Deep Reflectance Maps

Konstantinos Rematas
KU Leuven
krematas@esat.kuleuven.be

Tobias Ritschel
University College London
t.ritschel@ucl.ac.uk

Mario Fritz
MPI Informatics
mfritz@mpi-inf.mpg.de

Efstratios Gavves
University of Amsterdam
E.Gavves@uva.nl

Tinne Tuytelaars
KU Leuven
Tinne.Tuytelaars@esat.kuleuven.be

Abstract

Undoing the image formation process and therefore decomposing appearance into its intrinsic properties is a challenging task due to the under-constrained nature of this inverse problem. While significant progress has been made on inferring shape, materials and illumination from images only, progress in an unconstrained setting is still limited. We propose a convolutional neural architecture to estimate reflectance maps *of specular materials in natural lighting conditions. We achieve this in an end-to-end learning formulation that directly predicts a reflectance map from the image itself. We show how to improve estimates by facilitating additional supervision in an* indirect *scheme that first predicts surface orientation and afterwards predicts the reflectance map by a learning-based sparse data interpolation.*

In order to analyze performance on this difficult task, we propose a new challenge of Specular MAterials on SHapes with complex IllumiNation (SMASHINg) using both synthetic and real images. Furthermore, we show the application of our method to a range of image editing tasks on real images.

1. Introduction

A classic computer vision task is the decomposition of an image into its intrinsic shape, material and illumination. The physics of image formation are well-understood: the light hits a scene surface with specific orientation and material properties and is reflected to the camera. Factoring an image into its intrinsic properties, however, is very difficult, as the same visual result might be due to many different combinations of intrinsic object properties.

For the estimation of those properties, a common practice is to assume one or more properties as known or simplified and try to estimate the others. For example, traditional

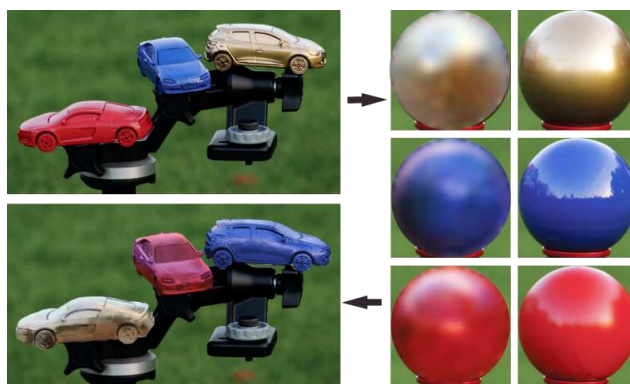Project: http://homes.cs.washington.edu/~krematas/DRM/

Figure 1. *Top:* Input 2D image with three cars of unknown shape and material under unknown natural illumination. *Right:* Our automatically extracted reflectance map and the reference. *Bottom:* Transfer of reflectance maps between the objects.

approaches to intrinsic images or shape-from-shading assume lambertian materials, or point lights. Furthermore, to simplify the problem, shape is often either assumed to be known in the form of a 3D model, or it is restricted to simple geometry such as spheres.

In this work, we extract reflectance maps [14] from images of objects with complex shapes and specular material, under complex natural illumination. A reflectance map holds the orientation-dependent appearance of a fixed material under a fixed illumination. It does not attempt to factor out material and/or illuminant and should not be confused with a reflection map that contains illumination [6] or with surface reflectance [5].

Under the assumptions of a constant material, no shadows, a distant light source and a distant viewer, the relation of surface orientation and appearance is fully described by the reflectance map. It can represent all illuminants and all materials, in particular specular materials under high-frequency natural illumination. Therefore, besides allowing for a better understanding and analysis of 2D imagery, the ability

to estimate reflectance maps lends itself to a broad spectrum of applications, including material transfer, inpainting, augmented reality and a range of image editing methods.

The input of our system is a 2D image where an object from a known class, (*e.g.*, cars), was segmented (Fig 1) and output is a reflectance map. To this end, we propose two different approaches: The first approach directly estimates a reflectance map from the input image using an end-to-end learning framework based on CNNs and upconvolutions. The second approach decomposes the process in two steps, enabling the use of additional supervision in form of object surface normals at training time. For the second approach we first predict per-pixel surface normals, which we use to compute sparse reflectance maps from the visible normals of the objects. Given the sparse reflectance map, we introduce a learned sparse data-interpolation scheme in order to arrive at the final reflectance map. In summary, we make the following five key contributions:

- First end-to-end approach to infer reflectance maps from a 2D image of complex shapes of specular materials under natural illumination.

- Dataset based on synthetic images and real photographs that facilitates the study of this task.

- A CNNs/upconvolutional architecture to learn the complex mapping from the spatial 2D image to the spherical domain.

- A CNN addressing a data-interpolation task of sparse unstructured data.

- Demonstration of our approach on a range of real images and a range of image-based editing tasks.

2. Related Work

Factoring images into their constituting components is an important goal of computer vision. It is inherently hard, as many combinations of factors can result in the same image. Having a decomposition available would help solving several important computer graphics and computer vision problems.

Factoring Images. Classic *intrinsic images* factor an image into illuminant and reflectance [3]. Similarly, *shape-from-shading* decomposes into reflectance and shading, eventually leading to an orientation map or even a full 3D shape. Larger-scale acquisition of reflectance [5] and illumination [6] have allowed to compute their statistics [7] helping to better solve inverse and synthesis problems.

Recently, factoring images has received renewed interest. Lombardi and Nishino [24] as well as Johnson and Adelson [15] have studied the relation of shape, reflectance and natural illumination. A key idea in their work is, that under natural illumination, appearance and orientation are in a much more specific relation (as used in Photometric stereo [13]) than for a single point light, where many similar appearance for totally different orientations can be present. They present different optimization approaches that allow for high-quality estimation of one component if at least one other component is known. In this work, we assume that the object is made of a single material and its object class and its segmentation mask are known. However, we do not aim at factoring out illuminant, reflectance and shape, but keep the combination of reflectance and illuminant and only factor it from the shape. Further factoring the reflectance map produced in our approach into material and illuminant would be complemented by methods such as [24] or [15].

Baron and Malik [2] factors shaded images into shape, reflectance and lighting, but only for scalar reflectance, *i.e.* diffuse albedo and for limited illumination frequencies. In a very different vein, Internet photo collections of diffuse objects can be used to produce a rough 3D shape that serves extracting reflectance maps in a second step [11].

A recent approach by Richter *et al.* [31] first estimates a diffuse reflectance map using approximate normals and then refines the normal map using the reflectance map as a guide. Different from our approach, they assume diffuse surfaces to be approximated using 2nd-order spherical harmonics (SH) and learn to refine the normals from the reflectance map using a regression forest. We compare the reflectance maps produced by our approach to reflectance maps using an SH basis which are limited to diffuse materials.

Computer Graphics. While appearance is considered view-independent in intrinsic images, view-dependent shading is described by *reflectance maps* [14]. In computer graphics, reflectance maps are popular and known as *lit spheres* [33] or *MatCaps* [32]. They are used to capture, transfer and manipulate the orientation-dependent appearance of photorealistic or artistic shading. A special user interface is required, to map surface orientation to appearance at sparse points in an image, from which orientations are interpolated for in-between pixels to fill the lit sphere (*e.g.* [30] manually aligned a 3D model with an image to generate lit spheres). Small diffuse objects in a single cluttered image were made to appear specular or transparent using image manipulations with manual intervention [16]. Our approach shares the simple and effective lit half-sphere parametrization but automates the task of matching orientation and appearance.

Deep Learning. In recent years convolutional neural networks (CNNs) have shown strong performance across different domains. In particular, the strong models for object recognition [18] and detection [10] can be seen as a layer-wise encoder of successively improved features. Based on ideas of encoding-decoding strategies similar to auto-encoders, convolutional decoders have been developed [37, 20] to decode condensed representations back to images.
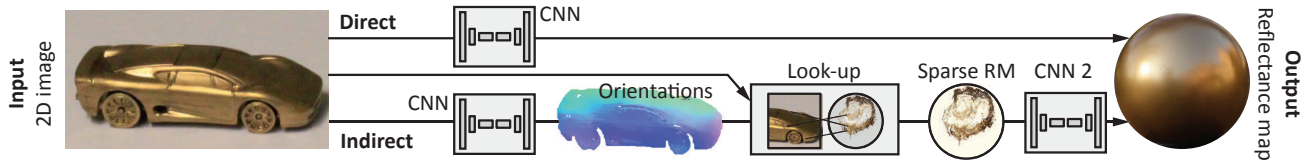
Figure 2. Overview of our approach, that comprises two variants: A direct one and an indirect one extracting surface orientations.

This has led to fully convolutional or deconvolutional techniques that have seen wide applicability for tasks where there is a per-pixel prediction target. In [25, 12], this paradigm has been applied to semantic image segmentation. In [1] image synthesis was proposed given object class, view and view transformations as input and synthesizing segmented new object instances as output. Similarly, [19] propose the *deep convolutional inverse graphics networks* with an encoder-decoder architecture, that given an image can synthesize novel views. In contrast, our approach achieves a new mapping to an intrinsic property – the reflectance map.

Deep lambertian networks [34] apply deep belief networks to the joint estimation of a reflectance, an orientation map and the direction of a single point light source. They rely on Gaussian Restricted Boltzmann Machines to model the prior of the albedo and the surface normals for inference from a single image. In contrast, we address specular materials under general illumination, but without factoring material and illuminant.

Another branch of research proposes to use neural networks for depth estimation [9, 21, 23], normal estimation [8, 35, 21], intrinsic image decomposition[27, 38] and lightness[28]. Wang *et al*. [35] show that a careful mixture of deep architectures with hand-engineered models allow for accurate surface normal estimation. Observing that normals, depth and segmentations are related tasks, [8] propose a coarse-to-fine, multi-scale and multi-purpose deep network that jointly optimizes depth and normal estimation and semantic segmentation. Likewise, [21] apply deep regression using convolutional neural networks for depth and normal estimation, whose output is further refined by a conditional random field. Going one step further, [23] propose to embed both the unary and the pairwise potentials of a conditional random field in a unified deep network. In contrast, our goal is not normal, but rather reflectance map estimation. In particular, our "direct approach" makes do without any supervision of normal information, while the "indirect approach" has normals as a by-product. In addition, our new challenge dataset captures reflectance maps and normals for the specular case, which are not well represented in prior recordings – in particular as also range sensors have difficulties on specular surfaces.

3. Model

Motivation We address a challenging inverse problem that is highly underconstrained. Therefore, any solution needs to mediate between evidence from the data and prior expectations – in our case over reflectance maps. In the general settings of specular materials and natural illuminations, modeling prior expectations over reflectance maps – let alone obtaining a parametric representation – seems problematic. This motivated us to follow a data-driven approach in an end-to-end learning framework, where the dependence of reflectance maps on object appearances is learnt from a substantial number of synthesized images of a given object class.

Overview The goal of our network is the estimation of the reflectance map of an object depicted in a single RGB image (Fig. 2). This is equivalent to estimating how a sphere [33] with the same material as the object would look like from the same camera position and the same illumination. From the estimated reflectance map, we can make the association between surface orientation and appearance. This allows surface manipulation and transfer of materials and illumination between objects or even scenes.

We propose two approaches to estimate reflectance maps: a direct (Sec. 3.2) and an indirect one (Sec. 3.3). Both have a general RGB image as input and a reflectance map as an output. The indirect method also produces a conjoint per-pixel normal map.

Both variants are trained from and evaluated on the SMASHINg dataset introduced in detail in Sec. 4.1. For now, we can assume the training data to consists of pairs of 2D RGB images (domain) and reflection maps (range) in the parametrization explained in Sec. 3.1. This section now explains the two alternative approaches in detail.

3.1. Reflectance Map Representation

A reflectance map $L(\omega) \in \mathcal{S}^+ \to \mathbb{R}^3$ [14] is a map from orientations ω in the positive half-sphere $\mathcal{S}^+$ to the RGB radiance value L leaving that surface to a distant viewer. It combines the effect of illumination and material. For the case of a mirror sphere it captures illumination [6] but is not limited to it. It also does not capture surface reflectance [5], which would be independent of illumination, but joins the two.

There are multiple ways to parameterize orientation ω. Here Horn [14] used positional gradients which are suitable for an analytic derivation but less attractive for computation as they are defined on the infinite real line. We instead parameterize the orientation simply by s, t the normalized surface normal's x and y components. Dropping the z coordinate is equivalent to drawing a sphere under orthographic projection with exactly this reflectance map as seen right in Fig. 2. Note, that orientations of surfaces in an image only cover the upper half-sphere, so we only need to parameterize a half-sphere, avoiding to deal with spherical functions, *e.g.* spherical harmonics [31], that reduce the maximal frequency, only allowing for diffuse materials.

3.2. Direct approach: End-to-end model for prediction of reflectance maps

In the direct approach, we learn a mapping between the object image and its reflectance map, following a convolutional-upconvolutional architecture.
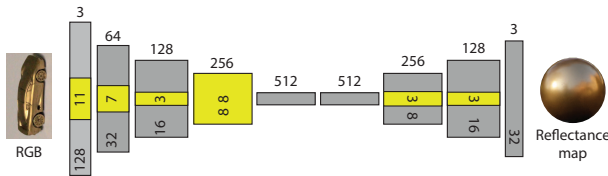

Figure 3. Architecture of the direct approach. The yellow boxes represent the filters' size.

The full architecture can be seen if figure Fig. 3. Starting from a series of convolutional layers, followed by batch normalization, ReLU and pooling layers, the size of the input feature maps is reduced to 1×1. After continuing with two fully connected layers, the feature maps are upsampled until the output size is 32×32 pixels. In all convolutional layers a stride of 1 is used and padded with zeros such that the output has the same size as the input. The final layer uses an euclidean loss between the RGB values for the predicted and the ground truth reflectance map.

In a typical CNN regression architecture, there is a spatial correspondence between input and output, *e.g.* in normal or depth estimation or semantic segmentation. In our case, the network needs to learn how to "encode" the input image so it can correspond to a specific reflectance map. This task is particularly challenging as the model has to learn not only how to place the image pixels to locations in the sphere (change from image to directional domain), but also to impute and interpolate appearance for unobserved normals.

3.3. Indirect approach: Reflectance maps from inferred normals and sparse interpolation

The indirect approach proceeds in four steps: *i)* estimating per-pixel orientation maps from the RGB image. *ii)*

upsampling the orientation map to the full available input image resolution. *iii)* changing from the image domain into the directional domain, producing a sparse reflectance map. *iv)* predicting a dense reflectance map from the sparse one.

The first and fourth step are model by CNN architectures, while the second and third step are prescribed transformations, related to the parametrization of the reflectance map. We will detail each step in the following paragraph.
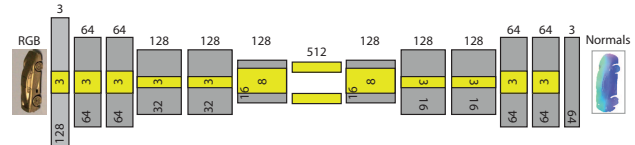

Figure 4. Architecture of the normal step of our indirect approach. The middle elements correspond to the fully convolutional filters.

Orientation estimation Our goal in the first step is to predict a surface orientation map from the RGB image. Thanks to our parametrization of the directional domain to coordinates in a flat 2D image of a lit sphere, the task is slightly simpler than finding full orientation. We seek to find the s and t parameters according to our reflectance map parameterization.

We train a CNN to learn the s, t coordinates. The architecture of the network is shown in Fig. 4. The network is fully convolutional as in [25] and it consist of a series of convolutional layers followed by ReLU and pooling layers that reduce the spatial extend of the feature maps. After the fully convolutional layers, there is a series of upconvolutional layers that upscale the feature representation to half the original size. Finally, we use two euclidean losses between the prediction and the L_2 normalized ground truth normals. The first one takes into account the x, y, z coordinates of the normals, while the second only the x, y.

Orientation upsampling The orientations are estimated at a resolution of $n = 128 \times 128$, so the number of appearance samples is in the order of ten-thousands. Most input images however are of much higher resolution with millions of pixels. A full-resolution orientation map is useful for resolving all appearance details in the orientation domain. The appearance of one orientation in the reflectance map can be related to all high-resolution image pixels (millions). Also intended applications performing shape manipulation in the 2D image (cf. Sec. 5.2) will benefit from a refined map. To produce a high-resolution orientation map, we used joint bilateral upsampling [17] as also done in range image [4]. Once we have an estimation of the object's normals, they can be mapped to a sphere and associated with appearance.

Change-of-domain We now reconstruct a sparse reflectance map from the orientation map and the input image.

This is a prescribed mapping transformation: The pairs of appearance L_i and orientation ω_i in every pixel are unstructured samples of the continuous reflectance map function $L(\omega)$ we seek to recover. Our goal now is to map these samples from the image to the directional domain, constituting the reflectance map. The most straightforward solution is to perform scattered data interpolation, such as

$$L(\omega) = (\sum_{i=1}^{n} w(\langle\omega, \omega_i\rangle))^{-1} \sum_{i=1}^{n} w(\langle\omega, \omega_i\rangle)L_i, \quad (1)$$

where $w(x) = \exp(-(\sigma \cos^{-1}(x))^2)$ is an RBF kernel.

In practice however, the orientation estimates are noisy and the requirements of a global reflectance map (infinite illumination, orthographic view, no shadows) are never fully met, asking for a more robust estimate. We found darkening due to shadows to be the largest issue in practice. Therefore, we perform a $\max$ operation over all samples closer than a threshold $\epsilon = \cos(5°)$ instead of an average, as in

$$L(\omega) = \max\{w(\langle\omega, \omega_i\rangle)L_i\}, \quad w(x) = \begin{cases} 1 & \text{if } x > \epsilon \\ 0 & \text{otherwise.} \end{cases}$$

If one orientation is observed under different amounts of shadow, only the one that is not in shadow will contribute – which is the intended effect. Still, the map resulting from this step is sparse due to normals that were not observed in the image as seen in Fig. 5 (left). This requires imputing and interpolating the sparse data in order to arrive at a dense estimate.

SparseNet: (Sparse-to-dense) Learning-based approach to sparse data interpolation The result of the previous step is a sparse reflectance map. It is noisy due to errors from incorrect normal estimation and has missing information at orientations that were not observed in the image. Note, that the latter is not a limitation of the normal estimation, but even occurs for ground truth surface orientations: If an orientation is not present, its appearance remains unknown.

A simple solution is to use Eq. 1 which already provides a dense output. We propose a learning-based approach to predict a dense reflectance maps from a sparse and noisy one. Accordingly, the network is trained on pairs of sparse and dense reflectance maps. The sparse ones are created using the first three steps explained (orientation from CNN, upsampling, Change-of-domain) on synthetic data where the target reflectance map is known by rendering a sphere.

The employed CNN architecture is shown in Fig. 5. Input is the sparse reflectance map and output the dense one. We use the output of the convolutional layers as additional cue. After each upconvolution layer, we concatenate its output with the feature map from the respective convolution layer. Again an L_2 loss between the output and the dense reference reflectance map is used.
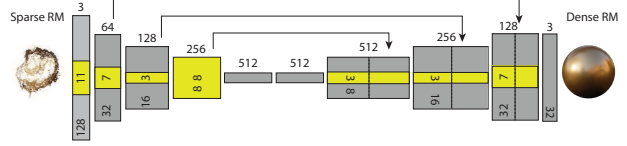

Figure 5. Architecture of the reflectance map step of our indirect approach.

4. The SMASHINg Challenge

We propose the Specular MAterials on SHapes with complex IllumiNation (SMASHINg) challenge. It includes a dataset (Sec. 4.1) of real as well as synthetic images, groundtruth reflectance maps and normals (where available), results from different methods to reconstruct and a set of metrics (Sec. 4.2) that we propose to evaluate and compare performance. At the time of publication we will make the data, baselines, our methods as well as the performance metrics publicly available.

4.1. Dataset

Our dataset combines synthetic images (Sec. 4.1.1), photographs (Sec. 4.1.2) and images from the web (Sec. 4.1.3) of cars. All images are segmented into foreground and background.

4.1.1 Synthetic images

Synthetic images are produced with random *i)* views, *ii)* 3D shapes, *iii)* materials, *iv)* illumination and *v)* exposure (Fig. 6) . The view is sampled from a random position around the object, looking at the center of the object with a FOV of $40°$. The 140 3D shapes come from the free 3D Warehouse repository, indexed by Shapenet [22]. For each sample the object orientation around the y axis is randomized. Illumination is provided by 40 free HDR environment maps. The exposure is sampled over the "key" parameter of Reinhard et al.'s photographic tone mapper [29] between 0.4 and 0.6. For materials, the MERL BRDF database [26] containing 100 materials is used. Overall 60 k sample images from that space are generated. We define a training-test split so that no shape, material or illumination is shared between the training and test set.

Figure 6. Our dataset comprises synthetic images with random view, 3D shape, material, illumination and exposure.

4.1.2 Photographs

As real test images, we have recorded photos of six toy cars that were completely painted with a single car lacquer, placed in four different lighting conditions and photographed from five different views, resulting a total of 120 images. For the corresponding ground truth reflectance maps, we placed in the same locations spheres painted with the same material. A qualitative evaluation, including examples of such real images is found in Sec. 5.2.

4.1.3 Internet Images

In order to provide an even more challenging test set, we collect an additional 32 car images from Internet search. Here we do not have access to groundtruth normals or reflectance maps, but the test provides a realistic test case for imaged-based editing methods. Again, we have manually segmented out the body of the car. This allows the study of single material normal and reflectance map prediction[1].

4.2. Methods and Metrics
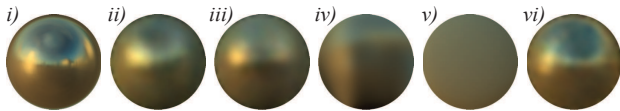
We include six different methods to reconstruct reflectance maps: *i)* ground truth, *ii)* our direct , *iii)* our indirect approach, *iv)* an approach that follows our indirect one, but does not use a CNN for sparse interpolation but an RBF reconstruction as described in Eq. 1 (RBF), *v)* spherical harmonics (SH) where project the ground truth reflectance map to the SH domain, *vi)* an indirect approach where the estimated normals are replaced by ground truth normals.

We employ two different metrics to assess the quality of reflectance map estimation. The first is plain L_2 error between all defined pixels of the reflectance map in RGB and the second the SSIM structural difference [36].

5. Experiments

We evaluate our proposed end-to-end direct approach to reflectance maps on the new SMASHINg Challenge and compare it to the indirect approach in its different variants. We start with a quantitative evaluation (Sec. 5.1) followed by qualitative results in (Sec. 5.2) including a range of image-based editing tasks.

5.1. Quantitative Results

Setup. Our quantitative results are summarized in Tbl. 1. We provide results for our *Direct* method that learns to predict reflectance maps directly from the image in an end-to-end scheme, as well as several variants of our *Indirect*

---

[1]For the Internet Images we used networks that were trained on synthetic data from segmented meshes to contain only the body.

approach that utilizes intermediate result facilitated by supervision through normals at training time. The variants of the indirect scheme are based on our normal estimate, but differ in their second stage that has to perform a type of data interpolation to arrive at a dense reflectance map, given the sparse estimate. For such interpolation scheme, we investigate the proposed learning-based approach *Indirect (SparseNet)* as well as using radial basis function interpolation *Indirect (RBF)*. Furthermore, we provide best case analysis by using ground-truth normals in the indirect approach *Indirect (GT Normals)* (only possible for synthetic data) and computing a diffuse version of the ground-truth by means of spherical harmonics *GT (SH)*. The latter gives an upper bound on the result that could be achieved by methods relying on a diffuse material assumption.

Table 1. Results for the different methods defined in Sec. 4.2.

Method	Synthetic		Real	
	MSE	DSSIM	MSE	DSSIM
Direct	.0019	.0209	.0120	.0976
Indirect (SparseNet)	.0018	.0180	.0143	.0991
Indirect (RBF)	.0038	.0250	.0116	.0814
Indirect (GT Normals)	.0008	.0111	—	—
GT (SH)	.0044	.0301	.0114	.0914

Reflectance Map Analysis. Overall, we observe consistency among the two investigated metrics in how they rank approaches. We obtain accurate estimations for the synthetic challenge set for our *direct* as well as the best *indirect* methods. The quantitative findings are underpinned by the visual results, e.g. showing the predicted reflectance maps in Fig. 7. The performance on the real images is generally lower with the error roughly increasing by one order of magnitude. Yet, the reconstruction still preserve rich specular structures and give a truthful reconstruction of the represented material.

In more detail, we observe that the best *direct* and *indirect* approach perform similar on the synthetic data, although *direct* did not use the normal information during training. For the real examples, this form of additional supervision seems to pay off more and even the simpler interpolations scheme RBF achieves best results in the considered metrics. Closer inspection of the results clearly shows limitations of image-based metrics. While the RBF-based technique yields a low error, it frequently fails to generate well localized highlight features on the reflectance map (see also illustration in Sec. 4.2). We encourage the reader to visit the supplementary material, where a detailed visual comparison for all methods is provided.

The ground-truth baselines give further insights into improvements over prior diffuse material assumptions and the future potential of the method. The *GT (SH)* baseline shows that our best methods improve over a best-case diffuse estimate with a large margin on in the DSSIM metric – highlight-

ing the importance of considering more general reflectance maps. The error metric is again affected by the aforementioned issues. The *Indirect (GT Normals)* illustrates a best case analysis of the indirect approach where we provide ground-truth normals. The results show that there is potential to double the performance by having better normal estimation in the first stage.

Table 2. Normals estimation of indirect approach on synthetic data.

	Mean	Median	RMSE
L_2	14.3	9.1	20.6
Dual	13.4	8.2	19.8
Dual up	13.3	8.2	19.9

Normal Analysis. Tbl. 2 quantifies the error in the normal estimate by the first stage of our indirect approach. This experiment is facilitated by the synthetic data where normals are available by the rendering pipeline. L_2 corresponds to a network using the euclidean loss on the x, y, z components of the normals, while dual uses the two losses described in Sec. 3.3. Up refers to a network trained on upsampled normals. Both, the dual loss and joint upsampling improve the estimation of normals. Despite providing more data to the down-stream computation, the employed upsampling procedure does not decrease – but rather slightly increase the accuracy of the normals. While this analysis is conducted on synthetic data, we found that our models predict very convincing normal estimation even in the most challenging scenario that we consider, *e.g.* Fig. 7 and Fig. 9.

5.2. Qualitative Results

Automatically extracting reflectance maps – together with the normal information we get as a by-product – facilitate a range of image-based editing applications, such as material acquisition, material transfer and shape manipulation. In the following, we present several example applications. The supplementary material contains images and videos that complement our following presentation.

Reflectance Map and Normal Estimation. Typical results of estimated reflectance maps are presented in Fig. 7, also showing the quality of the predicted normals. The first row shows two examples on synthetic images, the second and third row show examples on real images and the last row shows examples of web images (no reference reflectance map is available here). Notice how the overall appearance, reflecting the interplay between material and the complex illumination, is captured by our estimates. In most examples, highlights are reproduced and even a schematic structure of the environment can be seen in the case of very specular materials.

Material Acquisition for Virtual Objects. Fig. 8 shows synthesize image (column 2-5) that we have rendered from 3D models using the reflectance map automatically acquired from the images in column 1. Here, we use ambient occlusion [39] to produce virtual shadows. This application shows how material representations can be acquired from real objects and transferred to a virtual object. Notice how the virtual objects match in material, specularity and illumination to the source image on the left.

Appearance Transfer. In order to transfer appearance between objects of a scene, we estimate reflectance maps for each object independently, swapped the maps, and then use the estimated normals to re-render the objects using a normal lookup from the new map. To preserve details such as shadows and textures, we first re-synthesize each object with its original reflectance map, save the per-pixel difference in LAB color space, re-synthesize with the swapped reflectance map and add back the difference in LAB. An example is shown in Fig. 9. Despite the uncontrolled conditions, we achieve photorealistic transfer of the appearance – making it hard to distinguish source from target.

Shape Manipulation. As we estimate reflectance maps and surface normals, this enables various manipulation and re-synthesis approaches that work in the directional or normal domain. Here, the surface orientation is changed, *e.g.* using a painting interface and new appearance for the new orientation can be sampled from the reflectance map. Again, we save and restore the delta of the original reflectance map value and the re-synthesized one to keep details and shadows. An example is shown in Fig. 10. The final result gives a strong sense of 3D structure while maintaining an overall consistent appearance w.r.t. material and scene illumination.

6. Discussion

While our approach addresses a more general setting than previous methods, we still make certain assumptions and observe limitations: *i)* Up to now, all our experiments were conducted on cars and we assume that the car is segmented or in uniform background. Yet, our approach is learning-based and should – in principle – adapt to other classes in particular given dedicated training data. *ii)* We assume that the object is made out of a single material and up to now we cannot handle multiple parts or textures. *iii)* We assume distant illumination and therefore light interaction of close by objects or support surfaces (*e.g.* road) cannot be accurately handled by our model. *iv)* Due to our target representation of reflectance maps, the illumination is "baked in" and surface reflectance and illumination cannot be edited separately. *v)* Our quantitative evaluations are limited due to the absence of reliable (*e.g.* perceptual) metrics of reflectance maps.
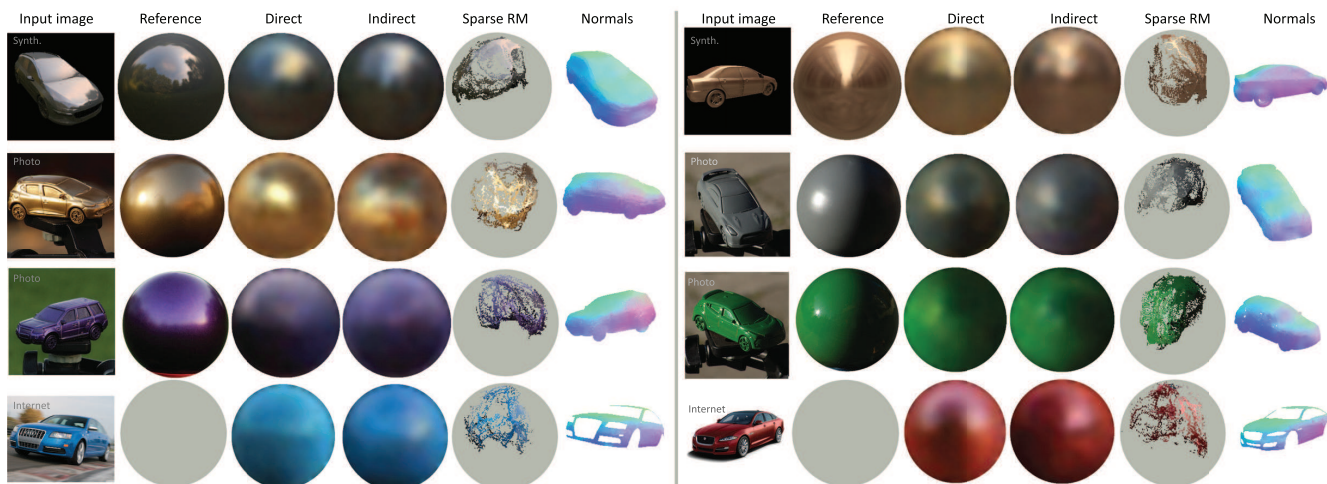
Figure 7. Results of different variants and steps of our approach *(left to right)*. Input image, GT RM, RM result of the direct approach, RM result of the indirect approach, the sparse RM input produced in the indirect variant, and the normals produced by the indirect variant as well. Each result is annotated to come from the synthetic, photographed or Internet part of our database. For the Internet-based part, no reference RM is available. Please see the supplemental material for exhaustive results in this form.
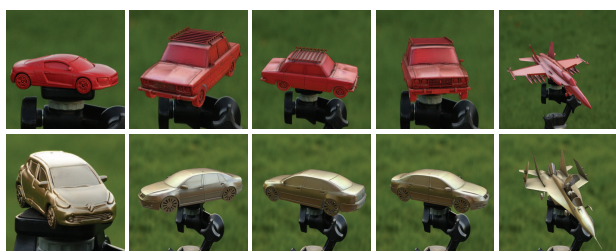

Figure 8. Transfer of reflectance maps from real photographs *(1st col.)* to virtual objects *(other col.'s)* of the same and other shape. The supplemental video shows animations of those figures.

Figure 9. Appearance transfer application: Images on the diagonal are the original input. Off-diagonal images have the appearance of the input in its column combined with the input shape of its row.

7. Conclusion

We have presented an approach to estimate reflectance maps from images of complex shapes with specular materials under complex natural illumination. Our study is facilitated by a new benchmark of synthetic, real and web images of in-

Figure 10. Shape manipulation application. A user has drawn to manipulate the normal map extracted from our indirect approach. The reflectance map and the new normal map can be used to simulate the new shape's appearance.

creasing difficulty, that we will make available to the public. Our approach features the first mapping using end-to-end learning from image to the directional domain as well as an application of neural networks to learning-based sparse data interpolation. We show how to incorporate additional supervision by normal information that increase accuracy as well as results in normal estimations as a byproduct. Our results show truthful reflectance maps in all three investigated scenarios and we demonstrate the applicability on several image-based editing tasks.

References

[1] A.Dosovitskiy, J.T.Springenberg, and T.Brox. Learning to generate chairs with convolutional neural networks. In *CVPR*, 2015. 3

[2] J. T. Barron and J. Malik. Shape, illumination, and reflectance from shading. *PAMI*, 2015. 2

[3] H. G. Barrow and J. M. Tenenbaum. Recovering intrinsic scene characteristics from images. *Comp. Vis. Sys.*, 1978. 2

[4] D. Chan, H. Buisman, C. Theobalt, and S. Thrun. A noise-aware filter for real-time depth upsampling. In *Proc. M2SFA*, 2008. 4

[5] K. J. Dana, B. Van Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and texture of real-world surfaces. *ACM Trans, Graph.*, 18(1), 1999. 1, 2, 3

[6] P. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. *SIGGRAPH*, 1998. 1, 2, 3

[7] R. O. Dror, T. K. Leung, E. H. Adelson, and A. S. Willsky. Statistics of real-world illumination. In *CVPR*, 2001. 2

[8] D. Eigen and R. Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *ICCV*, 2015. 3

[9] D. Eigen, C. Puhrsch, and R. Fergus. Depth map prediction from a single image using a multi-scale deep network. In *NIPS*, 2014. 3

[10] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, 2014. 2

[11] T. Haber, C. Fuchs, P. Bekaer, H.-P. Seidel, M. Goesele, H. P. Lensch, et al. Relighting objects from image collections. In *CVPR*, 2009. 2

[12] B. Hariharana, P. Arbelaez, R. Girshick, and J. Malik. Hypercolumns for object segmentation and fine-grained localization. In *CVPR*, 2015. 3

[13] A. Hertzmann and S. M. Seitz. Example-based photometric stereo: Shape reconstruction with general, varying BRDFs. *PAMI*, 27(8), 2005. 2

[14] B. K. Horn and R. W. Sjoberg. Calculating the reflectance map. *App. Opt.*, 18(11), 1979. 1, 2, 3, 4

[15] M. K. Johnson and E. H. Adelson. Shape estimation in natural illumination. In *CVPR*, 2011. 2

[16] E. A. Khan, E. Reinhard, R. W. Fleming, and H. H. Bülthoff. Image-based material editing. *ACM Trans. Graph.*, 25(3), 2006. 2

[17] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele. Joint bilateral upsampling. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 26(3), 2007. 4

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*. 2012. 2

[19] T. D. Kulkarni, P. Kohli, J. Tenenbaum, and V. Mansinghka. Deep convolutional inverse graphics network. In *NIPS*, 2014. 3

[20] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proc. ICML*, 2009. 2

[21] B. Li, C. Shen, Y. Dai, A. van den Hengel, and M. He. Depth and surface normal estimation from monocular images using regression on deep features and hierarchical CRFs. In *CVPR*, 2015. 3

[22] Y. Li, H. Su, C. R. Qi, N. Fish, D. Cohen-Or, and L. J. Guibas. Joint embeddings of shapes and images via cnn image purification. *ACM Trans. Graph.*, 2015. 5

[23] F. Liu, C. Shen, and G. Lin. Deep convolutional neural fields for depth estimation from a single image. In *CVPR*, 2015. 3

[24] S. Lombardi and K. Nishino. Reflectance and illumination recovery in the wild. *PAMI*, 2015. 2

[25] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015. 3, 4

[26] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Trans. Graph.*, 2003. 5

[27] T. Narihira, M. Maire, and S. X. Yu. Direct intrinsics: Learning albedo-shading decomposition by convolutional regression. In *ICCV*, 2015. 3

[28] T. Narihira, M. Maire, and S. X. Yu. Learning lightness from human judgement on relative reflectance. In *CVPR*, 2015. 3

[29] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. Photographic tone reproduction for digital images. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 21(3), 2002. 5

[30] K. Rematas, T. Ritschel, M. Fritz, and T. Tuytelaars. Image-based synthesis and re-synthesis of viewpoints guided by 3d models. In *CVPR*, 2014. 2

[31] S. Richter and S.Roth. Discriminative shape from shading in uncalibrated illumination. In *CVPR*, 2015. 2, 4

[32] Right Hemisphere. ZBruhs MatCap, 2015. 2

[33] P.-P. J. Sloan, W. Martin, A. Gooch, and B. Gooch. The lit sphere: A model for capturing NPR shading from art. In *Graphics interface*, 2001. 2, 3

[34] Y. Tang, R. Salakhutdinov, and G. Hinton. Deep lambertian networks. *arXiv preprint arXiv:1206.6445*, 2012. 3

[35] X. Wang, D. F. Fouhey, and A. Gupta. Designing deep networks for surface normal estimation. In *CVPR*, 2015. 3

[36] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *Proc. Sign., Sys. and Comp.*, volume 2, 2003. 6

[37] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus. Deconvolutional networks. In *CVPR*, 2010. 2

[38] T. Zhou, P. Krähenbühl, and A. Efros. Learning data-driven reflectance priors for intrinsic image decomposition. In *ICCV*, 2015. 3

[39] S. Zhukov, A. Iones, and G. Kronin. An ambient light illumination model. In *Proc. EGSR*. 1998. 7