Vision, Modeling, and Visualization (2025)
B. Egger and T. Günther (Eds.)

# Neural Acquisition & Representation of Subsurface Scattering

Arjun Majumdar[1] Raphael Braun[1] Hendrik Lensch[1]

[1]University of Tübingen, Germany

**Abstract**

*We present a method to acquire and estimate the sub-surface scattering properties of light transport at a highly detailed level by learning the pixel footprint response at each point on the object surface. The reconstruction leverages 3D scanning techniques as input to a U-Net CNN. A stereo projector-camera setup using phase-shifted profilometry (PSP) patterns efficiently captures the data for a variety of scattering objects. Reconstructing dense pixel footprints allows for relighting with arbitrary high-resolution projector patterns. The final output is a relit color image. Qualitative and quantitative comparison against illuminated real-world captured images demonstrate that the predicted footprints are almost identical to the actual responses. The same model is trained for multiple views across multiple objects such that the learned representations can be used to generalize to unseen sub-surface scattering materials as well.*

## 1. Introduction

In light transport *subsurface scattering* refers to the phenomena of light entering a surface, being scattered inside the material and leaving it at another location than where it entered. All non metal materials allow light to enter to various depths, which depends on the individual material. In case of glass like materials light passes through without being scattered. On the other side there are materials like wood, paper or concrete, which are considered opaque, as the light only penetrates to microscopic depths before being scattered back. Those extreme cases can be easily described with analytic BSSRDFs, for glass a clear dielectric model and for opaque materials a diffuse model. However in between there are materials such as wax, marble, organic objects, fruits, leaves, flowers, skin and others, which are not as straight forward. There are analytic models, that describe light transport inside the material as diffusion process [WJMLH23], which in cases of geometry in form of semi-infinite slabs have exact solutions. Recently deep learning based approaches appeared which push the idea of closed form solutions to arbitrary curved surfaces [VKJ19] while maintaining the efficiency of surface based models. Alternatively it is always possible to simulate the volumetric light transport inside a material via path tracing [FHL*18, CFS*18], which however requires a detailed volumetric representation of the optical properties inside an object to get correct results for complex materials such as alabaster or any object with complex volumetric composition. Creating or measuring such a volumetric representation is challenging, as only the scattering result on the surface are directly observable, not the underlying density fluctuations and phase function changes. Instead of describing the complex light transport using surface and volume representation we use an image based rendering approach, where the impact of every light ray on the scene is captured as image. Here relighting an object just means adding together the correct images with the correct scaling factors. In practice it is however very costly to capture such an image dataset. The non-local nature of Subsurface Scattering requires the acquisition of thousands of images, even when the effect of multiple rays are captured in a single picture.

In this paper, we propose a novel, completely data-driven method for image-based modeling of subsurface scattering objects over multiple views and across multiple objects, which after training can predict the required images for relighting without the need for excessive acquisitions. Instead, high-frequency phase-shifted sinusoidal patterns are used as input to a U-Net-based Convolutional Neural Network (CNN) [RFB15]. These patterns are the same patterns that are used to capture precise 3D geometry of the object during 3D scanning. Our U-Net predicts a per pixel anisotropic footprint in image space expressing the subsurface scattering response to incident illumination at that point on the surface. It learns to generalize to unseen views and objects. Furthermore, our method can be trained from images of a single object and, unlike other methods, does not require to learn from many other objects, nor is it probabilistically generative in nature. As a result, it will not suffer from prior distribution bias. Also, since it consists of a single U-Net network, training it is relatively cheap, efficient and easy.

We perform comprehensive qualitative and quantitative evaluations of our learned footprint responses to render the object under various projection patterns and compare the resulting images with the camera captured ground truth to demonstrate the performance of our model on objects with challenging material properties and geometry. Since the focus of our method is on learning and esti-

mating the footprint response of each surface pixel, we perform extensive evaluations on relit results.

In summary, we claim the following contributions:

- We learn anisotropic pixel footprint responses to capture subsurface scattering properties by only using projected high-frequency phase-shift patterns which can at the same time be used for 3D scanning.
- A neural network estimates footprint responses for each point on the object which then are used to perform relighting with spatially varying light patterns in RGB colorspace of the image.
- Our neural network architecture is able to generalize across multiple views and multiple objects, which we verify both qualitatively and quantitatively.

We compute image based sub-surface scattering (SSS) estimation. The actual material parameters for modeling the BSSRDF is not inferred. We only learn anisotropic SSS footprint responses which encodes both the geometry and the material properties and can be used for image based relighting.

## 2. Related Works

The acquisition of material properties has a long-standing tradition, including the measurement of subsurface scattering parameters or general relightable representations, e.g. reflectance fields [DHT*00]. While initial works sampled directly the impulse response to incident illumination, fixed wavelet noise patterns [PD03], compressive sensing [PML*09] or adaptive acquisition schemes [SCG*05, GTLL, OK10, ORK12] exploiting the directionally duality of the light transport have been employed for faster acquisition of general light transport matrices. The required number of acquisition patterns still remains substantial. After some initial training over multiple objects, any new object and new view in our approach only requires about six input images to recover the spatially varying pixel footprint for subsurface scattering.

After the introduction of the first practical BSSRDF model [JMLH01] the specific estimation of subsurface scattering parameters has been performed using dedicated point-wise illumination [JMLH01, WMP*06], For entire objects, Goesele et al. [GLL*04] have presented a system, where a single laser beam scans every surface point to measure the resulting camera footprint, requiring thousands of individual measurements.

A few neural network-based relighting approaches specifically focus on relighting of subsurface scattering objects [ZSS21, LTL*22, YGF*23, ZSB*23, DME*25] from a smaller set of images. However, those approaches assume distant illumination and often use one-light-source-at-a-time (OLAT) as training data.

Structured light scanning patterns, specifically high-frequency patterns, have been used to separate the observed reflected radiance wrt. being caused by local or global illumination [NKGR06]. There are existing methods that leverage high-frequency PSP for sub-surface scattering materials in 3D scanning: [FHL*08, CSL08, CLFS07, GN12, MST15], etc. However, they only focus on removing the unwanted effect of subsurface scattering on the depth estimation rather than measuring and representing the subsurface scattering per se for relighting. Geiger et al. [GHK22] utilize structured light projection to get correct topography for scattering objects as light undergoing volume scattering inside the object results in erroneous outputs. These errors are studied using Monte Carlo simulations, together with an additional method to correct the errors by quantifying the light propagation. Other approaches, e.g. [KLP*96, KKS*23], explicitly exploit the properties of high-frequency binary patterns and polarization filters to directly measure local isotropic scattering parameters of human skin.

To the best of our knowledge, we for the first time use high-frequency PSP images as the only input to reconstruct spatially-varying, potentially anisotropic scattering footprints by a neural network to obtain a more accurate relightable model.

A neural architecture for efficiently sampling Bidirectional Scattering Surface Reflectance Distribution Functions (BSSRDF) on complex 3D surfaces is proposed by Vicini et al. [VKJ19]. The system includes three networks: 1) A Conditional VAE (CVAE), trained on outputs from a slow but accurate volumetric path tracer, learns to generate similar scattering samples much faster by conditioning on input parameters. 2) A scale factor regression MLP adjusts the CVAE output to account for material-dependent absorption differences. 3) A preprocessing feature network transforms raw inputs into a learned feature space, enabling more effective conditioning for the CVAE and MLP.

Our method, by comparison, does not depend on any pre-trained VAE method and learns the anisotropic footprint responses directly from the input images. Since it does not depend on multiple networks and just consists of a single U-Net CNN architecture, training is more efficient and easier. VAEs model complex data distributions by maximizing the evidence lower bound (ELBO) loss and minimizing the reconstruction loss [RGB*22]. They might add bias due to the noisy, uninformative prior that is used in KL-loss computation, whereas our method learns the actual latent representation, thereby remaining bias-free. Lastly, VAEs might also suffer from posterior collapse [WBC21, DWW20, LTGN19], which our CNN does not. Our primary focus is to learn from the camera-captured ground truth anisotropic footprint responses as faithfully as possible.

## 3. Background

Our acquisition of subsurface scattering representations is based on N-step phase-shifted profilometry (PSP) which typically is applied to perform accurate 3D scanning.

Assuming a camera/projector setup with a linearized projector, a set of virtual sinusoidal patterns can be created and then projected onto the object in the scene, captured by the corresponding offset camera. These virtual sinusoidal patterns in the projector space can be represented as:

$$I^p(x^p, y^p) = a^p + b^p cos(2\pi f_0^p x^p + \delta_n), \quad (1)$$

where $(x_p, y_p)$ represents the projector pixel coordinate, $a^p$ depicts the direct component of intensity, $b^p$ represents the amplitude and $f_0^p$ is the frequency of the sinusoidal fringe in period per pixel. $\delta_n$ is the phase-shift index and can be represented as $\delta_n = \frac{2\pi}{N} n$ for the standard N-step PSP [ZFH*18].

The resulting camera images can then be expressed as:

$$I_n(x,y) = A(x,y) + B(x,y)cos\left(\phi(x,y) - \frac{2\pi}{N}\right) \quad (2)$$

where, $A(x,y)$ is the average intensity for the pattern brightness and background illumination, $B(x,y)$ is the intensity modulation pertaining to pattern contrast and surface reflectivity and $n$ is the phase-shift index ranging from $n = 0, 1, ...N$.

The wrapped phase map can be computed by using the equation for each point (x,y) in camera or image space:

$$\phi(x,y) = \tan^{-1}\frac{\sum_{n=0}^{N-1} I_n(x,y)\sin(2\pi n/N)}{\sum_{n=0}^{N-1} I_n(x,y)\cos(2\pi n/N)} \quad (3)$$

The wrapped phase $\phi(x,y)$ is in the range $[-\pi,\pi]$. Since there are three unknowns: $A(x,y)$, $B(x,y)$ and $\phi(x,y)$, we usually require three images to compute them [ZFH*18, ZHZ*16]. To convert the wrapped phase to unwrapped or continuous phase map, we use the equation:

$$\Phi(x,y) = \phi(x,y) + 2\pi k(x,y), \quad (4)$$

where $k(x,y)$ is the fringe integer number that indicates the fringe orders. The majority of the phase-unwrapping algorithms deal with computing this fringe integer number as correctly, quickly and faithfully as possible [ZHZ*16, ZFH*18]. The unwrapped phase directly correlates to the disparity between the camera and the projector in a stereo setup and therefore is related to depth.

Broadly, PSP is divided into spatial and temporal phase unwrapping (TPU), where in the former, neighboring pixels' information is used to unwrap a pixel, whereas in the latter, only time-varying information is used to unwrap a pixel. Consequently, temporal phase unwrapping is more robust and is the preferred way as it has a better ability to measure surface discontinuities [HK21].

However, for challenging objects, especially for subsurface scattering objects, traditional PSP TPU is insufficient due to the global effects of subsurface scattering disturbing the acquired lower unit frequency phase maps, which are required to unwrap the higher frequencies [CSL08, CLFS07]. An alternate approach for such challenging materials and complex geometrical objects employs interferometry-based hierarchical phase-unwrapping. Instead of projecting and capturing a unit-frequency, they use the principles of interferometry wherein a synthetic wavelength is used for unwrapping the highest-frequency phase map, which is free from artifacts due to global effects such as subsurface scattering. Additionally, high-frequency sinusoidal patterns can also be leveraged to decompose a captured scene into its direct and global components [NKUR23]. Here, the recovered amplitude $B$ is directly related to the local component due direct illumination turned on or off, while the recovered average $A$ corresponds to the global component representing all global illumination effects such as ambient illumination, interreflections and, relevant for us, subsurface scattering.

U-Net is a Convolutional Neural Network (CNN) which was originally proposed for medical domain based image segmentation task by outputting a mask for the given input images [RFB15].

## 4. Method

Our image and learning-based model of subsurface scattering assumes the additive nature of light. We estimate the anisotropic pixel footprint response at each point of a subsurface scattering object by using a set of high-frequency (4 pixels per cycle) PSP images. We employ a camera/projector stereo setup. A BenQ X1300I 3D DLP projector of 1080x1920 pixel resolution projects different patterns onto the subsurface object in the scene, which is then captured by a FLIR Oryx 10GigE camera of 3000x3200 resolution. Fig. 1 shows an example of a scattering candle: for 3-step PSP, three phase-shifted horizontal patterns with a phase-shift of $\frac{2\pi}{3}$ are visualized together with their corresponding unwrapped phase map. We use this 3D scanned output to obtain a pixel-wise mapping for each camera pixel to its corresponding projector pixel. This correspondence can be used for 3D reconstruction and is used during relighting to know which projector pixel illuminates which camera coordinate directly.

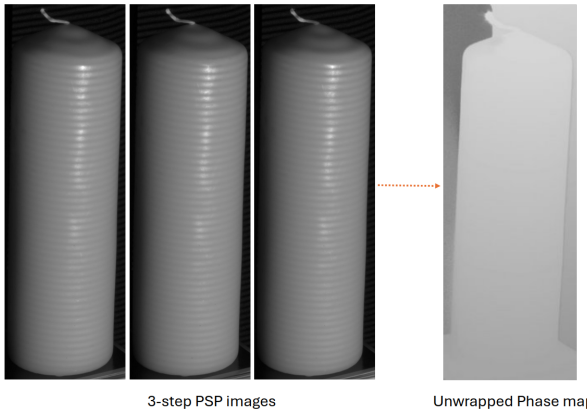



**Figure 1:** *3-step phase-shifted proliferometry (PSP) example. The same pattern can be used to measure disparity as well as for recovering spatially varying scattering footprints.*

For one view of an object we capture the pixel responses due to subsurface scattering explicitly and use them as ground truth for training the U-net. We capture the subsurface scattering-based pixel footprint responses by projecting a grid of dots $(m,m)$ onto the object. Here, $m$ denotes the distance in pixels in both the x and y axes in the projector space. Capturing each image with just one projected dot onto the surface of the object is clearly infeasible due to time and space complexity reasons. Therefore, we parallelize the capture of multiple surface scattering points by projecting multiple dots onto the object and then capturing all of them in a single camera image. The total number of acquired images is therefore $m \times m$. For full ground truth acquisition, we shift the grid by 1 pixel such that in the end, we have covered the entire object's surface area at least once. The pixel distance $m$ needs to be chosen such that the observed footprints clearly do not overlap. If $m$ is small, i.e., 40 pixels or less, then the light from the neighboring adjacent coordinates might affect the response for the current coordinate and thereby lead to a biased captured response. We use $(55,55)$ as standard for all our acquisitions. An example acquisition for a bar of soap is

shown in Fig. 2. The two images depict the shifted projected and captured patterns.

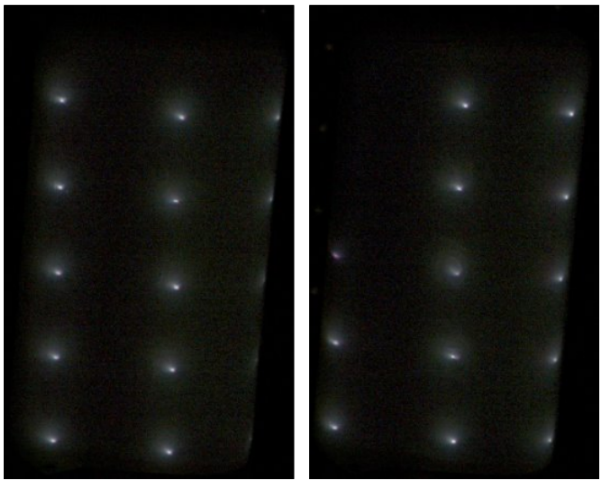


**Figure 2:** *Camera captured target ground truth. As training data we acquire densely sampled scattering footprints by illuminating individual dots with a shifting grid.*

The camera captured images are in raw, mosaiced, luminance-only data format. We apply the following data pre-processing steps on them before using them for our deep learning training: (i) Median filtering: we apply channel-wise filtering for *Red*, $Green_1$, $Green_2$ and *Blue* channels to get denoised outputs (ii) Demosaicing: [MHC04], [MAC07], [Get11] algorithms are used to obtain demosaiced RGB colored images. (iii) We use Segment Anything Model [KMR*23] to get a mask of the object while excluding the background so that only footprint responses on the object are used for training the U-Net and no other non-scattering response can harm the learned representations (iv) high-frequency PSP input images are captured as input for U-Net, which never sees the target dots

The processed data is split into train-test sets at a 85:15 ratio.

### 4.1. Learning Footprint Prediction

For estimating the pixel footprints from the horizontal and vertical high-frequency PSP input we use an Attention U-Net architecture [RFB15, OSF*18]. It learns to output the anisotropic footprint response at that point as schematically shown in Figure 3. Please note that during inference, the U-Net never sees any of the footprint response. It learns to estimate these only from the high-frequency PSP input patterns at each point of the object.

A vanilla U-Net was not able to faithfully learn the anisotropic footprint responses as the anisotropy seems to vary for each surface point on the object for each view and across objects. Therefore, we shifted to an Attention U-Net variant [OSF*18] where the attention gate allows the U-Net to focus on targets of varying shapes and sizes. For all of our experiments, we use the Attention U-Net variant only. Details about the deep learning training is provided in the supplementary.

Due to the extreme distribution in the footprints with very few rather bright spots, some anisotropic decay and otherwise many rather small values which are overlapping with the camera's noise floor, it is a challenge to determine a proper loss function.

Initially the Attention U-Net was trained with a Mean-Squared Error (MSE) cost function:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5}$$

Since the majority of each input patch is dark and thereby the resulting output of the network might also be proportionately small, the MSE often produces very small values. As a result, the gradient signal might be insufficient to update the parameters of the network. Trying to alleviate this, we also experimented with Mean Squared Logarithmic Error (MSLE) cost function:

$$\text{MSLE} = \frac{1}{n}\sum_{i=1}^{n}(log(y_i) - log(\hat{y}_i))^2 \tag{6}$$

This log transformation leads to good loss values and proper gradient signals, but it truncates the smooth anisotropic falloff as it punishes small deviations heavily. Consequently, the learned footprint responses don't produce the smooth blurring visualization effects, which is a typical characteristic of sub-surface scattering.

Finally, we settled with an inverse-Gaussian weighted MSE cost function with the following weighting scheme: $P' = P + \epsilon$, $W = w_{\max} \cdot \exp\left(-\frac{(\mathbf{P}')^2}{2\sigma^2}\right)$ and $W = \text{clamp}(\mathbf{W}, 1.0, w_{\max})$ where $\mathbf{P} \in \mathbb{R}^{B\times C\times H\times W}$ is the target subsurface scattering footprint, $\sigma = 0.1$ controls the decay rate of the weighting function, $\epsilon = 10^{-8}$ is a small numerical shift to avoid division by zero, $w_{\max} = 100.0$ is the maximum weight value, and $\mathbf{W}$ are the resulting inverse Gaussian weights that assign higher values to pixels with smaller target values.

The weighted MSE cost function used for all of our trainings are:

$$\text{inv wtd Gauss MSE} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{W_i}\cdot y_i - \mathbf{W_i}\cdot \hat{y}_i)^2 \tag{7}$$

The Attention U-Net is simultaneously trained on the footprints of a couple of different objects, typically providing one or two views of the pixel-grid pattern of Figure 2. The resulting network, however, generalizes to other views and other objects with different material properties.

In total, we train the network for 150 epochs on roughly 29k pairs of sinusoidal tiles and footprints in total. Although this might vary depending on the size of the object. We use Adam (lr = $1 \times 10^{-4}$) as our gradient-descent optimizer. We monitor both train and test losses for the train-test splits that help in monitoring the progress of the neural network training, diagnostics and troubleshooting.

### 4.2. Inference and Relighting

For a novel view of an object, n-step PSP patterns are acquired with in total 54 images. Only the six highest frequency images (horizontal and vertical with three phase shifts each) will be used for the

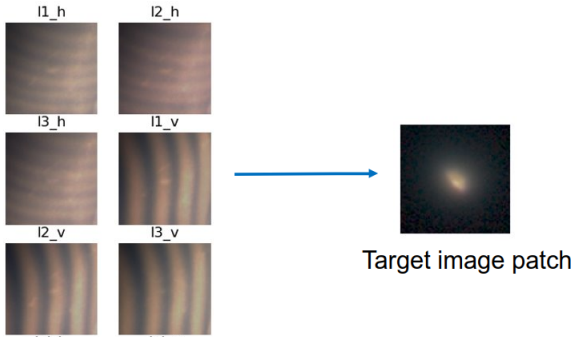



**Figure 3:** *Attention U-Net data flow. Given horizontal and vertical PSP pattern around camera pixel $x,y$ as input, the U-Net predicts the corresponding anisotropic footprint.*

| Object+View | Dataset | MSE* | PSNR | SSIM | LPIPS* |
|---|---|---|---|---|---|
| Soap-front | Train | 4.07 | 29.59 | 0.997 | 6.33 |
| Soap-front | Test | 4.09 | 29.59 | 0.997 | 5.99 |
| Soap-back | Train | 3.89 | 27.43 | 0.997 | 3.95 |
| Soap-back | Test | 3.81 | 27.48 | 0.997 | 3.73 |
| Orange-front | Train | 4.09 | 28.89 | 0.996 | 3.76 |
| Orange-front | Test | 4.83 | 28.85 | 0.996 | 3.45 |
| Orange-back | Train | 3.6 | 33.4 | 0.998 | 4.85 |
| Orange-back | Test | 3.4 | 33.41 | 0.998 | 4.14 |
| Leaf-front | Train | 3.75 | 23.82 | 0.996 | 0.93 |
| Leaf-front | Test | 3.78 | 23.76 | 0.996 | 0.91 |
| Leaf-back | Train | 6.25 | 26.91 | 0.996 | 14.4 |
| Leaf-back | Test | 6.01 | 26.93 | 0.996 | 13 |

**Table 1:** *Quantitative metrics on footprint reconstruction. *(MSE$\times 10^{-7}$, LPIPS$\times 10^{-4}$).*

footprint prediction, the other 48 images are just used for hierarchical temporal phase unwrapping for phase-shift profilometry fringe pattern projections.

We remove the background that did not receive sufficient projector illumination. For all remaining camera pixels, the U-Net predicts a separate footprint by cropping the corresponding input tiles around the pixel.

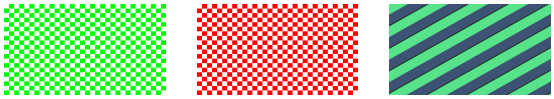
For relighting, the unwrapped camera-projector pixel correspondence map is used to identify which projector pixel (with subpixel accuracy) illuminates the corresponding camera pixel. The bilinearly interpolated project pixel's color $p(x,y)$ will be used as a color weight to multiply the camera pixel's footprint. Finally, the weighted footprint is splatted into the final framebuffer until all camera pixels are processed.

We perform white balance color correction by projecting and capturing pure red, green and blue images and computing a color correction matrix using Least Squares/Pseudo-inverse method against a reconstructed red, green and blue relit image. The camera captured pure red, green, blue and white images have known colors which are normalized to a maximum intensity of 1.0. We compute a white balance transformation matrix by solving a system of least squares where $I_{\text{rec}} \cdot \mathbf{X} = I_{\text{cam}}$ where the pixels of the reconstructed red, green and blue images in the vector $I_{\text{rec}}$ times the unknown color correction matrix $\mathbf{X}$ equals the captured pixel colors $I_{\text{cam}}$. This transformation matrix $\mathbf{X}$ is later multiplied to every reconstructed output images to obtain color corrected results. During acquisition we ensure that the cameras white balance settings stay constant to ensure the transferability of $\mathbf{X}$ between acquisitions.

## 5. Results

To demonstrate the performance of our approach we investigate both the quality of the estimated footprint as well as the resulting relit output.

We can observe that the network does learn the desired anisotropic footprint responses very well in general and does not overfit to the training set, but also generalizes to the test set. To relight the object with different virtual projector patterns of (1080, 1920) resolution, we use the following three images shown in Fig. 4.



(a) green-white (b) red-white (c) gradient

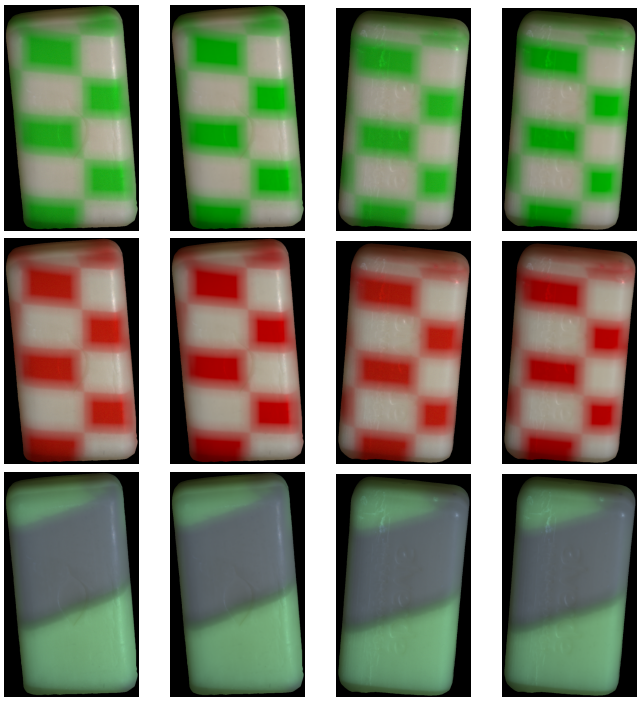
**Figure 4:** *Virtual Projector Patterns*

Fully relit results are presented for a commercially available soap in Figure 5, a leaf obtained from a nearby forest in Figure 7 and an orange fruit in Figure 6. An interesting side note for the captured leaf is that its rate of decomposition is very fast which means that it starts to deform in both shape and size just after being plucked from its tree. This can pose to be a problem if the SAM mask was generated before, while the rest of the acquisition takes time which might be compounded when acquiring multiple views of the object.

Relighting using the predicted footprints after the described color calibration (see Section 4.2) leads to images that are difficult to distinguish from ground truth. Specifically, the subsurface scattering-induced smooth transition between differently colored patches and the object-dependent color shifts are well captured. In addition, all geometric intricacies are reproduced. With image based rendering, all material properties, including highlights are baked in into the footprints. They would not move if one would rotate the virtual object.

We also provide quantitative evaluations for the relit results compared objects captured from multiple views in Table 2. These numbers underline the very close match between the relit images based on the reconstructed footprints and the ground truth.

Finally, we use the pre-trained attention U-Net, which has been trained on the training objects from multiple views for an unseen soap (Soap2) also for multiple views (see Figure 8 and supplemental materials). Even though the U-Net has never seen any input of the second soap during training, all object-specific features, such as the appearance of the imprint as well as the brighter veins and the scattering around them are perfectly reconstructed. The corresponding metrics are listed in Table 3.



(a) Target (b) Output (c) Target (d) Output

**Figure 5:** *Comparison: Target & Relit Images - Soap front view (first two columns), back view (last two columns). Note how the smooth transition between different colors as well as the indents on the front are accurately reproduced.*

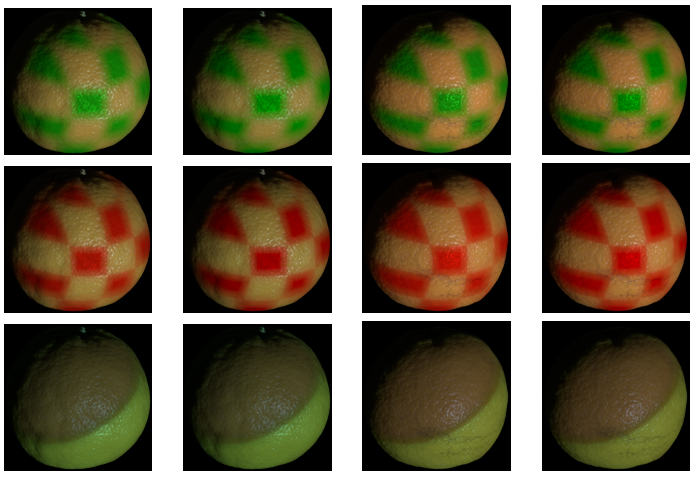
| Object+View | Pattern | MSE* | PSNR | SSIM | LPIPS* |
|---|---|---|---|---|---|
| Soap-front | red-white | 6.14 | 52.00 | 0.995 | 5.31 |
| Soap-front | green-white | 8.42 | 49.39 | 0.994 | 7.84 |
| Soap-front | gradient | 5.75 | 47.82 | 0.999 | 2.84 |
| Soap-back | red-white | 9.52 | 50.02 | 0.996 | 4.85 |
| Soap-back | green-white | 6.22 | 52.04 | 0.995 | 7.41 |
| Soap-back | gradient | 4.70 | 46.44 | 0.999 | 2.18 |
| Orange-front | red-white | 1.62 | 46.52 | 0.997 | 2.18 |
| Orange-front | green-white | 9.5 | 48.36 | 0.998 | 1.9 |
| Orange-front | gradient | 1.0 | 55.81 | 0.999 | 3.18 |
| Orange-back | red-white | 3.6 | 54.44 | 0.998 | 1.07 |
| Orange-back | green-white | 8.56 | 49.83 | 0.998 | 1.25 |
| Orange-back | gradient | 0.8 | 54.01 | 0.999 | 0.27 |

**Table 2:** *Quantitative metrics on relit results.* $^*(MSE\times10^{-6}, LPIPS\times10^{-3})$.

We use a frozen U-Net pre-trained on Soap, Orange and Leaf to relite a geometrically challenging 3D printed object to show that our method generalizes to such objects as well, which is provided in the supplementary.

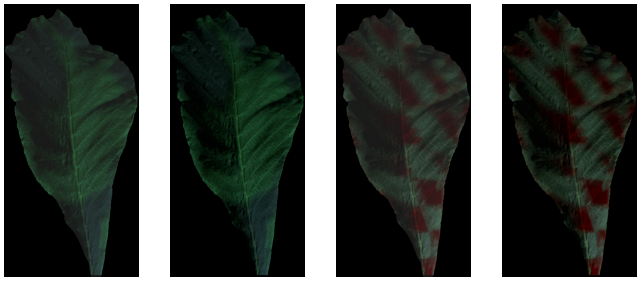
## 6. Conclusions

To obtain a relightable representation for subsurface scattering objects, we present a novel technique that can predict the per-pixel spatially varying pixel footprints from just six input images cap-



(a) Target (b) Output (c) Target (d) Output

**Figure 6:** *Comparison: Target & Relit Images - Orange front view (first two columns), back view (last two columns). Here, the color shift due to the orange peel as well as the peels structure are captured accurately.*

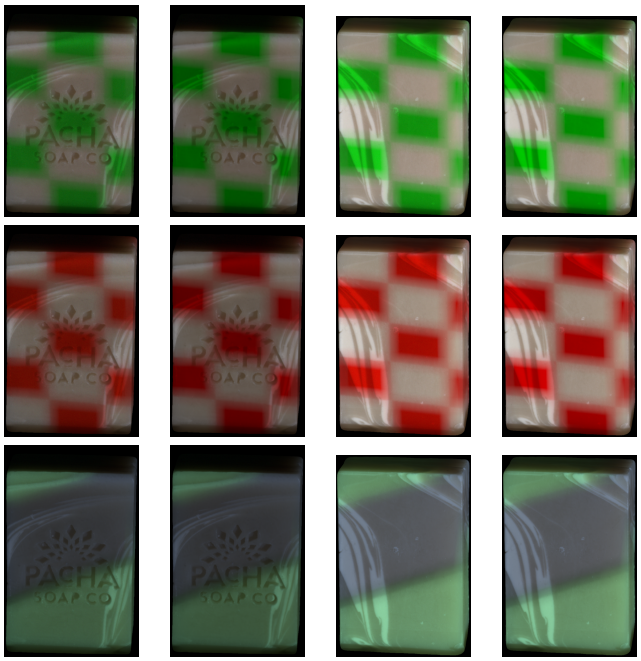


(a) Target (b) Output (c) Target (d) Output

**Figure 7:** *Comparison: Target & Relit Images - Leaf front view*

| Object+View | Pattern | MSE* | PSNR | SSIM | LPIPS* |
|---|---|---|---|---|---|
| Soap-front | red-white | 1.45 | 37.54 | 0.997 | 3.9 |
| Soap-front | green-white | 1.37 | 37.28 | 0.995 | 5.0 |
| Soap-front | gradient | 0.29 | 43.04 | 0.999 | 1.3 |
| Soap-back | red-white | 1.28 | 48.18 | 0.998 | 1.1 |
| Soap-back | green-white | 1.22 | 48.09 | 0.997 | 1.5 |
| Soap-back | gradient | 0.22 | 52.24 | 0.999 | 0.33 |

**Table 3:** *Quantitative metrics on relit images for the unseen object Soap2.* $^*(MSE\times10^{-5}, LPIPS\times10^{-3})$.



(a) Target (b) Output (c) Target (d) Output

**Figure 8:** *Comparison: Target & Relit Images - **Unseen** Soap2 front & back view. No images of this object had been used during training.*

tured with high-frequency horizontal and vertical sinusoidal patterns. A trained U-Net translates those phase-shifted profilometry patterns, which at the same time can be used for 3D scanning. The training of the U-Net requires a paired dataset of sinusoidal patterns with corresponding impulse response scattering footprints. But those could be acquired for a completely different view or on different objects, while the CNN generalized well to novel views and novel objects. The predicted scattering functions accurately model the actual scattering behaviour, resulting in relit images that are hard to distinguish from real-world captured objects. While our approach allows for subsurface scattering representations from very few measurements, in future work, a couple of further effects will need to be addressed: This should cover the separation of surface reflections and subsurface scattering, scattering effects that extend beyond the current size of the estimated footprint, as well as angular dependencies.

## 7. Acknowledgement

The work described in this paper was funded by: Sony Europe Limited, Germany. The authors thank Prasan Ashok Shedligeri, Zoltan Facius and Alexander Gatto for their support. The author would also like to extend their gratitude and thanks to the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Arjun Majumdar.

## References

[CFS*18] CHRISTENSEN P., FONG J., SHADE J., WOOTEN W., SCHUBERT B., KENSLER A., FRIEDMAN S., KILPATRICK C., RAMSHAW C., BANNISTER M., RAYNER B., BROUILLAT J., LIANI M.: Renderman: An advanced path-tracing architecture for movie rendering. *ACM TOG 37* (2018). 1

[CLFS07] CHEN T., LENSCH H. P. A., FUCHS C., SEIDEL H.-P.: Polarization and phase-shifting for 3d scanning of translucent objects. In *2007 IEEE Conference on Computer Vision and Pattern Recognition* (2007), pp. 1–8. doi:10.1109/CVPR.2007.383209. 2, 3

[CSL08] CHEN T., SEIDEL H.-P., LENSCH H. P. A.: Modulated phase-shifting for 3d scanning. In *2008 IEEE Conference on Computer Vision and Pattern Recognition* (2008), pp. 1–8. doi:10.1109/CVPR.2008.4587836. 2, 3

[DHT*00] DEBEVEC P., HAWKINS T., TCHOU C., DUIKER H.-P., SAROKIN W., SAGAR M.: Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (2000), pp. 145–156. 2

[DME*25] DIHLMANN J.-N., MAJUMDAR A., ENGELHARDT A., BRAUN R., LENSCH H.: Subsurface scattering for gaussian splatting. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems* (2025). 2

[DWW20] DAI B., WANG Z., WIPF D.: The usual suspects? Reassessing blame for VAE posterior collapse. In *Proceedings of the 37th International Conference on Machine Learning* (13–18 Jul 2020), III H. D., Singh A., (Eds.), vol. 119 of *Proceedings of Machine Learning Research*, PMLR, pp. 2313–2322. URL: https://proceedings.mlr.press/v119/dai20c.html. 2

[FHL*08] FUCHS C., HEINZ M., LEVOY M., SEIDEL H.-P., LENSCH H. P.: Combining confocal imaging and descattering. In *Computer Graphics Forum* (2008), vol. 27, Wiley Online Library, pp. 1245–1253. 2

[FHL*18] FASCIONE L., HANIKA J., LEONE M., DROSKE M., SCHWARZHAUPT J., DAVIDOVIČ T., WEIDLICH A., MENG J.: Manuka: A batch-shading architecture for spectral path tracing in movie production. *ACM TOG 37* (2018). 1

[Get11] GETREUER P.: Malvar-He-Cutler Linear Image Demosaicking. *Image Processing On Line 1* (2011), 83–89. https://doi.org/10.5201/ipol.2011.g_mhcd. 4

[GHK22] GEIGER S., HANK P., KIENLE A.: Improved topography reconstruction of volume scattering objects using structured light. *Journal of the Optical Society of America A 39* (2022). 2

[GLL*04] GOESELE M., LENSCH H. P., LANG J., FUCHS C., SEIDEL H.-P.: Disco: acquisition of translucent objects. In *ACM SIGGRAPH 2004 Papers*. 2004, pp. 835–844. 2

[GN12] GUPTA M., NAYAR S. K.: Micro phase shifting. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (2012), pp. 813–820. doi:10.1109/CVPR.2012.6247753. 2

[GTLL] GARG G., TALVALA E.-V., LEVOY M., LENSCH H. P.: Symmetric photography: Exploiting data-sparseness in reflectance fields. 2

[HK21] HE X., KEMAO Q.: A comparative study on temporal phase unwrapping methods in high-speed fringe projection profilometry. *Optics and Lasers in Engineering 142* (2021). doi:https://doi.org/10.1016/j.optlaseng.2021.106613. 3

[JMLH01] JENSEN H. W., MARSCHNER S. R., LEVOY M., HANRAHAN P.: A practical model for subsurface light transport. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques* (2001), pp. 511–518. 2

[KKS*23] KIKUCHI K., KATSUYAMA M., SHIBATA T., HARDEBERG J., YUASA T., AIZU Y.: Development of measurement system for subsurface scattering light of skin and analysis of its age-related changes. In *Biomedical Imaging and Sensing Conference* (2023), vol. 12608, SPIE, pp. 123–126. 2

[KLP*96] KIENLE A., LILGE L., PATTERSON M. S., HIBST R., STEINER R., WILSON B. C.: Spatially resolved absolute diffuse reflectance measurements for noninvasive determination of the optical scattering and absorption coefficients of biological tissue. *Applied optics 35*, 13 (1996), 2304–2314. 2

[KMR*23] KIRILLOV A., MINTUN E., RAVI N., MAO H., ROLLAND C., GUSTAFSON L., XIAO T., WHITEHEAD S., BERG A. C., LO W.-Y., DOLLÁR P., GIRSHICK R.: Segment anything. *arXiv:2304.02643* (2023). 4

[LTGN19] LUCAS J., TUCKER G., GROSSE R. B., NOROUZI M.: Don't blame the elbo! a linear vae perspective on posterior collapse. In *Advances in Neural Information Processing Systems* (2019), Wallach H., Larochelle H., Beygelzimer A., d'Alché-Buc F., Fox E., Garnett R., (Eds.), vol. 32, Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper_files/paper/2019/file/7e3315fe390974fcf25e44a9445bd821-Paper.pdf. 2

[LTL*22] LYU L., TEWARI A., LEIMKÜHLER T., HABERMANN M., THEOBALT C.: Neural radiance transfer fields for relightable novel-view synthesis with global illumination. In *European Conference on Computer Vision* (2022), Springer, pp. 153–169. 2

[MAC07] MENON D., ANDRIANI S., CALVAGNO G.: Demosaicing with directional filtering and a posteriori decision. *IEEE Transactions on Image Processing 16*, 1 (2007), 132–141. doi:10.1109/TIP.2006.884928. 4

[MHC04] MALVAR H., HE L.-W., CUTLER R.: High-quality linear interpolation for demosaicing of bayer-patterned color images. vol. 3, pp. iii – 485. doi:10.1109/ICASSP.2004.1326587. 4

[MST15] MORENO D., SON K., TAUBIN G.: Embedded phase shifting: Robust phase shifting with embedded signals. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), pp. 2301–2309. doi:10.1109/CVPR.2015.7298843. 2

[NKGR06] NAYAR S. K., KRISHNAN G., GROSSBERG M. D., RASKAR R.: Fast separation of direct and global components of a scene using high frequency illumination. In *ACM SIGGRAPH 2006 Papers*. 2006, pp. 935–944. 2

[NKUR23] NAYAR S. K., KRISHNAN G., UNIVERSITY C., RASKAR R.: *Fast Separation of Direct and Global Components of a Scene Using High Frequency Illumination*, 1 ed. Association for Computing Machinery, New York, NY, USA, 2023. URL: https://doi.org/10.1145/3596711.3596765. 3

[OK10] O'TOOLE M., KUTULAKOS K. N.: Optical computing for fast light transport analysis. *ACM Trans. Graph. 29*, 6 (2010), 164. 2

[ORK12] O'TOOLE M., RASKAR R., KUTULAKOS K. N.: Primal-dual coding to probe light transport. *ACM Trans. Graph. 31*, 4 (2012), 39–1. 2

[OSF*18] OKTAY O., SCHLEMPER J., FOLGOC L. L., LEE M., HEINRICH M., MISAWA K., MORI K., MCDONAGH S., HAMMERLA N. Y., KAINZ B., GLOCKER B., RUECKERT D.: Attention u-net: Learning where to look for the pancreas. In *Medical Imaging with Deep Learning* (2018). URL: https://openreview.net/forum?id=Skft7cijM. 4

[PD03] PEERS P., DUTRÉ P.: Wavelet environment matting. In *Proceedings of the 14th Eurographics workshop on Rendering* (2003), pp. 157–166. 2

[PML*09] PEERS P., MAHAJAN D. K., LAMOND B., GHOSH A., MATUSIK W., RAMAMOORTHI R., DEBEVEC P.: Compressive light transport sensing. *ACM Transactions on Graphics (ToG) 28*, 1 (2009), 1–18. 2

[RFB15] RONNEBERGER O., FISCHER P., BROX T.: U-net: Convolutional networks for biomedical image segmentation. *CoRR abs/1505.04597* (2015). URL: http://dblp.uni-trier.de/db/journals/corr/corr1505.html#RonnebergerFB15. 1, 3, 4

[RGB*22] REIZINGER P., GRESELE L., BRADY J., VON KÜGELGEN J., ZIETLOW D., SCHÖLKOPF B., MARTIUS G., BRENDEL W., BESSERVE M.: Embrace the gap: Vaes perform independent mechanism analysis. In *Advances in Neural Information Processing Systems* (2022), Koyejo S., Mohamed S., Agarwal A., Belgrave D., Cho K., Oh A., (Eds.), vol. 35, Curran Associates, Inc., pp. 12040–12057. URL: https://proceedings.neurips.cc/paper_files/paper/2022/file/4eb91efe090f72f7cf42c69aab03fe85-Paper-Conference.pdf. 2

[SCG*05] SEN P., CHEN B., GARG G., MARSCHNER S. R., HOROWITZ M., LEVOY M., LENSCH H. P.: Dual photography. In *ACM SIGGRAPH 2005 Papers*. 2005, pp. 745–755. 2

[VKJ19] VICINI D., KOLTUN V., JAKOB W.: A learned shape-adaptive subsurface scattering model. *ACM TOG 38* (2019). 1, 2

[WBC21] WANG Y., BLEI D., CUNNINGHAM J. P.: Posterior collapse and latent variable non-identifiability. In *Advances in Neural Information Processing Systems* (2021), Ranzato M., Beygelzimer A., Dauphin Y., Liang P., Vaughan J. W., (Eds.), vol. 34, Curran Associates, Inc., pp. 5443–5455. URL: https://proceedings.neurips.cc/paper_files/paper/2021/file/2b6921f2c64dee16ba21ebf17f3c2c92-Paper.pdf. 2

[WJMLH23] WANN JENSEN H., MARSCHNER S. R., LEVOY M., HANRAHAN P.: *A Practical Model for Subsurface Light Transport*, 1 ed. Association for Computing Machinery, New York, NY, USA, 2023. URL: https://doi.org/10.1145/3596711.3596747. 1

[WMP*06] WEYRICH T., MATUSIK W., PFISTER H., BICKEL B., DONNER C., TU C., MCANDLESS J., LEE J., NGAN A., JENSEN H. W., ET AL.: Analysis of human faces using a measurement-based skin reflectance model. *ACM Transactions on Graphics (ToG) 25*, 3 (2006), 1013–1024. 2

[YGF*23] YU H.-X., GUO M., FATHI A., CHANG Y.-Y., CHAN E. R., GAO R., FUNKHOUSER T., WU J.: Learning object-centric neural scattering functions for free-viewpoint relighting and scene composition. *arXiv preprint arXiv:2303.06138* (2023). 2

[ZFH*18] ZUO C., FENG S., HUANG L., TAO T., YIN W., CHEN Q.: Phase shifting algorithms for fringe projection profilometry: A review. *Optics and Lasers in Engineering 109* (2018). doi:doi:10.1016/j.optlaseng.2018.04.019. 2, 3

[ZHZ*16] ZUO C., HUANG L., ZHANG M., CHEN Q., ASUNDI A.: Temporal phase unwrapping algorithms for fringe projection profilometry: A comparative review. *Optics and Lasers in Engineering 85* (2016). doi:OpticsandLasersinEngineering. 3

[ZSB*23] ZHU S., SAITO S., BOZIC A., ALIAGA C., DARRELL T., LASSNER C.: Neural relighting with subsurface scattering by learning the radiance transfer gradient. *arXiv preprint arXiv:2306.09322* (2023). 2

[ZSS21] ZHENG Q., SINGH G., SEIDEL H.-P.: Neural relightable participating media rendering. *Advances in Neural Information Processing Systems 34* (2021), 15203–15215. 2