

Supplemental Materials for DeepLight: Learning Illumination for Unconstrained Mobile Mixed Reality

Chloe LeGendre^{1,2 *} Wan-Chun Ma¹ Graham Fyffe¹ John Flynn¹

Laurent Charbonnel¹ Jay Busch¹ Paul Debevec¹

¹Google Inc. ²USC Institute for Creative Technologies

1. Sphere Image Extraction Method

Our reference spheres seen in the training data are designed to be held in a fixed position in the camera’s FOV, but in practice their position wanders and jitters somewhat during data collection. To address this, we extract images of the three spheres for each video frame as follows. We compute optical flow for each frame against the following frame, and compute the gradient of the flow field. We modulate the gradient magnitude by 0.1^θ^2 where θ is the angle in radians between the gradient vector and the direction to a manually marked sphere center (within 5 pixels of the actual center). This produces a noisy image with bright rings around the spheres due to the difference in flow between the spheres and the background. We find the highest confidence bright circles in this image using template matching within a tolerance of the marked spheres and over a range of radii, and average the detected positions and radii across all frames of each video to eliminate jitter.

2. Additional Quantitative Results

Table 1. Average L_1 loss by BRDF: diffuse (d), mirror (m), and matte silver (s), and RGB angular error θ_{rgb} for diffuse (columns), for our network trained with different loss terms (rows). We compare ground truth images with those *rendered* using our HDR lighting inference, for *seen* indoor and outdoor locations.

Loss terms	$L_1(d)$		$L_1(s)$		$L_1(m)$		$\theta_{\text{rgb}}^{\circ}(d)$	
	SI	SO	SI	SO	SI	SO	SI	SO
$L_1(m,d,s) + L_{\text{adv}}$	0.11	0.09	0.12	0.10	0.18	0.14	8.0	10.4
$L_1(m,d,s)$	0.11	0.08	0.12	0.10	0.15	0.12	8.0	10.6
$L_1(m)$	0.25	0.15	0.18	0.13	0.14	0.11	8.6	13.2
$L_1(s)$	0.14	0.10	0.12	0.10	0.23	0.18	7.9	11.3
$L_1(d)$	0.11	0.08	0.14	0.12	0.30	0.25	7.9	10.4

3. Performance Evaluation

We report speed and accuracy for networks trained to predict lower resolution lighting (8×8 or 16×16) in Table 2.

*Work completed while interning at Google.

We report the lighting inference time on a variety of mobile devices in Table 3.

Table 2. Average inference time on a NVIDIA Quadro K1200 GPU (N) and Google Pixel 3 XL mobile CPU (P) and L_1 loss by BRDF: diffuse (d), mirror (m), and matte silver (s), and RGB angular error θ_{rgb} for diffuse (columns), for variants of our network trained for different output lighting resolutions [$n \times n$] and size of latent representations. Our baseline network is $[32 \times 32]$ -256.

Network	$L_1(d)$		$L_1(s)$		$L_1(m)$		$\theta_{\text{rgb}}^{\circ}(d)$		Time (ms)	
	UI	UO	UI	UO	UI	UO	UI	UO	N	P
Baseline	0.12	0.13	0.13	0.13	0.17	0.16	9.8	10.8	6.53	80.0
$[16 \times 16]$ -256	0.11	0.08	0.12	0.10	0.17	0.13	6.5	8.9	6.06	62.7
$[16 \times 16]$ -128	0.22	0.13	0.13	0.13	0.17	0.15	8.5	9.9	5.93	61.4
$[8 \times 8]$ -128	0.11	0.12	0.12	0.12	0.13	0.13	14.1	15.6	5.76	50.9

Table 3. Average inference time on various mobile devices / Qualcomm Snapdragon systems-on-a-chip (QS SoCs) CPUs for our standard (baseline) lighting estimation network.

Mobile device	QS SoC	Time (ms)	Rate (fps)
ASUS ROG	845 (2.96 GHz Kryo 385)	57.1	17.5
Samsung Galaxy S9+	845 (2.8 GHz Kryo 385)	59.8	16.7
Google Pixel XL	821 (2.15 GHz Kryo) [†]	57.8	17.3
Google Pixel 2 XL	835 (2.35 GHz Kryo 280)	102.1	9.8
Google Pixel 3 XL	845 (2.5 GHz Kryo 385)	80.0	12.5

4. Mobile Demonstration App

We develop a mobile app to demonstrate one way to use the proposed lighting inference for rendering virtual objects. The app is based on Google’s Android framework and uses the augmented reality platform ARCore for plane finding and virtual object anchoring.

The lighting inference provides omnidirectional illumination, and there are many possible ways to use this inference for rendering. For example, each pixel in the HDR lighting map is essentially a directional light, and we can simply use all pixels to light a virtual object. However this would be time-consuming as a 32×32 illumination map

[†]QS 821 has been shown to out-perform QS 835 for single-threaded floating point operations [1].

has already more than 800 directional lights. This trivial method is impractical for shadow computation as well, as it would require the same number of shadow buffer textures [11] for shadow mapping. Thus, in our demonstration app we employ a well-known technique of pre-computed radiance transfer [10] (PRT) for approximating the global illumination computation. As the name suggests, PRT pre-computes lighting and transfer functions in the frequency domain. By doing so, the evaluation of the rendering equation can be reduced to a dot product between the lighting and transfer function coefficients. A typical choice for the frequency series is spherical harmonics (SH). We encode both functions with the first nine SH coefficients per color channel, up to 2nd order SH. We also implement environment mapping [2] for realistic sharp specular, mirror-like reflections, and blend these sharp reflections with the PRT rendered result [6]. Other real-time rendering techniques could also be employed, such as irradiance mapping for dynamic, animated objects [8]. Furthermore, our generic lighting representation could be converted to a set of n point or directional light sources for real-time rendering, using a sampling technique as in [4].

We choose *Lullaby*, an open source graphics library for mobile devices, as the rendering engine for our demo application. For every input image from the live camera feed, the demo app uses the neural network to infer the lighting environment in camera space. Subsequently, the SH coefficients of the inferred lighting are computed on CPU and then fed to the GPU. Finally, the SH PRT is evaluated in the fragment shader pass. See the supplemental video for several live recordings of the demonstration app.

5. Off-line Rendering Technique

For our virtual object relighting comparisons (main paper Fig. 1 and 7, and supplemental materials Fig. 4 and 5), we use global illumination (GI) rendering techniques and image-based lighting (IBL) [3], using the Arnold renderer. For all images, we use Arnold’s aiStandardSurface shader for both the diffuse and matte silver virtual bunnies, shader parameters in Tables 4 and 5 respectively. These parameters were selected to best visually match the reflectance field basis images for spheres coated with the same paint as the real, 3D printed bunnies. For these off-line renders, virtual objects were set on a virtual planar surface manually placed by an artist to best match the background plate, after first setting the virtual camera’s approximate FOV. We use the Arnold aiShadowMatte shader for the virtual planes, with parameters in Table 6, which allows virtual objects to cast shadows onto them. Furthermore, this shader enables indirect diffuse and specular shading of virtual objects using colors sampled from the background image at the location where the placed virtual plane projects into the background image. Thus, the light bouncing off the virtual plane

is tinted the color of the background image, which is then used to light the underside of any virtual object placed on the surface.

Table 4. Diffuse gray object aiStandardSurface shader parameters.

parameters	values
base weight	0.700
base color RGB	[0.649, 0.700, 0.700]
specular weight	0.300
specular roughness	0.600
IOR	1.7

Table 5. Matte silver object aiStandardSurface shader parameters.

parameters	values
base weight	1.000
base color RGB	[0.588, 0.600, 0.600]
metalness	1.000
specular roughness	0.470

Table 6. Virtual plane object aiShadowMatte shader parameters.

parameters	values
use background image	on
shadow color RGB	[0.000, 0.000, 0.000]
shadow opacity	1.000
diffuse indirect	on
diffuse use background image	on
diffuse intensity	1.000
specular indirect	on
specular intensity	0.200
specular roughness	0.200
IOR	1.5

6. Additional Comparisons with Previous Work

Extending the qualitative comparisons with the state-of-the-art techniques, in Fig. 2 we show examples of *unseen indoor* input images and spheres of three BRDFs (diffuse, matte silver, and mirror) rendered using image-based relighting with our measured reflectance bases and our lighting inference and that of Gardner et al. [5], along with the ground truth sphere appearances. In Fig. 3 we show *unseen outdoor* input images, along with ground truth sphere images and renderings produced using our lighting inference and that of Hold-Geoffroy et al. [7].

Extending the virtual object relighting comparisons, in Fig. 4 we show *unseen indoor* input images (Fig. 4a) along with photographs of 3D-printed “real” bunnies, painted with two different finishes: diffuse (Fig. 4b) and matte silver (Fig. 4f). Using GI rendering with IBL [3], we render a virtual object using a ground truth HDR panorama captured by photographing a mirror sphere with multiple exposures using a Canon 5D Mark III DSLR camera (Fig. 4c and g), and using our lighting inference (Fig. 4d and h) and that of Gardner et al. [5](Fig. 4d and i). In Fig. 5, we show *unseen outdoor* input images and the same set of comparisons, except that we compare with the lighting inference

of Hold-Geoffroy et al. [7]. For the ground truth HDR panoramas, missing sun intensity was reconstructed using the techniques of [9]. These extended sets of renderings demonstrate that inferring lighting that generates accurate hard shadows is challenging for all approaches, including ours, indicating another avenue for future work.

Perceptual User Study. We presented 40 study participants with image triples: a *reference image* of a real, 3D-printed bunny in a scene (center), flanked by two rendered bunny composites. These were selected from three categories: bunnies rendered using our lighting inference, inference from the previous state-of-the-art [5]/[7], and ground truth HDR image-based lighting (IBL). Images were from paper Figs. 1 and 7, and supp. Figs. 4 and 5, representing bunnies of two BRDFs in 17 environments. Participants were asked “Which bunny looks more like the one in the reference image?”, choosing between {ours vs. [5]/[7]}, {ours vs. GT HDR IBL}, or {[5]/[7] vs. GT HDR IBL} ($n=34$ each, question order and left/right image randomized). For ours vs. [5]/[7], 69% of responses indicated that renders using our lighting estimates looked more like the reference, with 38% indicating “much more” (Fig. 1 top). For ours vs. GT HDR IBL, 29% of responses indicated renders using lighting inference looked equally like or more like the reference, while for [5]/[7] vs. GT HDR IBL, 5% indicated as such (Fig. 1 bottom).

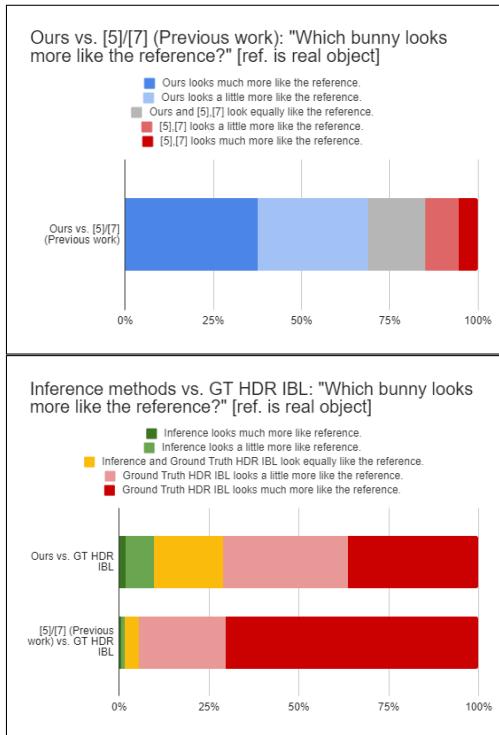
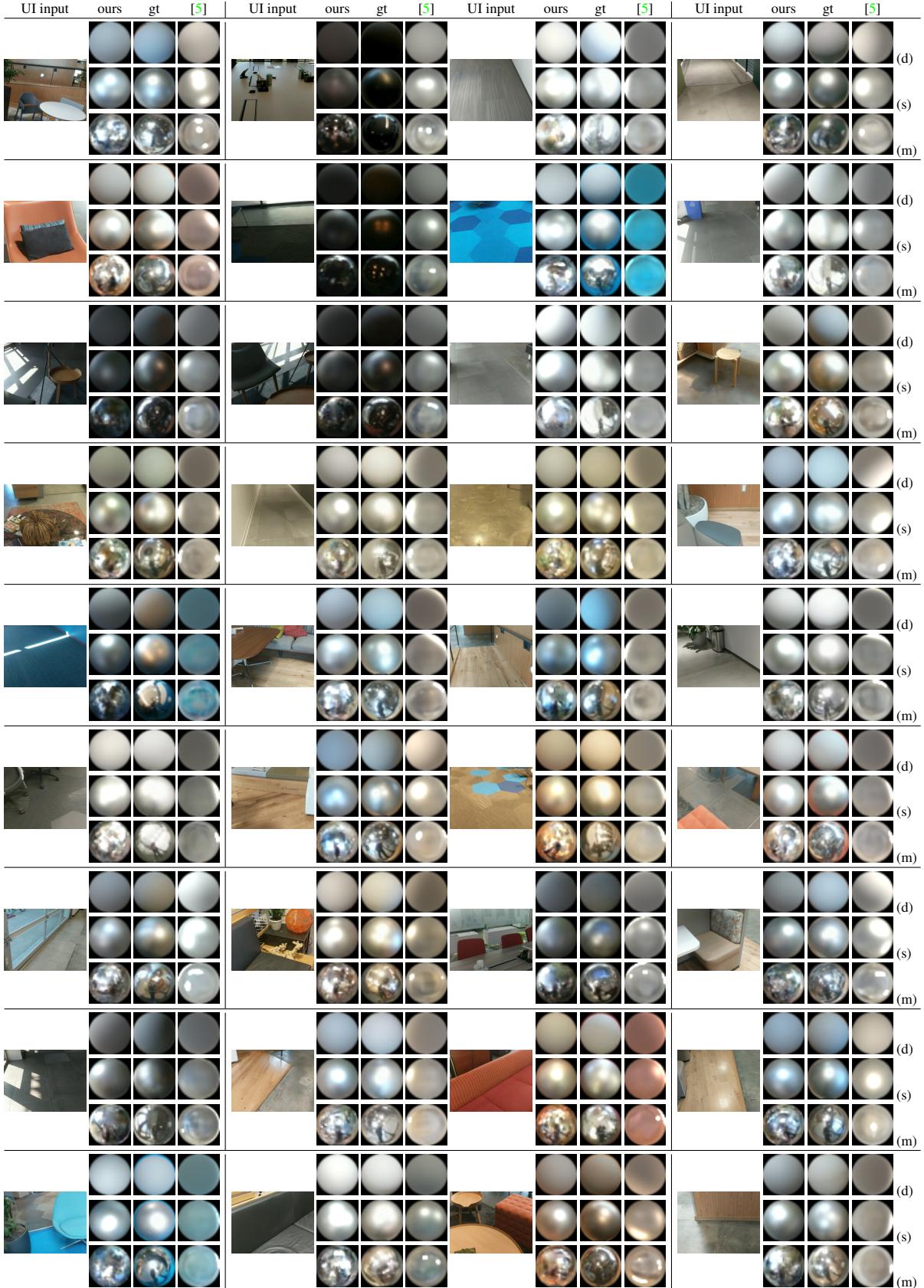
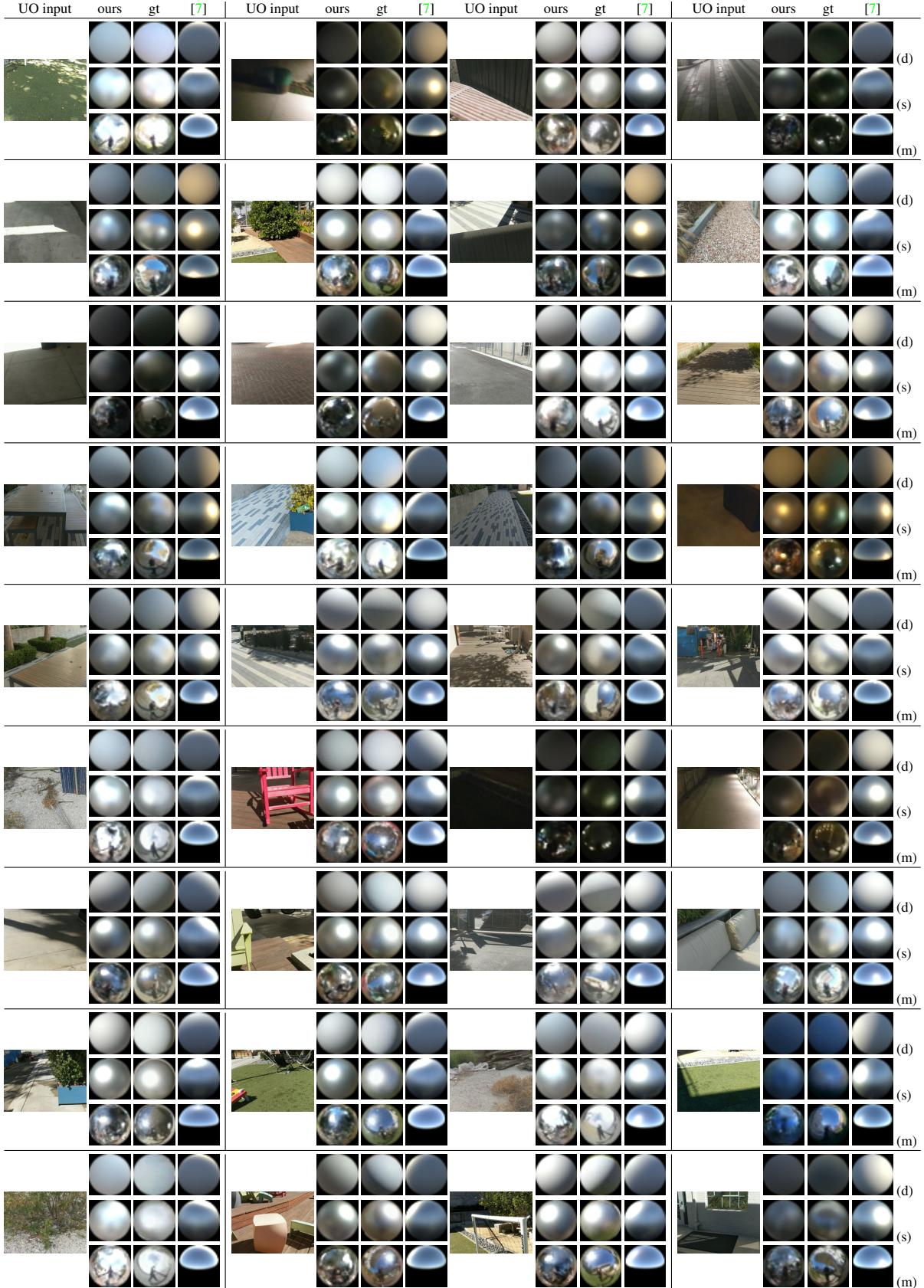


Figure 1. Results from user study (percentages of responses).

References

- [1] Qualcomm snapdragon 835 performance preview. <https://www.anandtech.com/show/11201/qualcomm-snapdragon-835-performance-preview/>. Accessed: 2018-11-18. 1
- [2] J. F. Blinn and M. E. Newell. Texture and reflection in computer generated images. *Communications of the ACM*, 19(10):542–547, 1976. 2
- [3] P. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 189–198. ACM, 1998. 2
- [4] P. Debevec. A median cut algorithm for light probe sampling. In *ACM Siggraph 2005 Posters*, page 66. ACM, 2005. 2
- [5] M.-A. Gardner, K. Sunkavalli, E. Yumer, X. Shen, E. Gambaretto, C. Gagné, and J.-F. Lalonde. Learning to predict indoor illumination from a single image. *ACM Trans. Graph.*, 36(6):176:1–176:14, Nov. 2017. 2, 3, 4, 6
- [6] R. Green. Spherical harmonic lighting: The gritty details. In *Archives of the Game Developers Conference*, volume 56, page 4, 2003. 2
- [7] Y. Hold-Geoffroy, K. Sunkavalli, S. Hadap, E. Gambaretto, and J.-F. Lalonde. Deep outdoor illumination estimation. In *IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, 2017. 2, 3, 5, 7
- [8] R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment maps. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH ’01, pages 497–500. ACM, 2001. 2
- [9] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010. 3
- [10] P.-P. Sloan, J. Kautz, and J. Snyder. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. *ACM Trans. Graph.*, 21(3):527–536, July 2002. 2
- [11] L. Williams. Casting curved shadows on curved surfaces. In *Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH ’78, pages 270–274, New York, NY, USA, 1978. ACM. 2





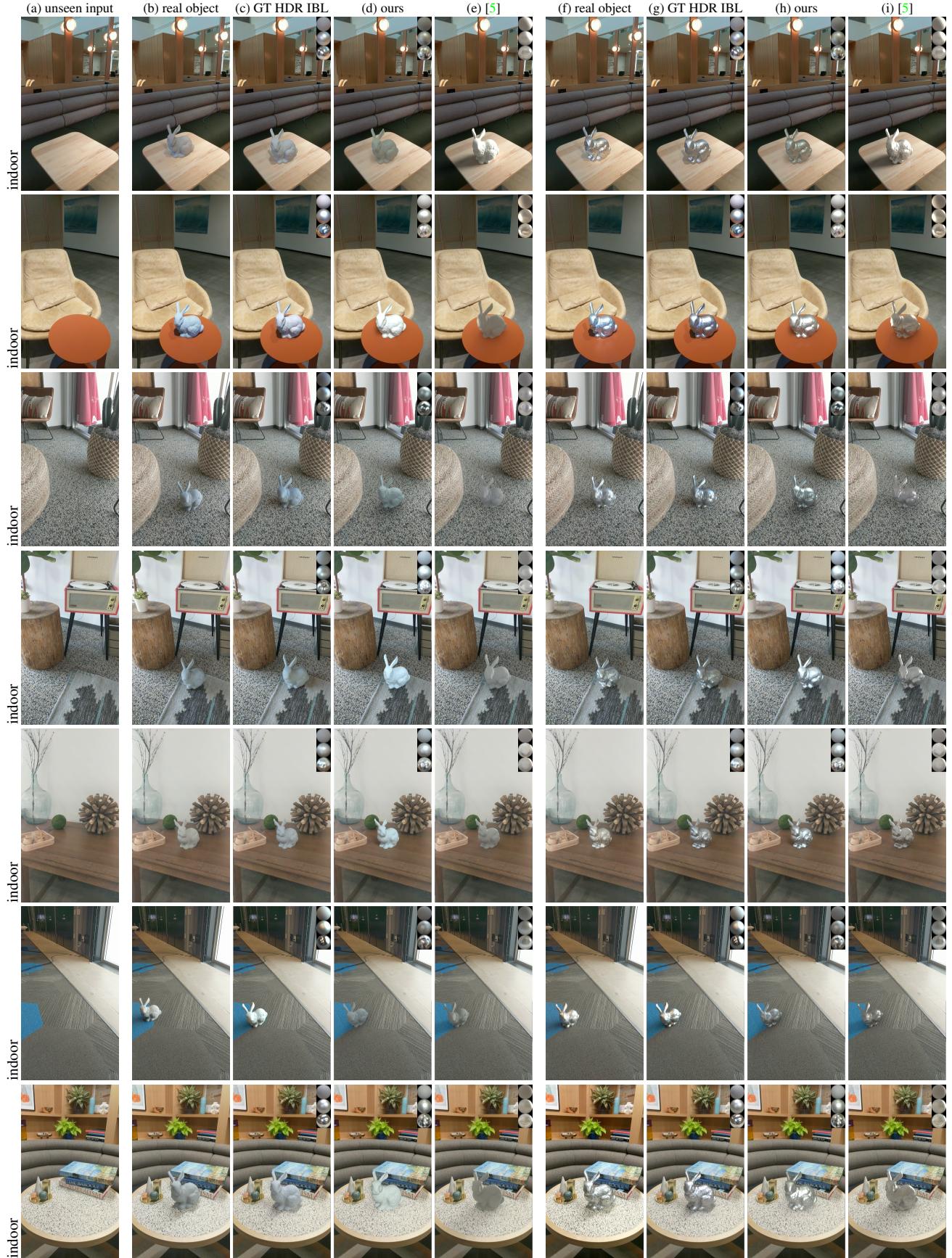


Figure 4. For each input image (a), we photograph a real 3D-printed bunny placed in the scene for two different BRDFs (b, f) and capture ground truth HDR panoramas at the bunny's location. Using GI rendering with IBL, we render a virtual bunny into the scene using ground truth lighting (c, g), our lighting inference (d, h), and that of the state-of-the-art methods for indoor [5] scenes (e, i).

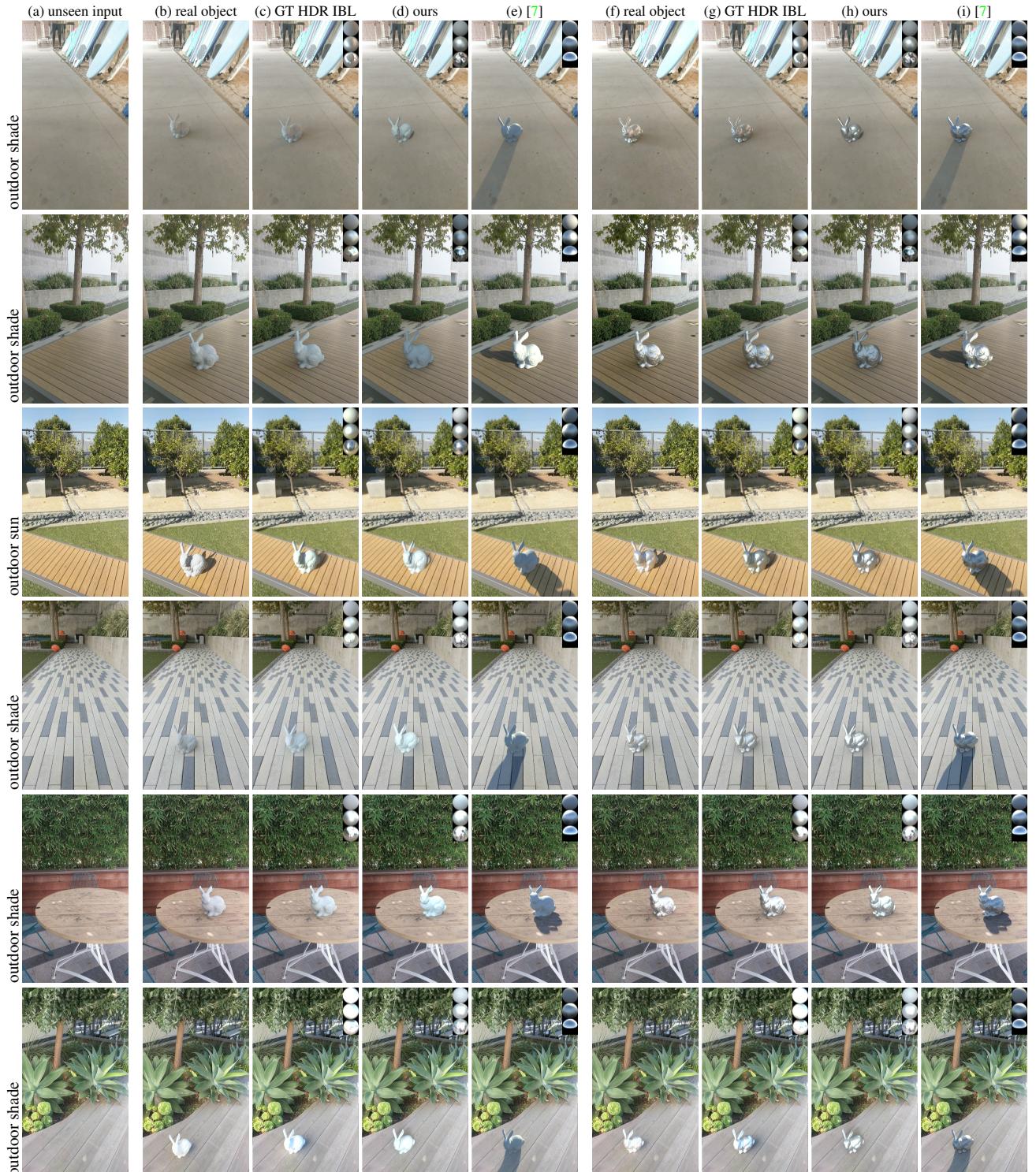


Figure 5. For each input image (a), we photograph a real 3D-printed bunny placed in the scene for two different BRDFs (b, f) and capture ground truth HDR panoramas at the bunny's location. Using GI rendering with IBL, we render a virtual bunny into the scene using ground truth lighting (c, g), our lighting inference (d, h), and that of the state-of-the-art methods for outdoor [7] scenes (e, i).