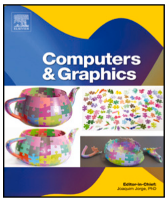
Computers & Graphics 133 (2025) 104475

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag



Special Section on STAG 2024

Fast and accurate neural reflectance transformation imaging through knowledge distillation☆

Tinsae G. Dulecha [a], Leonardo Righetto [a], Ruggero Pintus [b], Enrico Gobbetti [b], Andrea Giachetti [a],*

[a] University of Verona, Verona, Italy
[b] CRS4, Cagliari, Italy

ARTICLE INFO

Keywords:
Reflectance Transformation Imaging (RTI)
Neural RTI
DisK-NeuralRTI
Relighting
Neural relighting
Web based visualization

ABSTRACT

Reflectance Transformation Imaging (RTI) is very popular for its ability to visually analyze surfaces by enhancing surface details through interactive relighting, starting from only a few tens of photographs taken with a fixed camera and variable illumination. Traditional methods like Polynomial Texture Maps (PTM) and Hemispherical Harmonics (HSH) are compact and fast, but struggle to accurately capture complex reflectance fields using few per-pixel coefficients and fixed bases, leading to artifacts, especially in highly reflective or shadowed areas. The NeuralRTI approach, which exploits a neural autoencoder to learn a compact function that better approximates the local reflectance as a function of light directions, has been shown to produce superior quality at comparable storage cost. However, as it performs interactive relighting with custom decoder networks with many parameters, the rendering step is computationally expensive and not feasible at full resolution for large images on limited hardware. Earlier attempts to reduce costs by directly training smaller networks have failed to produce valid results. For this reason, we propose to reduce its computational cost through a novel solution based on Knowledge Distillation (DISK-NeuralRTI). Starting from a teacher network that can be one of the original Neural RTI methods or a more complex solution, DISK-NeuralRTI can create a student architecture with a simplified decoder network that preserves image quality and has computational cost compatible with real-time web-based visualization of large surfaces. Experimental results show that we can obtain a student prediction that is on par or more accurate than the existing NeuralRTI solutions with up to 80% parameter reduction. Using a novel benchmark of high-resolution Multi-Light image collections (RealRTIHR), we also tested the usability of a web-based visualization tool based on our simplified decoder for realistic surface inspection tasks. The results show that the solution reaches interactive frame rates without the necessity of using progressive rendering with image quality loss.

1. Introduction

Reflectance Transformation Imaging (RTI) is a computational photography technique widely used, especially in the Cultural Heritage (CH) domain, to interactively inspect a surface by varying its illumination to reveal details. RTI techniques create a relightable image encoding capable of producing pixel colors given a light configuration — in the most typical case, light intensity and direction. RTI encodings of objects are generated from captured data by analyzing how surface appearance changes under varying light conditions using Multi-Light Image Collections (MLICs), which in the most common case consist of a series of photographs taken from a fixed camera position, with each image illuminated by a differently positioned light [1].

The most popular and widely utilized encodings are Polynomial Texture Mapping (PTM, [2]), which is based on fitting the captured pixel value to second-order polynomial functions of the light direction components, and Hemispherical Harmonics (HSH) [3], exploiting the hemispherical basis defined from the shifted associated Legendre polynomials. These methods are fast to evaluate and very compact, as they represent per-pixel reflectance fields with few scalar components. They are, thus, the de facto standard for storing, transmitting, and interactively relighting MLICs. These methods, however, are low-frequency and often fail to suitably represent the subtle illumination effects generated by the intertwining of complex local geometric and appearance

☆ This article is part of a Special issue entitled: 'Special Section STAG2024' published in Computers & Graphics.
* Corresponding author.
E-mail addresses: tinsae.dulecha@univr.it (T.G. Dulecha), leonardo.righetto@univr.it (L. Righetto), ruggero@crs4.it (R. Pintus), gobbetti@crs4.it (E. Gobbetti), andrea.giachetti@univr.it (A. Giachetti).

https://doi.org/10.1016/j.cag.2025.104475
Received 11 April 2025; Received in revised form 31 October 2025; Accepted 3 November 2025
Available online 5 November 2025

surface features [4]. For these reasons, see Section 2, a variety of solutions have been proposed to improve their quality. Recently, neural network-based relightable image encodings [5–7] have been demonstrated to greatly enhance the effectiveness of traditional techniques, due to their ability to learn interpolable representations. In this context, NeuralRTI [7] has introduced an effective autoencoder-based solution, which has later been optimized for interactive relighting [8,9].

NeuralRTI, however, employs a decoder with numerous parameters, requiring many thousands of arithmetic operations per pixel to generate the relighted images, which hampers performance and makes them less suitable for real-time interactive object exploration, especially with high-resolution acquisitions. Consequently, recent works focused on improving viewer integration and efficiency by manually optimizing the network's layer count, speeding up the decoding with custom shaders, and implementing a level-of-detail management system that supports fine-grained adaptive rendering through dynamic resampling in the latent feature space [9].

The resulting viewer facilitates the interactive neural relighting of large images, but interactive performances are only achievable through progressive rendering in typical setups, making the user experience far from optimal. A reduction of the decoder complexity could solve this issue, but experimental tests [7,8] have shown that reducing the number and size of the layers, keeping the same training procedure, results in limited performance boosts without quality degradation. This is because, during training, the limited capacity of the small decoder to model complex reflectance functions makes it difficult to minimize the strongly non-linear loss effectively. As a result, the decoder fails to generalize well, leading to blurry or inaccurate outputs.

Building on previous work on network compression (Section 2), this work introduces a knowledge distillation technique called DisK-NeuralRTI for compressing the NeuralRTI decoder. Knowledge distillation helps by guiding the small decoder (the *student network*) with the outputs of a well-trained, larger model (the *teacher network*), simplifying the learning task and enabling better training convergence and performance. As a result, the method enables the production of high-quality relighted images with a limited fraction of the original decoding parameters, making it possible to perform a smooth interactive relighting even in the case of large images and limited computational power. To the best of our knowledge, this is the first work applying this approach to the RTI relighting domain. The results show that the resulting solution outperforms the manual tuning and is highly effective, making the Neural RTI encoding usable in practical settings.

This paper is the extension of a work published in the proceedings of STAG 2024 [10]. While, in the original paper, we applied the novel training procedure just to perform data reduction on the original NeuralRTI model, this work also evaluates the effect of using different teacher architectures. Furthermore, we improve the evaluation by introducing four additional datasets for testing both relighting quality and rendering efficiency for different types of surfaces and materials. Finally, we also evaluate the time needed to train the networks and show how to speed up the training procedure for large images and costly teacher networks.

The rest of the article is organized as follows: Section 2 presents related work on surface relighting with RTI, neural relighting, and network compression; Section 3 describes the proposed knowledge distillation solution with improved teachers and a lightweight student network. The validation of the proposed approach on the RealRTI and SynthRTI benchmarks are reported in Section 4, while Section 5 presents the benchmark comprising high resolution Multi-Light Image collections and the results obtained with our approach on it in terms of accuracy and interactive relight performances, as well as tests on strategies for downsampling of input pixels for the speed-up of the training procedure. Finally, Section 6 summarizes our findings and discusses potential future works.

2. Related work

RTI is routinely used in the CH domain to analyze surface properties [1] and is also employed in other domains, such as manufacturing [11] and quality assessment [12]. The goal of RTI is mainly to support interactive relighting, simulating the inspection of a surface with a manually controlled illumination direction. Technically, it relies on a pixel-space encoding of the reflectance behavior of the surface, depending on both shape and material properties, estimated from a Multi-Light image collection. The three main characteristics that these encodings must possess are compactness, to simplify end-to-end storage and transfer of relightable image data; smooth interpolation/approximation, to provide the illusion of continuous control of light direction; and speed, to support interactive relighting of high-resolution images on high-pixel-count displays without quality degradation. In the following, we briefly summarize the approaches that have been used to achieve these goals using shape and material separation (Section 2.1), classical (Section 2.2), and neural (Section 2.3) RTI techniques, before discussing which neural network compression methods proposed in the literature are more appropriate for our use case (Section 2.4).

2.1. Shape and material separation

A first approach to provide a compact and fast encoding of the reflectance field is to separate the shape and material components to generate a physically-based representation of the interaction between the light and the imaged object. In MLIC scenarios, such decoupled representations typically combine per-pixel maps of normals and Spatially-Varying Bidirectional Reflectance Distribution Functions (SV-BRDFs). Recovering this representation from MLIC data via photometric stereo and BRDF fitting is, however, challenging, as single-view, multi-illumination setups capture only a sparse slice of the BRDF, leading to the need for multi-view acquisitions or the use of strong analytical or learned priors [13–15]. Moreover, and most importantly, while decoupling shape and material can be effective, these methods are difficult to derive from commonly available sampled data and do not generalize easily across diverse object classes and material behaviors, for instance, semitransparent and multilayered objects [13,15]. As a result, relighting approaches instead directly approximate the reflectance by directly mapping lighting parameters (mostly direction) to final observed values, bypassing any explicit separation of shape and material [1,16].

2.2. Classical RTI

Polynomial Texture Mapping (PTM, [2,16]) and Hemispherical Harmonics (HSH, [3]) are the earlier and still most widely used compact, low-complexity reflectance field encodings proposed for relighting. They fit simple parametric functions of the light direction components to the local MLIC pixel values. They can render relighted images given a novel input direction using only a few arithmetic operations per pixel, but can only reproduce relatively low-frequency, smooth behaviors [1]. While the method achieves good results, especially for reflective surfaces, the behavior is similar to PTM and HSH when using a number of modes compatible with interactive reconstruction [17]. Low-frequency reconstructions were improved by separately modeling matte behaviors and high-frequency ones. In particular, several authors [16,18,19] have proposed using PTM or HSH for matte modeling, and a separate detail map to approximate the difference between the matte model and the original images. The storage cost of the detail coefficient map is, however, high. The multi-scale structure of the reflectance field was also harnessed by introducing Discrete Modal Decomposition (DMD) [17]. In [20], Radial Basis Function (RBF) interpolation of the original data has been proposed as an alternative idea, but the method requires run-time access to the original image data and cannot provide interactive relighting. It was later combined with Principal Component

Analysis (PCA) compression of the image stack and RBF interpolation in light space to improve efficiency at the cost of a slight reduction in quality [4].

Thanks to their versatility, compression rate, and decoding speed, PTM, HSH, and/or RBF+PCA are the encodings used in most of the publicly available web-based tools for image-based relighting, e.g., *WebRTIViewer* [21], *Pixel+ Viewer* [22], *Marlie* [23], *Relight* [4,24], and *OpenLIME* [25,26]. In this work, we aim to provide a plug-in solution based on neural compression (see Section 2.3), to improve quality, especially for high-frequency reflectance components, at storage and timing costs similar to standard solutions.

2.3. Neural-based RTI

In recent years, neural networks have proven effective for compression, nonlinear approximation, and interpolation of large datasets, and these properties have found applications in rendering tasks [27–29]. Specifically, the NeuralRTI method [7] was introduced as a direct alternative to traditional RTI representations. The method exploits a fully connected asymmetric autoencoder to represent the original per-pixel information with a low-dimensional vector. The network is trained end-to-end to accurately reproduce the training pixel values. After training, the encoder is discarded, and the resulting decoder can be used to relight the image using a novel light direction combined with stored latent features. Pistellato et al. [30] proposed a modified approach, training the decoder with PCA-compressed data to deal with a large number of input images. Due to non-linearity, the method improves over the other classical RTI solutions in reproducing complex light scattering properties and shadows, but at the cost of a computational complexity of the decoding at least two orders of magnitude higher. This fact impacts the interactivity of practical applications, particularly in the Cultural Heritage domain, where the classical techniques are still the most popular ones.

To address this issue, Righetto et al. [8] introduced a modified version of the original Neural RTI. Through a series of experiments, they manually reduced the complexity while maintaining relighting quality. However, this reduction was insufficient to ensure interactive relighting for high-resolution images in the case of limited computational power or screens with large pixel counts. Simplified decoders were tested but did not converge to good-quality representations due to the complexity of the error landscape. For this reason, they later improved interactivity by performing the decoding directly within pixel shaders and by using an adaptive multi-resolution renderer to meet rendering frequency requirements during interactive exploration [9]. This solution, however, does not provide full-quality images during relighting or when panning over high-resolution images on high-pixel-count displays , since the presented image is computed at a lower scale and upscaled at presentation time. In this work, we aim to reduce the decoding complexity by exploiting automated solutions to build reduced networks.

2.4. Neural network compression

A reasonable idea to speed up decoding is to exploit automated network compression techniques recently proposed in the literature to reduce per-pixel computing costs. These techniques are categorized into four major categories: parameter pruning, low-rank factorization, network quantization, and knowledge distillation.

Parameter pruning methods [31–33] concentrate on locating and eliminating redundant or non-essential parameters from models. These approaches can achieve significant compression rates and reduce the number of arithmetic operations. However, they often require transforming fully-connected layers into sparsely-connected configurations. This change can complicate the decoding process in GPU shaders and may hinder performance, particularly for small networks like NeuralRTI.

Low-rank factorization methods [34,35] employ matrix and tensor decomposition techniques to pinpoint the essential parameters in convolutional neural networks (CNNs). However, these strategies generally produce remarkable outcomes primarily for moderately large to very large networks, which are many orders of magnitude larger than the NeuralRTI decoder.

Network quantization techniques [36–38] lessen the number of bits needed to represent each weight, resulting in a compressed network. This size reduction may also accelerate processing by enhancing cache efficiency, but, on its own, it can only achieve a moderate speed-up when limited to data types supported by GPUs.

Lastly, knowledge distillation [39] focuses on training a more compact (student) model to imitate the behavior of a larger (teacher) model, transferring knowledge from the expansive network to the smaller one, while preserving prediction accuracy. Consequently, the student model learns to replicate the predictions of the teacher. Initially intended for classification tasks, this method has also been utilized in regression scenarios (e.g., [40,41]). This approach is effective for regressing complex nonlinear responses because the teacher model captures high-level patterns and smooth approximations of the target function that are difficult for a small model to learn directly. By mimicking the teacher's outputs, the student learns a simplified version of the complex response surface, making training more stable and enabling better generalization despite its limited capacity.

In our work, we use knowledge distillation to create a lightweight RTI decoder trained to imitate the original RTI by following the behavior of a more complex teacher network. To the best of our knowledge, we are the first to implement knowledge distillation in the context of neural relighting based on MLICs.

3. DisK-NeuralRTI

In this work, we introduce a knowledge distillation technique to improve the efficiency of neural relighting. While the method is generally applicable to reduce the size and complexity of general neural network solutions, we specialize it here for NeuralRTI [8,9]. In the following, we will first provide Background on the original NeuralRTI network (Section 3.1). We will then discuss the structure and parameterization of the student (Section 3.2) and teacher (Section 3.3), as well as the training strategy for performing the optimization (Section 3.4).

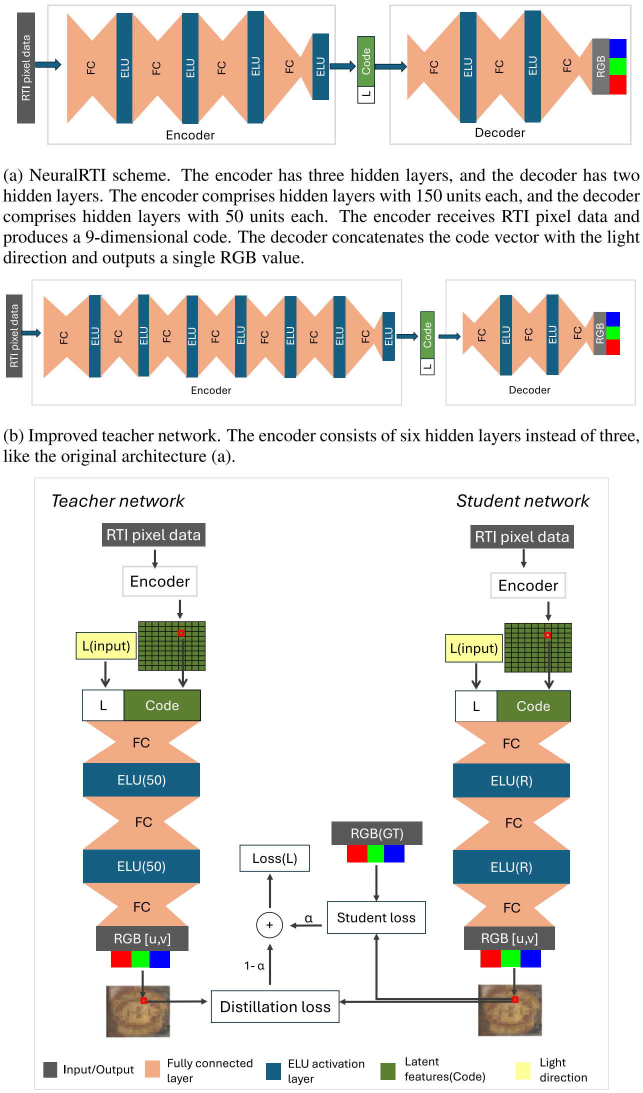
3.1. Background: The NeuralRTI network

The NeuralRTI network, as introduced in the work by Righetto et al. [8,9], is illustrated in Fig. 1(a).

In this model, the encoder comprises four layers, each with an Exponential Linear Unit (ELU) activation function. It processes per-pixel RTI data, which includes pixel values for the sampled lighting directions, and compresses this data into nine-dimensional latent space features. The decoder network, shown on the right in Fig. 1(a), includes two hidden layers, each with 50 units. It takes the concatenation of the pixel encoding and a 2D vector representing the light direction as input. The decoder's output is the predicted RGB pixel value illuminated from the specified light direction. The network is trained end-to-end on the pixels of the original MLIC (all, or a subset), minimizing the mean squared error between the predicted pixel values and the ground truth ones across the specified light directions.

Once the training phase is complete, the encoder is used to generate the final latent features for each pixel. Latent feature values are stored as relightable image data, and the encoder is then discarded. To compute relighted images, the learned decoder is then used to process the latent features, combined with the light directions interactively set by the user, to produce the final output.

The goal of this work is to use knowledge distillation to train a simple version of this network in a way that is not achievable directly by training it on the raw MLIC data.



(a) NeuralRTI scheme. The encoder has three hidden layers, and the decoder has two hidden layers. The encoder comprises hidden layers with 150 units each, and the decoder comprises hidden layers with 50 units each. The encoder receives RTI pixel data and produces a 9-dimensional code. The decoder concatenates the code vector with the light direction and outputs a single RGB value.

(b) Improved teacher network. The encoder consists of six hidden layers instead of three, like the original architecture (a).

(c) DisK-NeuralRTI. The encoder has the same architecture for student and teacher networks. The student network decoder contains two layers, each with an R number of units. We tested it with R values of 10 and 20.

Fig. 1. Network architecture for original NeuralRTI (top) and DisK-NeuralRTI (bottom).

3.2. Student network

The student network (Fig. 1(c), right) is designed to simplify only the decoder component, since the encoder size does not impact interactive relighting. We design it to be a version of the NeuralRTI model with the same encoder but a smaller decoder.

In the NeuralRTI model, the total number of decoder weights (W) and biases (B) is given by $W = (K+2) \times N + N \times N + N \times 3$ and $B = N + N + 3$. The K latent code values are stored per pixel, and setting $K = 9$ leads to a compression rate similar to standard PTM. Decoding complexity is mostly due to the number of decoder weights and biases. With decoder layers of size $N = 50$, as in the original network, the number of weights and related multiplications is $W = 3200$, and the number of biases and related additions is $B = 103$.

To achieve a sensible speed-up, we want to decrease the number of units in each hidden decoder layer from the original 50 to significantly smaller values, aiming to reduce the total number of parameters in the decoder while ensuring interactive relighting for large images, even with limited hardware resources.

Setting the decoder layer size to $N = 20$ reduces the number of weights and multiplications to $W = 680$ and the number of biases and sums to 43, achieving an 80% reduction in computation and memory fetches. A layer size of $N = 10$, leads, instead, to $W = 240$ weights and multiplications, and 23 biases and sums, giving a 92% reduction in computational cost and memory pressure. The theoretical speed-up is thus very significant: between ≈5× and ≈10× for these configurations.

3.3. Teacher network

The teacher network must first be able to learn a representation from the raw data directly, and then, once trained on the raw data, serve as guidance to the student network during the distillation phase. Since the original NeuralRTI autoencoder was proven to be able to perform relighting by learning from MLICs, we used it as a teacher in our original conference publication [10]. Since the efficient relighting is provided by the student network decoder, we are not constrained to use the original, manually optimized, NeuralRTI architecture as a teacher.

We tested several designs for a more complex teacher network, changing the number of encoding layers, increasing the layers' size, adding skip connections, and also trying to add the light direction information in the input pixel information, concatenated to the color data. The best results in preliminary testing on a subset of the MLIC considered were obtained with the solution shown in Fig. 1(b), as also illustrated in Section 4.

3.4. Training

Irrespective of the teacher used, we train the student network on all pixels of the original MLIC (or a subset of it) by minimizing the following loss function:

$$L = \frac{1}{n}\sum_{k=1}^{n} \alpha\|P_s - P_{gt}\|_k^2 + (1-\alpha)\|P_s - P_t\|_k^2 \qquad (1)$$

This function is a weighted combination of two components: the student loss, which measures the difference between the student's predictions (P_s) and the ground truth pixel values (P_{gt}), and the distillation loss, which measures the difference between the teacher's predictions (P_t) and the student's predictions (P_s). The parameter α determines the weight of each loss component. The distillation loss captures instead the discrepancy between the student and teacher models. Minimizing this loss during training enhances the student model's ability to replicate the teacher's predictions accurately.

The basic idea behind the approach is that training a very compact model through distillation should be more effective than training it directly on the original data. Fitting original data with the larger teacher network is easier than fitting it with the smaller student network, thanks to the larger number of parameters. At the same time, the teacher model's outputs are typically smoother/less noisy and may contain richer information than the exact regression target values coming from the original images. During distillation, the teacher model can thus provide hints about the underlying distribution of the data, which can guide the student model to learn more effectively to fit the original data and generalize better.

If the input MLIC is composed of multiple high-resolution images, the cost of training may be very high, especially when using a more complex teacher network. However, as many of the pixels, especially in the close neighborhood, often contain very similar information, there is no need to use all the pixels to train the model on all of them. We therefore experimented with learning DisK-NeuralRTI from only a subset of them. Currently, we subsample the input images before training using uniform random sampling to avoid introducing biases due to alignment with image features that could appear using a regular subsampling pattern. Section 4 illustrates the effects of training on reduced data in terms of speed and quality of achieved results.

4. Results and evaluation

A first set of experimental tests was aimed at assessing the benefits of the new training approach, determining a reasonable layer size for the simplified decoder, and showing the quality improvements obtained with the new teacher network. These tests were performed on existing low-resolution relighting benchmarks proposed in previous works [7], namely SynthRTI and RealRTI.

SynthRTI [42] is a collection of 51 synthetic MLICs rendered with the Blender Cycles engine. It is divided into two parts: SingleMaterial, featuring 24 collections created from three surfaces with 8 different materials applied, and MultiMaterial, including 27 collections created with the same 3 surfaces painted with 9 material combinations each. The simulated materials make it possible to evaluate how well the relighting techniques handle a wide range of reflective behaviors. Each collection is split into two sets of images, corresponding to two separate groups of light directions. The first set, called *Dome*, corresponds to a multi-ring light dome configuration with 49 directional lights arranged in concentric rings in the l_x, l_y plane at 5 different elevation angles $(10, 30, 50, 70, 90$ degrees). The second, called *Test*, corresponds to other 20 light directions at 4 intermediate elevation angles $(20, 40, 60, 80$ degrees). We used the *Dome* subset to train the networks and the *Test* set to evaluate the quality of the relighted images.

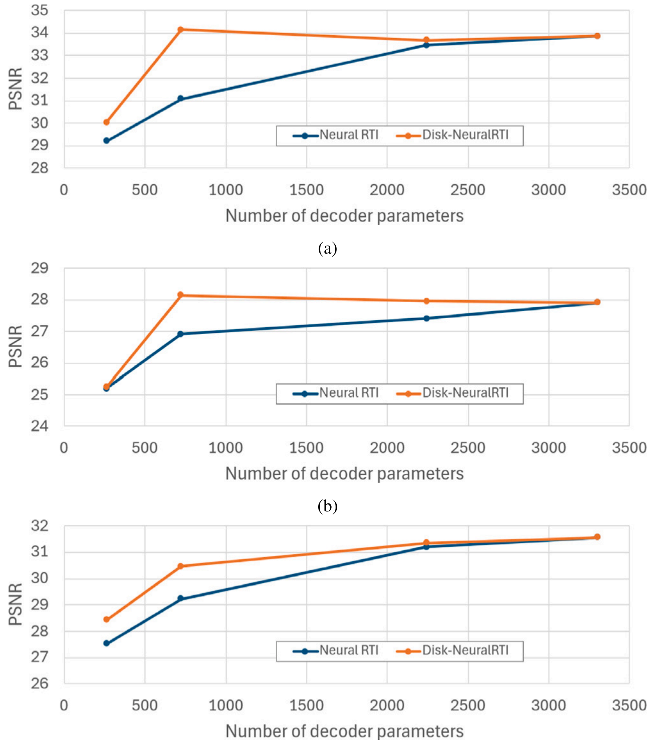
RealRTI [43] includes 12 multi-light image collections derived from real CH-related acquisitions, but cropped and resized to allow fast processing/evaluation. MLICs were acquired with both light domes and handheld RTI protocols [1] and feature surfaces with different complexity in shape and material. In the original paper [7], the testing protocol for validating the relighting was based on a leave-one-out training and testing methodology. We, instead, decided to use the same testing protocol of SynthRTI, removing 5 test images at different elevations for each collection (used then as the test set) and training the relightable images on the remaining ones. This makes the test faster but similarly informative and challenging.

For all the benchmarks, in this extended paper, we also evaluated additional metrics, namely LPIPS [44] and DeltaE [45], to assess the ability of the method to preserve perceptual similarity and chromatic information.

4.1. Setup and parameter tuning

For all the NeuralRTI training procedures, we used 90% of the total (or subsampled) RTI data pixels for training and reserved 10% for validation. We chose the Adam optimization algorithm [46], with a batch size of 64, a learning rate of 0.01, a gradient decay factor of 0.9, and a squared gradient decay factor of 0.99. For the distillation tasks, we determined with a preliminary test a single value of the parameter α and used it for all the benchmarks. In detail, we evaluated the average relighting quality on five real captures from the ReaRTI dataset, varying the value of α in the range [0.1–0.9]. It was possible to observe that the method is not very sensitive to variations of α in the range [0.1..0.7], and there is a drop in quality only when alpha exceeds 0.8. We therefore set $\alpha = 0.6$.

The network implementation leverages the PyTorch library. After training, per-pixel latent codes are converted to single bytes using offset/scale mapping and then stored as image byte planes. The decoder parameters (weights and biases) and the metadata are saved in a JSON file. The JSON header and the images with the latent codes can be used by an interactive web viewer, which is based on a custom shader that reads the decoder data and executes the decoding code, providing a fast estimation of the imaged surface relighted under novel, arbitrary light directions [8,9].



(a)

(b)

(c)

Fig. 2. Line charts showing the relighting quality (PSNR) as a function of the number of decoder parameters for the SynthRTI Multi-Material benchmark (a), the SynthRTI Multi-Material benchmark (b), the RealRTI benchmark (c). Training with DisK-NeuralRTI results in metrics close to or better than the teacher for the 723 parameters version.

4.1.1. Student performance vs. decoder size

A first series of experiments was aimed at evaluating the effect of the Knowledge Distillation approach applied with the original network as the teacher and at finding a reasonable target size for the lightweight decoder. Figs. 2(a), 2(b), and 2(c) represent the effect of the reduction of the decoder parameters on the quality of the relighting (measured by the PSNR of the comparison of relighted images and reference test images). Using the standard NeuralRTI training (blue lines), the quality becomes poor with the standard when the decoder layers are reduced from the original 50 units (3303 decoder parameters) to 40 (2243), 20 (723) and 10 (263) units. For the DisK-NeuralRTI training, instead, the results are still very good with 723 parameters, even better than the original in the case of synthetic data (orange lines). The decrease of the PSNR for smaller layers suggests that 20 can be a nearly optimal solution. Tables 1, 2, and 3 show the average PSNR and SSIM values of the comparisons between the ground truth test images and the ones relighted with different methods. Looking at the first three columns, it is possible to appreciate the large improvement provided by the proposed training procedure based on knowledge distillation relative to the standard training of a decoder of the same size, also discussed in the conference paper [10].

Fig. 3 shows an example of quality improvements deriving from the proposed compression strategy.

4.1.2. Improvements of the relighting quality with the DisK-NeuralRTI approach and the enhanced teacher

Using the original (3303 parameters) and the lightweight (723) decoder, we performed extensive tests on the benchmarks, also increasing the complexity of the encoder/teacher as described in Section 3.

Table 1
Average PSNR/SSIM values for the relighting of test images of SynthRTI SingleMaterial collections. DisK-NeuralRTI provides very good results with a per-pixel encoding size of 9 parameters (as PTM and PCA/RBF) and a sufficiently small number of shared decoding parameters per image. With a layer size of 20 elements. it provides better metrics than the teacher networks. Bold figures indicate the best values. Figures in parentheses indicate the network layers' size.

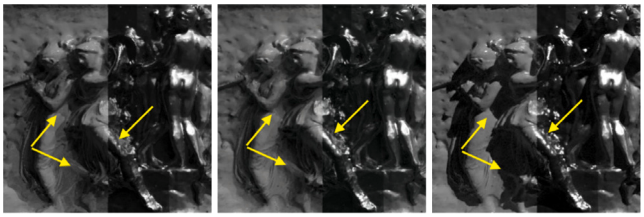
	NeuralRTI	NeuralRTI	DisK-NeuralRTI	NeuralRTI-IT	DisK-NeuralRTI-IT	PTM	HSH 3ord	PCA/RBF
	(50)	(20)	(20)	(50)	(20)			
Canvas	41.42/0.99	39.88/0.99	40.89/0.99	**46.15/0.99**	44.68/0.99	29.03/0.98	41.24/0.99	34.2/0.99
Tablet	29.13/0.88	26.45/0.83	30.53/0.90	**31.03/0.91**	29.78/0.89	23.79/0.81	29.92/0.87	25.87/0.80
Bas-relief	31.02/0.89	26.91/0.83	30.97/0.90	**34.08/0.94**	31.31/0.90	24.47/0.81	28.82/0.86	25.55/0.86
Average	33.86/0.92	31.08/0.88	34.13/0.93	**37.09/0.95**	35.26/0.93	25.76/0.87	33.33/0.91	28.54/0.88

Table 2
Average PSNR/SSIM values for the relighting of test images of SynthRTI MultiMaterial collections. The 20-elements layer compression achieves better results than the teacher network. Figures in parentheses indicate the network layers' size.

	NeuralRTI	NeuralRTI	DisK-NeuralRTI	NeuralRTI-IT	DisK-NeuralRTI-IT	PTM	HSH 3 ord	PCA/RBF
	(50)	(20)	(20)	(50)	(20)			
Canvas	33.33/0.96	31.94/0.95	32.63/0.95	**35.94/0.97**	33.21/0.95	25.17/0.93	30.03/0.95	27.95/0.95
Tablet	24.29/0.77	23.58/0.75	24.96/0.79	**25.57**/0.80	24.91/0.79	20.56/0.79	24.24/**0.84**	20.89/0.76
Bas-relief	26.08/0.83	25.20/0.80	26.83/0.84	**27.98/0.86**	26.48/0.83	22.34/0.76	25.10/0.79	21.54/0.81
Average	27.90/0.85	26.91/0.83	28.14/0.86	**29.83/0.88**	28.20/0.86	22.69/0.83	26.46/0.86	23.46/0.84

Table 3
Average PSNR/SSIM values for the relighting of test images of RealRTI collections. Values differ from [7] as we changed the testing protocol (see text). Figures in parentheses indicate the network layers' size.

	NeuralRTI	NeuralRTI	DisK-NeuralRTI	NeuralRTI-IT	DisK-NeuralRTI-IT	PTM	HSH 3 ord	PCA/RBF
	(50)	(20)	(20)	(50)	(20)			
Item 1	38.61/0.96	38.13/0.96	38.47/0.96	36.92/0.92	37.01/0.92	35.01/0.98	**39.66/0.98**	36.87/0.98
Item 2	36.49/0.95	36.53/0.95	36.42/0.95	**38.89/0.97**	38.50/0.97	27.66/0.96	37.88/0.98	32.11/0.96
Item 3	31.85/0.94	25.68/0.87	30.17/0.94	**36.65/0.96**	34.88/0.96	25.12/0.89	28.84/0.90	31.14/0.94
Item 4	33.49/0.95	30.17/0.92	30.86/0.91	**37.38/0.97**	36.81/0.97	25.29/0.95	32.18/0.98	32.70/0.98
Item 5	34.87/0.89	33.69/0.9	32.06/0.89	**39.80/0.95**	38.02/0.93	30.99/0.85	32.17/0.88	25.16/0.88
Item 6	38.64/0.95	38.63/0.95	37.50/0.94	38.72/0.95	34.27/0.92	33.64/0.93	**39.69/0.96**	29.91/0.89
Item 7	29.97/0.92	16.8/0.68	26.82/0.92	**39.17/0.95**	38.84/0.95	32.32/0.97	30.06/0.96	29.28/0.95
Item 8	29.60/0.87	25.24/0.8	26.87/0.85	**35.96/0.92**	35.49/0.91	29.43/0.89	29.75/0.90	27.92/0.88
Item 9	22.35/0.60	22.36/0.63	21.94/0.63	24.62/0.72	**27.37/0.80**	20.92/0.72	22.01/0.72	20.70/0.68
Item 10	23.21/0.74	22.45/0.71	23.00/0.75	**30.54/0.90**	29.25/0.89	16.90/0.60	19.92/0.66	17.99/0.65
Item 11	28.39/0.88	30.58/0.89	30.20/0.90	32.28/0.93	**35.44/0.95**	29.01/0.90	28.16/0.86	27.28/0.87
Item 12	31.33/0.90	30.53/0.89	31.31/0.89	**36.83/0.94**	33.92/0.93	29.42/0.89	30.32/0.88	28.44/0.88
Average	31.56/0.88	29.23/0.85	30.46/0.88	**35.65/0.92**	34.98/0.93	27.95/0.88	30.77/0.88	28.29/0.87



(a) NeuralRTI (20) (b) DisK-NeuralRTI (20) (c) Ground Truth

Fig. 3. (a) Relight with a test light direction of the SynthRTI multi-material set using the NeuralRTI(20) model. (b) Relight with the same light direction obtained with the DisK-NeuralRTI(20) compressed model. (c) Ground truth image corresponding to the test direction. It is possible to see (see arrows) that the layer size reduction with the original training (a) results in the loss of accuracy of the specular reflections and shadows. The image in (b) presents fewer artifacts compared with the ground truth (c).
Source: From [10].

The results presented in Tables 1, 2, and 3 show a comparison of the relighting quality obtained on the three low-resolution benchmarks, using the quality metrics employed in the original NeuralRTI paper [7].

In particular, we compare the "classical" PTM, HSH and PCA/RBF methods with the original Neural RTI [7] trained with both original (NeuralRTI (50)) and smaller decoder (NeuralRTI (20)), the student network with reduced complexity trained with the original NeuralRTI architecture (DisK-NeuralRTI), the Neural RTI version with additional layers (NeuralRTI-IT) and the student network with lighter decoder trained with this last version (DisK-NeuralRTI-IT).

The results consistently show that the increased encoder complexity gives strong advantages relative to the original one and that the training of the student network using knowledge distillation results in a very small decrease in the accuracy of the relight when the size of decoder layers is reduced to 20. The relighting with the DisK-NeuralRTI-IT methods is much better than the original NeuralRTI, despite the decoder compression. Especially on the RealRTI benchmark, the reduction of the decoder size has a limited impact on the quality metrics.

The new results on RealRTI with the improved teacher also seem to demonstrate that, differently from what was achieved in our previous conference publication [10], it is also possible to strongly enhance the accuracy of the relighting relative to third-order HSH.

Fig. 4 shows a visual example of improvements obtained with the improved DisK-NeuralRTI method relative to the original NeuralRTI with the heavier decoder. While the latter creates evident artifacts in the shadowed parts (highlighted by the yellow arrow) and cannot reproduce accurately small highlights (like the one indicated by the cyan arrow), the improved DisK-NeuralRTI provides a result quite close to the ground truth without artifacts.

The better performances of the novel method, more evident on challenging objects with complex shapes and materials, are also illustrated in Fig. 5, showing relighted images from a test direction for item 9 of the RealRTI benchmark. The original NeuralRTI with the heavy decoder and standard training fails in reproducing well the aspect of the golden metal and presents evident blending artifacts in the coin's shadow.

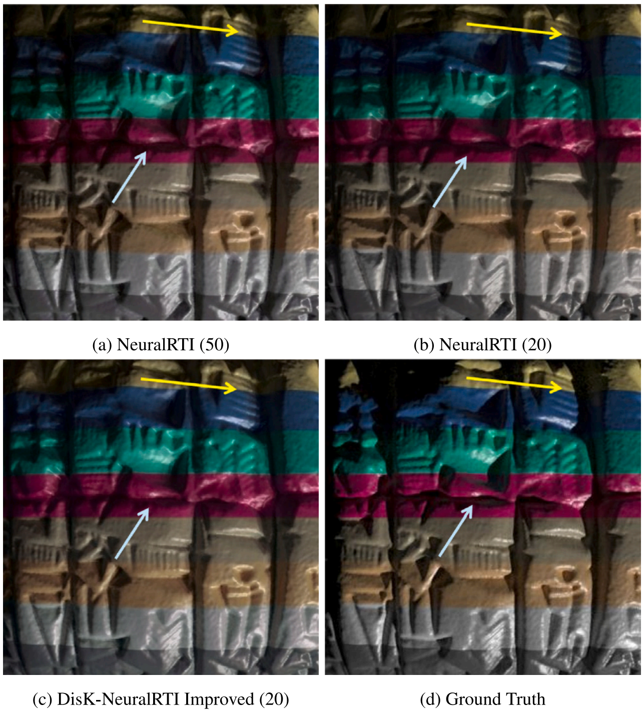

(a) NeuralRTI (50)
(b) NeuralRTI (20)
(c) DisK-NeuralRTI Improved (20)
(d) Ground Truth

Fig. 4. (a) Relight with a test light direction of the SynthRTI multi-material set using the NeuralRTI (50) model. (b) Relight with the same light direction obtained with the NeuralRTI(20) compressed model with standard training. (c) Relight with the same light direction obtained with the NeuralRTI(20) compressed model trained with improved teacher and Knowledge Distillation. The last result is the only one avoiding artifacts in shadows (yellow arrow) and non-realistic highlight (cyan arrow) compared to the ground truth (d).

Applying the DisK-NeuralRTI method, keeping the original network as the teacher, seems to improve the aspect of the golden metallic areas, but provides a poorer representation of the highlights. The architecture with a lightweight decoder trained with the improved teacher results in a better reproduction of material properties and highlights and removes the evident artifacts in the shadow (Fig. 5). Using the DisK-NeuralRTI with the improved teacher, it is possible to strongly improve the results obtained with the original model and reproduce quite well the complex material behavior of metallic surfaces (see the accurate reproduction of the highlights indicated by the green arrows). An evident improvement is also seen in the absence of blending artifacts in the cast shadow (indicated by the red arrows).

The better perceived quality of the NeuralRTI-based relighting with the novel teacher, and the effective compression obtained with the Knowledge Distillation-based approach are also evident looking at the comparisons performed using LPIPS and DeltaE, reported in Tables 4, 5, and 6.

In particular, it is possible to see that the network with the improved teacher provides, on average, a large improvement in the metrics with respect to the classical methods and the original NeuralRTI. The compression with DisK-NeuralRTI results in a non-negligible decrease of the LPIPS, which remains, however, comparable with the original network on SynthRTI and significantly better than the original NeuralRTI on RealRTI. In the case of DeltaE, measuring the preservation of the chromaticity in the relighting, the Neural method works quite well even though it does not process separately the color channels like PTM and HSH, with the improved teacher providing much better scores and, surprisingly, negligible worsening, or even improvements (on RealRTI) after the compression.

5. Evaluation of high resolution image relighting and the RealRTIHR dataset

For the evaluation of the relighting in a realistic application setting, we created a novel dataset with MLICs made of large images coming from real studies of cultural heritage artifacts featuring different material properties and shape complexity. On these images, it is possible not only to evaluate the objective quality of the relighted images, but also to test the interactive performance of the viewer application and the potential issues related to the training of the network.

5.1. *The RealRTIHR dataset*

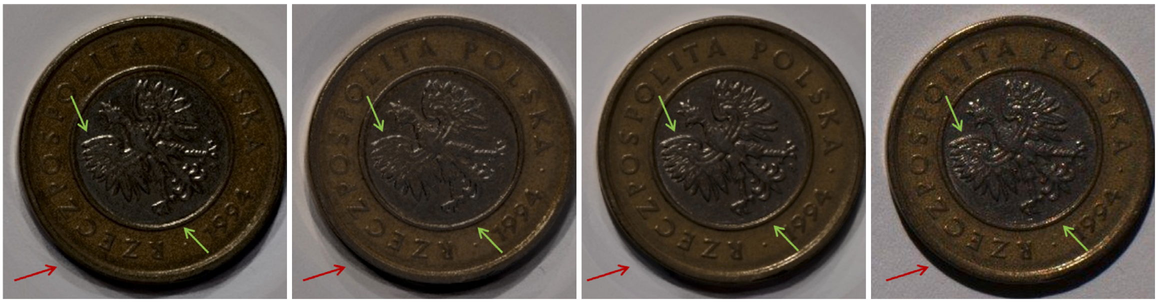
The dataset is composed of high-resolution MLICs captured in across multiple cultural heritage projects for different purposes, featuring different shape and material characteristics:

- A **lead sheet** found in the 1960s in Caesarea Maritima, during the excavations of an Italian archaeological mission (Fig. 6(a)). The item is now at the Archaeological Museum of Milan and was acquired to study the engraved inscriptions. The metallic surface is not flat and presents different degrees of roughness. The MLIC data were obtained with a light dome (47 LED) and a Nikon D810 DSLR camera. Original images were cropped to a resolution equal to 4328 × 2436 (10.5 Mp).
- **A small panel** (34 × 25 cm.) from the retable of St. Bernardino, painted in oil on a wooden support and dated 1455 (Fig. 6(b)). The item is now housed and displayed at the Pinacoteca Nazionale in Cagliari. The surface features different brilliant colors, variations in shininess, and relieved structures. The MLIC data have been captured with a 36.3 Megapixel DSLR FX Nikon D810 Camera with a 50 AF Nikkor Lens and a handheld white LED (5500K) that spans the entire visible spectrum. Images were cropped to 3811 × 2451 (9.34 Mp).
- **A larger panel** (54 × 36 cm.) from the same polyptych, featuring a golden arched frame with an image of Christ in pity, supported by an angel. It has been acquired with the same setup as the previous one. Images were cropped to 4117 × 3427 (14.1 Mp).
- **An ancient textile fragment** coming from a Viking Age burial mound at Oseberg in south Norway (Fig. 6(d)). Data is courtesy of Tomasz Łojewski (AGH University of Science and Technology, Kraków). The interesting aspect of the surface is the presence of fine patterns creating shadows in the matte surface of the tissue. Images have a resolution of 6240 × 4160 (25.9 Mp).
- **A bronze panel** representing a female figure. It is a copy of a bronze panel of Lorenzo Ghiberti's Paradise Door of the Florence Baptistery (Fig. 6(e)). The item was cast with a Cu90-Sn10 alloy, a type of bronze with very good corrosion resistance and durability. An artificial patination was applied to the surface by using Iron (III) Chloride to give the surface a brownish appearance. The item has been created for the Scan4Reco European project [20,47] The MLIC data have been captured with a 36.3 Megapixel Nikon D810 DSLR FX Nikon D810 Camera and a Handheld light and have been cropped to a resolution 4576 × 1488 (6.8 Mp).

For the benchmarking of novel view generation, we split the original MLIC data into a training and test set. This was done keeping an approximately uniform sampling in the train data and separating a minimum of 5 images with varying elevation for the test set.

5.1.1. *Relighting quality evaluation on RealRTIHR*

We used the training sets to fit all the classical and neural relighting models, and the test light directions to create the novel images to be compared against the ground-truth images, exactly as done for the SynthRTI and RealRTI benchmarks.



(a) NeuralRTI

(b) DisK-NeuralRTI

(c) DisK-NeuralRTI Improved

(d) Ground Truth

Fig. 5. Relight of a challenging object from the RealRTI benchmark. The relight obtained with the original Neural RTI method (a) reproduces the metallic behavior, but the golden part appears dark, the highlights are exaggerated with respect to the ground truth (d), and the cast shadow presents blending artifacts. Using this model to train a compressed decoder, we lose most of the highlights while the artifacts in the shadows are still there. The training of the lightweight decoder with the improved teacher, however, result in a relighted image with highlights and colors quite close to the ground truth and with reduced artifacts (c).

Table 4
Average LPIPS/DeltaE values for the relighting of test images of SynthRTI SingleMaterial collections. Bold figures indicate the best values. Figures in parentheses indicate the network layers' size.

	NeuralRTI (50)	NeuralRTI (20)	DisK-NeuralRTI (20)	NeuralRTI-IT (50)	DisK-NeuralRTI-IT (20)	PTM	HSH 3ord	PCA/RBF
Canvas	0.019/0.88	0.027/1.06	0.020/0.89	**0.014/0.39**	0.019/0.48	0.075/3.23	0.036/0.91	0.038/2.63
Tablet	0.098/2.45	0.110/2.71	0.124/2.72	**0.065**/2.32	0.094/**2.23**	0.188/4.15	0.111/2.48	0.167/3.73
Bas-relief	0.080/2.4	0.091/2.37	0.103/2.39	**0.039/1.40**	0.080/1.95	0.171/4.53	0.091/2.32	0.153/3.77
Average	0.065/1.91	0.076/2.05	0.082/2.00	**0.039/1.37**	0.064/1.55	0.145/3.97	0.079/1.90	0.119/3.38

Table 5
Average LPIPS/DeltaE values for the relighting of test images of SynthRTI MultiMaterial collections. Bold figures indicate the best values. Figures in parentheses indicate the network layers' size.

	NeuralRTI (50)	NeuralRTI (20)	DisK-NeuralRTI (20)	NeuralRTI-IT (50)	DisK-NeuralRTI-IT (20)	PTM	HSH 3 ord	PCA/RBF
Canvas	0.036/2.30	0.046/2.69	0.041/2.48	**0.022/1.65**	0.038/2.35	0.092/5.08	0.065/2.80	0.068/4.03
Tablet	0.120/4.89	0.185/6.73	0.147/5.24	**0.101**/5.23	0.119/4.56	0.217/6.53	0.129/**4.26**	0.237/7.11
Bas-relief	0.097/3.95	0.115/4.34	0.120/4.28	**0.076/3.71**	0.106/3.78	0.202/5.62	0.125/3.87	0.23/6.13
Average	0.084/3.71	0.115/4.59	0.103/4.00	**0.066/3.53**	0.088/3.56	0.17/5.74	0.106/3.64	0.178/5.76

Table 6
Average LPIPS/DeltaE values for the relighting of test images of RealRTI collections. Bold figures indicate the best values. Figures in parentheses indicate the network layers' size.

	NeuralRTI (50)	NeuralRTI (20)	DisK-NeuralRTI (20)	NeuralRTI-IT (50)	DisK-NeuralRTI-IT (20)	PTM	HSH 3 ord	PCA/RBF
Item 1	0.017/1.59	0.027/1.85	0.022/1.60	**0.017/1.88**	0.019/1.94	0.078/6.60	0.079/6.76	0.020/1.90
Item 2	0.032/1.70	0.043/1.77	0.039/1.79	0.029/1.42	0.037/1.57	0.045/3.49	**0.021/1.41**	0.054/2.86
Item 3	0.077/4.92	0.080/4.56	0.064/3.65	**0.024/2.08**	0.030/2.37	0.160/12.71	0.161/13.93	0.058/3.25
Item 4	0.025/3.03	0.033/3.52	0.028/3.60	**0.012/1.81**	0.015/2.04	0.110/16.73	0.086/12.60	0.019/2.48
Item 5	0.062/2.06	0.075/2.38	0.077/2.57	**0.042/1.52**	0.058/1.76	0.091/2.57	0.104/2.49	0.092/4.94
Item 6	0.041/1.60	0.040/1.46	0.048/1.58	0.035/1.44	**0.035**/1.47	0.055/2.06	**0.027/1.35**	0.069/3.13
Item 7	0.072/3.89	0.712/15.00	0.083/5.79	**0.037/1.07**	0.048/1.24	0.070/2.67	0.045/2.60	0.095/3.75
Item 8	0.102/4.39	0.122/7.03	0.129/5.87	0.067/**1.76**	0.076/1.83	0.095/3.21	**0.062**/2.54	0.128/3.75
Item 9	0.114/5.99	0.137/6.04	0.138/6.43	**0.098**/6.12	0.103/**4.02**	0.210/6.83	0.138/6.05	0.180/7.03
Item 10	0.147/4.96	0.163/5.48	0.142/5.12	**0.078/3.46**	0.089/3.72	0.268/10.80	0.170/7.05	0.217/8.95
Item 11	0.093/3.59	0.089/2.69	0.084/2.61	0.055/2.50	**0.049/2.06**	0.184/7.15	0.174/7.02	0.138/3.66
Item 12	0.158/2.35	0.188/2.52	0.196/2.42	**0.086/1.71**	0.155/2.06	0.214/2.60	0.168/2.37	0.258/3.11
Average	0.078/3.34	0.142/4.53	0.088/3.59	**0.048**/2.23	0.060/**2.17**	0.132/6.45	0.103/5.51	0.111/4.07

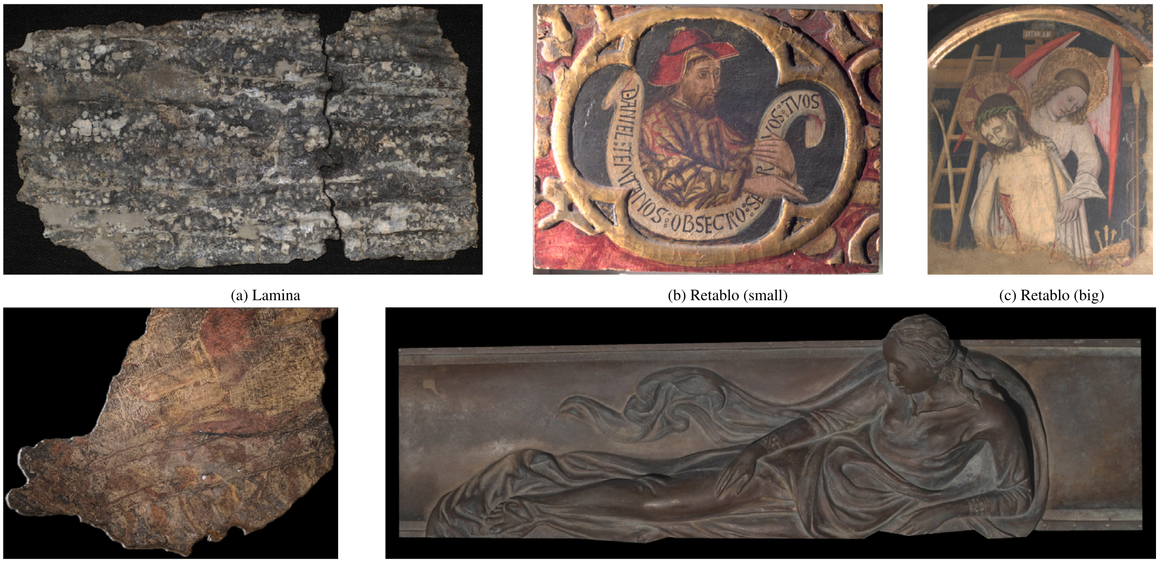
PSNR and SSIM scores obtained with the different techniques are reported in Table 7, while LPIPS and DeltaE values are shown in Table 8.

The quality provided by the neural methods is consistently better than that provided by the classical ones.

The average PSNR obtained with our method is approximately 20% higher than the one obtained with third-order HSH which is a huge difference. The average perceptual loss (LPIPS) provided by NeuralRTI-IT is halved compared to the corresponding result obtained with third-order HSH.

The improved teacher (IT) allows the method to enhance the results obtained by the previous architecture, and the metrics obtained with the lightweight decoder with layers made of 20 elements, trained with the knowledge distillation approach from the improved teacher (DisK-NeuralRTI-IT) are close to those obtained with the original NeuralRTI, and even better for SSIM and DeltaE.

The improvements in the relighting quality with the compressed NeuralRTI method are evident when compared with the best classic RTI method (Fig. 7). The third-order HSH cannot reproduce the highlights and the difference in the reflectance of different materials (a). The same relighting performed with the compressed DisK-NeuralRTI method (b)



(a) Lamina

(b) Retablo (small)

(c) Retablo (big)

(d) Textile fragment

(e) Relieved bronze panel

Fig. 6. Example images of the real-world Cultural Heritage MLICs used for benchmarking. (a): lead sheet found in Cesarea Marittima, Israel. (b), (c): Panels from the retable of St. Bernardino (1455), Cagliari, Italy. (d); Textile fragment from the Oseberg find. (e) Relieved bronze panel, copy of Lorenzo Ghiberti's Paradise Door.

Table 7
Average PSNR/SSIM of the methods on the different high-resolution datasets. The quality of the neural relight is far better with the neural model relative to classical techniques. and the compression with DisK-NeuralRTI improves performance to a level supporting interactivity while not affecting the rendering quality.

	NeuralRTI (50)	NeuralRTI (20)	DisK-NeuralRTI (20)	NeuralRTI-IT (50)	DisK-NeuralRTI-IT (20)	PTM	HSH 3 ord	PCA/RBF
Lamina	37.29/0.92	33.00/0.83	36.47/0.94	37.57/0.92	**38.33/0.94**	32.35/0.86	34.53/0.88	31.50/0.84
Retablo_small	31.68/0.84	26.84/0.72	31.04/0.82	**32.10/0.85**	31.07/0.82	23.06/0.76	24.92/0.77	24.34/0.76
Retablo_big	37.57/0.95	33.37/0.92	38.33/0.95	**38.40/0.96**	36.44/0.95	27.99/0.92	29.54/0.92	29.14/0.92
Textile fragment	29.94/0.92	29.49/0.90	30.08/0.92	**31.25/0.94**	31.11/**0.94**	29.95/0.92	30.36/0.92	29.61/0.91
Bronze panel	35.57/0.93	33.95/0.92	34.37/0.92	**35.52**/0.93	33.80/**0.94**	32.96/0.92	32.54/0.89	32.18/0.91
Average	34.41/0.91	31.33/0.86	34.06/0.91	**35.06/0.92**	34.24/**0.92**	29.22/0.85	30.34/0.85	29.35/0.85

Table 8
Average LPIPS/DeltaE of the methods on the different high-resolution datasets.

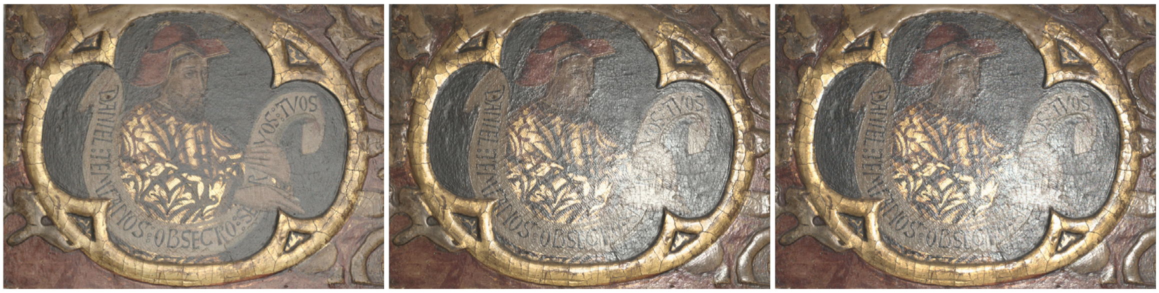
	NeuralRTI (50)	NeuralRTI (20)	DisK-NeuralRTI (20)	NeuralRTI-IT (50)	DisK-NeuralRTI-IT (20)	PTM	HSH 3 ord	PCA/RBF
Lamina	**0.014/1.11**	0.049/1.46	0.016/1.17	0.018/1.25	0.022/1.12	0.047/1.68	0.024/1.34	0.099/1.84
Retablo_small	0.014/3.61	0.030/4.84	0.017/3.80	**0.012/3.50**	0.017/3.81	0.113/6.73	0.065/5.71	0.103/6.74
Retablo_big	0.010/2.01	0.021/2.34	0.014/1.93	**0.006/1.61**	0.017/2.00	0.059/3.45	0.038/2.89	0.066/3.48
Textile fragment	0.021/4.17	0.031/5.07	0.023/4.20	**0.015/3.44**	0.020/3.54	**0.015**/3.95	0.006/3.50	0.030/4.88
Bronze panel	0.031/2.72	0.035/2.77	0.043/2.79	0.030/**2.58**	**0.029**/2.82	0.059/2.79	0.048/3.17	0.072/3.64
Average	0.018/2.72	0.033/3.30	0.023/2.78	**0.016/2.48**	0.021/2.66	0.059/3.72	0.036/3.32	0.074/4.12

results in a quite accurate highlight simulation and appears quite similar to the reference image (c). While the more costly NeuralRTI model produces a further, but very slight, quality increase, its complexity cannot ensure full-scale rendering at interactive rates for common viewport sizes (see Section 5.1.2). Thus, DisK-NeuralRTI has the highest quality among the real-time rendering methods.

5.1.2. Interactive relighting performances

NeuralRTI rendering is integrated into the OpenLIME web-based image viewer [9,26]. The implementation is based on a custom WebGL 2 shader, running the decoding algorithm in parallel on every pixel on the computer's graphics card. The shader loads the decoder's parameters stored after the data-specific training as external variables. The corresponding latent space is a matrix of dimension $H \times W \times K$, where $H \times W$ is the image resolution, and K is the number of features. The shader also loads the matrix elements, previously stored as a set of RGB JPEG images, as samplers. and executes the decoding operations: i.e., scalar product between weights and input vector, sum with the biases, and application of the activation function.

The number of operations required in this procedure is large if the decoder has too many parameters, and it could be unfeasible to use it for large images and large viewports on low-end computers. Particular strategies have been adopted to compensate for thie [9]. The image is split into tiles that are relighted independently, and only the visible tiles on the screen must be rendered. Moreover, the resolutions of the screen of the relighted image are decoupled. When the user interactively modifies the light direction or performs a zoom operation, the application tries to preserve the rendering speed by decreasing the resolution at which the decoding is performed to match a target value (e.g., 20 fps), and, finally, an upscaled version of the relighted image is displayed on the screen. When the user stops moving, the full resolution rendering is restored. Fig. 8(a) shows this effect on the web viewer: a snapshot captured during a zoom operation is blurred due to the low-resolution decoding and upsampling. Using the DisK-NeuralRTI encoding, there is

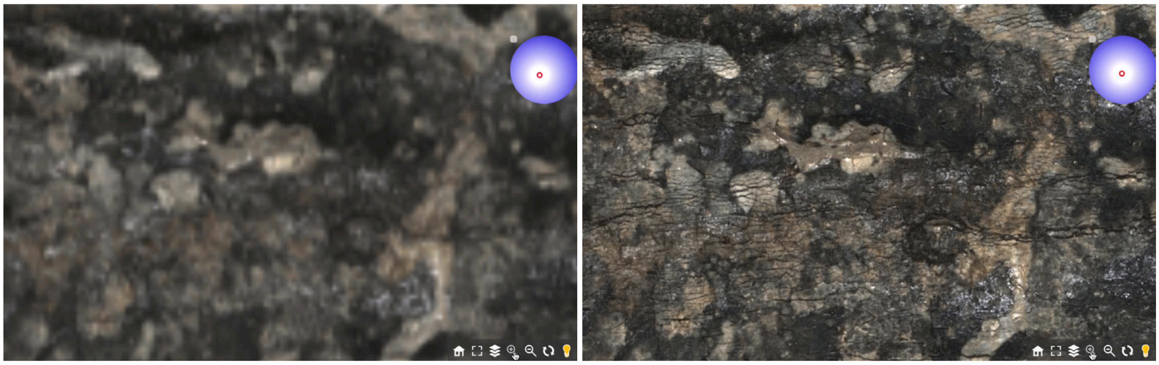


(a) HSH 3rd order

(b) DisK-NeuralRTI-IT (20)

(c) Ground truth

Fig. 7. Relighting of the Retablo (small) surface with a test light direction not included in the training set. Third-order HSH, despite the use of a heavier per-pixel encoding, fails in representing the correct reflectance behavior (a). The DisK-NeuralRTI-IT result (b) is, instead, quite close to the reference image (c).



(a) NeuralRTI-50

(b) DisK-NeuralRTI (20)

Fig. 8. Using the adaptive multiresolution rendering of OpenLIME, the system dynamically adapts the rendered images' resolution to guarantee interactivity. (a) Snapshot captured in a zooming interaction with the non-compressed NeuralRTI visualization of the Lamina surface. The image is heavily blurred. (b) Snapshot captured in a similar zooming interaction with the compressed version. Images are always sharp.
Source: From [10].

Table 9
Average fps values calculated during relighting of the three high-resolution datasets.
Source: From [10].

	Lamina (4328 × 2436)	Retablo small (3811 × 2851)	Retablo big (4117 × 3427)
DisK-NeuralRTI (20)	29.68	28.09	22.16
NeuralRTI (50)	1.60	1.42	1.10

no need for downsampling, and the snapshot captured during a similar zooming (b) is perfectly sharp.

To demonstrate that the NeuralRTI decoder with 20 units per layer and a total of 723 parameters achieves real-time relighting on low-end machines, we performed some tests with the RealRTIHR high-resolution images and the OpenLIME decoder with the image tiling and adaptive resolution options disabled. We repeated interactive image relighting in sequence with different decoder sizes, collecting the fps values and estimating averages. The evaluation was done on a MacBook Pro laptop of 2019 (1,4 GHz Intel Core i5 quad-core, graphics card Intel Iris Plus Graphics 645 1536 MB, RAM 8 GB 2133 MHz LPDDR3). The operating system was macOS Sonoma version 14.3.1 (23D60), and the web browser was Google Chrome (version 128.0.6613.138).

Table 9 shows the average refresh rate in frames per second in the interactive relighting obtained on the RealRTIHR data with the two decoder sizes. Only with the lighter ones, interactive performance is achieved without adaptive resolution.

To evaluate the size of the images that can be relighted on this platform without losing real time performances with no resolution loss we tested the relight of cropped relightable images of different size, approximately increasing the pixel number or one order of magnitude every step, from a 1000 × 1000 image (1 million pixels) to a 5000 × 2000 image (10 million pixels).

Fig. 9 illustrates the performance comparison between NeuralRTI (3303 parameters) and DisK-NeuralRTI (723 parameters) for relighting tasks, measured in frames per second (fps) as a function of the number of pixels recomputed per frame. The evaluation is performed without adaptive resolution reduction [9], and the cost includes the full processing load managed by the web renderer. The maximum achievable frame rate of 60 Hz, set by the display hardware. DisK-NeuralRTI sustains smooth, interactive performance, without any resolution reduction, above 30 fps across a wide range of resolutions (1M to 10M pixels), whereas NeuralRTI remains below 20 fps and drops below interactive thresholds once the pixel count exceeds 2 million, roughly equivalent to Full HD resolution. Video recordings of interactive relighting of RealRTIHR items on the laptop with NeuralRTI(50) and DisK-NeuralRTI(20) with and without the adaptive resolution tricks can be seen at the project link https://tgdulecha.github.io/Disk-NeuralRTI/.

5.1.3. Training performance and effect of input image subsampling

A possible drawback of increasing the encoder complexity and performing the student training is that the time required to train the final decoder may not be negligible for practical applications. We therefore analyzed the time required to complete the training of the teacher and student networks for the full pixel set of large images, as well as the effects of downsampling the set of training images, which can speed up the training, on the relight quality. Table 10 summarizes the training times obtained on the 10MP Lamina dataset with the full pixel set and different regular subsampling percentages, obtained on a computer with a 1.4 GHz Intel Core i5 quad-core processor, 32 GB of RAM, and a NVIDIA GeForce RTX 2080 Ti graphics card.

Table 10
Training times of the DisK-NeuralRTI model with the improved teacher with the full pixel set and regularly subsampled data at different ratios.

	Percentage of image pixels			
	1.56%	6.25%	25%	100%
PSNR/SSIM(Teacher)	32.48/0.86	33.8/0.87	35.21/0.90	37.57/0.92
PSNR/SSIM(DisK)	34.47/0.88	35.7/0.91	34.5/0.90	38.33/0.94
Teacher training time (min.)	6	24	53	100
Student training time (min.)	6	21	83	132
Total training time (min.)	12	45	136	232

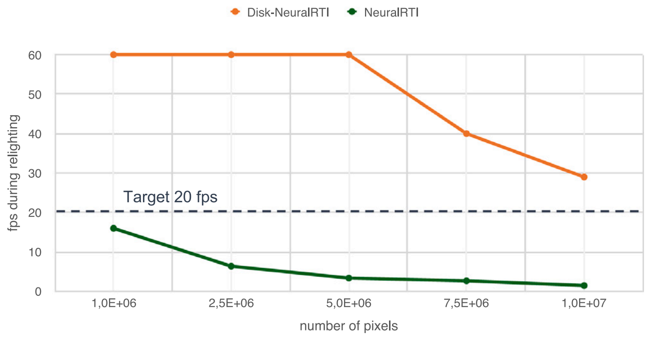


Fig. 9. Comparison between speed of exploration with NeuralRTI and DisK-NeuralRTI relighting for different numbers of recomputed pixels. The cost includes all the operations performed by the web viewer, and speed is capped at 60 Hz, which is the maximum supported refresh rate in these settings. DisK-NeuralRTI consistency achieves over 30 Hz for 1M–10M recomputed pixels per frame, while NeuralRTI is consistently below 20 fps, and not interactive starting from 2M pixels (corresponding to a standard Full HD display).

We can observe that the training times can be large on low-end machines, and a regular subsampling, e.g., 1 pixel every 8 × 8 tile, can provide a more efficient training, but results in a decrease of the achievable quality, which, however, remains still better than HSH. We plan to investigate smarter pixel sampling strategies and perform further work on training optimization as future work.

6. Discussion

The original NeuralRTI has shown its capability to preserve the compression rate of previous classic RTI solutions with a much improved reproduction of real reflectance properties of surfaces, especially high-frequency ones [7]. However, the relatively costly custom decoder used by the technique to perform the relighted image rendering may create an annoying latency for high-resolution images on low-end devices. Previous work aimed at integrating the method in an online viewer solved the issue by adapting the resolution of the rendered window to the desired frame rate during the interaction [9], at the cost of detail loss during light or camera movements, especially on large screen displays driven by commodity graphics boards. This situation is very common, for instance, when experts analyze models on laptop/mobile devices or when relightable image viewers are used for museum exploration on nowadays 4K, or even 8K, touch screens.

Using the proposed network compression approach based on Knowledge Distillation, we showed how to strongly reduce the decoding time, making it possible to render large images in real time with interactive performance on standard PCs without lowering the resolution.

Previous works have demonstrated that regressing the reflectance behavior by directly training a small network on raw reflectance data provides suboptimal results, due to the difficulty of the error landscape. Our work is the first attempt to apply automated network compression approaches to NeuralRTI, and it has given promising results.

Our tests show that the compressed encoding can guarantee smooth interactive relighting with a higher resolution than the one displayed on 4K UHD screens using low-end hardware. This makes it practical for professional cultural heritage and engineering applications.

DisK-NeuralRTI encodings can be generated from the same input data used by traditional RTI approaches, such as PTM and HSH. This compatibility allows DisK-NeuralRTI to serve as a drop-in replacement for these methods in relighting frameworks, offering significantly higher visual quality while maintaining comparable storage, transmission, and rendering performance

While the approach proposed in this work works well and the results obtained are promising, it is also useful to point out the limitations of our work and show directions for future improvements. A first one is related to the choice of the teacher network. Our initial experiments [10] used the same original NeuralRTI model as a teacher network. This architecture was initially designed with a light encoder and decoder architecture to achieve acceptable training times and interactive relighting. In this work, we have shown that by applying knowledge distillation from a deeper teacher architecture, it is possible to improve the quality of the results further or make the decoding even more efficient. Future work may improve in this area by also evaluating further modifications, both in the entire teacher network and in the encoder size of the student network, which is not used at run-time.

A second potential issue is related to the training time. Like other learning-based methods, NeuralRTI takes longer to generate a representation compared to traditional fitting-based approaches such as PTM or HSH, as it must optimize the many parameters of a non-linear function through loss minimization computed on input data. With distillation, the cost is increased, since we need to first optimize the teacher network to later optimize the desired student network. Using a more complex teacher further increases this cost. However, this is not a major concern for end-user applications, since training is done only once, is typically not time-critical, as opposed to exploration, and is still faster than acquiring a complex reflectance field of an object, especially when using GPU-accelerated nodes. In this article, we have shown that it is possible to reduce learning times by restricting training to a subset of the pixels, exploiting the redundancy present in the images. The selected method, which just performs data-independent subsampling, achieves performances higher than non-neural competitors, but decreases the quality relative to the best results achievable when training with all pixels.

To address this, future work will focus on optimizing the selection of training pixels based on their information content. Similar strategies have shown promising results in learning-based compression techniques for volume rendering [48,49].

7. Conclusion

We have shown how knowledge distillation can create more efficient Neural Reflectance Transformation Imaging (RTI) decoders for interactive object exploration in cultural heritage applications. While neural representations have demonstrated in the past superior image quality at storage costs comparable to traditional models like PTM or HSH, their decoding costs have often hindered their practical usage for real-time high-resolution exploration of large models on high-pixel-count displays. In contrast to previous manual attempts to tune network size, our approach leverages a knowledge distillation framework, where a smaller student network is trained to mimic the output of a larger, more complex teacher network, resulting in a compressed model that retains high-quality relighting capabilities.

The adoption of neural distillation has been shown to overcome the limitations of manually tuning the decoding network complexity, a process that often involves trade-offs between quality and efficiency.

To evaluate our optimization strategy, we extended the evaluation framework used in our previous work [10], also by incorporating four new datasets, introducing a comprehensive benchmark dubbed *RealRTIHR*. This benchmark covers a diverse range of surface types and material properties, and is designed to assess both relighting quality and computational efficiency under realistic usage conditions. The resulting compact model, when integrated in interactive web-based viewers, allows it to explore large, high-resolution relightable images interactively on standard hardware while preserving a high relighting quality during interaction. This contribution represents a significant step toward making neural relightable image representations more accessible and deployable in real-world cultural heritage contexts, where rendering performance, storage efficiency, and visual accuracy are all critical.

CRediT authorship contribution statement

Tinsae G. Dulecha: Writing – original draft, Software, Methodology, Conceptualization. **Leonardo Righetto:** Visualization, Conceptualization. **Ruggero Pintus:** Writing – review & editing, Data curation. **Enrico Gobbetti:** Writing – review & editing, Writing – original draft, Validation, Supervision. **Andrea Giachetti:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was partially funded by the consortium iNEST funded by the EU Next-GenerationEU PNNR M4C2 Inv1.5 – D.D. 1058 23/06/2022, ECS00000043 and by the project REFLEX (PRIN2022) funded by the EU Next-GenerationEU PNRR M4C2 Inv. 1.1. EG and RP also acknowledge the contribution of Sardinian Regional Authorities under project XDATA (RAS Art9 LR 20/2015). We thank Prof. Attilio Mastrocinque and the Civic Archaeological Museum of Milan for access to the lead sheet, the National Archaeological Museum of Cagliari for access to the retablos, Tomasz Łojewski (AGH University of Science and Technology, Kraków) for the Oseberg find artifact, and Opificio Delle Pietre Dure for the bronze panel. We thank Fabio Bettio (CRS4), Fabio Marton (CRS4), and Federico Ponchio (ISTI-CNR) for their work in developing OpenLIME and for the integration of Neural RTI rendering within the framework. We acknowledge ISCRA for awarding this project access to the LEONARDO supercomputer, owned by the EuroHPC Joint Undertaking, hosted by CINECA (Italy).

Data availability

Data will be made available on request.

References

[1] Pintus R, Dulecha TG, Ciortan I, Gobbetti E, Giachetti A. State-of-the-art in multi-light image collections for surface visualization and analysis. Comput Graph Forum 2019;38(3):909–34.
[2] Malzbender T, Gelb D, Wolters H. Polynomial texture maps. In: Proc. SIGGRAPH. 2001, p. 519–28.
[3] Gautron P, Křivánek J, Pattanaik SN, Bouatouch K. A novel hemispherical basis for accurate and efficient rendering. Render Tech 2004;2004:321–30.
[4] Ponchio F, Corsini M, Scopigno R. A compact representation of relightable images for the web. In: Proc. ACM web3D. 2018, p. 1:1–1:10.
[5] Ren P, Dong Y, Lin S, Tong X, Guo B. Image based relighting using neural networks. ACM TOG 2015;34(4):1–12.
[6] Xu Z, Sunkavalli K, Hadap S, Ramamoorthi R. Deep image-based relighting from optimal sparse samples. ACM TOG 2018;37(4):126.
[7] Dulecha TG, Fanni FA, Ponchio F, Pellacini F, Giachetti A. Neural reflectance transformation imaging. Vis Comput 2020;36:2161–74.
[8] Righetto L, Bettio F, Ponchio F, Giachetti A, Gobbetti E. Effective interactive visualization of neural relightable images in a web-based multi-layered framework. In: Proc. GCH. 2023, p. 57–66.
[9] Righetto L, Khademizadeh M, Giachetti A, Ponchio F, Gigilashvili D, Bettio F, Gobbetti E. Efficient and user-friendly visualization of neural relightable images for cultural heritage applications. ACM JOCCH 2024;17(4):54:1–24.
[10] Dulecha TG, Righetto L, Pintus R, Gobbetti E, Giachetti A. Disk-NeuralRTI: Optimized neuralrti relighting through knowledge distillation. In: Proc. GCH. 2024, p. 1–10.
[11] Nurit M, Le Goïc G, Lewis D, Castro Y, Zendagui A, Chatoux H, Favrelière H, Maniglier S, Jochum P, Mansouri A. HD-RTI: An adaptive multi-light imaging approach for the quality assessment of manufactured surfaces. Comput Ind 2021;132:103500.
[12] Coules H, Orrock P, Seow CE. Reflectance transformation imaging as a tool for engineering failure analysis. Eng Fail Anal 2019;105:1006–17.
[13] Guarnera D, Guarnera GC, Ghosh A, Denk C, Glencross M. BRDF representation and acquisition. Comput Graph Forum 2016;35(2):625–50.
[14] Zhang X, Srinivasan PP, Deng B, Debevec P, Freeman WT, Barron JT. NeRFactor: neural factorization of shape and reflectance under an unknown illumination. ACM TOG 2021;40(6):237:1–237:18.
[15] Pintus R, Ahsan M, Zorcolo A, Bettio F, Marton F, Gobbetti E. Exploiting local shape and material similarity for effective SV-BRDF reconstruction from sparse multi-light image collections. ACM JOCCH 2023;16(2):39:1–31.
[16] Zhang M, Drew MS. Efficient robust image interpolation and surface properties using polynomial texture mapping. EURASIP J Image Video Process 2014;2014(1):25.
[17] Pitard G, Le Goïc G, Mansouri A, Favrelière H, Desage S-F, Samper S, Pillet M. Discrete modal decomposition: a new approach for the reflectance modeling and rendering of real surfaces. Mach Vis Appl 2017;28(5–6):607–21.
[18] Drew MS, Hel-Or Y, Malzbender T, Hajari N. Robust estimation of surface properties and interpolation of shadow/specularity components. Image Vis Comput 2012;30(4–5):317–31.
[19] Fornaro P, Bianco A, Kaiser A, Rosenthaler L. Enhanced RTI for gloss reproduction. Electron Imaging 2017;29:66–72.
[20] Giachetti A, Ciortan IM, Daffara C, Marchioro G, Pintus R, Gobbetti E. A novel framework for highlight reflectance transformation imaging. Comput Vis Image Underst 2018;168:118–31.
[21] Palma G, et al. WebRTI Viewer. 2019, URL: http://vcg.isti.cnr.it/rti/webviewer.php [Online; accessed 20 September 2024].
[22] Vanweddingen V, Proesmans M, Hameeuw H, Vandermeulen B, der Perre AV, Vastenhoud C, Lemmers F, Watteeuw L, Gool LV. Pixel-Plus Viewer. 2020, URL: https://www.heritage-visualisation.org/pixelplusviewer.html [Online; accessed 20 September 2024].
[23] Jaspe Villanueva A, Ahsan M, Pintus R, Giachetti A, Gobbetti E. Web-based exploration of annotated multi-layered relightable image models. ACM JOCCH 2021;14(2):24:1–31.
[24] Ponchio F, et al. Relight. 2024, URL: http://vcg.isti.cnr.it/relight/ [Online; accessed 19 October 2024].
[25] OpenLime Team. OpenLime: Open Layered IMage Explorer. 2025, URL: https://github.com/cnr-isti-vclab/openlime and https://github.com/crs4/openlime [Online; accessed 06 August 2025].
[26] Ponchio F, Bettio F, Marton F, Pintus R, Righetto L, Giachetti A, Gobbetti E. OpenLIME: An open and flexible web framework for creating and exploring complex multi-layered relightable image models. In: Proc. digital heritage. 2025, p. 1–10.
[27] Ren P, Dong Y, Lin S, Tong X, Guo B. Image based relighting using neural networks. ACM TOG 2015;34(4):111:1–111:12.
[28] Xu Z, Sunkavalli K, Hadap S, Ramamoorthi R. Deep image-based relighting from optimal sparse samples. ACM TOG 2018;37(4):126:1–126:13.
[29] Tewari A, Fried O, Thies J, Sitzmann V, Lombardi S, Sunkavalli K, Martin-Brualla R, Simon T, Saragih J, Nießner M, et al. State of the art on neural rendering. Comput Graph Forum 2020;39(2):701–27.

[30] Pistellato M, Bergamasco F. On-the-go reflectance transformation imaging with ordinary smartphones. In: Proc. ECCV workshops, part i. 2023, p. 251–67.

[31] Han S, Mao H, Dally WJ. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. 2015, arXiv preprint arXiv:1510.00149.

[32] Guo Y, Yao A, Chen Y. Dynamic network surgery for efficient DNNs. In: NeurIPS. 2016;29.

[33] Yao S, Zhao Y, Zhang A, Su L, Abdelzaher T. Deepiot: Compressing deep neural network structures for sensing systems with a compressor-critic framework. In: Proceedings of the 15th ACM conference on embedded network sensor systems. 2017, p. 1–14.

[34] Tai C, Xiao T, Zhang Y, Wang X, et al. Convolutional neural networks with low-rank regularization. 2015, arXiv preprint arXiv:1511.06067.

[35] Bhattacharya S, Lane ND. Sparsification and separation of deep learning layers for constrained resource inference on wearables. In: Proc. ACM ENSS. 2016, p. 176–89.

[36] Gong Y, Liu L, Yang M, Bourdev L. Compressing deep convolutional networks using vector quantization. 2014, arXiv preprint arXiv:1412.6115.

[37] Courbariaux M, Bengio Y, David J-P. BinaryConnect: Training deep neural networks with binary weights during propagations. In: NeurIPS. Vol. 28, 2015.

[38] Hubara I, Courbariaux M, Soudry D, El-Yaniv R, Bengio Y. Binarized neural networks. In: NeurIPS. Vol. 29, 2016.

[39] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. 2015, arXiv preprint arXiv:1503.02531.

[40] Takamoto M, Morishita Y, Imaoka H. An efficient method of training small models for regression problems with knowledge distillation. In: Proc. MIPR. IEEE; 2020, p. 67–72.

[41] Saputra MRU, De Gusmao PP, Almalioglu Y, Markham A, Trigoni N. Distilling knowledge from a deep pose regressor network. In: Proc. ICCV. 2019, p. 263–72.

[42] University of Verona. SynthRTI. 2020, URL: https://github.com/Univr-RTI/SynthRTI [Online; accessed 06 August 2025].

[43] University of Verona. RealRTI. 2020, URL: https://github.com/Univr-RTI/RealRTI [Online; accessed 06 August 2025].

[44] Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR. 2018, p. 586–95.

[45] Mokrzycki W, Tatol M. Colour difference ΔE – A survey. Mach Graph Vis 2011;20(4):383–411.

[46] Kingma DP, Ba J. Adam: A method for stochastic optimization. 2014, arXiv preprint arXiv:1412.6980.

[47] Giachetti A, Ciortan I, Daffara C, Pintus R, Gobbetti E. Multispectral RTI analysis of heterogeneous artworks. In: Proc. GCH. 2017, p. 19–28.

[48] Gobbetti E, Iglesias Guitián J, Marton F. COVRA: A compression-domain output-sensitive volume rendering architecture based on a sparse representation of voxel blocks. Comput Graph Forum 2012;31(3pt4):1315–24.

[49] Díaz J, Marton F, Gobbetti E. Interactive spatio-temporal exploration of massive time-varying rectilinear scalar volumes based on a variable bit-rate sparse representation over learned dictionaries. Comput Graph 2020.