

NFPLight: Deep SVBRDF Estimation via the Combination of Near and Far Field Point Lighting

LI WANG, LIANGHAO ZHANG, FANGZHOU GAO, YUZHEN KANG, and JIAWAN ZHANG*, Tianjin University, China

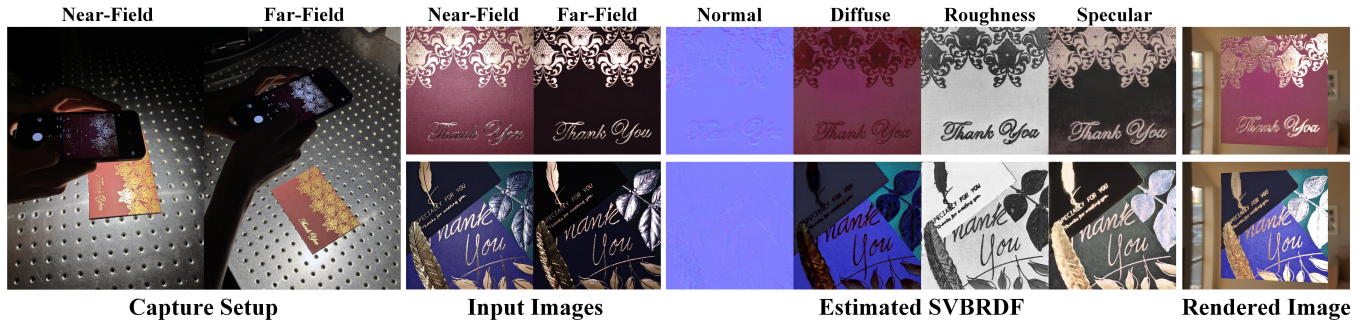


Fig. 1. We proposed a method for estimating SVBRDF that utilizes a novel capture setup. This setup separately acquires two material appearance images under near-field and far-field conditions. In comparison to solely using near-field point lighting, the far-field image provides more comprehensive information on specular reflectance. The integration of near-field and far-field images effectively resolves the ambiguity of SVBRDF estimation from a limited number of images. Here, we present two results for real scenes. Besides, both quantitative evaluations and qualitative visualizations demonstrate that our method achieves a superior quality compared to the state-of-the-art methods.

Recovering spatial-varying bi-directional reflectance distribution function (SVBRDF) from a few hand-held captured images has been a challenging task in computer graphics. Benefiting from the learned priors from data, single-image methods can obtain plausible SVBRDF estimation results. However, the extremely limited appearance information in a single image does not suffice for high-quality SVBRDF reconstruction. Although increasing the number of inputs can improve the reconstruction quality, it also affects the efficiency of real data capture and adds significant computational burdens. Therefore, the key challenge is to minimize the required number of inputs, while keeping high-quality results. To address this, we propose maximizing the effective information in each input through a novel co-located capture strategy that combines near-field and far-field point lighting. To further enhance effectiveness, we theoretically investigate the inherent relation between two images. The extracted relation is strongly correlated with the slope of specular reflectance, substantially enhancing the precision of roughness map estimation. Additionally, we designed the registration and denoising modules to meet the practical requirements of hand-held capture. Quantitative assessments and qualitative analysis have demonstrated that our method achieves superior SVBRDF estimations compared to previous approaches. All source codes will be publicly released.

*Corresponding authors.

Authors' address: Li Wang, li_wang@tju.edu.cn; Lianghao Zhang, lianghaozhang@tju.edu.cn; Fangzhou Gao, gaofangzhou@tju.edu.cn; Yuzhen Kang, yu_zhen@tju.edu.cn; Jiawan Zhang, jwzhang@tju.edu.cn, Tianjin University, China.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM 0730-0301/2024/12-ART1 <https://doi.org/10.1145/3687978>

CCS Concepts: • **Computing methodologies** → **Reflectance modeling**.

Additional Key Words and Phrases: Material Reflectance Modeling, SVBRDF, Deep Learning, Rendering

ACM Reference Format:

Li Wang, Lianghao Zhang, Fangzhou Gao, Yuzhen Kang, and Jiawan Zhang. 2024. NFPLight: Deep SVBRDF Estimation via the Combination of Near and Far Field Point Lighting. *ACM Trans. Graph.* 43, 6, Article 1 (December 2024), 11 pages. <https://doi.org/10.1145/3687978>

1 INTRODUCTION

Reconstructing the high-quality spatial-varying bi-directional reflectance distribution function (SVBRDF) under the as convenient as possible acquisition condition has been a long-standing problem in computer graphics. One of the most accessible methods is to capture a few appearance images by the mobile phone with a flash light. However, the estimated result is inherently ambiguous because different combinations of SVBRDF parameters can yield the same appearance.

Given the learned prior from data, the plausible SVBRDF can be produced from single-image input [Deschaintre et al. 2018; Guo et al. 2021; Zhou and Kalantari 2022; Sartor and Peers 2023; Wang et al. 2023]. However, the limited information from single-image input makes it challenging to recover high-quality results. The addition of more input images can effectively improve the quality of reconstruction [Deschaintre et al. 2019; Gao et al. 2019; Guo et al. 2020; Zhu et al. 2023]. However, increasing the number of inputs puts significant pressure on the real data capture process. Additionally, the added images contain a large amount of redundant information, which increases the computational burden but provides very limited help in improving the quality of reconstruction. Therefore, the key

challenge is to maximize the effective information in each input image, thus minimizing the required number of inputs for high-quality SVBRDF reconstruction.

In this paper, we present a novel co-located capture strategy to obtain two highly complementary appearance images, thereby improving the efficiency of input information. The key observation is that a far-field point light can be approximately considered as distant lighting. As shown in Fig. 2, since the surface normal of planar materials is almost upward, under the far-field central lighting condition, the captured appearance image is dominated by specular reflectance. Conversely, under the near-field central lighting, the captured image reflects a combination of both specular and diffuse reflectance characteristics. These two appearance images exhibit a significant difference and also encompass the essential information for SVBRDF estimation. Therefore, their combined inputs can efficiently reduce the ambiguity between different reflectance parameters. To extract more effective information from the input images, we explore the potential information present in their relationship. Through theoretical derivation, we find that the subtraction of the two images can reflect the slope of specular reflectance, which is highly related to the roughness map. We extract this relation and design a two-branch network to facilitate better SVBRDF estimation.

Technically, to achieve the hand-held capture of input images, we have the following designs. Firstly, under the conditions of unknown camera intrinsic and extrinsic parameters, we introduce an automatic correspondence detection method to assist in registering two input images to the same concerned material region. Secondly, under far-field conditions, the uncontrollable positional deviations caused by hand-held capture become more noticeable. Inspired by the latest research of relighting [Bieron et al. 2023], we design a specific denoising method where we treat the removal of these deviations as a relighting procedure. It can achieve denoising at the input stage, thereby reducing the impact of noise on the material recovery process. We evaluated our methods on both synthetic data and real-world data, and the results demonstrated that our method can produce more precise SVBRDF than the existing methods.

In summary, we have the following contributions:

- We present a novel two-image co-located capture strategy that combines near-field and far-field point lighting.
- We found the subtraction of these two images reflects the slope of specular reflectance, which closely corresponds to the roughness map. Subsequently, we designed a two-branch network to leverage this relation for better SVBRDF estimation.
- We introduced the auto registration and denoising techniques, ensuring the available two-image capture under hand-held conditions.

2 RELATED WORK

We discuss near-planar SVBRDF estimation methods under sparse input and categorize them based on the lighting condition.

2.1 Point Lighting

The rapid development of deep learning techniques enables SVBRDF reconstruction from sparse point-lighting images. Deschaintre et al.

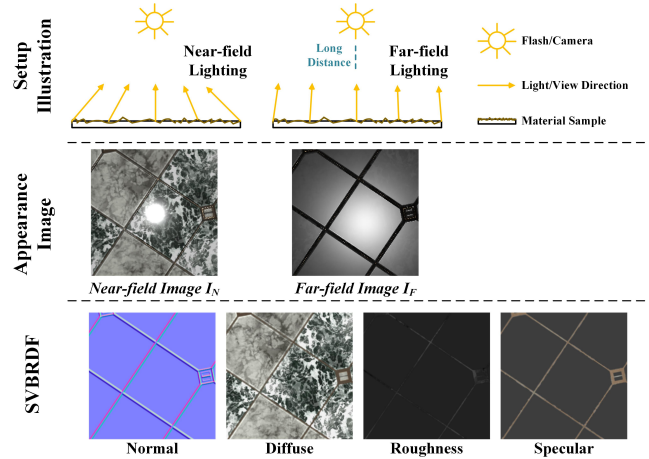


Fig. 2. We present two material appearance images under the central co-located point lighting at different capture distances. Under near-field condition, lighting directions on the material sample surface change rapidly. Therefore, the appearance image combines both diffuse and specular reflectance. Under the far-field, lighting directions are approximately parallel to the surface. The appearance image is dominated by specular reflectance.

[2018] and Li et al. [2018] respectively proposed SVBRDF dataset and built a deep network to estimate SVBRDF from a single image. By exploring novel network architectures and training strategies, subsequent research [Guo et al. 2021; Zhou and Kalantari 2021; Vecchio et al. 2021] further improved the quality of SVBRDF estimation. Zhou et al. [2022] and Fischer et al. [2022] introduced the meta-learning to incorporate test-time optimization into the training process, thereby avoiding overfitting issues. Wang et al. [2023] integrated two-level basis materials into deep learning, reducing the ambiguity in single-image estimation. Zhou et al. [2023] proposed a relightable material generator that can be trained using only a real flash-photograph dataset. Guo et al. [2023] adopted a divide-and-conquer solution to address ultra-high resolution SVBRDF capture. However, the significant lack of reflectance information in single-image input hinders the reconstruction of high-quality SVBRDF.

To facilitate SVBRDF estimation from multiple input images, Deschaintre et al. [2019] introduced an order-independent fusing layer to extend their single-image work. Moreover, they presented a fine-tuning method for large-scale SVBRDF capture [Deschaintre et al. 2020]. Zhu et al. [2023] designed a two-branch network to learn the lighting effects in images, thereby eliminating the need for precise lighting position calibration across multiple images. Additionally, Gao et al. [2019] employed an auto encoder to embed the SVBRDF into latent space, thus performing latent-space optimization. Guo et al. [2020] further constrained the latent space using StyleGAN2 network [Karras et al. 2020]. These methods all capture input images under near-field lighting. In contrast, our method takes the combination of near-field and far-field lighting as input. It offers a higher information density, thus achieving high-quality material recovery with a lower number of input images.

2.2 Extended Lighting

Ren et al. [2011] employed a linear light source to scan planar materials and reconstruct their reflectance properties. Chen et al. [2014] extended this method to the anisotropic material. Additionally, Aitala et al. [2013] utilized an LCD panel as a programmable light source to capture isotropic reflectance via frequency domain analysis. Riviere et al. [2016] further explored the use of planar lighting in the mobile device contexts. Beyond manually-designed planar lighting patterns, Kang et al. [2018; 2019] utilized an asymmetric deep autoencoder to simultaneously train both the SVBRDF recovery network and the lighting patterns tailored to the physical capture device. Ma et al. [2021] further advanced the method to accommodate free-form scanning contexts. Zhang et al. [2023] explored the application of learned planar lighting patterns for estimating SVBRDF from a single image. Additionally, Rodriguez-Pardo et al. [2023] leveraged a flatbed scanner to produce the type of diffuse illumination, which is used for the direct capture of diffuse map. Although extended lighting offers a broader sampling of material appearance compared to point lighting, the integrated appearance with lighting pattern is challenging to be decoupled into several distinct lighting space samples when only a limited number of images are available. Moreover, during the material acquisition process, these methods usually require more complex physical devices to carry the lighting pattern than the point-lighting methods.

2.3 Unknown environment Lighting

Li et al. [2017] proposed a self-augmented strategy to leverage a limited number of labeled datasets to train a single-image material inference network under unknown environment lighting conditions. Ye et al. [2018] further extended this work by using only unlabeled datasets for training. Martin et al. [2022] proposed a hybrid method to capture material in the wild. Although unknown environment lighting conditions are more common in real-world capture, their instability usually leads to insufficient expression of material appearance. Recently, large-scale diffusion model has been applied to many image generation fields. Sartor et al. [2023] introduced a generative diffusion model into SVBRDF estimation. Vecchio et al. [2023] further generated tileable materials, which can be used to capture material under the unknown environment lighting conditions. These method can provide a high-quality material estimation, but it does not guarantee a pixel-perfect match. Additionally, the training and inference cost of large-scale model is extremely huge.

3 METHOD

3.1 Problem Statement.

Our goal is to estimate spatial varying material reflectance properties from two color images. We assume that the material sample is a nearly planar surface with geometric details that the normal map can model. Moreover, the Cook-Torrance BRDF model [Cook and Torrance 1981] with GGX microfacet distribution [Walter et al. 2007] is chosen to represent the reflectance properties. Therefore, SVBRDF can be represented by four material maps: normal map n , diffuse map d , roughness map r , and specular map s . Furthermore, two images are captured by a mobile phone camera with a co-located flash light. The capturing positions are the near-field top

view and the far-field top view, respectively. Note that we only require a rough capture position and do not calibrate the intrinsic and extrinsic parameters of the camera. Our method aims at learning a mapping function F to recover material maps $M = \{n, d, r, s\}$ from two distinct inputs: the image captured under near-field lighting, denoted as *Near-field Image* I_N , and the image captured under far-field lighting, denoted as *Far-field Image* I_F , as follows:

$$M = F(I_N, I_F). \quad (1)$$

I_N and I_F are both captured at the central position of material sample and differ only in the terms of capture distance. It is evident that there exists some relations reflecting the material reflectance feature between their appearance variations. However, the challenges lie in determining what this relation is, and how to use this relation to serve for material reflectance estimation. Additionally, under the hand-held capture, collecting available images I_N and I_F is challenging. Firstly, since these two images are captured under two shots with different distance, their contents are not aligned to a certain material area. Secondly, the unstable capture position under hand-held conditions hinders exploring the fixed relation between the two images. Thirdly, due to the decay of lighting intensity with distance, *Far-field Image* typically exhibits low brightness, leading to a lower signal-to-noise ratio (LSNR).

3.2 Algorithm

To solve the above challenges, we propose NFPLight, as shown in Fig. 3. The core component is SVBRDF prediction part, including a relation extraction module and a prediction network. The former module effectively derives the relation between *Near-field Image* and *Far-field Image* via a sequence of explicit computations. The latter applies these relational features in the network to serve for material recovery. Furthermore, NFPLight includes a real data preprocessing component designed to correct real-world data captured in hand-held conditions. It ensures the real data are transformed into reliable inputs for the SVBRDF prediction component. In the following sections, we discuss the details.

3.2.1 Relation Extraction Module. According to the rendering equation, two input images captured under a single point light source can be represented as follows:

$$\begin{aligned} I_N &= L_N(D_N + S_N)\cos\theta_N, \\ I_F &= L_F(D_F + S_F)\cos\theta_F. \end{aligned} \quad (2)$$

where L_N and L_F represent the lighting intensity, D_N and D_F represent the diffuse reflectance component, S_N and S_F represent the specular reflectance component, while $\cos\theta_N$ and $\cos\theta_F$ indicate the relation between lighting direction and normal direction. Variables with the subscript N are associated with the near-field context, and those with subscript F correspond to the far-field. Note that Eq.2 is a typical rendering equation. As mentioned in Sec. 3.1, we use the Cook-Torrance to express it, where $D_{N/F} = \frac{d}{\pi}$, $S_{N/F} = \frac{DGF}{4(\omega_i \cdot n)(\omega_o \cdot n)}$. To explore the underlying relationship, we attempt subtraction, as follows:

$$I_N - I_F = L_N(D_N + S_N)\cos\theta_N - L_F(D_F + S_F)\cos\theta_F. \quad (3)$$

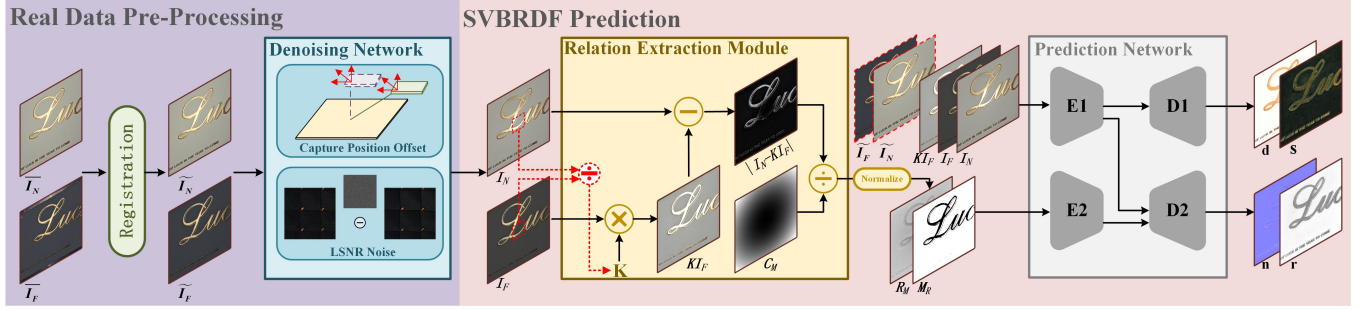


Fig. 3. The overview of our method NFPLight. It has two parts. The first one is the real data pre-processing part. It solves the problems under hand-held real capture, such as registration and denoising. The second part is the primary SVBRDF prediction part. In this part, the relation map is initially extracted by some explicit calculations, and then it combines some other images as the input of a two-branch prediction network, thus estimating the SVBRDF. Here, I_N, I_F represent the original captured images. \tilde{I}_N, \tilde{I}_F represent two images after registration. To simplify the notation, we denote the terms in Eq. 5 as $K = K_L K_\theta$; $C_M = \cos\theta_N |\cos\theta_N - \cos\theta_F|$.

Firstly, L_N and L_F are only affected by lighting distance. Given that the distance variations at different surface points have negligible impact, L_N and L_F can be treated as two homogeneous values. Considering the central lighting setting, the lighting directions at the center of the two images are both perpendicular to the surface, where the parameters exist $D_N = D_F, S_N = S_F$ and $\cos\theta_N = \cos\theta_F$. Therefore, the intensity scale K_L can be computed by the ratio of central values:

$$K_L = \frac{L_N}{L_F} = \frac{I_N[H/2, W/2]}{I_F[H/2, W/2]}. \quad (4)$$

where, H and W are respectively the height and width of I_N and I_F .

Secondly, once the capture configuration is determined, $\cos\theta_N$ and $\cos\theta_F$ become fixed. When assuming the normal ideally perpendicular to the surface, it allows us to pre-compute their ratio $K_\theta = \cos\theta_N / \cos\theta_F$. Therefore, we derive an intrinsic relationship as follows, which we define as the relation map R_M .

$$R_M = \frac{|I_N - K_L K_\theta I_F|}{\cos\theta_N |\cos\theta_N - \cos\theta_F|} = L_N \left| \frac{S_N - S_F}{\cos\theta_N - \cos\theta_F} \right|. \quad (5)$$

By substituting the above defined K_L, K_θ, I_N, I_F into Eq. 5 and simplifying, we obtain the form on the right side. It indicates the rate of change in specular reflectance along with the variation of incident angles, essentially reflecting the slope of microfacet distributions. In the GGX model, the roughness parameter controls the shape of microfacet distributions. Therefore, there exists a close correlation between the roughness map and the relation map. Besides, since the analysis is conducted in the slope domain, the color information of I_N and I_F in Eq. 2 becomes irrelevant, prompting us to use their mean value for the above calculations. Moreover, despite the relation map being coupled with the influence of surface normal due to the unknown normal map, it effectively removes the impact of albedo, significantly reducing the ambiguity in estimating SVBRDF. Two representative results of relation map extraction are shown in Fig. 4. The first example is an ideal planar sample, adhering to our assumption of surface normal. The structure and local intensity variations of the extracted relation map show a high similarity to the roughness map. The second one is a nearly planar sample that exists variations in normal. Its local intensity variations are still

correlated with the roughness map, but the overall structure aligns with the normal map. Additionally, it is evident from both examples that the effect of diffuse reflection is effectively removed.

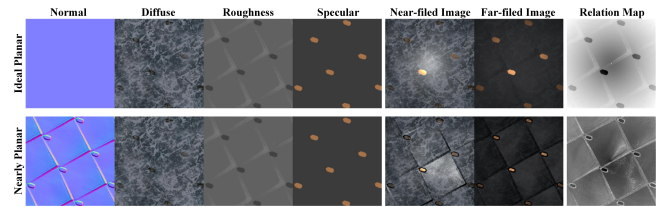


Fig. 4. Relation Map Extraction. We present two results of relation map extraction. The first example is an ideal planar sample, adhering to our assumption of surface normal. The second, a nearly planar sample, does not meet the assumption.

3.2.2 Prediction Network. To better utilize the relation map to facilitate the decoupling of SVBRDF parameters, we design a two-branch neural network to recover SVBRDF, as shown in Fig. 3. It has two sets of encoders and decoders. The first encoder processes inputs I_N, I_F , and $K I_F$, with its encoded features enabling the first decoder to estimate d and s . Due to the stable relationship in lighting intensity between I_N and $K I_F$, the addition of $K I_F$ ensures that the reflectance feature extraction is not affected by intensity variations. The second encoder processes R_M and M_R , and the second decoder takes the combination of both encoder's features as input to predict n and r . Considering the challenge of low dynamic range (LDR) capture, the relation extraction in over-exposure regions is unreliable. Therefore, a mask M_R is generated to identify the regions where intensity exceeds a threshold close to one. Although this is not a precise split, the identified regions must include the over-exposure regions. Therefore, it can assist the network in understanding the confidence of the relation map. Additionally, during the real-data processing under the unknown lighting position, the non-denoised images \tilde{I}_N, \tilde{I}_F are incorporated into the input of the first encoder to compensate for the lost information due to the prediction error of denoising network.

3.2.3 Registration Module. Given that the material sample is a planar surface, the registration of these two images can be performed by a homography transformation. We offer two methods to find correspondences between two images. For materials with rich textures, we employ LoFTR [Sun et al. 2021] for automatic detection of correspondences. In practice, we choose the whole *Near-field Image* as the concerned material area and transfer *Far-field Image* into the same area by homography. If the automatic detection fails, we alternatively place a square bounding box on the material, thereby finding correspondences by identifying corner points. Refer to the supplementary materials for more details.

3.2.4 Denoising Module. Given the irregular characteristics of capture position errors and LSNR noise, we employ a deep neural network to predict denoised images. To train this network, we simulate both types of noise. For the positional offset from the center caused by hand-held capture, we use three separate Gaussian distributions to model errors along the x-axis, y-axis, and z-axis. In practice, we empirically set the standard deviation of the x-axis and y-axis to 0.2, and a slightly larger deviation of 1.0 for the z-axis. For LSNR noise, we begin with a log-normal distribution to simulate real-world variations in noise intensity. The values sampled from this distribution serve as the standard deviations in Gaussian distributions that model the random LSNR noise. In our implementation, the mean values of the log-normal distribution are set at 0.001 for the *Near-field Image* and 0.005 for the *Far-field Image*, both with a uniform standard deviation of 0.03.

4 IMPLEMENTATION

In this section, we discuss the necessary implementation details of our method, including the training details and capture details. The source code and pre-trained model will be publicly released.

4.1 Training Details

Our method includes two networks: a denoising network and a prediction network. They both employ advanced image-to-image translation network NAFNet [Chen et al. 2022] as the core architecture. There are 4-layer encoder-decoder pairs for each U-net architecture, with skip connections between all layers. Each layer is composed of a stack of NAFBlocks with a base feature width of 32. The denoising network processes 12-channel inputs, including a 6-channel original *Near-field Image* \tilde{I}_N and *Far-field Image* \tilde{I}_F , and a 6-channel logarithmic image flattening the dynamic range $[0, 1]$ of them. The output is a 6-channel denoised version of *Near-field Image* I_N and *Far-field Image* I_F . The inputs of prediction network are detailed in in Sec.3.2.2, where a logarithmic transformation is applied to $KI_F, I_F, I_N, \tilde{I}_F, \tilde{I}_N$ to flatten dynamic range. All inputs, except for the 1-channel M_R and R_M , consist of three channels each. The outputs of Encoder 1 (E1) and Encoder 2 (E2) are concatenated and fed into Decoder 2 (D2), where the skip connections within D2 are performed in the same concatenated manner. The final output of prediction network is 10-channel SVBRDF map, including 3-channel normal, 3-channel diffuse, 1-channel roughness, and 3-channel specular.

We implemented NFPLight in PyTorch [Paszke et al. 2019] and used Adam optimizer [Kingma and Ba 2014] for two-stage training.

In the first stage, the denoising network and prediction network were individually trained for 400K iterations with learning rates initially set at $5e-4$ and gradually reduced to $1e-5$ following a cosine annealing schedule. Each network was supervised using L1-norm: the denoising network with ground-truth (GT) denoised images, and the prediction network with GT SVBRDF maps. During training of this stage, \tilde{I}_N and \tilde{I}_F are replaced with placeholders I_N and I_F . In the second stage, we fine-tune both networks for 100K iterations by joint training, which is supervised by L1-norm and rendering loss, similar to [Guo et al. 2021]. It has the same decay-schedule learning rate ranging from $1e-4$ to $1e-5$. The training data is sourced from a public SVBRDF dataset by Deschaintre et al. [2018]. We follow the default train-test split. All input images has a resolution of 256×256 during training. The total training time takes about 2 days on a single NVIDIA RTX 4090 graphics card.

4.2 Capture Details

The primary challenge in capturing real data is to find the required z-axis shooting distance. For ease of use, the distance in our experimental setup is defined relative to the material surface’s size. Based on the findings in Sec.5.2.3, the capture distance for the *Near-field Image* is determined by the maximum field of view (FOV) of the capture device. Therefore, under this certain FOV, all captured images meet the required distance for the *Near-field Image*. Similarly, since the distance for the *Far-field Image* is a fixed multiple of the near-field distance, when we adjust our zoom to this specific multiple and capture an area approximately the same as that in the *Near-field Image*, the shooting distance at that time will exactly meet the requirements for *Far-field Image*.

5 EXPERIMENTS

To evaluate the quality of the SVBRDF recovered by our method, we conducted a comparative experiment against the state-of-the-art (SOTA). In addition, we also conducted ablation studies to analyze the effects of various components in our method.

5.1 Comparison Experiments

We compared our method against SOTA multiple-image SVBRDF estimation methods, including DIR [Gao et al. 2019], FSC [Deschaintre et al. 2019], and MGan [Guo et al. 2020]. Due to Zhu et al. [2023] focusing on estimating SVBRDF without calibration and their source code not being publicly available, we didn’t compare against theirs. Instead, we compared our method against Gao et al. [2019] and Guo et al. [2020], which are precisely calibrated and can serve as substitutes for the comparison against Zhu et al. [Zhu et al. 2023]. Additionally, our comparative evaluation includes the latest single-image SVBRDF estimation method [Wang et al. 2023]. These results are obtained through the source code provided by their authors. The optimization-based methods, DIR and MGan, utilize estimation results of FSC as their initialization. For a more comprehensive comparison, we also obtained the synthetic results of the planar lighting method LPL [Zhang et al. 2023] by contacting the authors.

5.1.1 Comparison on Synthetic Data. Firstly, we conducted a numerical analysis on a collection of 122 synthetic scenes sourced from Deschaintre et al. [2018; 2019]. Note that these test scenes

Table 1. Numerical comparison on 122 synthetic scenes. We evaluate the quality of estimated SVBRDF in terms of RMSE. The re-renderings (Ren.) for each SVBRDF are performed on 30 random lighting directions and evaluated by both RMSE and LPIPS. The lowest errors are highlighted in bold. The upper part is a fair comparison against SOTA methods, the middle part is an enhanced comparison by providing our near-far-combination inputs for multi-image methods, and the bottom part is a challenging comparison against multi-image methods with 20 inputs and planar lighting method.

Methods	RMSE↓					LPIPS↓	
	N	D	R	S	Ren.	Ren.	
FSC	0.0578	0.0855	0.1636	0.0631	0.0775	0.2521	
DIR	0.0544	0.0247	0.1490	0.0593	0.0729	0.1608	
MGAN	0.0558	0.0247	0.1408	0.0572	0.0697	0.1731	
DeepBasis	0.0537	0.0306	0.1574	0.0561	0.0550	0.1528	
Ours	0.0195	0.0135	0.0316	0.0284	0.0310	0.0631	
FSC+	0.0726	0.1176	0.1858	0.0559	0.0857	0.3147	
DIR+	0.0700	0.0594	0.1578	0.0639	0.0591	0.1379	
MGAN+	0.0699	0.0354	0.1472	0.0433	0.0574	0.1360	
FSC-20	0.0407	0.0836	0.1429	0.0606	0.0738	0.2267	
DIR-20	0.0225	0.0152	0.1219	0.0594	0.0475	0.0642	
MGAN-20	0.0301	0.0224	0.1007	0.0514	0.0472	0.0871	
LPL	0.0359	0.0201	0.0486	0.0262	0.0452	0.1132	

were never utilized during training. During the rendering of input images, near-far-field relative capture distances were set to 2.414 and 10, respectively, corresponding to the common FOV of approximately 60° . The rationale behind selecting these capture distances is detailed in Sec.5.2.3. The quality of reflectance parameters was evaluated using Root Mean Square Error (RMSE), and the re-rendered images were assessed using both RMSE and Learned Perceptual Image Patch Similarity (LPIPS) [Zhang et al. 2018]. The re-renderings were performed under 30 random lighting and viewing directions. The numerical evaluation results are shown in Table 1. The upper part of the table provides a fair comparison with the SOTA point-lighting methods. The metrics demonstrate that our method achieves superior quality in estimated SVBRDF and in the quality of re-rendered images. Furthermore, in the middle part of Table 1, we conduct an enhanced comparison by providing our novel combination of near-field and far-field images as the inputs for SOTA multi-image method. Due to the training of DIR[Gao et al. 2019] and MGAN[Guo et al. 2020] being performed on the material domain, their generated latent space can be directly applied to near-far inputs. The enhanced comparison demonstrates that simply combining our near-far capture setting with latent-space method can't effectively improve reconstruction quality. The utilization of explicit or implicit correlation between near-far-field images from our method is crucial. Additionally, the bottom part shows a challenging comparison against previous multi-image methods with 20 input images and the learned planar lighting method [Zhang et al. 2023]. Even with the use of more input images or complex extended lighting in their methods, ours still yields comparable SVBRDF estimations and enhanced quality in re-rendered images.

Furthermore, we provided a visual comparison in Fig. 10. Due that near-field point lighting only induce specular reflectance in limited areas of material, it is insufficient to present the overall specular features when appearance measurements are sparse. This challenge is amplified in materials with strong diffuse reflectance, where specular reflection information in appearance images is significantly reduced. As a result, these methods like FSC, DIR, MGAN,

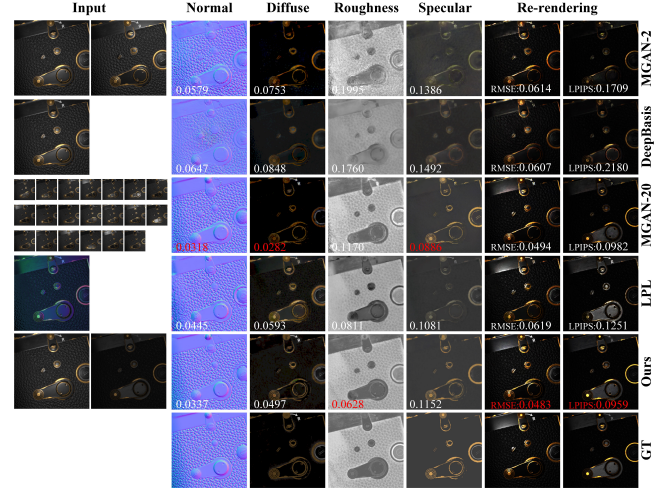


Fig. 5. The visual and numerical evaluation on Sartor et al. [2023].

and DeepBasis struggle to accurately reconstruct normal, roughness, and specular maps. Even with the use of our novel combination of near-field and far-field inputs, these SOTA multi-image methods like FSC+, DIR+ and MGAN+ still cannot resolve the ambiguity in parameter recovery, as shown in Fig. 11. Additionally, in Fig.12, facing a more complex material, even when the number of input images is increased to 20, methods such as FSC-20, DIR-20, MGAN-20 are still unable to effectively recover the details of the roughness and specular map. In LPL result, this problem is alleviated because the specular reflectance responses are largely activated by planar lighting. However, decoupling albedo maps is challenging given that material and complex lighting patterns are highly entangled in captured pixels. In contrast, far-field lighting can effectively activate the overall specular reflectance of the material while avoiding the demultiplexing of lighting patterns. Consequently, our approach provides more accurate SVBRDF estimations and yields higher quality in re-rendered images.

Finally, we also conducted experiments on a diverse range of data sources [Sartor and Peers 2023] and [Vecchio and Deschaintre 2024], which contain 30 and 65 different test scenes, respectively. The visual and numerical results shown in Fig. 5 and Fig. 6 lead to the same conclusion: our method outperforms previous methods (MGAN-2, DeepBasis) in a fair comparison and demonstrate comparable performance in a challenging comparison (MGAN-20, LPL).

5.1.2 Comparison on Real Data. To perform evaluation, we gathered 40 calibrated real scenes, similar to Guo et al. [2020]. Besides inputs, each scene contains 10 reference images to evaluate the quality of novel lighting. To obtain the calibrated camera position for comparison evaluation, additional markers surrounding the real material need to be captured. This requirement restricts the maximum FOV of the camera for near-field material capture. Consequently, a relatively lower FOV of approximately 40° is used, corresponding to near-far-field relative distances of 4 and 12, respectively. The real-world near-far-field distances for capturing these scenes are about

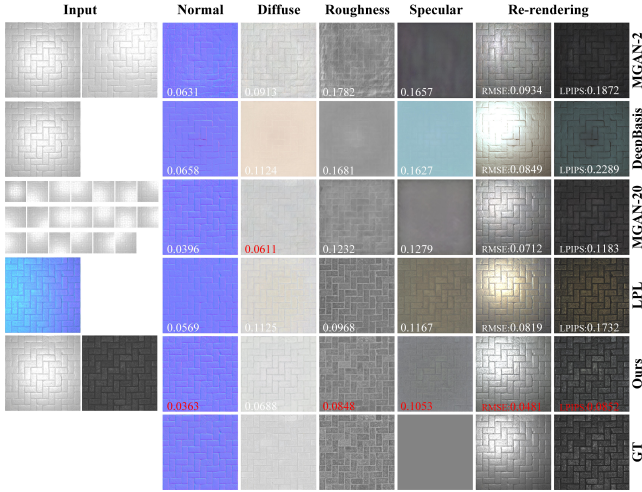


Fig. 6. The visual and numerical evaluation on Vecchio et al. [2024].

Table 2. Numerical Evaluation on 40 Real Scenes. Each scene contains 10 novel-lighting reference images, and we evaluate the re-rendering images by RMSE and LPIPS. The left part is a fair comparison against SOTA methods, the middle part is an enhanced comparison by providing our near-far-combination inputs for multi-image methods, and the right part is a challenging comparison against multi-image methods with 20 inputs.

Methods	Fair Cmp.		Enhanced Cmp.			Challenging Cmp.		
	RMSE↓	LPIPS↓	Methods	RMSE↓	LPIPS↓	Methods	RMSE↓	LPIPS↓
FSC	0.1543	0.3545	FSC+	0.1543	0.3719	FSC-20	0.1484	0.3789
DIR	0.1018	0.2728	DIR+	0.0992	0.2262	DIR-20	0.0833	0.2192
MGAN	0.0921	0.2408	MGAN+	0.0800	0.2048	MGAN-20	0.0756	0.2120
DeepBasis	0.0986	0.2585						
Ours	0.0760	0.1695	Ours	0.0760	0.1695	Ours	0.0760	0.1695

15cm and 45cm, respectively. The numerical comparison results are shown in Table 2, indicating that our estimated SVBRDF can be rendered closer to the reference. Additionally, the real calibration process aligned all other views into top-view by homography transformation. Because input images of enhanced comparison are both captured at top-view with different distance, their aligned errors are significantly lower than errors of arbitrary multi-view calibration in challenging comparison. Therefore, we noticed that from enhanced to challenging comparisons, the results of multi-image methods do not improve consistently. In Fig. 15-15, we present a total of 6 visual results for fair comparison, enhanced comparison, and challenging comparison, respectively. Benefiting from the significant activation of specular reflectance in *Far-field Image* and the extraction of the relation map, our method can robustly estimate high-quality specular and roughness maps. Therefore, compared to other methods, our approach can effectively decouple material ambiguity and produce more accurate and cleaner SVBRDF maps. More visual results of real scenes are available in supplementary material.

5.2 Ablation Studies

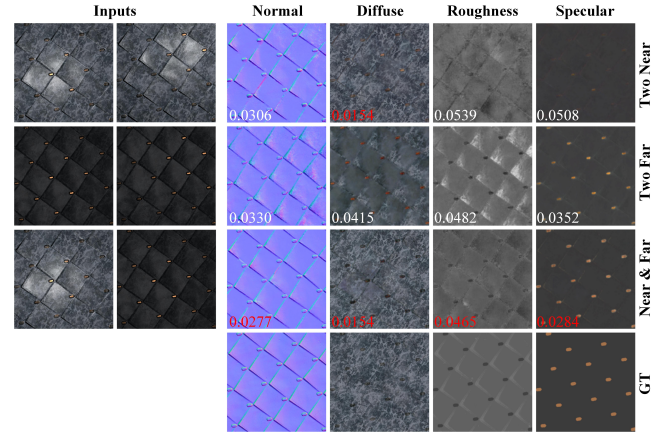


Fig. 7. The Effect of Near-field and Far-field Images. Numerical evaluations using RMSE metrics for 122 synthetic data are displayed on the corresponding parameter maps. The lowest RMSE values are highlighted in red. Two near inputs effectively show diffuse reflectance properties, helping estimate normal and diffuse maps but poorly recover roughness and specular maps due to incomplete specular activation. Two far inputs have the opposite effect. A combination of near and far inputs leverages both strengths.

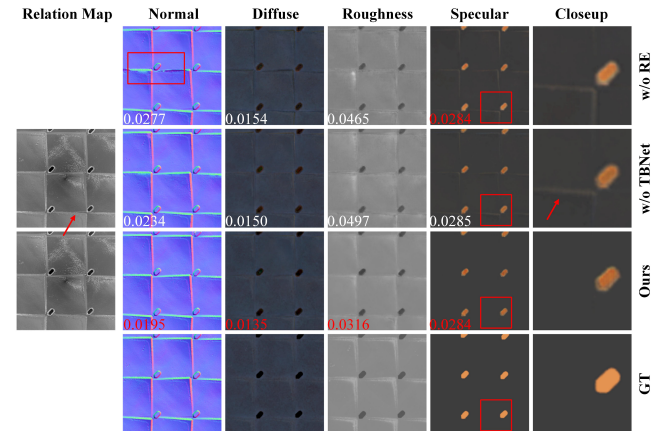


Fig. 8. The Effect of Relation Map and Two-Branch Network. Numerical evaluations using RMSE metrics for 122 synthetic data are displayed on the corresponding parameter maps. The lowest RMSE values are highlighted in red. Closeup column displays a detailed view of the red square area in the specular map. Compared to the first row, the second row has a better normal map estimation by integrating a relation map. In the third row, the utilization of two-branch network avoids the feature extracted from the relation map to misguide the estimation of diffuse and specular maps.

5.2.1 The Effect of Near-field and Far-field Images. To evaluate the effectiveness of combining near-field and far-field images as inputs, we constructed the following three sets of sub-experiments with different input modes. The first set employed two near-field images as inputs, the second utilized two far-field images, and the third combined both near-field and far-field images as inputs. For a fair comparison, each experiment used the original single-branch network, and there is no addition of relation extraction module. We

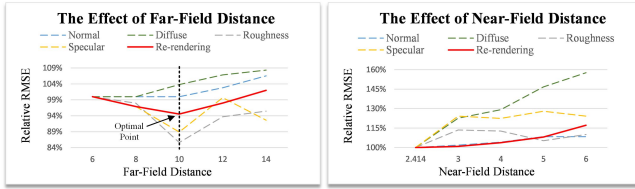


Fig. 9. The Effect of Capture Distance. All parameters are evaluated by RMSE metrics on 122 synthetic data. We employed relative RMSE to compactly display the all parameter variation. For the two charts, we take the first value in each as the standard value, and other values are divided by it to obtain a percentage. The results indicate that, within the limits permitted by the camera’s FOV, the smaller near-field distance is the better, and under a certain near-field distance, a relatively optimal far-field distance exists.

retrained the model with different input modes for each experiment, and the training strategy were maintained completely consistent. The results are shown in Fig. 7. Two near inputs, with abundant spatial variations in lighting/viewing directions, more easily show the diffuse reflectance properties, thus providing favorable conditions to estimate normal and diffuse maps. However, the lack of sufficient activation on specular reflectance leads to failed detail estimation in roughness and specular maps. Conversely, two far-field images produce the opposite effect. Thus, combining both types of images effectively utilizes complementary information, enabling a high-quality reconstruction of SVBRDF parameters.

5.2.2 The Effect of Relation Extraction and Two-Branch Network. To evaluate the effect of two-branch network, we conducted an experiment using a single branch to process all inputs (denoted as *w/o TBNet*). Furthermore, we used only *Near-field Image* and *Far-field Image* as the input for single-branch network to evaluate the effect of relation extraction (denoted as *w/o RE*). The results are shown in Fig.8. According to the comparison between *w/o TBNet* and *Ours*, two-branch network ensures the feature extracted from relation map can not be utilized by the estimation of diffuse and specular maps. This reduces ambiguity among these SVBRDF maps, as indicated by two red arrows in Fig.8. Moreover, the comparison between *w/o TBNet* and *w/o RE* demonstrates that the extracted relation map also improves the estimated quality of normal map.

5.2.3 The Effect of Capture Distance. To facilitate discussion, we define the captured region of material sample as a square. The half of side length of this square is considered as a standard unit for measurement. All distances mentioned subsequently adhere to this standard unit, and the models of different experiments are re-trained under the corresponding configuration. In experiments evaluating far-field capture distances, the near-field distance was fixed at 2.414 according to the common FOV of mobile phones, while the far-field distance ranged from 6 to 14. The results, shown in the left part of Fig.9, indicate that increasing the far-field distance improves the reconstructed quality of roughness and specular due to better specular reflectance capture. However, this reduces the spatial variations in lighting directions, negatively impacting the accuracy of normal and diffuse estimations. The errors are also coupled into the estimation of roughness and specular, collectively influencing

the rendering results. Therefore, it is evident that increasing the far-field distance does not uniformly enhance results. There exists an optimal distance, and it is 10 in the current experimental setup. Additionally, we evaluated the effect of near-field distance. In these experiments, with the far-field distance fixed at 10, the near-field distance varied from 2.414 to 6. As shown in the right part of Fig.9, increasing near-field distance worsens normal and diffuse estimations, reducing re-rendering quality. In summary, within the limits permitted by the camera’s FOV, the smaller the near-field distance is the better, and under a certain near-field distance, a relatively optimal far-field distance exists.

5.2.4 The Effect of Denoising Network. To evaluate the impact of various noise types on denoising network performance, we individually modulated the intensity of each noise type, producing input images with varying noise levels for 122 synthetic datasets. These predicted denoised images were then evaluated using the RMSE metric against the Ground Truth. The numerical results are displayed in the upper part of Fig.13. It is evident that within the maximum error range used in our training, although denoising accuracy degrades as noise intensity increases, the overall rate of change is still relatively slow. This demonstrates that within this predefined range, the denoising network maintains robust performance. Additionally, we present two visual examples of denoising results in the lower part of Fig.13. The left example shows a low-roughness material primarily affected by noise from lighting position offsets. Conversely, the right example presents a high-roughness material, which suffers significantly from LSNR noise due to its far-field image’s low brightness. A comparison with the Ground Truth (GT) illustrates that the denoising network effectively alleviates both types of noises.

6 LIMITATION AND FUTURE WORK

Although *Far-Field Image* has a lower lighting intensity compared to *Near-Field Image*, it still experiences over-exposure issues when facing mirror-like materials. The loss of information in over-exposed areas results in noticeable artifacts within the reconstructed SVBRDF. We illustrate an example of mirror-like material in Fig.14. Furthermore, we provide a comparison with other SOTA methods. Notably, the planar lighting approach used by LPL [Zhang et al. 2023] effectively mitigates over-exposure issues. Nonetheless, this method also introduces complications related to the decoupling of lighting patterns. In future work, exploring the integration of planar lighting and point lighting could potentially enable the accurate reconstruction of high-quality SVBRDF for mirror-like materials.

7 CONCLUSION

We have proposed NFFlight, a novel method to estimate SVBRDF using the combination of near-field and far-field input images. The addition of *Far-field Image* offers a comprehensive activation of specular reflectance, which complements the lost information in *Near-field Image*. Based on these two inputs, we found that their subtraction reflects the slope of specular reflectance, which is highly related to the roughness estimation. Furthermore, we designed a two-branch network to fully leverage this relation for better SVBRDF estimation. Additionally, to obtain the available two input images during hand-held capture, we introduce an auto registration module

and a denoising network. Extensive experiments conducted on both synthetic scenes and real-world captured images have demonstrated that our method outperforms SOTA techniques, yielding higher-quality SVBRDF reconstruction results.

ACKNOWLEDGMENTS

We would like to thank anonymous reviewers for their comments, Qishen Wang for his assistance in data collection, Xin Tong for his support. This work is partially supported by National Natural Science Foundation of China (62172295).

REFERENCES

- Miika Aittala, Tim Weyrich, and Jaakko Lehtinen. 2013. Practical SVBRDF capture in the frequency domain. *ACM Trans. Graph.* 32, 4, Article 110 (jul 2013), 12 pages. <https://doi.org/10.1145/2461912.2461978>
- James Bieron, Xin Tong, and Pieter Peers. 2023. Single Image Neural Material Relighting. In *ACM SIGGRAPH 2023 Conference Proceedings*. 1–11.
- 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 Trans. Graph.* 33, 4, Article 117 (jul 2014), 11 pages. <https://doi.org/10.1145/2601097.2601180>
- Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. 2022. Simple baselines for image restoration. In *European conference on computer vision*. Springer, 17–33.
- Robert L. Cook and Kenneth E. Torrance. 1981. A Reflectance Model for Computer Graphics (*SIGGRAPH '81*). Association for Computing Machinery, New York, NY, USA, 307–316. <https://doi.org/10.1145/800224.806819>
- Valentin Deschaintre, Miika Aittala, Fredo Durand, George Drettakis, and Adrien Bousseau. 2018. Single-Image SVBRDF Capture with a Rendering-Aware Deep Network. *ACM Trans. Graph.* 37, 4, Article 128 (jul 2018), 15 pages. <https://doi.org/10.1145/3197517.3201378>
- Valentin Deschaintre, Miika Aittala, Frédo Durand, George Drettakis, and Adrien Bousseau. 2019. Flexible svbrdf capture with a multi-image deep network. In *Computer graphics forum*, Vol. 38. Wiley Online Library, 1–13.
- Valentin Deschaintre, George Drettakis, and Adrien Bousseau. 2020. Guided fine-tuning for large-scale material transfer. In *Computer Graphics Forum*, Vol. 39. Wiley Online Library, 91–105.
- Michael Fischer and Tobias Ritschel. 2022. Metappearance: Meta-Learning for Visual Appearance Reproduction. *ACM Trans. Graph.* 41, 4 (2022).
- 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. 38, 4, Article 134 (jul 2019), 15 pages. <https://doi.org/10.1145/3306346.3323042>
- Jie Guo, Shuichang Lai, Chengzhi Tao, Yuelong Cai, Lei Wang, Yanwen Guo, and Ling-Qi Yan. 2021. Highlight-Aware Two-Stream Network for Single-Image SVBRDF Acquisition. *ACM Trans. Graph.* 40, 4, Article 123 (jul 2021), 14 pages. <https://doi.org/10.1145/3450626.3459854>
- Jie Guo, Shuichang Lai, Qinghao Tu, Chengzhi Tao, Changqing Zou, and Yanwen Guo. 2023. Ultra-High Resolution SVBRDF Recovery from a Single Image. *ACM Transactions on Graphics* (2023).
- Yu Guo, Cameron Smith, Miloš Hašan, Kalyan Sunkavalli, and Shuang Zhao. 2020. MaterialGAN: Reflectance Capture Using a Generative SVBRDF Model. 39, 6, Article 254 (nov 2020), 13 pages. <https://doi.org/10.1145/3414685.3417779>
- Kaizhang Kang, Zimin Chen, Jiaping Wang, Kun Zhou, and Hongzhi Wu. 2018. Efficient reflectance capture using an autoencoder. *ACM Trans. Graph.* 37, 4 (2018), 127.
- Kaizhang Kang, Cihui Xie, Chengan He, Mingqi Yi, Minyi Gu, Zimin Chen, Kun Zhou, and Hongzhi Wu. 2019. Learning efficient illumination multiplexing for joint capture of reflectance and shape. *ACM Trans. Graph.* 38, 6 (2019), 165–1.
- Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and Improving the Image Quality of StyleGAN. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- Xiao Li, Yue Dong, Pieter Peers, and Xin Tong. 2017. Modeling Surface Appearance from a Single Photograph Using Self-Augmented Convolutional Neural Networks. *ACM Trans. Graph.* 36, 4, Article 45 (jul 2017), 11 pages. <https://doi.org/10.1145/3072959.3073641>
- Zhengqin Li, Kalyan Sunkavalli, and Manmohan Chandraker. 2018. Materials for Masses: SVBRDF Acquisition with a Single Mobile Phone Image. In *Computer Vision – ECCV 2018*, Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Eds.). Springer International Publishing, Cham, 74–90.
- Xiaohe Ma, Kaizhang Kang, Ruisheng Zhu, Hongzhi Wu, and Kun Zhou. 2021. Free-form scanning of non-planar appearance with neural trace photography. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–13.
- 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 (Proc. EUROGRAPHICS 2022)* to appear, to appear (2022), to appear.
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32 (2019).
- Peiran Ren, Jiaping Wang, John Snyder, Xin Tong, and Baining Guo. 2011. Pocket Reflectometry. 30, 4, Article 45 (jul 2011), 10 pages. <https://doi.org/10.1145/2010324.1964940>
- J. Riviere, P. Peers, and A. Ghosh. 2016. Mobile Surface Reflectometry. *Computer Graphics Forum* 35, 1 (2016), 191–202. <https://doi.org/10.1111/cgf.12719>
- Carlos Rodriguez-Pardo, Henar Dominguez-Elvira, David Pascual-Hernandez, and Elena Garces. 2023. Umat: Uncertainty-aware single image high resolution material capture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5764–5774.
- Sam Sartor and Pieter Peers. 2023. MatFusion: A Generative Diffusion Model for SVBRDF Capture. In *SIGGRAPH Asia 2023 Conference Papers*. 1–10.
- Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. 2021. LoFTR: Detector-Free Local Feature Matching With Transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 8922–8931.
- Giuseppe Vecchio and Valentin Deschaintre. 2024. MatSynth: A Modern PBR Materials Dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 22109–22118.
- Giuseppe Vecchio, Rosalie Martin, Arthur Roullier, Adrien Kaiser, Romain Rouffet, Valentin Deschaintre, and Tamy Boubekeur. 2023. Controlmat: a controlled generative approach to material capture. *arXiv preprint arXiv:2309.01700* (2023).
- 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.
- Bruce Walter, Stephen R. Marschner, Hongsong Li, and Kenneth E. Torrance. 2007. Microfacet Models for Refraction through Rough Surfaces (*EGSR'07*). Eurographics Association, Goslar, DEU, 195–206.
- Li Wang, Lianghao Zhang, Fangzhou Gao, and Jiawan Zhang. 2023. DeepBasis: Hand-Held Single-Image SVBRDF Capture via Two-Level Basis Material Model. In *SIGGRAPH Asia 2023 Conference Papers*. 1–11.
- Wenjie Ye, Xiao Li, Yue Dong, Pieter Peers, and Xin Tong. 2018. Single image surface appearance modeling with self-augmented cnns and inexact supervision. In *Computer Graphics Forum*, Vol. 37. Wiley Online Library, 201–211.
- Lianghao Zhang, Fangzhou Gao, Li Wang, Minjing Yu, Jiamin Cheng, and Jiawan Zhang. 2023. Deep SVBRDF Estimation from Single Image under Learned Planar Lighting. In *ACM SIGGRAPH 2023 Conference Proceedings*. 1–11.
- Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Xilong Zhou, Miloš Hašan, Valentin Deschaintre, Paul Guerrero, Yannick Hold-Geoffroy, Kalyan Sunkavalli, and Nima Khademi Kalantari. 2023. PhotoMat: A Material Generator Learned from Single Flash Photos. In *SIGGRAPH 2023 Conference Papers*.
- Xilong Zhou and Nima Khademi Kalantari. 2021. Adversarial Single-Image SVBRDF Estimation with Hybrid Training. In *Computer Graphics Forum*, Vol. 40. Wiley Online Library, 315–325.
- Xilong Zhou and Nima Khademi Kalantari. 2022. Look-Ahead Training with Learned Reflectance Loss for Single-Image SVBRDF Estimation. 41, 6, Article 266 (nov 2022), 12 pages. <https://doi.org/10.1145/3550454.3555495>
- Pengfei Zhu, Shuichang Lai, Mufan Chen, Jie Guo, Yifan Liu, and Yanwen Guo. 2023. SVBRDF Reconstruction by Transferring Lighting Knowledge. In *Computer Graphics Forum*, Vol. 42. Wiley Online Library, e14973.

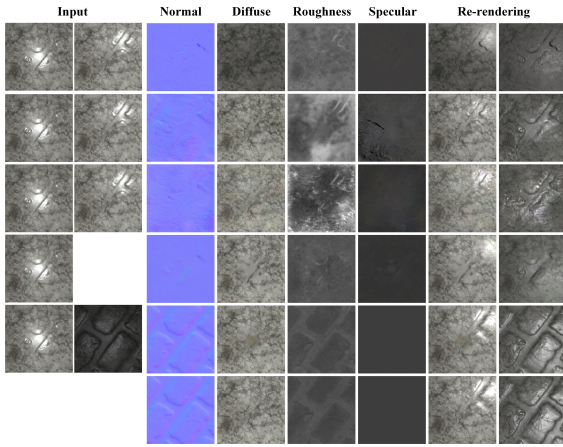


Fig. 10. Fair comparison on synthetic data against FSC of Deschaintre et al. [2019], DIR of Gao et al. [2019], MGAN of Guo et al. [2020], DeepBasis of Wang et al. 2023. The last row is the Ground Truth (GT).

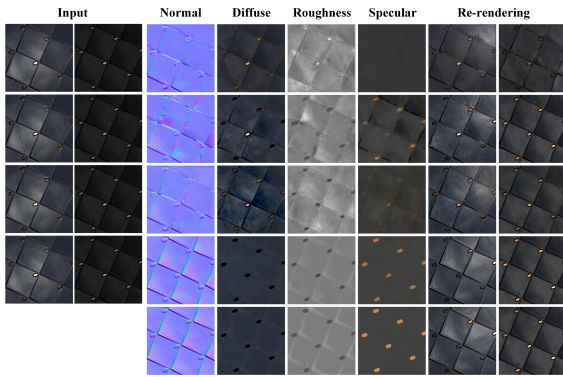


Fig. 11. Enhanced Comparison on Synthetic Data. We provide our novel combination of near-field and far-field inputs to FSC+, DIR+, MGAN+.

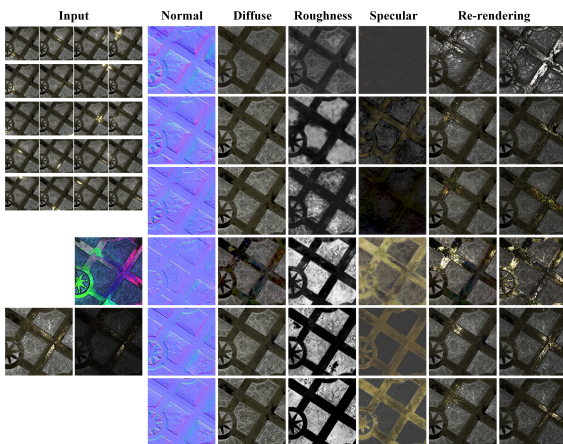


Fig. 12. Challenging comparison on synthetic data against multi-image methods with 20 inputs and planer-light method LPL of [Zhang et al. 2023].

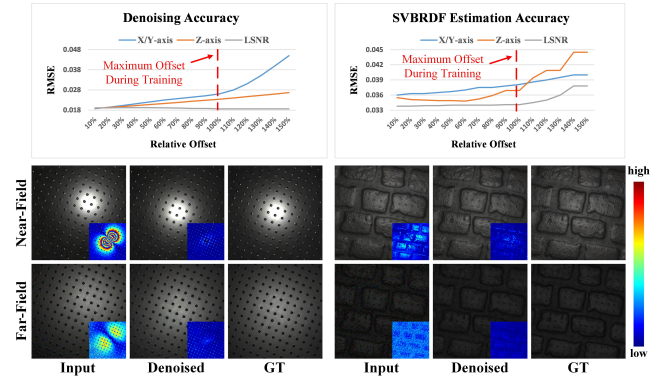


Fig. 13. The Evaluation of Denoising Accuracy. The upper part is the evaluation of denoising accuracy and SVBRDF estimation accuracy, with the variation in lighting position offset and LSNR noise intensity. Position offsets are categorized into two groups, X/Y-axis and Z-axis, to account for their differing standard deviations. To unify the scale of variations in different noise intensities, we adopt relative noise offsets as a measure of noise variation. For each type of noise, we use the maximum allowable offset in our experiments as the baseline (100 percent). The results show that within our allowed maximum error range, the increase in noise has a relatively small impact on denoising accuracy. The bottom part shows two representative materials along with their input images, denoised outputs, and the corresponding GT images. Additionally, error maps comparing these results with the GT are provided to enhance visualization. The results demonstrate that the denoising network mitigates issues caused by lighting position offsets and LSNR noise.

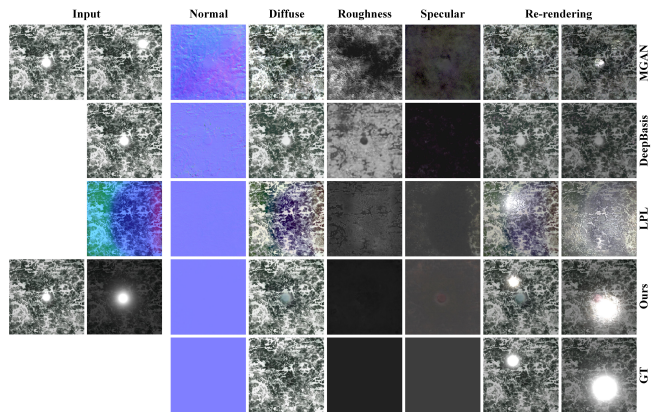


Fig. 14. Failure Case. This sample is a mirror-like material that has an extremely low roughness value. This leads to the over-exposure issue in both *Near-Field Image* and *Far-Field Image*, resulting in artifacts in the reconstructed SVBRDF. We also provide the results estimated by MGAN[Guo et al. 2020], DeepBasis[Gao et al. 2019], and LPL[Zhang et al. 2023].

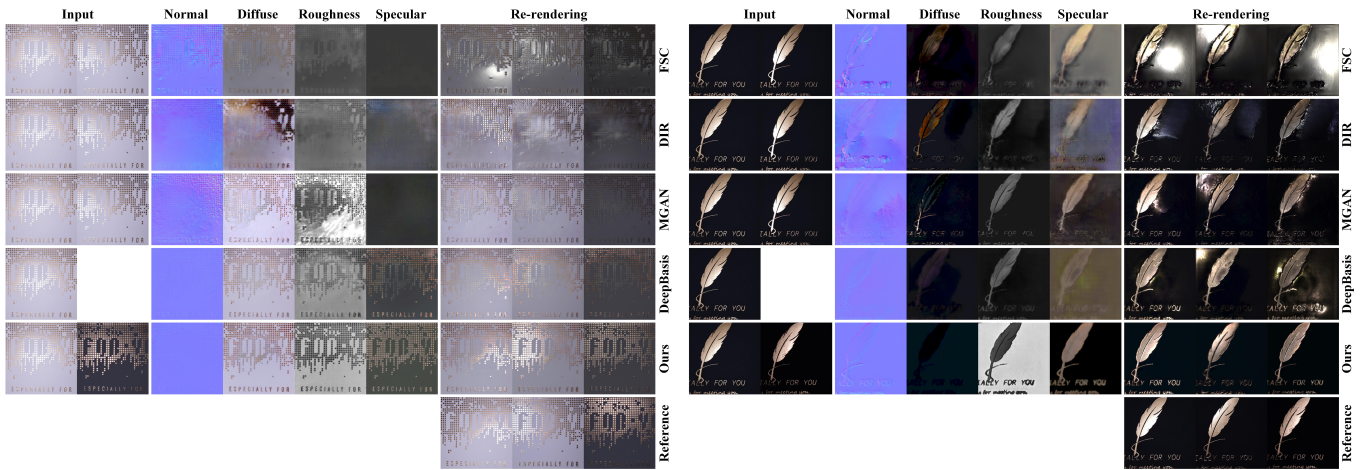


Fig. 15. Fair Comparison on Real Data. We compare our results against FSC of Deschaintre et al. [2019], DIR of Gao et al. [2019], MGAN of Guo et al. [2020] and DeepBasis of Wang et al. 2023.

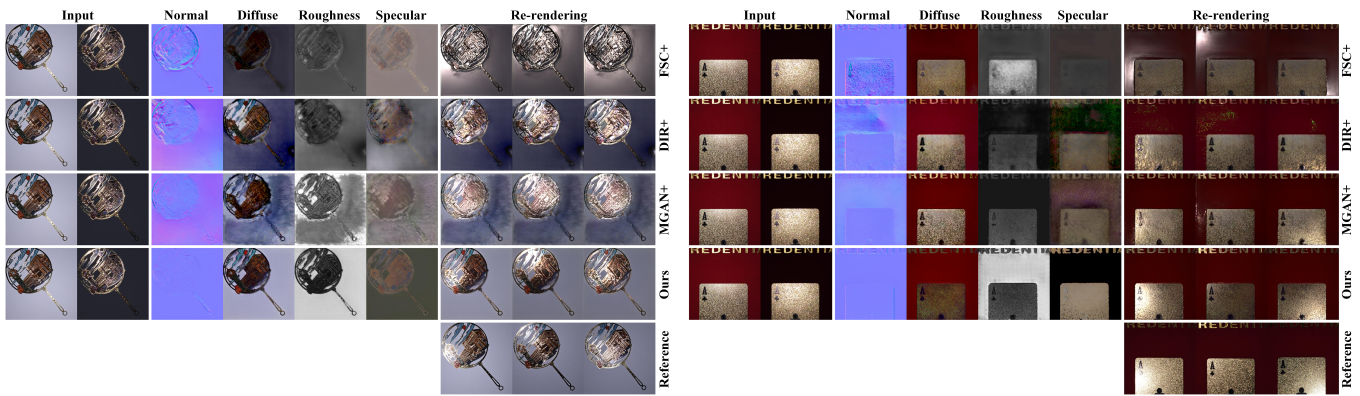


Fig. 16. Enhanced Comparison on Real Data. We provide our novel combination of near-field and far-field inputs to FSC+, DIR+, MGAN+.

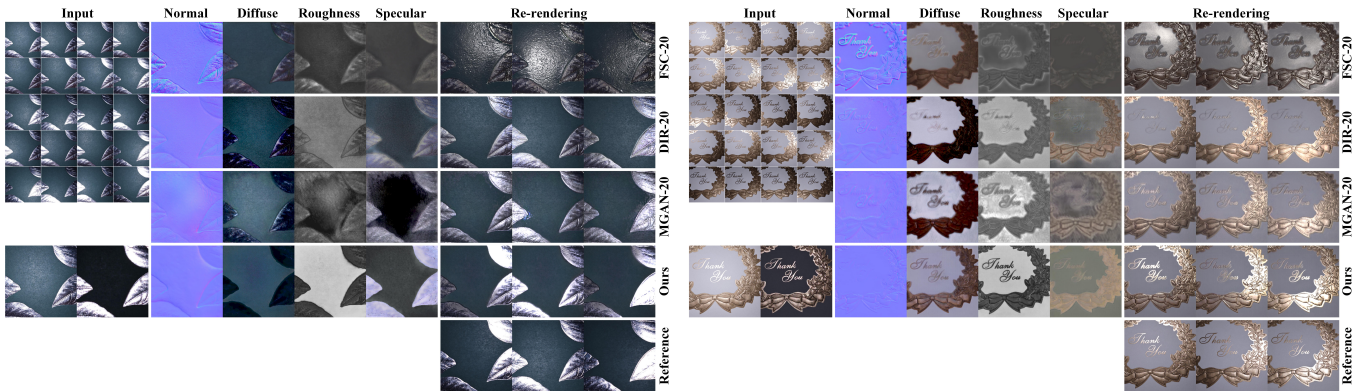


Fig. 17. Challenging comparison on real data against multi-image methods with 20 inputs.