

MatSynth: A Modern PBR Materials Dataset

Giuseppe Vecchio
 University of Catania

giuseppe.vecchio@phd.unict.com

Valentin Deschaintre
 Adobe Research

deschain@adobe.com

Metadata:

- Name: Marble_37
- Source: sharetextures
- License: CC0
- Maps:
 - basecolor
 - diffuse
 - displacement
 - height
 - metallic
 - normal
 - roughness
 - specular
- Category: Marble
- Tags:
 - architectural
 - interior
 - luxury
 - marble
- Method: approximation
- Height Factor: 0.001
- Height Mean: 0.0
- Version Data: 2023-10-31
- Description: ""
- Physical Size: N/A



Figure 1. **Samples of materials from the dataset.** Our dataset contains 4, 069 high-quality, 4K, tileable materials with permissive licences. Each material is augmented, rendered and supplemented by metadata containing its origin, tags, categories, method of creation, and more.

Abstract

We introduce *MatSynth*, a dataset of 4, 000+ CC0 ultra-high resolution PBR materials. Materials are crucial components of virtual relightable assets, defining the interaction of light at the surface of geometries. Given their importance, significant research effort was dedicated to their representation, creation, and acquisition. However, in the past 6 years, most research in material acquisition or generation relied either on the same unique dataset, or on company-owned huge library of procedural materials. With this dataset, we propose a significantly larger, more diverse, and higher resolution set of materials than previously publicly available. We carefully discuss the data collection process and demonstrate the benefits of this dataset for material acquisition and generation applications. The complete data further contains metadata with each material’s origin, license, category, tags, creation method, and, when available, descriptions and physical size, as well as 3M+ renderings of the augmented materials, in 1K, under various environment lightings. The *MatSynth* dataset is released through the project page at: <https://www.gvecchio.com/matsynth>.

1. Introduction

Materials are a key component of the rendering pipeline, describing the appearance of objects, for example in terms of colors, shininess or metallicity. The design of high-quality virtual materials is a complex endeavor, usually requiring significant artistic expertise [30]. To reproduce existing objects and surfaces, material acquisition and generation have been active research fields for decades [14]. More recently, the field turned to Machine Learning to tackle the ill-posedness of few images (lightweight) material acquisition [5, 23], and to explore new opportunities in material generation [18, 50]. Since 2018, most of the field has relied on a single dataset [5] or, when possible, on company-owned large material libraries [1, 28, 44, 50]. As company owned libraries are typically orders of magnitude bigger than what is publicly available, they enable very high quality results and experiments with higher resolutions. These materials are, however, not available to the community, underlining the need for larger, more diverse, public datasets to enable further research. Such material data are also particularly useful for synthetic data generation, for example, for object acquisition [8, 24], material-based selection [38], and numerous applications further from material understanding and authoring [34, 40, 42].

To bridge the gap between private and public datasets, we present MatSynth, a dataset of 4,069 tileable materials, available in 4K. Alongside the material properties themselves, we render 3,417,960 renderings with various scales cropping, rotation, and environment illuminations to facilitate future learning. The materials all come with permissive licences and were sourced from various libraries [2, 4, 32, 37, 41] and checked for duplication among newly gathered data and against previous datasets [5]. We further ensure that all materials are represented by the same set of parameters and that they are tileable. In particular, each material is represented by 7 textures/maps: Base color, Diffuse, Normal, Height, Roughness, Metallic and Specular, ensuring compatibility with both Diffuse/Specular (e.g. [5]) and Base Color/Metallic (e.g. [28]) workflows. Some materials in our dataset also have an opacity map to handle, for example, surfaces with holes. Finally, we process the previously available dataset to enrich it with metallic, base color, and height properties, facilitating their combinations for a total of $\sim 6,000$ materials available for research, with clear licenses attached.

To validate the benefit of these additional data, we train recent methods with publicly available implementation and evaluate both qualitatively and quantitatively the results quality for different single image material capture methods [5, 43] and material generation [45], showing that the additional data enables higher quality acquisition and improved diversity in the generation results.

In summary, we gather a new material dataset with the following properties and evaluate its impact on material related tasks:

- 4,069 unique 4K PBR materials with permissive licences.
- 683,592 augmented materials with rotations and crops
- 3,417,960 renderings of the augmented materials under various illumination.
- Associated metadata.

2. Related Work

We describe the main existing material dataset and related work leveraging it. We also discuss material acquisition/generation approaches using company-owned data or unsupervised learning.

2.1. Existing dataset

The only large-scale, publicly available, spatially-varying material dataset contains 1628 distinct, pre-augmentation materials [5] with a research-only license. 512x512 [6] and 2K [7] versions of the materials were later published. It is based on random variations from a set of 155 procedural materials gathered from Allegorithmic Substance Share in 2017, distributed in 9 categories: paint(6), plastic(5), leather(13), metal(35), wood(23), fabric(6), stone(25), ceramic tiles(29), ground(13). Other datasets used in the

field contain 100 and 62 measured BRDFs [10, 29], 170 SVBRDFs licensed from VRay[23, 46] or more recently 211 materials from online libraries [17]. Adobe 3D Assets [1]’s 11,000+ procedural assets have been used to unlock significant quality improvement in the methods that could benefit from it for acquisition, generation, retrieval and captioning [9, 28, 44, 50]. This data is however difficult to license for Machine Learning, limiting its use to company internal users. In contrast, we propose a systematic gathering of non-duplicate, tileable, publicly available materials with permissive licenses from various online libraries, assembling 4,000+ materials and their metadata, significantly increasing the variety and amount of resources available for material-related research.

2.2. Material acquisition

A typical use of synthetic material data is acquisition method training. Given one or a few images, these methods aim at recovering material properties. As the complete acquisition of a material behaviour typically requires hundreds to thousands of measurements [14], few images acquisition is a particularly ill-posed problem, which can strongly benefit from machine learning. In the past few years many methods were proposed to recover materials from one [5, 16, 17, 23, 28, 43, 44] or a few images [6, 7, 11, 12, 18], entirely relying on synthetic data for training. While most of the field relied on synthetic data, Zhou et al. [49] proposed to use a mix of synthetic and real data, and Henzler et al. [20] leverage a weakly supervised training combined with test-time fine tuning for stationary material acquisition. Some of these work use the large scale Adobe 3D Assets library [1, 28, 44] showing high quality results, suggesting that access to larger public materials datasets will be critical for further progress in the material acquisition field, for both higher quality, higher diversity and higher resolution results.

2.3. Material generation

Recent years saw fast quality improvement in generative tasks [22], inspiring research around material generation. In generative workflow in particular, data has been a key enabler for higher quality and diversity [3, 35]. In the context of materials, this lead generative research to rely on private large material libraries [1], making contributions to the field challenging for academics or companies which do not own such data. Generative models for materials leveraging supervised GANs were first proposed [18, 51]. More recently, a few approaches tackled the problem using text and image conditional diffusion models [44, 45] or tried to simplify the required data by relying solely on weakly supervised flash images during the training [52]. Aside from two [45, 52], all related works discussed in this paragraph were trained on private data [1]. A concurrent work [19] follows a sim-

ilar data gathering approach as ours, but do not make their dataset available to this day, with no plans for it announced. We believe that our dataset is a significant step towards reducing the gap between publicly and privately available material datasets, facilitating future research.

An orthogonal research direction, focused on procedural material generation [15, 21] requires access to complete procedural materials, which is out of our dataset’s scope.

2.4. Data generation

Aside from the flat material domain, different synthetic datasets, mostly of interior scenes [26, 36, 38, 53], have been proposed to enable further research for tasks such as inverse rendering and intrinsic image decomposition [25, 54], relighting [13], guided assets retrieval [48] or material similarity and selection [31, 38]. More generally, synthetic data has been an important resources for various world generation approaches [34, 40, 42], providing ground truth properties to bootstrap trainings, generalising to real images. Larger materials datasets availability will significantly simplify the creation of variations of existing scene datasets, increasing the diversity of appearances, for example through similar material retrieval or random assignment, as done recently with private material libraries [31, 38].

3. The Dataset

The MatSynth dataset is designed to support modern learning-based techniques for a variety of material-related tasks including, but not limited to, material acquisition, material generation, and synthetic data generation, e.g., for retrieval or segmentation. It improves on existing available datasets by providing a large collection (~ 2.5 times larger than previously available datasets) of non-duplicate, high-quality, high-resolution realistic materials alongside useful metadata. Each material in the dataset is represented by a set of reflectance maps (basecolor, diffuse, normal, height, roughness, metallic, and specular), complemented by fine-grained annotations including the source and link to the asset, its license, and original author (when available), tags, the creation method, whether is stationary or not, a timestamp for versioning, a description (when available), and the material physical size (when available).

In this section, we describe the MatSynth dataset, its gathering and analyse its content.

3.1. Materials Collection

To gather the MatSynth dataset we extensively collect data from multiple online sources operating under the CC0 and CC-BY licensing framework. This collection strategy allows to capture a broad spectrum of materials, from commonly used ones to more niche or specialized variants while guaranteeing that the data can be used for a variety of usecases. Materials under CC0 license were collected

from AmbientCG [2], CGBookCase [4], PolyHeaven [32], ShateTexture [37], and TextureCan [41]. The dataset also includes limited set of materials from the artist Julio Silet, distributed under CC-BY license. We collected over 6000 materials which we meticulously filter to keep only tileable, 4K materials. This high resolution allows us to extract many different crops from each sample at different scale for augmentation. Additionally, we discard blurry or low-quality materials (by visual inspection). The resulting dataset consists of 3736 unique materials which we augment by blending semantically compatible materials (e.g.: snow over ground). In total, our dataset contains 4069 unique 4K materials. We ensure that these new materials are not duplicates of those previously available in the existing material dataset [5].

3.2. Data Annotations

The dataset is composed of material maps (Basecolor, Diffuse, Normal, Height, Roughness, Metallic, Specular and, when useful, opacity) and associated renderings under varying environmental illuminations, and multi-scale crops. We adopt the OpenGL standard for the Normal map (Y-axis pointing upward). The Height map is given in a 16-bit single channel format for higher precision.

In addition to these maps, the dataset includes other annotations providing context to each material: the capture method—whether it is photogrammetry, procedural generation, or approximation—; list of descriptive tags; source name (website); source link ; licensing and a timestamps for eventual future versioning. For a subset of materials, when the information is available, we also provide the author name (387), text description (572) and a physical size, presented as the length of the edge in centimeters (358).

3.3. Data Processing

We carefully evaluate and select the collected data. We first visually inspect each material, looking for evident defects in the maps (e.g.: blurred maps, baked in shadows or highlights, unrealistic reflectance properties). The remaining materials have been automatically checked using CLIP [33] embeddings and contrastive prompts as described in [47]. In particular, we evaluated each rendered material for quality with the following contrastive pairs: quality (high-quality/low-quality), sharpness (sharp/blurry), noisiness (clean/noisy), and realism (natural/synthetic).

We check for materials duplication by cross comparing the CLIP [33] embeddings of all materials in the dataset, flagging values close to 1 as possible duplicate. Each duplicate candidate was then manually reviewed for false positives before being removed.

We ensure that all materials respect the assumed y-axis pointing upward by differentiating the height map into a normal map and counting the number vectors with a dis-

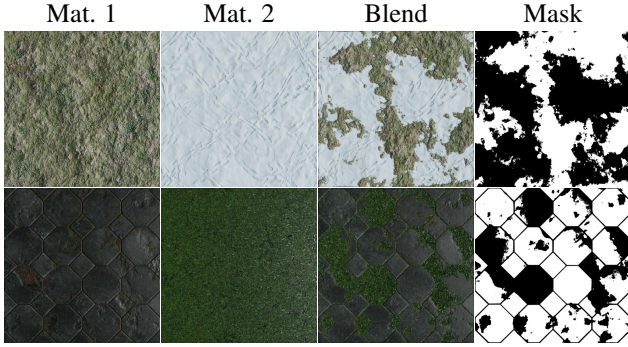


Figure 2. **Samples of height based material blends.** We show renderings of both materials being blended, the result of our blending, and the height-based computed mask.

cordant y component. For materials with a discordant count above 30%, we manually check the normal map and invert the y component when necessary. We perform a final visual inspection of the renders to ensure a correct normal orientation for all materials in the dataset.

Finally, we automatically compute an estimate of the correct displacement factor for each material, by optimizing a displacement factor parameter to match the normal vectors derived from the scaled height map to the ground truth normal map. This ensures a good match between the normal and the displacement scale when rendering. We provide the correctly scaled height map, as well as the original and the scaling parameters (scale factor and mid-value compatible with the Blender “Displacement” node).

3.4. Data Augmentation

Material blending. We further enrich the dataset by blending semantically compatible materials. First, we define a set of compatible classes in terms of base layer (the base of the material) and top layer (the material to be layered on top of it). Second, we compute a blending mask based on the combination of the two height maps. To do so, we randomly sample a threshold between the minimum and maximum values of the base layer material, we shift the top material height mean to the sampled threshold, and intersect the two height maps taking the material with the max height value for each pixel. We finally blur the blending mask with a 9 pixels Gaussian kernel to smooth the transition between the two materials. Each map of the blended material is computed as $M = M_{base} \times mask + M_{top} \times (1 - mask)$. We include the blending mask for each blended material in the dataset. Fig 2 shows examples of blended materials, the original materials and the blending masks. We generate a total of 332 blends. The full list of class pairs used for blending, and the number of blends, is provided in the supplemental material.

Rotation and crops and illumination. The collected 4069

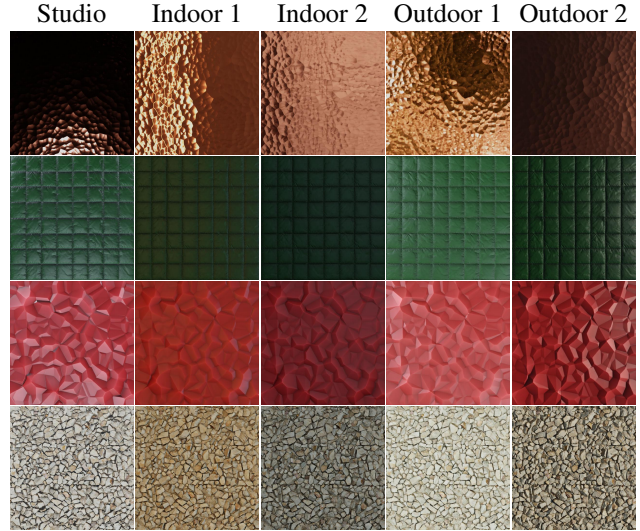


Figure 3. **Renderings under various environment maps.** We show four materials (Metal, Leather, Plastic and Pebbles) from the dataset rendered under the 5 chosen environment maps.

materials are augmented by rotation and cropping. In particular, we extract 168 unique crops for each material, at different scales. The crops are extracted in the following way. The original material is rotated 8 times (0° , 45° , 90° , 135° , 180° , 225° , 270° , 315°)—for angles that are not multiple of 90° , we first tile the material, apply the rotation, and finally crop the center of the material at 4K resolution. For each rotation, we extract 16 non-overlapping crops at 1K resolution, 4 at 2K resolution, 1 at 4K resolution, for a total of 21 crops per rotation. All crops are resized to a resolution of 1024×1024 pixels. Each crop is rendered under different environment illuminations (2 outdoors, 2 indoors, and 1 studio light), for a total of 840 renders per material. For each render, a random rotation is applied to the environment map. The resulting dataset consists of 683,592 material crops and 3,417,960 renders. Fig. 3 shows the five renderings under different environment illumination for 4 different materials from the dataset. We include the 5 environment maps used for rendering in the supplemental materials.

Rendering. We render each material displacing the mesh using the available Height map, thus introducing realistic cast shadows in the rendering. However, to maintain a pixel match with the maps, required for supervised SVBRDF estimation tasks, we opt for an orthographic rendering. This introduces yet another issue, as specular reflections are lost on perfectly flat surfaces due to rays being parallel when rendering with an orthographic camera. To restore specular reflection while keeping pixel match and cast shadows, we split the rendering in two separate passes. Similar to the data used to train ControlMat [44], we first render the diffuse component of the material using an orthographic cam-

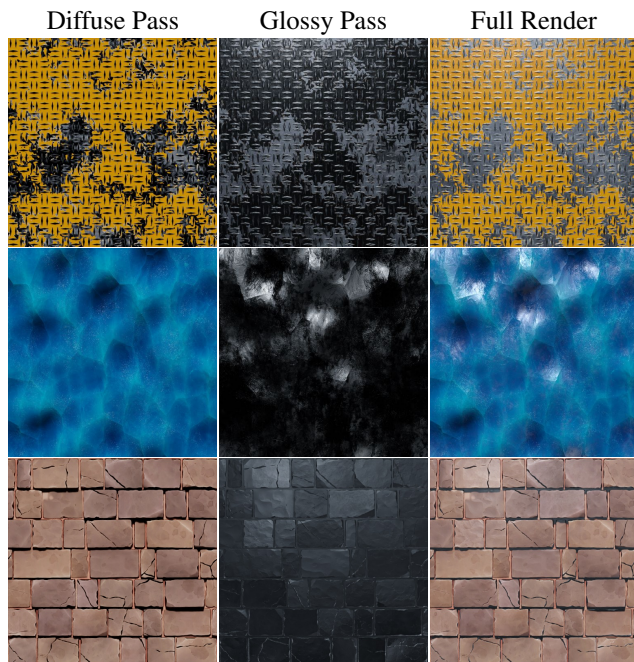


Figure 4. **Render samples using the two-pass strategy.** This ensures that the maps and the rendering are well aligned, avoiding parallax effects, but preserving specular highlights.

era with displacement enabled, we then render the specular component using a perspective camera, while disabling the displacement. The two renders are then added back to get the final render. Fig. 4 shows the output of the two render passes and the final result. We evaluate each augmentation in an ablation study in the Supplemental Materials.

3.5. Compatibility with previous datasets

To ensure compatibility and convenient combination between the previously existing large scale material dataset [5] and ours, we process it in the same way we did ours. We first extract the base color and metallic maps from the existing Diffuse and Specular maps. We extract 5 crops for each material due to the 2048x2048 resolution of the existing dataset. As the existing materials are not tileable we do not apply the rotation augmentation, and directly render each crop under 5 different lighting. This resulting in an additional 40,700 renderings of 8,140 material crops. We will provide these materials and renderings alongside our dataset with clear annotation of origin.

3.6. Dataset Statistics

We report here statistics of our dataset in terms of categories, tags, creation methodology, stationarity, and number of asset per source.

Categories. We report the category histogram in Fig. 5, showing a wide diversity of appearances, and compensat-

ing for some gaps in the distribution of previously available materials.

Tags. We have 21,737 tags from 1,239 unique tags associated with the materials. For example, the top five most represented tags are floor(924), dirty(543), man (504), wood (474), brown (444) and wall(424). Each material has between 0 (371 assets didn't have associated tags) and 20 tags with an average of 5.74, a standard variation of 4.35 and a median of 5.

Creation methodology. Our assets are created through 4 different approaches: manual approximation (299), photometry/acquisition (651), procedural generation (2689), blends (332), unknown (98).

Stationarity. Our dataset contains 3061 stationary and 1008 non-stationary materials. We compute stationarity by extracting 8 random crop of the basecolor and height map and comparing the CLIP cross-similarity matrix. If more than 12.5% of the cross-similarities are below a threshold (we empirically found 0.9 to work well), the material is not considered stationary.

Assets per source. We source 1425 materials from AmbientCG [2], 278 materials from CGBookCase [4], 387 materials from PolyHeaven [32], 647 materials from ShateTexture [37], 571 materials from TextureCan [41], and 224 materials from the artist Julio Sillet. We add 332 blended materials and 205 variations of materials from AmbientCG [2].

4. Experimental Results

In this section, we first introduce the dataset evaluation methodology and test set, we then briefly present the different methods used for evaluation and their training strategy. Finally, we show quantitative and qualitative results.

4.1. Test dataset

We build a static test set, with the aim of simplifying the comparison of future work, by handpicking 5 unique samples for each category. We further ensure that none of the chosen samples is a simple variation (e.g.: color or scale) of other items in the training set to avoid any bias. We merge this dataset, consisting of 65 materials, with the test set from Deschaintre et al. [7], for a total of 89 test materials.

4.2. Evaluation Methodology

To evaluate our dataset we propose a performance comparison between different state of the art methods when trained on the on the dataset by Deschaintre et al. [5], and on MatSynth.

Evaluation is carried on two main materials-related tasks: estimation and generation. For both tasks, we look at methods with publicly available implementations. In particular, we evaluate Deschaintre et al. [5] and SurfaceNet [43] for materials capture performance, and MatFuse [45] for

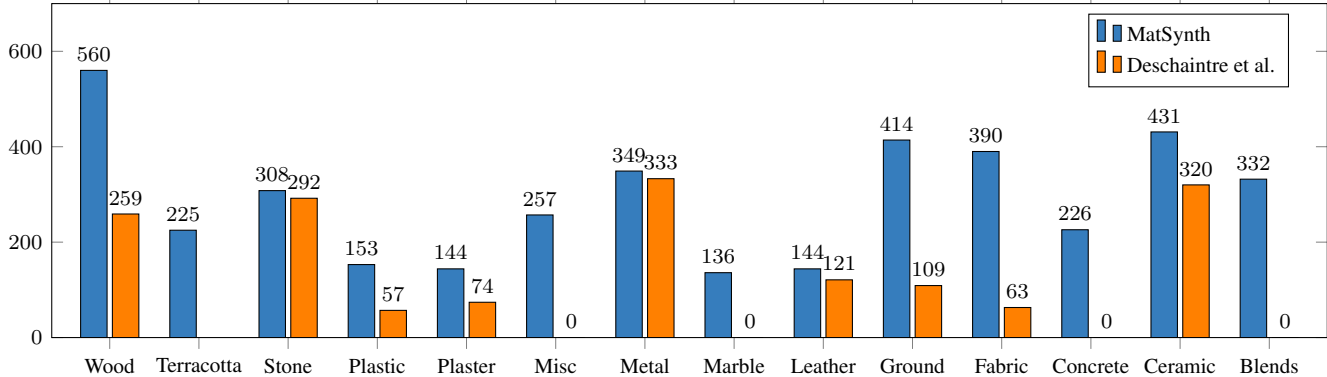


Figure 5. **Categories distribution of MatSynth and Deschaintre et al. [5]’s dataset.** As there are no duplicates in the two datasets, the two datasets can be joined to combine their benefits. We show here that our dataset significantly enriches the previously available data, reducing the inequalities in the category distributions of the previous dataset, in particular for Fabric or Ground categories.

Image Type	Deschaintre ([5])	Deschaintre (MatSynth)	SurfaceNet ([43])	SurfaceNet (MatSynth)
Renderings	0.172	0.160	0.161	0.135
Diffuse	0.100	0.093	0.119	0.094
Normal (Cos dist)	0.576	0.573	0.600	0.544
Roughness	0.322	0.215	0.221	0.162
Specular	0.119	0.080	0.086	0.057

Table 1. **Quantitative results.** We report the RMSE \downarrow between predicted and ground-truth maps and renderings (averaged over renderings under 5 environment lightings), except for Normal maps for which we report the cosine error \downarrow . We see that the additional data provided by MatSynth enables higher quality material acquisition both in terms of parameters and renderings comparisons.

material generation. In both cases, we focus on the diffuse/specular rendering workflow for an easier comparison with the current state of the art. We additionally propose an evaluation of SurfaceNet trained using the basecolor/metallic workflow in Supplemental material, to set a new baseline for future work shifting to this workflow.

4.3. Training Strategy

All methods are trained on PC with a single NVIDIA RTX 4090 GPU. Training of Deschaintre et al. [5] and SurfaceNet [43] is carried out on a flash-lit version of the dataset, as these methods are designed for SVBRDF estimation from flash-lit pictures.

Deschaintre et al. (2018) We train the method, using the author’s publicly available Tensorflow implementation, on a combination of our dataset and the original one. Following the original paper, we train the model for 400,000 steps (72 hours) with a batch size of 8 and a learning rate of 2^{-5} .

SurfaceNet Similarly to the training of Deschaintre et al. [5], we use the author’s original Pytorch implementation and the combined dataset. We also follow the original paper and train the model for 400,000 steps (96 hours) with a batch size of 8 and a learning rate of 2^{-5} . We include the adversarial discriminator after 100,000 steps.

Diffusion Model We evaluate the quality improvement

of generation provided by our dataset on the recent MatFuse [45]. We evaluate the performance both when fine-tuning the diffusion model and when training from scratch. For the first experiment, we use the same multi-encoders compression model, pre-trained on the Deschaintre et al. [5]’s dataset, but fine-tune the latent diffusion model itself. In particular, the diffusion model is fine-tuned for 150,000 iterations with a batch size of 20 and a learning rate of 10^{-5} . We use a linear learning rate warm-up until step 10,000. For the second experiment, we train the entire model, on our dataset. In particular, we train the autoencoder model for 1,000,000 iterations with a batch size of 4 and a learning rate of 10^{-4} . We include the adversarial discriminator after 300,000 steps. Finally, we train the diffusion model for 500,000 iterations with a batch size of 16 using an AdamW [27] optimizer, with a learning rate value of 10^{-4} , with a linear learning rate warm-up until step 10,000. For inference, following the original paper, we denoise using a DDIM sampling schedule [39] with 50 timesteps.

4.4. SVBRDF Estimation Results

We evaluate the quality improvement using our dataset for two methods: Deschaintre et al. [5] and SurfaceNet[43].

We retrain both of these methods on our dataset and report the RMSE for each material map (cosine distance

Image Type	Metric	Deschaintre ([5])	Deschaintre (MatSynth)	SurfaceNet ([43])	SurfaceNet (MatSynth)
Renderings	SSIM	0.532	0.560	0.494	0.613
	LPIPS	0.560	0.294	0.395	0.281

Table 2. **Perceptual quantitative results.** We report here the perceptual metrics for the Renderings, rendered under 5 environment lightings (which can be interpreted as natural images) SSIM \uparrow and LPIPS \downarrow . Similar to the RMSE evaluation, additional data significantly improves the rendering matching’s quality.

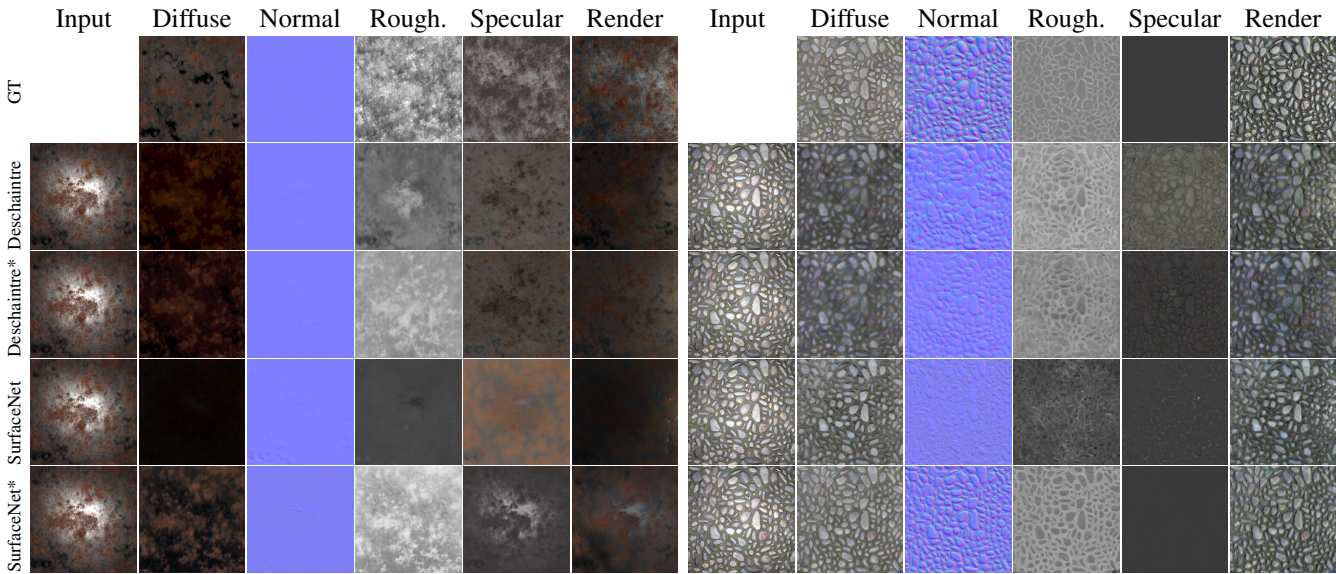


Figure 6. **Qualitative material acquisition comparison on synthetic data.** We compare Deschaintre et al. [5] and SurfaceNet [43] trained only on Deschaintre et al. [5]’s dataset against the same methods trained on MatSynth (marked with *). We can see that the fine-tuned versions better match the Ground truth, in particular for SurfaceNet in the Normal and Roughness maps.

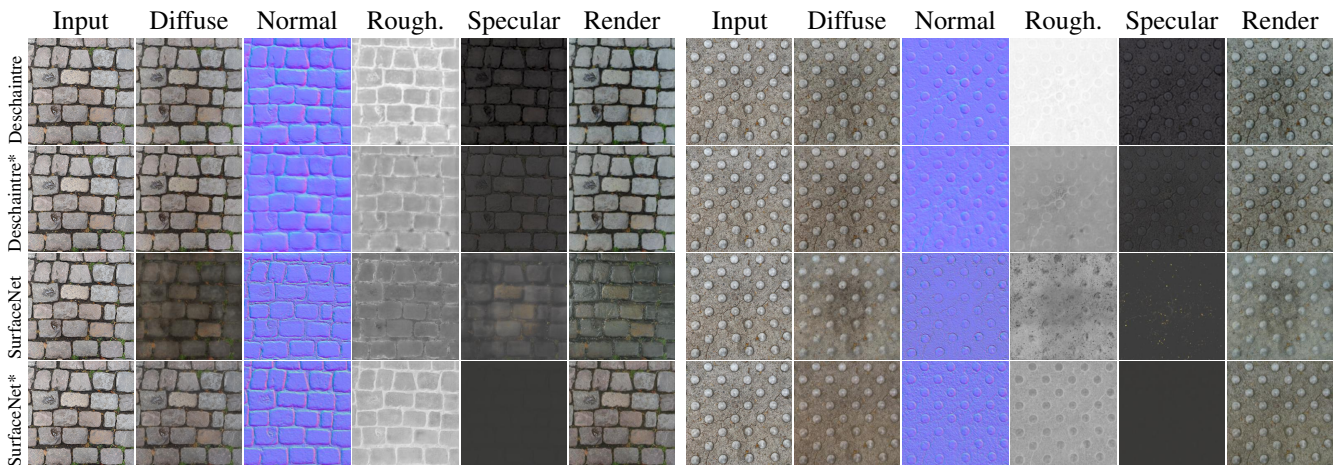


Figure 7. **Qualitative material acquisition comparison on real photographs data.** We compare Deschaintre et al. [5] and SurfaceNet [43] trained only on Deschaintre et al. [5]’s dataset against the same methods trained on MatSynth (marked with *). Here the fine-tuned versions better match the input picture, in particular for SurfaceNet.



Figure 8. **Comparison on Material Generation.** We show renderings of 10 randomly generated materials, both with and without our dataset. These results are directly sampled from the network and have not been filtered. As expected, additional data help achieve better and more diverse results.

for normals), as well as average RMSE and more perceptual metrics (LPIPS and SSIM) between renderings under 5 different environment illumination. We report the results in Table 1 and Table 2, showing improvement across the board, in particular for more challenging maps such as Roughness and Specularity.

We further show a qualitative comparison on synthetic data in Figure 6, as well as real real data captured with a phone camera in Figure 7, demonstrating that the additional data helps achieve better material capture quality, in particular for Surface Net [43] which can leverage the new data best through its discriminator.

4.5. Materials Generation Results

We evaluate material generation using MatFuse [45], trained on Deschaintre et al. [5]’s dataset only, and comparing it against a version fine-tuned with MatSynth and one fully trained on it. As material maps have different distributions than natural images, we report FID on ren-

	original	fine-tuned	full training
Wood	36.1%	25.72%	7.5%
Ceramic	1.1%	12.51%	8.5%
Stone	21.5%	14.8%	7.3%
Metal	3.45%	8.4%	6.4%
Fabric	0.33%	7.34%	7.8%
Ground	1.8%	16.2%	6.4%
Others	35.72%	15.03%	56.1%

Table 3. **Generated materials class distribution.** As expected, we observe that after fine-tuning the generative model create more diverse materials. We classify our results using CLIP [33] based zero-shot classification.

derings under environment lighting. For fairness, we only use Deschaintre et al. [5]’s dataset to compute the Ground-Truth distribution and compute the FID for each method using 1,000 random samples. Before fine-tuning we find a FID score (lower is better) of 239.9, 210.3 after fine-tuning, and 89.84 when training from scratch. We believe the performance difference to be caused by the retraining of the autoencoder using MatSynth, while the other experiments used the original VQ-VAE from MatFuse, trained on Deschaintre et al. [5]’s dataset only.

We further evaluate the benefit of the additional data for diversity and quality. We show in Table 3 the percentage of samples belonging to different classes, and show a qualitative comparison in Fig. 8. Both evaluations show a wider diversity of results, and Fig. 8 shows that the generation results are more realistic after fine-tuning or when training on MatSynth from scratch.

5. Conclusion

While there still exists a gap between real materials photographs and rendered synthetic materials, having access to a large, high-quality, database of materials is crucial to facilitate further research around their authoring, understanding, and use in automatic data generation [42]. With MatSynth, we take a significant step to improve material data accessibility, with 4,069 permissively licensed assets, processed and annotated, alongside millions of augmented renderings. We demonstrate that this additional data immediately improves existing method’s acquisition quality and generation diversity. The materials in the dataset are tileable and in 4K resolution, enabling future research on ultra-high resolution for materials. This is of particular interest as most material artists typically work at 4K+ resolutions.

6. Acknowledgments

We thank Renato Sortino for providing additional computational power to render the dataset and run trainings.

References

- [1] Adobe. Substance 3D Assets. <https://substance3d.adobe.com/assets/>, 2023. 1, 2
- [2] AmbientCG. <https://www.ambientcg.com/>, 2023. 2, 3, 5
- [3] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020. 2
- [4] CGBookCase. <https://www.cgbookcase.com/>, 2023. 2, 3, 5
- [5] Valentin Deschaintre, Miika Aittala, Frédo 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, 2018. 1, 2, 3, 5, 6, 7, 8
- [6] Valentin Deschaintre, Miika Aittala, Frédo Durand, George Drettakis, and Adrien Bousseau. Flexible svbrdf capture with a multi-image deep network. *Computer Graphics Forum (Eurographics Symposium on Rendering Conference Proceedings)*, 38(4):13, 2019. 2
- [7] Valentin Deschaintre, George Drettakis, and Adrien Bousseau. Guided fine-tuning for large-scale material transfer. *Computer Graphics Forum (Proceedings of EGSR)*, 39(4), 2020. 2, 5
- [8] Valentin Deschaintre, Yiming Lin, and Abhijeet Ghosh. Deep polarization imaging for 3d shape and svbrdf acquisition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 1
- [9] Valentin Deschaintre, Julia Guerrero-Viu, Diego Gutierrez, Tamy Boubekeur, and Belen Masia. The visual language of fabrics. *ACM Transactions on Graphics*, 42(4):1–15, 2023. 2
- [10] Jonathan Dupuy and Wenzel Jakob. An adaptive parameterization for efficient material acquisition and rendering. *Transactions on Graphics (Proceedings of SIGGRAPH Asia)*, 37(6):274:1–274:18, 2018. 2
- [11] Michael Fischer and Tobias Ritschel. Metappearance: Meta-learning for visual appearance reproduction. *ACM Trans Graph (Proc. SIGGRAPH Asia)*, 41(4), 2022. 2
- [12] 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. 2
- [13] David Griffiths, Tobias Ritschel, and Julien Philip. Outcast: Single image relighting with cast shadows. *Computer Graphics Forum*, 43, 2022. 3
- [14] D. Guarnera, G.C. Guarnera, A. Ghosh, C. Denk, and M. Glencross. Brdf representation and acquisition. *Computer Graphics Forum*, 35(2):625–650, 2016. 1, 2
- [15] 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. 3
- [16] 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), 2021. 2
- [17] Jie Guo, Shuichang Lai, Qinghao Tu, Chengzhi Tao, Changqing Zou, and Yanwen Guo. Ultra-high resolution svbrdf recovery from a single image. *ACM Trans. Graph.*, 42(3), 2023. 2
- [18] Yu Guo, Cameron Smith, Miloš Hašan, Kalyan Sunkavalli, and Shuang Zhao. MaterialGAN: Reflectance capture using a generative svbrdf model. *ACM Trans. Graph.*, 39(6), 2020. 1, 2
- [19] Zhen He, Jie Guo, Yan Zhang, Qinghao Tu, Mufan Chen, Yanwen Guo, Pengyu Wang, and Wei Dai. Text2Mat: Generating Materials from Text. In *Pacific Graphics Short Papers and Posters*. The Eurographics Association, 2023. 2
- [20] Philipp Henzler, Valentin Deschaintre, Niloy J Mitra, and Tobias Ritschel. Generative modelling of brdf textures from flash images. *ACM Trans Graph (Proc. SIGGRAPH Asia)*, 40(6), 2021. 2
- [21] Yiwei Hu, Paul Guerrero, Milos Hasan, Holly Rushmeier, and Valentin Deschaintre. Generating Procedural Materials from Text or Image Prompts. In *ACM SIGGRAPH 2023 Conference Proceedings*, 2023. 3
- [22] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020. 2
- [23] Xiao Li, Yue Dong, Pieter Peers, and Xin Tong. Modeling surface appearance from a single photograph using self-augmented convolutional neural networks. *ACM Trans. Graph.*, 36(4), 2017. 1, 2
- [24] Zhengqin Li, Zexiang Xu, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. Learning to reconstruct shape and spatially-varying reflectance from a single image. In *SIGGRAPH Asia 2018 Technical Papers*, page 269. ACM, 2018. 1
- [25] 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. 3
- [26] Zhengqin Li, Ting-Wei Yu, Shen Sang, Sarah Wang, Meng Song, Yuhua Liu, Yu-Ying Yeh, Rui Zhu, Nitesh Gundavarapu, Jia Shi, et al. Openrooms: An open framework for photorealistic indoor scene datasets. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7190–7199, 2021. 3
- [27] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 6
- [28] 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. 1, 2
- [29] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Transactions on Graphics (TOG)*, 22(3):759–769, 2003. 2

- [30] Wes McDermott. *The PBR Guide*. Allegorithmic, 2018. 1
- [31] Maxine Perroni-Scharf, Kalyan Sunkavalli, Jonathan Eisenmann, and Yannick Hold-Geoffroy. Material swapping for 3d scenes using a learnt material similarity measure. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2034–2043, 2022. 3
- [32] PolyHaven. <https://www.polyhaven.com/>, 2023. 2, 3, 5
- [33] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021. 3, 8
- [34] Alexander Raistrick, Lahav Lipson, Zeyu Ma, Lingjie Mei, Mingzhe Wang, Yiming Zuo, Karhan Kayan, Hongyu Wen, Beining Han, Yihan Wang, Alejandro Newell, Hei Law, Ankit Goyal, Kaiyu Yang, and Jia Deng. Infinite photorealistic worlds using procedural generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12630–12641, 2023. 1, 3
- [35] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2):3, 2022. 2
- [36] Mike Roberts, Jason Ramapuram, Anurag Ranjan, Atulit Kumar, Miguel Angel Bautista, Nathan Paczan, Russ Webb, and Joshua M. Susskind. Hypersim: A photorealistic synthetic dataset for holistic indoor scene understanding. In *International Conference on Computer Vision (ICCV) 2021*, 2021. 3
- [37] ShareTextures. <https://www.sharetextures.com/>, 2023. 2, 3, 5
- [38] Prafull Sharma, Julien Philip, Michaël Gharbi, Bill Freeman, Fredo Durand, and Valentin Deschaintre. Materialistic: Selecting similar materials in images. *ACM Trans. Graph.*, 42(4), 2023. 1, 3
- [39] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. 6
- [40] Zhihang Song, Zimin He, Xingyu Li, Qiming Ma, Ruibo Ming, Zhiqi Mao, Huaxin Pei, Lihui Peng, Jianming Hu, Danya Yao, and Yi Zhang. Synthetic datasets for autonomous driving: A survey. 2023. 1, 3
- [41] TextureCan. <https://www.texturecan.com/>, 2023. 2, 3, 5
- [42] Giuseppe Vecchio. *Accelerating reality: virtual environment generation for outdoor robot navigation using deep learning*. PhD thesis, Università degli studi di Catania, 2023. 1, 3, 8
- [43] Giuseppe Vecchio, Simone Palazzo, and Concetto Spampinato. SurfacerNet: Adversarial svbrdf estimation from a single image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12840–12848, 2021. 2, 5, 6, 7, 8
- [44] Giuseppe Vecchio, Rosalie Martin, Arthur Roullier, Adrien Kaiser, Romain Rouffet, Valentin Deschaintre, and Tamy Boubekour. ControlMat: A Controlled Generative Approach to Material Capture. *arXiv preprint arXiv:2309.01700*, 2023. 1, 2, 4
- [45] Giuseppe Vecchio, Renato Sortino, Simone Palazzo, and Concetto Spampinato. MatFuse: Controllable Material Generation with Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 2, 5, 6, 8
- [46] V-Ray. V-Ray:MaterialLibrary. <http://www.vray-materials.de/>, 2023. 2
- [47] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2555–2563, 2023. 3
- [48] K. Yan, F. Luan, M. Hašan, T. Groueix, V. Deschaintre, and S. Zhao. Psdr-room: Single photo to scene using differentiable rendering. In *ACM SIGGRAPH Asia 2023 Conference Proceedings*, 2023. 3
- [49] Xilong Zhou and Nima Khademi Kalantari. Adversarial single-image svbrdf estimation with hybrid training. *Computer Graphics Forum*, 2021. 2
- [50] 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*, New York, NY, USA, 2022. Association for Computing Machinery. 1, 2
- [51] 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*, pages 1–9, 2022. 2
- [52] 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*, New York, NY, USA, 2023. Association for Computing Machinery. 2
- [53] Jingsen Zhu, Fujun Luan, Yuchi Huo, Zihao Lin, Zhihua Zhong, Dianbing Xi, Rui Wang, Hujun Bao, Jiayang Zheng, and Rui Tang. Learning-based inverse rendering of complex indoor scenes with differentiable monte carlo raytracing. In *SIGGRAPH Asia 2022 Conference Papers*. ACM, 2022. 3
- [54] Rui Zhu, Zhengqin Li, Janarбек Matai, Fatih Porikli, and Manmohan Chandraker. Irisformer: Dense vision transformers for single-image inverse rendering in indoor scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2822–2831, 2022. 3