LIGHT SAMPLING FIELD AND BRDF REPRESENTATION FOR PHYSICALLY-BASED NEURAL RENDERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Physically-based rendering (PBR) is key for immersive rendering effects used widely in the industry to showcase detailed realistic scenes from computer graphics assets. A well-known caveat is that producing the same is computationally heavy and relies on complex capture devices. Inspired by the success in quality and efficiency of recent volumetric neural rendering, we want to develop a physically-based neural shader to eliminate device dependency and significantly boost performance. However, no existing lighting and material models in the current neural rendering approaches can accurately represent the comprehensive lighting models and BRDFs properties required by the PBR process. Thus, this paper proposes a novel lighting representation that models direct and indirect light locally through light sampling strategy in a learned light sampling field. We also propose BRDF models to separately represent surface/subsurface scattering details to enable complex objects such as translucent material (i.e., skin, jade).We then practice our proposed representations with an end-to-end physically-based neural face skin shader, which takes a standard face asset (i.e., geometry, albedo map, and normal map) and an HDRI for illumination as inputs and generates a photo-realistic rendering as output. Extensive experiments showcase the quality and efficiency of our PBR face skin shader, indicating the effectiveness of our proposed lighting and material representations.

1 INTRODUCTION

Physically-based rendering (PBR) provides a shading and rendering method to accurately represent how light interacts with objects in virtual 3D scenes. Whether working with a real-time rendering system in computer graphics or film production, employing a PBR process will facilitate creating images that look like they exist in the real world for a more immersive experience. Industrial PBR pipelines take the guesswork out of authoring surface attributes like transparency since their methodology and algorithms are based on physically accurate formulas and resemble real-world materials. This process relies on onerous artist tuning and high computational power in a long production cycle. In recent years, academia has shown incredible success using differentiable neural rendering in extensive tasks such as view synthesis (Mildenhall et al., 2020), inverse rendering (Zhang et al., 2021a), and geometry inference (Liu et al., 2019). Driven by the efficiency of neural rendering, a natural next step would be to marry neural rendering and PBR pipelines. However, none of the existing neural rendering representations supports the accuracy, expressiveness, and quality mandated by the industrial PBR process.

A PBR workflow models both specular reflections, which refers to light reflected off the surface, and diffusion or subsurface scattering, which describes the effects of light absorbed or scattered internally. Pioneering works of differentiable neural shaders such as Softras (Liu et al., 2019) adopted the Lambertian model as BRDF representation, which only models the diffusion effects and results in low-quality rendering. NeRF (Mildenhall et al., 2020) proposed a novel radiance field representation for realistic view-synthesis under an emit-absorb lighting transport assumption without explicitly modeling BRDFs or lighting, and hence is limited to a fixed static scene with no scope for relighting. In follow-up work, NeRV (Srinivasan et al., 2020) took one more step by explicitly modeling directional light, albedo, and visibility maps to make the fixed scene relightable. The indirect illumination was achieved by ray-tracing under the assumption of one bounce of incoming light. However, this lighting model is extremely computationally heavy for real-world environ-

ment illumination when more than one incoming directional lights exist. To address this problem, NeRD (Boss et al., 2020) and PhySG (Zhang et al., 2021a) employ a low-cost global environment illumination modeling method using spherical Gaussian (SG) to extract parameters from HDRIs. Neural-PIL (Boss et al., 2021) further proposed a pre-trained light encoding network for a more detailed global illumination representation. However, it is still a global illumination representation assuming the same value for the entire scene, which is not true in the real world, where illumination is subjected to shadows and indirect illumination bouncing off of objects in different locations in the scene. Thus it's still an approximation but not an accurate representation of the environment illumination. Regarding material (BRDF) modeling, all the current works adopt the basic rendering parameters (such as albedo, roughness, and metalness) defined in the rendering software when preparing the synthetic training data. However, they will fail in modeling intricate real-world objects such as participating media (e.g., smoke, fog) and translucent material (organics, skins, jade), where high scattering and subsurface scattering cannot be ignored. Such objects require more effort and hence attract more interest in research in their traditional PBR process.

In this work, we aim to design accurate, efficient lighting/ illumination and BRDF representations to enable the neural PBR process, which will support high-quality and photo-realistic rendering in a fast and lightweight manner. To achieve this goal, we propose a novel lighting representation - light sampling field to model both the direct and indirect illumination from HDRI environment maps. Our light sampling field representation faithfully captures the direct illumination (incoming from light sources) and indirect illumination (summary of all indirect incoming lighting from surroundings) given an arbitrary sampling location in a continuous field. Accordingly, we propose BRDF representations in the format of surface specular, surface diffuse, and subsurface scattering for modeling real-world object material. This paper mainly evaluates the proposed representations with a novel volumetric neural physically-based shader for human facial skin. We trained with an extensive high-quality database, including real captured ground truth images as well as synthetic images for illumination augmentation. We also introduce a novel way of integrating surface normals into volumetric rendering for higher fidelity. Coupled with proposed lighting and BRDFs models, our light transport module delivers pore-level realism in both on- and underneath-surface appearance unprecedentedly. Experiments show that our light sampling field is robust enough to learn illumination by local geometry, and such an effect can usually only be modeled by ray tracing. Therefore, our method compromises neither efficiency nor quality with the light sampling field when compared to ray tracing.

The main contributions of this paper are as follows: 1) A novel volumetric lighting representation that accurately encodes the direct and indirect illumination positionally and dynamically given an environment map. Our local representation enables efficient modeling of complicated shading effects such as inter-reflectance in neural rendering for the first time as far as we are aware. 2) A BRDF measurement representation that supports the PBR process by modeling specular, diffuse, and subsurface scattering separately. 3) A novel and lightweight neural PBR face shader that takes facial skin assets and environment maps (HDRIs) as input and efficiently renders photo-realistic, high-fidelity, and accurate images comparable to industrial traditional PBR pipelines such as Maya. Our face shader is trained with an image database consisting of extensive identities and illuminations. Once trained, our models will extract lighting models and BRDFs from input assets, which works well for novel subjects/ illumination maps. Experiments show that our PBR face shader significantly outperforms the state-of-the-art neural face rendering approaches regarding quality and accuracy, which indicates the effectiveness of the proposed lighting and material representations.

2 Related Work

Volumetric Neural Rendering Volumetric rendering models the light interactions with volume densities of absorbing, glowing, reflecting, and scattering materials (Max, 1995). A neural volumetric shader trains a model from a set of images and queries rendered novel images. The recent state-of-the-art was summarized in a survey (Tewari et al., 2020). In addition, Zhang et al. (2019) introduced radiance field as a differentiable theory of radiative transfer. Neural Radiance Field (NeRF) (Mildenhall et al., 2020) further described scenes as differentiable neural representation and the following raycasting integrated color in terms of the transmittance factor, volume density, and the voxel diffuse color. Extensions to NeRF were developed for better image encoding (Yu et al., 2020), ray marching (Bi et al., 2020a), network efficiency (Lombardi et al., 2021; Yariv et al., 2020), realistic shading (Suhail et al., 2022) and volumetric radiative decomposition (Bi et al., 2020b; Rebain

et al., 2020; Zhang et al., 2021c; Verbin et al., 2021). In particular, NeRV (Srinivasan et al., 2020), NeRD (Boss et al., 2020; 2021) decomposed the reconstructed volume into geometry, SVBRDF and illumination given a set of images even under varying lighting condition. RNR (Chen et al., 2020b) assumed the environmental illumination as distant light and is able to decompose the scene into an albedo map with a 10-order spherical harmonics of incoming directions.

Portrait and Face Relighting Early single image relighting techniques utilize CNN-based image translation (Nalbach et al., 2017; Thies et al., 2019). Due to lack of 3D models, image translation approaches cannot recover surface materials or represent realistic high fidelity details, thus neural volumetric relighting approaches are widely adopted recently. The face relighting utilized strong priors. Zhou et al. (2019) fitted a 3D face model to the input image and obtains refined normal to help achieve relighting. Chen et al. (2020a) relit the image by using spherical harmonics lighting on a predicted 3D face. Hou et al. (2022) introduced a shadow mask estimation module to achieve novel face relighting with geometrically consistent shadows. With high-quality volumetric capture in lightstage (Debevec et al., 2000) to obtain training data, this trend achieved the following: regression of a one-light-at-a-time (OLAT) image for relighting (Meka et al., 2019), encoding the feature tensors for a Phong shading into UV space and relighting using an HDRI map (Meka et al., 2020), a neural renderer that can predict the non-diffuse residuals (Zhang et al., 2021b). Similar to our approach most, Sun et al. (2021) inferred both light transport and density, and enabled relighting and view synthesis from a sparse set of input images.

3 METHODS

3.1 PRELIMINARIES

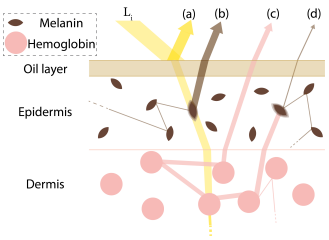


Figure 1: **Skin Scattering Model.** The incident light L_i scattered from skin consists of oil layer specular reflection (a) or penetrate one or more scattering layers (b, c, d) at some point underneath the oil layer. The difference of melanin distribution or color results in the diversity of skin appearance. In humans, the distribution and color of the melanin is the primary determinant of skin appearance. Darker and denser melanin in the epidermis leads to darker skin color and vise versa.

We use the rendering equation Kajiya (1986) to estimate the radiance L at a 3D point $\mathbf{x}$ with the outgoing direction ω_o: $L(\mathbf{x}, \boldsymbol{\omega}_o) = \int_{\boldsymbol{\omega}_i \in \Omega^+} f(\mathbf{x}, \boldsymbol{\omega}_o, \boldsymbol{\omega}_i) L_i(\mathbf{x}, \boldsymbol{\omega}_i)(\boldsymbol{\omega}_i \cdot \boldsymbol{n}_x) d\boldsymbol{\omega}_i$, where $f(\mathbf{x}, \boldsymbol{\omega}_o, \boldsymbol{\omega}_i)$ is the BRDF representation and $L_i(\mathbf{x}, \boldsymbol{\omega_i})$ measures the radiance of incident light with direction $\boldsymbol{\omega_i}$.

Generally, the incident light could be categorized as direct vs. indirect light. Fig. 1 illustrates an example of direct light (path (a), (b)) and indirect light (path (c)) in human face skin, where subsurface scattering happens in the deeper dermis layer, causing the nearby area to receive indirect light. Therefore, the rendering formulation can be split into separate components with direct and indirect lighting:

$$L(\mathbf{x}, \boldsymbol{\omega}_o) = \int_{\boldsymbol{\omega}_i \in \Omega^+} f_s(\mathbf{x}, \boldsymbol{\omega}_o, \boldsymbol{\omega}_i) L_i^d(\mathbf{x}, \boldsymbol{\omega}_i)(\boldsymbol{\omega}_i \cdot \boldsymbol{n}_x) d\boldsymbol{\omega}_i + \int_{\boldsymbol{\omega}_i \in \Omega} f_{ss}(\mathbf{x}, \boldsymbol{\omega}_o, \boldsymbol{\omega}_i) L_i^{id}(\mathbf{x}, \boldsymbol{\omega}_i)(\boldsymbol{\omega}_i \cdot \boldsymbol{n}_x) d\boldsymbol{\omega}_i \quad (1)$$

where f_s, f_{ss} represent the BRDF evaluation of surface and subsurface, respectively. Following the render equation, we design a lightweight physically-based rendering method by learning different BRDF representations and modeling the direct and indirect lights, as we will introduce in the next few sections.

3.2 ILLUMINATION AND BIDIRECTIONAL REFLECTANCE DISTRIBUTION LEARNING

We propose a lighting model and material model to construct the light sampling field and estimate the BRDF representations, as will be introduced in the next few sections.

3.2.1 LIGHTING MODEL

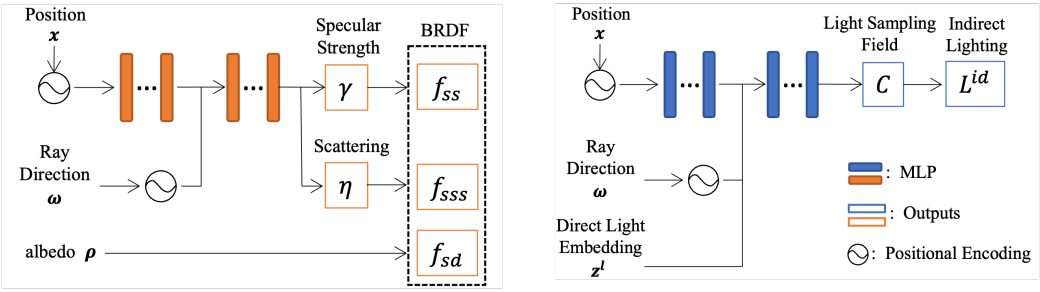
Considering interaction between light and different skin layers, we propose a novel HDRI light modeling that decomposes lighting into direct illumination component and indirect illumination component. An illustration of HDRI is shown in Fig. 2a.

(a) Importance Light Sampling.

(b) Light Sampling Field Visualization

Figure 2: **Lighting Model.** (a) We use importance sampling to sample heavily on pixels with high intensity. (b) We use various density of sampling to bring attention to locations with data as in the Light Sampling Field.

Importance Light Sampling for Direct Illumination. Direct radiance comes from HDRI map directly. In order to compute contribution of pixels in HDRI to different points in our radiance field, we use a SkyDome to represent direct illumination by projecting the HDRI environment map onto a sphere. Each pixel on the sphere is regarded as a distant directional light source. Hence, direct lighting is identical at all locations in the radiance field. Such representation preserves specular reflection usually achieved by ray-tracing methods. We take two steps to construct the representation. Firstly, we uniformly sample a point grid of size N on the sphere, where each point is a light source candidate (Fig. 2a). Secondly, we apply our unique importance light sampling method to filter valid candidates by two thresholds: 1) an intensity threshold that clips intensity of outliers with extreme values; 2) an importance threshold that filters out outliers in textureless regions.



(a) Material Network

(b) Light Sampling Field Network

Figure 3: **Material and Light Sampling Field Network.** (a) Our material network takes in 3D position and ray direction as inputs and predicts the specular strength and skin scattering parameters of our BRDF model. (a) The Light Sampling Field network applies similar differentiable network but with the light sampling created by our importance light sampling technique as another input to predict SH coefficients.

Light Sampling Field for Indirect Illumination. Indirect illumination models the incoming lights reflected or emitted from surrounding objects, which is achieved by ray-tracing with the assumption of limited bounce times in the traditional PBR pipeline. Inspired by the volumetric lightmaps used in Unreal Engine (Karis & Games, 2013), which stores precomputed lighting in sampled points and use them for interpolation at runtime for modeling indirect lighting of dynamic objects. We adopt a continuous light sampling field for accurately modeling the illumination variation in different scene positions. We use Spherical Harmonics(SH) to model the global effects of indirect illumination in each sample location. Our local SH representation enables inter-reflection and subsurface scattering effects with neural lighting transport. We visualize our light sampling field with selected discrete sample points in Fig. 2b.

Learning Light Sampling Field. We design a network (Fig. 3b) to predict the spherical coefficient C_k^m of a continuous light field. The inputs of this network are the lighting embeddings z^l, the positional encoding of 3D location $\mathbf{x}$ and ray direction $\boldsymbol{\omega}$. Conditioned on the lighting representations, the network succeeds in predicting the accurate lighting as well as providing a dynamic illumination environment. Fig. 7 evaluates our lighting model.

3.2.2 MATERIAL MODEL

The choice of reflectance parameters usually relies on artists' tuning and therefore requires high computational power and results in long production cycle, but sometimes even not ideal. To tackle the problem, we propose a lightweight material network (Fig. 3a) to estimate the BRDF parameters including specular strength $\gamma \in \mathbb{R}$ and skin scattering $\eta \in \mathbb{R}^3$ through learnable parameters. The parameters are crucial to represent the reflectance property of both surface and subsurface. Together with input albedo ρ, we can construct a comprehensive BRDF to mode surface reflection and subsurface scattering. It consists of a surface specular component f_{ss}, a surface diffuse component f_{sd} and a subsurface scattering component f_{sss}. Refer to Sec. 3.3 for a detailed light transport and Fig.6 for an evaluation of modeling subsurface scattering.

3.3 LIGHT TRANSPORT

Light transport defines a light path from the luminaire to the receiver. We introduce the light transport that connects our lighting model and material model in Fig. 4. We also detail the rendering equation in this section to matching the light transport along the light path.

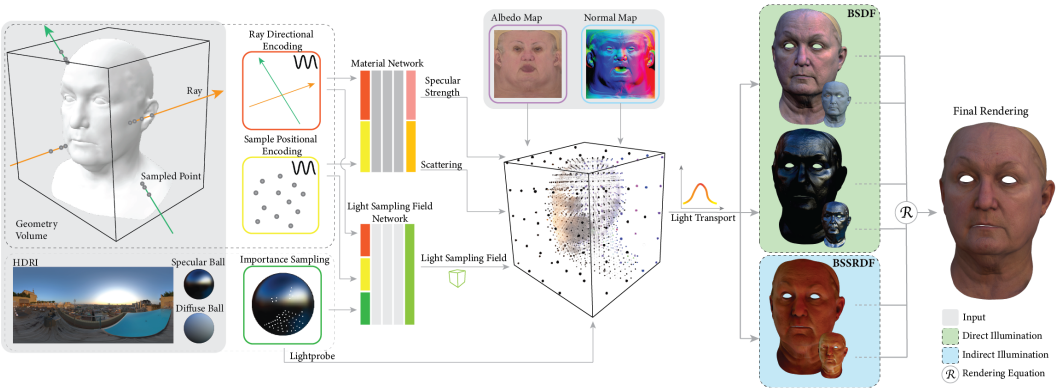


Figure 4: **Light Transport.** Starting from sampling lights from HDRI and 3D points close to geometry along the light path, we feed encoded locations, view direction to Material network and Light Sampling Field network (with extra direct illumination code). The networks output material and indirect illumination code. Our PBR equation generates radiance at each sample based on our material model and light transport in the volume further composites all the radiance on the light path to the receiver.

Volume Casting. Our lighting model includes direct illumination and indirect illumination. Light transports from explicit lights and casts along the light path. We therefore define the light transport along the light path, $\boldsymbol{p}$, in the volume by the following:

$$L(\mathbf{x}, \boldsymbol{\omega}) = \int_0^{\infty} \tau\left(t\right) \cdot \sigma\left(\mathbf{p}(t)\right) \cdot L\left(\mathbf{p}(t), \boldsymbol{\omega}\right)\, dt \tag{2}$$

where $L(\mathbf{x}, \boldsymbol{\omega})$ is the total radiance at $\mathbf{x}$ along the light path direction $\boldsymbol{\omega}$. $\sigma(\mathbf{x})$ is the volume density, which is converted from the input geometry. $\tau(x) = \exp(-\int_0^x \sigma(\boldsymbol{p}(t))dt)$ is the visibility that indicates whether the location $\mathbf{x}$ is visible on the light path, where $\boldsymbol{p}(t)$ represents the 3D location on the light path along $\boldsymbol{\omega}$ from $\mathbf{x}$ and defined as $\boldsymbol{p}(t) = \mathbf{x} + t\boldsymbol{\omega}$.

We use explicit geometry to construct a more reliable density field. More specifically, we locate the intersection coordinates, $\mathbf{x_0}$, between any arbitrary light path and the input geometry. For any point $\mathbf{x}$ along the light path $\boldsymbol{\omega}$, we define the density σ as the following central Gaussian distribution

function: $\sigma(\mathbf{x}) = \alpha_\sigma \cdot \exp\left(-d_G(\mathbf{x})^2/(2\delta^2)\right)$, where $d_G(\mathbf{x})$ is the distance between $\mathbf{x}$ and the intersection $\mathbf{x_0}$, α_σ and δ are two scalars that determine the magnitude and standard deviation of the gaussian distribution.

Material Scattering. The light transport between the scene objects and light sources is characterized by rendering equation. Eqn. 1 introduce the rendering covering direct illumination and indirect illumination. Also, classified by reflection location, the reflected radiance $L(\mathbf{x}, \boldsymbol{\omega}_o)$ has two componenets: surface reflectance and subsurface volume scattering. To have a comprehensive light transport representation, we further develop the equation with the dissection of light as well as the specialized BSSRDF components:

$$L(\mathbf{x}, \boldsymbol{\omega}_o) = \underbrace{\int_{\boldsymbol{\omega}_i \in \Omega_+} f_{ss}(\mathbf{x}, \boldsymbol{\omega}_o, \boldsymbol{\omega}_i) L_i^d(\mathbf{x}, \boldsymbol{\omega}_i) |\boldsymbol{\omega}_i \cdot \mathbf{n_x}| d\boldsymbol{\omega}_i}_{\text{surface specular reflectance with direct light}} + \underbrace{\int_{\boldsymbol{\omega}_i \in \Omega_+} f_{sd}(\mathbf{x}, \boldsymbol{\omega}_o, \boldsymbol{\omega}_i) L_i^d(\mathbf{x}, \boldsymbol{\omega}_i) |\boldsymbol{\omega}_i \cdot \mathbf{n_x}| d\boldsymbol{\omega}_i}_{\text{surface diffuse reflectance with direct light}}$$
$$+ \underbrace{\int_{\boldsymbol{\omega}_i \in \Omega} f_{sss}(\mathbf{x}, \boldsymbol{\omega}_o, \boldsymbol{\omega}_i) L_i^{id}(\mathbf{x}, \boldsymbol{\omega}_i) |\boldsymbol{\omega}_i \cdot \mathbf{n_x}| d\boldsymbol{\omega}_i}_{\text{subsurface scattering with indirect light}} \tag{3}$$

here $L_i^d(\mathbf{x}, \boldsymbol{\omega}_i)$ and $L_i^{id}(\mathbf{x}, \boldsymbol{\omega}_i)$ are the incoming direct and indirect radiance from direction $\boldsymbol{\omega}_i$ at point $\mathbf{x}$, respectively. f_{ss}, f_{sd} and f_{sss} are the different counterparts of light transport parameterized by material representations. We show the complete light transport in Eqn. 4 consists of our light and material representation.

$$L(\mathbf{x}, \boldsymbol{\omega}_o) = \underbrace{\gamma_\mathbf{x} \cdot \int_{\boldsymbol{\omega}_i \in \Omega_+} L_i^d(\mathbf{x}, \boldsymbol{\omega}_i) |\boldsymbol{\omega}_o \cdot R(\boldsymbol{\omega}_i, \mathbf{n_x})|^e d\boldsymbol{\omega}_i}_{\text{surface specular reflectance with direct light}} + \underbrace{\frac{\rho_\mathbf{x}^s}{\pi} \cdot \int_{\boldsymbol{\omega}_i \in \Omega_+} L_i^d(\mathbf{x}, \boldsymbol{\omega}_i) |\boldsymbol{\omega}_i \cdot \mathbf{n_x}| d\boldsymbol{\omega}_i}_{\text{surface diffuse reflectance with direct light}}$$
$$+ \underbrace{\frac{\rho_\mathbf{x}^{ss} + \eta_\mathbf{x}}{\pi} \cdot \int_{\boldsymbol{\omega}_i \in \Omega} L_i^{id}(\mathbf{x}, \boldsymbol{\omega}_i) |\boldsymbol{\omega}_i \cdot \mathbf{n_x}| d\boldsymbol{\omega}_i}_{\text{subsurface scattering with indirect light}} \tag{4}$$

where $\gamma_\mathbf{x}$ and $\eta_\mathbf{x}$ are the predicted specular strength and scattering from material network at $\mathbf{x}$. $R(\boldsymbol{\omega}_i, \mathbf{n_x})$ denotes the reflection direction of $\boldsymbol{\omega}_i$ at the surface with normal $\mathbf{n_x}$ and e denotes the specular exponent. Also, $\rho_\mathbf{x}^s$ and $\rho_\mathbf{x}^{ss}$ are surface and subsurface albedo at $\mathbf{x}$ sampled from the input albedo map correspondingly with geometry.

4 Experiments and Evaluation

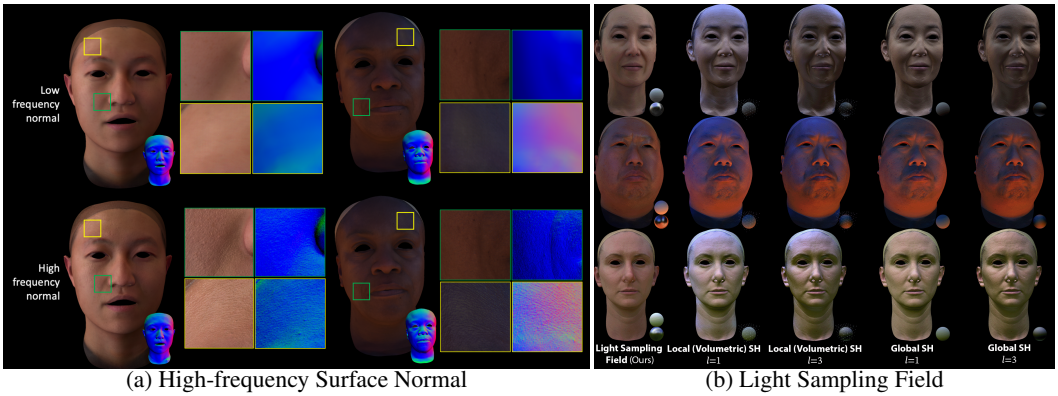
4.1 Training Dataset

Our training dataset is composed of a synthetic image dataset and a Lightstage-scanned image dataset. Under every combination of 21 face assets and 101 HDRI+86 OLAT illumination, we used a professionally-tuned Maya face shader to render 40-view colored images in *synthetic dataset*. *Lightstage-scan dataset* consists of 16-view captured colored images of 48 subjects in 27 expressions under white illumination. We carefully selected the combinations to cover diverse age, skin color, gender and illumination. Further details can be found in Appendix. B.

4.2 Ablation Studies

High-frequency Surface Normal. To investigate the effectiveness of high-frequency surface normal as input, we performed an ablation study by replacing high-frequency input normal map with a low-frequency one extracted from coarse geometry (Fig. 5a). Given other input assets unchanged, rendered results with low-frequency normal map shows incapability of extracting high-frequency information from input images. Therefore, high-frequency normal map is crucial in achieving pore-level sharpness even though training data contains high-frequency details on images.

Light Sampling Field. With same input, we performed ablation study by only replacing learned indirect diffuse light component with pre-calculated SH representation at level 1 and 3. To further strength our claim, we assigned locally computed SH representation to different sampled points to mark potential influence of position on achieving accurate illumination. However, local SH at level 3 (best among the four) still fails in delivering illumination as accurate as our method. The render results validate our predicted light sampling field is a better lighting model than global or local pre-calculated spherical harmonics.



(a) High-frequency Surface Normal (b) Light Sampling Field

Figure 5: **Ablation Studies.** (a) shows rendering with high- and low-frequency normal map input with zoom-in details. (b) shows rendering under different lighting models: our Light Sampling Field, local SH ($l = 1, 3$), and global SH ($l = 1, 3$).

4.3 Evaluation and Analysis

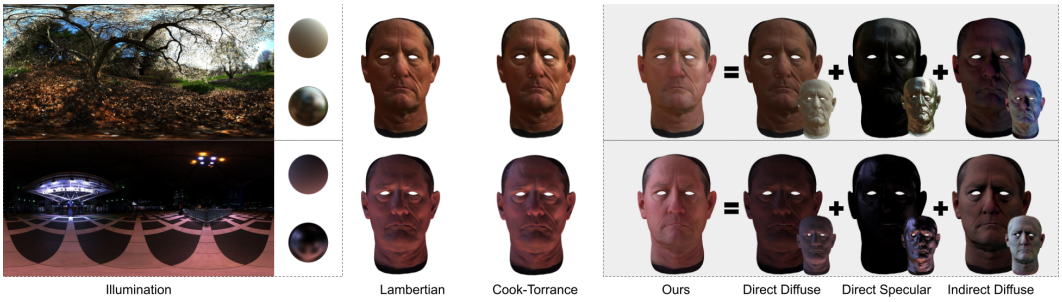


Figure 6: **Material Modeling.** Under HDRI illumination in column (a), we show appearance of (b) Lambertian BRDF , (c) Cook-Torrance BRDF, (d) ours, which is composed of direct diffuse, direct specular and indirect diffuse components.

Material Model Evaluation. The proposed material models surface scattering and subsurface scattering. We conduct a comparison with only surface scattering, BRDF, shown in Fig. 6. To illustrate the difference of surface scattering and our material modeling, we present two BRDF materials and also our proposed material with layer-by-layer decomposition. Cook-Torrance BRDF has more specular effect than Lambertian yet saving rubber-alike appearance from uncanny valley. Our method predicts subsurface scattering from material network to achieve vivid looking around nose tip and ears by depicting red blood cell and vessel color.

Lighting Model Evaluation. We claim that our lighting model delivers a more complete and accurate illumination representation than SH Lighting (level 1), SG and Neural-PIL. In supporting our claim, we rendered a uniform white ball within a scene illuminated by different lighting models (Fig. 7). Our result delivers the highest fidelity of illumination with the widest spectrum of light and a lighting field sensitive to lighting distribution.

Inverse Rendering. Instead of input geometry and texture maps, our method can still achieve novel lighting rendering with only multi-view images (under various illumination). We implemented

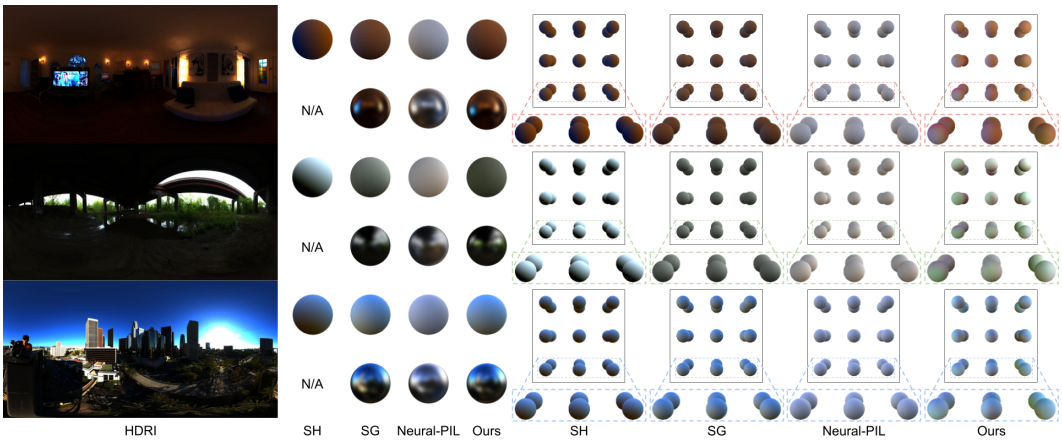


Figure 7: **Lighting Model Evaluation.** We rendered a uniform white ball and a grid of such balls with different lighting models: SH ($l = 1$), SG, Neural-PIL, and ours. We separated diffuse lighting and specular lighting by turning off corresponding component in rendering equation.

an additional MLP to predict a density field instead of converting from input geometry. We present our results under novel illumination in Fig. 8.

Figure 8: **Inverse Rendering.** We provide multi-view images as input and predict a density field for light transport. Given albedo, we achieve rendering under novel illumination.

4.4 Qualitative Results

Qualitative Comparison. In Fig. 10, we compare the rendering results of FRF, NeLF, SIPR, Neural-PIL and our methods under HDRI and OLAT illuminations. The proposed method not only casts clear hard shadow that ensures full occlusion by face geometry, but also showcases soft shadow due to indirect illumination. We also provide Maya rendering, top-notch industrial renderer, as reference. Neural-PIL and NeLF neither model directional light nor were trained on OLAT (one-light-at-a-time) data. We therefore compare them only under HDRI illumination. Regarding functionality, SIPR are image-based relighting methods that model scene in 2D, while FRF, Neural-PIL and NeLF model in 3D. Hence, SIPR cannot be queried for novel viewpoints in 3D. Neural-PIL inherits per-scene-per-train manner from NeRF and does not generalize to different geometries. Thus, Neural-PIL is unable to query novel subjects. We further present more qualitative results in Fig. 13, zoom-in comparison for both HDRI and OLAT illumination in Fig. 14 and performance of our method on other novel datasets in Fig.15.

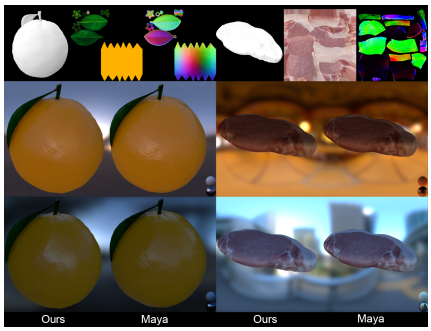


Figure 9: **Rendering on general object assets.**

Rendering on General Object Assets. We picked orange and meat in addition to face subjects to show that our method is generalizable on diverse multi-layered objects in Fig. 9. Through testing

(a) HDRI illumination (b) OLAT illumination

Figure 10: **Qualitative comparison with baseline methods.** (a) Under HDRI illumination, our rendering results achieve highest sharpness and non-rubber-alike appearance; (b) OLAT illumination simulates strong directional lighting. While our robust direct light modeling enables hard shadow casting, our indirect light modeling softens shadow at boundary realistically.

on different organic material, we show consistent sharpness and realistic appearance, especially in accomplishing difficult specular reflection on orange and accurate soft shadow on meat.

4.5 Quantitative Results

We evaluated PSNR, SSIM and LPIPS in quantitative measurements (Table 1) with Maya rendering as benchmark. Specifically, LPIPS evaluates perception on VGG-19 network. Ours outperforms other baseline methods in all three metrics.

Baseline	FRF	SIPR	NeLF	Neural-PIL	Ours
PSNR ↑	29.83	33.43	33.80	33.22	**36.69**
SSIM ↑	0.4489	0.8507	-	0.9153	**0.9250**
LPIPS (VGG) ↓	0.6017	0.1265	0.7538	0.1082	**0.0664**

Table 1: **Facial rendering metrics at** 800×800 **resolution.** We do not provide SSIM score of NeLF due to slight view difference.

5 Conclusion

We demonstrate that the prior neural rendering representation for physically-based rendering fail to accurately model environment lighting or capture subsurface details. In this work, we propose a differentiable light sampling field network that models dynamic illumination and indirect lighting in a lightweight manner. In addition, we propose a flexible material network that models subsurface scattering for complicated materials such as human face. Experiments on both synthetic and real-world datasets demonstrate that our light sampling field and material network collectively improve the rendering quality under complicated illumination compared with prior works. In the future, we will focus on modeling more complicated materials such as translucent materials and participated media. We will also collect datasets based on general objects and apply them for extensive tasks such as inverse rendering.

REFERENCES

Sai Bi, Zexiang Xu, Pratul Srinivasan, Ben Mildenhall, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. Neural reflectance fields for appearance acquisition. *arXiv preprint arXiv:2008.03824*, 2020a.

Sai Bi, Zexiang Xu, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. Deep reflectance volumes: Relightable reconstructions from multi-view photometric images. *arXiv preprint arXiv:2007.09892*, 2020b.

Mark Boss, Raphael Braun, Varun Jampani, Jonathan T Barron, Ce Liu, and Hendrik Lensch. Nerd: Neural reflectance decomposition from image collections. *CoRR*, 2020.

Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan Barron, and Hendrik Lensch. Neural-pil: Neural pre-integrated lighting for reflectance decomposition. *Advances in Neural Information Processing Systems*, 34, 2021.

Yajing Chen, Fanzi Wu, Zeyu Wang, Yibing Song, Yonggen Ling, and Linchao Bao. Self-supervised learning of detailed 3d face reconstruction. *IEEE Transactions on Image Processing*, 29:8696–8705, 2020a.

Zhang Chen, Anpei Chen, Guli Zhang, Chengyuan Wang, Yu Ji, Kiriakos N Kutulakos, and Jingyi Yu. A neural rendering framework for free-viewpoint relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5599–5610, 2020b.

Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pp. 145–156, 2000.

Andrew Hou, Michel Sarkis, Ning Bi, Yiying Tong, and Xiaoming Liu. Face relighting with geometrically consistent shadows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4217–4226, 2022.

James T. Kajiya. The rendering equation. *SIGGRAPH Comput. Graph.*, 20(4):143–150, aug 1986. ISSN 0097-8930. doi: 10.1145/15886.15902. URL https://doi.org/10.1145/15886.15902.

Brian Karis and Epic Games. Real shading in unreal engine 4. *Proc. Physically Based Shading Theory Practice*, 4:3, 2013.

Shichen Liu, Tianye Li, Weikai Chen, and Hao Li. Soft rasterizer: A differentiable renderer for image-based 3d reasoning. *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2019.

Stephen Lombardi, Tomas Simon, Gabriel Schwartz, Michael Zollhoefer, Yaser Sheikh, and Jason Saragih. Mixture of volumetric primitives for efficient neural rendering. *arXiv preprint arXiv:2103.01954*, 2021.

Nelson Max. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1(2):99–108, 1995.

Abhimitra Meka, Christian Haene, Rohit Pandey, Michael Zollhöfer, Sean Fanello, Graham Fyffe, Adarsh Kowdle, Xueming Yu, Jay Busch, Jason Dourgarian, et al. Deep reflectance fields: high-quality facial reflectance field inference from color gradient illumination. *ACM Transactions on Graphics (TOG)*, 38(4):1–12, 2019.

Abhimitra Meka, Rohit Pandey, Christian Häne, Sergio Orts-Escolano, Peter Barnum, Philip David-Son, Daniel Erickson, Yinda Zhang, Jonathan Taylor, Sofien Bouaziz, et al. Deep relightable textures: volumetric performance capture with neural rendering. *ACM Transactions on Graphics (TOG)*, 39(6):1–21, 2020.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision*, pp. 405–421. Springer, 2020.

Oliver Nalbach, Elena Arabadzhiyska, Dushyant Mehta, H-P Seidel, and Tobias Ritschel. Deep shading: convolutional neural networks for screen space shading. In *Computer graphics forum*, volume 36, pp. 65–78. Wiley Online Library, 2017.

Daniel Rebain, Wei Jiang, Soroosh Yazdani, Ke Li, Kwang Moo Yi, and Andrea Tagliasacchi. Derf: Decomposed radiance fields. *arXiv preprint arXiv:2011.12490*, 2020.

Pratul P Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T Barron. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. *arXiv preprint arXiv:2012.03927*, 2020.

Mohammed Suhail, Carlos Esteves, Leonid Sigal, and Ameesh Makadia. Light field neural rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8269–8279, 2022.

Tiancheng Sun, Kai-En Lin, Sai Bi, Zexiang Xu, and Ravi Ramamoorthi. Nelf: Neural light-transport field for portrait view synthesis and relighting. *arXiv preprint arXiv:2107.12351*, 2021.

Ayush Tewari, Ohad Fried, Justus Thies, Vincent Sitzmann, Stephen Lombardi, Kalyan Sunkavalli, Ricardo Martin-Brualla, Tomas Simon, Jason Saragih, Matthias Nießner, et al. State of the art on neural rendering. In *Computer Graphics Forum*, volume 39, pp. 701–727. Wiley Online Library, 2020.

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

Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. Ref-nerf: Structured view-dependent appearance for neural radiance fields. *arXiv preprint arXiv:2112.03907*, 2021.

Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. *Advances in Neural Information Processing Systems*, 33, 2020.

Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelnerf: Neural radiance fields from one or few images. *arXiv preprint arXiv:2012.02190*, 2020.

Cheng Zhang, Lifan Wu, Changxi Zheng, Ioannis Gkioulekas, Ravi Ramamoorthi, and Shuang Zhao. A differential theory of radiative transfer. *ACM Transactions on Graphics (TOG)*, 38(6): 1–16, 2019.

Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5453–5462, 2021a.

Xiuming Zhang, Sean Fanello, Yun-Ta Tsai, Tiancheng Sun, Tianfan Xue, Rohit Pandey, Sergio Orts-Escolano, Philip Davidson, Christoph Rhemann, Paul Debevec, et al. Neural light transport for relighting and view synthesis. *ACM Transactions on Graphics (TOG)*, 40(1):1–17, 2021b.

Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *arXiv preprint arXiv:2106.01970*, 2021c.

Hao Zhou, Sunil Hadap, Kalyan Sunkavalli, and David W Jacobs. Deep single-image portrait relighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7194–7202, 2019.

A IMPLEMENTATION DETAILS

In order to construct the density field σ, we set α_σ and δ to be 10 and 0.5, respectively. We further compare the rendering results of other values and visualize in Fig. 11. In the constructed radiance field, to sample rays, we draw 1024 random rays per batch which correspond to 1024 pixels to be rendered. Along each ray, we sample 64 points for shading model. The low-frequency location of 3D points and direction of ray are transformed to high-frequency input via positional encoding and directional encoding respectively (Mildenhall et al., 2020). The length of encoded positional and directional embedding is 37 and 63 respectively for both material network and light sampling field network. Also, the 2400-length lighting feature sampled from HDRI is fed into the material network. The output length of material MLP is 4 for each queried 3D point: specular strength $\gamma \in \mathbb{R}$, and scattering $\eta \in \mathbb{R}^3$. The light sampling field network outputs encoded spherical coefficients $C_k^m \in \mathbb{R}^{12}$ for indirect lighting. We show that the level of frequency is enough to generate sharp images as demonstrated in results. During light transport, we obtain a weighted value for each ray based on τ distribution among the sampled points along the ray. After introducing values from pre-processed albedo and normal maps, the value of each component on rays are gathered and visualized as an image with pixels representing their intensities following our rendering Eqn. 4. Finally, the rendered RGB values are constrained with ground truth by a MSE loss. In our application, MLP modules can converge in $50,000$ iterations (2.6 hour) on a single Tesla V100, with decent results on the same level of details as reference images.

B DATASETS

Figure 12: **Training Dataset.**

Our training dataset consists of 1) a lightstage-scanned multi-view colored image dataset under white illumination (Fig. 12 (a)), 2) a synthetic multi-view colored image dataset rendered via a professionally-tuned Maya face shader (Fig. 12 (b)). In the following sections, we introduce data composition and settings for synthetic data rendering (Sec. B) and lightstage scanning (Sec. B).

Synthetic Image Dataset Our input rendering assets to Maya renderer are 101 HDRI environment maps, 86 OLAT environment maps and 21 face assets: the HDRI data covers various illuminations from outdoor open area to small indoor area; OLAT environment map cover 86 different regions in SkyDome as directional light; 21 subjects cover a variety of skin color, age and gender. We rendered 40 fixed-view RGBA-space images under all combinations of illumination and face assets. In total, we acquired $37,160$ images in 800×800-pixel resolution for our synthetic image dataset.

Lightstage-scanned Real Image Dataset Synthetic data is insufficient to train a model with output rendering close to real world. We utilized lightstage to additionally capture multi-view images

for 48 subjects under uniform white illumination. Besides diverse skin color, gender and age of subjects, the lightstage-scanned dataset consists of 27 expressions of each subject. We set up 16 fixed cameras covering different viewpoint of frontal view. In total, we acquired $20,736$ images in 3008×4096-pixel resolution for our real image dataset.

C MORE RESULTS AND ANALYSIS

C.1 QUALITATIVE RESULTS

In Fig. 13, we provide comprehensive qualitative results. For each row, we exhibit rendering inputs including geometry, albedo map, object normal map and followed by six rendered results. The first three are different expressions under all white illumination, and the last three are neutral expression under different illuminations. With the proposed photo-realistic neural renderer, we can perform rendering at any arbitrary resolution while preserving the rendering quality.

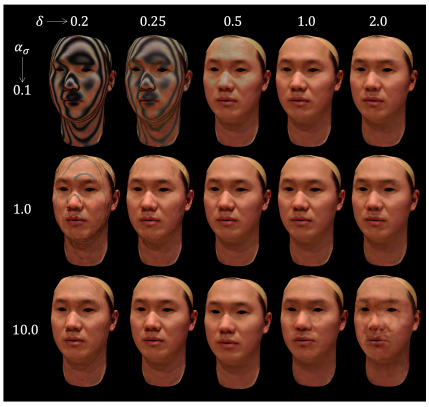


Figure 11: **Desnity Field Construction.** We present a comparison of hyper-parameters (α_σ and δ) in density field construction.

We further compare our rendering with Maya, industry-level rendering engine, under HDRI or OLAT illumination and present zoom-in pore-level details in Fig. 14. The zoom-in inspections show comparable or even better rendering results in their sharpness or illumination. Skin wrinkles and forehead specularity are particularly rich and sharp from the proposed method. At testing time, with the same rendering assets and queries (2500×2500 in resolution), our method requires the training images with only 800×800 in resolution. Under OLAT settigns, Our method casts hard shadow and soft shadow as accurate as Maya.

In addition to the rendering results, we also compare the rendering speed in Table 2. With the same specification of output and environment, our method is able to achieve up to 47-49 times faster with engineering acceleration (e.g. multi-thread processing)

Resolution	800×800	1440×1440	2880×2880	5760×5760
Maya	130s	370s	1857s	7848s
Ours	**3s**	**8s**	**31s**	**158s**

Table 2: Facial rendering speed benchmark on different resolution

To demonstrate that high-quality training data does not determine robustness of method, we evaluated our trained model over other available resources in Fig. 15. We converted three datasets to match our input as following,

Triplegangers We used FaceX 3DMM to fit and align geometries; albedo map was directly transferred from source; normal map was inferred using (Wang et al., 2018).

3D Scan Store We used 3DMM to fit and align geometries; albedo and normal maps were provided and not further processed.

FaceScape Geometries and albedo maps were provided and not further processed. Normal map was unavailable and not used.

Our method delivers promising fidelity on all three testing datasets. First, our method is capable of adapting to different topology in mesh. For example, mesh in Triplegangers dataset has denser vertices in front face than behind while mesh in FaceScape has more uniform density, but our trained-once model performs equally well in both testing datasets. Second, our method does not sacrifice fidelity when input normal is unavailable during testing. Thanks to explicit geometry volume and

robust model, pixels in albedo map are precisely mapped onto the surface and therefore, yields no noise in rendering.

Finally, we acquired and tested our pre-trained model over GeneratedPhotos dataset containing only low-frequency albedo and normal map in fig. 16. The dataset is generated by some anonymous method with only single-view online image as input. Our method not only outputs clear silhouette but also shows realistic pre-level detail when high-frequency input is unavailable.

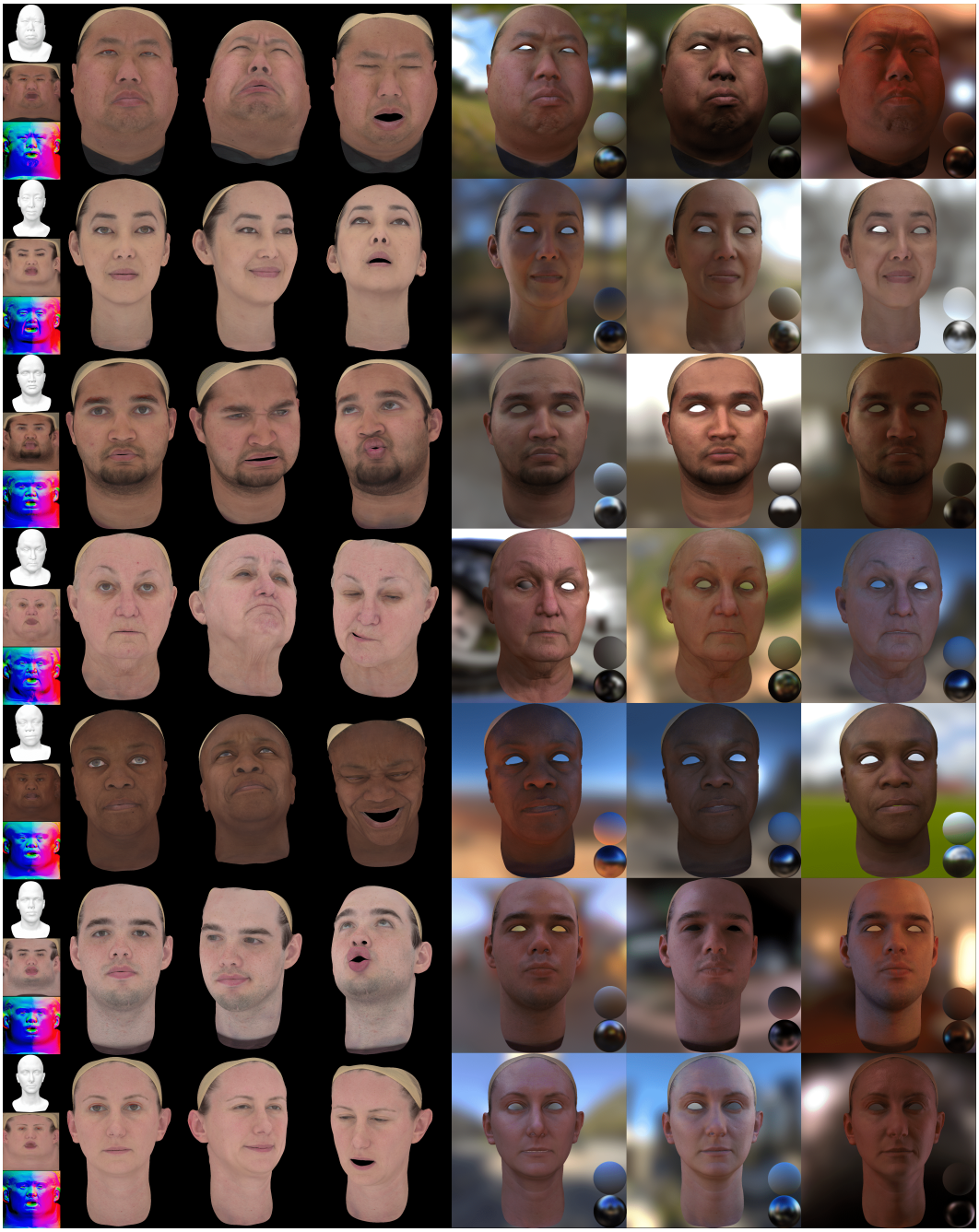
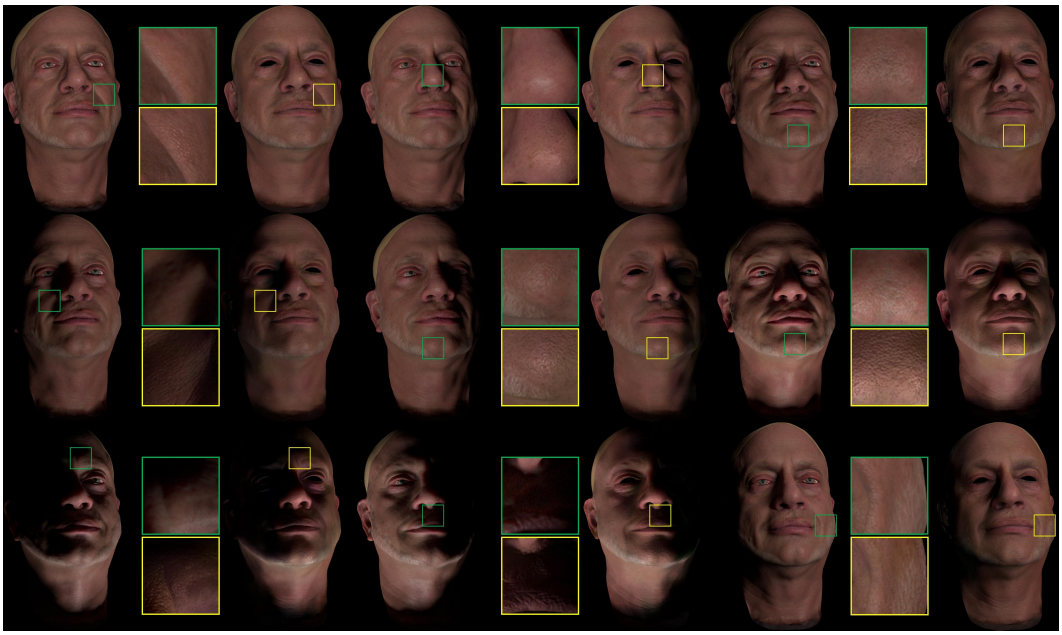


Figure 13: **Qualitative results.** In each row, We present our input (col. 1), three novel view rendering with neutral expression (col. 2), two other expressions under white illumination (col. 3, 4), and three frontal view rendering with neutral expression under different HDRI illumination (col. 5, 6, 7). Geometry, albedo map and normal map in UV space are list in the first column from top to bottom.

(a) HDRI



(b) OLAT

Figure 14: Qualitative comparison between ours and Maya rendering with zoom-in under (a) HDRI and (b) OLAT. Left image and green window in each pair is Maya rendering. Right image and yellow window in each pair is ours. Our result achieves realistic appearance of skin texture in addition to accurate diffuse and ambient illumination effect.

Figure 15: Testing performance of our trained-once models on other open source datasets. To demonstrate fidelity and generalization of our method, we further test our pre-trained model on Triplegangers, FaceScape and 3D Store datasets. Besides accurate shading, our rendering illustrates sharp detail in pore, wrinkle and hair.

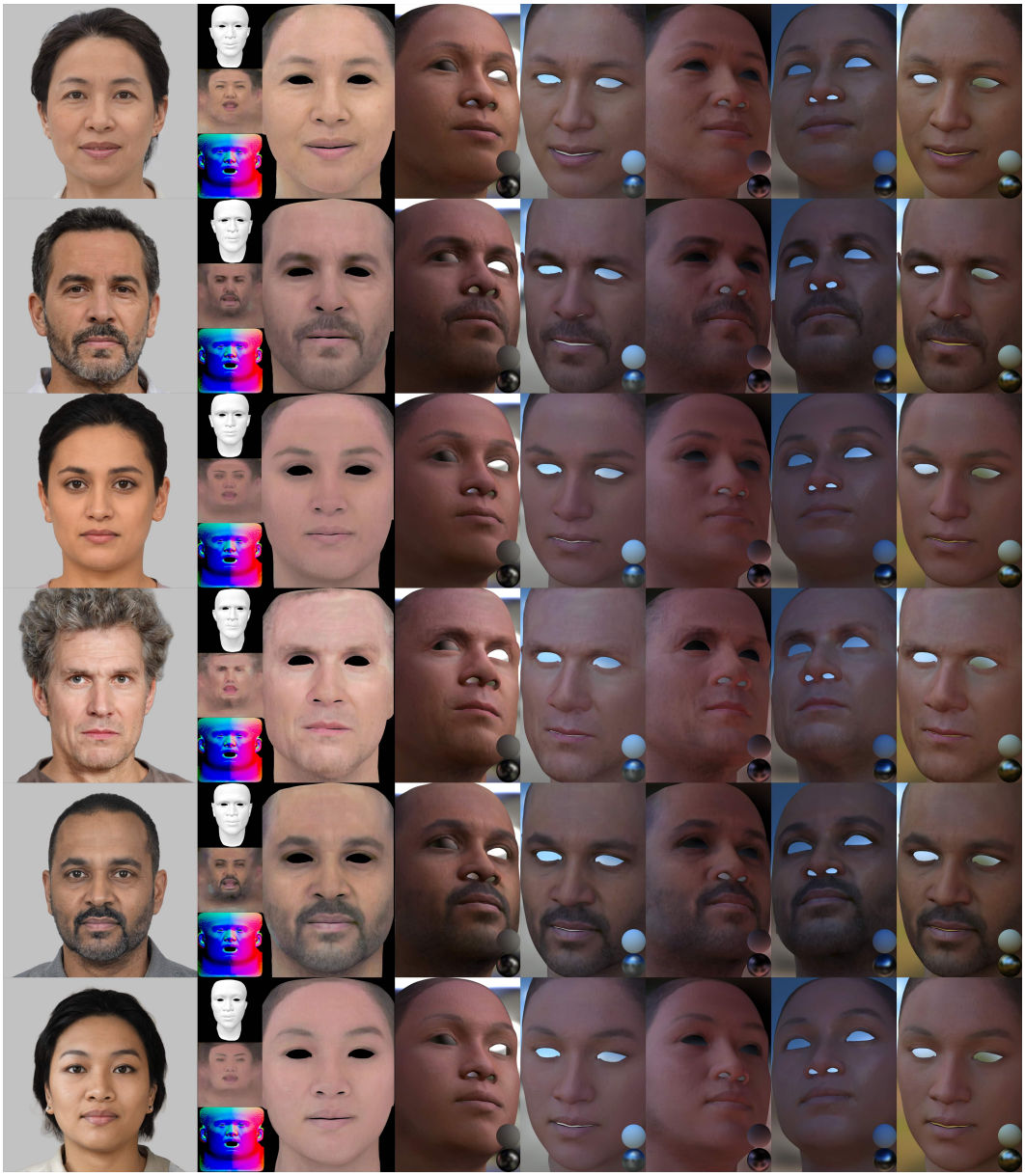


Figure 16: Testing performance of our trained-once models on GeneratedPhotos dataset (in-the-wild). Our model demonstrates high fidelity when accurate scans of albedo and normal map are unavailable in this test.