ESR-NeRF: Emissive Source Reconstruction Using LDR Multi-view Images

Jinseo Jeong[1] Junseo Koo[1] Qimeng Zhang[2] Gunhee Kim[1]
[1]Seoul National University [2]Korea University
jinseo@vision.snu.ac.kr, junseo.koo@vision.snu.ac.kr, zoe1024@korea.ac.kr, gunhee@snu.ac.kr
https://jinseo.kr/ESR-NeRF

Abstract

Existing NeRF-based inverse rendering methods suppose that scenes are exclusively illuminated by distant light sources, neglecting the potential influence of emissive sources within a scene. In this work, we confront this limitation using LDR multi-view images captured with emissive sources turned on and off. Two key issues must be addressed: 1) ambiguity arising from the limited dynamic range along with unknown lighting details, and 2) the expensive computational cost in volume rendering to backtrace the paths leading to final object colors. We present a novel approach, ESR-NeRF, leveraging neural networks as learnable functions to represent ray-traced fields. By training networks to satisfy light transport segments, we regulate outgoing radiances, progressively identifying emissive sources while being aware of reflection areas. The results on scenes encompassing emissive sources with various properties demonstrate the superiority of ESR-NeRF in qualitative and quantitative ways. Our approach also extends its applicability to the scenes devoid of emissive sources, achieving lower CD metrics on the DTU dataset.

1. Introduction

Extensive research has focused on reconstructing 3D object structures [16, 43, 47, 86], material properties [18, 29, 65], and lighting [15, 33, 34, 74, 79] from 2D images, applicable across domains including 3D graphics and augmented reality [62, 63, 69, 72]. This endeavor not only facilitates the creation of life-like virtual objects but also streamlines the process of scene manipulation [27, 58, 61, 73]. Recent advancements [24, 30, 36, 71] have built on Neural Radiance Fields (NeRF) [40] successes in novel view synthesis [3, 4, 45, 81, 91]. Significant progress in relighting [37, 38, 50] has facilitated scene editing via manipulating the reconstructed light sources. However, existing methods predominantly deal with the scenes lit by distant sources, like environment maps or collocated flashlights. Notably, NeRF-based inverse rendering has yet to

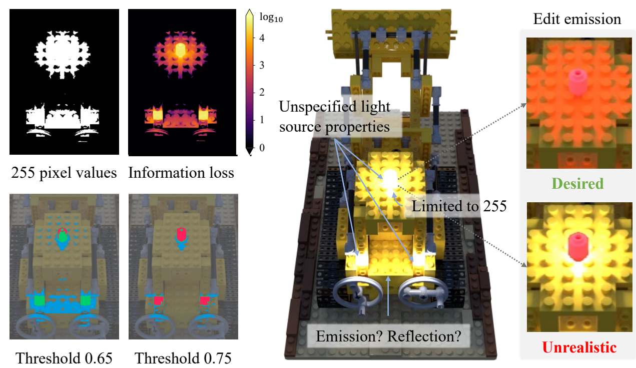


Figure 1. Challenges posed by emissive sources in LDR images. Green, red, and blue in thresholded images respectively show true positives, false negatives, and false positives of source identification. Thresholding values are scaled down divided by 255. The contrast between light on and off pixel values is more pronounced in surroundings than emissive sources. Inaccurate reconstruction of emissive sources disrupts scene editing, causing reflection areas to stay static while only the source colors change.

consider scenes with multiple emissive sources, a common real-world illumination condition.

Emissive sources in a scene introduce critical challenges: (i) ambiguity in decomposing scene components and (ii) high computational costs for analyzing the causes of pixel colors. This ambiguity stems from difficulties in identifying emissive source regions, as illustrated in Fig. 1. Contrary to prior setups [6–8, 66, 85, 93], we allow the possibility of numerous emissive sources throughout the scene. In standard photographs with pixel values from 0 to 255, the distinction between emissive sources and nearby reflection areas is challenging. As shown in Fig. 1, relying solely on pixel value thresholding is insufficient for differentiating between emissive sources and their reflections. Naive inverse path tracing is impractical, due to the computational costs rising exponentially with the number of ray bounces in volume rendering. This can cause inaccuracy in emissive source reconstruction, yielding unrealistic illumination in reflective areas as users manipulate emissive sources.

To address these challenges, we introduce ESR-NeRF

(Emissive Sources Reconstructing NeRF), a novel approach capable of reconstructing any number of emissive sources by progressively discovering reflection areas. We assume that the scenes are observed in two lighting conditions: one with all emissive sources active and the other with them inactive. Our approach utilizes neural networks as learnable functions for representing ray-traced fields. By training networks to satisfy each light transport segment, we sidestep the computational overhead of ray tracing associated with ray bounces. In this work, we exclusively use low dynamic range (LDR) images, setting us apart from prior mesh-based methods that rely on high dynamic range (HDR) images [2, 19, 48, 76].

Our experiments encompass synthetic and real scenes, ranging from single to multiple lighting configurations with complex reflections. The scenes vary in light source counts, color, and intensity. Qualitative and quantitative evaluations show ESR-NeRF's superiority over state-of-the-art NeRF-based re-lighting methods. Furthermore, Chamfer Distance (CD) metrics on the DTU dataset [23] indicate ESR-NeRF's competitive performance in scene reconstruction, even without emissive sources.

We summarize our contributions as follows.

1. Our work presents the first NeRF-based inverse rendering that can deal with the scenes with any number of emissive sources, challenging the distant light assumption of previous research.
2. Unlike existing mesh-based methods relying on HDR images, we use LDR images for the first time, overcoming the poor representation of emissive sources.
3. We provide a benchmark dataset designed to evaluate the performance of emissive source reconstruction.
4. Our method is applicable to the scenes with or without emissive sources, achieving superior mesh reconstruction results on the DTU dataset.

2. Related work

Neural Rendering. Advancements in implicit representations [51, 60] and volume rendering [39] have significantly enhanced neural rendering capabilities, enabling the reconstruction of scene components from 2D images. One of the key directions is mesh extraction [44, 70, 77, 78, 80, 97], with methods like NeuS [68] and VolSDF [87] utilizing signed distance function (SDF) values for volume rendering. Recently, the efficient computation of volume rendering has become a focal point due to the substantial computational cost associated with network inference for ray color calculation [41, 49, 88]. Several methods propose to directly predict ray color using the 4D light fields concept [1, 52, 55] or leveraging voxel grids for fast inference of spatial features [5, 11, 12, 14, 31, 56]. NeuralRadiosity [17] shares similarity with our method, as it predicts ray-traced values instead of explicitly tracing individual rays.

	Voxurf	TensoIR	Path Tracing	ESR-NeRF
Big O	n	$n \cdot d$	$(n \cdot d)^{b+1}$	$n^2 \cdot d$
Indirect illumination	✘	✔	✔	✔
BRDF decomposition	✘	✔	✔	✔
Emissive source control	✘	✘	✔	✔

Table 1. Computational cost comparison for inverse rendering methods. n is the number of sampled points along a ray, d is the number of scattering rays, and b is the number of ray bounces.

However, they primarily focus on calculating the final object color when all scene information is available. In contrast, our inverse rendering approach aims to reconstruct emissive sources within a scene, addressing the ambiguities introduced by their presence in LDR images.

Inverse Rendering. A growing emphasis revolves around the decomposition of materials represented by spatially varying bidirectional reflectance distribution functions (SVBRDF) [46, 83, 99]. To lessen the computational burden in inverse rendering [25, 54, 96, 98], several methods have adopted neural networks as lookup tables [9] or computational caches [54, 90, 95]. While NeRV [54] utilizes caching visibility and NeILF++ [90] adopts caching surface point radiance with the inter-reflection loss for incident radiance, our method diverges by focusing on tracing radiance origins. Specifically, we aim to identify emissive sources within a scene, moving beyond the simplification of incident radiance calculations. Several methods rely on diverse known lighting configurations to exploit variations in object appearances [59, 64, 84, 89]. Toggling emissive sources on and off resembles the common one-light-at-a-time (OLAT) technique, as seen in NLT [94] and ReNeRF [82]. However, our setting does not need to know light source properties and to toggle lights individually. Instead, we allow for toggling all lights together. Recent works have also jointly reconstruct the mesh, materials, and lighting [20, 35, 42, 57]. They tackle with images captured under a single unknown lighting condition [92, 95], assuming that radiance already encodes global illumination [75, 96]. However, they confine to the scenes illuminated by far-distant lights, constrained to an 8-bit color spectrum. Our work considers the presence of multiple emissive sources within a scene captured in LDR images, questioning the prevailing notion that radiance fields trained with the image rendering loss faithfully represents global illumination. While some methods [2, 19, 32, 48, 76] deal with the scenes featuring emissive sources, they work outside the volume rendering framework and depend on HDR input images, assuming prior knowledge of scene geometry.

3. Preliminaries

Surface Representation. Analgous to NeRF [40], neural network f_θ predicts SDF values at arbitrary 3D spatial locations. NeuS [68] integrates surface representa-

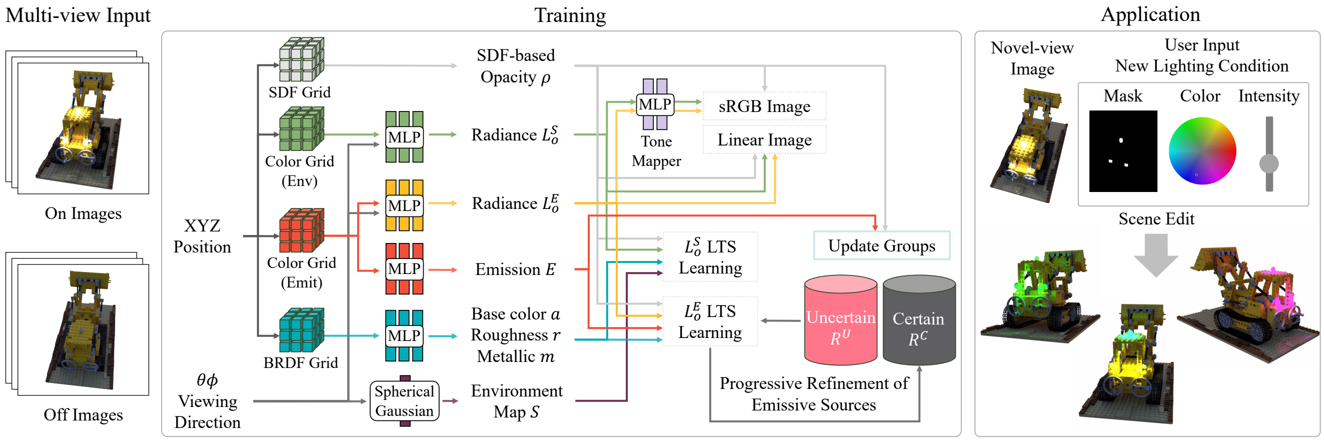


Figure 2. The pipeline of emissive source reconstruction. Given LDR images with emissive sources on and off, scene components are reconstructed by synthesizing training images and enforcing LTS requirements. Emissive sources are progressively refined via categorizing training rays into uncertain and certain groups. The scenes can be edited with new lighting conditions using reconstructed emissive sources.

tion into volume rendering using the SDF-based opacity $\rho(x) = \max(\frac{-\frac{d\Phi_s}{dx}(f(x))}{\Phi_s(f(x))}, 0)$. Here $\Phi_s(x) = (1 + e^{-sx})^{-1}$ is the sigmoid function where s controls the sharpness of surfaces. The color of a ray can be calculated as

$$\hat{C}(r) = \int_0^\infty T(r(t))\rho(r(t))L_o(r(t), \omega_o)\, dt, \tag{1}$$

where $\hat{C}(r)$ denotes the predicted ray color, $r(t; c, \omega_o) = c - t \cdot \omega_o$ is the ray with camera center c along direction ω_o, $T(r(t)) = \exp\left(\int_0^t -\rho(r(u))\, du\right)$ is the transmittance, and $L_o(r(t), \omega_o)$ is the outgoing radiance. Henceforth, we use x to denote a point in $r(t; c, \omega_o)$ for notational simplicity.

Light Transport in Volume Rendering. Extracting light sources necessitates analyzing the causes affecting the final ray colors. Kajiya's rendering equation [26] factorizes the outgoing radiance $L_o(x, \omega_o)$ into emission and reflections:

$$L_o(x, \omega_o) = E(x) + \int_\Omega L_i(x, \omega_i)R(x, \omega_o, \omega_i; b)d\omega_i, \tag{2}$$

where $E(x)$ is the emission, $R(x, \omega_o, \omega_i; b)$ represents the SVBRDF parametrized by parameters b with Lambert cosine multiplied, and $L_i(x, \omega_i)$ is the incident radiance. In volume rendering, computing the incident radiance at point x is akin to evaluating Eq. 1, with x serving as the camera center. By iteratively factorizing the outgoing radiance in the incident radiance, the contribution of a path length i for a pixel can be decomposed as in Eq. 3, where $\mathcal{H}_i = \Pi_{j=1}^{i-1} T(x_j)\rho(x_j)R(x_j, \omega_{j-1}, \omega_j)$ is the path throughput, $S(\omega_i)$ is the environment map strength in direction ω_i, and $V(x, \omega_i) = exp\left(\int_0^\infty -\rho(r(u; x_i, -\omega_i))\, du\right)$ is the visibility of the environment map at point x along direction ω_i:

$$\begin{aligned} P_i = \int_{l_1}\int_\Omega \cdots \int_{l_{i-1}}\int_\Omega \Big(\int_{l_i} T(x_i)\rho(x_i)E(x_i)\, dt_i + \\ S(\omega_{i-1})V(x_{i-1}, \omega_{i-1})\Big)\mathcal{H}_i dt_1 d\omega_1 \cdots dt_{i-1} d\omega_{i-1}. \end{aligned} \tag{3}$$

Extending the analysis to longer light paths, or equivalently, increasing the number of ray bounces, leads to exponential growth in computation complexity. This poses a challenge when attempting to decompose the influence of unknown emissive sources, as their ability to produce strong reflections makes ignoring indirect illumination infeasible.

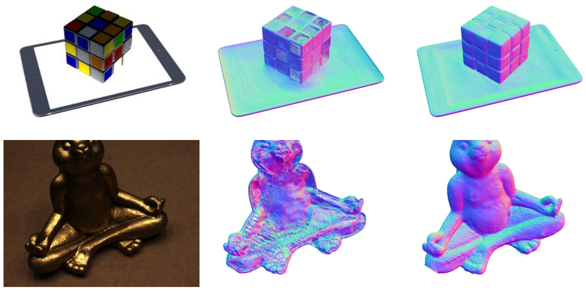
4. Methodology

None of the previous works address the reconstruction of emissive sources from LDR multi-view images. Sec. § 4.1 through § 4.5 detail our method, ESR-NeRF, which reconstructs emissive sources without prior knowledge of scene geometry, materials, or lighting specifics (including their location, number, or colors). We also show how these reconstructed sources can be used for scene editing in § 4.5.

4.1. Learnable Tone-mapper

Throughout the paper, we use $\mathcal{R}$ to represent camera rays, C for pixel values, and a binary flag $\mathbb{I}$ to indicate whether an image is captured with emissive sources on or off.

To extract HDR values from LDR images, we employ the softplus activation for outgoing radiance prediction and apply a clipping and gamma function τ [21] for the rendering loss such that $\hat{C}_\tau(r) = \tau(\hat{C}(r))$. Unlike previous NeRF-based works [25, 37, 54, 57] that limit radiance to the range of $[0, 1]$, our approach allows for any positive radiance values. Yet, it creates difficulties in differentiating between the surface weight $T(x)\rho(x)$ and the magnitude of radiance value $L_o(x, \omega_o)$, since it allows for the possibility of assigning extreme radiance to the points with low surface weights to render same ray colors. Such ambiguity poses challenges, particularly in dark and high-contrast scenes, aggravating surface reconstruction (see Fig. 3). To address this, we introduce a learnable tone-mapper $m_\theta : \mathbb{R}_+^3 \rightarrow [0, 1]^3$, that takes positionally encoded HDR linear values as input:



(a) Image (b) w/o m_θ (c) ESR-NeRF

Figure 3. Reconstructed surfaces with the learnable tone-mapper.

$$\hat{C}_{m_\theta}(r) = \int_0^\infty T(x)\rho(x)m_\theta(L_o(x,\omega_o))\,dt, \tag{4}$$

$$L_o(x,\omega_o) = L_o^S(x,\omega_o) + L_o^E(x,\omega_o)\cdot \mathbb{I}, \tag{5}$$

where $L_o^S(x,\omega_o)$ is radiance when emissive sources are turned off, while $L_o^E(x,\omega_o)$ stands for radiance added to the scene by emissive sources. Our rendering loss is then formulated as follows, with λ_τ as a hyper-parameter:

$$\mathcal{L}_{\text{render}} = \sum_{r\in\mathcal{R}}(\|C(r)-\hat{C}_{m_\theta}(r)\|_2^2+\lambda_\tau\|C(r)-\hat{C}_\tau(r)\|_2^2). \tag{6}$$

4.2. Learning of Light Transport Segments

The computational complexity of object appearance analysis in volume rendering is notably high, as shown in Eq. 3. We take an alternative approach by leveraging neural networks to represent ray-traced fields, rather than explicitly tracing every rays. Our distinct contribution to inverse rendering lies in precise adjustment of radiance. Specifically, we impose constraints on the predicted radiance to satisfy each light transport segments. The light transport segments (LTS) loss, $\mathcal{L}_{lts}$, plays a pivotal role in our method:

$$\mathcal{L}_{lts}^S = \sum_{x,\omega_o}\|L_o^S(x,\omega_o) - \hat{L}_o^S(x,\omega_o)\|_2^2, \tag{7}$$

$$\mathcal{L}_{lts}^E = \sum_{x,\omega_o}\|L_o^E(x,\omega_o) - \hat{L}_o^E(x,\omega_o)\|_2^2, \tag{8}$$

$$\hat{L}_o^S(x,\omega_o) = \int_\Omega \underbrace{S(\omega_i)V(x,\omega_i)R(x,\omega_o,\omega_i)}_{\text{direct illumination by an environment map}}\,d\omega_i + \int_\Omega\int_0^\infty \underbrace{T(x')\rho(x')L_o^S(x',-\omega_i)\,dt'\,R(x,\omega_o,\omega_i)}_{\text{indirect illumination by an environment map}}\,d\omega_i. \tag{9}$$

$$\hat{L}_o^E(x,\omega_o) = \underbrace{E(x)}_{\text{emission}} + \int_\Omega\int_0^\infty \underbrace{T(x')\rho(x')L_o^E(x',-\omega_i)\,dt'\,R(x,\omega_o,\omega_i)}_{\text{direct \& indirect illumination by emissive sources}}\,d\omega_i. \tag{10}$$

We ensure consistency between the radiance directly predicted by the network $L_o(x,\omega_o)$ and the radiance achievable based on the scene context $\hat{L}_o(x,\omega_o)$. Previous approaches have focused on matching $\hat{L}_o(x,\omega_o)$ to training views, overlooking the relations to $L_o(x,\omega_o)$. This hinders the restoration of HDR radiance by supervising scene components to LDR training views. In contrast, our LTS loss enables volumetric energy *transfer* of radiance, adjusting outgoing radiance based on their interrelations.

To implement this concept, we train six dedicated networks for SDF $f(x)$, SVBRDF parameters $b(x)$, emission $E(x)$, environment map $S(\omega_i)$, outgoing radiances $L_o^S(x,\omega_o)$ and $L_o^E(x,\omega_o)$, to adhere to these LTS requirements. For the environment map, we represent it using 48 Spherical Gaussians [67] : $\sum_{k=1}^M \mu_k e^{\lambda_k(\omega_i\cdot\xi_k-1)}$, followed by the softplus activation. $\mu\in\mathbb{R}^3$, $\lambda\in\mathbb{R}_+$, and $\xi\in\mathbb{S}^2$ respectively denote the lobe amplitude, sharpness, and axis.

4.3. Progressive Discovery of Reflection Areas

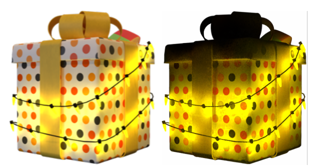


Figure 4. Left: Image with active emissive sources. Right: Identified emissive sources w/o progressive discovery of reflection areas.

Relying solely on LTS is insufficient for addressing ambiguity arising from low pixel values of emissive sources and intense reflections in adjacent regions, often leading to confusion between emission and reflection. The right image in Fig. 4 shows self-emitting objects restored with the naive LTS loss. While emissive sources are small, large areas affected by them are also identified as emissive sources. We propose a reflection-aware progressive approach for precise identification of emissive sources. By leveraging LTS learning, we extend the regions that can be regarded as reflection areas. Fig. 5 illustrates our progressive algorithm.

Reflection-Aware Emission Refinement. Since surface points are unknown and are updated during learning, we opt to utilize rays rather than surface points. This process involves categorizing training rays into two groups: uncertain ($\mathcal{R}^U$) and certain ($\mathcal{R}^C$). The certain group contains the rays confidently identified as reflection, aiding the transfer of radiance energy to nearby points. For the points in the certain group, we use the Eq. 11 instead of Eq. 10 to exclusively attibute outgoing radiances to reflections. Satisfying the LTS loss on the certain group results in adjusting the outgoing radiances of influential points, as illustrated in Fig. 5(a):

$$\hat{L}_o^E(x,\omega_o) = \int_\Omega\int_0^\infty T(x')\rho(x')L_o^E(x',-\omega_i)\,dt'\,R(x,\omega_o,\omega_i)\,d\omega_i. \tag{11}$$

The uncertain group includes the rays indicating the areas that are undetermined yet as reflection or emission. Using Eq. 12 to compute $\hat{L}_o^E(x,\omega_o)$, this group adjusts emissions $E(x)$ based on the radiance updates by the certain group, where "sg" represents the stop-gradient:

$$\hat{L}_o^E(x,\omega_o) = E(x)+ \\ \text{sg}\left(\int_\Omega \int_0^\infty T(x')\rho(x')L_o^E(x',-\omega_i)\,dt' R(x,\omega_o,\omega_i)\,d\omega_i\right). \tag{12}$$

As shown in Fig. 5(b), this leads to increased emissions for the regions whose radiances are adjusted to account for the reflections in the certain group. Conversely, emissions decrease for the regions where there is little change in outgoing radiance, but incident radiances are increased by surrounding influential points.

Ray Group Management. As emissions and radiances are adjusted, the groups are dynamically updated at predefined training intervals through the following process. Within the uncertain group, we evaluate the expected emission strength of rays, retaining only those above a threshold k_i. Rays below this threshold are then merged to the certain group:

$$\mathcal{R}_i^U = \{r \mid \max_{RGB}\left(\int_0^\infty T(x)\rho(x)E(x)\,dt\right) \geq k_i, r \in \mathcal{R}_{i-1}^U\}, \tag{13}$$

$$\mathcal{R}_i^C = \left(\mathcal{R}_{i-1}^U - \mathcal{R}_t^U\right) \cup \mathcal{R}_{i-1}^C. \tag{14}$$

Subsequently, newly added rays to the certain group can be used to localize influential points and update their outgoing radiances. This iterative process progressively refines the separation between reflective and emissive regions, attaining more accurate identification of emissive sources.

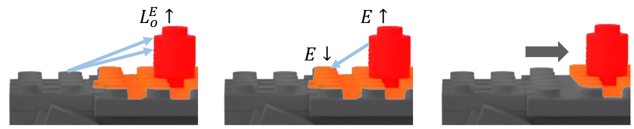
LTS Loss Decomposition. The LTS loss, as detailed in Eq. 15, can be decomposed using a stop-gradient operation to refine the adjustment process.

$$\mathcal{L}_{lts}^E = \sum_{x,\omega_o}(\lambda_l \|\text{sg}(L_o^E(x,\omega_o)) - \hat{L}_o^E(x,\omega_o)\|_1 + \\ \lambda_r \|L_o^E(x,\omega_o) - \text{sg}(\hat{L}_o^E(x,\omega_o))\|_1). \tag{15}$$

We prioritize λ_l to enhance the update of scene context, affecting other points' radiance given the predicted $L_o(x,\omega_o)$. λ_r prevents severe deviation of every $L_o(x,\omega_o)$ within the current scene context. This aligns with our focus on HDR source reconstruction from LDR images, addressing under-represented information in training data.

4.4. Training Details

We employ the Voxurf architecture [80] as backbone and adopt the simplified Disney BRDF model [10] for SVBRDF representation, with parameters including base color $\in [0,1]^3$, roughness $\in [0,1]$, and metallic $\in [0,1]$. The learnable tone-mapper, structured as a two-layer MLP, is utilized for the rendering loss only. Initially, we pre-train our networks using the rendering loss, subsequently integrating the basic LTS loss (Eq. 7 and Eq. 8) into our training regimen.



(a) LTS on R^C (b) LTS on R^U (c) Ray Group Managing

Figure 5. Illustration of the progressive emissive source reconstruction with reflection awareness. Gray color represents the areas belonging to the certain group, while the red (emissive sources) and orange (their reflections) areas belong to the uncertain group.

This phase transitions to the reflection-aware progressive training scheme, where we adopt the ℓ_1 loss due to its empirical stability in refining emissive source reconstruction. We use a smoothing regularization to promote local consistency in normals, BRDFs, and emissions. To ensure view-consistent labeling of 3D points as either reflective or emissive, we implement the emission suppression loss for points beloning to the certain group:

$$\mathcal{L}_{supp}^E = \sum_{r \in R_t^C} \|\int_0^\infty T(x)\rho(x)E(x)\,dt\|_2^2, \tag{16}$$

The threshold k_i linearly increases with each time step t, utilizing a grid search within a range of $[10^{-3}, 10^{-5}]$ to find the slope. We construct mini-batches via stratified sampling within each group. For a detailed description of our training procedure, please refer to Appendix.

4.5. Scene Editing

Reconstructed emissive sources enable scene editing; users select emissive sources using binary masks $M_{j=1...N}$ and specify lighting conditions using colors $c_{j=1...N}$ and intensities $i_{j=1...N}$ within the HSV color space [53].

We identify the rays in the uncertain group that match M by projecting expected surface points p of the rays onto the camera with the pose $\mathbf{R}|\mathbf{t}$:

$$p = \int_0^\infty T(x)\rho(x)x\,dt, \tag{17}$$

$$\mathbb{I}_j^{hit}(x) = \text{interp}(M_j, p') > 0, \text{ where } p' = \mathbf{K}[\mathbf{R}|\mathbf{t}][p|1]^T. \tag{18}$$

For the rays satisfying $\mathbb{I}_j^{hit}(x)$, we apply the designated lighting conditions. The new emission values are computed by substituting the original hue (H) and saturation (S) of $E(x)$ with the user-specified color c_j and adjusting the value (V) of $v(x)$ with the new intensity i_j:

$$E(x) = \text{hsv\_to\_rgb}\left([c_j|\,(v(x)\times i)]\right)\cdot \mathbb{I}_j^{hit} + E(x)\cdot \neg\mathbb{I}_j^{hit}. \tag{19}$$

These modifications influence scene appearance by optimizing the loss in Eq. 20. During this process, all networks, except for $L_o^E(x,\omega_o)$, are frozen:

$$\mathcal{L}_{edit} = \sum_{x,\omega_o}\|L_o^E(x,\omega_o) - \text{sg}(\hat{L}_o^E(x,\omega_o))\|_2^2. \tag{20}$$

	White colored												Vivid colored											
	Lego		Gift		Book		Cube		Billboard		Balls		Lego		Gift		Book		Cube		Billboard		Balls	
	IoU	MSE	IoU	MSE	IoU	MSE	IoU	MSE	IoU	MSE	IoU	MSE	IoU	MSE	IoU	MSE	IoU	MSE	IoU	MSE	IoU	MSE	IoU	MSE
Twins	0.22	20.19	**0.49**	8.59	0.63	3.91	**0.95**	31.83	0.69	1.12	0.90	0.06	0.25	6.96	**0.24**	6.09	0.55	2.63	0.95	10.64	0.09	0.75	0.83	**0.04**
NeILF++	0.43	20.88	0.07	9.38	**0.95**	4.64	0.93	32.67	0.01	1.95	**0.91**	0.80	0.30	7.65	0.09	6.86	**0.95**	3.36	0.94	11.49	0.02	1.57	0.92	0.78
TensoIR	**0.71**	**20.13**	0.15	**8.55**	**0.95**	**3.87**	**0.95**	**31.73**	**0.76**	**1.11**	**0.95**	**0.05**	**0.33**	**6.93**	0.15	**6.05**	**0.95**	**2.59**	**0.96**	**10.60**	**0.77**	**0.74**	**0.95**	**0.03**
ESR-NeRF	**0.81**	**8.38**	**0.60**	**3.49**	**0.96**	**1.19**	**0.97**	**17.87**	**0.84**	**0.46**	**0.95**	**0.04**	**0.51**	**5.48**	**0.59**	**2.50**	**0.96**	**0.51**	**0.97**	**7.94**	**0.88**	**0.26**	**0.94**	**0.03**

Table 2. Results of emissive source identification. ESR-NeRF outperforms state-of-the-art re-lighting methods in reconstructing emissive sources, regardless of their color. The IoU measures the source area identification (a higher value is better), and the MSE quantifies the difference between reconstructed images and HDR ground truth images (a lower value is better).

5. Experiments

We assess ESR-NeRF in reconstructing emissive sources by focusing on both identification and intensity restoration. To showcase its effectiveness, we conduct a range of experiments, including scene editing, ablation studies, illumination decomposition, and surface reconstruction, providing both quantitative and qualitative results.

5.1. Experiment Settings

We curate 6 diverse synthetic scenes, each with 200 training images evenly distributed between on and off lighting conditions. To evaluate the robustness of our approach against light colors, we consider two distinct settings of white colored and vivid colored emissive sources, resulting in a total of 12 scenes. The vivid colors are selected with full saturation in the HSV color space. We measure source identification and radiance reconstruction using IoU and MSE metrics on novel view test images, comparing against ground truth data from Blender-rendered emission masks and EXR files. The emission strengths, the maximum EXR file values, range from 2 to 200. For quantitative scene editing evaluation, we alter the white-colored sources to various colors—red, green, blue, cyan, magenta, yellow—and adjust intensities to half or double their original values. Qualitative results include scene editing for vividly colored sources and real scenes captured with a Fuji 100s camera using Philips smart bulbs as emissive sources. Quantitative assessments are based on 50 test images from novel camera poses, except for MSE measured for 25 test

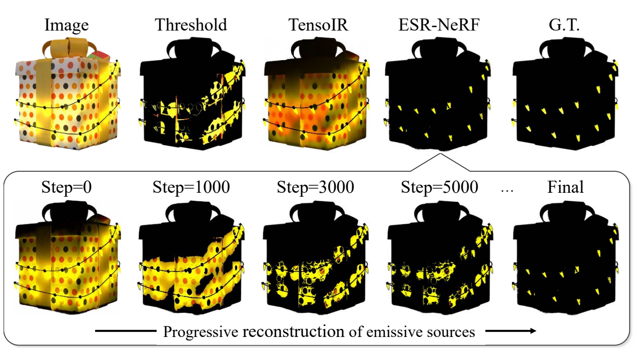


Figure 6. Comparison of identified emissive sources. ESR-NeRF excels through the reflection-aware progressive refinement.

	NV		NV + I		NV+ C		NV + I + C	
	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS
Twins	36.52	0.0141	**27.91**	**0.0252**	**31.02**	**0.0252**	**28.21**	**0.0310**
NeRF-W	36.44	0.0142	24.77	0.0417	-	-	-	-
NeILF++	24.40	0.0556	24.71	0.0579	24.06	0.0750	23.24	0.0770
TensoIR	**38.04**	**0.0103**	27.28	0.0418	26.36	0.0505	25.18	0.0531
PaletteNeRF	33.66	0.0233	23.27	0.0483	24.44	0.0646	22.58	0.0703
ESR-NeRF	**38.79**	**0.0083**	**29.99**	**0.0193**	**31.73**	**0.0196**	**31.63**	**0.0199**

Table 3. Scene editing results. NV: novel view synthesis, I: intensity editing, and C: color editing. A higher PSNR or lower LPIPS value is better.

images. We denote the best performance with blue and the second-best with green. Additionally, we utilize the DTU dataset [23] to evaluate ESR-NeRF's performance in surface reconstruction tasks where emissive sources are absent.

Baselines. We select two state-of-the-art re-lighting methods, TensoIR [25] and NeILF++ [90], that do not require prior lighting information. For thorough evaluation, we also implement a simple method, Twins, where separate models are trained under light on and off conditions. The Twins utilize the radiance discrepancies between the on and off models to distinguish and adjust emissive sources. For scene editing, we add NeRF-W [38] and PaletteNeRF [28] as baselines. Both NeRF-W and Twins adopt the Voxurf [80] architecture for fair comparison. For methods unable to individually control emissive sources, all sources are adjusted together to match the last lighting condition by a user. For the DTU dataset, we include state-of-the-art surface reconstruction methods that use object masks, such as NeuS [68] and Voxurf, as well as Neural-PBIR [57], that jointly reconstructs surfaces, materials, and environment maps.

5.2. Results

Emissive Source Recosntruction. Tab. 2 shows that our approach excels in accurately identifying emissive source regions and restoring their intensity, regardless of the source color. While TensoIR and NeILF++ can restore emissions by modifying their physical rendering equations, they suffer from emissive source ambiguity, leading to near-zero IoU performance (see Appendix). For a comprehensive comparison, we report the best performance of the baseline methods using thresholding on the reconstructed emission strength at 0.01 intervals. ESR-NeRF consistently outper-

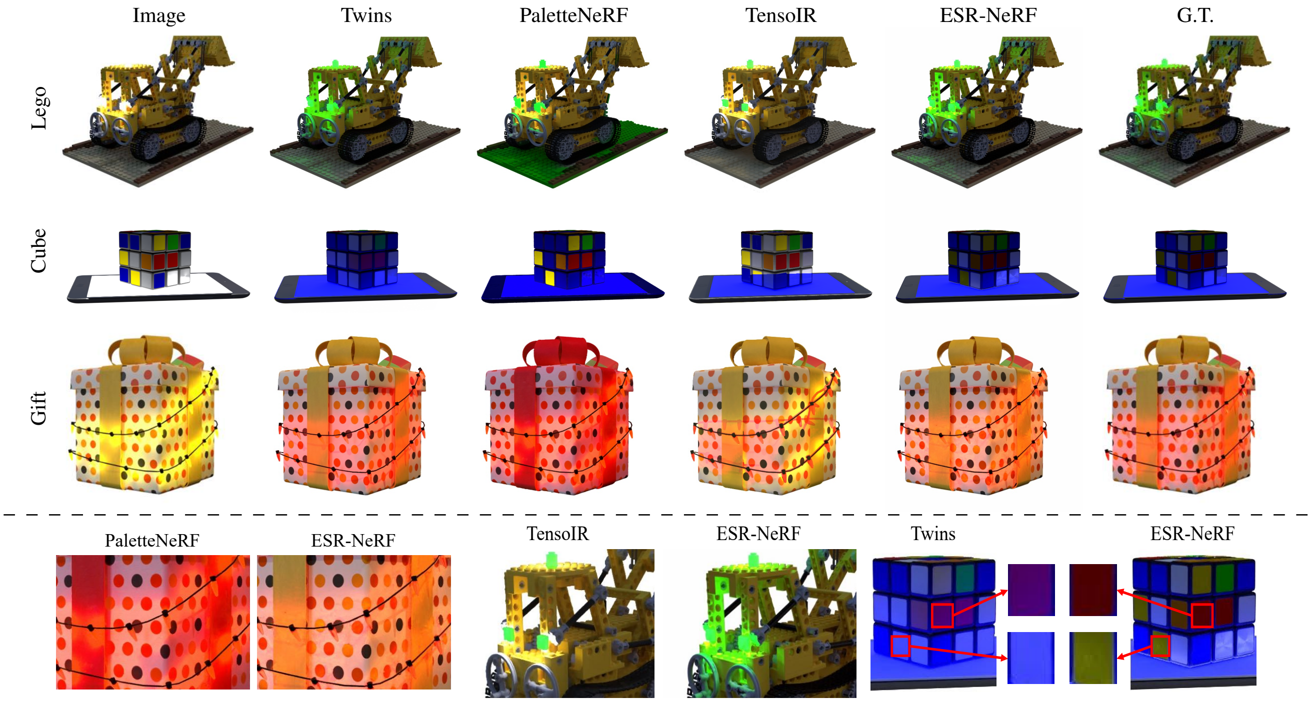


Figure 7. Comparison of scene editing. ESR-NeRF provides precise source control and faithfully represents reflection effects. For easy comparison in the Cube scene of low intensity, the bottom-right images are presented with a 40% increased brightness.

forms the baselines in identifying emissive source regions across all scenes. Our method also achieves significantly lower MSE values for restoring LDR to HDR images compared to the baselines, demonstrating its effectiveness of handling the ill-posed nature of the scenes with emissive sources. This is visually confirmed in Fig. 6, where ESR-NeRF surpasses the baselines in a complex scene with numerous small light bulbs.

Scene Editing. Tab. 3 and Fig. 7 showcase the scene editing results under novel lighting conditions. Baseline methods struggle to adapt to lighting changes due to their inability to reconstruct emissive sources accurately. For example, in the Lego scene, TensoIR fails to adjust the illumination in surrounding regions when the color of emissive sources is changed, and in the Cube scene, both the hidden iPad screen and the cube surface covered by the user input mask change together. Twins introduces blue light onto yellow and red surfaces, leading to unintended white and purple appearances, even though there should be no reflection. PaletteNeRF, which manipulates scenes through re-colorization, lacks precise control over illumination, as seen in the synchronous color changes in the yellow ribbon and lighting. In contrast, ESR-NeRF demonstrates superior performance in scene editing outshining all baselines thanks to the accurate identification of emissive sources, as detailed in Table 3. ESR-NeRF effectively balances source reconstruction and novel view synthesis, ensuring high performance in both tasks. NeRF-W is excluded from color adjustments since it doesn't support direct color change through interpolating latent variables learned with light on and off conditions.

Fig. 8 to 9 present additional examples of emissive source reconstruction and scene editing results. Fig. 10 shows results on real scenes, for which due to the impracticality of precise control over smart bulb colors, we offer emission reconstruction results with pseudo ground truth data. Our method effectively identifies emissive sources in real scenes, while it faces challenges in capturing complex

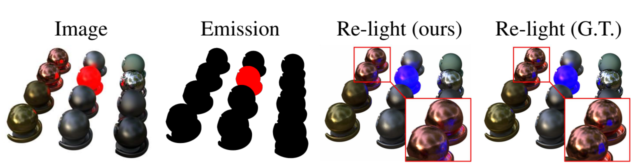


Figure 8. Reconstructed emitter and re-lighting at novel view.

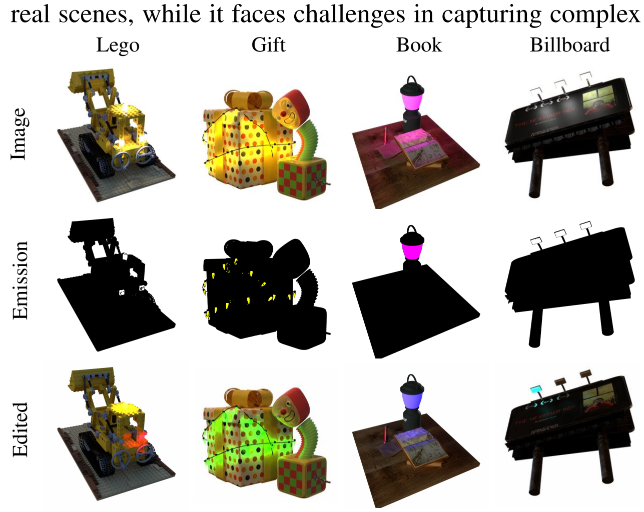


Figure 9. Results of source reconstruction and scene editing.

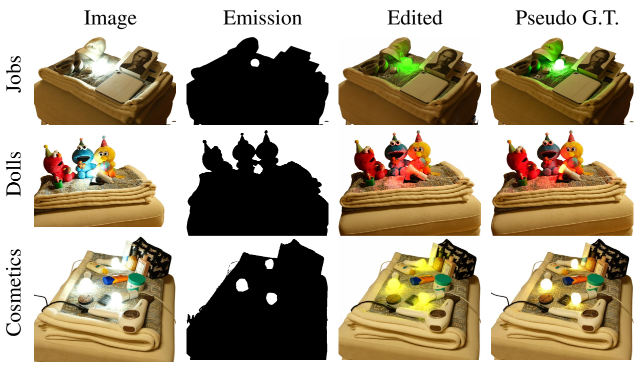


Figure 10. Source reconstruction and scene editing on real scenes.

reflections within light bulbs, as evident in the bright spot at the center of the bulbs in the ground truth edit results.

Ablation Analysis. Progressive refinement with the stop-gradient operation in Eq. 15 improves the identification of emissive sources and reduces MSE values. Without m_θ, surface reconstructions become unreliable, complicating the accurate reconstruction of emissive sources. This issue is evident from the CD metrics and illustrated in Fig. 3. Further analyses are provided in Appendix.

Illumination Decomposition. Fig. 11 demonstrates ESR-NeRF's decomposition of scene illumination into direct and indirect lighting from an environment map, as well as emissions and their reflections. The shadow behind the yellow ribbon in the direct figure and the illumination in the indirect figure showcase ESR-NeRF's ability to model both direct and indirect illumination. The reflection figure shows that our method accurately captures how emissive sources contribute to reflections on nearby regions.

Surface Reconstruction. Interestingly, our approach can be applied to the scenes without emissive sources to enhance surface reconstruction, as evidenced by the lower CD values in Tab. 4 on the DTU dataset. For this experiment, we use Eq. 7 to 10 without our progressive refinement technique. ESR-NeRF's ability to adjust interrelated outgoing radiances helps prevent surface formations where radiances cannot be produced, considering the predicted scene context. Additional visualizations of the normals, BRDF, and environment maps are provided in Appendix.

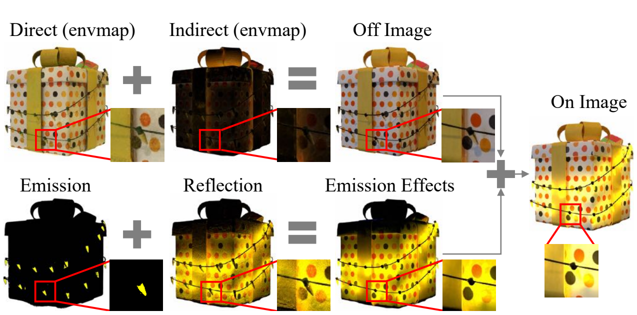


Figure 11. An example of illumination decomposition.

Scan	NeuS	Voxurf	Neural-PBIR	ESR-NeRF
24	0.83	0.65	**0.57**	**0.58**
37	0.98	0.74	0.75	**0.71**
40	0.56	**0.39**	**0.38**	**0.38**
55	0.37	**0.35**	0.36	**0.33**
63	1.13	0.96	1.04	**0.93**
65	0.59	0.64	0.73	**0.57**
69	**0.60**	0.85	**0.65**	0.78
83	1.45	1.58	**1.28**	**1.18**
97	**0.95**	1.01	0.97	**0.95**
105	0.78	**0.68**	0.76	**0.58**
106	**0.52**	0.60	**0.53**	0.54
110	1.43	1.11	**0.84**	**1.08**
114	0.36	0.37	0.38	**0.33**
118	0.45	0.45	0.46	**0.40**
122	**0.45**	0.47	0.49	**0.44**
mean	0.77	0.72	**0.68**	**0.65**

Table 4. Results of surface reconstruction via the Chamfer distance on the DTU dataset. A lower value is better.

	White		Vivid	
	IoU ↑	MSE ↓	IoU ↑	MSE ↓
w/o progressive	0.40	9.92	0.41	3.93
w/o sg	0.71	6.45	0.60	3.47
ESR-NeRF	0.86	5.24	0.81	2.79

	DTU CD ↓
w/o m_θ	0.93
w/o LTS	0.71
ESR-NeRF	0.65

Table 5. Ablation studies on the sruface reconstruction (left) and the emissive source reconstruction (right).

6. Conclusion

We present ESR-NeRF as the first NeRF-based inverse rendering method for the scenes with emissive sources. Our approach uses LDR images, eliminating the need of HDR images to reconstruct emissive sources. Furthermore, we demonstrate the application of reconstructed sources in scene editing, enabling color and intensity modifications.

Limitations. Future work could explore using a single lighting condition to disentangle emissive sources, environmental lighting, and object texture. It is also promising to address the challenge of volume ray tracing in unbounded scenes to extend to indoor scenes. Additionally, LTS based re-lighting may be weak in representing new colors that traverse unobserved light paths during training. An alternative approach could be extracting emission texture maps and modifying it using the engines such as Blender [13] or Mitsuba [22]. More details on alternative re-lighting methods and radiance fine-tuning are provided in Appendix.

7. Acknowledgements

This work was supported by Samsung Electronics MX, Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(RS-2023-00274280), and Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2019-0-01082, SW StarLab; No. 2022-0-00156, Fundamental research on continual meta-learning for quality enhancement of casual videos and their 3D metaverse transformation). Gunhee Kim is the corresponding author.

References

[1] Benjamin Attal, Jia-Bin Huang, Michael Zollhöfer, Johannes Kopf, and Changil Kim. Learning neural light fields with ray-space embedding. In *CVPR*, 2022. 2

[2] Dejan Azinovic, Tzu-Mao Li, Anton Kaplanyan, and Matthias Nießner. Inverse path tracing for joint material and lighting estimation. In *CVPR*, 2019. 2

[3] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *ICCV*, 2021. 1

[4] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *CVPR*, 2022. 1

[5] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Zip-nerf: Anti-aliased grid-based neural radiance fields. In *ICCV*, 2023. 2

[6] Sai Bi, Zexiang Xu, Pratul Srinivasan, Ben Mildenhall, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. Neural reflectance fields for appearance acquisition. In *arXiv*, 2020. 1

[7] Mark Boss, Varun Jampani, Kihwan Kim, Hendrik Lensch, and Jan Kautz. Two-shot spatially-varying brdf and shape estimation. In *CVPR*, 2020.

[8] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T. Barron, Ce Liu, and Hendrik P.A. Lensch. Nerd: Neural reflectance decomposition from image collections. In *ICCV*, 2021. 1

[9] Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan T. Barron, and Hendrik P.A. Lensch. Neural-pil: Neural pre-integrated lighting for reflectance decomposition. In *NeurIPS*, 2021. 2

[10] Brent Burley and Walt Disney Animation Studios. Physically-based shading at disney. In *SIGGRAPH*, 2012. 5

[11] Bowen Cai, Jinchi Huang, Rongfei Jia, Chengfei Lv, and Huan Fu. Neuda: Neural deformable anchor for high-fidelity implicit surface reconstruction. In *CVPR*, 2023. 2

[12] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *ECCV*, 2022. 2

[13] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018. 8

[14] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *CVPR*, 2022. 2

[15] Mathieu Garon, Kalyan Sunkavalli, Sunil Hadap, Nathan Carr, and Jean-Francois Lalonde. Fast spatially-varying indoor lighting estimation. In *CVPR*, 2019. 1

[16] Wenhang Ge, Tao Hu, Haoyu Zhao, Shu Liu, and Ying-Cong Chen. Ref-neus: Ambiguity-reduced neural implicit surface learning for multi-view reconstruction with reflection. In *ICCV*, 2023. 1

[17] Saeed Hadadan, Shuhong Chen, and Matthias Zwicker. Neural radiosity. In *ACM TOG*, 2021. 2

[18] Saeed Hadadan, Geng Lin, Jan Novák, Fabrice Rousselle, and Matthias Zwicker. Inverse global illumination using a neural radiometric prior. In *SIGGRAPH Conference Proceedings*, 2023. 1

[19] Bjoern Haefner, Simon Green, Alan Oursland, Daniel Andersen, Michael Goesele, Daniel Cremers, Richard Newcombe, and Thomas Whelan. Recovering real-world reflectance properties and shading from hdr imagery. In *3DV*, 2021. 2

[20] Jon Hasselgren, Nikolai Hofmann, and Jacob Munkberg. Shape, light, and material decomposition from images using monte carlo rendering and denoising. In *NeurIPS*, 2022. 2

[21] IEC. IEC 61966-2-1:1999. Technical report, International Electrotechnical Commission, 1999. 3

[22] Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, Merlin Nimier-David, Delio Vicini, Tizian Zeltner, Baptiste Nicolet, Miguel Crespo, Vincent Leroy, and Ziyi Zhang. Mitsuba 3 renderer, 2022. https://mitsuba-renderer.org. 8

[23] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engil Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *CVPR*, 2014. 2, 6

[24] Chaonan Ji, Tao Yu, Kaiwen Guo, Jingxin Liu, and Yebin Liu. Geometry-aware single-image full-body human relighting. In *ECCV*, 2022. 1

[25] Haian Jin, Isabella Liu, Peijia Xu, Xiaoshuai Zhang, Songfang Han, Sai Bi, Xiaowei Zhou, Zexiang Xu, and Hao Su. Tensoir: Tensorial inverse rendering. In *CVPR*, 2023. 2, 3, 6

[26] James T. Kajiya. The rendering equation. In *SIGGRAPH*, 1986. 3

[27] Petr Kellnhofer, Lars C Jebe, Andrew Jones, Ryan Spicer, Kari Pulli, and Gordon Wetzstein. Neural lumigraph rendering. In *CVPR*, 2021. 1

[28] Zhengfei Kuang, Fujun Luan, Sai Bi, Zhixin Shu, Gordon Wetzstein, and Kalyan Sunkavalli. Palettenerf: Palette-based appearance editing of neural radiance fields. In *CVPR*, 2023. 6

[29] Zhengqin Li, Mohammad Shafiei, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image. In *CVPR*, 2020. 1

[30] Zhengqin Li, Jia Shi, Sai Bi, Rui Zhu, Kalyan Sunkavalli, Miloš Hašan, Zexiang Xu, Ravi Ramamoorthi, and Manmohan Chandraker. Physically-based editing of indoor scene lighting from a single image. In *ECCV*, 2022. 1

[31] Zhaoshuo Li, Thomas Müller, Alex Evans, Russell H Taylor, Mathias Unberath, Ming-Yu Liu, and Chen-Hsuan Lin. Neuralangelo: High-fidelity neural surface reconstruction. In *CVPR*, 2023. 2

[32] Zhen Li, Lingli Wang, Mofang Cheng, Cihui Pan, and Jiaqi. Yang. Multi-view inverse rendering for large-scale real-world indoor scenes. In *CVPR*, 2023. 2

[33] Ruofan Liang, Huiting Chen, Chunlin Li, Fan Chen, Selvakumar Panneer, and Nandita Vijaykumar. Envidr: Implicit differentiable renderer with neural environment lighting. In *ICCV*, 2023. 1

[34] Fujun Luan, Shuang Zhao, Kavita Bala, and Zhao Dong. Unified shape and svbrdf recovery using differentiable monte carlo rendering. In *EGSR*, 2021. 1

[35] Linjie Lyu, Ayush Tewari, Thomas Leimkuehler, Marc Habermann, and Christian Theobalt. Neural radiance transfer fields for relightable novel-view synthesis with global illumination. In *ECCV*, 2022. 2

[36] Linjie Lyu, Ayush Tewari, Marc Habermann, Shunsuke Saito, Michael Zollhöfer, Thomas Leimkühler, and Christian Theobalt. Diffusion posterior illumination for ambiguity-aware inverse rendering. In *ACM TOG*, 2023. 1

[37] Alexander Mai, Dor Verbin, Falko Kuester, and Sara Fridovich-Keil. Neural microfacet fields for inverse rendering. In *ICCV*, 2023. 1, 3

[38] Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *CVPR*, 2021. 1, 6

[39] Nelson Max. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1(2):99–108, 1995. 2

[40] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 1, 2

[41] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. In *ACM TOG*, 2022. 2

[42] Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Müller, and Sanja Fidler. Extracting triangular 3d models, materials, and lighting from images. In *CVPR*, 2022. 2

[43] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *CVPR*, 2020. 1

[44] Michael Oechsle, Songyou Peng, and Andreas Geiger. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *ICCV*, 2021. 2

[45] Julian Ost, Issam Laradji, Alejandro Newell, Yuval Bahat, and Felix Heide. Neural point light fields. In *CVPR*, 2022. 1

[46] Rohit Pandey, Sergio Orts-Escolano, Chloe LeGendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Seann Fanello. Total relighting: Learning to relight portraits for background replacement. In *ACM TOG*, 2021. 2

[47] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *CVPR*, 2019. 1

[48] Julien Philip, Sébastien Morgenthaler, Michaël Gharbi, and George Drettakis. Free-viewpoint indoor neural relighting from multi-view stereo. In *ACM TOG*, 2021. 2

[49] Martin Piala and Ronald Clark. Terminerf: Ray termination prediction for efficient neural rendering. In *3DV*, 2021. 2

[50] Viktor Rudnev, Mohamed Elgharib, William Smith, Lingjie Liu, Vladislav Golyanik, and Christian Theobalt. Nerf for outdoor scene relighting. In *ECCV*, 2022. 1

[51] Vincent Sitzmann, Julien N.P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *NeurIPS*, 2020. 2

[52] Vincent Sitzmann, Semon Rezchikov, Bill Freeman, Josh Tenenbaum, and Fredo Durand. Light field networks: Neural scene representations with single-evaluation rendering. In *NeurIPS*, 2021. 2

[53] Alvy Ray Smith. Color gamut transform pairs. In *ACM TOG*, 1978. 5

[54] Pratul P. Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T. Barron. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In *CVPR*, 2021. 2, 3

[55] Mohammed Suhail, Carlos Esteves, Leonid Sigal, and Ameesh Makadia. Light field neural rendering. In *CVPR*, 2022. 2

[56] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In *CVPR*, 2022. 2

[57] Cheng Sun, Guangyan Cai, Zhengqin Li, Kai Yan, Cheng Zhang, Carl Marshall, Jia-Bin Huang, Shuang Zhao, and Zhao Dong. Neural-pbir reconstruction of shape, material, and illumination. In *ICCV*, 2023. 2, 3, 6

[58] Jiaming Sun, Xi Chen, Qianqian Wang, Zhengqi Li, Hadar Averbuch-Elor, Xiaowei Zhou, and Noah Snavely. Neural 3D reconstruction in the wild. In *SIGGRAPH Conference Proceedings*, 2022. 1

[59] Tiancheng Sun, Kai-En Lin, Sai Bi, Zexiang Xu, and Ravi Ramamoorthi. Nelf: Neural light-transport field for portrait view synthesis and relighting. In *EGSR*, 2021. 2

[60] Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *NeurIPS*, 2020. 2

[61] Jiaxiang Tang, Hang Zhou, Xiaokang Chen, Tianshu Hu, Errui Ding, Jingdong Wang, and Gang Zeng. Delicate textured mesh recovery from nerf via adaptive surface refinement. In *ICCV*, 2023. 1

[62] Ayush Tewari, Ohad Fried, Justus Thies, Vincent Sitzmann, Stephen Lombardi, Kalyan Sunkavalli, Ricardo Martin-Brualla, Tomas Simon, Jason Saragih, Matthias Nießner, et al. State of the art on neural rendering. In *CGF*, 2020. 1

[63] Ayush Tewari, Justus Thies, Ben Mildenhall, Pratul Srinivasan, Edgar Tretschk, Wang Yifan, Christoph Lassner, Vincent Sitzmann, Ricardo Martin-Brualla, Stephen Lombardi, et al. Advances in neural rendering. In *CGF*, 2022. 1

[64] Marco Toschi, Riccardo De Matteo, Riccardo Spezialetti, Daniele De Gregorio, Luigi Di Stefano, and Samuele Salti. Relight my nerf: A dataset for novel view synthesis and relighting of real world objects. In *CVPR*, 2023. 2

[65] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *CVPR*, 2022. 1

[66] Dongqing Wang, Tong Zhang, and Sabine Süsstrunk. Nemto: Neural environment matting for novel view and relighting synthesis of transparent objects. In *ICCV*, 2023. 1
[67] Jiaping Wang, Peiran Ren, Minmin Gong, John Snyder, and Baining Guo. All-frequency rendering of dynamic, spatially-varying reflectance. *ACM TOG*, 2009. 4
[68] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *NeurIPS*, 2021. 2, 6
[69] Shaofei Wang, Katja Schwarz, Andreas Geiger, and Siyu Tang. Arah: Animatable volume rendering of articulated human sdfs. In *ECCV*, 2022. 1
[70] Yiming Wang, Qin Han, Marc Habermann, Kostas Daniilidis, Christian Theobalt, and Lingjie Liu. Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction. In *ICCV*, 2023. 2
[71] Zian Wang, Jonah Philion, Sanja Fidler, and Jan Kautz. Learning indoor inverse rendering with 3d spatially-varying lighting. In *ICCV*, 2021. 1
[72] Zian Wang, Wenzheng Chen, David Acuna, Jan Kautz, and Sanja Fidler. Neural light field estimation for street scenes with differentiable virtual object insertion. In *ECCV*, 2022. 1
[73] Zian Wang, Tianchang Shen, Jun Gao, Shengyu Huang, Jacob Munkberg, Jon Hasselgren, Zan Gojcic, Wenzheng Chen, and Sanja Fidler. Neural fields meet explicit geometric representations for inverse rendering of urban scenes. In *CVPR*, 2023. 1
[74] Henrique Weber, Mathieu Garon, and Jean-François Lalonde. Editable indoor lighting estimation. In *ECCV*, 2022. 1
[75] Haoqian Wu, Zhipeng Hu, Lincheng Li, Yongqiang Zhang, Changjie Fan, and Xin Yu. Nefii: Inverse rendering for reflectance decomposition with near-field indirect illumination. In *CVPR*, 2023. 2
[76] Liwen Wu, Rui Zhu, Mustafa B. Yaldiz, Yinhao Zhu, Hong Cai, Janarbek Matai, Fatih Porikli, Tzu-Mao Li, Manmohan Chandraker, and Ravi Ramamoorthi. Factorized inverse path tracing for efficient and accurate material-lighting estimation. In *ICCV*, 2023. 2
[77] Qianyi Wu, Xian Liu, Yuedong Chen, Kejie Li, Chuanxia Zheng, Jianfei Cai, and Jianmin Zheng. Object-compositional neural implicit surfaces. In *ECCV*, 2022. 2
[78] Qianyi Wu, Kaisiyuan Wang, Kejie Li, Jianmin Zheng, and Jianfei Cai. Objectsdf++: Improved object-compositional neural implicit surfaces. In *ICCV*, 2023. 2
[79] Tong Wu, Jia-Mu Sun, Yu-Kun Lai, and Lin Gao. De-nerf: Decoupled neural radiance fields for view-consistent appearance editing and high-frequency environmental relighting. In *SIGGRAPH ASIA Conference Proceedings*, 2023. 1
[80] Tong Wu, Jiaqi Wang, Xingang Pan, Xudong Xu, Christian Theobalt, Ziwei Liu, and Dahua Lin. Voxurf: Voxel-based efficient and accurate neural surface reconstruction. In *ICLR*, 2023. 2, 5, 6
[81] Yuanbo Xiangli, Linning Xu, Xingang Pan, Nanxuan Zhao, Anyi Rao, Christian Theobalt, Bo Dai, and Dahua Lin. Bungeenerf: Progressive neural radiance field for extreme multi-scale scene rendering. In *ECCV*, 2022. 1
[82] Yingyan Xu, Gaspard Zoss, Prashanth Chandran, Markus Gross, Derek Bradley, and Paulo Gotardo. Renerf: Relightable neural radiance fields with nearfield lighting. In *ICCV*, 2023. 2
[83] Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K. Wong. Ps-nerf: Neural inverse rendering for multi-view photometric stereo. In *ECCV*, 2022. 2
[84] Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K. Wong. S^3-nerf: Neural reflectance field from shading and shadow under a single viewpoint. In *NeurIPS*, 2022. 2
[85] Yao Yao, Jingyang Zhang, Jingbo Liu, Yihang Qu, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. Neilf: Neural incident light field for material and lighting estimation. In *ECCV*, 2022. 1
[86] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. In *NeurIPS*, 2020. 1
[87] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. In *NeurIPS*, 2021. 2
[88] Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. PlenOctrees for real-time rendering of neural radiance fields. In *ICCV*, 2021. 2
[89] Chong Zeng, Guojun Chen, Yue Dong, Pieter Peers, Hongzhi Wu, and Xin Tong. Relighting neural radiance fields with shadow and highlight hints. In *SIGGRAPH Conference Proceedings*, 2023. 2
[90] Jingyang Zhang, Yao Yao, Shiwei Li, Jingbo Liu, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. Neilf++: Inter-reflectable light fields for geometry and material estimation. In *ICCV*, 2023. 2, 6
[91] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020. 1
[92] Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *CVPR*, 2021. 2
[93] Kai Zhang, Fujun Luan, Zhengqi Li, and Noah Snavely. Iron: Inverse rendering by optimizing neural sdfs and materials from photometric images. In *CVPR*, 2022. 1
[94] Xiuming Zhang, Sean Fanello, Yun-Ta Tsai, Tiancheng Sun, Tianfan Xue, Rohit Pandey, Sergio Orts-Escolano, Philip Davidson, Christoph Rhemann, Paul Debevec, et al. Neural light transport for relighting and view synthesis. In *ACM TOG*, 2021. 2
[95] Xiuming Zhang, Pratul P. Srinivasan, Boyang Deng, Paul Debevec, William T. Freeman, and Jonathan T. Barron. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. In *ACM TOG*, 2021. 2
[96] Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. Modeling indirect illumination for inverse rendering. In *CVPR*, 2022. 2

[97] Y. Zhang, Z. Hu, H. Wu, M. Zhao, L. Li, Z. Zou, and C. Fan. Towards unbiased volume rendering of neural implicit surfaces with geometry priors. In *CVPR*, 2023. 2

[98] Quan Zheng, Gurprit Singh, and Hans-Peter Seidel. Neural relightable participating media rendering. In *NeurIPS*, 2021. 2

[99] Jingsen Zhu, Fujun Luan, Yuchi Huo, Zihao Lin, Zhihua Zhong, Dianbing Xi, Rui Wang, Hujun Bao, Jiaxiang Zheng, and Rui Tang. Learning-based inverse rendering of complex indoor scenes with differentiable monte carlo raytracing. In *SIGGRAPH ASIA Conference Proceedings*, 2022. 2