MatFusion: A Generative Diffusion Model for SVBRDF Capture –Supplemental Material–

Sam Sartor
College of William & Mary
Williamsburg, USA
slsartor@wm.edu

Pieter Peers
College of William & Mary
Williamsburg, USA
ppeers@siggraph.org

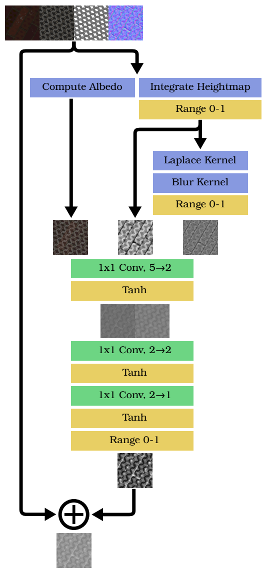

Figure 1: Network architecture used for the roughness augmentation.

ACM Reference Format:
Sam Sartor and Pieter Peers. 2023. MatFusion: A Generative Diffusion Model for SVBRDF Capture –Supplemental Material–. In *SIGGRAPH Asia 2023 Conference Papers (SA Conference Papers '23), December 12–15, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3610548.3618194

1 ROUGHNESS AUGMENTATION

As noted in the main document, we enrich the distribution of roughness maps by replacing the maps for a randomly selected subset of 20% of the 30,032 basis exemplars (i.e., ~6,000 training exemplars) based on the height map (normalized to [0..1]) and the (diffuse + specular) albedo map. We concatenate these four channels (height and RGB color) together with a Gaussian blurred ($\sigma = 3.0$) Laplacian of the height map. These 5 channels are fed into a randomly initialized dense neural network. This network consists of three dense layers, with tanh activations, that reduce the input channels from 5 to 2 and then to 1 output channel. See Figure 1 for a summary of the network architecture. As noted in the main document, we do not optimize the dense network, and use it as a random nonlinear transformation of height and albedo into roughness. Finally, we blend the normalized output R of the network and the material's original roughness map r as: $\sqrt{wR^2 + (1-w)r^2)}$, with $w = 0.3 + 0.6R$.

SA Conference Papers '23, December 12–15, 2023, Sydney, NSW, Australia

ACM ISBN 979-8-4007-0315-7/23/12…$15.00
https://doi.org/10.1145/3610548.3618194

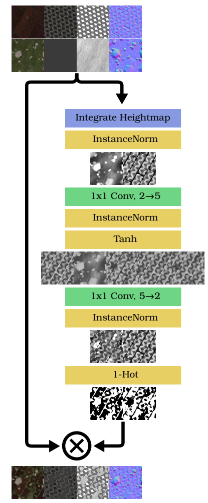

Figure 2: Network architecture used for the mixture augmentation.

2 MIXTURE AUGMENTATION

Many real-world materials are a piece-wise constant mixture of different materials. To further diversify the training set, we compute mixtures of the INRIA SVBRDFs and our basis SVBRDFs. For 2/3 of the cases we mix two materials, and for the remaining 1/3 we mix three materials. Each SVBRDF in the INRIA training set and our basis SVBRDFs is only used in a single mixture model. We leverage a randomly initialized dense neural network to compute the mixture weights (Figure 2 summarizes the architecture). This network takes as input, the diffuse+specular albedo maps bilinearly upsampled by a factor 2, concatenated with the corresponding height maps computed from the normal maps using the fast Poisson FFT method from Quéau et al. [2018]. We first apply an instance norm on the input vector, followed by two dense layers that keep the number of channels constant, each followed by another instance norm and activation (i.e., tanh and one-hot activation respectively for the 1st and 2nd dense layer). Finally, we downsample the output by a factor 2 yielding the final mixture weight maps.

3 TEST SET

Figure 3 shows all 50 materials from our test set. 31 are from the Deep Inverse Rendering [Gao et al. 2019] test set (marked green; this test set also includes SVBRDFs from the original INRIA [Deschaintre et al. 2018] test set), 6 are from Zhou and Kalantari's look-ahead method [Zhou and Kalantari 2022] (marked red), 11 from https://polyhaven.com (marked blue), and 2 from https://ambientcg.com (marked purple).

4 ADDITIONAL COMPARISONS ON "IN-THE-WILD" CAPTURED MATERIALS

Figure 4 and Figure 5 show additional qualitative comparisons of COLOCATED MatFusion and the adversarial network of Zhou and Kalantari [2021] and the look-ahead method of Zhou and Kalantari [2022] on *captured photographs*. These photographs do not contain reference SVBRDF maps or reference images under different lighting. Even for a human it is sometimes difficult to gauge what the material would exactly look like. Hence, these results are included to better understand how each method handles these difficult materials. From these results, we can make the following observations regarding the different recovered property maps:

Diffuse Albedo. A well-known problem with prior methods is the "burn-in" of specular reflections visible in the input photograph in the diffuse albedo (e.g., Figure 5, 1st, 2nd, and last example). In contrast, by virtue of being a generative model, MatFusion tends to produce much less "burn-in" and the resulting diffuse albedo maps are more clean.

Specular Albedo. We observe that many specular albedo maps produced by MatFusion are monochromatic (although not all). This is to be expected as most exemplars are composed of dielectric materials which produce white specular reflections. Exceptions are: Figure 4, 1st example exhibits non-white reflection on the metal part; the manual variant of the 1st example in Figure 5 exhibits a slight pink tone; and the render error variant of the last example in Figure 5 has a slight greenish tone. Furthermore, MatFusion tends to produce clean specular albedo maps that exhibit less correlation with the texture in the diffuse albedo maps. In contrast, the competing methods often correlates the texture in both albedo maps to correct errors in the normal maps (e.g., Figure 4, 2nd example, and Figure 5, 4th example), or, as noted before, bakes in the specular reflection in the diffuse albedo map.

Specular Roughness. Similar to specular albedo, prior methods correlate roughness strongly with diffuse albedo as well as exhibit "burn-in" (in fact, "burn-in" is more prevalent in the roughness maps than in the specular albedo maps). The correlation of diffuse texture and roughness is indicative that the networks are trying to correct shortcomings in other maps (mostly incorrect normals). While not free of correlation between the diffuse and roughness maps, the correlations are greatly reduced in our MatFusion results. Figure 5, 1st example, is a strong example of the difference between prior work and MatFusion on the "cleanliness" of the specular roughness parameters.

Normals. MatFusion produces better defined normals with more details (e.g., Figure 4, last example, and Figure 5, 1st, 3rd, and 4th example). Prior work often tends to display a non-uniform distribution of normal detail; more details in non-oversaturated specular reflections, and less detail in the diffuse-only observations. Because of the generative nature of MatFusion, a more homogeneously detailed normal map is synthesized.

REFERENCES

Valentin Deschaintre, Miika Aittala, Frédo Durand, George Drettakis, and Adrien Bousseau. 2018. Single-image SVBRDF capture with a rendering-aware deep network. *ACM Trans. Graph.* 37, 4 (2018).

Duan Gao, Xiao Li, Yue Dong, Pieter Peers, Kun Xu, and Xin Tong. 2019. Deep inverse rendering for high-resolution SVBRDF estimation from an arbitrary number of images. *ACM Trans. Graph.* 38, 4 (2019).

Yvain Quéau, Jean-Denis Durou, and Jean-Franccois Aujol. 2018. Normal Integration: A Survey. *Journal of Mathematical Imaging and Vision* 60, 4 (May 2018), 576–593.

Xilong Zhou and Nima Khademi Kalantari. 2021. Adversarial Single-Image SVBRDF Estimation with Hybrid Training. *Comp. Graph. Forum* (2021).

Xilong Zhou and Nima Khademi Kalantari. 2022. Look-Ahead Training with Learned Reflectance Loss for Single-Image SVBRDF Estimation. *ACM Trans. Graph.* 41, 6, Article 266 (nov 2022).

Figure 3: Test set. The materials are selected from the test sets of: Deep Inverse Rendering [Gao et al. 2019] (green), Zhou and Kalantari's [2022] look-ahead method (red), Polyhaven (blue), and AmbientCG (purple).

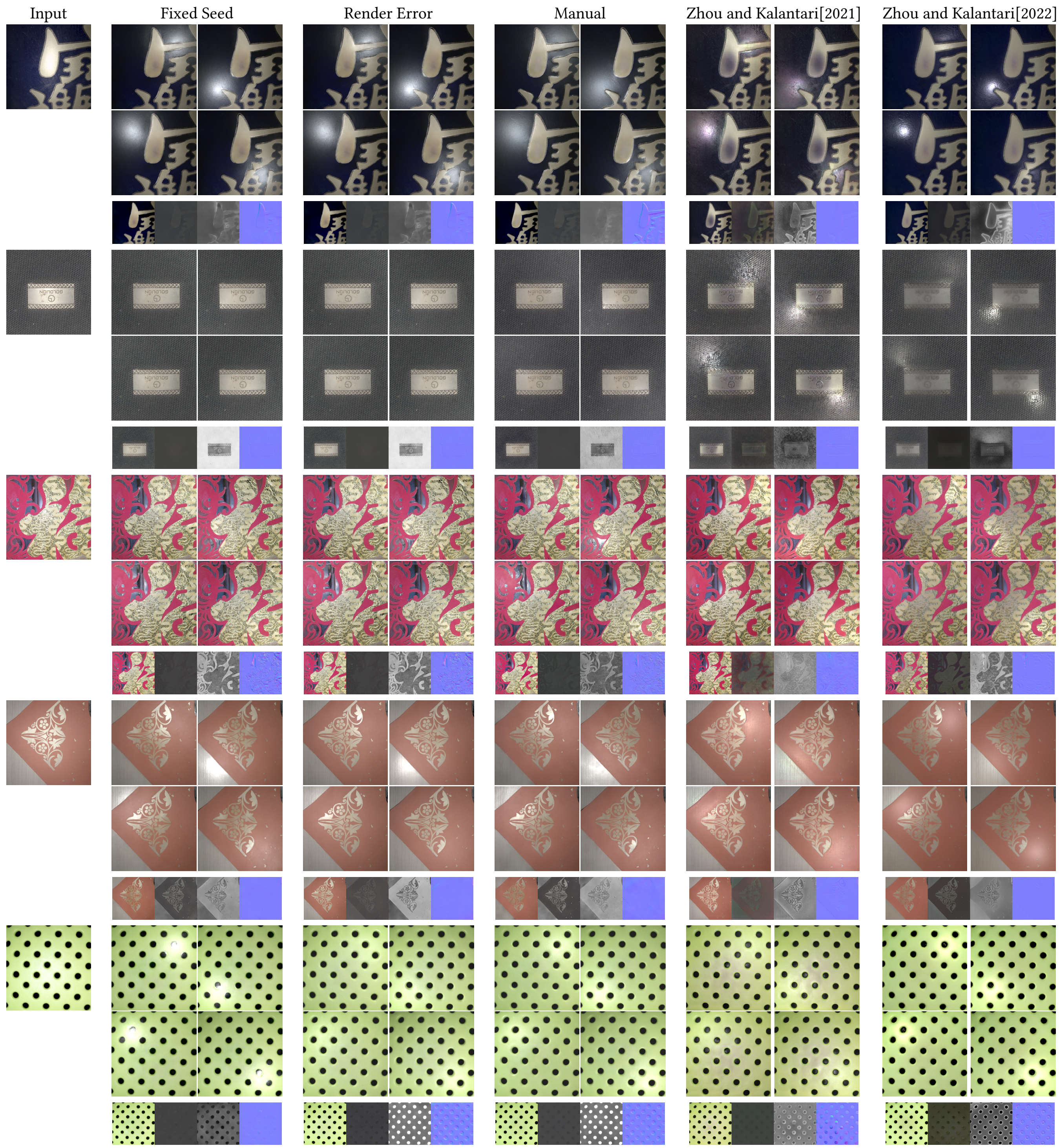

Figure 4: Qualitative comparison of MatFusion conditioned on colocated lighting (*fixed seed, render error,* and *manual* selection) against the adversarial direct inference of Zhou and Kalantari [2021] and the meta-leanring look-ahead method of Zhou and Kalantari [2022].

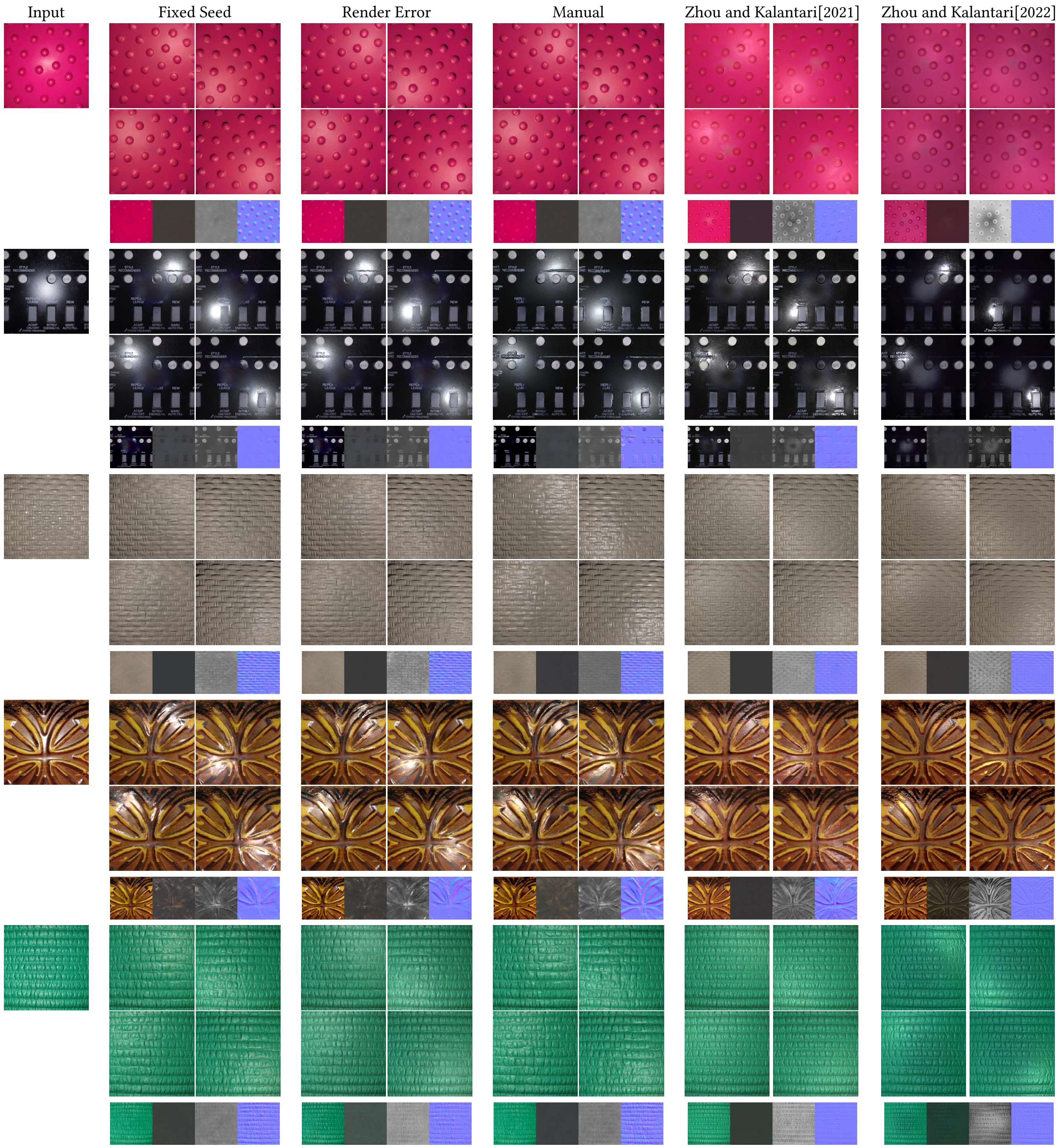

Figure 5: More Qualitative comparisons of MatFusion conditioned on colocated lighting (*fixed seed*, *render error*, and *manual* selection) against the adversarial direct inference of Zhou and Kalantari [2021] and the meta-leanring look-ahead method of Zhou and Kalantari [2022].