# Supplementary Material for 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

*In this supplementary document,* (A) *we provide more details in reconstructing our accidental light probes (ALPs),* (B) *include more qualitative results in both indoor and outdoor scenes, and* (C) *show additional evaluations including comparison to existing inverse rendering methods in a casual capture setup, additional ALPs for a controlled analysis of shape and reflectance, and evaluation on difficult object poses. In our* **supplementary video***, we provide an overview of our work.*

## A. Technical Details

**Losses in accidental light probe reconstruction.** Recall that for ALP reconstruction, our optimization problem is given by:

$$\min_{\pi,\alpha,f} \sum_{\{I_{\text{capture}}\}} \mathcal{L}\left(I_{\text{render}}(\pi,\alpha,f|L_i,S), I_{\text{capture}}\right), \tag{1}$$

where $\pi$ and $\alpha$ are the 6D pose and size to fit the scanned shape $S$ to multi-view camera coordinate frame, $f$ is the material parameterized by spatially-varying albedo $A$ and roughness $r$, $\{I_{\text{capture}}\}$ denotes a set of captured multi-view images, $I_{\text{render}}$ denotes the rendered image, and $L_i$ denotes the recorded incidental lighting. The loss function $\mathcal{L}$ consists of three parts:

$$\mathcal{L} = \mathcal{L}_{\text{RGB}} + \mathcal{L}_{\text{mask}} + \beta_1\,\mathcal{L}_{\text{albedo-reg}} + \beta_2\,\mathcal{L}_{\text{roughness-reg}}, \tag{2}$$

where $\mathcal{L}_{\text{RGB}}$ and $\mathcal{L}_{\text{mask}}$ are pixel-wise losses for RGB images and masks, and $\mathcal{L}_{\text{albedo-reg}}$ and $\mathcal{L}_{\text{roughness-reg}}$ denotes regularization for the material. For $\mathcal{L}_{\text{RGB}}$, we follow [4] to use $L_1$ loss in tone-mapped image space with sRGB transfer function [8] for stable optimization. We use $L_2$ loss for $\mathcal{L}_{\text{mask}}$. Similar to [4, 10], the material regularization for smoothness is defined as

$$\mathcal{L}_{\text{albedo-reg}} = \|A(x) - A(x+\Delta x)\|_1, \quad \mathcal{L}_{\text{roughness-reg}} = \|r(x) - r(x+\Delta x)\|_1 + \|r(x)\|_1, \tag{3}$$

where $x$ denotes a point at the object surface, and $\Delta x$ denotes a small deviation sampled from a normal distribution. We also encourage small roughness values since any inaccuracy (which can come from the violation of distant lighting assumption, inaccuracy of surface normals due to quantization in mesh triangles, etc.) in computing specular highlight reflection results in blurred highlights and thus higher estimation of roughness of the surface. As for shading, we also adopt the differentiable split-sum approximation [4] to shade the multi-view rendered images due to performance reason.

**Implementation details.** For ALP reconstruction, we use Adam [5] optimizer with default momentum parameters and learning rate 0.03. We optimize for 1500 iterations with 8 images per batch. We set $\beta_1 = 0.03$. The training images are cropped to be square and the resolution is 1000. We use $\beta_2 = 0.01$ for all soda cans and 0.1 for others. For joint pose-lighting estimation, we adopt a two-stage strategy. In the first stage, we use Adam optimizer with learning rate 0.1 for 500 iterations. We set the pose regularization weight $\lambda_1 = 1$ and the lighting regularization weight $\lambda_2 = 0.01$. For Monte Carlo estimation, we use 1 sample for shading with a cross-bilateral denoiser [7]. The small sample count is due to limited GPU memory. We found that the denoiser stabilizes the joint optimization but it introduces stripe-like artifacts in the environment map. Thus, in the second stage, we fix the 6D pose and keep optimizing the environment map for additional 1000 iterations using 200000 rays per iteration without the denoiser. We continuously sample sub-pixels (which provides continuously-changing surface normals) to remove alias in the environment map.

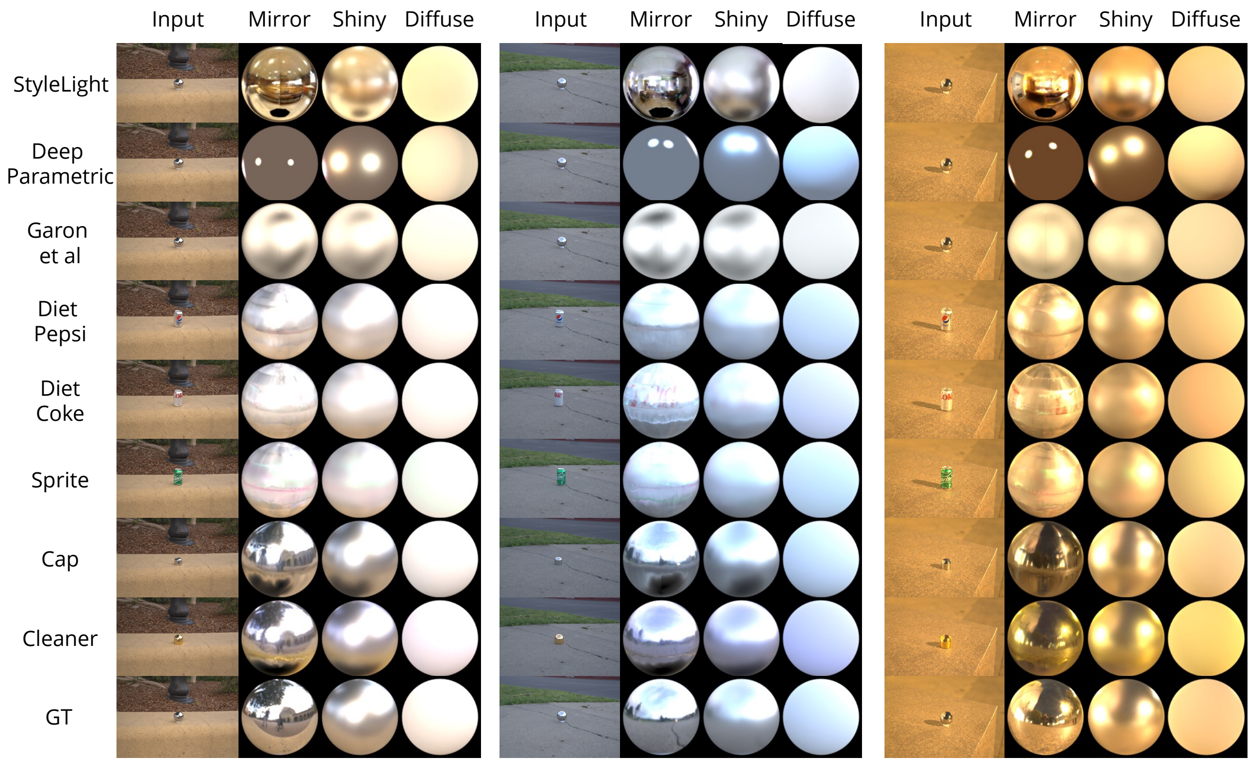


Figure 1. Additional qualitative results for outdoor scenes, comparing to StyleLight [9], Deep Parametric [2] and Garon et al [3].

## B. Additional Qualitative Results

We show qualitative results of our lighting estimation and relighting in Figure 1 and Figure 2 for outdoor and indoor scenes respectively. Similar to the observation in the main paper, our results are visually more consistent with the groundtruth relighted images compared to the baseline methods.

## C. Additional Evaluations

**Controlled experiments on reflectance and shapes of ALPs.** To explore how different reflectance properties and shapes of ALPs affect lighting estimation, we include four additional ALPs for controlled comparisons. We show these additional ALPs for controlled experiments in Fig. 3. Regarding reflectance, the estimated environment maps are more accurate and detailed when the ALPs are less rough, e.g., comparing Ornaments 1 and 2, while the albedo has little effect, e.g., comparing Ornaments 1 and 3. Regarding shape, although the golden bird statue is highly non-convex, it provides comparable results to the golden Ornament 1 which is a nearly perfect sphere, showing that our method can tolerate such shape variation.

**Dealing with extreme poses.** In our joint light-pose optimization, we address local minima by multi-initialization of orientation (currently 4) and keeping the one with the highest re-rendering PSNR. We show various poses for the Diet Pepsi can in Fig. 4. This combination of sampling-based and gradient-based methods leads to robust pose estimates, even for unusual poses like an upside-down can.

**Note on the ring.** We note that we also tried using a ring as an ALP. However, it is very challenging to have correct segmentation of the ring using the automatic segmentation algorithm [1], possibly due to its small size and highly concave shape. For the ring example we show in the main paper, we manually segment it and then use our method to estimate lighting.

**ALP reconstruction in casual capture setup.** Beside the light box capture setup, we also compare the ALP reconstruction using a casual setup, where a Diet Coke can is placed on a table (with some textured newspaper around it for solving camera poses) and we take multi-view photos for it. Example images of the casual setup photos are shown in Figure 5. We extend the

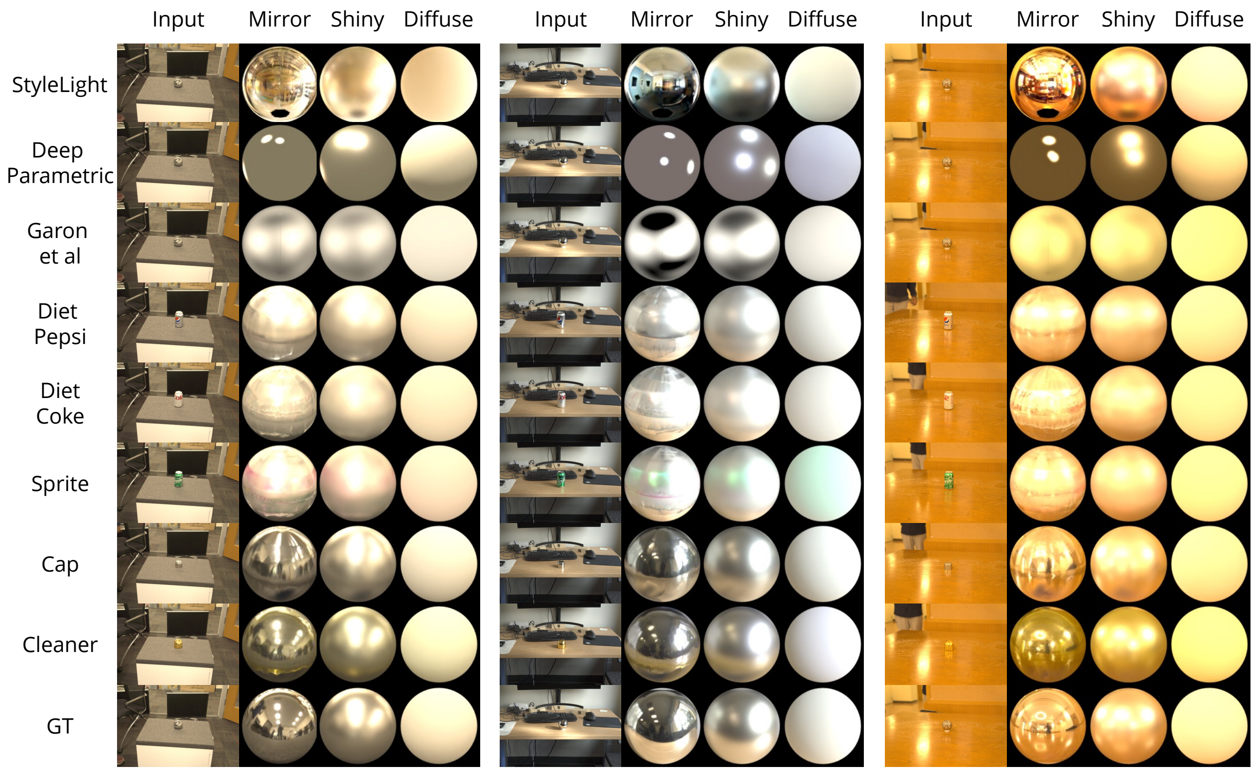


Figure 2. Additional qualitative results for indoor scenes. We compare our approach to StyleLight [9], Deep Parametric [2] and Garon et al [3] on relighting mirror, shiny and diffuse spheres.

| Method | Casual multi-view | | | Lightbox setup | | |
|---|---|---|---|---|---|---|
| | Mirror | Shiny | Diffuse | Mirror | Shiny | Diffuse |
| Nvdiffrec [6] | 12.23 | 10.99 | 9.14 | 6.99 | 5.06 | 3.59 |
| Nvdiffrecmc [4] | 8.76 | 7.37 | **4.71** | 6.55 | 4.60 | 3.84 |
| ALP (Ours) | **6.49** | **5.48** | 4.89 | **5.46** | **3.67** | **2.80** |

Table 1. Evaluation on our ALP model acquisition for a diet coke can under two different capture settings: casual multi-view images and our lightbox setup. We use the same lighting estimation approach for compared methods and report average angular error across all test scenes.

lightbox setup comparison from the main paper in Table 1 to include quantitative evaluation for lighting estimation using the reconstructed ALPs in the casual capture setup. In both setups, our reconstruction yields better lighting estimation.

While in the main paper we show a comparison for existing state-of-the-art inverse rendering methods [4, 6] and our reconstruction pipeline in a nutshell, here we show a more complete comparison of the reconstruction using the existing methods and ours under both the casual setup and the light box setup in Figure 6. We can see that our reconstruction method is consistently better under both casual setup and light box setup. We also note that our reconstruction using the casual setup still has the environment baked into the material textures. This is probably because the distant lighting assumption breaks due to near-field reflections. The incidental lighting on the ALP surface is spatially-varying instead of constant. Thus, on the bottom of the coke can, the actual incidental surface light field (where the reflected rays are yellowish) is different from the environment map, leading to yellowish texture baking.

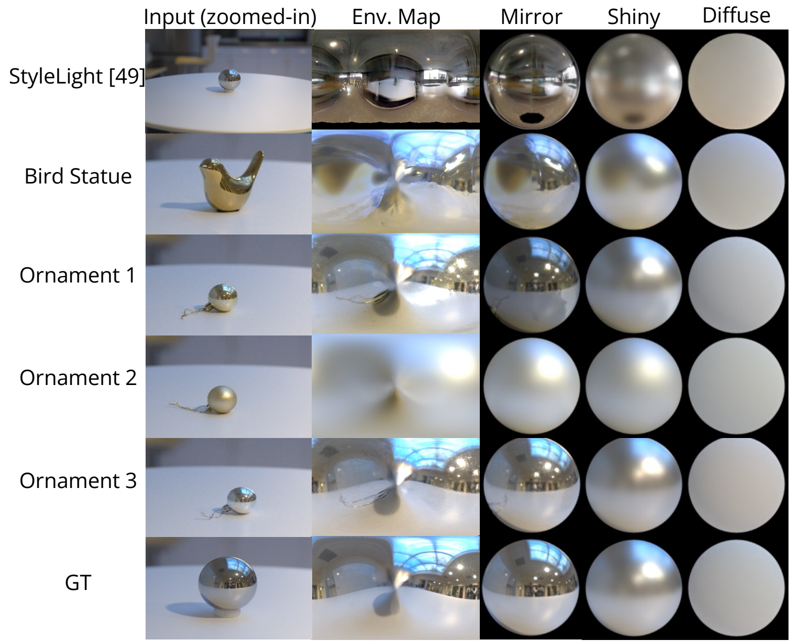


Figure 3. ALPs of different shapes (a golden Bird Statue vs. golden Ornament 1) and reflectance (Ornaments: 1 and 2 are different in roughness; 1 and 3 are different in albedo).


Figure 4. Estimated poses under varying pose conditions. The second row shows re-renderings (which have black backgrounds) overlayed on top of input images. They match accurately.

Figure 5. Casual multi-view images for ALP reconstruction. We compare our ALP reconstruction pipeline to [4, 6] using casually captured multi-view images of a diet coke can in contrast to using our lightbox setup that is referenced in the main paper.

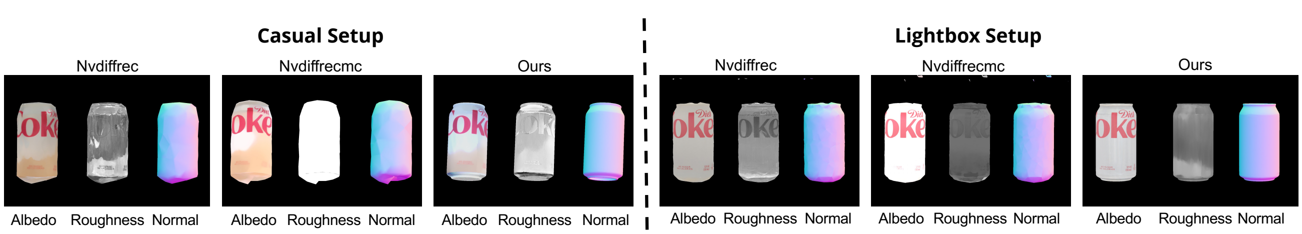


Figure 6. Qualitative comparison of the materials and geometry of a Diet Coke can using our method and [4, 6] in both setups.

## References


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

[2] Marc-André Gardner, Yannick Hold-Geoffroy, Kalyan Sunkavalli, Christian Gagné, and Jean-François Lalonde. Deep parametric indoor lighting estimation. In *ICCV*, 2019. 2, 3

[3] Mathieu Garon, Kalyan Sunkavalli, Sunil Hadap, Nathan Carr, and Jean-François Lalonde. Fast spatially-varying indoor lighting estimation. In *CVPR*, 2019. 2, 3

[4] Jon Hasselgren, Nikolai Hofmann, and Jacob Munkberg. Shape, light & material decomposition from images using monte carlo rendering and denoising. *arXiv:2206.03380*, 2022. 1, 3, 4

[5] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv:1412.6980*, 2014. 1

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

[7] Christoph Schied, Anton Kaplanyan, Chris Wyman, Anjul Patney, Chakravarty R Alla Chaitanya, John Burgess, Shiqiu Liu, Carsten Dachsbacher, Aaron Lefohn, and Marco Salvi. Spatiotemporal variance-guided filtering: real-time reconstruction for path-traced global illumination. In *Proceedings of High Performance Graphics*, 2017. 1

[8] Michael Stokes. A standard default color space for the internet-srgb. *http://www. color. org/contrib/sRGB. html*, 1996. 1

[9] Guangcong Wang, Yinuo Yang, Chen Change Loy, and Ziwei Liu. Stylelight: Hdr panorama generation for lighting estimation and editing. In *ECCV*, 2022. 2, 3

[10] Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM TOG*, 2021. 1