Invertible Neural BRDF
for Object Inverse Rendering

Zhe Chen, Shohei Nobuhara, and Ko Nishino

Kyoto University, Kyoto, Japan
`zchen@vision.ist.i.kyoto-u.ac.jp` `{nob,kon}@i.kyoto-u.ac.jp`
https://vision.ist.i.kyoto-u.ac.jp

Abstract. We introduce a novel neural network-based BRDF model and a Bayesian framework for object inverse rendering, *i.e.*, joint estimation of reflectance and natural illumination from a single image of an object of known geometry. The BRDF is expressed with an invertible neural network, namely, normalizing flow, which provides the expressive power of a high-dimensional representation, computational simplicity of a compact analytical model, and physical plausibility of a real-world BRDF. We extract the latent space of real-world reflectance by conditioning this model, which directly results in a strong reflectance prior. We refer to this model as the invertible neural BRDF model (iBRDF). We also devise a deep illumination prior by leveraging the structural bias of deep neural networks. By integrating this novel BRDF model and reflectance and illumination priors in a MAP estimation formulation, we show that this joint estimation can be computed efficiently with stochastic gradient descent. We experimentally validate the accuracy of the invertible neural BRDF model on a large number of measured data and demonstrate its use in object inverse rendering on a number of synthetic and real images. The results show new ways in which deep neural networks can help solve challenging radiometric inverse problems.

Keywords: Reflectance · BRDF · Inverse Rendering · Illumination Estimation

1 Introduction

Disentangling the complex appearance of an object into its physical constituents, namely the reflectance, illumination, and geometry, can reveal rich semantic information about the object and its environment. The reflectance informs the material composition, the illumination reveals the surroundings, and the geometry makes explicit the object shape. Geometry recovery with strong assumptions on the other constituents has enjoyed a long history of research, culminating in various methods of shape-from-X. Accurate estimation of reflectance and illumination is equally critical for a broad range of applications, including augmented reality, robotics, and graphics where the problem is often referred to as inverse rendering. This paper particularly concerns inverse rendering of object appearance rather than scenes, where the latter would require modeling of global light transport in addition to the complex local surface light interaction.

Even when we assume that the geometry of the object is known or already estimated, the difficulty of joint estimation of the remaining reflectance and illumination, persists. The key challenge lies in the inherent ambiguity between the two, both in color and frequency [35]. Past methods have relied on strongly constrained representations, for instance, by employing low-dimensional parametric models, either physically-based or data-driven (*e.g.*, Lambertian and spherical harmonics, respectively). On top of these compact parametric models, strong analytical but simplistic constraints are often required to better condition the joint estimation, such as a Gaussian-mixture on the variation of real-world reflectance and gradient and entropy priors on nonparametric illumination [25].

While these methods based on low-dimensional parametric BRDF models have shown success in object inverse rendering "in the wild," the accuracy of the estimates are inevitably limited by the expressive power of the models. As also empirically shown by Lombardi and Nishino [25], the estimation accuracy of the two radiometric constituents are bounded by the highest frequency of either of the two. Although we are theoretically limited to this bound, low-dimensional parametric models further restrict the recoverable frequency characteristics to the approximation accuracy of the models themselves. Ideally, we would like to use high-dimensional representations for both the reflectance and illumination, so that the estimation accuracy is not bounded by their parametric forms. The challenge then becomes expressing complex real-world reflectance with a common high-dimensional representation while taming the variability of real-world reflectance and illumination so that they can be estimated from single images. Nonparametric (*i.e.*, tabulated) BRDF representations and regular deep generative models such as generative adversarial networks [16] and variational autoencoders [18] are unsuitable for the task as they do not lend a straightforward means for adequate sampling in the angular domain of the BRDF.

In this paper, we introduce the *invertible neural BRDF* model (*iBRDF*) for joint estimation of reflectance and illumination from a single image of object appearance. We show that this combination of an invertible, differentiable model that has the expressive power better than a nonparametric representation together with a MAP formulation with differentiable rendering enable efficient, accurate real-world object inverse rendering. We exploit the inherent structure of the reflectance by modeling its bidirectional reflectance distribution function (BRDF) as an invertible neural network, namely, a nonlinearly transformed parametric distribution based on normalized flow [29,42,43]. In sharp contrast to past methods that use low-dimensional parametric models, the deep generative neural network makes no assumptions on the underlying distribution and expresses the complex angular distributions of the BRDF with a series of non-linear transformations applied to a simple input distribution. We will show that this provides us with comparable or superior expressiveness to nonparametric representations. Moreover, the invertibility of the model ensures Helmholtz reciprocity and energy conservation, which are essential for physical plausibility. In addition, although we do not pursue in this paper, this invertibility makes iBRDF also suitable for forward rendering applications due to its bidirectional, differentiable bijec-

tive mapping. We also show that multiple "lobes" of this nonparametric BRDF model can be combined to express complex color and frequency characteristics of real-world BRDFs. Furthermore, to model the intrinsic structure of the reflectance variation of real-world materials, we condition this generative model to extract a parametric embedding space. This embedding of BRDFs in a simple parametric distribution provides us a strong prior for estimating the reflectance.

For the illumination, we employ a nonparametric representation by modeling it as a collection of point sources in the angular space (*i.e.*, equirectangular environment map). Past methods heavily relied on simplistic assumptions that can be translated into analytical constraints to tame the high-dimensional complexity associated with this nonparametric illumination representation. Instead, we constrain the illumination to represent realistic natural environments by exploiting the structural bias induced by a deep neural network (*i.e.*, deep image prior [45]). We device this deep illumination prior by encoding the illumination as the output of an encoder-decoder deep neural network and by optimizing its parameters on a fixed random image input.

We derive a Bayesian object inverse rendering framework by combining the deep illumination prior together with the invertible neural BRDF and a differentiable renderer to evaluate the likelihood. Due to the full differentiability of the BRDF and illumination models and priors, the estimation can be achieved through backpropagation with stochastic gradient descent. We demonstrate the effectiveness of our novel BRDF model, its embedding, deep illumination prior, and joint estimation on a large amount of synthetic and real data. Both the quantitative evaluation and quantitative validation show that they lead to accurate object inverse rendering of real-world materials taken under complex natural illumination.

To summarize, our technical contributions consist of

– a novel BRDF model with the expressiveness of a nonparametric representation and computational simplicity of an analytical distribution model,
– a reflectance prior based on the embedding of this novel BRDF model,
– an illumination prior leveraging architectural bias of neural networks,
– and a fully differentiable joint estimation framework for reflectance and illumination based on these novel models and priors.

We believe these contributions open new avenues of research towards fully leveraging deep neural networks in solving radiometric inverse problems.

2   Related Work

Reflectance modeling and radiometric quantity estimation from images has a long history of research in computer vision and related areas, studied under the umbrella of physics-based vision, appearance modeling, and inverse rendering. Here we briefly review works most relevant to ours.

Reflectance Models. For describing local light transport at a surface point, Nicodemus [30] introduced the bidirectional reflectance distribution function (BRDF) as a 4D reflectance function of incident and exitant light directions. Since then, many parametric reflectance models that provide an analytical expression to this abstract function have been proposed. Empirical models like Phong [34] and Blinn models [5] offer great simplicity for forward rendering, yet fail to capture complex reflectance properties of real-world materials making them unsuitable for reflectance estimation. Physically-based reflectance models such as Torrance-Sparrow [44], Cook-Torrance [8] and Disney material models [7] rigorously model the light interaction with micro-surface geometry. While these models capture challenging reflectance properties like off-specular reflection, their accuracy is limited to certain types of materials.

Data-driven reflectance models instead directly model the BRDF by fitting basis functions (*e.g.*, Zernike polynomials [19] and spherical harmonics [4, 35] or by extracting such bases from measured data [27]. Nishino *et al.* introduce the directional statistics BRDF model [31, 32] based on a newly derived hemispherical exponential power distribution to express BRDFs in half-way vector representations and use their embedding as a prior for various inverse-rendering tasks [24, 25, 33]. Ashikhmin and Premoze [2] use a modified anisotropic Phong fit to measured data. The expressive power of these models are restricted by the underlying analytical distributions. Romeiro *et al.* [37] instead directly use tabulated 2D isotropic reflectance distributions. Although nonparametric and expressive, using such models for estimating the reflectance remains challenging due to their high-dimensionality and lack of differentiability. We show that our invertible neural BRDF model with comparable number of parameters achieves higher accuracy while being differentiable and invertible.

Reflectance Estimation. Joint estimation of reflectance and illumination is challenging due to the inherent ambiguity between the two radiometric quantities. For this reason, many works estimate one of the two assuming the other is known. For reflectance estimation early work such as that by Sato *et al.* [41] assume Torrance-Sparrow reflection and a point light source. Romeiro *et al.* [37] estimate a nonparametric bivariate BRDF of an object of known geometry taken under known natural illumination. Rematas [36] propose an end-to-end neural network to estimate the reflectance map. The geometry is first estimated from the image after which a sparse reflectance map is reconstructed. A convolutional neural network (ConvNet) was learned to fill the holes of this sparse reflectance map. The illumination is, however, baked into and inseparable from this reflectance map. Meka *et al.* [28] models the reflectance as a linear combination of Lambertian, specular albedo and shininess and regress each component with deep convolutional auto-encoders. Lombardi and Nishino [23] use a learned prior for natural materials using the DSBRDF model [31] for multi-material estimation. Kang *et al.* [17] and Gao *et al.* [12] estimate spatially-varying BRDFs of planar surfaces taken under designed lighting patterns with an auto-encoder. These methods are limited to known or prescribed illumination and cannot be applied to images taken under arbitrary natural illumination.

Illumination Estimation. Assuming Lambertian surfaces, Marschner *et al.* [26] recover the illumination using image-based bases. Garon *et al.* [14] encode the illumination with 5th-order spherical harmonics and regress their coefficients with a ConvNet. Gardner *et al.* [13] represent lighting as a set of discrete 3D lights with geometric and photometric parameters and regress their coefficients with a neural network. LeGendre *et al.* [21] trained a ConvNet to directly regress a high-dynamic range illumination map from an low-dynamic range image. While there are many more recent works on illumination estimation, jointly estimating the reflectance adds another level of complexity due to the intricate surface reflection that cannot be disentangled with such methods.

Joint Estimation of Reflectance and Illumination. Romeiro *et al.* [38] use non-negative matrix factorization to extract reflectance bases and Haar wavelets to represent illumination to estimate both. Their method, however, is restricted to monochromatic reflectance estimation and cannot handle the complex inter-play of illumination and reflectance across different color channels. Lombardi and Nishino [22, 25] introduce a maximum a posterior estimation framework using the DSBRDF model and its embedding space as a reflectance prior and gradient and entropy priors on nonparametric illumination. Our Bayesian frame-work follows their formulation but overcomes the limitations induced by the rigid parametric reflectance model and artificial priors on the illumination that leads to oversmoothing. More recently, Georgoulis *et al.* [15] extend their prior work [36] to jointly estimate geometry, material and illumination. The method, however, assumes Phong BRDF which significantly restricts its applicability to real-world materials. Wang *et al.* [46] leverage image sets of objects of different materials taken under the same illumination to jointly estimate material and illumination. This requirement is unrealistic for general inverse-rendering. Yu *et al.* [48] introduce a deep neural network-based inverse rendering of outdoor im-agery. The method, however, fundamentally relies on Lambertian reflectance and low-frequency illumination expressed in spherical harmonics. Azinović *et al.* [3] introduce a differentiable Monte Carlo renderer for inverse path tracing indoor scenes. Their method employs a parametric reflectance model [7], which restricts the types of materials it can handle. While the overall differentiable estimation framework resembles ours, our work focuses on high fidelity inverse rendering of object appearance by fully leveraging a novel BRDF model and deep priors on both the reflectance and illumination. We believe our method can be integrated with such methods as [3] for scene inverse rendering in the future.

3   Differentiable Bayesian Joint Estimation

We first introduce the overall joint estimation framework. Following [25], we formulate object inverse rendering, namely joint estimation of reflectance and illumination of an object of known geometry, from a single image as maximum a posteriori (MAP) estimation

$$\underset{\mathbf{R},\mathbf{L}}{\operatorname{argmax}}\, p(\mathbf{R}, \mathbf{L}|\mathbf{I}) \propto p(\mathbf{I}|\mathbf{R}, \mathbf{L})p(\mathbf{R})p(\mathbf{L})\,, \tag{1}$$

where $\mathbf{I}$ is the RGB image, $\mathbf{R}$ is the reflectance, and $\mathbf{L}$ is the environmental illumination. We assume that the camera is calibrated both geometrically and radiometrically, and the illumination is directional (*i.e.*, it can be represented with an environment map). Our key contributions lie in devising an expressive high-dimensional yet invertible model for the reflectance $\mathbf{R}$, and employing strong realistic priors on the reflectance $p(\mathbf{R})$ and illumination $p(\mathbf{L})$.

We model the likelihood with a Laplacian distribution on the log radiance for robust estimation [25, 47]

$$p(\mathbf{I}|\mathbf{R}, \mathbf{L}) = \prod_{i,c} \frac{1}{2b_I} \exp\left[-\frac{|\log I_{i,c} - \log E_{i,c}(\mathbf{R}, \mathbf{L})|}{b_I} \right], \qquad (2)$$

where b_I controls the scale of the distribution, $I_{i,c}$ is the irradiance at pixel i in color channel c, and $E_{i,c}(\mathbf{R}, \mathbf{L})$ is the expectation of the rendered radiance.

To evaluate the likelihood, we need access to the forward rendered radiance $E_{i,c}(\mathbf{R}, \mathbf{L})$ as well as their derivatives with respect to the reflectance and illumination, $\mathbf{R}$ and $\mathbf{L}$, respectively. For this, we implement the differential path tracing method introduced by Lombardi and Nishino [24].

4   Invertible Neural BRDF

Real-world materials exhibit a large variation in their reflectance that are hard to capture with generic bases such as Zernike polynomials or analytic distributions like that in the Cook-Torrance model. Nonparametric representations, such as simple 3D tabulation of an isotropic BRDF, would better capture the wide variety while ensuring accurate representations for individual BRDFs. On the other hand, we also need to differentiate the error function Eq. 1 with respect to the reflectance and also evaluate the exact likelihood of the BRDF. For this, past methods favored low-dimensional parametric reflectance models, which have limited expressive power. We resolve this fundamental dilemma by introducing a high-dimensional parametric reflectance model based on an invertible deep generative neural network. To be precise, the BRDF is expressed as a 3D reflectance distribution which is transformed from a simple parametric distribution (*e.g.*, uniform distribution). The key property is that this transformation is a cascade of invertible linear transforms that collectively results in a complex nonparametric distribution. The resulting parameterization is high-dimensional having as many parameters as a nonparametric tabulation, thus being extremely expressive, yet differentiable and also can be efficiently sampled and evaluated.

4.1   Preliminaries

The BRDF defines point reflectance as the ratio of reflected radiance to incident irradiance given the 2D incident and exitant directions, ω_i and ω_o, respectively,

$$f_r(\omega_i, \omega_o) = \frac{\mathrm{d}L_r(\omega_o)}{L_i(\omega_i)\cos\theta_i \mathrm{d}\omega_i}, \qquad (3)$$

where $\omega_i = (\theta_i.\phi_i)$, $\omega_o = (\theta_o.\phi_o)$, $\theta \in [0, \frac{\pi}{2}]$, $\phi \in [0, 2\pi]$.

We use the halfway-vector representation introduced by Rusinkiewicz [40]

$$\omega_h = \frac{\omega_i + \omega_o}{\|\omega_i + \omega_o\|} \,,\ \ \omega_d = \mathsf{R}_{\mathbf{b}}(-\theta_h)\mathsf{R}_{\mathbf{n}}(-\phi_h)\omega_i \,, \tag{4}$$

where $\mathsf{R}_{\mathbf{a}}$ denotes a rotation matrix around a 3D vector $\mathbf{a}$, and $\mathbf{b}$ and $\mathbf{n}$ are the binormal and normal vectors, respectively. The BRDF is then a function of the difference vector and the halfway vector $f_r(\theta_h, \phi_h, \theta_d, \phi_d)$. We assume isotropic BRDFs, for which ϕ_h can be dropped: $f_r(\theta_h, \theta_d, \phi_d)$. Note that f_r is a 3D function that returns a 3D vector of RGB radiance when given a unit irradiance.

4.2   BRDF As An Invertible Neural Network

Our goal in representing the reflectance of real-world materials is to derive an expressive BRDF model particularly suitable for estimating it from an image. Given that deep neural networks lend us complex yet differentiable functions, a naive approach would be to express the (isotropic) BRDF with a neural network. Such a simplistic approach will, however, break down for a number of reasons.

For one, materials with similar reflectance properties (*i.e.*, kurtosis of angular distribution) but with different magnitudes will have to be represented independently. This limitation can be overcome by leveraging the fact that generally each color channel of the same material shares similar reflectance properties. We encode the BRDF as the product of a normalized base distribution p_r and a color vector $\mathbf{c} = \{r, g, b\}$

$$f_r(\theta_h, \theta_d, \phi_d) = p_r(\theta_h, \theta_d, \phi_d)\mathbf{c} \,. \tag{5}$$

The color vector is left as a latent parameter and the base distribution is fit with density estimation during training as we explain later. This way, we can represent materials that only differ in reflectance intensity and base colors but have the same distribution properties by sharing the same base distribution. This separation of base color and distribution also lets us model each BRDF with superpositions of f_r, *i.e.*, employ multiple "lobes." This is particularly important when estimating reflectance, as real-world materials usually require at least two colored distributions due to their neutral interface reflection property (*e.g.*, diffuse and illumination colors) [20].

The bigger problem with simply using an arbitrary deep neural network to represent the BRDF is their lack of an efficient means to sample (e.g., GANs) and restricting latent parametric distribution (e.g., VAEs). For this, we turn to normalizing flow models [42, 43] which is one of a family of so-called invertible neural networks [1].

In particular, we represent the normalized base distribution using an extended Non-linear Independent Components Estimation (NICE) model [11, 29]. Given a tractable latent (*i.e.*, input) distribution $q(\mathbf{x})$ and a sequence of bijective transformations $f = f_1 \circ f_2 \circ \cdots \circ f_N$, we can get a new distribution $p_r(\mathbf{y})$ by
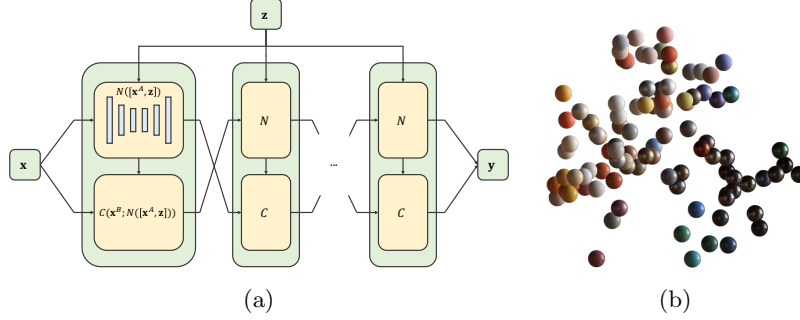
Fig. 1: (a) Architecture of the invertible neural BRDF model. The base distribution (p_r in Eq. 5) is represented with a normalizing flow [11,29], which transforms the input 3D uniform distribution $q(\mathbf{x})$ into a BRDF $p_r(\mathbf{y})$ through a cascade of bijective transformations. (b) We condition the parameters with a code $\mathbf{z}$, which let's us learn an embedding space of real-world BRDFs (shown in 2D with PCA). Similar materials are grouped together and arranged in an intuitive manner in this continuous latent space which can directly be used as a reflectance prior.

applying transformations $\mathbf{y} = f(\mathbf{x}; \mathbf{\Theta})$. Since f is bijective, under the change of variable formula, $q(\mathbf{x})$ and $p(\mathbf{y})$ are linked with

$$p_r(\mathbf{y}) = q(\mathbf{x}) \left| \det \left(\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\mathbf{y}} \right) \right| = q(f^{-1}(\mathbf{y}; \mathbf{\Theta})) \left| \det \left(\frac{\mathrm{d}f^{-1}(\mathbf{y}; \mathbf{\Theta})}{\mathrm{d}\mathbf{y}} \right) \right|, \qquad (6)$$

where $\left| \det(\frac{\mathrm{d}f^{-1}(\mathbf{y};\mathbf{\Theta})}{\mathrm{d}\mathbf{y}}) \right|$ is the absolute value of the Jacobian determinant. Such a sequence of invertible transformations f is called a normalizing flow. As long as f is complex enough, we can get an arbitrarily complicated $p_r(\mathbf{y})$ in theory. As shown in Fig. 1(a), for each layer f, we use the coupling transformation family C proposed in NICE and each layer is parameterized with a UNet N. The input $\mathbf{x}$ is split into two parts $\mathbf{x}^A$ and $\mathbf{x}^B$, where $\mathbf{x}^A$ is left unchanged and fed into N to produce the parameters of the transformation C that is applied to $\mathbf{x}^B$. Then $\mathbf{x}^A$ and $C(\mathbf{x}^B; N(\mathbf{x}^A))$ are concatenated to give the output. For the input latent distribution $q(\mathbf{x})$, we use a simple 3D uniform distribution.

There are some practical concerns to address before we can train this invertible neural BRDF on measured data. Since the base distribution $p_r(\theta_h, \theta_d, \phi_d)$ of a BRDF inherently has a finite domain ($\theta_h \in [0, \frac{\pi}{2})$, $\theta_d \in [0, \frac{\pi}{2})$, $\phi_d \in [0, \pi)$), it will be easier to learn a transformation mapping from a tractable distribution with the same domain to it. Thus we make $q(\theta_h, \theta_d, \phi_d) \sim \mathcal{U}^3(0, 1)$ and normalize each dimension of $p_r(\theta_h, \theta_d, \phi_d)$ to be in $[0, 1)$ before training. Then we adopt the piecewise-quadratic coupling transformation [29] as C to ensure that the transformed distribution has the same domain. This corresponds to a set of learnable monotonically increasing quadratic curve mappings from $[0, 1)$ to $[0, 1)$ each for one dimension.

4.3   Latent Space of Reflectance

Expressing the rich variability of real-world BRDFs concisely but with a differentiable expression is required to devise a strong constraint (*i.e.*, prior) on the otherwise ill-posed estimation problem (Eq. 1). As a general approach, such a prior can be derived by embedding the BRDFs, for instance, the 100 measured BRDFs in the MERL database [27], in a low-dimensional parametric space. Past related works have achieved this by modeling the latent space in the parameter space of the BRDF model (*e.g.*, linear subspace in the DSBRDF parameter space [25]). More recent works [12, 17] train an auto-encoder on the spatial map of analytical BRDF models to model spatially-varying BRDF. The latent space together with the decoder of a trained auto-encoder is then used for material estimation. Our focus is instead on extracting a tight latent space various real-world materials span in the nonparametric space of invertible neural BRDFs.

We achieve this by conditioning the invertible neural BRDF on a latent vector $\mathbf{z}$ (Fig. 1(a)) . We refer to this vector as the embedding code to avoid confusion with the latent distribution (*i.e.*, input) to the invertible BRDF. We jointly learn the parameters $\mathbf{\Theta}$ of iBRDF and its embedding code $\mathbf{z}$

$$\operatorname*{argmax}_{\mathbf{z}, \mathbf{\Theta}} \frac{1}{M} \sum_{i}^{M} \frac{1}{N} \sum_{j}^{N} \log p_r(\theta_h^{ij}, \theta_d^{ij}, \phi_d^{ij} | \mathbf{z}^i; \mathbf{\Theta}) . \tag{7}$$

Each trained embedding code in the embedding space is associated with a material in the training data (*e.g.*, one measured BRDF of the MERL database). This formulation is similar to the generative latent optimization [6] where the embedding code is directly optimized without the need of an encoder.

We treat each color channel of the training measured data as an independent distribution. Each distribution is assigned a separate embedding code which is initialized with a unit Gaussian distribution. Additionally, after each training step, we project $\mathbf{z}$ back into the unit hypersphere to encourage learning a compact latent space. This is analogous to the bottleneck structure in an auto-encoder. This conditional invertible neural BRDF is trained by maximizing the likelihood.

During inference, we fix $\mathbf{\Theta}$ and optimize the embedding code $\mathbf{z}$ to estimate the BRDF. In practice, we set the dimension of $\mathbf{z}$ to 16. In other words, the invertible neural BRDFs of real-world materials are embedding in a 16D linear subspace. Figure 1(b) shows the embedding of the 100 measured BRDFs in the MERL database after training. Materials with similar properties such as glossiness lie near each other forming an intuitive and physically-plausible latent space that can directly be used as a strong reflectance prior. Since all materials are constrained within the unit hypersphere during training, we thus define $p(\mathbf{R}) \sim \mathcal{N}(0, \sigma^2 I)$.

5   Deep Illumination Prior

Incident illumination to the object surface can be represented in various forms. With the same reason that motivated the derivation of iBRDF, we should avoid

unnecessary approximation errors inherent to models and represent the illumination as a nonparametric distribution. This can be achieved by simply using a latitude-longitude panoramic HDR image $\mathbf{L}(\theta, \phi)$ as an environment map to represent the illumination [10], which we refer to as the illumination map.

The expressive power of a nonparametric illumination representation, however, comes with increased complexity. Even a $1°$–sampled RGB illumination map (360×180 in pixels) would have 194400 parameters in total. To make matters worse, gradients computed in the optimization are sparse and noisy and would not sufficiently constrain such a high degree of freedom. Past work have mitigated this problem by imposing strong analytical priors on the illumination $p(\mathbf{L})$, such as sparse gradients and low entropy [25] so that they themselves are differentible. These analytical priors, however, do not necessarily capture the properties of natural environmental illumination and often lead to overly simplified illumination estimates.

We instead directly constrain the illumination maps to be "natural" by leveraging the structural bias of deep neural networks. Ulyanov *et al.* [45] found that the structure of a deep neural network has the characteristic of impeding noise and, in turn, represent natural images without the need of pre-training. That is, an untrained ConvNet can directly be used as a natural image prior. We adopt this idea of a deep image prior and design a deep neural network for use as a deep illumination prior.

We redefine the illumination map as $\mathbf{L} = g(\theta, \phi; \mathbf{\Phi})$, where $\mathbf{\Phi}$ denotes the parameters of a deep neural network g. We estimate the illumination map as an "image" generated by the deep neural network g with a fixed random noise image as input by optimizing the network parameters $\mathbf{\Phi}$. The posterior (Eq. 1) becomes

$$\underset{\mathbf{R}, \mathbf{\Phi}}{\operatorname{argmax}} \, p(\mathbf{R}, \mathbf{L}|\mathbf{I}) \propto p(\mathbf{I}|\mathbf{R}, g(\theta, \phi; \mathbf{\Phi}))p(\mathbf{R}) \,. \qquad (8)$$

The advantages of this deep illumination prior is twofold. First, the structure of a deep neural network provides impedance to noise introduced by the differentiable renderer and stochastic gradient descent through the optimization. Second, angular samples of the illumination map are now interdependent through the network, which mitigates the problem of sparse gradients. For the deep neural network g, we use a modified UNet architecture [39]. To avoid the checkerboard artifacts brought by transposed convolutional layer, we use a convolutional layer followed by a bilinear upsampling for each upsampling step. Additionally, to preserve finer details, skip connections are added to earlier layers.

6   Experimental Results

We conducted a number of experiments to validate the effectiveness of 1) the invertible neural BRDF model for representing reflectance of real-world materials; 2) the conditional invertible neural BRDF model for BRDF estimation; 3) the deep illumination prior, and 3) the Bayesian estimation framework integrated with the model and priors for single-image inverse rendering.
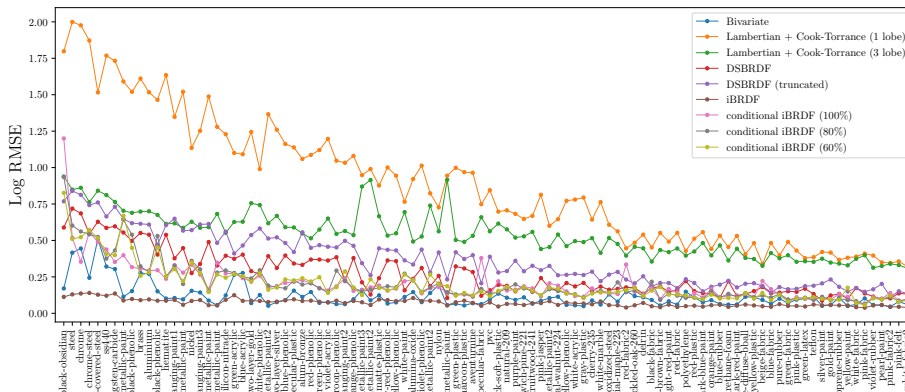
Fig. 2: Log RMSE of iBRDF and conditional iBRDF (*i.e.*, iBRDF with learned latent space) for 100 MERL materials. The iBRDF has higher accuracy than a nonparametric bivariate model. The conditional iBRDF models using 100%, 80%, and 60% of the leave-one-out training data achieves higher accuracy than other parametric models (*i.e.*, DSBRDF and Cook-Torrance). These results show the expressive power and generalizability of the invertible neural BRDF model.

6.1 Accuracy of Invertible Neural BRDF

To evaluate the accuracy of invertible neural BRDF, we learn its parameters to express measured BRDF data in the MERL database and evaluate the representation accuracy using the root mean squared error (RMSE) in log space [25]. As Fig. 2 shows, the invertible neural BRDF achieves higher accuracy than the nonparametric bivariate BRDF model. The conditional iBRDF, learned on 100%, 80%, and 60% training data all achieve high accuracy superior to other parametric models namely the DSBRDF and Cook-Torrance models. Note that all these conditional iBRDFs were trained without the test BRDF data. This resilience to varying amounts of training data demonstrates the robustness of the invertible neural BRDF model and its generalization power encoded in the learnt embedding codes. The results show that the model learns a latent space that can be used as a reflectance prior without sacrificing its expressive power.

6.2 Reflectance Estimation with iBRDF

Next, we evaluate the effectiveness of the invertible neural BRDF for single-image BRDF estimation. Unlike directly fitting to measured BRDF data, the BRDF is only partially observed in the input image. Even worse, the differentiable path tracer inevitably adds noise to the estimation process. The results of these experiments tell us how well the invertible neural BRDF can extrapolate
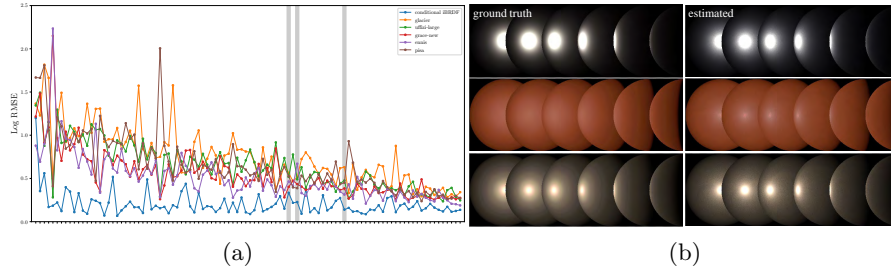
Fig. 3: (a) Log RMSE of iBRDF estimation from single images of a sphere with different materials under different known natural illumination. (b) Renderings of three samples of ground truth and estimated BRDFs. These results clearly demonstrate the high accuracy iBRDF can provide when estimating the full BRDF from partial angular observations under complex illumination.

unseen slices of the reflectance function given noisy supervision, which is important for reducing ambiguity between reflectance and illumination in their joint estimation.

We evaluate the accuracy of BRDF estimation for each of the 100 different materials in the MERL database rendered under 5 different known natural illuminations each taken from Debevec's HDR environment map set [9]. The BRDF for each material was represented with the conditional iBRDF model trained on all the measured data expect the one to be estimated (*i.e.*, 100% conditional iBRDF in Sec. 6.1). Fig. 3(a) shows the log-space RMS error for each of the combinations of BRDF and natural illumination. The results show that the BRDF can be estimated accurately regardless of the surrounding illumination. Fig. 3(b) shows spheres rendered with different point source directions using the recovered BRDF. The results match the ground truth measured BRDF well, demonstrating the ability of iBRDF to robustly recover the full reflectance from partial angular observations in the input image.

6.3   Illumination Estimation with Deep Illumination Prior

We examine the effectiveness of the deep illumination prior by evaluating the accuracy of illumination estimation with known reflectance and geometry. We rendered images of spheres with 5 different BRDFs sampled from the MERL database under 10 different natural illuminations. Fig. 4 shows samples of the ground truth and estimated illumination without and with the deep illumination. These results clearly demonstrate the effectiveness of the proposed deep illumination prior which lends strong constraints for tackling joint estimation.

6.4   Joint Estimation of Reflectance and Illumination

We integrate iBRDF and its latent space as a reflectance prior together with the deep illumination prior into the MAP estimation framework (Eq. 1) for object

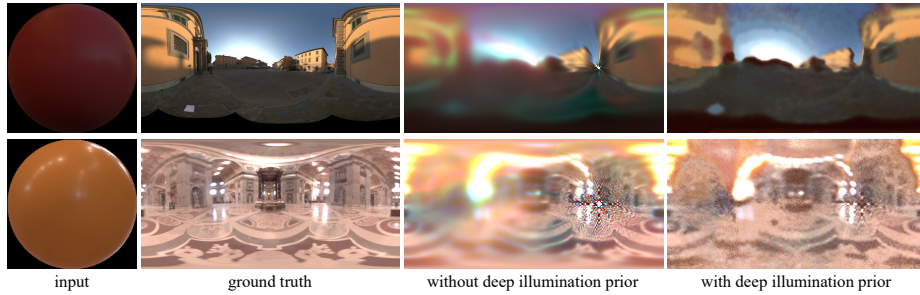| input | ground truth | without deep illumination prior | with deep illumination prior |

Fig. 4: Results of illumination estimation without and with the deep illumination prior. The illumination estimates using the prior clearly shows higher details that match those of the ground truth.

inverse rendering and systematically evaluate their effectiveness as a whole with synthetic and real images. First, we synthesized a total of 100 images of spheres rendered with 20 different measured BRDFs sampled from the MERL database under 5 different environment maps. Fig. 5(a) shows some of the estimation results. Qualitatively, the recovered BRDF and illumination match the ground truth well, demonstrating the effectiveness of iBRDF and priors for object inverse rendering. As evident in the illumination estimates, our method is able to recover high-frequency details that are not attainable in past methods. Please compare these results with, for instance, Fig. 7 of [25].

Finally, we apply our method to images of real objects taken under natural illumination. We use the Objects Under Natural Illumination Database [25]. Fig 5(b) shows the results of jointly estimating the BRDF and illumination. Our reflectance estimates are more faithful to the object appearance than those by Lombardi and Nishino, and the illumination estimates have more details (compare Fig. 5 with Fig. 10 of [25]), which collectively shows that our method more robustly disentangles the two from the object appearance. Note that the color shifts in the BRDF estimates arise from inherent color constancy, and the geometry dictates the recoverable portions of the environment. The estimates are in HDR and exposures are manually set to match as there is an ambiguity in global scaling. Please see the supplemental material for more results.

7   Conclusion

In this paper, we introduced the invertible neural BRDF model and an object inverse rendering framework that exploits its latent space as a reflectance prior and a novel deep illumination prior. Through extensive experiments on BRDF fitting, recovery, illumination estimation, and inverse rendering, we demonstrated the effectiveness of the model for representing real-world reflectance as well as its use, together with the novel priors, for jointly estimating reflectance and illumination from single images. We believe these results show new ways in which powerful
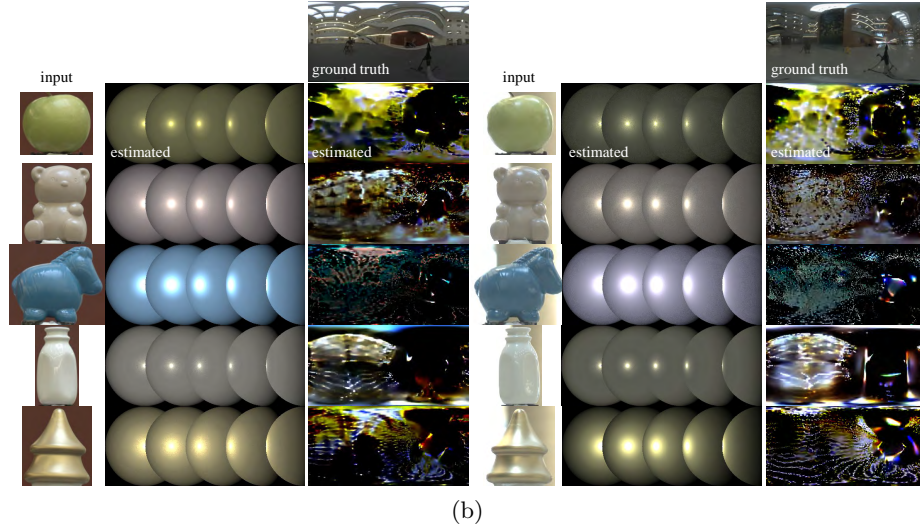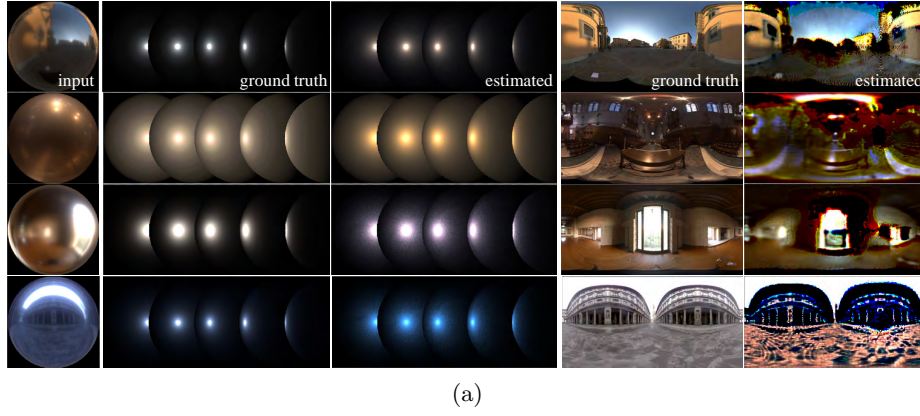
(a)

(b)

Fig. 5: Results of object inverse rendering using iBRDF and its latent space, and the deep illumination prior from (a) synthetic images and (b) real images. All results are in HDR shown with fixed exposure, which exaggerates subtle differences (*e.g.*, floor tile pattern in Uffizi). The results show that the model successfully disentangles the complex interaction of reflectance and illumination and recovers details unattainable in past methods (*e.g.*, [25]).

deep neural networks can be leveraged in solving challenging radiometric inverse and forward problems.

References

1. Ardizzone, L., Kruse, J., Wirkert, S., Rahner, D., Pellegrini, E.W., Klessen, R.S., Maier-Hein, L., Rother, C., Köthe, U.: Analyzing Inverse Problems with Invertible Neural Networks. In: International Conference on Learning Representations (2019)
2. Ashikhmin, M., Premoze, S.: Distribution-based BRDFs. Unpublished Technical Report, University of Utah **2**, 6 (2007)
3. Azinović, D., Li, T.M., Kaplanyan, A., Nießner, M.: Inverse Path Tracing for Joint Material and Lighting Estimation. In: Proc. Computer Vision and Pattern Recognition (CVPR), IEEE (2019)
4. Basri, R., Jacobs, D.: Photometric Stereo with General, Unknown Lighting. In: IEEE Int'l Conference on Computer Vision and Pattern Recognition. pp. 374–381 (2001)
5. Blinn, J.F.: Models of light reflection for computer synthesized pictures. In: Proceedings of the 4th annual conference on Computer graphics and interactive techniques. pp. 192–198 (1977)
6. Bojanowski, P., Joulin, A., Lopez-Paz, D., Szlam, A.: Optimizing the Latent Space of Generative Networks. In: International Conference on Machine Learning (2018)
7. Burley, B., Studios, W.D.A.: Physically-based Shading at Disney. In: ACM SIGGRAPH. vol. 2012, pp. 1–7 (2012)
8. Cook, R.L., Torrance, K.E.: A reflectance model for computer graphics. ACM Transactions on Graphics (TOG) **1**(1), 7–24 (1982)
9. Debevec, P.: Light Probe Gallary. http://www.debevec.org/Probes/
10. Debevec, P.: Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-based Graphics with Global Illumination and High Dynamic Range Photography. In: ACM SIGGRAPH 2008 Classes. pp. 32:1–32:10. SIGGRAPH '08, ACM, New York, NY, USA (2008). https://doi.org/10.1145/1401132.1401175, http://doi.acm.org/10.1145/1401132.1401175
11. Dinh, L., Krueger, D., Bengio, Y.: NICE: Non-linear Independent Components Estimation. CoRR (2014)
12. 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 (TOG) **38**(4), 134 (2019)
13. Gardner, M.A., Hold-Geoffroy, Y., Sunkavalli, K., Gagne, C., Lalonde, J.F.: Deep Parametric Indoor Lighting Estimation. In: The IEEE International Conference on Computer Vision (ICCV) (October 2019)
14. Garon, M., Sunkavalli, K., Hadap, S., Carr, N., Lalonde, J.F.: Fast Spatially-Varying Indoor Lighting Estimation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)
15. 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), 1932–1947 (2017)
16. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Nets. In: Advances in neural information processing systems. pp. 2672–2680 (2014)
17. Kang, K., Chen, Z., Wang, J., Zhou, K., Wu, H.: Efficient Reflectance Capture Using an Autoencoder. ACM Trans. Graph. **37**(4), 127–1 (2018)
18. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: International Conference on Learning Representations (2014)

19. Koenderink, J.J., van Doorn, A.J., Stavridi, M.: Bidirectional reflection distribution function expressed in terms of surface scattering modes. In: European Conference on Computer Vision. pp. 28–39 (1996)
20. Lee, H.C.: Illuminant Color from Shading, p. 340–347. Jones and Bartlett Publishers, Inc., USA (1992)
21. LeGendre, C., Ma, W.C., Fyffe, G., Flynn, J., Charbonnel, L., Busch, J., Debevec, P.: DeepLight: Learning Illumination for Unconstrained Mobile Mixed Reality. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5918–5928 (2019)
22. Lombardi, S., Nishino, K.: Reflectance and Natural Illumination from A Single Image. In: European Conference on Computer Vision. vol. VI, pp. 582–595 (2012)
23. Lombardi, S., Nishino, K.: Single Image Multimaterial Estimation. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 238–245. IEEE (2012)
24. Lombardi, S., Nishino, K.: Radiometric Scene Decomposition: Scene Reflectance, Illumination, and Geometry from RGB-D Images. In: International Conference on 3D Vision (2016)
25. Lombardi, S., Nishino, K.: Reflectance and Illumination Recovery in The Wild. IEEE transactions on pattern analysis and machine intelligence **38**(1), 129–141 (2016)
26. Marschner, S.R., Greenberg, D.P.: Inverse Lighting for Photography. In: Color and Imaging Conference. pp. 262–265. Society for Imaging Science and Technology (1997)
27. Matusik, W.: A data-driven reflectance model. Ph.D. thesis, Massachusetts Institute of Technology (2003)
28. Meka, A., Maximov, M., Zollhoefer, M., Chatterjee, A., Seidel, H.P., Richardt, C., Theobalt, C.: LIME: Live intrinsic material estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6315–6324 (2018)
29. Müller, T., Mcwilliams, B., Rousselle, F., Gross, M., Novák, J.: Neural Importance Sampling. ACM Trans. Graph. **38**(5), 145:1–145:19 (Oct 2019). https://doi.org/10.1145/3341156, http://doi.acm.org/10.1145/3341156
30. Nicodemus, F., Richmond, J., Hsia, J., Ginsberg, I., Limperis, T.: Geometric Considerations and Nomenclature for Reflectance. National Bureau of Standards (US) (1977)
31. Nishino, K.: Directional Statistics BRDF Model. In: 2009 IEEE 12th International Conference on Computer Vision. pp. 476–483. IEEE (2009)
32. Nishino, K., Lombardi, S.: Directional Statistics-based Reflectance Model for Isotropic Bidirectional Reflectance Distribution Functions. OSA Journal of Optical Society of America A **28**(1) (January 2011)
33. Oxholm, G., Nishino, K.: Shape and reflectance estimation in the wild. IEEE Transactions on Pattern Analysis and Machine Intelligence **38**(2), 376–389 (Feb 2016)
34. Phong, B.T.: Illumination for computer generated pictures. Communications of the ACM **18**(6), 311–317 (1975)
35. Ramamoorthi, R., Hanrahan, P.: A Signal-Processing Framework for Inverse Rendering. In: Computer Graphics Proceedings, ACM SIGGRAPH 01. pp. 117–128 (Aug 2001)
36. Rematas, K., Ritschel, T., Fritz, M., Gavves, E., Tuytelaars, T.: Deep Reflectance Maps. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4508–4516 (2016)

37. Romeiro, F., Vasilyev, Y., Zickler, T.: Passive Reflectometry. In: European Conference on Computer Vision. pp. 859–872. Springer (2008)
38. Romeiro, F., Zickler, T.: Blind Reflectometry. In: European conference on computer vision. pp. 45–58. Springer (2010)
39. Ronneberger, O., P.Fischer, Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). LNCS, vol. 9351, pp. 234–241 (2015)
40. Rusinkiewicz, S.M.: A New Change of Variables for Efficient BRDF Representation. In: Drettakis, G., Max, N. (eds.) Rendering Techniques '98. pp. 11–22. Springer Vienna, Vienna (1998)
41. Sato, Y., Wheeler, M.D., Ikeuchi, K.: Object shape and reflectance modeling from observation. In: Proceedings of SIGGRAPH'97. pp. 379–387 (1997)
42. Tabak, E.G., Turner, C.V.: A Family of Nonparametric Density Estimation Algorithms. Communications on Pure and Applied Mathematics 66(2), 145–164 (2013). https://doi.org/10.1002/cpa.21423, https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.21423
43. Tabak, E.G., Vanden-Eijnden, E.: Density estimation by dual ascent of the log-likelihood. Communications in Mathematical Sciences 8(1), 217—233 (2010)
44. Torrance, K.E., Sparrow, E.M.: Theory for off-specular reflection from roughened surfaces. Josa 57(9), 1105–1114 (1967)
45. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep Image Prior. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9446–9454 (2018)
46. Wang, T., Ritschel, T., Mitra, N.: Joint Material and Illumination Estimation from Photo Sets in the Wild. In: 2018 International Conference on 3D Vision (3DV). pp. 22–31. IEEE (2018)
47. Wilcox, R.R.: Introduction to robust estimation and hypothesis testing. Academic press (2011)
48. Yu, Y., Smith, W.A.: InverseRenderNet: Learning single image inverse rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019)