EBREnv: SVBRDF Estimation in Uncontrolled Environment Lighting via Exemplar-Based Representation

LI WANG, Tianjin University, China
JIAJUN ZHAO, Tianjin University, China
LIANGHAO ZHANG, Tianjin University, China
FANGZHOU GAO, Tianjin University, China
JIAWAN ZHANG*, Tianjin University, China

Fig. 1. Under a co-located camera and flashlight capture setup, we proposed a novel method to estimate SVBRDF in uncontrolled environment lighting. Our approach introduces an effective exemplar-based representation to enhance the prediction and utilization of environment lighting. Compared to previous co-located SVBRDF estimation methods, our method achieves high-quality on-site SVBRDF recovery without the need for an extremely low lighting intensity capture environment. Here, we present four estimated SVBRDF results from real scenes and their corresponding re-rendering images.

Recovering spatial-varying bi-directional reflectance distribution function (SVBRDF) from as few as possible captured images has been a challenging task in computer graphics. Benefiting from the co-located flashlight-camera capture strategy and data-driven priors, SVBRDF can be estimated from few input images. However, this capture strategy usually requires a controllable darkroom environment, ensuring the flashlight is a single light source. It is often impractical during on-site capture in real-world scenarios. To support SVBRDF estimation in an uncontrolled environment, the key challenge lies in the high-precise estimation of unknown environment lighting and its effective utilization on SVBRDF recovery. To address this issue, we proposed a novel exemplar-based environment lighting representation, which is easier to use for neural networks. These exemplars are a set of rendered images of selected materials under the environment lighting. By embedding the rendering process, our approach transforms environment lighting represented in the spherical domain into the sample-surface domain, thereby achieving the domain alignment with input images. This significantly reduces the network's learning burden, resulting in a more precise environment lighting estimation. Furthermore, after lighting prediction, we also present a dominant lighting extraction algorithm and an adaptive exemplar selection algorithm to enhance the guidance of environment lighting in SVBRDF estimation. Finally, considering the distant contribution of environment lighting and point lighting to SVBRDF recovery, we proposed a well-designed cascaded network. Quantitative assessments and qualitative analysis have demonstrated that our method achieves superior SVBRDF estimations compared to previous approaches. The source code will be released.

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

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

ACM Reference Format:
Li Wang, Jiajun Zhao, Lianghao Zhang, Fangzhou Gao, and Jiawan Zhang. 2025. EBREnv: SVBRDF Estimation in Uncontrolled Environment Lighting via Exemplar-Based Representation. In *SIGGRAPH Asia 2025 Conference Papers (SA Conference Papers '25), December 15–18, 2025, Hong Kong, Hong Kong*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3757377.3763905

1 Introduction

Conveniently recovering high-quality material reflectance properties from real world, such as spatially-varying BRDF (SVBRDF) remains challenging. The key lies in using common daily devices with minimal capture effort. The mobile phone, as the most common device, has attracted significant attention, as shown in Fig. 1. Its camera and nearly co-located flashlight provide a rich sampling of material appearance [Aittala et al. 2015; Gao et al. 2019; Wang et al. 2024], meanwhile avoiding the extra lighting calibration. Combining

*Corresponding authors.

Authors' Contact Information: Li Wang, Tianjin University, Tianjin, China, li_wang@tju.edu.cn; Jiajun Zhao, Tianjin University, Tianjin, China, jjzhao@tju.edu.cn; Lianghao Zhang, Tianjin University, Tianjin, China, lianghaozhang@tju.edu.cn; Fangzhou Gao, Tianjin University, Tianjin, China, gaofangzhou@tju.edu.cn; Jiawan Zhang, Tianjin University, Tianjin, China, jwzhang@tju.edu.cn.

SA Conference Papers '25, Hong Kong, Hong Kong

ACM ISBN 979-8-4007-2137-3/2025/12
https://doi.org/10.1145/3757377.3763905

with the learned prior from data, SVBRDF can be recovered from few inputs [Deschaintre et al. 2018; Guo et al. 2020; Wang et al. 2023; Zhou and Kalantari 2022]. However, co-located capture typically requires a darkroom environment to ensure that the flashlight is a single light source. This imposes a strict constraint on the capture environment, which is often unavailable in real-world scenarios.

When capturing SVBRDF in an uncontrolled environment, material appearance is also influenced by environment lighting. Although this could theoretically complement the limited information from co-located lighting, the random and unknown nature of environment lighting prevents its effective use for SVBRDF estimation. Furthermore, it also disrupts the stable activation pattern of material appearance from co-located lighting, increasing SVBRDF estimation difficulty. Benefited from the diffusion model, some works have attempted to ignore the environment lighting influence by directly generating SVBRDF from a single image [Vecchio et al. 2024] or the flash/no-flash pair [Sartor and Peers 2023]. However, these generative models cannot guarantee the semantical alignment with the input material sample. To leverage environment lighting, some methods [Boss et al. 2020] first predict environment lighting from flash/no-flash pair and then use it to guide more accurate SVBRDF estimation. However, traditional lighting representations like Spherical Gaussian (SG) used in their method, are ill-suited for direct prediction and effective network input, resulting in poor lighting estimation and inefficient guidance for SVBRDF recovery. The main reason lies in that environment lighting is represented in spherical domain, while material appearance image is expressed in sample-surface domain. This additional domain transformation complicates network training. Therefore, the key challenge is to explicitly relate these two domains, to enable easier environment lighting prediction and more accurate SVBRDF estimation.

In this paper, we present a novel exemplar-based environment lighting representation tailored for easier network use. The key observation is that, because the sample-surface domain is known and fixed for near-planar material reflectance estimation, we can transfer environment lighting into a set of appearance images in this domain by rendering with chosen materials. As shown in Fig. 2, compared to traditional representations in spherical domain, such as SG, our exemplar-based representation builds direct pixel-wise correlations with input appearance image. This simplifies the cross-domain regression problem of lighting prediction and utilization into a translation task within the same domain. Additionally, since our goal is to recover material properties, lighting estimation is used only to reduce ambiguity in material prediction. Thus, we only need to estimate the lighting's effect after convolution with the material, rather than the full environment lighting. As the material is unknown, we choose to estimate a set of known-material exemplars to approximate this effect, further simplifying network learning. After lighting prediction, to enhance its utilization on SVBRDF estimation, such as the residual computation between inputs and rendered images, the forward rendering needs to be supported. Given the exemplar materials are known, we propose an inverse rendering method to convert our well-predicted exemplar images back into traditional SG representation. Furthermore, to better guide subsequent material recovery, we design an adaptive selection algorithm to choose the most suitable exemplar images as the input of SVBRDF estimation network. Finally, considering that random environment lighting negatively influences specular reflectance estimation but not diffuse reflectance, we design a cascaded network to separately recover diffuse and specular reflectance.

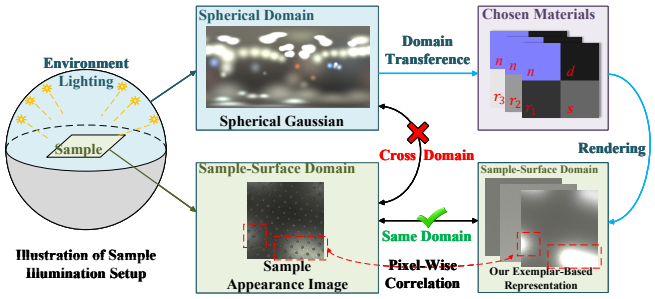

Fig. 2. We present an exemplar-based representation of environment lighting, including a set of rendered images under different chosen materials. These materials share uniform reflectance but differ roughness values, thereby carrying different lighting information. Compared to Spherical Gaussian representations, it provides a direct pixel-wise correlation with sample appearance image, as shown in red boxes, making it easier for lighting prediction and material recovery networks to use.

In summary, we have the following contributions:

- We present a novel exemplar-based lighting representation to build the explicit relation between spherical domain and sample-surface domain for easier network use.
- We present a method to extract SG from predicted exemplars and introduced an adaptive selection algorithm to better utilize the predicted lighting for subsequent material recovery.
- We designed a cascaded network to separately recover diffuse and specular reflectance, effectively isolating the negative impact of environment lighting.

2 Related Work

We review recent methods for material estimation and categorize them based on the illumination type.

2.1 Point Lighting

Given the convenience of mobile phone capture and rich reflectance details provided by a co-located flashlight, many works have attempted to recover SVBRDF. Deschaintre et al. [2018] and Li et al. [2018] respectively introduced SVBRDF datasets, enabling single-image estimation with deep priors. Based on the dataset, more effective network architectures are proposed, such as Highlight-Aware (HA) convolution [Guo et al. 2021], adversarial network [Vecchio et al. 2021; Zhou et al. 2023, 2022; Zhou and Kalantari 2021], two-level basis material models [Wang et al. 2023] and intermediate targets decomposition [Nie et al. 2025]. To further reduce the visual gap between synthetic and real data, meta-learning techniques [Fischer and Ritschel 2022; Zhou and Kalantari 2022] and recurrent neural network [2024a] are employed to perform test-time optimization. Additionally, Henzler et al. [2021] proposed a latent-based method to generate infinite stationary materials. Guo et al.

[2023] proposed an divide-and-conquer solution for high-resolution materials.

To overcome single-image insufficient information, Deschaintre et al. [2019] introduced a pooling layer to extend their work to support multi-image inputs. Gao et al. [2019] and Guo et al. [2020] utilized autoencoder [Hinton and Salakhutdinov 2006] and StyleGAN2 [Karras et al. 2020], respectively, to embed SVBRDF into latent space for inverse rendering optimization. Furthermore, Zhu et al. [2023] designed a two-branch network to learn lighting priors, thereby removing the need for precise lighting calibration across multiple images. Similarly, Luo et al. [2024b] introduced a Graph Convolutional Network to extract inter-image correlations for better initialization of latent-space optimization. Additionally, Wang et al. [2024] proposed a near-far-field capture strategy to enhance material capture efficiency. However, these methods rely on a controlled capture environment where the flashlight serves as the sole light source, limiting their on-site capture applicability. In contrast, our approach enables material capture under uncontrolled environment lighting, making it more practical for real-world applications.

2.2 Environment Lighting

Several approaches address SVBRDF estimation directly under environment lighting, removing the controlled lighting constraint. Li et al. [2017] proposed a self-augmented training strategy to address data scarcity, later extended to unlabeled data [Ye et al. 2018]. Martin et al. [2022] proposed a hybrid method to combine deep learning and numerical approaches. Recently, diffusion models have shown strong performance in image generation. Sartor et al. [2023] introduced a diffusion model to directly generate SVBRDF under unknown environmental lighting. Vecchio et al. [2024] extended this to generate tileable materials. Differently, Lopes et al. [2024] estimate SVBRDF by decomposing a texture generated by a diffusion model from a real-world image. Although these methods no longer require shooting under weak lighting intensity conditions, uncontrolled environmental lighting cannot reliably activate material reflectance, particularly specular reflectance. This increases the difficulty of model training and leads to usually lower material estimation quality from these methods.

To address the above problem, some methods capture two images per sample: one flash-on, one flash-off, both under environment lighting. This capture strategy was first introduced by Aittala et al. [2015]. They used the flash-on image for reflectance details and the flash-off image for structure guidance, but their method only works for stationary materials, restricting its generalizability. Sartor et al. [2023] also fine-tuned a variant of their diffusion model with two-shot images as input. Additionally, Boss et al. [2020] estimated SG environment lighting to explicitly utilize extra lighting information for better SVBRDF estimation. However, the cross-domain prediction from the sample-surface domain to the spherical domain poses a significant challenge for the network in predicting full SG parameters. As a compromise, they opted to predict only the amplitude of the SGs while keeping the other parameters fixed. This trade-off extremely limits the expressive power of the SG representation. In contrast, our exemplar-based method enables more accurate environment lighting estimation. Additionally, our SG extraction keeps forward rendering ability and full environment lighting expression, thereby enabling more comprehensive utilization of lighting information, leading to improved SVBRDF estimation quality.

3 Method

3.1 Problem Statement

Our goal is to estimate spatial varying material reflectance from flash/no-flash images under environment lighting. The material sample is assumed to be a nearly planar surface with geometric details modeled by a normal map. The reflectance properties are represented by Cook-Torrance BRDF model [Cook and Torrance 1982] with GGX microfacet distribution [Walter et al. 2007]. Therefore, they can be represented by four maps: normal map n, diffuse map d, roughness map r and, specular map s. Additionally, flash/no-flash pair images are captured by a mobile phone camera with a co-located flashlight at a short interval of time. Consequently, the environment lighting is assumed to remain consistent between these two shots. Our method aims at learning a mapping function F to recover material maps $M = \{n, d, r, s\}$ from flash image I_f and no-flash image I_{nf}, as follows:

$$\begin{aligned} M &= F(I_f, I_{nf}), \\ I_f = R(M, L_p + L_{env}), &\quad I_{nf} = R(M, L_{env}), \end{aligned} \tag{1}$$

where R is the rendering process, L_p is the point lighting, L_{env} is the environment lighting. Given the co-located central capture setting, L_p is inherently known to the network. Therefore, the key of inverse rendering in this setup lies in accurately estimating L_{env}. To achieve this, inspired by Wang et al. [2023] and Zhang et al. [2024], we propose an exemplar-based representation of L_{env}, as follows:

$$\{I_{ex_i}\}_{i=1}^{N} = \{R(M_{ex_i}, L_{env})\}_{i=1}^{N}, \tag{2}$$

where $\{I_{ex_i}\}_{i=1}^{N}$ represent a set of exemplar images rendered using exemplar materials $\{M_{ex_i}\}_{i=1}^{N}$. These materials share nearly identical and uniform reflectance properties, differing only in their uniform but varying roughness. Practically, we set diffuse to 0.1, specular to 0.2, and roughness is determined by Sec. 5.3.1. As shown in Fig. 2, due to pixel-wise correlations between exemplar and input images, $\{I_{ex_i}\}_{i=1}^{N}$ is easier predicted from I_f and I_{nf} than traditional representations. Therefore, different from Wang et al. [2023] and Zhang et al. [2024] which use fixed, heuristic exemplars for forward rendering under known lighting, we directly predict exemplar images using Lighting Net of Fig. 3, and the flash-only image is also predicted for subsequent guidance. However, there are several technical challenges in using this representation for SVBRDF estimation guidance. Firstly, the environment lighting represented in exemplar images cannot be directly utilized for forward rendering, restricting the use of rendering to provide richer information. Secondly, given that each material sample has a different roughness level, selecting the optimal exemplar images is crucial to achieving effective guidance in SVBRDF estimation. Finally, although environment lighting can be approximated through exemplar prediction, its inherent randomness still leads to unstable activation of material appearance, thereby increasing the complexity of network learning. Thus, designing an effective network structure to address this issue remains a challenge.

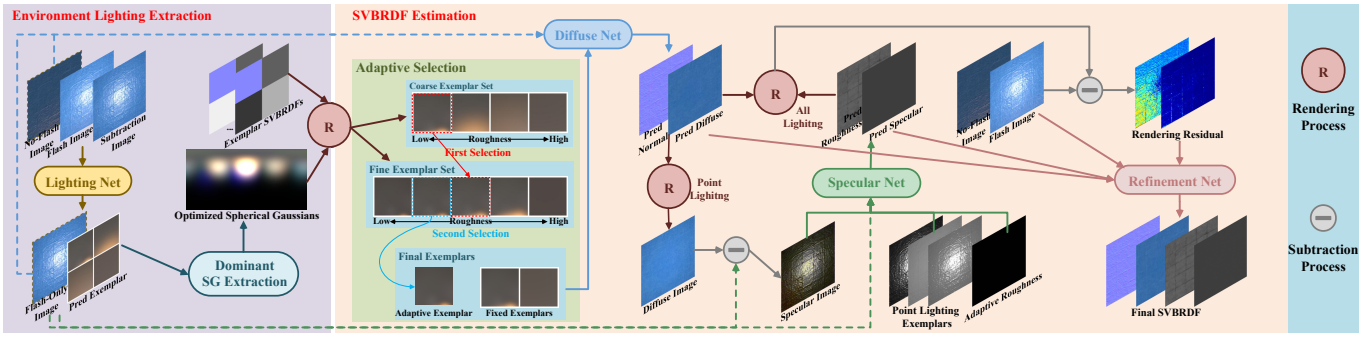

Fig. 3. Our method consists of two parts. The first part is the environment lighting extraction, which includes two steps: the prediction of environment lighting exemplar images and the extraction of dominant SGs. The flash-only image is also predicted alongside the exemplar images. After this part, the estimated environment lighting is obtained. The second part is SVBRDF estimation. To better utilize the extracted environment lighting, this part contains two special designs: adaptive exemplar selection and well-designed cascaded networks. Through the adaptive selection, the proper combination of exemplar images are determined for guiding the subsequent diffuse network. Considering the unstable specular reflectance activation of uncontrolled environment lighting, diffuse and normal maps are first predicted using inputs under environment lighting, however the roughness and specular maps are predicted with only point lighting images. Finally, the extracted SGs enable the computation of rendering residual, thereby improving the SVBRDF estimation quality by refinement network.

3.2 Algorithm

To address the above challenges, we propose a new pipeline, as shown in Fig. 3. Overall, It has two parts, including environment lighting extraction and SVBRDF estimation. In the former, we predict exemplar images and introduce dominant spherical gaussian extraction to enable its rendering ability. In the latter, we present an adaptive selection algorithm and a well-designed cascaded network to utilize predicted environment lighting for better SVBRDF estimation. In the following sections, we discuss the details.

3.2.1 Dominant Spherical Gaussian Extraction. To support forward rendering, our exemplar-based representation needs to be converted back into the spherical domain. Here, we adopt SG representation. Given the known exemplar materials and capture setting, extracting SG from exemplar images becomes a standard optimization-based inverse rendering problem. In this problem, its challenge lies in determining the appropriate initialization of SG parameters from exemplar images, including the axis, amplitude, sharpness, and number. Considering a single image pixel, theoretically its value is contributed from all hemispherical lighting. Our key observation is that practically, this value is always dominated by the lighting from the reflected direction of the viewing vector. Therefore, for this pixel, the reflected vector is a proper initialization of a SG axis. Furthermore, according to the microfacet theory, a lobe range centered on the reflection direction have the similarly significant impact on the value of the pixel, which means that several neighboring pixels may share same dominant SG. Based on the above analysis, we propose dominant spherical gaussian extraction, as illustrated in Fig. 4. Firstly, we employ a quad-tree algorithm to subdivide the exemplar images into patches. The subdivision terminates when the variance of the patch is below a certain threshold. For each patch, the central pixel is used to compute the SG parameters: the SG axis corresponds to the reflected vector of the viewing direction, the amplitude is equal to the pixel value, and the sharpness is selected from a pre-defined set based on the quad-tree level. Additionally, to address the limitation that the quad-tree can only divide patches in a fixed structure, we merge similar neighboring SGs to provide a more reasonable overall SG distribution, while reducing their significant redundancy. Taking the merged results as initialization of SGs, we perform the following optimization:

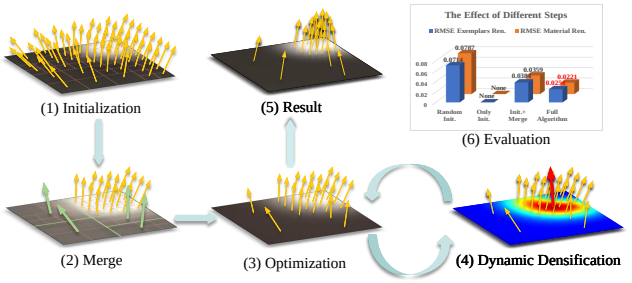

Fig. 4. Dominant Spherical Gaussian Extraction. (1): Extract per-patch dominant environment lighting. (2): Merge SGs obtained during initialization. (3, 4): Optimize the obtained SGs, and dynamically add SGs to areas with insufficient representation in the process. (5): Obtain the final dominant SG representation of the environment lighting. (6): Evaluate the effect of different steps on optimization performance. Note that Only Init. cannot be performed, due to huge GPU memory cost. The SG numbers are up to 64.

$$\arg\min_{SGs} \sum_{i}^{N} \mathcal{L}_{opt}(I_{ex_i}, R(SGs, M_{ex_i}))$$
$$\mathcal{L}_{opt}(y, \hat{y}) = \lambda(\|\hat{y} - y\|_2 + \|\log \hat{y} - \log y\|_2) \tag{3}$$

where $\mathcal{L}_{opt}$ is the loss function and λ is the weight corresponding to different exemplar. Finally, to further improve the expressiveness of SGs, we incorporate a dynamic densification mechanism during optimization. Specifically, we utilize the MS-SSIM [Wang et al. 2003] algorithm to identify the regions where SGs fail to capture sufficient details, and then dynamically add new SGs to these regions, thereby enriching the representation. To evaluate the expressiveness of our extracted dominant SGs, we compare them against the ground-truth

environment lighting, as shown in Fig. 5. The comparison demonstrates that our method achieves high similarity in the dominant regions of the environment lighting, which ensures precise forward rendering at the current camera setting. To balance time and quality, the optimization process is limited up to 400 iterations, taking about 5 seconds per sample test on a single RTX 3090 GPU.

3.2.2 Adaptive Exemplar Selection. After obtaining dominant SGs, theoretically, arbitrary exemplar images can be rendered. However, considering the computational burden, it is impractical to use an infinite number of exemplars as guidance for subsequent SVBRDF estimation. The key to efficient guidance lies in the pixel-wise correlations between the input and the exemplar image. Thus, the ideal exemplar image should closely match the spatial structure of the input image. A straightforward approach is to identify the exemplar by optimization under a structural similarity loss. However, integrating such an optimization algorithm into the training process would result in an impractical time cost. To solve this problem, we propose a coarse-to-fine candidate selection strategy, as shown in green box of Fig. 3. First, we uniformly sample several exemplars from the roughness range [0,1] to form a coarse set and select the most suitable exemplar based on structural similarity to the input no-flash image. Next, we refine the selection by constructing a fine-grained set from the neighbors of the selected exemplar in the roughness space. The final exemplar is chosen from this fine-grained set. Additionally, a single uniform exemplar often fails to provide sufficient environment lighting information for spatially varying material. To complement it, we include two fixed exemplars with low and high roughness levels. These fixed exemplars remain consistent across different material samples, enabling the network to more easily interpret lighting cues from this representation. Furthermore, they serve as anchors for the selected exemplar, reducing the network's learning burden and improving the robustness of the guidance.

3.2.3 Cascaded Network. The unstable activation of material appearance caused by random environment lighting primarily affects specular reflectance. Conversely, the direction-independent diffuse reflectance is often effectively activated in appearance. Inspired by previous cascaded network design [Li et al. 2020; Martin et al. 2022; Nie et al. 2025], we first estimate diffuse terms with the help of predicted environment lighting. Subsequently, we recover the specular terms only using the point lighting information, thereby avoiding the negative impact of random environment lighting. Specifically, we design a cascaded network to separate the SVBRDF estimation into three stages. Firstly, the diffuse network leverages no-flash image, predicted flash-only image and selected environment lighting exemplars to estimate normal and diffuse maps. Secondly, the diffuse image is rendered using the predicted normal and diffuse maps under point lighting. This enables the specular image extraction by subtracting the rendered diffuse image from the flash-only image. Therefore, specular network takes the specular image, the flash-only image, and point lighting exemplars selected using the same strategy as the environment ones, as inputs and predicts roughness and specular maps. Finally, the refinement network integrates the previously predicted SVBRDF maps, the flash/no-flash pair, and the computed rendering residual as inputs to recover the final SVBRDF.

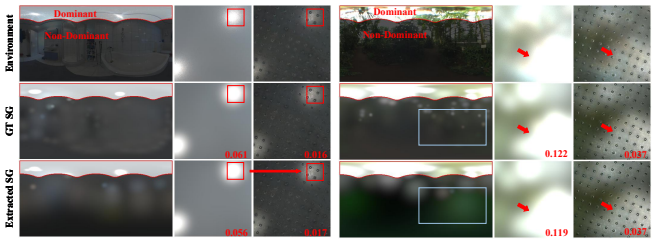

Fig. 5. The figure includes three rows: the sphere map of environment lighting, the GT SG obtained from sphere map, and the SG extracted from our exemplars. We render both the exemplar and the material, and calculate RMSE against GT rendering to evaluate the quality of lighting. The red markers highlight that our rendering results maintain high similarity to the ground truth both visually and numerically. Additionally, while there are noticeable differences in the non-dominant regions (blue boxes), these differences do not affect the rendering results.

4 Implementation

4.1 Network Architecture

In our method, there are four networks: lighting network, diffuse network, specular network and refinement network. All the networks are based on NAFNet [Chen et al. 2022], featuring a 4-layer encoder-decoder structure with skip connections and a single middle layer. Each layer adopts a stack of NAFBlocks with a base feature width of 36. The number of stacked NAFBlocks is 2, 2, 4, and 8 for the encoder layers, 2, 2, 2, and 2 for the decoder layers, and 12 for the middle layer. The inputs and outputs of each network are detailed in Sec. 3.2.3. Among them, all images and exemplars have three channels representing RGB channels. When they are fed into the network, an additional logarithmic transformation is performed on these images to flatten the dynamic range. The SVBRDF has 10 channels: 3 channels for normal, 3 channels for diffuse, 3 channels for specular and 1 channel for roughness.

4.2 Training Details

The training dataset is from MatSynth [Vecchio and Deschaintre 2024], which contains 5,700 meta SVBRDF entries. Using the augmentation strategy provided by authors, we generate 598,500 training samples with 256×256 resolution by applying rotation angles of 0°, 45°, 90°, 180°, and 270°. Additionally, we collected 265 sphere maps as environment lighting, processed into Spherical Gaussians (SGs) via an optimization procedure. Consequently, leveraging real-time rendering during training, we can randomly choose training samples from up to 598,500×265 combinations. Given that the optimization process of dominant SG extraction is time-consuming at training time, we adopt a two-stage training strategy to reduce time cost. Firstly, we train all networks using the ground-truth environment lighting for 5 epochs. The diffuse, specular, and refinement networks are trained sequentially, as each network in the cascade relies on the outputs of the preceding network as inputs. They have the same decay-schedule learning rate ranging from 5e-4 to 1e-6. At the second stage, we select 51,300 samples from the full training set and pre-compute the extracted SGs using our proposed method. The diffuse, specular, and refinement networks are then fine-tuned

Table 1. Numerical comparison on 86 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 top part is a comparison on near-field flash/no-flash capture strategy and the bottom part is a comparison on near-far-field capture strategy.

Methods	RMSE↓					LPIPS↓
	Norm.	Diff.	Rough.	Spec.	Ren.	Ren.
Comparison on Near-Field						
Matfusion	0.0823	0.1365	0.1604	0.0695	0.0658	0.1657
Two-Shot	0.0686	0.0503	0.1103	0.0457	0.0476	0.1166
Ours	**0.0351**	**0.0285**	**0.0718**	**0.0446**	**0.0394**	**0.0568**
Comparison on Near-Far-Field						
NFPLight	0.0406	0.0351	0.0615	0.0474	0.0350	0.0543
Ours	**0.0277**	**0.0199**	**0.0487**	**0.0278**	**0.0314**	**0.0342**

on this training subset for 10 epochs, with a learning rate from 5e-5 to 5e-7. The source code will be released.

5 Experiments

5.1 Comparison Experiments

We separately trained our method using two capture strategies: (1) the near-field flash/no-flash pairs (two shots) or (2) the near-far-field flash/no-flash pairs (four shots), proposed by NFPLight [Wang et al. 2024]. In the latter, all near-field images are replaced with near-far-field images, and the real data capture follows the method provided by the author. For the traditional near-field capture strategy, we compared our method against SOTA SVBRDF estimation methods with flash/no-flash input images, including Two-Shot [Boss et al. 2020] and MatFusion [Sartor and Peers 2023]. For the near-far-field strategy, we compared our method against NFPLight, using the predicted flash-only images generated by our lighting network as its inputs. All above results were obtained from the source code provided by the authors. For a fair comparison, we re-trained Two-Shot and NFPLight on our training dataset. Additionally, for NFPLight, we fine-tuned its base model using the outputs of our lighting network.

5.1.1 Comparison on Synthetic Data. We first numerically compared 86 synthetic scenes from MatSynth test set [Vecchio and Deschaintre 2024], which were not used in training. We also extensively compared results on other test datasets [Ma et al. 2023; Sartor and Peers 2023], detailed in supplementary materials. We assessed reflectance estimation quality by computing per-map RMSE, and evaluated re-renderings under 30 random lighting/viewing directions using RMSE and LPIPS [Zhang et al. 2018], as shown in Table 1. Compared to Two-Shot and MatFusion, our method achieved significant improvements in normal and diffuse estimation, leading to better overall rendering quality. Integrating the near-far-field strategy further improved all SVBRDF estimations, especially roughness.

Moreover, we also performed a visual comparison. Figure 7 compares our near-field results with prior flash/no-flash methods. As a generative model, MatFusion struggles to accurately capture fine reflectance details. Meanwhile, Two-Shot fails to estimate precise environment lighting, leading to incorrect lighting/reflectance decomposition, especially for diffuse maps. By accurately recovering and using environment lighting, our method can recover a cleaner diffuse map. We also compare near-far-field results with NFPLight in Fig. 8. While relying on our flash-only image prediction, NFPLight can normally work on uncontrolled environment scenes, it struggles to utilize the environment lighting information to complement the lost information due to central over-exposure issues. In contrast, our method fully leverages the additional information provided by the environment lighting, effectively mitigating the negative impacts of over-exposure and improving the quality of SVBRDF estimation.

Table 2. Numerical Comparison on 60 Real Scenes. Each scene contains 6 novel-lighting reference images, and we evaluate the re-rendering images by RMSE and LPIPS. The lowest errors are highlighted in bold. The left part is a comparison on near-field flash/no-flash capture strategy and the right part is a comparison on near-far-field capture strategy.

Comparison on Near-Field			Comparison on Near-Far-Field		
Methods	RMSE↓	LPIPS↓	Methods	RMSE↓	LPIPS↓
Matfusion	0.1684	0.3099	NFPLight	0.1424	0.2516
Two-Shot	0.1674	0.3167	Ours	**0.1266**	**0.1806**
Ours	**0.1514**	**0.2061**	-	-	-

5.1.2 Comparison on Real Data. To evaluate SVBRDF estimation quality on real scenes, we captured reference images under a controllable lighting environment using a mobile phone in professional mode with fixed camera settings (e.g., shutter speed, ISO) to ensure consistency between flash and no-flash captures. We captured input images with environment lights on and novel-lighting reference (flash-only) images with environment lights off for fair re-rendering comparison. A total of 60 real scenes were captured, with camera calibration following [Guo et al. 2020]. The numerical comparison results, shown in Table 2, demonstrate that our estimated SVBRDF quality surpasses that of previous methods. Furthermore, we provide a visual comparison. In Fig. 9, we first compare the near-field results of our method against Two-Shot and MatFusion. Consistent with the synthetic results, MatFusion, as a generative model, struggles to recover input-aligned reflectance details, leading to noticeable noise artifacts in these two samples. For Two-Shot, due to the inaccurate lighting estimation, their results often bake lighting effects into the diffuse or specular map, as indicated by the red arrows. In contrast, our method produces cleaner SVBRDF maps and significantly mitigates baked lighting effects. In Fig. 10, we provide a comparison on the near-far-field capture strategy against NFPLight. Since NFPLight is specifically designed for flash-only capture, it is sensitive to environmental lighting perturbations, even when utilizing our predicted flash-only images. This limitation leads to incorrect predictions, especially on the roughness map. Additionally, the central over-exposure issue in flash-only capture causes a loss of local details. In contrast, our method leverages the environment lighting information to recover these lost details, as illustrated in the red box of the right sample. In summary, whether the near-field or near-far-field strategy is used, our methods can effectively utilize environment lighting information to recover high-quality SVBRDF. Finally, we also test higher-resolution real-data results under outdoor environment lighting conditions, as shown in Fig. 11 and 12. More results are available in supplementary materials.

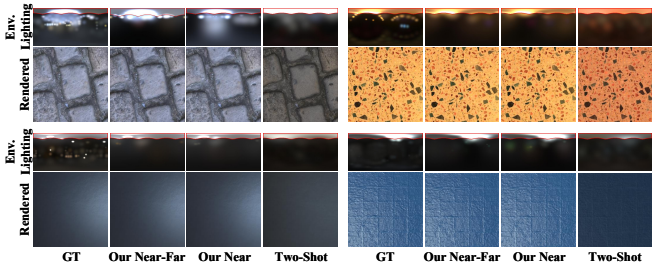

Fig. 6. Comparison of the estimated environment lighting between our near/near-far field method and Two-Shot. The red-bordered area on the sphere map indicates the dominant environment lighting. For reference, the first column of each sample includes the corresponding results from GT SG.

5.2 Comparison on Lighting Estimation

To evaluate our environment lighting estimation quality, we compare our near-field and near-far-field results against Two-Shot [Boss et al. 2020]. The visual comparison is shown in Fig. 6. Since directly predicting SGs is challenging, Two-Shot fixes most SG parameters and predicts only amplitude, losing many lighting details. In contrast, our algorithm predicts exemplar images and extracts the dominant environment lighting from it, greatly reducing the burden on the network and obtaining more precise estimation. Furthermore, we conducted a numerical comparison. To account for the impact of material roughness on lighting estimation, we categorized materials into three groups: low, medium, and high roughness levels. For each group, we evaluate the estimated quality of environment lighting by calculating the RMSE on rendered images of 86 test materials, as presented in Fig. 13(a). The results show that our estimation quality significantly outperforms Two-Shot across all roughness levels, and it is robust to variations in roughness.

5.3 Ablation Study

5.3.1 The Effect of Different Exemplar Combinations. As the number of exemplars increases, the lighting network's computational load increases, and its prediction accuracy decreases. Therefore, determining the optimal number of exemplars is crucial. Additionally, since different exemplars reveal different environment lighting details, selecting the optimal combination is also important. To find this optimal setup, we discretize the exemplar roughness into 9 values (0.1 to 0.9) and evaluate the resulting environment lighting accuracy for each combination. Since testing all 511 combinations is impractical, we employ a greedy algorithm to reduce complexity. First, we find the best single exemplar. Then, we fix it while searching for the best second exemplar to form an optimal pair. This process continues up to nine exemplars, reducing the number of experiments to 45. However, retraining the network 45 times is still too computationally expensive. To address this, instead of predicting exemplars, we use ground-truth SGs to render the input exemplars, significantly reducing the time cost. This ideal experiment identifies the optimal roughness combination for each exemplar count, as shown in blue in Fig. 13(b). Finally, we retrain the lighting network 9 times using these combinations. The results, presented by the green line in Fig. 13(b), indicate that the optimal number of exemplars is 4.

5.3.2 Ablation Study on SVBRDF Estimation. To evaluate the effectiveness of the adaptive exemplar selection and cascaded network design, we conducted four experiments: (1) training an end-to-end network using only flash/no-flash images as inputs (denoted as *w/o Cas.+Exp.*), (2) training an end-to-end network including our adaptive exemplars as additional inputs (denoted as *w/o Cas.*), (3) training a cascaded network using four fixed exemplars instead of our adaptive exemplars (denoted as *w/o Adp.*). (4) training a cascaded network and adaptive exemplars (denoted as *Full Model*). For consistency, each network was trained on near-far-field capture under the same training strategy. Numerical and visual evaluation results are shown in Fig. 14. Comparing *w/o Cas.+Exp.* and *w/o Cas.* demonstrates that introducing exemplars effectively decouples the environment lighting from estimated SVBRDF, as indicated at the red box. Furthermore, comparing *w/o Cas.* and *Full Model* shows that the cascaded network design better leverages lighting information to enhance overall quality. Finally, the comparison between *w/o Adp.* and *Full Model* highlights the effectiveness of the adaptive exemplar mechanism, especially in enhancing specular estimation.

5.3.3 The Effect of Environment Lighting Intensity. To evaluate the effect, we categorize environment lighting data into three levels of intensity: low, medium, and high. For each level, we evaluate the accuracy of our estimated SVBRDF on 86 synthetic data, with results shown in Fig. 15. With increased environment lighting intensity, some highly reflective materials exhibit over-exposure regions in the no-flash image. In these regions, all available input information is lost, resulting in poor SVBRDF estimation, especially on diffuse map. Apart from these cases, our method remains robust to variations in environment lighting intensity. More experiments regarding environment lighting are available in the supplementary material.

6 Limitation and Future Work

Although our method support near-far-field capture strategy, we haven't model the dynamic shadow caused by the variation of capture distance. When the environment light source is almost directly above the captured sample, significant shadow variance can occur, leading to incorrect decomposition of lighting and material, as shown in normal capture of Fig. 16. We currently mitigate the issue by manually blocking the dominant light source, reducing shadow variation, as shown in blocked capture of Fig. 16. In future work, simulating the dynamic near-far-field capture process and generating proper training data with dynamic shadows could further improve SVBRDF estimation quality. In addition, our method does not account for global illumination effect, such as self-shadows on height-variation objects (the rightest one of Fig. 16). Although our predicted flash-only images prevent these effects from corrupting the estimated normal, roughness and specular, the diffuse map still bakes artifacts(see red arrows). In contrast, methods like Sartor et al. [2023] and Vecchio et al. [2024] handle such effects better. While our method achieves higher accuracy, these works offer greater flexibility. For example, our approach requires a more complex capture setup, does not work with a single input image, and does not recover a height map. For future work, we are inspired by these methods to incorporate GI-enabled data to mitigate self-shadowing and to potentially simplify our capture requirements.

7 Conclusion

We propose a novel SVBRDF estimation method under uncontrolled environment lighting. Our novel exemplar-based lighting representation and dominant spherical gaussian extraction enable high-quality lighting estimation. Furthermore, with our adaptive exemplar selection algorithm and well-designed cascaded networks, environment lighting information can be effectively utilized to guide SVBRDF estimation. Additionally, our pipeline supports various co-located capture strategies, including traditional near-field two-shot capture and the latest near-far-field four-shot capture. Extensive experiments show our method achieves more accurate SVBRDF estimation than SOTA and yields higher re-rendering quality.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (62172295)

References

Miika Aittala, Tim Weyrich, and Jaakko Lehtinen. 2015. Two-shot SVBRDF capture for stationary materials. *ACM Transactions on Graphics (TOG)*. 34, 4 (2015), 1–13.

Mark Boss, Varun Jampani, Kihwan Kim, Hendrik P. A. Lensch, and Jan Kautz. 2020. Two-Shot Spatially-Varying BRDF and Shape Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 3981–3990.

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. 2022. Simple Baselines for Image Restoration. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 17–33.

Robert L Cook and Kenneth E. Torrance. 1982. A reflectance model for computer graphics. *ACM Transactions on Graphics (TOG)*. 1, 1 (1982), 7–24.

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 (TOG)*. 37, 4 (2018), 1–15.

Valentin Deschaintre, Miika Aittala, Fredo Durand, George Drettakis, and Adrien Bousseau. 2019. Flexible SVBRDF Capture with a Multi-Image Deep Network. *Computer Graphics Forum (CGF)* 38, 4 (2019), 1–13.

Michael Fischer and Tobias Ritschel. 2022. Metappearance - Meta-Learning for Visual Appearance Reproduction. *ACM Transactions on Graphics (TOG)*. 41, 6 (2022), 1–13.

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 Transactions on Graphics (TOG)*. 38, 4 (2019), 1–15.

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 Transactions on Graphics (TOG)*. 40, 4 (2021), 1–14.

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 (TOG)*. 42, 3 (2023), 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.

Philipp Henzler, Valentin Deschaintre, Niloy J. Mitra, and Tobias Ritschel. 2021. Generative modelling of BRDF textures from flash images. *ACM Transactions on Graphics (TOG)*. 40, 6 (2021), 284:1–284:13.

Geoffrey E Hinton and Ruslan R Salakhutdinov. 2006. Reducing the dimensionality of data with neural networks. *Science* 313, 5786 (2006), 504–507.

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)*. 8107–8116.

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 (TOG)*. 36, 4 (2017), 1–11.

Zhengqin Li, Mohammad Shafiei, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. 2020. Inverse Rendering for Complex Indoor Scenes: Shape, Spatially-Varying Lighting and SVBRDF From a Single Image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2472–2481.

Zhengqin Li, Kalyan Sunkavalli, and Manmohan Chandraker. 2018. Materials for Masses: SVBRDF Acquisition with a Single Mobile Phone Image. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 74–90.

Ivan Lopes, Fabio Pizzati, and Raoul de Charette. 2024. Material Palette: Extraction of Materials from a Single Image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Di Luo, Hanxiao Sun, Lei Ma, Jian Yang, and Beibei Wang. 2024b. Correlation-aware Encoder-Decoder with Adapters for SVBRDF Acquisition. In *SIGGRAPH Asia 2024 Conference Papers*. ACM, 1–10.

Xuejiao Luo, Leonardo Scandolo, Adrien Bousseau, and Elmar Eisemann. 2024a. Single-Image SVBRDF Estimation with Learned Gradient Descent. *Computer Graphics Forum (CGF)* 43, 2 (2024), i–iii.

Xiaohe Ma, Xianmin Xu, Leyao Zhang, Kun Zhou, and Hongzhi Wu. 2023. OpenSVBRDF: A Database of Measured Spatially-Varying Reflectance. *ACM Transactions on Graphics (TOG)*. 42, 6 (2023), 1–14.

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 (CGF)* 41, 2 (2022), 163–177.

Yongwei Nie, Jiaqi Yu, Chengjiang Long, Qing Zhang, Guiqing Li, and Hongmin Cai. 2025. Single-Image SVBRDF Estimation Using Auxiliary Renderings as Intermediate Targets. *IEEE Transactions on Visualization and Computer Graphics* 31, 9 (2025), 4908–4922.

Sam Sartor and Pieter Peers. 2023. MatFusion: A Generative Diffusion Model for SVBRDF Capture. In *SIGGRAPH Asia 2023 Conference Papers*. ACM, 1–10.

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 (CVPR)*.

Giuseppe Vecchio, Rosalie Martin, Arthur Roullier, Adrien Kaiser, Romain Rouffet, Valentin Deschaintre, and Tamy Boubekeur. 2024. ControlMat: A Controlled Generative Approach to Material Capture. *ACM Transactions on Graphics (TOG)*. 43, 5 (2024), 1–17.

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 (ICCV)*. 12820–12828.

Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. 2007. Microfacet Models for Refraction through Rough Surfaces. In *Proceedings of the Eurographics Symposium on Rendering*.

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 Transactions on Graphics (TOG)*. 43, 6 (2024), 1–11.

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*. ACM, 1–11.

Zhou Wang, Eero P Simoncelli, and Alan C Bovik. 2003. Multiscale structural similarity for image quality assessment. In *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, Vol. 2. IEEE, 1398–1402.

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 (CGF)* 37, 7 (2018), 201–211.

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/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 586–595.

Yuqing Zhang, Yuan Liu, Zhiyu Xie, Lei Yang, Zhongyuan Liu, Mengzhou Yang, Runze Zhang, Qilong Kou, Cheng Lin, Wenping Wang, and Xiaogang Jin. 2024. DreamMat: High-quality PBR Material Generation with Geometry- and Light-aware Diffusion Models. *ACM Transactions on Graphics (TOG)*. 43, 4 (2024), 1–18.

Xilong Zhou, Milos Hasan, 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 Proceedings*. ACM, 1–11.

Xilong Zhou, Milos Hasan, Valentin Deschaintre, Paul Guerrero, Kalyan Sunkavalli, and Nima Khademi Kalantari. 2022. TileGen: Tileable, Controllable Material Generation and Capture. In *SIGGRAPH Asia 2022 Conference Papers*. ACM, 1–9.

Xilong Zhou and Nima Khademi Kalantari. 2021. Adversarial Single-Image SVBRDF Estimation with Hybrid Training. *Computer Graphics Forum (CGF)* 40, 2 (2021), 315–325.

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.

Pengfei Zhu, Shuichang Lai, Mufan Chen, Jie Guo, Yifan Liu, and Yanwen Guo. 2023. SVBRDF Reconstruction by Transferring Lighting Knowledge. *Computer Graphics Forum (CGF)* 42, 7 (2023).

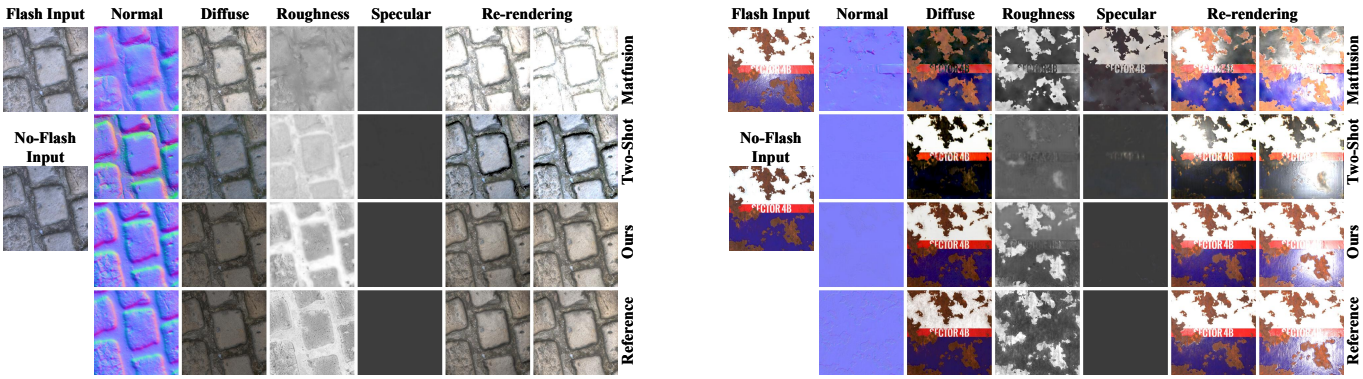

Fig. 7. Synthetic Comparison on Near-Field Capture Strategy. We compare our results against MatFusion of Sartor et al. [2023], Two-Shot of Boss et al. [2020].

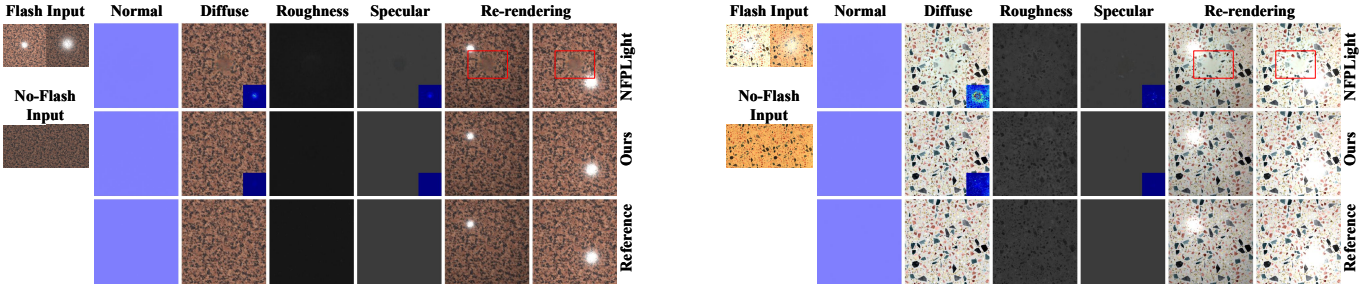

Fig. 8. Synthetic Comparison on Near-Far-Field Capture Strategy. We compare our results against NFPLight of Wang et al. [2024].

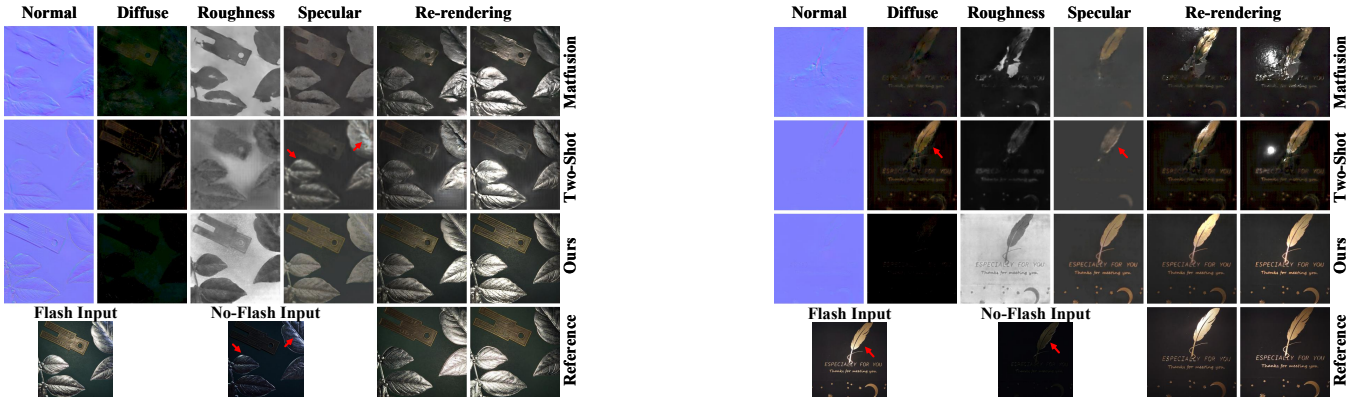

Fig. 9. Real Comparison on Near-Field Capture Strategy. We compare our results against MatFusion of Sartor et al. [2023], Two-Shot of Boss et al. [2020].

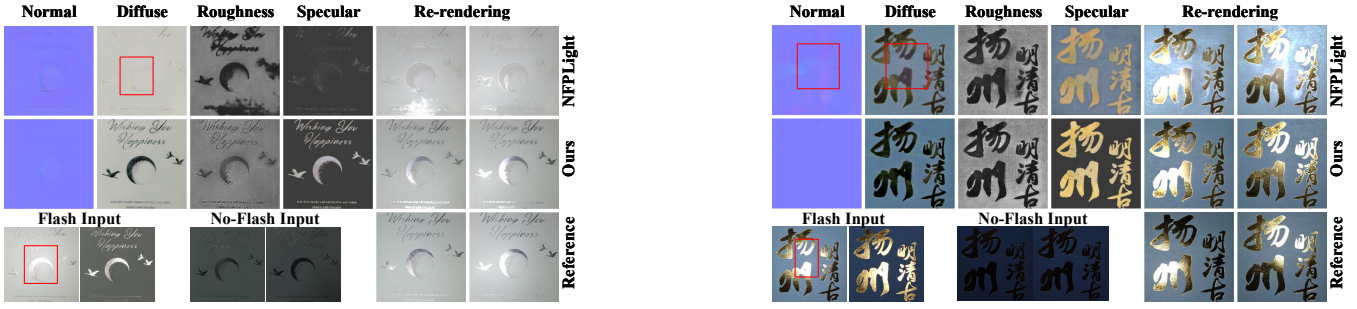

Fig. 10. Real Comparison on Near-Far-Field Capture Strategy. We compare our results against NFPLight of Wang et al. [2024].

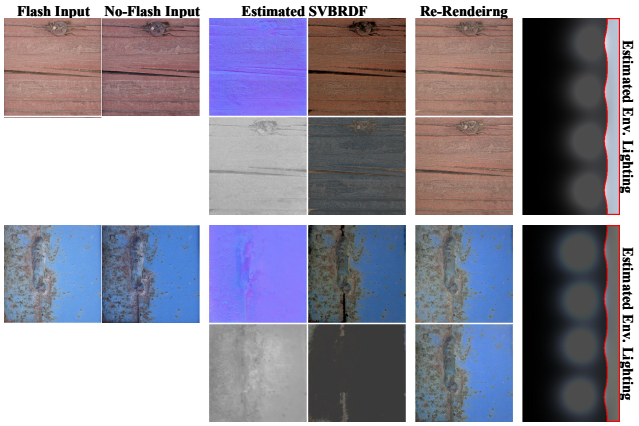

Fig. 11. Near-Field Real Material Estimation Results on Outdoor Environment. The re-rendering are performed under our estimated environment lighting and two novel point lighting.

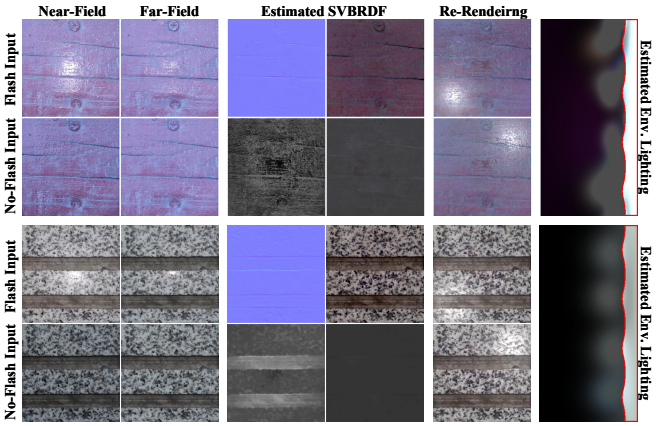

Fig. 12. Near-Far-Field Real Material Estimation Results on Outdoor Environment. The re-rendering are performed under our estimated environment lighting and two novel point lighting.

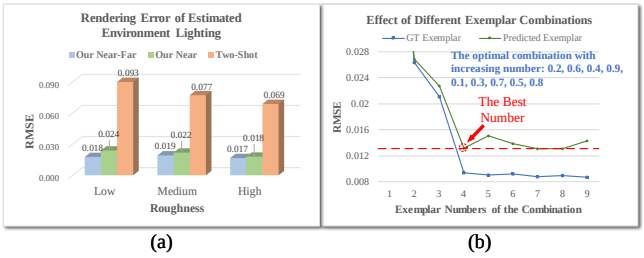

Fig. 13. (a): The blue and green bar represents the rendering RMSE of the environment lighting extracted by our near-far-field and near-field strategies, respectively, while the orange bar represents that of Two-Shot. (b): We calculated the accuracy of the optimal exemplar combination for environment lighting prediction with varying numbers of exemplars. The blue line shows results using GT SG to render exemplars, while the green line shows results from network-predicted exemplars.

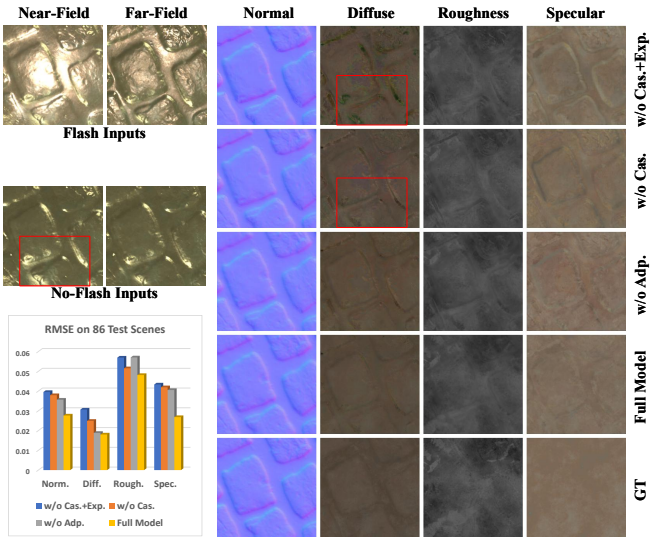

Fig. 14. Ablation Study on SVBRDF Estimation. Numerical evaluation results are shown at the left bottom. The RMSE metrics are computed on 86 synthetic scenes, and the lower values indicate better performance. The models are denoted as: (*Full Model*) the full model with the cascaded network and the adaptive exemplars; (*w/o Cas.*) without the cascaded network; (*w/o Adp.*) without the adaptive exemplars; and (*w/o Cas.+Exp.*) without both.

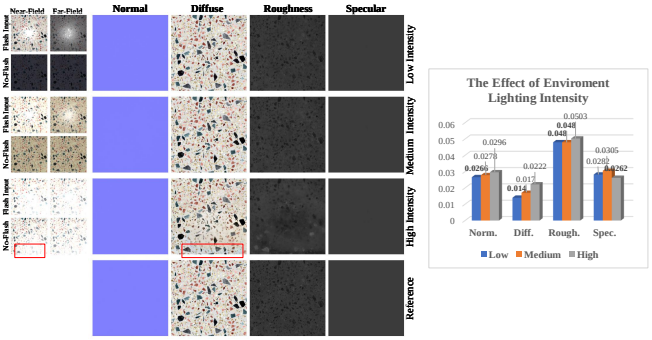

Fig. 15. The Effect of Environment Lighting Intensity. Numerical evaluation results are shown at the right. The RMSE metrics are computed on 86 synthetic scenes, and the lower values indicate better performance.

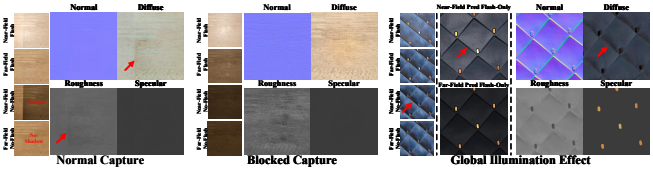

Fig. 16. Failure Case. This sample is illuminated by nearly top lighting. The mobile phone's movement during capture causes dynamic shadows between near-field and far-field images. These dynamic shadows lead to noticeable artifacts, as highlighted in the red rows of the diffuse and roughness maps (Normal Capture). Currently, we mitigate this problem by pre-blocking the dominant lighting, as shown in Blocked Capture.