Deep SVBRDF Estimation from Single Image under Learned Planar Lighting

Lianghao Zhang
Tianjin University
China
opoiiuiouiuy@tju.edu.cn

Fangzhou Gao
Tianjin University
China
gaofangzhou@tju.edu.cn

Li Wang
Tianjin University
China
li_wang@tju.edu.cn

Minjing Yu*
Tianjin University
China
minjingyu@tju.edu.cn

Jiamin Cheng
Tianjin University
China
cjm@tju.edu.cn

Jiawan Zhang*
Tianjin University
China
jwzhang@tju.edu.cn

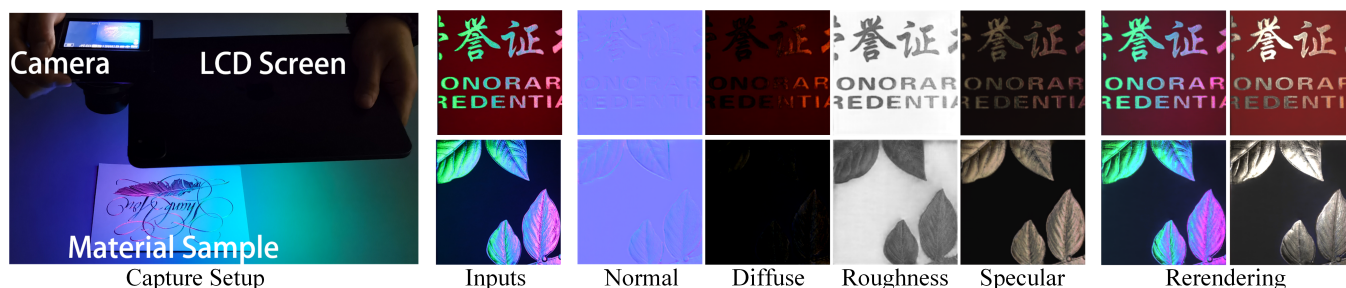| Capture Setup | Inputs | Normal | Diffuse | Roughness | Specular | Rerendering |

Figure 1: We propose a method that estimates SVBRDF from single image casually captured under a planar light source. The key idea of our method is to maximize the capture efficiency of single image using a learned lighting pattern and a proposed capture configuration. The material sample is captured with only a camera and a RGB LCD screen and without careful calibration. Here we show two SVBRDFs reconstructed by our method from a single image lit with the learned lighting pattern.

ABSTRACT

Estimating spatially varying BRDF from a single image without complicated acquisition devices is a challenging problem. In this paper, a deep learning based method was proposed to improve the capture efficiency of single image significantly by learning the lighting pattern of a planar light source, and reconstruct high-quality SVBRDF by learning the global correlation prior of the input image. In our framework, the lighting pattern optimization is embedded in the training process of the network by introducing an online rendering process. The rendering process not only renders images online as the input of network, but also efficiently back propagates gradients from the network to optimize the lighting pattern. Once trained, the network can estimate SVBRDFs from real photographs captured under the learned lighting pattern. Additionally, we describe an onsite capture setup that needs no careful calibration to capture the material sample efficiently. In particular, even a cell phone can be used for illumination. We demonstrate on synthetic and real data that our method could recover a wide range of materials from a single image casually captured under the learned lighting pattern.

CCS CONCEPTS

• **Computing methodologies** → **Reflectance modeling**; Computer vision.

KEYWORDS

Appearance Capture, SVBRDF, Single Image, Lighting pattern

*Corresponding authors.

1 INTRODUCTION

Accurately capturing the reflectance of real-world materials has been a longstanding problem in computer graphics and computer vision. The material properties of an opaque surface can be modeled as a 6D function called spatially varying bidirectional reflectance distribution function (SVBRDF). Traditional methods [Dana et al.

1999; Holroyd et al. 2010] devise sophisticated hardware for the exhaustive sampling of the six-dimensional space. But the customized hardware and tedious acquisition restrict the application.

Following the success of learning-based methods in computer vision, a growing body of methods estimate SVBRDF from a few or even a single image. Several of these methods [Li et al. 2017; Martin et al. 2022; Ye et al. 2018] attempted to reconstruct reflectance under environment lighting, yet imposed strong assumptions on SVBRDF, such as homogeneous or fixed material properties. Most methods [Deschaintre et al. 2018; Gao et al. 2019; Vecchio et al. 2021] employ active point lighting during capturing to use the strong specularity cues to estimate full SVBRDF. However, considering lighting domain sampling, the sparsity of directions sampled by point lighting (low sampling frequency) will incur aliasing when recovering the materials with narrow specular reflectance lobes (high appearance frequency). Thus, shiny surfaces, such as polished metal and ceramic tile, are challenging for point lighting-based methods [Guo et al. 2021].

Illuminating the material with lighting patterns is a common practice to improve the sampling efficiency. In contrast to using a single point light source, lighting patterns could offer a larger sampling subspace in lighting space. Many methods have been proposed to improve the capture efficiency with various lighting patterns, such as basis illumination [Aittala et al. 2013; Tunwattanapong et al. 2013], linear lighting [Chen et al. 2014; Ren et al. 2011] or learned lighting patterns [Kang et al. 2018]. However, as additional lighting information is packed into a single pixel, tens even hundreds of input photographs are required by these methods to disentangle the lighting and BRDF parameters. Besides, the acquisition process also usually needs careful calibration.

For common users, capturing reflectance features efficiently in fewer images with a simple acquisition process is more convenient. When only a single image is available, the designs of optimal lighting pattern and corresponding SVBRDF reconstruction algorithm are challenging if no explicit prior is posed on the SVBRDF of the captured sample. Existing methods [Deschaintre et al. 2018; Guo et al. 2020] learned global correlation prior among pixels using deep learning to reconstruct full SVBRDF from a single image. However, these methods suffer from the sampling sparsity caused by point lighting. Considering the acquisition process, planar lighting with lighting patterns is more suitable for casual capture to efficiently capture material properties because it could be easily implemented by a consumer device on everyone's desk, such as an Apple iPad or even a cell phone. Several methods [Aittala et al. 2013; Wang et al. 2011] proposed setups based on planar lighting for practical reflectance acquisition. However, careful calibration is still needed to register the camera, material sample, and planar light source, which is far beyond the capability of novice users.

In this paper, we propose a novel method that reconstructs high-quality SVBRDF from a single image casually captured under a planar light source, see Fig. 1. Our method introduces a deep learning-based framework to jointly learn the lighting pattern for acquisition and the global correlation prior for reflectance reconstruction. Specifically, the framework comprises a rendering process that embeds lighting pattern optimization in the training process and an image-to-image network to reconstruct full SVBRDF from a rendered image in training or captured image in testing utilizing

learned global correlation prior. Moreover, our method also describes an onsite capture setup that captures the material sample efficiently and needs no careful calibration. This setup comprises a pair of relatively fixed LCD screen and camera. A field-of-view (FOV) range of the camera is further derived to efficiently capture the most prominent reflectance features of the material sample. In this FOV range, our method is robust to the capture perturbations range caused by the lack of calibration. Extensive comparisons and experiments demonstrate that the proposed method can reconstruct a wide range of materials from a single photograph. The produced results are substantially superior to prior methods due to the following key technique contributions.

- An SVBRDF reconstruction framework that jointly learns the lighting pattern for acquisition and global correlation prior for SVBRDF estimation from a single photograph.
- An onsite capture setup that takes photographs efficiently and needs no careful calibration.

2 RELATED WORKS

Significant efforts have been made in capturing real-world materials. First, we discuss the SVBRDF reconstruction methods that take a single or a few photographs captured by standard cameras as inputs. Motivated by these reconstruction methods, those that estimate material properties based on lighting patterns are then discussed briefly.

2.1 SVBRDF from Sparse Images

Following the success of deep learning in various areas, a series of works attempts to reconstruct SVBRDF from sparse images.

In the context of multiple images, researchers have proposed a max-pooling-based network [Deschaintre et al. 2019] and latent spaces [Gao et al. 2019; Guo et al. 2020] in an inverse rendering framework to reconstruct SVBRDFs. Besides, some methods reconstruct large planar samples using detailed additional images [Deschaintre et al. 2020] or video [Ye et al. 2021]. However, registering multiple photographs usually needs careful calibration and is complicated for novice users.

For practicality, many methods have been proposed to estimate SVBRDF from a single image. Some methods assume the unknown lighting. Li et al. [2017] proposed a training strategy called *self-augmentation* to reduce the amount of labeled data required for training CNNs. Ye et al. [2018] further eliminated the requirement of labeled data. But both methods enforced strict spatial priors, that is, the specular albedo and roughness must be homogeneous. Martin et al. [2022] recently proposed a method to capture materials in the wild. However, their method cannot estimate the specular albedo map and approximate the roughness from the predicted normal.

Other methods ease the burden of estimation from a single image by using the strong specularity cues captured under active flash lighting. Several methods [Aittala et al. 2016; Henzler et al. 2021; Wen et al. 2022; Zhao et al. 2020] reconstruct SVBRDFs by assuming the material is stationary, which is a strong constraint.
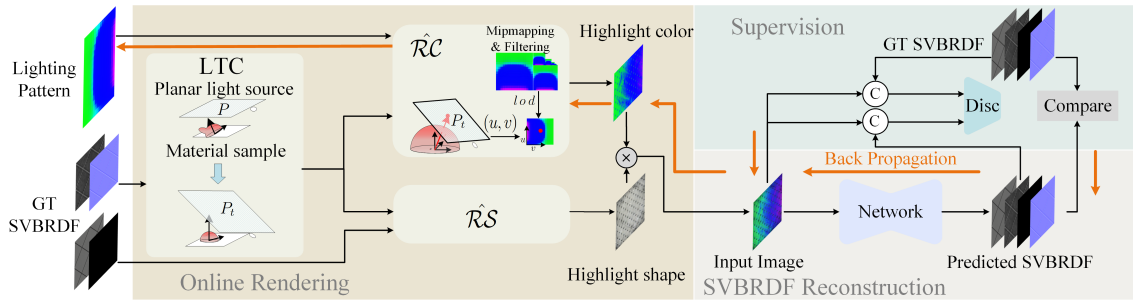
Figure 2: Our framework reconstructs SVBRDF from a single image lit with a planar light source. When training, the rendering process takes the ground truth SVBRDF as input and renders an image online. First, the BRDF distribution is transformed into clamped cosine distribution. The rendered image is then fed to the network for reconstructing the SVBRDF. Finally, under the supervision of the ground truth, the network parameters and lighting pattern are optimized jointly.

Li et al. [2018] proposed a CNN network to estimate SVBRDF considering the environment lighting but assuming fixed Fresnel reflectance. These methods assume explicit prior on SVBRDFs to simplify the problem. Several methods assume no explicit constraint on SVBRDFs to reconstruct more general materials. Some of these methods extract complementary features from distant regions [Deschaintre et al. 2018] or areas without highlight pollution [Guo et al. 2021] to handle the saturated pixels because of point lighting. Zhou et al. [2021] proposed a hybrid train strategy to narrow the data distribution gap between synthetic and real images. They introduced the generative adversarial loss for more details similar to SurfaceNet [2021]. Zhou et al. [2022] avoided overfitting in the test-time optimization to reproduce the appearance accurately. However, these methods suffer from the shortcoming of point lighting because most reflectance information is lacking in a single flash photograph. Instead, we employ planar lighting to efficiently pack additional information into a single image and reconstruct full SVBRDF.

2.2 Reflectance acquisition based on Lighting Patterns

Methods in this category illuminate material samples with different lighting patterns and reconstruct reflectance from captured measurements. Tunwattanapong et al. [2013] designed a setup based on a rotatable arc with controllable LEDs and illuminated the object with spherical harmonic illumination. Assuming the sample is planar, some methods [Ghosh et al. 2007; Nam et al. 2016] employed hemispherical illumination and designed various lighting patterns to capture reflectance efficiently. To maximize the capture efficiency, Kang et al. [2018] recently proposed a method that optimized the lighting patterns automatically rather than hand deriving. Kang et al. [2021; 2019] further extended this method to non-planar objects. Nevertheless, the capture setups of these methods required customized hardware.

Riviere et al. [2017] capture reflectance under natural lighting using polarization cues. Linear lighting has been employed to scan isotropic [Gardner et al. 2003; Ren et al. 2011] and anisotropic [Chen et al. 2014] material samples. The proposed method is most similar to those using planar lighting. Wang et al. [2011] designed a step-edge lighting pattern to capture dual-level appearance from

a single image, yet assuming homogeneous BRDF. Some methods [Ghosh et al. 2010; J. et al. 2015; Nogue et al. 2022] reduced the ambiguity by using gradient lighting patterns and polarization. Li et al. [2019] combined step-edge and gradient lighting to estimate SVBRDF from two images, but assuming the SVBRDF is piecewise. Aittala et al. [2013] proposed a portable capture setup to capture SVBRDF comprising a near-field LCD panel and a camera. However, their setup is carefully calibrated and the reconstruction needs hundreds of photographs as inputs. These hand-derived and predefined planar lighting patterns are neither suitable for SVBRDF reconstruction from a single image nor validated to be optimal on various test samples. Ma et al. [2021] optimized lighting pattern like Kang et al. [2018], yet requested a complex acquisition procedure. Inspired by Kang et al. [2018], we also jointly optimize the lighting pattern and network parameters while only taking a single image as input. They estimate per-point SVBRDF independently, which will not work for single-image SVBRDF reconstruction because of the lack of reflectance measurements and careful calibration. Different from them, we propose a rendering process that renders SVBRDFs online in the training process for global correlation prior learning, which bridges the spatial variances in appearance and lighting optimization.

3 METHOD

3.1 Problem Formulation and Method Overview

Our goal is to reconstruct high-quality reflectance from a single photograph of a near-planar surface lit by a planar light source. We assume the material properties of the planar surface for measurement can be well represented by Cook–Torrance [Cook and Torrance 1981] BRDF model. Hence, an SVBRDF comprises four material parameter maps $s := (n, k_d, r, k_s)$: diffuse albedo k_d, surface normal n, roughness r, and specular albedo k_s, for each point x on the surface. The planar light source is represented by a rectangle P with a lighting pattern L_i. According to the classical rendering process $\mathcal{R}(\cdot)$ of polygonal lighting, the intensity of each pixel $I(x)$ in the captured photograph I can be modeled as follows:

$$I(x) = \mathcal{R}(L_i, s) = \int_P f_r(\omega_i, \omega_o, s, x) L_i(\omega_i, x)(n(x) \cdot \omega_i) d\omega_i, \quad (1)$$

where $f_r(\cdot)$ is the BRDF function, ω_i and ω_o are the incident and view directions, respectively.

A simple method to reconstruct SVBRDF from a single photograph I is to train an image-to-image network G with parameters θ, which learns the global correlation prior and estimates the material maps $s = G_\theta(I)$. Given the ground truth material parameters s_{gt}, the network parameters are optimized to minimize the designed loss function $\mathcal{L}(\cdot, \cdot)$ as follows:

$$\theta^* = \arg\min_\theta \mathcal{L}(G_\theta(I), s_{gt}). \qquad (2)$$

However, we aim to leverage planar lighting to improve the capture efficiency of a single image. As previously mentioned, deriving the lighting pattern for acquisition is difficult with no explicit assumption on the measured material sample.

We address this problem by combining the rendering process $\mathcal{R}$ with an SVBRDF reconstruction network G to optimize the pattern L_i directly in the training process. More specifically, given the capture setup, the ground truth SVBRDF s_{gt} is used to render the input photograph of G with the lighting pattern L_i. The lighting pattern and network parameters are jointly optimized by evaluating the loss function. Hence, the training process of the proposed network can be expressed as:

$$L_i^*, \theta^* = \arg\min_{\theta, L_i} \mathcal{L}(G_\theta(\mathcal{R}(L_i, s_{gt})), s_{gt}). \qquad (3)$$

With this formulation, we propose a novel framework (Fig. 2) that comprises the following three modules: an online rendering process, an SVBRDF reconstruction network, and the supervision for the optimization of the lighting pattern and network parameters. The online rendering process will produce the input image of the network in the forward pass during training. The network will then reconstruct SVBRDF considering global correlations among the pixels of the input image. The gradients are propagated back to the network and passed to the lighting pattern through our online rendering process to minimize the discrepancy between predicted and ground-truth SVBRDF in supervision.

Providing the network with the image rendered physically is necessary (Eq. 1) for correct global correlations. Unfortunately, the integration is usually tackled by sampling directions over the hemisphere of a surface point and weighted sums of all results. This approach is costly, making it computationally infeasible for lighting pattern optimization. Instead, the proposed rendering process can approximate physically-based rendering and provide the input image online by leveraging and refining the recent progress in real-time rendering [Heitz et al. 2016] for SVBRDF reconstruction. Moreover, given the gradients back from the network defined on the rendered image, the online rendering process must efficiently propagate the gradients back to the lighting pattern. We also design the backpropagation pass by analyzing the quality and efficiency of the gradients through the rendering process. Integrated into modern autodifferentiation frameworks (AD), our rendering process can produce LoD-free gradients and update the lighting pattern conveniently.

Given the rendered input image, we employ the remarkable progress [Chen et al. 2022] in image restoration to learn the global correlation prior because the SVBRDF maps are essentially a multi-channel image. With the proposed online rendering process, the

network parameters and the lighting pattern are updated by the gradients propagated back from the supervision.

We also propose an onsite capture setup (Sec. 3.4) that not only captures reflectance efficiently but also constrains the capture perturbations for high-quality reconstruction. The setup describes the relative positions of lighting, camera, and material sample that need rough estimations of users. Given a real photograph $I_{L_i^*}$ lit by learned lighting pattern L_i^* captured using our setup, the SVBRDF can be directly reconstructed by the trained network $s = G_{\theta^*}(I_{L_i^*})$.

3.2 Online Rendering Process

We comprehensively describe the forward rendering and backpropagation passes of our online rendering process in this section.

Forward Rendering Pass. The proposed online rendering embeds the learnable lighting pattern in the forward pass of the training process by providing the network with rendered images. To achieve this goal, our forward rendering pass is built on the framework of textured polygonal rendering with linearly transformed cosine (TPR-LTC) [Heitz et al. 2016], which is proposed to approximate physically-based rendering from polygonal lights in real time. Using LTC transformation and assuming a constant lighting pattern ($L_i(\omega_i, x) = 1$), the integral over the planar domain P in Eq. 1 can be transformed to an integral of spherical cosine distribution $C(\omega_i)$ which can be solved analytically in real time. To render with heterogeneous lighting patterns, the rendering process is divided into **highlight shape** rendering $\hat{\mathcal{RS}}(\cdot)$ and **highlight color** rendering $\hat{\mathcal{RC}}(\cdot)$. $\hat{\mathcal{RS}}$ calculate Eq. 1 while assuming a constant lighting pattern and $\hat{\mathcal{RC}}$ is the division of Eq. 1 and $\hat{\mathcal{RS}}$ to reintroduce the influence of lighting pattern. This division defines a convolution between the lighting pattern L_i and a texture-space filter whose weight distribution is cosine-weighted BRDF value normalized over the polygonal domain P. After LTC transformation, the texture-space filter becomes $F(\omega_i) = \frac{C(\omega_i)}{\int_{P_t} C(\omega_i)d\omega_i}$ where P_t is transformed from the light rectangle P using same LTC transformation. The highlight color rendering then is approximated as lighting pattern filtering with level-of-details (LoDs) and texture fetching using fetch vectors in rendering.

However, the original TPR-LTC can cause artifacts in the rendered image when using polygonal lights with highly heterogeneous lighting patterns in rendering. Ideally, the texture fetching should fetch an appropriate predefined filter most matching $F(\omega_i)$ for rendering quality. But the fetched filters in the original TPR-LTC are different from $F(\omega_i)$ because the fetch vectors are expected to be always well-defined in texture space for arbitrary scene rendering. By contrast, our goal is to reconstruct SVBRDFs from real images given the relatively fixed capture setup. The network relies on the reflectance clues in the input image to predict SVBRDF, while the artifacts will mislead the network and lighting pattern with wrong global correlations. To address this problem, we propose a fetching strategy (Left in Fig. 3.) to fetch the filtered color in LoDs that ensures the rendered reflectance clues are sufficiently close to physically-based rendering. The key idea of our fetch strategy is to align the weight distribution peak of fetched filter and $F(\omega_i)$ because the peak is the most prominent feature of weight distribution with respect to material properties.
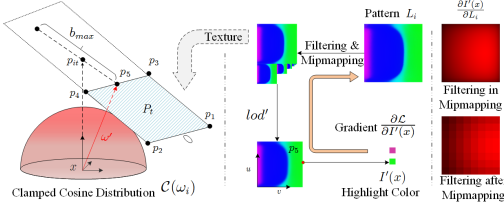
Figure 3: The forward and backward pass of the highlight color rendering of a surface point x. On the left, we illustrate the texture fetching when p_5 is selected. Then the highlight color is fetched by the fetching vector ω' and the selection of LoDs lod' from the mipmaps. The gradients are backpropagated to the pattern through texture fetching and mipmapping.

We sample the surface points p on P_t and select the point with the maximum weight value of $F(\omega_i)$. As the differential solid angle $d\omega_i$ can be expressed as a function of a differential area dA of lighting rectangle P_t [Drobot 2018], the weight distribution $F(\omega_i)$ is converted to $F(p)$:

$$F(p) = \frac{C(n_i \cdot \omega'(p))(n_i \cdot \omega'(p))}{d^2(p) \int_{P_t} C(\omega_i)d\omega_i},$$ (4)

where n_i is the normal of lighting planar, d is the distance from a point p on dA to the material surface point x and ω_i' means the direction $\overrightarrow{px}$.

Finding the maximum weight value in P_t is a non-linear optimization; thus, solving this problem is computationally unacceptable during training. Instead, we evaluate the weight value of five predefined points on P_t and select the point with the maximum weight value. The five predefined points comprise four vertices $\{p_1, ..., p_4\}$ of P_t and a point p_5 on P_t. The p_5 is related to the intersection p_{it} of lighting plane and the average direction of $C(\omega_i)$. When p_{it} is in the range of P_t, p_5 is equal to p_{it}. Otherwise, the point p_5 is calculated by directly clamping p_t to the range of P_t in texture space. The point p_5 represents the peak of $C(n_i \cdot \omega'(p))(n_i \cdot \omega'(p))$, while other points account for the variation of distance $d^2(p)$.

We apply a simple bias on the selection of LoD lod according to the distance from p_{it} to p_5, if p_5 is selected for fetching to avoid the ambiguity caused by the clamp:

$$lod' = lod + \frac{|\overrightarrow{p_{it}p_5}|}{b_{max}\frac{lod}{lod_{max}}}(lod_{max} - lod)$$ (5)

where lod_{max} is the max level of details and b_{max} is a threshold of distance $|\overrightarrow{p_{it}p_5}|$, beyond which the filter for lighting pattern prefiltering is a uniform filter. We still use the renormalized texture-space Gaussian following Heitz et al. [2016] for the weights of the filter. We argue that this approximation is sufficiently plausible for peak alignment because the distance variation is moderate for planar materials in our capture setup (Fig. 4). Furthermore, we implement the LoDs with mipmapping [Williams 1983] for online filtering compared with the offline lighting pattern filtering in real-time rendering.

The highlight shape rendering $\hat{RS} = \int_{P_t} C(\omega_i)d\omega_i$ can be calculated analytically in real time as the irradiance of the polygon P_t

[Baum et al. 1989]. However, this suffers from light leak artifacts due to the numerical instability of the transformation matrix in LTC near the grazing angle. The problem is especially acute in our case because we represent the surface variations with normal mapping instead of bumpy geometry. Thus, we apply a simple correction that simulates the negligible appearance near the grazing angle for most materials and inhibits the appearance when the incident or view angle is larger than $90°$. We rectify the rendered highlight shape by multiplying a weight calculated by a continuous function Δ about the view angle θ between ω_o and n:

$$\Delta(\theta) = \frac{e^{a \cdot cos\theta + b}}{e^{a \cdot cos\theta + b} + e^{-a \cdot cos\theta - b}}$$ (6)

The spatial variations of the lighting pattern are not constricted by the proposed rendering process and any lighting pattern can be recovered if desired by the optimization.

Backward Propagation Pass. In the backward pass of the framework, the backpropagation of gradients and searching space of the lighting pattern is decided by the online rendering. Following we design the backward propagation pass for efficient backpropagation and exploration of the lighting pattern.

First, the backpropagation efficiency of our rendering process will be limited if naively performing the texture-space filtering on each level MIP pyramid levels in $\hat{RC}$ (Filtering after Mipmapping) because the lower level of filtered LoDs than the selected one are not involved in rendering. The gradients that backpropagate through the rendering layer, thus, will have the same level of detail as the selected LoD. For instance, the gradients back through the rendering process will comprise many small uniform patches due to downsampling in mipmapping. Instead, we choose to generate high LoD by downsampling the filtered low LoD (Filtering in Mipmapping) to improve the efficiency of back propagation and produce LoD-free gradients. The key difference is that all lower filtered LoDs are involved in rendering, thus providing gradients at all levels. We visualize the gradients defined on the lighting pattern with two implementations of Filtered LoDs in the right of Fig 3.

Besides, the gradients from the diffuse component rendering defined on the lighting pattern only have minor variations and affect most areas of the lighting pattern since high LoDs are usually chosen for the rendering of diffuse components. By contrast, the specular component gradients can affect only a small area of lighting patterns in each step of training because the weight distributions of the specular lobe are more concentrated than that of the diffuse component. Consequently, the diffuse component gradients will limit the searching space due to the unbalance in the backpropagate efficiency of the two components. To search lighting patterns in a larger space, we truncate the gradients back from diffuse component rendering to maximize the backpropagate efficiency of the specular component.

Finally, calculating and gathering the gradients are essential to search lighting patterns using gradient-based optimization techniques. To achieve this, we leverage the Nvdiffrast [Laine et al. 2020], which is a high-performance and modular framework for differentiable rendering, to integrate proposed online rendering in AD frameworks such as TensorFlow and PyTorch.
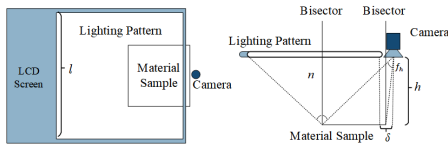
Figure 4: The top and front view of the capture setup. The camera is next to the LCD screen and the sample is placed in the range of the perpendicular bisectors of light planar edges and the camera.

3.3 SVBRDF reconstruction network

We pose SVBRDF reconstruction as an image-to-image translation task and leverage modern progress in computer vision to learn the global correlation prior from datasets and estimate SVBRDF. For the architecture of the reconstruction network for SVBRDF, we follow that of Chen et al. [2022], which is built for image restoration.

Loss Function. Our network is trained over a joint loss function that comprises a pixel-wise loss $\mathcal{L}_{pix}$ and a conditional adversarial loss $\mathcal{L}_{cadv}$:

$$\mathcal{L} = \lambda_{pix}\mathcal{L}_{pix} + \lambda_{cadv}\mathcal{L}_{cadv}, \tag{7}$$

where λ_{pix} and λ_{cadv} are hyper parameters for balancing the influence of each term. Herein, the first term measures the discrepancy of predicted and ground truth SVBRDF using l_1 normalization, and the second term aims to alleviate blurriness in results caused by pixel-wise loss. For adversarial training, following Vecchio et al. [2021], a simple patch discriminator D is employed to distinguish the predicted SVBRDF s and ground truth SVBRDF s_{gt} conditioned with the input image. We use WassersteinGAN loss [Arjovsky et al. 2017] as the second term to avoid mode collapse and reduce artifacts.

3.4 Acquisition Setup

In this section, we propose a simple acquisition setup for a novice user that can be implemented by common portable devices in daily life and needs no careful calibration. Our acquisition setup comprises a regular RGB LCD screen and a separate camera to capture the photograph. Color multiplexing could offer more optimization space and different light samples in different channels. This means more constraints for the reconstruction of channel-independent parameters (n, r). Besides, channel-dependent parameters (k_d, k_s) can also be well reconstructed since the pattern serves as a "base color pattern" and the color bias in appearance is caused by albedos.

For maximum user convenience, we assume that the camera optic axis is roughly perpendicular to the plane of material samples and the light planar is parallel to the material sample as illustrated in Fig. 4. Portable LCD screens are usually small (e.g., Apple iPad, cell phone). Therefore, we design a relative configuration to maximize the size of the material sample. For the expression simplicity, we assume the planar light source is a square that has the size $l \times l$, which can be easily extended to a rectangle. The camera is next to the LCD screen with a minor distance δ because of the physical camera size.

To efficiently capture the most prominent reflectance features with a planar light source, the material should be captured in mirror

directions [Aittala et al. 2013]. Thus, we satisfy this by assuming the material size is $\frac{l}{2} \times \frac{l}{2}$ and deriving a FOV range of the camera where the sample is captured in mirror directions. Specifically, the material sample is placed at the perpendicular bisectors of light planar edges and the camera. Given an arbitrary capture distance h from the camera to the sample, a FOV range (f_h, f_v) that satisfies the mirror capture configuration can be derived.

$$f_h \in (arctan\frac{\delta}{2h}, arctan\frac{\delta+l}{2h}), f_v \in (-arctan\frac{l}{2h}, arctan\frac{l}{2h}), \tag{8}$$

where f_h and f_v are the horizontal and vertical range respectively assuming the center of the image is $(0, 0)$. Notably, this configuration is weakly calibrated but is easy for users to estimate roughly. Furthermore, if the material sample has low roughness, then the FOV range that satisfies the mirror direction can be easily recognized by users to crop the captured photograph handily. In our capture setup, users can easily change the planar light source and camera and the network does not even need retraining if the light source and material sample keep the relative size ratio.

4 EVALUATION

We conduct qualitative and quantitative experiments in this section to validate our method on SVBRDF reconstruction.

4.1 Experiments

For comparison and ablation study, we use the training and test datasets provided by Deschaintre et al. [2018]. The SVBRDF reconstruction network and the discriminator are trained in an adversarial manner with a batch size of 6 for 400,000 iterations using Adam [Kingma and Ba 2014] optimizer. The weights of loss function are $\lambda_{pix} = 1, \lambda_{cadv} = 1e^{-3}$ in our experiments. The learning rate and other hyperparameters are set following Chen et al. [2022]. The lighting pattern for optimization has a size of 1024×1024 and is initialized with a uniform gray lighting pattern. We experimentally found that lighting pattern initialization has a minor influence on the final lighting pattern and accuracy of SVBRDF reconstruction.

We validate our method by comparing it with state-of-the-art methods on synthetic and real data. Specifically, we compare our method against RADN [Deschaintre et al. 2018], deep inverse rendering (DIR) [Gao et al. 2019], MaterialGAN [Guo et al. 2020], and Hybrid [Zhou and Kalantari 2021]. Besides, Guo et al. [2021] and Zhou et al. [2022] also suffer from the shortcomings of point lighting as mentioned in their failure cases. For fairness, a comparison against these methods is not performed, because the source codes are not publicly available. Moreover, we also compare our method with FFScanning [Ma et al. 2021] on the synthetic dataset, which also optimizes the lighting pattern for SVBRDF capturing using planar lighting. In the ablation study, we evaluate the impact of different lighting patterns and capture perturbations to validate the effectiveness of our method.

4.2 Results & Discussion

Results on Synthetic Data. We evaluate the estimated results and rerendered images of all methods using root mean squared error (RMSE). Moreover, we further evaluate rerendered images perceptually by learned perceptual image patch similarity (LPIPS). The quantitative results are reported in Table. 1. The error of predicted

Table 1: Quantitative comparison of reconstruction and re-rendering results between our methods and previous works. Column abbreviations correspond to "Diffuse", "Normal", "Roughness", "Specular", "Rendering".

Methods	Nrm.	Diff.	Rgh.	Spec.	Rend.	
					RSME	LPIPS
RADN	0.068	0.034	0.196	0.053	0.091	0.313
DIR	0.072	0.027	0.163	0.047	0.075	0.162
MaterialGAN	0.083	0.042	0.218	0.052	0.092	0.215
Hybrid	0.077	0.027	0.119	0.062	0.103	0.187
Ours	**0.033**	**0.011**	**0.035**	**0.021**	**0.042**	**0.100**

normal and roughness from our method are significantly lower than other methods. This finding is reasonable because the normal and roughness are constrained by the highlight shape across channels. As shown in Fig. 6, our method can reconstruct SVBRDFs of a wide range of materials. By contrast, because the specular reflectance of most surface areas is not captured, the materials with sharp specular reflectance are challenging to point lighting-based methods. For SVBRDFs with many surface variations, the planar lighting can provide light direction sampling directly and hint at the network using the lighting pattern with less ambiguity than point lighting. Besides, without the consideration of global correlations, only a single input image is very challenging for FFScanning even there is no capture perturbation in the input image.

Results on Real Data. All real images are captured by a consumer camera (Canon G7X Mark II). The planar light source in our method is implemented using an Apple iPad, and the collocated point light source in other methods is implemented by an extra flashlight placed next to the camera lens. The capture is aided by a simple paper frame (similar to Deschaintre et al. [2019]) to crop the same material sample patches for comparison. Fig. 7 compares the SVBRDFs reconstructed by all methods from real LDR photographs of three material samples. Because complex color and position calibrations are not involved in the capture, we adjusted the white balance of our rendered results to best match the flash photograph. SVBRDFs containing discontinuous material correlations are challenging to point light-based methods because of the sampling capacity of a point light source. By contrast, our method can reconstruct SVBRDFs much clearer than others, especially on normal and roughness, thanks to the additional information provided by the lighting pattern. Moreover, we also show two more results in Fig. 9 estimated from the same trained network and the automatically cropped input images according to the derived FOV range. The left material sample in the figure is illuminated by a cell phone. Our method is robust enough to estimate high-quality SVBRDF from automatically cropped input images.

Ablation study. Firstly, we validate the importance of lighting pattern optimization. We train the network on several datasets rendered with fixed, predefined lighting patterns comprising a point light, constant pattern, linear light, step-edge lighting pattern [Wang et al. 2011], and a hand-picked RGB pattern. The step-edge lighting pattern is proposed to capture the surface with statistically stationary variation and a homogeneous BRDF from a single image.

The hand-picked pattern contains low-frequency variations similar to the learned pattern and high-frequency variations for ablation. We also train and test the networks on datasets rendered with ray tracing using Mitsuba renderer [Jakob 2014] under all lighting patterns to further validate the efficiency of the learned lighting pattern. We show quantitative results in Table 3. and qualitative results in Fig. 8.

As shown in the table, our learned lighting pattern is the most efficient. The point lighting achieve the highest reconstruct error as expected because of the lowest sampling efficiency. Compared with other single-channel patterns, the learned pattern can achieve higher accuracy because of extra highlight color cues are provided in the input image as global correlations to the SVBRDF reconstruction network. Hence, the hand-picked pattern can achieve similar accuracy as the learned pattern, yet minor higher error because of the lack of joint optimization. Although there is little accuracy loss of the learned pattern on the raytraced dataset, it is enough for a common user to use our jointly trained network for SVBRDF reconstruction. The results in Fig. 10 show that the networks trained with online rendering and ray tracing produce similar results with the same photograph as input.

Table 2: Ablation study about lighting patterns. We show the quantitative results of the networks trained on datasets rendered with both online rendering and ray tracing. Since the point light rendering is physically-based, we do not provide another row in online rendering. The last row in online rendering shows the results predicted by the network jointly trained with a 256×256 lighting pattern for ablation. The data in column "Avg." corresponds to the average RSME value.

Methods	Nrm.	Diff.	Rgh.	Spec.	Avg.	Rend.	
						RSME	LPIPS
Online Rendering							
Cons	0.069	0.054	0.111	0.048	0.070	0.084	0.259
Linear	0.058	0.033	0.085	0.024	0.050	0.065	0.189
StepEdge	0.069	0.027	0.112	0.023	0.058	0.069	0.187
Picked	**0.032**	0.013	0.042	**0.021**	0.027	**0.041**	0.103
Ours	0.033	**0.011**	**0.035**	0.021	0.025	0.041	**0.100**
Ours(256)	0.035	0.012	0.037	0.024	0.027	0.042	0.107
Ray Tracing							
Point	0.071	0.022	0.161	0.033	0.072	0.076	0.251
Cons	0.049	0.027	**0.033**	0.020	0.032	0.064	0.197
Linear	0.058	0.031	0.095	0.027	0.053	0.063	0.175
StepEdge	0.060	0.017	0.055	0.022	0.038	0.059	0.171
Picked	0.034	0.013	0.040	**0.019**	0.026	0.043	0.110
Ours	**0.028**	**0.011**	0.036	0.021	**0.024**	**0.039**	**0.094**

Furthermore, we evaluate the impact of capture perturbations on synthetic data. Because there are ambiguities among lighting and camera configurations in the capture, we select three factors to evaluate comprising light source position, light planar direction, and light intensity. These factors shift, twist, and scale the global correlations of highlight color and shape in the input image, respectively. In Fig. 5, we visualize the relation of reconstruction accuracy with perturbations. The trends suggest that the normal and roughness estimation are mainly affected by shift and twist while the albedo is affected by the magnitude of global correlations. We argue that keeping the direction perturbations in a low range is easy for users.
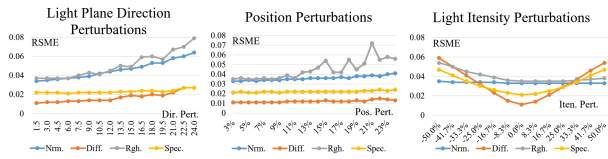
Figure 5: The impact of capture perturbations. In these diagrams, the horizontal coordinates represent the inclination angle of the lighting planar normal, the proportion of light position shift relative to the material size, and the proportion of lighting intensity bias relative to the intensity we used in training, respectively. We sample noises from a normal distribution that takes the horizontal coordinates as the sigma. Then the noise is applied as the according perturbation factor in testing.

In this circumstance, the position perturbations can be controlled by the derived FOV range. Hence, the main accuracy loss of our method is caused by the control of intensity. However, the intensity perturbations scale global correlation uniformly, which mainly affects albedo predictions. The results predicted with small-intensity perturbations can be corrected easily by adjusting the color balance.

Limitations. Our method is subject to several limitations. Different from other methods applying continuous lighting patterns [Aittala et al. 2013; Tunwattanapong et al. 2013], the material with almost perfectly specular reflectance properties is challenging to our method (Fig. 11). This problem can be potentially solved by careful calibration. Besides, as mentioned before, the albedo results recovered by our method have a global color bias according to the flash image due to the capture perturbations and lack of color calibration. Sometimes adjusting color balance fails because of other capture perturbations (last scene in Fig. 7). Probably the extra flash images should be provided to compensate for the color bias.

5 CONCLUSION

We propose an SVBRDF reconstruction method to maximize the capture efficiency of a causally captured image under a learned lighting pattern. In our method, we design a deep learning-based framework to jointly optimize the network parameters for learning global correlation prior and the lighting pattern for real photograph acquisition. Moreover, we also propose a convenient capture setup that captures the material efficiently in mirror directions. Extensive experiments demonstrate that the proposed method can produce accurate results on a wide range of materials.

ACKNOWLEDGMENTS

REFERENCES

Miika Aittala, Timo Aila, and Jaakko Lehtinen. 2016. Reflectance Modeling by Neural Texture Synthesis. *ACM Transactions on Graphics* 35, 4 (2016), 65.1–65.13.

M. Aittala, T. Weyrich, and J. Lehtinen. 2013. Practical SVBRDF capture in the frequency domain. *Acm Transactions on Graphics* 32, 4CD (2013), 1–12.

Martin Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein GAN. arXiv:1701.07875 [stat.ML]

D. R. Baum, H. E. Rushmeire, and J. M. Winget. 1989. Improving radiosity solutions through the use of analytically determined form-factors. In *Proceedings of the 16st Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1989*.

Guojun Chen, Yue Dong, Pieter Peers, Jiawan Zhang, and Xin Tong. 2014. Reflectance scanning: estimating shading frame and BRDF with generalized linear light sources. *ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2014* (2014).

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. 2022. Simple baselines for image restoration. *arXiv preprint arXiv:2204.04676* (2022).

Robert L. Cook and Kenneth E. Torrance. 1981. A reflectance model for computer graphics. *Acm Siggraph Computer Graphics* 15, 3 (1981), 307–316.

Kristin J Dana, Bram Van Ginneken, Shree K Nayar, and Jan J Koenderink. 1999. Reflectance and texture of real-world surfaces. *ACM Transactions On Graphics (TOG)* 18, 1 (1999), 1–34.

Valentin Deschaintre, Miika Aittala, Fredo Durand, George Drettakis, and Adrien Bousseau. 2018. Single-Image SVBRDF Capture with a Rendering-Aware Deep Network. *ACM Transactions on Graphics* 37, 4CD (2018), 128.1–128.15.

Valentin Deschaintre, Miika Aittala, Fredo Durand, George Drettakis, and Adrien Bousseau. 2019. Flexible SVBRDF Capture with a Multi-Image Deep Network. *Computer Graphics Forum* 38, 4 (2019).

Valentin Deschaintre, George Drettakis, and Adrien Bousseau. 2020. Guided Fine-Tuning for Large-Scale Material Transfer. *Computer Graphics Forum (Proceedings of the Eurographics Symposium on Rendering)* 39, 4 (2020). http://www-sop.inria.fr/reves/Basilic/2020/DDB20

Michal Drobot. 2018. Physically based area lights. In *GPU Pro 360 Guide to Lighting*. AK Peters/CRC Press, 177–210.

Duan Gao, Xiao Li, Yue Dong, Pieter Peers, Kun Xu, and Xin Tong. 2019. Deep Inverse Rendering for High-Resolution SVBRDF Estimation from an Arbitrary Number of Images. *ACM Trans. Graph.* 38, 4, Article 134 (July 2019), 15 pages. https://doi.org/10.1145/3306346.3323042

A. Gardner, C. Tchou, T. Hawkins, and Paul Debevec University of Southern California Institute for Creative Technologies Graphics Laboratory. 2003. Linear Light Source Reflectometry. *ACM Transactions on Graphics (TOG)* (2003).

A. Ghosh, S. Achutha, W. Heidrich, and M. O'Toole. 2007. BRDF Acquisition with Basis Illumination. In *IEEE International Conference on Computer Vision*.

A. Ghosh, T. Chen, P. Peers, C. A. Wilson, and P. Debevec. 2010. Estimating Specular Roughness and Anisotropy from Second Order Spherical Gradient Illumination. *Computer Graphics Forum* 28, 4 (2010), 1161–1170.

J. Guo, S. Lai, C. Tao, Y. Cai, and L. Q. Yan. 2021. Highlight-aware two-stream network for single-image SVBRDF acquisition. *ACM Transactions on Graphics* 40, 4 (2021), 1–14.

Yu Guo, Cameron Smith, Miloš Hašan, Kalyan Sunkavalli, and Shuang Zhao. 2020. MaterialGAN: reflectance capture using a generative SVBRDF model. *ACM Transactions on Graphics (TOG)* 39, 6 (2020), 1–13.

E. Heitz, J. Dupuy, S. Hill, and D. Neubelt. 2016. Real-time polygonal-light shading with linearly transformed cosines. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 41.1–41.8.

P. Henzler, V. Deschaintre, N. J. Mitra, and T. Ritschel. 2021. Generative Modelling of BRDF Textures from Flash Images. (2021).

Michael Holroyd, Jason Lawrence, and Todd Zickler. 2010. A coaxial optical scanner for synchronous acquisition of 3D geometry and surface reflectance. *ACM Transactions on Graphics (TOG)* 29, 4 (2010), 1–12.

J., Riviere, P., Peers, A., and Ghosh. 2015. Mobile Surface Reflectometry. *Computer Graphics Forum* (2015).

Wenzel Jakob. 2014. Mitsuba renderer 0.5. (2014). http://www.mitsuba-renderer.org/index_old.html

Kaizhang Kang, Zimin Chen, Jiaping Wang, Kun Zhou, and Hongzhi Wu. 2018. Efficient reflectance capture using an autoencoder. *ACM Transactions on Graphics* 37, 4CD (2018), 1–10.

Kaizhang Kang, Minyi Gu, Cihui Xie, Xuanda Yang, Hongzhi Wu, and Kun Zhou. 2021. Neural Reflectance Capture in the View-Illumination Domain. *IEEE Transactions on Visualization and Computer Graphics* (2021).

K. Kang, C. Xie, C. He, M. Yi, and H. Wu. 2019. Learning efficient illumination multiplexing for joint capture of reflectance and shape. *ACM Transactions on Graphics* 38, 6 (2019), 1–12.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

Samuli Laine, Janne Hellsten, Tero Karras, Yeongho Seol, Jaakko Lehtinen, and Timo Aila. 2020. Modular Primitives for High-Performance Differentiable Rendering. *ACM Transactions on Graphics* 39, 6 (2020).

Xiao Li, Yue Dong, Pieter Peers, and Xin Tong. 2017. Modeling surface appearance from a single photograph using self-augmented convolutional neural networks. *Acm Transactions on Graphics* 36, 4 (2017), 45.

Xiao Li, Peiran Ren, Yue Dong, Gang Hua, Xin Tong, and Baining Guo. 2019. Capturing Piecewise SVBRDFs with Content Aware Lighting. In *Advances in Computer Graphics: 36th Computer Graphics International Conference, CGI 2019, Calgary, AB,

Canada, June 17–20, 2019, Proceedings 36. Springer, 371–379.

Zhengqin Li, Kalyan Sunkavalli, and Manmohan Chandraker. 2018. Materials for Masses: SVBRDF Acquisition with a Single Mobile Phone Image. (2018).

X. Ma, K. Kang, R. Zhu, H. Wu, and K. Zhou. 2021. Free-form scanning of non-planar appearance with neural trace photography. *ACM Transactions on Graphics (TOG)* (2021).

Rosalie Martin, Arthur Roullier, Romain Rouffet, Adrien Kaiser, and Tamy Boubekeur. 2022. MaterIA: Single Image High-Resolution Material Capture in the Wild. *Computer Graphics Forum* 41, 2 (2022), 163–17715 pages. https://doi.org/10.1111/cgf. 14466

Giljoo Nam, Joo Ho Lee, Hongzhi Wu, Diego Gutierrez, and Min H. Kim. 2016. Simultaneous acquisition of microscale reflectance and normals. *Acm Transactions on Graphics* 35, 6 (2016), 185.

Emilie Nogue, Yiming Lin, and Abhijeet Ghosh. 2022. Polarization-imaging Surface Reflectometry using Near-field Display. In *Eurographics Symposium on Rendering. The Eurographics Association*, Vol. 2.

P. Ren, J. Wang, J. Snyder, T. Xin, and B. Guo. 2011. Pocket Reflectometry. *ACM Transactions on Graphics* 30, 4 (2011), 45.

Jérémy Riviere, Ilya Reshetouski, Luka Filipi, and Abhijeet Ghosh. 2017. Polarization imaging reflectometry in the wild. *ACM Transactions on Graphics (TOG)* 36, 6 (2017), 1–14.

B. Tunwattanapong, G. Fyffe, P. Graham, J. Busch, X. Yu, A. Ghosh, and P. Debevec. 2013. Acquiring reflectance and shape from continuous spherical harmonic illumination. *Acm Transactions on Graphics* (2013).

Giuseppe Vecchio, Simone Palazzo, and Concetto Spampinato. 2021. SurfaceNet: Adversarial SVBRDF Estimation from a Single Image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12840–12848.

C. P. Wang, N. Snavely, and S. Marschner. 2011. Estimating dual-scale properties of glossy surfaces from step-edge lighting. *Acm Transactions on Graphics* 30, 6CD (2011), 1–12.

Tao Wen, Beibei Wang, Lei Zhang, Jie Guo, and Nicolas Holzschuch. 2022. SVBRDF Recovery from a Single Image with Highlights Using a Pre-trained Generative Adversarial Network. In *Computer Graphics Forum*. Wiley Online Library.

Lance Williams. 1983. Pyramidal parametrics. In *Proceedings of the 10th annual conference on Computer graphics and interactive techniques*. 1–11.

Wenjie Ye, Yue Dong, Pieter Peers, and Baining Guo. 2021. Deep Reflectance Scanning: Recovering Spatially-varying Material Appearance from a Flash-lit Video Sequence. In *Computer Graphics Forum*, Vol. 40. Wiley Online Library, 409–427.

Wenjie Ye, Xiao Li, Yue Dong, Pieter Peers, and Xin Tong. 2018. Single Image Surface Appearance Modeling with Self-augmented CNNs and Inexact Supervision. *Computer Graphics Forum* 37, 7 (2018), 201–211.

Y. Zhao, B. Wang, Y. Xu, Z. Zeng, and N. Holzschuch. 2020. Joint SVBRDF Recovery and Synthesis From a Single Image using an Unsupervised Generative Adversarial Network. In *EGSR 2020*.

X. Zhou and N. K. Kalantari. 2021. Adversarial Single-Image SVBRDF Estimation with Hybrid Training. *Computer Graphics Forum: Journal of the European Association for Computer Graphics* 2 (2021), 40.

Xilong Zhou and Nima Khademi Kalantari. 2022. Look-Ahead Training with Learned Reflectance Loss for Single-Image SVBRDF Estimation. *ACM Transactions on Graphics (TOG)* 41, 6 (2022), 1–12.

Lianghao Zhang, Fangzhou Gao, Li Wang, Minjing Yu, Jiamin Cheng, and Jiawan Zhang
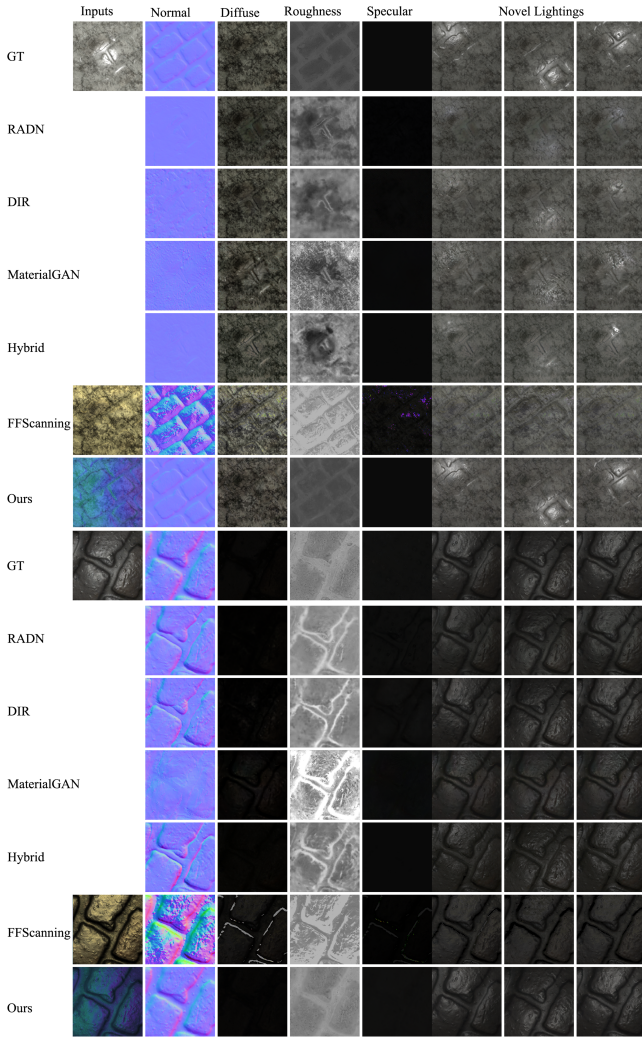

Figure 6: Comparison against state-of-the-art methods on two synthetic scenes. We show the reconstructed results and rerendered images under three novel point light sources. For the comparison configuration of FFScanning and more results on synthetic datasets, please refer to our supplementary materials. Note that there is no capture perturbation in the input images. Our method is robust to perturbations in real acquisition while FFScanning needs careful calibration to eliminate perturbations.

Figure 7: Comparison against previous methods on three real scenes. By leveraging a paper frame, we align the inputs of our and other methods by cropping the image. Here we compare the RADN, DIR, MaterialGAN, Hybrid. DIR and MaterialGAN only accept one image as the input for a fair comparison, and the default camera parameters are used in optimization.
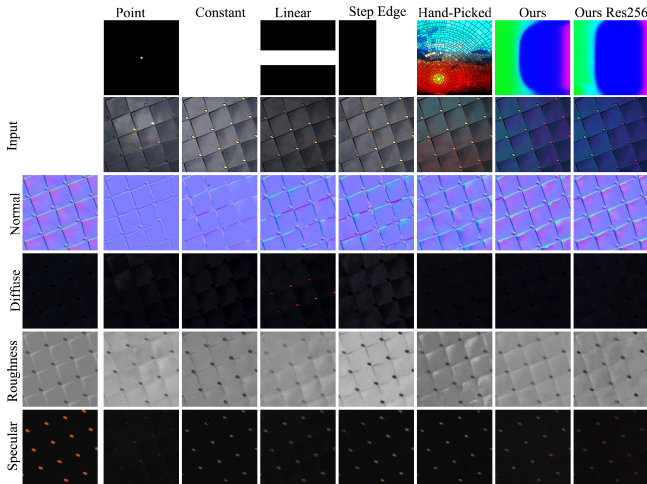
Figure 8: The impact of lighting pattern over SVBRDF estimation. The first column is ground truth SVBRDF. We visualize the lighting patterns in the first row and the SVBRDF estimated by the corresponding network trained using our online rendering process.
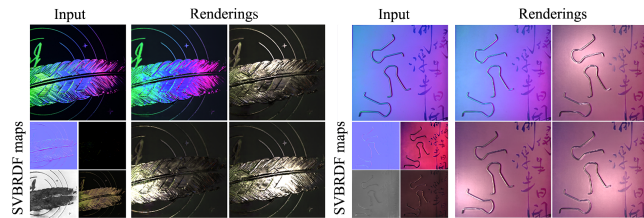

Figure 9: Here we show two more results reconstructed by the same trained network from a single real image. The input images are cropped automatically according to the derived FOV range without calibration. The left input image is captured under a cell phone (Huawei P40Pro) as illumination. Even so, our method can reconstruct high-quality material maps.
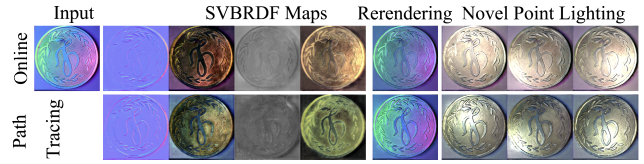

Figure 10: We compare the results of the network that was jointly trained with the online rendering process and the network trained by pre-rendered dataset using ray tracing and the learned lighting pattern. The results are similar with little color variance in diffuse and specular albedo, demonstrating the accuracy of our proposed online rendering process for lighting pattern optimization.
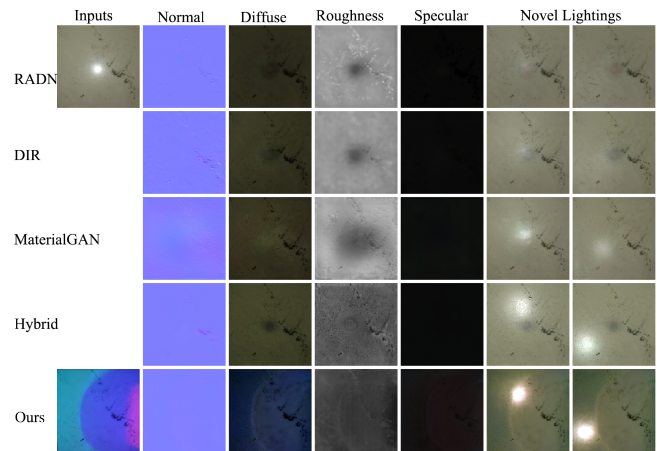

Figure 11: We compare our method against other methods on a mirror-like material. Our method is not able to properly reconstruct the material. However, the specularity reflectance of our results is still better than others.