0000-0001-6668-1556

A Neural Reflectance Field Model for Accurate Relighting in RTI Applications

SHAMBEL FENTE MENGISTU, FILIPPO BERGAMASCO, and MARA PISTELLATO, Dipartimento di Scienze Ambientali, Informatica e Statistica (DAIS), Università Ca'Foscari Venezia, Italy

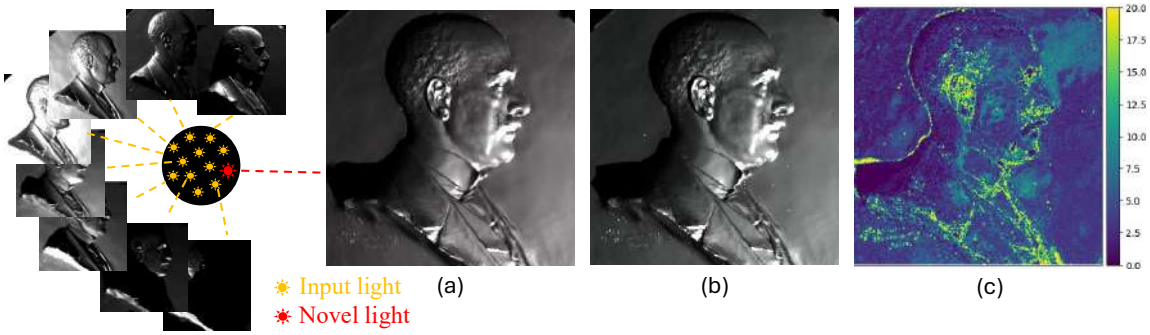

Fig. 1. Our method takes as input a set of images of an object captured from various lighting directions (in this example we used 80 input images) and produces images under novel light directions (a). Panel (b) shows the ground truth image for the selected novel light, not present in the training set, while (c) is the error map depicting Euclidean distance between ground-truth and reconstructed images in RGB space.

Reflectance Transformation Imaging (RTI) is a computational photography technique in which an object is acquired from a fixed point-of-view with different light directions. The aim is to estimate the light transport function at each point so that the object can be interactively relighted in a physically-accurate way, revealing its surface characteristics. In this paper, we propose a novel RTI approach describing surface reflectance as an implicit neural representation acting as a "relightable image" for a specific object. We propose to represent the light transport function with a Neural Reflectance Field (NRF) model, feeding it with pixel coordinates, light direction, and a latent vector encoding the per-pixel reflectance in a neighbourhood. These vectors, computed during training, allow a more accurate relighting than a pure implicit representation (i.e., relying only on positional encoding) enabling the NRF to handle complex surface shadings. Moreover, they can be efficiently stored with the learned NRF for compression and transmission. As an additional contribution, we propose a novel synthetic dataset containing objects of various shapes and materials created with a physically based rendering software. An extensive experimental section shows that the proposed NRF accurately models the light transport function for challenging datasets in synthetic and real-world scenarios.

CCS Concepts: • **Computing methodologies** → **Reflectance modeling**; **Computational photography**; *Appearance and texture representations*; Visual inspection; *Image-based rendering*.

Authors' Contact Information: Shambel Fente Mengistu, shambel.mengistu@unive.it; Filippo Bergamasco, filippo.bergamasco@unive.it; Mara Pistellato, mara.pistellato@unive.it, Dipartimento di Scienze Ambientali, Informatica e Statistica (DAIS), Università Ca'Foscari Venezia, 155, via Torino, Venice, Italy.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
Manuscript submitted to ACM

Additional Key Words and Phrases: reflectance transformation imaging (RTI), implicit neural representation, image-based relighting, interpolation, relighting network

ACM Reference Format:
Shambel Fente Mengistu, Filippo Bergamasco, and Mara Pistellato. 2024. A Neural Reflectance Field Model for Accurate Relighting in RTI Applications. *ACM Trans. Graph.* 0, 0, Article 0 (2024), 28 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Reflectance Transformation Imaging (RTI) [CHI 2023; Molly Hughes-Hallett and Messier 2021; Mytum and Peterson 2018] is a computational photography technique used to capture and store an object's surface characteristics. The produced data are employed in a variety of applications to provide tools for both visualization and analysis. Such applications range from quality inspection [Le Goïc et al. 2022; Pitard et al. 2017a; Zendagui et al. 2022], surface enhancement [Fattal et al. 2007], archaeology [Earl et al. 2010; Mytum and Peterson 2018] and cultural heritage [Mudge et al. 2006; Saha et al. 2022; Siatou et al. 2022]. The main goal of this technique is to allow non-contact and non-destructive visual inspection of the artifact by interactively relighting a picture of it observed from a fixed point of view. For this reason, RTI is commonly used to study objects that are mostly planar (on which full 3D scanning would not be beneficial) but exhibiting a surface rich in scratches, bumps, and regions with different shading properties. Classic examples include industrial surfaces with micro-scratches [Nurit et al. 2021], coins [Palma et al. 2014], and bas-reliefs [Barbosa et al. 2007].

RTI produces a "relightable image" of the object by deriving a model describing how the surface reflectance at each camera pixel is related to the incident illumination direction. Image-based relighting methods usually refer to this model as *light transport function* (or *reflectance field*) $T(\mathbf{x}, \omega)$, mapping incident light radiation from direction ω to the reflectance observed at pixel $\mathbf{x}$.

The RTI workflow consists of different stages. The first is known as *acquisition*, which involves imaging an object with multiple light directions from a fixed camera position. In this part, the light distribution is crucial to reveal details of surface geometry that would not appear in a single exposure. Different setups are used to acquire images, two examples are the single mobile light method, also known as Highlight-RTI [Giachetti et al. 2018] and Dome RTI [Corregidor et al. 2020; Earl et al. 2011; Pitard et al. 2017b; Saha et al. 2022]. The collected data is known as Multi-Light Image Collection (MLIC) or RTI source image set, representing essentially a sparse discrete sampling of the light transport function. The second step, which is at the core of RTI workflow, is referred to as *modeling* and involves the numerical definition of the light transport function as an interpolator of the MLIC. In this respect, RTI differs from other relighting techniques because it aims at (i) minimizing the bytes needed to store the function while (ii) maximizing the physical accuracy of images relit from light directions not observed during acquisition. The first principle is to allow post-examination in the absence of the physical object [CHI 2023; Palma et al. 2010], possibly by providing a lightweight web-based interface [Ponchio et al. 2019]. Physical accuracy is preferred against perceived visual quality because the aim is the visual inspection performed by a researcher interested in studying a particular artifact. Indeed, apart from relighting, it is common to perform digital enhancement of the object's color and texture to reveal surface information that is not apparent with visual empirical examination [Palma et al. 2010].

In this paper, we propose a novel approach to improve the modeling step of RTI. Driven by recent advances of coordinate-based neural representations for discrete low-dimensional signals, we model T as an Implicit Neural Representation (INR) with a Multi-Layer Perceptron (MLP) trained with the samples collected during the acquisition step. This idea was already proposed by Ren et al. [2015]. Still, it did not include any positional encoding of the input

which has been recently demonstrated to play a crucial role in capturing high frequencies [Tancik et al. 2020]. Unlike most INRs based only on input coordinates, we feed the model with an additional highly compressed per-pixel latent vector designed to capture local surface details while preserving the INR ability to model global illumination phenomena like self-shadows and inter-reflections accurately. Each latent vector is encoded by a CNN trained together with the INR, and stored with the network weights for subsequent relighting. Furthermore, we propose a novel synthetic dataset created using the physically based renderer Mitsuba 3 [Jakob et al. 2022]. Publicly available synthetic dataset benchmarks such as the one proposed by Dulecha et al. [2020] are limited in the size of MLICs, and miss objects with complex shadows and specularities. Besides, the light distribution does not typically span the full incident hemisphere as is in real-world Dome RTI, which limits its ability to generalize to new lighting directions. Our synthetic dataset avoids all these drawbacks by including large-size MLICs rendered in a dome-shaped virtual lighting configuration and incorporating objects of complex shapes. The dataset and the code used to generate it are publicly available[1]. An extensive experimental section highlights the proposed model relighting capabilities with a limited model size while preserving physical accuracy with novel lighting conditions.

2 RELATED WORK

Image relighting and view synthesis are popular topics in Computer Vision and Computer Graphics communities. Among them, most state-of-the-art methods aim to provide free relighting of a scene to obtain a realistic output. For example, some works focus on human portraits [Nishino and Nayar 2004; Pandey et al. 2021; Sun et al. 2019] and modern man-made objects [Loscos et al. 2000; Zhang et al. 2016]. RTI is focused on contactless visual inspection of surfaces acquired from a single viewpoint with varying light directions (the input is the MLIC), and the fidelity of the surface under study is of pivotal importance. In the following we review the relevant literature for the specific RTI application as well as general-purpose relighting approaches.

2.1 Classical RTI

Despite the ubiquitous usage of neural networks in modern computer vision and computer graphics applications, their application to RTI is not yet fully exploited. Almost all commercial RTI applications in cultural heritage, industry, medical imaging, and other fields are dominated by classical non-learning-based approaches [Pintus et al. 2019]. In cultural heritage which appears, by far, to be the application domain where RTI is more popular, Polynomial Texture Mapping (PTM) [Malzbender et al. 2001], Hemispherical Harmonics (HSH) [Gautron et al. 2004], and Discrete Modal Decomposition (DMD) [Pitard et al. 2017b] are well-established techniques designed to store a compact representation of the MLIC and interactively relight the images by interpolating a low-dimensional smooth function. Since seminal RTI applications based on PTM, several improvements were proposed with the goal of improving the interpolation functions [Drew et al. 2012; Toit 2008] or performing a robust regression of the sparse samples in the MLIC [Zhang and Drew 2014]. This line of research has led to sophisticated methods like the one presented by Ponchio et al. [2018], based on a joint interpolation-compression scheme combining a Principal Component Analysis (PCA) for data reduction with a Gaussian RBF to allow high fidelity of the relighted images with compact representation.

[1]Project repository: https://github.com/DAISCVprojects/NRF-RTI

2.2 Learning-based RTI

More recently, though limited in number, neural networks have been employed with great success in different RTI applications. One of the first attempts dates back to 2015 when Ren et al. [2015] used neural networks to model light transport as a non-linear function of light source position and pixel coordinates. This approach is remarkably close to modern coordinate-based neural representations but does not perform any positional encoding of the input. For this reason, the method requires scenes with hundreds of images to work accurately and fails to recover specularities and self-shadows.

A work by Xu et al. [2018] employed a convolution-deconvolution architecture to synthesize scene appearance under novel, distant illumination from the visible hemisphere but the amount of data to be stored (a 25-channel image) is considered too much for several RTI applications. Dulecha et al. [2020] proposed a feed forward Neural Network that exploits reflectance data in the MLIC to train a fully connected asymmetric autoencoder. The autoencoder compresses the original per-pixel reflectance data into a low-dimensional vector to be stored. Relighting is performed by the decoder, trained to reconstruct pixel values from the encoded vector and a user-specified light direction. The main limitation of their method is that it fails to model shadows and specular highlights accurately. Besides, the network size depends on the size of MLICs which limits the usage to large datasets acquired without a predefined light dome. Such method has been expanded by Righetto et al. [2024], where the authors optimize the model to enhance efficiency and embed it in an interactive web application. Recently, Pistellato and Bergamasco [2023] proposed an implicit neural representation to model the reflectance given a light direction and a compact vector describing the captured light distribution at each pixel. This is the first attempt at using an INR to map the light transport function, even if the pixel coordinate is not mapped directly by the model. The results are more accurate than encoder-decoder approaches but some image artifacts are still present, especially in shadowed areas.

2.3 Bidirectional Texture Function

A Bidirectional Texture Function (BTF) describes how a specific material appearance changes depending on the viewing direction, the 2D position on the surface and the illumination direction. Methods estimating the BTF are mainly designed to capture complex material properties and offer realistic rendering in graphics. While RTI can be interpreted as a fixed-view subset of BTF, its final goal differs from BTF since it aims at enhancing the surface relief and subtle details, allowing for virtual relighting and efficient compression of relightable scenes. BTF was first introduced by Dana et al. [1999], and few years later Filip and Haindl [2008] proposed a comprehensive survey on the topic. Recently, advancements in learning-based techniques led to the proposal of neural BTF. In particular, Rainer et al. [2019] proposed an asymmetric encoder-decoder architecture, where the encoder receives a per-pixel apparent BRDF (ABRDF) as input to produce a low-dimensional latent representation. The decoder takes this encoded latent representation, view direction, and camera direction to output a single RGB value. This method requires a viewing direction besides the light direction, which is not typically provided by the RTI acquisition process. In the work by Kuznetsov et al. [2021], the authors propose NeuMIP, a neural BTF for representing and rendering different material appearances at different scales, with the aim of integrating the method in a rendering engine. Despite their accuracy, such techniques have difficulties handling materials with significant shadows or detailed specular effects. The method proposed by Xue et al. [2024] aims at improving this aspect. In their work, Fan et al. [2023] proposed a lightweight architecture using biplane representation for BTFs where the decoder is trained once to represent a broad range of materials. Other recent methods proposing neural BTF representations are [Sztrajman et al. 2021; Zheng et al. 2021].

2.4 Photometric Stereo

Photometric stereo (PS) was originally introduced by Woodham [1989] and Silver [1980], and its main goal is to recover the surface shape from a combination of observed reflectance varying the lighting in multiple images. The original formulation was based on the simple Lambertian surface model, which was substituted with the more flexible bidirectional reflectance distribution function (BRDF), which parameters need to be carefully estimated [Goldman et al. 2005; Ikehata 2023; Li and Li 2022; Tiwari and Raman 2022]. A comprehensive survey on the topic is given by Ackermann et al. [2015]. During the years, several data-driven approaches have also been proposed for PS applications, for instance Santo et al. [2017] use the reflectance information under varying light directions to infer the per-pixel normals. A recent overview of data-driven models is given by Zheng et al. [2020].

Photometric stereo techniques share some commonalities with RTI: indeed, both methods follow similar acquisition setups to generate the MLICs, but PS is focused on retrieving the surface normals and follows different strategies to process the data. As a further step, results coming from PS (i.e. the BRDF parameters) can be exploited in an additional rendering layer to relight images from any target light direction. Tiwari and Raman [2022] propose to stack several hourglass neural networks that take two differently illuminated images and jointly retrieve surface normal, albedo, light estimation, and perform image relighting. Practically, this method is similar to single-image SVBRDF (Spatially Varying Bidirectional Reflectance Distribution Function) based relighting methods [Luo et al. 2024; Sang and Chandraker 2020; Tiwari et al. 2024; Yi et al. 2023a], except that it uses two differently illuminated input images under uncalibrated and self-supervised settings. Photometric Stereo is also used in applications such as quality control [Farooq et al. 2005; Smith et al. 1999], surface enhancement [Malzbender et al. 2006], industrial inspection [Ren et al. 2019], medical imaging [Parot et al. 2013], and cultural heritage preservation [Dessì et al. 2015; Yeh et al. 2016], to name a few.

2.5 Single Image Relighting

In recent years, the relighting problem has been also approached through methods taking as input a single image of the surface to be relighted. Such approaches leverage deep learning models to solve the ill-posed problem of relighting from a single image by directly computing the output image [Bieron et al. 2023] or passing through the surface SVBRDF estimation [Deschaintre et al. 2018; Luo et al. 2024; Zhou and Kalantari 2021]. Such methods simulate the rendering process and perform re-rendering using the estimated parameters to recover the relighted image. The method proposed by Luo et al. [2024] recovers the material's SVBRDF parameters through a learned gradient descent [Andrychowicz et al. 2016]. The network starts with an initial SVBRDF estimation and then applies an iterative update rule by minimizing the difference between the rendering of the final prediction SVBRDF and the rendering of the ground truth material maps under various view and light conditions. Then, the final estimated SVBRDF can be used to relight an image. Despite producing good relit images of planar surfaces, this method has several limitations. The first drawback is its poor generalization in situations where the input images are captured under conditions deviating from the training images. Secondly, the method handles only attached shadows, and images with large highlights and cast shadows severely hinder its efficacy of producing accurate relit images. The method proposed by Bieron et al. [2023] approached single image relighting differently, by directly generating the target visual appearance without using an intermediate SVBRDF representation. Similar to Luo et al. [2024], this method is also dependent on the view/light combinations seen during training, and it can not handle curved surfaces with cast shadows. The work proposed by Yi et al. [2023a] is designed for relighting in augmented reality applications and involves a weakly-supervised inverse rendering pipeline with different branches to factorise specular and diffuse components and estimate the surface normals. Other methods

propose to jointly estimate the SVBRDF parameters and relighting through a single image [Sang and Chandraker 2020; Tiwari et al. 2024].

Despite promising, single-image approaches are designed to be applied in specific applications and to estimate a limited number of materials from the training datasets. As discussed in the experimental section (see §5.4), they can not ensure the accuracy in real-world scenes acquired with standard RTI setups where complex surfaces with cast shadows are common and acquired with different illuminations. The described methods often apply both inverse rendering and re-rendering for reconstructing the image under novel lighting conditions, and sometimes additional information is required during the training process (the surface normals for example), while RTI applications require only the lights positions together with the MLIC (often this is given by the light dome by construction).

2.6 NeRF-based Relighting

Since its introduction in 2020, Neural Radiance Field (NeRF) [Mildenhall et al. 2020] revolutionized view synthesis techniques by explicitly mapping the radiance in a volume using a neural representation. In subsequent years, several works have been proposed to improve the reconstruction, handle complex reflections [Ge et al. 2023; Guo et al. 2021; Ma et al. 2023; Verbin et al. 2021] and synthesize faithful novel views. Such ideas have been extended to enable scene rendering from novel viewpoints under arbitrary lighting. The methods presented by Lyu et al. [2022] and Philip et al. [2021] edit the scene light starting from a multi-view capture with a single light condition, while Rudnev et al. [2022] propose to learn a neural relightable representation of outdoor scenes from a set of images capturing the same place from different viewpoints and at different times. Other works propose to have multi-light, multi-view acquisition as known inputs for the model to perform rendering under unseen lights [Toschi et al. 2023; Xu et al. 2023]. Srinivasan et al. [2021] use an MLP whose inputs are the 3D location and outputs are the volume density, surface normal, material parameters and visibility. This allows the model to render unseen views with arbitrary light conditions. The approach proposed by Yang et al. [2022] performs 3D neural reflectance field optimization from single-view images captured under different lights. The method recovers the geometry and BRDF of a scene for novel-view synthesis and relighting. Zeng et al. [2023] propose to use two MLPs: one to represent the acquired shape, and one to directly model local and global light at each point, without disentangling the light transport components. The method is trained on unstructured photographs of the scene captured from different viewpoints and different light positions.

Albeit interesting, radiance fields are designed to describe density and light radiation in a volume, that is a different goal with respect to RTI applications we target in this paper. Indeed, NeRF-based approaches require multiple views of the same object, which is not applicable for existing RTI data or for new data acquired with standard RTI setups. Also, most NeRF-based relighting methods are general purpose relighting used in real-world conditions or with different input data. Our proposed method is able to process old datasets acquired with light domes or mobile devices, while other general-purpose or multiview techniques could find difficulties with such limited data. Moreover, RTI applications are intrinsically characterized by efficient and lightweight implementations, since they are also used for compression and transmission purposes. For this reason RTI methods aim at using fewer parameters, while NeRF-based methods are computationally expensive in terms of memory usage and training time, as they need to optimize a higher number of parameters and require a higher number of samples.

3 NEURAL REFLECTANCE FIELD FOR RTI

We assume to have a sequence of N images $I_1 \dots I_N$, with size $w \times h$ pixels, each one captured with the same camera but with different light directions $\omega_1 \dots \omega_N \in \mathcal{S}^2$. Light is assumed to be infinitely far away so that a direction ω is a

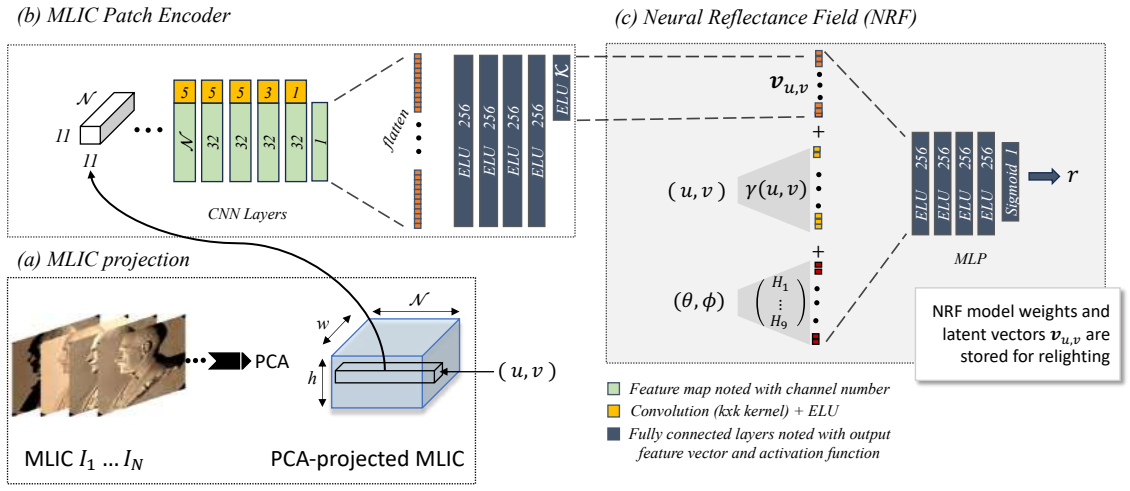

Fig. 2. Architecture of our proposed Reflectance Field model for RTI with its main components: (a) PCA-based projection, (b) MLIC patch encoder, and (c) Neural Reflectance Field with ELU (Exponential Linear Unit) activation function. See text for details.

unitary norm vector on the upper hemisphere. As introduced before, this data is known as MLIC and can be produced in different ways, including light domes or moving light sources paired with calibration targets.

The goal of RTI is to estimate the reflectance field $T(\mathbf{x}, \omega) : (\mathbb{R}^2, \mathcal{S}^2) \rightarrow \mathbb{R}^C$ describing the reflectance (expressed as a vector of C spectral bands) observed at each point $\mathbf{x}$ when the object is illuminated by light radiation propagating from direction ω. Once the reflectance field is known, the object can be relighted by evaluating T at a discrete set of pixels for any arbitrary light direction. Even if the MLIC is discrete, considering the reflectance field as a continuous function is useful because it allows synthesizing new images not only with arbitrary light directions but also on a different spatial domain (i.e. one could zoom in by evaluating T on a specific region of interest or upscale the output image by sampling at sub-pixel coordinates).

One of the main requirements of RTI is to construct a compact relightable representation of the object so it can be stored and/or transmitted efficiently. To this end, many methods (including learning-based ones) work by compressing the MLIC and then interpolating the discrete samples to create new images.

In this work, we build an estimation of T directly with a Multi-Layer Perceptron (MLP) $\Phi(\mathbf{x}, \omega, \mathbf{v_x}) : (\mathbb{R}^2, \mathcal{S}^2, \mathbb{R}^{\mathcal{K}}) \rightarrow \mathbb{R}^C$, where $\mathbf{x} = (u, v)$ denotes the pixel coordinates and ω is the light direction expressed in spherical coordinates with azimuth ϕ and elevation θ. The additional input $\mathbf{v_x}$ represents a latent vector encoding the reflectance in a local neighbourhood of $\mathbf{x}$ and will be further discussed in § 3.1. The size of the output $C \in \{1, 3\}$ denotes that the model can be configured to predict only the pixel luminance (so we have $C = 1$) or the full RGB triplet (that is $C = 3$). Since the output size only affects the last MLP layer, we decided to keep such flexibility and propose the two variants for our approach. In § 3.4 we discuss the two different output representations, that will be also analysed in the experimental section.

The complete architecture of our model is sketched in Fig. 2. The main component is the Neural Reflectance Field Φ (c) implemented as a simple MLP fed with the pixel position $\mathbf{x} = (u, v)$, light direction (θ, ϕ), and the latent vector $\mathbf{v}_{u,v}$ produced during the training phase by the MLIC Patch Encoder (b). In the first layer pixel position and light direction are projected on a higher dimensional space using periodic Fourier-based functions and Hemispherical Harmonics

(HSH) respectively, as discussed in § 3.3. The input MLIC has an initial shape of $w \times h \times N$, and before training it is compressed with PCA (Fig. 2,a) so that the PCA-projected MLIC has shape $w \times h \times \mathcal{N}$, with $N \geq \mathcal{N}$. Further details about the projection are described in § 3.2.

The MLIC Patch Encoder (Fig. 2,b) acts as an encoder taking in input an $s \times s \times \mathcal{N}$ patch[2] of the PCA-projected MLIC centered at pixel (u, v). This component is a fairly standard CNN followed by 4 fully connected layers: the convolutional layers allow the network to effectively exploit local coherence in the reflectance data and learn local surface details in a compact representation. The output is the latent vector $\mathbf{v}_{u,v}$ used to convey local reflectance information to the INR model.

The per-pixel latent vectors $\mathbf{v}_{u,v}$ and the model parameters are jointly estimated during the *training phase*, taking as input only the MLIC and the associated light directions. For each training iteration i the following steps are performed until convergence:

(1) The input MLIC is sampled repeatedly at a uniformly distributed random pixel location $\mathbf{p}^{(i)} = (u^{(i)}, v^{(i)})$, $u^{(i)} \sim \mathcal{U}(0, w)$, $v^{(i)} \sim \mathcal{U}(0, h)$ and light direction $\omega_j^{(i)}$, $j \sim \mathcal{U}(0, N)$.
(2) The $s \times s$ patch around $\mathbf{p}^{(i)}$ is extracted from the PCA-projected MLIC and fed into the MLIC Patch Encoder to produce the latent vector $\mathbf{v}_{u^{(i)},v^{(i)}}$.
(3) The values $u^{(i)}, v^{(i)}$ and $\omega_j^{(i)}$ are fed into the Neural Reflectance Field together with the computed latent vector $\mathbf{v}_{u^{(i)},v^{(i)}}$ to produce the output reflectance value $r^{(i)}$.
(4) A standard $L1$ loss is used to evaluate $r^{(i)}$ against $I_j(u^{(i)}, v^{(i)})$ and optimize the weights for both the Patch Encoder and NRF with backpropagation.

Once the model is fully trained, we can obtain the the per-pixel latent vectors simply evaluating $\mathbf{v}_{u,v} \forall u = 0 \dots w, \forall v = 0 \dots h$ and store the resulting vectors. In this way we obtain a $\mathcal{K}$-dimensional vector for each pixel and we can completely discard both the original MLIC and the Patch Encoder component, since the reflectance information is encoded in the latent vectors $\mathbf{v}_{u,v}$. Then, to perform relighting we need the following:

- All the computed latent vectors, that is a total of $w \times h \times \mathcal{K}$ values.
- The optimized weights of the NRF model (Figure 2, c).
- In the case the NRF outputs only the luminance, we also need the per-pixel chrominance values (or nothing, if the acquisition is grayscale).

To create a relighted image of the object with an arbitrary light direction $(\bar{\theta}, \bar{\phi})$ provided by the user, Φ is evaluated for all the pixels locations (u, v) with the associated $\mathbf{v}_{u,v}$ and the given light direction. The operation is fast as it implies only the forward pass of the NRF so it is suitable for real-time visualization. In the following we discuss additional details about specific sub-components of our proposed model.

3.1 The Importance of Having an Additional Input Latent Vector

We started by creating a "pure" INR model $\Phi'(u, v, \phi, \theta) \rightarrow \mathbb{R}^C$ taking as input a pixel coordinate $\mathbf{x} = (u, v)$ and light direction ω expressed in spherical coordinates with azimuth ϕ and elevation θ. We observed that regardless of the way we perform the positional encoding of the input and the number of weights, Φ tends to do an excellent job in reconstructing self-shadows and specularities while missing high-frequency details of the surface. On the other hand, INRs (as the one proposed by Pistellato and Bergamasco [2023]) that completely discard position information but take into account a per-pixel reflectance distribution vector exhibit great quality in surface details but poor performance

[2]We empirically fixed $s = 11$ as the patch size in all our experiments.

in modeling the shadows. We guess two possible reasons for that. First, light samples are a lot sparser than pixel samples, especially if the MLIC is captured with a light dome, and the distribution of training samples is important when training an INR. Second, albeit sparse, light direction may cause abrupt discontinuities in the produced image in the presence of self-shadows and specularities. Indeed, small variations in ω can cause a surface region to quickly saturate to full reflectance (specular highlight) or very low intensity (shadow). To overcome these limitations, we augmented the aforementioned approach proposing our reflectance field model Φ with an additional (trainable) $\mathcal{K}$-dimensional latent vector $\mathbf{v}_{u,v}$ encoding the per-pixel reflectance information in a local neighbourhood of (u, v). The parameter $\mathcal{K}$ is application-dependent and can be used to trade off reconstruction accuracy with model size, as shown in the experimental section. An extensive ablation study (§ 5.1) shows the advantages in terms of relighting accuracy when including such additional vector instead of applying the classic INR approach.

3.2 PCA-based MLIC Projection and Patch Encoding

The Patch Encoder component (Fig.2, b) may potentially process a patch extracted directly from the MLIC. Models like NeuralRTI [Dulecha et al. 2020] adopt this approach, but it carries some drawbacks because in this way the model architecture becomes highly dependent on the number of images N. This implies that the Patch Encoder must be created and trained from scratch for the specific RTI setup. Even if we accept that N is fixed (64 light sources are almost a standard value for light domes), the model still will be dependent on the ordering on which images are stacked (channel-wise) into the MLIC. To overcome this issue, we propose to project the pixel samples of the MLIC into the first $\mathcal{N}$ PCA bases before using it to encode the latent vectors. Specifically, the input data can be considered as a set of $w \times h$ N-dimensional vectors that can be linearly projected onto a set of $w \times h$ $\mathcal{N}$-dimensional vectors to be used as input of the Patch Encoder. If $\mathcal{N} = N$ there is no information loss in the process, but we are now completely independent of the light ordering. Indeed, any channel-wise shuffling of the original input MLIC would produce the same PCA-projected MLIC. If $\mathcal{N} < N$ we lose some (redundant) information but, in any case, we gain the possibility of using a pre-trained Patch Encoder regardless of the setup used to capture the MLIC. This will reduce the time and the computational burden when training the NRF, which must be "overfitted" to every object instance. Besides feeding the patch encoder a PCA projected MLICs instead of the raw MLICs has gained a slightly improved results as we will see in section 5.1.

3.3 Positional encoding of the input

Despite the fact that neural networks are universal function approximators, it is well-known that a positional encoding of the input to a higher-dimensional space allows the INR to train faster and better reproduce the high-frequency components of the signal [Tancik et al. 2020]. In our model, we encode both the pixel position and light direction onto orthonormal Fourier-like periodic bases as we now briefly discuss.

3.3.1 Mapping pixel coordinates. We follow Tancick et al. [2020] and map each pixel coordinate to a Fourier space of random frequencies. Let B be a Mx2 matrix where each value is sampled from a zero-mean Gaussian distribution with variance σ^2.

Normalized pixel coordinates $\bar{p} = \begin{pmatrix} \frac{u}{w} & \frac{v}{h} \end{pmatrix}^T$ are projected as follows:

$$\gamma(\bar{p}) = [\cos(2\pi B\bar{p}), \sin(2\pi B\bar{p})]^T. \tag{1}$$

In our experiments, we have chosen a sigma value of 0.3 for the multivariate isotropic Gaussian distribution and set the length of M to 10. This results in each pixel coordinate being mapped to a 20-dimensional space.

3.3.2 Mapping light direction vector. As done for pixel coordinates, we also project the light direction vectors into higher dimensional space, but in this case, we encode the light direction using the hemispherical harmonics functions. This is analogous to the Fourier space projection but on the hemisphere's surface. We prefer HSH over spherical harmonics (SH) because the incident and reflected lights are all distributed on an upper hemisphere, and full spherical information would therefore result in an over-parameterization of the input space.

We apply the HSH functions [Gautron et al. 2004] defined as follows:

$$H_l^m = \begin{cases} \sqrt{2}\tilde{K}_l^m \cos(m\phi)\tilde{P}_l^m(\cos\theta) & m > 0 \\ \sqrt{2}\tilde{K}_l^m \sin(-m\phi)\tilde{P}_l^{-m}(\cos\theta) & m < 0 \\ \tilde{K}_l^0 \tilde{P}_l^0(\cos\theta) & m = 0 \end{cases} \tag{2}$$

where $\tilde{P}_l^m$ and $\tilde{K}_l^m$ are the "shifted" associated Legendre Polynomials and the hemispherical normalization factors defined as

$$\begin{aligned} \tilde{P}_l^m(x) &= P_l^m(2x-1) \\ K_l^m &= \sqrt{\frac{(2l+1)(l-|m|)!}{2\pi(l+|m|)!}}. \end{aligned} \tag{3}$$

Then, we derive a set of basis functions as described in the following equations:

$$\begin{aligned} &H_i = H_l^m; i = ((l+1)l - m) + 1; \text{Order} = (l+1) : \\ &-Order1: \\ &H_1(\theta,\phi) = 1/\sqrt{(2\pi)} \\ &-Order2: \\ &H_2(\theta,\phi) = \sqrt{(6/\pi)}(\cos(\phi)\sqrt{(\cos(\theta) - \cos(\theta)^2)}) \\ &H_3(\theta,\phi) = \sqrt{(3/(2\pi))}(-1 + 2\cos(\theta)) \\ &H_4(\theta,\phi) = \sqrt{(6/\pi)}(\sin(\phi)\sqrt{(\cos(\theta) - \cos(\theta)^2)}) \\ &-Order3: \\ &H_5(\theta,\phi) = \sqrt{(30/\pi)}(\cos(2\phi)(-\cos(\theta) + \cos(\theta)^2)) \\ &H_6(\theta,\phi) = \sqrt{(30/\pi)}(\cos(\phi)(-1 + 2\cos(\theta))\sqrt{(\cos(\theta) - \cos(\theta)^2)}) \\ &H_7(\theta,\phi) = \sqrt{(5/(2\pi))}(1 - 6\cos(\theta) + 6\cos(\theta)^2) \\ &H_8(\theta,\phi) = \sqrt{(30/\pi)}(\sin(\phi)(-1 + 2\cos(\theta))\sqrt{(\cos(\theta) - \cos(\theta)^2)}) \\ &H_9(\theta,\phi) = \sqrt{(30/\pi)}((-\cos(\theta) + \cos(\theta)^2)\sin(2\phi)). \end{aligned} \tag{4}$$

Since hemispherical functions are limited to the upper hemisphere, Equation 2 is valid for $\theta \in [0, \pi/2], \phi \in [0, 2\pi]$. In our model, we use the first 9 basis functions $H_1 \ldots H_9$ defined on the first 3 bands of the HSH function in Equation 2. These 9 basis functions result in a 9-dimensional expansion of the 2-dimensional (θ, ϕ) light vector.

3.4 Luminance-chrominance Output Representation

A common design choice of most RTI methods involves the reflectance being modeled for each color channel separately, which incurs increased memory and computation costs. One source of redundancy in the MLIC is that the chromaticity of a particular pixel is fairly constant under varying light source directions; it is largely the luminance that varies. As a matter of fact, iridescence is a phenomenon pretty rare in common materials and mostly limited to ancient glass-based archeological artefacts [Emami et al. 2016]. For this reason, as previously discussed, in this paper we propose and analyze two alternative output representations, namely the full RGB triplet (3 values) or the simple luminance channel. For the latter, we take advantage of the color redundancy by converting the RGB representation to HLS color space and use only the luminance component (i.e. the L channel) to model the reflectance value of a pixel. To deal with colors, we store the pixel-wise average $\overline{H} = \frac{1}{N}\sum_{i=1}^{N} H_i$ and $\overline{S} = \frac{1}{N}\sum_{i=1}^{N} S_i$ for further processing. We observed that it is important to compute the averages excluding pixels with dark values and low saturation, as these values are irrelevant for restoring the color. One simple method is to threshold the L and S channels of the original images and then apply the mask on H and S values to compute the averages. This can be performed by tuning the threshold values on each MLIC, but in our experiments the same threshold value worked well for all data to restore the color. The HLS representation has some advantages over RGB representation, such as reduced training time and model size, thus it is suitable for simple materials or in situations where a higher compression rate is required. However, since it is not always possible for all materials to perfectly reconstruct the chrominance component from the averaged H and S channels, we also modeled the RGB representation to offer a better result by perfectly restoring the color components but slightly reducing the MLIC compression. In the experimental section we explore both alternatives: in general, the user can choose one model or the other, depending on the application specific requirements.

4 SYNTHETIC DATASET GENERATION

In this work, we provide a novel synthetic dataset to evaluate classic and learning-based RTI methods. Due to the limited number of publicly available benchmarks, our dataset can be considered a useful addition, especially for modeling real-world challenging phenomena such as shadows and specularity.

Our dataset consists of 9 models ranging from a simple planar surface to complex geometric shapes collected from different publicly available 3D model repositories [Laric 2023; Maggiordomo et al. 2020]. When necessary, each model was manually edited, aligned, and scaled to a consistent reference system. Together with the proposed model, we provide the pre-generated dataset of MLICs and the code to generate MLICs by rendering each selected object using the state-of-the-art renderer Mitsuba3 [Jakob et al. 2022]. We applied a database of real-world isotropic and anisotropic measured materials [Dupuy and Jakob 2018], exhibiting different surface scattering phenomena, such as smooth diffuse material, rough conductor material, and smooth plastic material. The measured materials collected from a database by Dupuy and Jakob [2018] are metallic paint, brown corrugated cardboard (smooth surface), matte painting, TeckWarp vinyl wrapping film, and white A4 paper. Sample images of our synthetic dataset for each geometric model applied with different materials are shown in Figure 3 (left). Images in MLICs are rendered by uniformly sampling points on a unit disc and then projecting the sampling points up to the hemisphere. This produces input directions with a

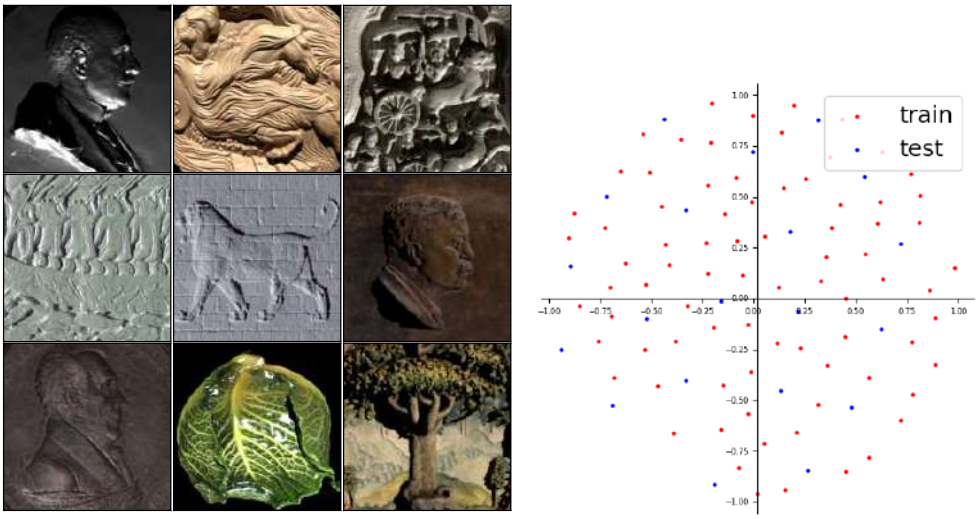

Fig. 3. Left: Sample images of our synthetic dataset with different materials. From left to right and top to bottom: Tondo, Horses, Ta Prohm, Bas-relief Bayon, Lion, Theodore, Andrew, Cabbage, Panel. Right: Distribution of generated light directions divided into train (red dots) and test set (blue).

cosine-weighted distribution along the normal, with light directions that are more frequent at around 45°, as commonly happens when we observe real objects. Figure 3 (right) shows the distribution of light directions with the train-test split.

5 EXPERIMENTAL EVALUATION

Our evaluation is based on two datasets: our synthetic dataset and the publicly available dataset presented in Dulecha et al. [2020]. Our dataset consists of nine challenging geometric models, each assigned with a different material, producing nine MLICs. Each MLIC contains 100 images of spatial resolution of 256×256. We kept 80 light directions (i.e. 80 images) for training and the remaining 20 for testing. We will refer to our dataset as *synthetic*. The dataset presented in Dulecha et al. [2020] consists of a synthetic and real dataset of different objects. In our experiments we considered the real-world dataset addressed as *RealRTI* and the Synthetic spatially varying multi-material dataset referred to *Synth-Multi*. RealRTI dataset involves 12 real-world scenes with different materials and geometry, acquired with RTI techniques (light domes or handheld RTI protocols). Since data comes from different sources (see the original paper for details), each object is captured with a different number of lights (from 47 to 72). We selected for each object 10 testing lights and used the remaining for training. SynthMulti dataset presents 27 captures of the same three surfaces rendered with nine material combinations and 16 tints (for further details see the original paper). We used 49 lights for training, that correspond to the set of lights called *Dome* by the authors, with lights organised in concentric rings at 5 different elevations. The remaining 20 lights (located in 4 intermediate elevations) have been used for testing, as proposed in the original paper. The described train and test division was fixed for all presented experiments.

We focused on these publicly available datasets because RealRTI includes typical real-world acquisitions, and SynthMulti allows extensive analysis of complex materials since the surfaces exhibit different properties in different regions. Indeed, such configuration is particularly challenging since, as highlighted by Dulecha et al. [2020], the variety of colours and materials may create problems in algorithms using global optimization of the reflectance.

Table 1. Quantitative results on our synthetic dataset for different input configurations. Note that the highlighted line denotes our full proposed method with RGB modeling.

Model Name	Light Mapping	Pixel Coord.	Encoded Feature Vector	PSNR	SSIM	RMSE	LPIPS
Proposed(RGB)	HSH	Used	MLIC Patch Enc.	31.92±5.93	0.912±0.028	0.049±0.039	0.054±0.017
Proposed(Lum)	HSH	Used	MLIC Patch Enc.	29.07±6.12	0.895±0.034	0.054±0.041	0.076±0.024
Fourier	Fourier	Used	MLIC Patch Enc.	27.24±5.08	0.826±0.033	0.078±0.057	0.114±0.025
No Pixel	HSH	Not Used	MLIC Patch Enc.	28.52±6.23	0.881±0.038	0.057±0.043	0.082±0.025
PCA Feature	HSH	Used	PCA	28.67±6.17	0.884±0.034	0.056±0.042	0.078±0.025
No Feature	HSH	Used	Not Used	23.85±4.20	0.677±0.110	0.082±0.062	0.306±0.087

For all the datasets, we trained our model splitting the MLIC pixel data of training lights into 90% training and 10% validation, uniformly sampled across pixel locations. Our model was implemented in PyTorch using the following hyper-parameters: Adam optimizer [Kingma and Ba 2014] with a batch size of 4096 samples, an initial learning rate of 0.001, decayed to 0.0001 after 25 epochs, for a total of 35 epochs. We trained our model on an NVIDIA GeForce RTX 4080, and the training takes approximately 2 hours for a single scene. We adopted four metrics to quantitatively evaluate the results, namely Root Mean Squared Error (RMSE), Peak Signal-to-Noise Ratio (PSNR), structure similarity index (SSIM) [Wang et al. 2004], and Learned Perceptual Image Patch Similarity (LPIPS) [Zhang et al. 2018].

5.1 Ablation Study

We first analyzed our method by carefully designing an ablation study to highlight the influence of each individual component in the proposed architecture. Table 1 shows the selected configurations together with the results obtained on the synthetic dataset. Our full model (denoted as Proposed) is shown in two different versions: the model predicting the RGB triplet and the one (below) predicting only the luminance value. We analyzed variations such as different encoding functions for light direction (HSH or Fourier, 1st column), the presence of pixel coordinates (2nd column), and the encoded feature vector (MLIC Patch Encoder, standard PCA projection or nothing, 3rd column) in input. The resulting setups will be identified with the names in the first column, where the first one (highlighted in the table) displays our complete RGB model.

As shown by the table, our RGB version offers slightly better performances when compared with the luminance-only approach, since the RGB model has in general a good luminance representation. Quantitative results are comparable with the luminance model, which still is better than all the other alternatives, offering a good solution if the acquisitions are grayscale or oriented towards compression. Adopting HSH function to project light direction significantly improves the quality of the relighted images with respect to the classical Fourier feature space: the average PSNR value is improved by 6% simply applying HSH. The *no pixel* model (3rd row in Table 1) shows the effects of removing the pixel coordinates from the NRF input: in this case, results are quite comparable to the full model but still worse. This also happens by substituting our trained MLIC patch encoder with a simple PCA vector, representing the first 8 principal components of the input MLIC (see 4th row of Table 1): results are similar but still, the full model allows for a better shadow representation. This is due to the fact that our MLIC patch encoding embeds useful information about the neighbourhood behaviour of each image location, allowing for a representation that is also spatially-aware to improve the local reflectance prediction.

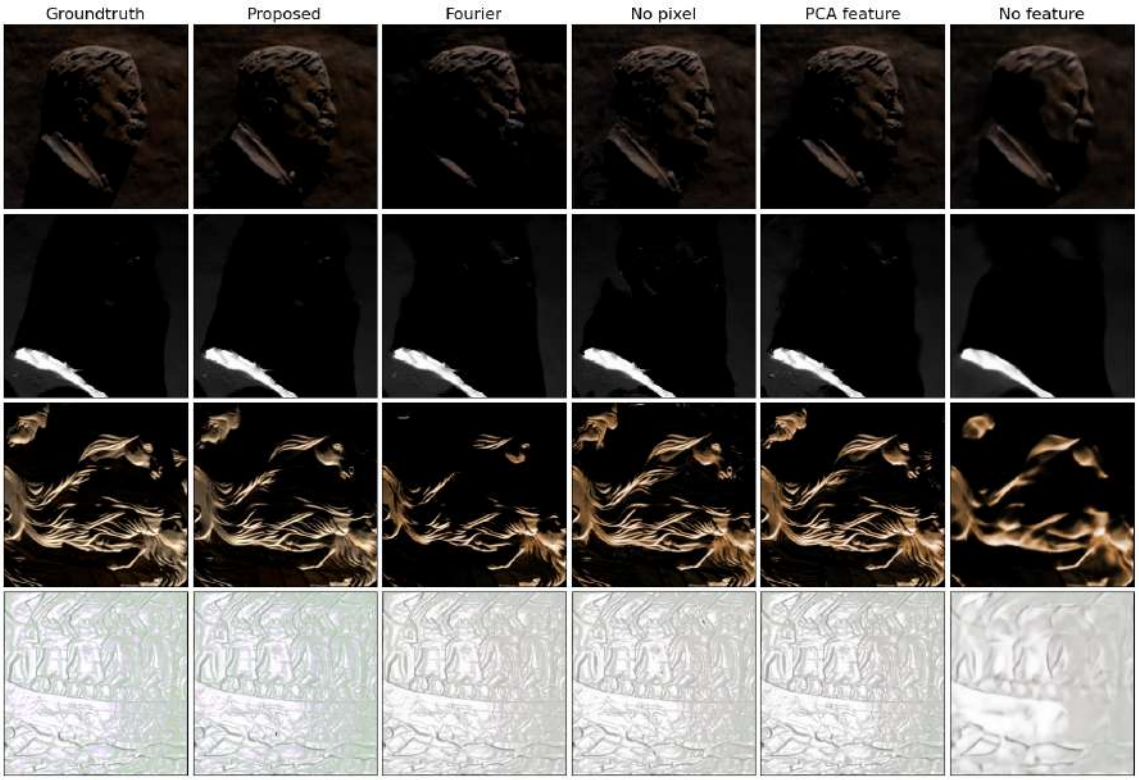

Fig. 4. Visual Results for the ablation study on images of the synthetic dataset. The first column shows the ground truth, while the other columns show the relighting results of different architecture combinations. Details of each configuration are reported in Table 1.

Finally, the last row in Table 1 shows the importance of including an encoded feature vector as input compared to a "pure" INR implementation with a comparable number of weights. In this case, the quality of relighted scenes drastically drops for all the metrics: PSNR is around 23 dB, while for all other cases, it is always greater than 27 dB. This result validates the fundamental role of the feature vector as input in our proposed NRF model. Note that the *no feature* model (last row) has a number of weights comparable with the other alternatives: indeed, a non-negligible aspect of RTI applications is data compression and efficiency.

To reinforce the quantitative results presented in Table 1, we also show some visual results in Figure 4. The complete *proposed* method (second column) produced images with detailed surfaces, smooth shadows, and specularities with respect to other alternatives. In particular, the *Fourier* configuration is not able to fully recover complex shadows, and the *No feature* version (last column) completely misses surface details.

5.2 Input and Model Size Analysis

In this section, we start analysing the model performances varying the number of input samples (i.e. the number of lights N). Figure 5 shows the resulting average PSNR and SSIM of our RGB model when increasing the training sample size from our synthetic dataset. To perform this experiment we exploited our synthetic setup to generate an increasing number of uniformly distributed training lights, and kept the same test lights as in the original dataset used in the

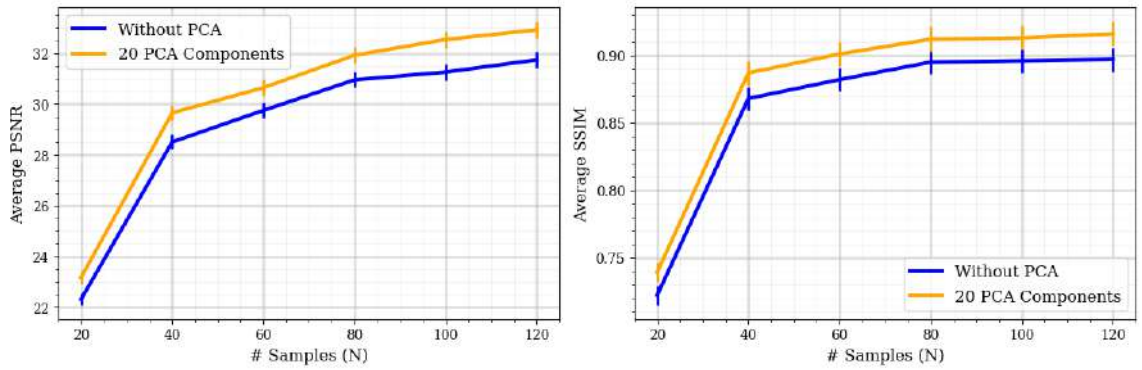

Fig. 5. Average PSNR and SSIM values with increasing input size on our synthetic dataset, when PCA module is not used and PCA module is applied with principal components set to 20.

other experiments to have comparable results. Additionally, to show the importance of applying PCA projection to the initial MLIC, we give as input the same training lights with and without performing PCA projection. Indeed, the orange curve represents the model in which we projected the MLIC into a fixed number of principal components $\mathcal{N} = 20$, while the blue curve represents models in which we directly input all the MLIC channels into the Patch Encoder. As we can observe, with 40 input lights we obtain an average PSNR greater than 28, that is still a reasonable result, while increasing the training images to 80 the model exhibits good relighting result (PSNR equal to 31), that is enough for most purposes. Another interesting result from this experiment is that using PCA projection gives quick convergence and better results since the principal components are able to compress all the relevant information. Projecting the MLIC input in the first $\mathcal{N}$ principal components allows for order invariance, and significantly reduces the Patch Encoder complexity: note that if we keep all the lights, the convolutional layer deals with up to 120-channels input. Potentially, if we had thousands of input lights, the input size would make the learning process simply impractical. Finally, we can observe that our method has a lower bound for $N = 20$, where the performance drops. This can be a limitation if we do not have more than 20 lights in our dataset, but in general RTI methods are not designed to minimize the number of lights, so the MLIC size is not a concern. Indeed, such methods are designed to provide the best tradeoff between model size, computational resources, and quality, so typical light domes are composed by far more then 20 lights. Note that in the case $N = \mathcal{N} = 20$ we still have a slight improvement for the PCA-projected data: this may be due to the fact that the PCA projection exhibits a better data representation, leading to a better model convergence.

The next experiment investigates the impact of $\mathcal{N}$, that is the number of principal components used as input. In Figure 6 we show the performance of our method increasing the principal components on RealRTI (right) and synthetic (left) datasets. We can observe that in both real and synthetic plots, the improvements between principal components 10 and 20 are very minimal, and increasing $\mathcal{N}$ to more than 30 in general tends to reduce performances. For this reason, in the complete proposed model we fixed $\mathcal{N} = 20$.

Finally, since a compact representation of the MLIC is vital for efficient real-time applications, we also analyzed the model performance and compression capability with respect to the size of the latent vector $\mathcal{K}$. Figure 7 shows the behaviour of our method for different latent vector sizes. The plots show average PSNR and SSIM for synthetic and real datasets varying the size of $\mathbf{v}_{u,v}$. For both datasets, we observe good quality with just 4 coefficients and stable

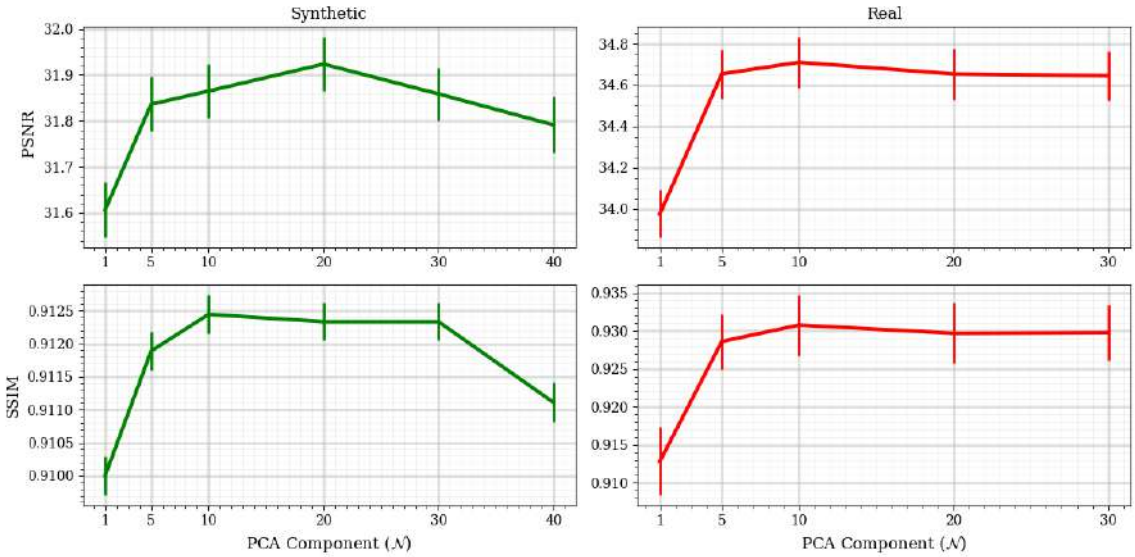

Fig. 6. Average PSNR and SSIM for our NRF model on our synthetic and RealRTI [Dulecha et al. 2020] datasets increasing the number of PCA components for input projection.

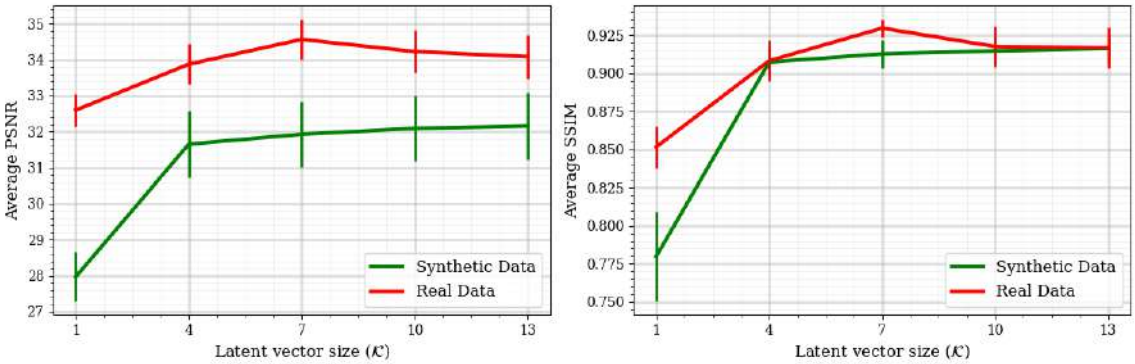

Fig. 7. Average PSNR and SSIM for our NRF model when increasing the latent vector size $\mathcal{K}$ (on x-axis). The two curves represent our synthetic and RealRTI [Dulecha et al. 2020] datasets respectively.

performance with sizes greater than 7. In all our experiments we set the size of $\mathbf{v}_{u,v}$ to 7, since this value is fair enough to obtain a good result.

5.3 Comparisons Against RTI Methods

We compared our proposed method against 5 different approaches that specifically target RTI applications. Among them, PTM (Polynomial Texture Mapping) [Malzbender et al. 2001] and RBF (Radial Basis Function) interpolation [Toit 2008] are two classic approaches; while the others are state-of-the-art learning-based methods, namely RelightNet [Xu et al. 2018], Neural reflectance transformation imaging (NeuralRTI) [Dulecha et al. 2020], and On-the-go reflectance

Table 2. Average value of different metrics for the test set computed on RealRTI, Synthetic and SynthMulti datasets when applying different RTI methods. Our proposed method is shown for two output variants, namely luminance only and full RGB triplet.

Dataset		PTM	RBF	RelightNet	NeuralRTI	onTheGoRTI	Proposed Lum	Proposed RGB
RealRTI	PSNR	27.69	28.75	24.22	30.21	28.78	33.26	34.65
	SSIM	0.796	0.771	0.780	0.860	0.857	0.917	0.929
	LPIPS	0.153	0.171	0.431	0.310	0.127	0.111	0.083
	RMSE	0.064	0.061	0.087	0.047	0.060	0.036	0.033
Synthetic	PSNR	20.49	27.08	21.25	25.13	25.73	29.07	31.92
	SSIM	0.706	0.809	0.694	0.750	0.795	0.895	0.912
	LPIPS	0.188	0.118	0.312	0.147	0.121	0.076	0.054
	RMSE	0.122	0.061	0.112	0.089	0.117	0.055	0.049
SynthMulti	PSNR	22.88	27.01	15.05	26.65	26.47	27.85	30.24
	SSIM	0.783	0.835	0.524	0.827	0.860	0.874	0.908
	LPIPS	0.145	0.094	0.490	0.093	0.077	0.073	0.071
	RMSE	0.089	0.055	0.186	0.061	0.062	0.055	0.047

transformation imaging (onTheGoRTI) [Pistellato and Bergamasco 2023]. In the following we summarise the features of the compared methods and how we trained each of them:

- **PTM** takes as input a set of single-view images with varying illumination and models the surface colour variation for each pixel independently with a biquadratic polynomial. The method is parametrised in N (number of input lights), that in all our experiments corresponds to the training set size of each scene.
- **RBF** is an interpolation method: we interpolated each pixel independently using the N input lights. For each scene we used the training set as given points and interpolated in the test light directions.
- **RelightNet** proposes to synthesize scene appearance under novel illumination from 5 selected images captured under pre-defined lights. The method uses a CNN to regress the relit image and is trained on a large synthetic dataset. The trained model weights are not available, so we trained from scratch for each scene using the same number of training samples as all other methods.
- **NeuralRTI** involves an encoder-decoder architecture. For our experiments the architecture has been modified according to the number of lights to accept an input of N images. Note that this paper also proposed the RealRTI and SynthMulti datasets, so train and test images are kept as proposed.
- **onTheGoRTI** accepts a variable number of input images, projects the MLIC with PCA followed by a MLP. For each dataset we used the same training and test images as discussed before and trained a model for each scene, as described in the original paper.

Comparisons against single-image SVBRDF-based and NeRF-based methods — which can be adapted for RTI applications albeit not being explicitly designed for that — are discussed in §5.4 separately.

Table 2 shows average values for single and multi-material synthetic and RealRTI datasets. The proposed model outperformed the other approaches for all the metrics, producing higher accuracy on all selected geometric models.

Some qualitative results from all compared methods are shown in Figures 8, 9, and 10 [3]. In Figure 8 we focus on synthetic data and show the per-pixel errors on relighted images. The ground truth image represents a lighting direction not seen during training, while all other images are the relighted outputs produced by the different methods and are associated with error maps showing per-pixel Euclidean distances in RGB space with respect to the ground truth. Figure 9 shows additional samples taken from multi-material and RealRTI datasets. In general, our approach outperforms the compared methods in both synthetic and real samples producing better surface details, physically plausible self-shadows, and specularities. For some instances, classic PTM and RBF methods completely fail on shadow and specular areas (see for instance the third column of Figure 9), and other techniques exhibit a clear shift in color representation. Additionally, learning-based methods sometimes show poor image quality or unrealistic artifacts. This is noticeable when looking at the coin's shadow (Figure 10), where we highlight a closeup of the top-left region of the image. Other methods fail to recover a sharp shadow, producing artifacts around the coin edges.

5.4 Comparison Against Single Image and NeRF-based Approaches

As discussed in Section 2, image relighting is a broad topic spanning different use cases (from photo editing of portraits to view synthesis, computer graphics, etc.) and multiple approaches, including SVBRDF-based and multi-view NeRF-based methods. Even if they do not target RTI explicitly, we choose a subset of state-of-the-art works for each category to have a comparison on how those perform in an RTI context.

We selected Single Image Neural Material Relighting (SINMR) [Bieron et al. 2023], Single-Image SVBRDF estimation with learned gradient descent (SI-SVBRDF) [Luo et al. 2024], and weakly-supervised Single-View Image Relighting (SVIR) [Yi et al. 2023b] in the category of single-image SVBRDF based relighting. Moreover, we selected relighting neural radiance fields with shadow and highlight hints [Zeng et al. 2023] in the category of NeRF-based relighting.

In particular, SINMR was trained from scratch (one model for all scenes) using the code provided by the authors and the training set images from all scenes in our synthetic dataset. Since we compared on our synthetic dataset, all data has the same FOV values (very similar to the original paper), and (gamma/log) was encoded correctly. Regarding NeRF approach, following the original paper, we trained one model for each scene from scratch using the training samples as defined at the beginning of §5. Since NeRF requires multiple views of the same object, we provided the only view available when in our RTI data. For SI-SVBRDF and SVIR we used pre-trained weights as published by the authors. Finally, for the methods requiring SVBRDF data, we bypassed the re-rendering that reconstructs the image from SVBRDF parameters and instead used our synthetic dataset captured under known lighting directions.

Table 3 summarizes the results achieved by each method on our synthetic dataset. Our model outperformed the NeRF-based by a slight margin, which is interesting because we provided just a single view of the object on a method designed to take advantage of the scene by modelling the radiance in a volume. On the other hand, the NeRF-based approach is an order of magnitude slower than our method to train (approx. 20 hours versus 2 hours on a GeForce 4080 GPU). Considering that it is an implicit neural representation that has to be trained for every object of a dataset, such a difference is not negligible in many applications where time efficiency is requested. Additionally, our model is simpler and smaller in size.

SINMR and SI-SVBRDF methods produced inaccurate results, probably because neither of these methods can handle complex cast shadows and large highlights (see in particular the first two columns of Fig. 11). Besides, there is a noticeable colour shift and artifacts on the relighted images. Both SINMR and SI-SVBRDF are not view/light configuration agnostic,

[3] Additional qualitative results for multi-material synthetic images are given in supplementary material.

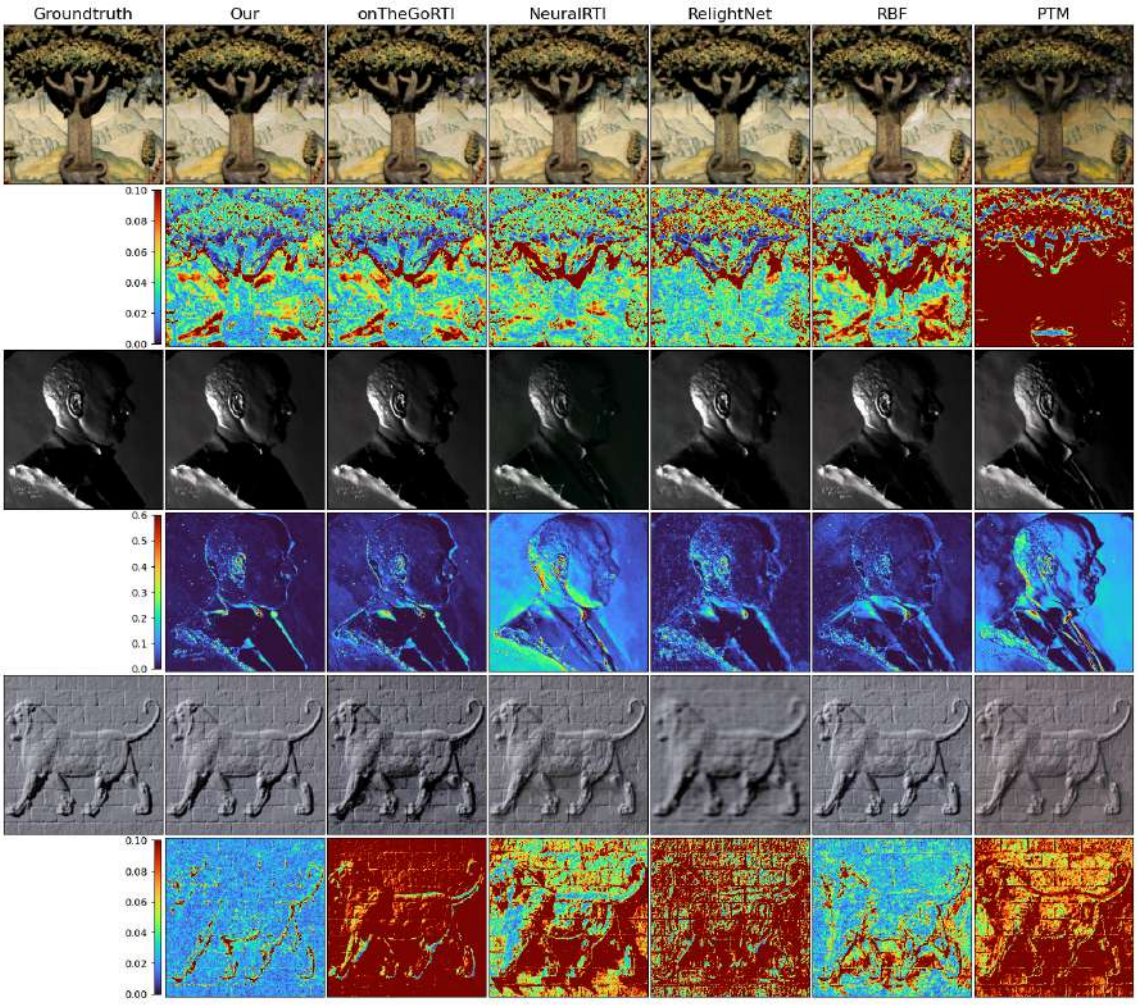

Fig. 8. Relighted images by different methods on Synthetic data. Rows 2,4,6 show Euclidean distances in RGB space between the ground truth and related images of the corresponding method.

and also their generalization capability on materials not seen during training is poor. Finally, the quantitative and qualitative results of SVIR suggest that this method is not suitable to be applied to specific RTI data, despite performing well in augmented reality application datasets.

In order to provide an additional analysis on the behaviour of NeRF in a realistic RTI scenario, we also compared our method with NeRF on the RealRTI dataset, with the idea of exploring its behaviour on real-world data acquired with standard RTI protocols. Note that in this case we avoid comparisons for other single-image relighting methods due to the values reported in previous experiment. Quantitative results for RealRTI dataset are reported in Table 4. We can observe that in this case the performances of NeRF are lower with respect to our technique, suggesting that the NeRF-based model could not be suitable for typical RTI setups. This can be caused by several factors, that are different material properties or less training images (recall that RealRTI training sets range from 37 to 62).

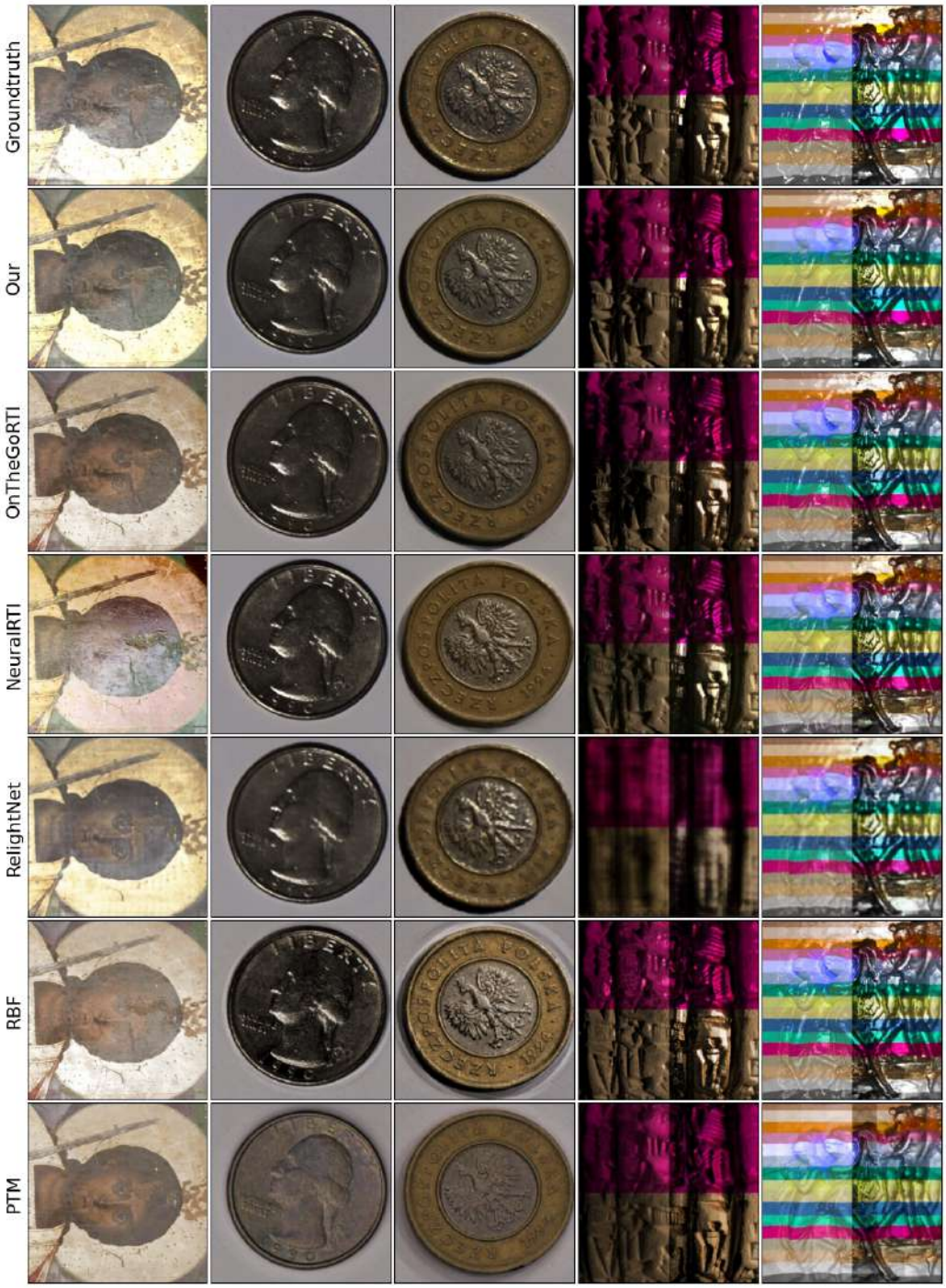

Fig. 9. Qualitative comparison for real-world and synthetic multi-material data.

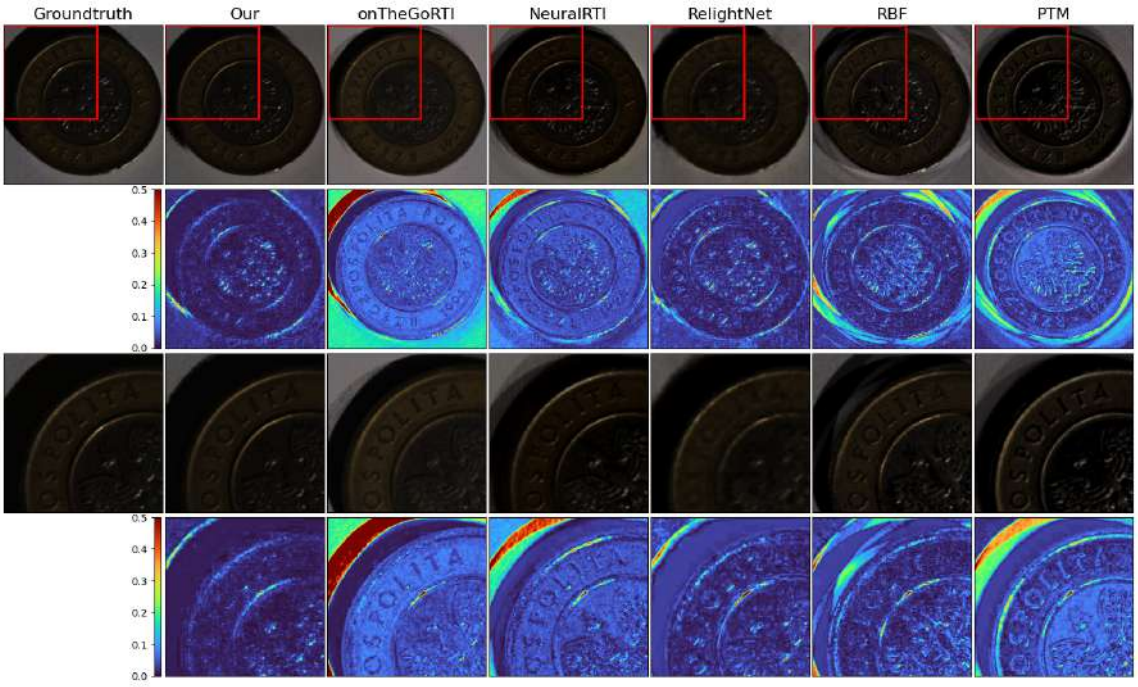

Fig. 10. Relighting results of a real-world scene for different methods. Lines 2 and 4 show per-pixel error maps in RGB space and the third row displays a shadow detail from the highlighted red region.

Table 3. Average metric values showing the comparison of our method against other NeRF and Single-image based relighting methods on our synthetic dataset and the respective model weight sizes.

Dataset	Model	Network Size (Bytes)	PSNR	SSIM	LPIPS	RMSE
Synthetic	Our	0.82 M	31.924	0.912	0.054	0.049
	NeRF-based	3.16 M	31.923	0.906	0.056	0.050
	SINMR	0.6 M	15.501	0.650	0.179	0.180
	SI–SVBRDF	640 M	14.541	0.460	0.200	0.317
	SVIR	16 M	12.273	0.307	0.290	0.365

Table 4. Average metric values showing the comparison of our method against NeRF on RealRTI dataset.

Dataset	Model	PSNR	SSIM	LPIPS	RMSE
RealRTI	Our	34.653	0.929	0.083	0.033
	NeRF-based	33.777	0.905	0.122	0.042

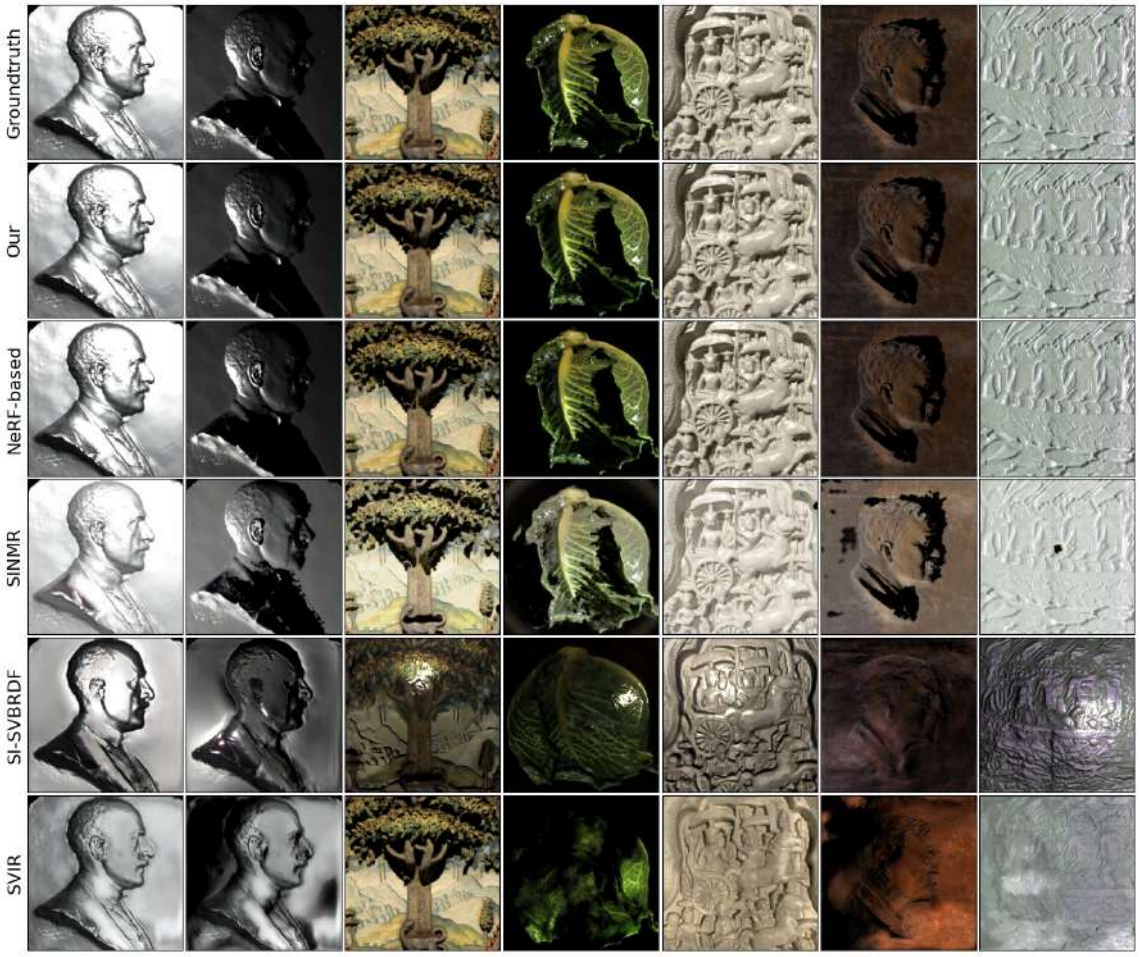

Fig. 11. Qualitative comparison of our method against other NeRF and Single-image based relighting on our synthetic dataset

5.5 Discussion on Model Sizes

In RTI applications, the amount of memory a model uses to represent the light transport function is important as well. For this reason, we conclude with a short discussion on how the different models used in the experiments compare with respect to the number of coefficients and storage size.

RBF requires much more space than all other methods because it simply stores the entire MLIC. PTM stores 6 coefficients for luminance and 2 for the chrominance for each pixel. RelightNet uses 9 sparse samples with a total of 45 coefficients (5 channels per sparse sample) per pixel and computes the relighting with an 11-layer convolutional encoder-decoder network, including fully connected layers to encode light direction ($\approx 12M$ weights total). In NeuralRTI the pixel data is encoded to a 9-dimensional vector, with the decoder network consisting of 3 fully connected layers with $3N$ units each, for a total of $3N^2 + 5N$ weights to store[4]. For onTheGoRTI, per-pixel information is stored in 8

[4] N is the number of lights (i.e. images of the MLIC).

values for luminance and 2 additional channels for chrominance. Their model is an MLP and requires 1252 values for network weights and $653K$ for PCA vectors (total $\approx 654K$ coefficients). The compared NeRF-based method requires 3.16 MB to store all the network weights, requiring a training time of about 20 hours. Finally, our proposed model requires 7 coefficients for storing the latent vector for each pixel. In the luminance-only version, we need to store 2 additional channels for average hue $\bar{H}$ and saturation $\bar{S}$, but the INR is slightly smaller since the last layer produces a single output, so the network weights are $\approx 205K$ for a total of $\approx 794K$ values to store for a 256×256 image (≈ 3.1 MB assuming single precision floating point numbers). In the RGB version (the one we used for comparison in Table 3), the INR requires 512 more weights for the last MLP layer, taking 0.82 MB of storage for the network only. Adding the 7 coefficients for each pixel, the total memory required for the RGB model is 2.66 MB for a 256×256 image.

Note that, as shown in Fig. 7, if we reduce the latent vector size to 4 coefficient we would get an average PSNR lowered by 1.5% with a memory usage of $\approx 598K$ values (2.3 MB).

5.6 Discussion on Dynamic Visualizations

To complete the qualitative comparisons, we generated a set of videos[5] to visually compare the relighting results in a dynamic setup. We selected the best performing methods, namely: NeRF-based, RBF, NeuralRTI, and onTheGoRTI. During the videos, the light moves around in a circle and progressively raises, increasing the elevation and tracing a spiral on the upper hemisphere. For each video we generated 900 frames in total, in order to obtain 30 second duration at 30 fps. Note that for our synthetic dataset we also have the ground truth for each individual frame, so we also plotted the per-pixel absolute error, while for other dataset that was not possible.

In Figure 12 we show a per-frame comparison of two objects from our synthetic dataset. Each plot shows the rendered light positions for each method, and the colour denotes the mean absolute error for that specific frame (the colour scale is kept the same for the entire row for comparison). The initial part of the videos corresponds to small light elevations, which are plotted as the outer region of the spiral in the 2D projection. We can observe that the NeRF approach has a very similar behaviour compared to our method, as already assessed by the similar average in Table 3, but exhibits a slightly worse performances for lower lights, probably due to the shadow complexity. We can also observe from the videos some artefacts located in correspondence of the shadow edges where transitions are faster. Such effects appear sometimes as irregular shadows for our method as well as for the NeRF approach, which in particular exhibits wrong shadows disconnected from the main one, creating a "bubble-like" effect. In Figure 13 two examples of such effect are reported. Regarding the comparison with other methods (RBF, NeuralRTI, onTheGoRTI), the videos show that they tend to generate less smooth transitions, with partial shadows appearing very quickly rather than gradually enlarging or reducing as the light moves. Moreover, for some methods the shadow appears brighter, and in some cases we observed that NeRF generates blurred images, losing surface details, or alters the overall scene brightness. Specific objects present unnatural artefacts in our method as well as for other approaches. For example, *item 7* object from RealRTI exhibits unnatural shadows at the beginning (1st second) and at the end of the video for our approach, while other methods produce a fast decreasing in the scene brightness (see NeRF at second 26, onTheGoRTI at second 4). This is probably due to the lack of data in the specific extreme light angles. Another example is *object 5* of our synthetic dataset, where both our model and NeRF exhibit some doubling artifacts in the shadows on the third leg from the left. This limitation could be caused by the limited receptive field of the convolutional encoder, so the object casting

[5] Available in the supplementary material.

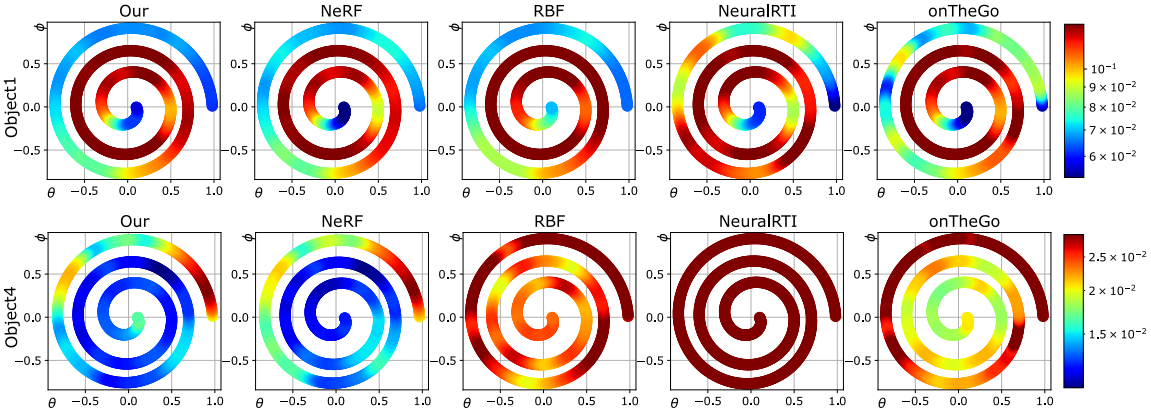

Fig. 12. Video frame comparison from our synthetic dataset. Each point corresponds to a light position on the upper hemisphere and its colour denotes the average absolute error for that specific predicted frame.

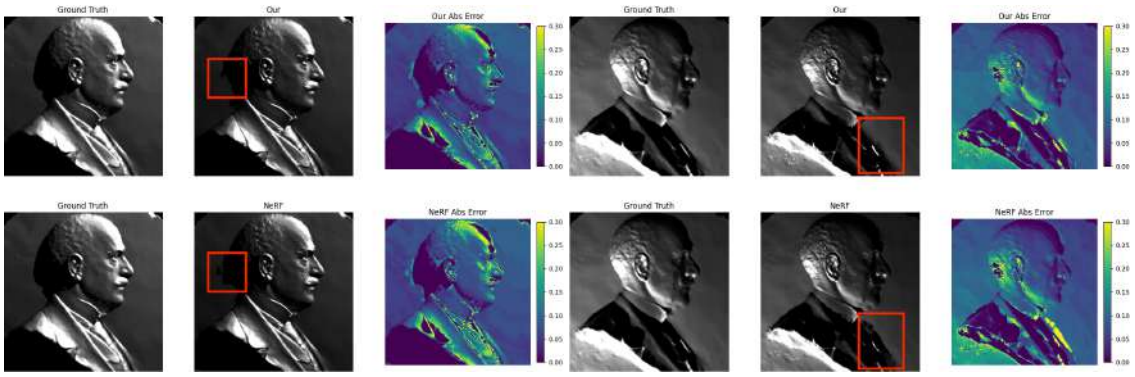

Fig. 13. Artifact comparison between our method and NeRF approach on two frames from the qualitative video comparison.

the shadow is not captured. An interesting improvement would be investigating other network structures (such as transformers) to capture non-local relationships on the surface and have a better shadow representation.

To conclude, despite our method still exhibits some artefacts, the qualitative comparison under dynamic illumination shows that the improvement with respect to other works is significant, especially in the shadow representation.

6 CONCLUSION

In this paper, we propose a novel Neural Reflectance Field method specifically designed for RTI applications. We feed the model with light and pixel coordinates, adding a spatial-aware latent vector $\mathbf{v}_{u,v}$ that encodes the local behaviour of a pixel. This specific encoding allows us to better grasp shadows and specular highlights by considering local information in a small region. Compared to other INR approaches, this results in a more accurate result with reduced storage space. Additionally, we propose a synthetic dataset designed to provide challenging MLIC data for RTI applications. The dataset, together with the code used to generate it, is publicly available, so that it can be used as a common benchmark for RTI and other related applications. Our proposed approach outperforms current state-of-the-art methods, especially in reproducing physically accurate shadows and specular highlights.

REFERENCES

Jens Ackermann, Michael Goesele, et al. 2015. A survey of photometric stereo techniques. *Foundations and Trends® in Computer Graphics and Vision* 9, 3-4 (2015), 149–254.

Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas. 2016. Learning to learn by gradient descent by gradient descent. *Advances in neural information processing systems* 29 (2016).

João Barbosa, J. Sobral, and Alberto Proença. 2007. Imaging techniques to simplify the PTM generation of a bas-relief. *VAST'07: Proceedings of the 8th International Symposium on Virtual Reality, Archaeology, and Cultural Heritage, UK, 2007* (01 2007), 28–31.

James Bieron, Xin Tong, and Pieter Peers. 2023. Single Image Neural Material Relighting. In *ACM SIGGRAPH 2023 Conference Proceedings* (Los Angeles, CA, USA) *(SIGGRAPH '23)*. Association for Computing Machinery, New York, NY, USA, Article 80, 11 pages. https://doi.org/10.1145/3588432.3591515

CHI 2023. *Cultural Heritage Imaging website.* Retrieved December 27, 2023 from https://culturalheritageimaging.org/Technologies/RTI/

Victoria Corregidor, Renato Dias, Norberto Catarino, Carlos Cruz, Luís Alves, and J. Cruz. 2020. Arduino-controlled Reflectance Transformation Imaging to the study of cultural heritage objects. *SN Applied Sciences* 2 (09 2020). https://doi.org/10.1007/s42452-020-03343-4

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

Valentin Deschaintre, Miika Aittala, Frédo Durand, George Drettakis, and Adrien Bousseau. 2018. Single-image SVBRDF capture with a rendering-aware deep network. *ACM Transactions on Graphics (TOG)* 37 (2018), 1 – 15. https://api.semanticscholar.org/CorpusID:46990252

R. Dessì, C. Mannu, G. Rodriguez, G. Tanda, and M. Vanzi. 2015. Recent improvements in photometric stereo for rock art 3D imaging. *Digital Applications in Archaeology and Cultural Heritage* 2, 2 (2015), 132–139. https://doi.org/10.1016/j.daach.2015.05.002 Digital imaging techniques for the study of prehistoric rock art.

Mark S. Drew, Yacov Hel-Or, Tom Malzbender, and Nasim Hajari. 2012. Robust estimation of surface properties and interpolation of shadow/specularity components. *Image and Vision Computing* 30, 4 (2012), 317–331. https://doi.org/10.1016/j.imavis.2012.02.012

Tinsae Dulecha, Filippo Fanni, Federico Ponchio, Fabio Pellacini, and Andrea Giachetti. 2020. Neural reflectance transformation imaging. *The Visual Computer* 36 (10 2020). https://doi.org/10.1007/s00371-020-01910-9

Jonathan Dupuy and Wenzel Jakob. 2018. An Adaptive Parameterization for Efficient Material Acquisition and Rendering. *Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 37, 6 (Nov. 2018), 274:1–274:18. https://doi.org/10.1145/3272127.3275059

Graeme Earl, Philip James Basford, Alexander Bischoff, Alan K. Bowman, C. Crowther, Jacob Dahl, Michael E. Hodgson, Leif Isaksen, Eleni Kotoula, Kirk Martinez, Hembo Pagi, and Kathryn E. Piquette. 2011. Reflectance Transformation Imaging Systems for Ancient Documentary Artefacts. In *EVA*. https://api.semanticscholar.org/CorpusID:16265320

Graeme Earl, Kirk Martinez, and Thomas Malzbender. 2010. Archaeological applications of polynomial texture mapping: analysis, conservation and representation. *Journal of Archaeological Science* 37 (2010), 2040–2050. https://api.semanticscholar.org/CorpusID:53641810

Mohammadamin Emami, Safiyeh Nekouei, Hossein Ahmadi, Christian Pritzel, and Reinhard Trettin. 2016. Iridescence in ancient glass: a morphological and chemical investigation. *International Journal of Applied Glass Science* 7, 1 (2016), 59–68.

Jiahui Fan, Beibei Wang, Milos Hasan, Jian Yang, and Ling-Qi Yan. 2023. Neural biplane representation for btf rendering and acquisition. In *ACM SIGGRAPH 2023 Conference Proceedings*. 1–11.

Abdul Rehman Farooq, Melvyn Lionel Smith, Lyndon Neal Smith, and Sagar Midha. 2005. Dynamic photometric stereo for on line quality control of ceramic tiles. *Computers in Industry* 56, 8 (2005), 918–934. https://doi.org/10.1016/j.compind.2005.05.017 Machine Vision Special Issue.

Raanan Fattal, Maneesh Agrawala, and Szymon Rusinkiewicz. 2007. Multiscale shape and detail enhancement from multi-light image collections. *ACM Trans. Graph.* 26, 3 (2007), 51.

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

Pascal Gautron, Jaroslav Krivanek, Sumanta Pattanaik, and Kadi Bouatouch. 2004. A Novel Hemispherical Basis for Accurate and Efficient Rendering. *ACM SIGGRAPH 2007 Papers - International Conference on Computer Graphics and Interactive Techniques*, 321–330.

Wenhang Ge, T. Hu, Haoyu Zhao, Shu Liu, and Yingke Chen. 2023. Ref-NeuS: Ambiguity-Reduced Neural Implicit Surface Learning for Multi-View Reconstruction with Reflection. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* (2023), 4228–4237. https://api.semanticscholar.org/CorpusID:257632404

Andrea Giachetti, Irina Mihaela Ciortan, Claudia Daffara, Giacomo Marchioro, Ruggero Pintus, and Enrico Gobbetti. 2018. A novel framework for highlight reflectance transformation imaging. *Computer Vision and Image Understanding* 168 (2018), 118–131. https://doi.org/10.1016/j.cviu.2017.05.014 Special Issue on Vision and Computational Photography and Graphics.

D.B. Goldman, B. Curless, A. Hertzmann, and S.M. Seitz. 2005. Shape and spatially-varying BRDFs from photometric stereo. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, Vol. 1. 341–348 Vol. 1. https://doi.org/10.1109/ICCV.2005.219

Yuanchen Guo, Di Kang, Linchao Bao, Yu He, and Song-Hai Zhang. 2021. NeRFReN: Neural Radiance Fields with Reflections. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 18388–18397. https://api.semanticscholar.org/CorpusID:244729083

Satoshi Ikehata. 2023. Scalable, detailed and mask-free universal photometric stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13198–13207.

Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, Merlin Nimier-David, Delio Vicini, Baptiste Nicolet Tizian Zeltner, Miguel Crespo, Vincent Leroy, and Ziyi Zhang. 2022. *Mitsuba 3 renderer.* https://mitsuba-renderer.org

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

Alexandr Kuznetsov. 2021. Neumip: Multi-resolution neural materials. *ACM Transactions on Graphics (ToG)* 40, 4 (2021).

Oliver Laric. 2023. *Three D Scans.* Retrieved October 28, 2023 from https://threedscans.com/

Gaëtan Le Goïc, Amen Benali, Marvin Nurit, Christophe Cellard, Laurent Sohier, Alamin Mansouri, Alexandre Moretti, and Romain Créac'hcadec. 2022. Reflectance transformation imaging for the quantitative characterization of experimental fracture surfaces of bonded assemblies. *Engineering Failure Analysis* 140 (2022), 106582.

Junxuan Li and Hongdong Li. 2022. Neural reflectance for shape recovery with shadow handling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 16221–16230.

Céline Loscos, George Drettakis, and Luc Robert. 2000. Interactive virtual relighting of real scenes. *IEEE Transactions on Visualization and Computer Graphics* 6, 4 (2000), 289–305.

Xuejiao Luo, Leonardo Scandolo, Adrien Bousseau, and Elmar Eisemann. 2024. Single-Image SVBRDF Estimation with Learned Gradient Descent. *Computer Graphics Forum* 43, 2 (April 2024). https://inria.hal.science/hal-04484036

Linjie Lyu, Ayush Tewari, Thomas Leimkühler, Marc Habermann, and Christian Theobalt. 2022. Neural radiance transfer fields for relightable novel-view synthesis with global illumination. In *European Conference on Computer Vision*. Springer, 153–169.

Li Ma, Vasu Agrawal, Haithem Turki, Changil Kim, Chen Gao, Pedro Sander, Michael Zollhofer, and Christian Richardt. 2023. SpecNeRF: Gaussian Directional Encoding for Specular Reflections. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2023), 21188–21198. https://api.semanticscholar.org/CorpusID:266375151

Andrea Maggiordomo, Federico Ponchio, Paolo Cignoni, and Marco Tarini. 2020. Real-World Textured Things: A repository of textured models generated with modern photo-reconstruction tools. *Computer Aided Geometric Design* 83 (2020), 101943. https://doi.org/10.1016/j.cagd.2020.101943

Thomas Malzbender, Dan Gelb, and Hans Wolters. 2001. Polynomial texture maps. *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics* 2001, 519–528. https://doi.org/10.1145/383259.383320

Tom Malzbender, Bennett Wilburn, Dan Gelb, and Bill Ambrisco. 2006. Surface enhancement using real-time photometric stereo and reflectance transformation. In *Proceedings of the 17th Eurographics Conference on Rendering Techniques* (Nicosia, Cyprus) *(EGSR '06)*. Eurographics Association, Goslar, DEU, 245–250.

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

Christina Young Molly Hughes-Hallett and Paul Messier. 2021. A Review of RTI and an Investigation into the Applicability of Micro-RTI as a Tool for the Documentation and Conservation of Modern and Contemporary Paintings. *Journal of the American Institute for Conservation* 60, 1 (2021), 18–31. https://doi.org/10.1080/01971360.2019.1700724 arXiv:https://doi.org/10.1080/01971360.2019.1700724

Mark Mudge, Thomas Malzbender, Carla Schroer, and Marlin Lum. 2006. New Reflection Transformation Imaging Methods for Rock Art and Multiple-Viewpoint Display. In *IEEE Conference on Visual Analytics Science and Technology*. https://api.semanticscholar.org/CorpusID:220915

Harold Mytum and John Robert Peterson. 2018. The Application of Reflectance Transformation Imaging (RTI) in Historical Archaeology. *Historical Archaeology* 52 (2018), 489–503. https://api.semanticscholar.org/CorpusID:165243134

Ko Nishino and Shree K Nayar. 2004. Eyes for relighting. *ACM Transactions on Graphics (TOG)* 23, 3 (2004), 704–711.

Marvin Nurit, Gaëtan Le Goïc, Stephane Maniglier, Pierre Jochum, Hermine Chatoux, and Alamin Mansouri. 2021. Improved visual saliency estimation on manufactured surfaces using high-dynamic reflectance transformation imaging. In *Fifteenth International Conference on Quality Control by Artificial Vision*, Vol. 11794. SPIE, 111–121.

Gianpaolo Palma, Monica Baldassarri, Maria Chiara Favilla, and Roberto Scopigno. 2014. Storytelling of a Coin Collection by Means of RTI Images: the Case of the Simoneschi Collection in Palazzo Blu. In *Museums and the Web 2014*, N. Proctor & R. Cherry (Ed.). Silver Spring. http://vcg.isti.cnr.it/Publications/2014/PBFS14

Gianpaolo Palma, Massimiliano Corsini, Paolo Cignoni, Roberto Scopigno, and Mark Mudge. 2010. Dynamic shading enhancement for reflectance transformation imaging. *Journal on Computing and Cultural Heritage (JOCCH)* 3, 2 (2010), 1–20.

Rohit Pandey, Sergio Orts Escolano, Chloe Legendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. 2021. Total relighting: learning to relight portraits for background replacement. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–21.

Vicente J Parot, Daryl Lim, Germán González, Giovanni Traverso, Norman S. Nishioka, Benjamin J. Vakoc, and N. Durr. 2013. Photometric stereo endoscopy. *Journal of Biomedical Optics* 18 (2013). https://api.semanticscholar.org/CorpusID:334534

Julien Philip, Sébastien Morgenthaler, Michaël Gharbi, and George Drettakis. 2021. Free-viewpoint indoor neural relighting from multi-view stereo. *ACM Transactions on Graphics (TOG)* 40, 5 (2021), 1–18.

R. Pintus, T. Dulache, I. Ciortan, E. Gobbetti, and A. Giachetti. 2019. State-of-the-art in Multi-Light Image Collections for Surface Visualization and Analysis. *Computer Graphics Forum* 38, 3 (2019), 909–934. https://doi.org/10.1111/cgf.13732

Mara Pistellato and Filippo Bergamasco. 2023. On-the-Go Reflectance Transformation Imaging with Ordinary Smartphones. In *Computer Vision – ECCV 2022 Workshops*, Leonid Karlinsky, Tomer Michaeli, and Nishino (Eds.). Springer Nature Switzerland, Cham, 251–267.

Gilles Pitard, Gaëtan Le Goïc, Alamin Mansouri, Hugues Favrelière, Maurice Pillet, Sony George, and Jon Yngve Hardeberg. 2017a. Robust anomaly detection using reflectance transformation imaging for surface quality inspection. In *Image Analysis: 20th Scandinavian Conference, SCIA 2017, Tromsø,*

Norway, June 12–14, 2017, Proceedings, Part I 20. Springer, 550–561.

Gilles Pitard, Gaëtan Le Goïc, Alamin Mansouri, Hugues Favreliere, Simon-Frédéric Désage, Serge Samper, and Maurice Pillet. 2017b. Discrete Modal Decomposition: a new approach for the reflectance modeling and rendering of real surfaces. *Machine Vision and Applications* 28 (08 2017). https://doi.org/10.1007/s00138-017-0856-0

Federico Ponchio, Massimiliano Corsini, and Roberto Scopigno. 2018. A Compact Representation of Relightable Images for the Web. In *Proceedings of the 23rd International ACM Conference on 3D Web Technology* (Poznań, Poland) *(Web3D '18)*. Association for Computing Machinery, New York, NY, USA, Article 1, 10 pages. https://doi.org/10.1145/3208806.3208820

Federico Ponchio, Massimiliano Corsini, and Roberto Scopigno. 2019. RELIGHT: A compact and accurate RTI representation for the web. *Graphical Models* 105 (2019), 101040.

Gilles Rainer, Wenzel Jakob, Avra Ghosh, and Tim Weyrich. 2019. Neural BTF Compression and Interpolation. *Computer Graphics Forum* 38 (2019). https://api.semanticscholar.org/CorpusID:84178815

Mingjun Ren, Gaobo Xiao, Limin Zhu, Wenhan Zeng, and David J. Whitehouse. 2019. Model-driven photometric stereo for in-process inspection of non-diffuse curved surfaces. *CIRP Annals* (2019). https://api.semanticscholar.org/CorpusID:149539162

Peiran Ren, Yue Dong, Stephen Lin, Xin Tong, and Baining Guo. 2015. Image based relighting using neural networks. *ACM Transactions on Graphics (TOG)* 34 (2015), 1 – 12. https://api.semanticscholar.org/CorpusID:207226292

Leonardo Righetto, Mohammad Khademizadeh, Andrea Giachetti, Federico Ponchio, Davit Gigilashvili, Fabio Bettio, and Enrico Gobbetti. 2024. Efficient and user-friendly visualization of neural relightable images for cultural heritage applications. *ACM Journal on Computing and Cultural Heritage* 17, 4 (2024), 1–24.

Viktor Rudnev, Mohamed Elgharib, William Smith, Lingjie Liu, Vladislav Golyanik, and Christian Theobalt. 2022. Nerf for outdoor scene relighting. In *European Conference on Computer Vision*. Springer, 615–631.

Sunita Saha, Amalia Siatou, Alamin Mansouri, and Robert Sitnik. 2022. Supervised segmentation of RTI appearance attributes for change detection on cultural heritage surfaces. *Heritage Science* 10 (10 2022). https://doi.org/10.1186/s40494-022-00813-3

Shen Sang and M. Chandraker. 2020. Single-Shot Neural Relighting and SVBRDF Estimation. In *ECCV*.

Hiroaki Santo, Masaki Samejima, Yusuke Sugano, Boxin Shi, and Yasuyuki Matsushita. 2017. Deep photometric stereo network. In *Proceedings of the IEEE international conference on computer vision workshops*. 501–509.

Amalia Siatou, Marvin Nurit, Yuly Castro, Gaëtan Le Goïc, Laura Brambilla, Christian Degrigny, and Alamin Mansouri. 2022. New methodological approaches in Reflectance Transformation Imaging applications for conservation documentation of cultural heritage metal objects. *Journal of Cultural Heritage* 58 (2022), 274–283. https://doi.org/10.1016/j.culher.2022.10.011

William M Silver. 1980. *Determining shape and reflectance using multiple images*. Ph. D. Dissertation. Massachusetts Institute of Technology.

M.L Smith, G Smith, and T Hill. 1999. Gradient space analysis of surface defects using a photometric stereo derived bump map. *Image and Vision Computing* 17, 3 (1999), 321–332. https://doi.org/10.1016/S0262-8856(98)00136-X

Pratul P. Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T. Barron. 2021. NeRV: Neural Reflectance and Visibility Fields for Relighting and View Synthesis. In *CVPR*.

Tiancheng Sun, Jonathan T. Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul Debevec, and Ravi Ramamoorthi. 2019. Single image portrait relighting. *ACM Transactions on Graphics* 38, 4 (July 2019), 1–12. https://doi.org/10.1145/3306346.3323008

Alejandro Sztrajman, Gilles Rainer, Tobias Ritschel, and Tim Weyrich. 2021. Neural BRDF representation and importance sampling. In *Computer Graphics Forum*, Vol. 40. Wiley Online Library, 332–346.

Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. 2020. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems* 33 (2020), 7537–7547.

Ashish Tiwari, Satoshi Ikehata, and Shanmuganathan Raman. 2024. MERLiN: Single-Shot Material Estimation and Relighting for Photometric Stereo. *arXiv preprint arXiv:2409.00674* (2024).

Ashish Tiwari and Shanmuganathan Raman. 2022. Deepps2: Revisiting photometric stereo using two differently illuminated images. In *European Conference on Computer Vision*. Springer, 129–145.

Wilna Du Toit. 2008. Radial basis function interpolation. https://api.semanticscholar.org/CorpusID:123907500

Marco Toschi, Riccardo De Matteo, Riccardo Spezialetti, Daniele De Gregorio, Luigi Di Stefano, and Samuele Salti. 2023. ReLight My NeRF: A Dataset for Novel View Synthesis and Relighting of Real World Objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 20762–20772.

Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T. Barron, and Pratul P. Srinivasan. 2021. Ref-NeRF: Structured View-Dependent Appearance for Neural Radiance Fields. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 5481–5490. https://api.semanticscholar.org/CorpusID:244920653

Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 4 (2004), 600–612. https://doi.org/10.1109/TIP.2003.819861

Robert J. Woodham. 1989. *Photometric method for determining surface orientation from multiple images*. MIT Press, Cambridge, MA, USA, 513–531.

Yingyan Xu, Gaspard Zoss, Prashanth Chandran, Markus Gross, Derek Bradley, and Paulo Gotardo. 2023. Renerf: Relightable neural radiance fields with nearfield lighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 22581–22591.

Zexiang Xu, Kalyan Sunkavalli, Sunil Hadap, and Ravi Ramamoorthi. 2018. Deep image-based relighting from optimal sparse samples. *ACM Transactions on Graphics (TOG)* 37 (2018), 1 – 13. https://api.semanticscholar.org/CorpusID:21711632

Bowen Xue, Shuang Zhao, Henrik Wann Jensen, and Zahra Montazeri. 2024. A Hierarchical Architecture for Neural Materials. In *Computer Graphics Forum*, Vol. 43. Wiley Online Library, e15116.

Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K Wong. 2022. $S^3 - NeRF$: Neural reflectance field from shading and shadow under a single viewpoint. *Advances in Neural Information Processing Systems* 35 (2022), 1568–1582.

Chia-Kai Yeh, Nathan Matsuda, Xiang Huang, Fengqiang Li, Marc Walton, and Oliver Cossairt. 2016. A Streamlined Photometric Stereo Framework for Cultural Heritage. In *ECCV Workshops*. https://api.semanticscholar.org/CorpusID:8256172

Renjiao Yi, Chenyang Zhu, and Kai Xu. 2023a. Weakly-supervised single-view image relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8402–8411.

Renjiao Yi, Chenyang Zhu, and Kai Xu. 2023b. Weakly-Supervised Single-View Image Relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 8402–8411.

Abir Zendagui, Gaëtan Le Goïc, Hermine Chatoux, Jean-Baptiste Thomas, Pierre Jochum, Stéphane Maniglier, and Alamin Mansouri. 2022. Reflectance Transformation Imaging as a Tool for Computer-Aided Visual Inspection. *Applied Sciences* 12, 13 (2022), 6610.

Chong Zeng, Guojun Chen, Yue Dong, Pieter Peers, Hongzhi Wu, and Xin Tong. 2023. Relighting neural radiance fields with shadow and highlight hints. In *ACM SIGGRAPH 2023 Conference Proceedings*. 1–11.

Edward Zhang, Michael F Cohen, and Brian Curless. 2016. Emptying, refurnishing, and relighting indoor spaces. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 1–14.

Mingjing Zhang and Mark S. Drew. 2014. Efficient robust image interpolation and surface properties using polynomial texture mapping. *EURASIP Journal on Image and Video Processing* 2014 (2014), 1–19. https://api.semanticscholar.org/CorpusID:7967206

R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 586–595. https://doi.org/10.1109/CVPR.2018.00068

Chuankun Zheng, Ruzhang Zheng, Rui Wang, Shuang Zhao, and Hujun Bao. 2021. A compact representation of measured BRDFs using neural processes. *ACM Transactions on Graphics (TOG)* 41, 2 (2021), 1–15.

Qian Zheng, Boxin Shi, and Gang Pan. 2020. Summary study of data-driven photometric stereo methods. *Virtual Reality & Intelligent Hardware* 2, 3 (2020), 213–221.

Xilong Zhou and Nima Khademi Kalantari. 2021. Adversarial Single-Image SVBRDF Estimation with Hybrid Training. *Computer Graphics Forum* 40 (2021). https://api.semanticscholar.org/CorpusID:232216960