Photorealistic Material Editing Through Direct Image Manipulation

Supplementary material

KÁROLY ZSOLNAI-FEHÉR, TU Wien
PETER WONKA, KAUST
MICHAEL WIMMER, TU Wien

ACM Reference Format:
Károly Zsolnai-Fehér, Peter Wonka, and Michael Wimmer. 2019. Photorealistic Material Editing Through Direct Image Manipulation: Supplementary material. 1, 1 (September 2019), 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 QUESTIONS & ANSWERS

Q: What training data was used for the neural renderer and the inversion network(s)?
A: The training set for the neural renderer is equivalent to the one used in Gaussian Material Synthesis [Zsolnai-Fehér et al. 2018]. Our inversion networks are formulated as the adjoint of this neural renderer, and hence, one of their key advantage is that they can be trained on the same dataset by swapping the inputs and outputs and applying the appropriate architectural changes discussed in the paper.

Q: What image editing operations are supported for the inputs?
A: To demonstrate the usefulness of our system, we endeavored to showcase a comprehensive set of creative operations, e.g., image stitching, colorization, changing the color balance, hue, image inpainting, grayscale transform, contrast enhancement, selectively blurring the specular highlights by hand, interpolation or mixing between two images and more. However, our goal is to be able to infer shader setups for inputs that stray outside our training set in many possible directions – therefore, we encourage artists to experiment with our system and come up with creative ideas beyond these transforms.

Q: Who should use this work, and who should use Gaussian Material Synthesis (GMS) instead?
A: By using our system, artists can reuse their image editing knowledge and apply it to material synthesis, even if they don't have any direct experience in this field. If one, or at most a handful of materials are sought, the modeling times of our proposed method are preferable to GMS. However, GMS incurs a significant fixed cost as recommendations are made by learning the artist's preferences. This typically pays off in the case of mass-scale material synthesis (i.e., equal to or more than a dozen materials) – in these cases, GMS should be preferable for the artist. In Table 1, we endeavored to simplify the process of choosing the appropriate class of methods for a prescribed application.

Authors' addresses: Károly Zsolnai-Fehér, TU Wien, Favoritenstrasse 9-11/193-02, Vienna, Austria, 1040, zsolnai@cg.tuwien.ac.at; Peter Wonka, KAUST, Al Khwarizmi Bldg 1, Thuwal, Kingdom of Saudi Arabia, 23955-6900, pwonka@gmail.com; Michael Wimmer, TU Wien, Favoritenstrasse 9-11/193-02, Vienna, Austria, 1040, wimmer@cg.tuwien.ac.at.

XXXX-XXXX/2019/9-ART
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Q: In some cases, there is no improvement after the first few iterations. How many iterations should the optimizer use in general?
A: We propose a two-stage system where first, our inversion networks propose a reasonably accurate initial solution, and in the next stage, it is used as an initial guess by the optimizer and undergoes further refinement. As the first stage executes within a few milliseconds, it can be used as-is for real-time applications where an approximate solution is tolerable. In production rendering environments where the artist can typically afford to wait 20 seconds for a more accurate solution, we recommend using both stages. Furthermore, since both the input and the output images are both available for the algorithm, the RMSE between the two can be compared. With a carefully chosen error threshold, this would result in a "best of both worlds" solution that only takes 20 seconds when necessary, and would execute in close to real time otherwise.

Q: Is the RMSE an appropriate error measure for this task?
A: RMSE is widely regarded as the standard way of measuring differences in BRDF modeling [Dupuy and Jakob 2018; Matusik et al. 2003]. There are indeed specialized cases, e.g., noise and blurring among other examples that require non-standard image quality metrics [Liu et al. 2013; Zhu and Milanfar 2009] – regardless, we have tried measuring the PSNR and produced per-channel greyscale images to record the SSIM [Wang et al. 2004] and have not found meaningful differences to RMSE in our test cases.

Q: Why does the proposed "best of 9" scheme work?
A: After training a neural network on a large and diverse set of training examples, if the test samples are not markedly different from the training set, proper measures are taken against overfitting and the layer architecture is chosen appropriately, the network is expected to perform well on unseen examples. However, in our case, the input images undergo a set of creative transforms by the artist and therefore, differ significantly from the images contained within the training set. As a result, in most cases, an exact match is impossible to attain through the given principled shader. To combat this, we have trained a set of neural networks with different tradeoffs that perform well on a disjoint set of target images. To demonstrate this

Name	In-scene editing	Exploration	Moderate-scale	Mass-scale
Direct interaction	✗	✗	✗	✗
BRDF relighting	✓	✗	✗	✗
Gaussian Material Synthesis	✗	✓	✓	✓
Photorealistic Material Editing (ours)	✗	✓	✓	✓

Table 1. The key advantage of BRDF relighting methods is the possibility of editing the materials directly within the final scene at the cost of forfeiting rapid exploration and mass-scale material synthesis. To help the artist explore many potential candidate materials, GMS supports variant generation through a 2D latent-space projection while our method offers close to real-time performance on image sequences. The yellow and green check marks showcase that our method outperforms in moderate-scale problems while GMS excels at mass-scale material synthesis.

effect, we have shown a set of example predictions in Fig. 4 in the paper that also reveals that in these cases, some inversion networks may predict results outside of the feasible domain. Due to the non-convex landscape of our principled shader, simply clamping back the parameters to the feasible domain may lead to undesirable results. This can be remedied by our "best of 9" scheme – since we have an atypical problem where both the predicted images and the target image are available, we can inexpensively determine and choose the best prediction of these inversion networks. We discussed the used architectures for all of these inversion networks in the Appendix section of the paper and we have included these network models in the supplementary materials as well.

Q: How does this work relate to "Generative Visual Manipulation on the Natural Image Manifold" [Zhu et al. 2016]?
A: This method uses a generative model to synthesize images, whereas our technique seeks a parameter setup to be used with a principled shader. In their work, the space of image editing operations is constrained, but in return, yields a large variability for their output images. Our technique strikes a different tradeoff where the space of editing operations is more forgiving, and produces outputs that must adhere to the rules of the principled shader, i.e., represent photorealistic materials. This design choice also necessitates our "best of 9" scheme to provide robust results. Furthermore, our optimization process involves invoking a neural renderer to produce the intermediate images to compare against the target image, and each of our stages are modular, i.e., can be used in isolation or combined together depending on the requirements of the artist (we discussed the details of this in the previous question).

2 SUPPLEMENTAL FILES

The submission contains a supplementary video with a high-level overview of our system and a discussion of the results. To maximize reproducibility, we also provide the full source code for the entirety of the project, the edited input images shown in the paper, a pre-trained neural renderer and all of our described inversion networks.

REFERENCES

Jonathan Dupuy and Wenzel Jakob. 2018. An Adaptive Parameterization for Efficient Material Acquisition and Rendering. *Transactions on Graphics (Proceedings of SIGGRAPH Asia)* (Dec. 2018).

Yiming Liu, Jue Wang, Sunghyun Cho, Adam Finkelstein, and Szymon Rusinkiewicz. 2013. A no-reference metric for evaluating the quality of motion deblurring. *ACM Trans. Graph.* 32, 6 (2013), 175–1.

Wojciech Matusik, Hanspeter Pfister, Matt Brand, and Leonard McMillan. 2003. A Data-Driven Reflectance Model. *ACM Transactions on Graphics* 22, 3 (July 2003), 759–769.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.

Jun-Yan Zhu, Philipp Krähenbühl, Eli Shechtman, and Alexei A Efros. 2016. Generative visual manipulation on the natural image manifold. In *European Conference on Computer Vision*. Springer, 597–613.

Xiang Zhu and Peyman Milanfar. 2009. A no-reference sharpness metric sensitive to blur and noise. In *2009 International Workshop on Quality of Multimedia Experience*. IEEE, 64–69.

Károly Zsolnai-Fehér, Peter Wonka, and Michael Wimmer. 2018. Gaussian Material Synthesis. *ACM Transactions on Graphics (Proc. SIGGRAPH)* (2018).