Guided Fine-Tuning for Large-Scale Material Transfer

Valentin Deschaintre, George Drettakis, Adrien Bousseau

Eurographics Symposium on Rendering 2020
C. Dachsbacher and M. Pharr
(Guest Editors)

Volume 39 (2020), Number 4

Guided Fine-Tuning for Large-Scale Material Transfer

Valentin Deschaintre[1,2,3], George Drettakis[1] and Adrien Bousseau[1]

[1] Université Côté d'Azur, Inria [2] Imperial College London
[3] Optis for Ansys

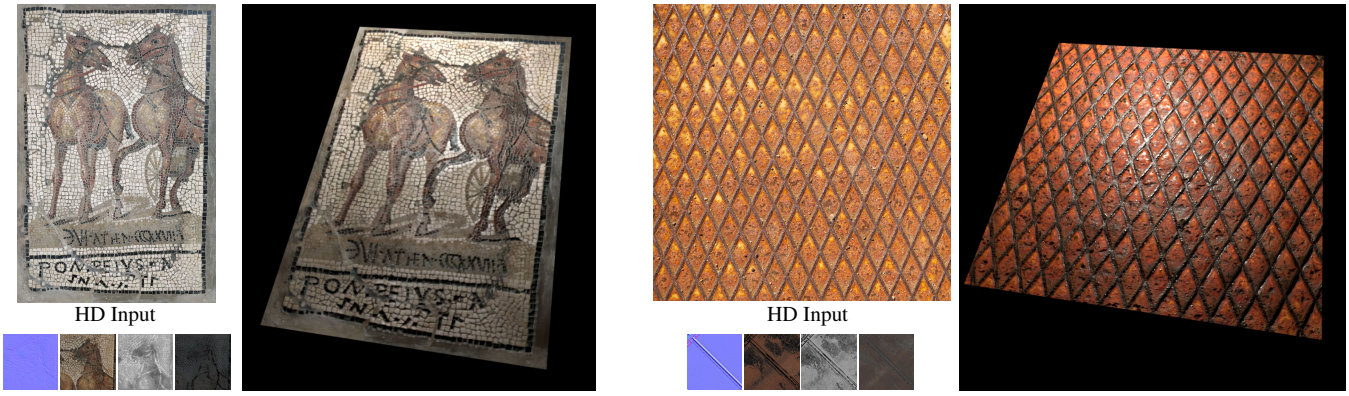

Figure 1: *Our method transfers the appearance of one or a few exemplar SVBRDFs to a target picture. This approach allows the capture of large planar surfaces taken with ambient lighting (far left), by extracting the SVBRDF exemplars from close-up flash pictures (lower left), as well as the creation of plausible SVBRDFs from internet pictures by using existing artist-designed materials as exemplars (right). Please see supplemental materials for high-resolution SVBRDF parameter maps and animated renderings of all our results, which give a much better impression of the material properties.*

Abstract

We present a method to transfer the appearance of one or a few exemplar SVBRDFs to a target image representing similar materials. Our solution is extremely simple: we fine-tune a deep appearance-capture network on the provided exemplars, such that it learns to extract similar SVBRDF values from the target image. We introduce two novel material capture and design workflows that demonstrate the strength of this simple approach. Our first workflow allows to produce plausible SVBRDFs of large-scale objects from only a few pictures. Specifically, users only need take a single picture of a large surface and a few close-up flash pictures of some of its details. We use existing methods to extract SVBRDF parameters from the close-ups, and our method to transfer these parameters to the entire surface, enabling the lightweight capture of surfaces several meters wide such as murals, floors and furniture. In our second workflow, we provide a powerful way for users to create large SVBRDFs from internet pictures by transferring the appearance of existing, pre-designed SVBRDFs. By selecting different exemplars, users can control the materials assigned to the target image, greatly enhancing the creative possibilities offered by deep appearance capture.

CCS Concepts

• **Computing methodologies** → **Reflectance modeling;** *Image processing;*

Keywords: material transfer, material capture, appearance capture, SVBRDF, deep learning, fine tuning

1. Introduction

Recent progress on lightweight appearance capture allows the recovery of plausible real-world spatially-varying reflectance distribution functions (SVBRDF) from just a few photographs of a surface. In particular, multiple methods take as input one or several photographs captured with a hand-held camera, where the co-located flash provides informative spatially-varying illumination over the measured surface sample [AWL15, AAL16, RPG16, HSL*17, DAD*18, LSC18, DAD*19, GLD*19]. However, near-field flash lighting greatly restricts the *scale* at which materials can be captured – typically a dozen centimeters wide using a cell phone held at a similar distance. Relying on a flash also prevents these methods from processing existing images captured under *unknown lighting*, such as textures downloaded from the Internet. Finally, another common limitation of the above methods is that they rely on black-box optimization or deep learning to infer SVBRDF parameters from few measurements, offering little *user control* on their output. We address all three limitations by proposing a *by-example* appearance capture method, which recovers SVBRDF parameter maps over large surfaces captured under environment lighting by transferring information from one or a few *exemplar SVBRDF patches* (Fig. 1), that can either be extracted from additional close-up flash photos, or come from a database of SVBRDFs.

Our technical solution to transfer material appearance from exemplars is surprisingly simple yet extremely effective. We build on a state-of-the-art SVBRDF capture deep network [DAD*18], which we re-train to take as input a single image captured under environment lighting, and output SVBRDF maps (normals, diffuse albedo, specular albedo, and roughness). Our key idea is to fine-tune this network on the provided exemplars, which strongly biases the network towards their specific SVBRDF values using the available color and texture cues. We then run this custom network on the target image, which effectively produces SVBRDF maps that contain similar values to the ones of the exemplars.

However, naively fine-tuning a large deep network on a small number of exemplars results in dramatic over-fitting, as the network quickly memorizes the spatial layout of the exemplars rather than learn material-specific filters that would generalize to the target image. We address this challenge by carefully augmenting the exemplar set. In particular, we generate a unique training image for each iteration of the fine-tuning by applying random geometric transformations on the exemplars, and by combining multiple transformed exemplars into a single collage via random masks. Our experiments demonstrate that this augmentation is critical to the success of the method.

We introduce two new applications that demonstrate the strength of our approach. Our *on-site acquisition* scenario is the first application to allow capture of plausible material properties of *large* surfaces with just a few photos. In this case, we capture a single photograph of a large surface as well as one or a few close-up flash photographs of its details. We then use an off-the-shelf network to extract SVBRDF maps from the flash photographs, and use our fine-tuned network to transfer this information to the large image, effectively acquiring SVBRDFs several meters wide. In our second scenario – *creative design* – we provide a powerful method for users to create realistic SVBRDFs from stock photos, simply using artist-created SVBRDFs downloaded from the Internet as exemplars. This demonstrates how our method allows fine control on the design process for SVBRDFs.

In summary, this paper makes the following contributions:

- We present a simple yet very effective algorithm to transfer material appearance from a few exemplars to a target image.
- We introduce a lightweight method to capture SVBRDFs of large planar surfaces, based on this algorithm.
- We introduce a novel workflow that allows material designers to create new SVBRDFs from existing photos and SVBRDF patches (*e.g.*, taken from online texture and SVBRDF repositories), using the same algorithm.

  Our code, data and supplemental material are available here: https://team.inria.fr/graphdeco/projects/large-scale-materials/

2. Related Work

Appearance capture and design is a vast and active research field; We refer to the survey by Guarnera et al [GGG*16] for a general introduction, and to the one by Dong [Don19] for a focus on methods based on deep learning. Here we discuss lightweight SVBRDF capture methods most similar to our approach, as well as related work on by-example image synthesis and deep learning.

Reconstructing multiple SVBRDF maps from one or a few pictures is an ill-posed problem, as the radiance observed in the pictures can be explained by a number of different combinations of SVBRDF parameters. Existing work tackled this challenge by incorporating domain-specific priors on the solution, either designed by hand or learned from large quantities of SVBRDF data. Example hand-designed priors include the assumption that the material sample is stochastic or self-repetitive [WSM11, AWL15, AAL16], or that the lighting exhibits natural statistics [DCP*14] and physical properties [RRFG17]. Data-driven methods seek to explain the observed data as a combination of known BRDFs [HSL*17, RWS*11], or more recently by training deep neural networks to predict SVBRDF parameter maps using synthetic data for supervision [DAD*18, LSC18, DAD*19, GLD*19]. While the above methods target planar surfaces like ours, some have also been extended to the problem of jointly capturing shape and material appearance, either using inverse-rendering optimization [BJTK18, NLGK18] or deep learning [LXR*18].

Most of these methods succeed in the task by targeting flash pictures captured at a small distance from planar material samples or small curved objects. In such a configuration, the flash produces a highlight at the center of the image as well as diffuse shading on its boundary, which provides information about the specular and diffuse behavior of the surface respectively, as well as complementary cues about normal variations. However, the use of a flash imposes three limitations for such methods. First, capturing large-scale surfaces would require the use of a large, powerful flash, defeating the purpose of these lightweight methods. Second, because flash lighting yields different visual cues in different places of the image, existing methods need to process the image in its entirety to aggregate all information, which is problematic for deep learning

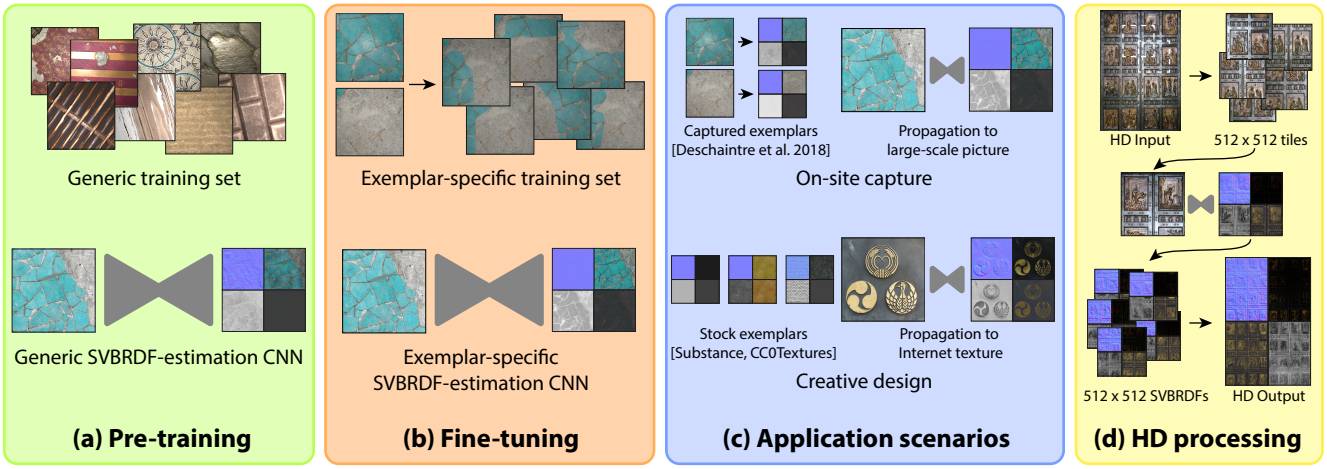

Figure 2: *Main steps of our method. We first pre-train an SVBRDF prediction network [DAD*18] on a large set of synthetic SVBRDF maps rendered under varying lighting (a). While this generic network produces plausible results, it often mis-interprets the material features in the absence of flash cues. Our key idea is to fine-tune the pre-trained network on renderings of user-provided SVBRDF exemplars (b). After fine-tuning, the resulting network combines generic pre-training knowledge with information from the exemplars. Here, this allows our method to interpret the cyan tiles as more shiny than the grey concrete. We demonstrate this approach on two application scenarios, either to acquire large-scale real-world surfaces by propagating small-scale exemplars (c, top), or to design new SVBRDFs by propagating existing SVBRDF maps over internet textures (c, bottom). While we train our network on images of* 512×512 *pixels, we process HD images of more than* 2048×2048 *pixels by processing small tiles individually, and by stitching their predicted SVBRDFs to generate the final output. This is made possible by the absence of strong local flash highlights in the input image.*

methods as the network resolution is limited by the GPU memory – related methods were typically demonstrated on images of 256×256 pixel resolution. In supplemental material we provide an example showing how the method by Deschaintre et al. [DAD*18], trained at low resolution, degrades when applied at higher resolution since the relative network footprint is reduced compared to the size and location of the flash highlights. Third, the reliance on co-located flash lighting prevents these methods from handling images taken in the wild with unknown lighting and arbitrary scale. Our approach lifts all these limitations thanks to SVBRDF exemplars that bias the interpretation of the image towards specific material values, effectively alleviating the need for the visual cues offered by flash lighting.

In contrast to the above methods, Li et al. [LDPT17] proposed a deep network capable of predicting SVBRDFs fom images captured under environment lighting, including images taken in the wild. However, environment lighting alone provides little in terms of visual cues of the complex behavior of SVBRDFs, which makes their results inferior to the ones obtained by more recent flash-based methods. In particular, their method assumes that the specular term of the BRDF does not vary spatially, while spatially-varying roughness greatly contributes to the richness of real-world materials. Li et al. [LDPT17] also introduced the concept of *self-augmentation*, which was further studied by Ye et al. [YLD*18]. The idea is to use the network output to build new training samples, effectively augmenting the *diversity* of SVBRDFs seen by the network. This strategy differs from ours, since our goal is rather to *specialize* the network to extract user-provided SVBRDF values, which we achieve by fine-tuning the network on specific exemplars.

Our use of exemplar images makes our problem akin to *image analogies* [HJO*01], where the goal is to copy the appearance of an exemplar onto a target. The image analogies framework has been applied to a variety of problems, such as image colorization [WAM02], style transfer [FJL*16], texture transfer [DBP*15]. All these methods share the strength of providing high-level control on their output thanks to the exemplars, a feature that we now provide in the context of SVBRDF capture and design. Closer to our application domain is the work by Melendez et al. [MGSJW12], who used patch-based texture synthesis to transfer diffuse albedo and depth variations from small material exemplars to large façade images. However, their approach assumes that every pixel of the target can be put in correspondence to similar pixels of the exemplar, which yields visual artifacts when the exemplars do not contain all the material variations of the target image (see Fig. 10). Several recent methods use deep learning for image-to-image translation problems in supervised [IZZE17, WLZ*18] or unsupervised settings [ZPIE17]. In particular, multiple methods combine dense correspondences with deep learning to achieve more robust colorization [HCL*18, HLC*19] and style transfer [LYY*17]. Our solution is simpler as it does not require explicit correspondences between the exemplars and the target. Instead, we train a deep material capture network to learn the mapping between the colors and textures of the exemplars and their SVBRDFs values, allowing us to apply this mapping on the target. In concurrent work, Texler et al. [TFK*20] used a similar strategy to specialize a style transfer network using a small number of style exemplars.

By complementing an input image with a few user-provided exemplars, our approach also relates to the interactive material design

system *AppGen* [DTPG11]. The main difference between the two approaches resides in the level of expertise required and control offered. While AppGen offers fine control on the local interpretation of an image thanks to user scribbles, it requires users to manually segment the different materials in the image, and to specify each specular BRDF. In contrast, users of our approach need only select exemplar SVBRDFs from an existing library, or acquire them using an existing lightweight method, and let our method automatically transfer BRDF values from the exemplars to the target image. Our on-site acquisition scenario also follows the same two-scale capture strategy as *Manifold Bootstrapping* [DWT*10], although we only need a few pictures of the surface at small and large scale where Dong et al. rely on specialized hardware to capture local BRDF samples, and on multiple photographs under varying lighting to capture global appearance.

Our technical solution for material transfer is inspired by the recent concept of *internal learning*, i.e., training a deep neural network on a specific image rather than on a large dataset. This intriguing idea first appeared in the seminal work of Ulyanov et al. [UVL18] on *deep image priors*, where a network trained to reconstruct a specific image was shown to denoise or inpaint that image. Subsequent work used image-specific training for various tasks, including unsupervised super-resolution [SCI18] and GAN-based image editing [BSP*19,SDM19]. Our approach differs, since while we fine-tune a deep network on a small set of images, we use the resulting network to *transfer* the knowledge it acquired on a different target image. Our work also relates to the *TileGAN* method of Frühstück et al. [FAW19], who train a conditional GAN to perform small-scale texture synthesis, and apply this GAN in a sliding-window fashion to produce large-scale images. However, training a GAN to synthesize a specific texture takes several days, while we show that it takes only a few minutes to fine-tune a generic material acquisition network to achieve successful material transfer. Our strategy can also be seen as a form of *few-shot learning*, that aims at adapting a pre-trained model to a new category of data given only a few examples of such data [LHM*19]. As mentioned above, in our context, a few minutes of fine-tuning on augmented exemplars is sufficient to achieve this adaptation.

3. Method

Fig. 2 provides a visual overview of our method to extract SVBRDF parameter maps for large-scale surfaces. The main steps include pre-training a deep SVBRDF prediction network on a varied set of SVBRDFs (Fig. 2a), fine-tuning this network on our exemplars (Fig. 2b), and finally using this exemplar-specific network to extract SVBRDFs similar to the exemplars over images of large surfaces, either captured on site or downloaded from the Internet (Fig. 2c). We first describe typical inputs to our method, before explaining how we pre-train and fine-tune the deep network to achieve material transfer.

3.1. Inputs

Our goal is to generate SVBRDF parameter maps for large-scale planar surfaces, such as walls, doors or furniture. To do so, our method takes two forms of input. First, a single picture of the surface of interest, captured under ambient indoor or outdoor lighting. Second, a series of SVBRDF patches that represent small parts of the surface, or of a similar material. To obtain these patches, we either capture close-up flash pictures of the surface and run an existing single-image SVBRDF method [DAD*18], or we select SVBRDFs from libraries of artist-designed materials [Ado19, Str19] (Fig. 2c).

As a pre-process, we split the large-scale image into tiles of 512×512 pixels. Our method processes each tile independently, and generates the final output by stitching these individual predictions (Fig. 2d, Sec. 3.5). Neighboring tiles have an overlap of 256 pixels to facilitate subsequent stitching of their SVBRDF maps. Applying the network in this sliding-window fashion ensures that our method has a constant memory footprint, and as such scales to images of arbitrary resolution. In contrast, while running the network in a fully-convolutional manner would also allow the processing of images of varied resolution [GLD*19], the memory consumption of the method would increase with resolution, and eventually saturate GPU memory.

Note also that we assume that all tiles receive approximately the same lighting, which is not the case for pictures taken with a flash as used in prior work [DAD*18, LSC18, DAD*19, GLD*19].

3.2. Neural network pre-training

Our method processes each tile of the input image independently to output four Cook-Torrance SVBRDF maps [CT82], corresponding to the normal, diffuse albedo, specular albedo, and specular roughness of each input pixel. We perform this task with the convolutional neural network proposed by Deschaintre et al. [DAD*18]. While the original network was trained with synthetic images rendered under flash lighting, we re-train it with images rendered under a random directional light to be robust to arbitrary lighting conditions in our inputs. We also mimic a simple white sky dome by adding a small multiple of the diffuse and specular albedos to the renderings, which we found to be necessary to prevent metallic materials to appear completely dark away from the specular highlight. Despite its simplicity, we found this lighting model to work well on real-world pictures, including textures downloaded from the internet (Sec. 4). We generated our training data with the same set of parametric SVBRDFs as Deschaintre et al., except that we render them at a higher resolution to train the network to process images of 512×512 pixels. In total, the network is pre-trained for 800.000 iterations, which took around 8 days on a 1080TI graphics card.

Pre-training the network on a large set of SVBRDFs not only accelerates the subsequent fine-tuning step, it also equips the network with general priors on material appearance, which complements the exemplar-specific priors learned during fine-tuning (see Fig. 6).

3.3. Neural network fine-tuning

A single image often does not provide enough information to recover SVBRDF parameters unambiguously, especially in the absence of flash highlights. The key idea of our work is to favor the SVBRDF parameter values present in the exemplars by fine-tuning the network on these images. In other words, we perform a number

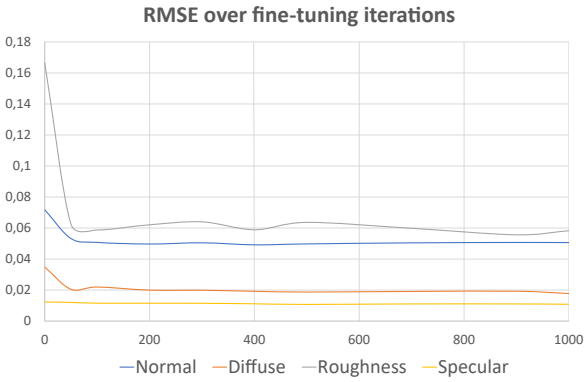

Figure 3: *Average RMSE of predicted maps for* 4 *synthetic SVBRDFs, using crops of these SVBRDFs as exemplars. The error quickly drops in less than* 100 *training iterations.*

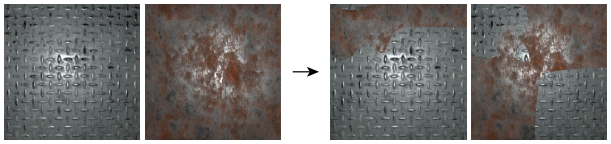

Figure 4: *We augment the input set of SVBRDF exemplars by composing them using a low-frequency random mask.*

of training iterations where we ask the network to predict exemplar SVBRDF maps given a rendering of that SVBRDF as input. The network thus becomes increasingly specialized in mapping the color and texture of the exemplar renderings to their normal and reflectance values. We used 1000 training iterations for all our results, which takes around 2 minutes on a 1080 Ti GPU and is largely sufficient to achieve successful transfer. Our numerical experiments suggest that most of the improvement occurs within a few hundred iterations (Fig. 3). Once fine-tuned, we run the network on each input tile to obtain its SVBRDF maps.

3.4. Exemplar augmentation

While extremely simple, the above procedure quickly overfits the network so much on the few exemplars that it does not generalize to input images having a different distribution of materials regions. Our solution to this challenge is to apply massive data augmentation on the exemplars to obtain a training set that retains their local appearance, but varies their overall layout. We achieve this goal by generating, for every training iteration, a unique SVBRDF that is composed of pieces of two randomly-selected different exemplars. We first apply random scaling and cropping on these exemplars, and then combine them according to a binary mask that we generate by thresholding a low-frequency Perlin noise (Fig. 4). We perform all these processing steps at training time in TensorFlow [AAB*15] to reduce storage and data transfer. When only one exemplar is provided, we only augment it with scaling and cropping. We use the same lighting model as for pre-training to render this training set.

3.5. Post-processing

The last step of our method consists in merging the predictions of all tiles into a large-scale SVBRDF. Since all tiles are processed using the same exemplars, neighboring tiles mostly agree in their predictions up to low frequency variations. We achieve a seamless composite by blending the tiles over their overlap using a Gaussian weighting kernel that gives a weight of 1 at the center of the tile and reaches almost 0 at its border. This mechanism allows our method to be applied on *high-resolution inputs of arbitrary aspect ratio*, as shown in our results of up to 2048×2048 pixels.

4. Evaluation

We first present results obtained by applying our method on our own photographs as well as on internet images. We then evaluate the impact of our fine-tuning and data augmentation strategies. Finally, we compare our method with alternative approaches on synthetic data for which we have ground truth SVBRDF maps. Please see supplemental materials for high-resolution SVBRDF parameter maps and animated renderings of all our results. We will release our code and data upon acceptance to ease reproduction.

4.1. Results

Our research was originally motivated by the need to quickly acquire the appearance of large-scale surfaces with minimal hardware. Following this first usage scenario, we used a smartphone to photograph a variety of planar objects. For each object, we first captured a single photograph of the entire surface under ambient lighting. We then captured 1-3 close-up flash photos of parts that exhibit characteristic material features. Finally, we ran the single-image network [DAD*18] to obtain SVBRDF exemplars for each close-up. Fig. 1 and 5 show a mosaic, tiled floors, and a sculpted wall captured on-site with this approach. Thanks to the exemplars provided, our method faithfully reproduces the varying shininess of the different tiles, and distinguishes rough stone from metal.

A second usage scenario of our method is to estimate the SVBRDF maps of existing pictures, using pre-designed SVBRDFs as exemplars of similar materials. Fig. 6 shows this on three internet images, processed with exemplars from libraries of artist-created procedural SVBRDFs [Ado19, Str19]. Our method transfers diffuse and specular reflectances of the exemplars across the surface while conforming to the input image. In this workflow, the user selects exemplars that correspond to the materials they would like to see over the large surface. For instance, by choosing appropriate exemplars, the golden part of the mural is successfully interpreted as having low roughness and yellow specular components, and rust is interpreted as a rough orange material. The last row of Fig. 6 illustrates the behavior of our method when part of the image is not covered by the provided exemplars. In this result, the exemplar guides the interpretation of the bricks, but not of the window. Nevertheless, our method also benefits from generic priors on material appearance learned during pre-training, here to interpret the dark window as more shiny than the brick.

While pre-designed SVBRDFs provide convincing material parameters, many come with normal maps that are either flat or

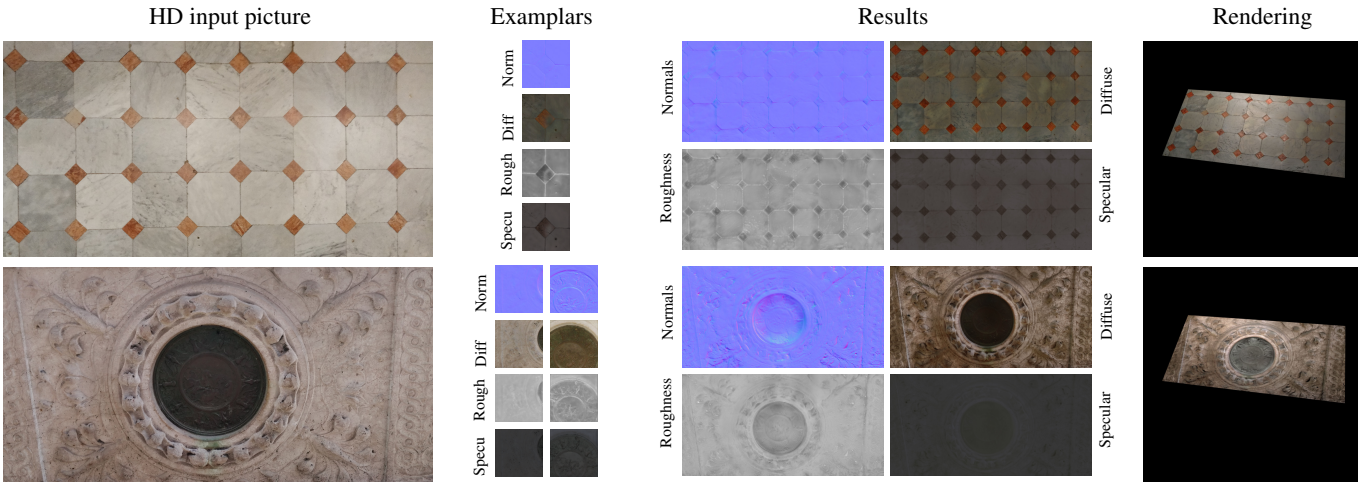

Figure 5: *Real-world surface captured on-site with our method. We used a single flash picture to capture the shininess of the tiles, which is propagated to all tiles of the large floor. We used two flash pictures for the second example, one for the diffuse stone and the other one for the more shiny metal disk. Please zoom on the .pdf to appreciate the high-resolution details of the individual SVBRDF maps. Images of resolution* 2048×1024.

weakly correlated to the target pictures. When this is the case, we ignore the normal map produced by the fine-tuned network and use the one produced by our pre-trained network instead. All results for which the exemplar normal map is not shown were obtained with this approach.

Fig. 7 further demonstrates the control that the input exemplars provide on the output SVBRDF. The input picture contains dark and yellow pixels with little in terms of visual cues of their respective shininess. We first selected a dark diffuse and a yellow metallic exemplar to achieve a golden appearance. We next show how changing the exemplar allows us to increase the roughness of the gold, or even to interpret the yellow pixels as diffuse paint. Finally, we also show how our method behaves in the presence of an outlier exemplar, which in this case gives a slight orange tint to the yellow pixels.

We show in Fig. 8 a visual comparison between real photographs of a surface and renderings of the SVBRDF created with our method. We used artist-designed SVBRDFs as exemplars for this comparison because the single-image method of Deschaintre et al. [DAD*18] fails to recover convincing maps from flash pictures of this complex surface (see supplemental materials for their result). This experiment shows that users can reproduce the desired overall appearance by guiding our method with adequate exemplars.

Finally, Fig. 14 showcases a variety of SVBRDFs created with our method, either via on-site acquisition or from stock photographs. Note that most of these results represent large, non-square surfaces encoded as high-resolution parameter maps, which contrasts with the small material samples often shown in related work.

4.2. Ablation study

We use the single-image network of Deschaintre et al. [DAD*18] as a backbone for SVBRDF prediction. Fig. 9 (first row) shows results of their method trained on our dataset of images rendered under random directional lighting. Without additional guidance, this method interprets the weathered golden door as made of rough plastic. Fig. 9 (second row) shows how fine-tuning this single-image network on two exemplars without data augmentation brings a golden appearance but distributes it uniformly over the surface. In our experiments, this tendency to produce uniform maps happens especially when the input exemplars are themselves uniform. By combining the exemplars to form random patterns, our data augmentation helps the transfer of the golden appearance to the least weathered parts of the door (Fig. 9, third row).

4.3. Comparisons

To our knowledge, our method is the first to offer by-example guidance for deep SVBRDF inference. We compare to related work on style transfer, as well as to single-image alternatives. We use synthetic SVBRDFs for these comparisons, which allows visual comparison to the ground truth maps, as well as numerical evaluation.

Qualitative comparisons. Our approach is related to the method by Melendez et al. [MGSJW12], which transfers diffuse albedo and displacement maps using patch-based texture synthesis akin to image analogies [HJO*01]. We reproduced this approach with the state-of-the-art patch-based synthesis algorithm of Fišer et al. [FJL*16], using the rendered SVBRDF as guidance. Note that since this algorithm was originally developed for style transfer, it assumes that the image to be synthesized only contains three color channels; to cope with this we ran their code on each SVBRDF parameter map separately. Fig. 10 shows results of this experiment;

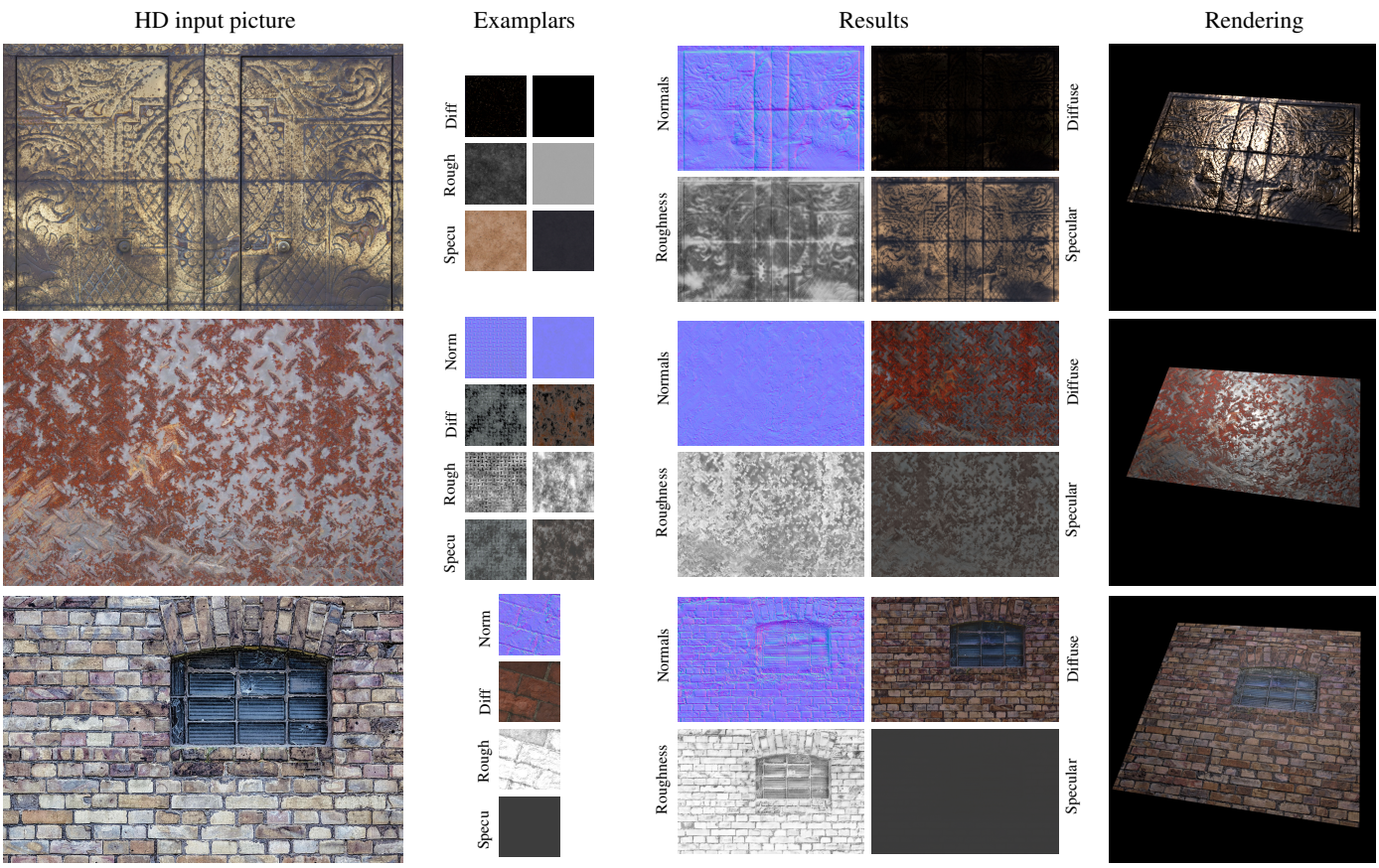

Figure 6: *Various SVBRDFs estimated from internet images. We selected artist-designed SVBRDF patches as exemplars for gold, paint, rust and bricks. Note how the shiny gold is well transferred to the yellow parts of the top panel, and how the diffuse rust is transferred to the brown parts of the middle plate. Note also that our method produces a plausible interpretation of the window (third row), even though the provided exemplar only contains bricks. Please zoom on the document to appreciate the high-resolution details of the individual SVBRDF maps. Images of resolution* 1536×1024.

	[DAD*18] No Flash	[LDPT17]	Few shot style transfer	Ours [DAD*18] exemplar	Ours GT exemplar
Normals	0.045		0.043	0.04	**0.039**
Diffuse	0.092		0.095	0.059	**0.028**
Roughness	0.215		0.195	0.142	**0.056**
Specular	0.016		0.015	0.021	**0.005**
Renderings	0.122	0.256	0.124	0.086	**0.071**

Table 1: *Numerical comparison to alternative methods using the RMSE metric (smaller is better), performed on synthetic SVBRDFs. Our method outperforms existing single-image algorithms thanks to the guidance of the exemplar (only one exemplar used). We only report the rendering error for [LDPT17] because this method outputs a different BRDF model than ours.*

Patch-based synthesis lacks variety in the maps due to the limited information contained in a single exemplar. While more advanced synthesis algorithms exist to interpolate between limited exemplars [DBP*15], our deep learning solution natively generalizes the exemplar to the entire large-scale image.

Fig. 10 also includes a comparison to AdaIN [HB17], a stylization algorithm based on deep learning that transfers statistics of deep features between an exemplar image and a target. Similarly to the above experiment, we applied the original implementation of the method on each SVBRDF parameter map separately. While this generic style transfer algorithm reproduces the overall color distribution of the maps, it misses many of the fine details.

Finally, we provide in Fig. 11 a visual comparison to the recent deep learning methods for single-image SVBRDF capture by Li et al. [LDPT17] and Deschaintre et al. [DAD*18]. While the method of Li et al. takes as input images captured under environment lighting, the original method of Deschaintre et al. assumes flash lighting, which is not compatible with the large-scale application scenarios we target. We thus re-trained their network on our training data to illustrate their performance on large-scale images taken without flash. Finally, for our method, we used the original method of Deschaintre et al. to recover SVBRDF exemplars from crops of the surface rendered under flash lighting, which emulates our on-site capture scenario. Both prior methods struggle to recover the shininess of the little metallic plates. Our method better recovers these small shiny parts thanks to the provided exemplar. In addition, our method can process large-scale images at high resolution, resulting in finer details in the SVBRDF maps.

Quantitative comparisons. Table 1 shows numerical comparisons to the single-image method of Deschaintre et al. [DAD*18] trained and tested on our data, and to the method by Li et al. [LDPT17] applied on images rendered under environment lighting. As in Fig. 11, we obtained exemplars for our method by providing crops of the ground truth SVBRDF rendered under flash lighting to the original method of Deschaintre et al. In this setup, a single exemplar is enough to outperform competitors. In addition, we also provide the performance of our method when guided by ground truth exemplars, which can be seen as an upper-bound on the quality it can achieve.

Finally, Table 1 (4th column) provides a numerical comparison to a version of our method inspired by the recent few-shot learning strategy proposed by Liu et al. [LHM*19], who build on AdaIN to transfer style from multiple exemplars provided at test time. We adapted their approach to our context by processing each SVBRDF exemplar with the encoder of Gao et al. [GLD*19] and by aggregating the resulting low-dimensional latent codes into a single code via max pooling. We next process this code with three fully-connected layers to produce parameters for several AdaIN layers that we use to transform the feature maps of the SVBRDF prediction network. The numerical evaluation reveals that the addition of AdaIN layers controled by the exemplars slightly improves performance over the baseline network of Deschaintre et al. [DAD*18] for some of the maps, but is largely inferior to our results obtained after fine-tuning this baseline on augmented exemplars.

4.4. Limitations

As with previous deep-learning based methods for material capture [DAD*18, LSC18], we cannot handle cast shadows, or any other phenomenon that requires more than a normal/bump map. Extending our approach to handle such cases, *e.g.*, using a displacement map, would require a much more complex differentiable renderer to handle 3D during training. Similarly, our SVBRDF model and renderer are not designed to handle non-local effects like subsurface scattering.

Despite the strong ability of deep learning to extract discriminative features, our method sometimes has difficulty distinguishing different materials that share similar colors and textures. This is the case in Fig. 12, where the shininess of the small metal disk is transferred to some of the stones that have a similar appearance in the input picture. Our method also assumes that the large-scale input is captured under largely uniform lighting. When this is not the case, large illumination gradients pollute the SVBRDF maps, as shown in Fig. 13. Nevertheless, our method is robust to localized highlights, as some occur in the training set (see synthetic materials in supplemental materials for typical examples).

Finally, while there is a theoretical limitation to the scale difference that our method can handle between the exemplar and large-scale input to correctly transfer the materials, we never encountered this problem in our tests.

5. Conclusion

Our method alleviates inherent limitations of flash-based material acquisition methods, namely limited scale, low resolution, and lack of user control. By complementing the input image with one or a few exemplars, our approach can recover SVBRDFs of much larger surfaces, at high resolution and arbitrary aspect ratio. Furthermore, our method greatly increases the creative freedom of material designers by letting them create plausible SVBRDFs from existing photographs with high-level control on their constituent materials. We achieved all these benefits thanks to a surprisingly simple fine-tuning strategy, which we believe to be directly applicable to other capture and design tasks based on deep learning.

Acknowledgments

We thank Simon Rodriguez for his help with video editing. This work was partially funded by an ANRT(http://www.anrt.asso.fr/en) CIFRE scholarship between Inria and Optis for Ansys, ERC Advanced Grant FUNGRAPH (No. 788065, http://fungraph.inria.fr), EPSRC Early Career Fellowship (EP/N006259/1) and by software donations from Adobe. The authors are grateful to Inria Sophia Antipolis - Méditerranée "Nef" computation cluster for providing resources and support (https://wiki.inria.fr/ClustersSophia/Clusters_Home).

References

[AAB*15] ABADI M., AGARWAL A., BARHAM P., BREVDO E., CHEN Z., CITRO C., CORRADO G. S., DAVIS A., DEAN J., DEVIN M., GHEMAWAT S., GOODFELLOW I., HARP A., IRVING G., ISARD M., JIA Y.,

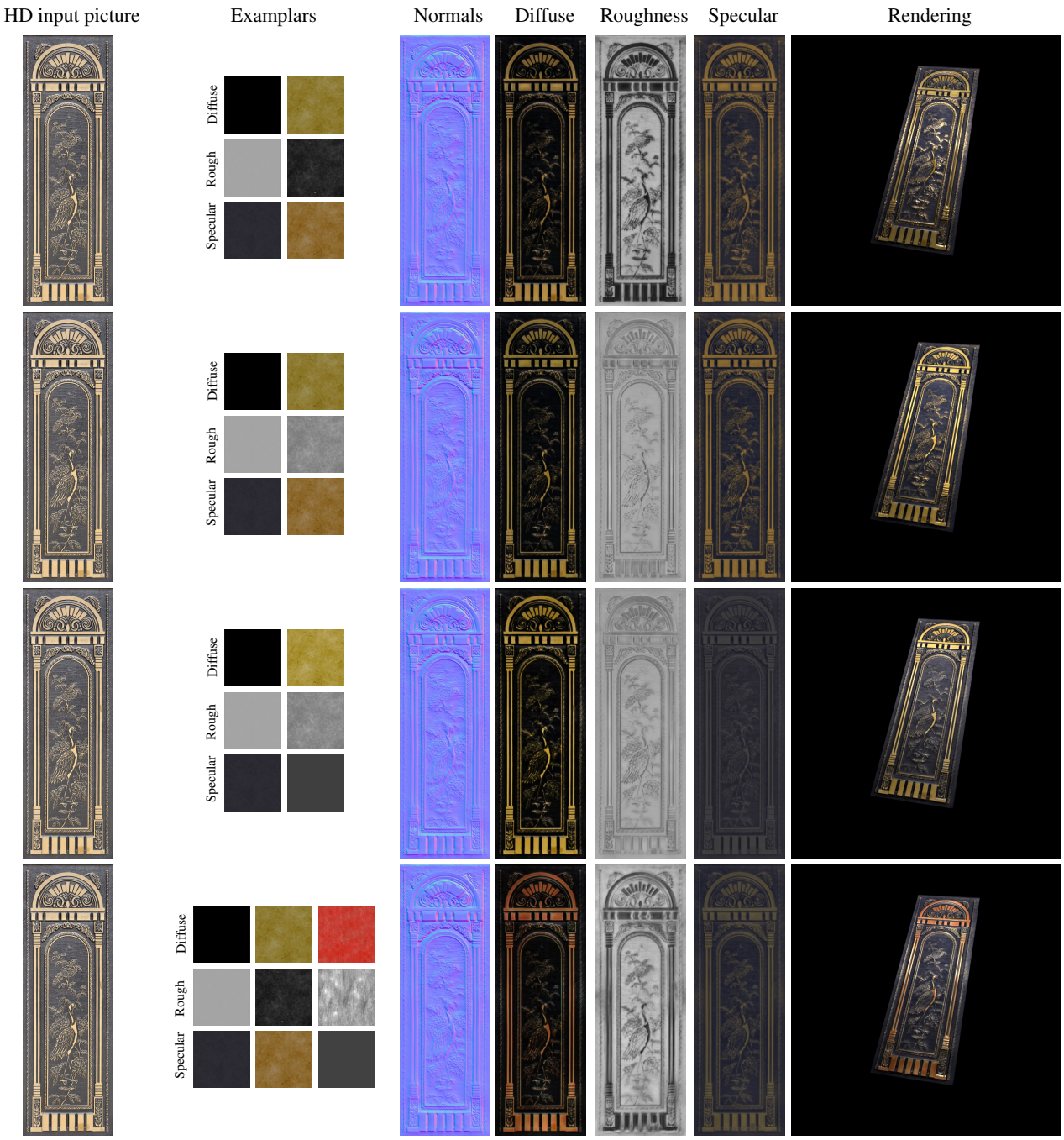

Figure 7: *Given the same input picture, we achieve different outcomes by changing the exemplars. In the first row, we provide an exemplar of a black diffuse material and an exemplar of a shiny yellow metal, which are successfully transferred to the dark and golden parts of the input picture respectively. In the second row, we increased the roughness of the yellow metallic exemplar, which is again successfully propagated to the golden parts of the input. In the third row, we replaced the metallic exemplar by a yellow diffuse material, which results in a SVBRDF where only the diffuse map contains yellow information. Finally, in the fourth row, we included an outlier red diffuse exemplar, which our method tends to mix with the yellow metal to produce a slightly orange diffuse map and a weaker specular map. Images of resolution* 512×1536.

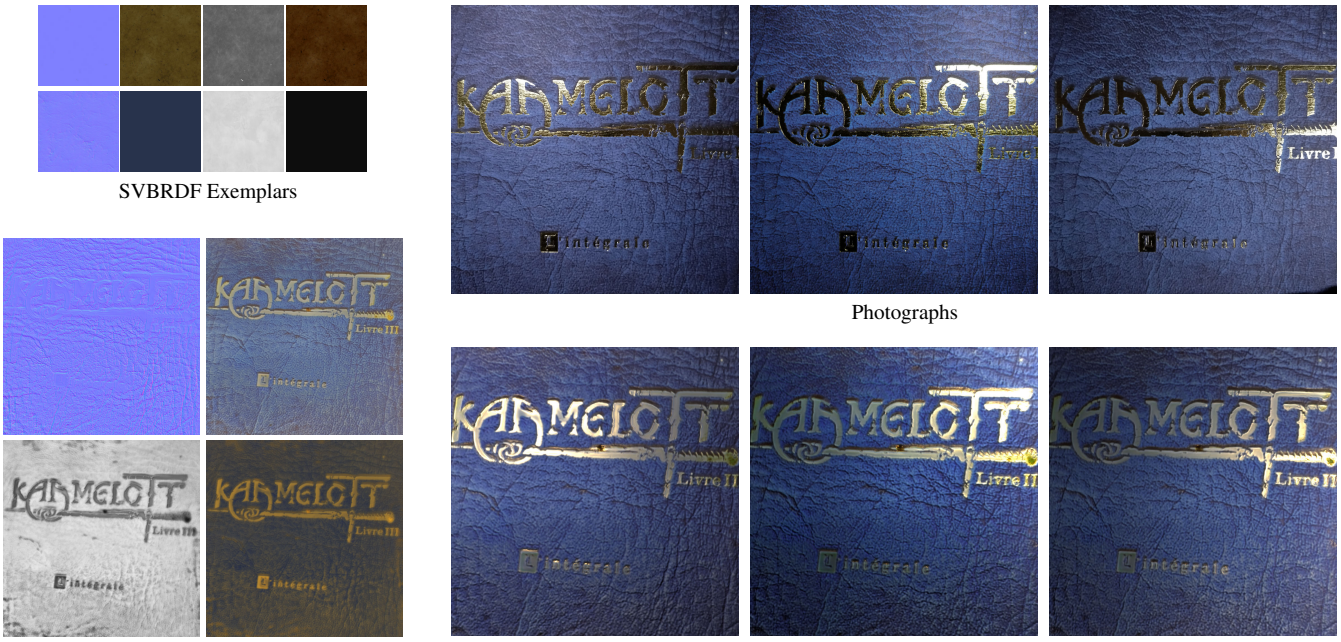

Figure 8: *Comparison to real-world photographs. We reproduced the appearance of a book cover using a single picture captured under environment lighting, and two exemplars of blue leather and golden material. The top row shows real-world pictures of the book under varying lighting, and the bottom row shows our renderings under similar lighting. A comparison with exemplars obtained with [DAD*18] and no exemplars is provided in supplemental material.*

JOZEFOWICZ R., KAISER L., KUDLUR M., LEVENBERG J., MANÉ D., MONGA R., MOORE S., MURRAY D., OLAH C., SCHUSTER M., SHLENS J., STEINER B., SUTSKEVER I., TALWAR K., TUCKER P., VANHOUCKE V., VASUDEVAN V., VIÉGAS F., VINYALS O., WARDEN P., WATTENBERG M., WICKE M., YU Y., ZHENG X.: TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org. 5

[AAL16] AITTALA M., AILA T., LEHTINEN J.: Reflectance modeling by neural texture synthesis. *ACM Transactions on Graphics (Proc. SIGGRAPH) 35*, 4 (2016). 2

[Ado19] ADOBE: Substance share, 2019. https://share.substance3d.com/. 4, 5

[AWL15] AITTALA M., WEYRICH T., LEHTINEN J.: Two-shot SVBRDF capture for stationary materials. *ACM Trans. Graph. (Proc. SIGGRAPH) 34*, 4 (July 2015), 110:1–110:13. doi:10.1145/2766967. 2

[BJTK18] BAEK S.-H., JEON D. S., TONG X., KIM M. H.: Simultaneous acquisition of polarimetric svbrdf and normals. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia) 37*, 6 (2018). 2

[BSP*19] BAU D., STROBELT H., PEEBLES W., WULFF J., ZHOU B., ZHU J., TORRALBA A.: Semantic photo manipulation with a generative image prior. *ACM Transactions on Graphics (Proc. SIGGRAPH) 38*, 4 (2019). 4

[CT82] COOK R. L., TORRANCE K. E.: A reflectance model for computer graphics. *ACM Transactions on Graphics 1*, 1 (1982), 7–24. 4

[DAD*18] DESCHAINTRE V., AITTALA M., DURAND F., DRETTAKIS G., BOUSSEAU A.: Single-image svbrdf capture with a rendering-aware deep network. *ACM Transactions on Graphics (SIGGRAPH Conference Proceedings) 37*, 128 (aug 2018), 15. 2, 3, 4, 5, 6, 7, 8, 10, 11, 13

[DAD*19] DESCHAINTRE V., AITTALA M., DURAND F., DRETTAKIS G., BOUSSEAU A.: Flexible svbrdf capture with a multi-image deep network. *Computer Graphics Forum (Proceedings of the Eurographics Symposium on Rendering) 38*, 4 (July 2019). 2, 4

[DBP*15] DIAMANTI O., BARNES C., PARIS S., SHECHTMAN E., SORKINE-HORNUNG O.: Synthesis of complex image appearance from limited exemplars. *ACM Transactions on Graphics (Proc. SIGGRAPH) 34*, 2 (2015). 3, 8

[DCP*14] DONG Y., CHEN G., PEERS P., ZHANG J., TONG X.: Appearance-from-motion: Recovering spatially varying surface reflectance under unknown lighting. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia) 33*, 6 (2014). 2

[Don19] DONG Y.: Deep appearance modeling: A survey. *Visual Informatics 3*, 2 (2019), 59 – 68. doi:https://doi.org/10.1016/j.visinf.2019.07.003. 2

[DTPG11] DONG Y., TONG X., PELLACINI F., GUO B.: Appgen: Interactive material modeling from a single image. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia) 30* (2011). 4

[DWT*10] DONG Y., WANG J., TONG X., SNYDER J., LAN Y., BEN-EZRA M., GUO B.: Manifold bootstrapping for svbrdf capture. *ACM Transactions on Graphics (Proc. SIGGRAPH) 29*, 4 (2010). 4

[FAW19] FRÜHSTÜCK A., ALHASHIM I., WONKA P.: TileGAN: Synthesis of large-scale non-homogeneous textures. *ACM Transactions on Graphics (Proc. SIGGRAPH) 38*, 4 (2019), 58:1–58:11. 4

[FJL*16] FIŠER J., JAMRIŠKA O., LUKÁČ M., SHECHTMAN E., ASENTE P., LU J., SÝKORA D.: StyLit: Illumination-guided example-based stylization of 3d renderings. *ACM Transactions on Graphics (proc. SIGGRAPH) 35*, 4 (2016). 3, 6, 12

[GGG*16] GUARNERA D., GUARNERA G. C., GHOSH A., DENK C., GLENCROSS M.: BRDF Representation and Acquisition. *Computer Graphics Forum* (2016). 2

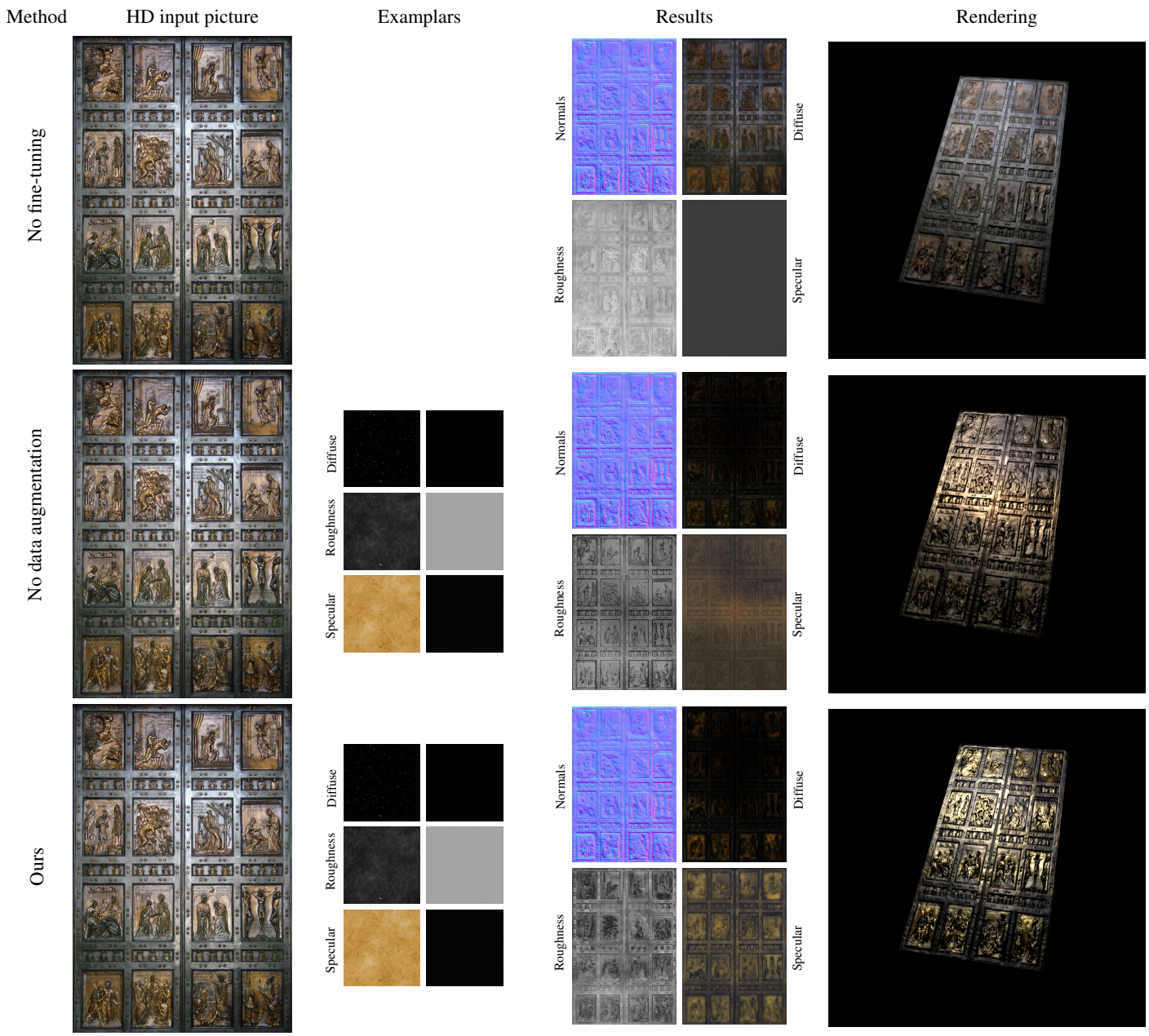

Figure 9: *Ablation study. The baseline single-image network of Deschaintre et al. [DAD*18] interprets this weathered golden door as made of rough plastic (first row). Fine-tuning this network on two exemplars without data augmentation yields a uniform golden appearance (second row). Thanks to data augmentation, our method successfully distinguishes the shiny golden parts from the more diffuse dark parts (third row). See supplemental materials for additional ablation results. Image of resolution* 1024×1536.

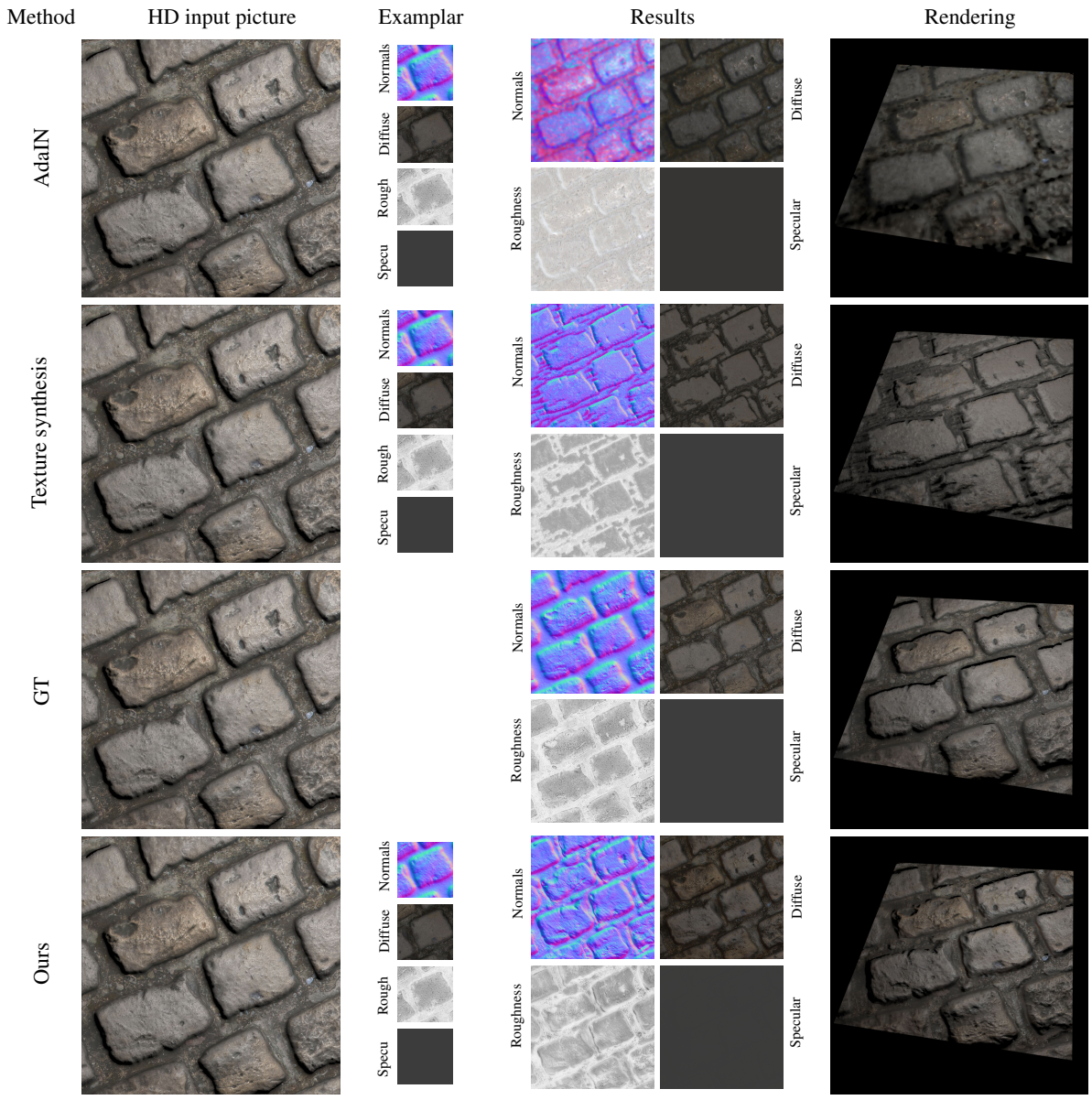

Figure 10: *Comparison to neural style transfer [HB17] and patch-based texture synthesis [FJL*16]. Our method better transfers details of the surface compared to prior work, which either only captures global statistics (1st row) or struggles to generalize from a limited exemplar (2nd row).*

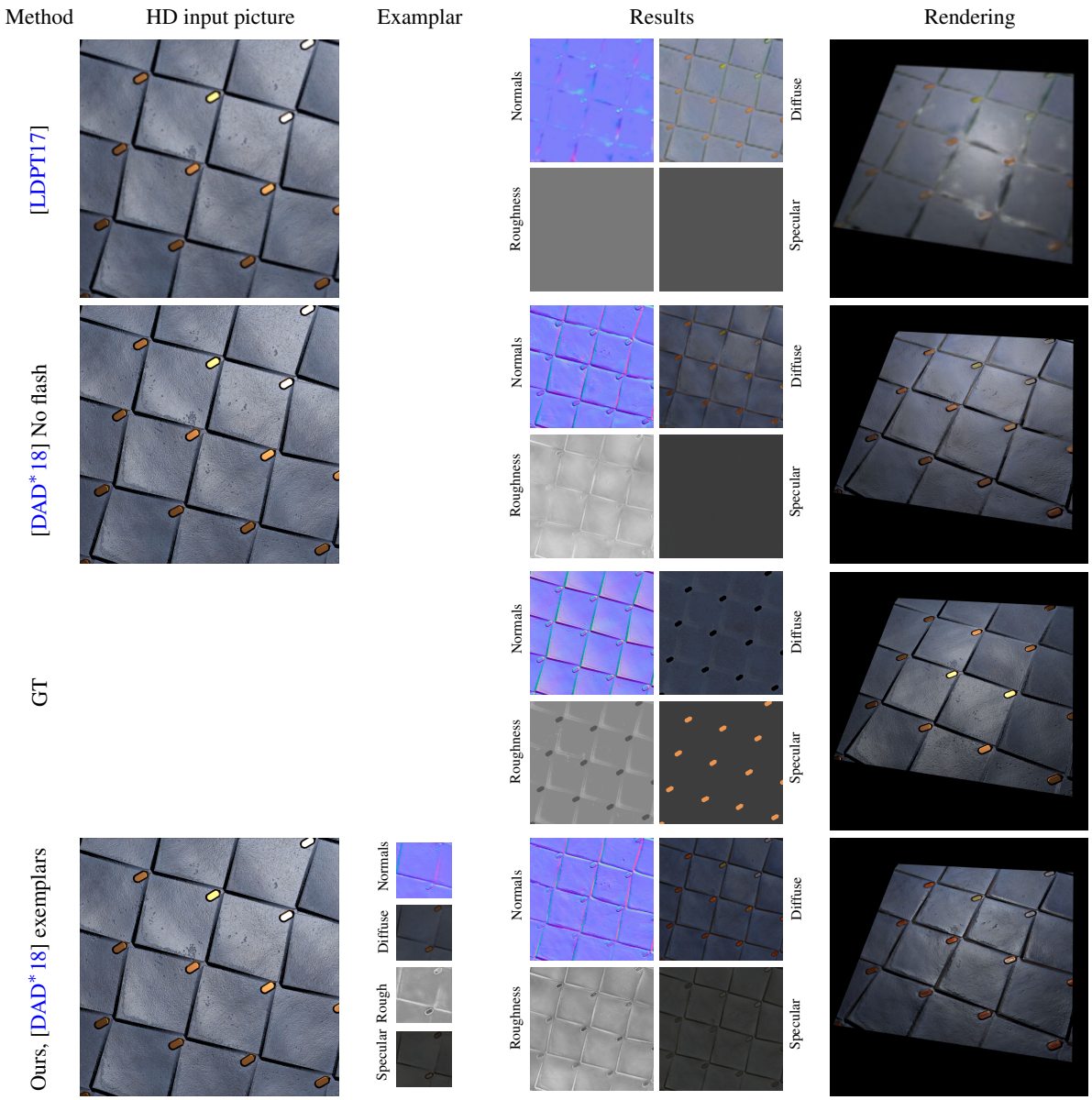

Figure 11: *Comparison to the single-image methods of [LDPT17] and [DAD*18]. Thanks to a small exemplar, our method recovers more pronounced normal maps than the one by [LDPT17], and also better captures the roughness of the small shiny metal plates, even though their specular strength remains underestimated. Also, since our method can process high-resolution images, it recovers finer details in the maps. Note that Li et al. use a different BRDF model than ours, so the values of their predicted maps shouldn't be directly compared to the ground truth maps.*

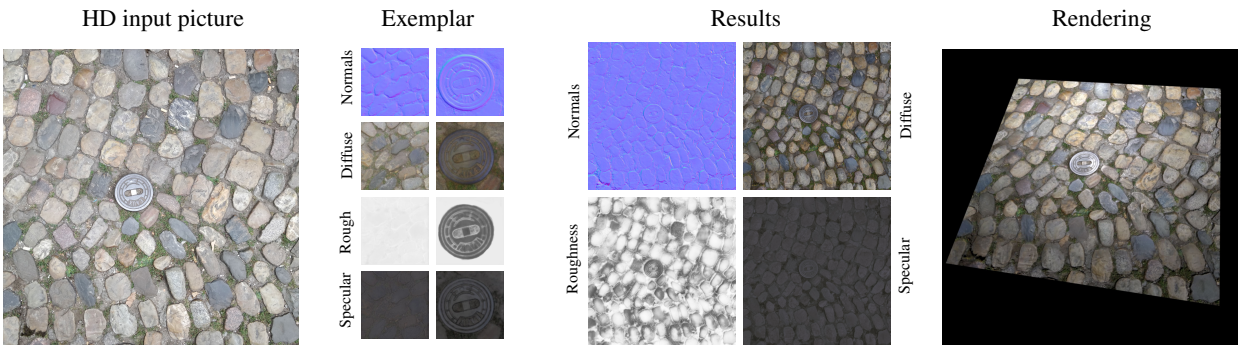

Figure 12: *Limitation. Our method can have difficulty distinguishing materials with similar colors and texture, such as this shiny metal disk that has a similar appearance to some of the dark rough stones.*

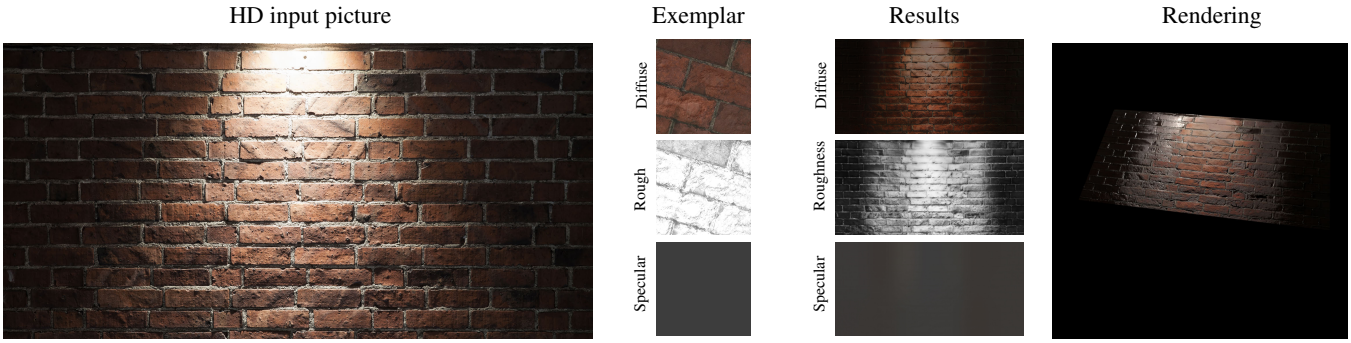

Figure 13: *Limitations. Our method is not designed to handle large illumination gradients over the surface.*

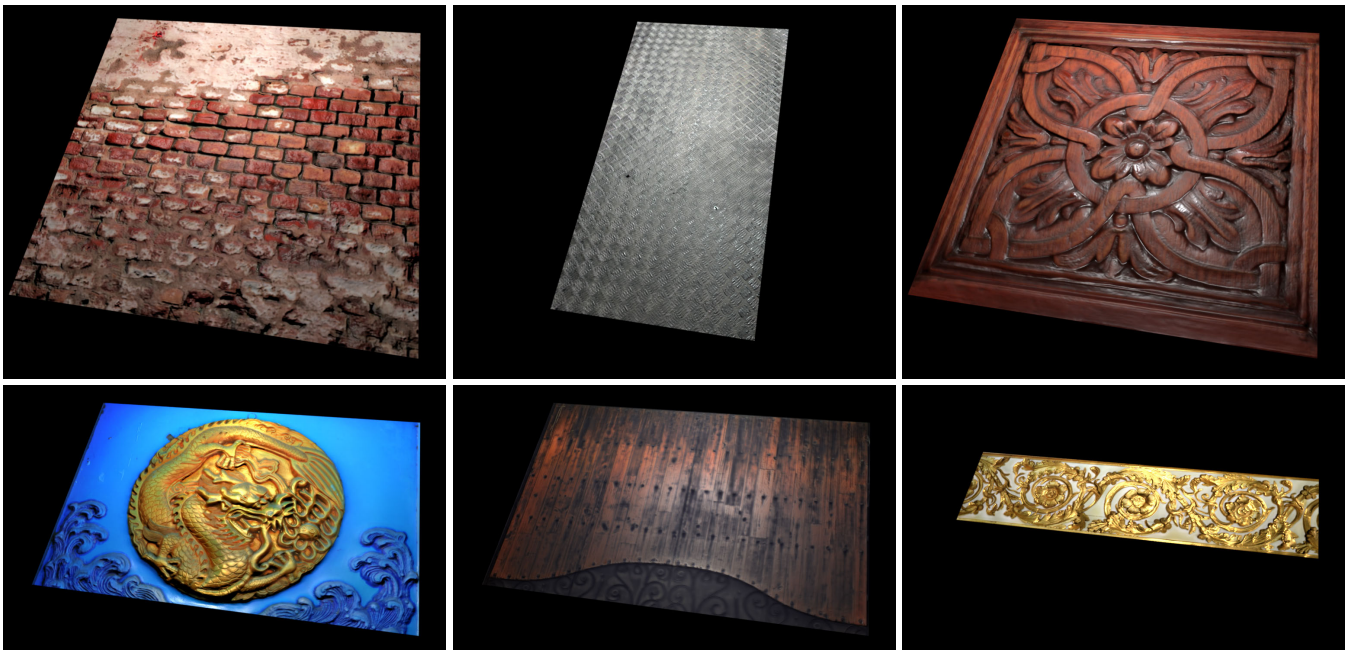

Figure 14: *A variety of surfaces captured or designed with our method. See supplemental materials for animated renderings.*

[GLD*19] 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. 38*, 4 (July 2019), 134:1–134:15. doi:10.1145/3306346.3323042. 2, 4, 8

[HB17] HUANG X., BELONGIE S.: Arbitrary style transfer in real-time with adaptive instance normalization. In *ICCV* (2017). 8, 12

[HCL*18] HE M., CHEN D., LIAO J., SANDER P. V., YUAN L.: Deep exemplar-based colorization. *ACM Transactions on Graphics (Proc. SIGGRAPH) 37*, 4 (2018). 3

[HJO*01] HERTZMANN A., JACOBS C. E., OLIVER N., CURLESS B., SALESIN D. H.: Image analogies. *ACM SIGGRAPH* (2001). 3, 6

[HLC*19] HE M., LIAO J., CHEN D., YUAN L., SANDER P. V.: Progressive color transfer with dense semantic correspondences. *ACM Trans. Graph. 38*, 2 (Apr. 2019). doi:10.1145/3292482. 3

[HSL*17] HUI Z., SUNKAVALLI K., LEE J. Y., HADAP S., WANG J., SANKARANARAYANAN A. C.: Reflectance capture using univariate sampling of brdfs. In *IEEE International Conference on Computer Vision (ICCV)* (2017). 2

[IZZE17] ISOLA P., ZHU J.-Y., ZHOU T., EFROS A. A.: Image-to-image translation with conditional adversarial networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (July 2017). 3

[LDPT17] LI X., DONG Y., PEERS P., TONG X.: Modeling surface appearance from a single photograph using self-augmented convolutional neural networks. *ACM Transactions on Graphics (Proc. SIGGRAPH) 36*, 4 (2017). 3, 7, 8, 13

[LHM*19] LIU M.-Y., HUANG X., MALLYA A., KARRAS T., AILA T., LEHTINEN J., KAUTZ J.: Few-shot unsupervised image-to-image translation. In *The IEEE International Conference on Computer Vision (ICCV)* (October 2019). 4, 8

[LSC18] LI Z., SUNKAVALLI K., CHANDRAKER M.: Materials for masses: SVBRDF acquisition with a single mobile phone image. *Proceedings of ECCV* (2018). 2, 4, 8

[LXR*18] LI Z., XU Z., RAMAMOORTHI R., SUNKAVALLI K., CHANDRAKER M.: Learning to reconstruct shape and spatially-varying reflectance from a single image. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)* (2018). 2

[LYY*17] LIAO J., YAO Y., YUAN L., HUA G., KANG S. B.: Visual attribute transfer through deep image analogy. *ACM Transactions on Graphics (Proc. SIGGRAPH) 36*, 4 (2017). 3

[MGSJW12] MELENDEZ F., GLENCROSS M., STARCK J., J. WARD G.: Transfer of albedo and local depth variation to photo-textures. pp. 40–48. doi:10.1145/2414688.2414694. 3, 6

[NLGK18] NAM G., LEE J. H., GUTIERREZ D., KIM M. H.: Practical svbrdf acquisition of 3d objects with unstructured flash photography. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia) 37*, 6 (2018). 2

[RPG16] RIVIERE J., PEERS P., GHOSH A.: Mobile surface reflectometry. *Computer Graphics Forum 35*, 1 (2016). 2

[RRFG17] RIVIERE J., RESHETOUSKI I., FILIPI L., GHOSH A.: Polarization imaging reflectometry in the wild. *ACM Transactions on Graphics (Proc. SIGGRAPH)* (2017). 2

[RWS*11] REN P., WANG J., SNYDER J., TONG X., GUO B.: Pocket reflectometry. *ACM Transactions on Graphics (Proc. SIGGRAPH) 30*, 4 (2011). 2

[SCI18] SHOCHER A., COHEN N., IRANI M.: "zero-shot" super-resolution using deep internal learning. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2018). 4

[SDM19] SHAHAM T. R., DEKEL T., MICHAELI T.: Singan: Learning a generative model from a single natural image. In *The IEEE International Conference on Computer Vision (ICCV)* (October 2019). 4

[Str19] STRUFFELPRODUCTIONS: Cc0textures, 2019. https://cc0textures.com/. 4, 5

[TFK*20] TEXLER O., FUTSCHIK D., KUČERA M., JAMRIŠKA O., SOCHOROVÁ R., CHAI M., TULYAKOV S., SÝKORA D.: Interactive video stylization using few-shot patch-based training. *ACM Transactions on Graphics (Proc. SIGGRAPH) 39*, 4 (2020). 3

[UVL18] ULYANOV D., VEDALDI A., LEMPITSKY V.: Deep image prior. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2018). 4

[WAM02] WELSH T., ASHIKHMIN M., MUELLER K.: Transferring color to greyscale images. *ACM Transactions on Graphics (Proc. SIGGRAPH) 21*, 3 (2002). 3

[WLZ*18] WANG T., LIU M., ZHU J., TAO A., KAUTZ J., CATANZARO B.: High-resolution image synthesis and semantic manipulation with conditional gans. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), pp. 8798–8807. 3

[WSM11] WANG C.-P., SNAVELY N., MARSCHNER S.: Estimating dual-scale properties of glossy surfaces from step-edge lighting. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia) 30*, 6 (2011). 2

[YLD*18] YE W., LI X., DONG Y., PEERS P., TONG X.: Single image surface appearance modeling with self-augmented cnns and inexact supervision. *Computer Graphics Forum 37*, 7 (2018), 201–211. 3

[ZPIE17] ZHU J.-Y., PARK T., ISOLA P., EFROS A. A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on* (2017). 3