# Accidental Light Probes

Hong-Xing Yu[1] Samir Agarwala[1] Charles Herrmann[2] Richard Szeliski[2] Noah Snavely[2]
Jiajun Wu[1] Deqing Sun[2]

[1]Stanford University [2]Google Research

## Abstract

*Recovering lighting in a scene from a single image is a fundamental problem in computer vision. While a mirror ball light probe can capture omnidirectional lighting, light probes are generally unavailable in everyday images. In this work, we study recovering lighting from* accidental *light probes (ALPs)—common, shiny objects like Coke cans, which often accidentally appear in daily scenes. We propose a physically-based approach to model ALPs and estimate lighting from their appearances in single images. The main idea is to model the appearance of ALPs by photogrammetrically principled shading and to invert this process via differentiable rendering to recover incidental illumination. We demonstrate that we can put an ALP into a scene to allow high-fidelity lighting estimation. Our model can also recover lighting for existing images that happen to contain an ALP*[*].

*I'd rather be Shiny.* — Tamatoa from Moana, 2016

Figure 1. (Left) From an image that has an accidental light probe (a Diet Pepsi can), we insert a virtual object (a Diet Coke can) with estimated lighting using the accidental light probe (Middle), and using estimated lighting from a recent state-of-the-art lighting estimation method [49] (Right). Note how our method better relights the inserted can to produce an appearance consistent with the environment (e.g., the highlight reflection and overall intensity).

## 1. Introduction

Traditionally, scene lighting has been captured through the use of light probes, typically a chromium mirror ball; their shape (perfect sphere) and material (perfect mirror) allow for a perfect measurement of all light that intersects the probe. Unfortunately, perfect light probes rarely appear in everyday photos, and it is unusual for people to carry them around to place in scenes. Fortunately, many everyday objects share the desired properties of light probes: Coke cans, rings, and thermos bottles are shiny (high reflectance) and curved (have a variety of surface normals). These objects can reveal a significant amount of information about the scene lighting, and can be seen as imperfect "accidental" light probes (e.g., the Diet Pepsi in Figure 1). Unlike perfect light probes, they can easily be found in casual photos or acquired and placed in a scene. In this paper, we explore using such everyday, shiny, curved objects as Accidental Light Probes (ALPs) to estimate lighting from a single image.

In general, recovering scene illumination from a single view is fundamental for many computer vision applications such as virtual object insertion [9], relighting [46], and photorealistic data augmentation [51]. Yet, it remains an open problem primarily due to its highly ill-posed nature. Images are formed through a complex interaction between geometry, material, and lighting [21], and without precise prior knowledge of a scene's geometry or materials, lighting estimation is extremely under-constrained. For example, scenes that consist primarily of matte materials reveal little information about lighting, since diffuse surfaces behave like low-pass filters on lighting during the shading process [38], eliminating high-frequency lighting information. To compensate for the missing information, the computer vision community has explored using deep learning to extract data-driven priors for lighting estimation [14, 44]. However, these methods generally do not leverage physical measurements to address these ambiguities, yet physical measurements can offer substantial benefits in such an ill-posed setting.

For images with ALPs, we propose a physically-based

[*]Project website: https://kovenyu.com/ALP

modeling approach for lighting estimation. The main idea is to model the ALP appearance using physically-based shading and to invert this process to estimate lighting. This inversion process involves taking an input image, estimating the ALP's 6D pose and scale, and then using the object's surface geometry and material to infer lighting. Compared to purely data-driven learning approaches that rely on diverse, high-quality lighting datasets, which are hard to acquire, our physically-based approach generalizes to different indoor and outdoor scenes.

To evaluate this technique, we collect a test set of real images, where we put ALPs in daily scenes and show that our approach can estimate high-fidelity lighting. We also demonstrate lighting estimation and object insertion based on existing images (Figure 1).

In summary, we make the following three contributions:

- We propose the concept of *accidental* light probes (ALPs), which can provide strong lighting cues in everyday scenes and casual photos.
- We develop a physically-based approach for lighting estimation for images with an ALP and show improved visual performance compared to existing light estimation techniques.
- We collect a dataset of ALPs and a dataset of images with ALPs and light probes in both indoor and outdoor scenes. We demonstrate that our physically-based model outperforms existing methods on these datasets.

## 2. Related Work

**Lighting estimation.** Traditional light probes capture omnidirectional lighting [9] but are usually absent in existing images. Researchers have used everyday objects like human faces [6, 23, 27, 46, 56] and eyes [33] to estimate lighting in portrait images. In contrast, we target images with high-reflectance objects. Other research focuses on lighting estimation from known, non-reflective objects. Weber et al. [53] and Park et al. [34] learn to regress illumination directly from homogeneous-material objects, while et al. [54] extend this to spatially-varying materials. These methods require large, diverse lighting data to generalize. Some approaches use RGBD video [35, 39] to estimate scene lighting, but we focus on lighting estimation from a single RGB image.

In addition to object-based lighting estimation, another popular line of work focuses on learning lighting estimation directly from images of scenes [12–14, 43, 44, 58, 59]. Many of these methods rely heavily on supervised training on synthetic data. As a result, they are sensitive to domain shifts between training and test data and, in particular, suffer from a synthetic-to-real domain gap. In contrast, our approach is based on physically principled modeling and is not vulnerable to this issue.

**Inverse rendering.** Our approach is closely related to inverse rendering methods that aim to jointly recover geometry, material, and lighting from images. Recent work in this area uses multi-view observations of an object with known camera poses to recover scene lighting and object properties [18, 32]. These methods jointly optimize geometry, material, and lighting and generalize to diverse scene settings. However, in single-view settings, the optimization problem for inverse rendering is highly ill-posed, and these methods often produce degenerate solutions.

Learning-based inverse rendering techniques have also gained popularity in material and geometry estimation tasks [30,42,52,57,61]. These methods include differential rendering as part of their training pipeline and can learn priors to model geometry and materials of scenes and objects. However, they are limited in their ability to generalize to a diverse set of scenes.

**Material reconstruction.** Material modeling and reconstruction have a long history in computer vision and graphics. Early papers [4, 15, 47] developed early analytical models of material reflection based on general experimental observations. More recent works [8, 20, 28] have attempted to directly solve for a general bidirectional reflectance distribution function (BRDF), which analytically defines how light is reflected at a given point on an object's surface; however, many of these techniques fail for highly specular or curved objects. For example, traditional BRDF acquisition [10, 31] requires a gonioreflectometer, which tries to precisely measure reflectance at different angles. This machine typically runs on flat objects and struggles on curved objects like Diet Coke cans. Modern approaches [62] use RGBD sensors and joint optimization on differentiably rendered objects and multi-view images [18, 32, 60]; our approach builds upon differentiable rendering to optimize material reconstruction and adapts them for ALPs.

## 3. Approach

Accidental Light Probes (ALPs) are daily metallic shiny objects, such as a soda can, a thermoflask, or a ring. Given a single image containing an ALP, we aim to recover the incidental illumination by inverting physically-based rendering, as shown in Fig. 2. Our main idea is that we can first acquire the shape and reconstruct the spatially-varying BRDF of the ALP offline (Fig. 2 top), and then optimize incidental lighting as well as the 6D pose of the ALP (Fig. 2 bottom).

### 3.1. Formulation

Our goal is to estimate high-fidelity lighting from the appearance of an ALP in a single image. We approach this goal through the perspective of inverse rendering, where the forward process is described by the rendering equation [21]:

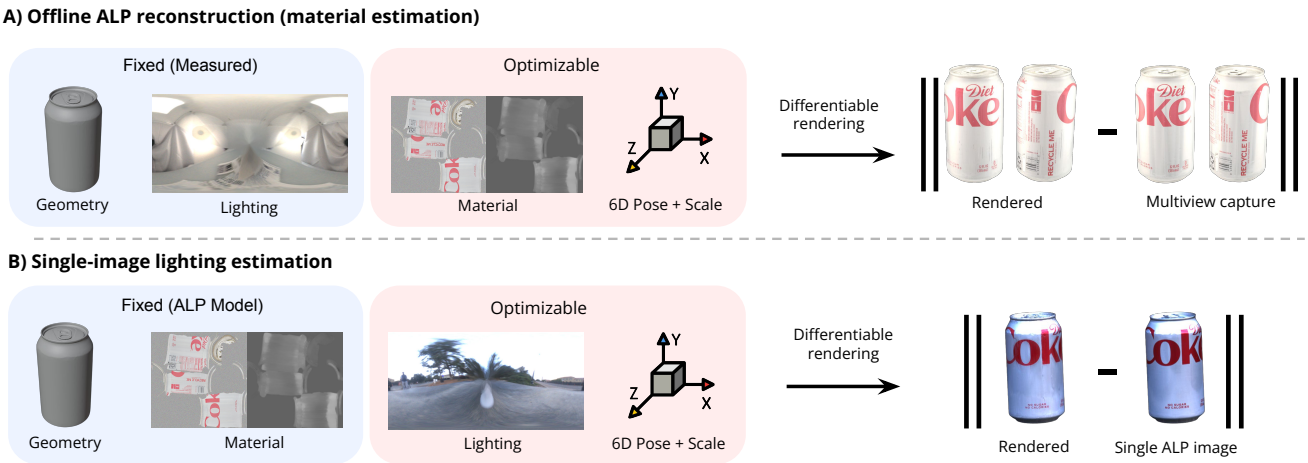


Figure 2. Our physically-based approach to lighting estimation consists of (A) offline Accidental Light Probe (ALP) reconstruction and (B) inference-time single-image lighting estimation. For (A), we use a capture-optimization hybrid method to reconstruct the ALP model with high fidelity. For (B), we formulate lighting estimation as a joint optimization of scale, 6D pose and environment lighting.

$$L(\omega_o) = \int_H L_i(\omega_i) f(\omega_i, \omega_o)(n \cdot \omega_i) d\omega_i, \tag{1}$$

where $L(\omega_o)$ is the outgoing radiance to direction $\omega_o$ (corresponding to pixel intensity), $L_i(\omega_i)$ is incidental radiance from direction $\omega_i$ (lighting), $f$ is the bidirectional reflectance distribution function (BRDF) at the surface location (material), $n$ is the normal direction (geometry), and $H$ is the upper hemisphere along the normal. Recovering lighting by inverting Eqn 1 is a highly ill-posed problem, as infinitely many combinations of geometry, material, and lighting can generate the same appearance in the image. Fortunately, for ALPs, we can pre-acquire prior physical knowledge of their shapes and materials as they are everyday objects. Thus, we can reduce the full inverse rendering problem to a joint estimation of 6D ALP pose and lighting, which is relatively more constrained and tractable:

$$\min_{\pi, L_i} \mathcal{L}\left(I_{\text{render}}(\pi, L_i | f, S), I_{\text{ref}}\right), \tag{2}$$

where $I_{\text{render}}$ is generated by a differentiable renderer that takes the shape $S$ (represented by a mesh), the 6D pose of the ALP $\pi$, the spatially-varying BRDF $f$, and the environment lighting $L_i$ as inputs. $I_{\text{ref}}$ denotes the observed single image. $\mathcal{L}$ denotes an image-space loss that we define in Section 3.3. Our physically-based formulation entails the high-fidelity acquisition of shape and spatially-varying material of the ALP, as well as a robust single-view joint optimization algorithm. We show an overview in Fig. 2 and elaborate the components in Section 3.2 and Section 3.3, respectively.

**Shading model.** We adopt physically-based rendering (PBR) [36] due to its principled photogrammetry and radiometry. Specifically, we consider metallic materials as they have little diffuse reflection. Diffuse reflection is undesirable as it behaves like a low-pass filter of lighting in the shading process [38], eliminating the physically recoverable lighting information. To model metallic material, we use a microfacet model [47] with a GGX distribution [48]:

$$f(\omega_i, \omega_o) = \frac{D \cdot F \cdot G}{4(n \cdot \omega_i)(n \cdot \omega_o)}, \tag{3}$$

where $D$ is the GGX normal distribution [48], $F$ is the Fresnel reflection, and $G$ is the geometric attenuation. We adopt Disney's parameterization [5], where the metallic material is modeled by its specular albedo $A$ and roughness $r$. Specifically, the specular albedo $A$ is used to model Fresnel reflection by Schlick's approximation [40] $F = A + (1 - A)(1 - |h \cdot \omega_o|)^5$, where $h = \frac{\omega_i + \omega_o}{|\omega_i + \omega_o|}$ denotes the half vector. The roughness $r$ controls the shape of the specular reflection lobe via the micro-normal distribution $D = \frac{r^4}{\pi(|n \cdot h|(r^4 - 1) + 1)^2}$ and the geometric attenuation $G = \frac{2|n \cdot \omega_i|}{|n \cdot \omega_i| + \sqrt{r^4 + (1 - r^4)|n \cdot \omega_i|^2}} \cdot \frac{2|n \cdot \omega_o|}{|n \cdot \omega_o| + \sqrt{r^4 + (1 - r^4)|n \cdot \omega_o|^2}}$.

**Lighting model.** To recover lighting for arbitrary conditions, we use an environment map to represent omnidirectional lighting and adopt image-based lighting for shading each pixel. For efficiency, we only consider direct lighting and use a differentiable rasterizer with deferred shading [24] to render $I_{\text{render}}$. This is inaccurate for concave objects with self-occlusion and self-reflections. To mitigate this without expensive global illumination, we include a soft visibility term to Eqn 1 to approximate it such that the shading output is modulated as $vL(\omega_o)$, where $v$ denotes the soft visibility that is optimized and treated as a surface texture.

### 3.2. Reconstructing ALPs

Recovering high-fidelity lighting by physically-based inverse rendering requires high-quality geometry and material reconstruction of the ALPs. While existing state-of-the-art inverse rendering methods can jointly optimize for geometry, material, and lighting from dense multi-view images [18, 32, 60], they still struggle for real metallic objects under arbitrary lighting due to high specularity (Fig. 4). Moreover, several challenges exist when the goal is not view synthesis but photogrammetrically correct reconstruction. For highly specular objects such as metallic ones, the reflected lights from the near field can lead to environment baking, as it breaks the distant light assumption (we show an example of the environment-baked material reconstruction in the supplementary material). The color ambiguity of material albedo and lighting is also not resolved. In addition, the geometry reconstruction quality heavily relies on the quality of object silhouettes in multi-view images.

To overcome these challenges, we reconstruct ALPs by a hybrid method. First, we use a light box with a turntable to control environment lighting for multi-view capture, and using a thin supporting stand to alleviate near-field reflections (setup shown in Fig. 3) and environment baking. Second, instead of optimizing the incidental lighting to the ALP under capture, we record it by a calibrated light probe to remove the color ambiguity between material and lighting. And third, we provide a high-quality shape using a range scanner [1] to reduce the geometry reconstruction down to 6D pose and size fitting. Thus, as demonstrated in the top row of Figure 2, our ALP reconstruction is cast as an optimization for its spatially-varying material and shape fitting:

$$\min_{\pi,\alpha,f} \sum_{\{I_{\text{capture}}\}} \mathcal{L}\left(I_{\text{render}}(\pi,\alpha,f|L_i,S), I_{\text{capture}}\right), \quad (4)$$

where $\pi$ and $\alpha$ are the 6D pose and size to fit the shape $S$ to multi-view camera coordinate frame solved by COLMAP [41], and $f$ is the material parameterized by spatially-varying albedo $A$ and roughness $r$. We show the reconstruction of a Coke can in Fig. 4. We include the optimization and loss details in the supplementary material.

### 3.3. Single-View Physically-Based Light Estimation

Given an image containing an ALP, we first extract an object segmentation mask for the ALP by manually cropping the image and then using an off-the-shelf foreground segmentation tool [2]; however, this could alternately be obtained by object detection [7] with salient object segmentation [37] or semantic segmentation [45]. We then retrieve the appropriate ALP model, containing its reflectance and geometric information. Yet, even given the ALP's 3D model and 2D segmentation in the input image, accurately aligning these two elements is still challenging. Traditional feature point detection and Perspective-n-Point methods do not work on textureless objects such as rings and thermoflasks. Additionally, modern learning-based single-view pose estimation methods [29, 50, 55] require diverse, realistic lighting to synthesize training data and do not generalize well outside the training distribution.

Figure 3. (Left) We use a light box with controllable lighting for our capture. To mitigate near-field reflections, we leverage a thin stand to support the object. (Right) To minimize environmental changes due to camera and photographer movement, we cover the lightbox with a cloth and use a turntable for multi-view capture.

Therefore, we formulate the lighting estimation and pose estimation as a joint estimation problem in Eqn 2, and we solve it via a differentiable rendering-based optimization which is generalizable to arbitrary scenes for both textured and textureless objects (see the bottom of Figure 2). Here we need a joint estimation as the appearance of a specular object (and thus the differentiable rendering gradient signals) is highly dependent on both the object pose and the environment lighting. We use Monte Carlo ray tracing with Visible Normal Distribution Function (VNDF) importance sampling [19] for unbiased shading.

**Losses and regularizations.** Our loss function used in Eqn 2 is given by:

$$\mathcal{L} = \mathcal{L}_{\text{RGB}} + \mathcal{L}_{\text{mask}} + \lambda_1 \mathcal{L}_{\text{pose-reg}} + \lambda_2 \mathcal{L}_{\text{light-reg}}, \quad (5)$$

where $\mathcal{L}_{\text{RGB}}$ denotes a $L_1$ loss on RGB images, $\mathcal{L}_{\text{mask}}$ denotes a combination of a $L_1$ loss and a Chamfer loss on masks [3], where the mask is given by the differentiable rasterizer [24]. $\mathcal{L}_{\text{pose-reg}}$ and $\mathcal{L}_{\text{light-reg}}$ denote a pose regularization and a lighting regularization with their weights $\lambda_1$ and $\lambda_2$, respectively.

Without multi-view constraints, the joint optimization problem has multiple local minima for the 6D pose; thus, we introduce a pose regularization and a lighting regularization. The pose regularization is given by:

$$\mathcal{L}_{\text{pose-reg}} = \|B(M_{\text{render}}) - B(M_{\text{ref}})\|_2^2 + \|q - q_{\text{ref}}\|_2^2, \quad (6)$$

where $M_{\text{render}}$ is the rendered mask, $B(M_{\text{render}})$ denotes the pixel-space barycenter of the mask, $q$ denotes the quaternion representation of the ALP orientation, and $q_{\text{ref}}$ denotes a common orientation (we use a front-facing canonical orientation

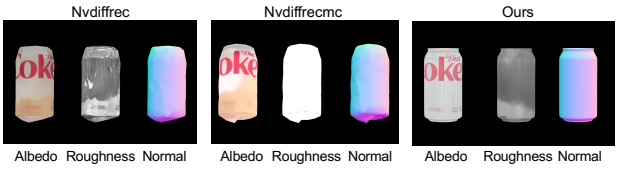


Figure 4. Visual comparison of ALP reconstruction from state-of-the-art optimization-based inverse rendering methods [18, 32] versus our hybrid method. Recent inverse rendering methods struggle on real textured metallic objects.

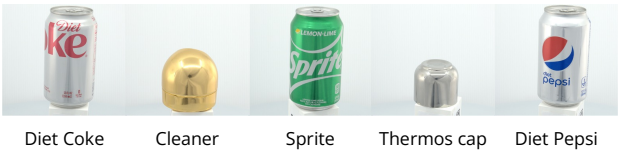


Figure 5. Close up of our ALP dataset.

obtained by aligning principal axes). The barycenter term prevents vanished gradient due to non-overlapping pose initialization, and the orientation term prevents hard-to-escape local minima like upside-down cans. We decay the weight of the pose regularization to zero through optimization. In addition, to further address local minima in 6D poses, we use multiple (4 in our experiments) orientation initialization and we keep the one with the highest re-rendering PSNR.

To accurately estimate omnidirectional lighting by inverting Eqn 1, we need to evaluate the Monte Carlo integral interval densely over light rays coming from all directions. However, from a single view, an ALP often only covers a limited subset of normal directions compared to a perfect sphere. Thus, light rays coming from a certain subset of directions contribute little to the appearance of the ALP. These directions are then under-sampled, and the lighting estimation for them is less informed and unconfident.

To mitigate this, we introduce a lighting smoothness regularization which "fills in" the less confident regions in the environment map by propagating the confident information from nearby directions. The regularization is given by:

$$\mathcal{L}_{\text{light-reg}} = \|L_i(\omega) - L_i(\omega + \Delta\omega)\|_1, \tag{7}$$

where $\Delta\omega$ denotes a small deviation of a solid angle sampled from a normal distribution, and $\omega$ is sampled uniformly in all solid angles. Note that in addition to propagating confident lighting estimates, the lighting regularization also helps improve pose estimation, since many pose estimation errors come from trying to fix mistakes in high-frequency [17] lighting changes, which light regularization alleviates.

# 4. Experiments

## 4.1. Setup

**Accidental Light Probes Dataset.** We acquire 5 common accidental light probes that have different shapes or spatially-varying BRDFs, including 3 soda cans (diet Coke, diet Pepsi, and Sprite), a thermos cap, and a solder tip cleaner. We show example images in Figure 5.

**Evaluation Dataset.** We collect a dataset of 10 indoor scenes and 13 outdoor scenes. The indoor and outdoor scenes are taken at different points of time, such as day and night, at different locations. We show examples in Figure 6. We place each of our ALPs in the scenes and capture HDR images of the ALPs. We also capture ground-truth lighting by a chromium ball (a perfect light probe).

**Baselines.** We compare our method to several state-of-the-art lighting estimation methods [12, 14, 49]. Unlike our method, all of these techniques utilize deep learning. Since [12, 14] do not have publicly available models, we asked their authors to run inference on our dataset.

## 4.2. Comparison to Baseline Methods

**Qualitative Results.** For all object insertion comparisons, we compute an environment map either through an ALP with our proposed method or by running the other baselines on a single image of the scene. Note that for the baselines, we use the image with the perfect light probe as input; this should provide a slight advantage to these techniques since the image with the perfect light probe contains the most information regarding scene lighting.

In Figure 7, we insert various objects into the scene and relight them using the computed environment maps; we then qualitatively compare the results. We demonstrate that our computed environment map produces significantly more accurate and compelling results than other single-image lighting estimation methods. In particular, note that our method is the only approach that can recover the overall tone of the lighting: other methods are either too yellow or gray.

In Figure 8, we show relighting results on perfect spheres of various finishes from all methods and ALPs on both indoor and outdoor scenes. Only our technique produces results similar to the ground truth for mirror finishes. We note that for all the three soda cans, the relighting on mirror spheres are slightly blurry, since their materials are much rougher than perfect mirror, which behaves as low-pass filters of lighting in the shading process [38]. We also note that for Sprite and diet Coke, there is some texture color baking in the recovered lighting due to imperfectly aligned 6D poses, which lead to high-frequency lighting artifacts to compensate the pixel-space misalignment. Our lighting regularization mitigates this type of artifacts, yet a highly robust algorithm remains as future work.

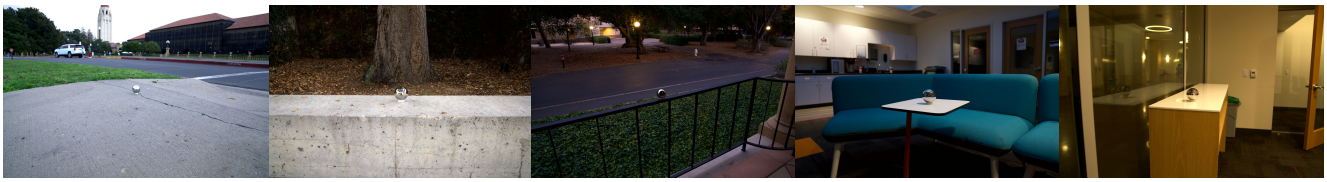

Figure 6. Examples of our collected dataset for evaluating lighting estimation under different illumination conditions, including indoor and outdoor scenes at daytime and nighttime.


Figure 7. Object insertion results on our test scenes for both indoor (first two rows) and outdoor (last two rows). We compare to Garon et al. [14], Deep parametric lighting [12], and StyleLight [49]. We center-crop the result images for better visualization.

**Quantitative Results.** In Table 1, we report quantitative results on relighting perfect spheres with various representative materials (mirror, shiny, diffuse). Similar to [26, 49], we compute angular error [11] and scale-invariant RMSE [16] to compare the relighted spheres from each technique to the ground truth relighting.

Quantitatively, for the relighting task, our method, applied to any of the ALPs, significantly outperforms the baselines. In particular, w.r.t. angular error, the Thermos cap provides a 3 to 4 times improvement over the best baseline.

### 4.3. Analysis

**Capture Setup.** We also analyze the quality of our reconstruction compared to two recent multi-view inverse rendering methods, Nvdiffrec [32] and Nvdiffrecmc [18] using our lightbox capture setups. Table 2 shows the results of using our lighting estimation pipeline with various reconstructions of a Diet Coke can. Our reconstruction performs the best and leads to a decrease of angular error by 20% or more. Both Nvdiffrec and Nvdiffrecmc are normally applied to multiview casual images, so for completeness, we also compute reconstructions and quantitative results for this setting (included in the supplementary materials). These reconstructions perform strictly worse than those computed from the lightbox setup. We also show a qualitative comparison of the geometry and materials in Figure 4, of each technique in its default setting, where ours are clearly better than the alternative methods. Table 2 shows that our ALP reconstruction pipelines give us better results than using current state-of-art inverse rendering methods to get our ALP models.

**Ablation for 6D Pose + Scale Estimation.** As mentioned in Sec. 3.3, the pose estimation problem for aligning a 3D model of an ALP and its appearance in a real image is challenging. The appearance of the object in the real image depends on both its pose and lighting; trying to jointly optimize these can introduce potential failure cases. In Sec 3.3, we describe several design choices w.r.t. the optimization and loss which address some of these failures cases. In Table 3,

| Method | Indoor | | | | | | Outdoor | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Angular Error*↓ | | | *Scale-invariant RMSE*↓ | | | *Angular Error*↓ | | | *Scale-invariant RMSE*↓ | | |
| | Mirror | Shiny | Diffuse | Mirror | Shiny | Diffuse | Mirror | Shiny | Diffuse | Mirror | Shiny | Diffuse |
| StyleLight [49] | 12.572 | 7.700 | 5.949 | 3.087 | 0.837 | 0.264 | 15.088 | 9.830 | 8.539 | 1.867 | 0.918 | 0.294 |
| Deep Param. [12] | 7.204 | 6.252 | 6.166 | 3.137 | 0.958 | 0.287 | 8.803 | 7.228 | 6.525 | 1.963 | 1.056 | 0.305 |
| Garon et al. [14] | 9.403 | 8.215 | 6.626 | 3.030 | 0.754 | 0.207 | 8.062 | 6.873 | 6.118 | 1.706 | 0.766 | 0.237 |
| Cleaner | 5.682 | 4.550 | 3.965 | **2.204** | **0.252** | 0.073 | 6.395 | 4.920 | 5.155 | **1.081** | 0.245 | 0.101 |
| Diet Coke | 4.733 | 3.405 | 3.067 | 2.901 | 0.550 | 0.101 | 6.011 | 3.877 | 2.587 | 1.460 | 0.501 | 0.136 |
| Diet Pepsi | 3.972 | 2.712 | 2.190 | 2.726 | 0.408 | 0.064 | 4.890 | 2.830 | **1.472** | 1.352 | 0.396 | 0.108 |
| Sprite | 5.952 | 4.445 | 3.767 | 2.913 | 0.556 | 0.112 | 7.023 | 4.923 | 3.892 | 1.468 | 0.513 | 0.154 |
| Thermos cap | **3.744** | **2.080** | **1.622** | 2.555 | 0.288 | **0.057** | **3.965** | **2.159** | 1.516 | 1.092 | **0.217** | **0.053** |

Table 1. Comparison to state-of-the-art single image lighting estimation methods: StyleLight [49], Deep Parametric [12] and Garon et al [14]. We evaluate them using relighting on different materials.

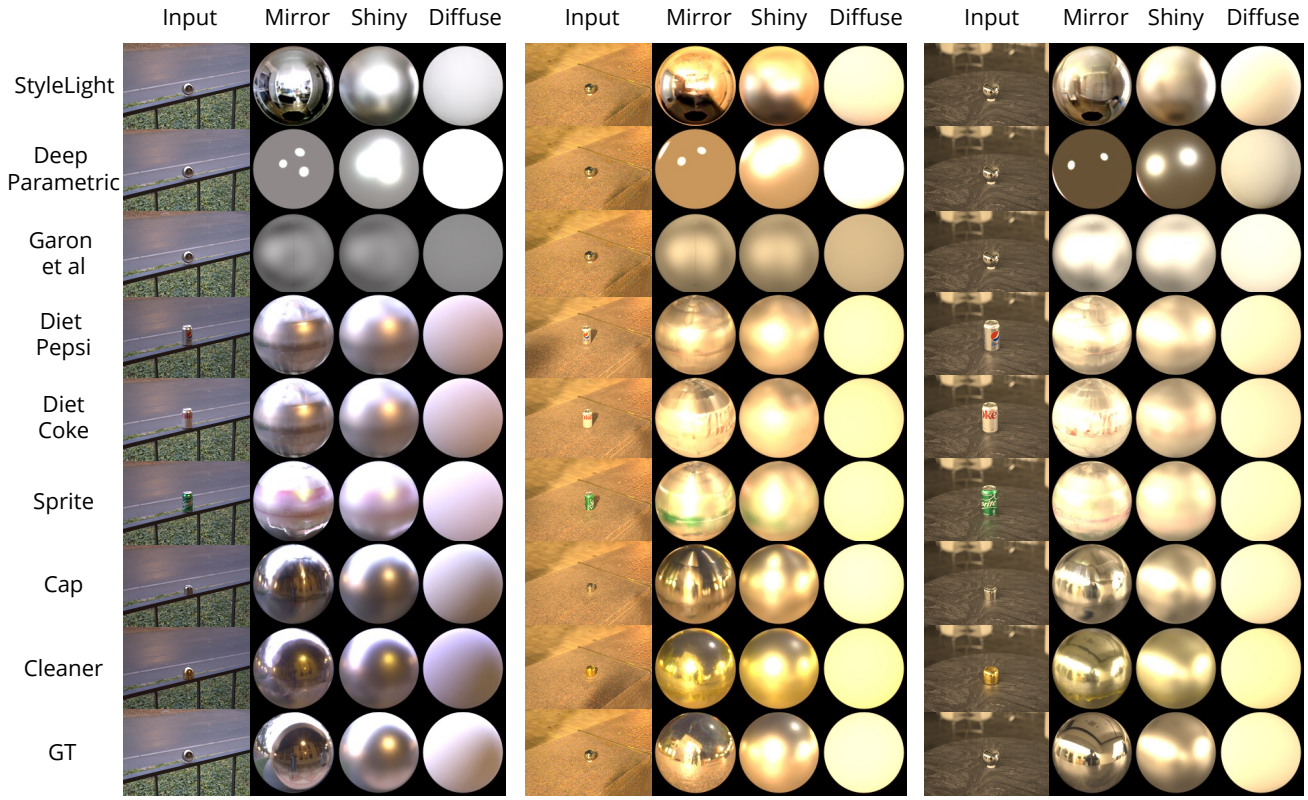


Figure 8. Qualitative comparison of relighting results in outdoor (left and center) and indoor (right) scenes. We compare our approach to StyleLight [49], Deep Parametric [12] and Garon et al [14] on relighting mirror, shiny and diffuse spheres.

we perform an ablation study on each of these decisions and report quantitative results. We show that all design decisions (Silhouette loss, Chamfer loss [3], joint optimization, pose, and light regularization) contribute to the final overall performance. We also show representative examples in Figure 9. They demonstrate how each design choice helps the pose estimation, which in return helps lighting estimation. In our supplementary material, we further showcase accurate estimations even under extreme object poses.

**Visualizing confident regions for ALPs.** As briefly discussed in Sec. 3.3, an ALP has a subset of surface normals compared to a perfect sphere light probe, which leads to under-sampled lighting directions. For example, cylindrical objects (Diet Coke can, ring, etc.) tend to sample well light rays perpendicular to the can while significantly undersampling light rays above and below the can. Since we use VNDF importance sampling which aligns well with our BRDF's density lobe, we visualize a "confidence map" as

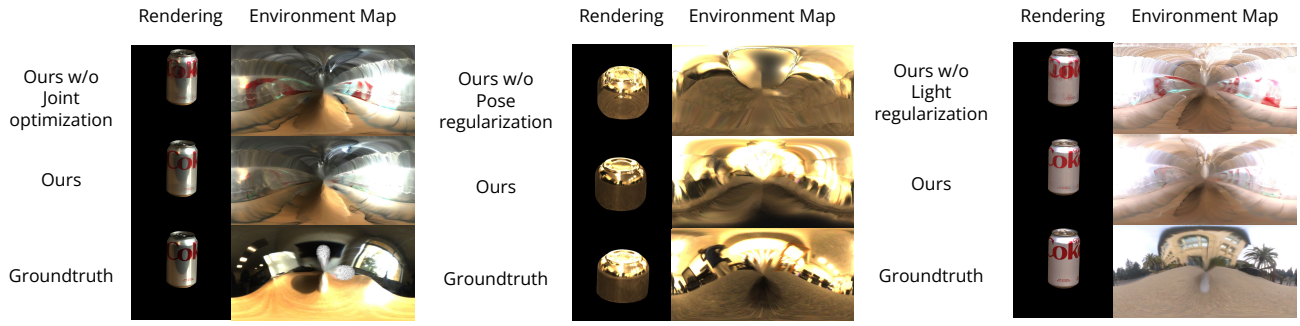


Figure 9. Qualitative ablation of the losses we use in our method. Each of our design choices contributes to improvements in pose and lighting optimization which can be observed qualitatively.

| Method | Mirror | Shiny | Diffuse |
|---|---|---|---|
| Nvdiffrec [32] | 6.99 | 5.06 | 3.59 |
| Nvdiffrecmc [18] | 6.55 | 4.60 | 3.84 |
| ALP (Ours) | **5.46** | **3.67** | **2.80** |

Table 2. Evaluation on our ALP model acquisition for a Diet Coke can using our lightbox setup. We compare our acquisition method to Nvdiffrec [32] and Nvdiffrecmc [18]. We use the same lighting estimation approach for compared methods and report average angular error across all test scenes.

| Method | Mirror | Shiny | Diffuse |
|---|---|---|---|
| Silhouette loss [3] | 6.812 | 4.976 | 3.919 |
| Ours w/o joint optimization | 5.401 | 3.726 | 3.044 |
| Ours w/o pose regularization | 5.962 | 4.180 | 3.338 |
| Ours w/o light regularization | 6.032 | 3.647 | 2.954 |
| Ours | **5.291** | **3.610** | **2.923** |

Table 3. Ablation study on our joint pose-lighting optimization. We compare to a baseline that uses a silhouette loss and a Chamfer loss [3], and variants of our approach. We show angular errors averaged on all test scenes.

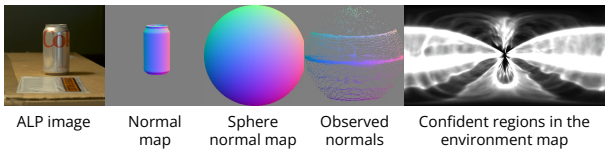


Figure 10. Visualization of sampling directions for a diet Coke can. See the text in 4.3 for a full description of these visualizations.

normalized sampling frequency. We show this confidence map in Figure 10 for a representative ALP (i.e., Diet Coke). This demonstrates that the visible surface of a Coke can from a single view only under-samples lighting directions from the top and the bottom.

In our supplementary material, we further show a controlled qualitative analysis of ALPs with different reflectance or shapes to demonstrate that our approach is tolerant to insignificant reflectance and shape variations.

**Discussion.** Our method shows strong promise for recovering scene lighting from a single image containing an ALP. One exciting potential application is improved image editing for in-the-wild images; however, to enable this for *any* image, we would either need to increase the number of ALPs or explore methods that enable us to dynamically edit one of the collected measurements (geometry or material). Another limitation is that we assume our input is an HDR image. However, we note that recent work has sought to convert LDR images to HDR [22,25], and HDR images have become more ubiquitous since many commercial mobile phones now support HDR capture.

## 5. Conclusion

In this paper, we introduced the use of accidental light probes to estimate environmental lighting from single images. We did this by first scanning common 3D objects and reconstructing their reflective properties. We then used differentiable rendering with a physically-based model to recover the unknown object pose and environment lighting when the object was placed (or naturally occurred) in an image. We created a new dataset of materials and geometry for several common, shiny, curved objects along with images showing these in a variety of indoor and outdoor environments. We demonstrate that our approach strongly outperforms previous approaches in realism and fidelity.

**Acknowledgements.** We would like to thank William T. Freeman for the invaluable discussion and for the photo credit, Varun Jampani for helping us with data collection, and Henrique Weber and Jean-François Lalonde for running their methods as comparisons for us. The work was done in part when Hong-Xing Yu was a student researcher at Google and has been supported by gift funding and GCP credits from Google and Qualcomm.

## References


[1] https://www.einscan.com/handheld-3d-scanner/. 4

[2] https://remove.bg. 4

[3] Alexandru O Balan, Leonid Sigal, Michael J Black, James E Davis, and Horst W Haussecker. Detailed human shape and pose from images. In *CVPR*, 2007. 4, 7, 8

[4] James F Blinn. Models of light reflection for computer synthesized pictures. In *Proceedings of the 4th annual conference on Computer graphics and interactive techniques*, pages 192–198, 1977. 2

[5] Brent Burley and Walt Disney Animation Studios. Physically-based shading at disney. In *SIGGRAPH*, 2012. 3

[6] Dan A Calian, Jean-François Lalonde, Paulo Gotardo, Tomas Simon, Iain Matthews, and Kenny Mitchell. From faces to outdoor light probes. In *CGF*, 2018. 2

[7] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *ECCV*, 2020. 4

[8] Kristin J Dana, Bram Van Ginneken, Shree K Nayar, and Jan J Koenderink. Reflectance and texture of real-world surfaces. *ACM Transactions On Graphics (TOG)*, 18(1):1–34, 1999. 2

[9] Paul Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *SIGGRAPH*, 1998. 1, 2

[10] Jonathan Dupuy and Wenzel Jakob. An adaptive parameterization for efficient material acquisition and rendering. *ACM Transactions on graphics (TOG)*, 37(6):1–14, 2018. 2

[11] Graham D Finlayson, Roshanak Zakizadeh, and Arjan Gijsenij. The reproduction angular error for evaluating the performance of illuminant estimation algorithms. *IEEE transactions on pattern analysis and machine intelligence*, 39(7):1482–1488, 2016. 6

[12] Marc-André Gardner, Yannick Hold-Geoffroy, Kalyan Sunkavalli, Christian Gagné, and Jean-François Lalonde. Deep parametric indoor lighting estimation. In *ICCV*, 2019. 2, 5, 6, 7

[13] 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. *arXiv:1704.00090*, 2017. 2

[14] Mathieu Garon, Kalyan Sunkavalli, Sunil Hadap, Nathan Carr, and Jean-François Lalonde. Fast spatially-varying indoor lighting estimation. In *CVPR*, 2019. 1, 2, 5, 6, 7

[15] Cindy M Goral, Kenneth E Torrance, Donald P Greenberg, and Bennett Battaile. Modeling the interaction of light between diffuse surfaces. *ACM SIGGRAPH computer graphics*, 18(3):213–222, 1984. 2

[16] Roger Grosse, Micah K Johnson, Edward H Adelson, and William T Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *ICCV*, 2009. 6

[17] Bruce Hartung and Daniel Kersten. Distinguishing shiny from matte. *Journal of Vision*, 2002. 5

[18] Jon Hasselgren, Nikolai Hofmann, and Jacob Munkberg. Shape, light & material decomposition from images using monte carlo rendering and denoising. *arXiv:2206.03380*, 2022. 2, 4, 5, 6, 8

[19] Eric Heitz. Sampling the ggx distribution of visible normals. *Journal of Computer Graphics Techniques (JCGT)*, 2018. 4

[20] Henrik Wann Jensen, Stephen R Marschner, Marc Levoy, and Pat Hanrahan. A practical model for subsurface light transport. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 511–518, 2001. 2

[21] James T Kajiya. The rendering equation. In *SIGGRAPH*, 1986. 1, 2

[22] Soo Ye Kim, Jihyong Oh, and Munchurl Kim. Deep sr-itm: Joint learning of super-resolution and inverse tone-mapping for 4k uhd hdr applications. In *ICCV*, 2019. 8

[23] Sebastian B Knorr and Daniel Kurz. Real-time illumination estimation from faces for coherent rendering. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 113–122. IEEE, 2014. 2

[24] Samuli Laine, Janne Hellsten, Tero Karras, Yeongho Seol, Jaakko Lehtinen, and Timo Aila. Modular primitives for high-performance differentiable rendering. *ACM TOG*, 2020. 3, 4

[25] Siyeong Lee, Gwon Hwan An, and Suk-Ju Kang. Deep recursive hdri: Inverse tone mapping using generative adversarial networks. In *ECCV*, 2018. 8

[26] Chloe LeGendre, Wan-Chun Ma, Graham Fyffe, John Flynn, Laurent Charbonnel, Jay Busch, and Paul Debevec. Deeplight: Learning illumination for unconstrained mobile mixed reality. In *CVPR*, 2019. 6

[27] Chloe LeGendre, Wan-Chun Ma, Rohit Pandey, Sean Fanello, Christoph Rhemann, Jason Dourgarian, Jay Busch, and Paul Debevec. Learning illumination from diverse portraits. In *SIGGRAPH Asia Technical Communications*. 2020. 2

[28] Hendrik PA Lensch, Jan Kautz, Michael Goesele, Wolfgang Heidrich, and Hans-Peter Seidel. Image-based reconstruction of spatial appearance and geometric detail. *ACM Transactions on Graphics (TOG)*, 22(2):234–257, 2003. 2

[29] Yi Li, Gu Wang, Xiangyang Ji, Yu Xiang, and Dieter Fox. Deepim: Deep iterative matching for 6d pose estimation. In *ECCV*, 2018. 4

[30] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2475–2484, 2020. 2

[31] Stephen R Marschner, Stephen H Westin, Eric PF Lafortune, and Kenneth E Torrance. Image-based bidirectional reflectance distribution function measurement. *Applied optics*, 39(16):2592–2600, 2000. 2

[32] Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Müller, and Sanja Fidler. Extracting triangular 3d models, materials, and lighting from images. In *CVPR*, 2022. 2, 4, 5, 6, 8

[33] Ko Nishino and Shree K Nayar. Eyes for relighting. *ACM TOG*, 2004. 2

[34] Jinwoo Park, Hunmin Park, Sung-Eui Yoon, and Woontack Woo. Physically-inspired deep light estimation from a homogeneous-material object for mixed reality lighting. *IEEE TVCG*, 2020. 2

[35] Jeong Joon Park, Aleksander Holynski, and Steven M Seitz. Seeing the world in a bag of chips. In *CVPR*, 2020. 2

[36] Matt Pharr, Wenzel Jakob, and Greg Humphreys. *Physically based rendering: From theory to implementation*. Morgan Kaufmann, 2016. 3

[37] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R Zaiane, and Martin Jagersand. U2-net: Going deeper with nested u-structure for salient object detection. *PR*, 2020. 4

[38] Ravi Ramamoorthi and Pat Hanrahan. A signal-processing framework for inverse rendering. In *SIGGRAPH*, 2001. 1, 3, 5

[39] Thomas Richter-Trummer, Denis Kalkofen, Jinwoo Park, and Dieter Schmalstieg. Instant mixed reality lighting from casual scanning. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 27–36. IEEE, 2016. 2

[40] Christophe Schlick. An inexpensive brdf model for physically-based rendering. In *Computer graphics forum*, 1994. 3

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

[42] Soumyadip Sengupta, Jinwei Gu, Kihwan Kim, Guilin Liu, David W. Jacobs, and Jan Kautz. Neural inverse rendering of an indoor scene from a single image. In *International Conference on Computer Vision (ICCV)*, 2019. 2

[43] Shuran Song and Thomas Funkhouser. Neural illumination: Lighting prediction for indoor environments. In *CVPR*, 2019. 2

[44] Pratul P Srinivasan, Ben Mildenhall, Matthew Tancik, Jonathan T Barron, Richard Tucker, and Noah Snavely. Lighthouse: Predicting lighting volumes for spatially-coherent illumination. In *CVPR*, 2020. 1, 2

[45] Robin Strudel, Ricardo Garcia, Ivan Laptev, and Cordelia Schmid. Segmenter: Transformer for semantic segmentation. In *ICCV*, 2021. 4

[46] Tiancheng Sun, Jonathan T Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul E Debevec, and Ravi Ramamoorthi. Single image portrait relighting. *ACM TOG*, 2019. 1, 2

[47] Kenneth E Torrance and Ephraim M Sparrow. Theory for off-specular reflection from roughened surfaces. *Josa*, 1967. 2, 3

[48] Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. Microfacet models for refraction through rough surfaces. In *Proceedings of the 18th Eurographics conference on Rendering Techniques*, 2007. 3

[49] Guangcong Wang, Yinuo Yang, Chen Change Loy, and Ziwei Liu. Stylelight: Hdr panorama generation for lighting estimation and editing. In *ECCV*, 2022. 1, 5, 6, 7

[50] He Wang, Srinath Sridhar, Jingwei Huang, Julien Valentin, Shuran Song, and Leonidas J Guibas. Normalized object coordinate space for category-level 6d object pose and size estimation. In *CVPR*, 2019. 4

[51] Zian Wang, Wenzheng Chen, David Acuna, Jan Kautz, and Sanja Fidler. Neural light field estimation for street scenes with differentiable virtual object insertion. In *ECCV*, 2022. 1

[52] Zian Wang, Jonah Philion, Sanja Fidler, and Jan Kautz. Learning indoor inverse rendering with 3d spatially-varying lighting. In *Proceedings of International Conference on Computer Vision (ICCV)*, 2021. 2

[53] Henrique Weber, Donald Prévost, and Jean-François Lalonde. Learning to estimate indoor lighting from 3d objects. In *3DV*, 2018. 2

[54] Xin Wei, Guojun Chen, Yue Dong, Stephen Lin, and Xin Tong. Object-based illumination estimation with rendering-aware neural networks. In *ECCV*, 2020. 2

[55] Yu Xiang, Tanner Schmidt, Venkatraman Narayanan, and Dieter Fox. Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes. *arXiv:1711.00199*, 2017. 4

[56] Renjiao Yi, Chenyang Zhu, Ping Tan, and Stephen Lin. Faces as lighting probes via unsupervised deep highlight extraction. In *ECCV*, 2018. 2

[57] Ye Yu and William AP Smith. Inverserendernet: Learning single image inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2

[58] Fangneng Zhan, Changgong Zhang, Wenbo Hu, Shijian Lu, Feiying Ma, Xuansong Xie, and Ling Shao. Sparse needlets for lighting estimation with spherical transport loss. In *ICCV*, 2021. 2

[59] Fangneng Zhan, Changgong Zhang, Yingchen Yu, Yuan Chang, Shijian Lu, Feiying Ma, and Xuansong Xie. Emlight: Lighting estimation via spherical distribution approximation. In *AAAI*, 2021. 2

[60] Kai Zhang, Fujun Luan, Zhengqi Li, and Noah Snavely. Iron: Inverse rendering by optimizing neural sdfs and materials from photometric images. In *CVPR*, 2022. 2, 4

[61] Yuxuan Zhang, Wenzheng Chen, Huan Ling, Jun Gao, Yinan Zhang, Antonio Torralba, and Sanja Fidler. Image gans meet differentiable rendering for inverse graphics and interpretable 3d neural rendering. In *International Conference on Learning Representations*, 2021. 2

[62] Michael Zollhöfer, Patrick Stotko, Andreas Görlitz, Christian Theobalt, Matthias Nießner, Reinhard Klein, and Andreas Kolb. State of the art on 3d reconstruction with rgb-d cameras. In *Computer graphics forum*, volume 37, pages 625–652. Wiley Online Library, 2018. 2