) \right) + c \mid x \in \mathcal{S} \right\} \quad (2)$$

where $s \in \mathbb{R}_{>0}$ is a scaling factor (bounding sphere radius) and $c \in \mathbb{R}^3$ is the center of the bounding sphere.

In our implementation, we employ a deformed cube for the mesh structure of \mathcal{M}_{env} to ensure even triangle distribution, analogous to cube mapping techniques. Direct optimization of $r(x)$ and radiance values $L_c(x)$ can be unstable. Hence, we utilize small MLPs with *permutohedral encoding* [Rosu and Behnke 2023b] to predict these values for a given $x \in \mathcal{S}$, where x is represented as a 3D unit vector when querying the encoding, to avoid poles. Laplacian loss $\mathcal{L}_{\text{lap}}(\mathcal{M}_{\text{env}})$ is further applied to smooth the background emitter mesh.

While the actual background rarely resembles a sphere, this approximation suffices in practice as a non-distant light representation. When vertices are positioned infinitely far away ($r \rightarrow 0$ in the limit), this representation approaches the traditional environment map, accommodating a mixture of near-field and far-field lighting scenarios. In addition, since this is simply a textured area light, we can importance sample it without additional effort.

4.2 Shape Reconstruction

4.2.1 Shape Initialization. Our pipeline uses an initialization stage to start with potentially imperfect predictions of object shape, reflectance, and illumination, similar to Neural-PBIR [Sun et al. 2023]. In a later stage, initialized by these predictions, we refine the initial results to obtain the final high-quality reconstruction. We found two initialization to perform the best in different cases. First, our pipeline utilizes the initial SDF reconstructed by PermutoSDF [Rosu and Behnke 2023b] due to its robustness in handling diffuse and rough glossy objects.

Second, for highly glossy objects where PermutoSDF can fail and produce holes, we implement a voxel-based visual hull algorithm [Laurentini 1994] to obtain the initial shape. We first estimate a set of masks based of the input images using a salient object segmentation model such as U²-Net [Qin et al. 2020]. Then we back-project the voxels within the scene bounding box to determine if they are within the foreground mask for a given view. We keep a voxel if the fraction of input views where it falls into the foreground mask is above a given threshold. This approach tends to produce fewer holes and broken geometry features, despite having less accurate boundaries than the result from PermutoSDF. The resulting occupancy grid can be transformed into an SDF using a Euclidean distance transform and serve as our initialization.

4.2.2 Neural Implicit Evolution. Specular highlights or inaccurate segmentation can create holes in our initial mesh, so our shape optimization routine needs to handle topological changes to correct these issues. Further, many objects have thin features such

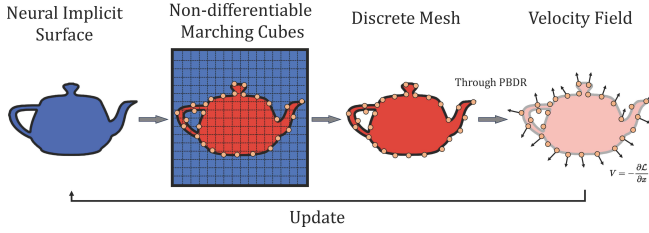


Fig. 3. **Neural Implicit Evolution (NIE)**. Here we illustrate our PBIR-NIE pipeline for optimizing the underlying geometry using neural implicit evolution (NIE) [Mehta et al. 2022]. We represent the geometry with a neural signed distance field (SDF) and employ a surface extraction algorithm, such as non-differentiable marching cubes, to obtain a discretized surface mesh. Next, we compute mesh vertex gradients using physics-based differentiable rendering (PBDR) and update the neural SDF through NIE with the obtained velocity field.

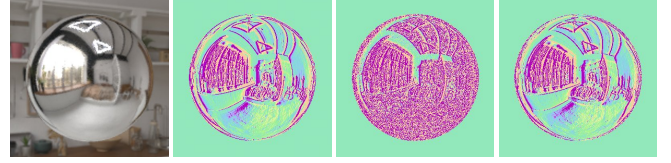
as cup handles. While representing the shape using a neural implicit function would be an ideal solution, only a few works [Cai et al. 2022; Bangaru et al. 2022; Vicini et al. 2022] have successfully combined this approach with a physically based differentiable renderer despite its popularity in neural surface reconstruction. The root of this problem is that accurately estimating the derivative of Equation 1 requires handling a boundary term [Zhang et al. 2020a], which requires tracking discontinuities at object boundaries, but it is particularly challenging for implicit surfaces.

Our pipeline **PBIR-NIE** incorporates neural implicit evolution (NIE) [Mehta et al. 2022] as the backbone for propagating surface gradients from mesh-based differentiable rendering to the underlying neural SDF. Given a neural implicit function $\Phi(x; \theta)$, where θ is the set of parameters of a neural network, we use its isosurface $\mathcal{M} = \{x \mid \Phi(x; \theta) = 0\}$ to represent the object we are reconstructing. NIE optimizes this surface by first extracting a mesh using non-differentiable Marching Cubes and obtains the vertex gradients w.r.t the training loss, $\frac{\partial L}{\partial x}$, through a differentiable renderer. It then constructs a flow field $V = -\frac{\partial L}{\partial x}$ and creates the target implicit function $\phi(x) = \Phi(x; \theta) - \Delta t (\nabla \Phi(x; \theta) \cdot V)$, where Δt is the time step. Finally, the implicit function $\Phi(x; \theta)$ is optimized by minimizing its difference with the target implicit function using gradient descent:

$$\min_{\theta} J(\theta) = \frac{1}{|\mathcal{M}|} \sum_{x \in \mathcal{M}} \|\phi(x) - \Phi(x; \theta)\|^2. \quad (3)$$

NIE is provably better than MeshSDF (further reading in [Mehta et al. 2022]) and it can be integrated into an existing differentiable renderer without requiring differentiable marching cubes; however, only minimal examples of differentiable rendering are shown in the original NIE work; our method is the first to use it in a full inverse rendering pipeline for geometry, materials and lighting.

We introduce two adjustments to NIE for reconstructing glossy objects. First, we clamp the magnitude of the flow field. When the roughness of the object is very low, $\frac{\partial L}{\partial x}$ becomes extremely high and disturbs the training stability. With the formulation of NIE, as long as the implicit surface is evolving towards the correct direction, the magnitude of the flow field is not important. Thus, we simply rescale $\frac{\partial L}{\partial x}$ so that its magnitude is at most ϵ_v .



(a) Rendering (b) Reference (FD) (c) AD w/o AS (d) Ours

Fig. 4. **Antithetic Sampling**. When dealing with highly glossy objects (such as a chrome ball in Fig. 4a), traditional BSDF sampling techniques may result in high variance gradients. In Fig. 4b, we compute finite differences and display the ground-truth gradient image corresponding to the shape translation of a glossy object with a low roughness value of 0.05. Without antithetic sampling (Fig. 4c), the gradient image appears noisy, leading to unstable training. However, by applying antithetic sampling (Fig. 4d), we achieve a significantly more reliable Monte Carlo gradient estimation with the same number of samples.

Second, although NIE is designed to work with level set functions (where the eikonal constraint $\|\nabla_x \Phi(x; \theta)\| = 1$ is not enforced), we found that the absence of this enforcement leads to degraded results at high learning rates. To speed up training, we add the eikonal constraint and enforce the implicit function to behave as a signed distance function. The commonly used eikonal loss $\mathcal{L}_{\text{Eik}} = \frac{1}{2} \|\nabla_x \Phi(x; \theta) - 1\|^2$ is in fact unstable, as demonstrated in [Li et al. 2010; Yang et al. 2023]. As a result, we use the DRLSE loss by Li et al. [2010]:

$$\mathcal{L}_{\text{DRLSE}}(s) = \begin{cases} \frac{1}{(2\pi)^2} (1 - \cos(2\pi s)), & \text{if } s \leq 1 \\ \frac{1}{2}(s - 1)^2, & \text{if } s > 1 \end{cases}, \quad (4)$$

where $s = \|\nabla_x \Phi(x; \theta)\|$. We usually use a weight of 1e-2 in our experiments.

4.3 Material

4.3.1 BRDF Model. We model the material using a simplified version of the Disney BRDF [Karis and Games 2013]. It consists of a diffuse lobe and an anisotropic microfacet specular lobe modeled with the GGX normal distribution function [Walter et al. 2007b]. It is also adopted by other works for inverse rendering [Sun et al. 2023; Luan et al. 2021; Bi et al. 2020]. Although we use an explicit mesh for rendering, the mesh’s connectivity and topology change constantly as the underlying implicit surface evolves. Thus, applying 2D textures to it is impractical. While we can encode the material properties using an MLP, querying them at each intersection point within the rendering loop requires turning off the megakernel mode in Mitsuba 3, resulting in significant performance regression. Instead, we store the predicted material properties at the mesh vertices and then interpolate them to estimate the material properties at the intersection points. The meshes extracted with marching cubes are very high-resolution, so the interpolation errors remain small. If higher-resolution textures are desired, we can run an additional stage with explicit UV mapping after the topology has stabilized.

4.3.2 Antithetic Sampling. Zhang et al. [2021a] found that using glossy or near-specular BRDFs with traditional sampling techniques results in high variance in gradient estimation (Fig. 4). This excessive noise negatively impacts shape optimization. The root of this issue

is that the derivative of the normal distribution function $\frac{dD}{d\omega_h}$ grows rapidly as the roughness decreases. The remedy they proposed is to create an antithetic sample ω'_h by mirroring the sampled half vector ω_h along the normal direction, which has the same D -value. Since $\frac{dD}{d\omega_h} = -\frac{dD}{d\omega'_h}$, the derivatives cancel out and the variance of the gradient estimation is reduced.

Since our goal is to reconstruct glossy objects, employing this technique is essential to ensure robustness. The antithetic sample requires us to trace an additional path. To simplify our implementation, instead of tracing two paths at each intersection, we render the image twice, once with normal BRDF sampling and once with antithetic sampling, and average them at the end. For multi-bounces, we only do antithetic sampling at the first bounce. The same random seed is used in the two passes to ensure correlation. This simplification is effective as demonstrated in Fig. 4d.

5 IMPLEMENTATION DETAILS

Our pipeline makes heavy use of neural networks for scene primitives. All of our neural networks share the same architecture: a small MLP with permutohedral encoding [Rosu and Behnke 2023a]. This ensures memory efficiency while maintaining expressiveness. We use 4 hidden layers of 32 neurons for the SDF network, and 2 hidden layers of 32 neurons for the material network. The emitted radiance $L_e(x)$ and inverted distance $r(x)$ in ENVMAP++ are predicted by two separate MLPs with 2 hidden layers of 32 neurons.

We use the Adam optimizer [Kingma and Ba 2014], with learning rates ranging from $1e-3$ to $1e-4$ depending on the scenes. We optimize most scenes with 2000 iterations with a batch size of 1. The resolution of the non-differentiable marching cubes during the NIE step is 256^3 . As mentioned in Sec. 4.1, we use a sphere mesh constructed as a deformed cube as the starting mesh. The resolution of each face of the cube is 32^2 .

We use a modified version of the prb_projective integrator from Mitsuba 3 with antithetic sampling support. By default, it supports paths of arbitrary depth using the Russian roulette stopping criterion.

6 RESULTS

To demonstrate the effectiveness of our method, we present reconstructions on synthetic input images and the real-world capture dataset Stanford-ORB [Kuang et al. 2024]. We compare the reconstructions obtained with our pipeline against three state-of-the-art baselines: NeRO [Liu et al. 2023], Neural-PBIR [Sun et al. 2023], and NeRF Emitter [Ling et al. 2024]. We demonstrate superior reconstruction quality in terms of geometry and lighting (Sec. 6.1). Additionally, we conduct ablation studies to evaluate several components of our pipeline (Sec. 6.2). Please refer to the supplement for more results.

6.1 Comparison with Baselines

Comparison with NeRO [Liu et al. 2023]. We conducted extensive evaluations comparing our method with NeRO [Liu et al. 2023], focusing on the recovery of texture details and relighting quality in scenes with glossy interreflections. Figs. 8, 9, and 12 illustrate these comparisons in various settings. Fig. 12 demonstrates our ability



Fig. 5. Our results on Stanford-ORB [Kuang et al. 2024] dataset.

to reconstruct detailed material using interreflection, thanks to the correct simulation of global illumination. NeRO fails at this task because it predicts indirect lighting using a neural network. Fig. 9 highlights the relighting quality on NeRO’s glossy synthetic objects BELL, CAT, and TEAPOT. Using the same geometry as NeRO, our method captures detailed highlights and reflections more effectively, demonstrating superior relighting performance. Lastly, Fig. 8, we present a detailed comparison of texture recovery through specular reflections. Two scenes involving a coin and a mushroom placed above a specular table demonstrate that our method successfully reconstructs fine texture details observed through specular reflections, whereas NeRO produces blurrier reconstructions.

Comparison with Neural-PBIR [Sun et al. 2023]. We compared our method with Neural-PBIR [Sun et al. 2023], focusing on the reconstruction quality of glossy objects. Due to the fact that Neural-PBIR’s initialization stage often fails for glossy objects, we specifically evaluated the mesh refinement stage. Fig. 10 shows the results for three different objects: SPOT, KNOT, and CROSS. Starting from the same initial geometry obtained from a visual hull, we refined the meshes using our physics-based inverse rendering (PBIR) approach. Our method demonstrates superior reconstruction quality, effectively capturing detailed highlights and reflections from the environment. This is evident in the sharper and more detailed insets of our results compared to those of Neural-PBIR. In contrast, Neural-PBIR struggles to accurately reproduce the glossy materials and fine geometric details, resulting in less realistic reconstructions.

Comparison with NeRF Emitter [Ling et al. 2024] and standard environment map. We evaluate our ENVMAP++ lighting representation against two baselines: NeRF Emitter [Ling et al. 2024], which uses a NeRF for background illumination, and the standard environment map commonly used in prior inverse rendering frameworks. Since NeRF emitter can only work under direct illumination and has no support for antithetic sampling as of writing, we compare these methods using diffuse objects under direct illumination. As shown in Fig. 11, we assess the joint optimization of object shape, material, and lighting for two scenes, SCULPTURE and DUCK. Both objects are placed inside an indoor room where the surrounding environment violates the infinite-distance assumption of the standard environment map. Our lightweight ENVMAP++ lighting representation enables

Table 1. Geometry, relighting, and view-interpolation quality on Stanford-ORB dataset [Kuang et al. 2024].

Method	Geometry			Novel Scene Relighting				Novel View Synthesis			
	Depth↓	Normal↓	Shape↓	PSNR-H↑	PSNR-L↑	SSIM↑	LPIPS↓	PSNR-H↑	PSNR-L↑	SSIM↑	LPIPS↓
PhySG [Zhang et al. 2021b]	1.90	0.17	9.28	21.81	28.11	0.960	0.055	24.24	32.15	0.974	0.047
NVDiffRec [Munkberg et al. 2022]	0.31	0.06	0.62	22.91	29.72	0.963	0.039	21.94	28.44	0.969	0.030
NeRD [Boss et al. 2021a]	1.39	0.28	13.7	23.29	29.65	0.957	0.059	25.83	32.61	0.963	0.054
NeRFactor [Zhang et al. 2021c]	0.87	0.29	9.53	23.54	30.38	0.969	0.048	26.06	33.47	0.973	0.046
InvRender [Zhang et al. 2022b]	0.59	0.06	0.44	23.76	30.83	0.970	0.046	25.91	34.01	0.977	0.042
NVDiffRecMC [Hasselgren et al. 2022]	0.32	0.04	0.51	<u>24.43</u>	<u>31.60</u>	0.972	0.036	<u>28.03</u>	<u>36.40</u>	0.982	0.028
Neural-PBIR [Sun et al. 2023]	0.30	0.06	0.43	26.01	33.26	0.979	0.023	28.83	36.80	0.986	0.019
PBIR-NIE (ours)	0.50	<u>0.05</u>	0.64	26.26	33.46	<u>0.977</u>	<u>0.028</u>	27.06	35.09	<u>0.983</u>	<u>0.023</u>

superior inverse rendering quality and geometry reconstruction, comparable to the computationally expensive NeRF Emitter results.

Stanford-ORB [Kuang et al. 2024] results. We validate our pipeline on the real-world inverse rendering dataset Stanford-ORB [Kuang et al. 2024]. The results demonstrate the effectiveness of our method in terms of geometry reconstruction, relighting, and view interpolation. Our method achieves the best performance on novel scene relighting, while being comparable to Neural-PBIR [Sun et al. 2023] on geometry reconstruction and novel view synthesis. Qualitative results are shown in Fig. 5, and quantitative results are summarized in Table 1.

6.2 Evaluations and Ablations

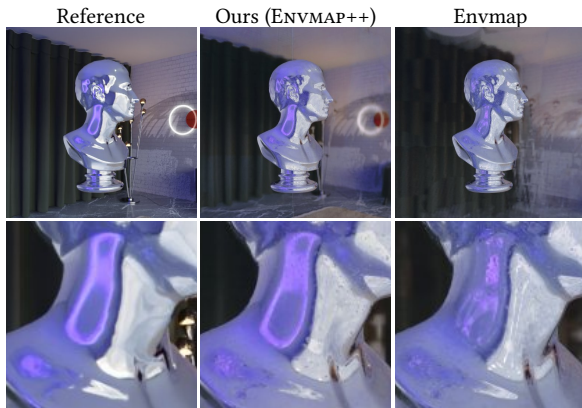


Fig. 6. **Ablation on ENVMAP++.** We evaluate the quality of glossy object appearance acquisition under non-distant background illumination using our proposed ENVMAP++ vs. standard environment map lighting.

Importance of ENVMAP++. In this experiment, we ablate the importance of our proposed ENVMAP++ for capturing glossy objects in a near-field environment. We use ground-truth geometry and compare the appearance of the glossy object with ENVMAP++ versus a standard environment map. As shown in Fig. 6, a highly glossy sculpture is placed in an indoor room, presenting strong parallax in the background. Our ENVMAP++ successfully captures this non-distant lighting, resulting in a more accurate appearance capture. In contrast, the baseline using standard environment map lighting representation fails to represent the near-field background illumination, leading to poorer quality.



(a) Reference (b) Initial (c) Large Steps (d) Ours (NIE)

Fig. 7. **Ablation on NIE [Mehta et al. 2022]** We demonstrate the necessity of NIE when given poor initialization.

NIE vs. Large Steps [Nicolet et al. 2021]. In this experiment, we ablate the NIE component of our pipeline by comparing the reconstructed geometry with fixed lighting and material against the reconstruction produced by the *Large Steps* algorithm [Nicolet et al. 2021]. The *Large Steps* algorithm operates on an explicit mesh with fixed topology, whereas *NIE* uses the explicit mesh as a proxy to optimize an implicit surface capable of topology changes. This capability is crucial when dealing with poor initializations that contain holes, as demonstrated in Fig. 7c and Fig. 7, where *NIE* outperforms *Large Steps*. Additionally, the gradient clamping step described in Sec. 4.2.2 helps stabilize the optimization process without creating spikes, as shown in Fig. 7c.

7 DISCUSSION AND CONCLUSION

Limitations and future work. Although ENVMAP++ provides a lightweight solution to address the non-distant illumination problem, its representational power is greatly limited by the underlying geometry: a deformed cube cannot possibly model complex backgrounds, especially when the background is very close to the object or when occlusion is present. Additionally, it cannot accurately model strong directional light sources due to the lack of directional information in the radiance texture. Nevertheless, we found it satisfactory for common scenes. Developing lightweight emitters with greater representational power would be of future interest.

Another issue we aim to further explore is the ambiguity between lighting and material. This ambiguity often manifests as "baking," for instance, reconstructed textures may incorrectly embed specular highlights. While humans can easily identify and resolve such artifacts, optimizers see these artifacts as just one of many possible solutions. To find the optimal solution, a strong prior in material and lighting is necessary.

Conclusion. In this work, we introduced an inverse rendering framework, PBIR-NIE, for reconstructing highly glossy objects' geometry, material, and surrounding illumination. By integrating neural implicit evolution, we achieved robust handling of complex object topology without the need for careful initialization. Our light-weight background representation, ENVMAP++, efficiently models both near-field and far-field background illumination, offering a more efficient solution for inverse rendering tasks. To tackle challenges with highly glossy BRDFs, we integrated an efficient variant of antithetic sampling, enabling faster convergence and more accurate reconstruction. Our pipeline delivers state-of-the-art results in geometry, material, and lighting estimation, enabling realistic view synthesis and relighting.

REFERENCES

- Miika Aittala, Tim Weyrich, and Jaakko Lehtinen. 2013. Practical SVBRDF capture in the frequency domain. *ACM Trans. Graph.* 32, 4 (2013), 110–1.
- Sai Bangaru, Tzu-Mao Li, and Frédo Durand. 2020. Unbiased Warped-Area Sampling for Differentiable Rendering. *ACM Trans. Graph.* 39, 6 (2020), 245:1–245:18.
- Sai Praveen Bangaru, Michael Gharbi, Fujun Luan, Tzu-Mao Li, Kalyan Sunkavalli, Milos Hasan, Sai Bi, Zexiang Xu, Gilbert Bernstein, and Fredo Durand. 2022. Differentiable rendering of neural sdfs through reparameterization. In *SIGGRAPH Asia 2022 Conference Papers*. 1–9.
- Sai Bi, Zexiang Xu, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. 2020. Deep reflectance volumes: Relightable reconstructions from multi-view photometric images. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III* 16. Springer, 294–311.
- 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.
- G. Cai, Y. Yan, Z. Dong, I. Gkioulekas, and S. Zhao. 2022. Physics-Based Inverse Rendering Using Combined Implicit and Explicit Geometries. *Computer Graphics Forum* 41, 4 (July 2022). <https://doi.org/10.1111/cgf.14592>
- Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. 2022. Tensor4: Tensorial radiance fields. In *European Conference on Computer Vision*. Springer, 333–350.
- Yue Dong, Guojun Chen, Pieter Peers, Jiawan Zhang, and Xin Tong. 2014. Appearance-from-motion: Recovering spatially varying surface reflectance under unknown lighting. *ACM Transactions on Graphics (TOG)* 33, 6 (2014), 1–12.
- Yue Dong, Jiaping Wang, Xin Tong, John Snyder, Yanxiang Lan, Moshe Ben-Ezra, and Baining Guo. 2010. Manifold bootstrapping for SVBRDF capture. *ACM Transactions on Graphics (TOG)* 29, 4 (2010), 1–10.
- 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. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4251–4260.
- Jon Hasselgren, Nikolai Hofmann, and Jacob Munkberg. 2022. Shape, light, and material decomposition from images using monte carlo rendering and denoising. *Advances in Neural Information Processing Systems* 35 (2022), 22856–22869.
- Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, Merlin Nimier-David, Delio Vicini, Tizian Zeltner, Baptiste Nicolet, Miguel Crespo, Vincent Leroy, and Ziyi Zhang. 2022b. *Mitsuba 3 renderer*. <https://mitsuba-renderer.org>.
- Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, and Delio Vicini. 2022a. DrJit: A Just-In-Time Compiler for Differentiable Rendering. *Transactions on Graphics (Proceedings of SIGGRAPH)* 41, 4 (July 2022). <https://doi.org/10.1145/3528223.3530099>
- Haian Jin, Isabella Liu, Peijia Xu, Xiaoshuai Zhang, Songfang Han, Sai Bi, Xiaowei Zhou, Zexiang Xu, and Hao Su. 2023. Tensorio: Tensorial inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 165–174.
- Brian Karis and Epic Games. 2013. Real shading in unreal engine 4. *Proc. Physically Based Shading Theory Practice* 4, 3 (2013), 1.
- Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics* 42, 4 (2023), 1–14.
- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- Zhengfei Kuang, Kyle Olszewski, Menglei Chai, Zeng Huang, Panos Achlioptas, and Sergey Tulyakov. 2022. Nerotic: Neural rendering of objects from online image collections. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–12.
- Zhengfei Kuang, Yunzhi Zhang, Hong-Xing Yu, Samir Agarwala, Elliott Wu, Jiajun Wu, et al. 2024. Stanford-ORB: A Real-World 3D Object Inverse Rendering Benchmark. *Advances in Neural Information Processing Systems* 36 (2024).
- Aldo Laurentini. 1994. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on pattern analysis and machine intelligence* 16, 2 (1994), 150–162.
- Chunming Li, Chenyang Xu, Changfeng Gui, and Martin D. Fox. 2010. Distance Regularized Level Set Evolution and Its Application to Image Segmentation. *IEEE Transactions on Image Processing* 19, 12 (Dec. 2010), 3243–3254. <https://doi.org/10.1109/TIP.2010.2069690>
- Jingwang Ling, Ruihan Yu, Feng Xu, Chun Du, and Shuang Zhao. 2024. NeRF as Non-Distant Environment Emitter in Physics-based Inverse Rendering. *arXiv preprint arXiv:2402.04829* (2024).
- 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:2305.17398* [cs]
- Fujun Luan, Shuang Zhao, Kavita Bala, and Zhao Dong. 2021. Unified shape and svbrdf recovery using differentiable monte carlo rendering. In *Computer Graphics Forum*, Vol. 40. Wiley Online Library, 101–113.
- Li Ma, Vasu Agrawal, Haithem Turki, Changil Kim, Chen Gao, Pedro Sander, Michael Zollhöfer, and Christian Richardt. 2023. SpecNeRF: Gaussian Directional Encoding for Specular Reflections. *arXiv preprint arXiv:2312.13102* (2023).
- Stephen Robert Marschner. 1998. *Inverse rendering for computer graphics*. Cornell University.
- Ishit Mehta, Manmohan Chandraker, and Ravi Ramamoorthi. 2022. A Level Set Theory for Neural Implicit Evolution Under Explicit Flows. In *Computer Vision – ECCV 2022*, Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (Eds.). Vol. 13662. Springer Nature Switzerland, Cham, 711–729. https://doi.org/10.1007/978-3-031-20086-1_41
- 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*.
- Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)* 41, 4 (2022), 1–15.
- 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.
- Giljoo Nam, Joo Ho Lee, Diego Gutierrez, and Min H Kim. 2018. Practical svbrdf acquisition of 3d objects with unstructured flash photography. *ACM Transactions on Graphics (TOG)* 37, 6 (2018), 1–12.
- Giljoo Nam, Joo Ho Lee, Hongzhi Wu, Diego Gutierrez, and Min H Kim. 2016. Simultaneous acquisition of microscale reflectance and normals. *ACM Trans. Graph.* 35, 6 (2016), 185–1.
- Baptiste Nicolet, Alec Jacobson, and Wenzel Jakob. 2021. Large Steps in Inverse Rendering of Geometry. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 40, 6 (Dec. 2021). <https://doi.org/10.1145/3478513.3480501>
- Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R Zaiane, and Martin Jagersand. 2020. U2-Net: Going deeper with nested U-structure for salient object detection. *Pattern recognition* 106 (2020), 107404.
- Ravi Ramamoorthi and Pat Hanrahan. 2001. A signal-processing framework for inverse rendering. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. 117–128.
- Edoardo Remelli, Artem Lukoianov, Stephan Richter, Benoit Guillard, Timur Bagautdinov, Pierre Baque, and Pascal Fua. 2020. Meshsdf: Differentiable iso-surface extraction. *Advances in Neural Information Processing Systems* 33 (2020), 22468–22478.
- Radu Alexandru Rosu and Sven Behnke. 2023a. Permutosdf: Fast multi-view reconstruction with implicit surfaces using permutohedral lattices. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8466–8475.
- Radu Alexandru Rosu and Sven Behnke. 2023b. PermutoSDF: Fast Multi-View Reconstruction with Implicit Surfaces Using Permutohedral Lattices. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Vancouver, BC, Canada, 8466–8475. <https://doi.org/10.1109/CVPR52729.2023.00818>
- 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.
- Cheng Sun, Guangyan Cai, Zhengqin Li, Kai Yan, Cheng Zhang, Carl Marshall, Jia-Bin Huang, Shuang Zhao, and Zhao Dong. 2023. Neural-PBIR Reconstruction of Shape, Material, and Illumination. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. 18000–18010. <https://doi.org/10.1109/ICCV51070.2023.01654>

- Eric Veach. 1997. *Robust Monte Carlo methods for light transport simulation*. Stanford University.
- 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.
- Delio Vicini, Sébastien Speierer, and Wenzel Jakob. 2022. Differentiable Signed Distance Function Rendering. *ACM Transactions on Graphics* 41, 4 (July 2022), 125:1–125:18. <https://doi.org/10.1145/3528223.3530139>
- Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. 2007a. Microfacet models for refraction through rough surfaces. In *Proceedings of the 18th Eurographics conference on Rendering Techniques*. 195–206.
- Bruce Walter, Stephen R. Marschner, Hongsong Li, and Kenneth E. Torrance. 2007b. Microfacet models for refraction through rough surfaces. In *Proceedings of the 18th Eurographics Conference on Rendering Techniques (Grenoble, France) (EGSR'07)*. Eurographics Association, Goslar, DEU, 195–206.
- Haoyuan Wang, Wenbo Hu, Lei Zhu, and Rynson WH Lau. 2024. Inverse Rendering of Glossy Objects via the Neural Plenoptic Function and Radiance Fields. *arXiv preprint arXiv:2403.16224* (2024).
- 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. *arXiv preprint arXiv:2106.10689* (2021).
- Liwen Wu, Sai Bi, Zexiang Xu, Fujun Luan, Kai Zhang, Iliyan Georgiev, Kalyan Sunkavalli, and Ravi Ramamoorthi. 2024. Neural Directional Encoding for Efficient and Accurate View-Dependent Appearance Modeling. (2024).
- Rui Xia, Yue Dong, Pieter Peers, and Xin Tong. 2016. Recovering shape and spatially-varying surface reflectance under unknown illumination. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 1–12.
- Huizong Yang, Yuxin Sun, Ganesh Sundaramoorthi, and Anthony Yezzi. 2023. StEik: Stabilizing the Optimization of Neural Signed Distance Functions and Finer Shape Representation. <https://doi.org/10.48550/arXiv.2305.18414> arXiv:2305.18414 [cs, math]
- Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. 2021. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems* 34 (2021), 4805–4815.
- Cheng Zhang, Zhao Dong, Michael Doggett, and Shuang Zhao. 2021a. Antithetic sampling for Monte Carlo differentiable rendering. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–12.
- Cheng Zhang, Bailey Miller, Kai Yan, Ioannis Gkioulekas, and Shuang Zhao. 2020a. Path-Space Differentiable Rendering. *ACM Trans. Graph.* 39, 4 (2020), 143:1–143:19.
- 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. 2021b. 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.
- Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. 2020b. NeRF++: Analyzing and Improving Neural Radiance Fields. <https://doi.org/10.48550/arXiv.2010.07492> arXiv:2010.07492 [cs]
- Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. 2021c. 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.
- Shuang Zhao, Wenzel Jakob, and Tzu-Mao Li. 2020. Physics-based differentiable rendering: from theory to implementation. In *ACM siggraph 2020 courses*. 1–30.
- Zhiming Zhou, Guojun Chen, Yue Dong, David Wipf, Yong Yu, John Snyder, and Xin Tong. 2016. Sparse-as-possible SVBRDF acquisition. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 1–12.
- 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. <https://doi.org/10.48550/arXiv.2304.08757> [cs]

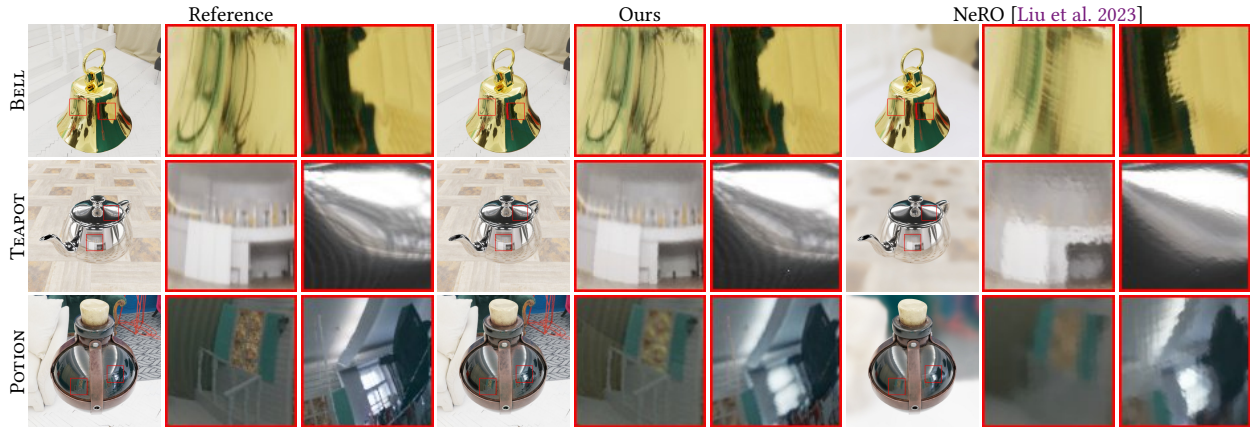


Fig. 8. **Comparison against NeRO [Liu et al. 2023] on material & lighting reconstruction.** We evaluate the quality of material and lighting reconstruction using NeRO’s glossy synthetic dataset. In this experiment, we use the same geometry as NeRO and compare NeRO’s stage II results with ours. Our PBIR-NIE demonstrates superior reconstruction quality on the appearance of glossy objects, capturing detailed highlights from the environment more effectively.

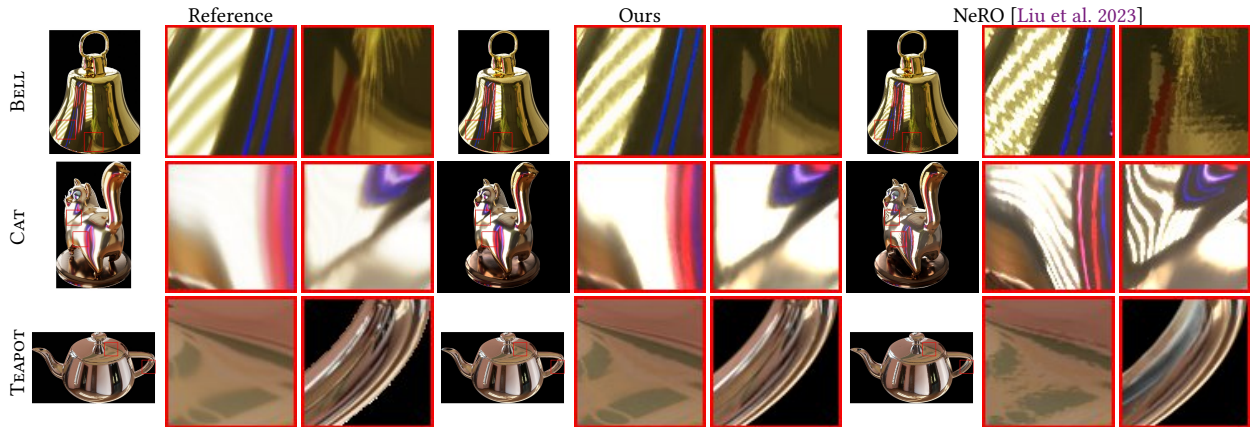


Fig. 9. **Relighting quality comparison with NeRO [Liu et al. 2023].** Similar to the above figure, but here we render under a novel view and lighting. Using the same geometry as NeRO, we compare the stage II results of NeRO with those of our PBIR-NIE. Ours relighting captures highlights on glossy objects in novel environments more accurately.



Fig. 10. **Comparison against Neural-PBIR [Sun et al. 2023] on glossy object reconstruction.** We evaluate the quality of glossy object reconstruction using Neural-PBIR and our method. Due to the fact that Neural-PBIR’s initialization stage often fails for glossy objects, we specifically evaluated the mesh refinement stage. In this experiment, we start with the same initial geometry obtained from a visual hull and compare the physics-based inverse rendering (PBIR) mesh refinement. Our PBIR-NIE demonstrates superior reconstruction quality, effectively capturing detailed highlights from the environment. In contrast, Neural-PBIR struggles to reproduce glossy materials and geometric details.



Fig. 11. **Comparison of lighting representations: NeRF Emitter [Ling et al. 2024], our ENVMAP++, and standard environment map.** In these two scenes, we jointly optimize the object shape, materials, and surrounding lighting with each lighting representation. We report the quantitative results — PSNR for the foreground-only region and the whole image (in the format 'PSNR: (foreground / whole) dB'), as well as Chamfer distance — above the corresponding image, and color-code the **best** and **second best** method accordingly. Our pipeline successfully recovers the object's geometry, while the standard environment map fails in these scenes due to the violation of the infinite-distance assumption and background parallax. NeRF Emitter uses a NeRF for background lighting, which is computationally expensive. In contrast, our ENVMAP++ is a lightweight representation, providing more efficient inverse rendering optimization.

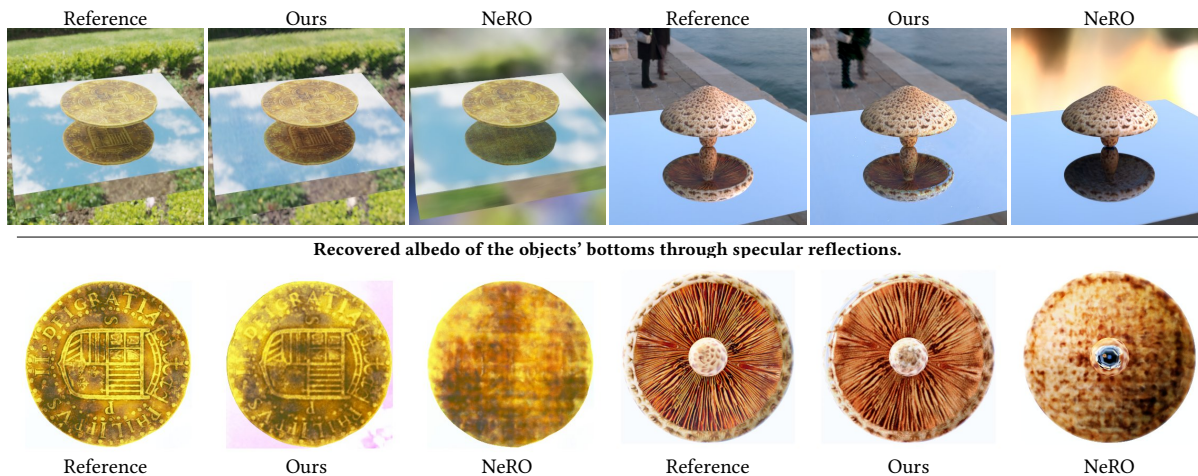


Fig. 12. **Comparison against NeRO [Liu et al. 2023] on glossy interreflections.** In these two scenes, an object (a coin or a mushroom) is placed above a specular table, where the input views can observe the bottom of the object through specular reflections on the table. Our pipeline successfully recovers the texture details on the object, while NeRO fails to and produces blurry reconstructions.

8 SUPPLEMENTAL RESULTS

The figures presented below serve as supplementary material to the comparisons discussed in Sec. 6.1. Fig. 13 provides various visualizations of a subset of our reconstructions of the Stanford-ORB dataset [Kuang et al. 2024]. Figures 14 and 15 provide additional comparisons with NeRO on their glossy synthetic dataset.



Fig. 13. Additional results of our pipeline on Stanford-ORB [Kuang et al. 2024] data. .

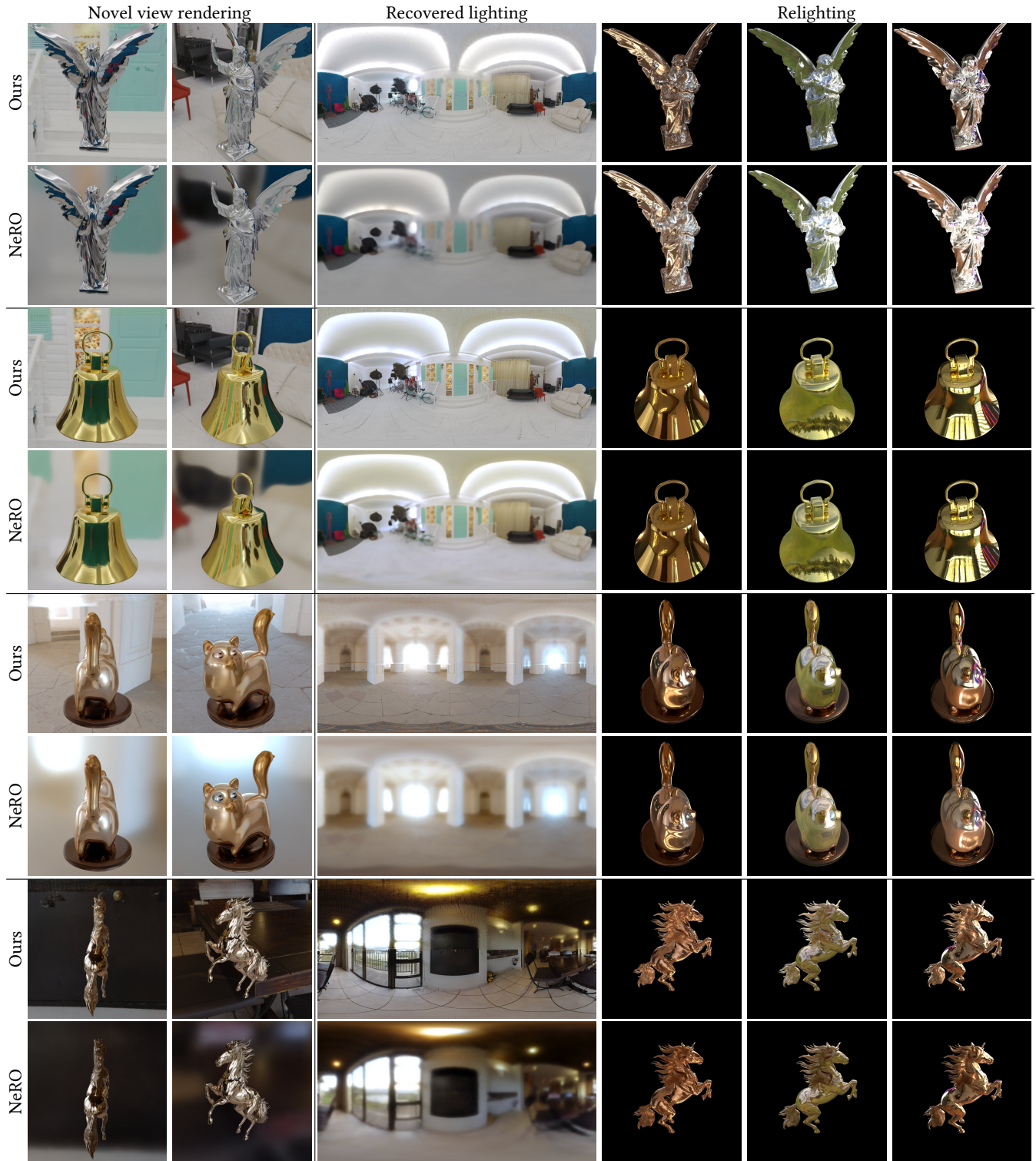


Fig. 14. Additional comparisons of our method against NeRO [Liu et al. 2023] on glossy synthetic data. Please zoom in to better compare the results.

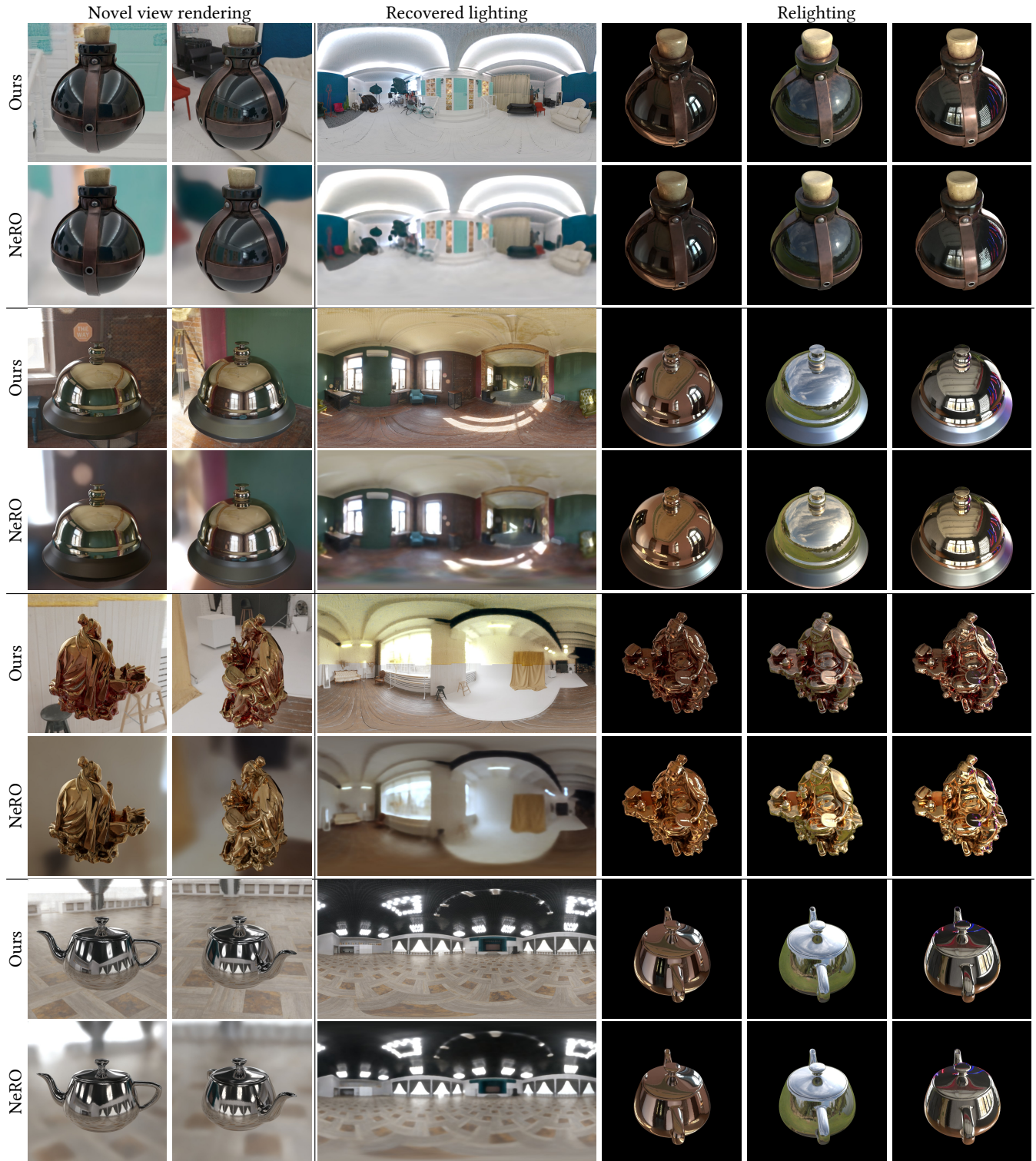


Fig. 15. Additional comparisons of our method against NeRO [Liu et al. 2023] on glossy synthetic data. Please zoom in to better compare the results.