EUROGRAPHICS 2026 / B. Masia and J. Thies
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 45 (2026), Number 2

ProjectiveShading: Inserting 3D Objects into Indoor Images with Complex Shadows

J. Luo[†1], X. Wu[2], N. Zhao[3], L. Wang[2], W. Li[1] and C. Richardt[4]

[1]University of Bath, United Kingdom
[2]Shandong University, China
[3]Adobe Research, United States
[4]Meta Reality Labs, Switzerland

Figure 1: *ProjectiveShading is a fully automatic approach for realistically inserting virtual objects into indoor images (left). It handles bi-directional shadow interactions between the real scene and the virtual object, and blends virtual shadows with complex textures, significantly outperforming previous methods. Compared to rendering with a single environment map (center), our approach (right) casts realistic real shadows and highlights onto the virtual object and synthesizes plausible virtual shadows. (Image source: Tanya Prodaan, Unsplash. Model source: Stanford Computer Graphics Laboratory.)*

Abstract

Realistically inserting virtual 3D objects into real-world images requires perceptually coherent shadowing of the object and background scene. Achieving this in single-view indoor scenes with sunlight is challenging due to complex, partially visible occluders and indirect lighting. Environment maps alone cannot produce realistic shadows on virtual objects, and any representation (scene parameters) used for rendering must be practically estimable. We introduce ProjectiveShading, the first automatic method for inverse- and re-rendering that handles bi-directional shadow interactions for realistic object composition. Our key innovation is the sunlight map, a 2D image encoding direct sunlight and arbitrary occlusions. It is generated from single-view estimations using off-the-shelf models and is compatible with standard rendering engines. We also propose algorithms to estimate sunlight direction and to blend virtual and real shadows while preserving background textures. Experiments on synthetic and in-the-wild images show our method outperforms previous approaches.

CCS Concepts
• ***Computing methodologies*** → ***Visibility; Image-based rendering;*** *Scene understanding; Mixed / augmented reality;*

1. Introduction

The photorealistic integration of virtual objects into real-world scenes is a long-standing challenge in visual computing, with broad applications in visual effects and mixed reality [KBG*15, NCL*21]. Shadows are critical to visual realism and coherence, as

† Corresponding author: jundanluo22@gmail.com

they provide geometric and photometric cues about spatial relationships and lighting in an image [MKK98, KOF13, Mam04, KKK06]. Our approach aims to automate hard-shadow synthesis in single-view virtual object composition: casting real shadows onto virtual objects and projecting virtual shadows into real environments.

The challenges arise from both rendering and inverse rendering. From the rendering side, shadow synthesis relies on lighting estimation. While recent deep models can estimate spatially varying lighting [GSH*19, LSR*20, WYLL22, BHY*23, CLP*23], simplifying the rendering process to a single global environment map [DM97] prevents the reproduction of expected shadows on virtual surfaces. This assumption of distant lighting ignores near-field illumination variation [RH01, LYX*24], e.g., due to occlusions, as demonstrated in Figure 1. Valença et al. [VZG*23] adopt a simplistic indirect lighting model that considers only sky illumination, making it unsuitable for indoor scenes. Besides, their method fails to handle non-Lambertian objects and secondary sunlight bounces even for Lambertian ones. From the inverse-rendering side, any representation (scene parameters) used for rendering must be practically estimable. Although explicit occluder geometry is required in the conventional approach to render shadows, reconstructing it is difficult with partial or missing observations in single-view images. Recent deep models [YGH*23, VZG*23] rely on direct shadow detection to represent occlusion, but their generalisation to indoor scenes is severely limited by the lack of indoor annotations.

To overcome these challenges, we reformulate the problem of shadowing virtual objects as the projection of a 2D shadow–illumination image. This formulation is motivated by a key geometric principle: shadows maintain a consistent shape in the sun's perspective (Figure 2), which enables shadow rendering to be treated as an image projection. It is further supported by intrinsic image decomposition theory [GRPCLM22], which shows that estimated shading naturally encodes both illumination and shadow information in general scenes. The key component of ProjectiveShading is the *sunlight map*, a novel high-resolution 2D representation that encodes spatial lighting and occlusion information in a lightweight way. This specialised illumination map acts as a projective textured light, ensuring compatibility with standard rendering engines like Blender [Ble25] without requiring custom shaders for various materials. In inverse rendering, we propose algorithms to recover the sunlight map in the sun's view from camera-view estimations using off-the-shelf models, including the essential step of estimating sunlight direction. As inaccurate estimation of indirect lighting often results in inconsistencies between rendered virtual shadows and background shadows, we further propose a shadow refinement algorithm that leverages learned image priors from a pre-trained generative diffusion model to refine both shadow intensity and the underlying albedo texture.

In summary, ProjectiveShading is the first automatic hard-shadow-aware composition approach for single-view indoor images. Our approach addresses both real and virtual hard shadows:

- Modelling occluded sunlight as a projectable sunlight map, enabling shadowing of virtual objects with proper light bounces as well as estimation from single views;
- A sunlight direction estimation algorithm exploiting geometric relationships between potentially unseen occluders, shadows, and light sources to improve the estimated sunlight map;
- A novel shadow refinement method using a pretrained diffusion model for seamlessly blending virtual shadows with real ones without fine-tuning the model; and
- Comprehensive validation on our proposed synthetic benchmark and real photos, demonstrating our method's state-of-the-art performance in achieving photorealistic rendering with shadows and overall scene realism.

Source code is released at
https://github.com/JundanLuo/ProjectiveShading/.

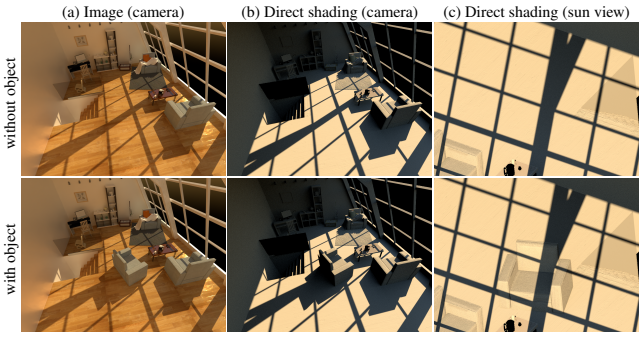

Figure 2: *Shadows from the camera and sun's perspectives. The image is shown in the camera's view (a), and direct shading of sunlight as seen from (b) the camera and (c) sun. Shadows appear deformed in views (a) and (b), but retain their shape in the sun's perspective (c). This shadow deformation can be resolved by orthographically projecting the shadow from the sun's perspective onto the virtual object. (Image source: our LightSOC dataset.)*

2. Related work

Lighting estimation. Lighting estimation is essential for virtual object composition. A thorough survey is presented by [KBG*15]. Some previous methods [UWH*03, Deb98] use mirror spheres to capture panoramic lighting, but these require specialised setups and are unsuitable for the single-view in-the-wild images we target. We focus on single-view lighting estimation. Previous works estimate a single environment map for the entire image [GSY*17, SK20, LMF*19, PCS*24, ZZY*21, LTH*23, YS19, LGND*24, SBXX25, WTC*25], while others estimate spatially-varying environment maps to capture localised lighting changes [GSH*19, SF19, LYS*21, LSR*20, TZW*22, WYLL22, GHGS*19, BHY*23, ZZLS21, ZLM*22]. Voxelised lighting representations [WPFK21, WXL24, ZLH*22], editable parametric representations [WGL22] and hybrid lighting models [KSH*14] have also been proposed. Despite progress in indoor lighting estimation, real shadow projection is often overlooked in virtual object rendering. This is the focus of our work.

Existing light source estimation methods have limitations, such as the requirement for user interaction [ZYL*17], segmentation of light sources [LSB*22], assumptions of point lights [JRM20], the need for panoramic input [ZCB*22], assumptions about constrained environments [SSI03], or designs specifically for outdoor

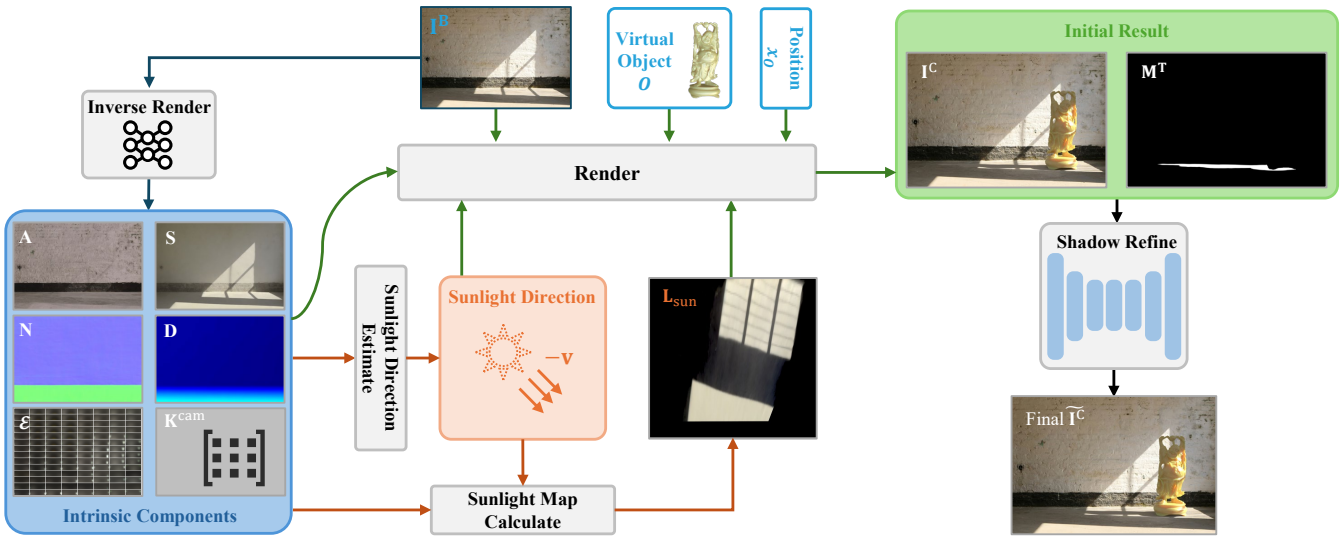

Figure 3: *Our proposed ProjectiveShading approach: an automatic inverse- and re-rendering approach. Given a single-view input image, a 3D object, and a user-specified placement, we estimate intrinsic components via inverse rendering, followed by estimation of the sunlight direction and sunlight map. The virtual object is then rendered into the reconstructed scene and composited onto the input image. Finally, the rendered virtual shadow is refined using our shadow refinement algorithm. (Image source: Utsman Media, Unsplash. Model source: Stanford Computer Graphics Laboratory.)*

scenes [HGSH*16, HGAL19, LEN12, CCC*18, YGH*21]. By contrast, our proposed sunlight estimation algorithm applies to general indoor single-view images without requiring additional inputs.

Shadow synthesis. Previous methods for virtual object shadowing have limited applicability. [CGC*03] only transfer pre-existing shadows. Some methods [XZP*13, KHFH11] require extensive human input for geometry and light-source estimation. Karsch et al. [KHFH11] model shadows via light shafts, which cannot handle irregular shapes. Other approaches [XZP*13, VZG*23] treat shadow casting as texture mapping, failing to handle light bounces accurately (see Figure 8). Yu et al. [YGH*23] propose an end-to-end deep model for warping shadows, but it heavily depends on training data and does not generalise to indoor scenes. Duchêne et al. [DRC*15] reconstruct unseen occluders to edit sunlight direction, but require multi-view input. Deep models for synthesising shadows of foreground crops [LLZ*20, LYW*24, WLT*21] are limited by the available training data. SpotLight [FCZM*26] requires user-defined uniform darkness and cannot handle overlaps between real and virtual shadows.

Intrinsic image decomposition. Intrinsic image decomposition [BBS14, LS18a, LS18b, LHL*20, LZLR24, LCY*24, CA23, ZDG*24, CA24, CPY*25, KSN23, JLYT23], surveyed by Garces et al. [GRPCLM22], estimates albedo, shading and geometry or other physical properties from images. Careaga et al. [CMA23] and Guo et al. [GZJ*21] utilise intrinsic images to harmonise or composite foreground image crops, enforcing physical consistency. Some methods [ZFCG*25, ZDG*24, LDHG*25, LGL*25, FCZM*26] condition diffusion models on intrinsic images of virtual objects to generate composition results. We construct our virtual scenes and sunlight maps based on intrinsic decomposition.

3. Projective shading methods

Our method is an automatic inverse- and re-rendering approach that requires only a single-view input image, a 3D object, and a user-specified 2D image coordinate for object placement (which determines the 3D proxy plane). The full pipeline is illustrated in Figure 3. A notation table is provided in the *supplement*. We first describe scene reconstruction (Section 3.1), including sunlight direction estimation (Section 3.1.2). In Section 3.2, we introduce our sunlight map, a projective textured light source representing occluded sunlight. Finally, we present shadow refinement (Section 3.3). We refer to components in the original input image as "real" or "background", and newly added elements as "virtual" or "foreground".

3.1. Inverse rendering with sunlight direction estimation

3.1.1. Scene representation

Given an input background image $\mathbf{I}^{\mathrm{B}}$ of a Lambertian scene and a user-provided 2D pixel location $\mathbf{x}_{\mathrm{o}}$ for object placement, we construct a 3D scene (Figure 4) with:

- A proxy plane $\mathcal{G}$ that receives shadows, modelled with normal $\mathbf{N}_{\mathrm{o}}$, distance ρ_{o} from the camera and albedo $\mathbf{A}^{\mathcal{G}}$,
- A virtual object O,
- A camera with intrinsic matrix $\mathbf{K}^{\mathrm{cam}}$, and
- Lighting modelled as the combination of a sunlight map $\mathcal{X}$ (Section 3.2.2) and an environment map $\mathcal{E}_{\mathrm{o}}$.

While renderers can shadow arbitrary objects or meshes, we adopt a proxy plane for simplicity. This proxy $\mathcal{G}$, defined by the normal and depth at the user-specified 2D object location $\mathbf{x}_{\mathrm{o}}$, allows arbitrary

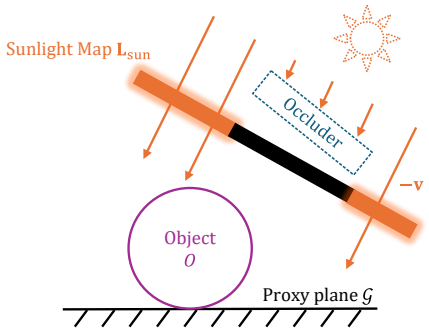

Figure 4: *Illustration of sunlight map placement. The dashed sun and occluder represent elements replaced by an orthographic area light projecting the sunlight map along the sunlight direction* $-\mathbf{v}$*. The sunlight map, a 2D image acting as a projector, encodes sunlight as bright regions and occlusion as zero intensity. It is placed perpendicular to the light direction, between the occluder and the shadow-receiving proxy plane.*

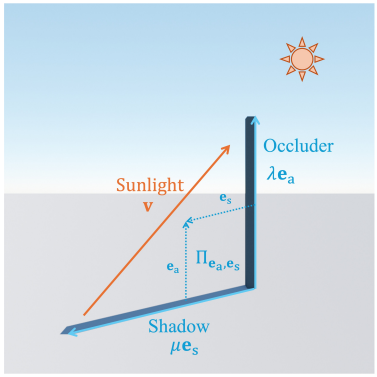

Figure 5: *The sunlight direction* $\mathbf{v}$ *lies in the plane* Π *spanned by the directions of the occluder* $\mathbf{e}_a$ *and its shadow* $\mathbf{e}_s$*.*

orientation. The environment map $\mathcal{E}_o$ represents indirect illumination and contributions from secondary light sources at $\mathbf{x}_o$.

In general, our scenes are reconstructed from standard intrinsic components in inverse rendering: depth $\mathbf{D}$, surface normals $\mathbf{N}$, albedo $\mathbf{A}$, shading $\mathbf{S}$, environment lighting $\mathcal{E}_o$, sunlight direction $\mathbf{v}$ (Section 3.1.2), and optionally the camera's intrinsic matrix $\mathbf{K}^{\text{cam}}$. Most of these are obtained using off-the-shelf models. We use MoGe [WXD*25] for depth, normals, and the intrinsic matrix; IntrinsicDiffusion [LCY*24] for albedo and shading; and Li et al. [LSR*20] for environment lighting estimation. Our contributions on estimating the sunlight direction and the sunlight maps are described next.

3.1.2. Sunlight direction estimation

The direction of sunlight is crucial for computing the sunlight map. We propose an algorithm to automatically estimate and refine the sunlight direction in general indoor scenes. It first coarsely estimates the direction from an environment map and then refines it based on the geometric relationship between the sunlight, occluder and shadows. This refinement algorithm is geometry-based rather than learning-based due to the lack of sufficient indoor sunlight training data. To the best of our knowledge, no publicly available indoor datasets provide sunlight annotations. Existing models do not support the estimation of indoor sunlight from a single image without additional inputs, such as user interactions [ZYL*17] or light source segmentation [LSB*22].

Initial sunlight direction estimation. Our approach assumes sunlight as the primary light source and thresholds the shading intensity image to distinguish sunlight and shadow regions. We aggregate the spatially-varying environment maps originally predicted by Li et al. [LSR*20] in the sunlight regions by averaging (Figure 6), and use the brightest pixel to define the initial sunlight direction $\mathbf{v}'$.

Refining the sunlight direction. As the sun is virtually infinitely far away, all sunlight rays are effectively parallel. Under parallel projection, parallel occluding edges (e.g., on a window frame) cast parallel shadow edges on the same planar surface. We propose a sunlight direction refinement algorithm that leverages this parallelism. As shown in Figure 5, consider a vertical occluder (e.g., a window frame) with direction $\mathbf{e}_a$, which casts a shadow line $\mathbf{e}_s$ on the proxy plane $\mathcal{G}$ due to the sunlight direction $\mathbf{v}$. The sunlight direction $\mathbf{v}$ lies in the plane Π spanned by $\mathbf{e}_a$ and $\mathbf{e}_s$ (see *supplement* for verification). Any sunlight direction within this plane Π casts a shadow parallel to $\mathbf{e}_s$ from the occluder $\mathbf{e}_a$. To ensure consistent shadow orientation between parallel real and virtual occluding edges, we project the initial estimate $\mathbf{v}'$ onto the plane Π to obtain the refined sunlight direction $\mathbf{v}$.

Hard shadows in indoor scenes are typically cast by occluding edges aligned with the surface normal of the shadow-receiving plane. For instance, shadows on floors are commonly cast by window frames or furniture resting on the floor. We use normal-aligned edges as anchor occluders to refine the coarse sunlight direction. Based on this, we search for shadow edges $\mathbf{e}_s$ in the background image and assume the corresponding occluder direction $\mathbf{e}_a$ using the average surface normal of the shadow edge's endpoints. We use a RANSAC-like algorithm [FB81] to exclude outliers that violate the occluder direction assumption, and project the initial sunlight direction onto the planes Π spanned by the inliers (see *supplement* for details). In practice, we interpolate the refined sunlight direction $\mathbf{v}$ with the proxy plane normal $\mathbf{N}_o$, limiting the deviation to a maximum of 70° to avoid infinite shadow lines when the sunlight is nearly parallel to the proxy plane $\mathcal{G}$.

The results are shown in Figure 6. By leveraging geometric relationships between shadows, occluders and receiving planes, our algorithm identifies anchored shadow edges and refines the sunlight direction, thus enhancing the realism of compositions.

3.2. Sunlight map estimation and rendering

3.2.1. Sunlight map representation

To model sunlight under occlusion, we propose the *sunlight map*. As illustrated in Figure 4 and Figure 7, the sunlight map is a 2D high-dynamic-range image that encodes where the sun's rays are

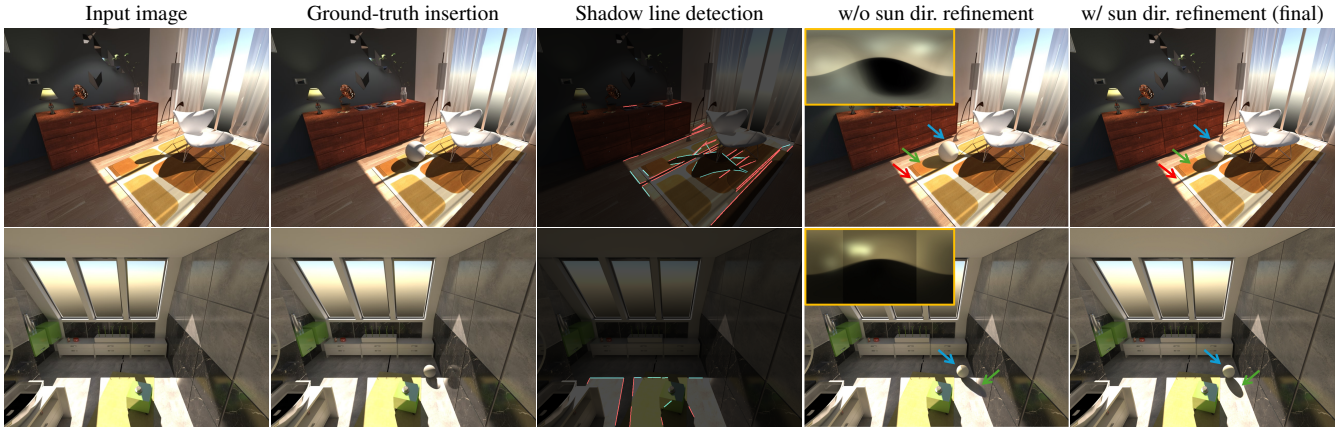

Figure 6: *Composition results with and without sunlight direction refinement. Selected shadow lines for refinement are marked in red, others in cyan (third column). The insets (fourth column) show the average environment maps [LSR*20] used for initial sunlight direction estimation (Li et al. [LSR*20] only estimate the visible hemisphere per surface pixel). Sunlight refinement improves the shape of shadows on curved surfaces (indicated by cyan arrows) and corrects shadow orientation on the floor (green arrows), even for the slanted window in the bottom row, which deviates from our surface-normal-oriented assumption. Without sunlight correction, inconsistent orientations between virtual shadows (green arrows) and real shadows (red arrows), cast by objects on the floor, are clearly visible. Table 2 shows quantitative ablation results. (Image source: our LightSOC dataset.)*

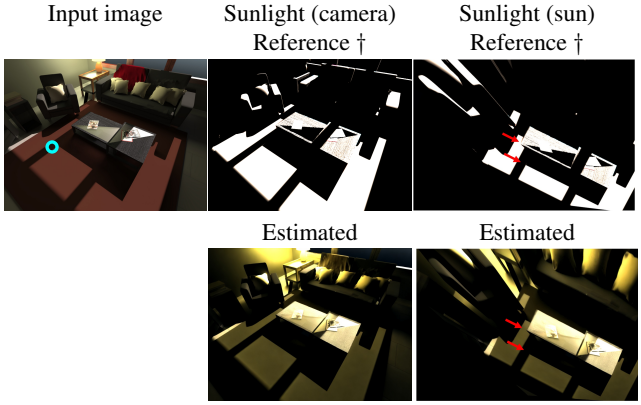

Figure 7: *Estimated sunlight map example. † Sunlight map derived from the ground-truth sun's direct shading and our proposed camera–sun view transformation, shown for reference since no true "ground truth" is accessible from the Blender renderer. The sunlight map is first computed in the camera view, then transformed to the sun's view using a planar homography induced by the proxy plane. This single-plane homography ensures that the occluder does not cover its corresponding shadow (red arrows) in the sun's view, allowing shadow projection. A user-selected location (circled in cyan) specifies the object placement and proxy plane. (Image source: our LightSOC dataset.)*

blocked or transmitted: zero values indicate occlusion, whilst bright values represent direct-light intensity. This sunlight map $\mathbf{L}_{\text{sun}}$ is positioned above the proxy plane $\mathcal{G}$ to project sunlight and shadows on virtual objects.

For rendering, the sunlight map $\mathbf{L}_{\text{sun}}$ is directly used as a light source (without conversion into other light primitives such as environment maps). It can be inserted into the reconstructed scene by texturing any planar orthographic emitter without requiring custom shaders to support various materials (Figure 8). In practice, we use Blender's standard area light with zero angular spread (i.e., orthographic projection) to emit the sunlight map. For completeness, we define sunlight as $\mathcal{X} = \{\mathbf{L}_{\text{sun}}, \mathbf{v}, \mathbf{t}\}$, where $\mathbf{v}$ is the sunlight direction (from the surface to the light) estimated in Section 3.1.2, and $\mathbf{t}$ is the translation of the sunlight map from the camera center. In principle, $\mathbf{t}$ can be set arbitrarily as long as the virtual object is illuminated; for convenience, we place it above the virtual object's center. Next, we describe how we estimate the sunlight map $\mathbf{L}_{\text{sun}}$, for a given light direction $\mathbf{v}$ and translation $\mathbf{t}$.

Comparison with shadow and texture mapping. The key difference between our method and previous shadowing algorithms is how occlusion is modelled: we treat it as zero-intensity *illumination*. Shadow mapping techniques resolve visibility by storing depth and performing depth comparisons [SWP11]. Some methods [VZG*23, XZP*13] first render sunlight without occlusion and then mask out direct shading in shadowed regions, treating occlusion as a zero-intensity diffuse texture. This design cannot render non-Lambertian materials, as shown in Figure 8. Even for Lambertian surfaces, indirect sunlight is exaggerated because secondary sunlight bounces are not tracked after rendering and thus remain untouched. By contrast, our approach enables more visually plausible and natural shadow rendering across different material types.

3.2.2. Sunlight map computation

We propose an algorithm for computing sunlight maps from shading and environment lighting estimation, which is applicable to both indoor and outdoor environments. A key observation, illus-

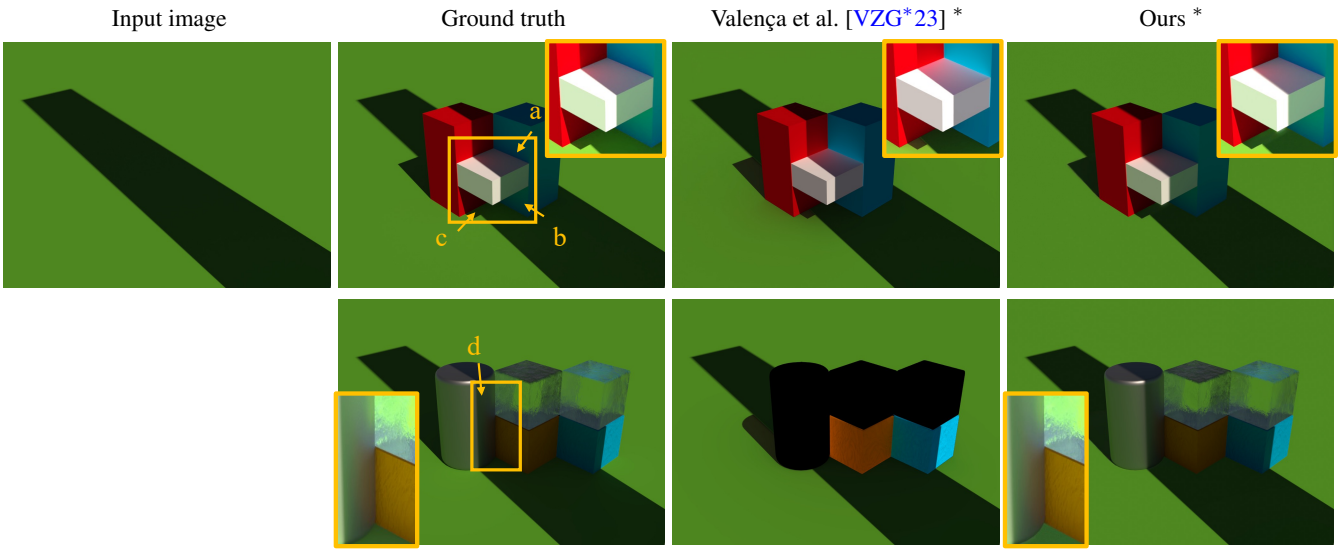

Figure 8: *Comparison of different rendered materials. Valença et al. [VZG*23] cannot render non-Lambertian materials such as metal and glass (bottom row). In addition, (a) sunlight bounces on Lambertian surfaces are improperly handled, causing light leakage artifacts; (b) colorful inter-reflections from the ground plane are lost because its albedo is ignored in scene reconstruction; and (c) our shadow refinement blends virtual shadows more smoothly with background shadows (Section 3.3). By modeling occlusion as zero-intensity light rays, our method allows the renderer to naturally handle diverse light transport, e.g., reflections and transmission through glass (d). * Both methods are provided with ground-truth lighting conditions and other required properties.*

trated in Figure 2, is that the shape of occlusion/shadows remains consistent in the sun's perspective, regardless of the inserted object. This motivates capturing the sunlight map $\mathbf{L}_{\text{sun}}$ in the sun's view for rendering. However, since intrinsic component estimation is available in the camera's view, we first compute the sunlight map in the camera's view and then transform it into the sun's view, as shown in Figure 7. Unless explicitly stated, computations are presented in the camera's view.

Sunlight map in the camera's view. The sunlight map $\mathbf{L}_{\text{sun}}$ is estimated from the direct shading map due to sunlight, which represents the light received by a surface without reflections or secondary bounces. We approximate direct shading by separating it from overall shading and environment shading. Assuming a Lambertian background scene, an image $\mathbf{I}$ is modelled as the element-wise product of albedo $\mathbf{A}$ and shading $\mathbf{S}$ using intrinsic image decomposition [GRPCLM22, BT78]. The shading $\mathbf{S}$ can be further decomposed into direct shading $\mathbf{S}^{\text{direct}}$ from the sun, and the environment shading $\mathbf{S}^{\text{env}}$ from other light sources and indirect light bounces from the sun:

$$\mathbf{I} = \mathbf{A} \odot \mathbf{S} = \mathbf{A} \odot (\mathbf{S}^{\text{direct}} + \mathbf{S}^{\text{env}}). \tag{1}$$

$\mathbf{S}^{\text{env}}_{\mathbf{x}}$ at location $\mathbf{x}$ can be obtained by integrating over the environment map $\mathcal{E}_{\mathbf{x}}$ (excluding direct sunlight from the sun's direction $\mathbf{v}$) with respect to the surface normal vector $\mathbf{N}_{\mathbf{x}}$ [BJ03, KBG*15] via:

$$\mathbf{S}^{\text{env}}_{\mathbf{x}} = \int_{\Omega} \mathcal{E}_{\mathbf{x}}(\omega) \max(0, \mathbf{N}_{\mathbf{x}} \cdot \omega)\, d\omega, \tag{2}$$

where $\mathbf{N}_{\mathbf{x}}$ is the surface normal at $\mathbf{x}$, $\mathcal{E}_{\mathbf{x}}(\omega)$ is the radiance from direction ω in the environment map and '$\cdot$' denotes the dot product.

Since we assume that the plane $\mathcal{G}$ is Lambertian and infinite, and render it with a single environment map, $\mathbf{S}^{\text{env}}_{\mathbf{x}}$ is constant and equal to $\mathbf{S}^{\text{env}}_{\text{o}}$, where subscript o denotes the user-specified location of the inserted object. Then, $\mathbf{S}^{\text{direct}}$ can be derived from Equation (1). The sunlight map in the camera's view is computed via Lambert's cosine law [Lam60]:

$$\mathbf{L}_{\text{cam},\mathbf{x}} = \begin{cases} \frac{\mathbf{S}^{\text{direct}}_{\mathbf{x}}}{\mathbf{N}_{\text{o}} \cdot \mathbf{v}}, & \text{if } \mathbf{N}_{\text{o}} \cdot \mathbf{v} > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

We use the normal $\mathbf{N}_{\text{o}}$ of the plane $\mathcal{G}$, ensuring that $\mathbf{L}_{\text{cam}}$ generates the same $\mathbf{S}^{\text{direct}}$ on this plane in the rendered foreground image $\mathbf{I}^{\text{F}}$.

Sunlight map in the sun's view. We transform the sunlight map $\mathbf{L}_{\text{cam}}$ from the camera's view to the sun's view $\mathbf{L}_{\text{sun}}$ using a planar homography induced by the proxy plane $\mathcal{G}$ [HZ03]. Unlike a full-scene view transformation, this homography prevents in-the-image occluders from covering their corresponding shadows on $\mathcal{G}$, as shown in Figure 7. Since only the shadows on $\mathcal{G}$ are properly prepared for projection, the sunlight map requires recomputation if the virtual object is moved to a different plane. This computation is detailed in the *supplement*.

3.2.3. Rendering and image compositing

We render the foreground image $\mathbf{I}^{\text{F}}$ of the inserted object and its shadows from our scene representation with Blender's Cycles path-tracing engine [Ble25]. We then composite the rendered foreground image $\mathbf{I}^{\text{F}}$ with the background image $\mathbf{I}^{\text{B}}$ to produce the composited image $\mathbf{I}^{\text{C}}$, as detailed in the *supplement*. Using ground-truth intrinsic components required by our approach, the renderings in Figure 9 represent the upper-bound performance of our method.

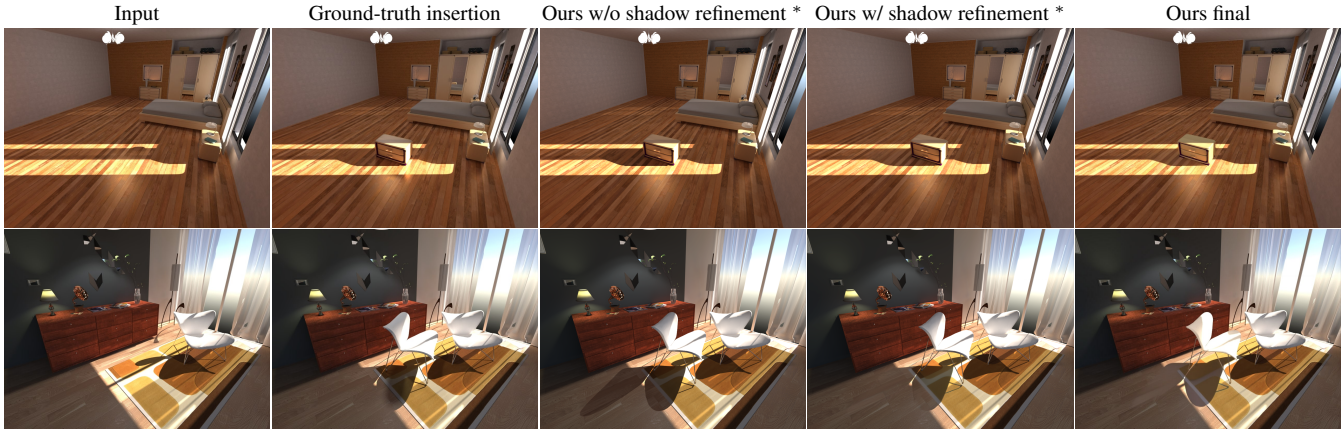

Figure 9: *Rendering results of our proposed ProjectiveShading. '*' denotes results using ground-truth intrinsic components, representing the upper-bound performance of our ProjectiveShading. Shadow refinement improves the consistency between the rendered virtual shadows and existing real shadows in the background. It is also useful when rendering using ground-truth components, as the environment lighting can change within the shadow regions (bottom row), which is insufficient to approximate using a single environment map. (Image source: our LightSOC dataset.)*

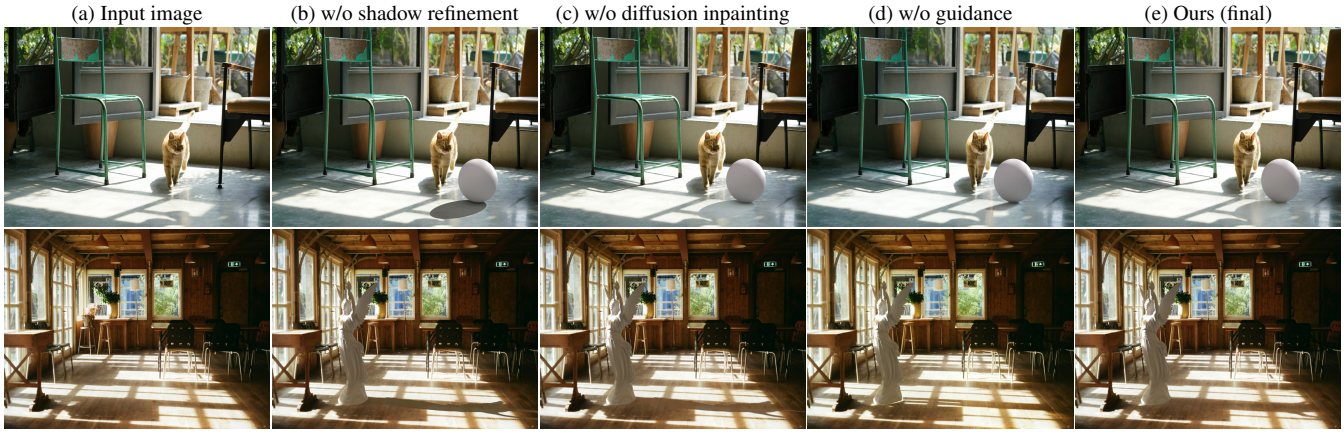

Figure 10: *Visual ablation of our two-step shadow refinement. (a) Input background image. (b) Raw rendered shadows from Blender. (c) Results after the first-step refinement. (d) Diffusion-based inpainting without guidance. While (c) achieves plausible shadow brightness and colour, it lacks smooth blending with background shadows and detailed albedo due to imperfect estimation. Without guidance from (c), (d) produces unrealistic shadows. The full shadow refinement (e), using (c) as guidance, significantly enhances the quality of the virtual shadows, overcoming issues caused by inaccurate estimation of the environment lighting and material. (Image source: Larry Zhao (top), Tanya Prodaan (bottom), Unsplash. Model source: Stanford Computer Graphics Laboratory.)*

3.3. Shadow refinement

Inconsistencies between original and rendered shadows can result in visually unconvincing composites, often caused by inaccurate intrinsic estimation or illumination changes introduced by the inserted objects. To improve realism, we propose an inpainting-based shadow refinement algorithm that leverages image priors learned in large-scale generative models. This algorithm consists of two steps: first, we globally adjust virtual shadows to align with nearby real shadows, similar to Valença et al. [VZG*23]; then, we refine shadows pixel-wise via guided inpainting.

For the second step, we use a pre-trained text-to-image latent diffusion model [RBL*22], selected for its strong generalisation ability and support for guided inpainting without task-specific fine-tuning. As trained on diverse image distributions, such generative models are not inherently constrained to produce physically correct results, even with explicit text prompts such as "realistic." To steer generation toward physically plausible shadows, we inject physical cues into the diffusion inpainting process through a guidance image. This guidance image is the result from the first-step refinement, where shadows are aligned using nearby real shadows, and it provides cues such as shadow shape from the renderer and coarse shadow brightness from the refinement step. The shadow mask $\mathbf{M}^{\mathrm{T}}$ defines the inpainting region. Following the guided inpainting tech-

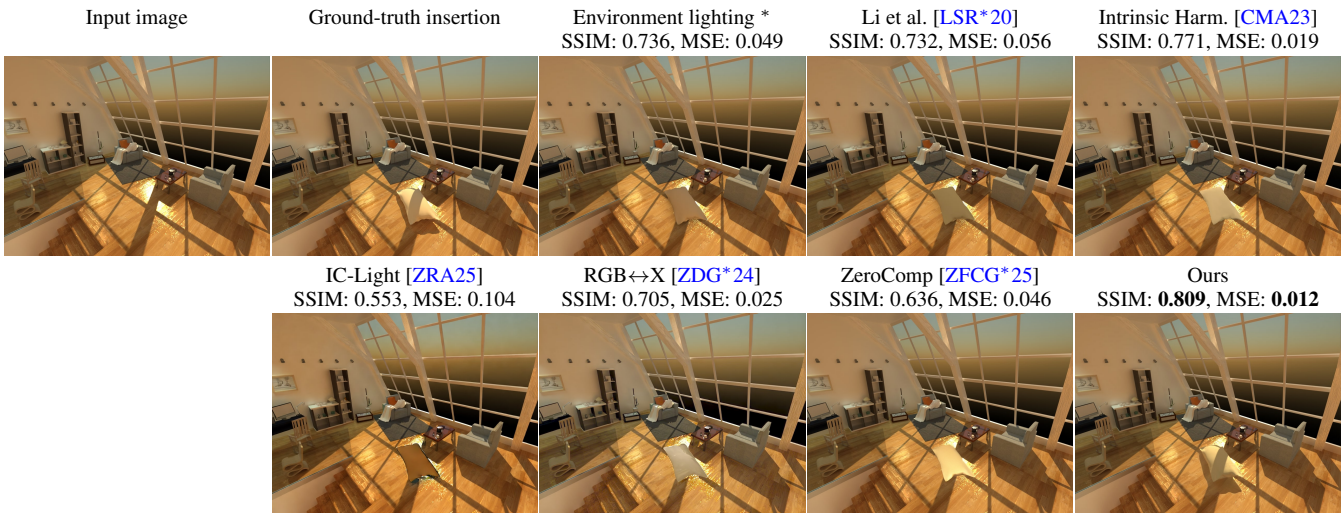

Figure 11: *Visual comparison of object insertion. '*' ground-truth components required for lighting and scene are provided. Only our approach can cast shadows onto the virtual object in this indoor image. The best scores are bold. (Image source: our LightSOC dataset.)*

nique SDEdit [MHS*21], we add noise to the guidance image for the last 20% of denoising steps, i.e., 20 out of 100 denoising steps. The noisy image is then denoised using the latent diffusion model with the text prompt "A photorealistic, highly detailed image with hard shadows" and a guidance scale of 7.5. The inpainted and refined shadow region is extracted from the denoised image $\tilde{\mathbf{I}}^{\mathrm{C}}_{\mathrm{sRGB}}$, and composited back into $\mathbf{I}^{\mathrm{C}}_{\mathrm{sRGB}}$ to obtain the final result. As shown in Figure 10, our guided diffusion inpainting refines virtual shadows and enhances realism.

4. Evaluation

4.1. Benchmarks and metrics

To evaluate our approach, we render a synthetic indoor dataset, *LightSOC*, tailored for virtual object insertion. Scenes and models are adapted from LightS [ETDBJ21] and rendered with Cycles, Blender's path-tracer [Ble25]. The dataset consists of 11 background scenes, each with multiple insertion results under primary sunlight, varying sky, and dimmer secondary lights (e.g., lamps). Each scene includes at least three Lambertian spheres and one additional object for evaluation, resulting in 63 test images in total.

OpenRooms [LYS*21] is an indoor dataset with pixel-wise ground-truth environment lighting maps on surfaces. We curated 140 test images likely containing primary sunlight and approximated the ground-truth sunlight direction as the brightest in the average environment map. These images were split into two folds of 70: one for tuning the sunlight direction refinement algorithm and the other for evaluation. More details are in the *supplement*.

On LightSOC, we use MSE and SSIM [WBSS04] to evaluate differences between rendered objects and ground truth within object mask regions (excluding shadows). Camera intrinsics and the depth of the closest ground-plane point are provided to ensure scale-invariant depth alignment and reduce perspective errors. On LightSOC and OpenRooms, sunlight direction estimation is evaluated using average and median angular errors, along with the percentage of errors below 10° and 20°.

4.2. Comparison

4.2.1. Comparison of inserted objects.

Since no previous work directly addresses our problem, we compare against the most relevant baselines from related areas:

- Environment lighting estimation and rendering [LSR*20];
- 2D image-based composition [CMA23, ZRA25]; and
- Neural rendering [ZDG*24, ZFCG*25].

For the environment lighting estimation method [LSR*20], we reimplemented the rendering pipeline in Blender using their estimated lighting, since the released code is not fully automatic. To ensure consistency, other scene properties (albedo, geometry and camera) match ours. As environment lighting estimation spans a large family of methods [GSY*17, PCS*24, ZZY*21, GSH*19, LSR*20, WTC*25], a baseline using ground-truth environment lighting for rendering is included to approximate upper-bound performance.

For 2D composition, we evaluate Intrinsic Harmonization [CMA23] and IC-Light [ZRA25]. The rendered result of Li et al. [LSR*20] (which is also used for rendering the environment lighting effects in our approach) is provided as the foreground crop. Neural rendering baselines include RGB↔X [ZDG*24] and ZeroComp [ZFCG*25]. Both are diffusion-based models conditioned on intrinsic properties. Background intrinsics are estimated using RGB↔X, and ground-truth intrinsics of the virtual objects are given.

Figures 11 and 12 show visual comparisons with baselines, and Table 1 reports quantitative results. All composited images have a long-side resolution of 1024 pixels. The virtual objects in Figure 12

Figure 12: *Visual comparison of object insertion into in-the-wild images: outdoor scenes (a–b) and indoor scenes (c–e). Our approach renders shadows on virtual objects and casts virtual shadows into the background, handling complex shadow interactions more faithfully than existing methods. Material characteristics: (a–b) spatially varying BRDF with various textures; (c) subsurface scattering; (d) mainly glossy; and (e) glossy with 0.5 roughness. (Image sources: Vero Eve (Unsplash), Victoria Strukovskaya (Unsplash), Utsman Media (Unsplash), David (Flickr), and Gianandrea Villa (Unsplash). Model sources: (a–b) from Hellz (FAB) and VladGl (Blender Swap); (c–e) from Stanford Computer Graphics Laboratory.)*

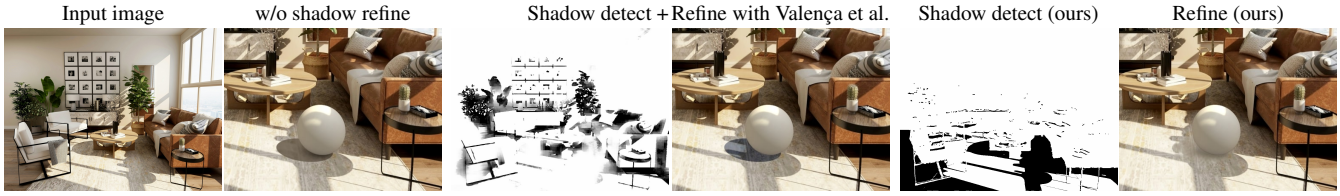

Figure 13: *Qualitative comparison of shadow refinement with Valença et al. [VZG*23]. Our rendered results are used as input for both methods, as their approach assumes sky-only environment lighting and is unsuitable for indoor scenes. Their shadow harmonization fails in detecting shadows in indoor environments and in blending virtual shadows with the background. Our shadow detection solely performs on the plane. (Image source: Collov Home Design, Unsplash.)*

Table 1: *Quantitative results of object composition on LightSOC. Errors are computed in the object mask regions (without shadows).* * *ground-truth components for lighting and scene are provided.*

Method	SSIM ↑	MSE ↓
Environment light *	0.724	0.049
Ours *	0.925	0.007
Careaga et al. [CMA23]	0.611	0.079
Li et al. [LSR*20]	0.695	0.062
Ours	0.728	0.045

Table 2: *Quantitative results of sunlight estimation on LightSOC and OpenRooms [LYS*21] datasets. With our proposed refinement method, the sunlight direction accuracy significantly improves across all metrics: average and median angular error, as well as percentile within 10°/20°.*

	LightSOC				OpenRooms			
Method	avg. ↓	med. ↓	10°(%) ↑	20°(%) ↑	avg. ↓	med. ↓	10°(%) ↑	20°(%) ↑
w/o ref.	32.5°	21.6°	9.1	36.4	47.7°	44.4°	10.0	34.3
Ours	27.0°	9.2°	54.5	72.7	43.5°	36.9°	20.0	41.4

have diverse geometries and are all non-Lambertian. In particular, the Stanford models (c–e) contain rich geometric detail and self-occlusion, stressing both visibility and light bounces.

Shadows on virtual surfaces are missing in almost all previous methods. RGB↔X is the only method that attempts this, though limited to outdoor scenes. It fails indoors and, even outdoors, produces shadows that are unrealistic and physically inconsistent with the occluder shape (b, Figure 12), in contrast to ours.

Virtual shadows on real surfaces are also missing in almost all previous methods. Li et al. [LSR*20], ZeroComp, and RGB↔X produce only soft shadows, lacking the hard cast shadows required under direct sunlight, and these soft shadows do not blend naturally with background shadows. Intrinsic Harmonization and IC-Light do not support casting any shadows into the background.

Material consistency is limited in diffusion-based neural rendering methods [ZDG*24, ZFCG*25]. Although ground-truth material properties of the virtual 3D objects are provided, RGB↔X introduces noticeable colour shifts (c and e, Figure 12). ZeroComp, on the other hand, often produces Lambertian-like results whilst the virtual objects are all non-Lambertian in Figure 12. Our results strictly preserve the provided material properties, a key advantage of classical rendering over these diffusion-based methods.

Our ProjectiveShading supports more realistic and complex shadow interactions.

4.2.2. Comparison of virtual shadows.

Valença et al. [VZG*23] model the sky as the sole environment lighting, making it unsuitable for indoor scenes. We use our results as the initial rendering for shadow refinement. Our shadow detection, based on thresholding the direct shading map, outperforms their proposed model, as demonstrated in Figure 13. Furthermore, our diffusion-based algorithm performs better in blending virtual shadows into the background. A key limitation of their model is its reliance on training data, which was only trained in outdoor environments. A comparison with additional examples is presented in the *supplement*.

4.3. Ablation and additional results

Ablation study. We evaluate the accuracy of our sunlight direction estimation with and without our proposed refinement qualitatively in Figure 6 and quantitatively in Table 2. Our refinement significantly improves over the initial estimate across all metrics. Figure 10 shows the visual quality improvement from our shadow refinement, which significantly improves shadow fidelity, as lighting estimation and intrinsic decomposition struggle with general and more complex Internet images.

Additional results. Figure 15 shows our results on a vertical wall, non-planar surfaces (with meshes reconstructed from Open3D [ZPK18]), and scenes lit by a spotlight. For the spotlight example, ground-truth intrinsic components and light direction are used, as our current method does not estimate spotlight direction. Our approach could be extended to other light sources with appropriate lighting estimation and projection models. More results are included in the *supplement*.

5. Limitations

The effectiveness of our method depends on the quality of geometry estimation and inverse rendering. Missing texture details in the estimated albedo can lead to blurring in virtual shadow regions if our

Figure 14: *Failure cases in sunlight direction estimation. While shadowing virtual objects enhances realism, the estimated sunlight direction is incorrect. Image sources: Amin Hasani (Unsplash) and Daniel McCullough (Unsplash).*

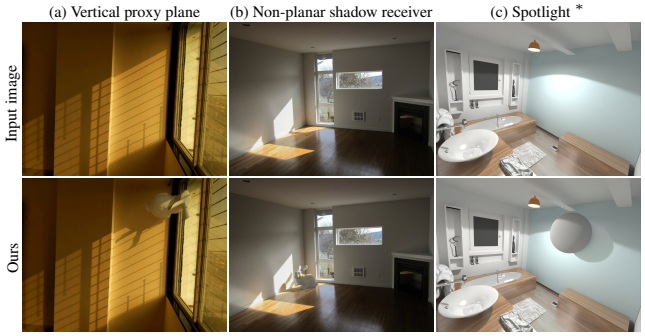

Figure 15: *Our insertion results. Our method supports arbitrary proxy plane orientations (e.g., the vertical wall in (a)) and shadowing non-planar surfaces (b). * indicates use of ground-truth intrinsic components, including spotlight direction. In (c), the highlight on the sphere is dim due to the improper use of orthographic projection for the spotlight, but the shadow on the sphere still enhances realism, demonstrating potential for extending our method to other light sources. (Image sources: Sreeja (Unsplash, left); erocsid (Flickr, middle); LightS [ETDBJ21] (right).)*

shadow refinement fails to recover them. Sunlight map estimation is influenced by the quality of HDR shading, the planarity of the proxy plane, the accuracy of the sunlight direction, and the visibility of background shadows. Sunlight direction refinement may fail when the environment lighting model significantly misidentifies the primary sunlight orientation or when hard shadow edges are not reliably detected. Some failure cases are shown in Figure 14. Finally, since our method leverages visible background shadows to infer sunlight-occlusion relationships, it supports shadowing virtual objects regardless of light source and occluder visibility, but struggles when the background shadows themselves are out-of-view. More discussions are included in the *supplement*.

6. Conclusion

We addressed the long-standing challenge of realistically inserting virtual objects into real-world images, particularly in complex indoor environments with intricate lighting conditions. ProjectiveShading achieves high realism by handling bi-directional shadow interactions through a fully automatic inverse- and re-rendering process. By rendering real-shadow effects on virtual objects, estimating primary sunlight direction, and blending newly cast virtual shadows into the background whilst accounting for material textures, our method significantly outperforms previous approaches in compositing objects. The effectiveness of ProjectiveShading is demonstrated through quantitative and qualitative evaluations on both synthetic ground-truth renderings and in-the-wild internet images. Our approach provides a novel solution for achieving realistic virtual object insertion in images of indoor environments.

7. Acknowledgments

This work was supported by the UKRI MyWorld Strength in Places Programme (SIPF00006/1). We would like to thank Zhengqin Li [LYS*21] and Ruicheng Wang [WXD*25] for their generous help in understanding the implementation details of their work. We also thank Elkhouly et al. [ETDBJ21] for releasing the source files of their LightS dataset.

References

[BBS14] BELL S., BALA K., SNAVELY N.: Intrinsic images in the wild. *ACM Trans. Graph. 33*, 4 (2014), 159:1–12. doi:10.1145/2601097.2601206. 3

[BHY*23] BAI J., HE Z., YANG S., GUO J., CHEN Z., ZHANG Y., GUO Y.: Local-to-global panorama inpainting for locale-aware indoor lighting prediction. *IEEE Trans. Vis. Comput. Graph. 29*, 11 (2023), 4405–4416. doi:10.1109/TVCG.2023.3320233. 2

[BJ03] BASRI R., JACOBS D. W.: Lambertian reflectance and linear subspaces. *IEEE Trans. Pattern Anal. Mach. Intell. 25*, 2 (2003), 218–233. doi:10.1109/TPAMI.2003.1177153. 6

[Ble25] BLENDER ONLINE COMMUNITY: *Blender – a 3D modelling and rendering package*. Blender Foundation, 2025. 2, 6, 8

[BT78] BARROW H. G., TENENBAUM J. M.: Recovering intrinsic scene characteristics from images. In *Computer Vision Systems* (1978), pp. 3–26. 6

[CA23] CAREAGA C., AKSOY Y.: Intrinsic image decomposition via ordinal shading. *ACM Trans. Graph. 43*, 1 (2023), 12:1–24. doi:10.1145/3630750. 3

[CA24] CAREAGA C., AKSOY Y.: Colorful diffuse intrinsic image decomposition in the wild. *ACM Trans. Graph. 43*, 6 (2024), 178:1–12. doi:10.1145/3687984. 3

[CCC*18] CHANG S.-H., CHIU C.-Y., CHANG C.-S., CHEN K.-W., YAO C.-Y., LEE R.-R., CHU H.-K.: Generating 360 outdoor panorama dataset with reliable sun position estimation. In *SIGGRAPH Asia Posters* (2018). 3

[CGC*03] CHUANG Y.-Y., GOLDMAN D. B., CURLESS B., SALESIN D. H., SZELISKI R.: Shadow matting and compositing. In *SIGGRAPH* (2003), pp. 494–500. doi:10.1145/1201775.882298. 3

[CLP*23] CHOI J., LEE S., PARK H., JUNG S.-W., KIM I.-J., CHO J.: MAIR: Multi-view attention inverse rendering with 3D spatially-varying lighting estimation. In *CVPR* (2023). 2

[CMA23] CAREAGA C., MIANGOLEH S. M. H., AKSOY Y.: Intrinsic harmonization for illumination-aware compositing. In *SIGGRAPH Asia* (2023), pp. 100:1–10. doi:10.1145/3610548.3618178. 3, 8, 9, 10

[CPY*25] CHEN X., PENG S., YANG D., LIU Y., PAN B., LV C., ZHOU X.: IntrinsicAnything: Learning diffusion priors for inverse rendering under unknown illumination. In *ECCV* (2025), pp. 450–467. 3

[Deb98] DEBEVEC P. E.: Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *SIGGRAPH* (1998). 2

[DM97] DEBEVEC P. E., MALIK J.: Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH* (1997). 2

[DRC*15] DUCHÊNE S., RIANT C., CHAURASIA G., MORENO J. L., LAFFONT P.-Y., POPOV S., BOUSSEAU A., DRETTAKIS G.: Multi-view intrinsic images of outdoors scenes with an application to relighting. *ACM Trans. Graph. 34*, 5 (2015), 164:1–16. doi:10.1145/2756549. 3

[ETDBJ21] ELKHOULY M. D., TSESMELIS T., DEL BUE A., JAMES S.: LightS: Light specularity dataset for specular detection in multi-view. In *ICIP* (2021). 8, 11

[FB81] FISCHLER M. A., BOLLES R. C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM 24*, 6 (1981), 381–395. doi:10.1145/358669.358692. 4

[FCZM*26] FORTIER-CHOUINARD F., ZHANG Z., MESSIER L.-E., GARON M., BHATTAD A., LALONDE J.-F.: SpotLight: Shadow-guided object relighting via diffusion. In *3DV* (2026). 3

[GHGS*19] GARDNER M.-A., HOLD-GEOFFROY Y., SUNKAVALLI K., GAGNE C., LALONDE J.-F.: Deep parametric indoor lighting estimation. In *ICCV* (2019), pp. 7175–7183. 2

[GRPCLM22] GARCES E., RODRIGUEZ-PARDO C., CASAS D., LOPEZ-MORENO J.: A survey on intrinsic images: Delving deep into Lambert and beyond. *International Journal of Computer Vision 130* (2022), 836–868. 2, 3, 6

[GSH*19] GARON M., SUNKAVALLI K., HADAP S., CARR N., LALONDE J.-F.: Fast spatially-varying indoor lighting estimation. In *CVPR* (2019), pp. 6908–6917. 2, 8

[GSY*17] GARDNER M.-A., SUNKAVALLI K., YUMER E., SHEN X., GAMBARETTO E., GAGNÉ C., LALONDE J.-F.: Learning to predict indoor illumination from a single image. *ACM Trans. Graph. 36*, 6 (2017), 176:1–14. doi:10.1145/3130800.3130891. 2, 8

[GZJ*21] GUO Z., ZHENG H., JIANG Y., GU Z., ZHENG B.: Intrinsic image harmonization. In *CVPR* (2021), pp. 16362–16371. 3

[HGAL19] HOLD-GEOFFROY Y., ATHAWALE A., LALONDE J.-F.: Deep sky modeling for single image outdoor lighting estimation. In *CVPR* (2019), pp. 6920–6928. 3

[HGSH*16] HOLD-GEOFFROY Y., SUNKAVALLI K., HADAP S., GAMBARETTO E., LALONDE J.-F.: Deep outdoor illumination estimation. In *CVPR* (2016), pp. 2373–2382. 3

[HZ03] HARTLEY R., ZISSERMAN A.: *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2003. 6

[JLYT23] JIN Y., LI R., YANG W., TAN R. T.: Estimating reflectance layer from a single image: Integrating reflectance guidance and shadow/specular aware learning. In *AAAI* (2023). 3

[JRM20] JIDDI S., ROBERT P., MARCHAND É.: Detecting specular reflections and cast shadows to estimate reflectance and illumination of dynamic indoor scenes. *IEEE Transactions on Visualization and Computer Graphics 28* (2020), 1249–1260. 2

[KBG*15] KRONANDER J., BANTERLE F., GARDNER A., MIANDJI E., UNGER J.: Photorealistic rendering of mixed reality scenes. *Comput. Graph. Forum 34*, 2 (2015), 643–665. doi:10.1111/cgf.12591. 1, 2, 6

[KHFH11] KARSCH K., HEDAU V., FORSYTH D., HOIEM D.: Rendering synthetic objects into legacy photographs. *ACM Trans. Graph. 30*, 6 (2011), 157:1–12. doi:10.1145/2070781.2024191. 3

[KKK06] KHANG B.-G., KOENDERINK J. J., KAPPERS A. M. L.: Perception of illumination direction in images of 3-D convex objects: Influence of surface materials and light fields. *Perception 35* (2006), 625–645. 2

[KOF13] KEE E., O'BRIEN J. F., FARID H.: Exposing photo manipulation with inconsistent shadows. *ACM Trans. Graph. 32* (2013), 28:1–28:12. 2

[KSH*14] KARSCH K., SUNKAVALLI K., HADAP S., CARR N., JIN H., FONTE R., SITTIG M., FORSYTH D.: Automatic scene inference for 3D object compositing. *ACM Transactions on Graphics (TOG) 33*, 3 (2014), 1–15. 2

[KSN23] KOCSIS P., SITZMANN V., NIESSNER M.: Intrinsic image diffusion for indoor single-view material estimation. In *CVPR* (2023), pp. 5198–5208. 3

[Lam60] LAMBERT J. H.: *Photometria Sive de Mensura et Gradibus Luminis, Colorum et Umbrae*. Eberhard Klett, 1760. 6

[LCY*24] LUO J., CEYLAN D., YOON J. S., ZHAO N., PHILIP J., FRÜHSTÜCK A., LI W., RICHARDT C., WANG T. Y.: IntrinsicDiffusion: Joint intrinsic layers from latent diffusion models. In *SIGGRAPH* (2024). doi:10.1145/3641519.3657472. 3, 4

[LDHG*25] LYU L., DESCHAINTRE V., HOLD-GEOFFROY Y., HAŠAN M., YOON J. S., LEIMKÜHLER T., THEOBALT C., GEORGIEV I.: IntrinsicEdit: Precise generative image manipulation in intrinsic space. *ACM Trans. Graph. 44*, 4 (2025), 106:1–13. doi:10.1145/3731173. 3

[LEN12] LALONDE J.-F., EFROS A. A., NARASIMHAN S. G.: Estimating the natural illumination conditions from a single outdoor image. *International Journal of Computer Vision 98*, 2 (2012), 123–145. doi:10.1007/s11263-011-0501-8. 3

[LGL*25] LIANG R., GOJCIC Z., LING H., MUNKBERG J., HASSELGREN J., LIN Z.-H., GAO J., KELLER A., VIJAYKUMAR N., FIDLER S., WANG Z.: DiffusionRenderer: Neural inverse and forward rendering with video diffusion models. In *CVPR* (2025). 3

[LGND*24] LIANG R., GOJCIC Z., NIMIER-DAVID M., ACUNA D., VIJAYKUMAR N., FIDLER S., WANG Z.: Photorealistic object insertion with diffusion-guided inverse rendering. In *ECCV* (2024). 2

[LHL*20] LUO J., HUANG Z., LI Y., ZHOU X., ZHANG G., BAO H.: NIID-Net: Adapting surface normal knowledge for intrinsic image decomposition in indoor scenes. *IEEE Transactions on Visualization and Computer Graphics 26*, 12 (2020), 3434–3445. 3

[LLZ*20] LIU D., LONG C., ZHANG H., YU H., DONG X., XIAO C.: ARShadowGAN: Shadow generative adversarial network for augmented reality in single light scenes. In *CVPR* (2020), pp. 8139–8148. 3

[LMF*19] LEGENDRE C., MA W.-C., FYFFE G., FLYNN J., CHARBONNEL L., BUSCH J., DEBEVEC P.: DeepLight: Learning illumination for unconstrained mobile mixed reality. In *CVPR* (2019). 2

[LS18a] LI Z., SNAVELY N.: CGIntrinsics: Better intrinsic image decomposition through physically-based rendering. In *ECCV* (2018). 3

[LS18b] LI Z., SNAVELY N.: Learning intrinsic image decomposition from watching the world. In *CVPR* (2018), pp. 9039–9048. 3

[LSB*22] LI Z., SHI J., BI S., ZHU R., SUNKAVALLI K., HAŠAN M., XU Z., RAMAMOORTHI R., CHANDRAKER M.: Physically-based editing of indoor scene lighting from a single image. In *ECCV* (2022). 2, 4

[LSR*20] LI Z., SHAFIEI M., RAMAMOORTHI R., SUNKAVALLI K., CHANDRAKER M.: Inverse rendering for complex indoor scenes: Shape,

spatially-varying lighting and SVBRDF from a single image. In *CVPR* (2020). 1, 2, 4, 5, 8, 9, 10

[LTH*23] LYU L., TEWARI A. K., HABERMANN M., SAITO S., ZOLLHÖFER M., LEIMKÜHLER T., THEOBALT C.: Diffusion posterior illumination for ambiguity-aware inverse rendering. *ACM Transactions on Graphics (TOG) 42* (2023). 2

[LYS*21] LI Z., YU T.-W., SANG S., WANG S., SONG M., LIU Y., YEH Y.-Y., ZHU R., GUNDAVARAPU N., SHI J., BI S., YU H.-X., XU Z., SUNKAVALLI K., HAŠAN M., RAMAMOORTHI R., CHANDRAKER M.: OpenRooms: An open framework for photorealistic indoor scene datasets. In *CVPR* (2021). 2, 8, 10, 11

[LYW*24] LIU Q., YOU J., WANG J., TAO X., ZHANG B., NIU L.: Shadow generation for composite image using diffusion model. In *CVPR* (2024). 3

[LYX*24] LING J., YU R., XU F., DU C., ZHAO S.: NeRF as non-distant environment emitter in physics-based inverse rendering. In *SIGGRAPH* (2024). 2

[LZLR24] LUO J., ZHAO N., LI W., RICHARDT C.: CRefNet: Learning consistent reflectance estimation with a decoder-sharing transformer. *IEEE Trans. Vis. Comput. Graph. 30*, 9 (2024), 6407–6420. doi:10.1109/TVCG.2023.3337870. 3

[Mam04] MAMASSIAN P.: Impossible shadows and the shadow correspondence problem. *Perception 33* (2004), 1279–1290. 2

[MHS*21] MENG C., HE Y., SONG Y., SONG J., WU J., ZHU J.-Y., ERMON S.: SDEdit: Guided image synthesis and editing with stochastic differential equations. In *ICLR* (2021). 8

[MKK98] MAMASSIAN P., KNILL D. C., KERSTEN D. J.: The perception of cast shadows. *Trends in Cognitive Sciences 2* (1998), 288–295. 2

[NCL*21] NIU L., CONG W., LIU L., HONG Y., ZHANG B., LIANG J., ZHANG L.: Making images real again: A comprehensive survey on deep image composition. arXiv:2106.14490, 2021. 1

[PCS*24] PHONGTHAWEE P., CHINCHUTHAKUN W., SINSUNTHITHET N., RAJ A., JAMPANI V., KHUNGURN P., SUWAJANAKORN S.: DiffusionLight: Light probes for free by painting a chrome ball. In *CVPR* (2024). 2, 8

[RBL*22] ROMBACH R., BLATTMANN A., LORENZ D., ESSER P., OMMER B.: High-resolution image synthesis with latent diffusion models. In *CVPR* (2022). 7

[RH01] RAMAMOORTHI R., HANRAHAN P.: An efficient representation for irradiance environment maps. In *SIGGRAPH* (2001), pp. 497–500. doi:10.1145/383259.383317. 2

[SBXX25] SHEN S., BAO Z., XU W., XIAO C.: IllumiDiff: indoor illumination estimation from a single image with diffusion model. *IEEE transactions on visualization and computer graphics* (2025). 2

[SF19] SONG S., FUNKHOUSER T.: Neural illumination: Lighting prediction for indoor environments. In *CVPR* (2019), pp. 6918–6926. 2

[SK20] SOMANATH G., KURZ D.: HDR environment map estimation for real-time augmented reality. In *CVPR* (2020), pp. 11293–11301. 2

[SSI03] SATO I., SATO Y., IKEUCHI K.: Illumination from shadows. *IEEE Trans. Pattern Anal. Mach. Intell. 25*, 3 (2003), 290–300. doi:10.1109/TPAMI.2003.1182093. 2

[SWP11] SCHERZER D., WIMMER M., PURGATHOFER W.: A survey of real-time hard shadow mapping methods. *Computer Graphics Forum 30* (2011). 5

[TZW*22] TANG J., ZHU Y., WANG H., CHAN J. H., LI S., SHI B.: Estimating spatially-varying lighting in urban scenes with disentangled representation. In *ECCV* (2022). 2

[UWH*03] UNGER J., WENGER A., HAWKINS T., GARDNER A., DEBEVEC P.: Capturing and rendering with incident light fields. In *Eurographics Symposium on Rendering* (2003), pp. 141–149. 2

[VZG*23] VALENÇA L., ZHANG J., GHARBI M., HOLD-GEOFFROY Y., LALONDE J.-F.: Shadow harmonization for realistic compositing. In *SIGGRAPH Asia* (2023), pp. 32:1–12. doi:10.1145/3610548.3618227. 2, 3, 5, 6, 7, 10

[WBSS04] WANG Z., BOVIK A. C., SHEIKH H. R., SIMONCELLI E. P.: Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing 13*, 4 (2004), 600–612. 8

[WGL22] WEBER H., GARON M., LALONDE J.-F.: Editable indoor lighting estimation. In *ECCV* (2022). 2

[WLT*21] WANG Y., LIU A., TUCKER R., WU J., CURLESS B. L., SEITZ S. M., SNAVELY N.: Repopulating street scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 5110–5119. 3

[WPFK21] WANG Z., PHILION J., FIDLER S., KAUTZ J.: Learning indoor inverse rendering with 3D spatially-varying lighting. In *ICCV* (2021). 2

[WTC*25] WANG L., TRAN D. M., CUI R., TG T., DAHL A. B., BIGDELI S. A., FRISVAD J. R., CHANDRAKER M.: Materialist: Physically based editing using single-image inverse rendering, 2025. URL: https://arxiv.org/abs/2501.03717, arXiv:2501.03717. 2, 8

[WXD*25] WANG R., XU S., DAI C., XIANG J., DENG Y., TONG X., YANG J.: MoGe: Unlocking accurate monocular geometry estimation for open-domain images with optimal training supervision. In *CVPR* (2025). 4, 11

[WXL24] WANG X., XIAO S., LIANG X.: LightOctree: Lightweight 3D spatially-coherent indoor lighting estimation. In *CVPR* (2024), pp. 4536–4545. 2

[WYLL22] WANG G., YANG Y., LOY C. C., LIU Z.: StyleLight: HDR panorama generation for lighting estimation and editing. In *ECCV* (2022). 2

[XZP*13] XING G., ZHOU X., PENG Q., LIU Y., QIN X.: Lighting simulation of augmented outdoor scene based on a legacy photograph. *Comput. Graph. Forum 32*, 7 (2013), 101–110. doi:10.1111/cgf.12217. 3, 5

[YGH*21] YU P., GUO J., HUANG F., ZHOU C., CHE H., LING X., GUO Y.: Hierarchical disentangled representation learning for outdoor illumination estimation and editing. In *ICCV* (2021), pp. 15313–15322. 3

[YGH*23] YU P., GUO J., HUANG F., CHEN Z., WANG C., ZHANG Y., GUO Y.: ShadowMover: Automatically projecting real shadows onto virtual object. *IEEE Trans. Vis. Comput. Graph. 29*, 5 (2023), 2379–2389. doi:10.1109/TVCG.2023.3247066. 2, 3

[YS19] YU Y., SMITH W. A.: InverseRenderNet: Learning single image inverse rendering. In *CVPR* (2019), pp. 3155–3164. 2

[ZCB*22] ZHI T., CHEN B., BOYADZHIEV I., KANG S. B., HEBERT M., NARASIMHAN S. G.: Semantically supervised appearance decomposition for virtual staging from a single panorama. *ACM Trans. Graph. 41*, 4 (2022). doi:10.1145/3528223.3530148. 2

[ZDG*24] ZENG Z., DESCHAINTRE V., GEORGIEV I., HOLD-GEOFFROY Y., HU Y., LUAN F., YAN L.-Q., HAŠAN M.: RGB↔X: Image decomposition and synthesis using material-and lighting-aware diffusion models. In *SIGGRAPH* (2024). 3, 8, 9, 10

[ZFCG*25] ZHANG Z., FORTIER-CHOUINARD F., GARON M., BHATTAD A., LALONDE J.-F.: ZeroComp: Zero-shot object compositing from image intrinsics via diffusion. In *WACV* (2025). 3, 8, 9, 10

[ZLH*22] ZHU J., LUAN F., HUO Y., LIN Z., ZHONG Z., XI D., WANG R., BAO H., ZHENG J., TANG R.: Learning-based inverse rendering of complex indoor scenes with differentiable Monte Carlo raytracing. In *SIGGRAPH Asia* (2022), pp. 6:1–8. doi:10.1145/3550469.3555407. 2

[ZLM*22] ZHU R., LI Z., MATAI J., PORIKLI F., CHANDRAKER M.: IRISformer: Dense vision transformers for single-image inverse rendering in indoor scenes. In *CVPR* (2022). 2

[ZPK18] ZHOU Q.-Y., PARK J., KOLTUN V.: Open3D: A modern library for 3D data processing, 2018. arxiv:1801.09847. 10

[ZRA25] ZHANG L., RAO A., AGRAWALA M.: Scaling in-the-wild training for diffusion-based illumination harmonization and editing by imposing consistent light transport. In *ICLR* (2025). 8, 9

[ZYL*17] ZHANG L., YAN Q., LIU Z., ZOU H., XIAO C.: Illumination decomposition for photograph with multiple light sources. *IEEE Transactions on Image Processing 26* (2017), 4114–4127. 2, 4

[ZZLS21] ZHU Y., ZHANG Y., LI S., SHI B.: Spatially-varying outdoor lighting estimation from intrinsics. In *CVPR* (2021). 2

[ZZY*21] ZHAN F., ZHANG C., YU Y., CHANG Y., LU S., MA F., XIE X.: EMLight: Lighting estimation via spherical distribution approximation. In *AAAI* (2021). 2, 8