No Shadow Left Behind
Supplementary Material

Edward Zhang[1] Ricardo Martin-Brualla[2] Janne Kontkanen[2] Brian Curless[1,2]

[1]University of Washington [2]Google

1. Example Proxy Model for a Real Scene

Figure 1 shows an example of the proxy model for a real scene. The proxy geometry and the images come from a single frame captured using the RGBD Kinect v2 device mounted on a tripod. The approximate lighting was captured using a Ricoh Theta S 360 camera with 5 exposures for high dynamic range placed approximately in the center of the scene, roughly pointed at the Kinect.

2. Example of a Synthetic Scene

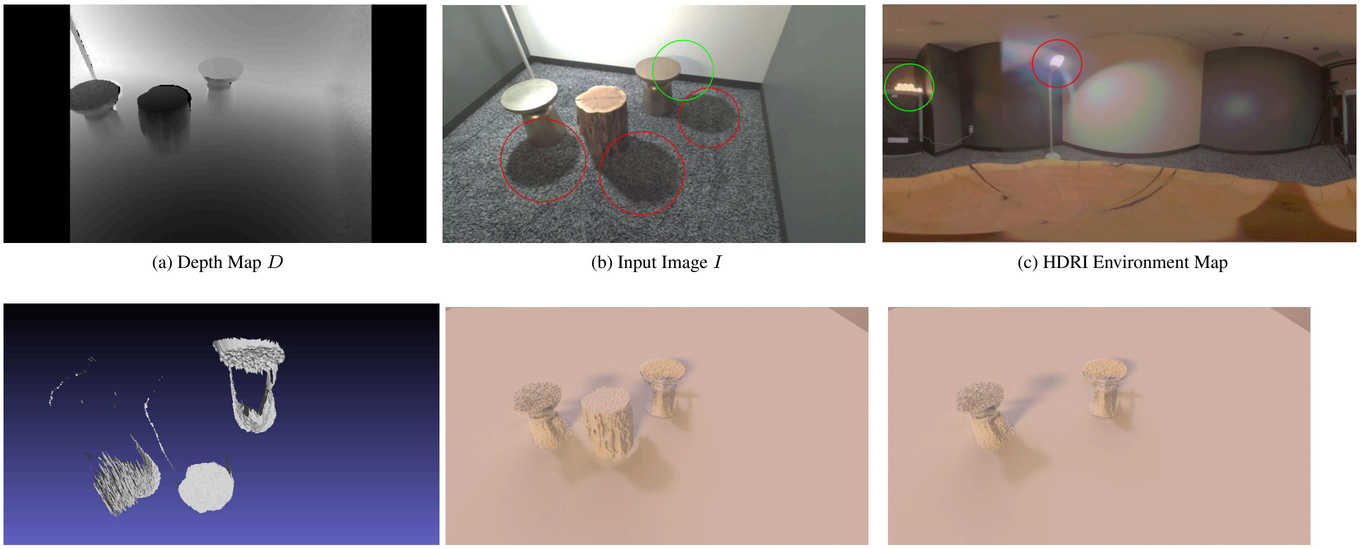
Figure 2 shows a full example of a synthetic scene, including the scene proxy. This scene is part of the test set.

3. Examples of Predicted Intermediates

We show the intermediates predicted by our network for the real and synthetic scenes from Figures 1-2. These intermediates include the predicted shadow mask, the predicted texture and lighting, and the predicted target lighting. For the synthetic scenes, these can be compared to the ground truth intermediates in Figure 2.

4. Comparison to Differential Rendering

Debevec's [2] differential rendering method does not apply directly to shadow removal, since a proxy model is



(a) Depth Map D

(b) Input Image I

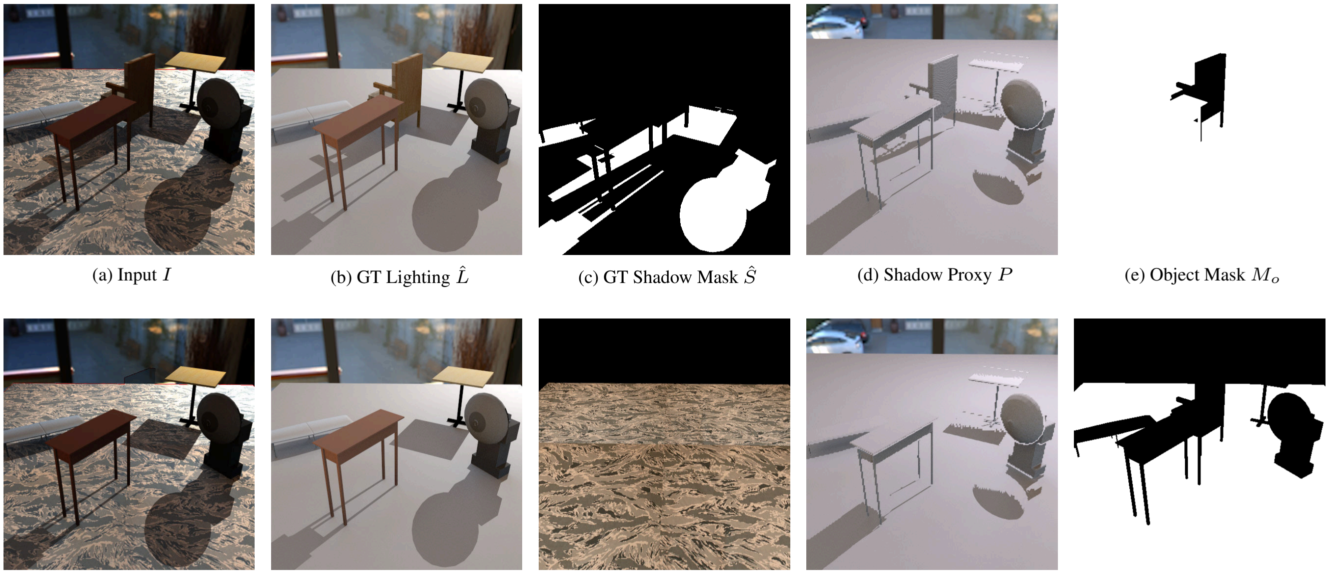
(c) HDRI Environment Map

(d) Geometry (top view)

(e) Shadow Proxy P

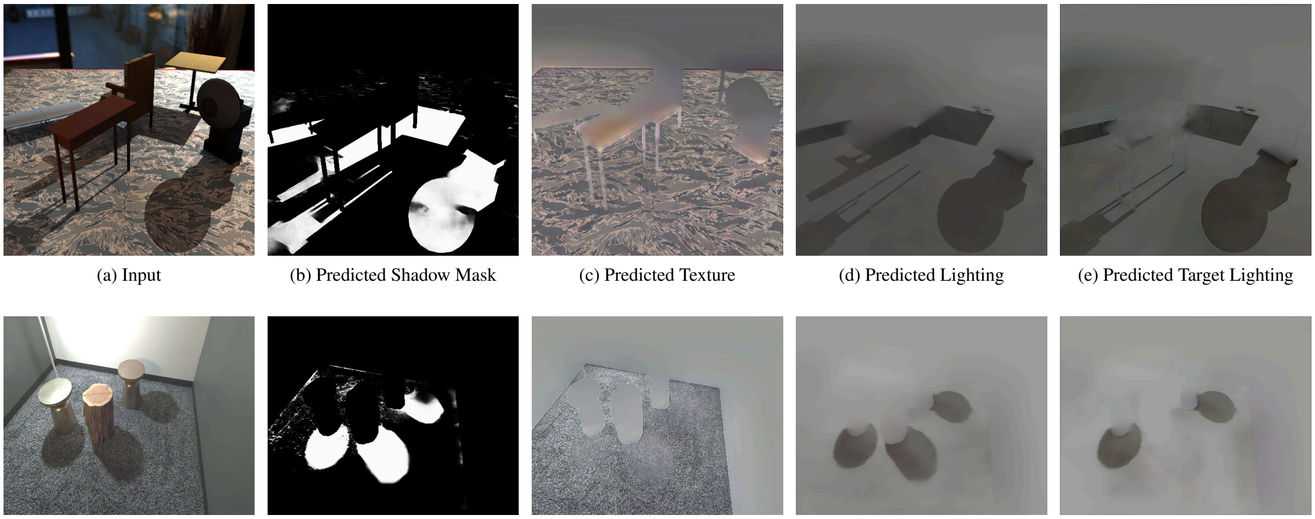
(f) Target Proxy P'

Figure 1: Example of a rough proxy model for a real scene. The raw inputs are shown in the top row, where I and D come from the RGBD sensor and the HDRI environment map comes from a 360 camera. The HDRI environment map (shown tonemapped) contains two light sources: a nearby lamp just visible in the input image that causes most of the shadows (red), and a set of LED lights further away that just barely cast a visible shadow (green). We show a top view of the proxy meshes of the three stools obtained from the depth sensor (the ground plane and wall geometry is omitted for clarity). The scanned geometry is not only incomplete, but also distorted, especially around the metal stools. The shadows in the proxy images P, P' are too rough to use directly (as shown in Figure 5, but they are good enough for our system to remove the necessary shadows.


(a) Input I (b) GT Lighting $\hat{L}$ (c) GT Shadow Mask $\hat{S}$ (d) Shadow Proxy P (e) Object Mask M_o

(f) GT Output $\hat{I}'$ (g) GT Target Lighting $\hat{L}'$ (h) GT Texture $\hat{T}$ (i) Target Proxy P' (j) Receiver Mask M_r

Figure 2: Synthetic scene components for a test scene, including network inputs and ground truth intermediates.

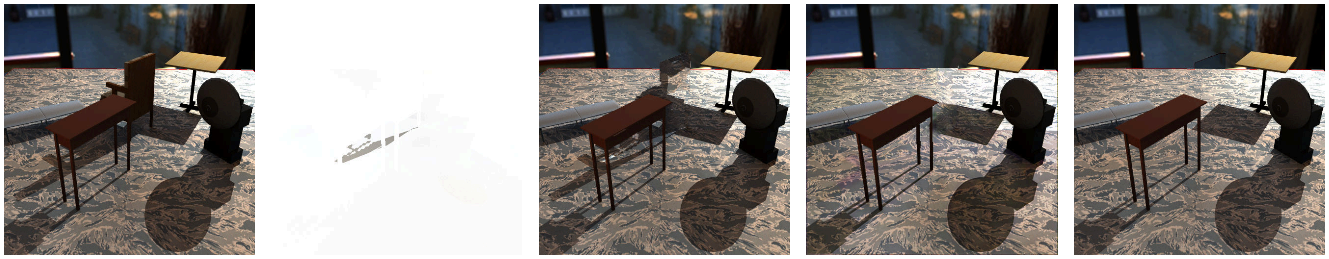

(a) Input (b) Predicted Shadow Mask (c) Predicted Texture (d) Predicted Lighting (e) Predicted Target Lighting

(f) Input (g) Predicted Shadow Mask (h) Predicted Texture (i) Predicted Lighting (j) Predicted Target Lighting

Figure 3: Predicted intermediates from our system for synthetic and real scenes.

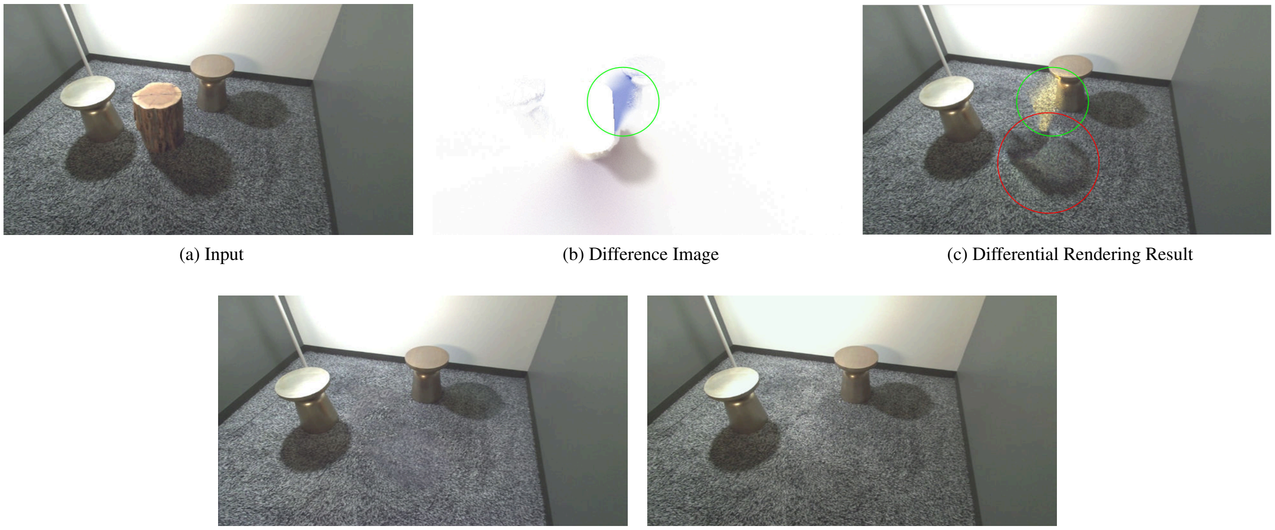
unlikely to be accurate enough for shadow removal. For example, applying differential rendering to the scene from Figure 2 results in an incomplete shadow removal because of the low-quality geometry in the proxy model, as can be scene in Figure 4. The ratio image shown here is computed as P/P' (*i.e.* the difference in log space). Applying the ratio image and inpainting gives $I_{\text{diff}} = g\left(\frac{I}{P/P'}, M_o\right)$.

Figure 5 shows an example of applying differential rendering to the real scene in Figure 1. Note that for this example, we also had to radiometrically calibrate our captured lighting to match with the input image. The inaccurate geometry in the proxy causes the extent of the removed stool's shadow to be underestimated. The scene lighting, consisting of one lamp visible in the image and one area light behind the camera, is poorly represented by the distant lighting model captured by our HDRI environment map. In the environment map, both light sources appear to be of similar intensity and therefore cast shadows of similar appear-



(a) Input (b) Difference Image (c) Differential Rendering Result (d) Our Result (e) Ground Truth

Figure 4: Classical differential rendering on a synthetic scene. The incomplete geometry causes the proxy model to underestimate the extent of the object's shadow.



(a) Input (b) Difference Image (c) Differential Rendering Result

(d) Our Result (e) Ground Truth

Figure 5: Classical differential rendering on a real scene. The proxy model's incomplete geometry underestimates the extent of the object's shadow (red), while the captured distant illumination drastically overestimates the intensity of the secondary light source (green).

ance. In reality, the area light is much further away and only casts a faint shadow. Thus differential rendering attempts to "remove" a shadow that is not truly present, resulting in a brightening of the image.

5. Intrinsic Decomposition Subsystem Comparisons

Our intrinsic decomposition subsystem builds on a body of related work. Figure 6 shows representative results comparing to recent works in shadow removal (Zou *et al.* [9], trained on ISTD [7] and on our dataset) and intrinsic image decomposition (Li and Snavely [5] using author's weights).

Unlike our differential rendering system, these methods do not use proxy renderings, which provide significant benefit. Moreover, shadow removal systems are generally trained on SRD[6] and ISTD[7], which are greatly restricted in lighting conditions and do not generally include the objects that cast the shadows. Shadow removal approaches trained on these datasets have difficulty generalizing to full scenes with more diverse lighting. When trained on our dataset, with more varied lighting and textures, the architecture and losses of these shadow removal methods fail to cleanly identify and remove shadows. Existing intrinsic decomposition works are generally trained on images of entire scenes (*e.g.* IIW[1]) or on single objects (*e.g.* MIT[3]), but typically focus on removing lighting variation due to surface orientation and not on cast shadows.

6. Ablation Study on Loss Terms

We show the importance of several of our loss terms in Table 1 and Figure 7. The multiscale loss was the most important loss term in our system – replacing it with an L1 loss caused many shadows to be left behind, resulting in a large drop in performance on the Shadow RMSE met-

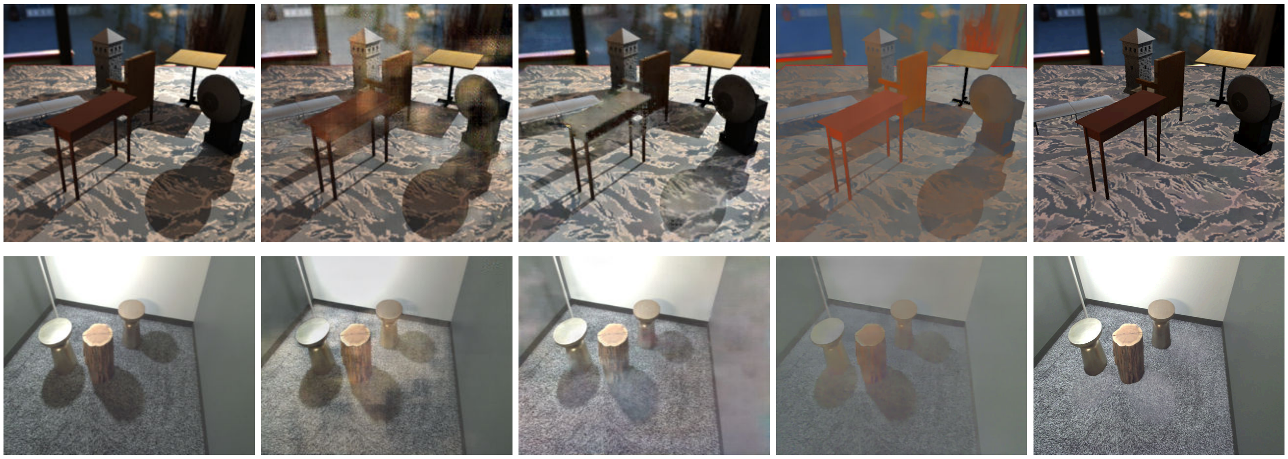


Figure 6: Decomposition of synthetic (top) and real (bottom) scenes. From left to right: input, shadow removal using Zou *et al.* [9] trained on ISTD, shadow removal using Zou *et al.* trained on our dataset, intrinsic image reflectance using Li and Snaveley [5], our reflectance. Note that the three compared methods only operate on the input image, while we also use a rendering of a rough proxy model.

	Synthetic			Real		
	RMSE	Shadow RMSE	Inpaint RMSE	RMSE	Shadow RMSE	Inpaint RMSE
Ours	**0.0248**	**0.0712**	**0.2143**	**0.0340**	**0.0616**	**0.0983**
No Sparse Gradient Prior	0.0290	0.0783	0.2161	0.0586	0.0673	0.0991
No Exclusion Losses	0.0292	0.0792	0.2182	0.0375	0.0683	0.0990
No Multiscale Loss	0.0277	0.0926	0.2185	0.0355	0.0672	0.0997

Table 1: Comparison of error rates under various ablations of our system.

ric. The sparse gradient prior is important to maintain texture fidelity, as without it texture details are sometimes assigned to lighting, and subsequently get removed as shadows. The exclusion losses perform a similar role but also prevent shadows from remaining in the texture image.

7. Additional Real Data Results

We show several additional examples of our system's results on real scenes, as well as comparisons with baselines, in Figures 8-9. The inpainting baseline uses Hifill [8] to inpaint pixels within the object mask as well as within the pixels of a supplied shadow mask. The Pix2Pix baseline is an image-to-image translation network [4] trained to take all of our inputs (including the proxy renderings) and output the pixels on the planar receiver, with the object itself inpainted with Hifill. Our method sometimes results in some overall color shifts (most evident in Figures 8b,9g), but perceptually our method is consistently better at removing shadows than either method (especially on the high-contrast texture in Figure 8e), and our decomposition-based inpainting scheme results in fewer artifacts in the inpainted regions.

References

[1] Sean Bell, Kavita Bala, and Noah Snavely. Intrinsic images in the wild. In *SIGGRAPH*, 2014. 3

[2] Paul Debevec. Rendering synthetic objects into real scenes. In *SIGGRAPH*, 1998. 1

[3] Roger Grosse, Micah K Johnson, Edward H Adelson, and William T Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *Int. Conf. Comput. Vis.*, 2009. 3

[4] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017. 4

[5] Zhengqi Li and Noah Snavely. Cgintrinsics: Better intrinsic image decomposition through physically-based rendering. In *ECCV*, 2018. 3, 4

[6] Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson WH Lau. Deshadownet: A multi-context embedding deep network for shadow removal. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017. 3

[7] Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018. 3

[8] Zili Yi, Qiang Tang, Shekoofeh Azizi, Daesik Jang, and Zhan Xu. Contextual residual aggregation for ultra high-resolution image inpainting. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020. 4

[9] Zhengxia Zou, Sen Lei, Tianyang Shi, Zhenwei Shi, and Jieping Ye. Deep adversarial decomposition: A unified framework for separating superimposed images. *CVPR*, 2020. 3, 4

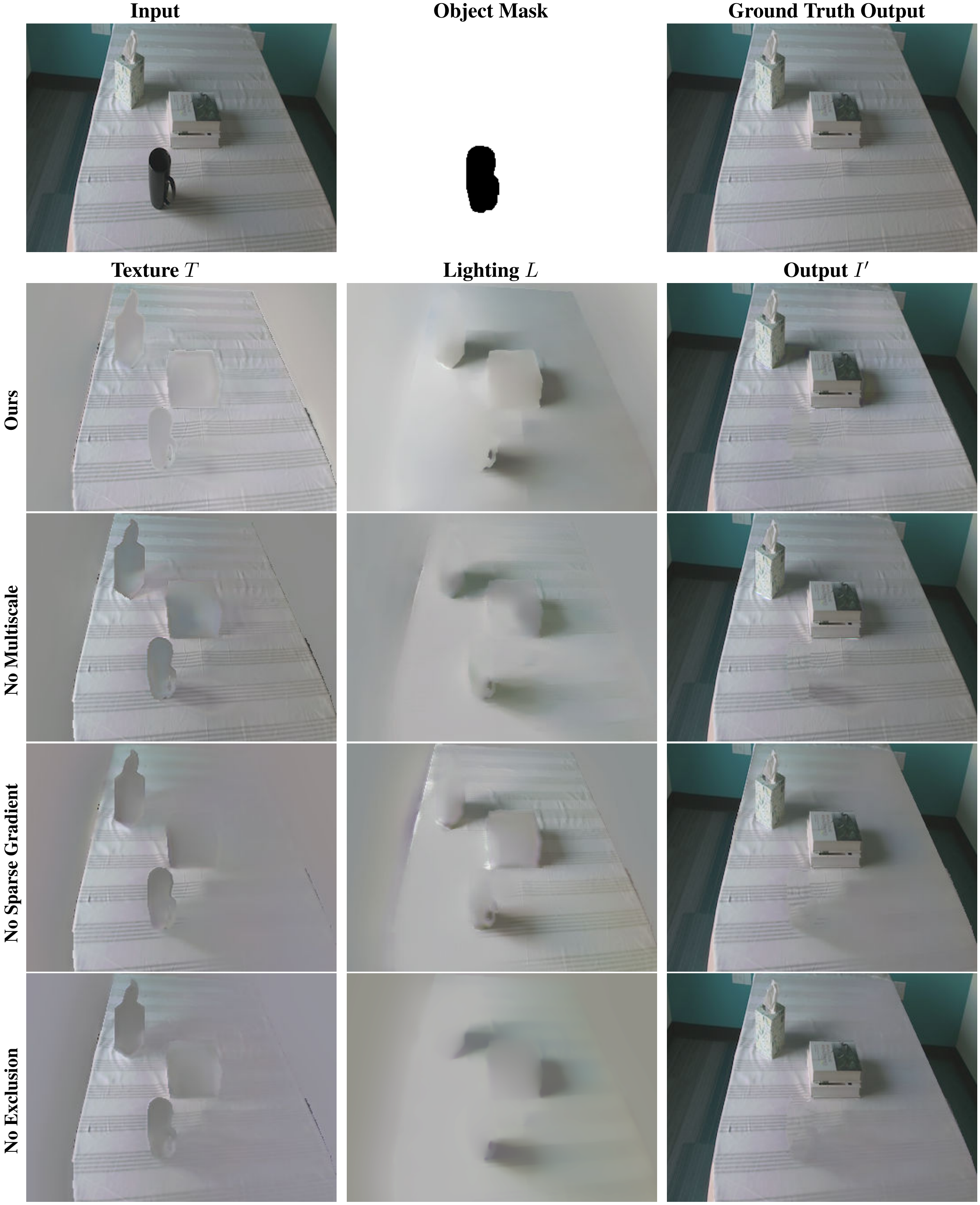


Figure 7: Effects of various loss terms on our results, with intrinsic decompositions. The multiscale loss results in better identification of the extents of shadows. The sparse gradient prior helps keep texture out of the lighting image (so that the texture is not removed by the shadow removal network). The exclusion losses perform a similar role but also prevent shadows from remaining in the texture image.

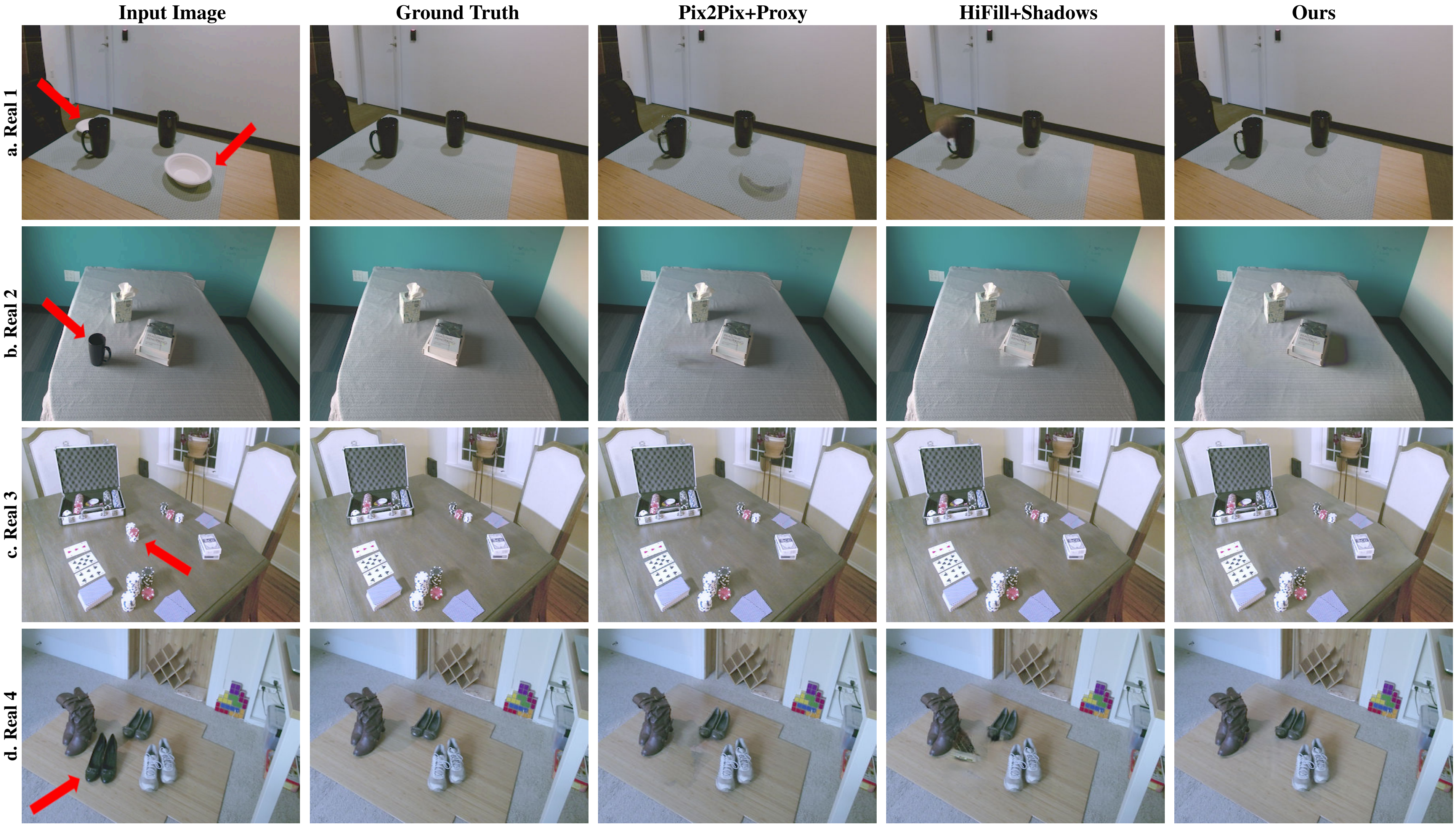


Figure 8: Additional results of shadow removal on real scenes. Removed object(s) are indicated by the red arrows.

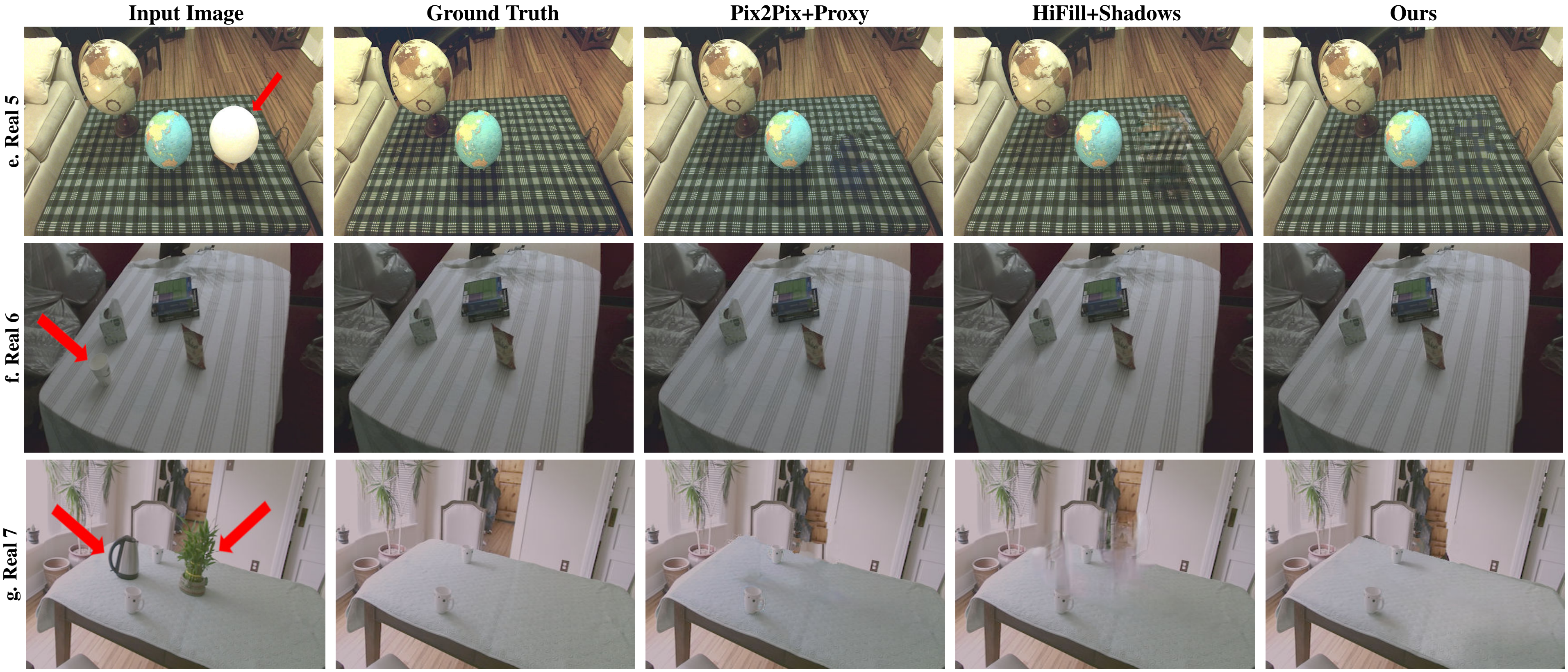


Figure 9: Additional results of shadow removal on real scenes. Removed object(s) are indicated by the red arrows.