# Invertible Neural BRDF for Object Inverse Rendering

## Supplementary Material

Zhe Chen, Shohei Nobuhara, and Ko Nishino

Kyoto University, Kyoto, Japan
zchen@vision.ist.i.kyoto-u.ac.jp {nob,kon}@i.kyoto-u.ac.jp
https://vision.ist.i.kyoto-u.ac.jp

**Abstract.** In this supplementary material, we show additional experimental results on
- in-depth comparison with Georgoulis et al. TPAMI 2017,
- reflectance estimation with iBRDF,
- illumination estimation with deep illumination prior, and
- joint estimatation of reflectance and illumination both for synthetic and real images.

## 1 Comparison with Georgoulis et al. TPAMI 2017 [15]

As we stated in the main manuscript, "Georgoulis et al. [15] extend their prior work [36] to jointly estimate geometry, material and illumination. The method, however, assumes Phong BRDF which significantly restricts its applicability to real-world materials." For this reason, their data are selectively of shiny material. Furthermore, the method consists of two steps in which the first step predicts a reflectance map from the input image and the second step decomposes the reflectance map into Phong parameters and an environment map. As our goal is fundamentally different, in that we jointly estimate an arbitrary BRDF and natural illumination directly from the input image albeit of an object with known geometry, our method is compared with the reflectance map decomposition network in [15].

Since we thoroughly evaluate our method's effectiveness on synthetic data in the main manuscript and in the following sections, we focus on comparing our method to the reflectance map decomposition on the real data of [15]. Fig. 1 shows comparisons of the estimated reflectance and illumination side-by-side with their results. We show results of rendered sphere of estimated illumination with mirror reflection (mirror) and rendered sphere of the estimated BRDF with a different illumination (nat. illum.). We omit comparison on rendered spheres of the estimated illumination with different known BRDF from the input as we believe they are misleading. As Lombardi and Nishino [25] showed, the reflectance and illumination are estimated up to the highest frequencies of either. As such,

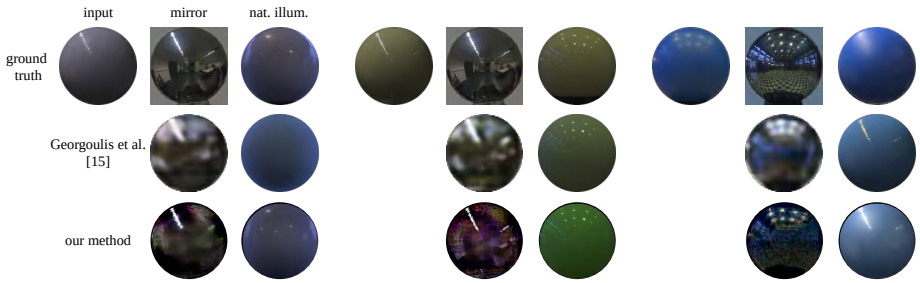



Fig. 1: Comparison with DeLightNet [15]. The illumination estimate is shown as a rendered mirror sphere (mirror) and the reflectance estimate is shown as a sphere rendered under a different natural illumination (nat. illum.) from the input image (input). Our results capture finer details of the BRDF and the illumination demonstrating more robust and accurate decoupling of the two from object appearance.

relighting using the illumination estimate with another BRDF would not accurately capture the true accuracy of the illumination estimate as that BRDF will attenuate the illumination esimate's frequency properties. Furthermore, neither the paper nor the code mentions which BRDFs were used to render these relit spheres, which prevents us from making a comparison.

Overall, judging from the sphere renderings of the estimated BRDF with a different illumination, our BRDF estimates qualitatively appear more accurate and faithful to the underlying reflectance of the input image as well as ground truth (e.g., higher frequency details of illumination estimates). Our method is a physically-based reconstruction, that decouples the reflectance and illumination of object appearance. In contrast, the method of [15] is a learned decomposition on tens of thousands of images, fundamentally bound by the combinations seen in the training data. Our method does not involve any learning, other than the BRDF model itself. We believe these two methods complement each other and can be used in conjunction, perhaps to obtain a quick learning-based initialization and then a physically-based decoupling for complex surfaces and environments that are rarely accurately represented with the Phong model. Table 1 shows quantitative comparison of 91 combinations of those we could identify (on the project web page–it is not clear what the remaining 9 are) among the 100 in [15]. For direct comparison to [15], we calculate the metrics in the log space defined by $\log(x + 1.0)$ as the original paper rather than $\log(x)$ as used in other parts of our paper. Note that there is a fundamental ambiguity in the scale between the recovered illumination and BRDF in addition to the color. The network in [15] recovers the parameters of an analytical reflectance model (i.e., Phong), not the radiance distribution of the BRDF, which implicitly avoids this scale ambiguity. In our case, it is hard to determine the scale difference. Instead, we multiply the rendering by a scale factor that minimizes the MSE between the recovery and the ground truth. Note that this post-processing does not affect

| | Mirror | | Nat. Illum. | |
|---|---|---|---|---|
| method | Log RMSE | DSSIM | Log RMSE | DSSIM |
| Gerogoulis et al. [15] | 0.933 | **0.365** | 1.110 | 0.186 |
| ours (w/ scale correction) | **0.744** | 0.369 | **0.340** | **0.179** |
| ours (w/o scale correction) | 0.926 | 0.369 | 0.932 | 0.179 |

Table 1: Mean Log RMSE and mean DSSIM errors of illumination estimates (mirror) and reflectance estimates (nat. illum.). Our method also quantitatively outperforms that of [15].

the properties of the recovered BRDF and illumination. As such the comparison is fair. As shown in Table 1, we achieve lower errors in all metrics, especially for the log RMSE, even without scale correction. These results show that our estimation matches the characteristics of the ground truth very well.

## 2 Reflectance Estimation with iBRDF

In the main manuscript, we validated the effectiveness of the invertibile neural BRDF for single-image BRDF estimation by showing the log RMSE errors of the accuracy of BRDF estimation for each of the 100 different materials in the MERL database rendered under 5 different known natural illuminations, i.e., total 500 tests (Fig. 3(a) of main manuscript). The results show that the BRDF can be estimated accurately regardless of the surrounding illumination. Here, in Fig. 2, in addition to Fig. 3(b) of the main manuscript, we show additional estimation results as spheres rendered with different point source directions using the recovered BRDF put side-by-side with the ground truth. The recovered BRDF renderings match the ground truth measured BRDF well, even when the illumination differs, demonstrating the ability of iBRDF to robustly recover the full reflectance from partial angular observations in the input image.

## 3 Illumination Estimation with Deep Illumination Prior

We validated the effectiveness of the deep illumination prior by showing samples of the ground truth and estimated illumination without and with the deep illumination prior in the main manuscript (Fig. 4). Here we show additional quantitative analysis. We rendered images of spheres with all the 100 different BRDFs of the MERL database under 15 different natural illuminations captured as HDR environment maps. Fig. 3 shows the relative log RMSE of the estimated illumination with (solid cirles) and without (dashed) the deep illumination prior for all 1500 combinations. We use relative log RMSE, i.e., the log RMSE normalized by the difference between the brightest and the dimmest point in the illumination, since HDR environment maps have different dynamic ranges. For all combinations, except for some involving matte materials, which results in smooth rather than structurally clear (see Fig. 4 of main manuscript) estimated


Fig. 2: Additional results of reflectance estimation from a single image with known illumination. The estimated reflectance match the ground truth (left most) well for all different illuminations (right five).

illumination that RMSE favors, as well as a handful of other combinations in the total of 1500, the deep illumination prior achieves higher accuracy of illumination estimates. Table 2 shows the mean relative log RMSE errors of the estimated illumination with and without the deep illumination prior. For all illumination, on average across different BRDFs, the deep illumination prior was effective in achieving more accurate illumination estimation. These results show that the deep illumination prior effectively constraints the optimization to recover accu-

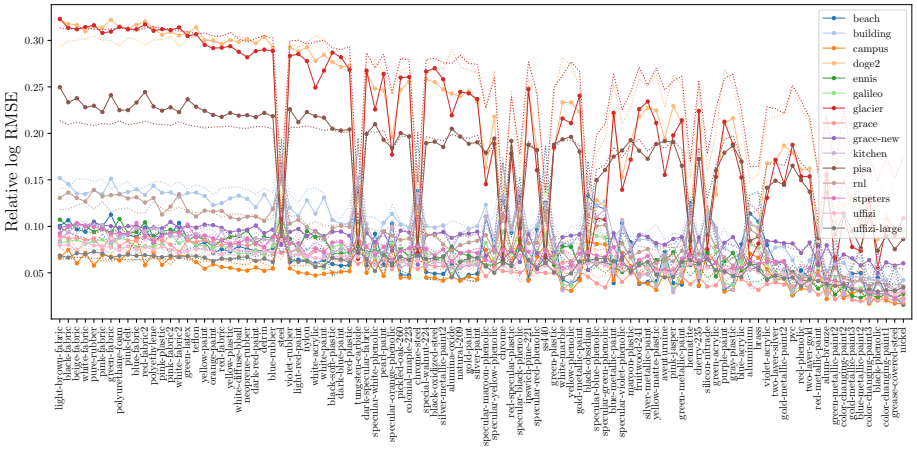



Fig. 3: Relative log-space RMSE errors of estimated illumination with (solid curve with circles) and without (dashed curve) deep illumination prior for 15 different natural illumination recovered from 100 different BRDFs. For most combinations of illumination and material, the RMSE when estimated with the deep illumination prior is lower, demonstrating the effectiveness of the prior.

rate, dense non-parametric representations of a wide variety of complex, natural illumination.

## 4 Joint Estimation of Reflectance and Illumination

We demonstrated the effectiveness of the proposed inverse rendering framework using the invertible neural BRDF, its embedding space, and the deep illumination prior on synthetic input images in Fig. 5(a) of the main manuscript. Here we show additional results including quantitative analysis. We rendered a total of 1500 images of spheres rendered with the 100 MERL BRDFs under 15 different environment maps, and used each as an input to our inverse rendering

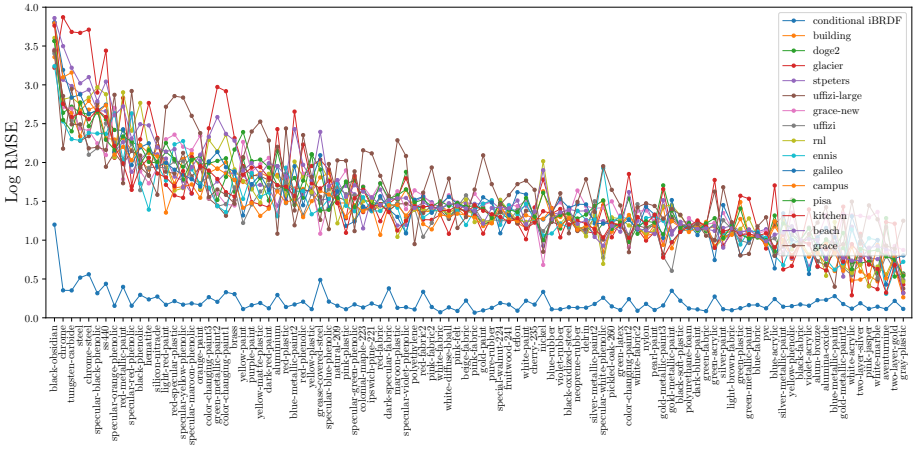
| environment map | beach | building | campus | doge2 | ennis | galileo | glacier |
|---|---|---|---|---|---|---|---|
| without prior | 0.077 | 0.110 | 0.058 | 0.243 | **0.066** | 0.071 | 0.252 |
| with prior | **0.073** | **0.101** | **0.053** | **0.216** | 0.068 | **0.065** | **0.207** |

| environment map | grace | grace new | kitchen | pisa | rnl | st peters | uffizi | uffizi large |
|---|---|---|---|---|---|---|---|---|
| without prior | 0.059 | **0.082** | 0.084 | 0.177 | 0.097 | 0.076 | 0.062 | 0.060 |
| with prior | **0.057** | 0.083 | **0.074** | **0.174** | **0.089** | **0.069** | **0.059** | **0.058** |

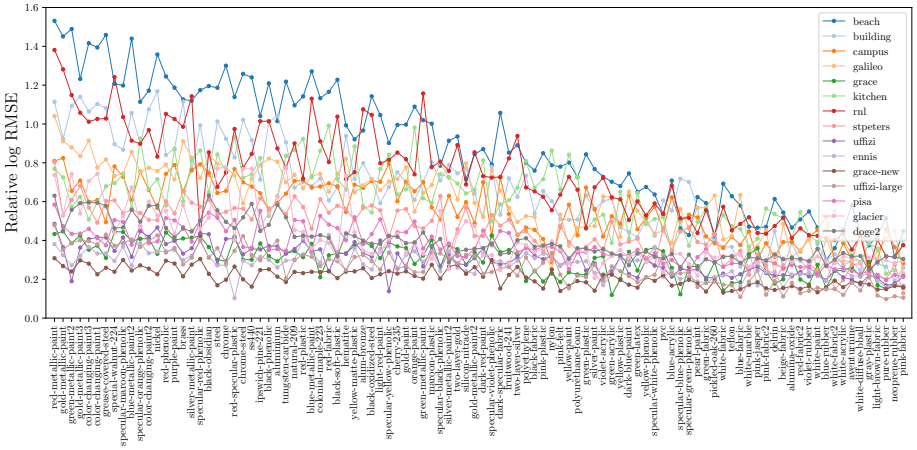
Table 2: Mean relative log-space RMSE errors of estimated illumination without and with the deep illumination prior. For every illumination, the use of the prior leads to lower error.

method. Fig. 4(a) shows the log-space RMSE of estimated BRDF, and Fig. 4(b) shows the relative log-space RMSE of estimated illumination for all the 1500 combinations of reflectance and illumination. The BRDF estimates are particularly accurate for most materials (about 90%) considering the fact that widely used "Lambertian + Cook-Torrance" reflectance model approaches log RMSE of 2 (Fig. 2 of main manuscript). The errors of illumination estimates also stay within reasonable range from the illumination estimation errors when the BRDF is known (Fig. 3. Note that the input, BRDF, and illumination estimates are all in high dynamic range, and small discrepancies in bright highlights can cause large RMSE errors. These results demonstrate the robustness of our inverse rendering method, the expressiveness of our invertible neurla BRDF model, and effectiveness of the deep illumination prior.

Finally, we show the remaining results on images of real objects taken under natural illumination from the Objects Under Natural Illumination Database [25]. When combined with Fig. 5(b) in the main manuscript, Fig. 5 shows all the results of jointly estimating the BRDF and illumination using input images in the database. The illumination is not as clear as those recovered from synthetic object appearance. This, however, is mainly attributed to the fact that real objects, especially those consisting of flat surfaces like the milk jug, only reflect a portion of the surrounding environment into the camera. Fig. 6 shows relative log-space RMSE errors of the illumination estimates. The recovered reflectance properties are consistent across results from different illumination, except for the color shifts due to the inherent ambiguity (especially apparent when the object is white). Baking this color constancy problem into the inverse rendering process is left as future work. The errors are larger than the synthetic case, which is also mainly caused by the partial observation captured in the input images. The characteristics of the illumination estimates where the object surface normals partially cover appear consistent across different objects for each environment. Overall, the illumination estimates are quantitatively and qualitatively reasonable, and the BRDF estimates realistic. Note that there are no ground truth measurements for the BRDF, and due to slight errors in geometric calibration of the dataset, direct relighting comparisons were not plausible. These results demonstrate the robustness and accuracy of our method applied to real scenes.




(a) Log-space RMSE of estimated reflectance




(b) Relative log-space RMSE of estimated illumination

Fig. 4: Log-space RMSE of jointly estimated reflectance and illumination using the conditional invertible neural BRDF and deep illumination prior for 1500 different combinations of 100 MERL BRDFs and 15 environmentmaps. The blue curve in (a) is the BRDF fit of conditional iBRDF for reference. These extensive results demonstrate the effectiveness of our proposed models and method (see main text).

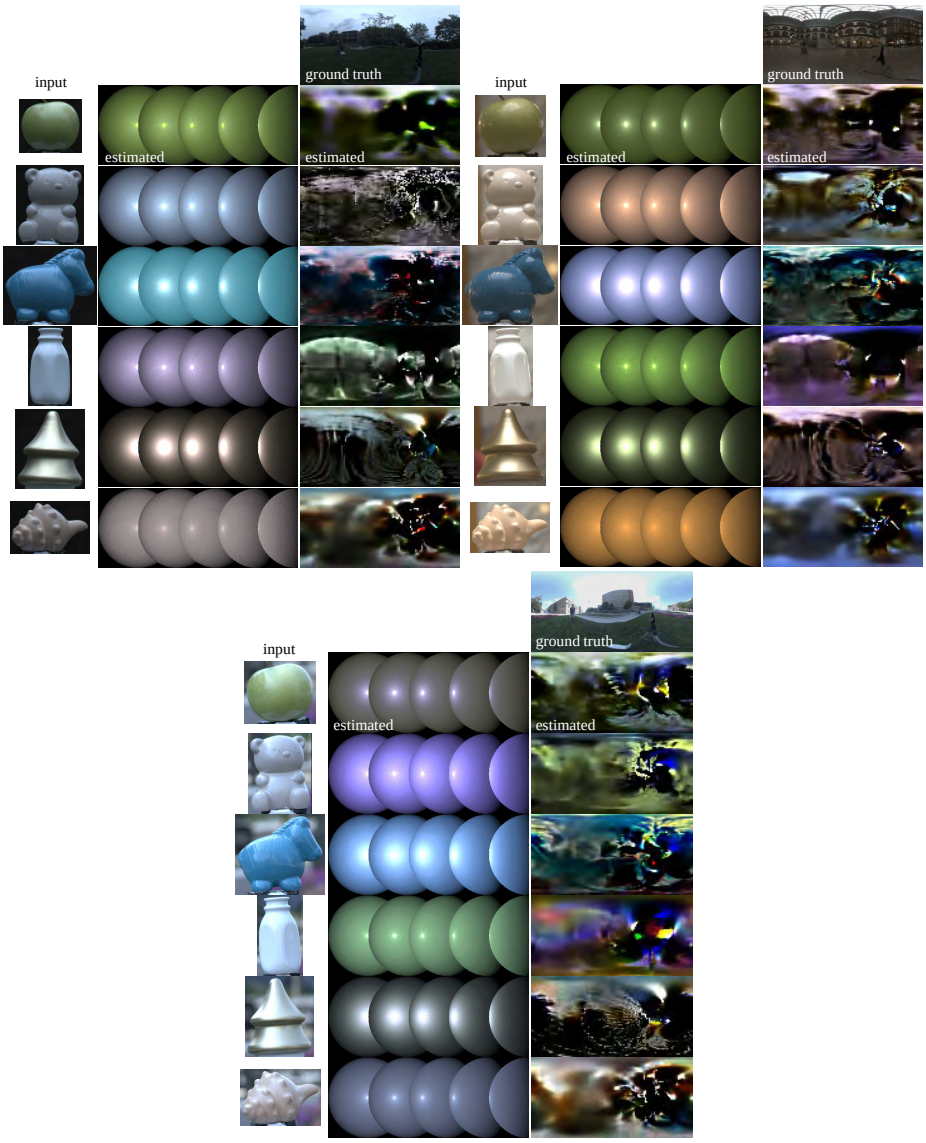



Fig. 5: Remaining results of joint estimation of reflectance and illumination from images of real objects [25].

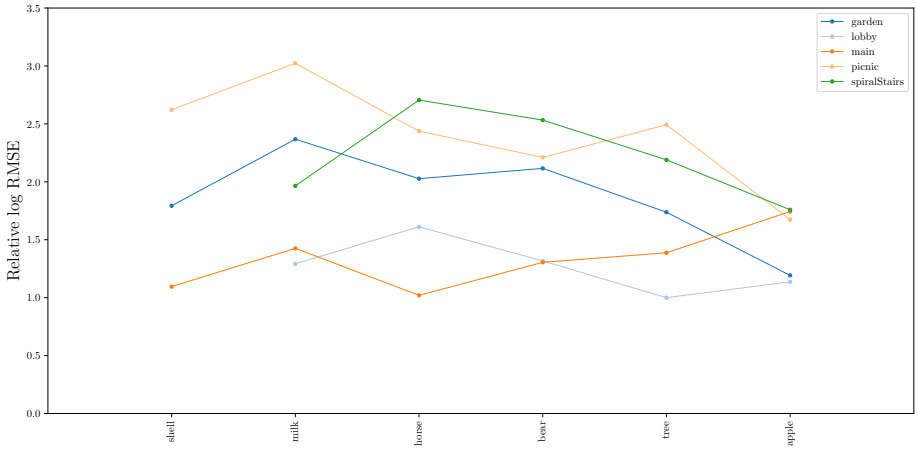



Fig. 6: Relative log-space RMSE errors of estimated illumination from joint estimation of Objects Under Natural Illumination Database [25].