

LumiNet: Latent Intrinsic Meets Diffusion Models for Indoor Scene Relighting

Xiaoyan Xing¹ Konrad Groh² Sezer Karaoglu¹ Theo Gevers¹ Anand Bhattad³

¹University of Amsterdam ²BCAI-Bosch ³Toyota Technological Institute at Chicago

<https://luminet-relight.github.io>

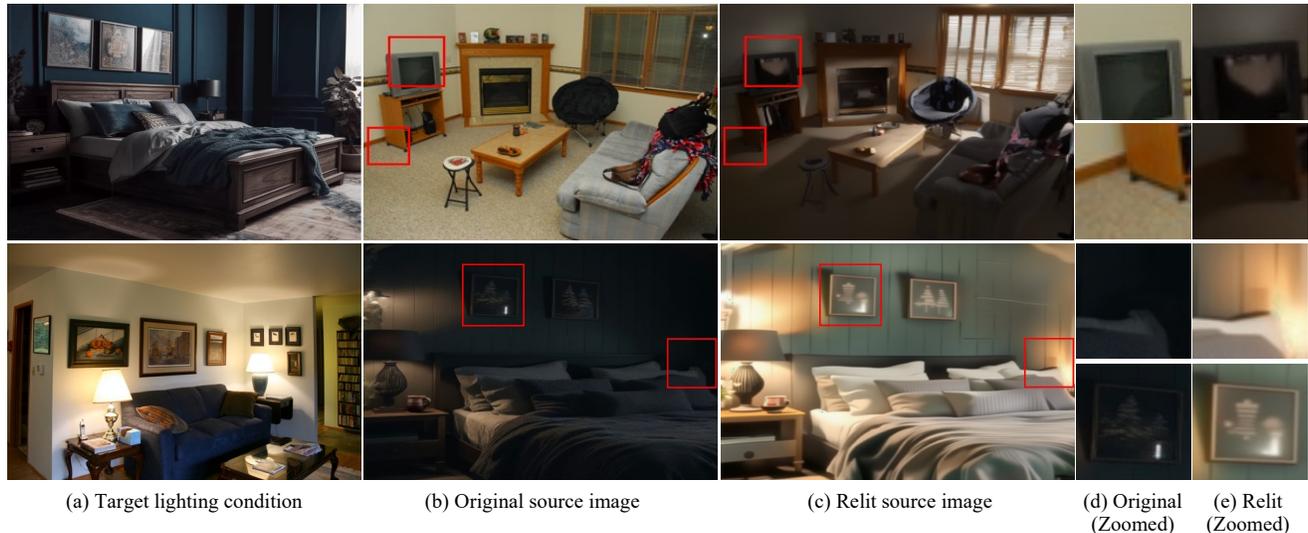


Figure 1. LumiNet transfers complex lighting conditions from a target image (a) to a source image (b), synthesizing a relit version of the source image (c) while preserving its geometry and albedo.

Abstract

We introduce *LumiNet*, a novel architecture that leverages generative models and latent intrinsic representations for transferring lighting from one image to another. Given a source image and a target lighting image, *LumiNet* generates a relit version of the source scene that captures the target’s lighting. Our approach makes two key contributions: a data curation strategy from the StyleGAN-based relighting model for our training, and a modified diffusion-based *ControlNet* that processes both latent intrinsic properties from the source image and latent extrinsic properties from the target image. We further improve lighting transfer through a learned adaptor that injects the target’s latent extrinsic properties via cross-attention and light-weight fine-tuning.

Unlike traditional *ControlNet*, which generates images with conditional maps from a single scene, *LumiNet* processes latent representations from two different images - preserving geometry and albedo from the source while transferring lighting characteristics from the target.

1. Introduction

Transferring lighting between indoor scenes has applications in cinematography, architectural visualization, and mixed reality. While neural rendering has advanced single-image relighting, transferring lighting across different scenes remains challenging due to complex geometry, materials, and illumination.

The difficulty lies in decomposing and transferring lighting across scenes with varying layouts and materials. Moreover, light must originate from luminaires, requiring an understanding of light sources. Indoor scenes exhibit complex phenomena such as interreflections and shadows [25]. Traditional inverse rendering struggles with model limitations and error propagation [10], while other methods require multi-view setups, focus on specific objects [5, 21], portraits [7, 15], or cannot handle complex lighting transfers [22, 26].

Recent work offers promising directions. Bhattad et al. [2] showed that StyleGAN’s latent space encodes lighting, but real-image transfer remains limited [1]. Zhang et al.

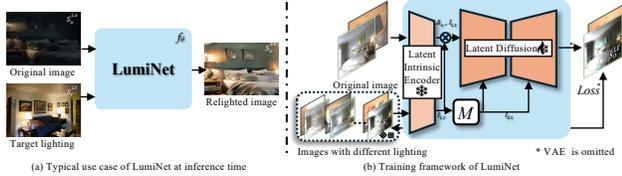


Figure 2. LumiNet’s Architecture and Training Pipeline.

[26] demonstrated that latent intrinsic decomposition captures albedo and illumination, though it struggles with complex scenes. Diffusion models [3, 17] with ControlNet [23] show strong conditional generation. DiffusionLight [14] and IC-Light [24] handle environment maps and portraits but not complex scenes.

We present LumiNet, a novel approach combining the strengths of these methods. By modifying ControlNet to operate on latent representations of scene intrinsics and extrinsics [26], we enable robust lighting transfer between arbitrary indoor scenes. First, we integrate a variational StyleGAN with real indoor data to alleviate mode collapse and address limited lighting variation. Second, we train a *Latent ControlNet* that transfers lighting features in latent space without explicit 3D or material modeling. Third, we introduce a lighting-aware adaptor that maps low-dimensional lighting vectors into high-dimensional codes, injected into a pretrained diffusion model by fine-tuning cross-attention layers to preserve lighting characteristics.

Our method relights challenging cases where source and target differ significantly (Fig. 1a–b), creating complex lighting phenomena like specular highlights, soft shadows, and interreflections (Fig. 1c–d). Extensive experiments show LumiNet outperforms previous methods, requiring only images as input, and surpasses prior SOTA by over 20% on the MIT Multi-Illumination dataset [13].

In summary, our contributions are:

- Novel framework: LumiNet combines latent intrinsic control with diffusion models for high-quality indoor scene relighting without 3D or multi-view inputs.
- Data curation: A variational StyleGAN approach generates diverse data from real indoor scenes.
- Generalizable relighting: Trained on same-scene pairs, LumiNet transfers lighting between different scenes.
- Plausible lighting effects: LumiNet reproduces specular highlights, shadows, and interreflections, validated by quantitative, qualitative, and user studies.

2. Data Preparation

Acquiring paired images of real-world scenes under different lighting conditions is extremely challenging, requiring controlled environments and extensive setups. We address this with a two-stage strategy: (1) a variational synthetic scene generation approach capturing key lighting patterns, and (2) curated in-the-wild images ensuring diverse and bal-

anced training data. This enables robust and photorealistic lighting transfer.

2.1. Variational Relit Scene Generation

StyLitGAN [2] generates relit images by interpolating StyleGAN’s latent space, mapping noise \mathbf{z} to style code \mathbf{w} and adding lighting direction \mathbf{d} . However, StyleGAN [6] suffers from mode collapse, producing similar outputs from different noise samples. GAN inversion mitigates this but is computationally slow [1].

To address this, we propose variational-StyLitGAN, mapping real images to latent space using a ConvNeXt-based [11] variational encoder $q_e(\mathbf{z}|\mathbf{x})$, then transforming \mathbf{z} into a style code \mathbf{w}^+ via a pretrained mapper, and reconstructing the scene image $\hat{\mathbf{x}}$ using a frozen StyLitGAN generator $p_d(\mathbf{x}|\mathbf{w}^+)$. We optimize:

$$\mathcal{L} = \underbrace{\text{MSE}(\mathbf{x}, \hat{\mathbf{x}}) + \text{LPIPS}(\mathbf{x}, \hat{\mathbf{x}})}_{\mathcal{L}_{\text{rec}}} + \underbrace{D_{\text{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \parallel \mathcal{N}(0, I))}_{\mathcal{L}_{\text{KL}}} \quad (1)$$

For dataset generation, we encode LSUN-bedroom images, map to \mathbf{w}^+ , and add lighting direction \mathbf{d} to generate seven lighting variations per scene. We further curate $\sim 1\text{K}$ high-quality images based on CLIP [16] similarity score

Although StyLitGAN provides good control, the domain gap with real images limits training solely on synthetic data. Thus, we mainly use this pipeline to enrich lighting variation diversity for training LumiNet.

2.2. In-the-Wild Training Data

To complement synthetic data, we use real-world datasets: MIIW [13] offers over 1,000 indoor scenes under 25 lighting conditions, capturing specular and direct lighting. Big-Time [9] contributes 460 scenes with 20–50 lighting variations via time-lapse captures. We also sample 1,000 images from LSUN Bedroom [20] for diversity. Overall, we train LumiNet on $\sim 2,500$ relit scenes and 1,000 unpaired LSUN scenes.

3. LumiNet

Our goal is to learn a generative model that transfers lighting between indoor scenes while preserving scene structure. The key challenge is modeling lighting interactions, addressed by leveraging latent intrinsic representations grounded in image formation theory.

3.1. Latent Intrinsic Extraction

Traditional pixel-space decomposition (e.g., albedo, normals) faces challenges: (1) perfect decomposition from monocular images is nearly impossible, and (2) full component recovery is expensive. We instead operate entirely in latent space.

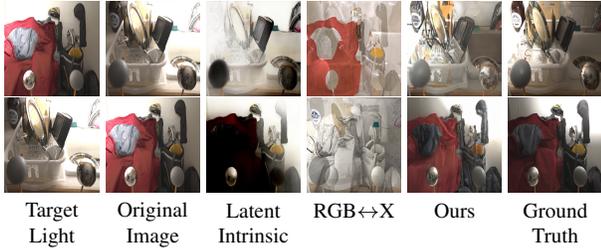


Figure 3. **Image Relighting Comparison on MIIW [13]**. Our method surpasses Latent Intrinsic [26], achieving superior relighting with finer geometric details and color. RGB↔X [22] fails with image prompting, and its text-prompt version lacks precise lighting control. Results are based on a fixed random seed and without any post-processing.

Using the pretrained model from Zhang et al. [26], given a scene pair $(S_o^{L_o}, S_o^{L_t})$ under lighting L_o and L_t , we extract latent intrinsic features $\mathcal{A}_o \in \mathbb{R}^{H \times W \times 128}$ and lighting codes $\{\mathcal{I}_{L_o}, \mathcal{I}_{L_t}\} \in \mathbb{R}^{16}$ using encoder f_λ .

3.2. Latent Intrinsic Control

Our illumination control consists of two parts. First, we implement control directly in latent space via a Latent Intrinsic ControlNet. We expand \mathcal{I}_{L_t} spatially and concatenate with \mathcal{A}_o to form $\{\mathcal{A}_o, \mathcal{I}_{L_t}\} \in \mathbb{R}^{H \times W \times 144}$, then process it through convolutions to obtain $\mathcal{L} \in \mathbb{R}^{H/2 \times W/2 \times 512}$.

Second, we enhance lighting control via cross-attention. A learned MLP $(3072 \rightarrow 4096 \rightarrow 4096 \rightarrow 4096 \rightarrow 3072)$ transforms \mathcal{I}_{L_t} into $\mathcal{I}_{E_t} \in \mathbb{R}^{3 \times 1024}$, matching text embedding dimensions. We omit text prompts to focus solely on image-based lighting control.

3.3. Training Objective

We train on same-scene lighting transfer using latent diffusion. Starting from $\epsilon(S^{L_t})$, we add noise to obtain $\epsilon(S^{L_t})_t$. The model predicts noise conditioned on timestep t , latent features $\{\mathcal{A}_o, \mathcal{I}_{L_t}\}$, lighting embedding \mathcal{I}_{E_t} , and scene S^{L_o} , optimizing:

$$\mathcal{L}_{\text{Lumi}} = \|\epsilon - \theta(\epsilon(S^{L_t})_t, t, \{\mathcal{A}_o, \mathcal{I}_{L_t}\}, \mathcal{I}_{E_t}, \epsilon(S^{L_o}))\|_2^2 \quad (2)$$

We train only the latent control network and cross-attention layers, freezing the rest of the diffusion and latent intrinsic models.

4. Experiment

We evaluate lighting transfer qualitatively (Fig. 3, Fig. 4) and quantitatively (Tab. 1), and assess perceptual quality through a user study (Tab. 2).

4.1. Quantitative Evaluation

We compare LumiNet against deep network methods (SA-AE [4], S3Net [19], Latent-Intrinsic [26]) and diffusion-

Methods	Labels	Raw Output		Color Correction	
		RMSE↓	SSIM↑	RMSE↓	SSIM↑
Input Img	-	0.384	0.438	0.312	0.492
SA-AE [4]	Light	0.288	0.484	0.232	0.559
SA-AE [4]	-	0.443	0.300	0.317	0.431
S3Net [19]	Depth	0.512	0.331	0.418	0.374
S3Net [19]	-	0.499	0.336	0.414	0.377
Latent-Intrinsic [26] ($\sigma = 0$)	-	0.326	0.232	0.242	0.541
Latent-Intrinsic [26]	-	0.297	0.473	0.222	0.571
RGB↔X [22]	-	0.256	0.476	0.253	0.470
Ours	-	0.180	0.647	0.144	0.673

Table 1. **Quantitative Evaluation - MIIW**. We evaluate our method on the multi-illumination dataset [13] with ground truth relights, achieving superior performance across all metrics. Unlike other approaches, our results require no post-processing, such as latent extrinsic search or flow-based cleanup, ensuring efficiency for large image pools.

Method	Surface Normal	Perceptual Relighting Quality		
	Median-AE ↓	I-PQ ↓	L-PQ ↓	P-PQ ↓
RGB↔X [22]	3.14	2.21	2.88	2.70
IC-Light-v2 [24]	3.42	3.06	2.57	2.74
Latent-Intrinsic [26]	3.61	2.24	2.52	2.40
Ours	2.74	1.71	1.30	1.40

Table 2. **Real-world Evaluation**. We evaluate surface normal consistency and conduct a user study inspired by [8], comparing our method against RGB↔X [22], IC-Light-v2 [24], and Latent-Intrinsic [26]. Users rated images generated by each method under the same target lighting (image or text prompt) on image quality (I-PQ), lighting quality (L-PQ), and prompt alignment (P-PQ). Our approach outperforms all baselines across all metrics, demonstrating strong and robust open-world relighting capabilities.

based RGB↔X [22] on the MIIW test set [13]. Following [26], we randomly sample an image and 12 reference lighting conditions, repeat experiments with different seeds, and report averaged results.

Tab. 1 shows two evaluations: raw output and color-corrected output (adjusting global RGB shifts). In both, LumiNet achieves SOTA performance on RMSE and SSIM, exceeding competing methods by over 20%.

Fig. 3 illustrates that LumiNet effectively transfers lighting effects (highlights, soft shadows) while preserving geometry and intrinsic properties. Latent-Intrinsic struggles with specific effects, and RGB↔X cannot transfer lighting across scenes, as it requires intrinsic channels from the same scene. Text-prompt-based methods are less suitable for fine-grained lighting control but included for room-level evaluations.

4.2. Geometry Consistency and User Study

We assess surface normal consistency and conduct a user study. Visual examples are shown in Fig. 4. For RGB↔X, only the text-prompt version is used, as irradiance-based relighting fails in this setting. For IC-Light [24], we use the latest FLUX-based IC-Light-v2.

We compute angular error (AE) between normals predicted from original and relit images (via RGB↔X). Tab. 2

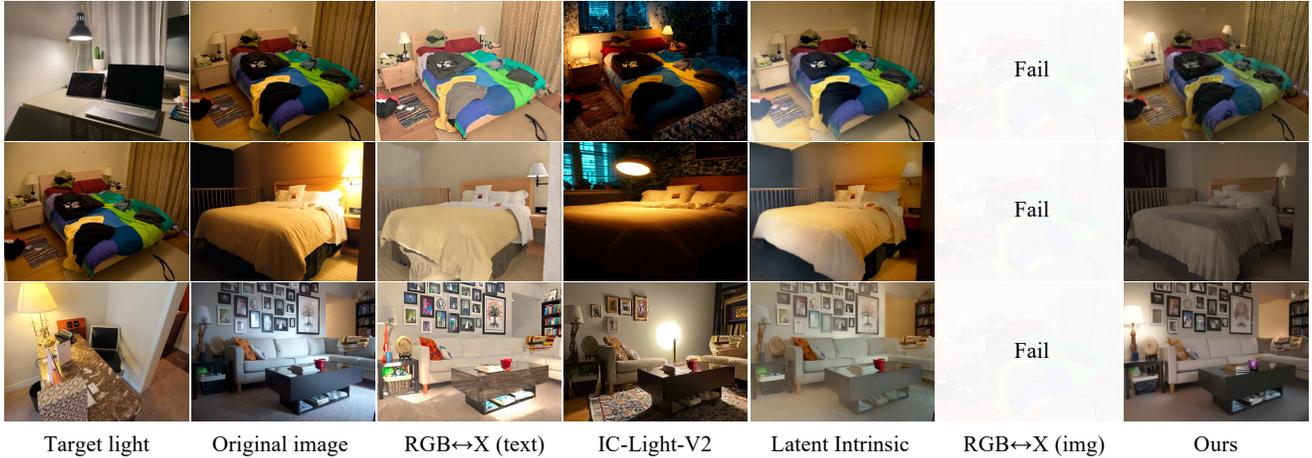


Figure 4. **In-the-wild image relighting visual comparison.** We evaluate LumiNet on diverse indoor scenes under various target lighting conditions, more in the supplemental. Both RGB↔X [22] and IC-Light-v2 [24] require text prompts to achieve relighting, where we use descriptions derived from the target lighting image (including actions like turning lights on/off, lamp placement, and scene type) as text prompts. In contrast, Latent Intrinsic [26] and our method rely solely on image input. When we pass the estimated irradiance from the target light image to RGB↔X’s intrinsic channels (RGB↔X image prompt), it fails to produce a meaningful image.



Figure 5. **Same Scene Under Various Lighting.** LumiNet can relight the same scene under different lighting conditions while preserving the overall layout, demonstrating effective disentanglement of intrinsic properties and lighting.

shows LumiNet preserves geometry best, with median AE $< 3^\circ$, outperforming others.

The user study with 31 participants evaluates (1) intrinsic preservation (I-PQ), (2) lighting realism (L-PQ), and (3) lighting alignment (P-PQ). Participants rank four methods; LumiNet consistently ranks first across metrics (Tab. 2), confirming perceptual quality.

5. Discussion

Our results demonstrate that photorealistic indoor relighting is achievable with a purely image-based, latent-space approach. Through latent intrinsic control and diffusion-based generation, LumiNet handles complex lighting phe-

nomena (cast shadows, specular highlights, indirect illumination) without requiring geometry or multi-view inputs.

Despite being trained only on same-scene pairs, LumiNet generalizes to cross-scene relighting, preserving structure and materials while transferring lighting between vastly different scenes. This stems from combining latent intrinsic features with conditional diffusion, enabling robust generalization.

Future work includes extending to dynamic scenes, improving 3D consistency, achieving real-time performance, and reducing artifacts without external cleanup like RF-Inversion. Our success suggests a broader shift toward latent-space manipulation over explicit physical modeling.

References

- [1] Anand Bhattad, Viraj Shah, Derek Hoiem, and DA Forsyth. Make it so: Steering stylegan for any image inversion and editing. *arXiv preprint arXiv:2304.14403*, 2023. 1, 2
- [2] Anand Bhattad, James Soole, and David A. Forsyth. Stylitgan: Image-based relighting via latent control. In *CVPR*, 2024. 1, 2
- [3] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 2020. 2
- [4] Zhongyun Hu, Xin Huang, Yaning Li, and Qing Wang. Sa-ae for any-to-any relighting. In *ECCV*. Springer, 2020. 3
- [5] Haian 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. 1
- [6] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, 2019. 2
- [7] 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. 1
- [8] Peter Kocsis, Julien Philip, Kalyan Sunkavalli, Matthias Nießner, and Yannick Hold-Geoffroy. Lightit: Illumination modeling and control for diffusion models. *arXiv preprint arXiv:2403.10615*, 2024. 3
- [9] Zhengqi Li and Noah Snavely. Learning intrinsic image decomposition from watching the world. In *CVPR*, 2018. 2
- [10] Zhengqi 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*, pages 555–572. Springer, 2022. 1
- [11] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *CVPR*, 2022. 2
- [12] I Loshchilov. Decoupled weight decay regularization. In *ICLR*, 2017. 1
- [13] Lukas Murmann, Michael Gharbi, Miika Aittala, and Fredo Durand. A multi-illumination dataset of indoor object appearance. In *ICCV*, 2019. 2, 3
- [14] Pakkapon Phongthawee, Worameth Chinchuthakun, Nontaphat Sinsunthithet, Amit Raj, Varun Jampani, Pramook Khungurn, and Supasorn Suwajanakorn. Diffusionlight: Light probes for free by painting a chrome ball. In *CVPR*, 2024. 2
- [15] Puntawat Ponglertnapakorn, Nontawat Tritrong, and Supasorn Suwajanakorn. Difareli: Diffusion face relighting. In *ICCV*, 2023. 1
- [16] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *ICML*. PMLR, 2021. 2
- [17] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, 2022. 2, 1
- [18] Katherine Xu, Lingzhi Zhang, and Jianbo Shi. Good seed makes a good crop: Discovering secret seeds in text-to-image diffusion models. *arXiv preprint arXiv:2405.14828*, 2024. 2
- [19] Hao-Hsiang Yang, Wei-Ting Chen, and Sy-Yen Kuo. S3net: A single stream structure for depth guided image relighting. In *CVPR*, 2021. 3
- [20] Fisher Yu, Yinda Zhang, Shuran Song, Ari Seff, and Jianxiong Xiao. Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*, 2015. 2
- [21] 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. 1
- [22] Zheng Zeng, Valentin Deschaintre, Iliyan Georgiev, Yannick Hold-Geoffroy, Yiwei Hu, Fujun Luan, Ling-Qi Yan, and Miloš Hašan. Rgb-x: Image decomposition and synthesis using material-and lighting-aware diffusion models. In *SIGGRAPH*, 2024. 1, 3, 4
- [23] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *ICCV*, 2023. 2
- [24] 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, 4
- [25] 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. *ACM ToG*, 40(1), 2021. 1
- [26] Xiao Zhang, William Gao, Seemantdar Jain, Michael Maire, David Forsyth, and Anand Bhattad. Latent intrinsics emerge from training to relight. In *NeurIPS*, 2024. 1, 2, 3, 4

LumiNet: Latent Intrinsic Meets Diffusion Models for Indoor Scene Relighting

Supplementary Material

Outline. We begin by detailing the training setup and implementation, followed by additional relit results and an illustration of the nearest-neighbor search.

A. Training details

We use Stable Diffusion 2.1 [17] as our base model to balance performance and training costs. To better preserve the details of the input images, we jointly estimate the denoised image and noise map at each denoising step (known as the v -prediction). Our method also applies to other objective functions, such as ϵ (only predicts the noise map). All training and testing are conducted on an 8-GPU NVIDIA A6000 Ada 48GB node. For the SD2.1 base model, we train on images with a resolution of 512×512 . An AdamW [12] optimizer with a learning rate of 4×10^{-5} and a decay rate of 0.9 is used. Training requires approximately 120 hours on a single GPU. At inference time, LumiNet outputs a relighted image (resolution: 512×512) in 5 seconds with 50 DDIM steps.

B. More results

Nearest neighbor search In complex real-world scene relighting, particularly for fine-grained light control, we observe that light transfer is highly sensitive to the choice of seed. To address this issue, we propose a nearest neighbor search over multiple seed candidates to identify the relit result that best approximates the target lighting.

Fig. S.1 illustrates the nearest neighbor search over multiple seed candidates to identify the relit result that best approximates the target lighting.

Additional relit images We present additional relit results in the following pages to demonstrate the robustness of our method under varying lighting conditions and across diverse scenes. These include an extension of our teaser figures (Fig. S.2), a scenario showcasing the effect of turning on ceiling lights (Fig. S.3), an example illustrating reduced ambient lighting in the room (Fig. S.4), and results depicting the effect of turning on lamps (Fig. S.5). We provide a detailed analysis of these phenomena in the figure captions and highlight the lighting effects using red bounding boxes within the figures.



Figure S.1. **Nearest Neighbor Search.** Diffusion models are sensitive to seed choice [18]. We observed that the choice of random seeds significantly impacts relighting quality. Here, we present sampled relights generated from 30 random seeds, sorted by their match to the target lighting image. Sorting is based on nearest-neighbor matching of the latent extrinsic (a low-dimensional lighting vector) to the target.



Figure S.2. Our LumiNet architecture transfers complex lighting conditions between indoor scenes using latent intrinsic representations while preserving scene layout, geometry, and albedo. Each scene shows an *original image* (left) paired with its *relighted* version (right) matching the target lighting shown at the top. Our method preserves scene structure and materials while accurately transferring lighting characteristics. **Left panel** demonstrates our method can adjust luminaires to match lighting conditions: it “knows” that to get more light in the right place in the room, it must switch on bedside lights (first row and second row) or table lamps (third row and fourth row), showing our model’s ability to handle direct illumination. Zoomed-in crops highlight the changes in images caused by relighting. In the first row, observe the added gloss on the wall behind the lamp in the top crop, as well as the effects on the side of the bed in the bottom crop, influenced by the invisible luminaire. In the second row, note the gloss removal on the side wall, as shown in the bottom crop. In the third row, you can see the reflection of the lamp on the large stationary glass window on the left, highlighted in the top crop. Finally, in the bottom row, observe the strong gloss added to the chair and the faint inter-reflection on the TV screen. **Right panel** shows natural lighting scenarios where bedside lamps are off. Top row’s crop shows suppressed specular reflections on the glass table and realistic lamp pole shadows added after relighting. Second row shows strong specular highlights on the wall clock and strong cast shadows from the AC unit. Third row captures soft ambient lighting with intricate specular details on window frames and appropriate surface sheen on furniture. Fourth row demonstrates the removal of bright light from the lamps and all indirect effects, including the recovery of sharp edges at the intersection of the ceiling and side walls.

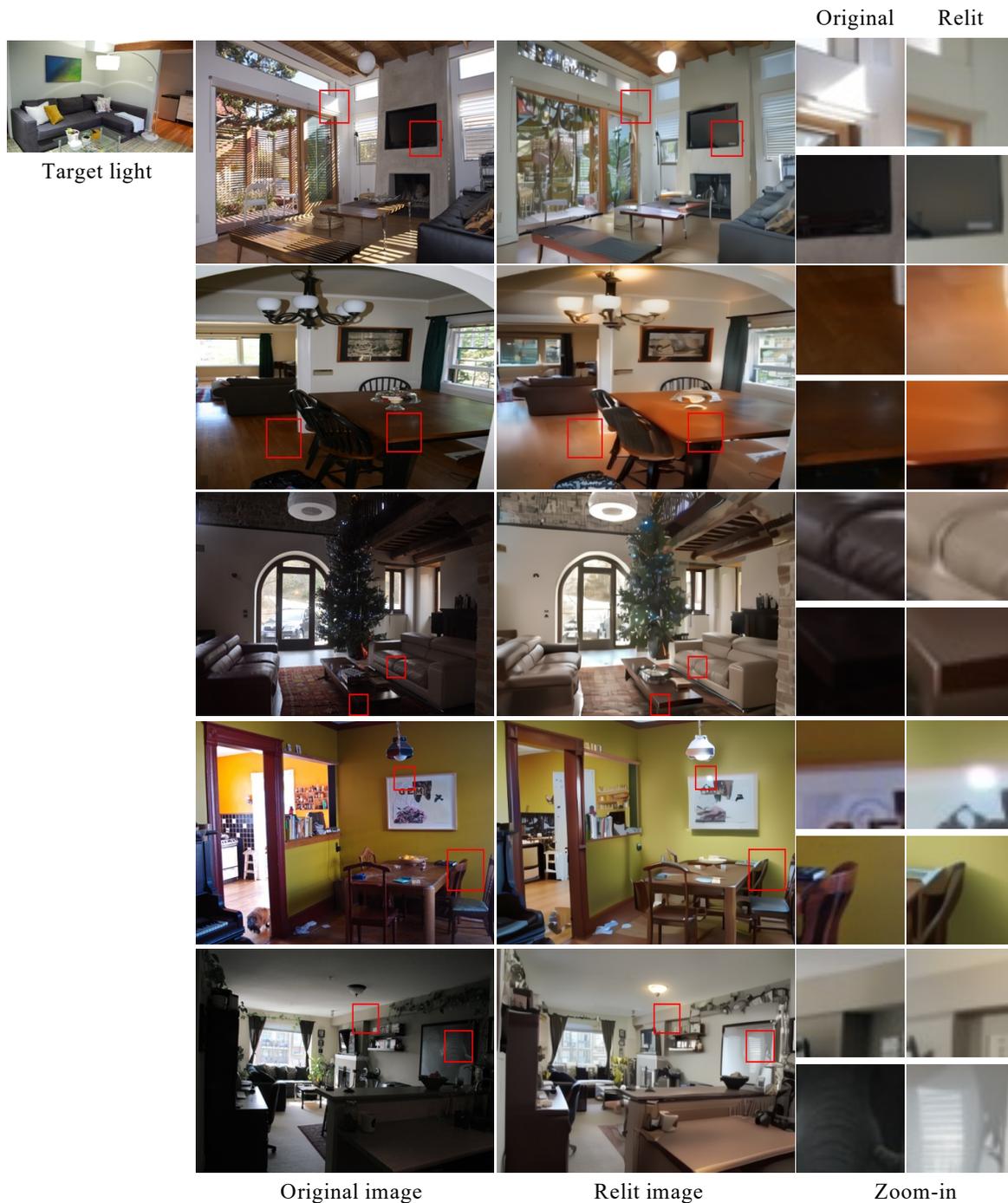


Figure S.3. **Additional Relit Images (switching on ceiling lamps)**. The target lighting is shown in the top-left image, where a ceiling lamp is switched on. Ceiling lamps are very rare in our training data; however, we find that LumiNet is still able to understand them and synthesize plausible relit images, as shown in the third column. In the first row, notice the suppression of gloss near the window at the top (see crop) and the added gloss due to inter-reflection on the TV screen. Also, note how the shaft lighting effect from the source image is suppressed. In the second row, observe how three ceiling lamps significantly brighten the room, with strong gloss visible on both the wooden floor and the dining table. In the third row, notice the sheen on the sofa and the edge of the coffee table, which become clearly visible after relighting. In the fourth row, see how the reflection of the lamp appears on the painting on the side wall. Also, note the shadow cast by the chair on the side wall below the painting. Finally, in the last row, observe how soft shadows along the edges of the ceiling and side wall are suppressed, while soft-light gloss becomes visible. Further, note the reflection on a mirror-like object in the bottom crop.

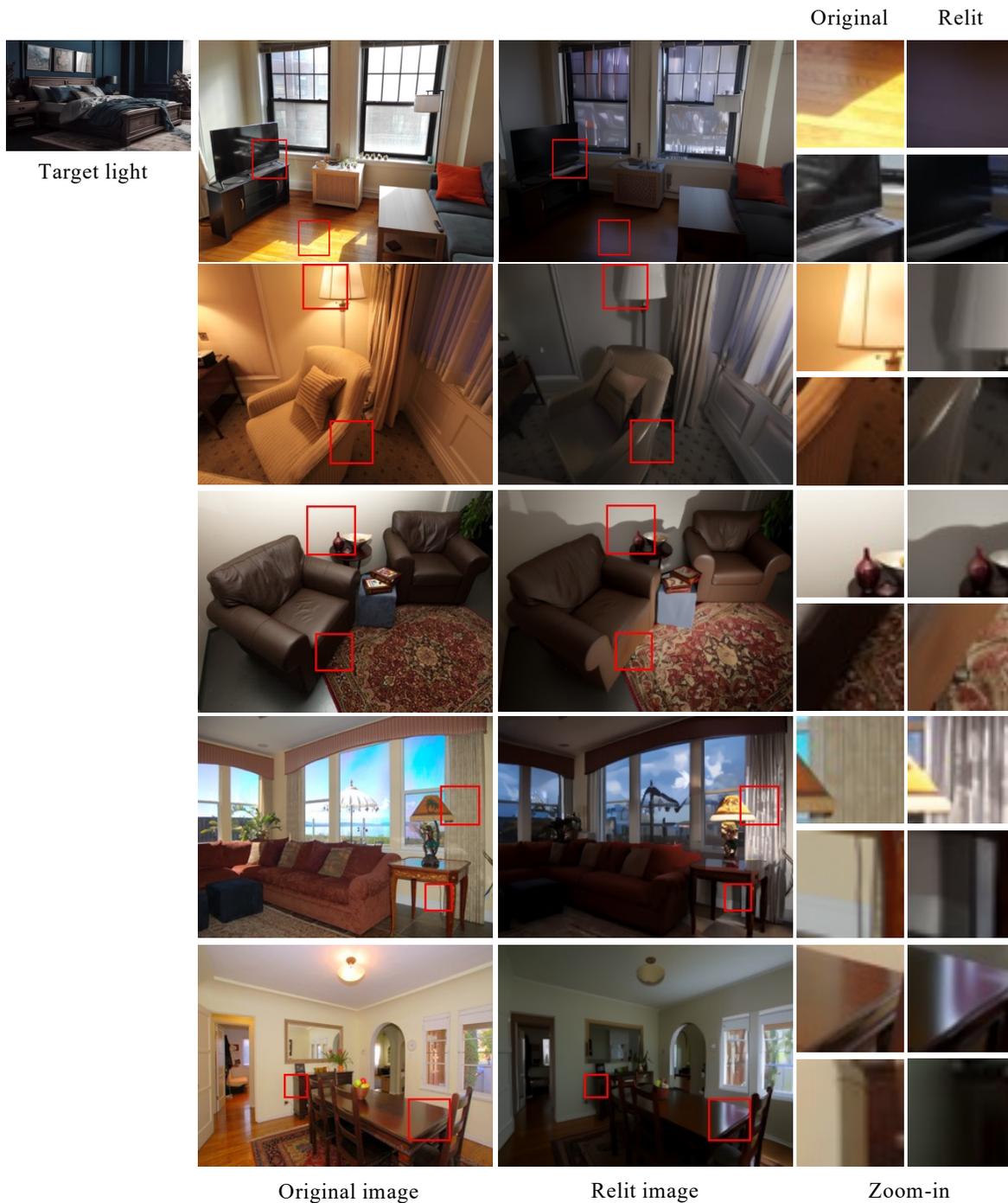


Figure S.4. **Additional Relit Images.** The target light is shown in the top-left image, where all lamps are switched off, and the only illumination comes from diffused natural light entering through a window on the right. The second column displays the source images to be relit to match the target light, while the third column presents the relit images. The final column highlights cropped regions before and after relighting, emphasizing the second-order lighting effects captured by LumiNet. In the top row (first relit image), note the table’s reflection in the TV and the strong gloss on the table from the directional window light. In the fourth row, observe how the sky changes to reflect the ambiance of the target light. In the last row, notice specular highlights on the table because of the direction light from the window. Also, notice the shadow cast by the cabinet in the bottom crop.

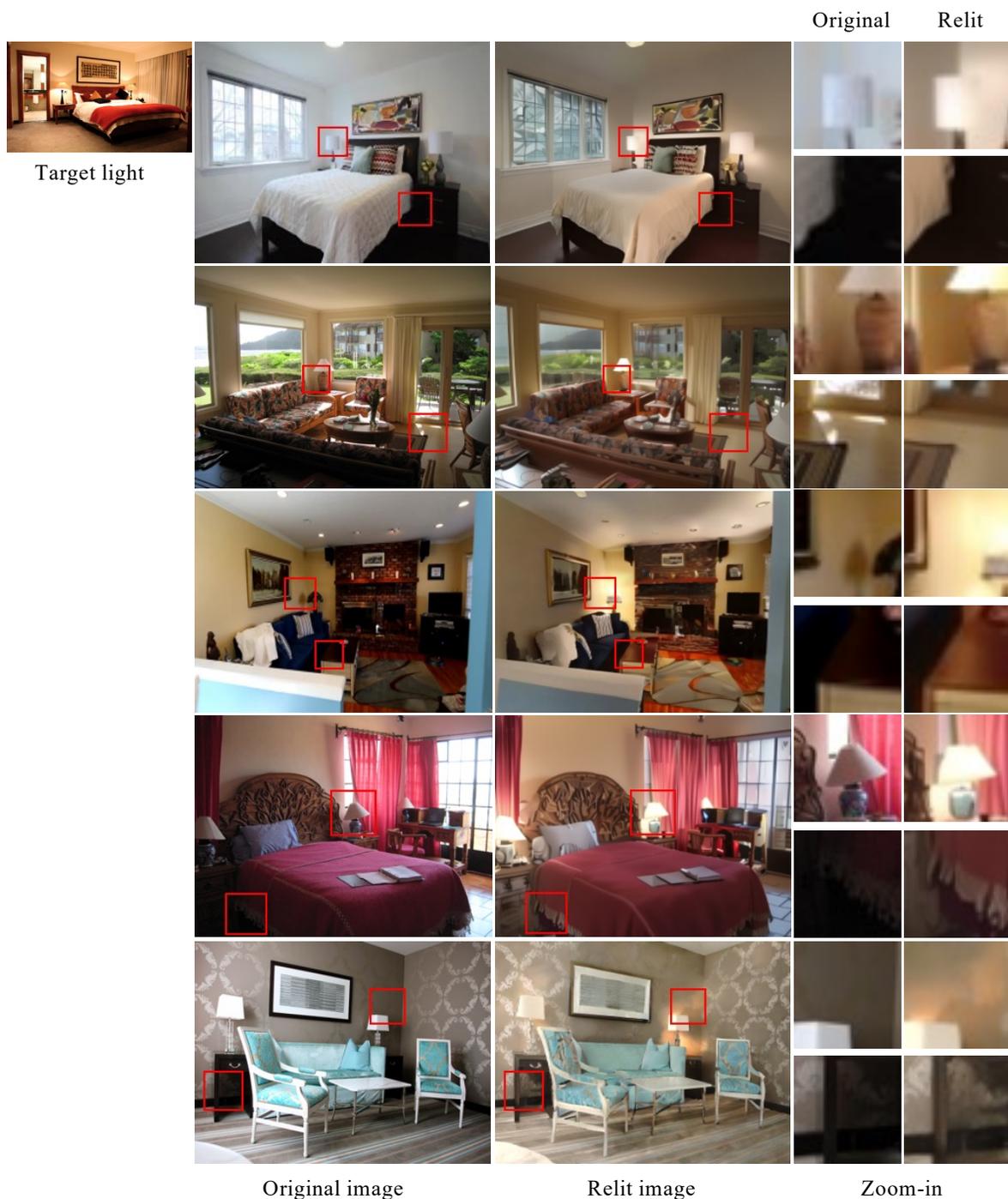


Figure S.5. **Additional Relit Images.** The target lighting is shown in the top-left image, where all lamps are switched on. The second column displays the source images to be relit to match the target lighting, where all lamps are switched off, and the third column presents the relit images. The final column highlights cropped regions before and after relighting. In the top row (first relit image), note the overall change in the room's color and the colored gloss added to the side of the bedsheets. In the second row, notice that the strong gloss on the carpet is removed. In the third row, switching on the side lamps removes the lamp shadow; also, observe the effect of the lamp on the ceiling and the gloss added to the edge of the table, as shown in the crop. In the fourth row, notice that the left side of the bed is now well-lit due to the lamp. Finally, in the last row, observe the gloss added to the wallpaper because of switching on the lamp