Single-Shot Neural Relighting and SVBRDF Estimation: Supplementary Material

Shen Sang[1,2] and Manmohan Chandraker[1]

[1] University of California, San Diego
[2] ByteDance Research

Our supplementary material includes further experiments and details that establish good generalization to real data for the proposed method, as follows:

- More reconstruction and relighting results on synthetic and real data (Sec. 1).
- Further examples demonstrating that the proposed method achieves relighting close to ground truth, on real data (Sec. 2).
- Comparisons on real data to prior state-of-the-art [5] in single-image shape and SVBRDF estimation (Sec. 3).
- Comparisons of our single-image normal estimation on real data, with state-of-the-art photometric stereo method of [3] that uses 96 images (Sec. 4).
- Further results, ablation studies, implementation details and explanation of the accompanying video (Secs. 5-9).

1 More examples on synthetic and real images

We include synthetic data to validate our method in Figure 1. In Figure 2, we include more real examples as a supplementary to Figures 3 and 4 in our main paper, to demonstrate that our method can generalize to real images although it is trained on a synthetic dataset.

2 Relighting comparisons against real data ground truth

We include examples for comparison of our neural relighting results and ground-truth in Figure 3. As mentioned in the main paper, we capture these ground-truth images using a gantry with a cellphone flash light attached to it. The specularities and shadows in these examples demonstrate that our method can produce single-image relighting close to the ground-truth on real data.

3 Further comparisons with prior state-of-the-art [5]

We qualitatively compare our shape and SVBRDF estimation performance on several real data examples, against the state-of-the-art method of [5] using its publicly available code. The results are shown in Figures 4 and 5. In each case, the image is acquired using a mobile phone camera with the flash enabled, under an unknown, arbitrary environment map. While ground truth is not available for

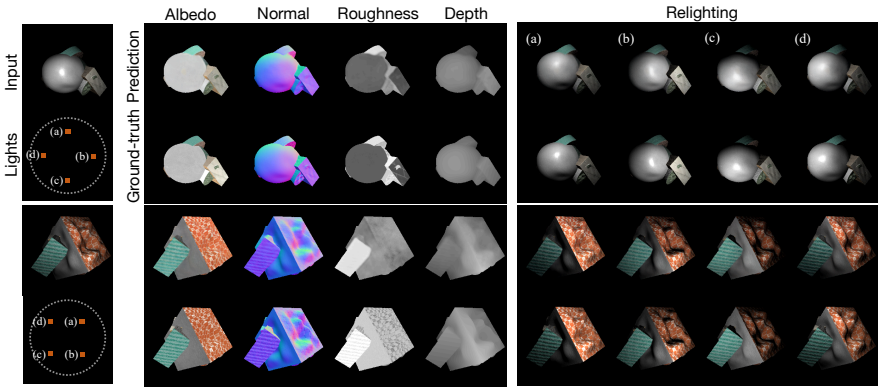

Fig. 1: Reconstruction and relighting results on synthetic data. The input images are shown in the first column, following four columns show the BRDF estimations. The remaining four columns show the relighting images under four new light sources. For each case, the first row is our estimation and the second row is the ground-truth. The new light source is shown as an orange point in a unit circle projected from a hemisphere.

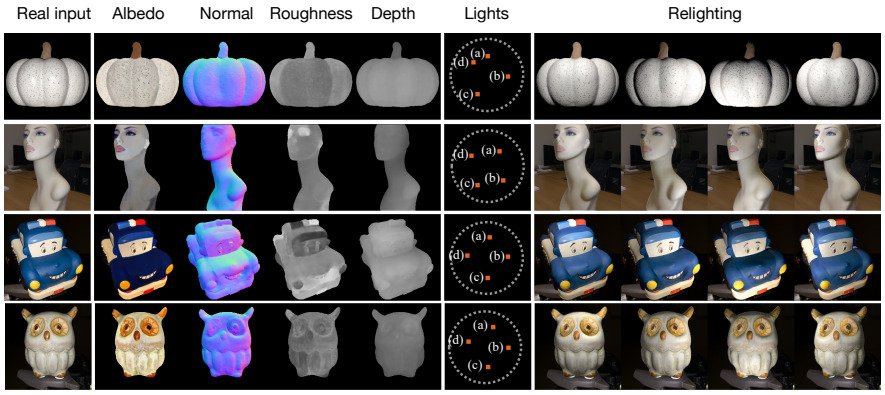

Fig. 2: More reconstruction and relighting results on real data captured under different illumination conditions.

these real images, we observe that the shape and material estimates from both the methods are quite plausible. We observe that the roughness estimates from our method are often better, which is perhaps due to feature sharing between the inverse and relighting decoders, allowing the features to encode cues about appearance variation under various lighting conditions.

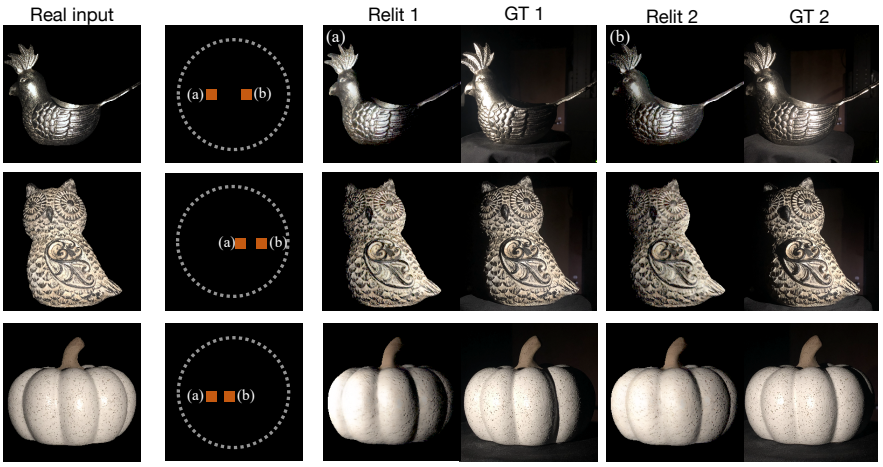

Fig. 3: Comparison between relighting results and ground-truth. The new light source is shown as an orange point in a unit circle projected from the hemisphere.

4 Normal estimation comparison

4.1 Normal estimation comparison with photometric stereo

We now compare with photometric stereo to produce high-quality normal estimates as pseudo ground truth. In Figure 6, we showcase our results on the Light Stage Data Gallery [4] and Gourd and Apple dataset [1]. We use the publicly available code for a state-of-the-art photometric stereo method, PS-FCN [3], which takes 96 images as input. Note that these real images are captured under directional lights while our training assumption is a point light source. Despite the input lighting not exactly matching our training data, our method can still produce good results using only a single image as input.

4.2 More comparisons with [2]

We add four more examples as supplementary for Figure 7 in our main paper, showing that our method outperforms [2] and produce more accurate normal estimations according to the differences in visual quality.

5 Reconstructed shapes

In Figure 8, we use Open3D [6] to synthesize different views of shapes reconstructed by our method, using a single real image as input. The synthesized views are quite accurate even though we use a single image of an object with complex SVBRDF under an unknown environment map, which indicates the quality of our shape estimation.

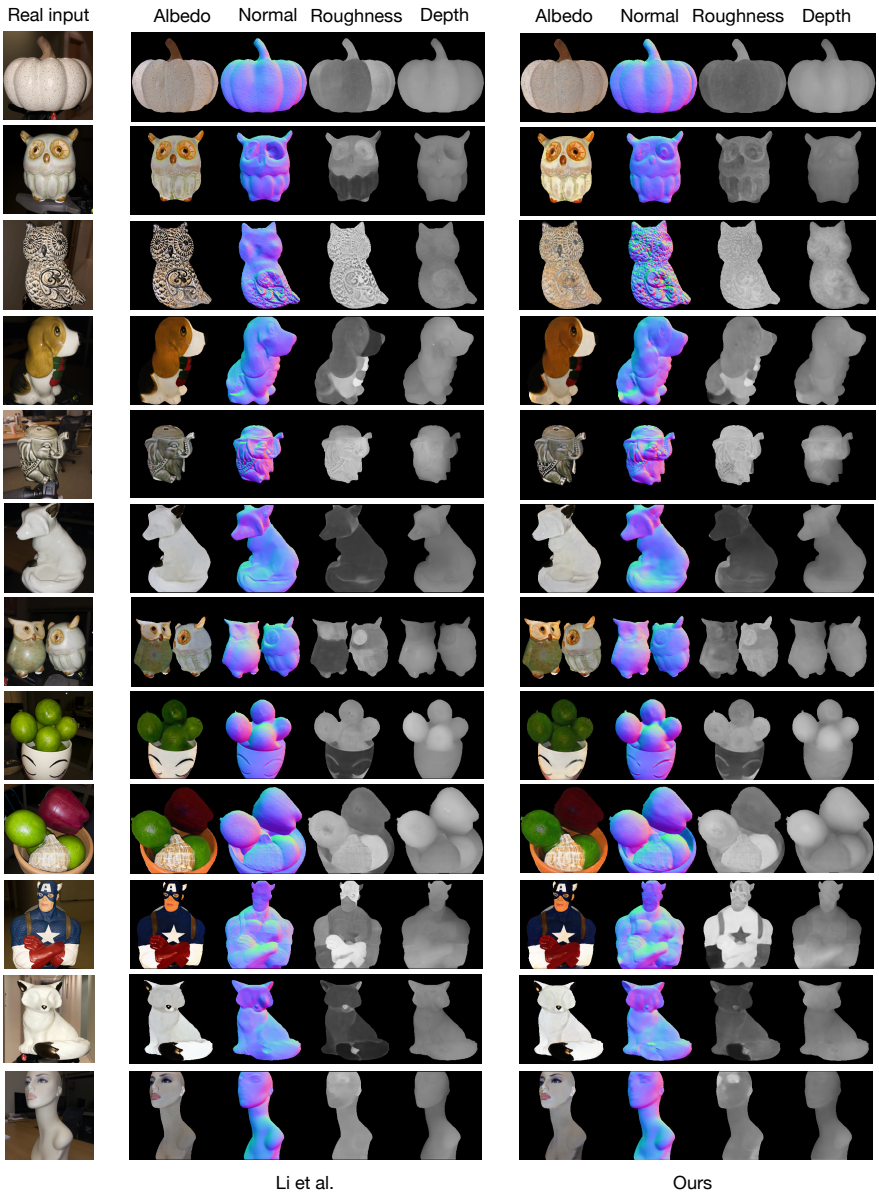

Fig. 4: Comparison with [5] on SVBRDF estimation on real images. While both methods produce high-quality estimates, some factors such as roughness are better estimated by our method.

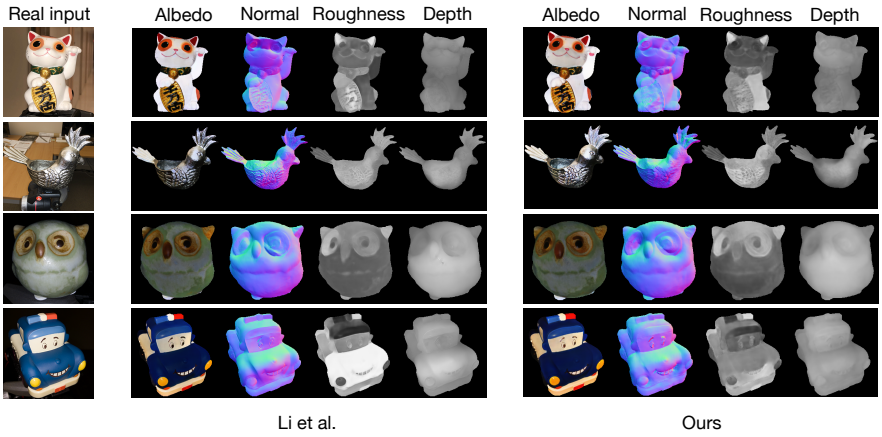

Fig. 5: Further examples for comparison with [5] on SVBRDF estimation, with similar observations as Figure 4. In some cases such as the *owl* in the third row, note the significantly better normals and roughness obtained using our method. This indicates the benefit of jointly learning the relighting task along with the shape and SVBRDF estimation task.

	A	N	R	D	Render	Relight	MS-SSIM
Multiple	1.45	3.41	4.20	1.57	1.19	0.99	0.89
Single	1.39	3.32	4.18	1.44	1.14	0.99	0.89

Table 1: Comparison of training using multiple relighting targets or single target. Error metrics as defined in the main paper.

6 Multiple relighting targets

We consider here the relative importance of the two jointly trained tasks. Using multiple relighting targets as ground truth for training is a way to introduce a larger weight for the relighting task. So, we train a network with the same architecture as the one in the main paper, but with relighting under 4 directions as ground-truth for each training iteration (as opposed to a single target in the main paper). As shown in Table 1, we find that with multiple targets involved in training, the relighting performance is not improved, while a penalty is incurred for SVBRDF reconstruction.

7 Environment map editing

Our model may also be fine-tuned to allow relighting by editing the environment map. Given a new environment map, we compute the corresponding spherical harmonic coefficients. Then, we replace the estimated environment coefficients

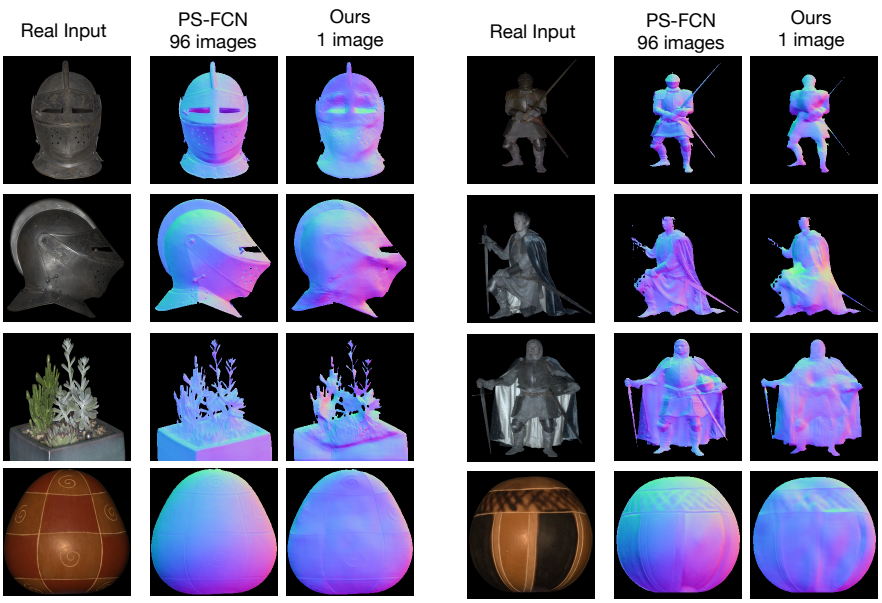

Fig. 6: Normal estimation on a real images from Light Stage Data Gallery [4] and Gourd and Apple dataset [1]. We use PS-FCN [3] to obtain pseudo ground-truth for comparison. Note that the results from the photometric stereo method come from 96 input images while our method use only a single image as input, while ours result is still reasonable compared with the pseudo ground-truth although the input images do not correspond to our training assumptions.

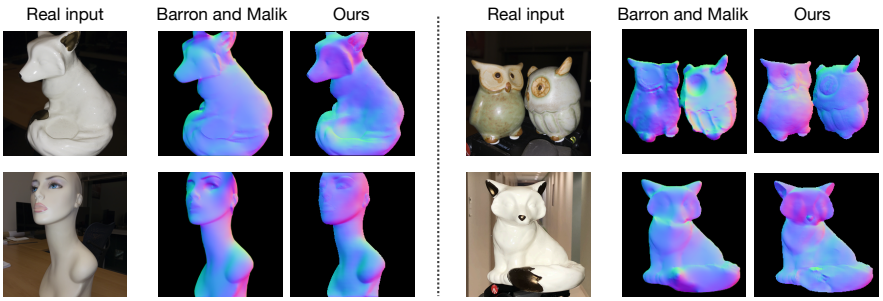

Fig. 7: Normal estimation on a real images with [2]. From the accuracy and visual quality of our normal, we show that our method is significantly better compared with that of [2].

with the new ones as input to the *RelightDecoder*. In Figure 9, we illustrate the modification in the architecture.

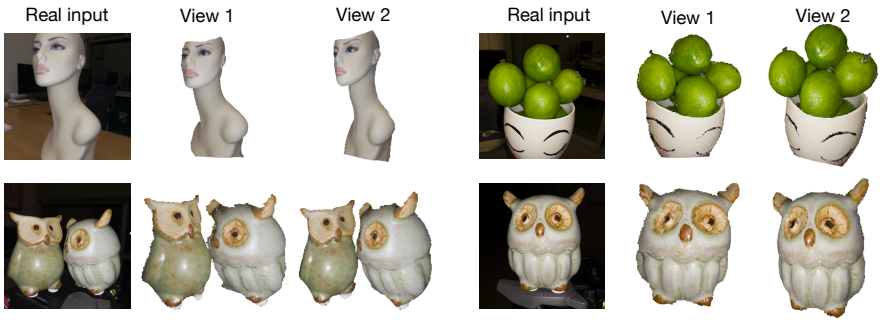

Fig. 8: Different views of reconstructed shapes using our method. Even with a single input image from a mobile phone camera, with complex SVBRDF and an unknown environment map, we estimate accurate shape and can synthesize novel views.

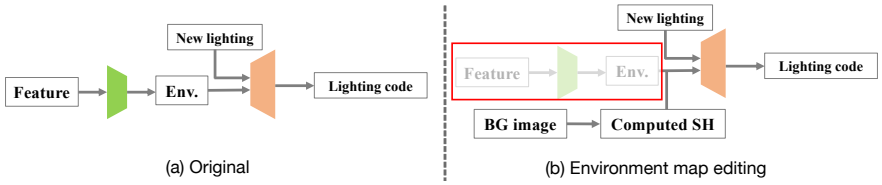

Fig. 9: Visualization of fine-tuning for environment illumination editing. **Left:** The original architecture used to generate the lighting code for relighting training. **Right:** The modified architecture for environment map editing. The estimated environment coefficients in the red box are replaced by the SH coefficients computed from the background image.

8 Cascade design

In our main paper, we provide quantitative results to show the effectiveness of our proposed method. In Figure 10, we provide both synthetic and real examples to show effectiveness of our cascade design qualitatively. From the synthetic and real examples, we observe that for both the reconstruction and relighting tasks, the results are gradually refined to yield higher accuracy in the latter stages.

9 Video

Relighting with moving point light We include videos to demonstrate our joint relighting and SVBRDF estimation. We showcase relighting results with real input images. The real images are acquired with a mobile phone camera. Results under a point light source, as well as including an arbitrary unknown environment map, show the high quality of our relighting under a wide range of light source positions.

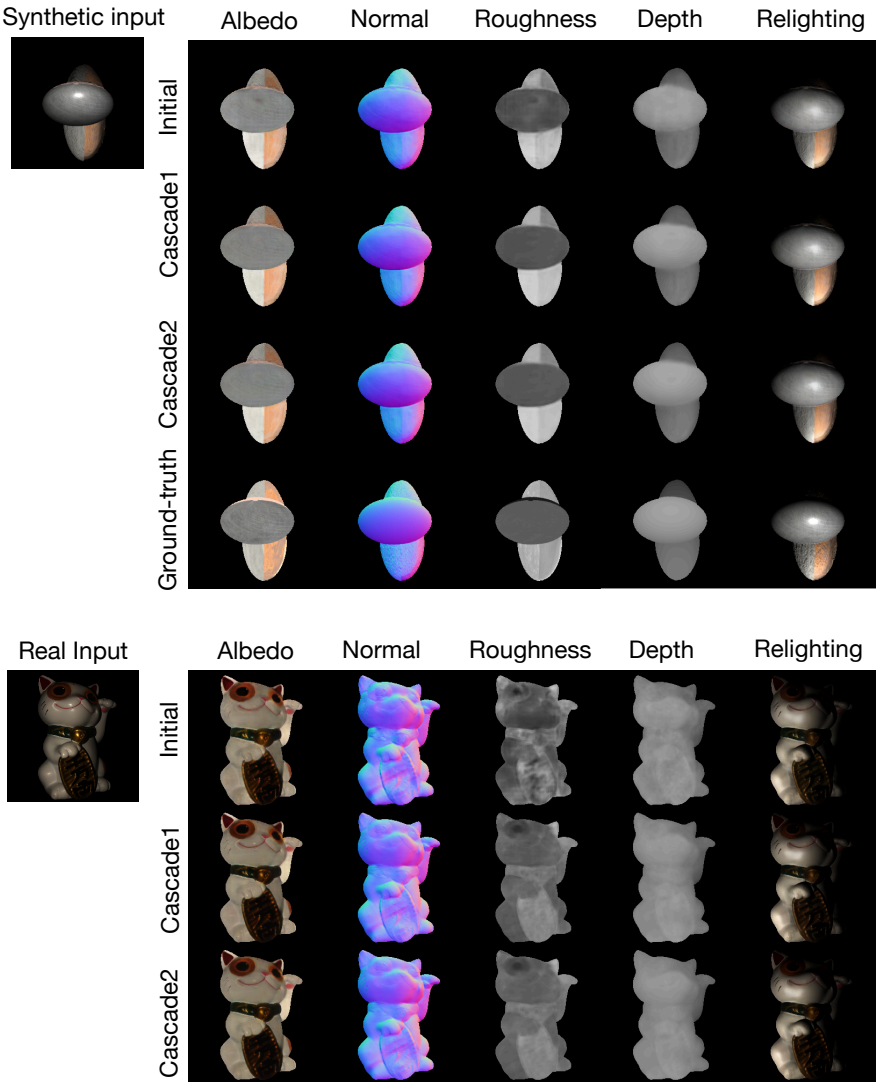

Fig. 10: The effectiveness of our cascade design. Each stage refines the outputs of the previous stage, for both the reconstruction and relighting tasks. Note that for shape and SVBRDF estimation, the results become less noisy, while specularities become more pronounced and refined.

Relighting under novel environment map We also demonstrate relighting outputs from our method with novel environment maps, by fine-tuning as described in Sec. 7. We include several examples with moving environment maps in our video, which demonstrate that our method can produce photorealistic

relighting results under a novel environment map with a single real image as input. In each case, we note that fine details are obtained by our relighting, such as color from the environment map reflected off the object surface.

Comparisons We also compare to a few alternative methods in our video. First, we compare our relighting method with two baselines – forward rendering on recovered shape and material parameters and image-to-image translation. We note that specularities and shadows are not recovered well by image-to-image translation, while our method estimates both accurately. In addition, we include a comparison with [5], which can produce high-quality SVBRDF estimation given a single image. Note that when shape or SVBRDF estimation are inaccurate, the errors persist in the relighting result of [5], while our method can still produce good relighting results. The comparisons show that our novel network design, feature sharing and joint training of the inverse and forward tasks lead to significant improvements for both tasks.

References

1. Alldrin, N., Zickler, T., Kriegman, D.: Photometric stereo with non-parametric and spatially-varying reflectance. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8. IEEE (2008)
2. Barron, J.T., Malik, J.: Shape, illumination, and reflectance from shading. IEEE transactions on pattern analysis and machine intelligence **37**(8), 1670–1687 (2014)
3. Chen, G., Han, K., Wong, K.Y.K.: PS-FCN: A flexible learning framework for photometric stereo. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 3–18 (2018)
4. Einarsson, P., Chabert, C.F., Jones, A., Ma, W.C., Lamond, B., Hawkins, T., Bolas, M., Sylwan, S., Debevec, P.: Relighting human locomotion with flowed reflectance fields. In: Proceedings of the 17th Eurographics Conference on Rendering Techniques. p. 183–194 (2006)
5. Li, Z., Xu, Z., Ramamoorthi, R., Sunkavalli, K., Chandraker, M.: Learning to reconstruct shape and spatially-varying reflectance from a single image. In: SIGGRAPH Asia 2018 Technical Papers. pp. 269:1–269:11. ACM (2018)
6. Zhou, Q.Y., Park, J., Koltun, V.: Open3D: A modern library for 3D data processing. arXiv:1801.09847 (2018)