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

Supplementary Material

Contents

1. Appendix

1.1. Implementation Details

Training Procedure. Our implementation builds upon Voxurf [17], excluding its dual-color network feature. We adhere to the coarse and fine processing stages described in Voxurf before initiating our LTS learning-based training strategy. Additionally, we compute ray colors using alpha masks to filter out points in empty space, aligning with practices in previous studies [2, 6, 17]. The LTS learning training procedure with progressive refinement approach is:

1. Initialize ray groups: Uncertain rays $R_0^U = R$ and certain rays $R_0^C = \emptyset$.
2. Form mini-batches using stratified sampling within each ray group.
3. Calculate the rendering loss, $\mathcal{L}_{render}$.
4. For rays in the mini-batch, uniformly sample 100 points to evaluate the LTS loss, $\mathcal{L}_{lts}$.
5. Compute the surface normal at sampled points.
6. For $L_o^E(x, \omega_o)$, sample an additional viewing direction on the upper hemisphere at these points.
7. For $\hat{L}_o^E(x, \omega_o)$, sample 256 rays on the upper hemisphere at these points to compute incident radiance.
8. Calculate $\mathcal{L}_{lts}$, considering the group membership of each point.
9. Update network parameters.
10. Adjust ray groups at specified training intervals.
11. Repeat steps 2 through 10 until training ends.

Discretization. Following NeuS [16], we approximate ray color computation using N discrete points sampled along the ray, denoted as $\{x_i = c - t_i\omega_o | i = 1, ..., N, t_i < t_{i+1}\}$:

$$\hat{C}(r) = \sum_{i=1}^{N} T_i \alpha_i L_o(x_i, \omega_o), \tag{1}$$

$$\alpha_i = \max\left(\frac{\Phi_s(f(x_i)) - \Phi_s(f(x_{i+1}))}{\Phi_s(f(x_i))}, 0\right), \tag{2}$$

$$T_i = \prod_{j=1}^{i-1}(1 - \alpha_j). \tag{3}$$

α is the discrete equivalent of the SDF-based opacity, ρ.

For reflections in $\hat{L}_o(x, \omega)$, we employ Monte Carlo sampling, uniformly sampling directions ω_i around the normal n at point x on the upper hemisphere. While the current implementation of ESR-NeRF doesn't include importance sampling for incident rays, incorporating it in future work for variance reduction may enhance overall performance.

$$\hat{L}_o(x, \omega_o) = E(x) + \frac{1}{M}\sum_{j=1}^{M}\left(\frac{L_i(x, \omega_j)R(x, \omega_o, \omega_j)}{\frac{1}{2\pi}}\right). \tag{4}$$

Simplified Diseny BRDF. We adopt the simplified Disney principled BRDF function [15], parameterized by base color b, metallic m, and roughness r.

$$\begin{aligned} R(x, \omega_o, \omega_i) = &\frac{D(h, n, r)F(\omega_o, h, b, m)G(\omega_o, \omega_i, h, r)}{4(n \cdot \omega_o)} \\ &+ (n \cdot \omega_i)(1 - m)\left(\frac{b}{\pi}\right), \end{aligned} \tag{5}$$

The half vector h is defined as $h = \frac{\omega_o + \omega_i}{\|\omega_o + \omega_i\|_2}$. Following NeILF++ [18], the normal distribution function D is approximated using Spherical Gaussian:

$$D(h, n, r) = \frac{1}{\pi r^4}\exp(\frac{2}{r^4}(h \cdot n - 1)), \tag{6}$$

The Fresnel term F is calculated as follows:

$$\begin{aligned} F(\omega_o, h, b, m) &= F_0 + (1 - F_0)(1 - (\omega_o \cdot h)^5), \\ \text{where } F_0 &= 0.04(1 - m) + bm, \end{aligned} \tag{7}$$

The geometry term G adopts the GGX function [1].

$$\begin{aligned} G(\omega_o, \omega_i, n, r) &= \frac{(n \cdot \omega_o)(n \cdot \omega_i)}{((n \cdot \omega_o)(1 - k) + k)((n \cdot \omega_i)(1 - k) + k)}, \\ \text{where } k &= \frac{r^2}{2}. \end{aligned} \tag{8}$$

For simplicity, our BRDF model incorporates the Lambert cosine term $(n \cdot \omega_i)$.

Gamma Correction. To ensure HDR linear color space for outgoing radiance, we apply the standard gamma correction as defined by IEC [4] to ray colors before calculating the

rendering loss. The gamma-corrected sRGB color, given a linear color C_{linear}, is computed as follows:

$$\tau(C_{\text{linear}}) = \begin{cases} 12.92C_{\text{linear}} & \text{if } C_{\text{linear}} \leq 0.0031308, \\ 1.055C_{\text{linear}}^{1/2.4} - 0.055 & \text{if } C_{\text{linear}} > 0.0031308. \end{cases} \tag{9}$$

RGB to HSV. For scene editing tasks, we utilize the HSV color model [11]. The hue ($H \in [0, 1]$), saturation ($S \in [0, 1]$), and value ($V \in \mathbb{R}_+$) are calculated using the following method:

$$\begin{aligned} M &= \max(R, G, B), \\ m &= \min(R, G, B), \\ C &= M - m. \end{aligned} \tag{10}$$

$$H = (H'/6.0) \mod 1.0,$$

$$H' = \begin{cases} 0 & \text{if } C = 0, \\ \frac{G-B}{C} & \text{if } M = R, \\ \frac{B-R}{C} + 2 & \text{if } M = G, \\ \frac{R-G}{C} + 4 & \text{if } M = B. \end{cases} \tag{11}$$

$$S = \begin{cases} 0 & \text{if } V = 0, \\ \frac{C}{V} & \text{otherwise}, \end{cases} \tag{12}$$

$$V = \max(R, G, B) \tag{13}$$

HSV to RGB. Once the color is replaced and intensity is adjusted in the HSV space, the conversion back to RGB is performed as:

$$\begin{aligned} m &= V - C, \\ H' &= H \times 6.0, \\ C &= S \times V, \\ X &= C \times (1 - |H' \mod 2 - 1|), \end{aligned}$$

$$(R', G', B') = \begin{cases} (C, X, 0) & \text{if } 0 \leq H' < 1, \\ (X, C, 0) & \text{if } 1 \leq H' < 2, \\ (0, C, X) & \text{if } 2 \leq H' < 3, \\ (0, X, C) & \text{if } 3 \leq H' < 4, \\ (X, 0, C) & \text{if } 4 \leq H' < 5, \\ (C, 0, X) & \text{if } 5 \leq H' < 6. \end{cases} \tag{14}$$

$$(R, G, B) = (R' + m, G' + m, B' + m). \tag{15}$$

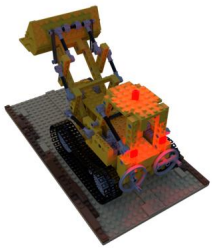
1.2. Dataset Details

Dataset Construction. This section outlines the dataset used for training and evaluation. Each scene in our dataset comprises 200 training images, with an equal split between two lighting conditions: "on" and "off". Emission masks are utilized as ground truth for emissive source identification, while EXR files with linear pixel values assess the accuracy of the reconstructed strength of emission and reflection. All data are rendered using the Cycles path tracing in Blender [3], with settings that could artificially alter scene illumination are disabled, such as incdient light clamping and the Filmic transform. For scene editing under novel lighting conditions, we introduce a variety of test scenarios, including intensity editing, color editing, and combined intensity and color editing, each with 50 images. We derive these scenarios from 25 unique camera positions from the novel view evaluation dataset, each under two different lighting conditions, Intensity adjustments are made relative to the original scene's emissive source strength, with "0" indicating "light off" and "1" matching the "light on" intensity. We test intensity adjustments at half (0.5) and double (2.0) the original levels. In scenes allowing individual source adjustments, we include an additional intensity condition where lights are selectively turned off (0.0). For color editing, we select six colors—red, green, blue, cyan, magenta, and yellow—to demonstrate the effects of various light source colors on scene illumination.

Scene Characteristics. Our scenes are meticulously crafted using assets from Blendswap and cgtrader, with licensing details and the count of emissive sources detailed in Tab. 1. Below, we describe the unique aspects of each scene.

- **LEGO**: This scene showcases three emissive sources, all starting with the same color and intensity. The intricate designs of the LEGO bricks create complex reflection effects. The emissive sources in these scenes are tested for both collective and individual adjustments.

(a) Lego (white)



(b) Lego (vivid)

- **Gift**: Featuring a gift box, a toy, and numerous small bulbs, this scene presents a challenge with its multitude of tiny light bulbs and extensive reflection areas.

(a) Gift (white)

(b) Gift (vivid)

Scene Name	Num Lights	License
Lego	3	By Heinzelnisse (CC-BY-NC): https://www.blendswap.com/blend/11490
Gift	29	By juan215 (Royalty Free): https://www.cgtrader.com/free-3d-models/household/household-tools/gift-box-aeb8f01e-929f-4041-9117-bcea21f3c813 By MiriamAHoyt (CC-0): https://blendswap.com/blend/21434
Book	1	By lakerice (CC-0): https://blendswap.com/blend/22197 By 3dfiles (CC-BY): https://blendswap.com/blend/28034 By bloknayrb (CC-BY): https://www.blendswap.com/blend/26172
Cube	1	By 4NDR31JK (CC-BY): https://www.blendswap.com/blend/30149 By sriniwasjha (CC-BY): https://blendswap.com/blend/18409
Billboard	6	By M0h4wkAD3 (CC-BY-NC-SA): https://blendswap.com/blend/27481
Balls	1	By elbrujodelatribu (CC-0): https://blendswap.com/blend/10120

Table 1. Number of emissive sources and licenses of objects used in scenes.

- **Book**: The Book scene features a single large light source consisting of a lamp, a book, and a pencil. The emphasis here is on identifying and restoring the very large emissive source.

(a) Book (white)

(b) Book (vivid)

- **Cube**: Comprising a tablet PC and a cube, this scene is marked by its sophisticated reflection effects, especially on the cube surfaces which varying albedo.

(a) Cube (white)

(b) Cube (vivid)

- **Billboard**: This scene includes two billboards, each equipped with three emissive sources, summing up to six sources. The lights are positioned to shine downwards from the billboards' tops. We adjust the emissive sources collectively and individually. Individual adjustments are performed for three light groups by pairing the light sources of the front and back billboards.
- **Balls**: This is the material balls scene in NeRF, with the modification of the red ball as an emissive source.

(a) Billboard (white)

(b) Billboard (vivid)

(a) Balls (white)

(b) Balls (vivid)

1.3. Baseline Implementation

In our evaluation, we compared against two leading relighting methods, TensoIR [6] and NeILF++ [18], known for their ability to operate without prior knowledge of scene components Additionally, we made Twins, a method focused on emissive source reconstruction without relying on inverse rendering techniques. For an in-depth analysis of scene editing capabilities, we include PaletteNeRF [7], which achieves scene modification through re-colorization, and NeRF-W [8], which adjusts scene illumination by interpolating between learned latent vectors. For surface reconstruction evaluations on the DTU dataset, we selected state-of-the-art methods such as Voxurf [17] and NeuS [16], alongside Neural-PBIR [14], which offers a joint reconstruction of surfaces, materials, and environment maps. We utilized the official implementations provided by the au-

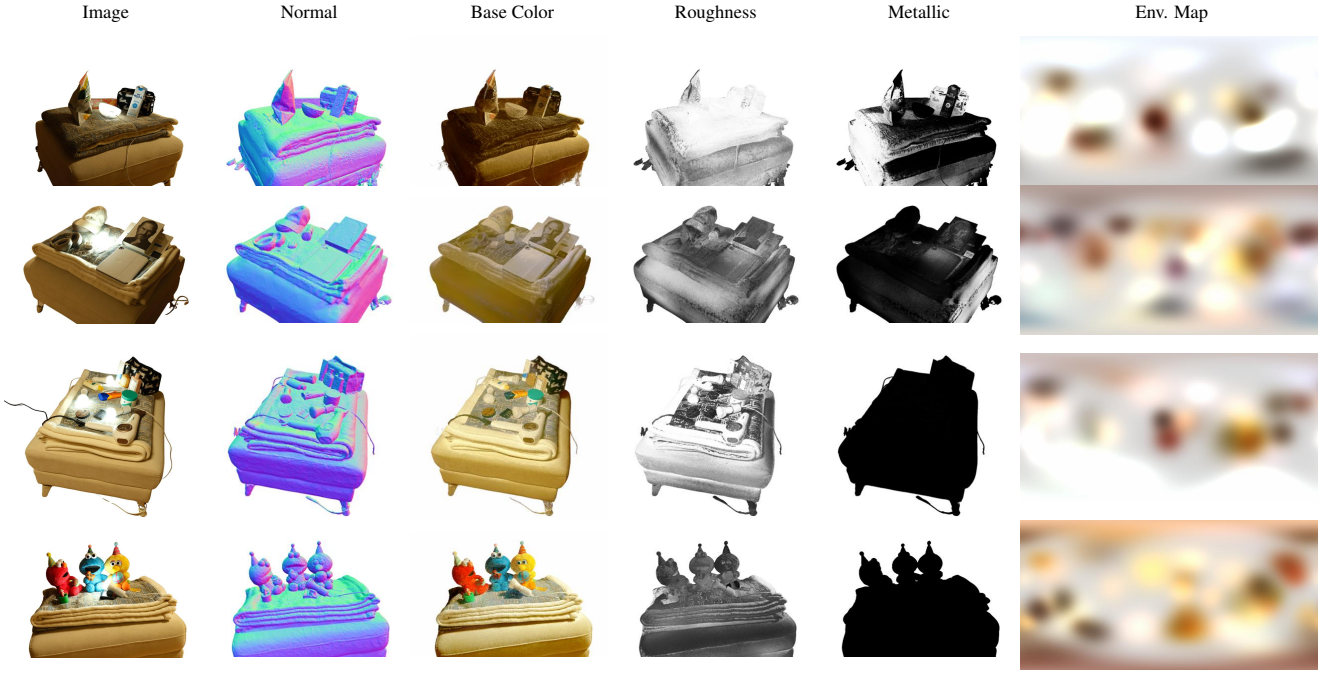


Figure 7. Decomposed scene components on real scenes

thors for all baselines, with the exception of NeRF-W. We used the official implementation codes provided by the authors for all baseline methods except for Twins and NeRF-W. Twins employs a dual-model strategy for 'light-on' and 'light-off' conditions, using radiance differences for emissive source identification and scene illumination editing. NeRF-W leverages two latent embeddings for similar purposes, focusing on intensity adjustments. Both Twins and NeRF-W are based on the Voxurf architecture to ensure a fair comparison with ESR-NeRF. For NeILF++, we omitted the use of prior scene information to align with methods that do not use geometry hints like object meshes or oriented point clouds. Neural-PBIR was excluded from emissive source reconstruction experiments as the code is not publicly available yet. Baseline performance data on the DTU dataset are borrowed directly from the Voxurf, NeuS, and Neural-PBIR papers.

1.4. Real Scene

We showcase the effectiveness of ESR-NeRF in identifying emissive sources in real-world scenes. Camera poses are estimated using COLMAP [10]. We use commercial smart light bulbs from Philips, which offer control over light colors. Since precise control over the color of the smart bulbs is infeasible, we provide qualitative results for emissive source identification and scene editing in real scenes. Fig. 7 presents the decomposed scene components, such as normal, base color, roughness, metallic, and the environment map. In Fig. 8, our method successfully identifies emissive sources, enabling scene illumination adjustments. Fig. 9 presents qualitative results for comparison with ground truth data. Although our model successfully identifies emissive sources, it encounters difficulties with complex reflections inside light bulbs, as indicated by the bright spots at the bulb centers in the ground truth edit images. Despite these challenges, ESR-NeRF stands out as the first NeRF-based inverse rendering method to address the reconstruction of emissive sources, enabling scene illumination modifications through the identification of light sources within a scene.

1.5. Reconstructed Scene Components

We present the reconstructed components of our synthetic scenes, including emissions, surface normals, and BRDF, in Fig. 13 and 14. We also provide the comparison of the reconstructed emission and BRDF performance among TensoIR, NeILF++, and ESR-NeRF in Fig. 10 and 11 for real scenes and Fig. 15 to 19 for synthetic scenes. TensoIR and NeILF++ encounter difficulties, as does ESR-NeRF, in capturing precise roughness, often resulting in shadows being baked into the albedo. This issue is exacerbated by a relatively dark environment map, in contrast to previous works, and is compounded by strong emissions and shadows. Nevertheless, while BRDF results are comparable, ESR-NeRF distinguishes itself in its primary goal: the accurate reconstruction of emissive sources We also provide the

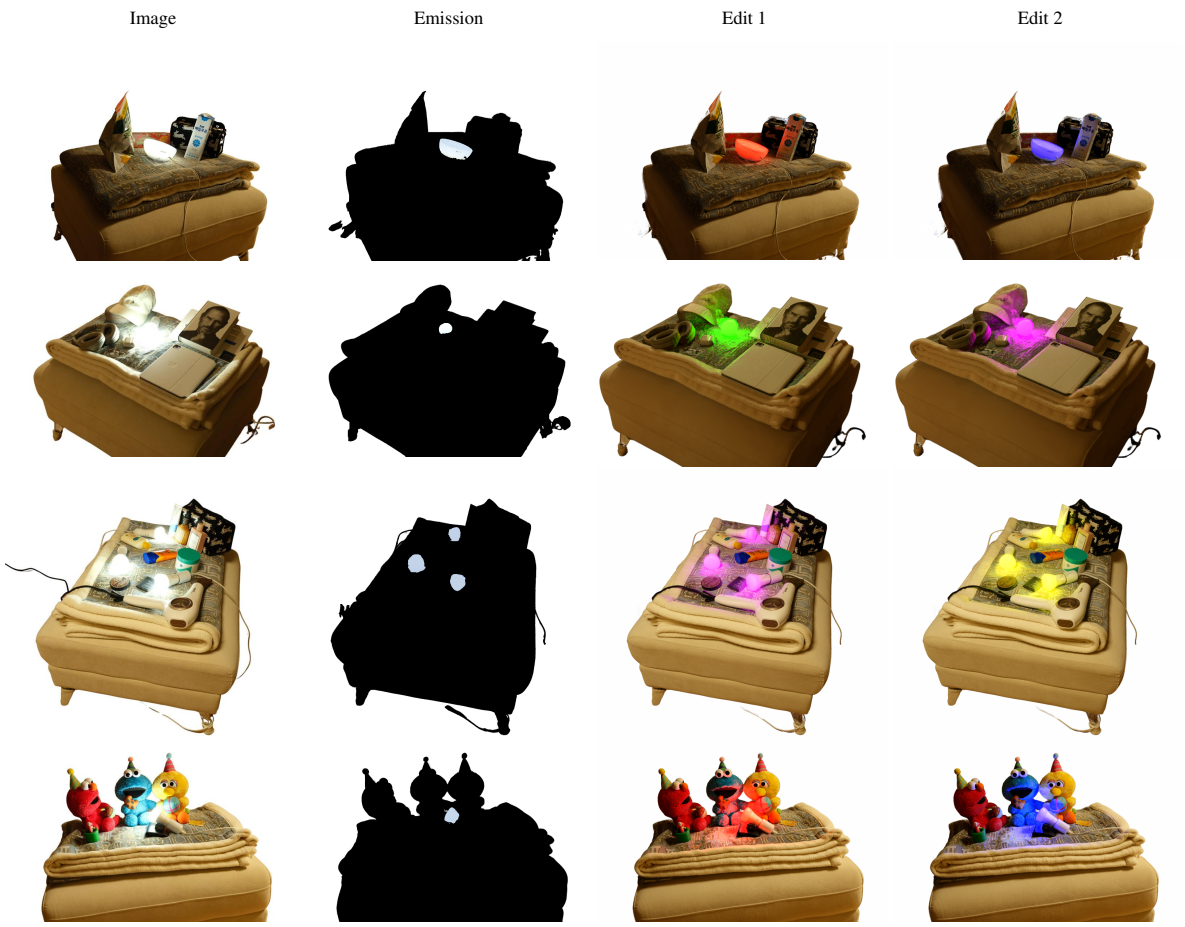


Figure 8. Identified emissive sources and edited results on real scenes.

reconstructed scene components on DTU dataset in Fig. 20 and 21.

1.6. Illumination Decomposition

We present additional results of decomposed illumination in Fig. 22. These visualizations offer insights into the effectiveness of ESR-NeRF in factorizing the scene illumination. The off image, for instance, is generated by merging direct and indirect illumination from the environment map, as shown in the first row. The second row illustrates the decomposition of emission effects, including both the emission and its reflection. Light-on images are created by adding the light-off and the emission effects images.

1.7. Scene Editing w&w/o Radiance Fine-tuning

Fig. 23 and 24 present additional scene editing examples, illustrating various scenarios including intensity and color edits, as well as their combination. As discussed in the conclusion section of the main paper, scene illumination can be adjusted without fine-tuning radiance fields, using alternative methods. Results on the right side of Fig. 23 to 24 are rendered by calculating only direct illumination from emissive sources for re-lighting, a technique commonly used in prior research [6, 19, 20], bypassing the fine-tuning of trained networks. This approach is particularly effective for scenes with vividly colored emissive sources, as shown in Fig. 25. To evaluate the effectiveness of direct illumination in scene editing, we provide quantitative results for each scenario in Tab. 4 and 5. Quantitative comparisons for scenes with vivid-colored emissive sources are detailed in Tab. 6 and 7.

1.8. Analysis of Learnable Tone-mapper

We eliminate the constraint on the range of radiance values to address the unbounded nature of emissive sources and their reflections. Instead of the commonly used sigmoid activation function in NeRF-based methods [2, 9, 12, 20, 21] for radiance prediction, we employ the softplus activation, extending the radiance range from $[0, 1]$ to $[0, \infty]$.

However, this modification may lead to inaccurate sur-

Figure 9. Qualitative results comparison with ground truth obtained using commercial smart bulbs.

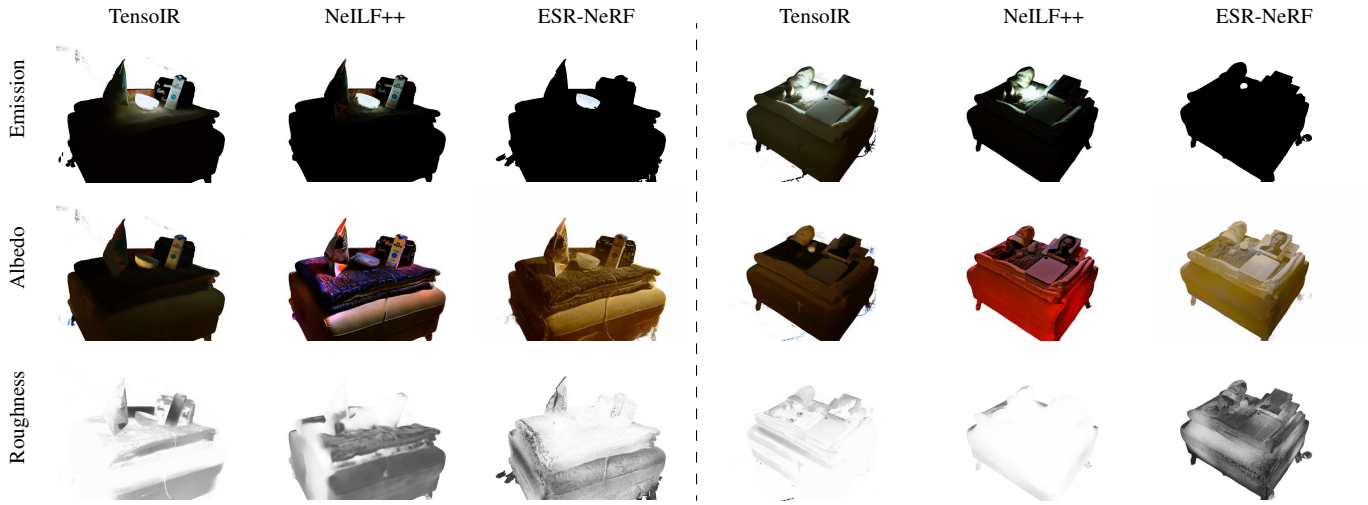


Figure 10. Comparison of identified emissive sources and decomposed BRDF.

face reconstructions, as highlighted in the main paper. Fig. 28 shows instances where surfaces become semi-transparen, lose structural details, and the rendered images significantly deviate from the ground truth, making the accurate reconstruction of emissive sources infeasible.

To address this issue, we introduce a learnable tone-mapper m_θ, taking positionally encoded HDR linear color as input and produce LDR sRGB colors outputs. Fig. 26 reveals that this tone-mapper helps in obtaining accurate surface normals and rendering photo-realistic images. Nonetheless, a trade-off exists between the quality of surface normals and rendered images, when using the learnable

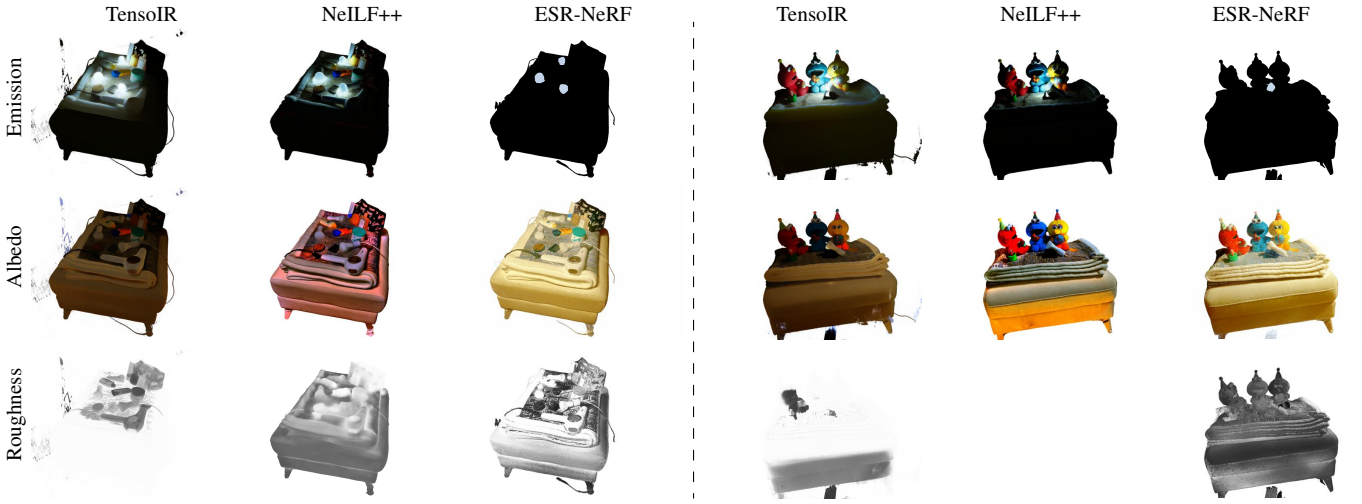


Figure 11. Comparison of identified emissive sources and decomposed BRDF.

tone-mapper. For example, a low λ_τ value, which indicates a heavier reliance on the tone-mapper in the rendering loss, may improve surface details but linear color values deviate significantly from expectations. This discrepancy occurs as the correlation between predicted linear colors and actual image pixel colors weakens with lower λ_τ values. Conversely, a higher λ_τ compromises surface reconstruction quality. Thus, setting λ_τ requires careful consideration of the balance between surface detail and color accuracy.

Interestingly, the choice of λ_τ also impacts the reconstruction of emissive sources in real scenes. A high λ_τ tends to result in lower intensity of reconstructed emissive sources. Re-lighting experiments in Fig. 12 show illumination effects confined to a narrow area compared to ground truth data. We suspect the camera may edit images for low contrast and apply color grading, particularly in HDR scenes. We used the Fuji 100s camera. A high λ_τ in the rendering loss could be problematic, as it aims to align gamma-corrected linear values with manipulated colors. Based on this insight, we slightly reduced λ_τ by 0.1 to enhance emission intensity (1.4 vs. 37.2) and expand reflections in re-lighting scenarios.

1.9. Near-zero IoU Results of Baselines

State-of-the-art re-lighting methods struggle with ambiguities surrounding emissive sources, often failing to accurately identify them. These methods typically cannot differentiate between reflections and emissions, leading to most regions being misclassified as emissive sources. This challenge is reflected in Tab. 3, where baseline methods exhibit near-zero IoU performance across various scenes. Despite extensive trinary grid searches with an interval of 0.01 for thresholding values to report the peak performance of baselines, ESR-NeRF consistently outperforms them. Additionally, our method's efficacy in classifying rays into the uncertain group for emissive source identification highlights its superiority in this task. This is further supported by additional results obtained using thresholding techniques applied to the baselines.

Figure 12. Scene edit results on jobs scene. Lower λ_τ results in stronger emission. The middle image is rendered with direct light for proving enhanced emission strength.

1.10. Failure Cases in Scene Editing

We also present failure cases in scene editing, discussing the limitations of the radiance fine-tuning method for re-lighting in §4.5 of the main paper. While ESR-NeRF effectively reconstructs and manipulates emissive sources, the radiance fine-tuning method for re-lighting has its limitations. These are depicted in Fig. 27, where we note that LTS learning-based radiance fine-tuning may be constrained to color adjustments within the training spectrum. In other words, using the LTS loss to transfer radiance within light transport segments may be weak in representing new colors that traverse unobserved light paths during training. For example, it can shift colors from yellow to green but not to blue. Additionally, the network's inherent smoothness capability may introduce illumination inaccuracies. In the last row in Fig. 27, changing only the top emissive source to red inadvertently affects the bulldozer's lower ceiling.

Exploring alternative rendering approaches could address these issues. We showcase scene editing results by computing direct illumination from emissive sources in

Fig. 25, enabling changes to any colors. Reconstructing emissive sources using ESR-NeRF, then extracting emission texture maps to use rendering engines like Blender [3] or Mitsuba [5] is also promising. Howver, the texture map extraction in NeRF-like methods often faces severe UV atlas fragmentation. Recent methods like Nuvo [13] offer some hope for feasible emission texture editing. We consider these avenues for future exploration

References

[1] Brent Burley and Walt Disney Animation Studios. Physically-based shading at disney. In *SIGGRAPH*, 2012. 1

[2] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *ECCV*, 2022. 1, 5

[3] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018. 2, 8

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

[5] 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

[6] 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. 1, 3, 5

[7] 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. 3

[8] 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. 3

[9] 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. 5

[10] Johannes L. Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *CVPR*, 2016. 4

[11] Alvy Ray Smith. Color gamut transform pairs. In *ACM TOG*, 1978. 2

[12] 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. 5

[13] Pratul P. Srinivasan, Stephan J. Garbin, Dor Verbin, Jonathan T. Barron, and Ben Mildenhall. Nuvo: Neural uv mapping for unruly 3d representations. *arXiv*, 2023. 8

[14] 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. 3

[15] Bruce Walter, Stephen R. Marschner, Hongsong Li, and Kenneth E. Torrance. Microfacet models for refraction through rough surfaces. In *Proceedings of the 18th Eurographics Conference on Rendering Techniques*, 2007. 1

[16] 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. 1, 3

[17] 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. 1, 3

[18] 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. 1, 3

[19] 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. 5

[20] Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. Modeling indirect illumination for inverse rendering. In *CVPR*, 2022. 5

[21] Quan Zheng, Gurprit Singh, and Hans-Peter Seidel. Neural relightable participating media rendering. In *NeurIPS*, 2021. 5

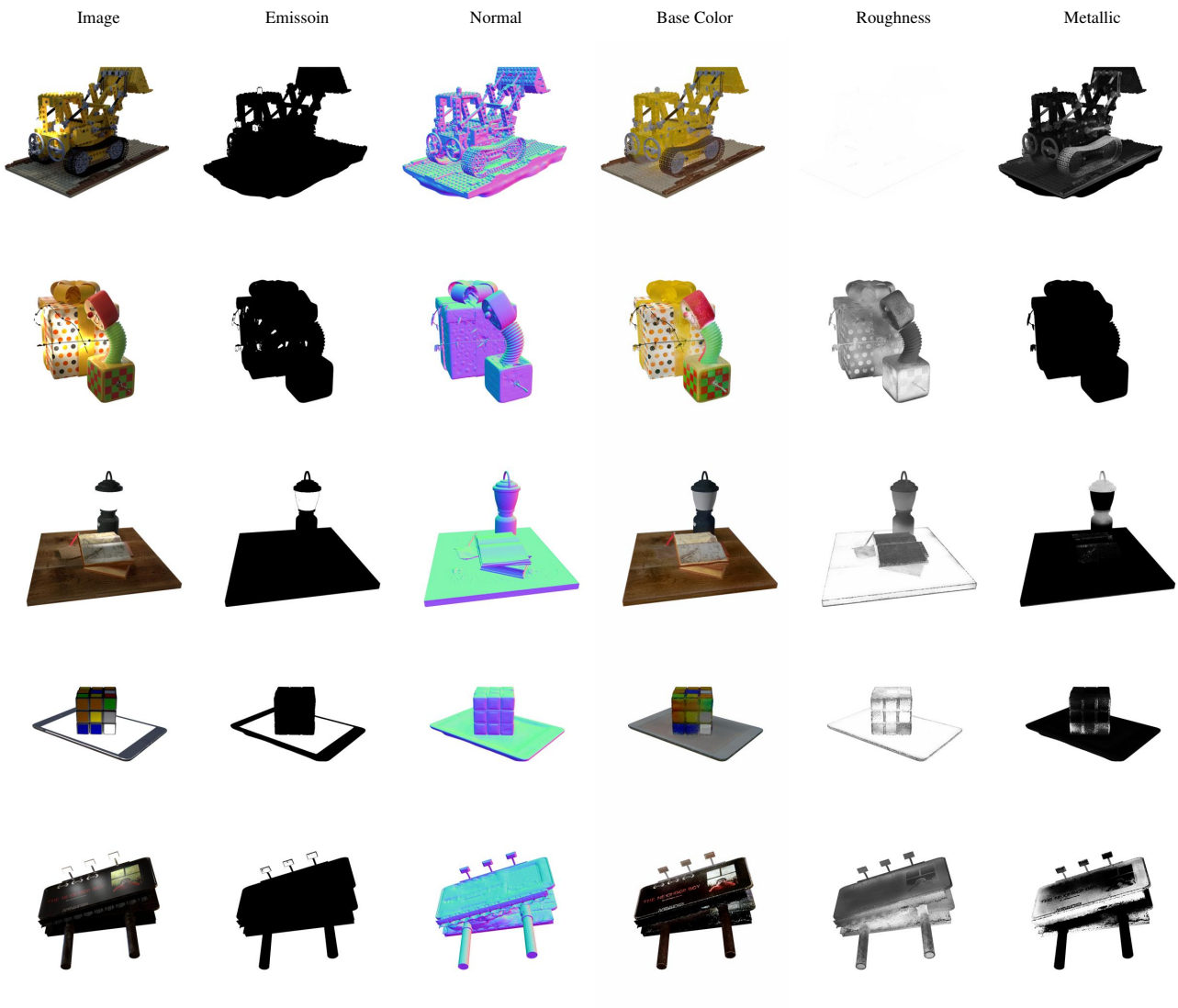


Figure 13. Decomposed scene components on scenes with white-colored emissive sources.

	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
w/o progressive	0.09	18.87	0.05	5.93	0.38	2.84	0.82	30.82	0.14	1.00	0.93	0.04	0.09	6.71	0.05	3.89	0.37	1.69	0.84	10.60	0.14	0.64	0.94	0.02
w/o sg	0.79	8.33	0.50	5.32	0.35	2.91	0.96	21.28	0.72	0.80	0.95	0.04	0.16	6.43	0.35	3.60	0.35	1.87	0.93	8.65	0.89	0.25	0.92	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. Per-scene metrics on emissive source reconstruction tasks. 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).

	White colored						Vivid colored					
	Lego	Gift	Book	Cube	Billboard	Balls	Lego	Gift	Book	Cube	Billboard	Balls
NeILF++	**0.00**	**0.01**	**0.04**	**0.39**	0.00	**0.07**	**0.00**	**0.01**	**0.04**	**0.39**	0.00	**0.07**
TensoIR	**0.00**	**0.01**	**0.04**	0.37	**0.01**	**0.07**	**0.00**	**0.01**	**0.04**	0.37	**0.01**	**0.07**
ESR-NeRF	**0.81**	**0.60**	**0.96**	**0.97**	**0.84**	**0.95**	**0.51**	**0.59**	**0.96**	**0.97**	**0.88**	**0.94**
NeILF++ (∗)	0.43	0.07	**0.95**	0.93	0.01	0.91	0.30	0.09	**0.95**	0.94	0.02	0.92
TensoIR (∗)	**0.71**	**0.15**	**0.95**	**0.95**	**0.76**	**0.95**	**0.33**	**0.15**	**0.95**	**0.96**	**0.77**	**0.95**
ESR-NeRF (∗)	**0.81**	**0.60**	**0.98**	**0.98**	**0.94**	**0.96**	**0.51**	**0.61**	**0.98**	**0.97**	**0.93**	**0.94**

Table 3. Results of emissive source identification. The IoU measures the source area identification (a higher value is better). The asterisk (*) denotes that thresholding is applied to reconstructed emission strengths.

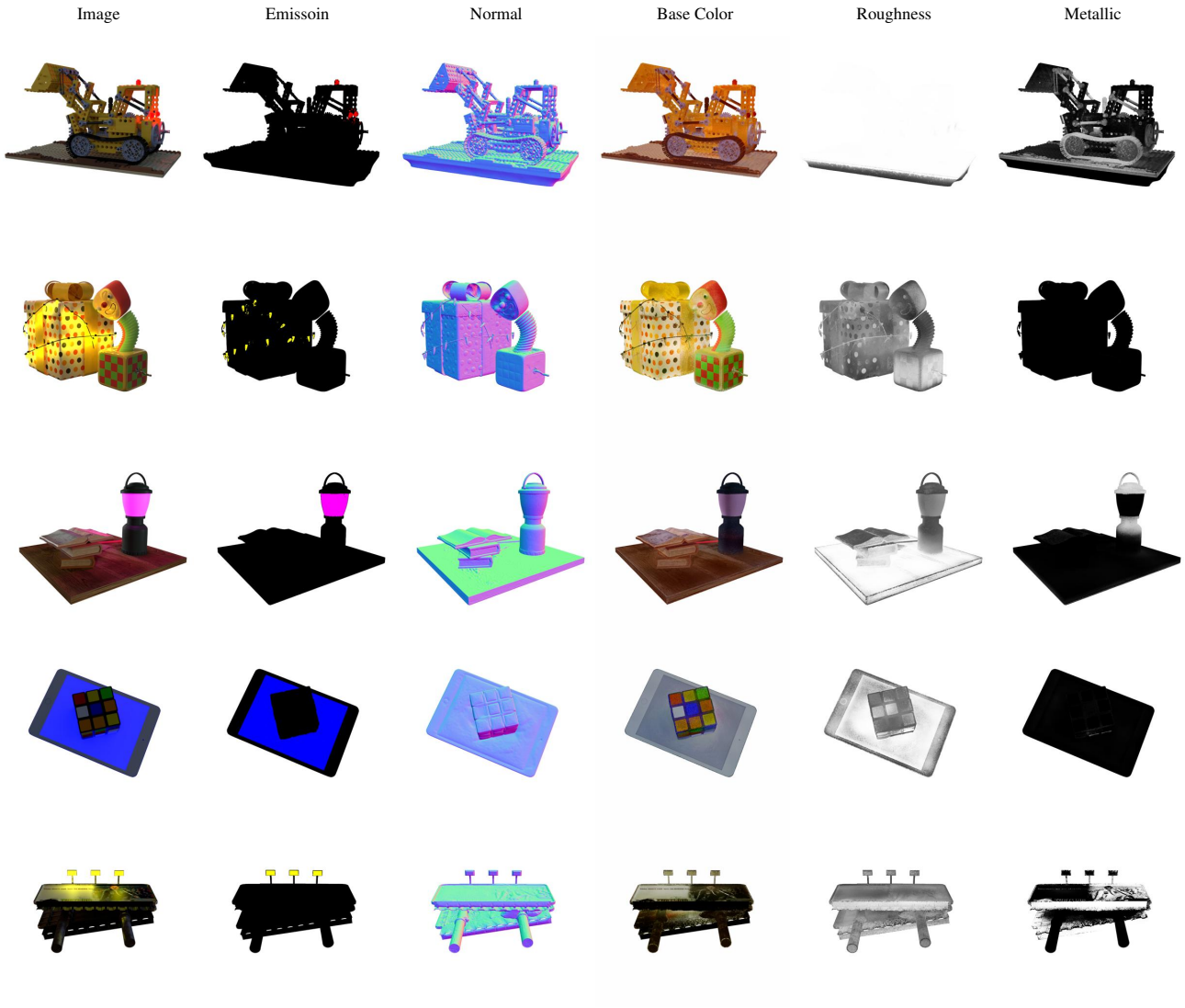


Figure 14. Decomposed scene components on scenes with vivid-colored emissive sources.

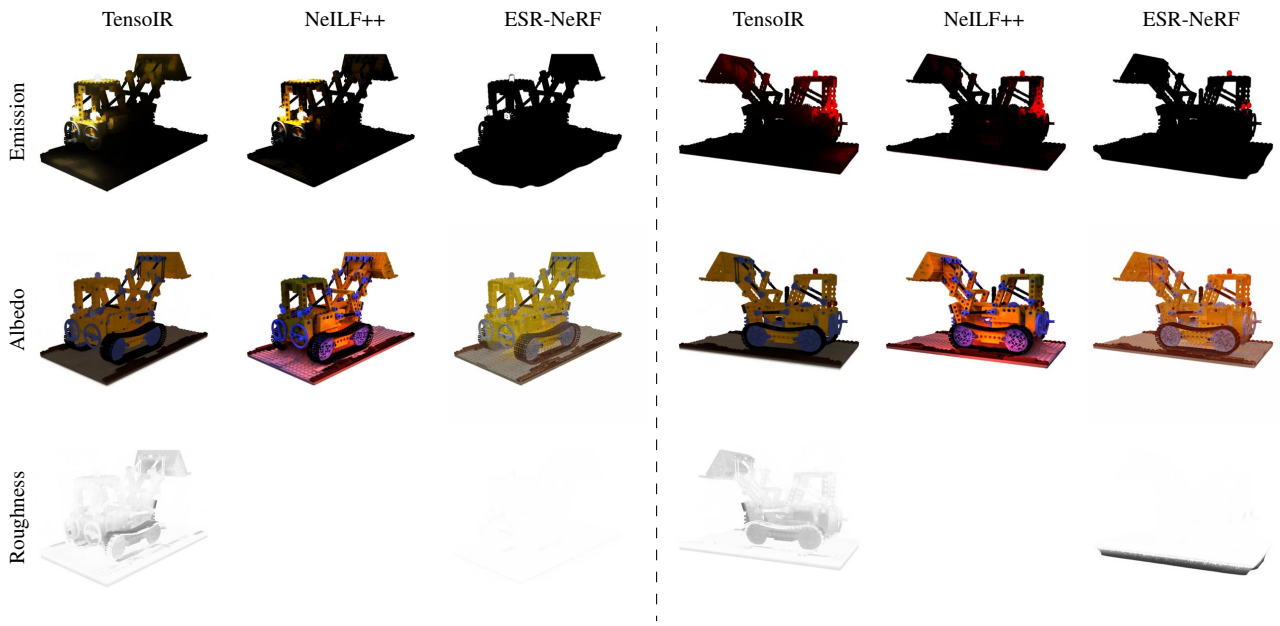


Figure 15. Comparison of identified emissive sources and decomposed BRDF on the Lego scene. Left: Lego white. Right: Lego vivid.

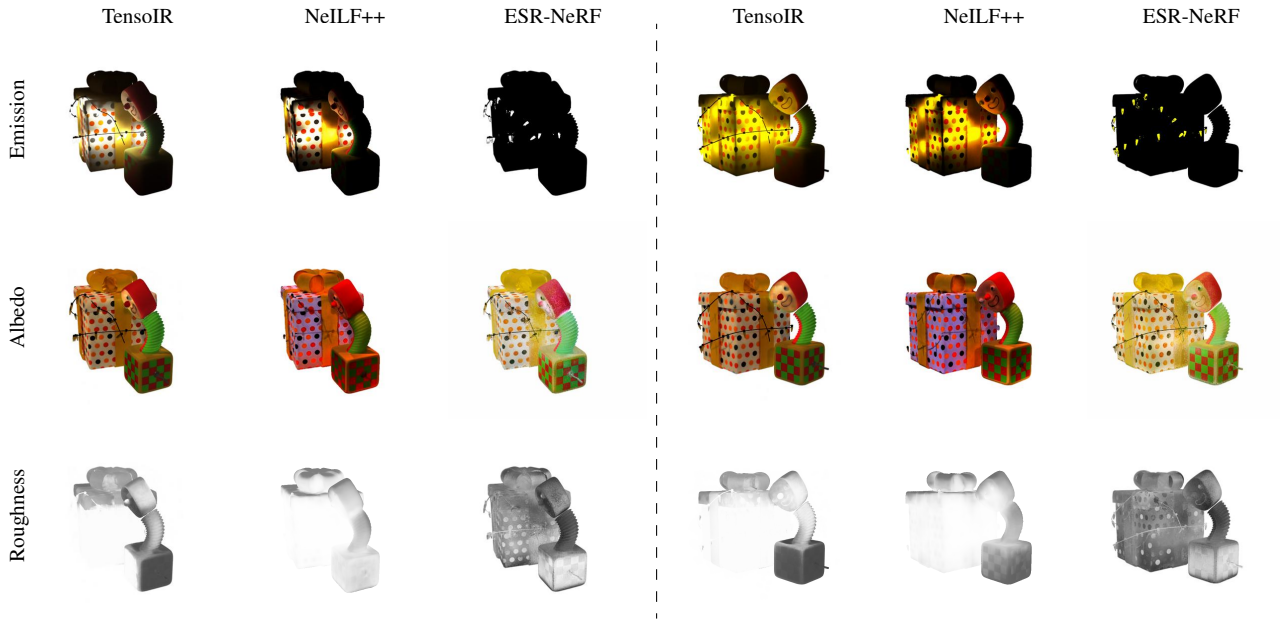


Figure 16. Comparison of identified emissive sources and decomposed BRDF on the Gift scene. Left: Gift white. Right: Gift vivid.

	White colored															
	Lego (C)		Lego (I)		Gift		Book		Cube		Billboard (C)		Billboard (I)		Balls	
	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS
NV	37.77	0.0082	37.77	0.0082	37.72	0.0060	44.95	0.0032	43.60	0.0022	36.23	0.0109	36.23	0.0109	32.49	0.0190
NV + I	32.53	0.0175	29.50	0.0261	27.27	0.0163	30.29	0.0166	31.47	0.0097	29.50	0.0188	30.31	0.0216	29.03	0.0281
NV + C	32.27	0.0220	29.93	0.0259	31.28	0.0140	34.92	0.0123	35.08	0.0083	31.17	0.0197	27.66	0.0322	31.55	0.0221
NV + I + C	29.12	0.0291	30.44	0.0248	31.02	0.0151	34.80	0.0128	34.04	0.0093	31.22	0.0200	31.94	0.0245	30.44	0.0239

Table 4. Editing performance on scenes with white-colored emissive sources by using the fine-tuning method. (C) denotes collective adjustments of emissive sources, while (I) represents individual adjustments of emissive sources. NV denotes novel view synthesis, I denotes intensity editing, and C denotes color editing. A higher PSNR or lower LPIPS value is better.

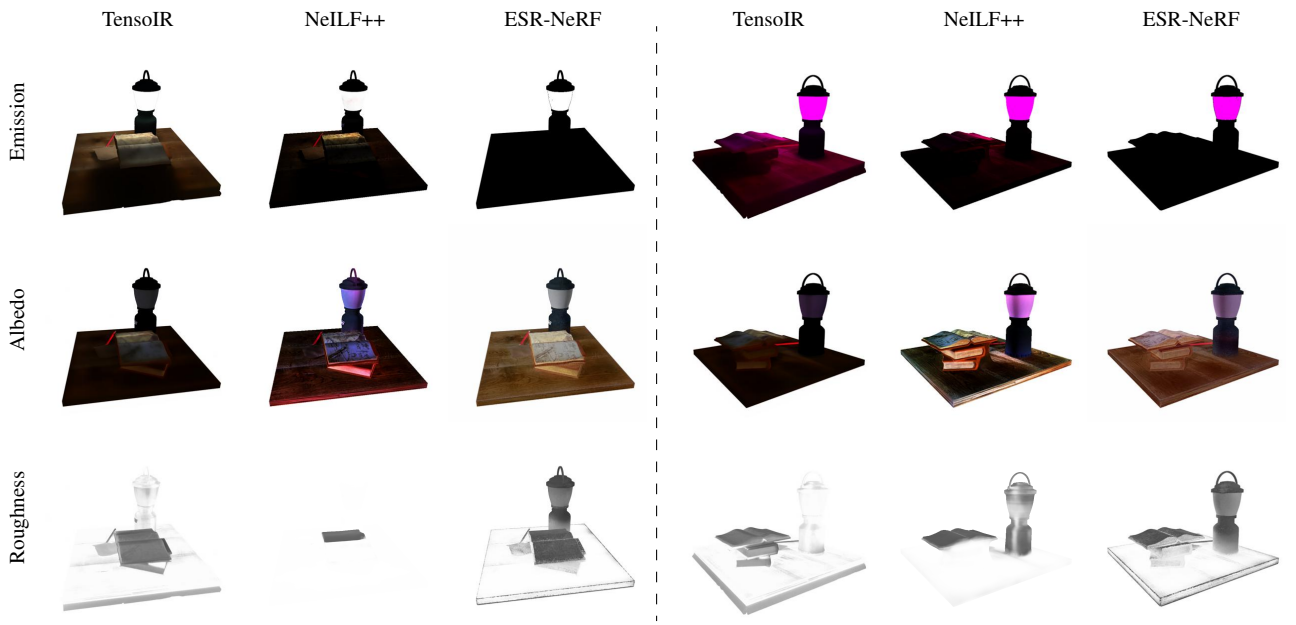


Figure 17. Comparison of identified emissive sources and decomposed BRDF on the Book scene. Left: Book white. Right: Book vivid.

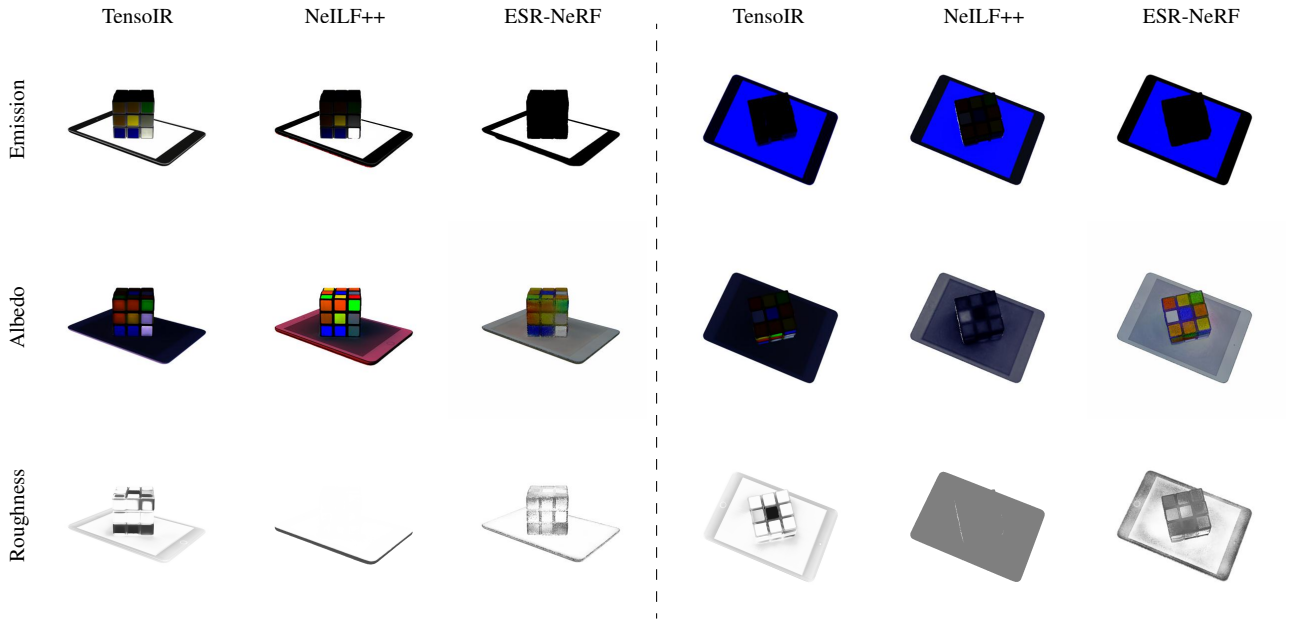


Figure 18. Comparison of identified emissive sources and decomposed BRDF on the Cube scene. Left: Cube white. Right: Cube vivid.

	White colored															
	Lego (C)		Lego (I)		Gift		Book		Cube		Billboard (C)		Billboard (I)		Balls	
	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS
NV	37.77	0.0082	37.77	0.0082	37.72	0.0060	44.95	0.0032	43.60	0.0022	36.23	0.0109	36.23	0.0109	32.49	0.0190
NV + I	27.77	0.0329	29.00	0.0293	22.77	0.0461	28.85	0.0327	24.71	0.0382	26.49	0.0422	31.14	0.0249	28.41	0.0411
NV + C	30.44	0.0292	29.96	0.0284	27.00	0.0316	33.71	0.0191	30.34	0.0229	29.90	0.0328	31.14	0.0320	30.74	0.0282
NV + I + C	30.17	0.0307	30.44	0.0275	27.49	0.0318	33.40	0.0203	27.67	0.0312	29.80	0.0313	32.92	0.0242	29.94	0.0311

Table 5. Editing performance on scenes with white-colored emissive sources by computing direct illumination from reconstructed emissive sources. (C) denotes collective adjustments of emissive sources, while (I) represents individual adjustments of emissive sources. NV denotes novel view synthesis, I denotes intensity editing, and C denotes color editing. A higher PSNR or lower LPIPS value is better.

Figure 19. Comparison of identified emissive sources and decomposed BRDF on the Billboard scene. Left: Billboard white. Right: Billboard vivid.

	Vivid colored											
	Lego (C)		Gift		Book		Cube		Billboard (C)		Balls	
	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS
NV	39.76	0.0062	38.31	0.0055	44.97	0.0033	45.13	0.0016	34.47	0.0165	32.78	0.0180
NV + I	35.00	0.0154	28.70	0.0163	32.86	0.0141	35.54	0.0072	28.56	0.0288	30.78	0.0231
NV + C	22.80	0.0916	26.00	0.0312	29.11	0.0371	22.50	0.0379	27.82	0.0347	30.05	0.0259
NV + I + C	23.28	0.0873	26.64	0.0290	28.15	0.0419	24.10	0.0361	27.12	0.0367	26.91	0.0329

Table 6. Editing performance on scenes with vivid-colored emissive sources by using the fine-tuning method. (C) denotes collective adjustments of emissive sources. NV denotes novel view synthesis, I denotes intensity editing, and C denotes color editing. A higher PSNR or lower LPIPS value is better.

	Vivid colored											
	Lego (C)		Gift		Book		Cube		Billboard (C)		Balls	
	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS
NV	39.76	0.0062	38.31	0.0055	44.97	0.0033	45.13	0.0016	34.47	0.0165	32.78	0.0180
NV + I	29.21	0.0330	24.97	0.0378	31.78	0.0216	27.44	0.0314	28.19	0.0356	27.20	0.0352
NV + C	27.41	0.0330	26.49	0.0318	33.95	0.0198	28.81	0.0260	28.51	0.0326	29.57	0.0301
NV + I + C	27.27	0.0348	26.80	0.0312	33.28	0.0232	24.62	0.0447	28.56	0.0329	24.89	0.0388

Table 7. Editing performance on scenes with vivid-colored emissive sources by computing direct illumination from reconstructed emissive sources. (C) denotes collective adjustments of emissive sources. NV denotes novel view synthesis, I denotes intensity editing, and C denotes color editing. A higher PSNR or lower LPIPS value is better.

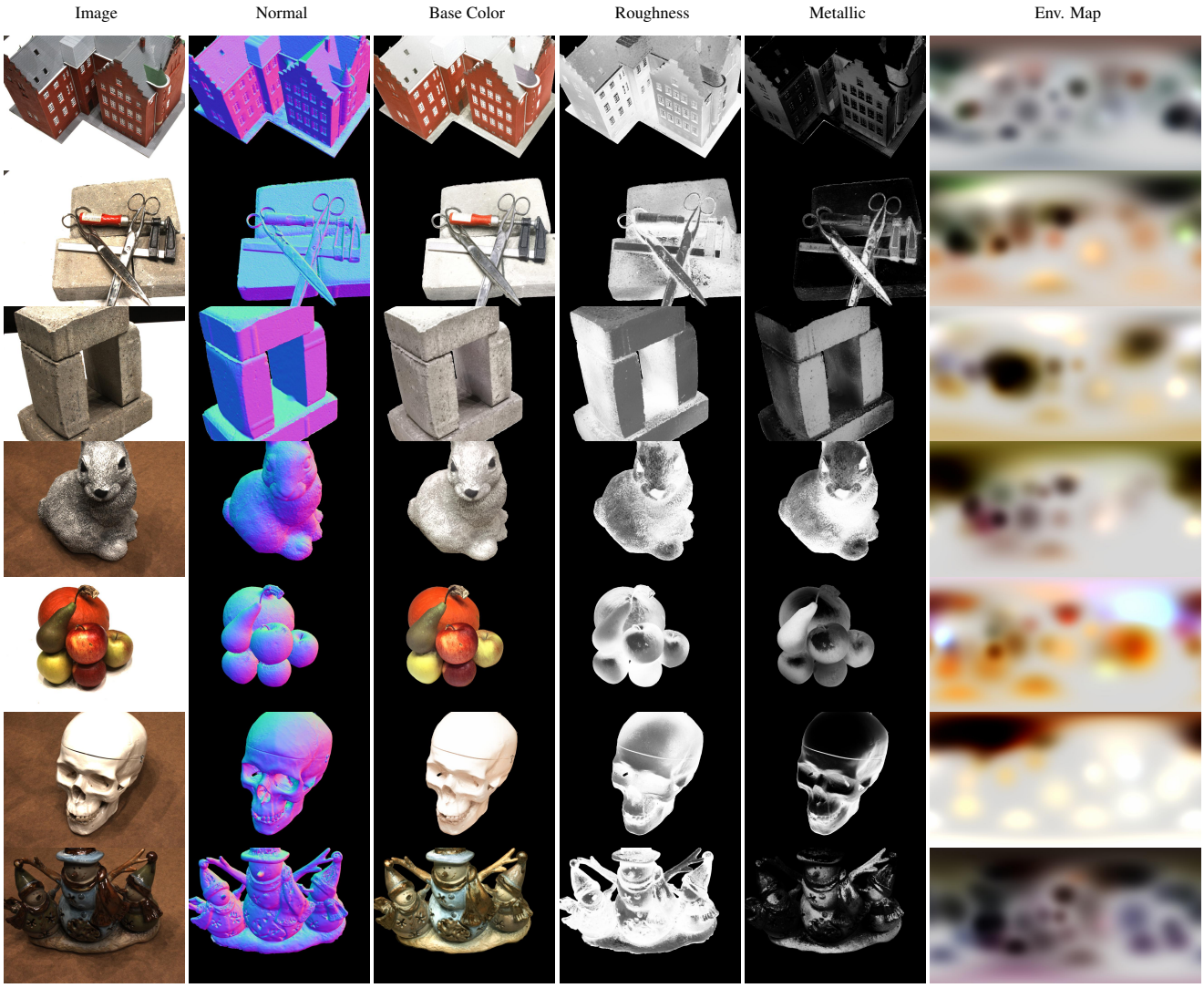


Figure 20. Decomposed scene components on DTU scenes without emissive sources.

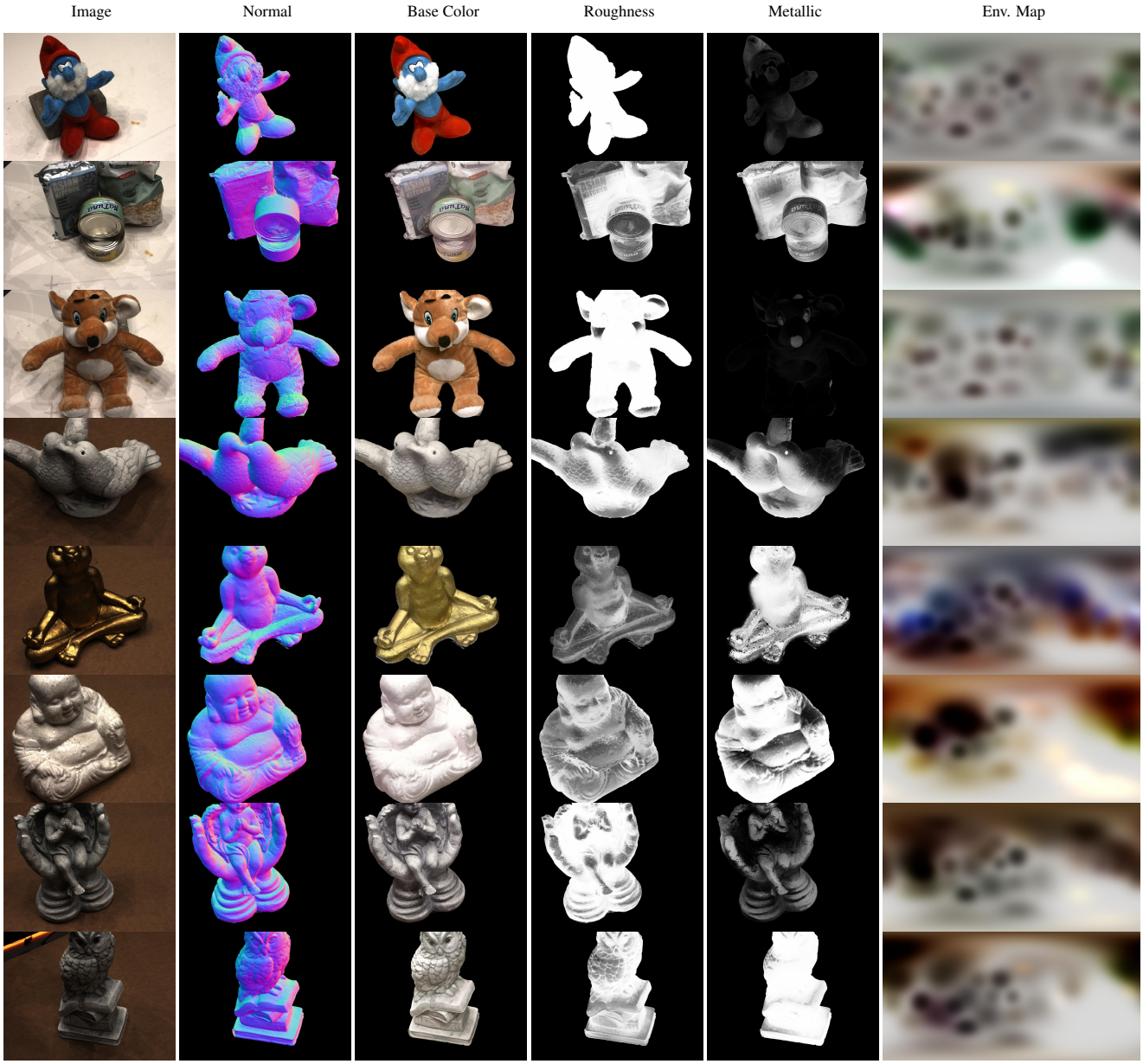


Figure 21. Decomposed scene components on DTU scenes without emissive sources.

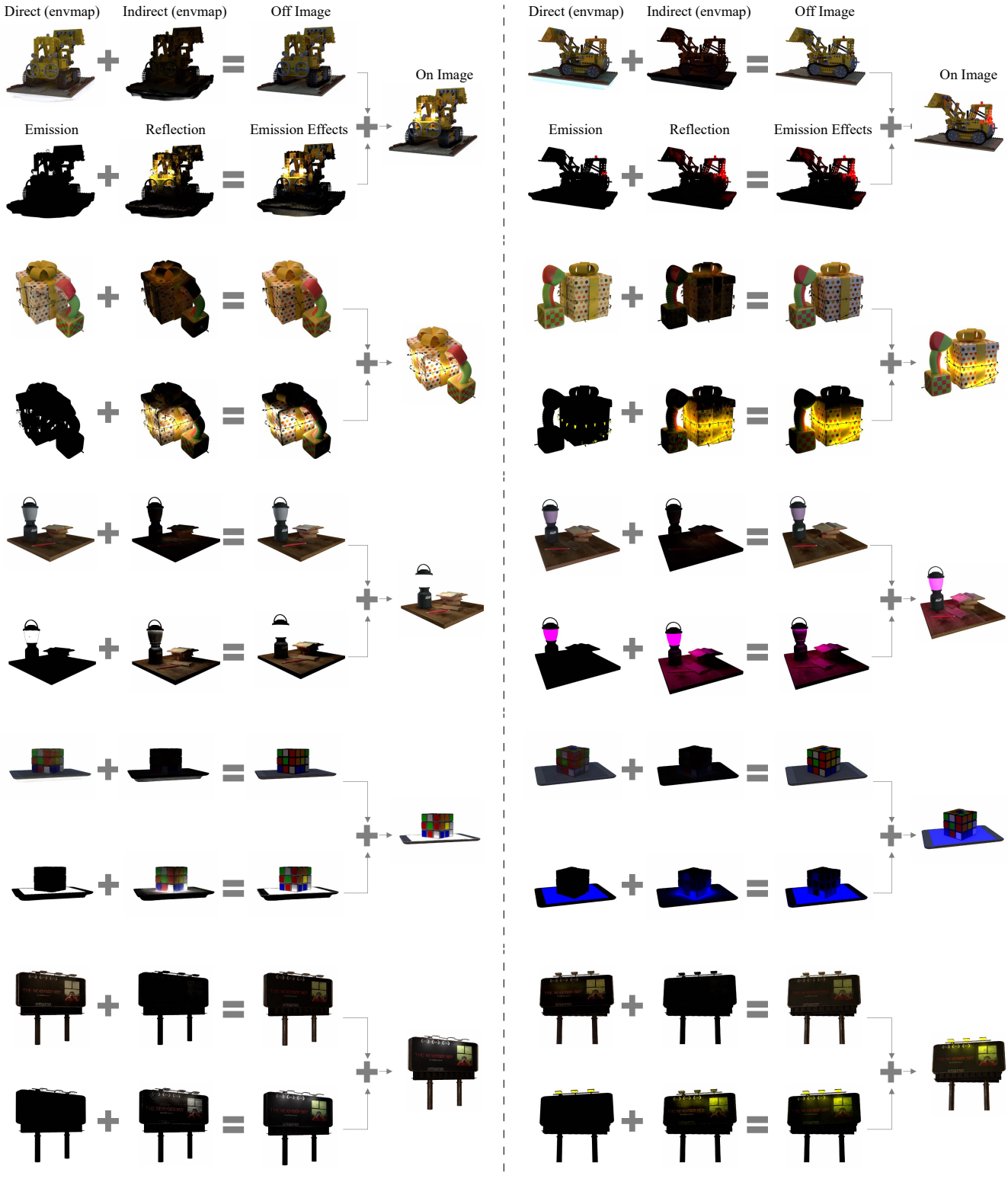


Figure 22. Illumination decomposition results. Left: scenes with white-colored, Right: scenes with vivid-colored emissive sources.

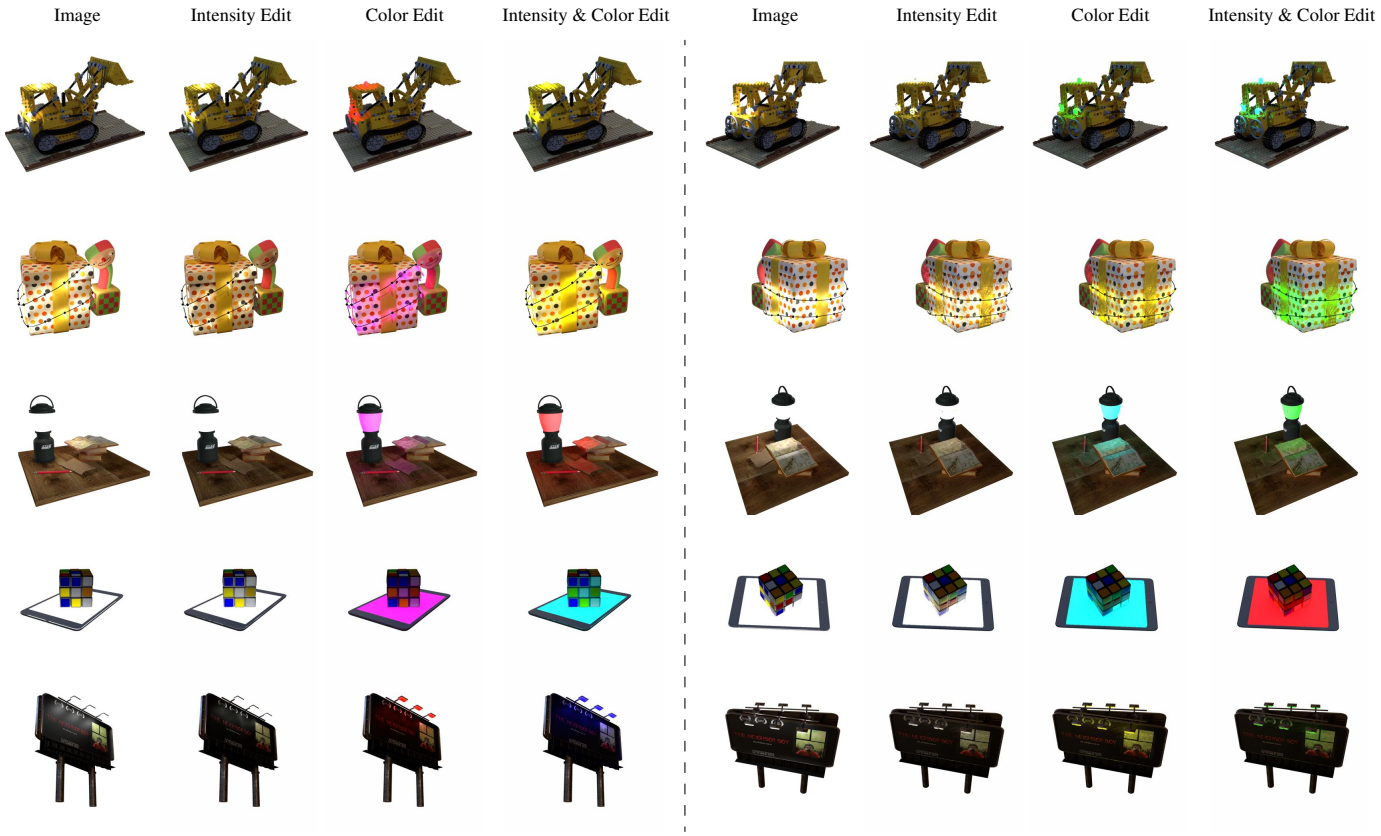


Figure 23. Re-lighting scenes containing white emissive sources. Left: through fine-tuning radiance fields, Right: computing direct illumination from reconstructed emissive sources.

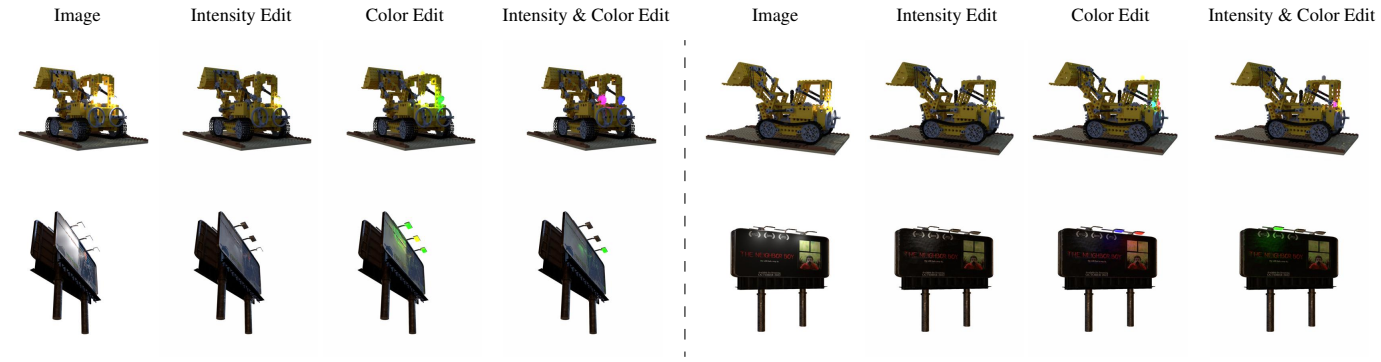


Figure 24. Individual emissive sources control. Left: through fine-tuning radiance fields, Right: computing direct illumination from reconstructed emissive sources.

Figure 25. Re-lighting scenes containing vivid-colored emissive sources by computing direct illumination from reconstructed emissive sources.

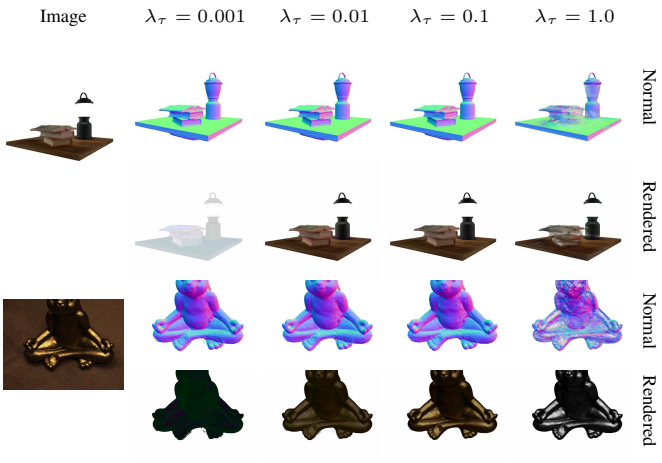


Figure 26. Reconstructed surface normals and rendered linear images with varying λ_τ values. Gamma correction is applied to linear images for easy comparison.

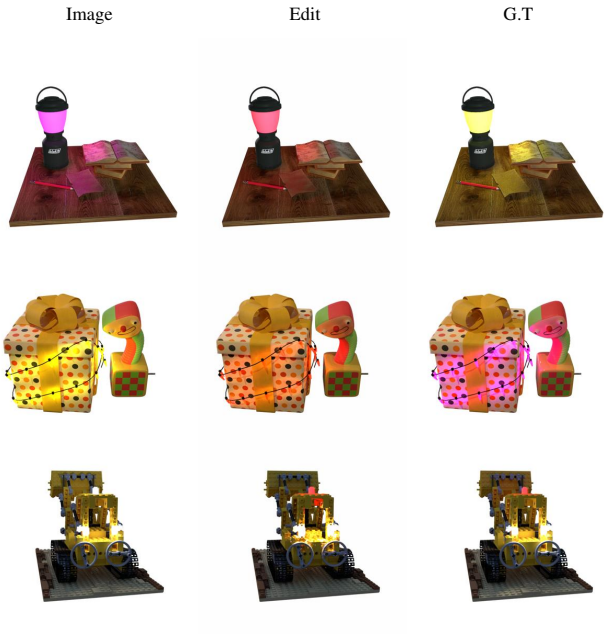


Figure 27. Failure cases for editing scene illumination using the radiance fine-tuning method.

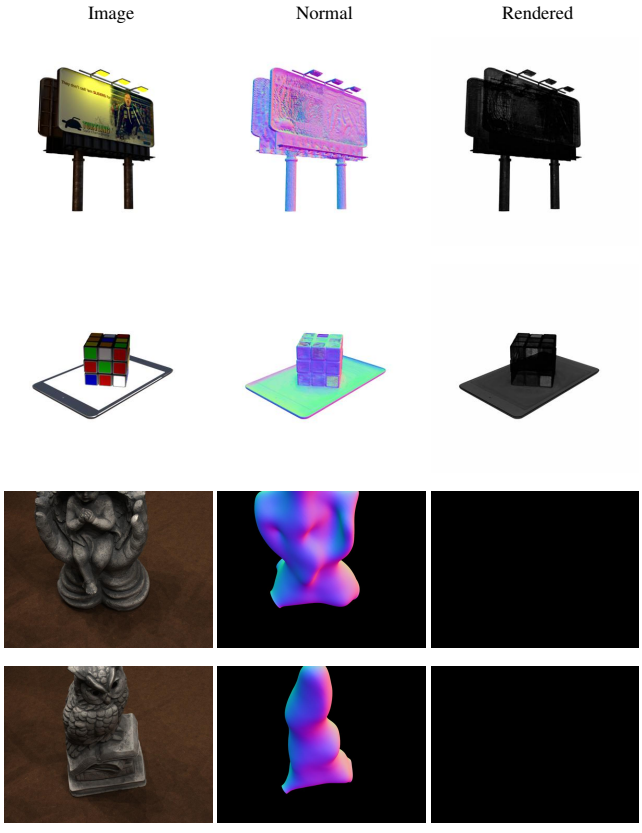


Figure 28. Erroneously reconstructed surfaces and rendered linear images when using softplus activation for radiances without utilizing the tone-mapper m_θ. Gamma correction is applied to linear images for easy comparison.