

LuxRemix: Lighting Decomposition and Remixing for Indoor Scenes

Supplementary Material

This supplemental document provides additional details for LuxRemix. We first describe the data preparation process in Section 6, then introduce additional details about our models in Section 7, and finally provide additional results and comparisons in Section 8.

Please see our project website <https://luxremix.github.io> for the following additional supplemental material:

- **Interactive Lighting Blend Explorer:**
A browser-based demo for interactive single-image light editing based on our lighting decompositions.
- **Single-Image Lighting Decomposition:**
Shows the decomposition of seven light sources and their flexible recombination for various relighting effects.
- **Multi-View Lighting Harmonization:**
Visualizes the decomposition and harmonization of multiple light sources across multiple views.
- **Real-time Remixable Lighting:**
Screen recordings from our modified Splatfacto renderer in the Nerfstudio viewer demonstrating our interactive relighting with 3D Gaussian splatting. Note that temporarily lower rendering resolutions are due to the dynamic resolution rendering implemented in the Nerfstudio viewer.

6. Data

Here we provide additional information for how we created the training data for LuxRemix. We render scenes from Avetisyan et al. [2] with procedurally generated light sources using Infinigen [72] into many one-light-at-a-time (OLAT) equirectangular images. We choose equirectangular 360° images to save on the rendering cost and allow dynamic sampling of diverse perspective camera viewpoints on the fly while training the model. Please see Table 3 for a comparison of the rendering and storage costs.

Table 3. **Rendering and storage comparison.** The perspective views are rendered at resolution 512×512 pixels, and the equirectangular 360° images are rendered at resolution 2048×1024 pixels. 64 samples per pixel are used for Blender’s Cycles path tracer [9].

| Setup | Render time | Storage |
|---------------------------|---------------------|-----------------|
| 20,000 perspective images | 140,000 sec | 17.2 GB |
| 1,000 360° images | 27,000 sec = 19% | 6.8 GB = 40% |

Rendering scenes once. To reduce our dataset size while maintaining flexibility during training, we store equirectangular 360° images instead of pre-rendering a large collection

of perspective views that our model ingests. Equirectangular representations compactly encode the full scene from a single viewpoint, allowing arbitrary camera perspectives to be sampled on the fly. In contrast, storing every possible perspective image incurs significant storage overhead and limits viewpoint diversity at training time. For example, training on 20K pre-rendered perspective images with resolution 512×512 pixels would require 17.2 GB of storage. Instead we could store 1K equirectangular images with a resolution of 2048×1024 pixels in just 6.8 GB and sample these from new viewpoints on the fly, assuming we can always sample at least 20 distinct views per equirectangular image. This design substantially reduces disk usage and data preparation cost while preserving the ability to generate diverse, view-consistent training inputs dynamically. We found it particularly useful to render the dataset once and use the stored data for the entire project.

Implementation details. As described in Section 6, we use Blender’s Cycles renderer [9] to generate equirectangular HDR images for each synthetic room. For each room, we sample four distinct locations to capture equirectangular images at a resolution of 2048×1024 pixels. Rendering is performed with 64 samples per pixel (spp), using the OptiX denoiser to enhance path-traced results. We used approximately 2,800 GPU hours on NVIDIA A100s to render the complete dataset. In total, we rendered approximately 49,600 different equirectangular views across over 12,000 synthetic rooms. The additional depth map and light mask are rendered for each 360° image to assist effective perspective view sampling. Combining all different light passes, the total dataset size is about 9 TB.

On-the-fly sampling. As our model is designed for regular perspective views, during training, the perspective views are sampled from the rendered equirectangular 360° images. For single-image light editing, we sample perspective views such that the target light source is visible within the field of view (FOV). Specifically, we use the equirectangular light mask to select a visible light source, and then use its center to guide the sampling range of azimuth and elevation for the perspective projection.

For multi-view lighting, after selecting an initial perspective view that contains the target light source within the field of view, we sample additional perspective views to incrementally increase the overall coverage of visible regions. These views are drawn from multiple source equirectangular 360° images and are chosen to ensure that consecutive views share overlapping regions, promoting consistency across frames. To guide this sampling, we project depth maps from the vari-

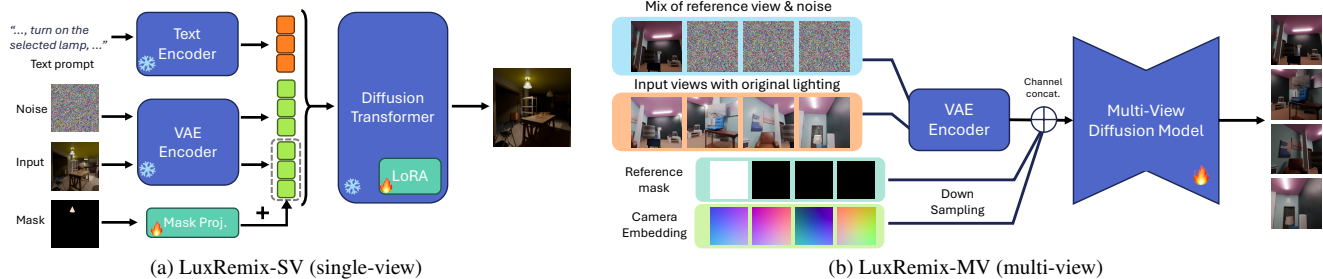


Figure 9. **Model Architectures.** (a) Single-view lighting decomposition and (b) multi-view lighting harmonization.

ous equirectangular images into a global coordinate system, allowing us to accurately identify co-visible areas within the room. Additionally, depth information is used to avoid sampling views that are too close to scene geometry or the camera, ensuring a diverse and meaningful set of training perspectives.

After sampling the perspective views, we generate training data pairs on-the-fly. For the OLAT decomposition task, we select a target OLAT image I_{target} and its corresponding full-light image I_{full} , forming the training pair $(I_{\text{full}}, I_{\text{target}})$. For the one-light-off task, we compute $I_{\text{one-off}} = I_{\text{full}} - c_{\text{target}} \cdot I_{\text{target}}$, where c_{target} is the scaling factor for the target OLAT image as indicated in the rendering metadata. The resulting pair $(I_{\text{full}}, I_{\text{one-off}})$ teaches the model to turn off the selected light source. To further augment the training set, we perform additional lighting composition by loading multiple OLAT images and adding them to the full-light image. This produces new, well-lit image variants such as $I'_{\text{full}} = I_{\text{full}} + \sum_{i=1}^N c'_i \cdot I_i$, enriching the diversity of the training data. To preserve the high dynamic range of the original images when converting to the sRGB color space for model training, we apply AgX tone mapping¹ to the sampled linear images, which has been shown to be effective in preserving more realistic highlight effects.

7. Models

7.1. Single-image Light Editing – LuxRemix-SV

Our single-image editing model builds upon a pretrained text-based image editing diffusion transformer (DiT) with an architecture similar to FLUX.1 Kontext [8]. Because full-parameter fine-tuning is prohibitively expensive, we adopt LoRA [38] for efficient lightweight adaptation. Importantly, the base DiT model already demonstrates strong capability for general instruction-based image editing (see Figure 11), which allows LoRA fine-tuning to specialize the model for the light editing task using only limited synthetic training data, and generalize to real-world images using the generative prior inherent in the base DiT model. Figure 9a illustrates the model design.

¹<https://github.com/EaryChow/AgX>

We use LoRA fine-tuning for two closely related light editing tasks: (1) OLAT decomposition and (2) turning off the target light. Below are the instruction templates we used for the two tasks:

1. “{trigger word: OLAT}. Darken the room to a night scene: eliminate all light sources and window light. The only light source should be the selected {light type}, on at {low,medium,high} brightness, according to the selection mask.”
2. “{trigger word: LTOFF}. Only turn off the selected {light type}, according to the selection mask. Keep the remaining light sources unchanged.”

We use two distinct trigger words to differentiate between the two closely related tasks. For the OLAT decomposition task, we further specify the desired brightness level “{low, medium, high} brightness”, for the target light source. Specifically, we instruct the model to generate LDR OLAT images corresponding to approximate exposure values of EV-4 (low), EV-2 (medium), and EV0 (high) relative to the HDR target after tone mapping. This approach encourages the model to better learn light transport and enables reconstruction of HDR OLAT images through multi-exposure fusion [21]. Figure 10 shows two examples of our additional brightness level control.

Other than text prompts, we also provide the light mask as a spatial prompt for the model. The mask is processed

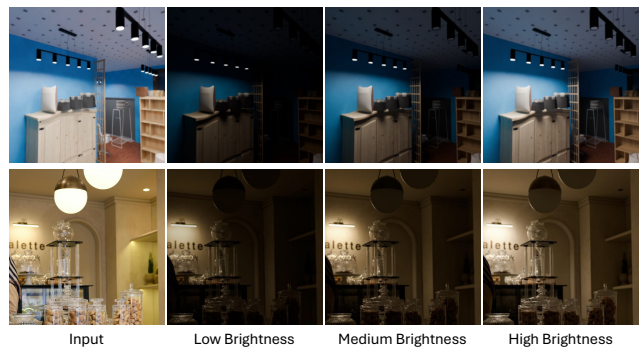


Figure 10. **Brightness Level Control.** LuxRemix-SV’s OLAT decomposition at different brightness levels.

by a single-layer MLP to match the DiT’s input latent dimensions, then added to the input condition image latents. This additional single-layer MLP is fine-tuned along with the LoRA parameters to further improve the model’s ability to understand the light mask.

Training details. We insert LoRA adapters into all attention blocks of the DiT model, using a LoRA rank of 32. Fine-tuning is performed for 3,000 iterations with a batch size of 192, leveraging the Prodigy optimizer [64] for adaptive learning rate adjustment and bias correction. Training was conducted on 48 NVIDIA A100-40GB GPUs, completing in approximately 12 hours.

7.2. Multi-view Harmonization – LuxRemix-MV

Our multi-view harmonization model extends the lighting decomposition from single to multiple views. While it is possible to use the single-view LuxRemix-SV model for multi-view propagation (e.g., by processing diptychs or grids of images), this becomes computationally infeasible for large image sets due to high resource demands. To address this, we use a scalable multi-view diffusion model based on a pretrained U-Net architecture (similar to CAT3D [30] and SEVA [99]) for efficient multi-view lighting propagation.

The base model is pretrained to generate novel views given a few reference images. However, our task does not require synthesizing entirely new views; instead, we have access to multi-view input images prior to lighting decomposition, and our goal is to propagate the lighting decomposition from sparse reference views to all other output views. To accomplish this, we concatenate the original input views and the sparse light-decomposed views, together with the corresponding Plücker ray embeddings and reference view masks, as inputs to the pretrained multi-view diffusion U-Net, by extending the model’s channels of the input projection layer, as illustrated in Figure 9b. To distinguish between OLAT decomposition and one-light-off editing, we also include a binary mask in the input condition: all zeros for OLAT decomposition, and all ones for one-light-off editing.

Training details. We conduct full-parameter fine-tuning of the U-Net for 30,000 iterations. The training progresses in three stages: For the first 15,000 iterations, the model is trained on 4-view input batches of size 192. In the next 10,000 iterations, the number of input views is increased to 8, with a batch size of 144. Finally, during the last 5,000 iterations, the model is further trained with 15-view input batches of size 96. We utilize the AdamW optimizer [57] with a learning rate of 5×10^{-5} . Training is performed on 48 NVIDIA A100-40GB GPUs and completes in approximately 28 hours.

Inference details. Our model is fine-tuned to handle up to 15 views per forward pass. To accommodate larger multi-view datasets, we employ a sequential multi-pass strategy.

The core idea is to process images in batches while maintaining consistency by conditioning on both the original reference views and previously generated frames. Let \mathcal{U} denote the set of unprocessed target frames and \mathcal{P} the set of processed frames. We use a distance metric $d(I_i, I_j)$ based on camera pose similarity to guide the selection process:

- **Pass 1:** We select target frames from \mathcal{U} that are spatially closest to the original source reference views \mathcal{R}^* .
- **Pass $k > 1$:** We iteratively select remaining frames from \mathcal{U} that are closest to any frame in the processed set \mathcal{P} .

For each pass, we construct a dynamic reference set \mathcal{R}_k that always includes the original source references \mathcal{R}^* . Additionally, if a target frame is selected due to its proximity to a previously generated frame $I_{\text{prev}} \in \mathcal{P}$, we optionally include I_{prev} as a secondary reference in \mathcal{R}_k . This approach allows the model to “chain” visual information from source views to distant targets, ensuring that subsequent passes remain consistent with the lighting decomposition established in earlier steps.

7.3. Gaussian Splatting with Lighting Control

Following the two-stage training pipeline described in Section 3.4, we extend the 3D Gaussian splatting model [42] to enable precise lighting control. We achieve this by fitting per-light RGB parameters for each Gaussian, utilizing the light-decomposed multi-view images.

Stage 1: Pretraining a standard 3DGS. We first establish the scene’s geometric structure by training a standard 3D Gaussian splatting model on the original multi-view images using gsplat [89]. Upon convergence, these pretrained Gaussians serve as the basis for incorporating additional parameters for lighting control.

Stage 2: Fitting per-light RGB parameters. We augment each Gaussian from the previous stage with a set of learnable lighting parameters $\mathbf{L}_i \in \mathbb{R}^{M \times 3}$, where M denotes the number of decomposed light sources (including ambient lighting). Modeled in a linear HDR space to represent the physical accumulation of light, these parameters allow us to render the image for a specific light source m , denoted as \hat{I}_m , via standard rasterization using the Gaussians’ existing geometry and the new learnable lighting parameters. We optimize these per-light RGB parameters as follows:

1. **Joint Optimization:** We jointly optimize the per-light RGB parameters \mathbf{L} alongside the shared geometry parameters. This step refines the pretrained Gaussians’ geometry and appearance to better align with the light-decomposed multi-view images. We train the Gaussians in this step for 4,000 iterations.
2. **Light Fitting (Frozen Geometry):** To prevent the model from explaining lighting residuals by altering geometry, we freeze all geometric parameters and focus solely on optimizing \mathbf{L} to fit the light-decomposed images for the remaining 2,000 iterations.

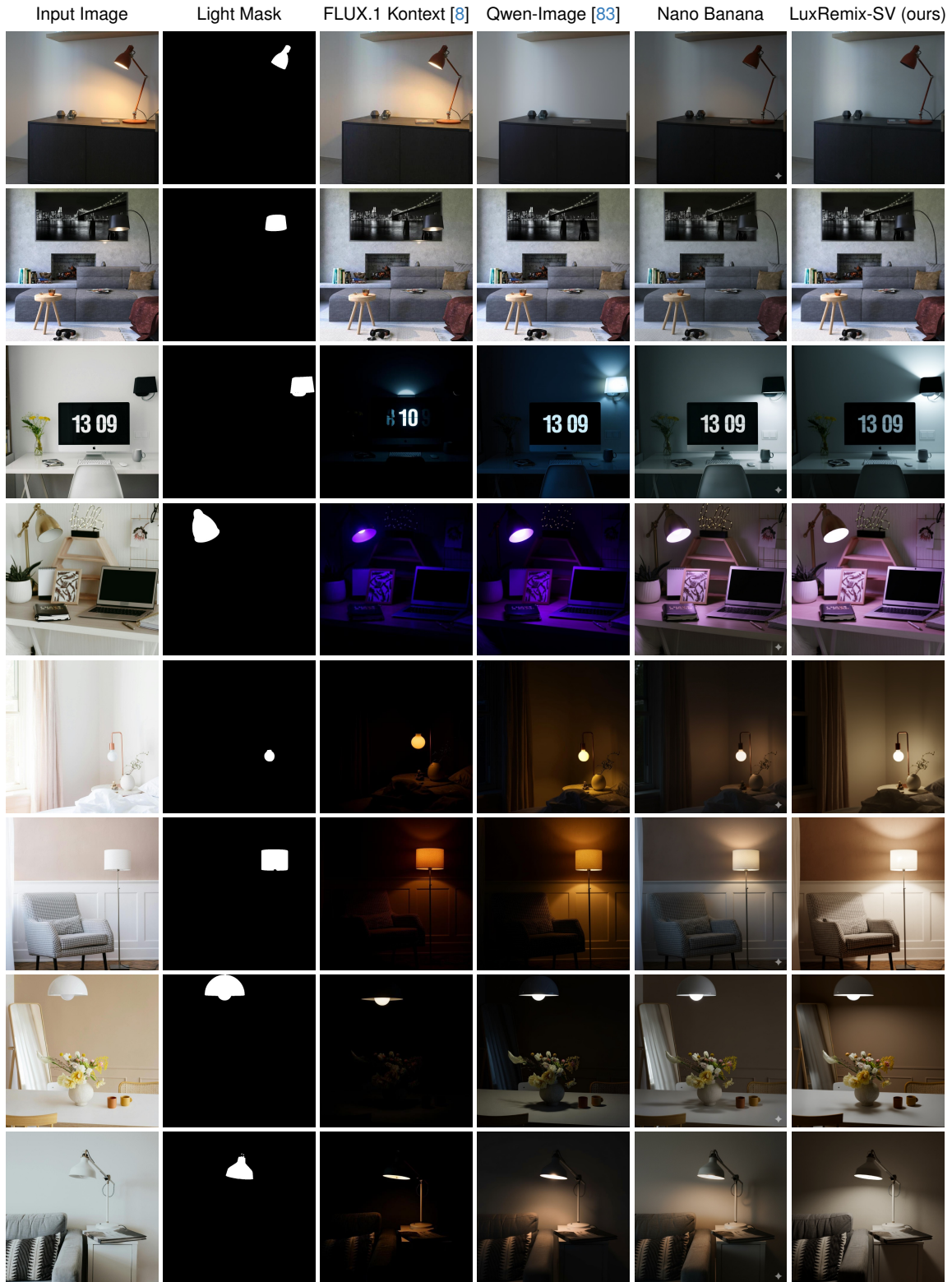


Figure 11. **Additional Comparisons of Single-Image Light Editing.** Given an input image and a light mask (left), we compare our light editing results with baseline methods for switching lights on and off. FLUX.1 Kontext [8] sometimes removes or modifies image details. Qwen-Image [83] sometimes removes lights entirely. Nano Banana produces plausible light edits. Our LuxRemix-SV model produces convincing results throughout with fine-grained controllability over individual light intensities and colors. Real-world images from Pexels.

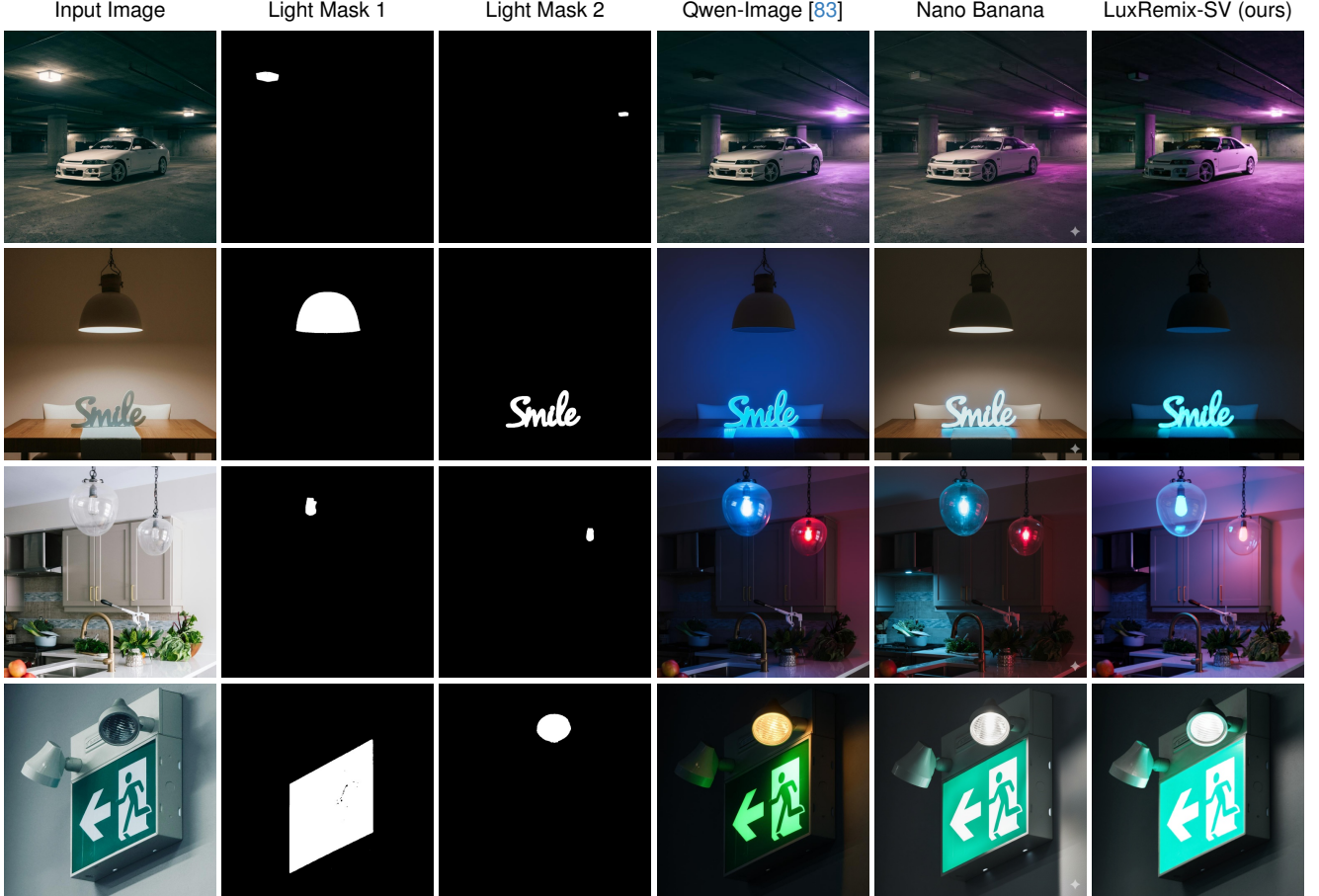


Figure 12. **Comparisons of Single-Image Light Editing with Multiple Masks.** Given an input image and two light masks (left), we compare our light editing results (right) with baseline image editing models. While Qwen-Image [83] and Nano Banana enable basic prompt-based light editing, only our LuxRemix-SV approach enables realistic and fine-grained light editing control via light mask conditioning. Real-world images from Pexels and Unsplash.

Training Objectives. The optimization in Stage 2 incorporates three losses: one for the fidelity of individual light renderings, one for the consistency of the recombined lighting, and one for the spatial smoothness of the Gaussian parameters.

The first term is the photometric loss between the rendered light image \hat{I}_m and the ground-truth light image I_m :

$$\mathcal{L}_{\text{olat}} = \frac{1}{M} \sum_{m=1}^M \left(\|\hat{I}_m - I_m\|_1 + \lambda \mathcal{L}_{\text{D-SSIM}}(\hat{I}_m, I_m) \right). \quad (2)$$

The second term is a composition consistency loss between the recombined lighting $\hat{I}_{\text{comp}} = \sum_m w_m \hat{I}_m$ and the original input image I_{ori} , where w_m is a learnable per-light scaling factor for light recombination. This ensures that the learned lighting coefficients, when recombined, accurately reconstruct the original appearance:

$$\mathcal{L}_{\text{comp}} = \|\mathcal{T}(\hat{I}_{\text{comp}}) - I_{\text{ori}}\|_1, \quad (3)$$

where $\mathcal{T}(x) = (x + \beta)^{\frac{1}{\gamma}}$ is a differentiable tone mapping function with a learnable gamma γ and offset β .

The third component is a spatial smoothness loss. To mitigate high-frequency noise where adjacent Gaussians learn divergent light responses, we impose a spatial smoothness regularizer based on K-Nearest Neighbors (KNN). For each Gaussian i , we penalize the deviation of its lighting coefficients from its K nearest spatial neighbors $\mathcal{N}(i)$:

$$\mathcal{L}_{\text{smooth}} = \frac{1}{NK} \sum_{i=1}^N \sum_{j \in \mathcal{N}(i)} \|\mathbf{L}_i - \mathbf{L}_j\|_2^2. \quad (4)$$

This encourages local consistency in light reflectance, helping to reduce sparkling artifacts during relighting. We apply this loss every 100 optimization iterations after the initial 4,000 iterations.

The final objective is the weighted sum of three losses:

$$\mathcal{L} = \mathcal{L}_{\text{olat}} + \lambda_{\text{comp}} \mathcal{L}_{\text{comp}} + \lambda_{\text{smooth}} \mathcal{L}_{\text{smooth}}. \quad (5)$$



Figure 13. **Single-View vs. Multi-View Consistency.** Lighting edits on a synthetic test scene.

After training, we can randomly blend multiple per-light renderings from the same set of Gaussians to get consistent relighting results. We include screen recording videos in the supplement to show our interactive relighting with 3DGS.

8. Additional Results and Comparisons

8.1. Single-Image Lighting Editing

We show additional editing results in Figures 11 and 12. In Figure 11, we show results of switching lights off and on, with various light colors. The baseline comparisons to FLUX.1 Kontext [8], Qwen-Image [83] and Gemini 2.5 Flash Image a.k.a. Nano Banana show prompt-based editing results trying to match our results. Figure 12 shows more advanced editing results with two editable light sources and a mixture of out-of-distribution light shapes and novel colors. In the supplemental material, we include a video demonstrating the decomposition of seven distinct light sources, and recombining them in many different ways.

8.2. Multi-view Lighting Harmonization

Figure 15 shows additional multi-view lighting harmonization results. Note how our single-view model LuxRemix-SV can recover individual OLAT lights with plausible shadows for the plant in the top scene. Our multi-view harmonization model LuxRemix-MV then propagates these single-view lighting decompositions consistently across views. For the middle scene, our model can cleanly disentangle the three hanging lights. And for the bottom scene, our model can switch on the standing lamp in the center of the scene (see “OLAT 3”). Note also that the ambient lighting in this scene correctly includes the out-of-view ceiling lights.

Table 4. **Real-World Evaluation on LightLab.** Quantitative comparison on 10 paired real-world examples.

| Method | PSNR \uparrow | SSIM \uparrow | LPIPS \downarrow |
|--------------------|-----------------|-----------------|--------------------|
| ScribbleLight [20] | 15.04 | 0.465 | 0.478 |
| Qwen-Image [83] | 19.20 | 0.735 | 0.196 |
| LightLab [59] | 21.80 | 0.791 | 0.144 |
| Ours | 20.93 | 0.771 | 0.168 |

8.3. Real-World Paired Evaluation on LightLab

To complement our synthetic evaluation, we include additional real-world paired benchmarking on the LightLab examples. For each method, we tune light editing parameters to best match each paired target image and report photometric metrics. Table 4 summarizes results on all 10 available paired examples. While LightLab achieves the strongest scores on this real-data benchmark, our method remains competitive despite being trained on synthetic data only.

8.4. Single-View versus Multi-View Consistency

We further compare per-frame single-view editing against our multi-view harmonization, as shown in Figure 13. Applying LuxRemix-SV independently to each view can introduce cross-view inconsistencies in shadows and lighting appearance. In contrast, LuxRemix-MV propagates decomposition jointly across views and yields more consistent multi-view results.

8.5. Failure cases

Figure 14 shows some failure cases of our single-image light decomposition and editing method. In our synthetic training data, the shape of the light spread is biased towards conical

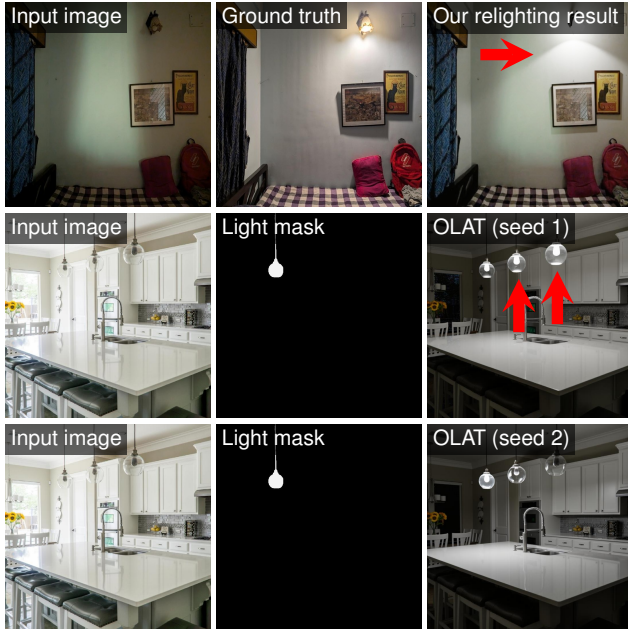


Figure 14. **Failure cases.** **Top:** the light spread shape is biased towards cones, diverging from the ground truth. **Middle:** decomposition occasionally ignores the provided light mask. **Bottom:** different random seeds can yield different OLAT decompositions.

shapes, which does not always match real-world lights. In multi-light images, our lighting decomposition sometimes fails to adhere to the provided light mask. Using a different random seed can result in different OLAT decompositions.

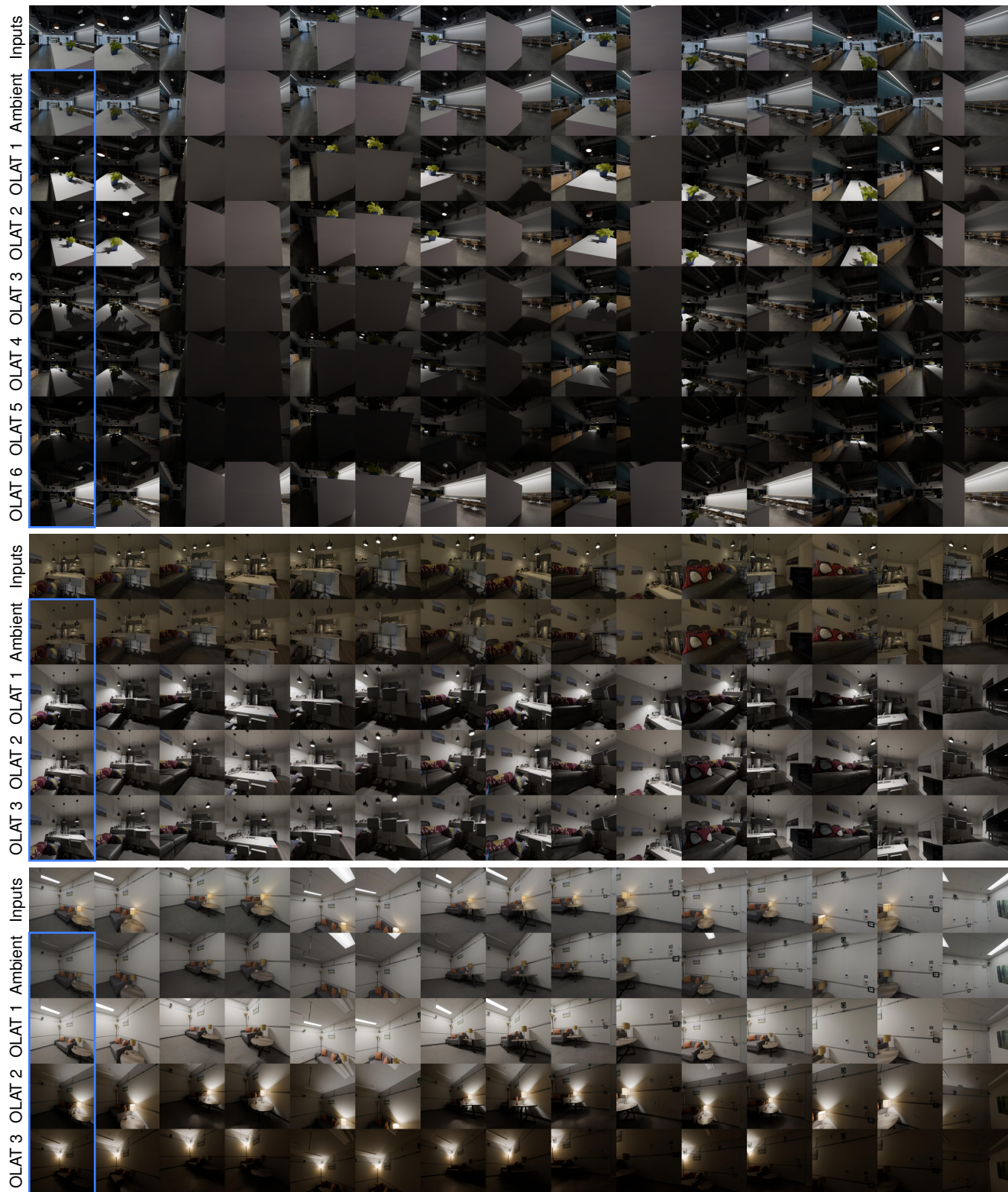


Figure 15. **Additional Multi-view Lighting Harmonization Results.** From the first input image, we decompose the lighting into ambient and OLAT components (blue), then propagate them consistently across all views (top row). Real-world captures from VR-NeRF [87].

References

- [1] Hadi Alzayer, Philipp Henzler, Jonathan T. Barron, Jia-Bin Huang, Pratul P. Srinivasan, and Dor Verbin. Generative multiview relighting for 3D reconstruction under extreme illumination variation. In *CVPR*, 2025. 2, 3
- [2] Armen Avetisyan, Christopher Xie, Henry Howard-Jenkins, Tsun-Yi Yang, Samir Aroudj, Suvam Patra, Fuyang Zhang, Duncan Frost, Luke Holland, Campbell Orme, Jakob Engel, Edward Miller, Richard Newcombe, and Vasileios Balntas. SceneScript: Reconstructing scenes with an autoregressive structured language model. In *ECCV*, 2024. 4, 13
- [3] Haiyang Bai, Jiaqi Zhu, Songru Jiang, Wei Huang, Tao Lu, Yuanqi Li, Jie Guo, Runze Fu, Yanwen Guo, and Lijun Chen. GaRe: Relightable 3D Gaussian splatting for outdoor scenes from unconstrained photo collections. In *ICCV*, 2025. 3
- [4] 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. 6
- [5] Anand Bhattad, James Soole, and D.A. Forsyth. StyLitGAN: Image-based relighting via latent control. In *CVPR*, 2024. 3
- [6] Sai Bi, Zexiang Xu, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. Deep reflectance volumes: Relightable reconstructions from multi-view photometric images. In *ECCV*, 2020. 2, 3
- [7] Zoubin Bi, Yixin Zeng, Chong Zeng, Fan Pei, Xiang Feng, Kun Zhou, and Hongzhi Wu. GS³: Efficient relighting with triple Gaussian splatting. In *SIGGRAPH Asia*, 2024. 3
- [8] Black Forest Labs. FLUX.1 Kontext: Flow matching for in-context image generation and editing in latent space. [arXiv:2506.15742](https://arxiv.org/abs/2506.15742), 2025. 7, 14, 16, 18
- [9] Blender Online Community. *Blender – a 3D modelling and rendering package*. Blender Foundation, 2025. 4, 13
- [10] Christophe Bolduc, Yannick Hold-Geoffroy, Zhixin Shu, and Jean-François Lalonde. GaSLight: Gaussian splats for spatially-varying lighting in HDR. In *ICCV*, 2025. 3
- [11] 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, 3
- [12] Mark Boss, Andreas Engelhardt, Abhishek Kar, Yuanzhen Li, Deqing Sun, Jonathan T. Barron, Hendrik P. A. Lensch, and Varun Jampani. SAMURAI: Shape and material from unconstrained real-world arbitrary image collections. In *NeurIPS*, 2022. 3
- [13] Iyaylo Boyadzhiev, Sylvain Paris, and Kavita Bala. User-assisted image compositing for photographic lighting. *ACM Trans. Graph.*, 32(4):36:1–12, 2013. 3
- [14] Chris Careaga and Yağız Aksoy. Intrinsic image decomposition via ordinal shading. *ACM Trans. Graph.*, 43(1):12:1–24, 2023. 2
- [15] Chris Careaga and Yagiz Aksoy. Physically controllable relighting of photographs. In *SIGGRAPH*, 2025. 3
- [16] Sumit Chaturvedi, Mengwei Ren, Yannick Hold-Geoffroy, Jingyuan Liu, Julie Dorsey, and Zhixin Shu. SynthLight: Portrait relighting with diffusion model by learning to re-render synthetic faces. In *CVPR*, 2025. 2, 3
- [17] Hongze Chen, Zehong Lin, and Jun Zhang. GI-GS: Global illumination decomposition on Gaussian splatting for inverse rendering. In *ICLR*, 2025. 3
- [18] Changwoon Choi, Juhyeon Kim, and Young Min Kim. IBL-NeRF: Image-based lighting formulation of neural radiance fields. *Comput. Graph. Forum*, 42(7):e14929, 2023. 3
- [19] JunYong Choi, SeokYeong Lee, Haesol Park, Seung-Won Jung, Ig-Jae Kim, and Junghyun Cho. MAIR++: Improving multi-view attention inverse rendering with implicit lighting representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 47(6):5076–5093, 2025. 3
- [20] Jun Myeong Choi, Annie Wang, Pieter Peers, Anand Bhattad, and Roni Sengupta. ScribbleLight: Single image indoor relighting with scribbles. In *CVPR*, 2025. 3, 7, 8, 18
- [21] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH*, pages 369–378, 1997. 5, 14
- [22] Youming Deng, Xueting Li, Sifei Liu, and Ming-Hsuan Yang. Physics-based indirect illumination for inverse rendering. In *3DV*, 2024. 3
- [23] Kang Du, Zhihao Liang, Yulin Shen, and Zeyu Wang. GS-ID: Illumination decomposition on Gaussian splatting via diffusion prior and parametric light source optimization. In *ICCV*, 2025. 3
- [24] Sylvain Duchêne, Clement Riant, Gaurav Chaurasia, Jorge Lopez Moreno, Pierre-Yves Laffont, Stefan Popov, Adrien Bousseau, and George Drettakis. Multiview intrinsic images of outdoors scenes with an application to relighting. *ACM Trans. Graph.*, 34(5):164:1–16, 2015. 3
- [25] Farshad Einabadi, Jean-Yves Guillemaut, and Adrian Hilton. Deep neural models for illumination estimation and relighting: A survey. *Comput. Graph. Forum*, 40(6):315–331, 2021. 3
- [26] Andreas Engelhardt, Amit Raj, Mark Boss, Yunzhi Zhang, Abhishek Kar, Yuanzhen Li, Deqing Sun, Ricardo Martin Brualla, Jonathan T. Barron, Hendrik P. A. Lensch, and Varun Jampani. SHINOBI: Shape and illumination using neural object decomposition via BRDF optimization in-the-wild. In *CVPR*, 2024. 3
- [27] Jiahui Fan, Fujun Luan, Jian Yang, Miloš Hašan, and Beibei Wang. RNG: Relightable neural Gaussians. In *CVPR*, 2025. 3
- [28] Fan Fei, Jiajun Tang, Ping Tan, and Boxin Shi. VMINer: Versatile multi-view inverse rendering with near- and far-field light sources. In *CVPR*, 2024. 3
- [29] Jian Gao, Chun Gu, Youtian Lin, Hao Zhu, Xun Cao, Li Zhang, and Yao Yao. Relightable 3D Gaussians: Realistic point cloud relighting with BRDF decomposition and ray tracing. In *ECCV*, 2024. 3
- [30] Ruiqi Gao, Aleksander Holynski, Philipp Henzler, Arthur Brussee, Ricardo Martin-Brualla, Pratul Srinivasan, Jonathan T. Barron, and Ben Poole. CAT3D: Create anything in 3D with multi-view diffusion models. In *NeurIPS*, 2024. 3, 5, 15
- [31] Elena Garces, Carlos Rodriguez-Pardo, Dan Casas, and Jorge Lopez-Moreno. A survey on intrinsic images: Delving deep into Lambert and beyond. *Int. J. Comput. Vis.*, 130:836–868, 2022. 2

- [32] David Griffiths, Tobias Ritschel, and Julien Philip. OutCast: Outdoor single-image relighting with cast shadows. *Comput. Graph. Forum*, 41(2):179–193, 2022. 3
- [33] Chun Gu, Xiaofei Wei, Zixuan Zeng, Yuxuan Yao, and Li Zhang. IRGS: Inter-reflective Gaussian splatting with 2D Gaussian ray tracing. In *CVPR*, 2025. 3
- [34] Ayaan Haque, Matthew Tancik, Alexei A. Efros, Aleksander Holynski, and Angjoo Kanazawa. Instruct-NeRF2NeRF: Editing 3D scenes with instructions. In *ICCV*, 2023. 7
- [35] Jon Hasselgren, Nikolai Hofmann, and Jacob Munkberg. Shape, light, and material decomposition from images using Monte Carlo rendering and denoising. In *NeurIPS*, 2022. 3
- [36] Kai He, Ruofan Liang, Jacob Munkberg, Jon Hasselgren, Nandita Vijaykumar, Alexander Keller, Sanja Fidler, Igor Gilitschenski, Zan Gojcic, and Zian Wang. UniRelight: Learning joint decomposition and synthesis for video relighting. In *NeurIPS*, 2025. 3
- [37] Chenxiao Hu, Meng Gai, Guoping Wang, and Sheng Li. Real-time global illumination for dynamic 3D Gaussian scenes. [arXiv:2503.17897](https://arxiv.org/abs/2503.17897), 2025. 3
- [38] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *ICLR*, 2022. 5, 14
- [39] Lianghai Huang, Wei Wang, Zhi-Fan Wu, Yupeng Shi, Huanzhang Dou, Chen Liang, Yutong Feng, Yu Liu, and Jingren Zhou. In-context LoRA for diffusion transformers. [arXiv:2410.23775](https://arxiv.org/abs/2410.23775), 2024. 7
- [40] Yingwenqi Jiang, Jiadong Tu, Yuan Liu, Xifeng Gao, Xiaoxiao Long, Wenping Wang, and Yuexin Ma. GaussianShader: 3D Gaussian splatting with shading functions for reflective surfaces. In *CVPR*, 2024. 3
- [41] Haiyan Jin, Yuan Li, Fujun Luan, Yuanbo Xiangli, Sai Bi, Kai Zhang, Zexiang Xu, Jin Sun, and Noah Snavely. Neural gaffer: Relighting any object via diffusion. In *NeurIPS*, 2024. 2, 3
- [42] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139:1–14, 2023. 3, 6, 15
- [43] Hoon Kim, Minje Jang, Wonjun Yoon, Jisoo Lee, Donghyun Na, and Sanghyun Woo. SwitchLight: Co-design of physics-driven architecture and pre-training framework for human portrait relighting. In *CVPR*, 2024. 3
- [44] Jeremy Klotz and Shree K. Nayar. Minimal sensing for orienting a solar panel. *Solar Energy*, 300:113833, 2025. 4
- [45] Peter Kocsis, Julien Philip, Kalyan Sunkavalli, Matthias Nießner, and Yannick Hold-Geoffroy. LightIt: Illumination modeling and control for diffusion models. In *CVPR*, 2024. 3
- [46] Peter Kocsis, Vincent Sitzmann, and Matthias Nießner. Intrinsic image diffusion for indoor single-view material estimation. In *CVPR*, 2024. 3
- [47] 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. 3
- [48] 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. 3
- [49] 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. 3
- [50] Zhengqin Li, Dilin Wang, Ka Chen, Zhaoyang Lv, Thu Nguyen-Phuoc, Milim Lee, Jia-Bin Huang, Lei Xiao, Cheng Zhang, Yufeng Zhu, Carl S. Marshall, Yufeng Ren, Richard Newcombe, and Zhao Dong. LIRM: Large inverse rendering model for progressive reconstruction of shape, materials and view-dependent radiance fields. In *CVPR*, 2025. 3
- [51] Ruofan Liang, Zan Gojcic, Huan Ling, Jacob Munkberg, Jon Hasselgren, Zhi-Hao Lin, Jun Gao, Alexander Keller, Nandita Vijaykumar, Sanja Fidler, and Zian Wang. DiffusionRenderer: Neural inverse and forward rendering with video diffusion models. In *CVPR*, 2025. 2, 3
- [52] Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, and Kui Jia. GS-IR: 3D Gaussian splatting for inverse rendering. In *CVPR*, 2024. 3
- [53] Zhihao Liang, Hongdong Li, Kui Jia, Kailing Guo, and Qi Zhang. GUS-IR: Gaussian splatting with unified shading for inverse rendering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 47(10):8364–8378, 2025. 3
- [54] Chih-Hao Lin, Jia-Bin Huang, Zhengqin Li, Zhao Dong, Christian Richardt, Tuotuo Li, Michael Zollhöfer, Johannes Kopf, Shenlong Wang, and Changil Kim. IRIS: Inverse rendering of indoor scenes from low dynamic range images. In *CVPR*, 2025. 2, 3
- [55] Yehonathan Litman, Fernando De la Torre, and Shubham Tulsiani. LightSwitch: Multi-view relighting with material-guided diffusion. In *ICCV*, 2025. 2, 3
- [56] Zhenyuan Liu, Yu Guo, Xinyuan Li, Bernd Bickel, and Ran Zhang. BiGS: Bidirectional Gaussian primitives for relightable 3D Gaussian splatting. In *3DV*, 2025. 3
- [57] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *ICLR*, 2019. 15
- [58] Jundan Luo, Duygu Ceylan, Jae Shin Yoon, Nanxuan Zhao, Julien Philip, Anna Frühstück, Wenbin Li, Christian Richardt, and Tuanfeng Y. Wang. IntrinsicDiffusion: Joint intrinsic layers from latent diffusion models. In *SIGGRAPH*, 2024. 2
- [59] Nadav Magar, Amir Hertz, Eric Tabellion, Yael Pritch, Alex Rav-Acha, Ariel Shamir, and Yedid Hoshen. LightLab: Controlling light sources in images with diffusion models. In *SIGGRAPH*, 2025. 2, 3, 8, 18
- [60] 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, 7
- [61] Yiqun Mei, Yu Zeng, He Zhang, Zhixin Shu, Xuaner Zhang, Sai Bi, Jianming Zhang, HyunJoon Jung, and Vishal M. Patel. Holo-relighting: Controllable volumetric portrait relighting from a single image. In *CVPR*, 2024. 3
- [62] Yiqun Mei, Mingming He, Li Ma, Julien Philip, Wenqi Xian, David M George, Xueming Yu, Gabriel Dedic, Ahmet Levant Taşel, Ning Yu, Vishal M. Patel, and Paul Debevec. Lux

- post facto: Learning portrait performance relighting with conditional video diffusion and a hybrid dataset. In *CVPR*, 2025. 3
- [63] Abhimitra Meka, Mohammad Shafiei, Michael Zollhöfer, Christian Richardt, and Christian Theobalt. Real-time global illumination decomposition of videos. *ACM Trans. Graph.*, 40(3):22:1–16, 2021. 3
- [64] Konstantin Mishchenko and Aaron Defazio. Prodigy: An expeditiously adaptive parameter-free learner. In *ICML*, 2024. 15
- [65] 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. 3
- [66] Merlin Nimier-David, Zhao Dong, Wenzel Jakob, and Anton Kaplanyan. Material and Lighting Reconstruction for Complex Indoor Scenes with Texture-space Differentiable Rendering. In *Eurographics Symposium on Rendering*, 2021. 2, 3
- [67] Rohit Pandey, Sergio Orts-Escolano, Chloe LeGendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. Total relighting: Learning to relight portraits for background replacement. *ACM Trans. Graph.*, 40(4), 2021. 3
- [68] Julien Philip, Michaël Gharbi, Tinghui Zhou, Alexei A. Efros, and George Drettakis. Multi-view relighting using a geometry-aware network. *ACM Trans. Graph.*, 38(4):78:1–14, 2019. 2, 3
- [69] Julien Philip, Sébastien Morgenthaler, Michaël Gharbi, and George Drettakis. Free-viewpoint indoor neural relighting from multi-view stereo. *ACM Trans. Graph.*, 40(5):194:1–18, 2021. 3
- [70] Yohan Poirier-Ginter, Alban Gauthier, Julien Philip, Jean-François Lalonde, and George Drettakis. A diffusion approach to radiance field relighting using multi-illumination synthesis. *Comput. Graph. Forum*, 43(4):e15147, 2024. 3
- [71] Poly Haven. Poly Haven: The public 3D asset library. <https://polyhaven.com/>, 2025. 4
- [72] Alexander Raistrick, Lingjie Mei, Karhan Kayan, David Yan, Yiming Zuo, Beining Han, Hongyu Wen, Meenal Parakh, Stamatis Alexandropoulos, Lahav Lipson, Zeyu Ma, and Jia Deng. Infinigen indoors: Photorealistic indoor scenes using procedural generation. In *CVPR*, 2024. 4, 13
- [73] Pramod Rao, Gereon Fox, Abhimitra Meka, Mallikarjun B R, Fangneng Zhan, Tim Weyrich, Bernd Bickel, Hanspeter Pfister, Wojciech Matusik, Mohamed Elgharib, and Christian Theobalt. Lite2Relight: 3D-aware single image portrait relighting. In *SIGGRAPH*, 2024. 3
- [74] Mengwei Ren, Wei Xiong, Jae Shin Yoon, Zhixin Shu, Jianming Zhang, HyunJoon Jung, Guido Gerig, and He Zhang. Relightful harmonization: Lighting-aware portrait background replacement. In *CVPR*, 2024. 3
- [75] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, 2022. 7
- [76] Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, and Giljoo Nam. Relightable Gaussian codec avatars. In *CVPR*, 2024. 3
- [77] Jonathan Schmidt, Simon Giebenhain, and Matthias Niessner. BecomingLit: Relightable Gaussian avatars with hybrid neural shading. In *NeurIPS*, 2025. 3
- [78] Yahao Shi, Yanmin Wu, Chenming Wu, Xing Liu, Chen Zhao, Haocheng Feng, Jian Zhang, Bin Zhou, Errui Ding, and Jingdong Wang. GIR: 3D Gaussian inverse rendering for relightable scene factorization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2025. 3
- [79] Shanlin Sun, Yifan Wang, Hanwen Zhang, Yifeng Xiong, Qin Ren, Ruogu Fang, Xiaohui Xie, and Chenyu You. Ouroboros: Single-step diffusion models for cycle-consistent forward and inverse rendering. In *ICCV*, 2025. 2
- [80] Alex Trevisan, Roni Paiss, Philipp Henzler, Dor Verbin, Rundi Wu, Hadi Alzayer, Ruiqi Gao, Ben Poole, Jonathan T. Barron, Aleksander Holynski, Ravi Ramamoorthi, and Pratul P. Srinivasan. SimVS: Simulating world inconsistencies for robust view synthesis. In *CVPR*, 2025. 3, 5
- [81] Junying Wang, Jingyuan Liu, Xin Sun, Krishna Kumar Singh, Zhixin Shu, He Zhang, Jimei Yang, Nanxuan Zhao, Tuanfeng Y. Wang, Simon S. Chen, Ulrich Neumann, and Jae Shin Yoon. Comprehensive relighting: Generalizable and consistent monocular human relighting and harmonization. In *CVPR*, 2025. 3
- [82] Shaofei Wang, Tomas Simon, Igor Santesteban, Timur Bagautdinov, Junxuan Li, Vasu Agrawal, Fabian Prada, Shou-I Yu, Pace Nalbony, Matt Gramlich, Roman Lubachersky, Chenglei Wu, Javier Romero, Jason Saragih, Michael Zollhoefer, Andreas Geiger, Siyu Tang, and Shunsuke Saito. Relightable full-body Gaussian codec avatars. In *SIGGRAPH*, 2025. 3
- [83] Chenfei Wu, Jiahao Li, Jingren Zhou, Junyang Lin, Kaiyuan Gao, Kun Yan, Sheng ming Yin, Shuai Bai, Xiao Xu, Yilei Chen, Yuxiang Chen, Zecheng Tang, Zekai Zhang, Zhengyi Wang, An Yang, Bowen Yu, Chen Cheng, Dayiheng Liu, Deqing Li, Hang Zhang, Hao Meng, Hu Wei, Jingyuan Ni, Kai Chen, Kuan Cao, Liang Peng, Lin Qu, Minggang Wu, Peng Wang, Shuting Yu, Tingkun Wen, Wensen Feng, Xiaoxiao Xu, Yi Wang, Yichang Zhang, Yongqiang Zhu, Yujia Wu, Yuxuan Cai, and Zenan Liu. Qwen-image technical report. [arXiv:2508.02324](https://arxiv.org/abs/2508.02324), 2025. 7, 8, 16, 17, 18
- [84] Liwen Wu, Rui Zhu, Mustafa B. Yaldiz, Yin hao 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, 3
- [85] Xiaoyan Xing, Konrad Groh, Sezer Karaoglu, Theo Gevers, and Anand Bhattad. LumiNet: Latent intrinsics meets diffusion models for indoor scene relighting. In *CVPR*, 2025. 3
- [86] Congrong Xu, Justin Kerr, and Angjoo Kanazawa. Splatfacto-W: A nerfstudio implementation of Gaussian splatting for unconstrained photo collections. [arXiv:2407.12306](https://arxiv.org/abs/2407.12306), 2024. 7
- [87] Linning Xu, Vasu Agrawal, William Laney, Tony Garcia, Aayush Bansal, Changil Kim, Samuel Rota Bulò, Lorenzo Porzi, Peter Kotschieder, Aljaž Božič, Dahua Lin, Michael Zollhöfer, and Christian Richardt. VR-NeRF: High-fidelity

- virtualized walkable spaces. In *SIGGRAPH Asia*, 2023. 6, 20
- [88] Kai Ye, Chong Gao, Guanbin Li, Wenzheng Chen, and Baoquan Chen. GeoSplatting: Towards geometry guided Gaussian splatting for physically-based inverse rendering. In *ICCV*, 2025. 3
- [89] Vickie Ye, Ruilong Li, Justin Kerr, Matias Turkulainen, Brent Yi, Zhuoyang Pan, Otto Seiskari, Jianbo Ye, Jeffrey Hu, Matthew Tancik, and Angjoo Kanazawa. gsplat: An open-source library for Gaussian splatting. *Journal of Machine Learning Research*, 26(34):1–17, 2025. 6, 15
- [90] Ye Yu, Abhimitra Meka, Mohamed Elgharib, Hans-Peter Seidel, Christian Theobalt, and William A. P. Smith. Self-supervised outdoor scene relighting. In *ECCV*, 2020. 3
- [91] Chong Zeng, Yue Dong, Pieter Peers, Youkang Kong, Hongzhi Wu, and Xin Tong. DiLightNet: Fine-grained lighting control for diffusion-based image generation. In *SIGGRAPH*, 2024. 2, 3
- [92] Jianshu Zeng, Yuxuan Liu, Yutong Feng, Chenxuan Miao, Zixiang Gao, Jiwang Qu, Jianzhang Zhang, Bin Wang, and Kun Yuan. Lumen: Consistent video relighting and harmonious background replacement with video generative models. [arXiv:2508.12945](https://arxiv.org/abs/2508.12945), 2025. 3
- [93] Zheng Zeng, Valentin Deschaintre, Iliyan Georgiev, Yannick Hold-Geoffroy, Yiwei Hu, Fujun Luan, Ling-Qi Yan, and Miloš Hašan. RGB \leftrightarrow X: Image decomposition and synthesis using material- and lighting-aware diffusion models. In *SIGGRAPH*, 2024. 2
- [94] Longwen Zhang, Qixuan Zhang, Minye Wu, Jingyi Yu, and Lan Xu. Neural video portrait relighting in real-time via consistency modeling. In *ICCV*, 2021. 3
- [95] Libo Zhang, Yuxuan Han, Wenbin Lin, Jingwang Ling, and Feng Xu. PRTGaussian: Efficient relighting using 3D Gaussians with precomputed radiance transfer. In *Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2024. 3
- [96] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Scaling in-the-wild training for diffusion-based illumination harmonization and editing by imposing consistent light transport. In *ICLR*, 2025. 2, 3
- [97] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 6
- [98] Xiaoming Zhao, Pratul P. Srinivasan, Dor Verbin, Keunhong Park, Ricardo Martin Brualla, and Philipp Henzler. IllumiNeRF: 3D relighting without inverse rendering. In *NeurIPS*, 2024. 3
- [99] Jinghao Zhou, Hang Gao, Vikram Voleti, Aaryaman Vasishtha, Chun-Han Yao, Mark Boss, Philip Torr, Christian Rupprecht, and Varun Jampani. Stable virtual camera: Generative view synthesis with diffusion models. In *ICCV*, 2025. 3, 5, 15
- [100] Yujie Zhou, Jiazi Bu, Pengyang Ling, Pan Zhang, Tong Wu, Qidong Huang, Jinsong Li, Xiaoyi Dong, Yuhang Zang, Yuhang Cao, Anyi Rao, Jiaqi Wang, and Li Niu. Light-a-video: Training-free video relighting via progressive light fusion. In *ICCV*, 2025. 3
- [101] Yongyang Zhou, Fang-Lue Zhang, Zichen Wang, and Lei Zhang. RTR-GS: 3D Gaussian splatting for inverse rendering with radiance transfer and reflection. In *ACM Multimedia*, 2025. 3
- [102] Jingsen Zhu, Yuchi Huo, Qi Ye, Fujun Luan, Jifan Li, Dianbing Xi, Lisha Wang, Rui Tang, Wei Hua, Hujun Bao, and Rui Wang. I²-SDF: Intrinsic indoor scene reconstruction and editing via raytracing in neural SDFs. In *CVPR*, 2023. 3
- [103] Rui Zhu, Zhengqin Li, Janarбек Matai, Fatih Porikli, and Manmohan Chandraker. IRISformer: Dense vision transformers for single-image inverse rendering in indoor scenes. In *CVPR*, 2022. 3