RESEARCH STATEMENT

Xilong"Logan" Zhou (1992zhouxilong@gmail.com)

My research goal is to bridge the gap between virtual and real world, and develop practical systems to accurately represent the appearance and realism of our world. This encompasses a variety of fascinating graphics and vision tasks, such as 3D generation, appearance acquisition, relighting, neural rendering and inverse rendering. My PhD research focuses on a sub-branch of these tasks, with the emphasis on material acquisition and generation.

In the graphics rendering pipeline, the reflectance property of materials is an important factor determining the appearance of the scene. Spatially-varying bidirectional reflectance distribution function (SVBRDF) codifies the material property by describing how the light is reflected from the surface. There are two general ways to obtain SVBRDFs: material acquisition, with the goal of extracting SVBRDFs from the target, and material generation, aiming at synthesizing SVBRDFs without relying on reference. In traditional methods, technicians use either hardware to acquire materials or procedural substance graphs for material generation. Although these methods are accurate and interactive, they are time-consuming and impractical to non-professional users.

With the advent of deep learning techniques, many learning-based approaches have been proposed for lightweight material acquisition [1, 2, 3, 4, 5, 6, 7, 8] and generation [9, 10]. Specifically, these approaches focus on extracting SVBRDFs from a casually captured single photo and generating SVBRDFs efficiently. Since obtaining ground truth SVBRDFs is challenging for real data, these systems are mostly trained on synthetic data. Unfortunately, synthetic data exhibit a huge distribution gap to the real-world data, causing unrealistic acquisition and generation results.

To address these limitations, **my PhD research strives to explore practical and robust systems with the goal of mitigating the gap between real and synthetic data and obtaining realistic materials.** This research statement is organized as follows: the first section discusses my completed and in-progress research on realistic material acquisition and generation, and the second section covers my future agenda.

1 Completed and Ongoing Research

1.1 Material Acquisition

Hybrid training strategy The material priors of most single-shot SVBRDF acquisition methods are trained on synthetic datasets, causing limited generalization to real examples. To address this issue, in our work published at Eurographics 2021 [11], we propose a novel hybrid training strategy to train our system under the supervision of both synthetic and real datasets. Our key observation is that a pair of real images of the same material captured with different flash lights can be used for supervising the network. We collect a real dataset containing pairs of same materials under different flash lights, and use one of the images as the input and the other as the ground truth. Our system trained on hybrid dataset demonstrates better performance (Figure 1) in handling real examples than state-of-the-art methods.

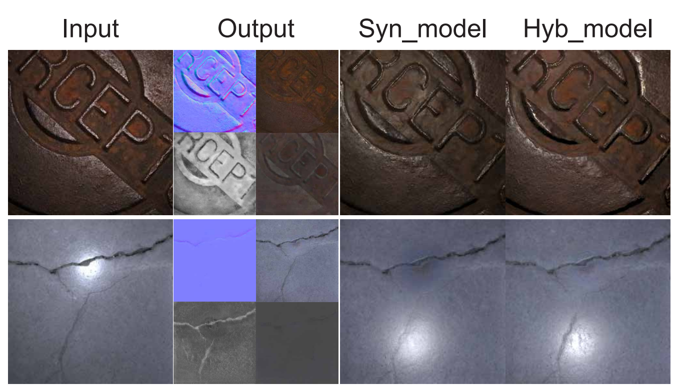

Figure 1: Our system trained on hybrid data can better reproduce the appearance of real data than synthetically trained model.

Semi-procedural material prior To incorporate editability and tileability to the acquisition system, the existing methods utilize procedural substance graphs as material priors. However, such synthetic priors are still complex and have limitations in representing real samples. To tackle this problem, in our work published at CGF 2023 [12], we design a lightweight semi-procedural material prior that can acquire editable and tileable materials without relying on any synthetic dataset or node graphs. Inspired by the traditional substance graphs, we utilize a specialized lightweight network to convert a set of input noises/patterns to SVBRDFs that matches the target materials under the style similarity metric. By manipulating the input, users can obtain tileable materials with varying fine details while following the style of the target materials. This prior can avoid highlight burn-in artifacts (Figure 2) and enable control over the results at low computational and storage cost.

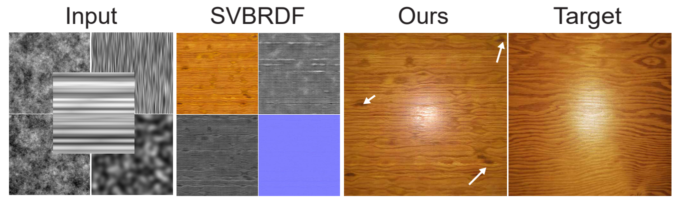

Figure 2: Our lightweight semi-procedural material prior can produce a result matching the appearance of this target wood: our result even shows knots in the wood pattern, despite the initial grayscale maps having no knot-like features.

1.2 Material Generation

Material generators trained on real flash photos The existing material generators are built upon different architectures (such as GAN and transformer) to synthesize materials. One of our work, TileGen [13], published at Siggraph Asia 2022, is a category-specific GAN-based material generator producing tileable and editable materials. However, TileGen and other existing generators are all trained on synthetic dataset, which significantly limits the realism of the sampled materials. To address this limitation, we propose PhotoMat [14] , the first realistic material generator trained exclusively on real flash photos. Training generators on real data is challenging since the supervision of material maps is unavailable for real photos. We achieve this goal by performing both generation and acquisition within a single framework and splitting the problem into two parts. First, we train a generator for a neural material that is rendered with a learned relighting module to synthesize relit realistic materials. Then we utilize a map estimator to decode SVBRDF from the neural material. We demonstrate that PhotoMat surpasses all existing material generators in its ability to generate photo-realistic materials (Figure 3).

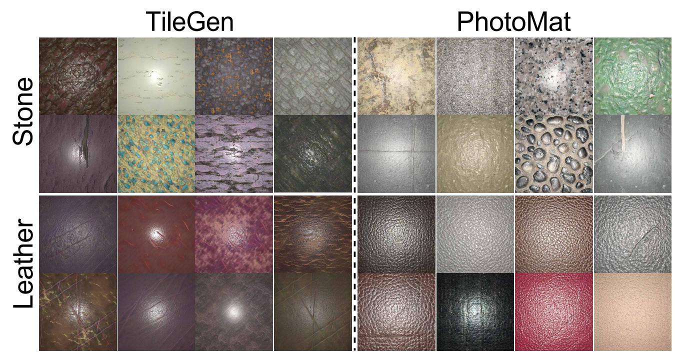

Figure 3: The comparison between PhotoMat and TileGen. As is shown here, the sampled materials of PhotoMat are more photo-realistic than TileGen, demonstrating the effectiveness of our method.

Extension of PhotoMat This is an in-progress work in collaboration with Adobe Research. PhotoMat has demonstrated the possibility of training a material system without direct supervision. However, the training dataset of PhotoMat is limited to flash photos only, and the scale of such dataset is relative small. Therefore, in this project, we relax the data capture constraint, extending from known flashlight to unknown environmental lighting, and aim at training a realistic material generator on a general real dataset such as million-level Laion Dataset. This would require three carefully-designed key components: an effective curation techniques to filter Laion Dataset, an environment light estimator for real materials, as well as a robust neural renderer under environment lighting.

2 Future Research Agenda

My long-term research goal is to develop systems that can accurately represent our real world and generate realistic contents. To fulfill this objective, in addition to the previously discussed research, there are a variety of interesting and challenging tasks I would like to explore in the future.

Figure 4: An example of "material picker".

Scene-level inverse rendering Several of my completed research target on realistic material acquisition from planar surfaces, with the potential extension to indoor scenes. One prospective avenue is scene-level inverse rendering, with a compelling application described as "material picker". Similar to "color picker" through which users obtain the RGB value by selecting different pixels on an image, "material picker" enables users to acquire SVBRDFs from any selected small patch on an indoor scene lit under unknown lighting (Figure 4). In the existing scene-level inverse rendering methods [15, 16, 17], their systems are trained under the supervision of synthetic assets, causing the limited realism and low computational efficiency. To circumvent this problem, it would be promising to develop a realistic "material picker" that is exclusively trained on real indoor dataset. This would require several key components including the estimation of indoor lighting, dataset preparation and robust acquisition and generation system.

3D content generation Most 3D content generators are trained on 3D datasets [18, 19, 20, 21] or well-curated 2D datasets [22, 23] containing images with similar scale, orientation and categories. Unfortunately, although the existing 3D datasets are carefully designed, they still exhibit several limitations compared to common 2D image dataset such as lack of realism, limited diversity and relatively small scale. In addition, most large-scale 2D datasets are non-curated, consisting of diverse objects with unknown scale and orientation. My goal is to train a 3D content generator on the 2D non-curated datasets in order to generate more diverse and realistic content. An existing work [24] has demonstrated the possibility of training such system, but there still exist plenty of room for improvement in the future. For example, we could combine and leverage the benefits of diffusion model and GAN or include stronger 3D priors extracted from 2D dataset.

3D intrinsic generation Another area of interest is investigating generators which are capable of generating lighting-disentangled intrinsic properties of 3D objects or scenes. Existing 3D generators commonly use implicit mesh representation and synthesize final RGB image via volume rendering, where lighting, geometry and material properties are coupled with each other. This would limit the seamless integration of generated content into the graphics rendering pipeline. To address this limitation, it would be promising to design a system that can generate 3D intrinsic properties such as albedo, normal and roughness that are disentangled with lighting. Furthermore, to process further, we could potentially train such 3D intrinsic generators on some well-curated 2D datasets. This would necessitate robust priors on lighting estimation, intrinsic decomposition and a differentiable neural renderer.

References

[1] Valentin Deschaintre, Miika Aittala, Fredo Durand, George Drettakis, and Adrien Bousseau. Single-image svbrdf capture with a rendering-aware deep network. *ACM Trans. Graph.*, 37(4):128:1–128:15, 2018.

[2] Zhengqin Li, Kalyan Sunkavalli, and Manmohan Chandraker. Materials for masses: SVBRDF acquisition with a single mobile phone image. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part III*, volume 11207 of *Lecture Notes in Computer Science*, pages 74–90, 2018.

[3] Wenjie Ye, Xiao Li, Yue Dong, Pieter Peers, and Xin Tong. Single image surface appearance modeling with self-augmented cnns and inexact supervision. In *Computer Graphics Forum*, volume 37, pages 201–211. Wiley Online Library, 2018.

[4] Jie Guo, Shuichang Lai, Chengzhi Tao, Yuelong Cai, Lei Wang, Yanwen Guo, and Ling-Qi Yan. Highlight-aware two-stream network for single-image svbrdf acquisition. *ACM Trans. Graph.*, 40(4), jul 2021.

[5] Rosalie Martin, Arthur Roullier, Romain Rouffet, Adrien Kaiser, and Tamy Boubekeur. Materia: Single image high-resolution material capture in the wild. *Computer Graphics Forum*, 41(2):163–177, 2022.

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

[7] Yezi Zhao, Beibei Wang, Yanning Xu, Zheng Zeng, Lu Wang, and Nicolas Holzschuch. Joint svbrdf recovery and synthesis from a single image using an unsupervised generative adversarial network. 2020.

[8] Liang Shi, Beichen Li, Miloš Hašan, Kalyan Sunkavalli, Tamy Boubekeur, Radomir Mech, and Wojciech Matusik. Match: Differentiable material graphs for procedural material capture. *ACM Trans. Graph.*, 39(6):1–15, December 2020.

[9] Yu Guo, Cameron Smith, Miloš Hašan, Kalyan Sunkavalli, and Shuang Zhao. Materialgan: Reflectance capture using a generative svbrdf model. *ACM Trans. Graph.*, 39(6):254:1–254:13, 2020.

[10] Paul Guerrero, Milos Hasan, Kalyan Sunkavalli, Radomir Mech, Tamy Boubekeur, and Niloy Mitra. Matformer: A generative model for procedural materials. *ACM Trans. Graph.*, 41(4), 2022.

[11] Xilong Zhou and Nima Khademi Kalantari. Adversarial Single-Image SVBRDF Estimation with Hybrid Training. *Computer Graphics Forum*, 2021.

[12] Xilong Zhou, Miloš Hašan, Valentin Deschaintre, Paul Guerrero, Kalyan Sunkavalli, and Nima Khademi Kalantari. A semi-procedural convolutional material prior. In *Computer Graphics Forum*. Wiley Online Library, 2023.

[13] Xilong Zhou, Milos Hasan, Valentin Deschaintre, Paul Guerrero, Kalyan Sunkavalli, and Nima Khademi Kalantari. Tilegen: Tileable, controllable material generation and capture. In *SIGGRAPH Asia 2022 Conference Papers*, SA '22, New York, NY, USA, 2022. Association for Computing Machinery.

[14] Xilong Zhou, Milos Hasan, Valentin Deschaintre, Paul Guerrero, Yannick Hold-Geoffroy, Kalyan Sunkavalli, and Nima Khademi Kalantari. Photomat: A material generator learned from single flash photos. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–11, 2023.

[15] Kai Yan, Fujun Luan, Miloš Hašan, Thibault Groueix, Valentin Deschaintre, and Shuang Zhao. Psdr-room: Single photo to scene using differentiable rendering. In *SIGGRAPH Asia 2023 Conference Papers*, pages 1–11, 2023.

[16] Yu-Ying Yeh, Zhengqin Li, Yannick Hold-Geoffroy, Rui Zhu, Zexiang Xu, Miloš Hašan, Kalyan Sunkavalli, and Manmohan Chandraker. Photoscene: Photorealistic material and lighting transfer for indoor scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18562–18571, 2022.

[17] Zhengqin Li, Mohammad Shafiei, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2475–2484, 2020.

[18] Jiahao Li, Hao Tan, Kai Zhang, Zexiang Xu, Fujun Luan, Yinghao Xu, Yicong Hong, Kalyan Sunkavalli, Greg Shakhnarovich, and Sai Bi. Instant3d: Fast text-to-3d with sparse-view generation and large reconstruction model. *arXiv preprint arXiv:2311.06214*, 2023.

[19] Yinghao Xu, Hao Tan, Fujun Luan, Sai Bi, Peng Wang, Jiahao Li, Zifan Shi, Kalyan Sunkavalli, Gordon Wetzstein, Zexiang Xu, and Kai Zhang. Dmv3d: Denoising multi-view diffusion using 3d large reconstruction model, 2023.

[20] Minghua Liu, Chao Xu, Haian Jin, Linghao Chen, Zexiang Xu, Hao Su, et al. One-2-3-45: Any single image to 3d mesh in 45 seconds without per-shape optimization. *arXiv preprint arXiv:2306.16928*, 2023.

[21] Ruoshi Liu, Rundi Wu, Basile Van Hoorick, Pavel Tokmakov, Sergey Zakharov, and Carl Vondrick. Zero-1-to-3: Zero-shot one image to 3d object, 2023.

[22] Titas Anciukevičius, Zexiang Xu, Matthew Fisher, Paul Henderson, Hakan Bilen, Niloy J Mitra, and Paul Guerrero. Renderdiffusion: Image diffusion for 3d reconstruction, inpainting and generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12608–12618, 2023.

[23] Eric R. Chan, Connor Z. Lin, Matthew A. Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas Guibas, Jonathan Tremblay, Sameh Khamis, Tero Karras, and Gordon Wetzstein. Efficient geometry-aware 3D generative adversarial networks. In *arXiv*, 2021.

[24] Ivan Skorokhodov, Aliaksandr Siarohin, Yinghao Xu, Jian Ren, Hsin-Ying Lee, Peter Wonka, and Sergey Tulyakov. 3d generation on imagenet. *arXiv preprint arXiv:2303.01416*, 2023.