

NeuMIP: Multi-Resolution Neural Materials

ALEXANDR KUZNETSOV, University of California, San Diego, USA

KRISHNA MULLIA, Adobe Research, USA

ZEXIANG XU, Adobe Research, USA

MILOŠ HAŠAN, Adobe Research, USA

RAVI RAMAMOORTHI, University of California, San Diego, USA

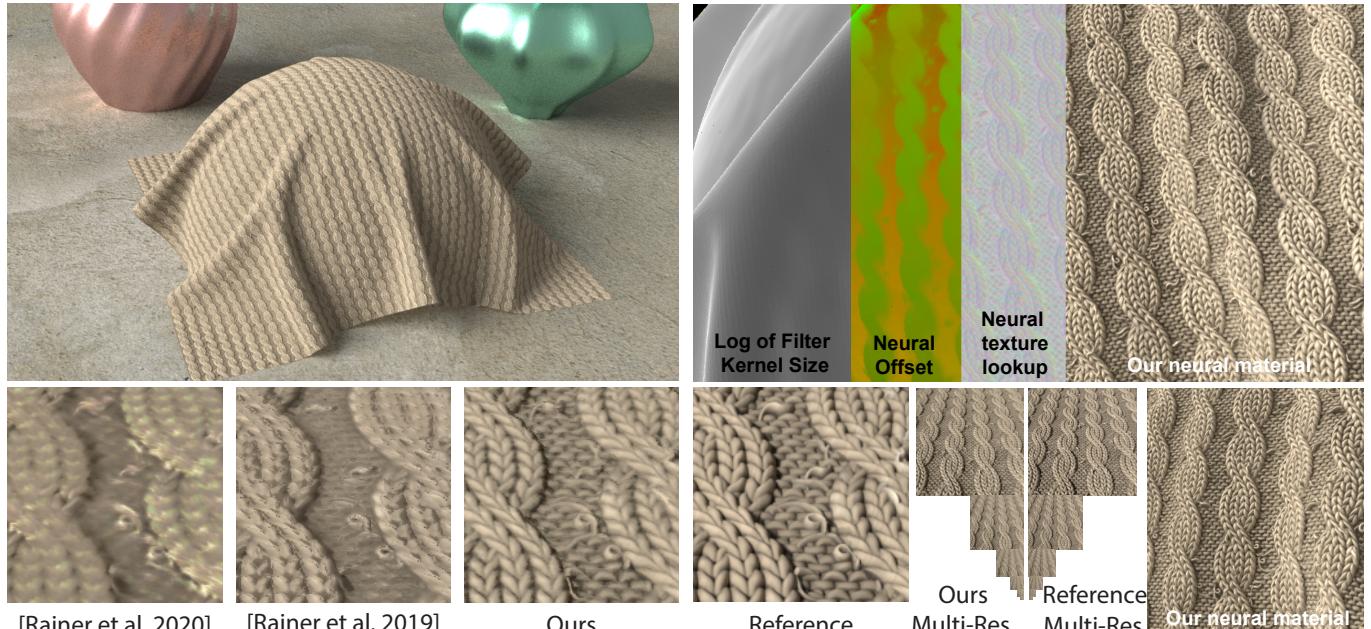


Fig. 1. Top left: Our multi-resolution neural material representing Twisted Wool, rendered seamlessly among standard materials using Monte Carlo path tracing. The neural representation is trained using hundreds of reflectance queries per texel, across multiple resolutions, and is independent of the underlying input, which could be based on displaced geometry (in this example), fiber geometry, measured data, or others. **Top right:** The stages of our pipeline: computing a kernel size based on pixel coverage, evaluating a neural offset module for improved handling of parallax effects, evaluating a neural texture pyramid to obtain a local feature vector, and applying a small fully-connected neural network to obtain a reflectance value usable in a standard renderer. **Bottom left:** Comparison of our result to previous techniques and to a reference path-traced from the ground-truth geometry. **Bottom right:** Our results match the reference across resolutions. One additional lighting / camera angle shown.

We propose NeuMIP, a neural method for representing and rendering a variety of material appearances at different scales. Classical prefiltering (mipmapping) methods work well on simple material properties such as diffuse color, but fail to generalize to normals, self-shadowing, fibers or more complex microstructures and reflectances. In this work, we generalize traditional mipmap pyramids to pyramids of neural textures, combined with

Authors' addresses: Alexandr Kuznetsov, University of California, San Diego, La Jolla, CA, USA, alkuznet@eng.ucsd.edu; Krishna Mullia, Adobe Research, San Jose, CA, USA, mulliala@adobe.com; Zexiang Xu, Adobe Research, San Jose, CA, USA, zexu@adobe.com; Miloš Hašan, Adobe Research, San Jose, CA, USA, mihasan@adobe.com; Ravi Ramamoorthi, University of California, San Diego, La Jolla, CA, USA, ravir@cs.ucsd.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2021 Copyright held by the owner/author(s).
0730-0301/2021/8-ART175
<https://doi.org/10.1145/3450626.3459795>

a fully connected network. We also introduce neural offsets, a novel method which enables rendering materials with intricate parallax effects without any tessellation. This generalizes classical parallax mapping, but is trained without supervision by any explicit heightfield. Neural materials within our system support a 7-dimensional query, including position, incoming and outgoing direction, and the desired filter kernel size. The materials have small storage (on the order of standard mipmapping except with more texture channels), and can be integrated within common Monte-Carlo path tracing systems. We demonstrate our method on a variety of materials, resulting in complex appearance across levels of detail, with accurate parallax, self-shadowing, and other effects.

CCS Concepts: • Computing methodologies → Rendering; Ray tracing; Neural networks; Image processing.

Additional Key Words and Phrases: materials, BTF, neural, mipmap, multi-resolution

ACM Reference Format:

Alexandr Kuznetsov, Krishna Mullia, Zexiang Xu, Miloš Hašan, and Ravi Ramamoorthi. 2021. NeuMIP: Multi-Resolution Neural Materials. *ACM Trans. Graph.* 40, 4, Article 175 (August 2021), 13 pages. <https://doi.org/10.1145/3450626.3459795>

1 INTRODUCTION

The world is full of materials with interesting small-scale structure: a green pasture consisting of millions of individual blades of grass, a scratched and partially rusted metallic paint on a car, a knitted sweater or a velvet dress. The underlying mesostructure and microstructure phenomena are wildly variable: complex surface height profiles, fibers and yarns, self-shadowing, multiple reflections and refractions, and subsurface scattering. Many of these effects vary at different levels of detail: for example, we can see individual fibers of a fabric when we zoom in, but they morph into yarns and eventually disappear when we zoom out.

While computer graphics has made great strides in modeling these phenomena, this is usually at the cost of large computational expense and/or loss of generality. Many previous approaches were designed for a specific material at a particular level of detail, and evaluating those methods over a large patch becomes either slow or results in artifacts. In essence, when we zoom out, we integrate over a given patch. Theoretically, this can be achieved using Monte Carlo integration techniques by evaluating a large number of samples of the patch. However, the variance of such an estimator grows with the size of the patch, and the method quickly becomes impractical, requiring large effort to compute a function that typically becomes *simpler* under zoomed-out viewing conditions. Traditional mipmap techniques [Williams 1983] can erroneously average parameters such as normals that influence the final appearance non-linearly. A universal method for prefiltering a material (that is, finding the integral of the patch of material microstructure covered by a pixel) has remained a challenge, despite some methods that address this problem for specific kinds of specular surfaces [Dupuy et al. 2013; Jakob et al. 2014] and fabrics [Zhao et al. 2016].

Our goal is to develop a neural method to accurately represent a variety of complex materials at different scales, train such a method on synthetic and real data, and integrate it into a standard path-tracing system. Our neural architecture learns a continuous variant of a bidirectional texture function (BTF) [Dana et al. 1999], which we term multi-scale BTF (MBTF). This is a 7-dimensional function, with two dimensions each for the query location, incoming and outgoing direction, and one extra dimension for the filter kernel size. This framework can represent a complex material (with self-shadowing, inter-reflections, displacements, fibers or other structure) at very different scales and can smoothly transition between them.

Inspired by the mipmaping technique, we propose NeuMIP, a method that uses a set of learned power-of-2 feature textures to represent the material at different levels, combined with a fixed per-material fully connected neural network. The network takes as input the trilinearly interpolated feature vector queried from the texture pyramid, along with incoming and outgoing directions, and outputs a reflectance value.

We also introduce a neural offset system, which allows us to efficiently represent materials with prominent non-flat geometric

features. This is achieved by adjusting the texture query location through a learned offset, resulting in a parallax effect. This allows the rendering of intricate geometry without any tessellation. We obtain the appearance of a non-flat material without the cost of constructing, storing and intersecting the displaced geometry of the material.

Because our method can represent a wide variety of materials using the same architecture, adding support for a new material becomes a simple matter of creating a dataset of random material queries and optimizing the feature textures and network weights. This typically takes only about 45 minutes on a single GPU for 512^2 resolution of the bottom pyramid level, which is easily more efficient than explicitly generating or acquiring a full high-resolution BTF. As opposed to an explicit BTF, our representation only requires storage on the order of traditional mipmapped texture pyramids (typically with 7 instead of 3 channels per texel), while enabling easy prefiltering and multiscale appearance.

Recent related works [Rainer et al. 2020, 2019] also use a neural network to efficiently compress BTFs, and we build upon these methods. However, they do not support prefiltering with arbitrary kernel sizes. They also do not have an equivalent of our neural offset technique, limiting the methods to mostly flat materials. Finally, an advantage of our method is that no encoder needs to be trained, and as a result we do not need high-resolution BTF slices as inputs, instead only requiring about 200-400 random BTF queries per texel to train a material model; our decoder network is small and fast.

The major contributions of this work are as follows:

- A neural method which can represent a wide variety of geometrically complex materials at different scales, trained from random queries of the continuous 7-dimensional multiscale BTF. These queries can come from real or synthetic data.
- A neural offset technique for rendering complex geometric appearance including parallax effects without tessellation, trained in an unsupervised manner.
- Ability to learn appearance from a small number of queries per texel (200-400), due to an encoder-less architecture.

NeuMIP can be integrated into a Monte Carlo rendering engine, since each material query can be evaluated independently, allowing for light transport between regular and neural materials as shown in Figure 1.

2 RELATED WORK

In this section, we will briefly review previous work related to material representation and the use of deep learning in rendering.

Prefiltering and mipmaping. Efficiently rendering objects at different scales is one of the fundamental problems of computer graphics. A key challenge is efficiently finding an integral of a patch of the surface of the material, which is covered by a pixel. An overview of prefiltering methods was presented by Bruneton [2011]. Williams [1983] proposed the mipmap technique to create a pyramid of pre-filtered textures; this is a standard method found in most rendering engines. However, the prefiltering problem becomes challenging if we drop the assumption of flat and rough materials. Many techniques were proposed over the years to address these shortcomings;

however, the solutions tend to be approximate and/or focus on special cases or specific materials. Becker and Max [1993] introduced a method to smoothly blend between a BRDF, bump mapping, and displacement mapping. Han et al. [2007] use spherical harmonics and spherical vMFs to prefilter normal maps. Olano and Baker [2010] introduced LEAN mapping for filtering bump maps and normal maps. Dupuy et al. [2013] improve upon LEAN mapping to prefilter displacement mapped surfaces. Kaplanyan et al. [2016] propose a real-time method for prefiltering of normal distribution functions. Wu et al. [2019] make further improvements in multi-resolution rendering of heightfield surfaces, taking into account shadowing and inter-reflection.

Parallax mapping [Kaneko et al. 2001] is a classic technique for improving bump and normal mapping by adding an approximate parallax effect. The method works by computing a texture space offset based on the local height and normal value. Other parallax-aware methods [Oliveira and Policarpo 2005; Wang et al. 2005] can also enhance the geometry details in rendering. Our neural offset technique is inspired by these methods, but is learned unsupervised; that is, we do not feed any heightfields or normals into either the training or rendering, and in fact support materials where such heightfields/normals are not precisely defined.

Bidirectional texture functions. Dana et al. [1999] introduced the notion of the bidirectional texture function (BTF), a 6D function describing arbitrary reflective surface appearance. Given 2D location coordinates, incoming and outgoing directions, the BTF outputs a reflectance value. While prefiltering BTFs by mipmapping is theoretically simple, storing a discretized 6D function requires a large amount of memory. Therefore, many methods were developed to minimize storage requirements. A common solution [Kourelka et al. 2003; Müller et al. 2003] is to use PCA or clustered PCA to compress the function. A comprehensive overview of different techniques was published by Filip and Haindl [2008].

Neural reflectance. Recently, Rainer et al. [2019] proposed to use an autoencoder framework to compress BTF slices per texel (also termed apparent BRDFs or ABRDFs); the decoder takes incoming/outgoing directions as input in addition to the latent vector, and the autoencoder is trained per BTF. Later, they extended the work by unifying different materials into a shared latent space, so only a single autoencoder needs to be trained [2020]. Other works [Maximov et al. 2019; Rematas et al. 2016] have applied CNNs to regress 2D or 4D maps that express the appearance of real objects. Within the context of complex specular appearance, [Kuznetsov et al. 2019] used Generative Adversarial Networks (GANs) to generate reflectance functions perceptually similar to synthetic or measured input data, and rendered them using partial evaluation of the generator network. Inspired by these methods, we extend the neural textures to multi-resolution materials, and introduce the neural offset module, which greatly improves the quality of non-flat materials; we also find that neither an encoder nor a discriminator is needed in our case, and direct optimization of the feature textures works well.

Other neural material methods. Yan et al. [2017] use a neural network as a mapping function to convert fur parameters into a

Table 1. Notation used in the paper.

Symbol	Domain	Definition
\mathcal{H}		Hemisphere
ω_i	\mathcal{H}	Incoming direction
ω_o	\mathcal{H}	Outgoing direction
\mathbf{u}	\mathbb{R}^2	Surface position
σ	\mathbb{R}^+	Radius of query kernel
l	\mathbb{R}	Level of detail (log. of σ)
s	\mathbb{R}	Discrete level of detail
\mathbf{v}	\mathbb{R}^c	Feature vector
ψ	\mathbb{R}^{c_2}	Feature \mathbf{v} . for neural offset
c, c_2	\mathbb{R}	Feature vec. channel num.
T_s	$\mathbb{R}^2 \rightarrow \mathbb{R}^c$	Neural texture lookup
M	$\mathbb{R}^2, \mathbb{R}^+, \mathcal{H}, \mathcal{H} \rightarrow \text{RGB}$	Multiscale BTF
P	$\mathbb{R}^2, \mathbb{R}^+ \rightarrow \mathbb{R}^c$	Neural Texture Pyramid
F	$\mathbb{R}^n, \mathcal{H}, \mathcal{H} \rightarrow \text{RGB}$	Fully connected module
O	$\mathbb{R}^2, \mathcal{H} \rightarrow \mathbb{R}^2$	Neural offset module
T_{off}	$\mathbb{R}^2 \rightarrow \mathbb{R}^{c_2}$	Neural offset texture
F_{off}	$\mathbb{R}^n, \mathcal{H} \rightarrow \mathbb{R}$	Neural offset MLP
H	$\mathcal{H}, \mathbb{R} \rightarrow \mathbb{R}^2$	Ray depth to offset

participating media, simplifying the simulation. A neural network can also be used to accelerate the rendering in an unbiased way. Mueller et al. [2019; 2020] presented path-guiding methods which learn a sampling distribution function for rendering using normalizing flow networks. Nalbach et al. [2017] proposed deep shading, a technique which uses a CNN to achieve screen-space effects like ambient occlusion, subsurface scattering etc, from simple feature buffers. Thies et al. [2019] introduced the idea of a neural texture and using it inside deferred rendering.

Recently, Mildenhall et al. [2020] introduced neural radiance fields (NeRFs), a view synthesis framework based on differentiable volume rendering, which can fit the geometry and outgoing radiance of 3D objects or entire scenes from a number of training views, encoding the entire appearance in a fully-connected network (MLP). This work has stimulated a large number of follow-up efforts, including one [Bi et al. 2020] that uses a parametric reflectance model to represent volumetric appearance with an MLP. Our problem is simpler in that we only focus on material rather than geometric shape; however, it is more complex in other ways, as we support relighting and multi-resolution viewing, and require much faster queries to integrate in a full rendering system. We do not encode the entire reflectance in a single large MLP, and instead use more structure: a feature texture pyramid and a neural offset module combined with a much smaller MLP.

3 NEURAL MBTF REPRESENTATION

In this section we define the multi-resolution BTF (Sec. 3.1), and discuss our neural architecture (Sec. 3.2 and 3.3). We introduce a baseline version of our neural MBTF in Sec. 3.2 and then present our full model in Sec. 3.3 with a neural offset technique. In later sections, we will describe how to train our models and how to use them for rendering. Our notation is summarized in Table 1.

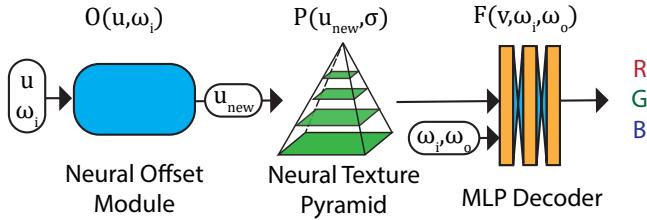


Fig. 2. Overview of our neural architecture. **Left:** The neural offset module (more detail in Fig. 4) takes a uv-space location and incoming direction, and predicts a new (offset) uv-space location to simulate parallax effects. **Middle:** The neural texture pyramid is queried using trilinear interpolation to obtain a 7-channel feature vector. **Right:** The feature vector, with incoming and outgoing directions, is fed to a material-specific multi-layer perceptron, which predicts the RGB reflectance value.

3.1 Multi-resolution BTF

Accurately computing the reflected radiance on a material surface with complex microgeometry – usually modeled by displaced geometry, fibers, volumetric scattering, etc. – is highly expensive and requires tracing complex geometry at a microscopic scale. Our approach is to precompute the reflectance of the material for a given location \mathbf{u} , radius of the footprint kernel σ , incoming direction ω_i , and outgoing direction ω_o . We denote this function $M(\mathbf{u}, \sigma, \omega_i, \omega_o)$, and call it a multi-resolution bidirectional texture function (MBTF). Below we define the MBTF more precisely.

Let $G(\mu, \sigma; \mathbf{x})$ be a normalized 2D Gaussian with mean μ and standard deviation σ , as a function of \mathbf{x} . Our MBTF is defined as follows:

$$M(\mathbf{u}, \sigma, \omega_i, \omega_o) = \int_{\mathbb{R}^2} G(\mathbf{u}, \sigma; \mathbf{x}) B(\mathbf{x}, \omega_i, \omega_o) d\mathbf{x}, \quad (1)$$

where $B(\mathbf{u}, \omega_i, \omega_o)$ can be seen as a traditional BTF at the finest material level. In other words, the MBTF value is the weighted average over the Gaussian kernel centered at \mathbf{u} of the exitant radiance in direction ω_o , assuming the material is lit by distant light of unit irradiance from direction ω_i .

Such an MBTF can be captured from real data, which we also support; however, in most of our results, we use Monte Carlo path tracing to define it using synthetic microstructure modeled on a reference plane. We compute the value $B(\mathbf{u}, \omega_i, \omega_o)$ by tracing a standard path estimator from direction $-\omega_o$ towards \mathbf{u} , assuming lighting from a distant light from direction ω_i with unit irradiance onto the reference plane. We use a distant light with a finite smoothing kernel in directions, to improve convergence of the path tracer and make multiple importance sampling applicable.

Note that our definition generalizes the standard radiometric definition of a BRDF, which is defined as outgoing radiance per unit incoming irradiance. Therefore, if the synthetic microstructure consists of a simple plane with a homogeneous BRDF, our BTF and MBTF will be equal to that BRDF for any position and kernel size. Also note that the BTF and MBTF will in general not be reciprocal due to occlusions and multiple light bounces.

3.2 Neural MBTF baseline

The 7-dimensional MBTF is too prohibitive to store directly. Nevertheless, this function (depending on the material) has redundancies,

which can be captured by a well-chosen neural network architecture. We propose a baseline neural architecture that can already model many complex materials. In particular, our baseline architecture consists of a neural texture pyramid P that encodes complex multi-scale spatially-varying material appearance and a material decoder network F that computes directional reflectance from the pyramid. This neural MBTF is expressed by

$$M(\mathbf{u}, \sigma, \omega_i, \omega_o) = F(P(\mathbf{u}, \sigma), \omega_i, \omega_o), \quad (2)$$

where $P(\mathbf{u}, \sigma)$ represents a neural feature lookup from the neural pyramid P and the decoder network F evaluates the final reflectance from the neural feature given input directions (ω_i and ω_o).

Neural texture pyramid. Instead of having one single neural texture of dimension $2^k \times 2^k \times c$ (where k is some positive integer and c the number of channels), we leverage a neural texture pyramid $P = \{T_s\}$, consisting of a set of neural textures T_s . Each T_s has a size $2^s \times 2^s \times c$, where each texel contains a c -channel latent neural feature, s denotes the discrete level of details, and we let $0 \leq s \leq k$. Such a pyramid structure enables encoding complex material appearance at multiple scales. This is similar to standard mipmapping for color (or BRDF) textures; however, our neural pyramid models challenging appearance effects caused by complex microgeometry. In addition, unlike traditional mipmaps, our per-level neural textures are independent from one another: there is no simple operation that produces all texture levels from the finest one. Each neural texture in the set is independently optimized, which ensures that the MBTF can represent appearance for the material at all levels.

We utilize trilinear interpolation to fetch a neural feature $P(\mathbf{u}, \sigma)$ at location \mathbf{u} and (continuous) level σ . Much like in classical mipmapping, we use standard bilinear interpolation to get features for the spatial dimensions of the location coordinate \mathbf{u} , followed by linear interpolation in the logarithmic space for the scale dimension, since the scales express powers of 2. Specifically, a neural feature v at (\mathbf{u}, σ) is computed by:

$$v = P(\mathbf{u}, \sigma) = w_1 T_{\lfloor l \rfloor}(\mathbf{u}) + w_2 T_{\lceil l \rceil}(\mathbf{u}), \quad (3)$$

where $l = \log_2(\sigma)$, $w_1 = (\lceil l \rceil - l)$, $w_2 = (l - \lfloor l \rfloor)$ and $\lfloor l \rfloor$, $\lceil l \rceil$ are floor and ceiling operations. $T_s(\mathbf{u})$ is the bilinearly-interpolated neural texture lookup from the level s with resolution $2^s \times 2^s$ with infinite tiling (wrap-around).

Material decoder. We design our material decoder F as a multi-layer perceptron (MLP) network to regress the final reflectance from the neural feature queried from the neural pyramid. The input of this network contains $c + 4$ values, consisting of c values from neural textures and 4 values from the directions (encoded as 2D points on the projected hemisphere). One of the design goals of our architecture is fast evaluation. Therefore, our multi-layer perceptron consists of only 4 layers. Each intermediate layer has 25 output channels. The final 3-channel output is used as the RGB reflectance value of the material, with an additional $\log(x + 1)$ compression applied to better learn high dynamic range. We use ReLU as the activation function for all layers, including the final layer where it clamps the final output to be non-negative. Note that, unlike [Thies et al. 2019] that uses a CNN to process neural textures for global context reasoning in scene rendering, our goal is modeling

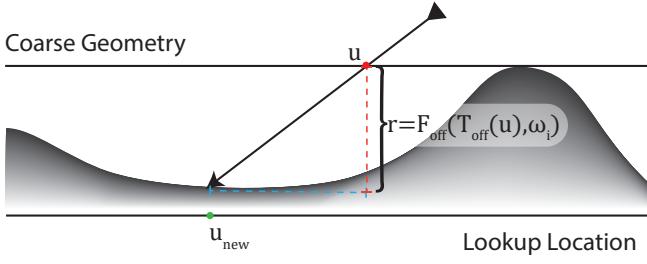


Fig. 3. An intuition for the fixed function stage of the neural offset module. Instead of predicting the uv-space offset directly, we predict the depth under the intersection with the reference plane, and convert it to the offset by assuming a locally flat surface. Note however that no actual heightfield is used to supervise the method, and for some materials (e.g. volumetric or made from fibers) such a heightfield is not even precisely defined.

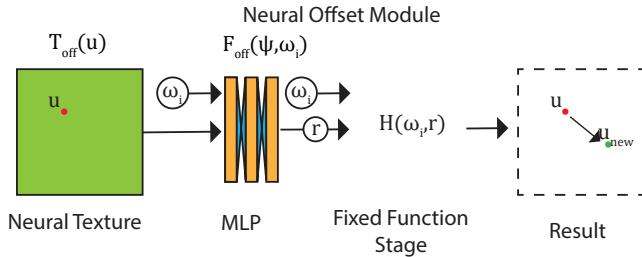


Fig. 4. Details of the neural offset module, which is shown in blue in Fig. 2. A bilinear query of the neural offset texture results in a feature vector, which (together with incoming direction) is fed into a multi-layer perceptron. The resulting scalar value is interpreted as a ray offset, which is converted into a 2D offset in texture space by the fixed function stage.

realistic material with locally complex microgeometry. Therefore, we enforce that our neural textures express the spatial information, encoding complex microgeometry effects; thus a light-weight MLP is able to efficiently decode the texture for our task and is also fast to evaluate.

3.3 Neural MBTF with Neural Offset

Our baseline neural MBTF can already model many real materials; however, it is very challenging for this network to handle highly non-flat materials that have significant parallax and occlusion effects. Although, by increasing its capacity, a big enough neural network can potentially approximate any function, this is not ideal, as bigger networks lead to longer rendering times; they also result in a slower training rate as the neural texture needs to learn correlated information across multiple pixels for different camera directions. We have also observed poor angular stability when learning non-flat materials in a generic way.

To improve results on complex non-flat materials, we introduce a neural offset module: a network that predicts a coordinate offset for the feature texture lookup. Instead of directly using the intersection location $\mathbf{u} \in \mathbb{R}^2$, we use the network module $O(\mathbf{u}, \omega_o)$ to calculate a new lookup location:

$$\mathbf{u}_{\text{new}} = \mathbf{u} + O(\mathbf{u}, \omega_o) \quad (4)$$

This is shown in Fig. 3. With the help of neural offsets, we can slightly adjust the lookup location of the texture depending on the viewing direction, achieving effects such as parallax.

Neural offset. Instead of directly regressing a 2D offset, we train a network to regress a 1D scalar value r representing the ray depth at the intersection which can be easily turned into the final 2D offset given the view direction ω_o (see Fig. 3). This makes our model more geometric-aware, easing the neural offset regression task. In particular, the neural offset module consists of 3 components: a neural offset texture T_{off} , an MLP F_{off} that regresses the ray depth r from T_{off} , and a final step H (a fixed function) that outputs the offset (See Fig. 4). The design of using a neural texture and an MLP is similar to our baseline MBTF network described above, except the texture look-up is just bilinear (no pyramid). Specifically, the ray depth r is computed by

$$r = F_{\text{off}}(\psi, \omega_o) = F_{\text{off}}(T_{\text{off}}(\mathbf{u}), \omega_o), \quad (5)$$

where $\psi = T_{\text{off}}(\mathbf{u})$ is the latent feature vector lookup in T_{off} on the initial location \mathbf{u} . The MLP F_{off} takes the latent vector ψ and the viewing direction ω_o as input; it again consists of 4 layers, and each layer outputs 25 channels (except for the last one), with ReLU activation functions in between. Given the estimated ray depth r , the offset is computed by

$$H(r, \omega_o) = \frac{r}{\max(\omega_z, .6)}(\omega_x, \omega_y), \quad (6)$$

where $\omega_x, \omega_y, \omega_z$ are the components of ω_o . Therefore, the final form of the neural offset query is:

$$\mathbf{u}_{\text{new}} = \mathbf{u} + H(r, \omega_o) = \mathbf{u} + \frac{r}{\max(\omega_z, .6)}(\omega_x, \omega_y). \quad (7)$$

The new lookup location \mathbf{u}_{new} can now be used in place of \mathbf{u} to lookup a latent vector in the neural texture pyramid in eq. 2. Note that the network can also learn the 2D offset function $O(\mathbf{u}, \omega_o)$ directly, but in our experiments the result was worse without the constraint of a hard-coded function $H(r, \omega)$. Figure 5 visualizes the predicted offset learned by the module on a highly non-flat material.

Full neural MBTF representation. Our full neural representation is modeled by prepending the the neural offset module to our baseline neural MBTF network. Basically, we use a neural offset module to get a new location \mathbf{u}_{new} (Eq. 7) by translating the original input \mathbf{u} . Then we use that \mathbf{u}_{new} to query a feature vector from the neural texture pyramid P (eq. 3). Finally, we use an MLP material decoder F , in conjunction with incoming/outgoing directions, to get the final reflectance value. Our full neural MBTF can be described as follows:

$$M(\mathbf{u}, \sigma, \omega_i, \omega_o) = F(P(\mathbf{u} + O(\mathbf{u}, \omega_o), \sigma), \omega_i, \omega_o). \quad (8)$$

This whole framework is fully differentiable, enabling end-to-end training that simultaneously optimizes the neural offset, the multi-scale pyramid, and the decoder. We train the neural offset module in an unsupervised way. We do not provide any ground truth offsets; in fact, for some materials (e.g. volumetric or fiber-based) there may be no clear surface and therefore no well-defined correct offset. The end-to-end training allows our full model to jointly auto-adjust the multiple neural components and leads to the best visual quality it can achieve. Please refer to section 4 for the training procedure.

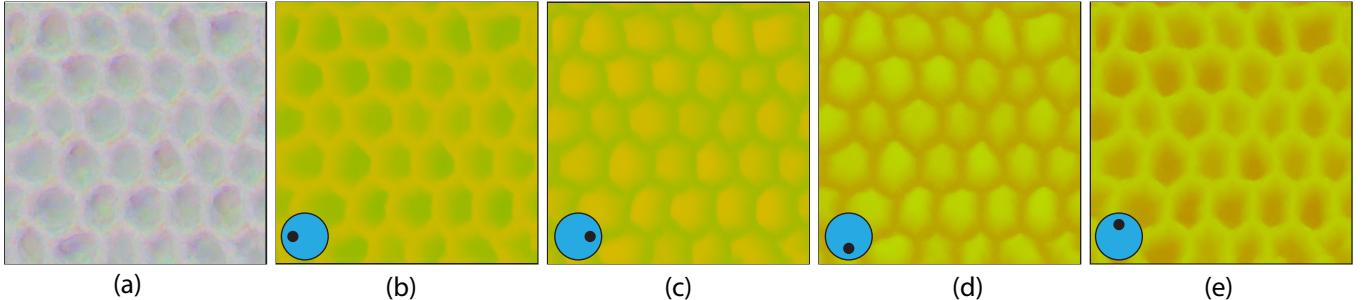


Fig. 5. Neural offset visualization on the *turtle shell* material (which is shown in Section 7 and the supplementary video). (a) The learned neural offset feature texture (3 out of 7 channels shown as RGB colors). (b-e) The computed 2D offset, color-coded using red/green for the two components. The blue circle shows the incoming ray direction on a projected hemisphere.

4 DATA GENERATION AND TRAINING

Synthetic data preparation. We generate synthetic MBTF data by first constructing the microgeometry, and using a CPU-based standard path tracer with a custom camera ray generator to render the required queries; we use smoothed directional lighting with unit irradiance on the reference plane, to ensure valid radiometric properties. In most cases, the microgeometry is constructed by meshing and displacing a high-resolution heightfield texture, and driving other reflectance properties (albedo, roughness, metallicity, micro-scale normal) from additional textures. The synthetic MBTF data is generated in about 30 minutes on a 32-core CPU, using 64 samples per query in a commercial rendering engine. In the basket weave example in Fig. 12, we use a fiber-level representation without any meshed surface, applying the fiber shading model of Chiang et al. [2016]; this material is precomputed using the PBRT renderer [Pharr et al. 2018].

Training. Our neural module is fully differentiable and can be trained end-to-end. As the input to the module we provide 7D queries consisting of the light direction, camera direction, uv-location and kernel radius. The network produces RGB colors for the given queries in a forward pass, and back-propagation updates the network weights and neural textures. One training batch consists of around a million queries (2^{20}); this number is much larger than the number of input MBTF queries per texel, so one batch updates many texels. We train the network until convergence (typically 30000 iterations). The training time is about 45 minutes for a 512^2 maximum resolution, and about 90 minutes for a 1024^2 maximum resolution, using a single NVIDIA RTX 2080 Ti GPU.

If we optimize neural feature vectors individually, this can result in noisy neural textures. As a result, objects rendered with such materials will have a noisy appearance, reminiscent of Monte Carlo noise. This is especially true for the neural offset texture. We have developed a technique to avoid those problems. During training, we apply a Gaussian blur with initial standard deviation of $\sigma_i = 8$ texels to the neural textures. As the training progresses, we relax σ exponentially over time with a half life $h = 3333$ iterations: $\sigma(t) = \sigma_i \cdot 2^{-t/h}$.

5 RENDERING

Our neural materials can be integrated into Monte Carlo rendering engines, so that light can seamlessly interact across regular objects and objects with neural materials. The results shown in the paper use an implementation on top of the Mitsuba rendering engine [Jakob 2010]. If a surface with our neural material is hit, we need to evaluate the neural module. We also need to sample an outgoing direction for indirect illumination. For simplicity, we sample indirect rays according to the cosine-weighted hemisphere, which is sufficient for our current examples. Note that for each shading point, we need to evaluate the material up to twice: for the light sample and for the BRDF sample.

There are multiple choices for implementing our neural module in a practical rendering system. To integrate with Mitsuba, our solution is to reuse the PyTorch code to ensure exactly matching outputs between training and rendering. We use material query buffers: in our integrator, we trace one bounce at a time. When encountering an intersection, we put the query in the buffer, and continue tracing. Finally, we send the whole buffer to PyTorch/GPU for batch evaluation.

We also implemented our method in an interactive path-tracer using the Nvidia OptiX framework. We re-implemented our network inference directly using the CUDA programming language (not using any external inference libraries). We call the MBTF evaluation function for each surface shading event directly, with no batching.

6 PERFORMANCE

A single evaluation of our method on a 1920×1080 image (one query for every pixel) on an RTX 2080 Ti takes only 5ms (compared to 92ms for Rainer et al. [2019]). The bulk of the time is taken by fully connected layers. Please refer to Fig. 9 for the timing breakdown of our method. The method can be divided into 2 stages. During the first stage, we perform the neural offset module and get the latent vector v , which we can cache. This is because the first stage requires only camera direction and is independent of the light direction. During the second stage, we just evaluate the fully connected network, F for a given light direction. So, if we want to evaluate the function on multiple lights, we need to reevaluate only the second stage.

Our OptiX version achieves real-time performance of 60 frames per second (one path per pixel per frame) at 1920×1080 resolution.

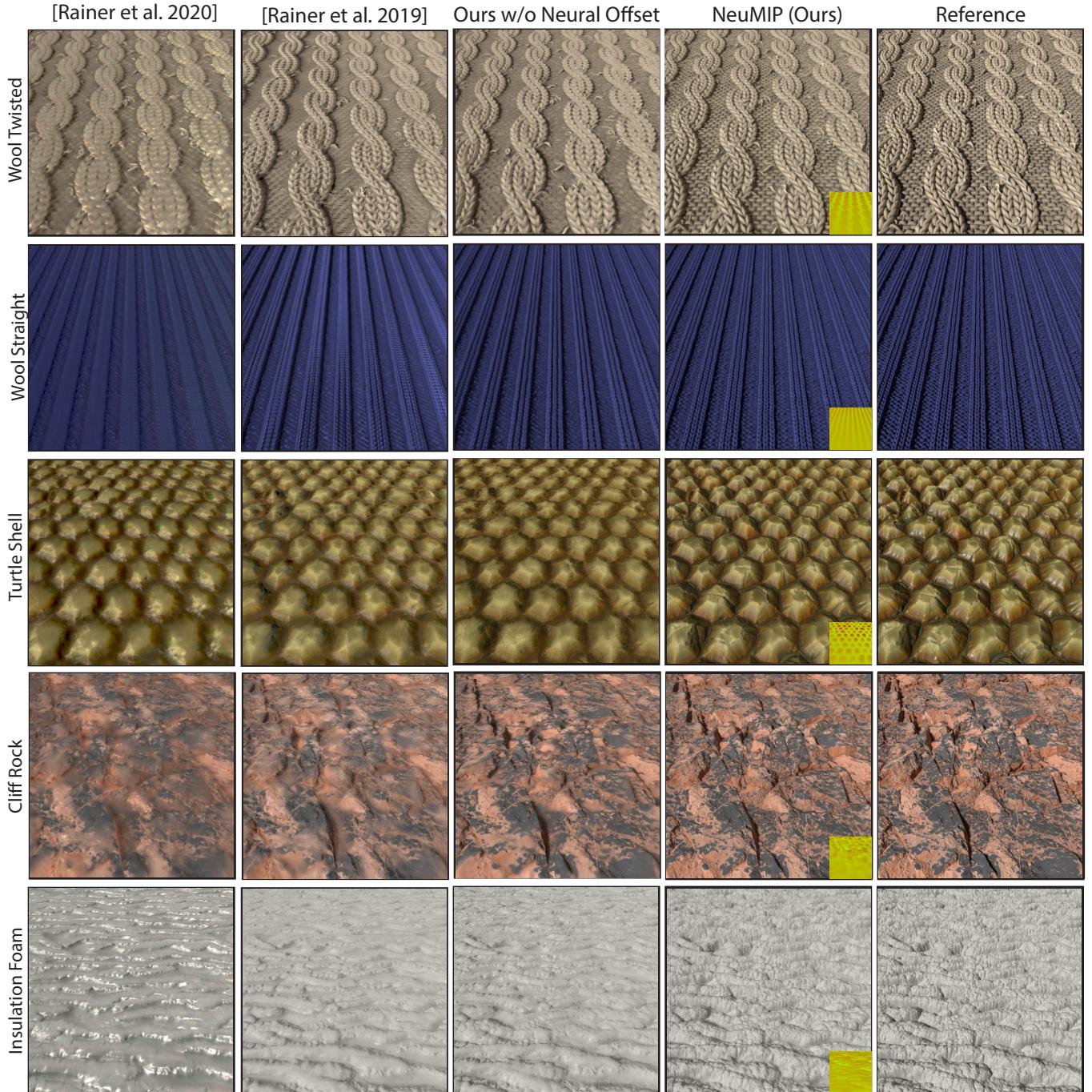


Fig. 6. **Comparison.** We show a comparison of several different methods: Rainer et al. [2020], Rainer et al. [2019], our baseline method (without neural offset), our full method, and reference computed by path-tracing of synthetic material microstructure. The materials are mapped to a plane, viewed at an angle and lit by a single light slightly to the left. Our baseline method already outperforms Rainer et al. Rainer et al. [2020] and [2019]. Our neural offset adds even better handling of parallax effects. The small insets in “Ours” show the color-coded 2D neural offset. The match with reference is close, with a minor loss in shadow contrast (the hard shadows form a reflectance discontinuity which is hard to learn perfectly).

Table 2. Errors for images in Fig. 6. Note that both our per-pixel MSE and perceptual LPIPS scores are consistently better.

	MSE $\times 10^{-3}$						LPIPS					
	Our w/o n.o.	Our	Rainer19	Rainer19/ our ratio	Rainer20	Rainer20/ our ratio	Our w/o n.o.	Our	Rainer19	Rainer19/ our ratio	Rainer20	Rainer20/ our ratio
Wool Twisted	11.31	3.98	12.79	3.21	19.03	4.78	0.333	0.151	0.376	2.49	0.501	3.32
Wool Straight	1.04	0.39	2.30	5.90	1.92	4.93	0.236	0.112	0.350	3.12	0.403	3.59
Turtle Shell	3.35	1.31	2.83	2.16	4.84	3.70	0.527	0.161	0.427	2.65	0.477	2.96
Cliff Rock	6.00	2.00	7.89	3.94	10.26	5.12	0.316	0.142	0.389	2.75	0.473	3.35
Insulation Foam	9.18	3.90	10.04	2.57	21.75	5.57	0.425	0.196	0.472	2.41	0.561	2.87

Table 3. MSE Errors (scaled $\times 10^{-3}$) for images rendered across multiple levels of detail, shown in Fig. 11. Note that the errors are generally becoming *lower* for more distant zooms, i.e. larger filter kernels.

Material LoD	0	1	2	3	4	5	6	7
Wool Twisted	3.98	2.10	1.47	1.07	0.74	1.16	1.17	1.08
Wool Straight	0.39	0.16	0.11	0.08	0.09	0.05	0.03	0.10
Turtle Shell	1.31	0.92	0.70	0.67	0.88	0.82	0.55	0.48
Cliff Rock	2.00	0.87	0.63	0.51	0.42	0.32	0.24	0.10
Insulation Foam	3.90	1.64	1.20	0.86	0.76	0.55	0.57	0.49

Table 4. The total number of network parameters. For fair comparison, we only count the number of weights in decoders for both Rainer et al.'s methods, as encoders are not needed for deployment in a rendering system. We use 7 channels per texel for both the neural offset texture and the feature texture pyramid.

	Ours	Rainer19	Rainer20
# of network weights	3332	35725	38269
Ratio of weights vs ours	N/A	10.7	11.5
# of texture channels	14	14	38

We have not used any specialized hardware like Tensor Cores, which might improve the performance further.

7 RESULTS

In this section, we showcase the abilities of our neural method to represent and render a range of material appearances, with training data coming from different input representations: displaced height-field geometry with varying reflectance properties, fiber geometry, and measured data.

Comparisons. In Figure 6, we show results rendered on a flat plane, with camera and directional light at an angle. We compare several different methods: Rainer et al. [2019], Rainer et al. [2020], our baseline method (without neural offset), our full method (with neural offset), and a ground truth computed by path-tracing of the synthetic material structure. The materials are mapped to a plane, viewed at an angle and lit by a single light slightly to the left. Our baseline method already outperforms encoders of Rainer et al. [2019] and Rainer et al. [2020] methods, despite being trained with fewer BTF queries. We believe this is due to our decoder-only architecture, which can adapt to the material and benefits from a stochastic distribution of the input BTF queries, and our improved training techniques (especially the progressively decaying spatial Gaussian blur). The multi-resolution nature of our solution also helps. On the other hand, Rainer et al. [2020]'s solution has the benefit of very fast encoding in case the queries are already in the

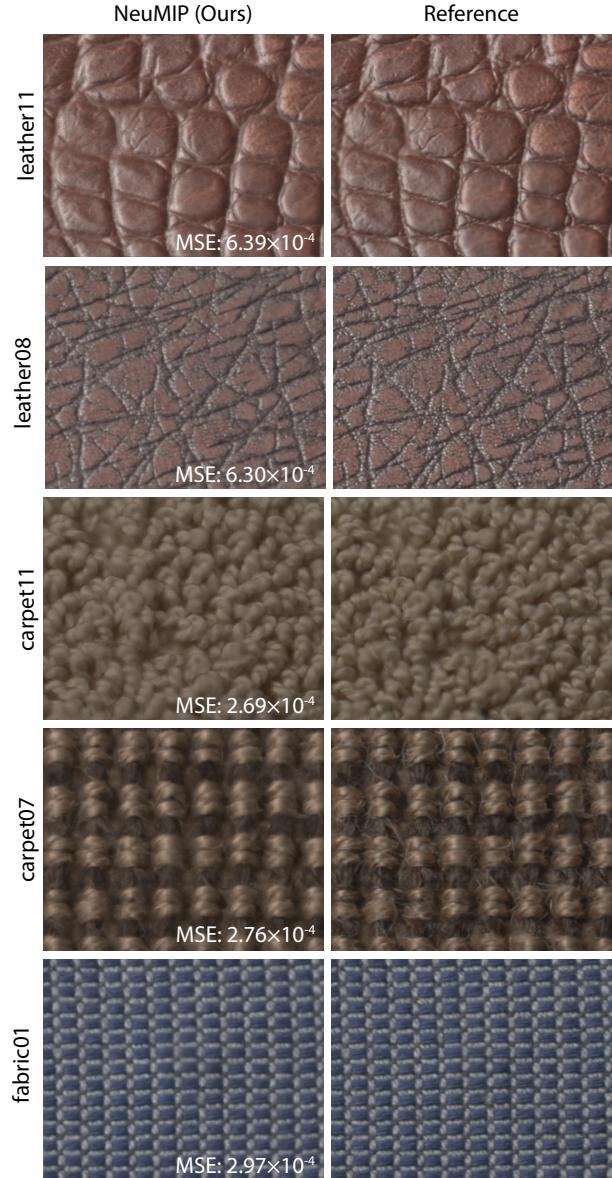


Fig. 7. **Real BTF data.** Our method is trained with a random set of BTF queries, and these can come from any source. We also observe good results when applying our method to real BTF data acquired from physical material samples.



Fig. 8. Examples of our method using real BTF data and rendered on non-flat cloth shape. Note that we use the original raw datasets, which are not tileable.

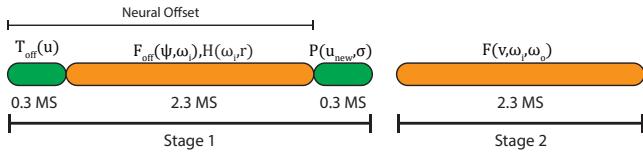


Fig. 9. Timing breakdown of the evaluation of our network on a 1920x1080 image. In total, our network takes 5ms to evaluate. The green color indicates the neural texture lookup section and the orange color indicates the MLP evaluation. Our network can be split into two stages. Stage 1 is independent of the light direction. So, for a different light direction, only Stage 2 needs to be reevaluated.

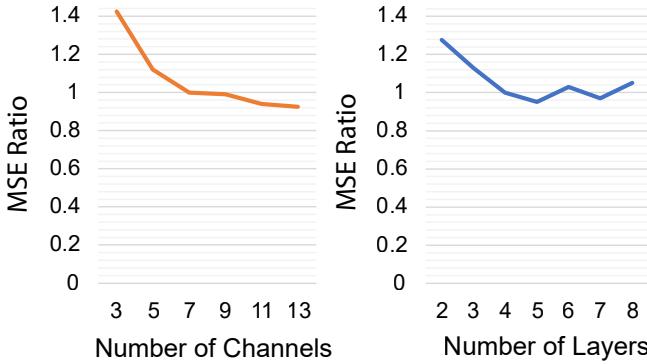


Fig. 10. Ablation studies. *Left*: the ratio of MSE errors of a given network and the default network with respect to the number of channels in two neural textures. We use two 7 channel textures in the rest of the paper. *Right*: the ratio of MSE errors of a given network and the default network with respect to the number of layers in the network. We use 4 layers in the rest of the paper.

required uniform format, and its performance could likely improve if the encoder was retrained on a different distribution of materials that matches our examples more closely.

Our neural offset adds even better handling of parallax effects on top of our baseline result. The small insets in “Ours” show the color-coded 2D neural offset. The match with reference is close, with a minor loss in shadow contrast (the hard shadows form a reflectance discontinuity which is hard to learn perfectly).

Real BTF results. Since our method is trained with a random set of BTF queries, these can come from any source. While we mostly focus on synthetically generated BTF queries, we also support fitting

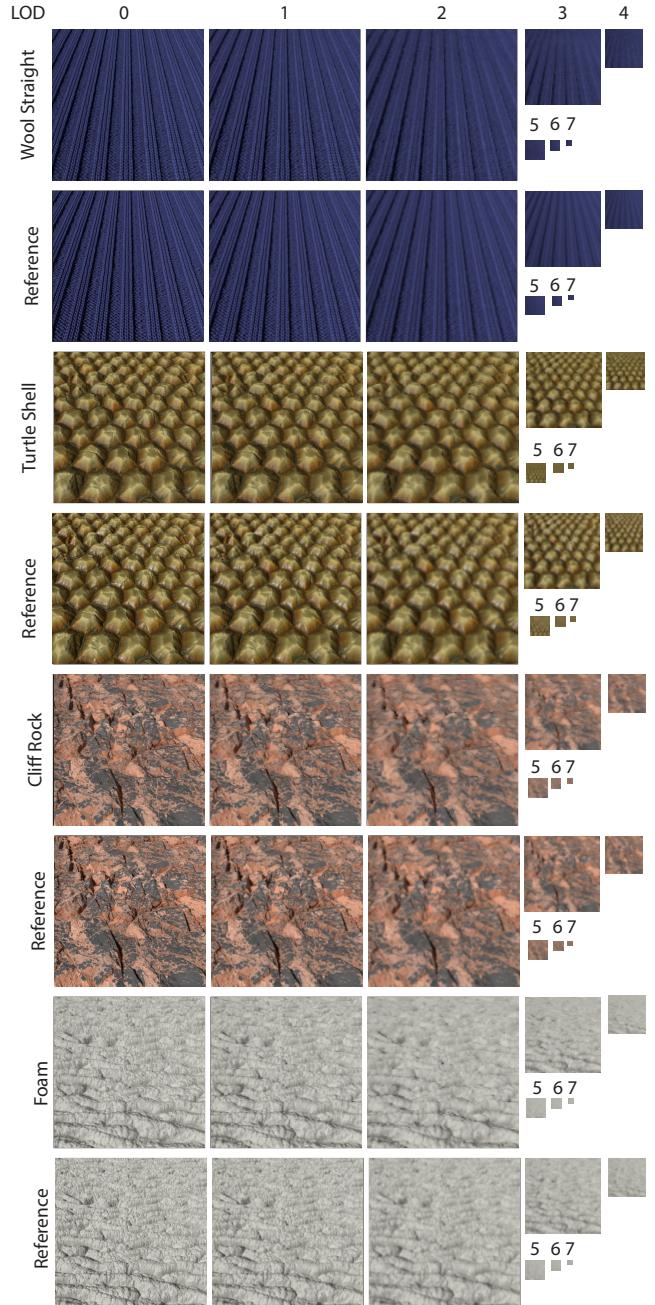


Fig. 11. Materials rendered at the different level of details. Because the camera is at a 45°, the actual LOD of the image is slightly higher at the tops of the images and slightly lower at the bottoms of the images.

real BTF data. We show this on the UBO 2014 dataset [Weinmann et al. 2014] (acquired from physical material samples). We fit the original raw data, without explicitly extracting a heightfield from it (which is essentially done implicitly by our neural offset approach). See Figure 7 for the results and the MSE errors and Figure 8 for examples of our model applied to non-flat geometry.

Quantitative evaluation. We also measure the numerical error for images in Fig. 6 when compared to the reference, both using a per-pixel MSE (Mean Squared Error) and perceptual LPIPS (Learned Perceptual Image Patch Similarity) scores. Our scores, shown in Table 2, are consistently better than both Rainer et al. [2019] and Rainer et al. [2020].

Multiresolution results. In Table 3, we also report the MSE scores for images rendered across multiple levels of detail, shown visually in Fig. 11. We observed that with higher (coarser) levels in the multiresolution hierarchy, the errors actually tend to decrease. This is because materials at a coarse level of detail tend to have fewer high-frequency details. As a result, it becomes even easier to optimize corresponding feature vectors from the neural texture pyramid and the network at those levels of detail.

Network Size. We also compare our neural network sizes to Rainer et al. [2020; 2019]’s in Table 4. Not only does our method have smaller errors, it has 10-11 times fewer weight parameters compared to the decoders in Rainer et al. [2020; 2019]. Our method also uses less than half the number of channels for the neural textures Rainer [2020]. This is because our decoder does not require universality, and because we divide our network into two stages via a neural offset module, increasing the network representative power compared to a simple MLP architecture.

Ablation Studies. While designing our architecture, we considered not only the quality of the results but the speed and the storage requirements of our network. As Fig. 10 demonstrates, the current design achieves a good trade-off. We use 2 neural textures with 7 channels each throughout the paper. Using fewer channels results in a much higher error and using more channels results in negligible improvements. Similarly, we chose to use MLPs with 4 layers. Using fewer layers results in much higher error. We also evaluated the importance of the proposed neural offset module as discussed previously (see Fig. 6).

Additional renderings. In Figure 12, we show several fabric-like materials rendered at multiple levels of zoom. The top three fabrics are modeled as heightfields with spatially varying reflectance properties. The textures driving the height and reflectance are from the Substance Source library [Adobe 2021], though any textures can be used. The last Basket Weave example is constructed from yarns using actual fiber geometry shaded with the model of Chiang et al. [2016] and does not use a heightfield. We show several camera and light directions in the right column. In Figure 13, we show a further selection of materials rendered with our method, showing different camera views and light directions. Please make sure to view their animated versions in the supplementary video, which is important to fully appreciate parallax/occlusion effects.

Limitations. One limitation of our current implementation is simple importance sampling. For more specular materials, this could be extended by fitting the parameters of a parametric pdf model (e.g. a weighted combination of a Lambertian and microfacet lobe of a given roughness) per texel. The benefit of this solution would be that no additional neural network is required to sample, and path continuations can be decided independent of network inference.

Another limitation is that very specular (e.g. glinty) materials would be hard to handle without blurring; this could be addressed by inserting randomness into our decoder, to synthesize some specular detail instead of attempting to fit it exactly.

8 CONCLUSION AND FUTURE WORK

We presented a neural architecture that can be trained to accurately represent a variety of complex materials at different scales. Our neural architecture learns a multi-scale bidirectional texture function (MBTF): a 7-dimensional function, with two dimensions each for the query location, incoming and outgoing direction, and one dimension for the filter kernel radius. As part of the architecture, we introduced a neural offset technique for rendering complex geometric appearance including parallax effects without tessellation, trained in an unsupervised manner. Our encoder-less architecture can be trained from a small number of random queries per texel (200-400). These queries can come from real or synthetic data. We show a number of results, demonstrating high quality appearance with accurate displacement, parallax, self-shadowing, and other effects. Due to our compact network architecture, we even achieve real-time performance in our OptiX implementation. We believe this approach will stimulate further research on neural representations of materials that are difficult or expensive to handle with classical methods.

The most exciting future work avenue, in our opinion, is to fully explore the set of material structures that can be procedurally generated as inputs to our method. Other interesting directions include more advanced importance sampling (e.g. by fitting multi-lobe parametric distributions per texel); the parameters of such a distribution could be stored in small additional textures. A straight-forward extension would be to support semitransparent materials by predicting alpha transparency in addition to reflectance. Yet another direction would be to make the method support glinty specular effects, perhaps by inserting an additional random vector into the decoder, and training it with a GAN loss to generate stochastic detail matching the input data distribution.

ACKNOWLEDGMENTS

We thank Xuezhang Wang for his immense help in developing the interactive OptiX renderer. We thank Sitian Chen and Steve Marschner for the fiber-level fabric microstructure data. This work was funded in part by NSF grant 1703957 and 1730158, ONR DURIP N000141912293, the Ronald L. Graham Chair and the UC San Diego Center for Visual Computing.

REFERENCES

- Adobe. 2021. Substance Source library. <https://source.substance3d.com>.
- Barry G Becker and Nelson L Max. 1993. Smooth transitions between bump rendering algorithms. In *SIGGRAPH 93*. 183–190.
- Sai Bi, Zexiang Xu, Pratul Srinivasan, Ben Mildenhall, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. 2020. Neural reflectance fields for appearance acquisition. *arXiv preprint arXiv:2008.03824* (2020).
- Eric Bruneton and Fabrice Neyret. 2011. A survey of nonlinear prefiltering methods for efficient and accurate surface shading. *IEEE Transactions on Visualization and Computer Graphics* 18, 2 (2011), 242–260.
- Matt Jen-Yuan Chiang, Benedikt Bitterli, Chuck Tappan, and Brent Burley. 2016. A Practical and Controllable Hair and Fur Model for Production Path Tracing. *Computer Graphics Forum* 35, 2 (2016), 275–283.

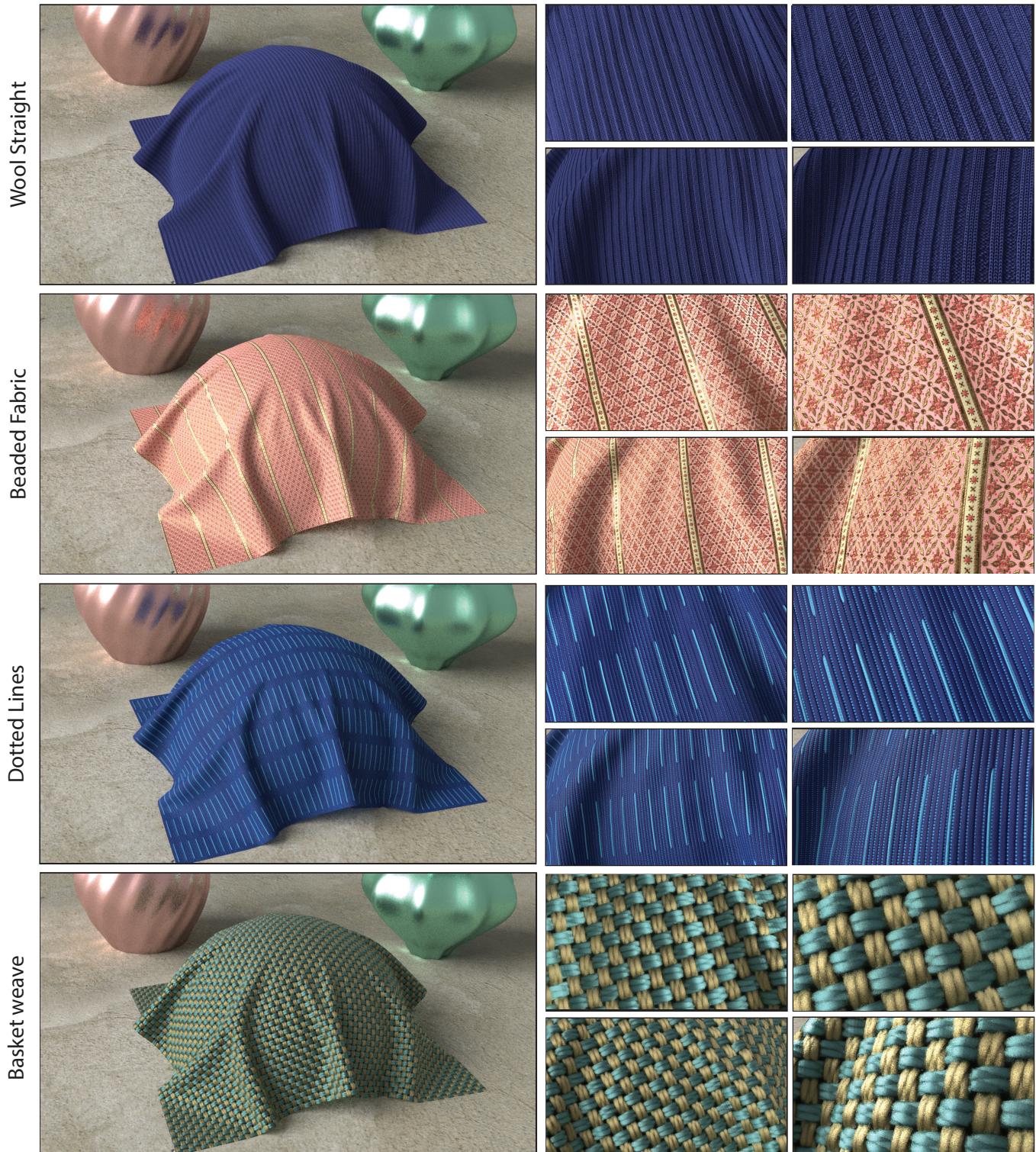


Fig. 12. **Fabric** materials rendered with our method. The fabrics are modeled as heightfields with spatially varying reflectance properties, except the last Basket Weave example, which is constructed from yarns using actual fiber geometry shaded with the model of Chiang et al. [2016] and does not use a heightfield. We show several camera, light directions, and multiple zoom levels. Please see the supplementary video for extensive animated versions.

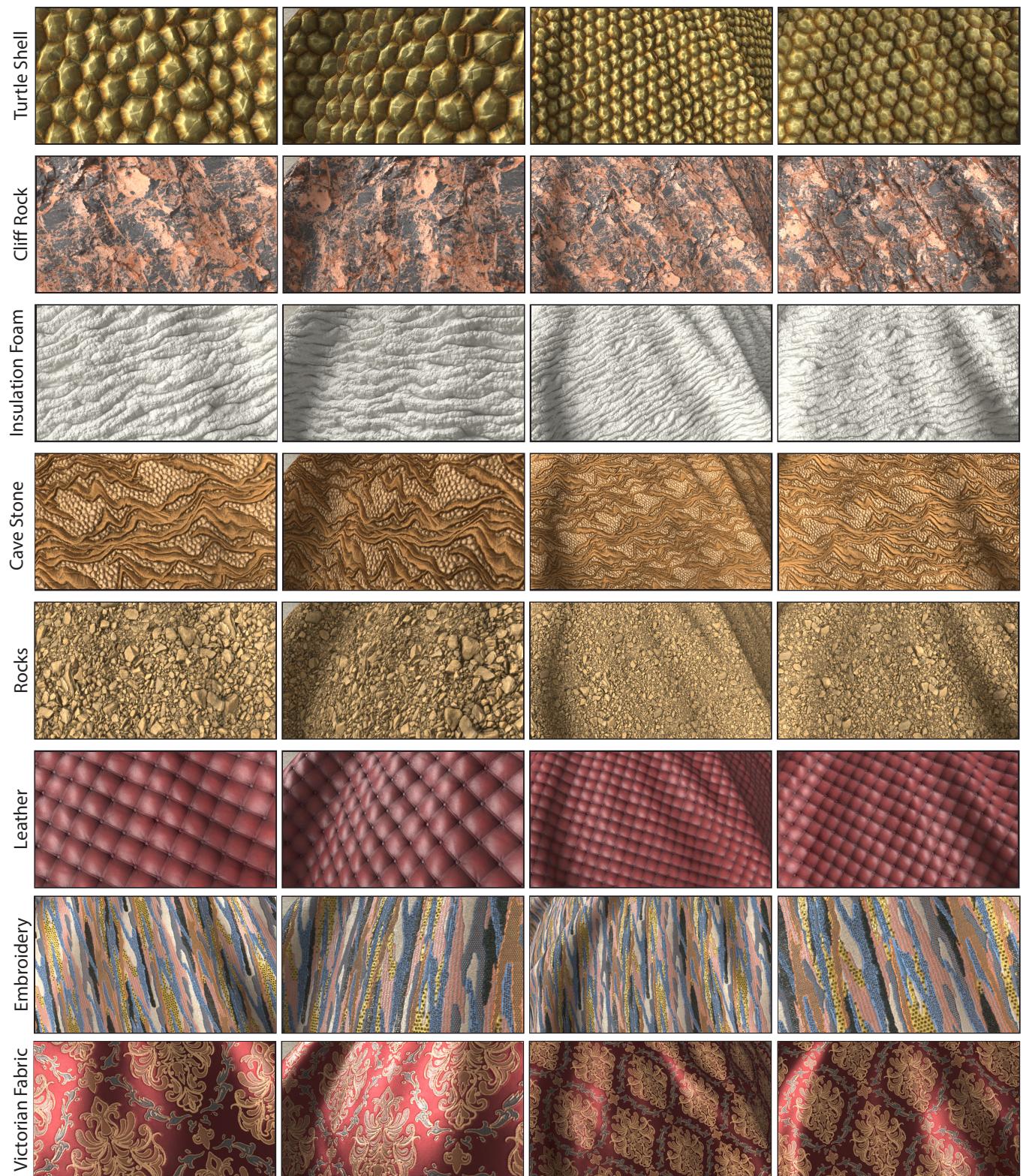


Fig. 13. An assortment of other materials rendered with our method, showing different camera views and light directions. Please make sure to view their animated versions in the supplementary video, which is important to fully appreciate parallax/occlusion effects.

Kristin J Dana, Bram Van Ginneken, Shree K Nayar, and Jan J Koenderink. 1999. Reflectance and texture of real-world surfaces. *ACM Transactions on Graphics (TOG)* 18, 1 (1999), 1–34.

Jonathan Dupuy, Eric Heitz, Jean-Claude Iehl, Pierre Poulin, Fabrice Neyret, and Victor Ostromoukhov. 2013. Linear efficient antialiased displacement and reflectance mapping. *ACM Transactions on Graphics (TOG)* 32, 6 (2013), 1–11.

Jiří Filip and Michal Haindl. 2008. Bidirectional texture function modeling: A state of the art survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 11 (2008), 1921–1940.

Charles Han, Bo Sun, Ravi Ramamoorthi, and Eitan Grinspun. 2007. Frequency Domain Normal Map Filtering. *ACM Transactions on Graphics (TOG)* 26, 3, 28.

Wenzel Jakob. 2010. Mitsuba renderer. <http://www.mitsuba-renderer.org>.

Wenzel Jakob, Miloš Hašan, Ling-Qi Yan, Jason Lawrence, Ravi Ramamoorthi, and Steve Marschner. 2014. Discrete stochastic microfacet models. *ACM Transactions on Graphics (TOG)* 33, 4 (2014), 1–10.

Tomomichi Kaneko, Toshiyuki Takahei, Masahiko Inami, Naoki Kawakami, Yasuyuki Yanagida, Taro Maeda, and Susumu Tachi. 2001. Detailed shape representation with parallax mapping. In *Proceedings of the ICAT 2001* (01 2001).

Anton S Kaplanyan, Stephan Hill, Anjul Patney, and Aaron E Lefohn. 2016. Filtering distributions of normals for shading antialiasing. In *High Performance Graphics*. 151–162.

Melissa L Koudelka, Sebastian Magda, Peter N Belhumeur, and David J Kriegman. 2003. Acquisition, compression, and synthesis of bidirectional texture functions. In *3rd International Workshop on Texture Analysis and Synthesis (Texture 2003)*. 59–64.

Alexandr Kuznetsov, Milos Hasan, Zexiang Xu, Ling-Qi Yan, Bruce Walter, Nima Khademi Kalantari, Steve Marschner, and Ravi Ramamoorthi. 2019. Learning generative models for rendering specular microgeometry. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 225.

Maxim Maximov, Laura Leal-Taixé, Mario Fritz, and Tobias Ritschel. 2019. Deep appearance maps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 8729–8738.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *ECCV 2020*.

Gero Müller, Jan Meseth, and Reinhard Klein. 2003. Compression and Real-Time Rendering of Measured BTFs Using Local PCA. In *VMV*. 271–279.

Thomas Müller, Brian McWilliams, Fabrice Rousselle, Markus Gross, and Jan Novák. 2019. Neural Importance Sampling. *ACM Transactions on Graphics (TOG)* 38, 5, Article 145 (2019), 19 pages.

Thomas Müller, Fabrice Rousselle, Alexander Keller, and Jan Novák. 2020. Neural Control Variates. *ACM Transactions on Graphics (TOG)* 39, 6, Article 243 (2020), 19 pages.

Oliver Nalbach, Elena Arabadzhyska, Dushyant Mehta, Hans-Peter Seidel, and Tobias Ritschel. 2017. Deep Shading: Convolutional Neural Networks for Screen-Space Shading. *Computer Graphics Forum (Proc. EGSR 2017)* 36, 4 (2017), 65–78.

Marc Olano and Dan Baker. 2010. Lean mapping. In *Proceedings of the 2010 ACM SIGGRAPH symposium on Interactive 3D Graphics and Games*. 181–188.

Manuel M Oliveira and Fabio Policarpo. 2005. An efficient representation for surface details. *UFRGS Technical Report* (2005).

Matt Pharr, Wenzel Jakob, and Greg Humphreys. 2018. PBRT v3. <https://github.com/mmp/pbrt-v3/>.

Gilles Rainer, Abhijeet Ghosh, Wenzel Jakob, and Tim Weyrich. 2020. Unified Neural Encoding of BTFs. *Computer Graphics Forum (Proceedings of Eurographics)* 39, 2 (2020), 167–178.

Gilles Rainer, Wenzel Jakob, Abhijeet Ghosh, and Tim Weyrich. 2019. Neural BTF Compression and Interpolation. *Computer Graphics Forum (Proceedings of Eurographics)* 38, 2 (2019), 235–244.

Konstantinos Rematas, Tobias Ritschel, Mario Fritz, Efstratios Gavves, and Tinne Tuytelaars. 2016. Deep reflectance maps. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4508–4516.

Justus Thies, Michael Zollhöfer, and Matthias Nießner. 2019. Deferred neural rendering: Image synthesis using neural textures. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–12.

Jiaping Wang, Xin Tong, John Snyder, Yanyun Chen, Baining Guo, and Heung-Yeung Shum. 2005. Capturing and rendering geometry details for BTF-mapped surfaces. *The Visual Computer* 21, 8 (2005), 559–568.

Michael Weinmann, Juergen Gall, and Reinhard Klein. 2014. Material Classification Based on Training Data Synthesized Using a BTF Database. In *ECCV 2014*. 156–171.

Lance Williams. 1983. Pyramidal parametrics. In *SIGGRAPH 83*. 1–11.

Lifan Wu, Shuang Zhao, Ling-Qi Yan, and Ravi Ramamoorthi. 2019. Accurate appearance preserving prefiltering for rendering displacement-mapped surfaces. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 137.

Ling-Qi Yan, Weilun Sun, Henrik Wann Jensen, and Ravi Ramamoorthi. 2017. A BSSRDF Model for Efficient Rendering of Fur with Global Illumination. *ACM Transactions on Graphics (TOG)* 36, 6 (2017), 208.

Shuang Zhao, Lifan Wu, Frédéric Durand, and Ravi Ramamoorthi. 2016. Downsampling scattering parameters for rendering anisotropic media. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 166.