EUROGRAPHICS 2026 / B. Masia and J. Thies
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 45 (2026), Number 2

High-Gloss SVBRDF Capture Using Bounce Light

Tomáš Iser[1,2] Andrei-Timotei Ardelean[1] Tim Weyrich[1]

[1]Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

[2]Charles University, Faculty of Mathematics and Physics, Czech Republic

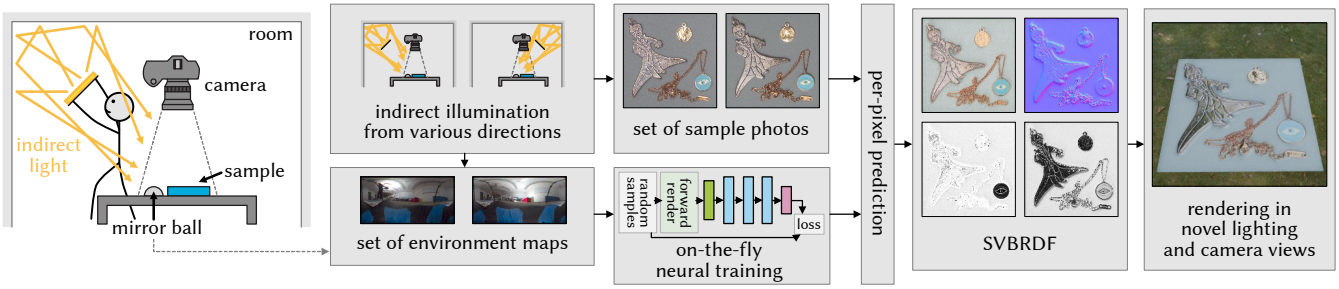

Figure 1: *Our approach for SVBRDF reconstruction makes use of indirect illumination to increase the amount of information retrieved from glossy materials. We train a neural network without a dataset of materials, fitting it to the observed environment maps. After training, the network can quickly predict high-resolution material maps for relighting.*

Abstract

Reflectance capture aims at the visual reproduction of an object under varying illumination. Past works differ substantially in their experimental overhead, from single- or few-image approaches, that employ significant (often learned) priors at the expense of biased reconstructions, to more accurate approaches that tend to be time-consuming, which to a good part is due to the need for carefully controlled illumination. Moreover, as we will show, the frequently employed point-light or directional lighting tends to clip highlights and under-sample the reflectance of glossy surfaces, leading to incorrect reconstructions under previously unseen illumination. Our work aims to strike a new balance, combining a low-overhead capture methodology with a fast (neural) model fit. A key feature of our approach is the use of handheld, indirect bounce light that enables a convenient capture methodology, limits the dynamic range of the reflectance (effectively avoiding highlight clipping) and ensures contiguous hemispherical incidence, even with few images, eliminating under-sampling of highly specular reflectance lobes. Moreover, our approach does not require training on pre-existing material datasets and thus is not restricted by the choice of dataset, and its inference scales linearly with the number of pixels, scaling exceptionally well to large image sizes. As a result, our method enables high-resolution capture of a spatially-varying reflectance distribution function (SVBRDF) from a small set of casually captured, indirectly lit photographs, making high-quality material acquisition practical even on consumer hardware.

Overall, we believe that our method occupies a unique trade-off between acquisition effort, model assumptions and resulting quality, and it has the potential to transform areas that routinely use handheld point-light sources, such as the popular reflectance transformation imaging (RTI), leading to more faithful reproductions of artefacts and their surface characteristics.

CCS Concepts

• ***Computing methodologies*** → ***Reflectance modeling;*** *Image-based rendering; Learning paradigms;*

1. Introduction

Reflectance capture aims at reproducing how an object looks under varying illumination. Past approaches can be roughly divided into: (1) *image-based methods* that simply record how an object responds to a set of carefully controlled illuminations; and (2) *model-based approaches* that try to infer explicit parameters of a reflectance model that are consistent with observations under varying lighting and/or viewing directions. Many methods from both types of approaches have been presented with substantial differences in their applicability and experimental overhead.

Single- or few-image methods, which employ significant priors (that are often machine-learned), are either restricted in the material classes they support [AWL15, AAL16, DAD*18, VMR*24] or use generative mechanisms to fill in or hallucinate unobserved information based on pretraining with carefully curated material databases. While valuable for content creation in computer graphics, these restrictions limit their utility for applications that rely on faithful reproductions of the original real samples, such as conservation and study of historic artifacts.

At the other extreme are methods that aim for high accuracy of reproduction but are time-consuming, which to a good part is due to the need for carefully controlled illumination and viewing conditions, requiring dedicated hardware and lengthy acquisition times.

Along that spectrum, *reflectance transformation imaging* (RTI) has become a hugely popular methodology in practice, widely adopted in archaeology and the museums sector. Its popularity stems from a good balance between acquisition effort and added value of the resulting datasets: a stationary camera observes a mostly-planar object surface that is successively lit from various directions, using either a handheld flash light or a dome-mounted arrangement of LED light sources; a glossy sphere next to the object records the direction from which a handheld light source shone. The captured datasets can be viewed in an interactive software that allows a user to virtually relight the object's appearance from various angles, as well as employ further enhancement through image processing, including inference of approximate surface normals.

The majority of reflectance capture methodologies, including RTI, employ point-light or directional lighting. As we will show, this tends to both clip highlights and under-sample the reflectance of glossy surfaces, leading to incorrect reconstructions under previously unseen illumination.

Our work aims to strike a new balance, combining a low-overhead capture methodology with a fast (neural) model fit. Similar to other methods, we repeatedly capture a largely planar object with a stationary camera across multiple illuminations. However, we do not illuminate the object by shining light at it directly but instead we employ *indirect bounce light*: a human operator carries a small light source that shines *away* from the object, against the walls of a room. This creates a smoothly lit environment, ensures contiguous hemispherical incidence, and avoids high dynamic range and the risk of highlight clipping in the camera sensor. These lighting environments are recorded by a mirrored sphere that is next to the object, and even for a small number of images, the large extent of the indirect light sources ensure good coverages of the hemisphere. As we will discuss in Section 3, this further eliminates under-sampling of highly specular reflectance lobes. Bounce light captures were also used in prior works, mainly to digitally relight product and portrait photography to ensure a flattering illumination that is not directly achievable in the room itself [MTB*06, MDKD16].

Our method outputs the object's appearance parameters at each position (pixel). We model surface reflectance as a spatially-varying bidirectional reflectance distribution function (SVBRDF; based on the simplified Disney BRDF model). In order to fit a BRDF to each image pixel, we introduce a lightweight per-pixel multi-layer perceptron (MLP) that is *trained on the fly* without requiring any pre-collected dataset. Once the data capture is completed, that network overfits to the observed environment maps using synthetically generated random ground truths, enabling rapid training within minutes. The trained network then infers per-pixel BRDFs within microseconds, that is, mere seconds for entire images. Considering data I/O and the need for semi-automatic verification of light source directions by state-of-the-art RTI software (`https://github.com/cnr-isti-vclab/relight`) [PCS18], this leads to reconstruction times that rival those of classical RTI.

We argue that our method is similarly practical to RTI, and we show that it offers greatly improved quality, particularly for glossy objects. Our individual contributions are as follows.

- A novel, practical methodology for reflectance capture that, unlike previous work, remains stable and maintains excellent quality even for very glossy materials.
- We demonstrate that the problem has local minima and that generic differentiable rendering is not viable.
- A robust neural SVBRDF estimator with no principled limits on texture resolutions; the BRDF is estimated for each pixel independently, with no convolutional layers.
- No reliance on any pre-collected datasets, as the neural estimator is always newly trained on the fly.

2. Background and related work

During rendering, we are solving the light transport equation (LTE) [PJH23]. How much light is reflected from a position $\mathbf{x}$ on an opaque, non-emissive surface, is determined by:

$$L_o(\mathbf{x},\theta_o,\phi_o) = \int_0^{2\pi}\int_0^{\frac{\pi}{2}} \underbrace{f(\mathbf{x},\theta_o,\phi_o,\theta_i,\phi_i)}_{\text{SVBRDF (6 dim.)}} L_i(\mathbf{x},\theta_i,\phi_i)\,|\cos\theta_i|\,|\sin\theta_i|\,\mathrm{d}\theta_i\mathrm{d}\phi_i\,, \quad (1)$$

where $\mathbf{x}$ is the two-dimensional surface position such as the texture coordinate, $L_o(\theta_o,\phi_o)$ is the outgoing radiance leaving the material at spherical coordinates θ_o,ϕ_o (polar angle, azimuth), and $L_i(\theta_i,\phi_i)$ is the incoming radiance arriving from the scene. The SVBRDF is a spatially-varying extension of BRDF $f(\theta_o,\phi_o,\theta_i,\phi_i)$, and determines how the material looks at each position.

2.1. Relighting and reflectance transformation imaging (RTI)

The problem above (Eq. 1) can be simplified by assuming a *fixed camera position* (constant view angles θ_o,ϕ_o) and *varying illumination direction* (varying θ_i,ϕ_i). In that case, only a 4-dimensional function $f(\mathbf{x},\theta_i,\phi_i)$ needs to be estimated. This is called *relighting* or *reflectance transformation imaging* (RTI) and the goal is to render a new image of the same object under the same view but with different lighting. The concept is of high importance in fields like cinematography (face relighting, initiated by [DHT*00], and later improved by neural models [HSB*22, HCT*24]) or cultural heritage (digitizing and studying artifacts or art [WLK*24]).

Traditional RTI is based on photographing the object many times, illuminated by directional lights from various angles that can

be simply calibrated using spheres [MDA02]. The reflectance can be represented, for example, as *polynomial texture maps* (PTM) whose coefficients can be numerically fitted [MGW01]. In general, for fixed θ_o, ϕ_o, we can write the reflectance as: [PCS18]

$$f(\mathbf{x}, \theta_i, \phi_i) = \sum_k a_k(\mathbf{x}) w_k(\theta_i, \phi_i) \,, \tag{2}$$

where a_k are the fitted coefficients for each pixel, and w_k are directional weights. The exact formulation of w_k differs based on the underlying model, such as PTM, *hemispherical harmonics* (HSH) [ZSD14], or interpolating the original photos via *radial basis functions* (RBF) [GCD*17].

Relighting methods are often affordable and since they do not rely on any underlying BRDF model, they are robust to self-shadowing and complicated surface geometries. The main downside is that they cannot render the object from different viewpoints, so they cannot be used in arbitary scenes in traditional rendering pipelines. They also require a significant number of input photos to ensure good interpolations, and they naturally suffer from aliasing on glossy surfaces (see Fig. 2 for a diagram and Fig. 3 for photos). The problem of aliasing persists in all methods that use a simple point-light illumination (including Sec. 2.3), and it was also highlighted by prior work [GAHO07, AWL13].

2.2. SVBRDF estimation

While the camera view is fixed in RTI, we generally want the camera position to be dynamic as well, so we need to capture the full SVBRDF $f(\mathbf{x}, \theta_o, \phi_o, \theta_i, \phi_i)$. Due to its importance for photorealistic rendering, this problem has been studied extensively, leading to approaches that differ in their tradeoff between acquisition complexity and fidelity. The traditional BRDF measurement device, the gonioreflectometer, enables precise measurements by carefully calibrating and controlling both the light source and the sensor. With a sufficiently dense sampling strategy, the device is able to capture the BRDF of the material with high fidelity; however, a single capture may take several hours [F*97, WLL*09].

Many publications proposed captures that require hardware to project illumination patterns, such as LCD screens [AWL13], light-stage cubes [KCW*18, KXH*19], RGB LED arrays [MKZ*21], or a custom single-point BRDF measurement device with embedded sensors and illuminants [DWT*10]. Some of these also use deep networks [KCW*18, KXH*19]; however, our method uses a significantly more lightweight capture setup, and our network is not trained as an autoencoder but directly outputs Disney BRDF parameters.

Other methods that require additional resources are those using polarized light (such as from LCD and CRT screens) and/or polarization filters [GCP*09, GCP*10, RRFG17], or methods that require an on-site calibration with a custom BRDF calibration chart [RWS*11], or setups with RGB-D cameras [WWZ16].

In contrast, our method enables *in situ* SVBRDF captures without specialized equipment: we only require a camera, light source, and a mirror ball. Methods that capture SVBRDF and that we directly compete with are those that rely on just a mobile phone camera with a flash light. Recently, this concept gained popularity via neural networks, which we explore in the following section.

2.3. Neural SVBRDF estimation from flash photography

Neural SVBRDF estimation is of significant importance due to the simplicity of its use. In a recent survey [KHM*24], over 50 publications on that topic were identified. In order to capture spatially varying materials from a small number of photographs, most methods assume various constraints such as: collocation of a point-light with the sensor [AWL15, ZWX*20], stationarity of the material [AWL15, AAL16, HDMR21, ZWX*20], availability of an extensive dataset for learning a prior [DAD*18, DAD*19, GLD*19, GSH*20, LSC18, LXR*18], or a combination of the above.

A common strategy among approaches based on neural networks is to train a relatively large model on synthetic data and use *one or more flash images* during inference. These photos can be taken with a handheld phone camera, making it easy to capture images with a collocated light source. One of the first single-image SVBRDF estimation metods [DAD*18] introduced a training strategy based on a synthetic dataset of 20000 SVBRDFs (after augmentations). The material maps are predicted using a U-Net [RFB15], which is enhanced through additional connections of global features. In a following work by the same authors [DAD*19], the method was improved to support an arbitrary number of input images during inference. A different approach for leveraging multiple images, using inverse rendering, was proposed [GLD*19]. As a reasonable starting point is required for the optimization, the approach leverages the single-image method as initialization. Concurrently, single-image appearance modeling was also presented by others at that time [LDPT17, YLD*18].

Later, adversarial training [GPAM*20] was successfully used to improve the results in low data settings for both material maps generation [HDMR21, ZHD*23] and SVBRDF estimation [VPS21]. A prominent work in this category is MaterialGAN [GSH*20], which adapts the StyleGAN2 [KLA*20] architecture for material maps. By training the generative model, the method learns an expressive latent space, which is optimized using inverse rendering to match the input photographs. This strategy is similar to the one employed before [GLD*19], except it does not require a specific initialization. More recently, optimizing the material maps through

Figure 2: *Aliasing in the SVBRDF capture is caused by a combination of a narrow reflection lobe of glossy materials and directional illumination that does not cover all angles. In our method, we use indirect illumination, which is smooth and low-frequency in the angular domain, covering a larger portion of the hemisphere.*

inverse rendering was shown to benefit from a tailored neural network architecture that takes into account the correlation between different maps [LSM*24]. For single-image SVBRDF estimation, several recent methods build on the success of diffusion models [VMR*24, SP23, YYSF24].

Overall, these methods seek to address the fact that estimating the SVBRDF from a single image or a few flash photographs is an ill-posed problem, by relying on a data-driven prior. Differently, we propose to make the input images more informative through indirect illumination, enabling a more accurate reconstruction.

2.4. Uncontrolled illumination and 3D reconstruction

Our method works under uncontrolled indirect illumination, so we also explore works in that setting. Several prior publications focused on BRDF estimation under completely unknown natural lighting, but the entire object had to be of a single uniform material [ON14, GRR*18, MMZ*18, RL22]. For SVBRDF, [MRR*22] showed a neural network trained on synthetic datasets that only needs a single input photo, but has limited support for metalic and roughness estimations.

Using uncontrolled illumination is more common in methods that reconstruct not only the appearance but also an *arbitrary 3D shape*, typically from multiple views of the scene, although single-image estimation is also possible [BM15]. Compared to neural radiance fields (NeRF) [MST*22] and NeuS [WLL*21], which model radiance and SDF, respectively, several methods explored reconstructions for 3D geometries, which include material properties. Such works include [DCP*14, ZCD*16] (appearance reconstruction, known geometry), [XDPT16] (shape and SVBRDF reconstruction from a rotating object), [ZLW*21] (shape and SVBRDF reconstruction, but only a constant specular lobe), [VSJ22] (shape and SVBRDF, known environment map, using differentiable rendering, limited to albedo and roughness), [BJB*21, YF23] (shape and SVBRDF, require a large-scale dataset and long training), [LWL*23] (shape and BRDF of mirror-like objects), or [WPH*24] (uncertainty in SVBRDF acquisition).

These methods that support arbitrary 3D shapes are often slower and have to solve other challenges such as pose estimation and geometry alignment across the individual captures. In contrast, our method focuses on mostly-planar surfaces and is very fast at outputting pixel-perfect SVBRDF texture estimates with an arbitrarily high image resolution (limited just by the resolution of the camera). Especially, we can reconstruct high-resolution normal maps, which is not equivalent to reconstructing a low-resolution 3D geometry.

3. Analysis

In this section, we analyze the motivation behind our approach. Our goal is to estimate the SVBRDF $f(\mathbf{x}, \theta_o, \phi_o, \theta_i, \phi_i)$ of a planar material sample. Geometry features (such as embossing and scratches) should be embedded in the SVBRDF itself, which is a concept known from normal mapping. We use a per-pixel pipeline, so we assume that each position $\mathbf{x}$ is independent and can be estimated separately.

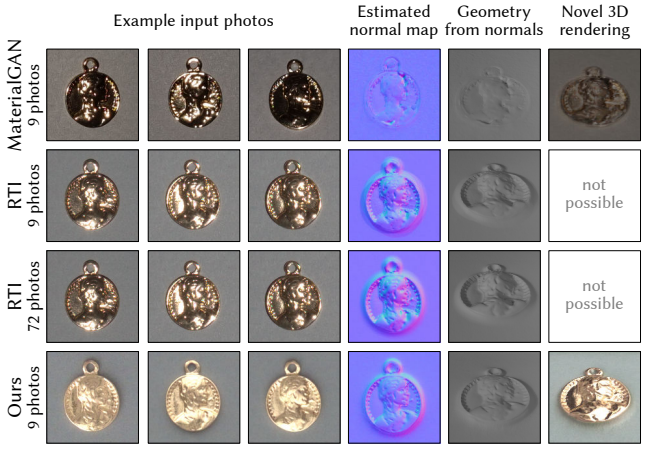

Figure 3: *Examples of photos captured for different methods, and the resulting reconstructed normal maps, geometry, and novel view rendering. The medallion is cropped from a larger capture (Fig. 14). Notice how directional lighting (top 3 rows) leads to aliasing, overexposed highlights, high dynamic range, and loss of information. With our indirect illumination (bottom row), the object remains well-lit in all photos and the dynamic range is reduced.*

Mirror reflection and directional lights For a given camera direction (θ_o, ϕ_o), a perfect mirror material would have $f(\mathbf{x}, \theta_o, \phi_o, \theta_i, \phi_i) \neq 0$ only for a single incoming direction. If the sample is illuminated by directional lights or point lights, the mathematical probability that we will align a light exactly in that direction is zero. Photographs would be black with nothing to observe.

Glossy materials and directional lights For materials that are not perfect mirrors but still are highly glossy, their reflectance lobe is very narrow (illustrated in Fig. 2). In Fig. 3, we can see that photographs of such objects under directional lights are mostly dark, and the pixels whose reflection hit a light source are significantly overexposed and clipped by the camera sensor. Underexposing the photos would improve the highlights but would render the rest of the sample completely black. Stacking multiple photos with different intensities of the camera flash would be technically challenging. Reconstructing a glossy material from flash-lit images is a difficult task that (1) requires a dense grid of illumination angles, or (2) relies on hallucinating the unseen reflections via neural networks that were pre-trained on large datasets, or (3) needs to consider additional inputs such as captures with natural illuminants [RPG16].

Mirror reflection and bounce light In our method, we use indirect illumination (bounce light). Unlike directional light, bounce light is *continuous* and covers a larger portion of the environment above the sample (illustrated in Fig. 2, right). We can simply move the bounce light around a room N times to create different environments: $\{E_j(\theta, \phi)\}_{j=1}^{N}$. Each time, we also capture a new photograph of the sample: $\{I_j\}_{j=1}^{N}$. Each point on the unknown sample reflects to a direction that we have to find and which corresponds to the surface normal vector. If we assume a mirror material, for each pixel $\mathbf{x}$ of the sample there is exactly one reflection direction

Figure 4: *Each photo contains a small mirror sphere next to the sample (Figs. 1, 13). By reprojecting the pixels of the sphere, we get an environment map of the incident illumination from all angles.*

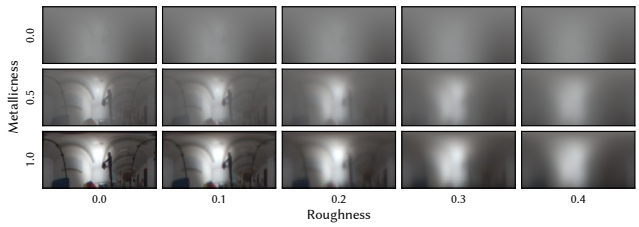

Figure 5: *Environment illumination from the point of view of a material sample with the Disney Principled BRDF. With increasing roughness, the reflection lobe widens and causes blurring. With decreasing metallicness, the reflection softens and becomes diffuse.*

(θ_r, ϕ_r) that contributes to the pixel color:

$$\forall \mathbf{x} \;\; \exists (\theta_r, \phi_r) \;\; \forall j \in \{1, \ldots, N\}: \;\; I_j(\mathbf{x}) = \rho(\mathbf{x}) \cdot E_j(\theta_r, \phi_r) \,, \quad (3)$$

where $\rho(\mathbf{x}) \in [0,1]$ denotes albedo of the mirror at position $\mathbf{x}$. Finding the unknown ρ and (θ_r, ϕ_r) for each pixel $\mathbf{x}$ becomes an optimization problem:

$$\forall \mathbf{x} \;\; (\rho(\mathbf{x}), \theta_r(\mathbf{x}), \phi_r(\mathbf{x})) = \arg\min_{\rho,\theta,\phi} \sum_{j=1}^{N} \left| I_j(\mathbf{x}) - \rho(\mathbf{x}) \cdot E_j(\theta, \phi) \right| . \quad (4)$$

The reason we need $N > 1$ environments is because a single real environment cannot be guaranteed to have a unique radiance at each direction, so the problem would be underdetermined. Note that the optimization problem is three-dimensional, but it can be reduced by dividing the individual sum elements of Eq. (3) with each other:

$$\forall j, j' \in \{1, \ldots, N\}, j \neq j' : \quad \frac{I_j}{I_{j'}} = \frac{\rho}{\rho} \cdot \frac{E_j(\theta_r, \phi_r)}{E_{j'}(\theta_r, \phi_r)} \,,$$

where ρ cancels out, resulting in a two-dimensional problem:

$$(\theta_r(\mathbf{x}), \phi_r(\mathbf{x})) = \arg\min_{\theta,\phi} \sum_{j \neq j'} \left| I_j(\mathbf{x}) \cdot E_{j'}(\theta, \phi) - I_{j'}(\mathbf{x}) \cdot E_j(\theta, \phi) \right| . \quad (5)$$

This Eq. (5) can be solved algorithmically. The function space contains plateaus and local minima, so gradient descent will usually not converge to the correct solution. However, it can be solved by a simple brute-forced grid search. This, of course, assumes that we know each environment illumination E_j. For that, we place a mirror sphere next to the sample, so each photograph I_j also contains a mirror reflection of the entire environment E_j (Fig. 4).

Principled BRDF and bounce light Solving the problem for mirror-like materials was mostly just a thought experiment. Instead of a mirror reflection, we model the material using the well-standardized *Disney Principled BRDF model* [Bur12]. To ensure the problem is tractable, we limit our method to the 7 core parameters of the model: base color (3 channels, RGB), roughness (1 channel), metallicness (1 channel), and the normal vector (2 channels, normalized). Our goal is to find the 7 parameters for each pixel, based on the N observations $\{I_j\}_{j=1}^{N}$. As the material is no longer a mirror, we cannot use Eq. (5) anymore. Instead, we need to solve the integral LTE for each parameter combination by rendering the pixels. Finding the parameters from the images is then equivalent to *inverse rendering*. We could pre-compute the intensity of each pixel $\{I_j(\mathbf{x})\}_{j=1}^{N}$ and then perform a grid-search (see Fig. 5 for how such a pre-computed table would look). But the dimensionality is too high for a lookup table to be feasible: 100 subdivisions in 7 dimensions would require 364 TB of storage assuming each element was as small as 4 bytes.

Differentiable rendering Solving high-dimensional inverse rendering problems is often done via differentiable rendering. To compute the partial derivatives (gradient), rendering frameworks such as Mitsuba 3 [JSRV22, JSR*22] rely on automatic differentiation and just-in-time compilation. We ran an experiment with Mitsuba 3 and the Adam [KB17] numerical optimization algorithm. This approach has two major downsides. First, as we show in Fig. 6, the optimization tends to converge to solutions that are far from optimum, most likely because the function space contains plateaus and local minima. As a result, the solution looks correct under the reference illumination, but fails when rendered under novel lighting. Second, differentiable rendering is orders of magnitude slower than our approach: the backpropagation pass needs to be computed for each pixel in the image in each iteration, requiring multiple samples per pixel. In contrast, as we demonstrate in the following section, our training pass is completely independent on the texture resolution and the inference pass only has a single iteration per pixel.

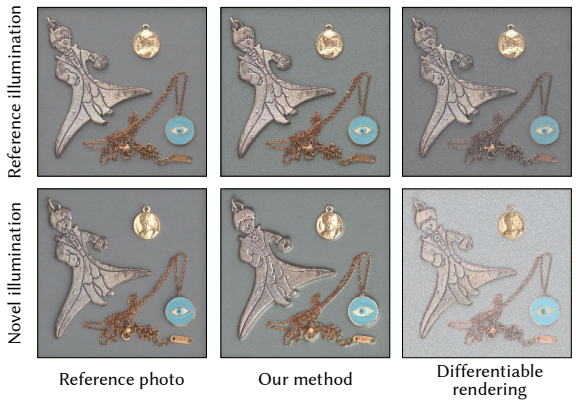

Figure 6: *Differentiable rendering tends to converge to a local minimum that looks correct for the data it has seen, but fails when viewed under novel illumination conditions.*

4. Our method

Following the analysis in the previous section, we propose to solve the SVBRDF estimation problem by using indirect bounce light and training a neural network that *learns the inverse* of the rendering function. That is, given N observations with N bounce lights, it outputs the SVBRDF parameters.

4.1. Architecture

Our goal is to create a neural network that can be trained very fast and does not have any dependencies on pre-collected datasets. By that we completely avoid the necessity to collect material samples, whether real or synthetic, as there is no need for them anywhere in the training procedure. As a consequence, instead of a convolutional or diffusion network that would require a dataset, we design a simple multi-layer perceptron (MLP) that works for each pixel separately, without any information from the neighbors.

For that to be viable, we assume that: (1) all input images $\{I_j\}_{j=1}^N$ are pixel-aligned, i.e., camera does not move; (2) inter-reflections and self-shadowing between pixels are ignored; (3) the environment illumination is located in infinity and the camera rays are collimated (orthographic projection), so that all pixels are illuminated and observed under the same geometry.

Our network $\mathcal{Q}\colon \mathbb{R}^{3\times N} \to \mathbb{R}^7$ takes for each pixel N observed RGB reflectances and outputs the 7 parameters of the *Disney Principled BRDF model* (as described in Sec. 3). The inference runs on the captured photos pixel by pixel and as a result we assemble a 7-channel SVBRDF texture that has the same resolution as the input photos. For visualization purposes and to ensure compatibility with rendering pipelines, we split the 7-channel textures into the commonly used *base color texture* (RGB), *roughness texture* (single-channel), *metallicness texture* (single-channel), and *normal map* (encoded as RGB, but internally stored as 2 channels since the 3D vectors are normalized). Note that we assume isotropic materials and thus do not store anisotropy information.

Execution For N photos with $H \times W$ pixels each (N illuminations, H height, W width), $\{I_j\}_{j=1}^N, I_j \in \mathbb{R}^{H\times W\times 3}$, the output SVBRDF textures are also $H \times W$ pixels large, and the network inference $\mathcal{Q}$ is executed $H \times W$ times in total. To ensure efficiency, we always try to fill the available GPU memory and process as many pixels as possible in a single run. If the images are larger than GPU memory, the execution is split into batches of pixels. Since each pixel is independent, we can process images with an infinitely high resolution with no visible seams in the textures. In Fig. 7, we show that the inference time scales linearly with pixel count and input photos. Inference on $N = 9$ images, 2048×2048 pixels each, takes *less than one second* on consumer hardware.

Layers Our MLP $\mathcal{Q}$ is a tiny network implemented with 3 hidden layers whose goal is to learn the mapping between reflectances and SVBRDF parameters. The network input is a tensor of $N \times 3$ values ($N \times$ RGB observations), which are in the $[0, \infty)$ range, with most values being $[0, 1]$ since our indirect illumination naturally prevents overexposed highlights. Before feeding the values to the network, we take the logarithm in order to increase the contrast of dark regions, then apply an affine transform to roughly align the values to the $[-1, 1]$ range. We observed that if the captured photos are extremely noisy (such as due to weak illumination), the logarithm exaggerates the noise too much, in which case it can be disabled. All hidden layers have 1024 neurons and are followed by a `ReLU` activation and a 1D batch normalization. The last layer of the network outputs 7 values (SVBRDF parameters) followed by a `Sigmoid` activation to ensure these values are normalized. Finally, the normalized values are remapped and split to the individual SVBRDF parameters (base color, roughness, metallicness, normal map).

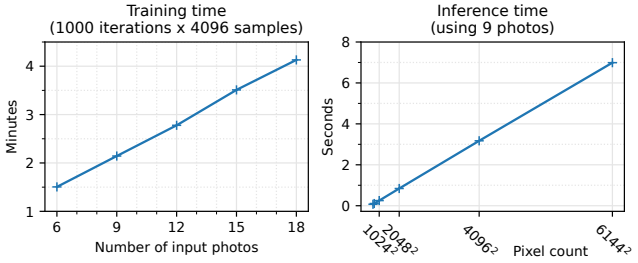

Figure 7: *Our network trains in a few minutes, scaling linearly with the number of input photos N. Inference takes less than 1 second for a 2048×2048 image, and it is linear in the number of pixels.*

4.2. Training

As mentioned, we do not train our network $\mathcal{Q}$ on a pre-collected dataset; instead, we train it on-the-fly on a large set of *randomly sampled* material parameters. During the training, the N environment maps are *fixed*, so whenever we capture new N photographs for which we want to estimate the SVBRDF, the network is trained *again from scratch*. However, our training times are very short (Fig. 7), only requiring a few minutes on consumer hardware, so retraining the network is not a problem. Hypothetically, if one could ensure that the environment illumination is fixed and repeatable, such as in a purposefully built measurement device, the training could be only performed once during an initial calibration. For in-the-wild measurements that we primarily target, the network is trained for each session independently.

The training is performed for T iterations with the batch size of M (usually around $T \approx 1000$ and $M \approx 4096$). In each training iteration $t \in \{1, \ldots, T\}$, we randomly generate a set of M material parameters $P_t \subset \mathbb{R}^{7\times M}$ that roughly cover the manifold of possible 7-channel parameters in the BRDF model. We sample the albedo and roughness uniformly in the range of possible values. The metallic attribute is sampled uniformly for half of the batch and takes an extreme value of 0 or 1 for the other half; this is done to improve the coverage of the training set given that most real materials exhibit either fully metallic (1) or fully dielectric (0) properties, and their interpolation is not physically meaningful, even though it makes sense in the BRDF modeling context. Finally, the normal vectors are randomly sampled such that their azimuthal angles uniformly cover the $[0, 2\pi]$ range, whereas the polar angle is restricted to $[-\frac{\pi}{4}, \frac{\pi}{4}]$ to avoid light being reflected below the horizon.

The rendering function, in our case implemented using the Mit-

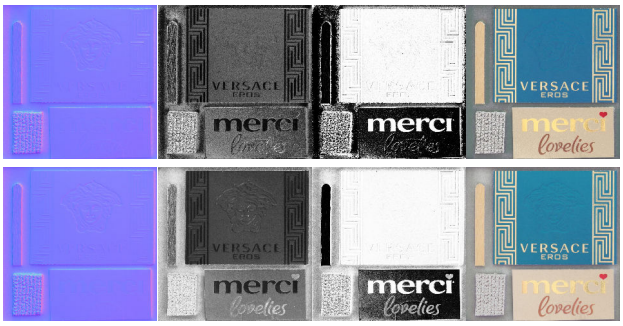

Figure 8: *Comparison between the material maps obtained without any noise augmentation during training (top) and our results, trained with noise scale* $\sigma = 0.05$ *(bottom).*

suba 3 renderer [JSR*22], is then used to compute N RGB synthetic observations, resulting in a set of rendered observations $R_t \subset \mathbb{R}^{N\times 3\times M}$. At this point, we have a set of parameters $P_t \subset \mathbb{R}^{7\times M}$ and their corresponding observations $R_t \subset \mathbb{R}^{N\times 3\times M}$. The network is trained to learn the inverse mapping, from R_t to P_t, by back-propagating through it using an L_1 loss function that compares the predicted parameters $\mathcal{Q}(R_t)$ to the ground truth P_t in the batch. All parameters are weighted equally except for normals, which receive a weight of 10, as they affect most the quality of relighting. To avoid instabilities pertaining to computing the loss function in a spherical coordinate system, we represent the normals as two free scalars (x,y) and assume $z = 1$. Then, to obtain the true normal, we simply normalize the $(x,y,1)$ vector. To train $\mathcal{Q}$, we use the Adam optimizer [KB17] with a learning rate of 10^{-3}.

We empirically observe that naively training the network in this way quickly overfits to the fixed environment maps and material model. This is the expected and desired behavior if we assume an ideal capture, devoid of any sources of error, and that the assumed BRDF model exactly matches the photos as in Sec. 5. However, when reconstructing SVBRDFs from real-world captures (Sec. 6), the photos are susceptible to noise and furthermore the real materials are not perfectly matching the Disney Principled BRDF model. As we visualize in Fig. 8 (top row), predictions from real photos would suffer from noise and bias. To mitigate this issue, for real captures we add a small amount of training noise (sampled from a normal distribution with standard deviation between $\sigma = 0.05$ and 0.10) to the input RGB values with a probability of 85%. Fig. 8 shows that this addition improves the training, and it makes the neural network more robust, resulting in cleaner material maps.

5. Synthetic evaluation

To evaluate our approach, we primarily refer to 3 methods that allow relighting materials: MaterialGAN [GSH*20], as an established deep-learned method for estimating the SVBRDF from a set of flash photographs; Correlation-Aware [LSM*24] as a more recent work that follows a similar philosophy with significant changes to the neural network architecture and optimization strategy; and finally RTI, as a widely-used relighting method.

To verify that our method is viable, we first test it on synthetic data for which the ground truth is known. Large datasets with ground truth are, for example, OpenSVBRDF [MXZ*23] and MatSynth [VD24], which is a collection of many datasets. We chose a subset of 200 materials from MatSynth, specifically materials that have a CC0 license and are not from [DAD*18] as both MaterialGAN [GSH*20] and Correlation-Aware [LSM*24] used it for training. Out of the 200 materials, 100 are randomly sampled from the dataset after the filtering described above. The other materials are selected by taking the top 100 items after sorting the materials by a glossiness score: the amount of pixels with high metallicness and low roughness; this selection is made to ensure that glossy materials form a sufficiently large part of our evaluation.

The materials are represented by base color, metallicness, roughness, and normal map textures, and we assume the same Principled BRDF model as in the rest of this paper. The ground truth materials are placed in virtual environments and rendered, resulting in synthetic input images, which are then used to fit the following reconstruction methods: MaterialGAN, Correlation-Aware, and our method, all of which take multiple images as input. We target an accurate reconstruction of the material properties, which requires several observations of the material; therefore, we use 9 input images in our experiments and do not compare to single-image methods. RTI is out of the scope for this synthetic evaluation because it does not allow re-rendering with arbitrary novel illuminations.

5.1. Generating synthetic inputs

Collocated flash images MaterialGAN and Correlation-Aware use an identical format for the inputs: a flat material sample is rendered in Mitsuba 3 from 9 different angles with a perspective camera with a collocated light source simulating a phone flash. Just like in a real capture setup, we assume the exact camera position is unknown, so we render positional markers around the sample that the methods use to reconstruct the camera position. We follow the capture instructions and Python implementations of the original method authors.

Images for our method For our method, we need to model indirect bounce light, which requires a more involved setup to simulate. To remain as close as possible to a real acquisition environment, we use a 3D model of a furnished room illuminated by a moving area light. We moved and rotated the area light 9 times to illuminate the room in different ways, pointing at the walls and furniture, just like during a real capture session. The modeling and rendering was done in Blender, producing 9 different environment maps, which are then used to render the actual samples. That is, the final renderings are done in Mitsuba 3 to ensure consistency between the inputs of the three methods.

5.2. Results

We take the output SVBRDF textures estimated by each of the three methods and perform a comparison to the ground truth. Since our method and the other methods use a slightly different SVBRDF representation, we do not compare all the textures pixel-by-pixel as that would be meaningless. Instead, we (1) compare the normal

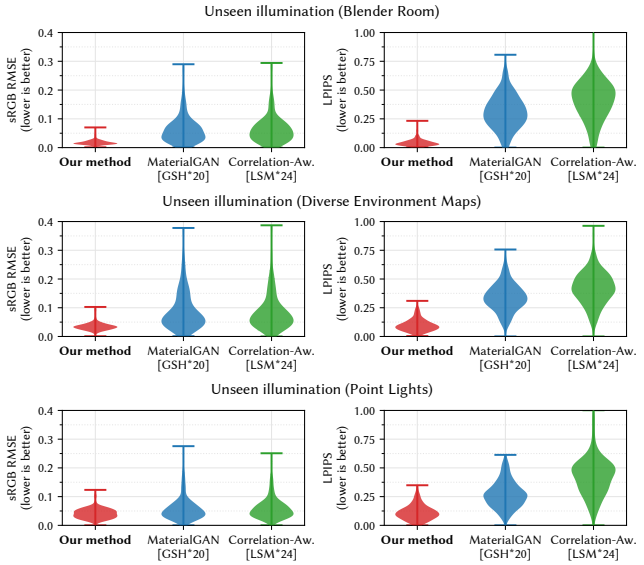

Figure 9: *Relighting errors on synthetic SVBRDF maps. Ground truth images are compared to renderings using SVBRDF textures predicted by our method, by MaterialGAN [GSH*20], and by Correlation-Aware [LSM*24].*

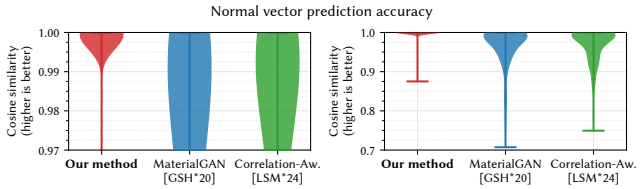

Figure 10: *Accuracy of normal vector predictions on synthetic SVBRDF maps. Ground truth normal maps are compared to ones reconstructed using our method, MaterialGAN [GSH*20], and Correlation-Aware [LSM*24]. Note that the two plots display the same data, differing just in the scale of the vertical axis.*

vectors and (2) render the reconstructed materials under previously unseen illuminations and compare the renders to ground truths. For a comprehensive evaluation of the materials through rendering, we make three categories of novel illumination: *Blender Room* – where we move and rotate the area light 9 more times to create new environment maps in a similar fashion to our training setup, *Diverse Environment Maps* – using both outdoors and indoors environment maps collected from the internet, and *Point Lights* – obtained by moving a point light at different positions above the material, creating similar images to the training views of MaterialGAN and Correlation-Aware.

The quantitative comparison of our method method with the baselines is presented using violin plots in Figure 9 and Figure 10, aggregating the results for 3 metrics: RMSE, LPIPS for renderings under novel illuminations, and cosine similarity for the predicted normals. As shown in Fig. 10, the normals predicted using our approach are significantly closer to the ground truth, with virtually all

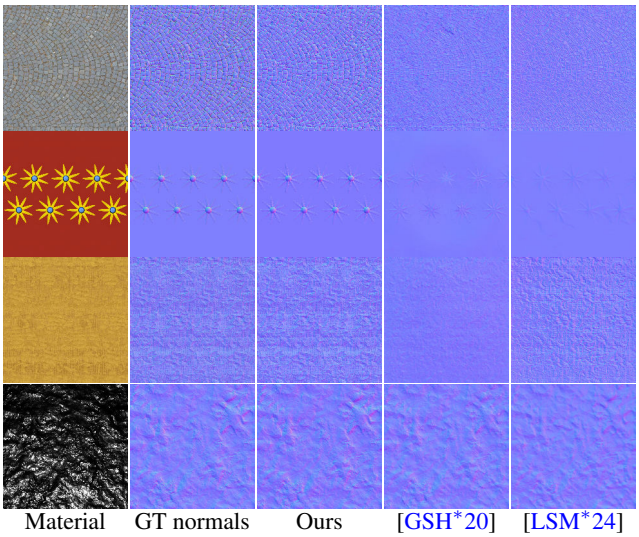

Figure 11: *Qualitative comparison on synthetic data: ground truth (GT), our method, MaterialGAN, and Correlation-Aware. Two of the materials are highly glossy: the second row is a fully metallic Christmas ornament, and the fourth row is a metal foil. For relighting renders from all three methods, see the supplement.*

predictions having a cosine similarity higher than 0.99. The average cosine similarity for our method is 0.998, which corresponds to an error of approximately 3.6 degrees. The advantage of our method over the baselines is also visible in the rendering-based metrics, as our approach yields better metrics on all three illumination categories considered.

We include in Fig. 11 a qualitative comparison of the normals produced by our method and the baselines. Note that our method is trained on a random distribution of SVBRDF parameters and has never seen the synthetic dataset, yet it achieves a very high accuracy. Please find additional comparisons, which include renderings for novel illuminations, in the supplementary material.

6. Evaluation on real captures

We also perform a qualitative comparison on real captures, during which we test a large set of materials from various objects, in three different rooms, using two different illuminants, and we also compare some of our results to MaterialGAN, Correlation-Aware, and RTI techniques.

6.1. Acquisition setup

We use an acquisition setup following the diagram in Fig. 1 and photos in Fig. 13. The material samples are placed on a flat surface. Next to the samples is a reflective, metallic sphere, which is captured as part of the photographs and is later used to reconstruct the environment maps $\{E_i\}_{i=1}^{N}$ (Fig. 4). The camera is attached to a tripod and aimed down toward the sample. Ideally, the camera is sufficiently high (about 1.3 m in our case) and lens with high focal length is used such that the rays from the sample are more or less

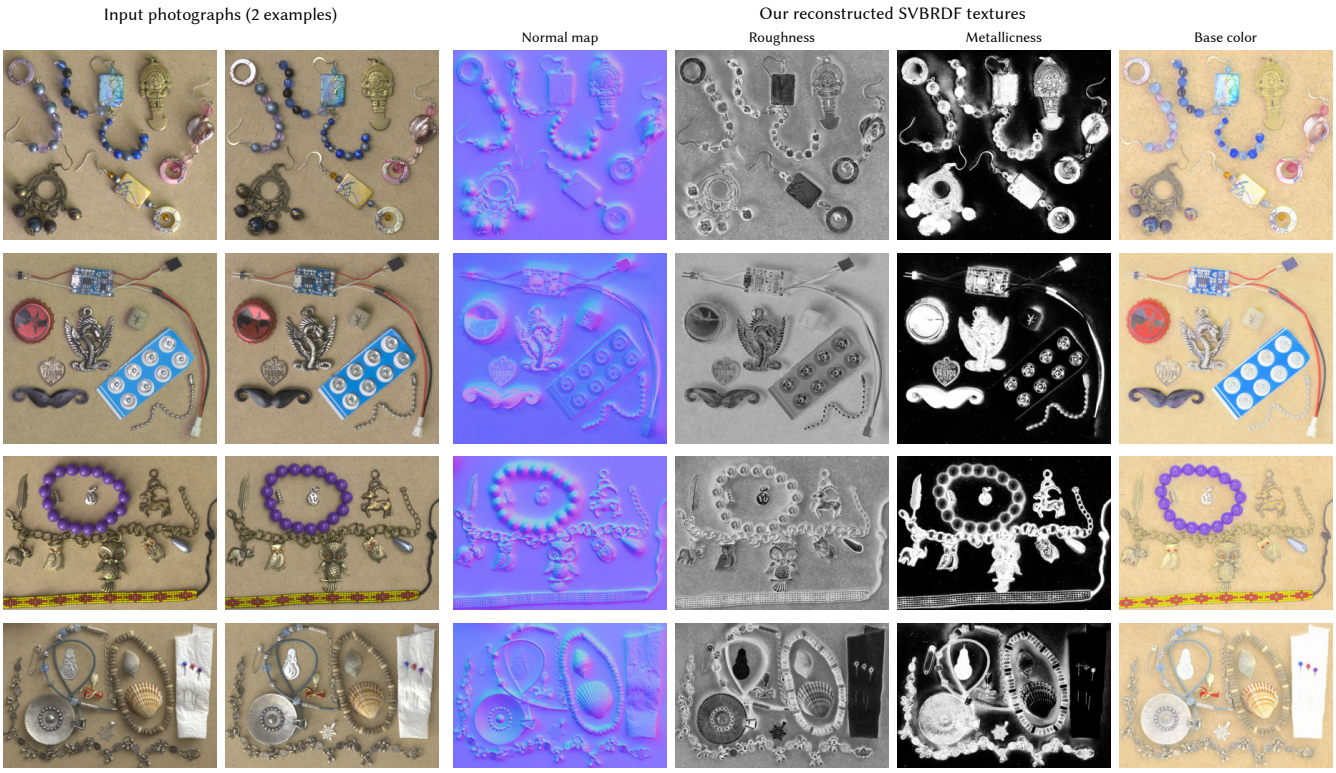

Figure 12: *Diverse results using our method (high resolution in supplement). Each scene is reconstructed from 9 photographs (2 shown).*

Figure 13: *Photographs of our simple acquisition setup*

parallel, resulting in a projection close to orthographic. We have successfully tested the setup with two different DSLR cameras and two different lenses, Canon EOS 6D Mark II with Canon EF 24–105mm f/4L IS USM lens (set to 105 mm focal length), and Canon EOS 250D with Canon EF 85mm f/1.8 USM lens.

To illuminate the room, we used an LED floodlight aimed toward an arbitrary direction, away from the material; we use an inexpensive floodlight, as only a moderate amount of power is required. For each photograph, the light is directed in a different way, changing the environment illumination. Hypothetically, if the lighting conditions could not be easily controlled, the environment could remain static and instead the material could be rotated (e.g., by using a turntable), but we have not tried this approach.

Since the camera does not move between captures, our photographs are directly pixel-aligned, and we do not require markers for registration. As an additional benefit, we do not suffer from misalignments due to parallax, which is in contrast to other methods that move the camera around the material and then have to do a re-projection of the captures. To avoid confusion, we note that markers can still be seen in some of our captures, but these markers were used in other to generate comparisons to other methods that require them, specifically MaterialGAN and Correlation-Aware.

6.2. Sample captures “in the wild”

In Fig. 12, we show a subset of our results in real-world scenes. To highlight the robustness of our approach, we captured 30 different objects featuring roughly 70 different materials in total. These scenes were captured “in the wild” with the minimal version of our setup (Fig. 13), without any positional markers that would be needed to generate comparisons to other methods. The captures were performed in two different rooms: one which features a green reflector to demonstrate resilience against tinted environments, and one which is very small, severely challenging the environment map illumination assumption; nonetheless, we obtained reasonable results in both scenarios.

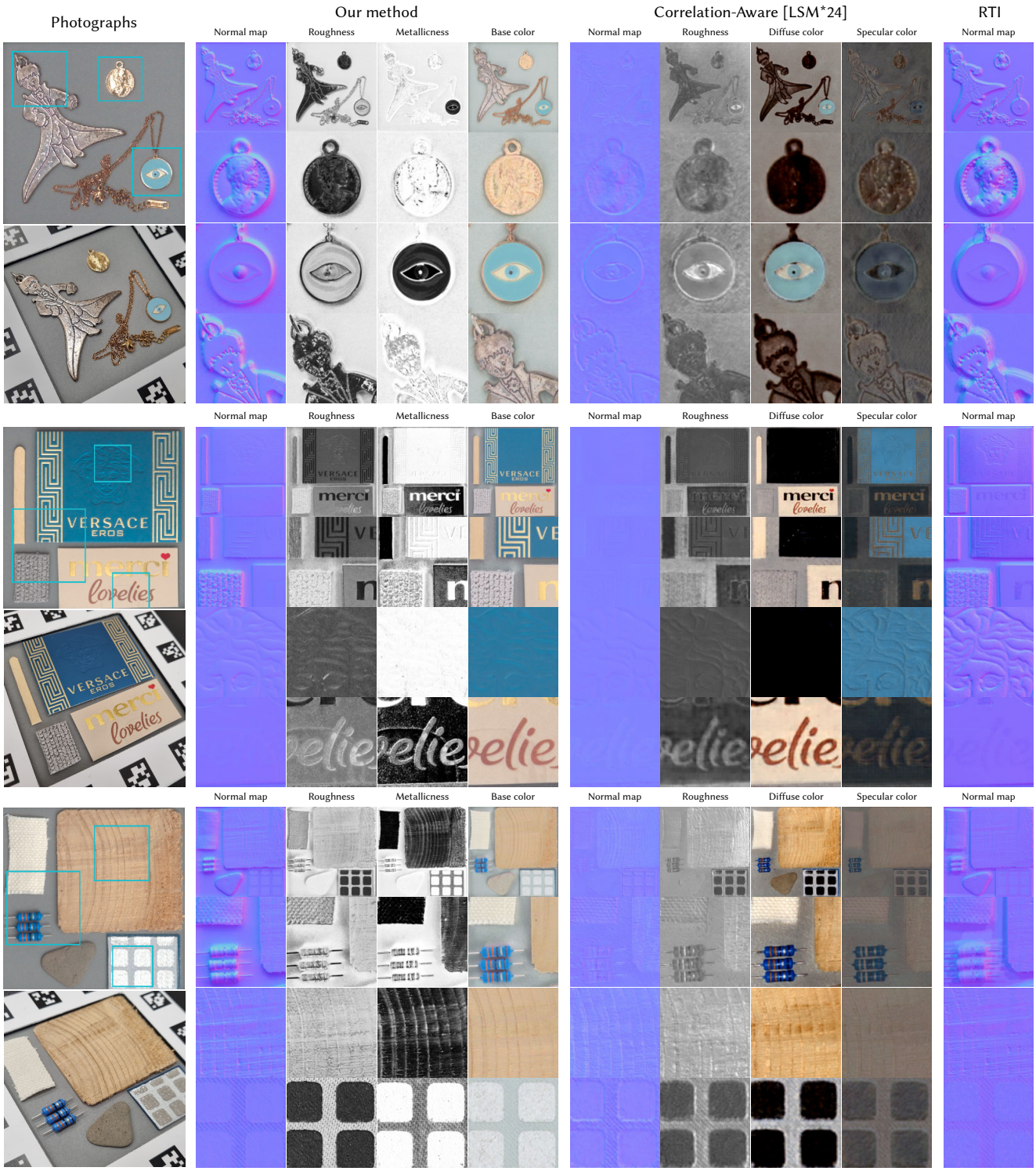

Figure 14: *Visualization of the material maps obtained using our method compared to a recent deep-learning-based SVBRDF estimation method [LSM*24], and RTI normal reconstruction. All results in this figure were based on 9 input photographs (not visualized). Notice the high fidelity of our normal maps compared to the baseline. A larger-scale comparison including MaterialGAN [GSH*20] is available in the supplementary material.*

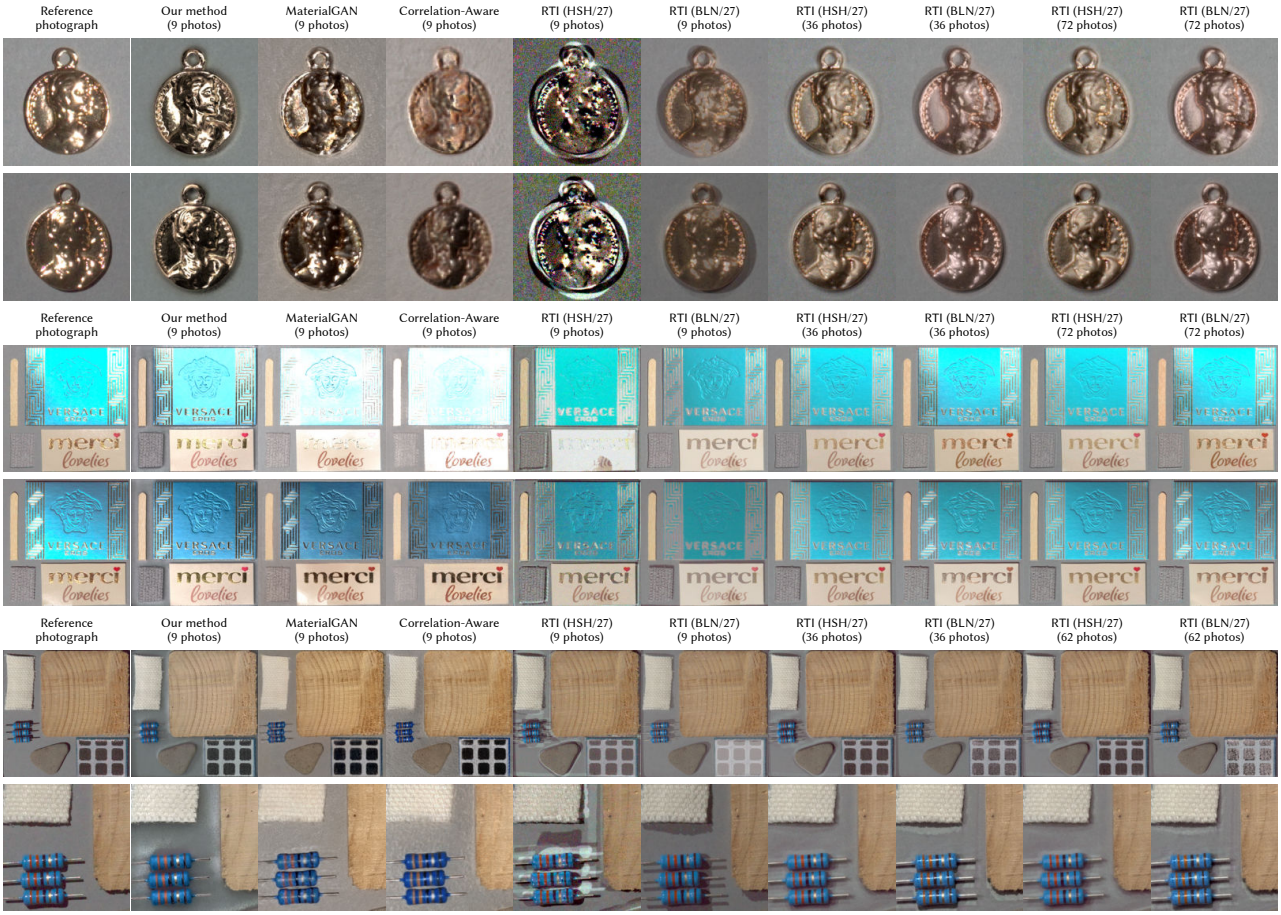

Figure 15: *Qualitative comparison for the relighting scenario (synthesis under the same top-down camera view but with novel directional illumination), under which all methods can be evaluated. The environment maps, RTI flash light positions, and novel light positions are visualized in the supplement. Please note that even though the photos were white balanced, some color inaccuracies were inevitable as the individual methods were captured under different spectra (LED floodlight vs. camera flashes).*

6.3. Captures with comparisons

In Figs. 14 and 15, we show additional results compared to other methods. To ensure fair comparisons, we had followed the acquisition protocol of each method. For example, MaterialGAN and CorrelationAware require a frame with ArUco markers; and RTI requires dark spheres around the sample to detect the flash direction. All the samples were placed in a custom-made frame with registration markers, multiple spheres, and color checkers, such that all methods could be aligned together and each method was given the benefit of its ideal capture regime.

In Fig. 14, we illustrate the material maps estimated by the different methods. Since [LSM*24] uses a different parametrization and RTI does not rely on a parametric model, the results are not 1-to-1 comparable. However, all approaches can deliver normal maps, which are a powerful indicator of the method's ability to accurately recover the properties of the material. The figure suggests that both our method and RTI recover plausible normal maps, whereas Correlation-Aware tends to predict flatter geometry and inverted structures. In contrast to RTI, our representation is easily amenable to novel environment illumination and can be directly incorporated into most modern rendering systems.

In Fig. 15, our method is compared with the baselines by rendering the material under novel illumination. We perform the evaluation using a point-light source, which is advantageous for all baselines, as this conforms with the captures they use during fitting. Nevertheless, our method generally produces images of higher quality, which more closely adhere to the novel lighting conditions. This is especially noticeable for glossy materials, such as the medallion, the *Versace* box, and the *Merci* label. MaterialGAN and Correlation-Aware offer good results mostly for diffuse surfaces, or materials with low normal variation. RTI yields impressive renderings; however, it requires significantly more images compared to our method, and, due to the image-based representation used, the results cannot be rendered from novel views (consider Fig. 16).

Model limitations We note here that none of the approaches, neither ours nor the ones we compare to, explicitly model self-

Figure 16: *Unlike RTI methods, where the view direction is fixed, our SVBRDF estimation allows rendering the sample both from a novel view (camera angle) and with a novel illumination. A short video sequence of rotating the camera and sample around the environment is available in the supplement.*

occlusions and shadows. That can lead to ghost artifacts when reconstructing taller objects with non-flat geometry (see the last row of Fig. 15). RTI's image-based representation can contend with this effect more easily as it simply bakes shadows into the corresponding light directions, but methods that directly reconstruct SVBRDF textures cannot do that. A related limitation is that by representing the geometry of the materials as normal maps, we do not consider secondary light bounces within the material, which can hurt the prediction accuracy. Finally, we observe that our method tends to reconstruct very dark materials as metallic, even when they are dielectric (noticeable in Fig. 12). We suspect that to be caused by a singularity in the BRDF model: a black metallic object observed from the top is indistinguishable from a black dielectric.

Capture limitations Our capture setup assumes that the camera rays are collimated (orthographic projection) and that the environment illumination is infinitely far, which would limit the relative sizes of captured objects compared to the room. The latter limitation could be tackled with a non-distant environment emitter [LYX*24]. Alternatively, one could consider extending the capture to natural outdoor illumination in a timelapse fashion [SMPR07].

Scaling with the number of input photographs We study how our method scales with the number of input photographs (and, implicitly, the number of environment illuminations) in Fig. 17. This study is required as, generally, the SVBRDF reconstruction problem is ill-posed. However, we can see that the ambiguity quickly decreases with the number of input images as our indirect illumination creates a rich signal, so the quality of our results also converges very fast. Fig. 17 shows that reasonable results can be obtained with as few as 6 images. The key to use as few photographs as possible is to point the indirect illumination at different sides of the environment in each photograph, otherwise the normal maps would not be properly reconstructed.

7. Conclusion

Capturing material appearance is a long-studied topic that received an increased attention with recent neural SVBRDF methods trained

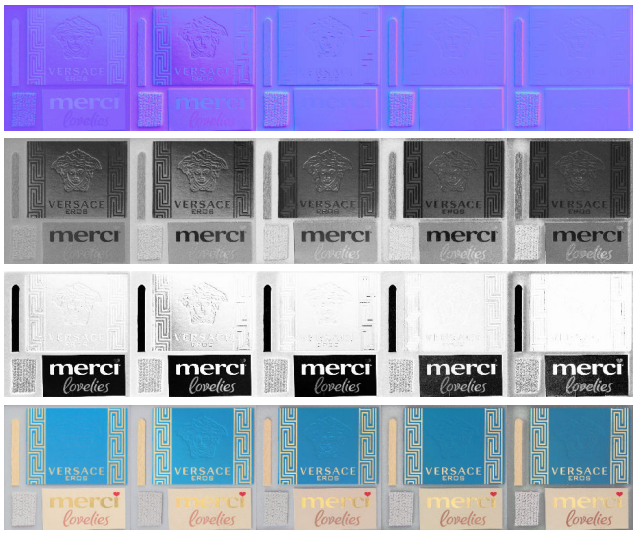

Figure 17: *Material maps obtained when running our method with different number of photographs.*

on large datasets. Still, capturing glossy materials remained a technical challenge because of aliasing and missing the directions where the light sources were located. We showed that walking around the sample while aiming a light source *away* from it (instead of directly at the sample) results in low-frequency indirect lighting and lower dynamic range, providing rich information about the material's properties. We designed a small neural network that can be trained to directly predict material properties, deliberately overfitted to the given environment illuminations without any pre-collected datasets. Our approach is validated on both synthetic and real scenes, which we capture with several types of illumination in order to ensure that the results are comparable across different methods. Our results show consistency even for samples that are glossy and have a surface geometry that we can reconstruct from our high-fidelity normal maps. That allows 3D rendering of the samples from various camera angles and under varying illuminations.

We see the key importance of our method in demonstrating that capturing with indirect bounce light is significantly more robust than previous techniques; and showing that neural SVBRDF estimators can be trained on the fly, within minutes on a consumer GPU, without relying on huge datasets.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956585. This work was further supported by the Charles University grant SVV-260822.

References

[AAL16] AITTALA M., AILA T., LEHTINEN J.: Reflectance modeling by neural texture synthesis. *ACM Transactions on Graphics 35*, 4

(July 2016), 1–13. URL: https://dl.acm.org/doi/10.1145/2897824.2925917, doi:10.1145/2897824.2925917. 2, 3

[AWL13] AITTALA M., WEYRICH T., LEHTINEN J.: Practical SVBRDF Capture In The Frequency Domain. *ACM Transactions on Graphics 32*, 4 (July 2013), 1–12. URL: https://dl.acm.org/doi/10.1145/2461912.2461978, doi:10.1145/2461912.2461978. 3

[AWL15] AITTALA M., WEYRICH T., LEHTINEN J.: Two-shot SVBRDF capture for stationary materials. *ACM Transactions on Graphics 34*, 4 (July 2015), 1–13. URL: https://dl.acm.org/doi/10.1145/2766967, doi:10.1145/2766967. 2, 3

[BJB*21] BOSS M., JAMPANI V., BRAUN R., LIU C., BARRON J. T., LENSCH H. P. A.: Neural-PIL: Neural Pre-Integrated Lighting for Reflectance Decomposition, Oct. 2021. arXiv:2110.14373 [cs]. URL: http://arxiv.org/abs/2110.14373. 4

[BM15] BARRON J. T., MALIK J.: Shape, Illumination, and Reflectance from Shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence 37*, 8 (Aug. 2015), 1670–1687. URL: http://ieeexplore.ieee.org/document/6975182/, doi:10.1109/TPAMI.2014.2377712. 4

[Bur12] BURLEY B.: Physically-Based Shading at Disney, 2012. URL: https://disneyanimation.com/publications/physically-based-shading-at-disney/. 5

[DAD*18] DESCHAINTRE V., AITTALA M., DURAND F., DRETTAKIS G., BOUSSEAU A.: Single-image SVBRDF capture with a rendering-aware deep network. *ACM Transactions on Graphics 37*, 4 (Aug. 2018), 1–15. URL: https://dl.acm.org/doi/10.1145/3197517.3201378, doi:10.1145/3197517.3201378. 2, 3, 7

[DAD*19] DESCHAINTRE V., AITTALA M., DURAND F., DRETTAKIS G., BOUSSEAU A.: Flexible SVBRDF Capture with a Multi-Image Deep Network. *Computer Graphics Forum 38*, 4 (July 2019), 1–13. URL: https://onlinelibrary.wiley.com/doi/10.1111/cgf.13765, doi:10.1111/cgf.13765. 3

[DCP*14] DONG Y., CHEN G., PEERS P., ZHANG J., TONG X.: Appearance-from-motion: recovering spatially varying surface reflectance under unknown lighting. *ACM Transactions on Graphics 33*, 6 (Nov. 2014), 1–12. URL: https://dl.acm.org/doi/10.1145/2661229.2661283, doi:10.1145/2661229.2661283. 4

[DHT*00] DEBEVEC P., HAWKINS T., TCHOU C., DUIKER H.-P., SAROKIN W., SAGAR M.: Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques - SIGGRAPH '00* (Not Known, 2000), ACM Press, pp. 145–156. URL: http://portal.acm.org/citation.cfm?doid=344779.344855, doi:10.1145/344779.344855. 2

[DWT*10] DONG Y., WANG J., TONG X., SNYDER J., LAN Y., BEN-EZRA M., GUO B.: Manifold bootstrapping for SVBRDF capture. *ACM Transactions on Graphics 29*, 4 (July 2010), 1–10. URL: https://dl.acm.org/doi/10.1145/1778765.1778835, doi:10.1145/1778765.1778835. 3

[F*97] FOO S. C., ET AL.: *A gonioreflectometer for measuring the bidirectional reflectance of material for use in illumination computation*. PhD thesis, Citeseer, 1997. 3

[GAHO07] GHOSH A., ACHUTHA S., HEIDRICH W., O'TOOLE M.: Brdf acquisition with basis illumination. In *2007 IEEE 11th International Conference on Computer Vision* (2007), IEEE, pp. 1–8. 3

[GCD*17] GIACHETTI A., CIORTAN I. M., DAFFARA C., PINTUS R., GOBBETTI E.: Multispectral RTI Analysis of Heterogeneous Artworks, 2017. Artwork Size: 10 pages ISBN: 9783038680376 ISSN: 2312-6124 Pages: 10 pages Publication Title: Eurographics Workshop on Graphics and Cultural Heritage. URL: https://diglib.eg.org/handle/10.2312/gch20171288, doi:10.2312/GCH.20171288. 3

[GCP*09] GHOSH A., CHEN T., PEERS P., WILSON C. A., DEBEVEC P.: Estimating Specular Roughness and Anisotropy from Second Order Spherical Gradient Illumination. *Computer Graphics Forum 28*, 4 (June 2009), 1161–1170. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1467-8659.2009.01493.x, doi:10.1111/j.1467-8659.2009.01493.x. 3

[GCP*10] GHOSH A., CHEN T., PEERS P., WILSON C. A., DEBEVEC P.: Circularly polarized spherical illumination reflectometry. *ACM Transactions on Graphics 29*, 6 (Dec. 2010), 1–12. URL: https://dl.acm.org/doi/10.1145/1882261.1866163, doi:10.1145/1882261.1866163. 3

[GLD*19] GAO D., LI X., DONG Y., PEERS P., XU K., TONG X.: Deep inverse rendering for high-resolution SVBRDF estimation from an arbitrary number of images. *ACM Transactions on Graphics 38*, 4 (Aug. 2019), 1–15. URL: https://dl.acm.org/doi/10.1145/3306346.3323042, doi:10.1145/3306346.3323042. 3

[GPAM*20] GOODFELLOW I., POUGET-ABADIE J., MIRZA M., XU B., WARDE-FARLEY D., OZAIR S., COURVILLE A., BENGIO Y.: Generative adversarial networks. *Communications of the ACM 63*, 11 (Oct. 2020), 139–144. URL: https://dl.acm.org/doi/10.1145/3422622, doi:10.1145/3422622. 3

[GRR*18] GEORGOULIS S., REMATAS K., RITSCHEL T., GAVVES E., FRITZ M., VAN GOOL L., TUYTELAARS T.: Reflectance and Natural Illumination from Single-Material Specular Objects Using Deep Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence 40*, 8 (Aug. 2018), 1932–1947. URL: https://ieeexplore.ieee.org/document/8014479/, doi:10.1109/TPAMI.2017.2742999. 4

[GSH*20] GUO Y., SMITH C., HAŠAN M., SUNKAVALLI K., ZHAO S.: MaterialGAN: reflectance capture using a generative SVBRDF model. *ACM Transactions on Graphics 39*, 6 (Dec. 2020), 1–13. URL: https://dl.acm.org/doi/10.1145/3414685.3417779, doi:10.1145/3414685.3417779. 3, 7, 8, 10

[HCT*24] HE M., CLAUSEN P., TAŞEL A. L., MA L., PILARSKI O., XIAN W., RIKKER L., YU X., BURGERT R., YU N., DEBEVEC P.: Diffrelight: Diffusion-based facial performance relighting. In *SIGGRAPH Asia 2024 Conference Papers* (New York, NY, USA, 2024), SA '24, Association for Computing Machinery. URL: https://doi.org/10.1145/3680528.3687644, doi:10.1145/3680528.3687644. 2

[HDMR21] HENZLER P., DESCHAINTRE V., MITRA N. J., RITSCHEL T.: Generative modelling of BRDF textures from flash images. *ACM Transactions on Graphics 40*, 6 (Dec. 2021), 1–13. URL: https://dl.acm.org/doi/10.1145/3478513.3480507, doi:10.1145/3478513.3480507. 3

[HSB*22] HOU A., SARKIS M., BI N., TONG Y., LIU X.: Face relighting with geometrically consistent shadows. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2022), pp. 4217–4226. 2

[JSR*22] JAKOB W., SPEIERER S., ROUSSEL N., NIMIER-DAVID M., VICINI D., ZELTNER T., NICOLET B., CRESPO M., LEROY V., ZHANG Z.: Mitsuba 3 renderer, 2022. 5, 7

[JSRV22] JAKOB W., SPEIERER S., ROUSSEL N., VICINI D.: Dr.Jit: A Just-In-Time Compiler for Differentiable Rendering. *Transactions on Graphics (Proceedings of SIGGRAPH) 41*, 4 (July 2022). doi:10.1145/3528223.3530099. 5

[KB17] KINGMA D. P., BA J.: Adam: A Method for Stochastic Optimization, Jan. 2017. arXiv:1412.6980 [cs]. URL: http://arxiv.org/abs/1412.6980, doi:10.48550/arXiv.1412.6980. 5, 7

[KCW*18] KANG K., CHEN Z., WANG J., ZHOU K., WU H.: Efficient reflectance capture using an autoencoder. *ACM Transactions on Graphics 37*, 4 (Aug. 2018), 1–10. URL: https://dl.acm.org/doi/10.1145/3197517.3201279, doi:10.1145/3197517.3201279. 3

[KHM*24] KAVOOSIGHAFI B., HAJISHARIF S., MIANDJI E., BARAVDISH G., CAO W., UNGER J.: Deep SVBRDF Acquisition and Modelling: A Survey. *Computer Graphics Forum 43*, 6 (Sept. 2024), e15199. URL: https://onlinelibrary.wiley.com/doi/10.1111/cgf.15199, doi:10.1111/cgf.15199. 3

[KLA*20] KARRAS T., LAINE S., AITTALA M., HELLSTEN J., LEHTINEN J., AILA T.: Analyzing and Improving the Image Quality of StyleGAN. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2020), pp. 8107–8116. ISSN: 2575-7075. URL: https://ieeexplore.ieee.org/abstract/document/9156570, doi:10.1109/CVPR42600.2020.00813. 3

[KXH*19] KANG K., XIE C., HE C., YI M., GU M., CHEN Z., ZHOU K., WU H.: Learning efficient illumination multiplexing for joint capture of reflectance and shape. *ACM Transactions on Graphics 38*, 6 (Dec. 2019), 1–12. URL: https://dl.acm.org/doi/10.1145/3355089.3356492, doi:10.1145/3355089.3356492. 3

[LDPT17] LI X., DONG Y., PEERS P., TONG X.: Modeling surface appearance from a single photograph using self-augmented convolutional neural networks. *ACM Transactions on Graphics 36*, 4 (Aug. 2017), 1–11. URL: https://dl.acm.org/doi/10.1145/3072959.3073641, doi:10.1145/3072959.3073641. 3

[LSC18] LI Z., SUNKAVALLI K., CHANDRAKER M.: Materials for Masses: SVBRDF Acquisition with a Single Mobile Phone Image. In *Computer Vision – ECCV 2018*, Ferrari V., Hebert M., Sminchisescu C., Weiss Y., (Eds.), vol. 11207. Springer International Publishing, Cham, 2018, pp. 74–90. URL: https://link.springer.com/10.1007/978-3-030-01219-9_5, doi:10.1007/978-3-030-01219-9_5. 3

[LSM*24] LUO D., SUN H., MA L., YANG J., WANG B.: Correlation-aware Encoder-Decoder with Adapters for SVBRDF Acquisition. In *SIGGRAPH Asia 2024 Conference Papers* (Tokyo Japan, Dec. 2024), ACM, pp. 1–10. URL: https://dl.acm.org/doi/10.1145/3680528.3687594, doi:10.1145/3680528.3687594. 4, 7, 8, 10, 11

[LWL*23] LIU Y., WANG P., LIN C., LONG X., WANG J., LIU L., KOMURA T., WANG W.: NeRO: Neural Geometry and BRDF Reconstruction of Reflective Objects from Multiview Images. *ACM Transactions on Graphics 42*, 4 (Aug. 2023), 1–22. URL: https://dl.acm.org/doi/10.1145/3592134, doi:10.1145/3592134. 4

[LXR*18] LI Z., XU Z., RAMAMOORTHI R., SUNKAVALLI K., CHANDRAKER M.: Learning to reconstruct shape and spatially-varying reflectance from a single image. *ACM Transactions on Graphics 37*, 6 (Dec. 2018), 1–11. URL: https://dl.acm.org/doi/10.1145/3272127.3275055, doi:10.1145/3272127.3275055. 3

[LYX*24] LING J., YU R., XU F., DU C., ZHAO S.: Nerf as a non-distant environment emitter in physics-based inverse rendering. In *ACM SIGGRAPH 2024 Conference Papers* (New York, NY, USA, 2024), SIGGRAPH '24, Association for Computing Machinery. URL: https://doi.org/10.1145/3641519.3657404, doi:10.1145/3641519.3657404. 12

[MDA02] MASSELUS V., DUTRÉ P., ANRYS F.: The free-form light stage. In *ACM SIGGRAPH 2002 conference abstracts and applications* (San Antonio Texas, July 2002), ACM, pp. 262–262. URL: https://dl.acm.org/doi/10.1145/1242073.1242275, doi:10.1145/1242073.1242275. 3

[MDKD16] MURMANN L., DAVIS A., KAUTZ J., DURAND F.: Computational bounce flash for indoor portraits. *ACM Transactions on Graphics 35*, 6 (Nov. 2016), 1–9. URL: https://dl.acm.org/doi/10.1145/2980179.2980219, doi:10.1145/2980179.2980219. 2

[MGW01] MALZBENDER T., GELB D., WOLTERS H.: Polynomial texture maps. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 2001), SIGGRAPH '01, Association for Computing Machinery, pp. 519–528. URL: https://doi.org/10.1145/383259.383320, doi:10.1145/383259.383320. 3

[MKZ*21] MA X., KANG K., ZHU R., WU H., ZHOU K.: Free-form scanning of non-planar appearance with neural trace photography. *ACM Transactions on Graphics 40*, 4 (Aug. 2021), 1–13. URL: https://dl.acm.org/doi/10.1145/3450626.3459679, doi:10.1145/3450626.3459679. 3

[MMZ*18] MEKA A., MAXIMOV M., ZOLLHOFER M., CHATTERJEE A., SEIDEL H.-P., RICHARDT C., THEOBALT C.: LIME: Live Intrinsic Material Estimation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT, USA, June 2018), IEEE, pp. 6315–6324. URL: https://ieeexplore.ieee.org/document/8578759/, doi:10.1109/CVPR.2018.00661. 4

[MRR*22] MARTIN R., ROULLIER A., ROUFFET R., KAISER A., BOUBEKEUR T.: MaterIA: Single Image High-Resolution Material Capture in the Wild. *Computer Graphics Forum 41*, 2 (May 2022), 163–177. URL: https://onlinelibrary.wiley.com/doi/10.1111/cgf.14466, doi:10.1111/cgf.14466. 4

[MST*22] MILDENHALL B., SRINIVASAN P. P., TANCIK M., BARRON J. T., RAMAMOORTHI R., NG R.: NeRF: representing scenes as neural radiance fields for view synthesis. *Communications of the ACM 65*, 1 (Jan. 2022), 99–106. URL: https://dl.acm.org/doi/10.1145/3503250, doi:10.1145/3503250. 4

[MTB*06] MOHAN A., TUMBLIN J., BODENHEIMER B., GRIMM C., BAILEY R.: Table-top computed lighting for practical digital photography. In *ACM SIGGRAPH 2006 Courses on - SIGGRAPH '06* (Boston, Massachusetts, 2006), ACM Press, p. 3. URL: http://portal.acm.org/citation.cfm?doid=1185657.1185742, doi:10.1145/1185657.1185742. 2

[MXZ*23] MA X., XU X., ZHANG L., ZHOU K., WU H.: Opensvbrdf: A database of measured spatially-varying reflectance. *ACM Trans. Graph. 42*, 6 (Dec. 2023). URL: https://doi.org/10.1145/3618358, doi:10.1145/3618358. 7

[ON14] OXHOLM G., NISHINO K.: Multiview Shape and Reflectance from Natural Illumination. In *2014 IEEE Conference on Computer Vision and Pattern Recognition* (Columbus, OH, USA, June 2014), IEEE, pp. 2163–2170. URL: https://ieeexplore.ieee.org/document/6909674, doi:10.1109/CVPR.2014.277. 4

[PCS18] PONCHIO F., CORSINI M., SCOPIGNO R.: A compact representation of relightable images for the web. In *Proceedings of the 23rd International ACM Conference on 3D Web Technology* (Poznań Poland, June 2018), ACM, pp. 1–10. URL: https://dl.acm.org/doi/10.1145/3208806.3208820, doi:10.1145/3208806.3208820. 2, 3

[PJH23] PHARR M., JAKOB W., HUMPHREYS G.: *Physically based rendering: from theory to implementation*, fourth edition ed. The MIT Press, Cambridge London, 2023. URL: https://pbr-book.org/4ed/contents. 2

[RFB15] RONNEBERGER O., FISCHER P., BROX T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Cham, 2015), Navab N., Hornegger J., Wells W. M., Frangi A. F., (Eds.), Springer International Publishing, pp. 234–241. doi:10.1007/978-3-319-24574-4_28. 3

[RL22] RHEE C.-H., LEE C. H.: Estimating Physically-Based Reflectance Parameters From a Single Image With GAN-Guided CNN. *IEEE Access 10* (2022), 13259–13269. URL: https://ieeexplore.ieee.org/document/9696243/, doi:10.1109/ACCESS.2022.3147483. 4

[RPG16] RIVIERE J., PEERS P., GHOSH A.: Mobile Surface Reflectometry. *Computer Graphics Forum 35*, 1 (Feb. 2016), 191–202. URL: https://onlinelibrary.wiley.com/doi/10.1111/cgf.12719, doi:10.1111/cgf.12719. 4

[RRFG17] RIVIERE J., RESHETOUSKI I., FILIPI L., GHOSH A.: Polarization imaging reflectometry in the wild. *ACM Transactions on Graphics 36*, 6 (Dec. 2017), 1–14. URL: https://dl.acm.org/doi/10.1145/3130800.3130894, doi:10.1145/3130800.3130894. 3

[RWS*11] REN P., WANG J., SNYDER J., TONG X., GUO B.: Pocket reflectometry. *ACM Transactions on Graphics 30*, 4 (July 2011), 1–10. URL: https://dl.acm.org/doi/10.1145/2010324.1964940, doi:10.1145/2010324.1964940. 3

[SMPR07] SUNKAVALLI K., MATUSIK W., PFISTER H., RUSINKIEWICZ S.: Factored time-lapse video. In *ACM SIGGRAPH 2007 Papers* (New York, NY, USA, 2007), SIGGRAPH '07, Association for Computing Machinery, p. 101–es. URL: https://doi.org/10.1145/1275808.1276504, doi:10.1145/1275808.1276504. 12

[SP23] SARTOR S., PEERS P.: MatFusion: A Generative Diffusion Model for SVBRDF Capture. In *ACM SIGGRAPH Asia 2023 Conference Proceedings* (Sydney NSW Australia, Dec. 2023), ACM, pp. 1–10. URL: https://dl.acm.org/doi/10.1145/3610548.3618194, doi:10.1145/3610548.3618194. 4

[VD24] VECCHIO G., DESCHAINTRE V.: Matsynth: A modern pbr materials dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024). 7

[VMR*24] VECCHIO G., MARTIN R., ROULLIER A., KAISER A., ROUFFET R., DESCHAINTRE V., BOUBEKEUR T.: ControlMat: A Controlled Generative Approach to Material Capture. *ACM Transactions on Graphics 43*, 5 (Oct. 2024), 1–17. URL: https://dl.acm.org/doi/10.1145/3688830, doi:10.1145/3688830. 2, 4

[VPS21] VECCHIO G., PALAZZO S., SPAMPINATO C.: SurfaceNet: Adversarial SVBRDF Estimation from a Single Image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 12840–12848. 3

[VSJ22] VICINI D., SPEIERER S., JAKOB W.: Differentiable signed distance function rendering. *ACM Transactions on Graphics 41*, 4 (July 2022), 1–18. URL: https://dl.acm.org/doi/10.1145/3528223.3530139, doi:10.1145/3528223.3530139. 4

[WLK*24] WILK Ł., LECH P., KLEBOWSKI M., BELDYGA M., OSTROWSKI W.: Application of a stand-alone RTI measuring system with an integrated camera in cultural heritage digitisation. *Journal of Archaeological Science: Reports 53* (Feb. 2024), 104318. URL: https://linkinghub.elsevier.com/retrieve/pii/S2352409X23004935, doi:10.1016/j.jasrep.2023.104318. 2

[WLL*09] WEYRICH T., LAWRENCE J., LENSCH H. P. A., RUSINKIEWICZ S., ZICKLER T.: Principles of Appearance Acquisition and Representation. *Foundations and Trends® in Computer Graphics and Vision 4*, 2 (Aug. 2009), 75–191. URL: https://www.nowpublishers.com/article/Details/CGV-022, doi:10.1561/0600000022. 3

[WLL*21] WANG P., LIU L., LIU Y., THEOBALT C., KOMURA T., WANG W.: Neus: learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *Proceedings of the 35th International Conference on Neural Information Processing Systems* (Red Hook, NY, USA, 2021), NIPS '21, Curran Associates Inc. 4

[WPH*24] WIERSMA R., PHILIP J., HAŠAN M., MULLIA K., LUAN F., EISEMANN E., DESCHAINTRE V.: Fast and Uncertainty-Aware SVBRDF Recovery from Multi-View Capture using Frequency Domain Analysis, June 2024. arXiv:2406.17774 [cs]. URL: http://arxiv.org/abs/2406.17774. 4

[WWZ16] WU H., WANG Z., ZHOU K.: Simultaneous localization and appearance estimation with a consumer rgb-d camera. *IEEE Transactions on Visualization and Computer Graphics 22*, 8 (2016), 2012–2023. doi:10.1109/TVCG.2015.2498617. 3

[XDPT16] XIA R., DONG Y., PEERS P., TONG X.: Recovering shape and spatially-varying surface reflectance under unknown illumination. *ACM Transactions on Graphics 35*, 6 (Nov. 2016), 1–12. URL: https://dl.acm.org/doi/10.1145/2980179.2980248, doi:10.1145/2980179.2980248. 4

[YF23] YUAN L., FUJISHIRO I.: Multiview SVBRDF capture from unified shape and illumination. *Visual Informatics 7*, 3 (Sept. 2023), 11–21. URL: https://linkinghub.elsevier.com/retrieve/pii/S2468502X23000311, doi:10.1016/j.visinf.2023.06.006. 4

[YLD*18] YE W., LI X., DONG Y., PEERS P., TONG X.: Single Image Surface Appearance Modeling with Self-augmented CNNs and Inexact Supervision. *Computer Graphics Forum 37*, 7 (Oct. 2018), 201–211. URL: https://onlinelibrary.wiley.com/doi/10.1111/cgf.13560, doi:10.1111/cgf.13560. 3

[YYSF24] YUAN L., YAN D., SAITO S., FUJISHIRO I.: DiffMat: Latent diffusion models for image-guided material generation. *Visual Informatics 8*, 1 (Mar. 2024), 6–14. URL: https://www.sciencedirect.com/science/article/pii/S2468502X24000019, doi:10.1016/j.visinf.2023.12.001. 4

[ZCD*16] ZHOU Z., CHEN G., DONG Y., WIPF D., YU Y., SNYDER J., TONG X.: Sparse-as-possible svbrdf acquisition. *ACM Trans. Graph. 35*, 6 (Dec. 2016). URL: https://doi.org/10.1145/2980179.2980247, doi:10.1145/2980179.2980247. 4

[ZHD*23] ZHOU X., HASAN M., DESCHAINTRE V., GUERRERO P., HOLD-GEOFFROY Y., SUNKAVALLI K., KALANTARI N. K.: PhotoMat: A Material Generator Learned from Single Flash Photos. In *ACM SIGGRAPH 2023 Conference Proceedings* (Los Angeles CA USA, July 2023), ACM, pp. 1–11. URL: https://dl.acm.org/doi/10.1145/3588432.3591535, doi:10.1145/3588432.3591535. 3

[ZLW*21] ZHANG K., LUAN F., WANG Q., BALA K., SNAVELY N.: PhySG: Inverse Rendering with Spherical Gaussians for Physics-based Material Editing and Relighting. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021). 4

[ZSD14] ZHANG M., S DREW M.: Efficient robust image interpolation and surface properties using polynomial texture mapping. *EURASIP Journal on Image and Video Processing 2014*, 1 (Dec. 2014), 25. URL: https://jivp-eurasipjournals.springeropen.com/articles/10.1186/1687-5281-2014-25, doi:10.1186/1687-5281-2014-25. 3

[ZWX*20] ZHAO Y., WANG B., XU Y., ZENG Z., WANG L., HOLZSCHUCH N.: *Joint SVBRDF Recovery and Synthesis From a Single Image using an Unsupervised Generative Adversarial Network*. The Eurographics Association, 2020. URL: https://doi.org/10.2312/sr.20201136. 3