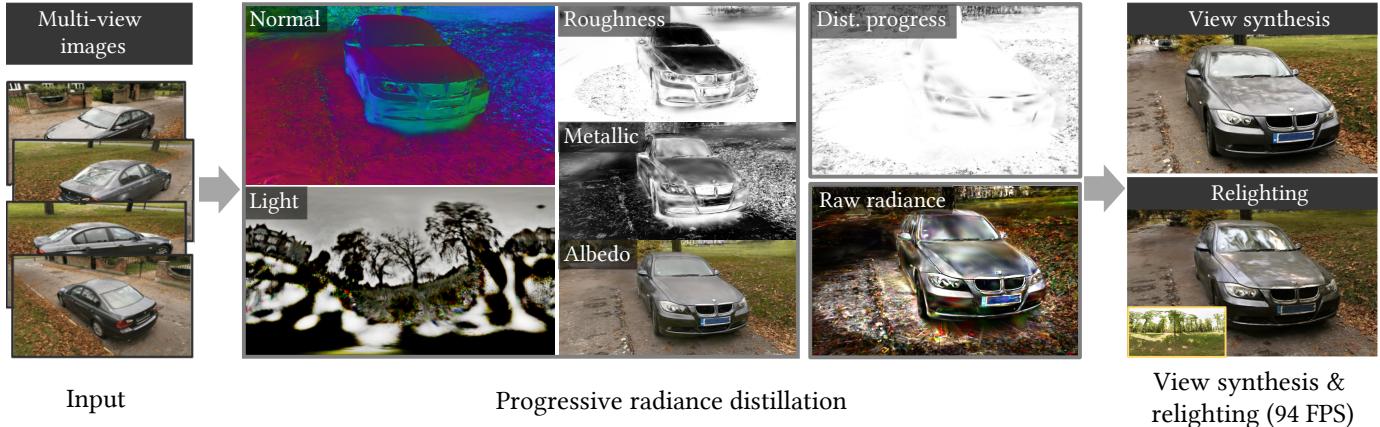


Progressive Radiance Distillation for Inverse Rendering with Gaussian Splatting

Keyang Ye, Qiming Hou, and Kun Zhou



We propose progressive radiance distillation for geometry-light-material decomposition from multi-view images. Our rendering model combines physically-based rendering with Gaussian-based radiance field rendering using a distillation progress map, which outperforms state-of-the-art methods in both novel view synthesis and relighting. The scene is *sedan* from [1], a real-world capture. We use a spherical mask to restrict distillation to foreground objects, which is visible in the distilled layers but does not manifest as noticeable artifacts in final results.

Abstract—We propose *progressive radiance distillation*, an inverse rendering method that combines physically-based rendering with Gaussian-based radiance field rendering using a distillation progress map. Taking multi-view images as input, our method starts from a pre-trained radiance field guidance, and distills physically-based light and material parameters from the radiance field using an image-fitting process. The distillation progress map is initialized to a small value, which favors radiance field rendering. During early iterations when fitted light and material parameters are far from convergence, the radiance field fallback ensures the sanity of image loss gradients and avoids local minima that attracts under-fit states. As fitted parameters converge, the physical model gradually takes over and the distillation progress increases correspondingly. In presence of light paths unmodeled by the physical model, the distillation progress never finishes on affected pixels and the learned radiance field stays in the final rendering. With this designed tolerance for physical model limitations, we prevent unmodeled color components from leaking into light and material parameters, alleviating relighting artifacts. Meanwhile, the remaining radiance field compensates for the limitations of the physical model, guaranteeing high-quality novel views synthesis. Experimental results demonstrate that our method significantly outperforms state-of-the-art techniques quality-wise in both novel view synthesis and relighting. The idea of progressive radiance distillation is not limited to Gaussian splatting. We show that it also has positive effects for prominently specular scenes when adapted to a mesh-based inverse rendering method.

Index Terms—Novel view synthesis, relighting, Gaussian splatting, NeRF, real-time rendering.

K. Zhou is the corresponding author. All authors are with State Key Lab of CAD & CG, Zhejiang University, Hangzhou 310058, China.
E-mail: kunzhou@acm.org

I. INTRODUCTION

INVERSE rendering, a pivotal discipline within computer graphics and vision, solves the rendering equation [2] to decompose observed images into physically-based components such as lighting conditions, material properties, and geometry. When interpreted as a regression problem, though, the rendering equation poses significant mathematical ambiguity. Light and material can be challenging to factorize as they always appear in the same integration and infinitely many combinations can yield the same pixel colors. Such ambiguity frequently allows approximation errors from idealized light-material assumptions to leak into model parameters, which can cause visible artifacts when synthesizing new images.

This ambiguity problem can be side-stepped by dropping the light-material decomposition in favor of a *radiance field*, a learned function parameterized over spatial position and view direction. Such a function is significantly more robust to learn, leading to high-fidelity novel view synthesis. This has been demonstrated by the seminal works of NeRF (Neural Radiance Field) [3] and 3DGS (3D Gaussian Splatting) [4]. The robustness comes at a cost, though, as light and material remain entangled and cannot change independently during image synthesis.

Many attempts [5]–[9] have been made to apply successful radiance field function representations to inverse rendering, among which several approaches [8], [10] keep the original uninterpreted radiance alongside light-material integration as an indirect light approximation and train their models from

scratch. Such a radiance term alleviates novel view synthesis quality issues caused by the limitations of physically-based rendering models. However, function representations are orthogonal to the ambiguity problem and additive uninterpreted radiance amplifies it by adding yet another ambiguity source.

We tackle the ambiguity problem by treating radiance fields and physically-based models as interchangeable rather than complimentary. Specifically, we relate them analogously to the *distillation* concept in physics, where different components are gradually extracted from a liquid as the temperature rises. Compared to joint optimization from scratch, which attempts to factorize all components at once, distillation effectively isolates the ambiguities between different components, resulting in a more accurate decomposition.

We propose *progressive radiance distillation*, an inverse rendering method using Gaussian splatting. It initializes from a pre-trained Gaussian-based radiance field guidance, and distills physically-based light and material parameters from the radiance field using a hybrid image-fitting process. Specifically, for each training image, we render one image using radiance field and another using physically-base parameters. The loss function is computed on a final predicted image, linearly interpolated from the two renderings using a learned *distillation progress map* for weights, which guarantees an image quality better or equal to the raw radiance baseline by optimizing the radiance distribution.

The distillation progress map is initialized to a small value, which favors the radiance field and only introduces a minuscule amount of physically-based rendering. During early iterations when fitted physical parameters are still far from convergence, the radiance field fallback ensures the sanity of image loss gradients and avoids local minima that attracts under-fit states. As light and material parameters converge to a comparable level of fitting accuracy, the physical model gradually takes over and the distillation progress increases correspondingly. In presence of light paths unmodeled by the physical model, the distillation progress never reaches 100% on affected pixels and the learned radiance field stays in the final rendering. With this designed tolerance for physical model limitations, we prevent unmodeled color components from leaking into light and material parameters, alleviating ambiguity-related artifacts for better relighting. Meanwhile, the remaining radiance field compensates for the limitations of the physical model, guaranteeing high-quality novel views synthesis. To further reduce ambiguity, we choose to fit specular and diffuse parameters separately as opposed to the more common joint optimization. In presence of the radiance field fallback, this allows each optimization stage to focus on the pixels where the fitted components are most prominent. We choose to fit the specular component first as its physical model out-competes learned radiance faster, making it easier to distill.

Experimental results on several published datasets demonstrate that our method significantly outperforms state-of-the-art techniques quality-wise in both novel view synthesis and relighting, especially enhancing the reconstruction accuracy for scenes with prominent specular reflections. Ablation studies shows that our stage-wise radiance distillation is effective

at avoiding early-stage local minima and alleviates light-material ambiguities. We also explore the possibility of applying our progressive distillation paradigm to mesh-based methods, demonstrating its generalization potential for inverse rendering.

II. RELATED WORK

Inverse rendering [11]–[13], the task of decomposing the appearance of objects in observed images into the underlying geometry, material properties, and lighting conditions, has been extensively explored in both computer graphics and vision. Making a full decomposition involves two important considerations: how to represent geometry and appearance, and how to address the ill-posed problem.

As a widely used geometry representation in graphics pipelines and game engines, meshes represent explicit surfaces and support high performance rendering, which sparks a surge of research in mesh-based differentiable rendering [14]–[17]. However, due to the depth discontinuities that arise when rendering a mesh to an image, and the difficulty posed in changing their topology and avoiding self-intersections [18], mesh-based methods require good geometry initialization, have limited ability for fine-tuning inaccurate topology [19], and are unstable to the ambiguity caused by surface reflections [20]. To alleviate these issues, implicit representations, typified by signed distance field (SDF), have emerged as alternatives [21], [22]. SDF is typically parameterized by a Multi-Layer Perceptron (MLP), with its zero level set corresponding to the geometric surface. Due to its continuity and independence from topological constraints, shape optimization becomes more flexible. However, the computation of SDF-based ray-surface intersection usually relies on time-consuming ray marching, leading to a long training time. Additionally, SDF tends to provide smooth normals, failing to capture some subtle details. NDR [23] and Neural-PBIR [24] use SDF as a geometry initialization, extract coarse meshes from the SDF, and fine-tune the meshes through differentiable rendering. They achieve high reconstruction quality on diffuse-dominant objects with a moderate training time while still struggle with reflective objects. They tend to use fragmented geometry within the surface to fit view-dependent changes. The fixed topology limits further fine-tuning after geometry extraction.

Similar to SDF, radiance fields methods, such as NeRF [3] and 3DGS [4], also construct a implicit and optimizable geometry. NeRF-based methods [5], [6], [25]–[28] replace the emissive radiance with material properties predicted by MLPs and build a fully differentiable rendering pipeline. However, the predicted volume density is underconstrained, leading to inaccurate normal reconstruction. Some methods [7], [29], [30] replace the density with signed distance and perform volume rendering near the zero level set, but still remain the same shortcomings as SDF-based methods. Recently, 3D Gaussians became a popular representation of radiance fields. Gaussian-based inverse rendering methods [8]–[10], [31], [32] inherit the highly efficient training and rendering from vanilla 3DGS, but they are difficult to perfectly reconstruct reflective surfaces.

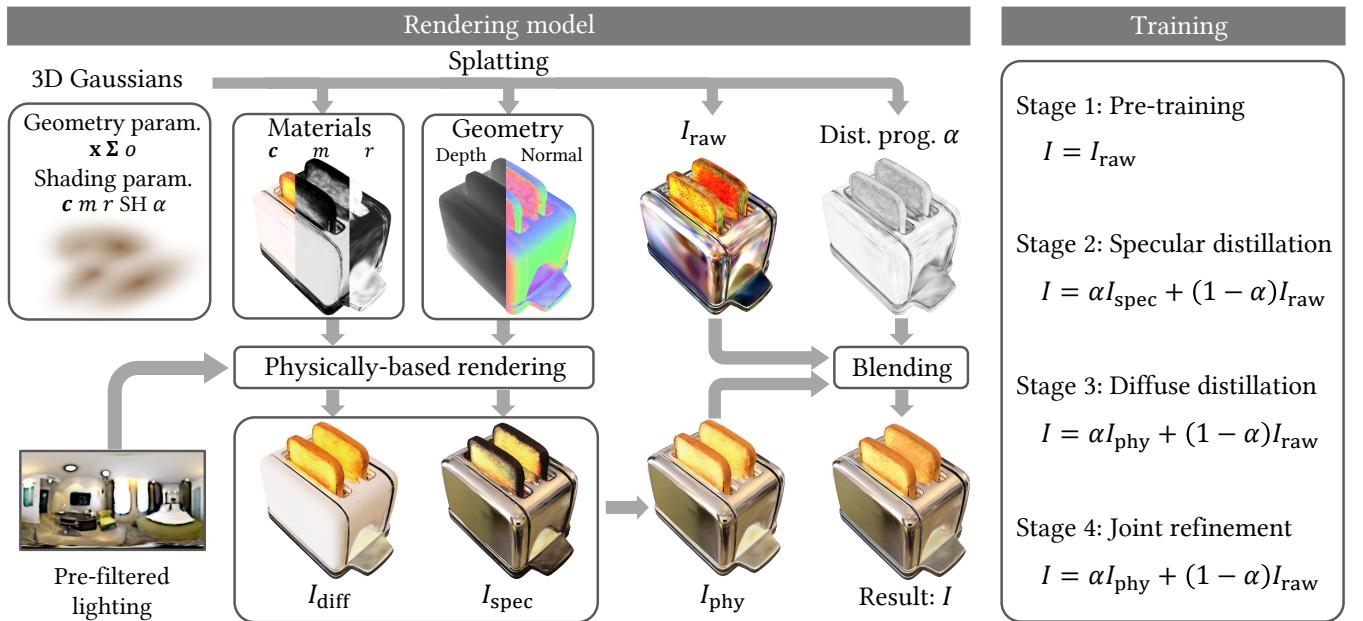


Fig. 1: The rendering model and training stages of our method. We adopt a deferred shading pipeline. First, geometry parameters (position: \mathbf{x} , covariance: Σ and opacity: o) and shading parameters (albedo: c , metallic: m , roughness: r , SH colors y_l^m and distillation progress: α) are splatted into several screen space maps, including material maps (albedo, metallic and roughness), geometry maps (depth and normal), the raw radiance map I_{raw} and the distillation progress map α . Then, we send the maps into a physically-based shader to produce diffuse and specular color maps: I_{diff} and I_{spec} . The physical term I_{phy} combines I_{diff} and I_{spec} , and then gets blended with I_{raw} using α as weight to produce the final result. The various shading parameters are distilled from I_{raw} using a four-stage training process. To mitigate ambiguity, we subset and specialize the rendering model for each individual stage.

3DGS-DR [33] proposed a deferred rendering pipeline with normal propagation, achieving detailed normal reconstruction for reflective surfaces. We choose 3D Gaussians as our geometry representation and incorporate contributions from 3DGS-DR, achieving accurate and efficient reconstruction for objects with various roughness.

For the appearance, we choose the widely used Cook-Torrance microfacet model [34] with albedo, roughness and metallic as its parameters. To keep the real-time performance, our rendering is based on image based lighting [35]. Other rendering methods, such as Monte-Carlo path tracing [36], adopted by NDRMC [17] and Neural-PBIR [24], may achieve a more thorough decomposition but at the cost of decreased efficiency.

The inherent ambiguity has been a recurrent problem since the beginning of inverse rendering research. Most previous methods attempt to address this difficulty by adding priors and regularizations. Specifically, these methods minimize the loss function, which consists of a image loss and other regularization terms, to optimize a physically based rendering model from scratch. In contrast, we alleviate ambiguities by carefully controlling the training process to optimize different components in isolation. To this end, we design a hybrid model linearly combining radiance fields with physical models using optimizable distillation progress as the final rendering model, and gradually increase the distillation progress to drive a full factorization using image-fitting. Similarly, some

prior works also utilize a trained radiance field to reconstruct material/lighting parameters. For example, [24] initializes the BRDF parameters by fitting the trained radiance field to accelerate convergence. [37] recover density from the trained radiance field to initialize the optimization of physical medium parameters. However, they only use the radiance field to initialize the purely physically based model optimization. Besides, some prior works [8], [10] also combine a emissive radiance term with the physical model, regarding it as an additive indirect lighting approximation. However, they are directly added together and jointly optimized from scratch, which can potentially increase ambiguity. In contrast, our distillation process starts from a trained radiance field model, and our linear interpolation design allows the radiance term to remain stable until its weight decreases to nearly zero. Compared with state-of-the-art methods, our hybrid rendering model and progressive distillation mechanism produce superior results in both novel view synthesis (NVS) and relighting, contributing valuably to inverse rendering.

III. PROGRESSIVE RADIANCE DISTILLATION

A. Rendering Model

The general workflow of 3DGS [4] consists of fitting a set of spatial Gaussians carrying learned parameters. Images are synthesized by splatting and blending Gaussian properties to pixels, and the same image synthesis model is used for training

and rendering. We extend the pipeline for inverse rendering while following the general workflow.

We categorize the per-Gaussian parameters into *geometry parameters* and *shading parameters*. Geometry parameters consist of the position, rotation, scaling and opacity of each Gaussian. When combined, they completely define the Gaussian-to-pixel mapping, a linear function that splats and blends the per-Gaussian shading parameters to screen space. We then compute the final pixel colors from the splatted screen space parameters using deferred shading [38].

We will focus on the final shading step as it is most relevant to our paper. Our full rendering model, the version applied after inverse rendering converges, is specified:

$$I(x, \omega_o) = \alpha_x I_{\text{phy}}(x, \omega_o) + (1 - \alpha_x) I_{\text{raw}}(x, \omega_o). \quad (1)$$

Here $I(x, \omega_o)$ is the final radiance parameterized over spatial position x and view direction ω_o . $I_{\text{phy}}(x, \omega_o)$ is the physically-based term and $I_{\text{raw}}(x, \omega_o)$ is the radiance field term, which we fine-tune during our distillation. α_x is the distillation progress value. It is a learned function of position x and we put the variable in the subscript for clarity. All similar subscripts follow this meaning.

The raw radiance term follows the original 3DGS paper [4] and uses an order-4 SH (Spherical Harmonics) [39] function:

$$I_{\text{raw}}(x, \omega_o) = \sum_{l=0}^3 \sum_{m=-l}^l y_{l,x}^m Y_l^m(\omega_o), \quad (2)$$

where $Y_l^m(\omega_o)$ is the spherical harmonic basis function. $y_{l,x}^m$ are the corresponding coefficients. Note that the SH series are evaluated on the view direction ω_o .

The physically-based term is based on the Cook-Torrance microfacet model [34]:

$$I_{\text{phy}}(x, \omega_o) = (1 - m_x) I_{\text{diff}}(x, \omega_o) + I_{\text{spec}}(x, \omega_o), \quad (3)$$

$$I_{\text{diff}}(x) = \frac{c_x}{\pi} \int_{\Omega} V(x, \omega_i) L(\omega_i)(\mathbf{n}_x \cdot \omega_i) d\omega_i, \quad (4)$$

$$I_{\text{spec}}(x, \omega_o) = \int_{\Omega} \rho(\omega_i, \omega_o; c_x, r_x, m_x) L(\omega_i)(\mathbf{n}_x \cdot \omega_i) d\omega_i, \quad (5)$$

where c_x , r_x , m_x are the diffuse albedo map, roughness map and metallic map, respectively. $L(\omega_i)$ is the learned incident radiance from incoming direction ω_i , which we approximate as position-independent. \mathbf{n}_x is the surface normal at x . $V(x, \omega_i)$ is a baked visibility term and $\rho(\omega_i, \omega_o; c, r, m)$ is the microfacet BRDF (Bidirectional Reflectance Distribution Function).

All learned functions presented of position x are implemented as per-Gaussian shading parameters. The light term $L(\omega_i)$ is implemented as an environment map. The diffuse-term integration is approximated as a SH triple product [40], during which $L(\omega_i)$ is projected to SH. The visibility $V(x, \omega_i)$ is pre-computed over a regular grid, with each cell computing

a set of SH coefficients by projecting a splatted opacity cube-map. During training, we compute V once during initialization and keep it frozen throughout the iterations. While we fine-tune geometric parameters, we focus on shading and the visibility changes during training are minimal. As our visibility approximation is low frequency, we opt to drop it from the specular term. The specular component is computed using a split-sum process [41] with pre-filtered environment light, detailed in the supplement material.

B. Training

We split training into four stages: pre-training, specular distillation, diffuse distillation, and joint refinement. Each stage is an independent SGD (Stochastic Gradient Descent) loss-minimization session run for a fixed number of epochs. All learned parameters are shared throughout all stages, though each stage may freeze a subset. All stages share an image loss term \mathcal{L}_{rgb} identical to baseline 3DGS defined on final radiance $I(x, \omega_o)$. For simplicity, we omit the base image loss and focus our discussion on additional, stage-specific loss terms.

a) Stage 1: Pre-training.: In this stage, we fix $\alpha_x = 0$ and only I_{raw} is trained. This stage is identical to the 3DGS baseline. We effectively have:

$$I(x, \omega_o) = I_{\text{raw}}(x, \omega_o). \quad (6)$$

At the end of this stage, the basic Gaussian geometry and raw radiance are close to convergence, though we keep fine-tuning geometry parameters such as Gaussian rotation, scale and opacity in the following stages.

b) Stage 2: Specular distillation.: With the baseline I_{raw} trained, we bump α_x to 0.01 as initialization and fix $m_x = 1$ to start fitting specular effects. We distill the specular term first because it is less ambiguous alongside I_{raw} and takes over faster. The output radiance in this stage is effectively:

$$I(x, \omega_o) = \alpha_x I_{\text{spec}}(x, \omega_o) + (1 - \alpha_x) I_{\text{raw}}(x, \omega_o). \quad (7)$$

Note that $m_x = 1$ completely disables the diffuse term, as indicated by Eq. (3). We follow the normal propagation and color sabotage steps in 3DGS-DR [33] for normal reconstruction. Since the physically-based specular reflection approximates high-frequency view-dependent effects better than the SH-based I_{raw} , the alpha map naturally increases in low-roughness areas with solely an image-fitting loss. At the end of this stage, normal and roughness of specular objects, plus the portion of light environment map visible under specular reflection, are essentially distilled to convergence.

The second row of Fig. 2 illustrates the effect of a converged specular distillation. At this point, the diffuse reflection remains in I_{raw} and the distillation progress on diffuse-only pixels remains close to 0. With the improved specular fitting, the overall model already demonstrates improved novel view synthesis capabilities. However, to further decompose material parameters for relighting, we still need to distill the diffuse term.

Before starting the next stage, we bake the visibility term $V(x, \omega_i)$ using the current Gaussian geometry.

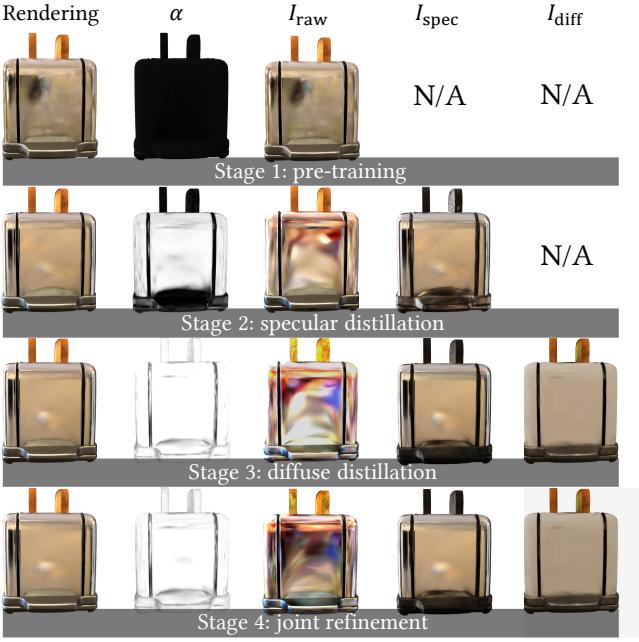


Fig. 2: Visualization of the distillation progress map and different rendering components generated by each optimization stage.

c) Stage 3: Diffuse distillation.: We unfreeze m_x in this stage and the final output radiance matches the full model in Eq. (1). However, unlike the specular stage, the image loss alone is insufficient to differentiate the diffuse reflection from the similarly low-frequency raw SH radiance. As a solution, we introduce an extra loss term \mathcal{L}_α to promote physical modeling:

$$\mathcal{L}_\alpha = \text{MSE}(I_{\text{mask}}, \alpha_x), \quad (8)$$

where I_{mask} is an object mask. I_{mask} is 1 on pixels covered by any object and 0 otherwise. Note that our method does not require per-object masks. We describe how to handle real-world data with no readily-available masks in the supplementary material. With a soft constraint pulling α to 1, explainable portions of raw radiance are distilled into diffuse reflection, and specular reflection adapts to the changed m_x accordingly. At the end of this stage, raw radiance naturally shift to unmodeled light paths such as inter-reflection for high- α pixels.

We follow previous works [10], [23] and add $\mathcal{L}_{\text{light}}$ to penalize coloration and make the light environment map as close to neutral white as possible. This is achieved by minimizing the difference between individual RGB channel values and their mean using the following equation:

$$\mathcal{L}_{\text{light}} = \sum_{i=1}^3 |l_i - \frac{1}{3} \sum_{j=1}^3 l_j|, \quad (9)$$

where l_i and l_j represent the individual RGB components of the light. The final loss function for SGD is $\mathcal{L} = \lambda_1 \mathcal{L}_{\text{rgb}} + \lambda_2 \mathcal{L}_\alpha + \lambda_3 \mathcal{L}_{\text{light}}$, where $\lambda_1 = 1.0$, $\lambda_2 = 0.08$, $\lambda_3 = 0.003$ are tunable meta-parameters.

d) Stage 4: Joint refinement.: In this final stage, we slightly reduce the learning rate of α and fine-tune all other parameters. For real scenes, we also freeze the light L and drop the I_{mask} term from \mathcal{L}_α :

$$\mathcal{L}_\alpha = \text{MSE}(1, \alpha_x), \quad (10)$$

Note that switching to Eq. (10) would distill the entire image, including background pixels.

The real scene-specific designs are motivated by relighting quality, as we prefer background pixels to respond consistently with foreground pixels in presence of light changes. To achieve that, we need to encourage a diffuse fitting on the previously neglected background pixels. However, such pixels are poorly explained by our shading model and we have to anticipate high fitting errors. We therefore freeze the light L to cut off the main gradient path that facilitates cross-pixel error propagation, which limits background fitting errors to their own pixels.

IV. RESULTS AND EVALUATION

We conduct comprehensive experiments on a workstation with an i7-13700KF CPU, 32GB memory and an NVIDIA RTX 4090 GPU, to demonstrate the effectiveness and efficiency of our approach. We also perform ablation studies to validate our core progressive distillation design.

a) Dataset.: We conduct evaluation on several datasets encompassing scenes with various materials, including three synthetic datasets, Shiny Blender [1], Glossy Synthetic [7] and TensoIR Synthetic [43], and one real captured dataset, Stanford ORB [42]. Shiny Blender and Glossy Synthetic mainly contains objects with low roughness while TensoIR Synthetic and Stanford ORB contains diffuse-dominant objects. All four datasets provide per-image object masks.

b) Implementation details.: Before the specular distillation stage, we initialize the roughness r and all channels of albedo c to 0.99. We apply a sigmoid activation to material properties c , r and m to restrict their range to between 0 and 1. We apply an exponential activation to the light environment map to better reflect its HDR (High Dynamic Range) nature. I_{phy} is tone-mapped by the ACES Filmic tone mapping curve [44] and converted to the sRGB color space before image loss computation. The normal of each Gaussian is defined as its shortest axis facing the camera, which follows 3DGS-DR [33]. The environment light is represented as a $128 \times 128 \times 6$ cube map. The full rendering is based on a deferred shading pipeline, including two passes. In the first pass, maps of c , r , m , n , d , α , and I_{raw} are splatted. Here d is the depth map, computed by splatting a per-Gaussian depth, defined as the distance from its center to the camera. The positions used for querying the visibility term for each pixel are computed from the blended d . In the second pass, the output radiance is computed from the aforementioned maps and pre-filtered environment light as described in Sec. III-A.

c) Baselines and metrics.: We compare our method against the following baselines: **3DGS**: original 3D Gaussian Splatting [4]; **GShader** [8]: a method that shades each Gaussian with a reflection-aware shader. **RGS** [31]: a method that

TABLE I: Dataset-averaged image quality comparison for novel views synthesis.

Methods	Shiny Blender			Glossy Synthetic			TensoIR Synthetic			Stanford ORB		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Ref-NeRF	33.13	0.961	0.080	27.50	0.927	0.100	38.42	0.974	0.026	35.43	0.983	0.028
3DGS	30.36	0.947	0.084	26.50	0.917	0.092	39.07	0.988	0.016	36.34	0.990	0.013
GShader	31.97	0.958	0.067	27.54	0.922	0.087	38.78	0.980	0.021	36.86	0.992	0.153
GS-IR	26.63	0.878	0.150	23.79	0.827	0.161	34.98	0.962	0.043	34.21	0.968	0.038
RGS	29.19	0.940	0.100	25.64	0.911	0.099	39.01	0.981	0.017	35.70	0.988	0.016
NDR	28.49	0.930	0.116	25.69	0.900	0.113	30.11	0.941	0.058	33.15	0.983	0.022
NDRMC	31.24	0.954	0.092	28.62	0.947	0.059	30.03	0.949	0.083	36.44	0.982	0.028
TensoIR	28.01	0.905	0.131	25.88	0.904	0.112	35.09	0.976	0.04	34.66	0.980	0.020
ENVIDR	33.46	0.968	0.046	29.58	0.952	0.056	30.76	0.943	0.051	34.78	0.981	0.022
3DGS-DR	34.09	0.971	0.057	30.14	0.953	0.058	39.24	0.988	0.017	36.33	0.990	0.013
Ours	34.43	0.973	0.054	30.32	0.958	0.051	39.11	0.986	0.017	36.50	0.990	0.012

TABLE II: Normal and light reconstruction quality (evaluated by MAE° and LPIPS respectively) comparisons on the Shiny Blender Dataset.

	GShader	GS-IR	NDR	ENVIDR	3DGS-DR	Ours
MAE° ↓	23.31	31.24	17.02	4.618	4.871	4.621
LPIPS ↓	0.621	0.687	0.636	0.615	0.511	0.521

performs ray tracing to render Gaussians. **GS-IR** [9]: a method that uses deferred shading and physically-based rendering model similar to ours. **3DGS-DR** [33]: a method focusing on mirror reflection rendering with 3DGS. **NDR** [23]: a method that extracts meshes from optimized SDFs (Signed Distance Functions) and render meshes using image based lighting. **NDRMC** [17]: a method that extends NDR by replacing the image based renderer with Monte Carlo renderer and using a differentiable denoiser to reduce variance. **ENVIDR** [45]: a SDF-based method using neural rendering for inverse rendering. **TensoIR** [43]: a NeRF-based method that using MLPs to cache visibility and indirect lighting. **Ref-NeRF** [1]: a NeRF-based method that improves the quality of novel view synthesis of reflective objects.

For the task of novel view synthesis, we compare with all the above-mentioned methods. For relighting tasks, 3DGS, 3DGS-DR and Ref-NeRF lack or only have limited relighting capability, so we only compare with GShader, GS-IR, RGS, NDR, NDRMC, ENVIDR and TensoIR. We present quantitative results measured with three standard metrics: PSNR, SSIM and LPIPS. The evaluation results of NDRMC [17] are rendered using 2000 samples per pixel (spp) for novel views and 64 spp with denoiser for relighting, consistent with their paper.

We note that previous works [26], [42] align each predicted image to ground-truth via channel-wise scale factors for relighting comparison, which aims to eliminate the inherent scale ambiguity in inverse rendering problems. This scheme may lead to different scale factors for different views of the same scene, which is unreasonable and impractical in real rendering applications. A more reasonable way is to calculate and use the same scale factor for the same scene. Therefore, we propose to align the ground-truth environment map to the predicted environment map via channel-wise scale factors, and use these factors to scale the new environment map for relighting. All

quantitative and qualitative results related to relighting below are based on this adjustment. We also report quantitative results of relighting using different scaling schemes in the supplementary material. In addition, we use Mean Angular Error in degrees (MAE°) to evaluate the accuracy of normal reconstruction. We evaluate environment map reconstruction accuracy with LPIPS to mitigate the inherent ambiguity.

A. Comparisons with baselines

a) *Novel view synthesis*.: Table I presents the quantitative comparison results on three datasets. Our method demonstrates a clear advantage in terms of image quality on synthetic datasets and also shows comparable results on real datasets. For diffuse-dominant scenes, the differences in image quality among various methods are quite small. The visual comparisons on synthetic and real datasets are shown in Fig. 3, and the supplementary material provides results on more scenes (mainly diffuse objects).

The results from 3DGS-DR [33] and ENVIDR [45] demonstrate good quality in most scenes. However, 3DGS-DR only models mirror reflection and its quality decreases when a scene contains a wide range of roughness, such as the excessively sharp reflections in the bands of *potion*. ENVIDR [45] is based on SDFs and can over-smooth surface normal, as shown in *luyu*. It can also produce regularly-patterned noise as in *toaster*. NDR [23] may generate incorrect geometry, leading to incorrect reflections, as in *car*. It also performs poorly for specular objects with high diffuse albedo, as in *cactus*. GShader [8] and RGS [31], both based on per-Gaussian shading, tend to blur specular reflections. Our method produces satisfactory results for objects with various materials and stays closer to the ground truth.

b) *Decomposition*.: We visualize and compare the reconstruction results of normal, albedo, and lighting, and showcase our fully decomposition results.

Correct normals are crucial for relighting. We show the qualitative comparison in Fig. 4 and dataset-averaged mean angular error of estimated normal on the Shiny Blender Dataset in Table II. The results from 3DGS-DR [33] show some noise in *teapot*. ENVIDR [45] produces good normals, but over-smooths geometry details at normal discontinuities due to its smoothness SDF prior. This manifested as bumps

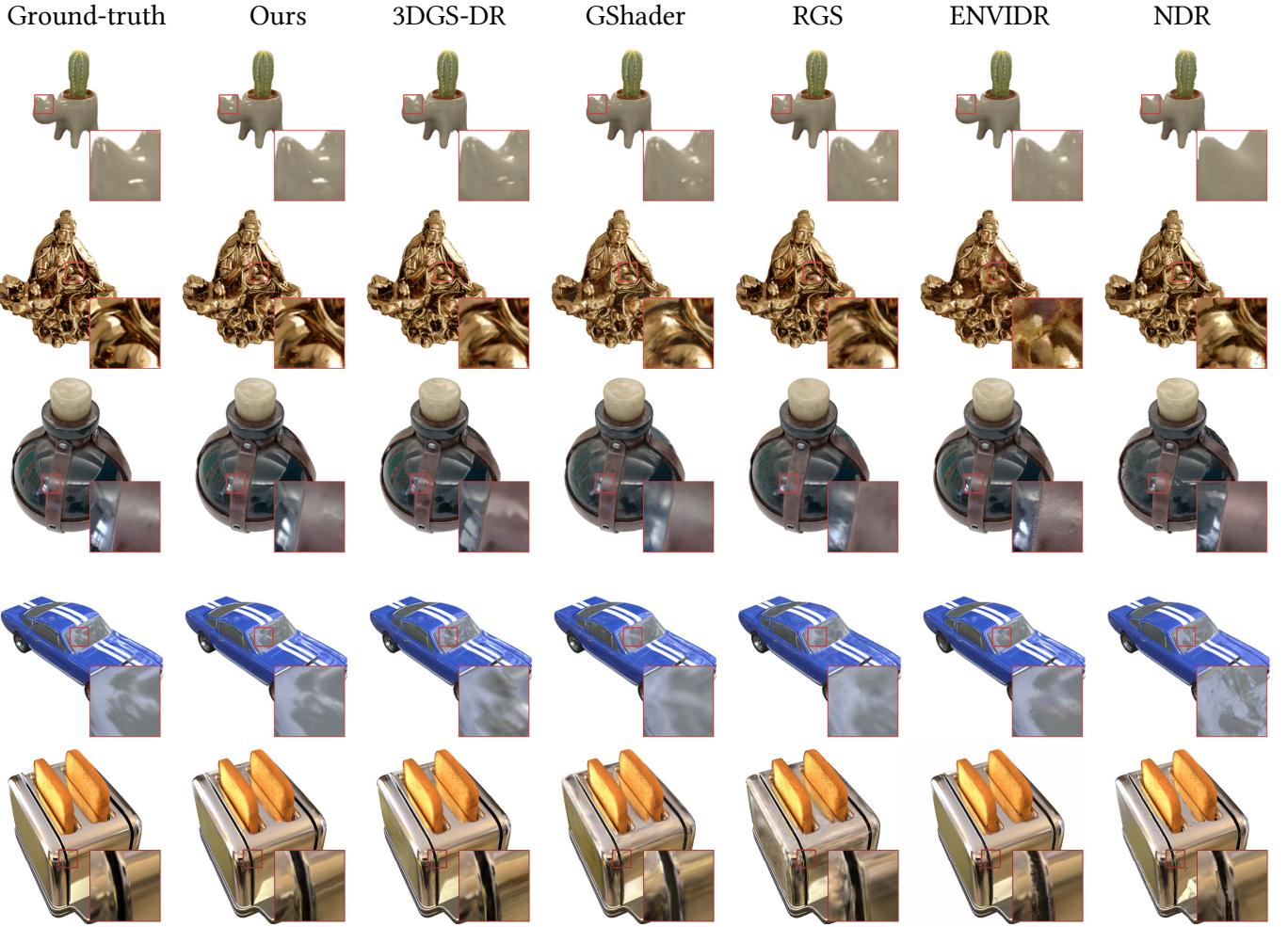


Fig. 3: Qualitative comparisons of novel view synthesis results. From top to bottom: cactus (from Stanford ORB [42]), lulu and potion (from Glossy Synthetic [7]), car and toaster (from Shiny Blender [1]).

in the concave parts at the top of *potion*, just below the bottle neck. Other methods generally produced noisy normals, especially for reflective surfaces. Our method can achieve normal reconstruction quality comparable with ENVIDR, while retaining sharp boundaries.

Albedo, by definition, is the base color of an object and should not contain any reflection. Due to differences in rendering models among different methods, as well as the inherent ambiguity between light and material, we only conduct qualitative comparison in Fig. 6. We mainly focus on whether the fitted albedo has reflections removed, rather than how closely it matches the reference image. Based on this criterion, ENVIDR [45], NDR [23], and our method all produce reasonable albedo maps, while other methods fail to disambiguate albedo from reflections, especially on smooth surfaces.

Correctly separating lighting is essential for distilling the correct physical model. The qualitative and quantitative comparison results are shown in Fig. 5 and Table II. To compensate for the fundamental light-albedo ambiguity, we equalize the total energy of each environment map pair before comparison, and use the less scaling-sensitive LPIPS metric for evaluation. As shown in Fig. 5, *helmet*, *teapot* and *toaster* exhibit dis-

tinctive specular reflections, and *gnome* is diffuse dominant. 3DGS-DR [33] predicts accurate lights for specular objects, but fails in *gnome* because the method is mainly designed for mirror reflection. GShader [8] can only reconstruct a few texels for each Gaussian, leading to a noisy image with many holes. SDF (Signed Distance Function) based methods (ENVIDR [45] and NDR [23]), struggle with light decomposition accuracy, leading to messy patterns in the lighting estimation. Our method reconstructs environment maps with almost full directional coverage for reflective objects, and also accurately locates the primary light sources for diffuse-dominant scenes.

We show more decomposition results in Fig. 7. Our method accurately identifies surfaces with different roughness as well as metallic and non-metallic reflections. The distillation progress map α is close to 1 in most pixels, but slightly lower in pixels where the physical model does not fit well, such as between the slices of bread in *toaster* and the spherical handle of *tbell*. The raw radiance I_{raw} compensates for indirect light, which improves the quality in novel view synthesis.

c) *Relighting*.: Relighting is our primary goal. Without resolving the ambiguities between material and lighting or reconstructing accurate geometry, relighting will exhibit significant errors. We show the quantitative and qualitative results

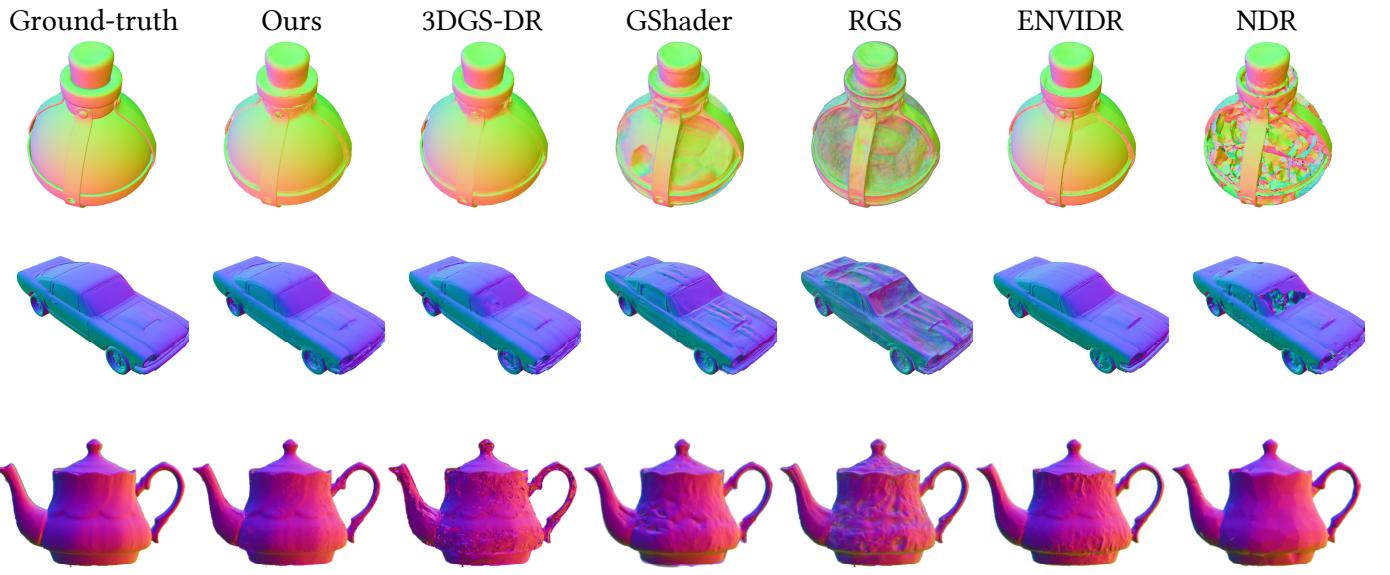


Fig. 4: Qualitative comparisons of normal estimated by different methods. From top to bottom: potion (from Glossy Synthetic [7]), car (from Shiny Blender [1]) and teapot (from Stanford ORB [42]).

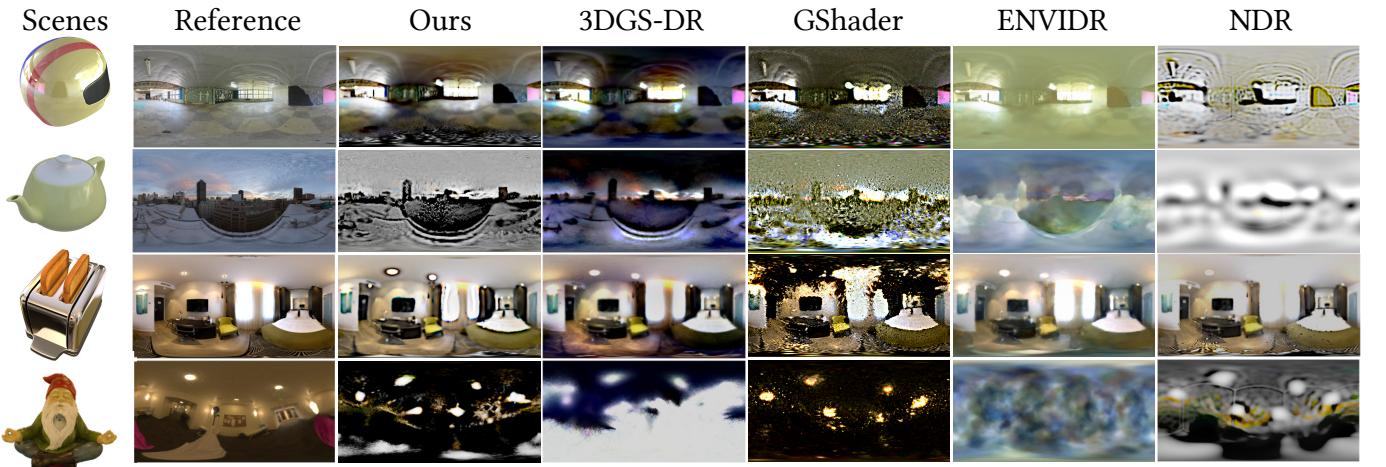


Fig. 5: Qualitative comparisons of environment maps estimated by different methods. From top to bottom: helmet, teapot and toaster (from Shiny Blender [1]), gnome (from Stanford ORB [42]).

in Table III and Fig. 8. Our method demonstrates a significant advantage across all datasets. The results from GShader [8], RGS [31], and GS-IR [9] differ significantly from ground truth due to their inability to produce correct geometry and resolve the ambiguity between albedo and reflection. As shown in *potion* of Fig. 8, these methods erroneously retain reflections from the original lighting in relighted images. For diffuse-dominant objects, the ambiguity of geometry-material-lighting is smaller, making it a simpler task. As demonstrated in *gnome*, almost all methods produce plausible results. The results from NDR [23] show the impact of unstable geometry on relighting. As shown in *helmet*, NDR has demonstrated mostly-correct reflections except for where the estimated geometry ends up bumpy. ENVIDR [45] shows robust relighting results. However, due to its use of MLPs for lighting representation,

one has to synthesize a considerable amount of training data for each relighting environment. The overall process consumes more than an hour per environment map. Additionally, its reflections tend to miss details. Thanks to our rendering model and carefully designed training process, we are able to resolve most of the ambiguities and demonstrate high-quality relighting results.

d) Efficiency.: Table IV lists the dataset-averaged training time and rendering frame rate of all tested methods. Frame rate values are computed as the reciprocal of averaged frame render time to reduce the impact of unfairly large numbers on simple scenes. Additionally, we list the consumed time for GS-IR [9] and our method to bake visibility (7 minutes for GS-IR and 8 minutes for ours) in the caption. NeRF-based methods, including Ref-NeRF [1], TensoIR [43] and



Fig. 6: Qualitative comparisons of albedo estimated by different methods. Due to differences in rendering models among different methods, as well as the inherent ambiguity of light and material, the comparison is only for reference. From top to bottom:potion (from Glossy Synthetic [7]), toaster (from Shiny Blender [1]) and teapot (from Stanford ORB [42]).

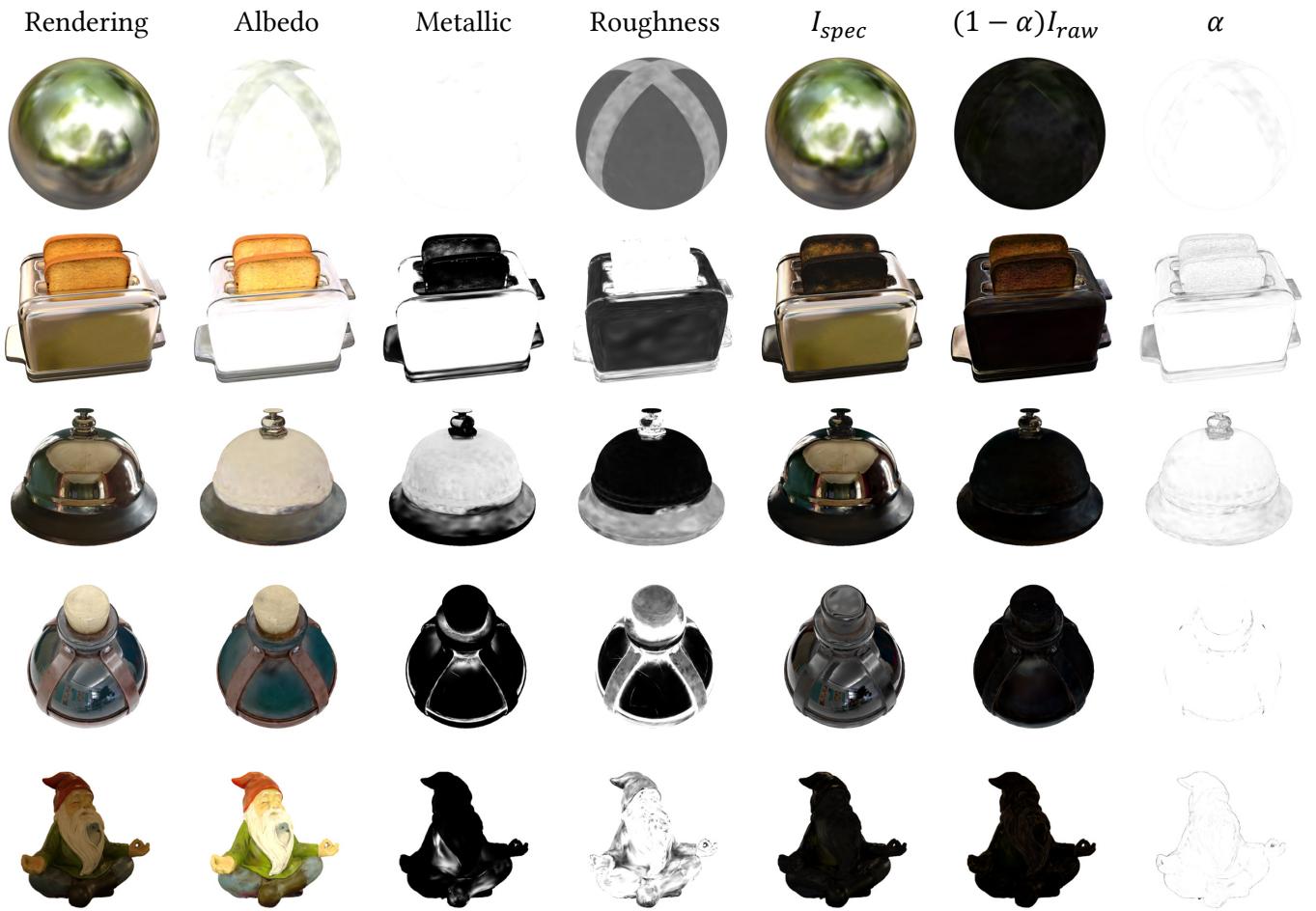


Fig. 7: Decomposition results of our method. We display $(1 - \alpha)I_{\text{raw}}$ instead of I_{raw} to help illustrate its contribution to final results. From top to bottom: ball and toaster (from Shiny Blender [1]), tbell and potion (from Glossy Synthetic [7]), gnome (from Stanford ORB [42]).

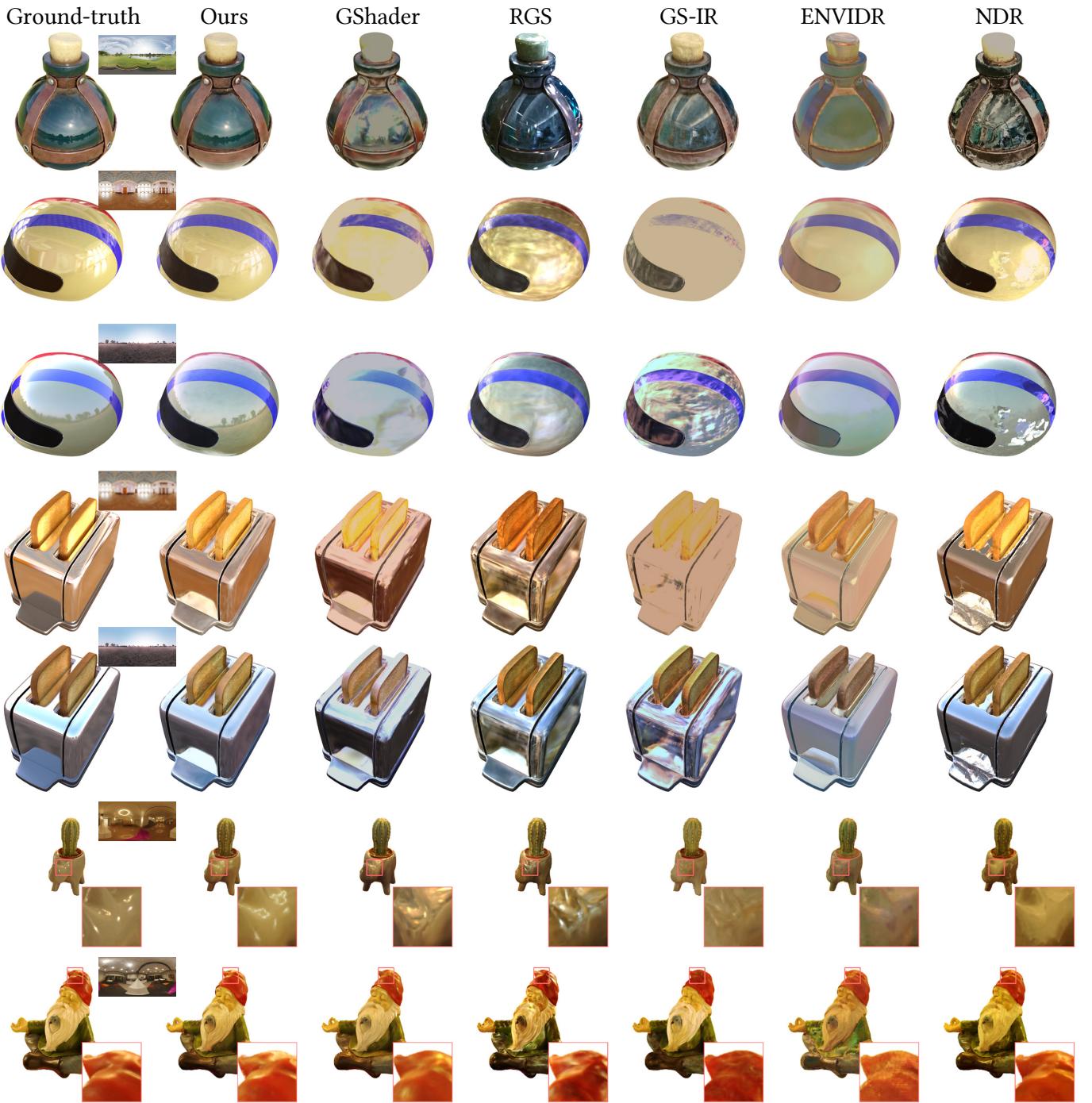


Fig. 8: Qualitative comparison of relighting results predicted by different methods. We scale the light for each method to cancel out inherent light-material ambiguity. From top to bottom: potion (from Glossy Synthetic [7]), helmet and toaster (from Shiny Blender [1]. Each takes up two rows), cactus and gnome (from Stanford ORB [42]).

ENVIDR [45] require more than one hour training time, and do not support real-time rendering. Gaussian splatting based methods, including RGS [31], GShader [8], GS-IR [9], 3DGSDR [33], and vanilla 3DGS [4] can be trained to convergence within one hour, and are able to render with high frame rates except for RGS, which is based on the more expensive ray tracing. Similarly, NDRMC [17] is based on differentiable Monte Carlo renderer with denoiser. Although accurate path tracing may achieve higher quality, it comes at the cost of

longer training time and reduced real-time performance. Our method requires moderate training time and shows high real-time performance among relighting-supporting methods. I_{raw} compensates unmodeled light paths, enabling our rendering model based on image based lighting to achieve quality comparable to path tracing methods on novel view synthesis.

TABLE III: Quantitative comparison of relighting results using PSNR, SSIM, and LPIPS metrics. Results are averaged over 20 different viewpoints with four different environment maps on Shiny Blender and Glossy Synthetic datasets. For the Stanford ORB dataset, relighting results are evaluated on the provided 20 image-envmap pairs. The environment maps are scaled to offset the inherent material-light ambiguity.

Methods	Shiny Blender			Glossy Synthetic			TensoIR Synthetic			Stanford ORB		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
GShader	20.26	0.881	0.130	21.66	0.870	0.120	25.61	0.927	0.066	29.72	0.957	0.038
GS-IR	19.37	0.862	0.171	19.80	0.808	0.160	25.52	0.904	0.084	28.05	0.947	0.046
RGS	17.75	0.841	0.151	17.08	0.807	0.155	25.38	0.916	0.082	26.49	0.939	0.046
NDR	21.81	0.891	0.133	19.82	0.826	0.153	26.59	0.906	0.091	29.07	0.953	0.041
NDRMC	21.92	0.904	0.143	21.42	0.872	0.134	26.86	0.907	0.113	30.88	0.962	0.040
TensoIR	19.01	0.862	0.171	18.71	0.812	0.156	28.23	0.938	0.088	30.01	0.957	0.036
ENVIDR	20.90	0.905	0.102	19.13	0.859	0.124	25.58	0.924	0.087	26.75	0.942	0.049
Ours	23.68	0.925	0.083	23.10	0.895	0.091	29.69	0.941	0.057	30.92	0.960	0.033

TABLE IV: Averaged training time and rendering frame rates. Time for baking visibility of GS-IR and our method are 7 and 8 minutes, respectively. The FPS of NDRMC [17] is evaluated by Blender Cycles with 64 samples per pixel (spp) and denoising.

	Ref-NeRF	NDRMC	TensoIR	ENVIDR	RGS	NDR	GShader	GS-IR	3DGS-DR	3DGS	Ours
Training time	19h	4.2h	3.9h	3.2h	41min	78min	60min	19+7 min	13min	6min	39+8 min
FPS	0.05	6 (64 spp)	4	3	4	32	34	182	281	294	221

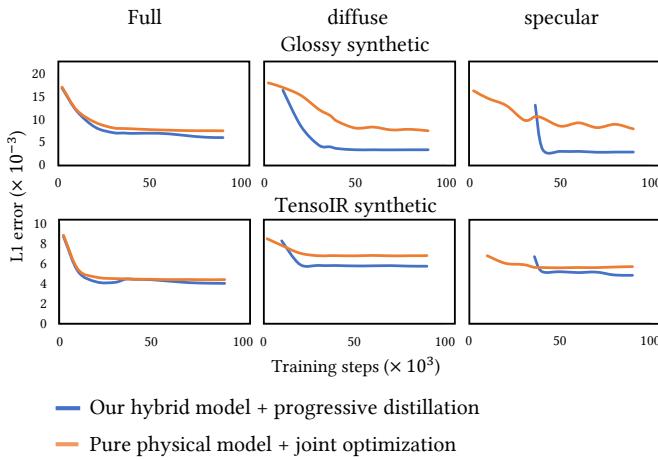


Fig. 9: L1 errors in the full rendering, diffuse and specular components as the training step increases. The first row is evaluated on Glossy synthetic dataset [7] and the second row is evaluated on TensoIR synthetic dataset [43].

TABLE V: Ablation study. Dataset-averaged image quality with selectively disabled algorithm components.

Ablations	Novel view synthesis			Relighting		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
w/o α	32.57	0.958	0.089	22.17	0.878	0.112
w/o sep.	33.04	0.965	0.064	23.09	0.909	0.095
w/o \mathcal{L}_α	34.45	0.972	0.054	19.88	0.861	0.169
w/o I_{raw}	32.97	0.964	0.066	23.41	0.912	0.087
Ours	34.43	0.973	0.054	23.68	0.925	0.083

B. Ablation Study

We conduct ablation studies to validate our core designs: the use of the distillation progress map and the separation of specular and diffuse distillation. The quantitative and qualitative

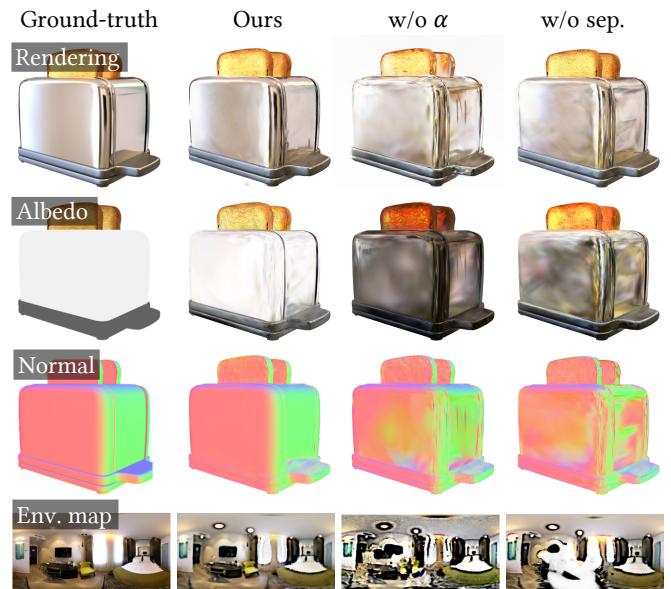


Fig. 10: Ablation studies. w/o α means directly using $I_{\text{raw}} + I_{\text{phy}}$ as the rendering model. w/o sep. means no separation of the specular and diffuse distillation.

results are shown in Table V and Fig. 10, which demonstrate the significance of our contribution.

a) *Joint optimization from scratch vs. progressive distillation.*: To intuitively demonstrate the benefits of progressive distillation, we analyze the distillation process through visualizing the reconstruction quality of the diffuse/specular components and the full rendering during optimization. As shown in Fig. 9, compared to jointly optimizing a pure physical model ($I(x, \omega_o) = I_{\text{phy}}(x, \omega_o)$) from scratch (referred to as JNT method), our method has similar specular and full rendering errors in early iterations but converges fast with

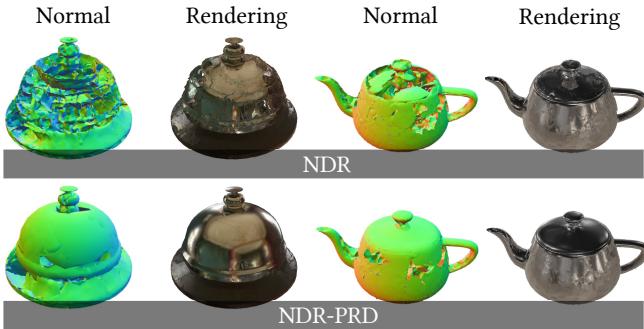


Fig. 11: Qualitative comparison of the original NDR [23] and modified NDR incorporating progressive radiance distillation (denoted as NDR-PRD). Our carefully designed optimization contributes to more accurate normal reconstruction for reflective objects.

TABLE VI: Quantitative comparison of the original NDR [23] and modified NDR (denoted as NDR-PRD) that uses our optimization strategy as described in Sec. III-B. Results are evaluated with PSNR metric.

	Shiny Blender		Glossy Synthetic		Stanford ORB	
	NDR	NDR-PRD	NDR	NDR-PRD	NDR	NDR-PRD
Novel view	28.49	30.39	25.69	26.49	33.15	33.05
Relighting	21.81	22.09	19.82	20.09	29.07	28.43

smaller errors, which demonstrates that our distillation strategy effectively reduce ambiguities between different components. Consequently, the error in diffuse/specular components decreases rapidly after a few iterations, rather than slowly decreasing with fluctuations as seen in the JNT method. Moreover, although the JNT method shows significantly larger errors in the diffuse/specular components compared to our method, the error in full rendering is not as pronounced. This indicates that the diffuse component is contaminated with specular reflections, which will lead to a decrease in relighting quality, similar to the *potion* results of RGS and GS-IR in Fig. 8.

b) *Impact of the distillation progress map.*: We carefully designed a training process without using the distillation progress map. Specifically, the training still begins with a pre-training stage, during which I_{raw} and basic geometry parameters are optimized. In the second stage (and also the subsequent stages), the output radiance is:

$$I(x, \omega_o) = I_{\text{phy}}(x, \omega_o) + I_{\text{raw}}(x, \omega_o). \quad (11)$$

Normal propagation and color sabotage are still performed in this stage to help geometry reconstruction. In the third stage, we use a loss term to encourage I_{raw} to be zero, pushing the model to be physical. In the fourth stage, we drop that loss term and decrease the learning rate of the I_{raw} to fine tune everything. We believe this experiment design maximizes the capability of novel view synthesis and relighting in the absence of an explicit distillation progress. We refer to this experiment design as *w/o* α .

Without the guidance of the distillation progress map, the complicated physical model and light need to be optimized

from the messy starting normal, during which each pixel will back-propagate a considerable amount of gradient to lighting and material attributes. This greatly increases the risk of getting stuck in local optima. As illustrated in Fig. 10, the environment map ends up containing several copies of the blackboard and the chair, because the optimization tends to copy reflected objects to explain incorrect normal. Our method initializes the distillation progress map with small values, which diverts most of the bad early gradient to the raw radiance, alleviating its negative impact on parameters to be estimated. When the fitting loss becomes favorable for the physical model, the distillation progress value starts to rise, simultaneously suppressing the contribution of raw radiance, which further reduces ambiguity and promotes the raw-to-physical transition. With the guidance of the distillation progress map, optimization can robustly proceed until specular reflections are fully distilled from raw radiance at the end of the second stage. In the third diffuse stage, we just *move* the remaining parts from being represented by raw radiance to being represented by diffuse reflection.

c) *Impact of the separation of specular and diffuse distillation.*: Here we only keep the non-diffuse Stages 1, 2 and 4 in the original design, and replace Eq. (7) in Stage 2 with the full rendering model Eq. (1). We refer to this experiment design as *w/o sep*. As shown in Fig. 10, the albedo map ends up entangled with random fragments of reflection, and the environment map missed some details. Although specular reflection and diffuse reflection can be theoretically differentiated using their angular frequency in the view-domain, the diffuse term propagates significant gradient to light and geometry, unlike the relatively independent raw radiance. Therefore, reflections that cannot be explained by low frequency effects may get forced into the diffuse term through erroneous normals, albedo, and lighting adjustments, disrupting optimization. Our method distills specular reflection first, and the converged results serve as a soft pinning for the lighting and geometry parameters, thereby avoiding the risk of a misguided optimizer using the diffuse term to explain specular reflection.

d) *Impact of \mathcal{L}_α .*: We use an extra loss term \mathcal{L}_α to promote physical modeling. As shown in Table V, the \mathcal{L}_α term has a negligible impact on novel view synthesis. Being introduced after a converged specular stage, it primarily drives the transition of low-frequency representations from raw radiance to diffuse. Without \mathcal{L}_α , raw radiance lingers on diffuse objects yet cannot respond to changes in lighting, leading to a significant decline in relighting quality.

e) *Impact of I_{raw} .*: We retain I_{raw} in the final model to compensate for unfactorizable light paths and brake the distillation progress when it would lead to view interpolation regressions, which preserves the novel view quality not inferior to vanilla 3DGS. I_{raw} can be regarded as a self-emissive term, which can be easily supported in off-the-shelf engines with shaders. Furthermore, as shown in Table V, it has low weights in general with negligible impacts on relighting plausibility.

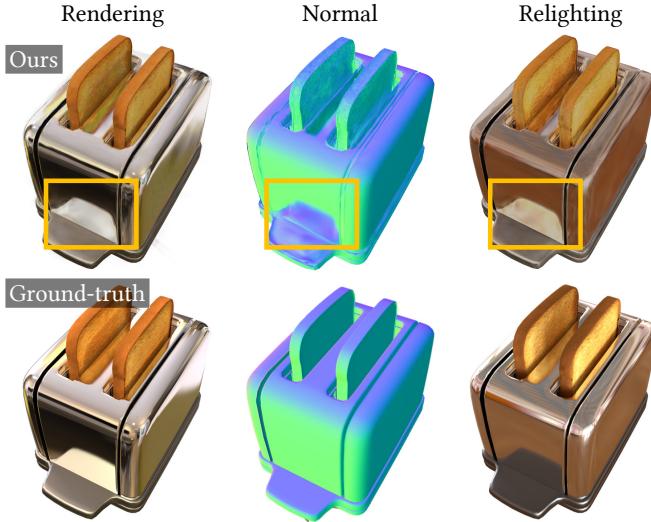


Fig. 12: Limitations. Inter-reflections mislead normal reconstruction, resulting in incorrect relighting.

C. Discussion and Limitations

Note that the idea of progressive radiance distillation is not limited to 3D Gaussian splatting, and it is possible to generalize it to NeRF representations. To explore this possibility, we adapted the rendering model and optimization strategy of NDR [23], a representative mesh-based inverse rendering method, to incorporate our radiance distillation. NDR extracts meshes from SDF and uses the same PBR model as ours for inverse rendering. The output radiance of NDR is $I(x, \omega_o) = I_{AO}I_{phy}$ throughout its entire training process, where I_{AO} is a learned ambient occlusion texture. We remove I_{AO} and introduce I_{raw} and α as new learned textures, and employ Eq. (1) as the rendering model. We modify NDR to follow the same four stages as our method, with the output radiance of each stage consistent with Sec. III-B. We denote the modified NDR as NDR-PRD. The detailed implementation of NDR-PRD is described in the supplementary material.

The dataset-averaged PSNR metric of novel view synthesis and relighting are shown in Table VI. Quantitatively, NDR-PRD achieves slightly better results on the Shiny Blender and Glossy Synthetic dataset, but performs slightly worse on the Stanford ORB dataset. This could be attributed to the discontinuities in mesh based optimization, leading to a more challenging normal reconstruction, thus offsetting the benefits provided by our method. However, for a few scenes with specular reflections, NDR-PRD produces significantly better results, as shown in Fig. 11. We provide more experimental results in the supplementary material.

The primary limitation of our method is the inability to handle indirect lighting explicitly, which could result in incorrect normals on surfaces with specular inter-reflections, and indirect light represented by raw radiance retaining the original lighting conditions when relighting, as shown in Fig. 12. Besides, we use 3-order SH to compress the visibility at grid points, which is more suitable for soft shadows on diffuse surfaces. As shown in Fig. 6, shadows may be baked into the albedo. Performing ray tracing for inverse rendering like

RGS [31] may solve these problems but significantly increase the rendering time. Another limitation arises from deferred shading, which affects the rendering quality of transparent or translucent objects like glass.

V. CONCLUSION

We have presented an inverse rendering method based on 3D Gaussian splatting, featuring a rendering model combining the radiance field rendering and physically-based rendering with a distillation progress map, and a stage-wise distillation strategy alleviating the inherent ambiguity of light-material decomposition. Extensive experiments demonstrate our competitive visual quality and real-time performance in both novel view synthesis and relighting. Our progressive radiance distillation also shows the potential for generalization to mesh-based inverse rendering. It will be interesting to fully explore the impact of combining our method with existing approaches.

APPENDIX A BRDF MODEL

For I_{spec} , the specular component of the physically-based term I_{phy} , we adopt the Cook-Torrance microfacet specular shading model:

$$\rho(\omega_i, \omega_o) = \frac{DFG}{4(\omega_o \cdot \mathbf{n})(\omega_i \cdot \mathbf{n})}, \quad (12)$$

where D , G and F are respectively the GGX [46] normal distribution function (NDF), the geometric attenuation function and the approximated Fresnel term. Their specific expressions are as follows:

$$F = F_0 + (1 - F_0)(1 - (\mathbf{h} \cdot \omega_o))^5. \quad (13)$$

$$G(\mathbf{n}, \omega_o, \omega_i, k) = G_{sub}(\mathbf{n}, \omega_o, k)G_{sub}(\mathbf{n}, \omega_i, k), \quad (14)$$

$$G_{sub}(\mathbf{n}, \mathbf{v}, k) = \frac{\mathbf{n} \cdot \mathbf{v}}{(\mathbf{n} \cdot \mathbf{v})(1 - k) + k}, \quad (15)$$

$$D(\mathbf{n}, \mathbf{h}, a) = \frac{a^2}{\pi((\mathbf{n} \cdot \mathbf{h})^2(a^2 - 1) + 1)^2}, \quad (16)$$

where \mathbf{h} is the half-way vector, a unit vector in the direction of $\omega_o + \omega_i$. $a = r^2$, $k = r^4/2$ are scalars derived from the roughness r . and F_0 is the basic reflection ratio:

$$F_0 = m * \mathbf{c} + (1 - m) * 0.04, \quad (17)$$

where m is the metallic map and \mathbf{c} is the albedo.

We adopt the split-sum [41] system to approximate the I_{spec} as:

$$I_{spec}(x, \omega_o) \approx \int_{\Omega} L(\omega_i)D(\omega_i, \omega_r, r^2)d\omega_i \cdot \int_{\Omega} \frac{DFG}{4(\omega_o \cdot \mathbf{n})}d\omega_i, \quad (18)$$

where ω_r is the reflection direction $2(\omega_o \cdot \mathbf{n})\mathbf{n} - \omega_o$. The first part of the convolution is referred to as the pre-filtered environment maps, which are implemented as a mip-map pyramid, using lower resolution for higher roughness. The second part is the BRDF integration map, which is independent of lighting and can be precomputed into a lookup table. After

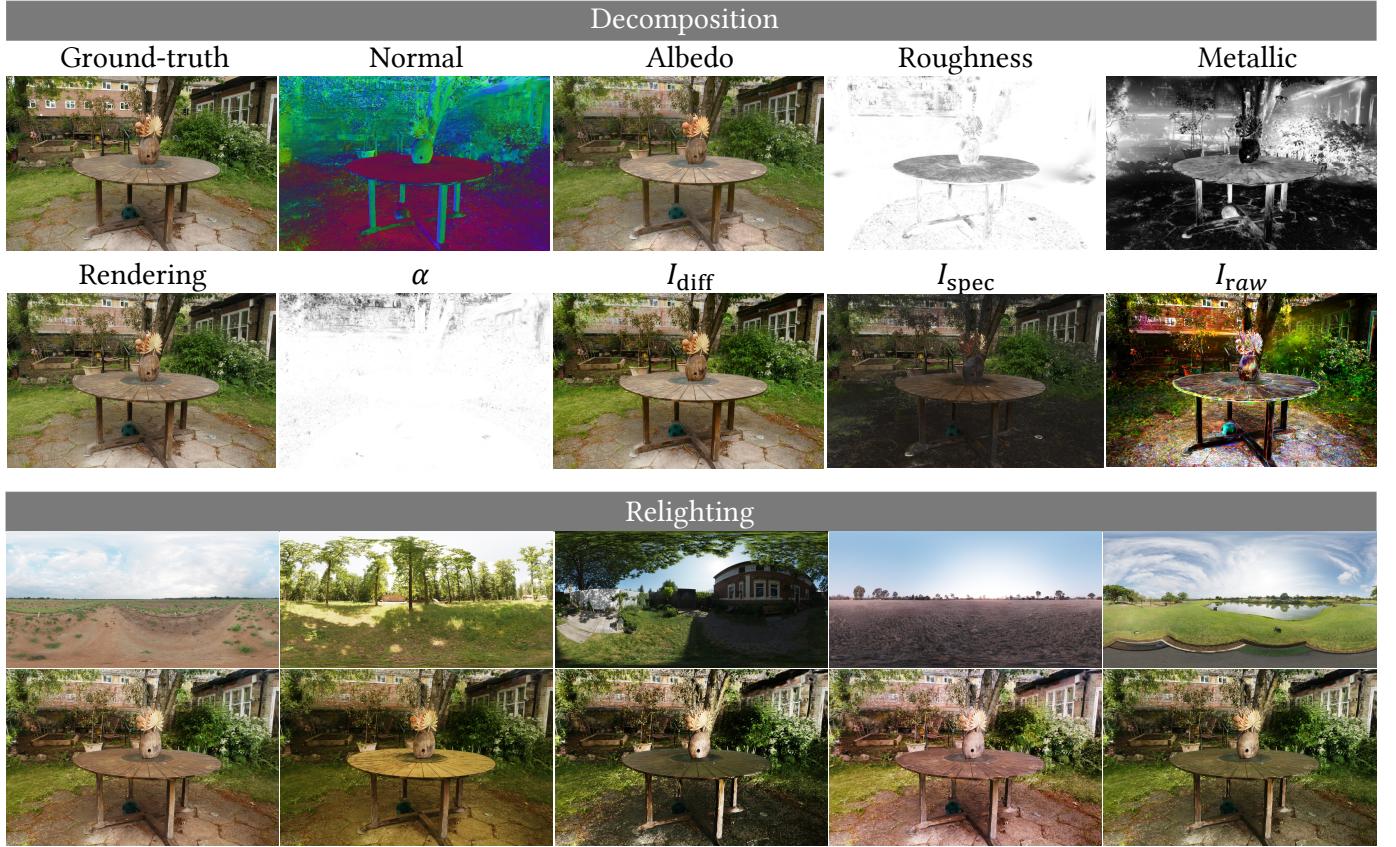


Fig. 13: Decomposition and relighting results of garden.

precomputation, real-time physically-based rendering can be achieved, and the entire process is differentiable.

In the specular distillation stage, we fix $m = 1$, which simplifies Eq. (17) to $F_0 = \mathbf{c}$. Therefore, the albedo \mathbf{c} at this stage actually serves as the basic reflection ratio, and it will be repurposed as the actual albedo in the diffuse distillation stage.

For the diffuse part I_{diff} , we implement the visibility term as a shadow field [47]. The final rendering step starts by projecting both the environment light and the visibility to order-3 SH basis:

$$L(\omega_i) = \sum_{l=0}^2 \sum_{m=-l}^l l_l^m Y_l^m(\omega_i), \quad (19)$$

$$V(x, \omega_i) = \sum_{l=0}^2 \sum_{m=-l}^l v_{l,x}^m Y_l^m(\omega_i), \quad (20)$$

where $Y_l^m(\omega_i)$ is the spherical harmonics basis, l_l^m and $v_{l,x}^m$ are the corresponding coefficients. We perform SH triple product [40] to project $V(x, \omega_i)L(\omega_i)$ back to the order-3 SH basis:

$$V(x, \omega_i)L(\omega_i) = \sum_{l=0}^2 \sum_{m=-l}^l p_{l,x}^m Y_l^m(\omega_i), \quad (21)$$

where $p_{l,x}$ is the a tensor product of l_l^m , $v_{l,x}^m$ and the triple product constant tensor C . Dropping x for clarity and merging l, m pairs into linear indices i, j and k , we get:

$$p_i = \sum_j \sum_k C_{ijk} l_j v_k, \quad (22)$$

$$C_{ijk} = \int_{\Omega} Y_i(\omega_i) Y_j(\omega_i) Y_k(\omega_i) d\omega_i. \quad (23)$$

As explained by [48], the remaining cosine term $(\mathbf{n} \cdot \omega_i)$ and the hemi-spherical integration domain are azimuthally symmetric and the corresponding order-3 SH coefficients ρ_l^m can be computed symbolically using Zonal Harmonics (ZH) rotation. The irradiance $\int_{\Omega} V(x, \omega_i) L(\omega_i) (\mathbf{n}_x \cdot \omega_i) d\omega_i$ is thus a dot product between ρ_l^m and k_l^m . In conclusion, we approximate I_{diff} as:

$$I_{\text{diff}}(x) \approx \frac{\mathbf{c}_x}{\pi} \sum_{l=0}^2 \sum_{m=-l}^l p_l^m \rho_l^m, \quad (24)$$

or in the expanded triple-product form:

$$I_{\text{diff}}(x) \approx \frac{\mathbf{c}_x}{\pi} \sum_i \sum_j \sum_k C_{ijk} \rho_i l_j v_k, \quad (25)$$

APPENDIX B VISIBILITY BAKING

To obtain the SH visibility term, we must first build a shadow field. We implement it as a regular grid ($120 \times 120 \times$

TABLE VII: Quantitative comparison of relighting results using PSNR, SSIM, and LPIPS metrics. Environment map scaling is disabled to better compare the absolute decomposition accuracy.

Methods	Shiny Blender			Glossy Synthetic			TensoIR Synthetic			Stanford ORB		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
GShader	16.22	0.868	0.152	16.32	0.838	0.135	22.90	0.922	0.067	26.23	0.952	0.043
GS-IR	16.77	0.849	0.185	19.36	0.802	0.162	22.52	0.888	0.094	25.01	0.935	0.050
RGS	14.08	0.831	0.160	14.70	0.823	0.152	20.88	0.906	0.094	22.01	0.903	0.060
NDR	21.73	0.893	0.137	16.97	0.809	0.162	18.68	0.872	0.113	26.69	0.949	0.044
NDRMC	21.42	0.872	0.155	18.46	0.831	0.147	21.47	0.913	0.108	26.64	0.941	0.046
TensoIR	17.01	0.838	0.142	16.12	0.821	0.154	22.97	0.922	0.067	26.22	0.947	0.049
ENVIDR	19.55	0.898	0.106	17.11	0.844	0.129	18.61	0.863	0.124	24.17	0.935	0.051
Ours	22.48	0.924	0.086	20.81	0.892	0.090	22.65	0.921	0.066	26.24	0.950	0.045

TABLE VIII: Quantitative comparison of relighting results using PSNR, SSIM, and LPIPS metrics. Each predicted image is aligned to ground-truth via channel-wise scaling before computing metrics, as described in [26], [42].

Methods	Shiny Blender			Glossy Synthetic			TensoIR Synthetic			Stanford ORB		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
GShader	20.99	0.889	0.125	21.80	0.870	0.120	26.86	0.930	0.063	30.32	0.958	0.037
GS-IR	20.43	0.865	0.166	20.17	0.813	0.157	26.24	0.907	0.083	29.02	0.949	0.045
RGS	18.61	0.847	0.148	17.39	0.809	0.151	26.41	0.919	0.080	27.32	0.941	0.045
NDR	22.65	0.896	0.131	20.02	0.827	0.153	27.10	0.907	0.091	29.44	0.953	0.040
NDRMC	23.91	0.911	0.151	22.13	0.879	0.126	27.81	0.907	0.110	31.52	0.967	0.039
TensoIR	19.17	0.887	0.150	18.97	0.824	0.143	28.58	0.944	0.081	30.60	0.960	0.033
ENVIDR	21.40	0.906	0.101	19.27	0.860	0.124	26.21	0.911	0.114	26.64	0.942	0.048
Ours	24.77	0.928	0.079	23.30	0.895	0.091	30.29	0.942	0.056	31.32	0.961	0.033

120 in our experiments) that fills the scene bounding box, where visibility SH coefficients are computed and saved on each grid point. Specifically, we first set all Gaussian colors to completely black and splat them into a white background from six cube face cameras centered at the grid point to form a cubemap. Then the rendered cubemaps are projected to the order-3 SH basis for coefficients. For any 3D point in the scene, its SH visibility can thus be queried from the grid using trilinear interpolation.

APPENDIX C MASKLESS REAL-WORLD SCENES

Most inverse rendering methods depend on accurate per-object masks [5], [23], [26], whereas our approach only relies on masks in the diffuse distillation stage, in the \mathcal{L}_α term we introduced to promote physical modeling. However, there is an alternative design for \mathcal{L}_α that simply encourages the α of each Gaussian to approach 1:

$$\mathcal{L}_\alpha^* = \sum_i |1 - \alpha_i|, \quad (26)$$

where i is the Gaussian index. We find that using \mathcal{L}_α^* might lead to slightly lower quality results for synthetic data (the average PSNR decreases by approximately 0.1 for novel view synthesis). However, \mathcal{L}_α^* , combined with our rendering model, allows our method to be applied to any real-world scenes without relying on an object mask. We can also manually control which areas need to be fully decomposed and which areas retain the raw radiance rendering. Our rendering model can seamlessly concatenate these areas.

Take the real-world scene *garden* provided by MipNeRF360 [49] as an example. The decomposition and relighting results are shown in Fig. 13. We use a sphere domain to approximately cover the table and consider objects outside the sphere as background. In the diffuse distillation stage, we use Eq. (26) to promote the decomposition of Gaussians inside the sphere. In specular and diffuse distillation stage, we use the loss term $\mathcal{L}_{bkg} = \sum_i \alpha_i^2$, where i loops over the indices of Gaussians outside the specified sphere, to encourage the background to fall back to raw radiance, which prevents it from interfering with environment light optimization. The results demonstrate the robustness of our method on in-the-wild data, making it more practical and user-friendly. Note that physically-based shading parameters of Gaussians outside the sphere domain generally become unconstrained and can be ignored. Another real-world scene we present is the *sedan* from [1], used as our teaser figure. It employs the same processing method.

APPENDIX D QUANTITATIVE RESULTS OF RELIGHTING WITH DIFFERENT SCALING SCHEMES

We show the PSNR, SSIM and LPIPS metrics of relighting results without scaling the environment light in Table VII, which reflects the absolute decomposition accuracy to some extent. Our method demonstrates a clear advantage on the Shiny Blender and Glossy Synthetic datasets, and achieves results very close to the best.

We also show relighting results with per-image scaling as described in [26], [42] in Table VIII. Our method outperforms alternatives regardless of the scaling schemes employed.

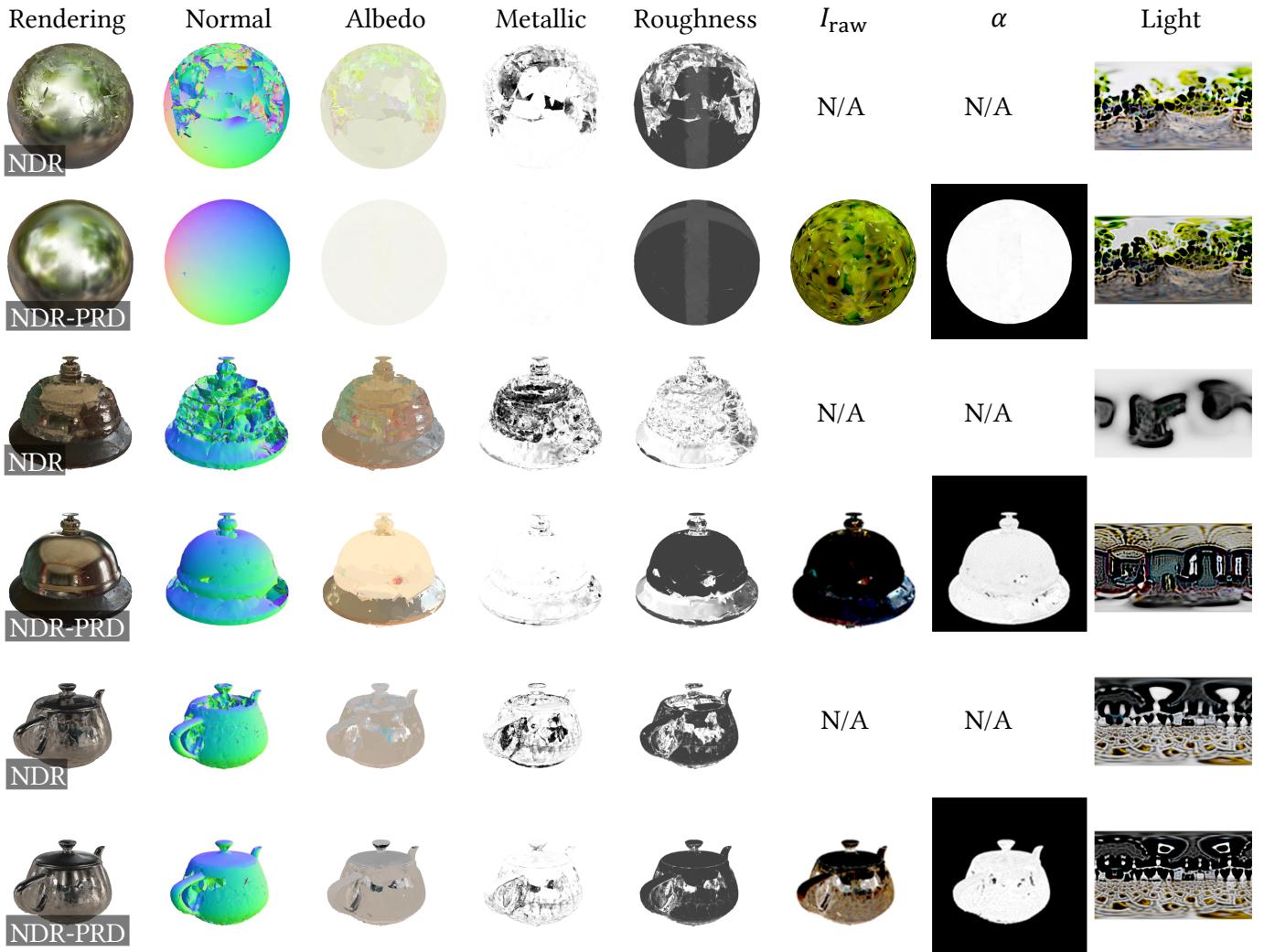


Fig. 14: Decomposition results of NDR [23] and NDR-PRD.

Fig. 15: Material editing. Left: the original *potion* object. Middle: the same object with Matte material, simulated by flipping its roughness $r' = 1 - r$. Right: also replace its albedo with a volumetric wooden texture.

procedure as described in Sec. 3.2. Any parameters unrelated to output radiance, including learning rate and loss weights, remain unchanged.

a) *Pre-training*: In this stage, meshes are extracted from the SDF and a position map is rendered from meshes. Following the original NDR [23], we encode the positions and take them as the MLP input to predict the raw radiance I_{raw} . The MLP architecture remains identical to the original NDR. We simply add three additional output channels for raw radiance. The same image loss is computed by comparing I_{raw} with ground-truth.

b) *Specular distillation*: The optimization in this stage is essentially the same as in the previous stage, except that the output radiance is replaced with the following:

$$I(x, \omega_o) = \alpha_x I_{\text{spec}}(x, \omega_o) + (1 - \alpha_x) I_{\text{raw}}(x), \quad (27)$$

where α is the progressive distillation map, which is also an additional MLP output. The PBR model used in NDR is identical to our method. The only difference is that NDR uses a $512 \times 512 \times 6$ cubemap, whereas we use $128 \times 128 \times 6$. Here we bumped our cubemap resolution of NDR-PRD to match NDR.

APPENDIX E IMPLEMENTATION DETAILS OF NDR-PRD

To apply progressive radiance distillation to NDR [23], we use the same rendering model given by Eq. (1) in Sec. 3.1 of the main paper, and develop a four-stage optimization

c) *Diffuse distillation*: Before starting this stage, following NDR, we extract meshes from the SDF and perform UV unwrapping. All attributes are then converted from MLP volume representations to optimizable 2D textures. The output radiance in this stage is the same as in our full rendering model. We use \mathcal{L}_α as an additional loss term to promote physically-based modeling.

d) *Joint refinement*: In this stage, we drop \mathcal{L}_α but keep other regularization terms as the original NDR. The total number of training iterations for the four stages is equal to that of the original NDR.

Fig. 14 compares the decomposition results of NDR-PRD with the original NDR. Starting from a converged raw radiance, we are able to avoid early-stage local minima and our distillation appears significantly more sane for highly specular objects.

APPENDIX F MATERIAL EDITING

In addition to relighting, our method also supports material editing. Fig. 15 illustrates one such example. Our high normal accuracy and clean distillation minimizes lingering fragments of the original texture, renders a convincing and thorough material change.

APPENDIX G MORE COMPARISONS

Fig. 16 compares our method with GShader [8] on the *sedan* scene. Our method produces smoother surface normal and managed to deduce the car material as almost completely specular. This allows us to reproduce the new-light coloration more faithfully. In contrast, GShader-relighted images exhibit visible remnants of the original reflection, with distinguishable tree-leaf patterns.

Fig. 17 compares relighting results evaluated on TensoIR Synthetic dataset [43], which contains all diffuse-dominant objects. Compared to all baselines, our method achieves realistic highlights and soft shadows close to the ground-truth (e.g. shadows under the jaw of *armadillo* on the second row, and highlights reflected by the plate on the sixth row). TensoIR [43] achieves sharper shadows by caching the visibility using a MLP, but it fails to capture any highlights reflections and suffers from noticeable aliasing issues. NDRMC [17] achieves more accurate shadows and indirect lighting by using a differentiable Monte Carlo renderer, but produces fragmented meshes for slightly smoother surfaces (e.g. pot of *ficus* on the third and forth row). Other methods fail to capture shading details and RGS [31] produces failure results under certain lighting.

Fig. 18 compares novel view synthesis and relighting between our method and state-of-the-art techniques on more scenes.

REFERENCES

- 
- Fig. 16: Comparison with GShader [8] on the *sedan* scene.
- [1] D. Verbin, P. Hedman, B. Mildenhall, T. Zickler, J. T. Barron, and P. P. Srinivasan, “Ref-nerf: Structured view-dependent appearance for neural radiance fields,” in *Proceedings of the IEEE/CVF CVPR*. IEEE, 2022, pp. 5481–5490.
 - [2] J. T. Kajiya, “The rendering equation,” in *Proceedings of the 13th annual conference on Computer graphics and interactive techniques*, 1986, pp. 143–150.
 - [3] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” in *Proceedings of ECCV*, 2020, pp. 405–421.
 - [4] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, “3d gaussian splatting for real-time radiance field rendering,” *ACM Trans. Graph.*, vol. 42, no. 4, July 2023.
 - [5] X. Zhang, P. P. Srinivasan, B. Deng, P.Debevec, W. T. Freeman, and J. T. Barron, “Nerfactor: Neural factorization of shape and reflectance under an unknown illumination,” *ACM Trans. Graph.*, vol. 40, no. 6, 2021.
 - [6] M. Boss, R. Braun, V. Jampani, J. T. Barron, C. Liu, and H. Lensch, “Nerd: Neural reflectance decomposition from image collections,” in *Proceedings of the IEEE/CVF CVPR*, 2021, pp. 12684–12694.
 - [7] Y. Liu, P. Wang, C. Lin, X. Long, J. Wang, L. Liu, T. Komura, and W. Wang, “Nero: Neural geometry and brdf reconstruction of reflective objects from multiview images,” *ACM Trans. Graph.*, vol. 42, no. 4, jul 2023.
 - [8] Y. Jiang, J. Tu, Y. Liu, X. Gao, X. Long, W. Wang, and Y. Ma, “Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces,” 2023.
 - [9] Z. Liang, Q. Zhang, Y. Feng, Y. Shan, and K. Jia, “Gs-ir: 3d gaussian splatting for inverse rendering,” 2024.
 - [10] Y. Shi, Y. Wu, C. Wu, X. Liu, C. Zhao, H. Feng, J. Liu, L. Zhang, J. Zhang, B. Zhou, E. Ding, and J. Wang, “Gir: 3d gaussian inverse rendering for relightable scene factorization,” 2023.
 - [11] H. Barrow and J. Tenenbaum, “Recovering intrinsic scene characteristics from images,” *Recovering Intrinsic Scene Characteristics from Images*, 01 1978.
 - [12] R. Ramamoorthi and P. Hanrahan, “A signal-processing framework for inverse rendering,” in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH ’01. New York, NY, USA: Association for Computing Machinery, 2001, p. 117–128. [Online]. Available: <https://doi.org/10.1145/383259.383271>
 - [13] Y. Sato, M. Wheeler, and K. Ikeuchi, “Object shape and reflectance modeling from observation,” *proceedings of ACM Siggraph ’97 (Computer Graphics)*, 04 1997.
 - [14] G. Patow and X. Pueyo, “A Survey of Inverse Rendering Problems,” *Computer Graphics Forum*, vol. 22, no. 4, pp. 663–687, 2003.
 - [15] H. Kato, D. Beker, M. Morariu, T. Ando, T. Matsukawa, W. Kehl, and A. Gaidon, “Differentiable rendering: A survey,” *arXiv preprint arXiv:2006.12057*, 2020.
 - [16] D. Azinovic, T.-M. Li, A. Kaplyanyan, and M. Nießner, “Inverse path tracing for joint material and lighting estimation,” in *Proceedings of the IEEE/CVF CVPR*, 2019, pp. 2447–2456.
 - [17] J. Hasselgren, N. Hofmann, and J. Munkberg, “Shape, light, and material decomposition from images using monte carlo rendering and denoising,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 22 856–22 869, 2022.
 - [18] B. Nicolet, A. Jacobson, and W. Jakob, “Large steps in inverse rendering of geometry,” *ACM Transactions on Graphics (TOG)*, vol. 40, no. 6, pp. 1–13, 2021.
 - [19] F. Luan, S. Zhao, K. Bala, and Z. Dong, “Unified shape and svbrdf recovery using differentiable monte carlo rendering,” in *Computer Graphics Forum*, vol. 40, no. 4. Wiley Online Library, 2021, pp. 101–113.
 - [20] Y. Quéau, J. Mélou, F. Castan, D. Cremers, and J.-D. Durou, “A variational approach to shape-from-shading under natural illumination,” in *Energy Minimization Methods in Computer Vision and Pattern Recognition*

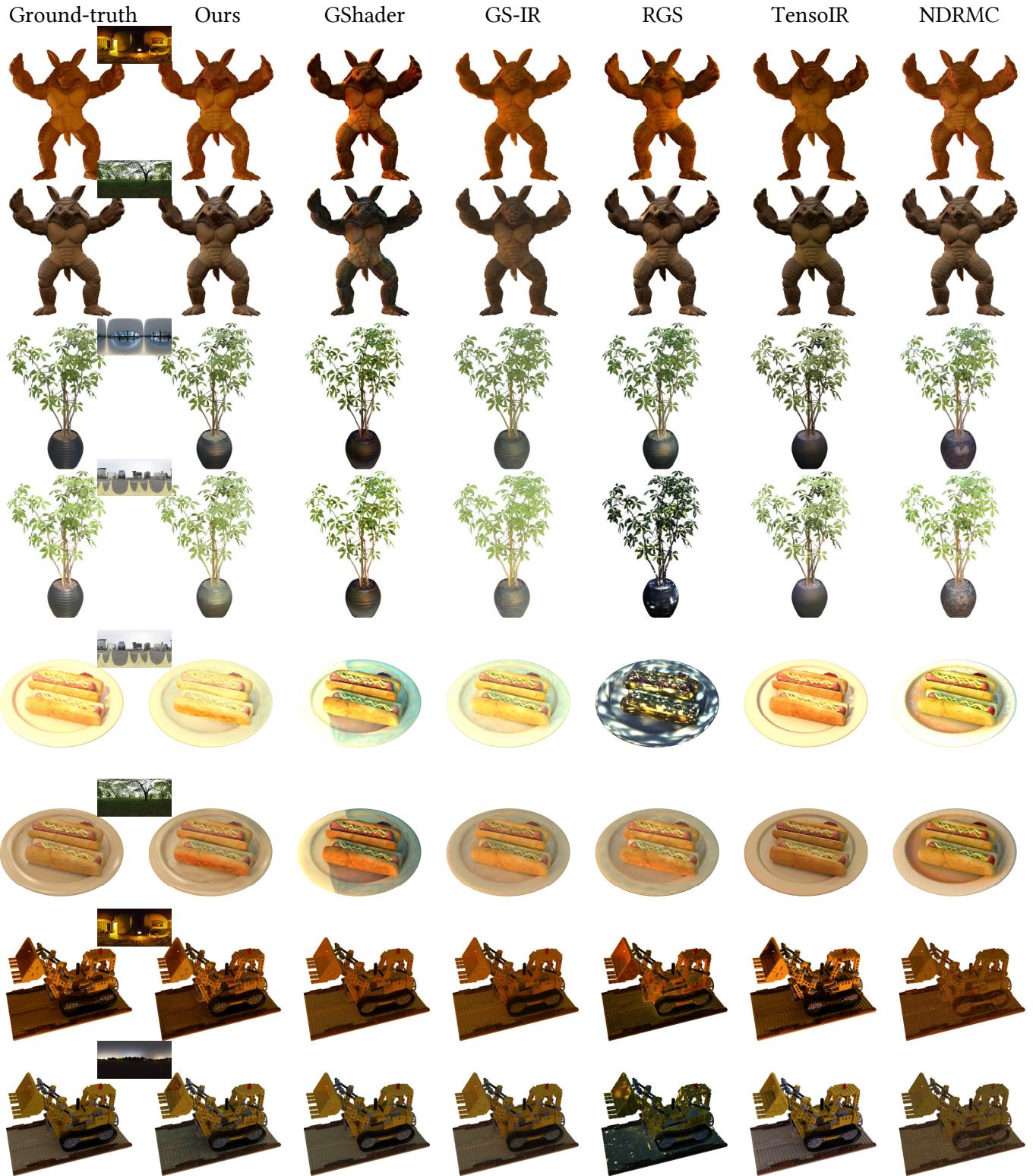


Fig. 17: More relighting results on TensoIR Synthetic dataset [43]. All objects are diffuse-dominant.

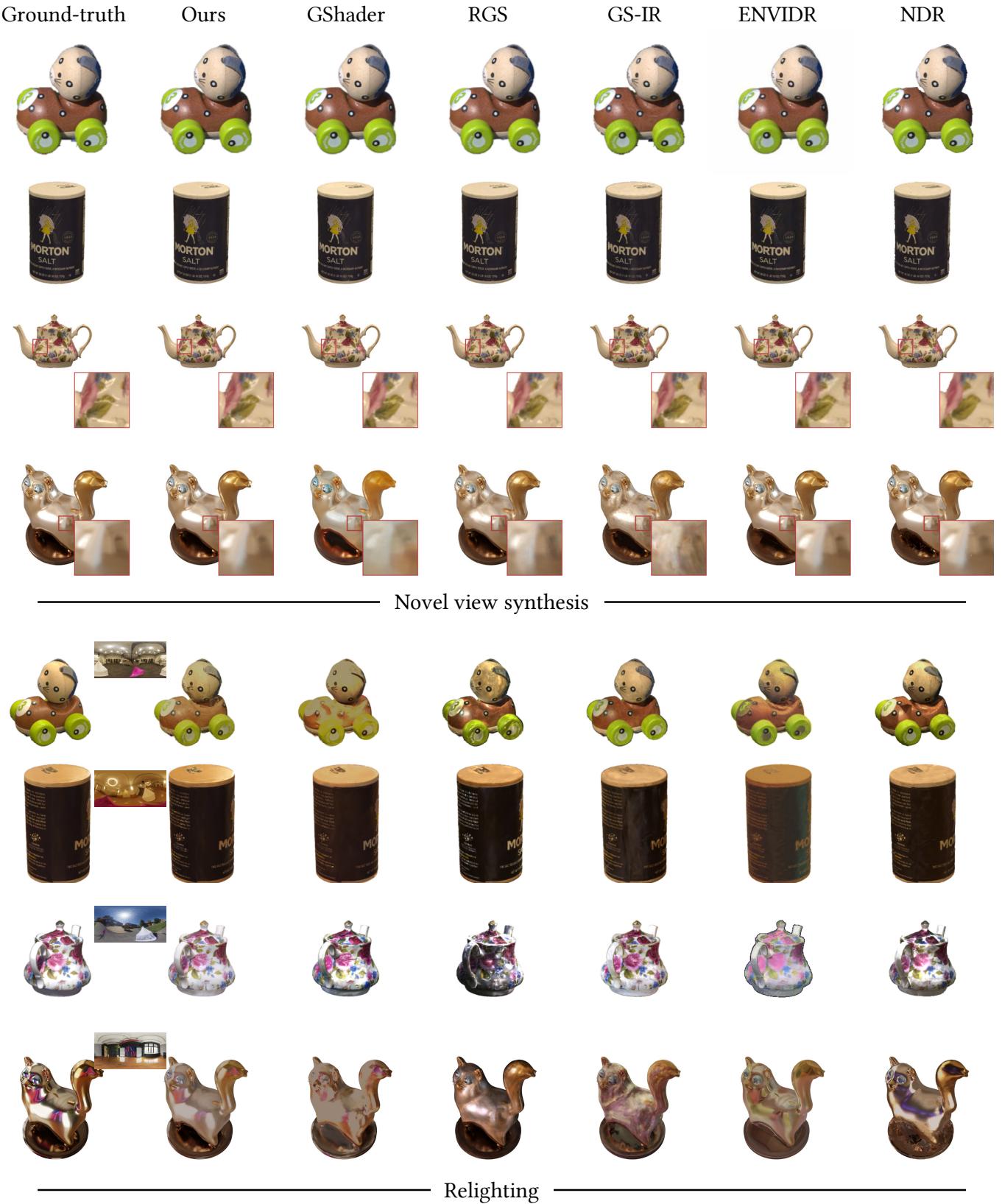


Fig. 18: More comparisons. Note that captions are placed below corresponding subfigures here.

- nition: 11th International Conference, EMMCVPR 2017, Venice, Italy, October 30–November 1, 2017, Revised Selected Papers 11. Springer, 2018, pp. 342–357.
- [21] M. Niemeyer, L. Mescheder, M. Oechsle, and A. Geiger, “Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 3504–3515.
- [22] L. Yariv, Y. Kasten, D. Moran, M. Galun, M. Atzmon, B. Ronen, and Y. Lipman, “Multiview neural surface reconstruction by disentangling geometry and appearance,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 2492–2502, 2020.
- [23] J. Munkberg, J. Hasselgren, T. Shen, J. Gao, W. Chen, A. Evans, T. Müller, and S. Fidler, “Extracting triangular 3d models, materials, and lighting from images,” in *Proceedings of the IEEE/CVF CVPR*, 2022, pp. 8280–8290.
- [24] C. Sun, G. Cai, Z. Li, K. Yan, C. Zhang, C. Marshall, J.-B. Huang, S. Zhao, and Z. Dong, “Neural-phr reconstruction of shape, material, and illumination,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 18046–18056.
- [25] P. P. Srinivasan, B. Deng, X. Zhang, M. Tancik, B. Mildenhall, and J. T. Barron, “Nerv: Neural reflectance and visibility fields for relighting and view synthesis,” in *Proceedings of the IEEE/CVF CVPR*, 2021, pp. 7495–7504.
- [26] K. Zhang, F. Luan, Q. Wang, K. Bala, and N. Snavely, “Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting,” in *Proceedings of the IEEE/CVF CVPR*, 2021, pp. 5453–5462.
- [27] Y. Zhang, J. Sun, X. He, H. Fu, R. Jia, and X. Zhou, “Modeling indirect illumination for inverse rendering,” in *Proceedings of the IEEE/CVF CVPR*, 2022, pp. 18643–18652.
- [28] A. Mai, D. Verbin, F. Kuester, and S. Fridovich-Keil, “Neural microfacet fields for inverse rendering,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 408–418.
- [29] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang, “Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction,” *arXiv preprint arXiv:2106.10689*, 2021.
- [30] M. Oechsle, S. Peng, and A. Geiger, “Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5589–5599.
- [31] J. Gao, C. Gu, Y. Lin, H. Zhu, X. Cao, L. Zhang, and Y. Yao, “Relightable 3d gaussian: Real-time point cloud relighting with brdf decomposition and ray tracing,” 2023.
- [32] T. Wu, J.-M. Sun, Y.-K. Lai, Y. Ma, L. Kobbelt, and L. Gao, “Deferrededges: Decoupled and editable gaussian splatting with deferred shading,” 2024.
- [33] K. Ye, Q. Hou, and K. Zhou, “3d gaussian splatting with deferred reflection,” in *SIGGRAPH Asia 2024 Conference Papers*, 2024, p. TBD.
- [34] R. L. Cook and K. E. Torrance, “A reflectance model for computer graphics,” *ACM Transactions on Graphics (ToG)*, vol. 1, no. 1, pp. 7–24, 1982.
- [35] S. Chan, H.-Y. Shum, and K.-T. Ng, “Image-based rendering and synthesis,” *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 22–33, 2007.
- [36] T. Whitted, “An improved illumination model for shaded display,” *Commun. ACM*, vol. 23, no. 6, p. 343–349, jun 1980. [Online]. Available: <https://doi.org/10.1145/358876.358882>
- [37] M. Nimier-David, T. Müller, A. Keller, and W. Jakob, “Unbiased inverse volume rendering with differential trackers,” *ACM Transactions on Graphics (TOG)*, vol. 41, no. 4, pp. 1–20, 2022.
- [38] M. Deering, S. Winner, B. Schediwy, C. Duffy, and N. Hunt, “The triangle processor and normal vector shader: a vlsi system for high performance graphics,” in *Proceedings of SIGGRAPH ’88*. New York, NY, USA: ACM, 1988, p. 21–30.
- [39] P.-P. Sloan, J. Kautz, and J. Snyder, “Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments,” in *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, 2002, pp. 527–536.
- [40] P.-P. Sloan, “Stupid spherical harmonics (sh) tricks,” in *Game developers conference*, vol. 9, 2008, p. 42.
- [41] S. Hill, S. McAuley, L. Belcour, W. Earl, N. Harrysson, S. Hillaire, N. Hoffman, L. Kerley, J. Patry, R. Piecké *et al.*, “Physically based shading in theory and practice,” in *ACM SIGGRAPH 2020 Courses*, 2020, pp. 1–12.
- [42] Z. Kuang, Y. Zhang, H.-X. Yu, S. Agarwala, S. Wu, and J. Wu, “Stanford-orb: A real-world 3d object inverse rendering benchmark,” 2023.
- [43] H. Jin, I. Liu, P. Xu, X. Zhang, S. Han, S. Bi, X. Zhou, Z. Xu, and H. Su, “Tensoir: Tensorial inverse rendering,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 165–174.
- [44] W. Arrighetti, “The academy color encoding system (aces): A professional color-management framework for production, post-production and archival of still and motion pictures,” *Journal of Imaging*, vol. 3, no. 4, p. 40, 2017.
- [45] R. Liang, H. Chen, C. Li, F. Chen, S. Panneer, and N. Vijaykumar, “Envidr: Implicit differentiable renderer with neural environment lighting,” in *Proceedings of the IEEE/CVF ICCV*, 2023, pp. 79–89.
- [46] B. Walter, S. R. Marschner, H. Li, and K. E. Torrance, “Microfacet models for refraction through rough surfaces,” in *Proceedings of the 18th Eurographics conference on Rendering Techniques*, 2007, pp. 195–206.
- [47] K. Zhou, Y. Hu, S. Lin, B. Guo, and H.-Y. Shum, “Precomputed shadow fields for dynamic scenes,” in *ACM SIGGRAPH 2005 Papers*, 2005, pp. 1196–1201.
- [48] R. Ramamoorthi and P. Hanrahan, “On the relationship between radiance and irradiance: determining the illumination from images of a convex lambertian object,” *JOSA A*, vol. 18, no. 10, pp. 2448–2459, 2001.
- [49] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, “Mip-nerf 360: Unbounded anti-aliased neural radiance fields,” in *Proceedings of the IEEE/CVF CVPR*, 2022, pp. 5470–5479.