Materials for Masses: SVBRDF Acquisition with a Single Mobile Phone Image

Zhengqin Li* Kalyan Sunkavalli† Manmohan Chandraker*

*University of California, San Diego †Adobe Research, San Jose

1 Further Experimental Analysis

Error distribution on test set To provide better intuition into our quantitative results, we plot the distributions of prediction errors for diffuse albedo ($\mathcal{L}_d$), normals ($\mathcal{L}_n$), roughness ($\mathcal{L}_r$) and relighting ($\mathcal{L}_{rec}$) in Figure 1. Then, we sort the BRDF reconstruction results in the test set according to $\mathcal{L}_d + \mathcal{L}_n + \mathcal{L}_{rec}$ and illustrate the estimation and relighting quality for a random material picked from various percentiles of the above error distribution. The qualitative comparison is shown in Figure 2.

Even though our network is trained end-to-end, we observe physically meaningful trends in Figure 1. For instance, the materials that correspond to lower error percentiles tend to have flat normals, uniform diffuse color and wide specular lobes. On the other hand, materials with higher errors tend to have more complex normals, stronger local variations in diffuse color and roughness, or more prominent highlights. This demonstrates the benefits of our network design which considers the underlying problem structure. We also observe that normal and diffuse color estimates are quite accurate even at error percentiles higher than 50, which contributes to reasonable relighting results under *novel* lighting even at high error percentiles.

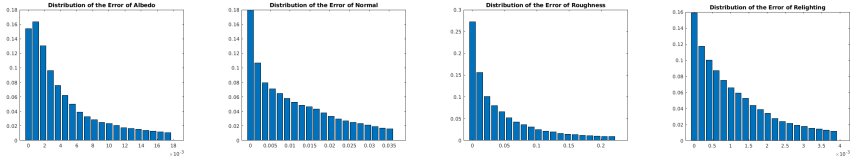

Fig. 1. From the left to the right, error distributions of diffuse albedo, normal, roughness and relighting.

2 Further Results on Real Data

Comparison with photometric stereo as reference In Figure 3, we compare the normals estimated by our method with that of [1], using the normal map from

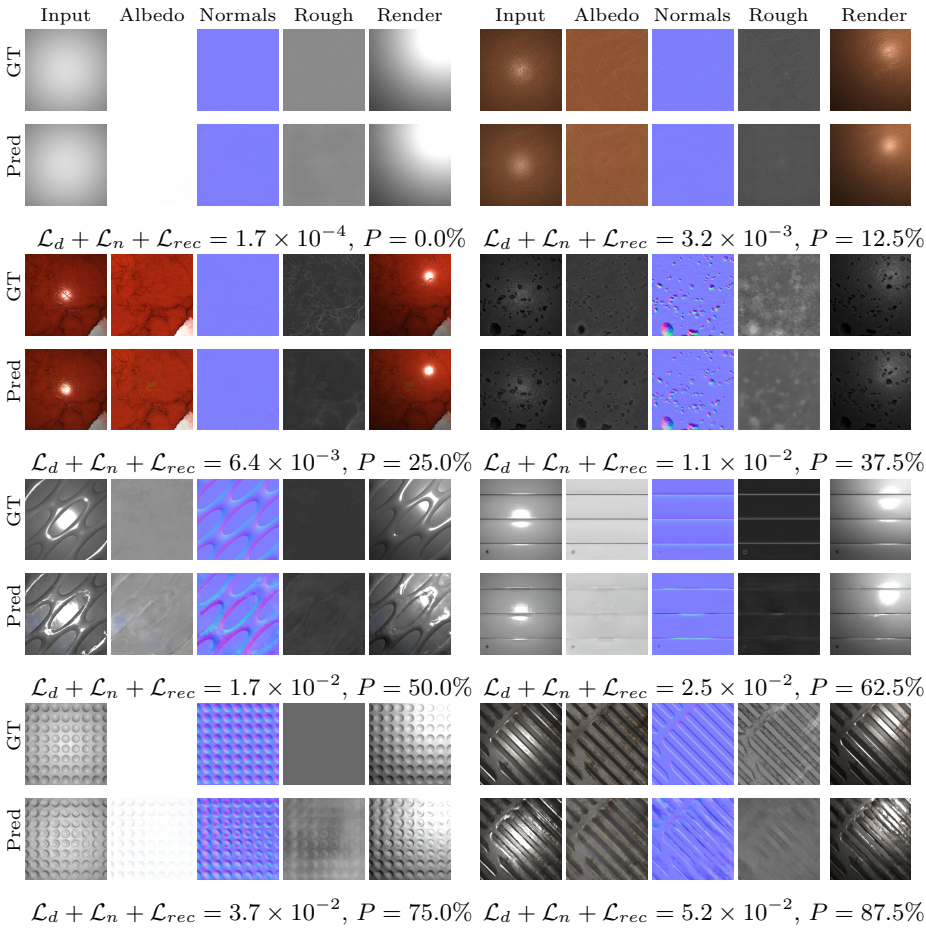

Fig. 2. SVBRDF estimation results sorted according to the prediction error. The error here is defined as $\mathcal{L}_d + \mathcal{L}_n + \mathcal{L}_{rec}$. We do not consider $\mathcal{L}_r$ here roughness has relatively smaller influence towards the final appearance of the surface. Here, P denotes the percentage of samples in the test set with error higher than the considered sample.

photometric stereo as reference. In the main paper, we use the photometric stereo method of [2]. Here, we instead use a simpler but more robust method. We acquire images of a material sample under 52 different directional point light sources. We abandon the 5 most brightest observations and 5 darkest observations and use the rest for a Lambertian photometric stereo. We find such a method to be quite robust to shadows, as well as the effects of complex BRDF such as glossiness or specularity. In comparison, we observe that our CNN is able to capture very fine details in the normal map, in particular, better than the method of [1]. For instance, note the detail within the grooves of the material in the second and third rows. This demonstrates the efficacy of the proposed method for normal and SVBRDF estimation.

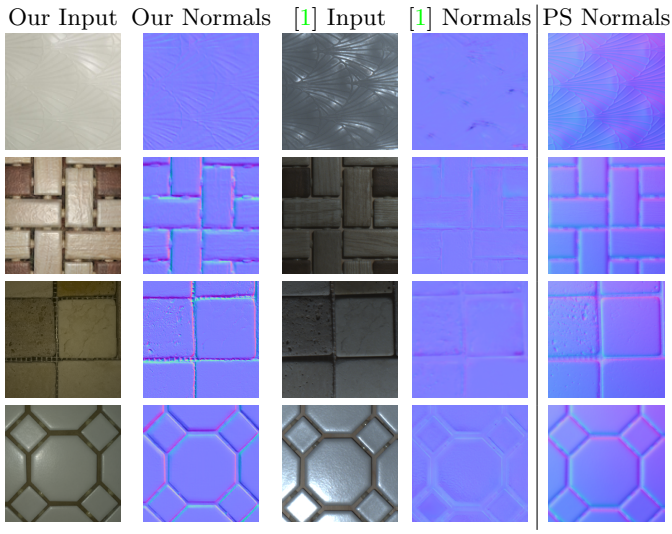

Fig. 3. Comparison of normal maps using our method and [1], with photometric stereo as reference. Even with a lightweight acquisition system, our network predicts high quality normal maps.

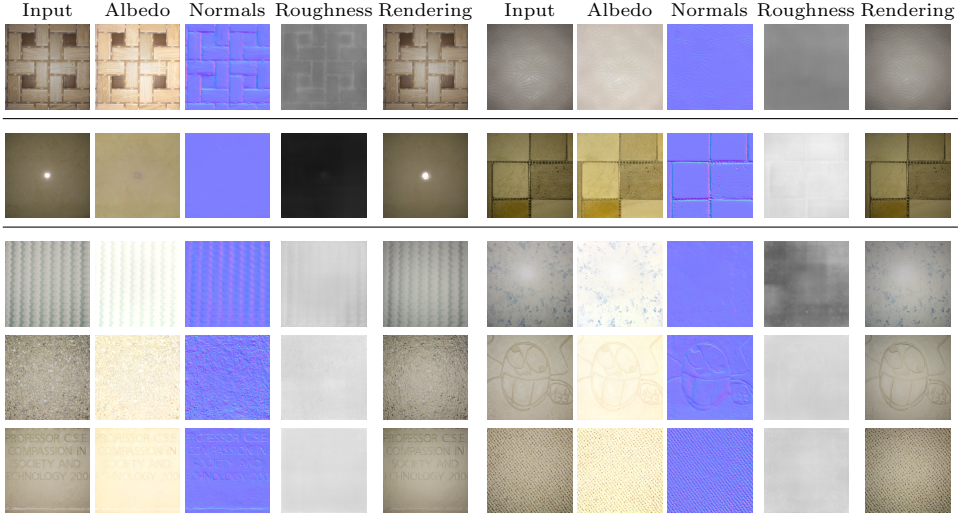

Fig. 4. SVBRDF estimation results on real data. All images are acquired using a handheld mobile phone camera, where the z-axis of the camera is only approximately perpendicular to the sample surface. The inaccuracy in positional calibration of the camera is visible in the input image of the example in the second row of the first column, where the highlight is not exactly in the center of the image. However, our method still obtains reasonable normal and SVBRDF estimation results in all cases. The images in the first row are obtained using an iPhone 6s, the second row using a Huawei P9 while the next three rows using a Lenovo Phab 2. This demonstrates that our algorithm can handle new unknown devices quite well.

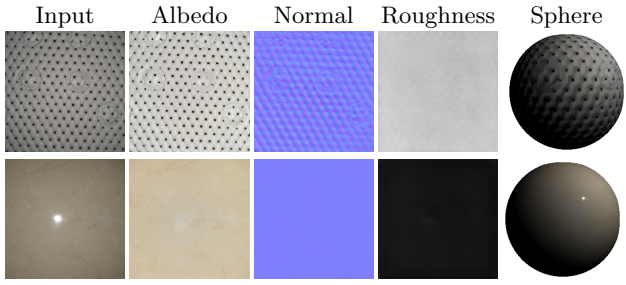

Fig. 5. Rendering of the estimated real spatially-varying BRDF on a sphere, under a very different oblique lighting direction.

Further real data results in unconstrained environments In Figure 4, we show several more examples of surface normal and BRDF estimation with real data using the proposed method. The images are acquired in unconstrained settings with the camera flash enabled, for several different material types derived from wood flooring, tiles, carpets and so on. In all rows, the mobile phone is hand-held and only approximately parallel to the surface. In each case, we observe that the recovered normals, as well as the diffuse albedo and specular components of the spatially-varying BRDF appear qualitatively correct. In some cases, such as the second row of the first column, we observe that even very tight specular lobes are well-estimated, as evident from the lobe's compactness in the relighted image. The first row is imaged using an iPhone 6s, the second row with a Huawei P9 and the last three rows with a Lenovo Phab 2. Even though we do not calibrate the mobile phone, our network generalizes quite well to new devices.

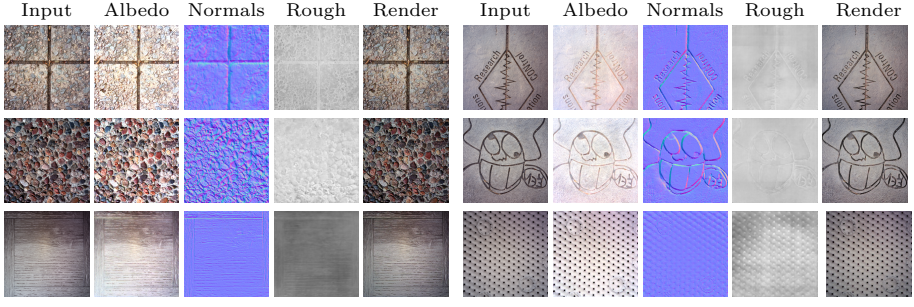

Fig. 6. BRDF estimation results on real data acquired in bright outdoor illumination.

Real data results in outdoor illumination Although we train out network with indoor lighting plus flash light, our method generalizes reasonably well to outdoor illumination. We acquire images in bright outdoor conditions with the flash enabled. We use the same network `clsCRF-pt` from the main paper, that assumes dominant point and environment lighting. The SVBRDF estimation results are

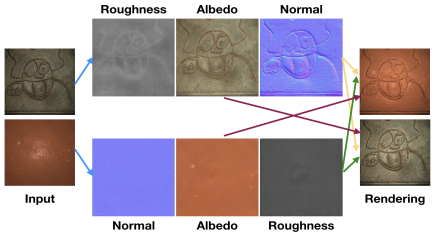

Fig. 7. A material editing example. Having reconstructed the SVBRDF and normals of the two samples, we swap the original geometry and material properties, then relight under novel illumination.

shown in Figure 6. While ground truth is not available, the recovered normal maps and BRDF terms are plausible, with well-delineated details. We believe that our dataset augmentation techniques have been effective in handling this scenario.

Material Editing We can edit the reconstructed SVBRDFs by transferring material properties. Figure 7 shows an example where we transfer BRDF properties across different material types and render in a novel lighting condition.

Visualization for relighting on a sphere under oblique illumination For another visualization of the normal and BRDF estimation on real data, we render the estimated material on a sphere illuminated under an oblique lighting direction that is very different from the input lighting. Recall that we only use an approximately planar patch of material as input. The BRDF estimation and relighted sphere are illustrated in Figure 5. We observe that the appearance of the sphere even under a novel lighting direction is quite reasonable.

Comparison with [1] on [1]'s dataset In the main paper, we compare the method of [1] on our data (Table 4 and Figure 7). A comparison on the data from [1] is complicated by errors in their provided ground truth. We illustrate this with a few examples in Figure 9. Note that while the plastic samples with circular bumps have been rotated, the ground truth normals should have pointed in the same direction. Similarly, for the wood samples, note that the visualized colors of normal directions in the grooves are the same, even though the input has been rotated. It seems that the same rotation has been applied to both diffuse and normal maps for data augmentation, without taking geometry into account.

We nevertheless try our model on test data from [1] and report the results in Table 1. We evaluate only diffuse albedos and normals, since microfacet models in the two works are different. Since [1] does not report quantitative numbers, we use their provided models to obtain the comparative numbers. For diffuse albedo, our method outperforms [10] for all material types without fine-tuning on their data. Our normal errors are "larger", but this comparison is not meaningful due to errors in their ground truths. Qualitative results are in Figure 8.

	Wood		Metal		Plastic		Summary	
	[1]	Ours	[1]	Ours	[1]	Ours	[1]	Ours
Albedo	0.0022	0.0019	0.0040	0.0035	0.0033	0.0007	0.0031	0.0020
Normal	0.0104	0.0173	0.0186	0.0187	0.0127	0.0103	0.0137	0.0162

Table 1. Quantitative comparison with [1] using its dataset, without fine-tuning. Please note the numbers are only representative, due to errors in ground truth of [1].

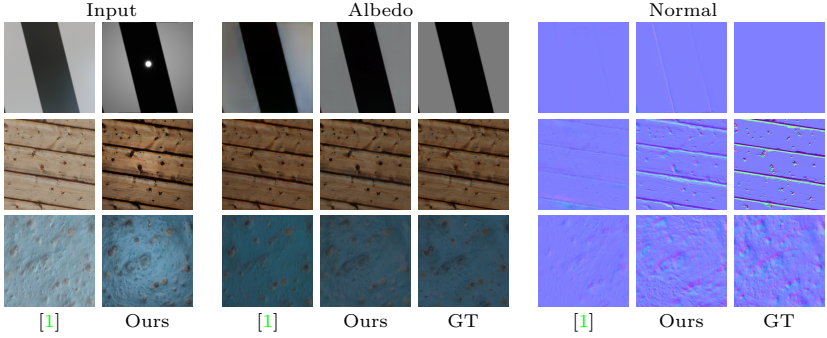

Fig. 8. Qualitative comparison with [1]. The diffuse colors have been normalized as in [1] for a fair comparison. We do not fine-tune our network on the dataset from [1].

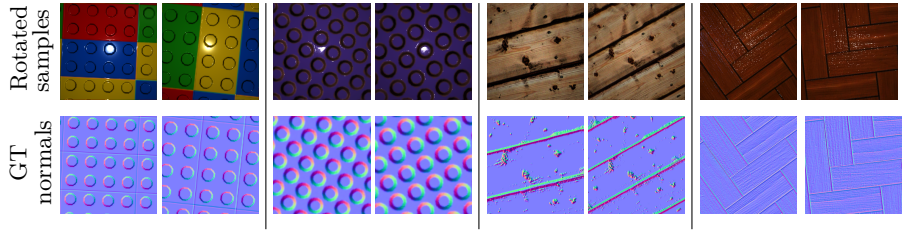

Fig. 9. Examples of errors in ground truth normals in the dataset of [1].

3 Microfacet BRDF Model

We use the microfacet BRDF model proposed in [3]. Let $\mathbf{d}_i$, $\mathbf{n}_i$, r_i be the diffuse color, normal and roughness, respectively, at pixel i and $I(\mathbf{d}_i, \mathbf{n}_i, r_i)$ be its intensity observed by the camera. Our BRDF model is defined as

$$I(\mathbf{d}_i, \mathbf{n}_i, r_i) = \mathbf{d}_i + \frac{D(\mathbf{h}_i, r_i)F(\mathbf{v}_i, \mathbf{h}_i)G(\mathbf{l}_i, \mathbf{v}_i, \mathbf{h}_i, r_i)}{4(\mathbf{n}_i \cdot \mathbf{l}_i)(\mathbf{n}_i \cdot \mathbf{v}_i)}, \tag{1}$$

where $\mathbf{v}_i$ and $\mathbf{l}_i$ are the view and light directions, while $\mathbf{h}_i$ is the half angle vector. Further, $D(\mathbf{h}_i, r_i)$, $F(\mathbf{v}_i, \mathbf{h}_i)$ and $G(\mathbf{l}_i, \mathbf{v}_i, \mathbf{h}_i, r_i)$ are the distribution, Fresnel and

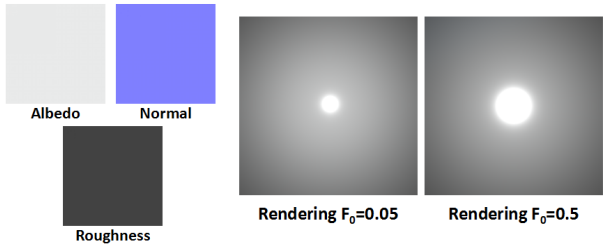

Fig. 10. An aluminum material rendered with different F_0. We obersve that when rendering with $F_0 = 0.5$ the area of specular highlight is much larger and better matches appearnaces of metals in the real world. For all other materials, we use $F_0 = 0.05$ as the most reasonable value.

geometric terms, respectively, which are defined as

$$D(\mathbf{h}_i, r_i) = \frac{\alpha_i^2}{\pi((\mathbf{n}_i \cdot \mathbf{h}_i)^2(\alpha_i^2 - 1) + 1)^2} \tag{2}$$

$$\alpha_i = r_i^2 \tag{3}$$

$$F(\mathbf{v}, \mathbf{h}) = (1 - F_0)2^{(-5.55473(\mathbf{v}\cdot\mathbf{h}) - 6.98316)\mathbf{v}\cdot\mathbf{h}} + F_0 \tag{4}$$

$$G(\mathbf{l}, \mathbf{v}, \mathbf{n}) = G_1(\mathbf{v}, \mathbf{n})G_1(\mathbf{l}, \mathbf{n}) \tag{5}$$

$$k_i = \frac{(r_i + 1)^2}{8} \tag{6}$$

$$G_1(\mathbf{v}, \mathbf{n}) = \frac{\mathbf{n} \cdot \mathbf{v}}{(\mathbf{n} \cdot \mathbf{v})(1 - k) + k} \tag{7}$$

$$G_1(\mathbf{l}, \mathbf{n}) = \frac{\mathbf{n} \cdot \mathbf{l}}{(\mathbf{n} \cdot \mathbf{l})(1 - k) + k}\,, \tag{8}$$

with F_0 the specular reflectance at normal incidence. For a dielectric material, the value of F_0 is determined by the index of refraction η:

$$F_0 = \frac{(1 - \eta)^2}{(1 + \eta)^2}. \tag{9}$$

For a conductor material, it is determined by the index of refraction η and the absorption coefficient κ:

$$F_0 = \frac{(1 + \eta)^2 + \kappa^2}{(1 - \eta)^2 + \kappa^2}. \tag{10}$$

When rendering our dataset, we set $F_0 = 0.5$ for *metal* and $F_0 = 0.05$ for other kinds of materials. Figure 10 shows an example of smooth aluminum material rendered with $F_0 = 0.05$ and $F_0 = 0.5$. We observe that the material rendered with $F_0 = 0.5$ has a much larger area of specular highlight, which matches appearances of metals in practice.

4 Details of Continuous DCRFs

We use continuous densely connected conditional random fields (DCRFs) for post-processing to remove artifacts caused by saturated highlights and noise in the prediction of the neural network [4,5]. We customize the DCRFs to better suit our problem of spatially-varying BRDF reconstruction. The distinguishing factor for our DCRF construction is the design of spatially varying weight maps that allow incorporating domain specific knowledge into the CRF inference. In the following, we will discuss the design and the intuition behind the usage of the weight map, as well as the details of training and inference for the DCRF.

Weight Maps of DCRFs We first discuss the DCRF for diffuse albedo prediction. Its energy function is defined as

$$\min_{\{\mathbf{d}_i\}}: \sum_{i=1}^{N} \alpha_i^d (\mathbf{d}_i - \hat{\mathbf{d}}_i)^2 + \sum_{i,j}^{N} (\mathbf{d}_i - \mathbf{d}_j)^2 \Big(\beta_1^d \kappa_1(\mathbf{p}_i; \mathbf{p}_j) + \beta_2^d \kappa_2(\mathbf{p}_i, \bar{\mathbf{I}}_i; \mathbf{p}_j, \bar{\mathbf{I}}_j) + \beta_3^d \kappa_3(\mathbf{p}_i, \hat{\mathbf{d}}_i; \mathbf{p}_j, \hat{\mathbf{d}}_j) \Big). \tag{11}$$

Here, the coefficient α_i^d is spatially varying. A larger α_i^d indicates greater confidence in the prediction from the neural network. Since we use a colocated point light source for illumination, an observation is that saturations caused by the specular highlight are usually in the middle of the image. Another observation is that since the flash illumination is white in color, the saturated pixels are usually white, which means the minimum of their RGB values will be large. Therefore, for regions near the center of the image or regions with specular highlights, we should have a smaller unary weight so that the DCRF may smooth out the artifacts. Based on these two observations, we define the weight map for the unary term α_i^d as

$$\alpha_i^d = \alpha_{i0}^d \max(1 - \exp(-\frac{\mathbf{p}_i^2}{\sigma_{d0}^2}), 1 - \exp(-\frac{(c_i^{min} - 1)^2}{\sigma_{d1}^2})) + \alpha_{i1}^d, \tag{12}$$

where c_i^{min} is the minimum of the three color channels at pixel i:

$$c_i^{min} = \min(R_i, G_i, B_i). \tag{13}$$

Here, α_{i0}^d and α_{i1}^d are two learnable parameters. We set $\alpha_{i1}^d = 0$ and $\alpha_{i0}^d = 1$ at the beginning of the training process. We set $\sigma_{d0} = 0.5$ and $\sigma_{d1} = 0.08$ through the whole training process. Figure 11 shows examples of the weight map for diffuse albedo prediction.

For normal prediction, we do not observe such strong correlation between the prediction error and the position or intensity of the image. Therefore, we just set

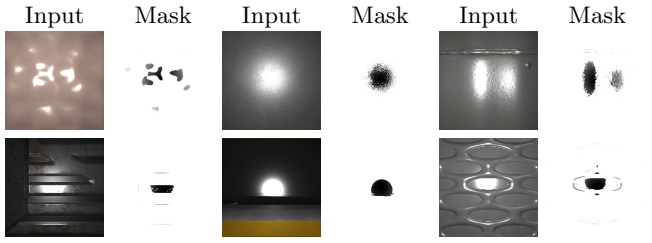

Fig. 11. The spatially varying weight α_i^d for the DCRF of diffuse albedo prediction.

a uniform weight for every pixel in the image. The energy function is defined as

$$\min_{\{\mathbf{n}_i\}}: \sum_{i=1}^{N} \alpha^n (\mathbf{n}_i - \hat{\mathbf{n}}_i)^2 + \sum_{i,j}^{N} (\mathbf{n}_i - \mathbf{n}_j)^2 \Big(\beta_1^n \kappa_1(\mathbf{p}_i; \mathbf{p}_j) + \beta_2^n \kappa_2(\mathbf{p}_i, \Delta\mathbf{d}_i; \mathbf{p}_j, \Delta\mathbf{d}_j) \Big), \tag{14}$$

where α^n, β_1^n and β_2^n are learnable parameters that trade-off relative confidences in the unary, a pairwise smoothness prior and a prior on correlation between normals and albedo boundaries.

Finally, for roughness prediction, the energy function is defined as

$$\min_{\{r_i\}}: \sum_{i=1}^{N} \alpha_{i0}^r (r_i - \hat{r}_i)^2 + \alpha_{i1}^r (r_i - \tilde{r}_i)^2 + \sum_{i,j}^{N} (r_i - r_j)^2 \Big(\beta_1 \kappa_1(\mathbf{p}_i; \mathbf{p}_j) + \beta_2 \kappa_2(\mathbf{p}_i, \mathbf{d}_i; \mathbf{p}_j, \mathbf{d}_j) \Big), \tag{15}$$

where $\hat{r}_i$ is the prediction from the network and $\tilde{r}_i$ is the prediction from a grid search. We find that the prediction from grid search is usually only accurate near the glossy regions, which means these regions should have a larger α_{i1}^r. Therefore, we define the weight map to be

$$\alpha_{i1}^r = \max(\exp(-\frac{\mathbf{p}_i^2}{\sigma_{r0}^2}), \exp(-\frac{c_m^i - 1}{\sigma_{r1}^2})), \tag{16}$$

where α_{i0}^r is constant across the whole image. Both α_{i0}^r and α_{i1}^r can be learned through back propagating the gradient. We set $\sigma_{r0} = 0.5$ and $\sigma_{r1} = 0.2$.

Hyperparameters for Training And Inference In order to increase the capacity of the DCRF model, we learn different sets of BRDF parameters for each type of material. During both inference and training time, we average the DCRF coefficients according to the output of our material classifier. Let $\{\theta_i\} = \{\{\alpha_i\}, \{\beta_i\}\}$ be the DCRF coefficients for one material. To enhance the robustness of our method, we re-parameterize the coefficients as

$$\bar{\theta_i} = \frac{\theta_i}{\sum_i \theta_i}. \tag{17}$$

We clip the DCRF coefficients to always be positive. We use the Adam optimizer to optimize the coefficients. The learning rate is set to 2×10^{-4} and we reduce it by half after every 2000 iterations. We adopt the method in [5] to train our DCRF model. The batch size is set to 32. We train the DCRF for diffuse albedo prediction over 4000 iterations and the DCRF for roughness and normal prediction over 3000 iterations. The standard deviations of Gaussian smooth kernels for the three DCRFs are shown in Table 2.

Gaussian Kernels of DCRF for Diffuse Albedo			
	$\mathbf{p}_i$	$\bar{\mathbf{I}}_i$	$\mathbf{d}_i$
κ_1	0.04	-	-
κ_2	0.06	0.2	-
κ_3	0.06	-	0.1

Gaussian Kernels of DCRF for Normal Map		
	$\mathbf{p}_i$	$\Delta\mathbf{d}_i$
κ_1	0.03	-
κ_2	0.06	0.1

Gaussian Kernels of DCRF for Roughness Map		
	$\mathbf{p}_i$	$\mathbf{d}_i$
κ_1	0.04	-
κ_2	0.06	0.2

Table 2. Standard deviations of the Gaussian smoothing kernels of the DCRFs for diffuse albedo, normal and roughness prediction.

5 Details of Dataset

In experiments, besides rotating and cropping the original high resolution spatially-varying materials, another important data augmentation is to scale the BRDF parameters for each patch before rendering them into images. For diffuse albedo, we uniformly sample scale coefficients in the range 0.8 to 1.4. For normal map, we sample the scale coefficients in the same way, apply the coefficients to the x and y components, then normalize the normal vector to be of unit length. For roughness, we sample the scale coefficients from a Gaussian distribution centered at 1, with standard deviation equal to 0.2. Empirically, we observe that such data augmentation can greatly improve the generalization ability of the network. For example, simply scaling the roughness parameter for each patch decreases the validation error for roughness prediction by 15%.

6 Video Comparison

In the supplementary material, we include a video clip in which we compare the relighting results using ground-truth BRDF parameters(*Ground truth*), our estimated BRDF parameters (*Our result*), BRDF estimation from [1](*Li et al.*) and from Allegorithmic substance B2M(*Bitmap2Material*), a commercial software for single image material capture. We compare 4 kinds of materials, specular stone, metal, wood and plastic. All of them are from our synthetic dataset. The inputs to our network are rendered under the illumination of point light source and environment map while the inputs to [1] and Allegorithmic substance B2M are rendered with environment map only. From the video clip, we can see that our relighting results are very close to the ground-truth, while [1] does not recover the specular highlight well. *Bitmap2Material* predicts BRDF parameters based on simple heuristics and is very sensitive to image gradients. Note that there are watermarks on the results of *Bitmap2Material*.

References

1. Li, X., Dong, Y., Peers, P., Tong, X.: Modeling surface appearance from a single photograph using self-augmented convolutional neural networks. ACM Trans. Graph. **36**(4) (July 2017) 45:1–45:11 1, 2, 3, 5, 6, 11
2. Hui, Z., Sankaranarayanan, A.C.: A dictionary-based approach for estimating shape and spatially-varying reflectance. In: International Conference on Computational Photography (ICCP). (2015) 2
3. Karis, B., Games, E.: Real shading in unreal engine 4. Proc. Physically Based Shading Theory Practice (2013) 6
4. Ristovski, K., Radosavljevic, V., Vucetic, S., Obradovic, Z.: Continuous conditional random fields for efficient regression in large fully connected graphs. In: AAAI. (2013) 8
5. Xu, D., Ricci, E., Ouyang, W., Wang, X., Sebe, N.: Multi-scale continuous crfs as sequential deep networks for monocular depth estimation. arXiv preprint arXiv:1704.02157 (2017) 8, 10