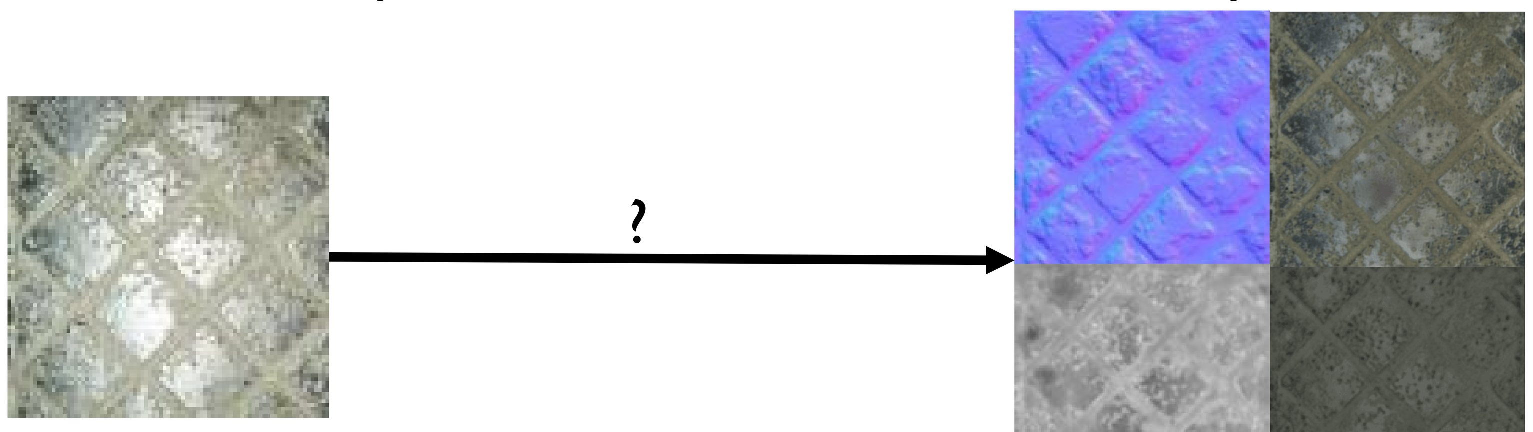
Material acquisition using deep learning

UNIVERSITÉ CÔTE D'AZUR · Inria · ANSYS OPTIS

Valentin Deschaintre

Material representation from a few pictures



Input picture

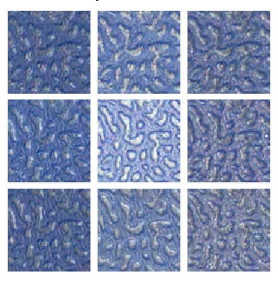
Cook-Torrance Reflectance parameters decomposition

The appearance of real-world objects results from complex interactions between light, reflectance, and geometry. Disentangling these interactions is at the heart of lightweight appearance capture, which aims at recovering reflectance functions from one or a few photographs of a surface.
Our goal is to acquire a material appearance from as little as one image while having the possibility to improve the results with more photographs.

Classical strategy is to introduce a priori assumptions about the space of plausible solutions:

- Use of video feed
- Target is made of a set of known materials

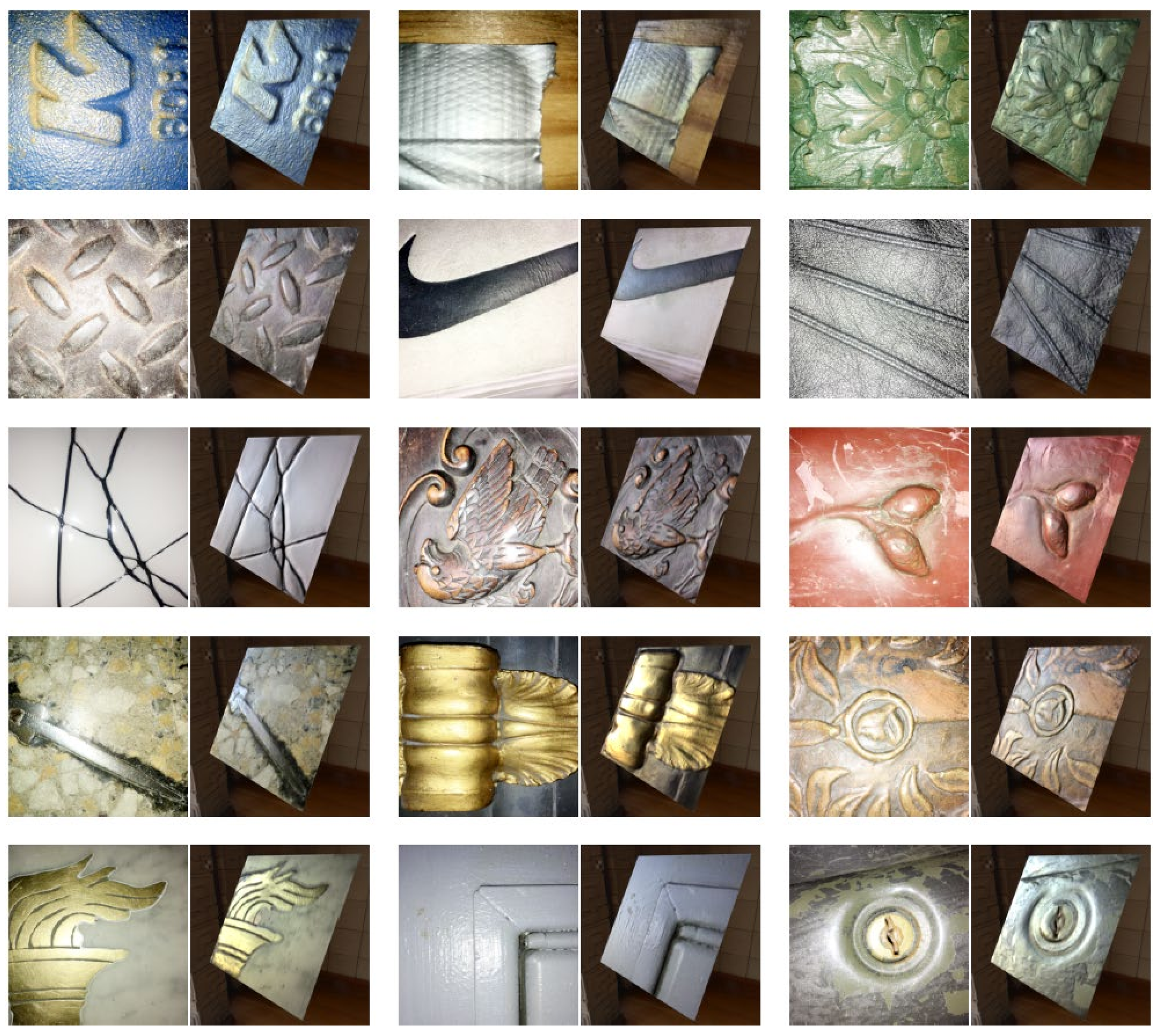
Ren et al., Pocket Reflectometry, 2011



- 2 images (flash and no flash)
- Target material is self-repetitive

Aittala et al., Two-shot SVBRDF Capture for Stationary Materials, 2015

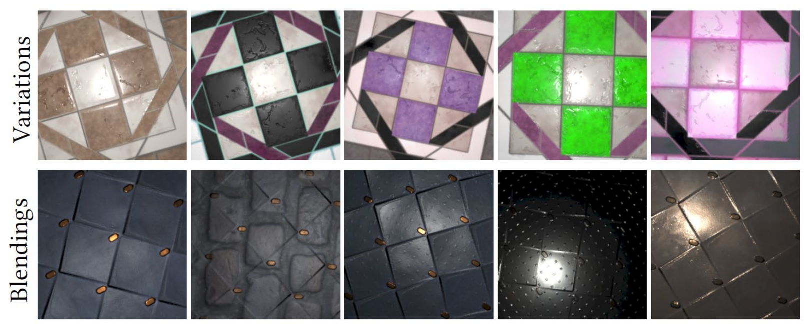
Results



A selection of results from our one image method on real-world photographs. In each image pair, the left image is a photograph of a surface, and the right image is a re-rendering with environment lighting of the material solved from that image.

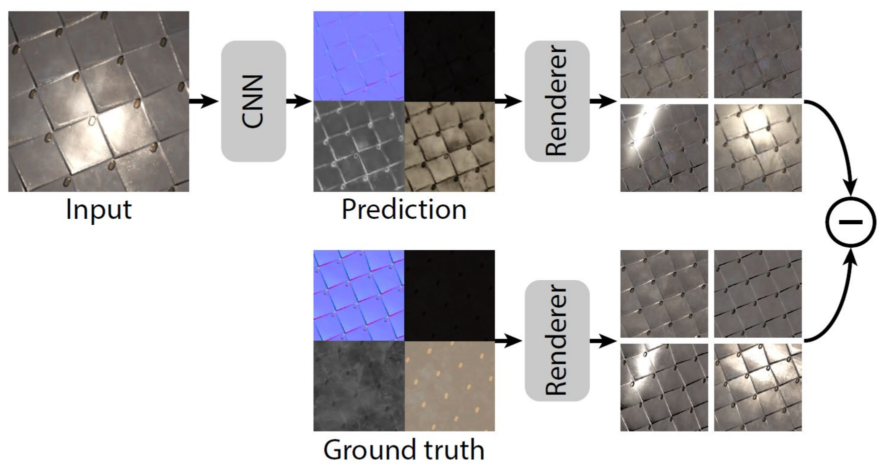
One image acquisition

Data generation



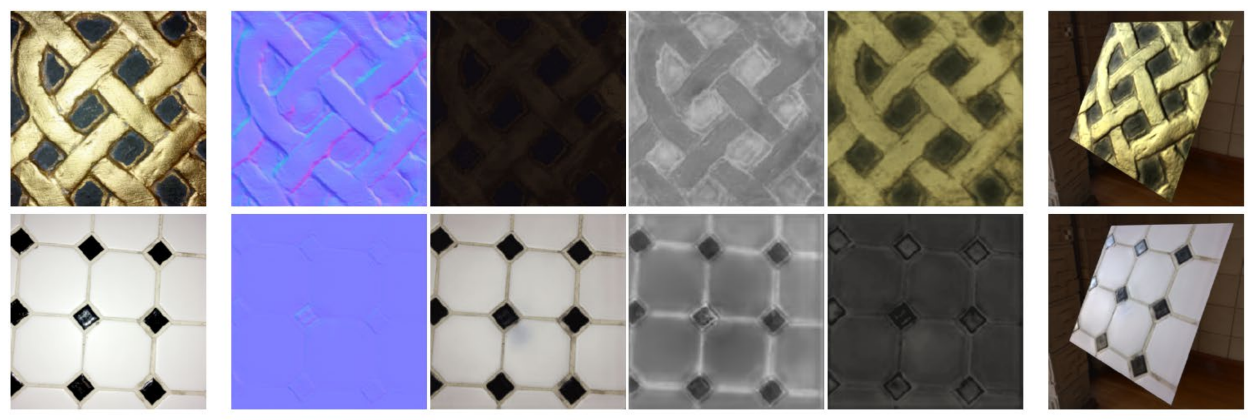
Supervised learning methods require large amount of data for training.
Given the lack of real world SVBRDF/Picture pairs database, we turn to procedural generation and parameters blending to generate our data on the fly during the training.

Training

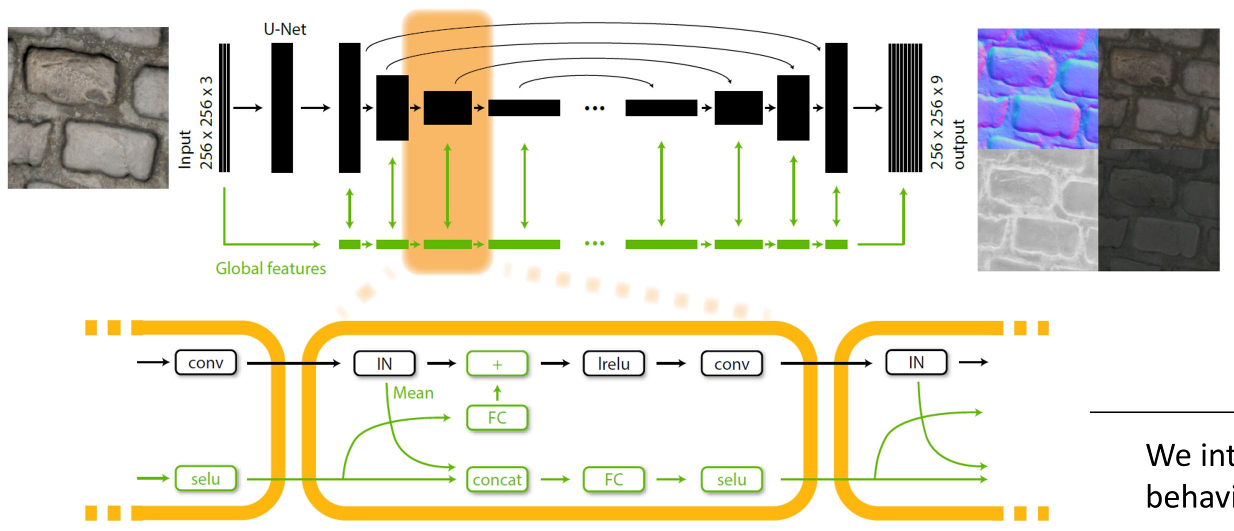


We leverage differentiable rendering algorithms to compare multiple lighting conditions rather than directly comparing the parameters maps. We are therefore independent of the parametrization and focus on the rendered appearance rather than individual parameters.

Inferred



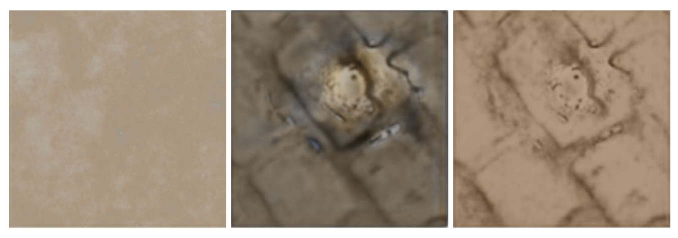
Using a single picture as input, our method produces 4 parameter maps corresponding to the Cook-Torrance model.



We use a classical Unet[1] architecture for image-to-maps transformation.
The input is taken through multiple scale convolutions to extract the most relevant information.
"Skip connections" are then used to provide spatial information and details to complete the extracted features.

[1] *Unet : O. Ronneberger, P.Fischer, and T. Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Medical Image Computing and Computer-Assisted Intervention (MICCAI) (LNCS),*

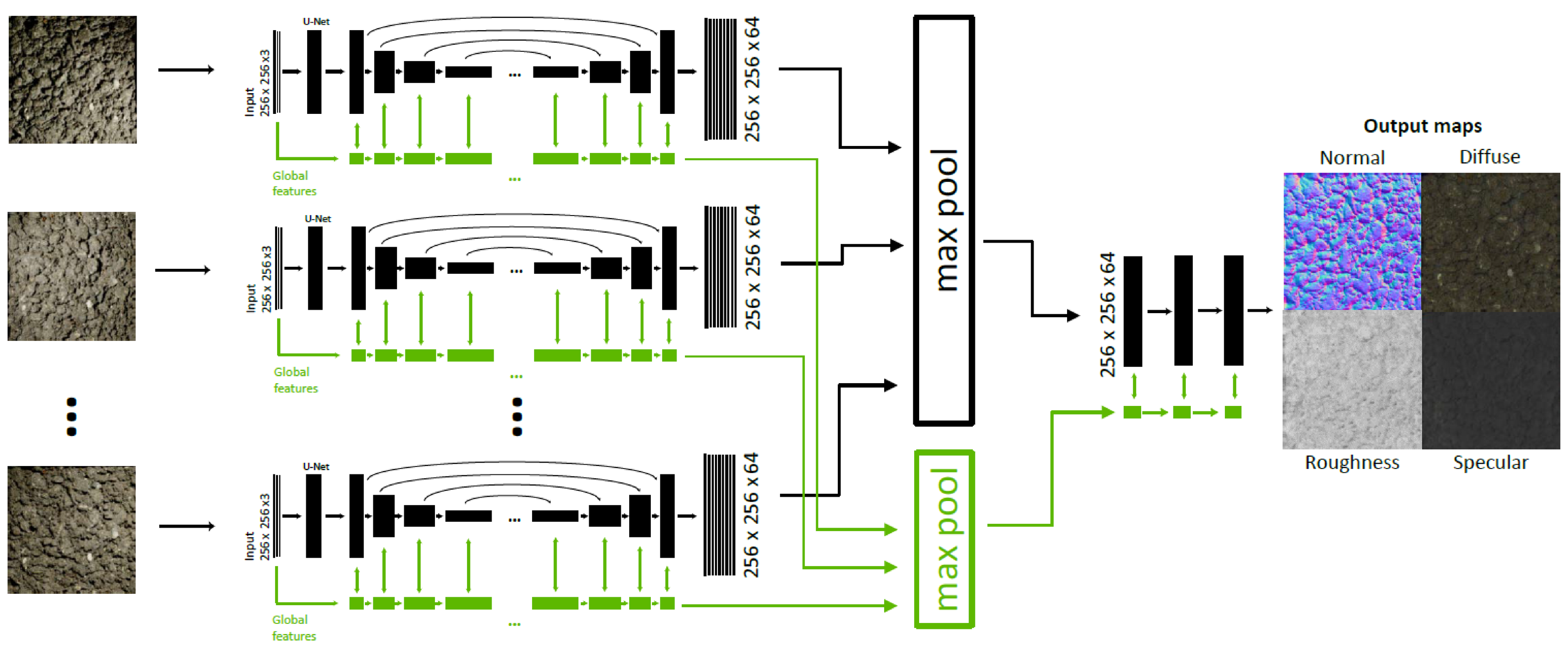
We introduce a global feature network providing more information about the global behaviour of the material to the network .



Inferred specular map without (a) and with (b) global feature network

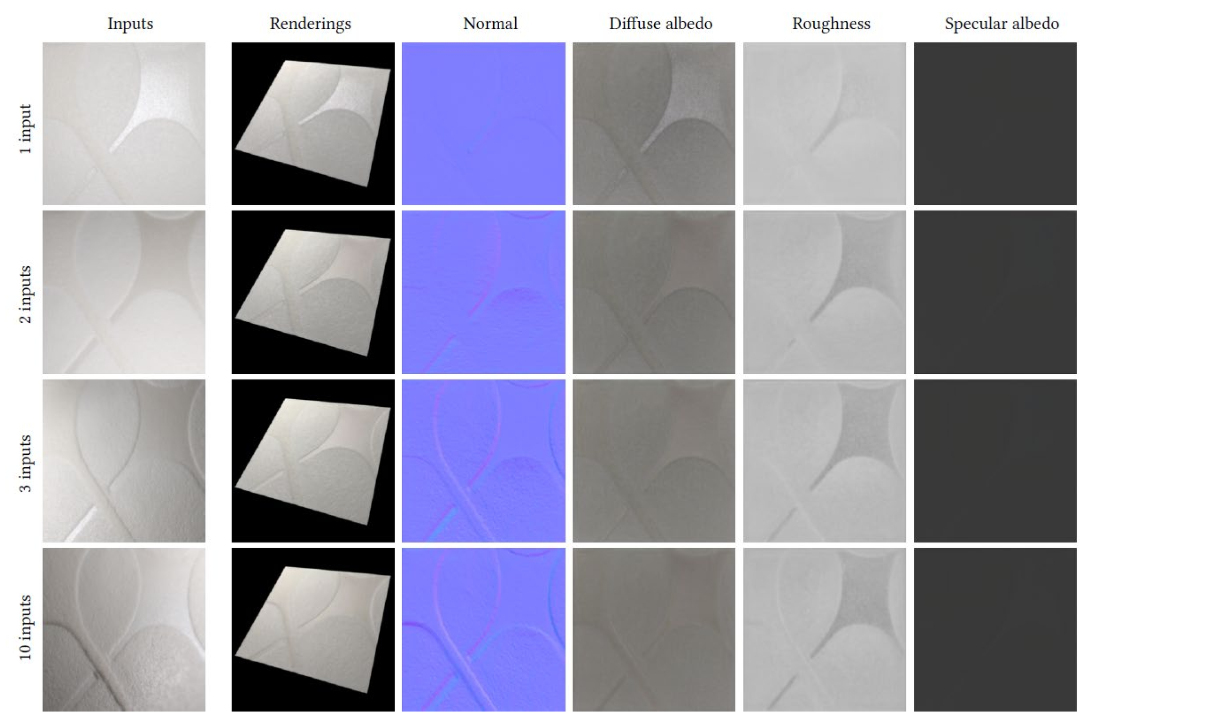
Multi image acquisition

Architecture



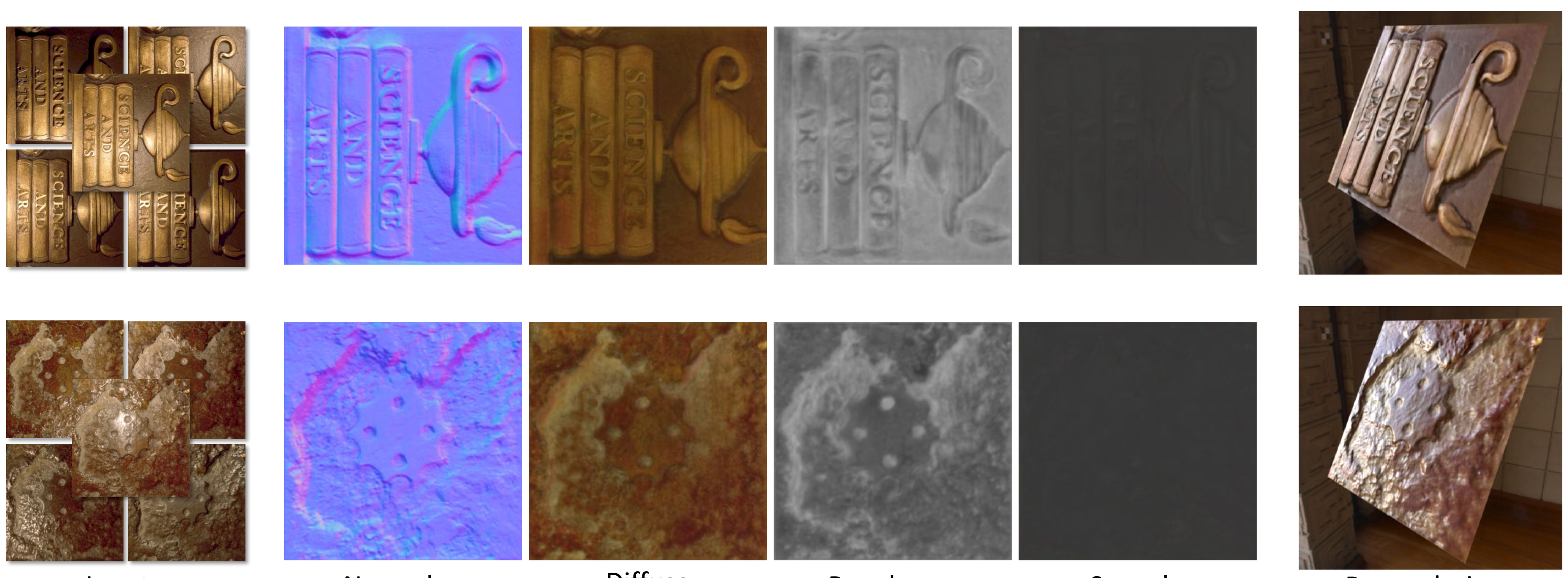
Each input image is processed by a copy of the one image method encoder-decoder to produce a feature map. While the number of images and network copies can vary, a pooling layer fuses the output maps to obtain a fixed-size representation of the material, which is then processed by a few convolutional layers to produce the SVBRDF maps.
With this architecture, the method is able to handle an arbitrary number of inputs from any light and view directions.

Evaluation



Evaluation on a measured BTF. Three images are enough to capture most of normal and roughness maps. Adding images further improves the result by removing lighting residual from the diffuse albedo, and adding subtle details to the normal and specular maps

Multi image results



Inputs · Normals · Diffuse · Roughness · Specular · Re-rendering

Conclusion

This poster is based on my thesis work[2, 3] in which I present a number of contributions to the lightweight material acquisition problem.
We introduce two network architectures allowing to conveniently capture real world materials with as little as one image, while being able to leverage more information if available.

Several key contributions were necessary to achieve high quality SVRBDF estimation:

- We build a flexible synthetic training data generation pipeline, compensating for the lack of real world paired SVBRDF dataset.
- The rendering loss focuses on the final appearance of materials to compare, rather than individual SVBRF parameters.
- Our global feature network allows to propagate global information at different scale throughout our main Unet network.
- We design an architecture capable of aggregating the information from an arbitrary number of photographs from any light and view direction.

[2] Valentin Deschaintre, Miika Aittala, Fredo Durand, George Drettakis, and Adrien Bousseau. Single-image svbrdf capture with a rendering-aware deep network. ACM Transactions on Graphics (SIGGRAPH Conference Proceedings), 37(128):15, aug 2018
[3] Valentin Deschaintre, Miika Aittala, Fredo Durand, George Drettakis, and Adrien Bousseau. Flexible SVBRDF capture with a multi-image deep network Computer Graphics Forum, 38 (4) (Jul 2019)