Computers & Graphics 127 (2025) 104186

Contents lists available at ScienceDirect

Computers & Graphics

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

Technical Section

Single-image reflectance and transmittance estimation from any flatbed scanner☆

Carlos Rodriguez-Pardo [a,b,c,*], David Pascual-Hernandez [d], Javier Rodriguez-Vazquez [e], Jorge Lopez-Moreno [d], Elena Garces [f]

[a] Politecnico di Milano, Italy
[b] CMCC Foundation - Euro-Mediterranean Center on Climate Change, Italy
[c] RFF-CMCC European Institute on Economics and the Environment, Italy
[d] Universidad Rey Juan Carlos, Spain
[e] Arquimea Research Center, Spain
[f] Adobe Research, United States of America

ARTICLE INFO

Keywords:
Material capture
Reflectance
Transmittance
Generative models
SVBSDF

ABSTRACT

Flatbed scanners have emerged as promising devices for high-resolution, single-image material capture. However, existing approaches assume very specific conditions, such as uniform diffuse illumination, which are only available in certain high-end devices, hindering their scalability and cost. In contrast, in this work, we introduce a method inspired by intrinsic image decomposition, which accurately removes both shading and specularity, effectively allowing captures with any flatbed scanner. Further, we extend previous work on single-image material reflectance capture with the estimation of opacity and transmittance, critical components of full material appearance (SVBSDF), improving the results for any material captured with a flatbed scanner, at a very high resolution and accuracy.

1. Introduction

Several industries, such as architectural and fashion design, or media and gaming, benefit from realistic digital replicas of physical materials. Yet, crafting these copies remains a laborious and slow task, demanding skilled artists, or sophisticated and expensive hardware [1,2]. Consequently, recent research has focused on devising affordable and user-friendly capture setups.

In this scenario, flatbed scanners have emerged as promising tools for high-resolution material capture [3], owing to their user-friendly nature and provision of uniform illumination conditions. High-end scanners can even offer a lighting type closely resembling diffuse illumination, usable directly as an albedo image [3]. Nevertheless, most scanners lack this functionality, with a majority featuring a single directional light that leads to undesirable micro-specular reflections, directional shading, and cast shadows (depicted in Fig. 1(a) and 2).

In this work, we address the drawbacks of prior approaches and introduce a technique for digitizing materials using any scanner, removing undesirable shading and specular highlights. We show that the naïve solution employing an image-to-image translation network [3,4] falls short for this purpose. Instead, we suggest employing a cycle-consistency loss in combination with a residual formulation inspired by intrinsic image decomposition methods [5].

In addition, a key contribution of our method is to expand the realism of the digital replica by including opacity and transmittance in the material model. These attributes are critical for thin-layer materials, like textiles, but have been neglected in current literature. We estimate the parameters of a Spatially-Varying Bidirectional Scattering Distribution Function (SVBSDF) that can reproduce complex effects of light as it passes through the material, thereby augmenting its realism in virtual environments.

We evaluate our method using extensive and thorough experiments, leveraging image-based metrics that measure the precision of each map individually, and render-aware metrics that measure the final appearance of the material in a global context. We further demonstrate that our method works with a variety of scanning devices, producing effective results even with less controllable devices such as smartphones.

☆ This article was recommended for publication by J. Dorsey.

* Corresponding author at: Politecnico di Milano, Italy.
E-mail address: carlos.rodriguezpardo.jimenez@gmail.com (C. Rodriguez-Pardo).

https://doi.org/10.1016/j.cag.2025.104186
Received 9 July 2024; Received in revised form 4 February 2025; Accepted 19 February 2025
Available online 1 March 2025

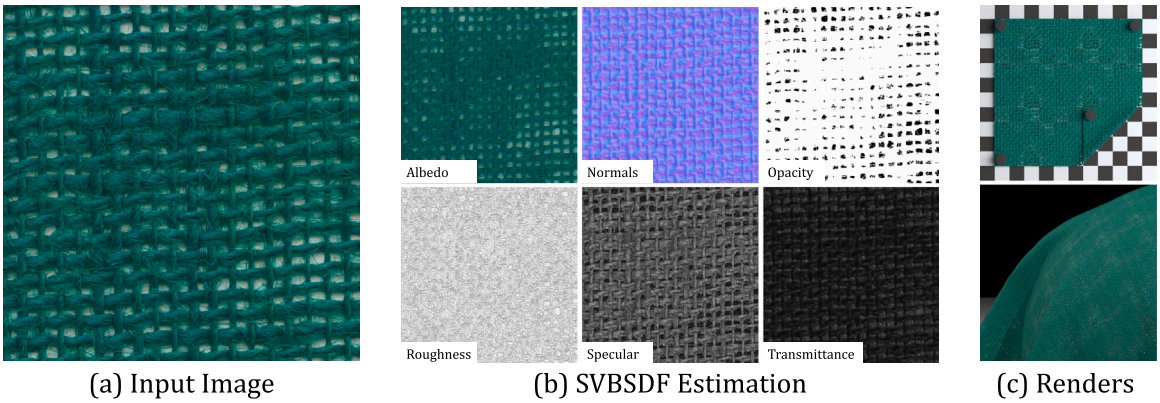

Fig. 1. From a single image captured with any flatbed scanner (a), our method estimates a set of high-resolution SVBSDF maps (b), which can be used in any render engine (c).

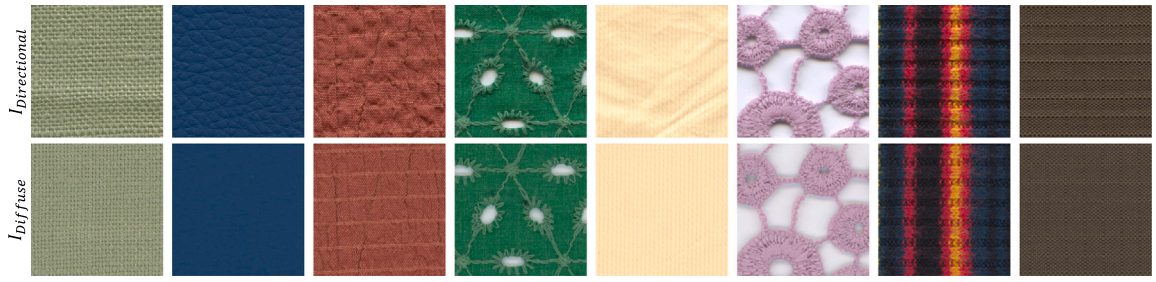

Fig. 2. Some materials in our test dataset, captured on the same flatbed scanner using directional and diffuse illuminations, better suited for material capture.

2. Related work

Single-image material capture. Estimating full reflectance properties of a material, using only a single image of it, is a challenging problem which has been tackled extensively in the literature in the recent years. These approaches can be categorized based on the estimation method, the imaging device employed, and the range of digitizable reflectance properties.

Neural style transfer [6] can be leveraged for single-image capture of stochastic materials, by matching the latent statistics of input images and renders of estimated SVBRDFs [7,8]. A more common approach is to train an image-to-image translation model which takes a single image as input and estimates the set of SVBRDF maps. Originally supervised using pixel-wise or render-aware losses [9–12], these methods have been improved by incorporating cascaded estimation [13, 14], adversarial losses [3,15–19], inference-time optimization [12], or refinement [20]. More recently, diffusion models have emerged as powerful material estimators, showing competitive results [21–24]. A complementary line of work uses procedural graphs for material estimation [25–28].

In terms of devices, the most common setup encompasses fronto-planar flash-lit images captured with a smartphone. Other setups trade this simplicity for quality, such as LCD screens [29–31], or high-end flatbed scanners [3]. Single-image material estimation methods typically estimate a reduced number of SVBRDF parameters, with the exception of [21], which also estimates opacity.

Our approach differs from previous work in two ways. First, we provide a generic framework for material capture from *any* flatbed scanner, with arbitrary directional illumination, effectively removing the limitations in [3]. Furthermore, to the best of our knowledge, our method is the first single-image material capture which can estimate a full SVBSDF of a material, incorporating important effects like transmittance and opacity while preserving a high level of accuracy and resolution.

Material delighting. Removing shading and specular highlights from images has been explored extensively in the literature, with particular focus on removing strong shadows [32–35] and human relighting [36–39]. In the context of BRDF estimation, material delighting has been explored by combining convolutional neural networks with Poisson optimization [4]. Our method also leverages material delighting for albedo estimation, by incorporating ideas from intrinsic image decomposition [5].

Cycle-consistent generative models. Learning to map from two distinct image domains for image-to-image translation tasks can be tackled through cycle-consistent generative models [40]. By introducing the cycle-consistency loss, these models enable accurate and diverse mappings. These have shown impressive results on a wide variety of applications, including stenography [41], voice conversion [42], medical imaging [43,44], face generation [45], or improving diffusion models [46,47]. Inspired by these methods, we leverage cycle-consistency to train a model capable of both material delighting and relighting, which showcases high accuracy in both tasks under several metrics (see Fig. 3).

2.1. Preliminaries: Material model

Building upon previous work [3], we use a physically-based material model based on microfacets reflectance [48], into which we incorporate additional parameters to enable transmittance effects. Our material model aggregates a diffuse component (i.e. the material albedo) $\mathbf{A} \in \mathbb{R}^{3\times x\times y}$, with a grayscale, isotropic specular GGX [49] lobe $s_{l,v} \in \mathbb{R}^{x\times y}$, which depends on the surface normal $\mathbf{N}$, its specularity $\mathbf{S}$ and roughness $\mathbf{R}$. The shading model $f^{\text{BSDF}}_{l,v} \in \mathbb{R}^{4\times x\times y}$ for a particular light l and camera v has an additional transparency term which depends on the material binary opacity $\mathbf{O} \in \mathbb{Z}_2^{x\times y}$ and its transmittance $\mathbf{T} \in \mathbb{R}^{x\times y}$, as follows:

$$f^{\text{BSDF}}_{l,v}(\mathbf{A},\mathbf{N},\mathbf{S},\mathbf{R},\mathbf{O},\mathbf{T}) = \mathbf{O}\cdot(\underbrace{\frac{\mathbf{A}}{\pi} + s_{l,v}(\mathbf{N},\mathbf{S},\mathbf{R})}_{\text{reflectance } f^{\text{BRDF}}_{l,v}} + \underbrace{(\mathbf{T}\cdot\mathbf{A})}_{\text{transmittance}}) \quad (1)$$

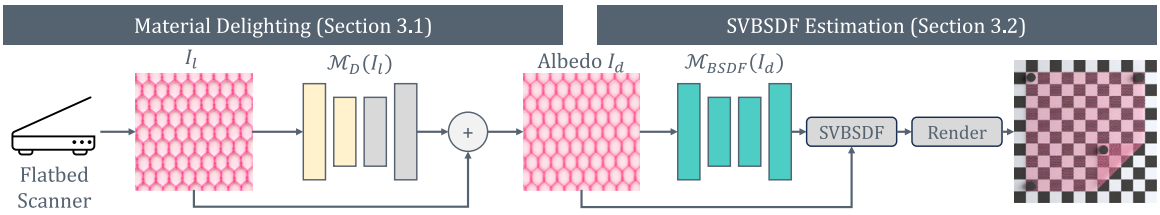

Fig. 3. From an image I_l captured with any flatbed scanner, we first estimate its albedo I_d using a residual generative model $\mathcal{M}_D$, which removes specular highlights and shading. Taking I_d as input, a second model $\mathcal{M}_{BSDF}$ estimates the rest of the SVBSDF, namely the surface normals, roughness, specular, transmittance, and opacity maps. These can be then rendered to generate photo-realistic images.

The transmittance is modeled as the base albedo **A** modulated by a gray scale value **T**. This assumes that the light scattered through the material is a linear attenuation of the reflectance wavelength (albedo). Finally, both reflectance and transmission are weighted by the binary operator **O**, which differentiates areas with partial and total transmission. Finally, both reflectance and transmission are weighted by the binary operator **O**, which differentiates areas with partial transmission from fully transparent pixels. The distinction between both **O** and **T**, being the former just a particular threshold on the continuous transmittance **T**, is due to its traditional use as a binary mask in several rendering methods, to reduce shader execution time by discarding pixels.

Although there are richer and more complex models for transmittance and sub-surface scattering phenomena (E.g.: [50]), we find that this thin-layer diffuse transmission model suffices to represent a large proportion of materials that can be captured with a scanner, while having low requirements for real time visualization and less memory consumption that a multi-channel transmission map.

3. Method

Our method takes as input a single image of the material and estimates its spatially-varying SVBSDF material parameters, including reflection and transmission per-pixel coefficients. The input image can be obtained with any capture device that provides mostly *uniform* lighting, such as the one provided by flatbed scanners. Our algorithm has two steps. In the first step, described in Section 3.1, we use a cycle-consistent residual generative network to *delight* the material and obtain an albedo-like reflectance map. After our processing, the resulting map lacks micro-reflections and shadows that might be originally present due to directional lighting hitting the material. In the second step, described in Section 3.2, we use this image as input of an attention-guided U-Net that estimate the remaining material maps, to convey reflection and transmission.

3.1. Material delighting

In this step, our goal is to estimate an albedo-like reflectance map $I_d \approx \mathcal{A}$ from a single image I_l of the material taken under any kind of uniform lighting. We term this process *delighting*, as we aim to remove specular reflections, shadings, or shadows. A straightforward solution to this problem would be to train an image-to-image translation approach with labeled data. However, as we demonstrate, this baseline approach does not achieve the desired level of accuracy due to the under-constrained nature of the problem and our relatively reduced training dataset (see Table 1). Therefore, we propose a more sophisticated architecture to improve this performance, which uses residual learning and a cycle-consistency loss. Inspired by intrinsic image decomposition [5], we formulate the *delighting* problem as estimating a residual layer $\mathcal{M}_R$ that adds to the albedo image to form a *lighted* image, $I_l = I_d + \mathcal{M}_R(I_d)$. Similarly, within our cycle-consistent architecture, the equivalent inverse operation also exists, and we term it *relighting*, $I_d = I_l + \mathcal{M}_D(I_l)$. Our residuals $\mathcal{M}_R(I_d)$ and $\mathcal{M}_D(I_l)$ are RGB images to make the estimation more flexible, thereby removing the assumption that either the source or reflected lights are white. Fig. 4 presents an overview of the architecture.

Loss function. Our loss for each branch of our cycle-consistency model is a combination of pixel-wise, perceptual, frequency, and adversarial losses,

$$\mathcal{L}_{\text{im}}(\cdot,\cdot) = \lambda_1 \mathcal{L}_1(\cdot,\cdot) + \lambda_{\mathcal{L}_{perc}} \mathcal{L}_{perc}(\cdot,\cdot) + \lambda_{\mathcal{L}_{freq}} \mathcal{L}_{freq}(\cdot,\cdot) + \lambda_{adv} \mathcal{L}_{adv}. \tag{2}$$

Following [1,3,51], for $\mathcal{L}_{perc}$ we use the AlexNet version of [52] and for $\mathcal{L}_{freq}$ we measure the Focal Frequency Loss [53]. For the adversarial loss, we follow the methodology specified in [40]. Then, we build our cycle-consistency loss and full loss as,

$$\mathcal{L}_{\text{cycle}}(I_d, I_l) = \underbrace{\mathcal{L}_{\text{im}}(I_d, \mathcal{M}_D(\mathcal{M}_R(I_d)))}_{\text{delighting}} + \underbrace{\mathcal{L}_{\text{im}}(I_l, \mathcal{M}_R(\mathcal{M}_D(I_l)))}_{\text{relighting}} \tag{3}$$

$$\mathcal{L}_{\text{full}}(I_d, I_l) = \mathcal{L}_{\text{im}}(I_d, \mathcal{M}_D(I_l)) + \mathcal{L}_{\text{im}}(I_l, \mathcal{M}_R(I_d)) + \lambda_{cycle} \mathcal{L}_{\text{cycle}}(I_d, I_l). \tag{4}$$

Architecture design. For the generator architectures, we follow the attention-guided U-Net design in [3], using a single decoder for each model, and removing the MLPs appended to the end of the architecture. For the discriminators, we follow previous work on texture synthesis [54,55] and use a 4-layer PatchGAN [56].

Data augmentation. We train the models using random cropping, with 128×128 resolution patches. Besides, we use random rescaling, enabling the model to generalize on the $(300, 1200)$ PPI range, covering most flatbed scanners. Finally, we use random horizontal and vertical flips to further enhance generalization.

Implementation details. We standardize each dataset of directional and diffuse images using their respective means and standard deviations, enabling the model to focus on relative differences and not on global average values. We use PyTorch [57] and Torchvision [58] for training, and Kornia [59] for data augmentation. We train the models using Adam [60] for 100 iterations, with an initial learning rate of 0.002, halved every 30 iterations. We leverage automatic gradient scaling and mixed precision training [61]. Both generators and discriminators are initialized using *orthogonal initialization* [62]. Training these models takes approximately 12 h on a NVidia 3060 GPU. After a Bayesian hyperparameter [63] optimization performed on a separate validation dataset, we set the loss weighting as $\lambda_{cycle} = 0.25, \lambda_{adv} = 0.15, \lambda_{perc} = 0.3, \lambda_{freq} = 0.2, \lambda_{\mathcal{L}_1} = 1$. Further details are included in the supplementary material.

3.2. SVBSDF estimation

To estimate the rest of the SVBSDF, we build upon the training methodology described in [3]. First, we expand their model to enable the estimation of opacity and transmittance maps, thus introducing two additional decoders to their attention-guided U-Net network, and expanding the loss function and discriminator architecture accordingly.

We further introduce additional minor changes to improve the estimation. Most notably, we parameterize the normal map so as to estimate θ, ϕ angles instead of full cartesian coordinates xyz, and following [1] adopt *elliptical grid mapping* [64] for additional performance gains. Finally, we use AdamW [65] and 256×256 crops for training, and perform minor hyperparameter changes, which are fully described in the supplementary material.

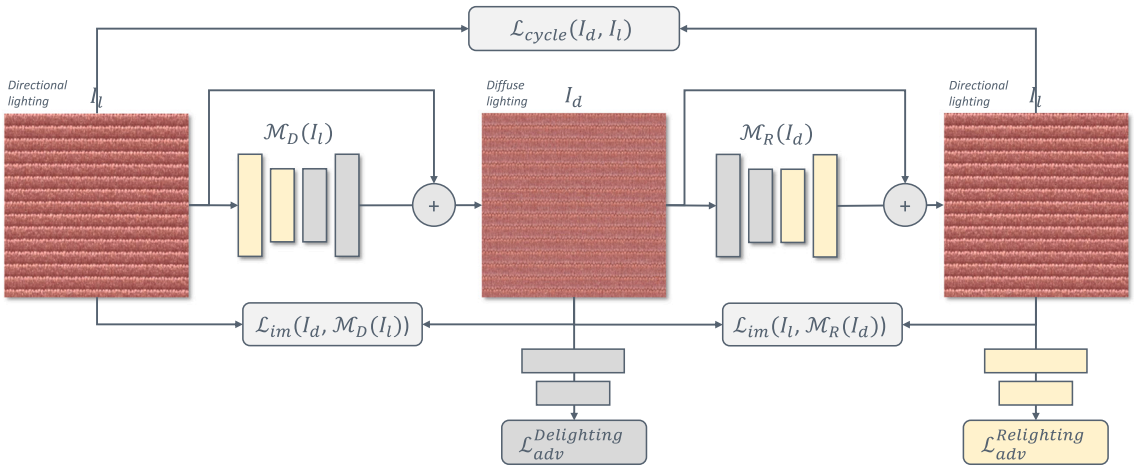

Fig. 4. Diagram of our cycle-consistent generative model capable of both material delighting and relighting.

4. Evaluation

4.1. Dataset

Using a high-end *EPSON V850 Pro* flatbed scanner, we capture 3830 10 × 10 cm material samples at 1200 PPI resolution. Note that this scanner can capture images using a standard, single LED strip illumination, like lower-end scanners, but also enables a higher-quality setup using a dual-light which provides diffuse-like illumination. A detailed description of the dataset is provided in the supplementary material. The later setup closely resembles fitted albedos [3], removing strong shades caused by wrinkles or mesostructure, hiding shadows casted to the scanner lid and eliminating specular highlights (as shown in Fig. 2). We thus capture two images for each material: I_l and I_d, preserving pixel-wise correspondence for every material under both illuminations. To augment this dataset, we further capture every material on their front and back sides, and rotate them by 90° to allow for generalization to multiple orientations. We also capture these materials on a custom gonioreflectometer, and leverage the methodology described in [1,66] to propagate the ground truth material parameters described in Eq. (1). We use 10% of this dataset for testing.

4.2. Metrics

To measure the performance of our models, we use a variety of metrics aimed at understanding the perceptual, pixel-wise, and render-aware accuracy of our estimations. First, to measure the errors of our generators, $\mathcal{M}_D$ and $\mathcal{M}_R$, we leverage traditional image quality metrics, as well as ΔE [67], which accurately measures color differences, and perceptually-motivated alternatives like FLIP [68] and LPIPS [52]. We also quantify per-map accuracy, leveraging pixel-wise $\mathcal{L}_1$ norms for the Albedo, Roughness, Specular, and Transmittance maps, angular distances $\mathcal{L}_\measuredangle$ for the surface normals, and the Jaccard index $\mathcal{L}_{Jacc}$ for the opacity maps. Following [3], we also report Pearson correlations ρ.

The previous metrics are useful to assess individual precision of the estimations. However, when reproducing a real material, it is of critical importance to understand how these parameters interact with each other in the integrated physically-based rendering space. Thus, we propose a set of metrics aimed to evaluate the accuracy of the full material model in terms of both reflectance and transmittance. For reflectance, we expand the $\mathcal{L}_{\text{BRDF}}$ metric in [3], which measures the perceptual error, with extra terms that account for material opacity, and cosine weighting and peak reflectance attenuation to account for human visual perception [69]. We measure the render-space reflectance estimation difference between the ground truth $\mathbf{M}_{GT}$ and predicted $\hat{\mathbf{M}}$ material as follows:

$$\mathcal{L}_{\text{BRDF}}(\mathbf{M}_{GT}, \hat{\mathbf{M}}) = \frac{1}{xy} \times \sum_{xy} \sqrt{\frac{1}{|S|} \sum_{(l,v)\in S} \sqrt[3]{\cos^2(\theta_l)\left(f_{l,v}^{\text{BRDF}}(\mathbf{A}_{GT}, \mathbf{N}_{GT}, \mathbf{S}_{GT}, \mathbf{R}_{GT}) \cdot O_{GT} - f_{l,v}^{\text{BRDF}}(\hat{\mathbf{A}}, \hat{\mathbf{N}}, \hat{\mathbf{S}}, \hat{\mathbf{R}}) \cdot \hat{O}\right)^2}} \tag{5}$$

where l, v are a set of 50 lights and viewing angles optimized for BRDF acquisition, gathered from [70].

For transmittance, we introduce a novel metric, $\mathcal{L}_{\text{BTDF}}$, which explicitly measures the error in the estimation of transmissive effects as follows:

$$\mathcal{L}_{\text{BTDF}}(\mathbf{M}_{GT}, \hat{\mathbf{M}}) = \frac{1}{xy} \sum_{xy} |\mathbf{T}_{GT} \cdot \mathbf{A}_{GT} \cdot \mathbf{O}_{GT} - \hat{\mathbf{T}} \cdot \hat{\mathbf{A}} \cdot \hat{\mathbf{O}}| \tag{6}$$

Finally, we define our final metric $\mathcal{L}_{\text{BSDF}}$ as a weighted combination of $\mathcal{L}_{\text{BRDF}}$ and $\mathcal{L}_{\text{BTDF}}$, setting $w_{\text{BRDF}} = \frac{1}{2}$ for simplicity:

$$\mathcal{L}_{\text{BSDF}} = w_{\text{BRDF}}\mathcal{L}_{\text{BRDF}} + (1 - w_{\text{BRDF}})\mathcal{L}_{\text{BTDF}} \tag{7}$$

This integrated metric is render-aware and perceptually validated and enables the comparison of different configurations of our models.

4.3. Ablation study

In this section, we present an ablation study to validate each of our components.

Delighting model

Table 1 presents the results of the study for our *Relighting* and *Delighting* flows. Our baseline is a pure regression-based model which uses only pixel-wise $\mathcal{L}_1$ losses. We progressively add components to this baseline, to study their impact. First, making the models generative by introducing $\mathcal{L}_{adv}$ to their training losses enables higher accuracy. Training $\mathcal{M}_D$ and $\mathcal{M}_R$ together, using our cycle-consistency loss $\mathcal{L}_{cycle}$, strongly improves accuracy across every metric, providing evidence that this is a key component for achieving high-quality albedo estimations. Further, using our residual learning approach, inspired by intrinsic decomposition, provides significant gains across every metric. Finally, incremental improvements are achieved by introducing $\mathcal{L}_{perc}$ and $\mathcal{L}_{freq}$, and our full data augmentation policy. Interestingly, we observe that relighting is typically a harder task. We believe that our cycle-consistent approach allows our model to generalize better because it works as a form of data augmentation, while residual learning improves training dynamics and makes the task easier to learn.

Fig. 5. Qualitative results of our material delighting framework. On the first two rows, we show images captured with flatbed scanners under diffuse (top) and directional (bottom) illumination. We use those as input to our delighting $\mathcal{M}_D$ and relighting $\mathcal{M}_R$ models, respectively, for which we show the results on the bottom rows.

Table 1
Results of our ablation study of our material delighting algorithm, across a variety of metrics. On the top row, we show the results when no delighting is applied. We use a color code to highlight best and cases.

	Relighting					Delighting				
	PSNR ↑	SSIM [71] ↑	LPIPS [52] ↓	ΔE ↓	ꟻLIP [68] ↓	PSNR ↑	SSIM [71] ↑	LPIPS [52] ↓	ΔE ↓	ꟻLIP [68] ↓
No Delighting	24.01	0.686	0.262	4.732	0.264	24.01	0.686	0.262	4.732	0.264
Baseline Delighting										
+ $\mathcal{L}_{adv}$	24.69	0.809	0.257	5.587	0.244	26.72	0.810	0.242	4.359	0.202
+ Cycle-Consistency	26.54	0.856	0.218	4.213	0.194	27.82	0.851	0.202	3.604	0.169
+ Residual	28.42	0.903	0.165	3.115	0.147	29.79	0.897	0.171	2.754	0.132
+ Full Loss	28.67	0.912	0.164	3.071	0.144	30.19	0.906	0.151	2.630	0.126
+ Aug. (Final Model)	28.48	0.907	0.161	3.012	0.138	31.41	0.933	0.136	2.261	0.111

Table 2
Results of previous work, and of our ablation study, on final digitization accuracy, on per-map and integrated metrics. We use a color code to highlight best and cases. Errors marked with * correspond to input images which we assume to be the ground truth albedos, hence $\mathcal{L}_1^A = 0$. †Note that [3] does not estimate transmittance nor opacity, instead we assume the materials are fully opaque.

	Pixel-Wise Errors						Correlations			Render-Aware		
	$\mathcal{L}_1^A$ ↓	$\mathcal{L}_\measuredangle$ ↓	$\mathcal{L}_1^R$ ↓	$\mathcal{L}_1^S$ ↓	$\mathcal{L}_1^T$ ↓	$\mathcal{L}_{Jac}^O$ ↑	ρ^R ↑	ρ^S ↑	ρ^T ↑	$\mathcal{L}_{BRDF}$ ↓	$\mathcal{L}_{BTDF}$ ↓	$\mathcal{L}_{BSDF}$ ↓
UMat [3], Diffuse Illumination †	0.000*	2.666	0.060	0.086	0.073	0.931	0.714	0.852	0.000	0.324	0.058	0.191
UMat [3], Directional Illumination †	0.051	2.813	0.062	0.089	0.073	0.931	0.701	0.841	0.000	0.384	0.089	0.237
Ours, w/o Delighting												
Diffuse Illumination	0.000*	2.221	0.055	0.081	0.017	0.998	0.731	0.899	0.937	0.223	0.026	0.125
Directional Illumination	0.051	2.771	0.061	0.086	0.025	0.958	0.711	0.877	0.897	0.344	0.059	0.202
Ours w/ Baseline Delighting									0.873			
+ $\mathcal{L}_{adv}$	0.037	2.981	0.062	0.086	0.034	0.952	0.701	0.877		0.345	0.067	0.206
+ Cycle-Consistency	0.032	2.692	0.058	0.084	0.028	0.981	0.713	0.889	0.895	0.303	0.051	0.177
+ Residual	0.026	2.474	0.057	0.086	0.021	0.992	0.722	0.881	0.921	0.276	0.038	0.157
+ Full Loss	0.024	2.441	0.057	0.081	0.023	0.991	0.721	0.898	0.919	0.261	0.035	0.148
+ Aug. (Final Model)	0.021	2.333	0.057	0.080	0.019	0.994	0.722	0.903	0.930	0.253	0.030	0.142

SVBSDF estimation accuracy

Table 2 presents the errors of our end-to-end digitization pipeline, including the error of not using the delighting model (*Ours w/o Delighting*) and comparison with related work UMat [3], the main available method in previous work that uses scanners as a capture device. Note that, to date, there is no previous work that estimates transmittance images. We also test different inputs using images captured under Diffuse *albedo-like* Illumination (therefore delighting operation would be necessary), and images with Directional Illumination.

As shown, our model behaves consistently better across every metric than UMat. Note that UMat does not estimate opacity nor transmittance (we set all their materials to be full opaque), which is heavily penalized by the integrated metric $\mathcal{L}_{\text{BSDF}}$. Interestingly, the relative improvements of our model are more visible on the normal map than on the roughness or specular maps, likely due to our proposed normal reparameterization. Minor training improvements like the surface normal reparameterization, some hyperparameter changes, and a larger dataset helped push accuracy further.

We also expand our previous ablation using render metrics. Notably, we observe that more accurate delighting does not only result in

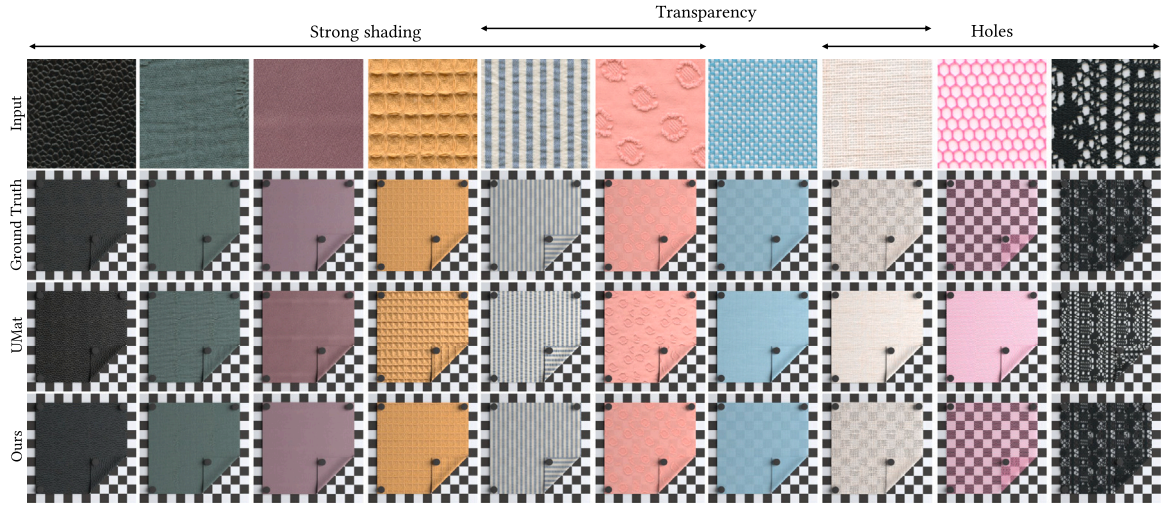

Fig. 6. Qualitative comparisons of our method with UMat [3] for a few representative materials in our test set, with strong shading (leftmost columns), transparency (middle) or holes (rightmost). We show the input image (top row), and renders using the ground truth materials (captured with a gonioreflectometer), the estimation of [3] and ours, on the second, third and fourth rows, respectively. Best viewed in color on a screen.

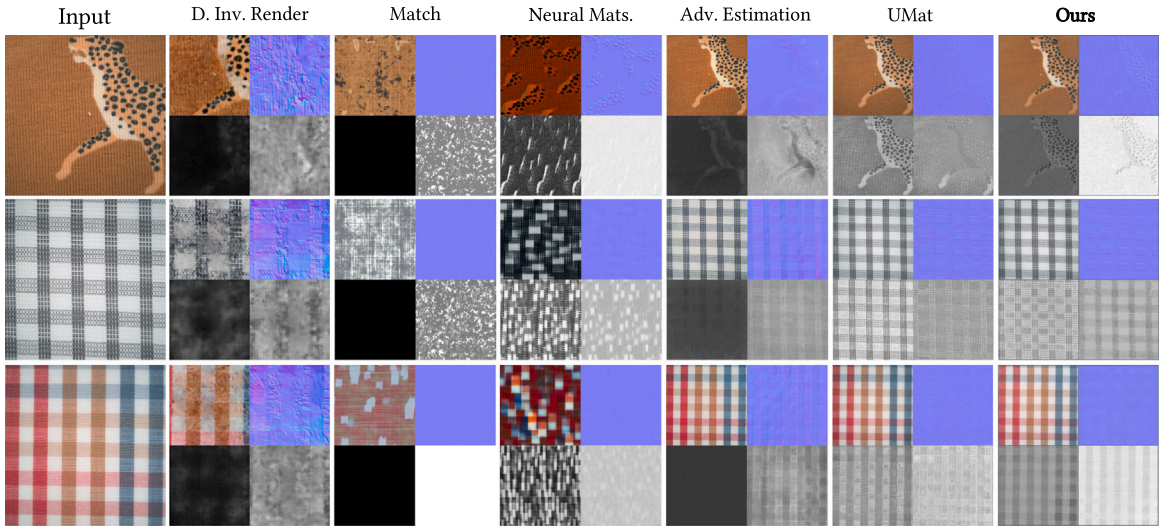

Fig. 7. Comparisons of our method with previous work on images captured with a smartphone, using ambient lighting (top row) and flash illumination (bottom two rows), at different levels of resolution. From left to right, we show input images, and the results of Deep Inverse Rendering [12], Match [25], Neural Materials [7], Adversarial SVBRDF Estimation [17], UMat [3], and ours. Note that we only show the four reflectance maps used by every method: albedo, normals, specular and roughness.

better albedo estimation, but the estimation of the remainder of the SVBSDF also becomes more precise. This is particularly visible on the surface normals, transmittance and opacity maps, while roughness or specularity estimations are generally less dependent on the delighting quality. Overall, our final model achieves the best results on every metric, with our cycle-consistency and residual approaches proving to be the most impactful components of the method.

4.4. Qualitative results

Fig. 5 shows qualitative results of our material delighting and relighting models, along with ground truth data. Our delighting model behaves accurately even in challenging cases, like the corduroy on the first column, the satin on the third or the suede leather on the last one. The predicted images contain no shading, wrinkles are hidden and shadows casted on the scanner lid are eliminated. It can be seen why the material relighting model is less accurate according to our metrics. Precisely introducing shadows, specular highlights or shading proves to be a more challenging task than removing them, and our relighting model sometimes misplaces or inaccurately estimates the intensity of these reflections. We believe that, during training using cycle-consistency, this helps the delighting model as this works as a short of data augmentation, as the delighting model is shown variations of the same material with different variations on shading and specularity.

Fig. 6 compares our outputs with UMat [3] for a variety of material types. We show the results on a render scene designed to better show the impact of material transparency. As shown, our model behaves accurately across many different materials, with precise and sharp albedo estimations, and realistic transparency and opacity predictions. These results highlight the importance of estimating these maps, as the results of UMat are less appealing and realistic in comparison.

In Fig. 7, we show results of different methods of material capture, for which we capture the input images using a smartphone. We include results on ambient lighting (top row) and flash-lit images (last two rows). The methods of Deep Inverse Rendering [12], Match [25],

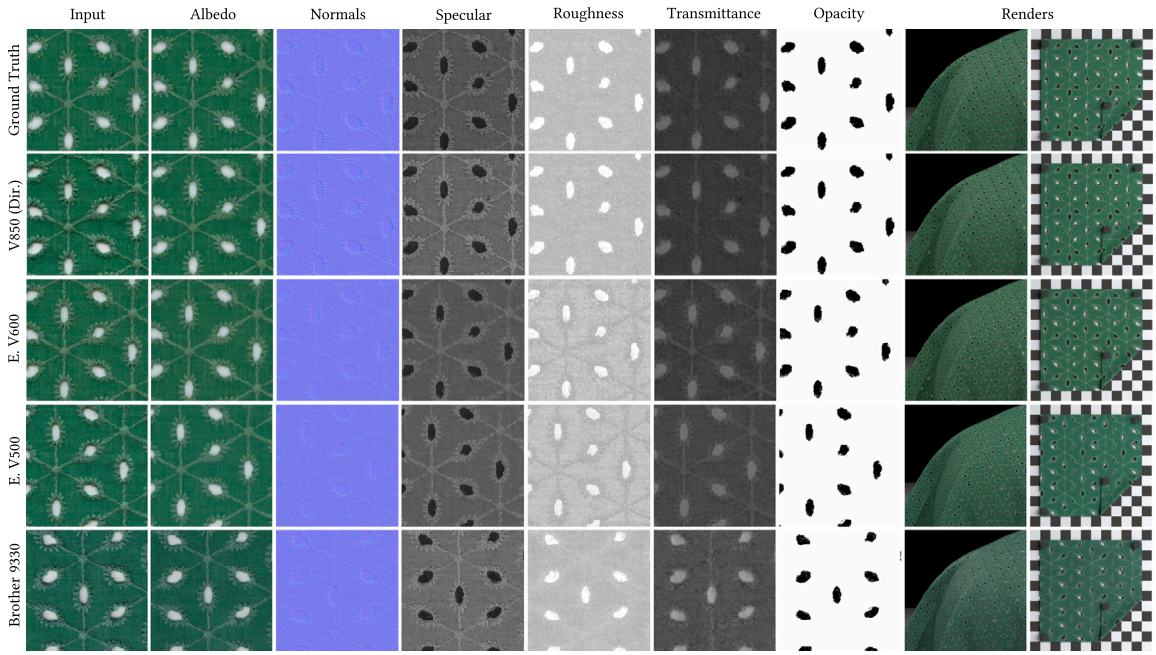

Fig. 8. Qualitative comparison between several flatbed scanners for the same material. Note that the images are not exactly pixel-wise coherent across scanners.

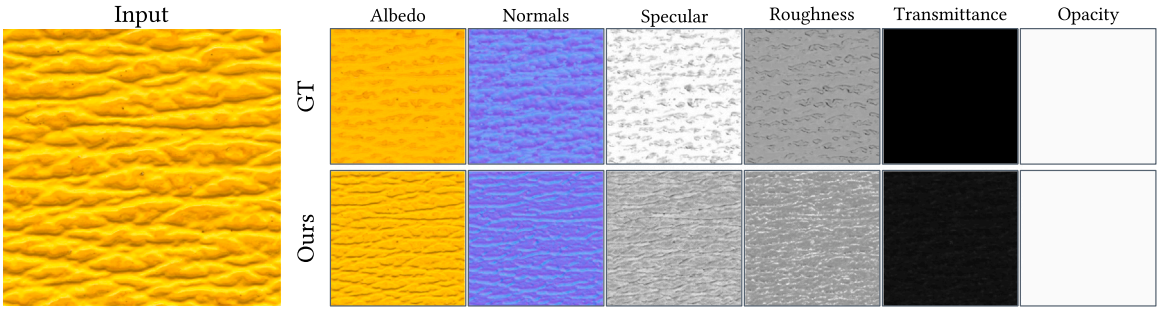

Fig. 9. A failure case of our method. For the input image on the right, we show the ground truth albedo, normals, specular, roughness, transmittance and opacity maps (top row), and our model estimations.

Neural Materials [7], Adversarial SVBRDF Estimation [17] all assume a smartphone capture, while UMat [3] and our method assume a flatbed scanner capture. Regardless on the illumination conditions, our model provides sharp and accurate estimations which better preserve the structure and color of the inputs compared to generative or optimization-based models [7,12,25]. Compared to UMat [3], our delighting framework enables more uniform and higher-quality albedos, and we achieve more globally coherent maps than [17]. Overall, our method proves robust to images captured with smartphones across a variety of illumination conditions, even if we never train our models with this type of data. Note that this comparisons are done in a qualitative fashion as there is no accessible ground truth for their SVBRDF maps.

In Fig. 8, we show the results achieved by our model on the same material, across a variety of flatbed scanners. This material is challenging, containing holes, wrinkles and fly-away fibers, all of which pose problems for digitization. The inputs on the first two rows were captured with a high-end EPSON V850 Pro scanner, for which we show the ground truth materials and renders (first), and the estimation on the directional light configuration on said scanner. On the third and fourth rows, we show the results on lower-end Epson V600 and V500 flatbed scanners, and the final row was captured with a Brother 9930 multi-functional fax machine, which also contains a budget scanner. As shown, our model estimations remain consistent across devices, regardless on the quality of the input scanner. The results for Brother 9930 struggle in terms of color due to calibration issues, but our SVBSDF estimation contains sharp, accurate reflectance maps. More results are shown in the supplementary material.

4.5. Failure cases and limitations

Our method inherits the limitations of using flatbed scanners as a capture device. This setup cannot be used for non-flat materials (eg the marble in a statue) or materials which cannot physically be placed into this setup (eg a wall). It is also limited by our material model of choice, which, while it is more expressive than those of previous work, it cannot accurately represent complex phenomena such as anisotropy, strong displacements, high reflectivity, or subsurface scattering. Also, in order to capture non-uniform multi spectral absorption, $\mathbf{T}$ would require an additional attenuation value for each wavelength channel. Finally, our model sometimes struggles with some complex materials for which a single image is not a sufficient cue to estimate its optical properties. Such is the case for the bright, thick leather we show in Fig. 9, which is the material in our test set with the highest $\mathcal{L}_{\text{BSDF}}$. This type of material is also uncommon in our training dataset, which also explains the reduced generalization.

CRediT authorship contribution statement

Carlos Rodriguez-Pardo: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **David Pascual-Hernandez:** Validation, Software, Formal analysis, Data curation. **Javier Rodriguez-Vazquez:** Software, Data curation. **Jorge Lopez-Moreno:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Elena Garces:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, 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.

Acknowledgment

V+Real, PID2021-122392OB-I00 funded by MCIN/AEI/10.13039/501100011033/FEDER, UE.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cag.2025.104186.

Data availability

Data will be made available on request.

References

[1] Garces E, Arellano V, Rodriguez-Pardo C, Pascual-Hernandez D, Suja S, Lopez-Moreno J. Towards material digitization with a dual-scale optical system. ACM Trans Graph 2023;42(4):1–13.
[2] X-RITE: Tac7. 2016, https://www.xrite.com/categories/appearance/total-appearance-capture-ecosystem/tac7. [Accessed: 28 April 2022].
[3] Rodriguez-Pardo C, Dominguez-Elvira H, Pascual-Hernandez D, Garces E. Umat: Uncertainty-aware single image high resolution material capture. In: Proceedings of the IEEE/CVF international conference on computer vision. 2023.
[4] Martin R, Meyer A, Pesare D. De-lighting a high-resolution picture for material acquisition.. In: EGSR (DL/i). 2019, p. 69–72.
[5] Garces E, Rodriguez-Pardo C, Casas D, Lopez-Moreno J. A Survey on Intrinsic Images: Delving Deep into Lambert and Beyond. Int J Comput Vis 2022.
[6] Gatys LA, Ecker AS, Bethge M. A neural algorithm of artistic style. 2015, arXiv preprint arXiv:1508.06576.
[7] Henzler P, Deschaintre V, Mitra NJ, Ritschel T. Generative modelling of BRDF textures from flash images. ACM Trans Graph (Proc SIGGRAPH Asia) 2021;40(6).
[8] Aittala M, Aila T, Lehtinen J. Reflectance modeling by neural texture synthesis. ACM Trans Graph (ToG) 2016;35(4):1–13.
[9] Deschaintre V, Aittala M, Durand F, Drettakis G, Bousseau A. Single-image svbrdf capture with a rendering-aware deep network. ACM Trans Graph 2018;37(4):1–15.
[10] Li Z, Sunkavalli K, Chandraker M. Materials for masses: SVBRDF acquisition with a single mobile phone image. In: Proceedings of the European conference on computer vision. 2018, p. 72–87.
[11] Ye W, Li X, Dong Y, Peers P, Tong X. Single image surface appearance modeling with self-augmented cnns and inexact supervision. In: Computer graphics forum. 37, (7):Wiley Online Library; 2018, p. 201–11.
[12] Gao D, Li X, Dong Y, Peers P, Xu K, Tong X. Deep inverse rendering for high-resolution SVBRDF estimation from an arbitrary number of images. ACM Trans Graph (ToG) 2019;38(4):134:1–134:15.
[13] Li Z, Xu Z, Ramamoorthi R, Sunkavalli K, Chandraker M. Learning to reconstruct shape and spatially-varying reflectance from a single image. ACM Trans Graph 2018;37(6):1–11.
[14] Sang S, Chandraker M. Single-shot neural relighting and svbrdf estimation. In: Proceedings of the European conference on computer vision. Springer; 2020, p. 85–101.
[15] Wen T, Wang B, Zhang L, Guo J, Holzschuch N. SVBRDF recovery from a single image with highlights using a pre-trained generative adversarial network. In: Computer graphics forum. Wiley Online Library; 2022.
[16] Guo J, Lai S, Tao C, Cai Y, Wang L, Guo Y, et al. Highlight-aware two-stream network for single-image svbrdf acquisition. ACM Trans Graph 2021;40(4):1–14.
[17] Zhou X, Kalantari NK. Adversarial single-image SVBRDF estimation with hybrid training. In: Computer graphics forum, vol. 40, no. 2. Wiley Online Library; 2021, p. 315–25.
[18] Zhou X, Hasan M, Deschaintre V, Guerrero P, Sunkavalli K, Kalantari NK. Tilegen: Tileable, controllable material generation and capture. In: SIGGRAPH Asia 2022 conference papers. 2022, p. 1–9.
[19] Vecchio G, Palazzo S, Spampinato C. SurfaceNet: Adversarial SVBRDF estimation from a single image. In: Proceedings of the IEEE/CVF international conference on computer vision. 2021, p. 12840–8.
[20] Luo X, Scandolo L, Bousseau A, Eisemann E. Single-image SVBRDF estimation with learned gradient descent. In: Computer graphics forum (proceedings of eurographics), vol. 43, no. 2. 2024.
[21] Vecchio G, Martin R, Roullier A, Kaiser A, Rouffet R, Deschaintre V, Boubekeur T. ControlMat: A controlled generative approach to material capture. 2023, arXiv preprint arXiv:2309.01700.
[22] Vecchio G, Sortino R, Palazzo S, Spampinato C. Matfuse: Controllable material generation with diffusion models. 2023, arXiv preprint arXiv:2308.11408.
[23] Yuan L, Yan D, Saito S, Fujishiro I. DiffMat: Latent diffusion models for image-guided material generation. Vis Informatics 2024;8(1):6–14.
[24] Sartor S, Peers P. MatFusion: a generative diffusion model for SVBRDF capture. In: ACM SIGGRAPH Asia conference proceedings. 2023, http://dx.doi.org/10.1145/3610548.3618194.
[25] Shi L, Li B, Hašan M, Sunkavalli K, Boubekeur T, Mech R, et al. Match: differentiable material graphs for procedural material capture. ACM Trans Graph 2020.
[26] Hu Y, He C, Deschaintre V, Dorsey J, Rushmeier H. An inverse procedural modeling pipeline for svbrdf maps. ACM Trans Graph 2022;41(2):1–17.
[27] Guo Y, Hašan M, Yan L, Zhao S. A bayesian inference framework for procedural material parameter estimation. In: Computer graphics forum. 39, (7):Wiley Online Library; 2020, p. 255–66.
[28] Jin W, Wang B, Hasan M, Guo Y, Marschner S, Yan L-Q. Woven fabric capture from a single photo. In: SIGGRAPH Asia 2022 conference papers. 2022, p. 1–8.
[29] Aittala M, Weyrich T, Lehtinen J. Practical SVBRDF capture in the frequency domain.. ACM Trans Graph (ToG) 2013;32(4). 110–1.
[30] Zhang L, Gao F, Wang L, Yu M, Cheng J, Zhang J. Deep SVBRDF estimation from single image under learned planar lighting. In: ACM SIGGRAPH 2023 conference proceedings. 2023, p. 1–11.
[31] Xu X, Lin Y, Zhou H, Zeng C, Yu Y, Zhou K, et al. A unified spatial-angular structured light for single-view acquisition of shape and reflectance. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023, p. 206–15.
[32] Qu L, Tian J, He S, Tang Y, Lau RW. Deshadownet: A multi-context embedding deep network for shadow removal. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, p. 4067–75.
[33] Vasluianu F-A, Seizinger T, Timofte R. WSRD: A novel benchmark for high resolution image shadow removal. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR) workshops. 2023, p. 1826–35.
[34] Fu L, Zhou C, Guo Q, Juefei-Xu F, Yu H, Feng W, et al. Auto-exposure fusion for single-image shadow removal. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021, p. 10571–80.
[35] Wang J, Li X, Yang J. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
[36] Lagunas M, Sun X, Yang J, Villegas R, Zhang J, Shu Z, et al. Single-image full-body human relighting. 2021, arXiv preprint arXiv:2107.07259.
[37] Yeh Y-Y, Nagano K, Khamis S, Kautz J, Liu M-Y, Wang T-C. Learning to relight portrait images via a virtual light stage and synthetic-to-real adaptation. ACM Trans Graph 2022;41(6):1–21.
[38] Wimbauer F, Wu S, Rupprecht C. De-rendering 3d objects in the wild. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022, p. 18490–9.
[39] Ji C, Yu T, Guo K, Liu J, Liu Y. Geometry-aware single-image full-body human relighting. In: European conference on computer vision. Springer; 2022, p. 388–405.
[40] Zhu J-Y, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. 2017, p. 2223–32.
[41] Chu C, Zhmoginov A, Sandler M. Cyclegan, a master of steganography. 2017, arXiv preprint arXiv:1712.02950.
[42] Kaneko T, Kameoka H, Tanaka K, Hojo N. Cyclegan-vc2: Improved cyclegan-based non-parallel voice conversion. In: ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing. IEEE; 2019, p. 6820–4.

[43] Yang H, Sun J, Carass A, Zhao C, Lee J, Prince JL, et al. Unsupervised MR-to-CT synthesis using structure-constrained cyclegan. IEEE Trans Med Imaging 2020;39(12):4249–61.
[44] Harms J, Lei Y, Wang T, Zhang R, Zhou J, Tang X, et al. Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. Med Phys 2019;46(9):3998–4009.
[45] Lu Y, Tai Y-W, Tang C-K. Attribute-guided face generation using conditional cyclegan. In: Proceedings of the European conference on computer vision. 2018, p. 282–97.
[46] Wu CH, De la Torre F. Unifying diffusion models' latent space, with applications to CycleDiffusion and guidance. 2022, arXiv preprint arXiv:2210.05559.
[47] Su X, Song J, Meng C, Ermon S. Dual diffusion implicit bridges for image-to-image translation. 2022, arXiv preprint arXiv:2203.08382.
[48] Burley B. Physically-based shading at disney. In: SIGGRAPH courses: practical physically based shading in film and game production. 2012.
[49] Walter B, Marschner SR, Li H, Torrance KE. Microfacet models for refraction through rough surfaces. In: Proceedings of the 18th eurographics conference on rendering techniques. 2007, p. 195–206.
[50] Burley B. Physically based shading in theory and practice: Extending the disney BRDF to a BSDF with integrated subsurface. In: In ACM SIGGRAPH 2015 courses. New York, NY, USA: Association for Computing Machinery; 2015.
[51] Rodriguez-Pardo C, Kazatzis K, Lopez-Moreno J, Garces E. NeuBTF: Neural fields for BTF encoding and transfer. Comput Graph 2023.
[52] Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE/CVF international conference on computer vision. 2018, p. 586–95.
[53] Jiang L, Dai B, Wu W, Loy CC. Focal frequency loss for image reconstruction and synthesis. In: Proceedings of the IEEE/CVF international conference on computer vision. 2021, p. 13919–29.
[54] Rodriguez-Pardo C, Garces E. SeamlessGAN: Self-Supervised Synthesis of Tileable Texture Maps. IEEE Trans Vis Comput Graphics 2022.
[55] Zhou Y, Zhu Z, Bai X, Lischinski D, Cohen-Or D, Huang H. Non-stationary texture synthesis by adversarial expansion. ACM Trans Graph 2018;37(4):1–13.
[56] Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE/CVF international conference on computer vision. 2017, p. 1125–34.
[57] Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, et al. Automatic differentiation in pytorch. 2017.
[58] Marcel S, Rodriguez Y. Torchvision the machine-vision package of torch. In: Proceedings of the 18th ACM international conference on multimedia. 2010, p. 1485–8.
[59] Riba E, Mishkin D, Ponsa D, Rublee E, Bradski G. Kornia: an open source differentiable computer vision library for pytorch. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2020, p. 3674–83.
[60] Kingma DP, Ba J. Adam: A method for stochastic optimization. In: International conference on learning representations. 2015.
[61] Micikevicius P, Narang S, Alben J, Diamos G, Elsen E, Garcia D, et al. Mixed precision training. 2017, arXiv preprint arXiv:1710.03740.
[62] Hu W, Xiao L, Pennington J. Provable benefit of orthogonal initialization in optimizing deep linear networks. 2020, arXiv preprint arXiv:2001.05992.
[63] Biewald L. Experiment tracking with weights and biases. 2020, URL https://www.wandb.com/. Software available from wandb.com.
[64] Fong C. Analytical methods for squaring the disc. 2015, arXiv preprint arXiv:1509.06344.
[65] Loshchilov I, Hutter F. Decoupled weight decay regularization. 2017, arXiv preprint arXiv:1711.05101.
[66] Rodriguez-Pardo C, Garces E. Neural photometry-guided visual attribute transfer. IEEE Trans Vis Comput Graphics 2021.
[67] Mokrzycki W, Tatol M. Colour difference E-A survey. Mach Graph Vis 2011;20(4):383–411.
[68] Andersson P, Nilsson J, Akenine-Möller T, Oskarsson M, Åström K, Fairchild MD. ꟻLIP: A Difference Evaluator for Alternating Images. Proc the ACM Comput Graph Interact Tech 2020;3(2):15:1–23.
[69] Lavoué G, Bonneel N, Farrugia J-P, Soler C. Perceptual quality of BRDF approximations: dataset and metrics. In: Computer graphics forum, vol. 40, no. 2. Wiley Online Library; 2021, p. 327–38.
[70] Nielsen JB, Jensen HW, Ramamoorthi R. On optimal, minimal BRDF sampling for reflectance acquisition. ACM Trans Graph 2015;34(6):1–11.
[71] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 2004;13(4):600–12.