SPOTLIGHT: Shadow-Guided Object Relighting via Diffusion

Frédéric Fortier-Chouinard[1] Zitian Zhang[1] Louis-Etienne Messier[1]
Mathieu Garon[2] Anand Bhattad[3] Jean-François Lalonde[1]
[1]Université Laval, [2]Depix Technologies, [3]Johns Hopkins University
https://lvsn.github.io/spotlight

Figure 1. When inserting a piece of virtual furniture (chair) in an image, (left) existing diffusion-based renderers produce static composites without lighting control. In contrast, (right) SPOTLIGHT enables *shadow*-guided lighting control: a user specifies the desired shadows of the inserted object (inset)—SPOTLIGHT realistically blends the shadows and relights the object appropriately, acting as a virtual spotlight for the inserted object. Through our proposed shadow conditioning, we show that existing diffusion-based renderers can be guided to achieve realistic, controllable relighting without requiring any additional training.

Abstract

Recent work has shown that diffusion models can serve as powerful neural rendering engines that can be leveraged for inserting virtual objects into images. However, unlike typical physics-based renderers, these neural rendering engines are limited by the lack of manual control over the lighting, which is often essential for improving or personalizing the desired image outcome. In this paper, we show that precise and controllable lighting can be achieved without any additional training, simply by supplying a coarse shadow *hint for the object. Indeed, we show that injecting only the desired shadow of the object into a pre-trained diffusion-based neural renderer enables it to accurately shade the object according to the desired light position, while properly harmonizing the object (and its shadow) within the target background image. Our method,* SPOTLIGHT*, is entirely training-free and leverages existing neural rendering approaches to achieve controllable relighting. We show that* SPOTLIGHT *achieves superior object compositing results, both quantitatively and perceptually, as confirmed by a user study, outperforming existing diffusion-based models specifically designed for relighting. We also demonstrate other applications, such as hand-scribbling shadows and full-image relighting, demonstrating its versatility.*

1. Introduction

Object relighting—the process of inserting a virtual object in an image and lighting it appropriately—has traditionally been confined to 3D graphics pipelines, where lighting is explicitly modeled and rendered. In contrast, object relighting in real photographs remains a fundamentally ill-posed problem since it requires inferring the physical properties of the scene: its geometry, materials, and illumination are not observed, and their complex interactions—such as specular reflections, soft shadows, or material-dependent reflectance—must be inferred or approximated.

Recent advances in image generative models, especially diffusion-based methods [57], have opened up new possibilities for object relighting [40, 81, 83]. These models encode strong priors over the appearance of scenes under varied lighting, allowing implicit reasoning about illumination, material, and geometry. Yet, precise and localized relighting control remains limited, particularly in the context of virtual object insertion, where a single object must be relit.

We focus on the problem of local lighting control for object compositing using generative models. Specifically, we aim to relight an inserted object such that the object conforms to a specified lighting direction and intensity around it, while blending it naturally with the background. Achieving such control is challenging: the relighting must align

with physical cues like shadows and shading, yet remain flexible enough to work across varied materials, perspectives, and image contexts. Diffusion models offer powerful priors, but they are not inherently controllable without carefully designed guidance.

We present SPOTLIGHT, a training-free method for plausible, realistic, and controllable object relighting by leveraging the rich prior knowledge encoded in pre-trained, diffusion-based neural renderer. Rather than obtaining specialized high-quality data and training a model specifically for the task of object relighting (e.g., [30, 79]), our key insight is to drive the relighting process by supplying a coarse representation of the desired *shadow* of the object to a general-purpose neural renderer. Not only is a 2D shadow an intuitive input for human control, but we also find that diffusion models can effectively utilize it. We call our method SPOTLIGHT because it guides the diffusion process in a manner analogous to a physical spotlight, selectively illuminating the object to insert in the image.

SPOTLIGHT exploits classifier-free guidance within existing pre-trained diffusion-based renderers and progressively harmonizes imperfect shadows with the background while relighting the object. Our guidance mechanism steers the rendering process and modifies the appearance of the inserted composite, resulting in local lighting control. This mechanism is applied to recent diffusion models pre-trained to render images from input intrinsic maps (depth, normals, albedo, etc.) [80, 83].

Fig. 1 shows an example of such shadow-based lighting control: the shadow of the chair is rendered using basic shadow mapping [20] according to the desired lighting direction. From this, SPOTLIGHT realistically blends the input guiding shadow within the background, and relights the virtual object to match the desired lighting.

In summary, our contributions are as follows. First, we introduce a framework that enhances pre-trained diffusion renderers with object relighting control without requiring additional training. We demonstrate superior quality, both perceptually through user studies and quantitatively, compared to competitive specialized models. Finally, we release an evaluation dataset tailored for lighting control in object insertion, which will be made available upon publication.

2. Related work

Conditional image generation. Diffusion-based generative models have become the de facto choice for image generation due to their ability to produce high-quality images [28, 57, 63, 64]. These models offer control mechanisms for various modalities, such as text [53, 57], image content [22, 58], intrinsic images [40, 44, 80, 83], classifier guidance [17] and cross-modal data [12, 77]. Notably, training-free guidance methods [3, 4, 21, 27, 47] have enabled image editing capabilities without specific training, forming the basis of our approach.

Image relighting modifies the global or local shading of an image without changing other properties such as geometry and materials. Previous methods for harmonization [35, 51, 56, 67] do not prioritize physically accurate shading of the objects. Recently, [52] utilized a single-view multi-illumination dataset [48] to enable direct control over the dominant lighting direction. Retinex theory has been leveraged [11, 72] for relighting indoor scenes. IC-Light [81] enforces consistency in appearance to relight portraits and various objects. LumiNet [73] and ScribbleLight [15] achieve image relighting through illumination transfer or user-defined scribbles, respectively. Some methods specialize in outdoor scenes, using geometric priors like depth [26] or normalized coordinates [32] or both [41]. These methods however cannot handle lighting on specific region or objects in the scene. Recent work on object relighting has evolved from consistency-based approaches in intrinsic images [9] to harmonization of foreground-background albedo [14]. Recently, diffusion models have been adapted for the task of rendering from partial intrinsic maps [40, 80, 83], which can be leveraged to achieve zero-shot object composition. However, these methods have limited lighting control. ZeroComp [83] provides no lighting control, RGB↔X [80] can only provide lighting control through a text prompt, and DiffusionRenderer [40] requires a prohibitively long and approximate lighting estimation process to annotate each sample in the training set. Other diffusion-based approaches, such as Neural Gaffer [30], DiLightNet [79] and IllumiNeRF [84] utilize HDR environment maps and multi-lighting renders, but are not designed for object composition in existing images. LightLab [46] achieves fine-grained control over existing light sources, but does not handle object insertion. Careaga et al. [13] propose a sophisticated inverse rendering method for data generation, wheras SPOTLIGHT is training-free. IntrinsicEdit [45] proposes a training-free method for lighting-based editing, but, unlike our approach, lacks explicit control over object-specific virtual lighting effects. We bridge this gap by introducing a shadow-guided strategy for object relighting that leverages existing models without any additional training.

Explicit lighting estimation approaches infer HDR lighting from a single image [23, 25, 37, 39, 50, 86], though they generally lack controllability. Parametric models [16, 24, 70] or GAN inversion methods [69] offer more control, but these methods require physics-based rendering engines for generating the composite image—here, we focus instead on neural renderers.

Shadow generation. Generative models for rendering realistic shadows [29, 42, 43, 82], or harmonizing rendered shadows with the image [68] have been proposed. Other methods enable controllable shadows for 3D [60] and 2D objects [61, 62]. Our approach can leverage exist-

ing shadow generation methods (e.g., [65]) and shows that diffusion-based neural renderers can be conditioned on such shadows to achieve object relighting.

Image intrinsics. Decomposing images into albedo and shading has long been studied [7, 14, 34, 49, 71]. Numerous works have also focused on estimating scene geometry, for example, depth [8, 31, 54, 74, 75, 78] and normals [5, 6, 19, 55, 76]. Recently, approaches such as [37, 86] jointly infer different intrinsic maps including shape, spatially-varying lighting, scene geometries, and materials. Conversely, generative models have shown a powerful internal understanding of intrinsic properties [2, 10, 18, 31, 44, 76]. We similarly leverage diffusion models' ability to interpret intrinsic maps to generate realistic images as a foundation for rendering image composites.

3. Background: diffusion renderers

Recent work has shown that diffusion models can be adapted to generate photorealistic images from intrinsic maps including materials (e.g., albedo, roughness, metalness), geometry (e.g., surface normals and depth), and shading. This capability was built using a ControlNet in ZeroComp [83] or full finetuning in RGB↔X [80]. We refer to this new class of conditional generation models as "diffusion renderers".

Specifically, ZeroComp [83], which we will use in sec. 5, is trained to generate images from albedo, normals, depth, and partially-masked shading. It can be used for compositing in a zero-shot manner by alpha-compositing the intrinsics of the background and the object, and by masking out a region on and around the object in the shading map. The network then generates the full image with realistic shading. The intrinsics of the object are rendered using simple graphics shaders, while those of the background are estimated using pre-trained networks (e.g., [8] for depth, see sec. 5.2). In the case of RGB↔X [80], their X→RGB network achieves object insertion by accepting as input alpha-composited albedo, normals, and a masked image. They then train an inpainting version of their network to generate the final image. Unlike traditional rendering, these methods offer no control over the lighting conditions.

In this work, we use such a pre-trained diffusion renderer and show that, by incorporating an *approximate* guiding shadow, we can steer the diffusion model to enable local lighting control, *without any additional training*. We now describe our approach in detail.

4. SpotLight

We employ a user-provided, guiding shadow to steer a pre-trained diffusion renderer (c.f. sec. 3) to follow the desired lighting direction. Our key insight is that even an approximate shadow encapsulates enough important lighting information and, by integrating this cue in the denoising process, we can simultaneously refine the composite appearance and its shadow, yielding a natural blend with the scene.

4.1. Approach overview

An overview of SpotLight is illustrated in fig. 2. In addition to the intrinsic maps for the backbone diffusion renderer, our approach requires two inputs: an object mask $\mathbf{m}_{\text{obj}}$ and a user-specified guiding shadow $\mathbf{m}_{\text{shw}}$ which guides the latent diffusion process. The guiding shadow can be obtained in several ways—in this paper, we experiment with fast rasterization [20], a trained shadow generation model [65], and hand-drawn scribbles. Without retraining the diffusion renderer, we steer its generative process by incorporating the guiding shadow into the latent space.

4.2. Blending shadows with the background

Provided the guiding shadow $\mathbf{m}_{\text{shw}}$, we first need to ensure that the diffusion model will properly blend the shadow with the background. We take inspiration from Blended Latent Diffusion [4] to progressively blend the noisy latents $\mathbf{z}_t$ at time t with a noised VAE-encoded image which contains the shadow roughly composited over the image, $\mathbf{g}$. In practice, we obtain $\mathbf{g}$ by compositing the object albedo and the guiding shadow on the background. The update made at each timestep to incorporate the desired shadow latents is

$$\tilde{\mathbf{z}}_t = (1-\beta\mathbf{m}_{\text{shw},\downarrow})\odot\mathbf{z}_t+(\beta\mathbf{m}_{\text{shw},\downarrow})\odot\text{noise}(\mathcal{E}(\mathbf{g}),t)\,, \quad (1)$$

where $\beta = 0.05$ is the shadow latent weight and $\mathbf{m}_{\text{shw},\downarrow}$ is the shadow mask, bilinearly downsampled to the latent resolution (64×64), with the edges opacity increased and dilated to maintain the desired shadow softness (see supp. for details). Noise is added to the encoded composite using the DDIM [64] noise scheduler. Different from [4], we do not use the slow test-time decoder optimization nor the progressive mask shrinking steps.

4.3. Enhancing lighting control

Simply running a diffusion renderer conditioned on the desired shadow direction yields perceptible, but subtle visual changes in the generated object (as will be shown in sec. 5.6). We propose to amplify the effect by running two parallel branches of the model: one which receives the shadow in the desired direction (positive branch), while the other has the shadow in the opposite direction (negative branch), obtained by shifting the light azimuth angle by $180°$. Although the opposite light direction works best for the negative branch, we found that casting no shadow also provides a good negative sample. From the outputs of the diffusion renderer (typically, v-predictions [59], i.e., linear combination of noise and estimated image), we then use classifier-free guidance to amplify the effect of the positive

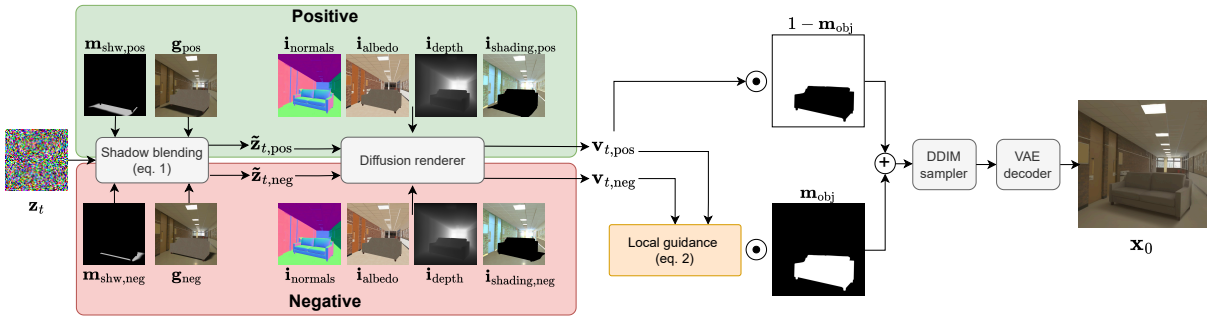


Figure 2. Overview of SPOTLIGHT. We leverage diffusion-based neural renderers, pre-trained to render images from input intrinsics (normals, albedo, depth, and partial shading). Our proposed framework consists of two parallel branches: a positive branch (green), where the user-provided guiding shadow aligns with the desired light direction, and a negative branch (red), where the shadow is aligned with an opposite light direction. Initially, a latent blending operation merges the noisy latents with a rough scene composition using the background image, object albedo, and a coarse shadow. Both branches are then processed by a diffusion renderer conditioned on intrinsic images. Finally, relighting of the object is amplified using local guidance.

shadow on the lighting on the object:

$$\begin{aligned}\tilde{\mathbf{v}}_t = (1 - \mathbf{m}_{\text{obj},\downarrow}) \odot \mathbf{v}_{t,\text{pos}} \\ + \mathbf{m}_{\text{obj},\downarrow} \odot (\mathbf{v}_{t,\text{neg}} + \gamma(\mathbf{v}_{t,\text{pos}} - \mathbf{v}_{t,\text{neg}}))\,,\end{aligned} \tag{2}$$

where $\gamma = 3.0$ is the guidance scale, $\mathbf{m}_{\text{obj},\downarrow}$ is the object mask bilinearly downsampled to the latent resolution (64×64). After eq. (2), we run a regular DDIM sampling step to obtain the latents $\mathbf{z}_{t-1}$.

4.4. Final image synthesis

After the guided diffusion process, the resulting latents $\mathbf{z}_0$ are decoded with the VAE decoder. The background preservation strategy of [83] is applied to ensure that only the object and its shadow are modified, yielding a composite image with natural local lighting.

5. Evaluation on 3D object compositing

Evaluating composition methods where lighting conditions can be controlled is challenging, as the quality of the results must be assessed across multiple lighting condition for the same scene. To address this, we build over the evaluation setup of [83] where 3D objects are realistically rendered and composited into background images using a physics-based renderer and ground-truth lighting extracted from HDR panoramas. We evaluate two scenarios: (1) "reference-based", which preserves the original scene lighting to assess the ability of methods to produce images close to the ground truth, and (2) "user-controlled", where the original lighting is modified (e.g., dominant light rotation) to simulate user-specified lighting conditions, requiring methods to adapt while maintaining a perceptually pleasing render. In this section, we describe how these scenarios are used for quantitative evaluation and a user study to assess the realism and controllability of the methods.

5.1. Evaluation dataset

We create datasets for the two scenarios using the Blender scenes graciously provided by the authors of [83]. We first improve the 3D object rendered shadow by replacing the planar shadow catcher by a scene mesh using estimated depth from Depth Anything V2 [75]. We also improve the ground truth intrinsic maps with antialiased versions.

"Reference-based" scenario. To assess the capacity of the methods to generate a composite in the ground truth lighting conditions, this scenario uses renders with the depth-warped HDR as in [83] to ensure lighting fidelity around the inserted object. We extract the dominant light direction and use it to render the guiding shadow using blender's EEVEE rasterization-based rendering engine, which employs real-time shadow mapping [20]. A total of 210 images featuring various objects are rendered, with example samples shown in fig. 3, in the "ground truth" column. This dataset is intended for quantitative evaluation (sec. 5.4).

"User-controlled" scenario. This dataset simulates a scenario where a user specifies a desired light position. 8 lighting directions are defined, spaced at $45°$ increments in azimuth around the object, and used to create the guiding shadows, again using blender's EEVEE engine. The same set of background and object combinations as in the "reference-based" version is used, resulting in a total of 1,680 images. This dataset is used in the user study (sec. 5.5) to evaluate the perceptual accuracy of renders under modified lighting conditions.

Note that these evaluation datasets are independent of the datasets used to train both the backbones employed in SPOTLIGHT (sec. 5.2) and the baselines (sec. 5.3).

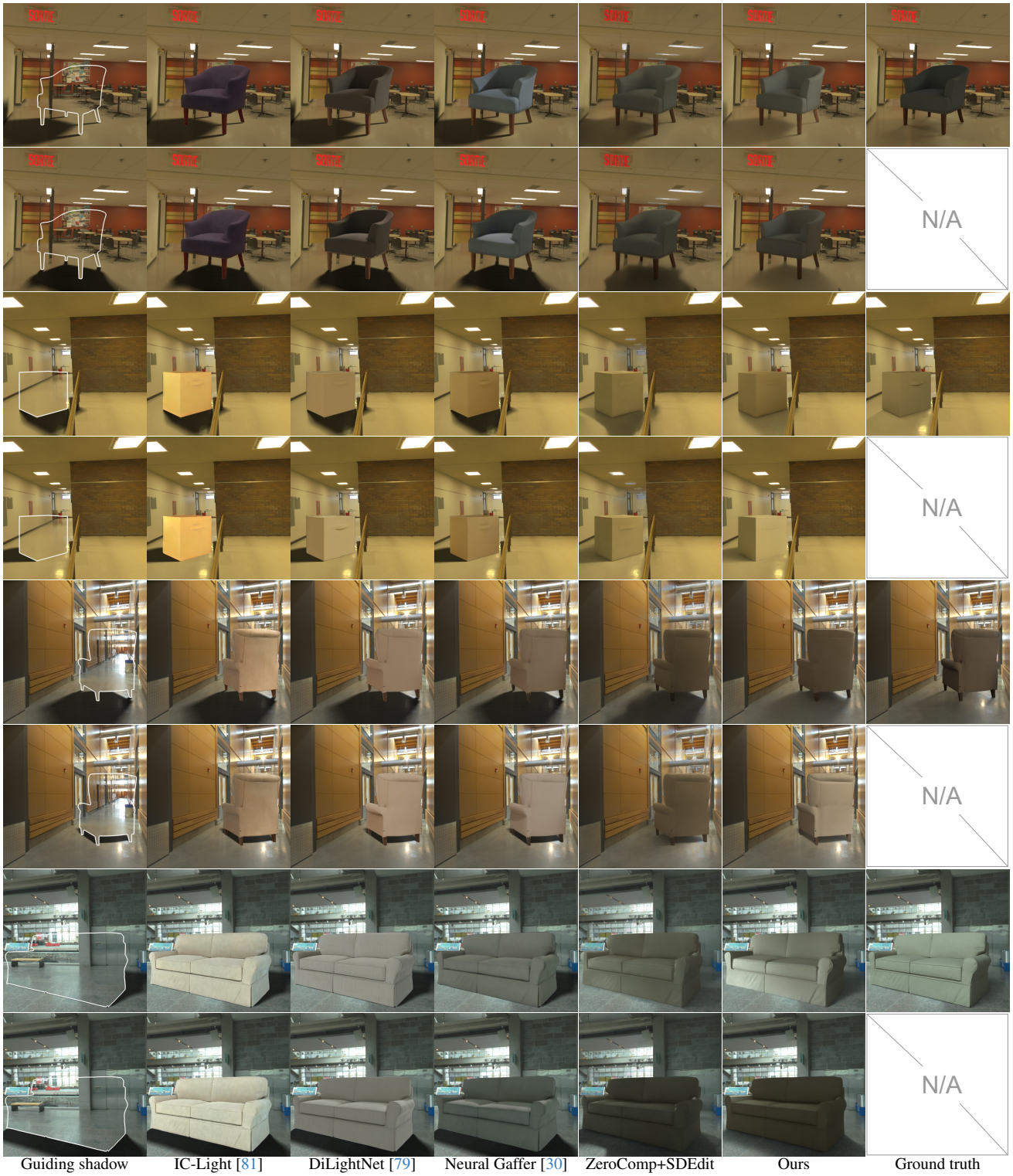


Figure 3. Qualitative comparison against the baselines on our evaluation dataset. For each scene, we show the dominant shadow direction from the "reference-based" dataset, where the ground truth is available, and one from the "user-controlled" dataset. Observe how SPOTLIGHT generates results that are visually closest to the ground truth (odd rows) and accurately match the guiding shadow (even rows) without changing the color of the object (IC-Light, DiLightNet), preserving the shape of the guiding shadow (ZeroComp+SDEdit), and generates more visible shading effects on the object (Neural Gaffer), all without being trained specifically for object relighting (DiLightNet, Neural Gaffer). Please zoom in and consult the supp. for additional qualitative results.

Method	PSNR↑	SSIM↑	RMSE↓	MAE↓	LPIPS↓
Light-conditioned methods					
DiLightNet [79]	24.67	0.948	0.064	0.022	0.042
Neural Gaffer [30]	28.44	0.963	0.042	0.015	0.038
Shadow-conditioned methods					
IC-Light [81]	26.87	0.959	0.054	0.019	0.040
ZeroComp+SDEdit [47]	26.00	0.938	0.053	0.025	0.079
SPOTLIGHT	30.68	0.974	0.033	0.012	0.030

Table 1. Quantitative results obtained by conditioning the methods on the ground truth dominant light direction. All metrics are computed on our "reference-based" evaluation dataset against the ground truth on the full image (see supp. for foreground- and background-only metrics). SPOTLIGHT surpasses all baselines. Results are color coded by best and second- best.

5.2. SPOTLIGHT specifics

As mentioned in sec. 3, our proposed SPOTLIGHT readily applies to existing diffusion-based neural renderers, such as ZeroComp [83] or RGB↔X [80]. Here, we provide details on ZeroComp, see supp. for RGB↔X.

We use the variant trained on Openrooms [38] provided by the authors. The masked shading map is obtained by dividing the background image by the albedo, and by masking out the shading on both the object and shadow regions. Given its superior results, we adopt this method as our default model for 3D object compositing. To extract intrinsics from real images, we use the same conditioned intrinsic estimators as in [83], namely StableNormal [76] for normals, IID [33] for albedo, and ZoeDepth [8] for depth.

5.3. Baselines

We compare SPOTLIGHT with two types of baselines, that is methods that are conditioned on: the guiding shadow (like SPOTLIGHT); or an explicit lighting representation, through a 360° environment map. We now describe per-method details. OpenImageDenoise [1] is also used to denoise the results of all methods to limit the impact of rendering noise present in the synthetic training data of some models.

5.3.1 Shadow-conditioned methods

These methods, like ours, use the user-specified guiding shadow $\mathbf{m}_{\text{shw}}$ as input (obtained with EEVEE, see sec. 5.1).

ZeroComp [83]+SDEdit [47]. We create a strong baseline by first running ZeroComp with the shading of the object masked out, and composite the guiding shadow $\mathbf{m}_{\text{shw}}$ over the background shading intrinsic map. To blend the shadows, we perform a refinement step, inspired by SDEdit [47], by noising this prediction for 50% of the timesteps, and rerunning ZeroComp for the remaining timesteps, this time with the shading of the whole image masked out.

IC-Light [81]. We evaluate the background-conditioned model of IC-Light, where the background image with the shadow is used for conditioning, with the object lit by a constant ambient lighting as input. No prompts are used.

5.3.2 Light-conditioned methods

Object relighting methods accept an explicit lighting representation as input in the form of a 360° environment map, which we generate using the parametric light direction (see supp.). Unfortunately, these methods focus solely on object relighting—they do not create any cast shadows outside the object. To prevent this absence of shadows from severely hampering the realism of their results, we composite the guiding shadow onto the background image.

DiLightNet [79] requires a prompt describing the object, which we obtain by feeding the object lit by constant ambient lighting and composited on a black background to BLIP-2 [36] with the prompt "Putting aside the black background, this object is a". DiLightNet also requires depth and radiance hints: we obtain depth from the 3D model of the object and render the radiance hints using the input environment map as outlined in [79]. The rendered object is then composited with the background.

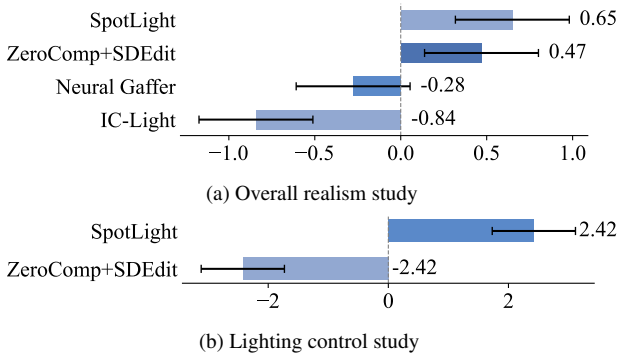
Neural Gaffer [30]. The object lit by constant ambient lighting is composited on a white background before feeding it to the network, as expected by the method. Since Neural Gaffer is trained at a lower resolution (256×256), its result is upsampled to 512×512 then composited on the full resolution background. We noticed Neural Gaffer tends to generate noticeable white boundary artifacts—we remove them through an erosion operation to avoid overly penalizing it.

5.4. Experimental results

Tab. 1 presents quantitative results comparing the various baselines (c.f. sec. 5.3) with SPOTLIGHT (c.f. sec. 5.2) on our "reference-based" dataset, where the ground truth is available (sec. 5.1). Here, all methods are conditioned on the ground truth dominant light direction using their specific light parametrization (sec. 5.3). Fig. 3 shows qualitative results for each method. We observe that SPOTLIGHT with the ZeroComp backbone outperforms the baselines on quantitative metrics and produces visually superior results.

5.5. User studies

Two user studies were conducted to assess the perceptual realism and controllability of SPOTLIGHT compared to baselines. In both cases, the Thurstone Case V Law of Comparative Judgement [66] is used to obtain a z-score for each method, where a higher value indicates higher human preference. Here, we give an overview of those user studies. Refer to the supp. for additionnal details and studies.



(a) Overall realism study

(b) Lighting control study

Figure 4. User study results, displaying the Thurstone Case V z-scores along with the 95% confidence intervals. SPOTLIGHT was preferred over baselines by users both in terms of (a) overall realism and (b) lighting control.

User study I: Overall realism Here, we use the "user-controlled" version of the dataset (see sec. 5.1) to evaluate the realism of results obtained by conditioning on a specific lighting direction, even if it is not aligned with the scene lighting conditions. We compare SPOTLIGHT against the three baselines with the best PSNR from tab. 1, namely Neural Gaffer [30], IC-Light [81], and ZeroComp [83]+SDEdit [47]. Since no ground truth is available, we design a two-alternative forced choice user study, where results obtained by two methods are shown side by side. The participants are asked to "Click on the image that looks the most realistic, by considering both the appearance of the object and its shadows." For each pair of methods, a set of 20 random images are chosen (we ensure images where shadows are clearly visible are selected), resulting in 120 pairs of images observed by each participant. We randomly shuffle the 120 pairs for each user, and the left-right order. Three sentinel images were added to discard observers that did not understand the task. Examples are shown in fig. 3.

User study I was completed by $N = 35$ observers. Fig. 4a shows that our method is the favorite, indicating that SPOTLIGHT achieves the highest realism, with a statistically significant difference to Neural Gaffer and IC-Light.

User study II: lighting control Since the goal of our method is to give full artistic control over the object relighting, we design a second two-alternative forced choice user study to evaluate lighting controllability separately from the realism. The users were shown two videos side by side and were asked to "Select the video where the light direction has the most impact on the lighting of the object, by considering both the appearance of the object and its shadows.". The videos show a virtual object relit by a rotating light source (see supp.). Here, we compared the two methods with the highest realism according to user study I: SPOTLIGHT and ZeroComp+SDEdit. Since this task aims to evaluate the local relighting within the object mask only, we replace the shadowed background with the output from our method.

Method	PSNR$_\uparrow$	SSIM$_\uparrow$	RMSE$_\downarrow$	MAE$_\downarrow$	LPIPS$_\downarrow$
$\gamma = 3, \beta = 0.05$ (Ours)	30.68	0.974	0.033	0.012	0.030
$\gamma = 1$ (no guidance)	31.69	0.976	0.029	0.011	0.029
$\gamma = 7$	28.68	0.966	0.043	0.015	0.036
$\beta = 0$	30.81	0.974	0.032	0.012	0.029
$\beta = 0.2$	29.24	0.969	0.039	0.014	0.034

Table 2. Impact of parameter selection on quantitative metrics. We observe that using no guidance ($\gamma = 1$), may provide better quantitative results. However, we also observe that these changes diminish the level of light control over the object. Our selected parameter combination provides good quantitative performance and adequate lighting control.

User study II was completed by $N = 8$ observers. As seen in fig. 4b, SPOTLIGHT achieves a statistically significant improvement in performance over ZeroComp+SDEdit, despite the low number of participants, which we attribute to the use of our local guidance strategy. Please see the supp. for more user study results and experimental details.

5.6. Effect of parameter selection

Our method uses two tunable parameters: the local guidance scale (γ in eq. (2)) and the shadow blending weight (β in eq. (1)). Tab. 2 reports metrics obtained on the "reference-based" dataset where the ground truth is available (c.f. sec. 5.1). Although using no guidance ($\gamma = 1$) results in better quantitative performance, we observe qualitatively in fig. 5 that this significantly reduces the visibility of generated shading on the inserted object, thereby reducing the impact of the desired local lighting control. Employing no shadow blending ($\beta = 0$) results in similar metric values but makes shadows much less visible. Please refer to the supp. for a qualitative comparison for the γ, β and the shadow softness (light radius) parameters.

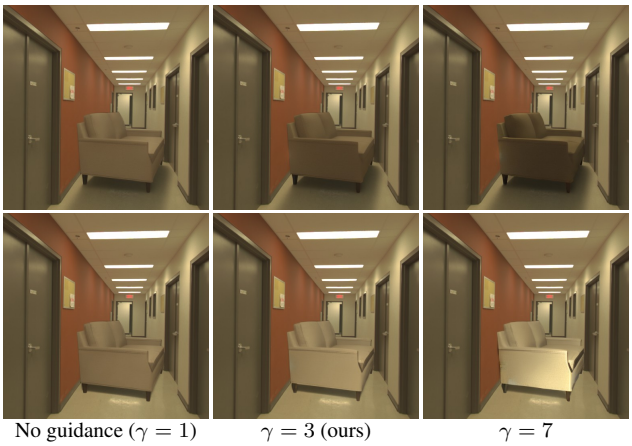
6. Applications

This section highlights additional capabilities derived from our framework to further demonstrate its versatility.

Scribbles as shadows. Instead of relying on shadow mapping in 3D, one can simply *draw* the desired shadow! Fig. 6 shows that SPOTLIGHT can realistically relight the object and refine its shadow even when it is a user-drawn scribble.

2D objects. SPOTLIGHT can be applied to real 2D objects from existing photographs by estimating their intrinsics, as in sec. 5.2, and generating desired shadows with shadow generation methods such as [65], as illustrated in fig. 7.

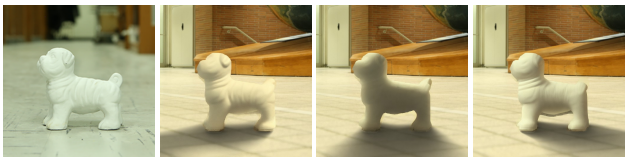
Reflective materials. When using a ZeroComp backbone pre-trained on the InteriorVerse dataset [85], SPOTLIGHT naturally generalizes to metallic objects and generates controllable specular reflections on the object, see fig. 8.



No guidance ($\gamma = 1$) $\gamma = 3$ (ours) $\gamma = 7$

Figure 5. Effect of varying the local guidance scale γ, on back (top) and front (bottom) lights. A low γ value (left) results in minimal variation in the relighting effect, whereas a high value (right) over-amplifies the reshading from the virtual light source. We empirically use $\gamma = 3$ as it offers the best balance between realism and control for indoor scene settings.

Figure 6. Rough sketches of shadows also provide a powerful lighting cue to SPOTLIGHT. By scribbling a shadow next to the object region (top), SPOTLIGHT is able to realistically relight the object and refine the shadow (bottom).



Input object Relit object on target background

Figure 7. SPOTLIGHT can also be applied to objects from 2D images using estimated intrinsics and shadows generated using [65].

Additional relighting results. SPOTLIGHT also shows other capabilities: relighting for more diverse objects and outdoor scenes, simulating multiple light sources by combining outputs with different guiding shadows, and full-image relighting. Due to space constraints, the implementation details and results are provided in the supp.

Figure 8. When using a ZeroComp backbone trained on InteriorVerse [85], SPOTLIGHT generates reflections on specular objects.

Incorrect user-defined shadow Correct shadow

Figure 9. User-specified (possibly inconsistent) shadows. Even when the input shadow contradicts the scene lighting (left), SPOTLIGHT produces a visually plausible composition; the user may adjust shadow direction (right).

7. Discussion

Limitation: physically incorrect shadows. Since SPOTLIGHT imposes no constraints on the input shadow, users may specify a shadow direction that potentially contradicts the real lighting in the background. In this case, SPOTLIGHT still attempts to render a visually plausible composition consistent with the given shadow, as shown in fig. 9.

We present SPOTLIGHT, a method for realistically controlling the local lighting of an object through a coarse shadow, compatible with diffusion-based intrinsic image renderers without any additional training. Our results demonstrate that fine-grained control over local lighting can be achieved while attaining realistic compositions.

While our method can realistically generate relighting results from a shadow and despite showing that this can be done in multiple ways (shadow mapping, 2D generation, hand-drawn scribbles), the process of generating a shadow from scratch could prove challenging to some. A promising future direction is to enable end-to-end shadow control using parametric light models, such as point or sphere lights. Additionally, similar approaches could be explored to enable global lighting adjustments for entire images.

Acknowledgements. This research was supported by FRQNT scholarship 328810, NSERC grants RGPIN 2020-04799 and ALLRP 586543-23, Mitacs and Depix. Computing resources were provided by the Digital Research Alliance of Canada. The authors thank Zheng Zeng, Yannick Hold-Geoffroy and Justine Giroux for their help as well as all lab members for discussions and proofreading help.

References

[1] Attila T. Áfra. Intel® Open Image Denoise, 2024. https://www.openimagedenoise.org. 6

[2] Anonymous. Intrinsic-controlnet : A generative rendering approach to render any real and unreal. In *Submitted to Int. Conf. Learn. Represent.*, 2024. under review. 3

[3] Omri Avrahami, Dani Lischinski, and Ohad Fried. Blended diffusion for text-driven editing of natural images. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022. 2

[4] Omri Avrahami, Ohad Fried, and Dani Lischinski. Blended latent diffusion. *ACM Trans. Graph.*, 42(4), 2023. 2, 3

[5] Gwangbin Bae and Andrew J Davison. Rethinking inductive biases for surface normal estimation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2024. 3

[6] Gwangbin Bae, Ignas Budvytis, and Roberto Cipolla. Estimating and exploiting the aleatoric uncertainty in surface normal estimation. In *Int. Conf. Comput. Vis.*, pages 13137–13146, 2021. 3

[7] Harry Barrow, J Tenenbaum, A Hanson, and E Riseman. Recovering intrinsic scene characteristics. *Comput. vis. syst*, 2 (3-26):2, 1978. 3

[8] Shariq Farooq Bhat, Reiner Birkl, Diana Wofk, Peter Wonka, and Matthias Müller. Zoedepth: Zero-shot transfer by combining relative and metric depth. *arXiv preprint arXiv:2302.12288*, 2023. 3, 6

[9] Anand Bhattad and David A Forsyth. Cut-and-paste object insertion by enabling deep image prior for reshading. In *Int. Conf. 3D Vis.*, 2022. 2

[10] Anand Bhattad, Daniel McKee, Derek Hoiem, and David Forsyth. Stylegan knows normal, depth, albedo, and more. *Adv. Neural Inform. Process. Syst.*, 36, 2024. 3

[11] Anand Bhattad, James Soole, and DA Forsyth. Stylitgan: Image-based relighting via latent control. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2024. 2

[12] Tim Brooks, Aleksander Holynski, and Alexei A Efros. Instructpix2pix: Learning to follow image editing instructions. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2023. 2

[13] Chris Careaga and Yağız Aksoy. Physically controllable relighting of photographs. In *Proc. SIGGRAPH*, 2025. 2

[14] Chris Careaga, S Mahdi H Miangoleh, and Yağız Aksoy. Intrinsic harmonization for illumination-aware compositing. In *ACM SIGGRAPH Asia Conf.*, 2023. 2, 3

[15] Jun Myeong Choi, Annie Wang, Pieter Peers, Anand Bhattad, and Roni Sengupta. Scribblelight: Single image indoor relighting with scribbles. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2025. 2

[16] Mohammad Reza Karimi Dastjerdi, Jonathan Eisenmann, Yannick Hold-Geoffroy, and Jean-François Lalonde. Everlight: Indoor-outdoor editable HDR lighting estimation. In *Int. Conf. Comput. Vis.*, 2023. 2

[17] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In *Adv. Neural Inform. Process. Syst.*, 2021. 2

[18] Xiaodan Du, Nicholas Kolkin, Greg Shakhnarovich, and Anand Bhattad. Generative models: What do they know? do they know things? let's find out! *arXiv preprint arXiv:2311.17137*, 2023. 3

[19] Ainaz Eftekhar, Alexander Sax, Jitendra Malik, and Amir Zamir. Omnidata: A scalable pipeline for making multi-task mid-level vision datasets from 3d scans. In *Int. Conf. Comput. Vis.*, pages 10786–10796, 2021. 3

[20] Elmar Eisemann, Michael Schwarz, Ulf Assarsson, and Michael Wimmer. *Real-Time Shadows*. A. K. Peters, Ltd., 2011. 2, 3, 4

[21] Dave Epstein, Allan Jabri, Ben Poole, Alexei Efros, and Aleksander Holynski. Diffusion self-guidance for controllable image generation. In *Adv. Neural Inform. Process. Syst.*, 2023. 2

[22] Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H Bermano, Gal Chechik, and Daniel Cohen-Or. An image is worth one word: Personalizing text-to-image generation using textual inversion. *arXiv preprint arXiv:2208.01618*, 2022. 2

[23] Marc-André Gardner, Kalyan Sunkavalli, Ersin Yumer, Xiaohui Shen, Emiliano Gambaretto, Christian Gagné, and Jean-François Lalonde. Learning to predict indoor illumination from a single image. *ACM Trans. Graph.*, 9(4), 2017. 2

[24] Marc-André Gardner, Yannick Hold-Geoffroy, Kalyan Sunkavalli, Christian Gagné, and Jean-François Lalonde. Deep parametric indoor lighting estimation. In *Int. Conf. Comput. Vis.*, 2019. 2

[25] Mathieu Garon, Kalyan Sunkavalli, Sunil Hadap, Nathan Carr, and Jean-François Lalonde. Fast spatially-varying indoor lighting estimation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019. 2

[26] David Griffiths, Tobias Ritschel, and Julien Philip. Outcast: Single image relighting with cast shadows. *Comput. Graph. Forum*, 43, 2022. 2

[27] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021. 2

[28] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Adv. Neural Inform. Process. Syst.*, 2020. 2

[29] Yan Hong, Li Niu, and Jianfu Zhang. Shadow generation for composite image in real-world scenes. In *Assoc. Adv. of Art. Int.*, 2022. 2

[30] 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 *Adv. Neural Inform. Process. Syst.*, 2024. 2, 5, 6, 7

[31] Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, and Konrad Schindler. Repurposing diffusion-based image generators for monocular depth estimation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2024. 3

[32] Peter Kocsis, Julien Philip, Kalyan Sunkavalli, Matthias Nießner, and Yannick Hold-Geoffroy. Lightit: Illumination modeling and control for diffusion models. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2024. 2

[33] Peter Kocsis, Vincent Sitzmann, and Matthias Nießner. Intrinsic image diffusion for single-view material estimation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2024. 6

[34] Balazs Kovacs, Sean Bell, Noah Snavely, and Kavita Bala. Shading annotations in the wild. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017. 3

[35] Jean-Francois Lalonde and Alexei A Efros. Using color compatibility for assessing image realism. In *Int. Conf. Comput. Vis.*, 2007. 2

[36] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *Int. Conf. Mach. Learn.*, 2023. 6

[37] 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 *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020. 2, 3

[38] Zhengqin Li, Ting-Wei Yu, Shen Sang, Sarah Wang, Meng Song, Yuhan Liu, Yu-Ying Yeh, Rui Zhu, Nitesh Gundavarapu, Jia Shi, et al. Openrooms: An end-to-end open framework for photorealistic indoor scene datasets. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2021. 6

[39] Ruofan Liang, Zan Gojcic, Merlin Nimier-David, David Acuna, Nandita Vijaykumar, Sanja Fidler, and Zian Wang. Photorealistic object insertion with diffusion-guided inverse rendering. *ECCV*, 2024. 2

[40] Ruofan Liang, Zan Gojcic, Huan Ling, Jacob Munkberg, Jon Hasselgren, Zhi-Hao Lin, Jun Gao, Alexander Keller, Nandita Vijaykumar, Sanja Fidler, et al. Diffusionrenderer: Neural inverse and forward rendering with video diffusion models. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2025. 1, 2

[41] Zhi-Hao Lin, Bohan Liu, Yi-Ting Chen, Kuan-Sheng Chen, David Forsyth, Jia-Bin Huang, Anand Bhattad, and Shenlong Wang. Urbanir: Large-scale urban scene inverse rendering from a single video. In *Int. Conf. 3D Vis.*, 2025. 2

[42] Daquan Liu, Chengjiang Long, Hongpan Zhang, Hanning Yu, Xinzhi Dong, and Chunxia Xiao. ARShadowGAN: Shadow generative adversarial network for augmented reality in single light scenes. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020. 2

[43] Qingyang Liu, Junqi You, Jianting Wang, Xinhao Tao, Bo Zhang, and Li Niu. Shadow generation for composite image using diffusion model. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2024. 2

[44] Jundan Luo, Duygu Ceylan, Jae Shin Yoon, Nanxuan Zhao, Julien Philip, Anna Frühstück, Wenbin Li, Christian Richardt, and Tuanfeng Wang. Intrinsicdiffusion: joint intrinsic layers from latent diffusion models. In *ACM SIGGRAPH Conf.*, 2024. 2, 3

[45] Linjie Lyu, Valentin Deschaintre, Yannick Hold-Geoffroy, Miloš Hašan, Jae Shin Yoon, Thomas Leimkühler, Christian Theobalt, and Iliyan Georgiev. Intrinsicedit: Precise generative image manipulation in intrinsic space. *arXiv preprint arXiv:2505.08889*, 2025. 2

[46] 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. *arXiv preprint arXiv:2505.09608*, 2025. 2

[47] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. SDEdit: Guided image synthesis and editing with stochastic differential equations. In *Int. Conf. Learn. Represent.*, 2022. 2, 6, 7

[48] Lukas Murmann, Michael Gharbi, Miika Aittala, and Fredo Durand. A multi-illumination dataset of indoor object appearance. In *Int. Conf. Comput. Vis.*, 2019. 2

[49] 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, 2019. 3

[50] Pakkapon Phongthawee, Worameth Chinchuthakun, Nontaphat Sinsunthithet, Varun Jampani, Amit Raj, Pramook Khungurn, and Supasorn Suwajanakorn. Diffusionlight: Light probes for free by painting a chrome ball. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2024. 2

[51] Francois Pitie, Anil C Kokaram, and Rozenn Dahyot. N-dimensional probability density function transfer and its application to color transfer. In *Int. Conf. Comput. Vis.*, 2005. 2

[52] Yohan Poirier-Ginter, Alban Gauthier, Julien Phillip, J-F Lalonde, and George Drettakis. A diffusion approach to radiance field relighting using multi-illumination synthesis. In *Comput. Graph. Forum*, page e15147. Wiley Online Library, 2024. 2

[53] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1 (2):3, 2022. 2

[54] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44 (3):1623–1637, 2020. 3

[55] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. In *Int. Conf. Comput. Vis.*, 2021. 3

[56] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *IEEE Comp. Graph. Appl.*, 21(5):34–41, 2001. 2

[57] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022. 1, 2

[58] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH Conf.*, pages 1–10, 2022. 2

[59] Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. In *International Conference on Learning Representations*, 2022. 3

[60] Yichen Sheng, Jianming Zhang, and Bedrich Benes. Ssn: Soft shadow network for image compositing. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2021. 2

[61] Yichen Sheng, Yifan Liu, Jianming Zhang, Wei Yin, A Cengiz Oztireli, He Zhang, Zhe Lin, Eli Shechtman, and Bedrich Benes. Controllable shadow generation using pixel height maps. In *Eur. Conf. Comput. Vis.*, 2022. 2

[62] Yichen Sheng, Jianming Zhang, Julien Philip, Yannick Hold-Geoffroy, Xin Sun, He Zhang, Lu Ling, and Bedrich Benes. Pixht-lab: Pixel height based light effect generation for image compositing. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2023. 2

[63] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Int. Conf. Mach. Learn.*, 2015. 2

[64] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *Int. Conf. Learn. Represent.*, 2021. 2, 3

[65] Onur Tasar, Clément Chadebec, and Benjamin Aubin. Controllable shadow generation with single-step diffusion models from synthetic data. *arXiv preprint arXiv:2412.11972*, 2024. 3, 7, 8

[66] Louis L Thurstone. A law of comparative judgment. In *Scaling*, pages 81–92. Routledge, 1927. 6

[67] Yi-Hsuan Tsai, Xiaohui Shen, Zhe Lin, Kalyan Sunkavalli, Xin Lu, and Ming-Hsuan Yang. Deep image harmonization. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017. 2

[68] Lucas Valença, Jinsong Zhang, Michaël Gharbi, Yannick Hold-Geoffroy, and Jean-François Lalonde. Shadow harmonization for realistic compositing. In *ACM SIGGRAPH Asia Conf.*, 2023. 2

[69] Guangcong Wang, Yinuo Yang, Chen Change Loy, and Ziwei Liu. Stylelight: HDR panorama generation for lighting estimation and editing. In *Eur. Conf. Comput. Vis.*, 2022. 2

[70] Henrique Weber, Mathieu Garon, and Jean-François Lalonde. Editable indoor lighting estimation. In *Eur. Conf. Comput. Vis.*, 2022. 2

[71] Jiaye Wu, Sanjoy Chowdhury, Hariharmano Shanmugaraja, David Jacobs, and Soumyadip Sengupta. Measured albedo in the wild: Filling the gap in intrinsics evaluation. In *Int. Conf. Comput. Photo.*, 2023. 3

[72] Xiaoyan Xing, Vincent Tao Hu, Jan Hendrik Metzen, Konrad Groh, Sezer Karaoglu, and Theo Gevers. Retinex-diffusion: On controlling illumination conditions in diffusion models via retinex theory. *arXiv preprint arXiv:2407.20785*, 2024. 2

[73] Xiaoyan Xing, Konrad Groh, Sezer Karagolu, Theo Gevers, and Anand Bhattad. Luminet: Latent intrinsics meets diffusion models for indoor scene relighting. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2025. 2

[74] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2024. 3

[75] Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything v2. *arXiv:2406.09414*, 2024. 3, 4

[76] Chongjie Ye, Lingteng Qiu, Xiaodong Gu, Qi Zuo, Yushuang Wu, Zilong Dong, Liefeng Bo, Yuliang Xiu, and Xiaoguang Han. Stablenormal: Reducing diffusion variance for stable and sharp normal. *arXiv preprint arXiv:2406.16864*, 2024. 3, 6

[77] Hu Ye, Jun Zhang, Sibo Liu, Xiao Han, and Wei Yang. Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arXiv:2308.06721*, 2023. 2

[78] Wei Yin, Chi Zhang, Hao Chen, Zhipeng Cai, Gang Yu, Kaixuan Wang, Xiaozhi Chen, and Chunhua Shen. Metric3D: Towards zero-shot metric 3D prediction from a single image. In *Int. Conf. Comput. Vis.*, 2023. 3

[79] Chong Zeng, Yue Dong, Pieter Peers, Youkang Kong, Hongzhi Wu, and Xin Tong. DiLightNet: Fine-grained lighting control for diffusion-based image generation. In *ACM SIGGRAPH Conf.*, 2024. 2, 5, 6

[80] 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 *ACM SIGGRAPH Conf.*, 2024. 2, 3, 6

[81] 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 *Int. Conf. Learn. Represent.*, 2025. 1, 2, 5, 6, 7

[82] Shuyang Zhang, Runze Liang, and Miao Wang. Shadowgan: Shadow synthesis for virtual objects with conditional adversarial networks. *Comp. Vis. Media*, 5:105–115, 2019. 2

[83] Zitian Zhang, Frédéric Fortier-Chouinard, Mathieu Garon, Anand Bhattad, and Jean-François Lalonde. Zerocomp: Zero-shot object compositing from image intrinsics via diffusion. In *Winter Conf. App. Comput. Vis.*, 2025. 1, 2, 3, 4, 6, 7

[84] Xiaoming Zhao, Pratul P. Srinivasan, Dor Verbin, Keunhong Park, Ricardo Martin Brualla, and Philipp Henzler. IllumiNeRF: 3D Relighting Without Inverse Rendering. In *NeurIPS*, 2024. 2

[85] Jingsen Zhu, Fujun Luan, Yuchi Huo, Zihao Lin, Zhihua Zhong, Dianbing Xi, Rui Wang, Hujun Bao, Jiaxiang Zheng, and Rui Tang. Learning-based inverse rendering of complex indoor scenes with differentiable monte carlo raytracing. In *ACM SIGGRAPH Asia Conf.*, 2022. 7, 8

[86] Rui Zhu, Zhengqin Li, Janarbek Matai, Fatih Porikli, and Manmohan Chandraker. Irisformer: Dense vision transformers for single-image inverse rendering in indoor scenes. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022. 2, 3