Sparse SVBRDF Acquisition via Importance-Aware Illumination Multiplexing

LIANGHAO ZHANG, Tianjin University, China
ZIXUAN WANG, Tianjin University, China
LI WANG, Tianjin University, China
FANGZHOU GAO, Tianjin University, China
RUYA SUN, Tianjin University, China
JIAWAN ZHANG*, Tianjin University, China

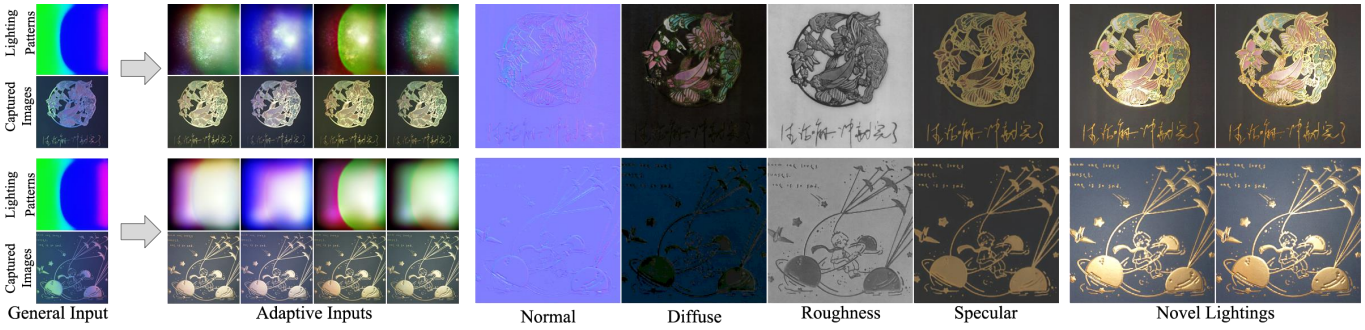

Fig. 1. Given an image captured under a general pattern, our method predicts adaptive lighting patterns to capture the salient part of reflectance responses. As illustrated, our predicted patterns not only consider previous sampling of the general pattern, but also adapt to material behaviors to reduce redundant sampling. We show material maps of two real materials of our method and re-renderings under novel environment and point lighting.

Reflectance acquisition from sparse images has been a long-standing problem in computer graphics. Previous works have addressed this by introducing either material-related priors or illumination multiplexing with a general sampling strategy. However, fixed lighting patterns in multiplexing can lead to redundant sampling and entangled observations, making it necessary to adaptively capture salient reflectance responses in each shot based on material behavior. In this paper, we propose combining adaptive sampling with illumination multiplexing for SVBRDF reconstruction from sparse images lit by a planar light source. Central to our method is the modeling of a sampling importance distribution on lighting surface, guided by the statistical nature of microfacet theory. Based on this sampling structure, our framework jointly trains networks to learn an adaptive sampling strategy in the lighting domain, and furthermore, approximately separates pure specular-related information from observations to reduce ambiguities in reconstruction. We validate our approach through experiments and comparisons with previous works on both synthetic and real materials.

CCS Concepts: • **Computing methodologies** → **Reflectance modeling**; Computer vision.

Additional Key Words and Phrases: Appearance Capture, SVBRDF, Lighting pattern

ACM Reference Format:
Lianghao Zhang, Zixuan Wang, Li Wang, Fangzhou Gao, Ruya Sun, and Jiawan Zhang. 2025. Sparse SVBRDF Acquisition via Importance-Aware Illumination Multiplexing. *ACM Trans. Graph.* 44, 6, Article 264 (December 2025), 13 pages. https://doi.org/10.1145/3763324

1 Introduction

Accurately reconstructing SVBRDF (Spatially Varying Bidirectional Reflectance Distribution Function) from sparse images has attracted growing attention, as it is crucial for photorealistic rendering. The inherently high dimensionality of SVBRDFs poses a significant challenge, due to the limited information available in sparse observations and the inherent ambiguity in reflectance decomposition.

For convenient acquisition, some methods use flash-lit images to capture material in a point-wise sampling manner. These approaches either introduce material-related priors to build data correlations [Deschaintre et al. 2018; Gao et al. 2019; Guo et al. 2021; Luo et al. 2024], or employ near-far capture setups to better activate specular reflections [Wang et al. 2024]. However, the specular signals captured under co-located point light sources remain sparse, which limits reconstruction performance.

*Corresponding author.

Authors' Contact Information: Lianghao Zhang, Tianjin University, China, lianghaozhang@tju.edu.cn; Zixuan Wang, Tianjin University, China, zixuan_wang@tju.edu.cn; Li Wang, Tianjin University, China, li_wang@tju.edu.cn; Fangzhou Gao, Tianjin University, China, gaofangzhou@tju.edu.cn; Ruya Sun, Tianjin University, China, 2023244177@tju.edu.cn; Jiawan Zhang, Tianjin University, China, jwzhang@tju.edu.cn.

ACM 1557-7368/2025/12-ART264
https://doi.org/10.1145/3763324

Our method instead follows another line of work [Kang et al. 2018; Ma et al. 2021; Zhang et al. 2023], which uses extended light sources and multiplexed lighting to efficiently sample specular reflections. While effective, these methods adopt fixed lighting patterns for all materials' sampling, ignoring material-specific reflectance behavior. Prior works on adaptive point sampling [Dupuy and Jakob 2018; Filip et al. 2013; Fuchs et al. 2007] have demonstrated that generalized sampling is less efficient than adaptive strategies, particularly under sparse observations. This inefficiency becomes more pronounced under multiplexed lighting, where each observation aggregates multiple specular lobes and diffuse components. Fixed patterns may lead to redundant sampling, especially for materials with various roughness scales and normal variations, thereby increasing ambiguities in SVBRDF decomposition. These limitations underscore the importance of adaptive, material-aware lighting strategies that enable informative observations and improve reconstruction accuracy.

In this paper, we propose a novel planar lighting-based method for SVBRDF reconstruction from sparse images lit by learned, adaptive lighting patterns. Our key insight to support adaptive multiplexing is the construction of a sampling importance distribution on the lighting surface leveraging the statistical nature of microfacet theory. This distribution identifies lighting surface regions that are globally important for capturing salient specular responses across all surface points of an SVBRDF.

Technically, we build the distribution by projecting the normal distribution function (NDF) onto the planar light source polygon, and then aggregating all BRDFs to model the overall importance distribution. While this provides a theoretically sound base for sampling structure, adaptive lighting patterns for different images should be sufficiently distinct for high-quality material decomposition. Building on this structure, we learn the adaptive color variations of lighting patterns through joint training of a lighting prediction network and an SVBRDF reconstruction network. However, we observed that joint training can lead to the over-reliance of reconstruction on spatial correlations, which hinders the lighting network from discovering distinct and informative sampling patterns. To address this, we propose a two-stage training strategy: the first stage employs a per-pixel reconstruction network to encourage representative sampling independent of spatial context, while the second stage reintroduces spatial correlations for sampling refinement and high-quality SVBRDF reconstruction. Moreover, to further reduce reconstruction ambiguity under multiplexed lighting, we also derive an approximate separation of specular-related appearance components by leveraging known lighting patterns.

We demonstrate our method on a simple and practical capture system, consisting of a fixed camera-LCD pair. This could be potentially implemented with consumer devices such as tablets or smartphones. Comprehensive experiments on synthetic and real photographs validate that our adaptive illumination multiplexing can produce superior results than previous methods from sparse images. In summary, our main contributions can be concluded as:

- a sampling importance distribution based on microfacet theory, enabling adaptive illumination multiplexing that accounts for material behaviors in SVBRDF acquisition;
- a two-stage framework that leverages this distribution to jointly predict adaptive lighting patterns and accurately reconstruct SVBRDFs;

2 Related Works

We provide a brief overview of learning-based methods on SVBRDF estimation from sparse images. Our approach is also related to techniques for optimal sampling in appearance acquisition, as lighting pattern is a form of sampling within the lighting domain.

2.1 SVBRDF Reconstruction from sparse images

Recently, the reconstruction of SVBRDF from sparse images has gained significant attention from researchers [Deschaintre et al. 2018; Gao et al. 2019; Zhang et al. 2023]. Broadly speaking, these methods can be categorized into two classes, as detailed below.

Introducing material-related priors. To address ambiguities caused by insufficient measurements, researchers have developed various priors grounded in material behavior. For a single input image, Li et al. [2017] pioneered the usage of deep learning-based data priors for partial SVBRDF reconstruction. Deschaintre et al. [2018] further created a large synthetic dataset to enable full SVBRDF maps reconstruction. Subsequent works have explored category-based priors [Li et al. 2018], highlight cues [Guo et al. 2021], texture-like-based priors [Aittala et al. 2016; Zhou et al. 2023], and meta-learning-based priors [Fischer and Ritschel 2022; Zhou and Kalantari 2022]. Besides, Zhou et al. [2021] and Giuseppe et al. [2021] introduced priors based on real data distribution from photographs, while Zhao et al. [2020] and Henzler et al.[2021] tackled unsupervised SVBRDF reconstruction assuming stationary materials. Recently, some methods introduced diffusion-based priors [Sartor and Peers 2023] or explicit modeling, such as basis assumption [Wang et al. 2023] and intermediate targets [Nie et al. 2024], to facilitate reconstruction. Despite these efforts, the appearance information in a single input image often remains insufficient for precise SVBRDF reconstruction.

Given sparse input images, Li et al. [2019] estimated SVBRDF from two images captured under planar lighting. Yet, they limited materials to piecewise homogeneous BRDFs for the application of step-edge lighting [Wang et al. 2011]. Since direct optimization from sparse captures can easily fall into local minimums, some methods utilized learning-based priors to reconstruct from sparse images. Deschaintre et al. [2019] proposed a max-pooling-based network, while Gao et al. [2019] and Guo et al. [2020] embedded a latent space with learned material priors into the inverse rendering pipeline. Luo et al. [2024] introduced a graph convolutional network to model correlation among images. However, accurately reconstructing SVBRDF from sparse flash photos remains challenging.

Improving capture efficiency. Another line of research focuses on efficiently capturing representative material measurements. A popular way to improve efficiency is using multiplexing illumination. Kang et al. [2018] developed a method to automatically learn lighting patterns for pixel-independent SVBRDF reconstruction, which was further extended to 3D objects [Kang et al. 2021, 2019; Ma et al. 2021] and a high-quality, real SVBRDF dataset [Ma et al. 2023a]. These methods leveraged full-domain illumination or video inputs to

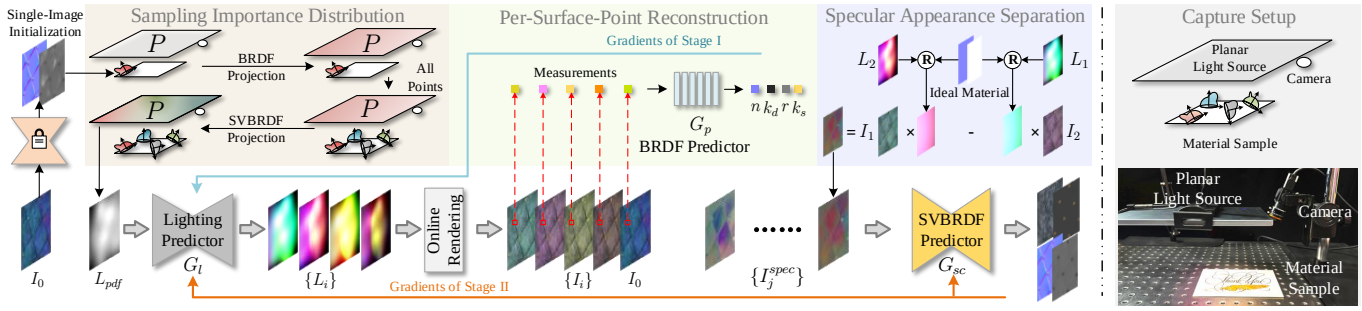

Fig. 2. Overview of our framework (left) and capture setup (right). The framework consists of three auxiliary modules (top left) and a main pipeline (bottom left). Specifically, given the first captured image, we initialize material behaviors using existing single-image methods. Then, a sampling importance distribution within the lighting rectangle is constructed based on NDFs. Then, adaptive lighting patterns are predicted by network G_l for other input images capturing or rendering. During training, the lighting predictor is jointly trained with a BRDF predictor G_p in the first stage and then refined by the joint training with the SVBRDF predictor G_{sc} in the second stage. We also explicitly separate specular-related appearance to aid reconstruction. Once trained, the SVBRDF maps can be obtained using lighting predictor for capture and the SVBRDF predictor for reconstruction.

regress lumitexels [Lensch et al. 2003a] from observations captured under learned patterns, achieving impressive results with several dozen captures. However, full-domain illumination is challenging to implement with off-the-shelf devices, and reconstruction quality degrades significantly with very sparse input shots. Instead, Zhang et al. [2023] used multiplexed lighting and spatial priors for single-image reconstruction, but struggling with high-frequency details due to ambiguities in the single input. Recently, Wang et al. [2024] proposed a near-far flash image pair for more specular responses, yet still suffers from low sampling-efficiency. Our specular information separation from observations is inspired by their relation map, while designed for illumination multiplexing with a fixed view.

Unlike all these approaches that learn a series of fixed lighting patterns or sampling strategies for all materials, our method adaptively learns patterns to reduce redundant sampling and focus on salient specular reflectance responses.

2.2 Optimal sampling for appearance acquisition

Optimal sampling has been explored to reduce acquisition efforts for various appearance representations (BRDF [Nielsen et al. 2015], SVBRDF [Lensch et al. 2003b], reflectance fields [Fuchs et al. 2007]). We discuss these methods by their adaptability to different materials.

General Sampling Strategy. These methods devise a universal sampling strategy for diverse materials. By expressing BRDFs as linear combinations of basis BRDFs, Nielsen et al. [2015] proposed a direct sampling pattern using Principal Component Analysis on a public dataset. Liu et al. [2023] applied meta-learning to automatically learn a direct sampling strategy. Various lighting patterns, including wavelet noise [Peers and Dutré 2005], Gaussian [Aittala et al. 2013] and spherical harmonics [Tunwattanapong et al. 2013], have been explored for illumination multiplexing during acquisition. Learned patterns have also been used for display photometric stereo [Choi et al. 2024]. Through introducing polarization, material properties can be effectively estimated using gradient patterns [Ghosh et al. 2010; Ma et al. 2007; Nogue et al. 2022; Riviere et al. 2016]. While these advancements are valuable, adaptive sampling strategies tailored to specific materials are often more efficient.

Adaptive sampling strategy. Several methods adapt to specific material behaviors and design sampling structures to improve capture efficiency. These structures can avoid oversampling of low-frequency diffuse components and increase the likelihood of capturing critical specular highlights. Lensch et al. [2003b] designed a function that reduces material parameter uncertainty to guide acquisition planning. Filip et al. [2013] proposed an adaptive slice model for anisotropic materials, which Vavra et al. [2018] later extended to accommodate additional slice types. More recently, Dupuy et al. [2018] utilized inverse mapping in importance sampling to adaptively parameterize BRDFs, which applies to acquisition, storage, and rendering. This approach inspired our use of the normal distribution function to represent sampling importance. However, these methods focus on direct sampling of materials.

Our adaptive acquisition strategy, by contrast, is designed for multiplexing illumination in sparse SVBRDF acquisition. Unlike point sampling where each observation corresponds to a single lighting direction, multiplexed lighting aggregates reflectance from many directions into one observation, introducing greater ambiguity. Hence, adaptive illumination multiplexing must jointly consider both the sampling structure and the distinctiveness of lighting patterns, in order to generate informative observations that support accurate SVBRDF reconstruction. Contrastingly, we address adaptive multiplexing and sparse reconstruction by constructing a physically-grounded sampling structure and designing a novel pipeline to jointly learn adaptive patterns and reconstruction.

Recently, some methods [Wiersma et al. 2025; Zhou et al. 2024] analyzed parameter uncertainty given multiple viewing/lighting conditions to guide acquisition. Although theoretically solid, sparse-image material reconstruction is still challenging for these methods.

3 Deep Adaptive Multiplexing

3.1 Problem Statement and Method Overview

Problem Statement. Before introducing our method, we first formulate the reconstruction problem. We assume that material appearance follows the Cook-Torrance reflectance model [Cook and Torrance 1981] with the GGX normal distribution [Walter et al. 2007].

An SVBRDF s is characterized by four parameter maps: normal n, diffuse albedo k_d, roughness r, and specular albedo k_s. Our method reconstructs these maps from m images ($\{I_i\}, i = 0, 1, ..., m-1$) captured under different patterns including a fixed pattern L_0 and adaptive patterns $\{L_i\}, i = 1, ..., m-1$. Each image can be formulated as follows:

$$I_i = \int_P f_r(\omega_i, \omega_o, s(x)) L_i(\omega_i, x)(n(x) \cdot \omega_i) d\omega_i, \tag{1}$$

where $f_r(\cdot)$ is the BRDF function, x is a surface point on the material, and ω_i, ω_o are incident and outgoing directions, respectively. Given all captured images, the SVBRDF maps are then reconstructed using a learned mapping function.

Method Overview. The primary goal of our method is to integrate adaptive sampling with illumination multiplexing, thereby minimizing redundant measurements and maximizing capture efficiency in the lighting domain. We address this by proposing a novel framework (Fig. 2), consisting of a main pipeline and three key auxiliary modules. The main pipeline (the lower part of Fig. 2) follows a straightforward idea: adaptive patterns $\{L_i\}$ are generated from an initial appearance I_0 by a lighting prediction network G_l. After capturing all input images ($I_0, \{I_i\}$) with a fixed LCD–camera pair (right of Fig. 2), the final SVBRDF maps are predicted by the reconstruction network G_{sc}. The two networks are jointly trained for optimal pattern prediction and reconstruction. However, it is challenging to directly learn the translation from the initial appearance domain to the lighting sampling domain and to reconstruct material from sparse appearance images.

To address this problem, our three key auxiliary modules (the upper part of Fig. 2) provide theoretical constraints and guidance to the main pipeline in both pattern design and material decomposition stages. The first two modules, sampling importance distribution building and per-surface-point reconstruction, support the design of adaptive patterns. The former constructs a physically grounded sampling structure, called the sampling importance distribution L_{pdf} (Sec. 3.2), to focus lighting sampling on salient reflectance responses. This distribution constrains G_l in pattern prediction by modeling the importance of each light source area in activating salient reflectance responses. The latter focuses learning on discovering optimal sampling strategies and capturing discriminative observations, leading to a two-stage training strategy for G_l (Sec. 3.3). For material decomposition, the specular appearance separation module (Sec. 3.4) provides specular cues by explicitly modeling relations among sparse images. It performs a plausible separation of specular appearance between each image pair and further produces specular-related maps $\{I_k^{spec}\}$ to aid *Gsc* in reconstruction.

3.2 Sampling Importance Distribution

In adaptive pattern prediction, learning a direct translation from the image domain to the lighting sampling domain is challenging due to the highly non-linear relationship between material properties and the full lighting domain. To address this, we leverage the statistical nature of microfacet theory to construct a physically grounded sampling-importance distribution of planar lighting, which bridges the initial appearance I_0 and the sampling structure. This distribution models the contribution of each lighting position x_l to activating salient reflectance responses, and serves as the foundation of our adaptive sampling algorithm. Here we detail the calculation of the sampling importance distribution L_{pdf} for an SVBRDF, which forms the basis of our adaptive sampling algorithm.

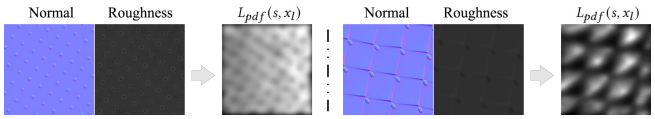

Fig. 3. Two examples of calculated sampling importance distribution L_{pdf}. As designed, Normal influences the position of high-importance areas while roughness influences the importance scale. Thus, the lower the roughness level, the more concentrated the importance distributions will be.

In the microfacet-based model, the statistical nature of facet normal is characterized by the NDF D, describing the probability distribution of normal orientations. This distribution can be viewed as an importance distribution of salient specular responses, which are high-frequency specular peaks that need more sampling, following the concept from Dupuy et al. [2018] on BRDF adaptive parameterization. In comparison, we utilize this distribution to model the sampling importance of an SVBRDF illuminated by a planar light source.

BRDF L'_{pdf} Projected on the Lighting Polygon. Under a fixed view, the NDF of a BRDF with roughness r is a 2D spherical distribution $D(r, \omega_h)$ where $\omega_h = \frac{\omega_i + \omega_o}{\|\omega_i + \omega_o\|}$. Since NDF is normalized over the hemisphere: $\int_\Omega D(r, \omega_h)(\omega_h \cdot n) d\omega_h$, we can derive the importance distribution over the lighting domain by replacing $d\omega_h$ with $d\omega_i$, and applying the Jacobian $\|J_i(\omega_i)\| = \|\frac{\partial \omega_h}{\partial \omega_i}\|$. Besides, we avoid the custom construction of hemisphere lighting by using a planar light source, which is easier to implement. Hence, the importance distribution should be projected to the lighting rectangle P. By replacing $d\omega_i$ with the differential area dA on P [Drobot 2018], the projected distribution L'_{pdf} of a BRDF can be formalized as follows:

$$L'_{pdf}(n, r, x, x_l) = \frac{D(r, \omega_h)(\omega_h \cdot n)(-n_l \cdot \omega_i)}{4(\omega_h \cdot \omega_o)\|x - x_l\|^2}, \tag{2}$$

where n_l is the normal of the lighting polygon, x is the position of a surface point. This distribution tells us which part of the planar light source is important to activate the salient specular responses, while which part only samples diffuse responses.

SVBRDF L_{pdf} calculation. The sampling importance distribution of an SVBRDF should represent globally optimal sampling direction of all surface points. Since each surface point of an SVBRDF is a BRDF, we can project *NDF*s of all surface points to the light source using Eq. 2. To maintain the statistical nature within the polygon, we normalize each projected distribution over the polygon P:

$$\widetilde{L'_{pdf}} = \frac{L'_{pdf}}{\int_P D(\omega_h)(\omega_h \cdot n) d\omega_h}, \tag{3}$$

where we omit the input parameters for simplicity. Then, the sampling importance distribution for an SVBRDF L_{pdf} is calculated by simply summing all these normalized distributions: $L_{pdf}(s, x_l) = \sum_x \widetilde{L'_{pdf}}(s(x), x_l)$, where $s(x)$ are material properties of this BRDF.

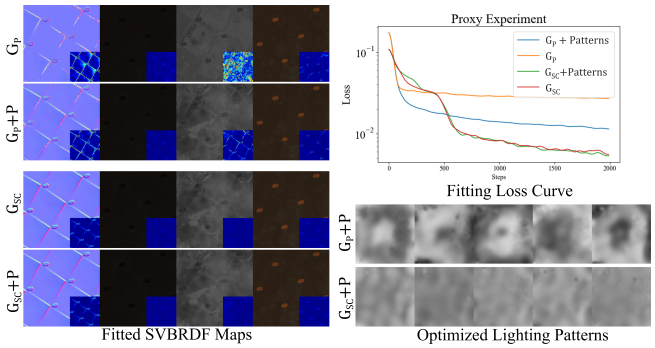

Fig. 4. The left part shows fitted SVBRDF maps of the proxy experiments alongside error maps comparing fitted and ground-truth maps. On the right, we show loss curves and optimized lighting patterns. These results indicate that prioritizing spatial correlation hampers the efficiency of lighting sampling strategy search during joint training of G_P and G_{sc}. "+P" means networks with pattern optimization.

Finally, we compress the dynamic range of the value to $[0, 1]$ using min-max normalization for numerical stability. The scale information of the dynamic range is lost by this normalization, and thus, we input the first captured image to G_l to complement lost information. In summary, this function represents the adaptive importance of an SVBRDF over the lighting domain, with planar lighting as the capture setup. We show two calculated L_{pdf} using the normal and roughness maps in Fig. 3.

Discussion of Bootstrapping. In acquisition, the material behavior is unknown before capture. In order to instantiate the sampling importance distribution, we need to establish an initial NDF by capturing the material with a fixed lighting pattern. Specifically, we use the learned pattern of Zhang et al. [2023] to capture the first image and estimate the normal and roughness maps. We argue that errors in initialization have a limited effect on reconstruction, since they are not used directly in reconstruction and the details of predicted patterns will be smoothed by materials. We also analyze the influence of initialization on reconstruction in ablation studies.

3.3 Two-Stage Sampling Strategy Searching

Based on the first captured image and the importance distribution, a straightforward method is to train the G_l and G_{sc} jointly —— G_l for sampling strategy searching and G_{sc} for spatial correlation priors learning and SVBRDF reconstruction. However, we observed that, during joint training, G_l tends to neglect optimal sampling strategy to capture representative measurements efficiently. This occurs because networks find it easier to predict material maps using spatial correlations rather than utilizing the unique characteristics of individual measurements of a single surface point. As a result, the gradients from G_{sc} to G_l for sampling learning are diminished.

To address this issue, we propose a two-stage searching strategy of G_l to find a more effective sampling approach. In the first stage, we introduce another network G_p that independently reconstructs each surface point of the SVBRDF. We then perform joint training on G_l and G_p. This setup eliminates spatial correlation during joint training, as G_p relies solely on the distinctiveness of the input measurements. By this, the first stage will force G_l to capture representative measurements for better demultiplexing and reconstruction instead of relying on spatial correlation heavily and then falling into a local minimum. After first stage training, we train the G_l and G_{sc} and refine the sampling strategy of G_l to re-introduce the spatial relation prior in the second stage.

To demonstrate this approach, we designed a simple proxy experiment that directly optimizes two groups of lighting patterns by training with G_{sc} and G_p respectively. This training is driven by simple L_1 loss between reconstructed and GT maps on a single SVBRDF. The optimized patterns equal to the accumulation of gradients propagated back from reconstruction networks G_{sc}/G_p to G_l. Additionally, we also optimize these two networks without pattern optimization to evaluate the reliance on spatial correlation in reconstruction. As shown in Fig. 4, when using fixed patterns, G_p has a low fitting accuracy compared to G_{sc} since the absence of spatial correlation. However, joint optimization with patterns improves the accuracy of G_p but not improve the accuracy of G_{sc}. This indicates the accuracy of G_{sc} depends largely on spatial correlation, while G_p needs to search sampling strategy for more representative measurements. The patterns optimized with G_p also confirm this, exhibiting greater variability compared to the other configuration.

3.4 Specular Appearance Separation

Despite efficient pattern design, directly recovering material from multiplexing-based observations is still ambiguous. We observe that, for planar objects, the irradiance is independent of material properties; under fixed capture geometry and known lighting patterns, it can be pre-computed for each image prior to material reconstruction. Based on this observation and the inherent simplicity of the diffuse component, we design a theoretical model to approximately separate specular-related appearance, aiding networks in albedo disentanglement.

Our model relies on that, in the rendering of Lambertian diffuse responses I_i^d, diffuse contribution is direction independent:

$$I_i^d = k_d \cdot L_i^d = k_d \int_P \frac{L_i(\omega_i, x) \cdot (n \cdot \omega_i)}{\pi} d\omega_i, \tag{4}$$

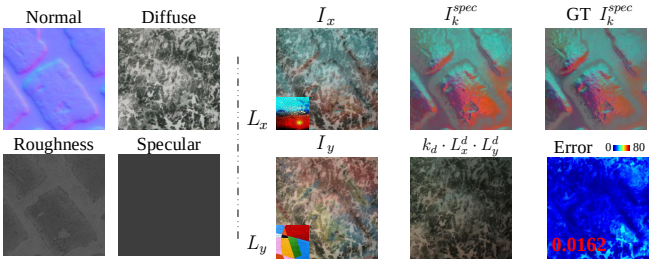

Fig. 5. Illustration of specular appearance separation. Given an SVBRDF, we render two images I_x, I_y under patterns L_x, L_y. The 4th-5th images in the first row show the derived specular appearance map I_k^{spec} and corresponding ground truth, which is rendered using variance surface normal. In the second row, $k_d * L_x^d * L_y^d$ is the scaled diffuse term for visual comparison, while in the last image, we also display the error map and RMSE metric.

where irradiance L_i^d is a scale to k_d and only calculated using the setup geometry and surface normals. Thus, given every two captured images I_x, I_y, we can assume a flat normal and scale up their diffuse terms to same level using known lighting patterns $I_x^d \cdot L_y^d = I_y^d \cdot L_x^d$. Then, we can easily subtract scaled images to eliminate diffuse albedo in appearance. The residual information is a pure specular-related appearance map I_k^{spec}:

$$I_k^{spec} = \int_P (L_x \cdot L_y^d - L_y \cdot L_x^d) f_r^s(\omega_i, \omega_o, s(x))(n \cdot \omega_i) d\omega_i, \quad (5)$$

where f_r^s is the specular term in reflectance function f_r. We ignore the input (ω_i, x) of patterns for simplicity.

As shown in figure, the textures in diffuse albedo are clearly removed and specular responses are left. In practice, we calculate I_k^{spec} given every input image, and thus $k = \frac{m*(m-1)}{2}$ in total, where m and k are the number of input images and appearance maps, respectively. Although assuming a flat normal, this explicit separation of specular information can significantly mitigate networks learning burden on disentanglement among albedos and lighting patterns.

3.5 Network Architecture and Loss Functions

As discussed, there are three networks in our framework: the lighting predictor G_l, the BRDF predictor G_p and the SVBRDF predictor G_{sc}. Our network structure is based on NAFNet [Chen et al. 2022] and G_{sc} extended it to two encoders and G_l is a shallower version, which is a well designed base structure for image-to-image problems. Two encoders in G_{sc} are used to extract features from input images $\{I_i\}$ and specular appearance maps $\{I_j^{spec}\}$, respectively, and the decoder takes concatenated features to reconstruct SVBRDF maps. The channel width of encoders is 32 in G_{sc}, while is 12 in G_l with extra output heads for each predicted pattern. For the G_p, we use the SIREN network [Sitzmann et al. 2020] which can represent signal implicitly. For online rendering during network training, we use LTC-based renderer proposed in [Zhang et al. 2023].

Overall, the training process of all networks can be expressed as:

$$\theta_l^*, \theta_{net}^* = \arg\min_{\theta_l, \theta_{net}} \mathcal{L}(s_{gt}, G_{net}(I_0, \{I_i\}, \{I_k^{spec}\})), \quad (6)$$

where $\mathcal{L}(\cdot,\cdot)$ is the loss function, s_{gt} denotes the ground-truth SVBRDF maps, θ_{net} are parameters of network G_{net}, and $\{I_i\}$ are rendered using predicted patterns with a differentiable rendering process: $\{I_i\} = \mathcal{R}(s_{gt}, \{L_i\})$. Adaptive patterns are predicted from I_0 and L_{pdf}: $\{L_i\} = G_l(I_0, L_{pdf})$. The G_{net} is BRDF predictor G_p in the first stage training and is SVBRDF predictor G_{sc} in the second stage. We ignore the network parameters in equations for simplicity. After training, real-world materials can be captured under the fixed pattern L_0 and adaptive patterns $\{L_i\}$. Once all the images $\{I_i\}$ are captured, the reconstruction results s can be obtained by simply running a feedforward pass through the reconstruction network: $s = G_{sc}(\{I_i\}, \theta_{sc}^*)$

During training, the constraints of G_l by the gradients from reconstruction networks are too weak to predict reasonable patterns. To regularize G_l, we propose two loss functions for G_l to understand the sampling importance distribution.

Importance consistency loss. To maintain the importance distribution in pattern prediction, the network should keep the structure of L_{pdf}. However, directly measuring the intensity difference between predicted patterns and the importance distribution can severely limit the search space, affecting multichannel variations within a single pattern and illumination complementarity across multiple patterns. Instead, we use Multi-Scale Structural Similarity Index Measure (MS-SSIM) [Wang et al. 2003] to measure the structure difference between the sum of all predicted patterns $\{I_i\}$ and L_{pdf}. By only keeping the contrast and structure terms at each scale of MS-SSIM, we allow for the learning of color variations to aid SVBRDF reconstruction. Besides, accounting for previous sampling is also vital in adaptive sampling methods, and hence, we add the first pattern to predicted patterns in metric calculation. The importance consistency loss $\mathcal{L}_c(\cdot,\cdot)$ can be summarized as following:

$$\mathcal{L}_c(\{L_i\}, L_{pdf}) = MSSSIM_{cs}(L_0 + \sum_n^i L_i, L_{pdf}). \quad (7)$$

Inverse importance loss. Additionally, we expect the network to understand the relation between the normal, roughness and the calculated L_{pdf}. To achieve this, we add another prediction head in G_l to estimate normal and roughness maps (n_{G_l}, r_{G_l}). The inverse importance loss $\mathcal{L}_{inv}$ measures L_1 distance between these maps and the corresponding ground-truth maps, where "inverse" indicates recovering the underlying roughness and normal from the constructed distribution.

Reconstruction loss. For reconstruction networks G_p and G_{sc}, we simply use the l_1 distance between predicted SVBRDF maps s and ground-truth maps s_{gt} as the reconstruction loss $\mathcal{L}_r$.

Consequently, the overall loss $\mathcal{L}$ for joint training is defined as:

$$\mathcal{L} = \lambda_c \mathcal{L}_c + \lambda_{inv} \mathcal{L}_{inv} + \lambda \mathcal{L}_r, \quad (8)$$

where $\lambda_c, \lambda_{inv}, \lambda_r$ are the weights of each loss. We set them to $1.0, 0.5, 1.0$ respectively in the first stage. The λ_{inv} is set to 0.05 in the second stage, allowing for refinement of G_l based on spatial correlation.

4 Capture Setup and Implementation Details

In this section, we will discuss the acquisition system prototype and calibrations, while the implementation details will also be discussed.

Setup and Calibrations. Theoretically, our sampling distribution can be applied to any setup geometry configurations, while we choose planar lighting for convenient construction. To implement our method, we build a simple setup consisting of a camera and an LCD, which are all mounted with gantries and communicated through a PC. We design the geometric relation in the setup roughly following Zhang et al. [2023], which proposed an ideal configuration to capture in mirror directions. Furthermore, we calibrated three aspects of the capture system, including the geometry, color, and non-linear radiometric function. Then, we apply calibration results in method training. Besides, the uneven emission distribution of LCD is challenging to measure and be applied in training. Instead, we approximate it with a simple pattern intensity decay, since the lack of emission calibration mainly affects the appearance of the

Table 1. Quantitative comparison with previous methods. The upper part shows DIR and MGAN with 5/20 inputs, and NFPLight and CorrAware with 2/4 flash inputs. The lower part presents LPL with a single input and its enhanced version LPL+, optimized with five lighting patterns.

Methods	Nrm.	Diff.	Rgh.	Spec.	Rend. RMSE	Rend. LPIPS
	Point lighting methods					
DIR_5	0.0305	0.0175	0.1340	0.0582	0.0627	0.0763
DIR_20	0.0225	0.0152	0.1219	0.0594	0.0533	0.0531
MGAN_5	0.0378	0.0254	0.1237	0.0550	0.0719	0.0956
MGAN_20	0.0301	0.0224	0.1007	0.0514	0.0559	0.0695
NFPLight	0.0279	0.0228	0.0388	0.0254	0.0375	0.0542
CorrAware	0.0497	0.0437	0.1119	0.1723	0.0905	0.1289
	Planar lighting methods					
Ours	**0.0193**	**0.0149**	**0.0303**	**0.0188**	**0.0318**	**0.0517**
LPL	0.0323	0.0188	0.0440	0.0226	0.0415	0.0854
LPL+	0.0293	0.0206	0.0337	0.0189	0.0418	0.0753
FreeForm+	0.0428	0.0909	0.2154	0.1558	0.1093	0.2624

material's right side. These surface points are illuminated by light emitted from the left portion of the source at oblique angles, whereas other regions are lit at smaller incident angles. Thanks to our explicit modeling of appearance in Sec. 3.4, we can apply this decay in rendering of L_i^d to correct reconstruction, which is experimentally chosen. Our calibration process is easy and needs no equipment movement, please refer to our supplementary materials for more detailed procedures.

Implementation Details. In network training and testing, we use the dataset proposed by Deschaintre et al. [2018; 2019] following previous works, containing $100k$ different materials for training and 122 for testing. Moreover, we use the Adam optimizer [Kingma and Ba 2014] with default hyper-parameters. G_l is trained for $200k$ iterations in the first stage and jointly trained with G_{sc} for $500k$ iterations. The initial learning rate of G_{sc} is set to $2e^{-4}$ and is gradually decreased using the cosine annealing schedule [Loshchilov and Hutter 2016]. Besides, we observed that the data distribution rendered by online rendering module has a gap to real images and this will cause accuracy degradation on real data. So we fine-tune the SVBRDF predictor G_{sc} for $50k$ iterations using the data rendered by trained G_l's patterns and Mitsuba3 [Jakob et al. 2022]. Note that following comparison results of our method on synthetic data are reconstructed with inputs rendered by Mitsuba3, which can demonstrate the superiority of our method on physically-based rendered data. The resolution we use in training and testing is 256×256.

5 Evaluation

In this section, we present detailed experiments. We first compare our method with prior works on both synthetic and real data (Sec. 5.1), and then conduct ablation studies on synthetic data to evaluate each component (Sec. 5.2). Training and evaluation use the datasets of Deschaintre et al. [2018; 2019], with five input images as discussed in the ablation studies.

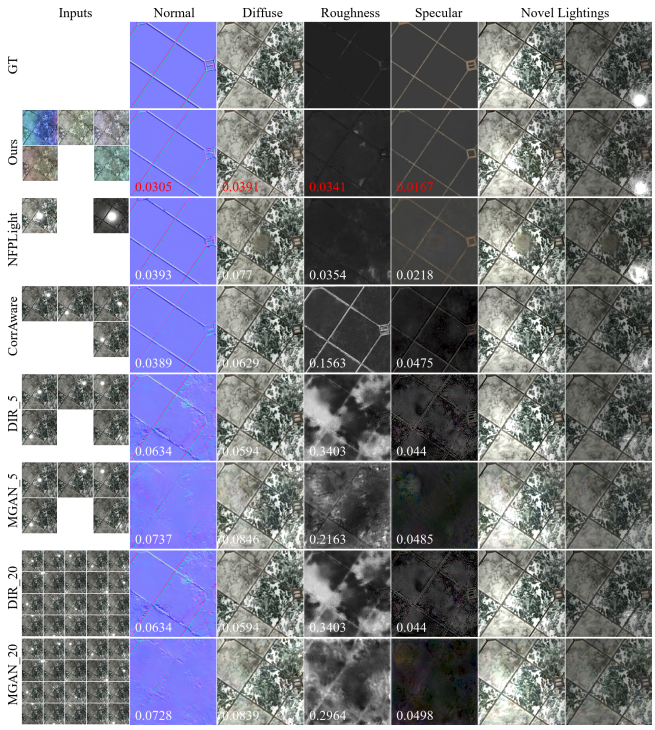

Fig. 6. Comparison results against point lighting methods on synthetic data. The input images of all methods are shown in 1st column and the reconstructed SVBRDF maps are shown in 2nd-5th columns. Besides, we also render material maps under three novel point lighting for visual comparison.

5.1 Comparison

We compare our method with previous approaches for SVBRDF reconstruction from sparse images, including point-lighting methods: DIR [Gao et al. 2019], MGAN [Guo et al. 2020], CorrAware [Luo et al. 2024], NFPLight [Wang et al. 2024], and planar-lighting methods: LPL [Zhang et al. 2023], FreeForm [Ma et al. 2021]. Note that we select FreeForm over other lumitexel-based methods (e.g., Kang et al. [Kang et al. 2018], OpenSVBRDF [Ma et al. 2023a], MatMoE [Ma et al. 2023b]) because these require full-domain illumination, which is essential for capturing specular lobes and pixel-independent reconstruction. While OpenSVBRDF achieves high accuracy, it designed a bottleneck for post-optimization and needed extra linear lighting-lit images, limiting its performance on sparse-shot reconstruction.

Unlike our method, both LPL and FreeForm optimize only one lighting pattern. To ensure a fair comparison, we retrained these methods to optimize five lighting patterns by adjusting the input channel of LPL and the linear projection layer of FreeForm. Besides, FreeForm is retrained based on our setup configuration. The retrained versions are labeled as LPL+ and FreeForm+, respectively. Note that LPL and LPL+ are also trained with calibration results. For visual comparison, all albedo maps in the result figures were gamma corrected, and all DIR and MGAN results were initialized with FSC [Deschaintre et al. 2019], which also reconstructs SVBRDFs from arbitrary flash photos.

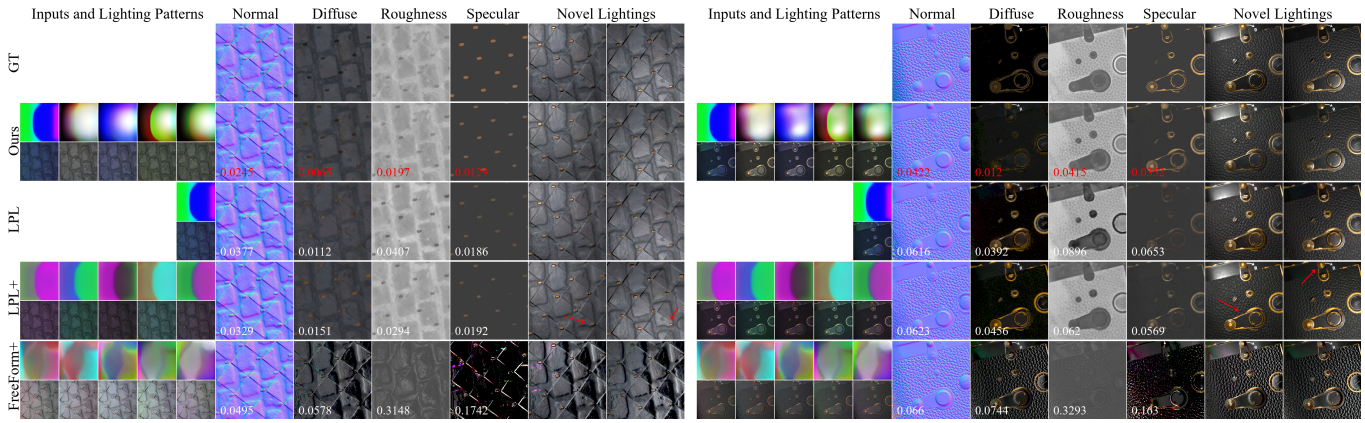

Fig. 7. Comparison results on two synthetic scenes against LPL, LPL+ and FreeForm+, which capture material under planar lighting. We show all lighting patterns and input images in the first column. Note that our patterns are adaptive and varying while the patterns of other methods are constant among different SVBRDFs. The right material is from MatFusion [Sartor and Peers 2023].

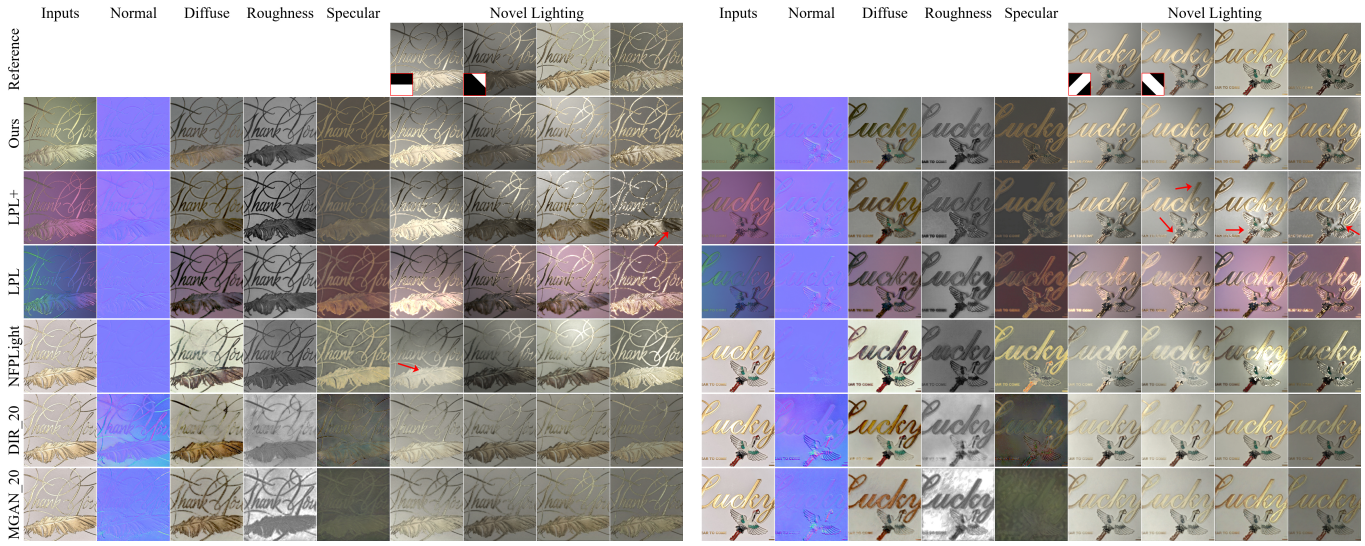

Fig. 8. Comparison on two real scenes: "Card", "Bird". We only show a single input image for each method, while others can be found in supplementary materials. In 6th-9th columns, we show captured reference images. In step edge lighting columns, we visualize the patterns for reference image captures in the down-left corner. All re-renderings under step-edge lighting are rendered with calibration results of our setup. We also mark some artifacts with red arrows.

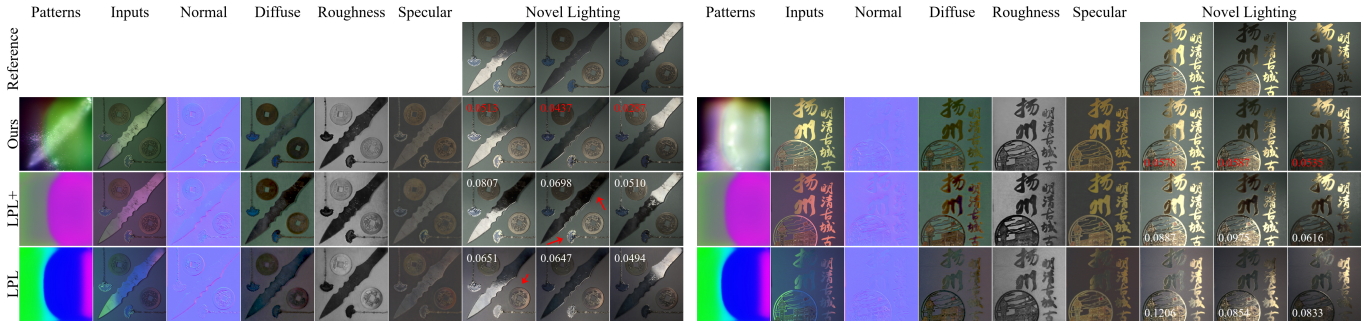

Fig. 9. Comparison against LPL and LPL+ on two real scenes. We show a pattern and corresponding input image of each method in 1st-2nd columns. All references and re-renderings are based on step-edge patterns. The errors are displayed in the down-right corner, while the lowest is marked with red.

Comparison on synthetic dataset. We provide both qualitative and quantitative comparisons with previous methods. Visual comparisons against point-lighting methods and planar-lighting methods are shown in Fig. 6 and Fig. 7, respectively. As illustrated in Fig. 6, point-lighting methods struggle with materials having narrow specular lobes, even with 20 flashlights, as they activate limited specular responses compared to our adaptive patterns. While NFPLight performs well with just two inputs, it assumes strict z-axis camera motion, requiring motorized equipment for ideal real captures. In Fig. 7, we demonstrate the effectiveness of our adaptive lighting patterns using two challenging scenes with large normal variations and albedo-roughness entanglement, which introduce severe ambiguities. Compared with general lighting patterns, our lighting patterns focus on sampling normal variations and low-roughness areas. Although using the same capture setup as LPL and FreeForm+, our method can decouple material parameters accurately while other methods misunderstand captured measurements.

For quantitative comparison, we evaluate the predicted maps using Root Mean Square Error (RMSE), and assess 20 re-rendered images using both RMSE and Learned Perceptual Image Patch Similarity (LPIPS). The re-rendered images are generated under 10 random point light sources and 10 step-edge lighting patterns, as detailed in the supplementary materials. The results in Tab. 1 show that our method outperforms previous methods on both material reconstruction and re-renderings. Flash images primarily capture diffuse responses, making roughness and specular reconstruction challenging for most point-lighting methods. Although LPL+ achieves similar accuracy on specular albedo, its fixed pattern causes redundant sampling and results in degradation on other material parameters. By contrast, although using the same setup with LPL+, our adaptive capture framework successfully balances all parameters and achieves the highest accuracy among all methods. For more comprehensive comparisons, we also provide comparisons on other datasets in supplementary materials [Ma et al. 2023a; Sartor and Peers 2023; Vecchio and Deschaintre 2024].

Comparison on real dataset. All real images of planar lighting methods are captured with our calibrated setup within a few seconds, while flash photos for point lighting methods are captured with a mobile phone (iPhone 15 Pro) and calibrated with markers. Although FreeForm achieves remarkable results in material scanning, sparse inputs are challenging for per-pixel reconstruction, so we compare only LPL and LPL+ on real data. For fairness, we fine-tuned the SVBRDF predictor of LPL and LPL+ for the same iterations using the dataset rendered with Mitsuba3. Due to different capture devices, we applied color correction to all flash images, normalizing the correction curves to avoid overexposure.

In Fig. 8, we show the reconstructed results of both point and planar lighting methods on two challenging real scenes. To validate the reconstruction quality, we also capture the material under novel point lighting and planar lighting as the reference images (step-edge patterns used in synthetic comparison). Because of errors in color correction and the built-in processing of smartphone when capturing flash images, we manually adjust the intensity of all re-renderings for better visual comparison. Although LPL and LPL+ can distinguish smooth and rough areas, their results have fewer details in all parameter maps and often reconstruct baked-in albedos or inferior roughness. Besides, they heavily suffer from uneven emission, such as global bias in albedos or normal, while our method can alleviate the influence thanks to the decay in explicit modeling. On the aspects of point lighting method, they often fail to distinguish large normal variations and reflectance responses, because of limited sampling efficiency of flash lighting. Moreover, their results tend to be blurred, which is caused by hand-held capture. In comparison, our method can capture sharp edges and details as a result of fixed-view capture, and we think this is more suitable for dark-room capture.

For a more comprehensive comparison, we also provide comparisons with only LPL and LPL+. Since all these methods used the same setup, we can calculate the RMSE between the re-renderings and the reference images. Note that there is no manual intensity adjustment in this comparison. As shown in Fig. 9, our method can reconstruct more accurate and detailed results, such as the sharp edges of pendants in the left scene and normal variations in the right scene. Despite the artifacts caused by uneven emission in LPL and LPL+, their models are struggling to disentangle parameters because of fixed sampling strategy. Contrastingly, our adaptive patterns can reduce redundant sampling and achieve better results.

5.2 Ablation Study

Here we analyze our method from four main aspects, consisting of key method components, the effectiveness of pattern design, the influence of material initialization and the color strategy choice of patterns. We also demonstrate our choice of the input number with

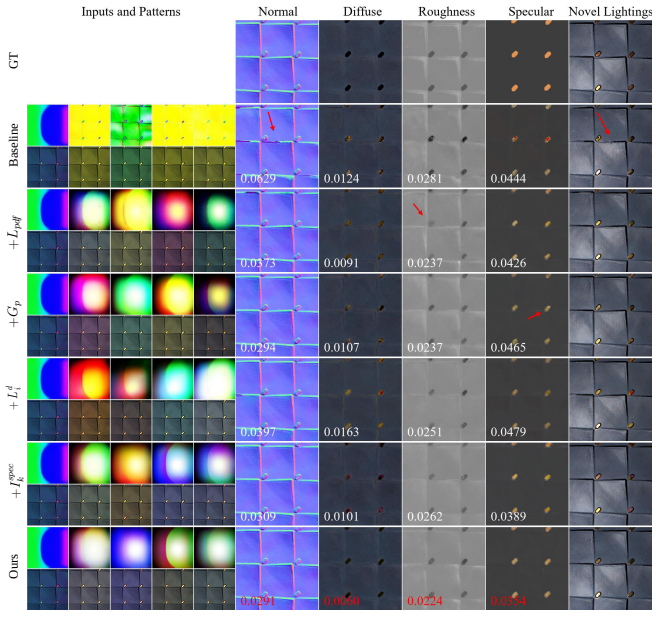

Fig. 10. Ablations on key components. We show input images and learned patterns in the first column, while other columns are SVBRDF results and re-renderings under a novel flash. Besides, we mark the artifacts with red arrows, and also calculate RMSE error on each parameter map and mark the lowest error with red texts. Note that pattern changes in the last two rows are auto learned given different inputs or networks.

Table 2. Quantitative results on key components and initialization. In the upper part, we display results of networks with gradually added key components. In the bottom part, we show results of our method tested with different material initialization. The metric is RMSE ↓.

Methods	Nrm.	Diff.	Rgh.	Spec.	Avg.
	Ablations on Key Components				
Baseline	0.0315	0.0195	0.0325	0.0221	0.0264
$+L_{pdf}$	0.0240	0.0161	0.0361	0.0210	0.0243
$+G_p$	0.0231	0.0150	0.0301	0.0205	0.0222
$+L_i^d$	0.0266	0.0171	0.0352	0.0231	0.0255
$+I_j^{spec}$	0.0224	0.0145	0.0312	0.0200	0.0220
Ours	0.0216	0.0120	0.0299	**0.0176**	0.0203
	Different Material Initialization				
GT Init.	**0.0214**	**0.0116**	**0.0291**	0.0177	**0.0200**
Diff. Init.	0.0226	0.0176	0.0475	0.0214	0.0273

experiments. Note that following experiments are conducted on synthetic data and networks are trained without calibration results.

Analysis on Key Components. Firstly, we analyze the different components of our method. The analysis begins with a baseline that directly predicts adaptive patterns from I_0 and estimates SVBRDF maps using a vanilla NAFNet on all input images. Building upon this baseline, we gradually introduce our technical components and compare reconstruction accuracy. Specifically, we first input our sampling importance distribution and add importance consistency loss to constrain pattern prediction ($+L_{pdf}$), and then incorporate the two-stage searching strategy using G_p to better learn pattern color variations ($+G_p$). Following this, we further conduct ablation studies on the decoupling aspect by evaluating three settings: feeding the approximated irradiance term L_i^d into the G_{sc} ($+L_i^d$), replacing it with specular appearance maps I_j^{spec} to verify their effectiveness over scale terms ($+I_j^{spec}$), and finally, using I_j^{spec} with two separate encoders for feature extraction (our full model).

The metrics in the upper part of Tab. 2 clearly demonstrate the effectiveness of our technical designs. While the baseline performs well in roughness prediction, it struggles to decouple other parameters and achieves the lowest overall accuracy. Introducing L_{pdf} reduces redundant sampling and alleviates ambiguities in other parameters, especially in normal estimation. The two-stage training with G_p further improves performance by refining the learning of color variations. Directly adding L_i^d to reconstruction degrades accuracy, as L_i^d lacks decoupled albedo information and distracts the reconstruction network. In comparison, explicitly separating specular information via I_j^{spec} helps resolve ambiguities in albedo decoupling. Finally, the full model uses separate encoders for feature extraction and achieves the best performance.

As shown in Fig. 10, the baseline produces lighting patterns that inefficiently sample the lighting domain and exhibit limited variation, leading to noticeable errors. Although the addition of L_{pdf} and two-stage training significantly improves normal prediction, albedo decoupling remains challenging. Our proposed specular appearance maps I_j^{spec} enhance specular prediction but introduce errors in other parameters. This is due to the confusion caused by the jointly input information (5 images I_i and 10 specular appearance maps I_j^{spec}, totaling 45 channels). As a result, separate feature extraction proves more effective.

Table 3. Quantitative results on different patterns. In the upper part, we display results network trained with other patterns. The bottom part shows the results of our Adaptive version. The metric is RMSE ↓

Methods	Nrm.	Diff.	Rgh.	Spec.	Avg.
ColorGrad	0.0252	0.0223	0.0432	0.0228	0.0284
MonoGrad	0.0256	0.0164	0.0361	0.0198	0.0245
PCA	0.0268	0.0179	0.0494	**0.0197**	0.0284
Picked	0.0251	0.0164	0.0527	0.0225	0.0292
Optimized	0.0351	0.0209	0.0365	0.0206	0.0283
Adaptive	**0.0231**	**0.0150**	**0.0301**	0.0205	**0.0222**

Different Pre-Defined Patterns. To demonstrate the effectiveness of adaptive pattern design, we compare our method with four fixed pattern groups for appearance capture: monochrome gradient patterns (MonoGrad), PCA-analyzed patterns (PCA), hand-picked patterns (Picked) and colored gradient patterns (ColorGrad). Specifically, MonoGrad is introduced by Ghosh et al.[2010], from which we select 5 representative patterns. PCA patterns are generated by applying PCA to our L_{pdf}, computed across SVBRDFs from the training set. The Picked group includes patterns from Zhang et al.[2023], Li et al.[2019], and three patterns chosen for their large color variations. ColorGrad refers to the colored gradient patterns proposed by Fyffe[2015]. Besides, we also compare with LPL+ (Optimized) in this experiment. To isolate the effect of pattern design, we compare our method (without I_j^{spec}, denoted as Adaptive) against a vanilla NAFNet trained with each fixed pattern group. Note that this Adaptive version of our method corresponds to the $+G_p$ setting in the previous ablation, where NAFNet is used for reconstruction.

In Tab.3, our adaptive patterns outperform others on most parameters, with a slight trade-off in specular albedo prediction. This is expected, as adaptive patterns reduce redundant sampling and focus more on parameter disentanglement. Interestingly, MonoGrad performs reasonably well and even surpasses Optimized patterns, as its gradient structure is grounded in physical principles. The Optimized baseline, by contrast, fails by seeking general patterns across all materials. Our method integrates optimization with physical priors, achieving the best performance. We also provide qualitative comparisons in Fig.11. In this test scene, the lighting patterns are convolved with materials exhibiting large normal variations and a range of roughness levels. Networks trained with fixed patterns struggle to capture material properties, especially for roughness in areas with large normal variation or in albedo disentanglement. In contrast, our adaptive lighting patterns emphasize the lower-right region and reduce redundancy, as normal variations make this area important in sampling. As a result, our method yields much cleaner material maps and more effectively decouples SVBRDF parameters compared to other configurations.

Influence of initial L_{pdf}. Here we analyze the influence of initial SVBRDF on patterns and reconstruction. Since previous metrics of

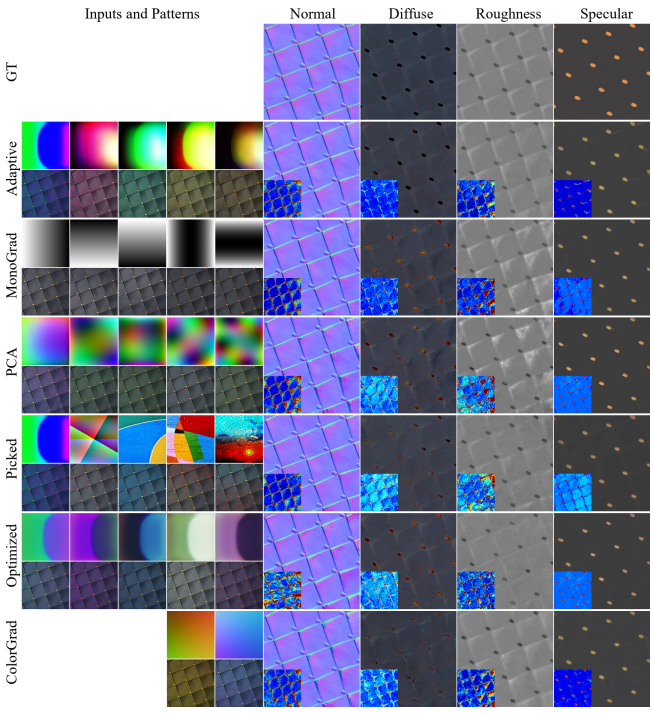

Fig. 11. Comparison of our method against the networks trained with different fixed lighting patterns. The first column shows lighting patterns and input images, while other columns are reconstructed results. We show the error maps between results and GT at the left-bottom corner.

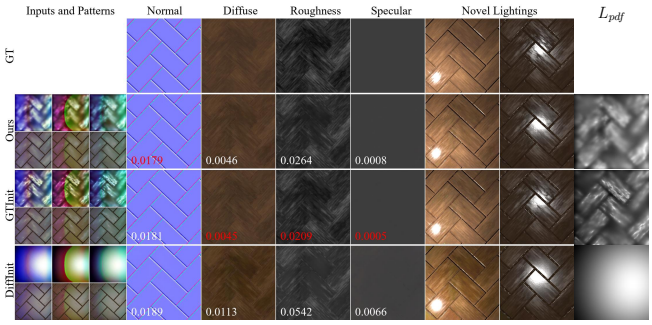

Fig. 12. Ablation on different L_{pdf} initialization. The first column shows predicted lighting patterns of difference L_{pdf} and corresponding rendered images, while follwoing four columns are reconstructed results. The last column shows the L_{pdf} for initialization. We show the RMSE errors at the left-bottom corner.

our method are reconstructed using LPL initialization, we instead use GT SVBRDFs (GT Init.) or a fixed rough material (Diff. Init.) as initialization to validate our method. The metrics (lower part in Tab. 2) demonstrate the robustness on LPL initialization and the effectiveness of adaptive sampling compared with fixed L_{pdf}. Furthermore, as shown in Fig. 12, although L_{pdf} constructed from LPL's results are different from which from GT, the overall sampling structure

Table 4. Quantitative comparison on different color strategy choices. The metric is RMSE ↓. We choose the Colored strategy choice in our experiments.

Methods	Nrm.	Diff.	Rgh.	Spec.	Avg.
Colored	**0.0317**	0.0188	**0.0346**	0.0209	**0.0265**
Gray	0.0362	**0.0186**	0.0368	**0.0194**	0.0278
Mono	0.0428	0.0251	0.0593	0.0247	0.0380
Mono_9	0.0344	0.0192	0.0443	0.0220	0.0300

is maintained and the difference is further filtered by pattern prediction. In comparison, the fixed L_{pdf} destroys the structure and activates relatively uniform appearance, resulting in degradation of results, especially on diffuse and roughness.

The color strategy choice of pattern. To validate the effectiveness of using colored lighting patterns for SVBRDF reconstruction, we jointly optimize three patterns and the reconstruction network under different color strategies: colored (Colored), monochrome (Mono), and grayscale (Gray). Since three colored or grayscale patterns correspond to nine channels, we also include a configuration with nine monochrome patterns (Mono_9) for a fair comparison. Quantitative and qualitative results are reported in Tab. 4 and Fig. 13, respectively. As expected, Mono performs the worst since single-channel patterns capture little information, while Mono_9 performs better due to more input images. In contrast, Gray and Colored patterns simultaneously capture all color channels. Colored sacrifices little accuracy in albedo prediction while providing cross-channel constraints for reconstructing channel-irrelevant parameters (e.g. normal, roughness), thus achieving the best performance.

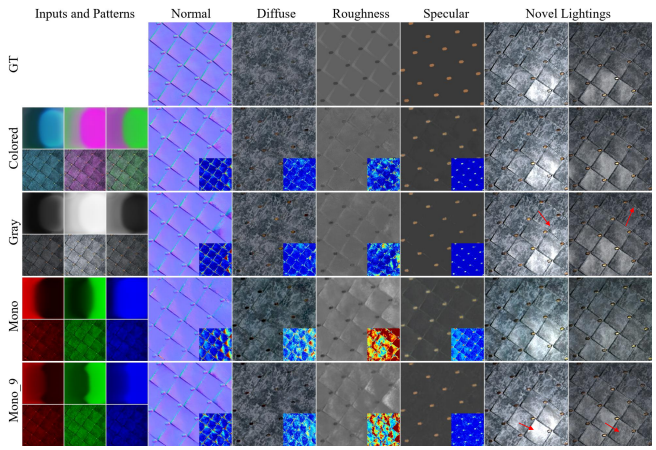

Fig. 13. Validation of the color strategy choice of patterns. The first column shows optimized lighting patterns with different color strategy configurations and their corresponding input images. For Mono_9, only three patterns are shown for simplicity. Artifacts are highlighted with red arrows or error maps in the bottom-right corner.

Number of Input Images. We trained our method with four settings of the input number: (2, 5, 7, 10), and visualized error curves

on SVBRDF reconstruction in Fig. 14. As shown, the errors reconstructed with two input images are highest, while five input images significantly improve accuracy. Notably, the error curves have fluctuations given more input images. This is because the input information is overwhelmed to the reconstruction network, where 45 specular appearance maps are generated given 10 input images. We argue that a dedicated network is needed for more inputs, which we think is beyond the scope of this paper. Thus, we choose 5 as the input number of our method.

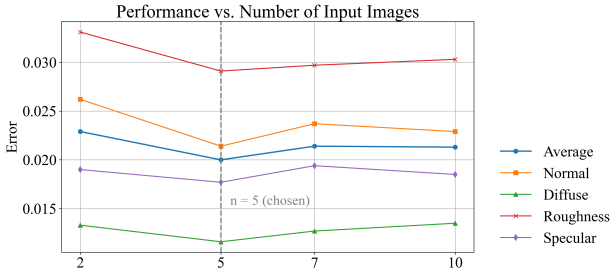

Fig. 14. Ablation on input numbers. We visualized the error curves for different material properties, using different colors. The x-axis labels represent the input numbers, and the dashed line indicates our current experiment setting.

6 Discussion and Limitations

Although our method produces high-quality results, it still suffers from several limitations. Firstly, we assume fixed input numbers, which have less flexibility. A more promising way is iteratively sampling according to importance with arbitrary inputs for high-quality reconstruction. Secondly, our prototype acquisition system can only capture small material samples because of the limited size of the planar light source, which is $9cm \times 9cm$ in our experiments. Although this constrains usage on larger samples, our approach is agnostic to the geometric configuration of the setup. Thus, the sampling importance distribution and lighting learn scheme can be extended to larger illumination sources like display-based setups. Besides, our method can only reconstruct isotropic materials currently. Designing adaptive illumination on complex appearance, such as anisotropic or translucent materials, is an interesting research direction. Finally, while uneven emission is hard to calibrate and model in training, the absence of this calibration can cause artifacts in reconstruction. As shown in Fig. 15, although we use a pattern decay in L_0^d rendering and avoid the artifacts in normal (slightly points to right) and roughness (relatively low), this will sometimes cause baked-in albedos.

7 Conclusion

In this paper, we propose a novel method that combines adaptive sampling with illumination multiplexing to maximize capture efficiency for sparse input images of a given SVBRDF. Given material behavior captured by a predefined lighting pattern, we model a sampling importance distribution of surface points on the planar light source for adaptive sampling. Based on this distribution, the sampling strategy of the lighting predictor is jointly trained with material predictors in a two-stream manner for a broader search space. We demonstrate that our adaptive multiplexing method captures salient reflectance responses efficiently and can reconstruct high-quality material maps with fine details from real captured photographs.

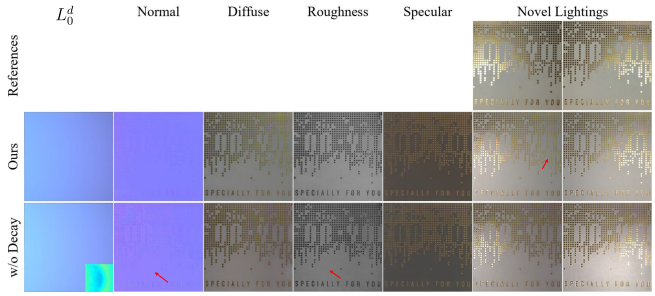

Fig. 15. Impact of uneven emission distribution on reconstruction. The first column shows L_0^d rendered with or without pattern decay, while the last two columns show references and re-renderings under novel point lighting. The normal in w/o Decay is slightly pointed to right and roughness is relatively low to compensate the absence of uneven distribution modeling. Our decay strategy can improve accuracy while introducing baked-in albedo.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant No. 62172295, which is gratefully acknowledged.

References

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

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

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. 2022. Simple Baselines for Image Restoration. In *Computer Vision – ECCV 2022*, Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (Eds.). Springer Nature Switzerland, Cham, 17–33.

Seokjun Choi, Seungwoo Yoon, Giljoo Nam, Seungyong Lee, and Seung-Hawn Baek. 2024. Differentiable Display Photometric Stereo. *CVPR* (2024).

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

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

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

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

Jonathan Dupuy and Wenzel Jakob. 2018. An adaptive parameterization for efficient material acquisition and rendering. *ACM Transactions on graphics (TOG)* 37, 6 (2018), 1–14.

Jiri Filip, Radomir Vavra, Michal Haindl, Pavel Zid, Mikulas Krupika, and Vlastimil Havran. 2013. BRDF slices: Accurate adaptive anisotropic appearance acquisition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1468–1473.

Michael Fischer and Tobias Ritschel. 2022. Metappearance: Meta-Learning for Visual Appearance Reproduction. *ACM Trans. Graph.* 41, 6, Article 245 (nov 2022), 13 pages. doi:10.1145/3550454.3555458

Martin Fuchs, Volker Blanz, Hendrik PA Lensch, and Hans-Peter Seidel. 2007. Adaptive sampling of reflectance fields. *ACM Transactions on Graphics (TOG)* 26, 2 (2007), 10–es.

Graham Fyffe and Paul Debevec. 2015. Single-shot reflectance measurement from polarized color gradient illumination. In *2015 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 1–10.

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

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

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

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

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

Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, and Delio Vicini. 2022. Dr. jit: A just-in-time compiler for differentiable rendering. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–19.

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

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

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

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

Hendrik PA Lensch, Jochen Lang, Asla M Sá, and Hans-Peter Seidel. 2003b. Planned sampling of spatially varying BRDFs. In *Computer graphics forum*, Vol. 22. Wiley Online Library, 473–482.

Hendrik P A Lensch, Jan Kautz, Michael Goesele, Wolfgang Heidrich, and Hans Peter Seidel. 2003a. Image-based reconstruction of spatial appearance and geometric detail. *ACM Transactions on Graphics (TOG)* (2003).

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

Xiao Li, Peiran Ren, Yue Dong, Gang Hua, Xin Tong, and Baining Guo. 2019. Capturing Piecewise SVBRDFs with Content Aware Lighting. In *Advances in Computer Graphics: 36th Computer Graphics International Conference, CGI 2019, Calgary, AB, Canada, June 17–20, 2019, Proceedings 36*. Springer, 371–379.

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

Chen Liu, Michael Fischer, and Tobias Ritschel. 2023. Learning to learn and sample brdfs. In *Computer Graphics Forum*, Vol. 42. Wiley Online Library, 201–211.

Ilya Loshchilov and Frank Hutter. 2016. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983* (2016).

Di Luo, Hanxiao Sun, Lei Ma, Jian Yang, and Beibei Wang. 2024. Correlation-aware Encoder-Decoder with Adapters for SVBRDF Acquisition. In *SIGGRAPH Asia 2024 Conference Papers* (Tokyo, Japan) *(SA '24)*. Association for Computing Machinery, New York, NY, USA, Article 131, 10 pages.

Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, Paul E Debevec, et al. 2007. Rapid Acquisition of Specular and Diffuse Normal Maps from Polarized Spherical Gradient Illumination. *Rendering Techniques* 9, 10 (2007), 2.

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

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

Xiaohe Ma, Yaxin Yu, Hongzhi Wu, and Kun Zhou. 2023b. Efficient Reflectance Capture With a Deep Gated Mixture-of-Experts. *IEEE Transactions on Visualization and Computer Graphics* (2023).

Yongwei Nie, Jiaqi Yu, Chengjiang Long, Qing Zhang, Guiqing Li, and Hongmin Cai. 2024. Single-Image SVBRDF Estimation Using Auxiliary Renderings as Intermediate Targets. *IEEE Transactions on Visualization and Computer Graphics* (2024).

Jannik Boll Nielsen, Henrik Wann Jensen, and Ravi Ramamoorthi. 2015. On optimal, minimal BRDF sampling for reflectance acquisition. *ACM Transactions on Graphics (TOG)* 34, 6 (2015), 1–11.

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

Pieter Peers and Philip Dutré. 2005. Inferring reflectance functions from wavelet noise. In *Proceedings of the Sixteenth Eurographics conference on Rendering Techniques*. 173–182.

J Riviere, P Peers, and A Ghosh. 2016. Mobile Surface Reflectometry. In *COMPUTER GRAPHICS forum*, Vol. 35. 191–202.

Sam Sartor and Pieter Peers. 2023. Matfusion: a generative diffusion model for svbrdf capture. In *SIGGRAPH Asia 2023 conference papers*. 1–10.

Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. 2020. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems* 33 (2020), 7462–7473.

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

Radomír Vávra and Jiří Filip. 2018. Adaptive slices for acquisition of anisotropic BRDF. *Computational Visual Media* 4 (2018), 55–69.

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, 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. In *Eurographics Symposium on Rendering Techniques*.

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

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, Article 274 (Nov. 2024), 11 pages.

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* (, Sydney, NSW, Australia,) *(SA '23)*. Association for Computing Machinery, New York, NY, USA, Article 85, 11 pages. doi:10.1145/3610548.3618239

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

Ruben Wiersma, Julien Philip, Miloš Hašan, Krishna Mullia, Fujun Luan, Elmar Eisemann, and Valentin Deschaintre. 2025. Uncertainty for SVBRDF Acquisition using Frequency Analysis. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*. 1–12.

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 (SIGGRAPH '23)*. doi:10.1145/3588432.3591559

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

Xilong Zhou, Miloš Hašan, Valentin Deschaintre, Paul Guerrero, Kalyan Sunkavalli, and Nima Khademi Kalantari. 2023. A Semi-Procedural Convolutional Material Prior. In *Computer Graphics Forum*. Wiley Online Library.

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

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

Zhiqian Zhou, Cheng Zhang, Zhao Dong, Carl Marshall, and Shuang Zhao. 2024. Estimating Uncertainty in Appearance Acquisition. In *EGSR (ST)*.