

BRDF Estimation of Complex Materials with Nested Learning

Raquel Vidaurre Dan Casas Elena Garces Jorge Lopez-Moreno
URJC, Madrid Desilico Labs

{raquel.vidaurre@urjc.es|dan.casas@urjc.es} {elena.garces@desilico.tech|jorge.lopez@desilico.tech}



Figure 1: From just two input images (left) our method is capable of estimating the BRDF parameters (right, synthesized from novel viewpoints) of complex materials.

Abstract

The estimation of the optical properties of a material from RGB-images is an important but extremely ill-posed problem in Computer Graphics. While recent works have successfully approached this problem even from just a single photograph, significant simplifications of the material model are assumed, limiting the usability of such methods. The detection of complex material properties such as anisotropy or Fresnel effect remains an unsolved challenge. We propose a novel method that predicts the model parameters of an artist-friendly, physically-based BRDF, from only two low-resolution shots of the material. Thanks to a novel combination of deep neural networks in a nested architecture, we are able to handle the ambiguities given by the non-orthogonality and non-convexity of the parameter space. To train the network, we generate a novel dataset of physically-based synthetic images. We prove that our model can recover new properties like anisotropy, index of refraction and a second reflectance color, for materials that have tinted specular reflections or whose albedo changes at glancing angles.

1. Introduction

One of the main problems in Computer Graphics is the creation of synthetic images which are indistinguishable from real world objects, a process known as rendering. Among the existing rendering techniques, physically-based rendering algorithms achieve plausible results by computing the way light bounces through the scene, an operation in which the right definition of a material is critical to convey

realistic results. To describe and encapsulate the material reflectance behavior, a common model is the a Bidirectional Reflectance Distribution Function (BRDF), but the acquisition and measurement of the BRDF of real materials is a big challenge due to the sheer number of uncontrolled optical phenomena affecting the light transport on the material surface. While there are specific techniques for this purpose [13, 14, 29, 12, 33], it is a slow, constrained and expensive process. A usual approach to circumvent this problem is to use a simplified analytic model controlled by a set of parameters [22, 35, 11, 9]. Despite the expressiveness of these models, when trying to reproduce a real world material, this process of tuning parameters by hand is a time consuming and error-prone task for digital artists.

Anisotropic models are capable of depicting a great variety of materials (Figure 2), however, choosing the appropriate model for the appearance is critical, and, sometimes, the flexibility required by an artist to tweak the parameters is not aligned with the facility of finding them in an optimization procedure. The approach of Aittala et al. [4] fits a learned model from Brady et al. [9] by means of a flash-no flash image pair and using non-linear optimization methods. In opposition to an empirical BRDF, in this work we chose a physically-based, industry-oriented model [1] widely adopted in the film and video game industries. Designed with artistic expressiveness as main goal, it has the key challenge of having a non-orthogonal, non-convex parameter space, making existing optimization methods unsuitable.

In this work we propose a novel real-time method based on neural networks to automatically estimate the BRDF parameters of an arbitrary material using just two low resolution images. We estimate ten parameters that model com-

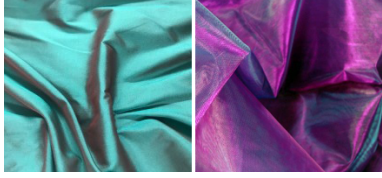


Figure 2: Examples of anisotropic materials

plex reflectance effects, including anisotropy. As this is a highly ill-posed problem, we make some assumptions: we assume that pictures are taken in a controlled (but low cost) environment and, we also assume planar objects with homogeneous materials.

Recent methods based on deep learning techniques have proved to be successful in fitting the parameters of isotropic material models from arbitrary single images [28]. However, real world materials like brushed metals or cloth cannot be modelled with simple isotropic models like Phong [21] or Ward [34].

This paper makes the following contributions:

- A method that estimates the parameters of an anisotropic BRDF from two shots of the material at 30° and 90° degrees.
- A hybrid deep learning architecture that combines feature learning with nested fully-connected layers that leverage dependencies in the parameter space of the model.
- A novel loss function that combines pure classification with gaussian smoothing to perform regression of the parameters.
- A novel synthetic dataset of planar samples of anisotropic materials rendered under multiple viewpoints.

2. Related Work

Anisotropic Materials The final appearance of anisotropic materials is determined by multiple factors including the mechanical properties of the material along with the viewer and light source orientation. Therefore, accurately capturing these complex interactions require the use of expensive gonireflectometers [13, 14], or controlled setups with various sampling strategies [29, 12, 33].

Similarly, a few parametric models exist to simulate anisotropic behaviors based on microfacets [35, 11], Wasserstein coordinates [8], or non-physically-based models with editable capabilities [22]. Among them, the closest to our work is the method of Aittala et al. [4]. Equally relying on an image pair, they use Levenberg-Marquardt to optimize the orthogonal parameters of a machine-learned

microfacets model [9]. In contrast, our anisotropic model is artist-friendly, meaning in practice that different set of parameters might yield the same appearance, and thus, making the problem unsuitable for conventional optimization methods. In this work, we handle this ambiguity by modeling the parameter relationships with a nested neural network that estimates the parameters of our model, given only two pictures of the material taken under two different orientations.

Other methods that model and capture anisotropy cannot be used in a wild and uncontrolled environment. We refer to the state-of-the-art paper of Guarnera et al. [16] for a complete analysis of anisotropic models and acquisition methods. We present the first low cost and easy to use a method that capture anisotropic materials using a highly expressive and artist-friendly model.

Isotropic Materials On the other hand, isotropic materials, have received much more attention due to their simplicity to capture [27], model and render in virtual environments [37]. Non deep-learning methods for isotropic models include: capturing SVBRDF and normal maps from two shots [4], single shots [3], dictionary-based [18], or based on neural texture synthesis [2]. Others take as input RGBD images like [36] or [26].

However, recent works that rely on deep learning techniques have proved to outperform conventional optimization. We can categorized these methods in two big groups according to whether the loss function is computed as: 1) the pixel-wise difference between the ground truth image and a rendered image computed with the parameters predicted by the model, i.e. the render loss ($rloss$) [19, 23, 38, 15, 28]; 2) the difference between the ground truth and the estimated parameters, i.e. the parameter loss ($ploss$) [20, 24, 28]. Methods that rely on the $rloss$ are in general more precise as the training learns how much a variation of parameters affect the desired results. However, they need the render function to be computed almost in real time. Our anisotropic model is computationally complex and takes several minutes to compute, making our network unfeasible to be trained with the $rloss$. We get inspiration from the work of Kim et al. [20], that takes as input an RGBD image, and uses the $ploss$ enhanced with color histogram-based regularizators to find the five parameters of an isotropic Ward model [34]. On the contrary, we use a nested architecture designed to reduce the ambiguity given by the non-orthogonality of the parameters. With the exception of Meka et al. [28], the other methods estimate the material parameters by regression. Instead, we pose the problem as a classification task which results in more robust estimations with greater explicability. Table 1 shows a summary of the deep learning based methods for isotropic models.

Method	Architecture(s)	Loss	Material Model	Output
[19]	U-Net	<i>rloss</i>	Phong [21]	Images: albedo, irradiance, speculars, and occlusions
[23]	[1] CNN+FC; [2] U-Net	<i>rloss</i>	SVBRDF Ward [34]	[1] RGB specular and roughness (4 params); [2] Images: albedo, normals
[38]	CNNResnet+FC	<i>rloss</i>	Nielsen [30]	15 parameters
[15]	[1] U-Net; [2] CNN+FC, U-Net	<i>rloss</i>	[1] R. Map [17] [2] Phong [32]	[1] image; [2] 7 params + image spherical illumination
[24]	CNN + FC	<i>rloss</i> , <i>ploss</i>	Directional Statistics BRDF [25]	normals, environment map, 108 material params
[28]	Stacked U-Net	<i>rloss</i> , <i>ploss</i>	Blinn-phong [7]	4 params (RGB albedo and specular coefficient)
[20]	CNN/Voxel	<i>ploss</i>	Isotropic Ward [34]	5 parameters
Ours	CNN+nestedFC	<i>ploss</i>	Maxwell (Ashikhmin-Shirley's) [5]	10 parameters

Table 1: Comparison of deep learning-based methods for isotropic materials. All the methods take as input a single image except the method of Kim et al. [20] that needs RGBD images, the method of Li et al. [23] which constrains the input to planar samples of metal, wood or plastic, and our method, that requires two captures. Note that our method is the first to fit a complex anisotropic model, while all the other methods use isotropic material models.

Datasets for learning-based methods. Any supervised machine learning-based method requires an annotated dataset for training. Given the variety of models, each method has developed its own dataset specific for the task at hand without publishing it, with a few exceptions. The dataset of OpenSurfaces [6] with perceptual Ward [31] parameters obtained via crowdsourcing, SynBRDF [20] containing 500K RGBD images rendered under a variety of natural illumination (5000 materials, 5000 shapes), Objects under Natural Illumination Database [25] with six objects taken under five natural illumination environments with calibrated ground-truth geometry and illumination. These datasets rely on simple isotropic models of materials. Our MaxBRDF dataset, which we plan to make available, provides a very complex set of planar samples of anisotropic materials.

3. BRDF Model Analysis

We focused our work in the standard BSDF of the commercial software Maxwell Render [1], the reason is twofold: first, Maxwell Render is an unbiased physically based rendering engine, capable of generating realistic synthetic images reproducing a real setup [38]. And second, Maxwell's underlying BSDF is designed for artists, with very expressive but non-convex and non-orthogonal parameters: this is a challenging scenario for classic optimization approaches, but a nice testbed to explore the potential of CNN architectures for parameter fitting.

Maxwell Render's underlying BSDF is a reflectance (BRDF) and transmittance (BTDF) model defined by twenty-five (25) parameters. Although artists usually combine this model with spatially varying (SV) textures, often

layered in a stack of multiple BSDFs, for the purpose of this work we match the material in isolation, although it could be extended to SVBRDF with further statistical analysis, such as the texture transfer technique described by Aittala *et al.* [4]. We focus our study in the reflectance component (BRDF) of the BSDF, as shown in Figure 3 and defined by the equations described next.

The BRDF defining the opaque reflectance is an adaptation of Ashikhmin-Shirley's model [5] able to represent complex behaviors like anisotropy and Fresnel reflections. The main difference is the elimination of the original diffuse term ρ_d , and the inclusion of a 90° -reflectance value which is dominant at glancing angles and the reversion to a purely diffuse model when the roughness is approaching 100%, as shown in Equation 1 which defines the BRDF model:

$$f(\omega_i, \omega_o) = \begin{cases} f_m(\omega_i, \omega_o) & R \leq 90 \\ \alpha f_m(\omega_i, \omega_o) + (1 - \alpha) \text{RGB}_0 \cos \theta_o & R > 90 \end{cases} \quad (1)$$

$$\alpha = 0.1|100 - R|$$

$$f_m(\omega_i) = (\text{RGB}_{90} (1 - \cos \theta_o) + \text{RGB}_0 \cos \theta_o) \rho_s(\omega_i, \omega_o) f_r(\omega_i) \quad (2)$$

$$f_r(\omega_i) = \frac{1}{2} \left(\frac{k^2 \cos^2 \theta_i + (N_d \cos \theta_i - 1)^2}{k^2 \cos^2 \theta_i + (N_d \cos \theta_i + 1)^2} + \frac{k^2 + (N_d - \cos \theta_i)^2}{k^2 + (N_d + \cos \theta_i)^2} \right) \quad (3)$$

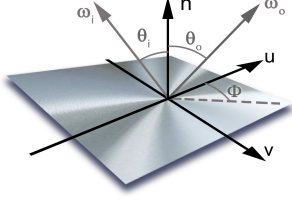


Figure 3: Diagram of Maxwell Render’s BRDF model. The vectors ω_i and ω_o are generally associated to incoming light and viewer direction.

$$\rho_s(\omega_i, \omega_o) = \frac{\sqrt{(n_u + 1) + (n_v + 1)}}{8\pi} \cdot \frac{(n \cdot h)^{n_u \cos^2 \phi + n_v \sin^2 \phi}}{(h \cdot k) \max((n \cdot \omega_i), (n \cdot \omega_o))} \quad (4)$$

Where f_r is the Fresnel component, and ρ_s is the specular component (See equations 3 and 4). Please note that the operations with $\text{RGB}_0, \text{RGB}_{90}$ are not representing a simple multiplication of RGB channels, as the luminance is controlled by the Fresnel component. Only the chromatic hue is conserved, yielding an interesting feature for artists to represent tinted Fresnel reflections in materials such as metals.

Based on these equations, a wide range of materials can be represented by artists tuning the following ten parameters:

- θ_0 : **Roughness (R)** determines the tiny imperfections that make a material reflect the light in a diffuse way. Its influence on the final appearance is very significant and most of the remaining parameters depend directly on its value. We can observe it for RGB_0 and RGB_{90} in Equation 1 and 2, for **anisotropy** in equation 5 and for IOR (N_d) in Equation 2, as the term ρ_s is multiplying the Fresnel term ($f_r(\omega_i)$).
- θ_{1-2} : **Anisotropy and Anisotropy Angle** address to how directional the light is reflected. Anisotropy is a common property in brushed metals. The n_u and n_v exponents depend on the value of R and **anisotropy**, and control the shape of the specular lobe in a similar fashion to the Ashikhmin-Shirley’s model [5] as defined in Equation 5. The **anisotropy angle** controls the direction of the effect, by setting the origin w.r.t. the Φ angle within the plane defined by the surface U,V vectors.

$$\begin{aligned} n_u &= (0.001 + 0.00999R)^{-3} \\ n_v &= (n_u^{-1/3} + \text{Anisotropy}^3)^{-1} \end{aligned} \quad (5)$$

- θ_3 : N_d is the Index Of Refraction (IOR). It determines how light propagates through the material and

the amount of reflection. A material with N_d set to a low value will have weak reflections, while high values will create a mirror-like effect. The complex part of the IOR (the refraction extinction coefficient k) is set to zero in our data, due to its negligible effect in opaque materials.

- θ_{4-6} : **Reflectance 0° (RGB_0)** refers to the color of a material, i.e. how light is reflected when it hits the surface and is seen from a front view. It is composed of 3 parameters, the reflectance color in RGB space.
- θ_{7-9} : **Reflectance 90° (RGB_{90})**, like the previous one, its three parameters determine the amount of light that is reflected, but in this case, when it is seen at 90° . The influence of each of the reflectance colors depends strongly on the roughness value R, but also on the N_d value. In general, the RGB_0 value is more prevalent than RGB_{90} on the final appearance, due to the form factor ($\cos \theta_o$) which adds weight to the camera-facing surface pixels and the use of a Lambert model for high values of R.

The equations show no interdependencies between Anisotropy and IOR(N_d), but in our analysis of the material appearance variation produced by fixing one parameter and varying the other for regular increments of R, the anisotropy-dependent term $\rho_s(\omega_i, \omega_o)$ produced a greater variation per pixel for a given camera response. Intuitively, for increasing values of N_d the existing reflections on the surface vary proportionally in luminance and color, while an equivalent differential variation on anisotropy, affects the shape of the reflection lobe ρ_s , drastically altering the shape and position of the reflections over the surface (see supplementary material).

Optimization Strategy Any attempt to fit this set of ten non-convex parameters will face the problem of local minima produced by first, multiple possible combinations producing the same spatially local values, and second by overlaid appearance effects masking each other. In an ideal case we would try to detect and isolate the effect of one parameter over the others at a given pixel, but with traditional optimization approaches, this has proven to be challenging.

On the early stage of the research, we tried to fit the three parameters of RGB_0 albedo using a classical optimization method, but the results were limited. These preliminary results and the nature of our parameter space, led us towards deep neural networks, more suitable for such scenarios, materializing the architecture described in the next section.

4. BRDF Estimation

Our goal is to provide a low-cost tool to easily capture the appearance of complex anisotropic materials. To this

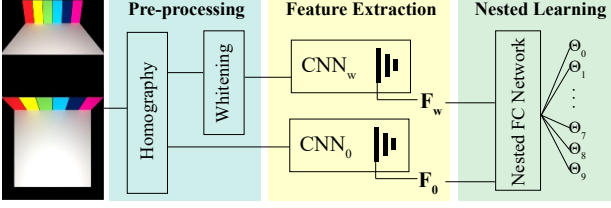


Figure 4: Pipeline of our method.

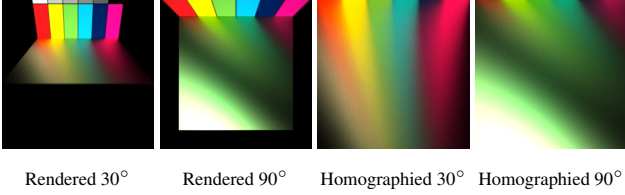


Figure 5: Two of the 14 views of a sample of our MaxBRDF dataset, and their corresponding *homographed* version after the pre-processing step described in Section 4.1

end, we train a combination of deep neural networks that, taking as input two shots of the target material, yields the values of the ten parameters of our anisotropic model described in Section 3. Figure 4 shows the pipeline of our method. First, an initial pre-processing step transforms the input pair into a more suitable representation by means of computing its planar homography and whitening the images. The *homographed* images are fed into two CNNs that extract a set of features, which are the input to our nested neural network that predicts the material parameters.

4.1. Pre-processing

Our input consists of two views of the material captured at 30° and 90°, as we show in Section 5.1, we empirically observed that these two views alone summarize most of the interesting features of the material. Before feeding the input image pair to the network, we perform a pre-processing step. We leverage the constraint to planar materials and find the homography that warps each input material to the orthogonal camera plane, resulting in a squared image which we referred to as *homographed* input. This step facilitates the training and inference of the network particularly in real images that may have slightly different backgrounds. This transformation is shown in figures 5 and 10 for synthetic and real data, respectively. Additionally, the images are whitened to remove color correlations which hinder the estimation of parameters such as *roughness* and the index of refraction N_d .

4.2. Parameter Loss

Before defining each of the modules of our pipeline, we describe the loss function that our network uses. Our goal

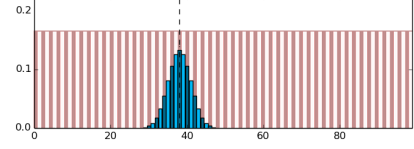


Figure 6: Input label after Gaussian smoothing (blueish bins), originally a one-hot value of 38 (dashed line).

is to predict the values for a set of parameters, a problem typically solved as a regression task [20, 23]. However, by posing the problem as a classification task [28], we reduce the complexity: from having to find a number within the whole domain of the real numbers, to a fixed set of unknown values. We empirically found that discretizing each parameter into 100 uniform bins provides enough accuracy and convergence of the models. In contrast to most of the previous methods, which encode the material albedo using RGB space [28], we leverage the *perceptually uniform* distribution of CIELAB space and encode the RGB_0 and RGB_{90} parameters in LAB. This enforces that a misclassified bin in this color space directly match human perception, improving the accuracy of the predictions, as we show in Section 5.

A vanilla classification loss function computes the cross-entropy loss per each parameter θ_i as

$$\mathcal{L}_{\theta_i}(X, Y) = - \sum_{j=1}^N y_j \log(x_j), \quad (6)$$

where $N = 100$ is the number of bins/classes, x_j is the probability for the class, and y_j is the true label codified using one-hot encoding. However, in our particular scenario of regression-by-classification, a weakness of the loss defined in Equation 6 is that *almost* correctly classified parameters are dismissed. In other words, vanilla classification loss makes sense when there is no relationship between the classes, however, in our case, small variations in the parameters produce imperceptible variations in the render image. Therefore, mispredicted classes located close by in the parameter space should be rewarded. To this end, we apply a Gaussian 1-dimensional discrete filter of radius nine to the one-hot encoded labels Y (see Figure 6) and redefine our loss as:

$$\hat{\mathcal{L}}_{\theta}(X, \hat{Y}) = - \sum_{j=1}^N \hat{y}_j \log(x_j) \quad (7)$$

where $\hat{y}_j$ is the value after application of the Gaussian smoothing to the whole vector Y . This change led to smooth results and patently improved the accuracy of our model.

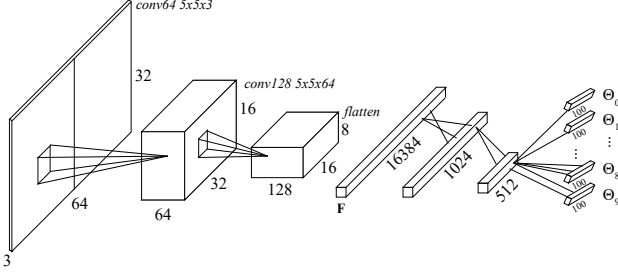


Figure 7: Architecture of each of the two CNN of our Feature Extraction module. We obtain the feature vectors $\mathbf{F}_0$ and $\mathbf{F}_w$ corresponding to the output of the second convolutional layer for a network trained with the original data CNN_0 and a network trained with the whitened data CNN_w .

4.3. Feature Extraction

The first module of our pipeline aims at extracting highly discriminative features from the input pair. It consists of two identical convolutional layers with 64 and 128 5×5 filters, respectively. Each layer is followed by a ReLU (Rectified Linear Unit) activation, a batch normalization, an LRN layer (Local Response Normalization) and a max-pooling layer with a stride of 2. After that, we apply three dense (or fully connected) layers, all of them followed, again, by a ReLU layer. Between all the layers, we apply dropout (with 75% of keeping values) to make the network more robust and avoid overfitting. A diagram of the network is shown in Figure 7. Note that, in this step, all the ten classifiers are part of the same network that learns shared weights in the hidden layers and diverges in the last fully connected layer. Such structure is usually known as multitask learning and it improves the accuracy, as the learning of similar but different tasks can help the network to recognize relations and adds some noise that helps generalization [10].

The total loss to train this network is given by the sum of the local losses (Equation 7) per parameter as follows:

$$\mathcal{L}_{total} = \frac{1}{P} \sum_{i=1}^P \hat{\mathcal{L}}_{\theta_i}, \quad (8)$$

where P is the number of parameters.

We train two identical end-to-end networks: the CNN_0 network, which takes the original homographed pair as input; and the CNN_w network, which inputs are additionally whitened. The performance of these two networks is reasonable, as our ablative study shows in Section 5, but still far from the ground truth. The reason is that the dependencies between the parameters are implicitly learned by the network, increasing the difficulty of the learning process. In light of these results, and as we know the parameter dependencies from the model definition in Section 3, we manually designed a nested fully-connected architecture that replaces

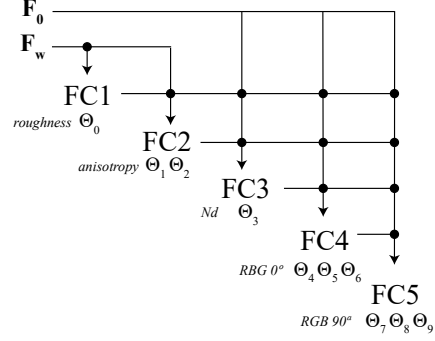


Figure 8: Nested Architecture. Each $\bullet$ represents a concatenation of features. Each FC contains three fully connected layers with the same structure as the CNN baseline architecture for feature extraction. Note that the feature vector $\mathbf{F}_w$ is used as input for FC1 and FC2, and $\mathbf{F}_0$ feeds FC3, FC4 and FC5.

the FC layers of the current CNN and combines the knowledge learned from both models, CNN_0 and CNN_w .

4.4. Nested Learning

Given the feature maps $\mathbf{F}_w$ and $\mathbf{F}_0$ obtained in the Feature Extraction module, we train a nested fully connected architecture hand-crafted specifically for the task of estimating the parameters of our reflectance model. Each fully connected block (FC) has the same structure as the fully connected block of the CNN of the Feature Extraction module, i.e. three layers with intermediate units of sizes 1024 and 512 with a softmax activation unit per parameter of size 100. Each FC layer is trained independently and differ in the input and the output as described below.

From the definition of our BRDF model, in Equation 1, we observe that there is a dependency between the parameters, i.e. some parameters overwrite others. We explicitly model this dependency in the way feature maps and outputs of the FC layers are concatenated. The connections are depicted in Figure 8, where the dots mean a concatenation operation between vectors that intersect. For example, the roughness parameter θ_0 takes as input the feature vector $\mathbf{F}_w$ and outputs 100 units; these units are therefore concatenated with the feature vector $\mathbf{F}_w$ to form the input of the layer FC2. FC2 will learn two parameters θ_1 and θ_2 , which are again concatenated with the first parameter θ_0 and the feature vector $\mathbf{F}_0$ to form the input of the layer FC3.

5. Evaluation and Results

5.1. Dataset and Training

MaxBRDF Dataset It is well known that training deep learning methods require large amount of annotated data,

however, annotating real world materials with their corresponding parameters of such complex model is unfeasible and error-prone even for skilled artists. We therefore present a new physically-based rendered dataset, the MaxBRDF dataset, with much richer annotations than existing works, which we use to train our networks.

Samples of the dataset are shown in supplementary material. We use Maxwell Render to generate 126K planar samples of 9K different materials varying our ten target parameters described in Section 3. For each material, we render 14 different views. First, with the light source set at 45° of elevation above the sample, four rotations of the material with camera at 45° , plus five orientations of the camera (15° , 30° , 45° , 60° and 90°), and finally, five additional views with the same camera directions but the light source at 90° . In the former, we add a color checker, as the way the material reflects this colors can help detect some important features. The scene is illuminated with a spherical emitter that simulates a bulb with power value of 1000W and efficacy of 17,6 lm/W. The overall setup is simple and designed to be reproducible with real objects and material samples.

Training We use the synthetic data to train our neural networks. In particular, only the views at 30° and 90° with the color checker are used for training. We found that these views are sufficiently representative of most materials and we empirically proved that they are optimal for this task. The 12 other views are used at test time to validate the generalization of the predicted material to novel views (see supplementary material for an example). We separate 2/3 of the data for the training, 1/6 for validation and 1/6 for test. For some of the models, We perform data augmentation by adding Gaussian noise with a mean of 0 and random standard deviation between 0 and 4. We evaluate the performance of this operation in the following section.

We first train the networks CNN_0 and CNN_w using the parameter loss of Equation 7 with Mini-batch Stochastic Gradient Descent, learning rate of 0.003 and random initialization of the weights. Then, we use the features learned in this first step to train the nested module. For this aim, the feature vectors are kept fixed (the weights of the convolutional networks aren’t updated) and the weights of the new fully-connected layers in the FC networks are randomly initialized and trained with a learning rate of 0.03.

5.2. Quantitative Evaluation

We quantitatively evaluate our method using the test set of our MaxBRDF dataset, from which we know both the ground truth parameters as well as the rendered images of each sample.

Figure 9 depicts the pixel-wise mean square error (MSE) of the rendered predictions and demonstrates that our approach (“nested LAB”) outperforms the baseline and other

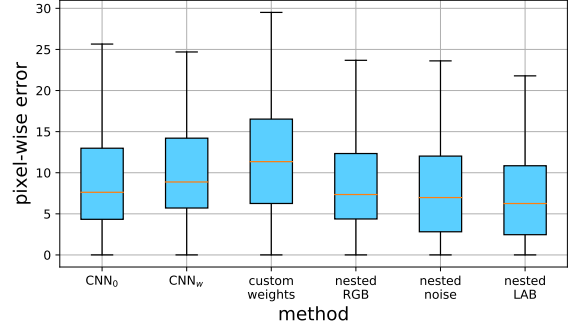


Figure 9: Pixel-wise error of the predictions on the testset of our MaxBRDF dataset. Each box depicts the results obtained using a different method. Nested LAB, described in Section 4, outperforms the rest.

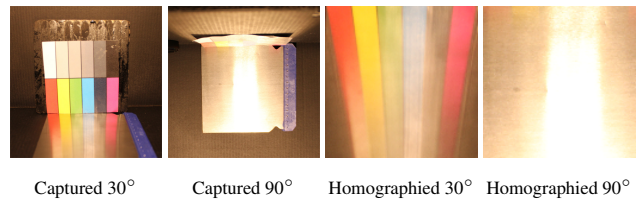


Figure 10: Sample of our test set of real materials. Using a low cost setup, we capture them from two viewpoints.

intermediate solutions. In particular, we show MSE per-pixel error for: first, our CNN_0 and CNN_w alone, described in Section 4.3. As expected, these two networks are good at extracting discriminative features of the input images, but struggle to provide accurate predictions because they have to implicitly learn parameter dependencies; we also experimented with assigning explicit weights (in Figure 9, “custom weights”) to each of the parameters to the loss function $\mathcal{L}_{total}$ defined in Equation 8, hoping that these weights would ease the learning process. However, results were unsatisfactory, probably due to the complex parameter dependency of our BRDF model; finally, we show results with our nested approach using different representations to compute the error of the reflectance parameters RGB_{90} and RGB_0 (in Figure 9, “nested RGB”, “nested noise” and “nested LAB”). We conclude that our nested approach using the LAB color space to encode RGB_{90} and RGB_0 , which includes Gaussian noise augmentation in the training set, outperforms any other method.

5.3. Qualitative Evaluation

The images of materials predicted by our method using the test set of MaxBRDF dataset are really accurate. Several images on how it generalizes to new viewpoints can be found in the supplementary material. However, our final goal is to create a method that estimates the parameters of

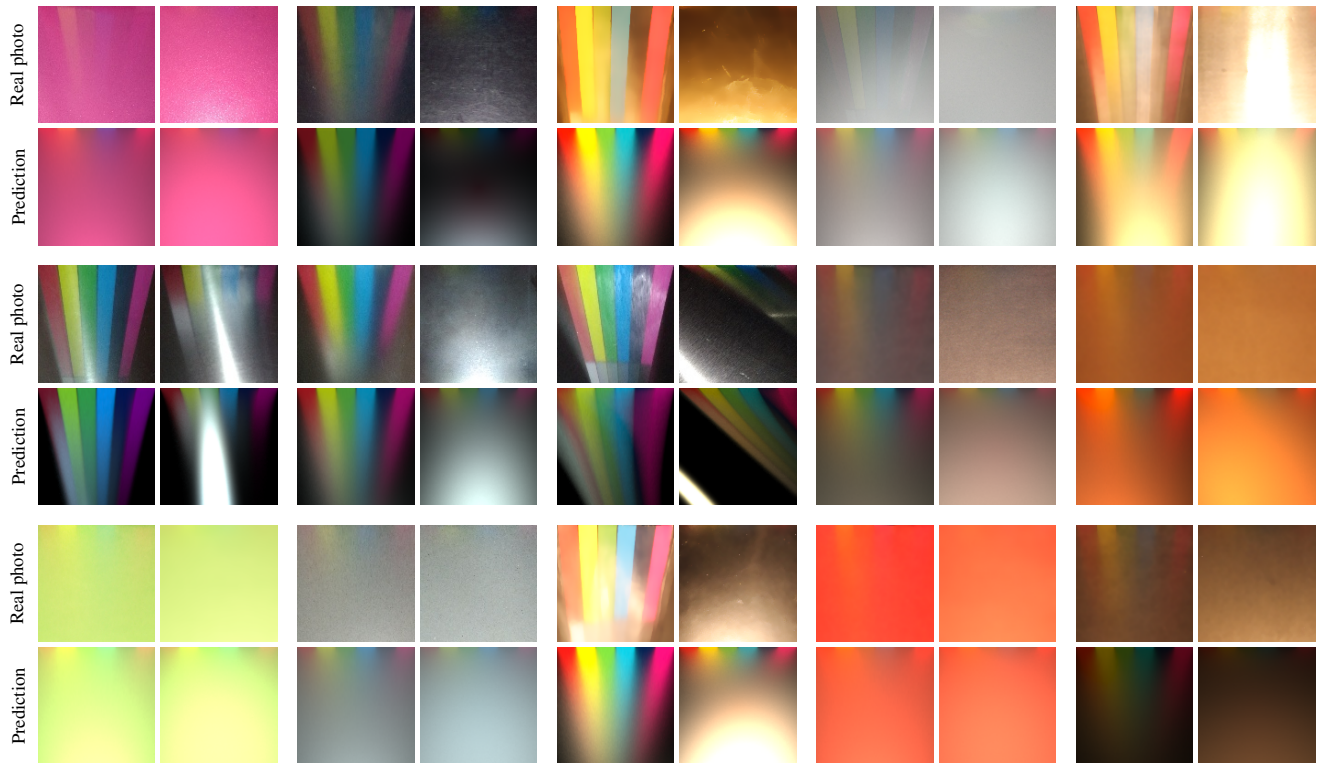


Figure 11: Qualitative evaluation on real materials. Each 4x4 block shows a pair of homographed images of a real material (captured at 30° and 90°), and the predicted material rendered in the same configuration. This results demonstrates that the proposed method is capable of capturing complex appearance behaviors such as strong highlights and anisotropies.

pictures of real materials.

Even though our new MaxBRDF dataset features a large variety of physically-based rendered materials, it does not contain any *real* image to use for testing.

We overcome this limitation by capturing a new test set featuring real-world materials in a low-cost setup, as shown in Figure 10. Capturing the ground truth properties of those materials is a challenging task, even with the help of a goniorreflectometer and BRDF fitting methods which are beyond the scope of this work. However, we can use these images to visually assess the quality of the predictions and evaluate the potential of our method for casual BRDF acquisition on the wild.

For each material, we take a photo with a handheld camera approximately at 30° and 90° in a controlled environment, where a unique light is located roughly at the same position as in the synthetic dataset. As shown in Figure 10, following our pre-processing step, each of these image pairs is homographed before being fed into our network.

Figure 11 shows 15 pairs of homographed real images of complex materials, and the corresponding renders of the predicted BRDF. We demonstrate that our method obtains highly accurate results, even for materials featuring complex anisotropic and specular behaviors.

Furthermore, in the supplementary video, we show highly realistic animations of the predicted BRDF of real materials, using a moving camera. This demonstrates that our method can be used to render novel views of complex materials captured with just two quick shots.

6. Conclusions

In this paper, we have presented the first method to estimate the complex appearance of anisotropic materials using a novel architecture based on nested neural networks. We have shown that given only two views of a material sample at 30° and 90° degrees, our model can extrapolate the estimations to the remaining fourteen views of our dataset. We have shown the performance of our method with quantitative and qualitative experiments on synthetic and real data as input, providing compelling results in both scenarios. We have also presented the first method to achieve non-convex parameter fitting by using a hybrid approach between feature learning with convolutional neural networks and classification-based regression using a nested fully connected architecture. We believe we are the first to provide such kind of solution to the problem outperforming end-to-end baseline architectures.

References

- [1] Maxwell renderer. <http://www.nextlimit.com/maxwell/>. Accessed: 2019-05-28. 1, 3
- [2] M. Aittala, T. Aila, and J. Lehtinen. Reflectance modeling by neural texture synthesis. *ACM Transactions on Graphics (TOG)*, 35(4):65, 2016. 2
- [3] M. Aittala, T. Weyrich, and J. Lehtinen. Practical SVBRDF Capture in the Frequency Domain. *ACM Transactions on Graphics (TOG)*, 32(4), 2013. 2
- [4] M. Aittala, T. Weyrich, and J. Lehtinen. Two-shot SVBRDF Capture for Stationary Materials. *ACM Transactions on Graphics (TOG)*, 34(4), 2015. 1, 2, 3
- [5] M. Ashikhmin and P. Shirley. An anisotropic phong brdf model. *J. Graph. Tools*, 5(2):25–32, Feb. 2000. 3, 4
- [6] S. Bell, P. Upchurch, N. Snavely, and K. Bala. Opensurfaces: A richly annotated catalog of surface appearance. *ACM Transactions on Graphics (TOG)*, 32(4):111, 2013. 3
- [7] J. F. Blinn. Models of light reflection for computer synthesized pictures. In *ACM SIGGRAPH Computer Graphics*, volume 11, pages 192–198. ACM, 1977. 3
- [8] N. Bonneel, G. Peyré, and M. Cuturi. Wasserstein Barycentric Coordinates: Histogram Regression Using Optimal Transport. *ACM Transactions on Graphics*, 35(4), Apr. 2016. 2
- [9] A. Brady, J. Lawrence, P. Peers, and W. Weimer. genbrdf: Discovering new analytic brdfs with genetic programming. *ACM Trans. Graph.*, 33(4):114:1–114:11, jul 2014. 1, 2
- [10] R. Caruana. Multitask learning. In *Learning to learn*, pages 95–133. Springer, 1998. 6
- [11] J. Dupuy, E. Heitz, J.-C. Iehl, P. Poulin, and V. Ostromoukhov. Extracting microfacet-based brdf parameters from arbitrary materials with power iterations. In *Computer Graphics Forum*, volume 34, pages 21–30. Wiley Online Library, 2015. 1, 2
- [12] J. Filip and R. Vávra. Template-Based Sampling of Anisotropic BRDFs. *Computer Graphics Forum*, 33(7):91–99, 2014. 1, 2
- [13] J. Filip, R. Vávra, M. Haindl, M. Krupika, V. Havran, et al. Brdf slices: Accurate adaptive anisotropic appearance acquisition. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1468–1473. IEEE, 2013. 1, 2
- [14] J. Filip, R. Vávra, and M. Havlíček. Effective acquisition of dense anisotropic brdf. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 2047–2052. IEEE, 2014. 1, 2
- [15] S. Georgoulis, K. Rematas, T. Ritschel, E. Gavves, M. Fritz, L. Van Gool, and T. Tuytelaars. Reflectance and Natural Illumination from Single-Material Specular Objects Using Deep Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. 2, 3
- [16] D. Guarniera, G. C. Guarniera, A. Ghosh, C. Denk, and M. Glencross. Brdf representation and acquisition. In *Computer Graphics Forum*, volume 35, pages 625–650. Wiley Online Library, 2016. 2
- [17] B. K. Horn and R. W. Sjöberg. Calculating the reflectance map. *Applied optics*, 18(11):1770–1779, 1979. 3
- [18] Z. Hui, K. Sunkavalli, J.-Y. Lee, S. Hadap, J. Wang, and A. C. Sankaranarayanan. Reflectance capture using univariate sampling of BRDFs. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2017. 2
- [19] C. Innocenti, T. Ritschel, T. Weyrich, and N. J. Mitra. Decomposing single images for layered photo retouching. In *Computer Graphics Forum (Proc. EGSR)*, volume 36, pages 15–25. Wiley Online Library, 2017. 2, 3
- [20] K. Kim, J. Gu, S. Tyree, P. Molchanov, M. Nießner, and J. Kautz. A lightweight approach for on-the-fly reflectance estimation. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2017. 2, 3, 5
- [21] E. P. LaFortune and Y. D. Willems. Using the modified phong reflectance model for physically based rendering. 1994. 2, 3
- [22] J. Lawrence, A. Ben-Artzi, C. DeCoro, W. Matusik, H. Pfister, R. Ramamoorthi, and S. Rusinkiewicz. Inverse shade trees for non-parametric material representation and editing. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 25(3), July 2006. 1, 2
- [23] X. Li, Y. Dong, P. Peers, and X. Tong. Modeling surface appearance from a single photograph using self-augmented convolutional neural networks. *ACM Transactions on Graphics (TOG)*, 36(4):45, 2017. 2, 3, 5
- [24] G. Liu, D. Ceylan, E. Yumer, J. Yang, and J.-M. Lien. Material editing using a physically based rendering network. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2280–2288. IEEE, 2017. 2, 3
- [25] S. Lombardi and K. Nishino. Reflectance and natural illumination from a single image. In *European Conference on Computer Vision*, pages 582–595. Springer, 2012. 3
- [26] S. Lombardi and K. Nishino. Radiometric scene decomposition: Scene reflectance, illumination, and geometry from rgb-d images. In *3D Vision (3DV), 2016 Fourth International Conference on*, pages 305–313. IEEE, 2016. 2
- [27] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Transactions on Graphics*, 22(3):759–769, July 2003. 2
- [28] A. Meka, M. Maximov, M. Zollhoefer, A. Chatterjee, H.-P. Seidel, C. Richardt, and C. Theobalt. Lime: Live intrinsic material estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2, 3, 5
- [29] A. Ngan, F. Durand, and W. Matusik. Experimental analysis of brdf models. In *Proceedings of the Sixteenth Eurographics Conference on Rendering Techniques, EGSR '05*, pages 117–126, Aire-la-Ville, Switzerland, Switzerland, 2005. Eurographics Association. 1, 2
- [30] J. B. Nielsen, H. W. Jensen, and R. Ramamoorthi. On optimal, minimal brdf sampling for reflectance acquisition. *ACM Transactions on Graphics (TOG)*, 34(6):186:1–186:11, November 2015. 3
- [31] F. Pellacini, J. A. Ferwerda, and D. P. Greenberg. Toward a psychophysically-based light reflection model for image synthesis. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 55–64. ACM Press/Addison-Wesley Publishing Co., 2000. 3
- [32] B. T. Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975. 3

- [33] R. Vávra and J. Filip. Minimal sampling for effective acquisition of anisotropic brdfs. In *Computer Graphics Forum*, volume 35, pages 299–309. Wiley Online Library, 2016. [1](#), [2](#)
- [34] G. J. Ward. Measuring and modeling anisotropic reflection. *ACM SIGGRAPH Computer Graphics*, 26(2):265–272, 1992. [2](#), [3](#)
- [35] A. Weidlich and A. Wilkie. Arbitrarily layered micro-facet surfaces. In *Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia*, pages 171–178. ACM, 2007. [1](#), [2](#)
- [36] H. Wu, Z. Wang, and K. Zhou. Simultaneous localization and appearance estimation with a consumer rgb-d camera. *IEEE transactions on visualization and computer graphics*, 22(8):2012–2023, 2016. [2](#)
- [37] Z. Xu, J. B. Nielsen, J. Yu, H. W. Jensen, and R. Ramamoorthi. Minimal brdf sampling for two-shot near-field reflectance acquisition. *ACM Transactions on Graphics (TOG)*, 35(6), 2016. [2](#)
- [38] Y. Yu and W. A. P. Smith. PVNN: A Neural Network Library for Photometric Vision. In *IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017. [2](#), [3](#)