DTDMat: A Comprehensive SVBRDF Dataset with Detailed Text Descriptions

Mufan Chen
Nanjing University, State Key Lab for Novel Software Technology
Nanjing, Jiangsu Province, China

Yanxiang Wang
Nanjing University, State Key Lab for Novel Software Technology
Nanjing, Jiangsu Province, China

Detao Hu
Nanjing University, State Key Lab for Novel Software Technology
Nanjing, Jiangsu Province, China

Pengfei Zhu
Nanjing University, State Key Lab for Novel Software Technology
Nanjing, Jiangsu Province, China

Jie Guo*
guojie@nju.edu.cn
Nanjing University, State Key Lab for Novel Software Technology
Nanjing, Jiangsu Province, China

Yanwen Guo
Nanjing University, State Key Lab for Novel Software Technology
Nanjing, Jiangsu Province, China

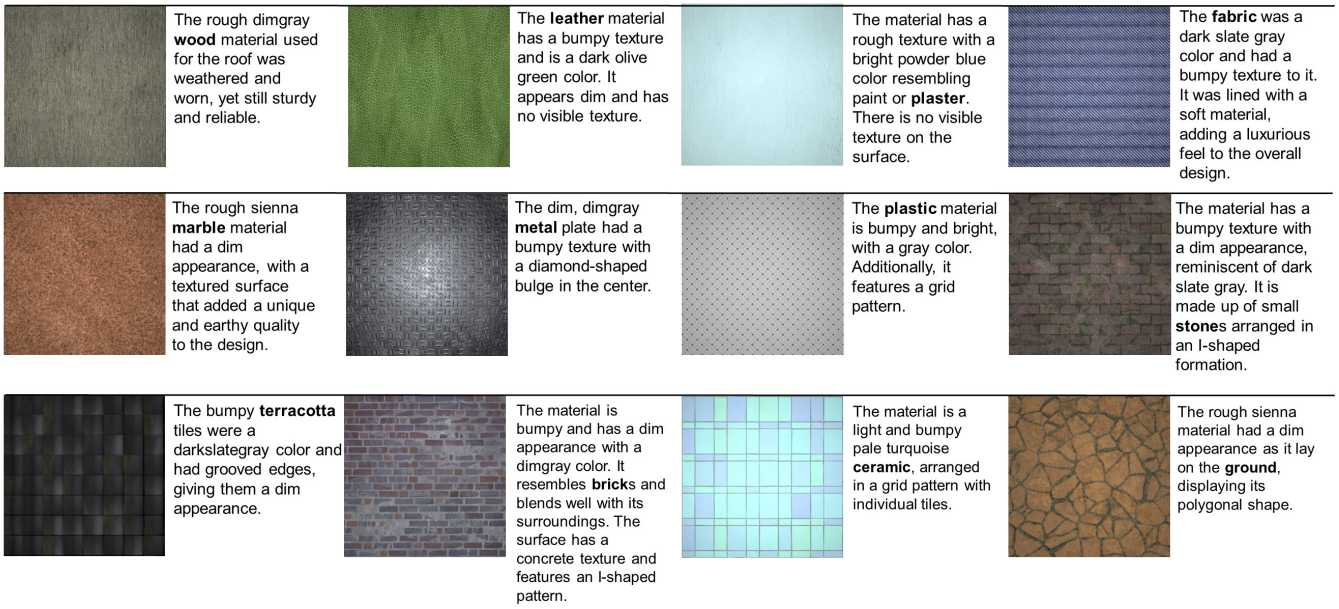


Figure 1: We present 12 different types of materials in DTDMat, with intrinsic attributes highlighted in bold. DTDMat encompasses a total of 20 intrinsic types, featuring 22 distinct texture patterns. Each material is accompanied by a detailed text description reflecting its properties, generated through our automatic annotation method.

Abstract

In this paper, we designed an automatic annotation tool to generate full descriptions to solve material datasets lacking essential text information challenge. This tool can extract six aspect tags from BRDF maps: intrinsic type, texture, color, roughness, lightness, and other relevant attributes. We applied this tool to both open-source material datasets and our own dataset to create DTDMat, which consists of 14,919 high-resolution Physically Based Rendering materials, each accompanied by a detailed text description. DTDMat covers 20 intrinsic material types and 22 texture structures. It stands out as the most diverse dataset in this domain and represents the largest texture dataset with associated text, offering a wide range of categories and diverse descriptions. We then trained a text-to-material generation framework based on DTDMat, yielding multiple generated BRDF maps that satisfy the input text.

CCS Concepts

• **Computing methodologies** → **Reflectance modeling**; *Machine learning*.

*Corresponding author

VRCAI '24, December 1–2, 2024, Nanjing, China

ACM ISBN 979-8-4007-1348-4/24/12
https://doi.org/10.1145/3703619.3706053

Keywords

PBR material generation, Diffusion Models, Text-to-Material

ACM Reference Format:

Mufan Chen, Yanxiang Wang, Detao Hu, Pengfei Zhu, Jie Guo, and Yanwen Guo. 2024. DTDMat: A Comprehensive SVBRDF Dataset with Detailed Text Descriptions. In *The 19th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry (VRCAI '24), December 1–2, 2024, Nanjing, China.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3703619.3706053

1 Introduction

Materials play a crucial role in the rendering pipeline, especially when it comes to creating realistic environments and objects. The Spatially Varying Bidirectional Reflectance Distribution Function (SVBRDF) describes the reflective properties of material surfaces, capturing how light interacts with different surface details. SVBRDF is widely used in Computer Graphics to reproduce accurate material reflectance, and it has significant applications in various fields, such as Gaming, Virtual Reality (VR), Furniture Modeling, and Cloth Design. In particular, Physically Based Rendering (PBR) materials are essential in VR, as they enable more immersive and visually convincing experiences by ensuring realistic lighting and shading interactions. Accurate material representation in VR not only enhances the realism of virtual environments but also plays a critical role in user immersion, making virtual worlds feel more tangible and lifelike.

Crafting virtual materials of superior quality requires proficiency across various aspects such as color, reflectivity, and metallic properties, often necessitating familiarity with diverse attributes and characteristics of materials. This endeavor demands considerable artistic acumen and expertise[McDermott 2018]. Since Goodfellow et al.[Goodfellow et al. 2020] propose Generative adversarial networks(GANs), generative tasks have been a popular researching direction. In particular, image generating tasks have a rapid development. Karras et al.[Karras et al. 2017, 2021] improve stability and quality on GANs. Parks et al.[Park et al. 2019] introduce style control in GANs. GANs are applied in different kinds of downstream tasks, like text-to-image[Ho et al. 2020; Ramesh et al. 2022] and even 3d shape generation[Li et al. 2019].

The realm of image generation from text highlights a pressing demand for datasets that facilitate material generation based on textual inputs[He et al. 2023; Memery et al. 2023]. Before the emergence of material generation tasks, datasets in the field of materials largely relies on a single synthetic dataset[Deschaintre et al. 2018], or company-owned large material libraries[Adobe 2024; Vecchio et al. 2023; Zhou et al. 2022] to solve material reconstruction scene[Guo et al. 2021; Kang et al. 2019; Ma et al. 2021]. These datasets only include SVBRDF maps lacking accompanying textual information. Newly available datasets[Deschaintre et al. 2023; Ma et al. 2023; Vecchio and Deschaintre 2024] pay attention to textual information, contributing to advancements in material research. MatSynth dataset [Vecchio and Deschaintre 2024] offers 4,069 PBR materials accompanied by tags rather than full textual descriptions. OpenSVBRDF[Ma et al. 2023] presents a measured SVBRDF database, featuring 1,000 high-quality near-planar materials spanning 9 intrinsic categories. The textual information in these datasets is insufficient to support training from text to material, comprising fragmented knowledge or isolated words. He et al.'s dataset[He et al. 2023] presents a PBR material dataset comprising over 2,500 samples. The dataset employs a text annotation method in which three tags are inserted into a fixed sentence pattern, resulting in overly uniform descriptive statements. The text annotation method only covers three aspects of materials, limited to textual knowledge. Text2fabric dataset[Deschaintre et al. 2023] encompasses over 3000 unique fabric textures, each accompanied by textual descriptions. However, the dataset is limited to fabric materials and entirely manually labeled, incurring significant costs and focusing solely on fabric intrinsic attributes.

The lack of diverse material datasets containing comprehensive textual descriptions poses a significant challenge. This gap between material and text data underscores a significant issue in the field. To address this disparity, we introduce DTDMat: A Comprehensive SVBRDF Dataset with Detailed Text Descriptions. DTDMat comprises 14,919 high-resolution Physically Based Rendering materials, each accompanied by crafted text descriptions reflecting their properties. These descriptions are generated through our automatic annotation tool. Each description within DTDMat encapsulates intrinsic type, texture, color, roughness, lightness, and other pertinent attributes, providing a holistic understanding of each material sample. Notably, DTDMat represents the largest texture dataset with associated text, offering a wide array of categories and diverse descriptions. DTDMat also emerges as the most diverse and largest-scale dataset in this domain, spanning 20 distinct material types and 22 texture structures,

We further leverage DTDMat to train a text-to-material generation framework, resulting in the generation of new materials that further enrich our dataset. Additionally, this framework can be utilized to generate materials desired by users based on input descriptions of the materials they seek. This significantly reduces dependency on artists and streamlines the complex process of material design.

In summary, our main contributions are as follows:

- A dataset containing over 14,919 high-resolution Physically Based Rendering materials, each accompanied by detailed text descriptions, covering 20 distinct material types and 22 texture structures.
- An automatic annotation method to generate accurate descriptions based on SVBRDFs, covering at least five aspects.

2 Related Work

2.1 Generative Models

Generative Adversarial Networks (GANs)[Goodfellow et al. 2020] study a collection of training examples and learn the probability distribution that generated them. Then GANs are able to generate more examples from the estimated probability distribution, displaying a strong generation capability to create high-fidelity images. Karras et al.[Karras et al. 2021] find careless signal processing causing aliasing in the generator network and derived generally applicable, small architectural changes that guarantee that unwanted information cannot leak into the hierarchical synthesis process. Zhu et al.[Zhu et al. 2019] propose the Dynamic Memory Generative Adversarial Network (DM-GAN) to generate high-quality images.

As for text-to-image tasks, the most popular methods are diffusion models[Ho et al. 2020; Ramesh et al. 2022]. There are more efficient sampling strategies[Liu et al. 2022; Song et al. 2020] to reduce the number of required sampling steps greatly and improve image generation performance. Rombach et al.[Rombach et al. 2022] propose performing denoising process in learned compact latent space instead of pixel space, achieve high-resolution synthesis results, called Stable Diffusion.

Control diffusion process is crucial for creations. There are research focusing on add lightweight multi-modal controllability without the requirements of extensive data and high computational power. Hu et al.[Hu et al. 2021] propose Low-Rank Adaptation(LoRA), which freezes the pretrained model weights and injects trainable rank decomposition matrices into each layer of the Transformer architecture. This method greatly reduced the number of trainable parameters for downstream tasks. Zhang et al.[Zhang et al. 2023] add spatial conditioning controls to large, pretrained text-to-image diffusion models, which locks the production-ready large diffusion models, and reuses their deep and robust encoding layers pretrained with billions of images as a strong backbone to learn a diverse set of conditional controls. Ye et al[Ye et al. 2023] presents an effective and lightweight adapter (IP-Adapter) to achieve image prompt capability for the pretrained text-to-image diffusion models, which decouples cross-attention mechanism that separates cross-attention layers for text features and image features.

The material reconstruction method can only restore the reflective properties of the material based on the input images. With the increasing popularity of diffusion models, researchers are now focusing on utilizing the generative approach in material studies to generate different SVBRDF maps directly.

Guo et al.[Guo et al. 2020] presents MaterialGAN, a deep generative convolutional network based on StyleGAN2, to reconstructing spatially-varying BRDFs from a small set of image measurements. Zhou et al.[Zhou et al. 2022] propose TileGen, a generative model for SVBRDFs that is specific to a material category, always tileable, and optionally conditional on a provided input structure pattern. Sam et al.[Sartor and Peers 2023] provides a backbone based on a diffusion model and three conditional SVBRDF diffusion models to address different types of incident lighting. Hu et al.[Hu et al. 2023] demonstrate a multi-modal node graph generation neural architecture for high-quality procedural material synthesis which can be conditioned on different inputs (text or image prompts), using a CLIP-based encoder. Guerrero et al.[Guerrero et al. 2022] presents MatFormer, a generative model that can produce a diverse set of high-quality procedural materials with complex spatial patterns and appearance. Vecchio et al.[Vecchio et al. 2023] propose ControlMat, a method which, given a single photograph with uncontrolled illumination as input, conditions a diffusion model to generate plausible, tileable, high-resolution physically-based digital materials. Xin et al.[Xin et al. 2024] introduce a novel diffusion-based generative framework (DreamPBR) designed to create spatially-varying appearance properties guided by text and multi-modal controls, providing high controllability and diversity in material generation. Memery et al.[Memery et al. 2023] presents a model for generating Bidirectional Reflectance Distribution Functions (BRDFs) from textual descriptions, enabling real-time material changes in 3D environments. The model is trained using a semi-supervised approach and further tuned unsupervised for generating MDL materials parameters.

Table 1: Comparison results between DTDMat and other datasets from various perspectives.

Name	Material	Year	Intrinsic	Texture	Tag	Text
DTDMat	14,919	2024	20	22	65,700	14,919
MatSynth	4,069	2024	14	-	21,737	572
He et al.'s	2,500	2023	-	-	3,000	2,500
Text2fabric	3,000	2023	1	-	-	15,000
OpenSVBRDF	1,000	2023	9	-	-	-

2.2 Material Datasets

Early material datasets[Deschaintre et al. 2018] are focused on material reconstruction tasks and don't include textual information. Only recently have new datasets[Deschaintre et al. 2023; He et al. 2023] been introduced that incorporate textual descriptions. Text2fabric PBR dataset by Deschaintre et al.[Deschaintre et al. 2023] is limited to fabrics. Subsequently, Vecchio et al.[Vecchio and Deschaintre 2024] introduce MatSynth dataset, which expands beyond fabrics to include material PBR includes 18 different attribute categories, along with corresponding tag information. He et al.[He et al. 2023] demonstrate a PBR material dataset with over 2500 samples and a texture annotation technique.

However, the textual information provided by these datasets is limited in both the quality of text descriptions or the diversity of materials. Additionally, the annotation process requires manual intervention, resulting in significant costs. Furthermore, the original diffusion model is trained on natural images, highlighting the significant reliance of text-to-material tasks on large-scale material text datasets.Therefore, we propose DTDMat along with an automated annotation tool to address this issue.

3 The Dataset

Datasets in the material domain are scarce, especially those related to text and materials. Most datasets only include material texture maps, with some providing related tags but lacking complete descriptive sentences. DTDMat is designed to support material generation from text and also could be used in other material related tasks, like material reconstruction. DTDMat is larger than previously available dataset with 1K high resolution and realistic materials along with an accurate description. In this section, we introduce details in DTDMat and compare our dataset with other public available dataset.

3.1 Composition of Dataset

DTDMat stands out as the largest material dataset with text, comprising 14,919 materials. we render each material under three distinct lighting conditions: directional light, point light, and environment light, generating a total of 89,514 render images using specular and metallic two workflows. For text information, each material is accompanied by at least five tags and a complete sentence describing its properties. While Text2fabric[Deschaintre et al. 2023] provides five sentences for each material, it incurs significant manual effort for description and ranking. In contrast, our automatic

method is more convenient and efficient without human intervention. He et al.[He et al. 2023] utilize a fixed sentence format where three tags are inserted to generate material descriptions. DTDMat's descriptions, generated by a Large Language Model (LLM), exhibit greater naturalness and diversity, transcending the constraints of predefined formats.

In DTDMat, each material is assigned from six aspects of labels, and a matching description is generated based on these labels. These six aspects are: intrinsic attribute, texture, lightness, roughness, color and others. In the next subsection, we will provide detailed information about our tags.

DTDMat encompasses 20 distinct intrinsic attributes, as illustrated in Figure 2, presenting a more expansive coverage in comparison to other datasets. Text2fabric[Deschaintre et al. 2023] solely focuses on fabric material, while OpenSVBRDF[Ma et al. 2023] provides 1,000 texture maps sourced from real-world references spanning 9 material categories. MatSynth[Vecchio and Deschaintre 2024] comprises 14 intrinsic attributes. DTDMat not only incorporates all previously existing intrinsic attributes from these datasets with substantially larger material samples but also introduces novel ones such as glass, lime, and crystal. Moreover, we introduce 22 different texture labels not found in other datasets, including both regular and irregular texture structures.

Table 1 displays the comparison results between DTDMat and other datasets from various perspectives. This further underscores that DTDMat boasts the largest scale in this field, covers the most intrinsic attributes, exhibits the richest texture structure categories, and provides abundant tag information and textual descriptions.

3.2 Material Tags

We design material tags from six different aspects: intrinsic attributes, texture, color, smoothness, brightness and others. The intrinsic attributes of a material determines its origin composition. It is a challenge to add categories of material properties. Fortunately, MatSynth dataset provides a reference for us. MatSynth has categorized the collected materials into 14 distinct categories and OpenSVBRDF also provides 9 different kinds of intrinsic attributes real-world materials' SVBRDFs with a network. We organized two datasets and add some new material intrinsic attributes, like lime, glass and crystal. Consequently, we expanded upon the dataset we organized, resulting in 20 categories of material intrinsic attributes. We display examples of all texture labels in the appendix.

DTDMat comprises a total of 22 texture labels, encompassing various types of texture structure. And DTDMat is the first dataset to provide texture patterns based on material samples rather than natural images. We reference texture labels from existing datasets such as DTD[Cimpoi et al. 2014] and He et al.[He et al. 2023], which are based on natural images. However, we make modifications by adding and removing certain texture labels based on our dataset. Ultimately, we identify the following 22 texture labels: bulge, circle, diamond, flat, flower, grid, grooved, herringbone, I-shaped, lined, marbled, mottled, pattern, peeling, pitted, polygonal, rectangular, scaly, stratified, woven, zigzagged, and unknown. The unknown labels contain materials whose textures are challenging to describe. DTDMat provides a more wider range of texture labels and each texture label has over 500 material samples averagely. The least number of texture label categories is scaly, which also has 44 material samples. In contrast to He et al.[He et al. 2023], who employ natural texture images from datasets like DTD[Cimpoi et al. 2014], DTDMat's classification is centered around material samples rather than natural images, making it better suited for research in the field of materials. The specific texture material quantity distribution information is placed in appendix.

In addition to the intrinsic attributes and texture labels, we also introduce labels for color, smoothness, lightness, and others. These tags are predicted through traditional methods rather than neural networks, which will be discussed in the next section.

Color is a visual effect of light produced by the eyes, brain, and our life experiences. Determining the color of a material is a challenging problem. A slight color change is hard for human to describe. He et al.[He et al. 2023] choose 14 basic colors that are commonly known to people: red, orange, yellow, green, cyan, blue, purple, pink, brown, grey, black, white, golden and beige and label materials by hand within 14 colors. We aim to develop a hands-free automated method that can encompass a wide range of colors. For material, color attribute is mostly related to Base Color map of the material. In Figure 3, we demonstrate different colors obtained through our method and corresponding material render images. The result shows our method could process a wide variety of color.

We define four different levels of roughness for the roughness tag: bumpy, rough, textured, and smooth. These labels are generated based on the Roughness map. Figure 4 shows four materials with different roughness levels: the rendered images are in the first row, and the roughness maps are in the second row. From left to right, their standard deviations are 0.22761, 0.05837, 0.00295, and 0.0. The material's surface has a diamond texture when examined closely, which differs from that of smooth material.

For brightness labels, we categorize them into two types: dim and bright. We ascertain the material's brightness from the Base Color map, determined by its luminance value.

In addition to the aforementioned tags obtained from SVBRDF maps or render images, we can also gather valuable information when collecting these materials such as material's name. We could obtain the functionality of material from their name, like roof, wall or floor. However, these names often include meaningless words, such as letters used for sorting and terms serving other purposes. Therefore, we implemented a filtering process to extract only the relevant words and retain them as miscellaneous tags.

4 Automatic Annotation Tool

Text2fabric dataset[Deschaintre et al. 2023] contains of 15,000 natural language descriptions with 3,000 fabric materials. It requires native English speakers with normal color vision and familiarity with fashion or design to describe materials. Participants underwent training and a qualification test, where overly simplistic descriptions were excluded. It found quality of descriptions highly dependent on the describer and established continuous data verification protocol involving manual auditing. It ended up with 15,461 valid descriptions and 3,706 invalid ones (19.3% rejection rate). The process is highly labor-intensive, involving substantial human effort in material description and manual verification for accuracy. We propose an automatic annotation tool to reduce human resource

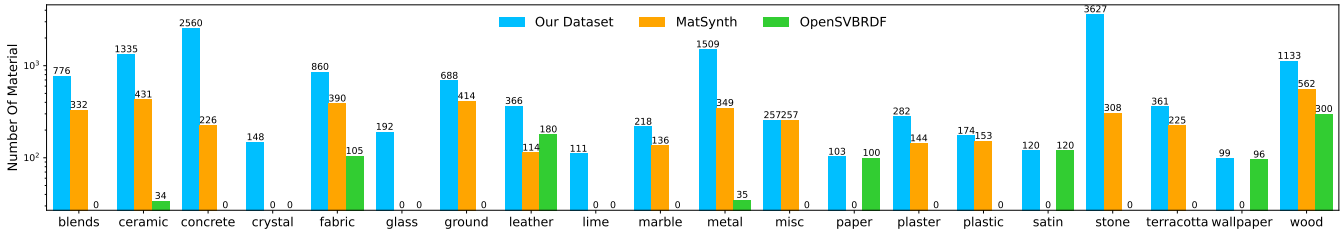


Figure 2: Distribution of intrinsic attributes across DTDMat, MatSynth[Vecchio and Deschaintre 2024], and OpenSVBRDF[Ma et al. 2023]. DTDMat offers a broader range of intrinsic attributes compared to any other dataset.

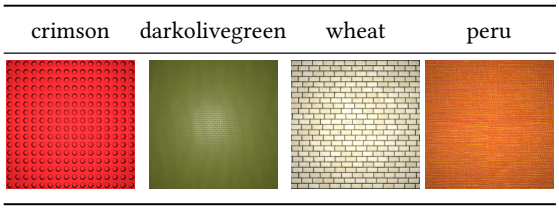


Figure 3: Different colors obtained through our method and corresponding material render images. The result shows our method could process a wide variety of color.

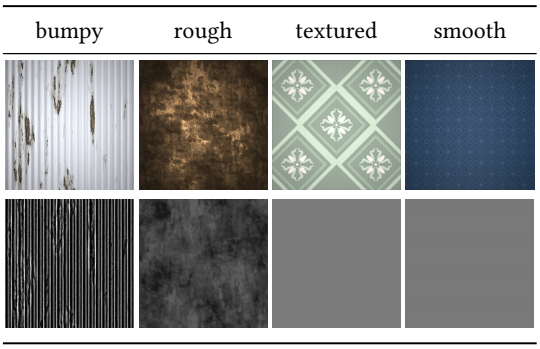


Figure 4: Four materials with different roughness levels are displayed: the rendered images are in the first row, and the roughness maps are in the second row.

costs, enabling annotation of various material types and generating descriptions of the materials. Our automatic annotation tool is primarily divided into two stages. The first stage is acquiring material tags from six aspects. Process SVBRDF textures based on material to generate six-dimensional label information, including: color, roughness, lightness, functionality, intrinsic type, texture and other. In the second stage, feeding the generated tags into the LLM model, ask it to produce a description of the material with our designed prompt. This section will provide a detailed explanation of how we obtain a detailed textual description from material.

4.1 Stage I: Tags Generation

The initial phase involves generating tags from materials. We have developed six specific tags: intrinsic property, texture, roughness, luminosity, color, and miscellaneous. Each tag is produced through a distinct methodology. Figure 5 provides a visual representation of this process using a material sample. Roughness tag is determined by the standard deviation of the Roughness map, it get "bumpy" tag

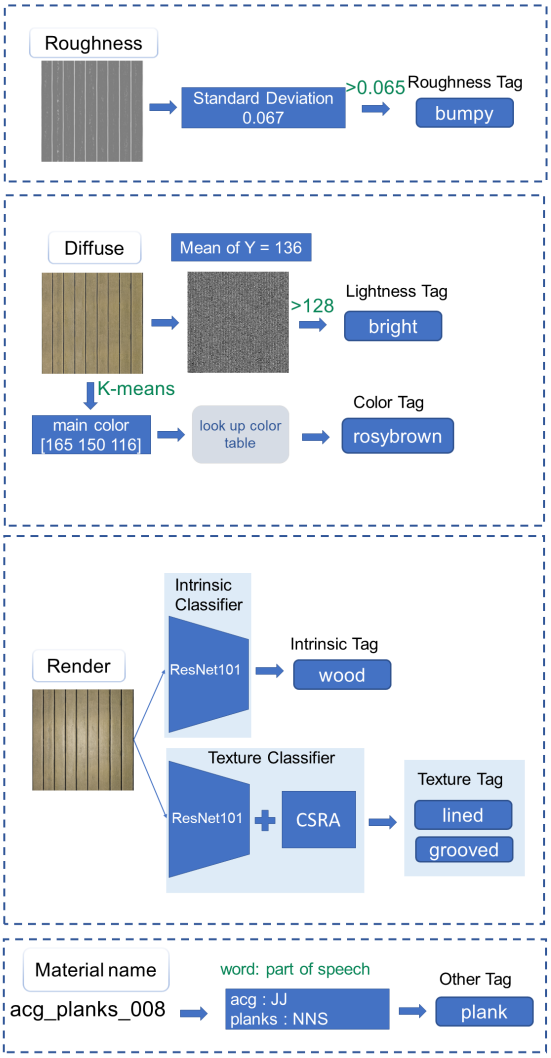


Figure 5: Overview of Stage I.

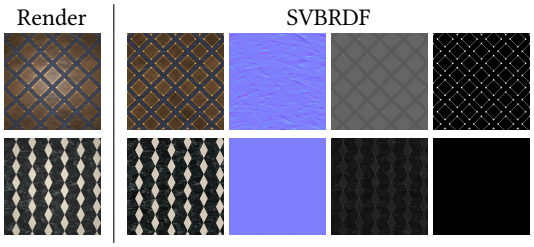


Figure 6: Materials with the same texture label "diamond" is influenced by different BRDF maps.

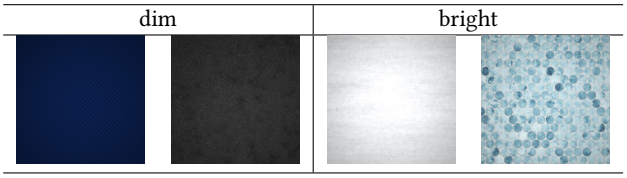


Figure 7: Rendered images under consistent lighting conditions.

for the standard deviation is greater than bumpy threshold. Lightness Tag is derived from the luminance level of the Base Color map. Mean of luminance is 138, which is greater than 128, so it get bright lightness tag. For color tag, we employ the K-means algorithm to determine the primary color of the Base Color map. Subsequently, we reference a color table to ascertain the closest matching name. In this specific instance, the predominant color is represented by the RGB values (165,150,116), which, upon consultation with the CSS3 color table, is identified as "rosybrown".

For the intrinsic attributes, we utilize dedicated classification networks based on rendered images of the material as input. This is because all BRDF maps jointly determine the intrinsic attributes of the materials, rather than just a single BRDF map. Then intrinsic attribute tag is determined by the highest scores obtained, signifying the dominant characteristics of the material.

Similarly, texture tags are not affected solely by one BRDF map. Figure 6 illustrates two scenarios in which materials with the same texture label "diamond" is influenced by different BRDF maps. In the first row, "diamond" texture label is influenced by the base color, roughness, and metallic maps. The second row is influenced by the base color and roughness maps. Therefore, we use the rendered image as the input for the texture label classification network to cover all possible situations. Besides, texture tags may encompass multiple descriptors. So any predicted scores surpassing a predefined threshold are deemed significant. In this particular sample, two texture tags were identified: lined and grooved. In the provided sample, the material is denoted as "acg_plank_008". Following the extraction of individual words and the omission of numerical sequences, the Natural Language Toolkit (NLTK) is employed for part-of-speech tagging. The analysis reveals that "acg" is tagged as an adjective (JJ), indicating its non-essential nature, and is thus disregarded. Conversely, "planks" is tagged as a noun (NN), signifying its significance, and is consequently retained as part of the miscellaneous tag.

In the intrinsic attribute block, we classify our materials into 20 different intrinsic types and train a label classifier using ResNet101[Deng et al. 2009; He et al. 2016], which determines the intrinsic type that materials belong to. The classifier requires a render image as inputs and outputs prediction scores for 20 classes of intrinsic properties. Subsequently, we assign the label of the intrinsic property to the material based on the category with the highest score.

Regarding the texture tag, we also employ ResNet101[He et al. 2016] trained on 22 texture classes that we designed. Given the potential for multiple texture tags to be applicable simultaneously, we add a CSRA (Cross-Scale Residual Attention)[Zhu and Wu 2021] module into ResNet to bolster the performance of multi-label classification. Texture tags are retained if their scores surpass the texture threshold.

Concerning the color label, we use k cluster algorithm to find the most related RGB value. We set the number of cluster centers to one or two and ran the K-means clustering algorithm ten times with different initial centroids, selecting the best clustering result. From this, we obtain the RGB values of the clustered colors. Then, we lookup web standard color table help us to find the name of exact RGB value, resulting in the color tag. If the color table does not contain this RGB value, we then assign the color name closest in Euclidean distance within the color space. In this method, we can distinguish subtle color shifts and find the exact color name with no human involvement. This method encompasses a wider spectrum of colors in contrast to He et al.'s approach, which focuses solely on a limited number of common color categories. Additionally, it does not require manual intervention to obtain specific color information.

For the roughness tag, we calculate the σ value, which is determined using the standard squared difference formula based on the roughness map. The σ indicates the magnitude of change. A larger *sigma* signifies a more pronounced alteration in roughness map. According to experimental testing, three thresholds are established to categorize the roughness of texture maps into four levels: bumpy, rough, textured, and smooth. A standard deviation greater than 0.065 is classified as bumpy, between 0.065 and 0.003 as rough, between 0.003 and 0.001 as textured, and below 0.001 as smooth. In lightness block, we first convert the RGB values of the Base Color map into YUV encoding, then compute the average luminance. A threshold is set, above which the material is considered bright, and below which it is considered dim. Figure 7 displays rendered images under consistent lighting conditions but varying brightness levels.

To enhance our tags, we add miscellaneous tag to incorporate additional information obtained during material collection, such as material names. Some redundant words, such as identifiers or abbreviations, need to be filtered out. Therefore, we design a process to filter these textual information, retaining only relevant vocabulary. We utilize NLTK[nlt 2024] to analyze the part-of-speech for each word. Based on their part-of-speech, we extract meaningful words while discarding irrelevant content. Nouns and adjectives are saved, while other parts-of-speech such as verbs and conjunctions are discarded.

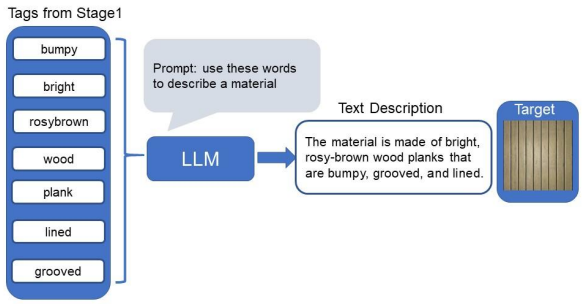


Figure 8: Overview of Stage II.

4.2 Stage II: Generating Textual Description

In Stage I, our automatic annotation tool generates relevant tags from six different aspects using various methods, ensuring that each category has at least one word. Our ultimate goal is to obtain descriptions of the materials. Therefore, the next subsection will introduce Stage II in our automatic annotation tool. Regarding the acquisition of textual descriptions for materials, Valentin et al.[Deschaintre et al. 2023] enlisted numerous participants in characterizing the appearance of 3,000 fabric materials. He et al. generated text based on color, texture and material labels. Fill three labels into a fixed format: "[color] [material] with/arranged in [texture] pattern", which constrained the diversity of sentence structures. We employed manual description methods initially. However, ensuring consistency in descriptions with materials proves challenging, with excessively prolonged annotation times and overly uniform descriptions from the same individual for different materials.

Faced with this dilemma, with the rise of Large Language Models (LLM), we turn to LLM to aid in generating descriptions. Thus, we needed to input the tag information generated in the first stage into the LLM, specifying appropriate prompts to guide the LLM in generating accurate descriptions. Figure 8 illustrates an example, where the tag information generated in the first stage forms the input to the LLM, with the prompt "use these words to describe a material". It can be observed that the output from the LLM aligns with the material, with the rendered image of the material placed on the far right as the target. It is worth noting that to prevent ambiguity in the input tag words, the term 'texture pattern' will be added after the input of texture tags.

Our automatic annotation tool generates six distinct labels from material and utilizes LLM to generate sentence, significantly enhancing sentence diversity. This tool eliminates the need for manual intervention and ensures greater accuracy in the annotation results compared to those obtained through purely manual annotation methods[Deschaintre et al. 2023]. Comparative experiments will be discussed in detail in the experimental section.

5 Experiments

5.1 Annotation Tool

We compare our automatic annotation tool with the annotation method used by He et al. [He et al. 2023]. The annotation method employed by He et al. initially involves manually labeling the material colors, which are restricted to a predefined set of 14 common colors. However, with our automatic annotation tool, we can process all colors defined in the CSS color table, encompassing a total of 147 colors (17 standard colors plus over 130 additional ones). He et al. labeled the colors in Figure 3 as red, green, white, and orange. The colors generated by our annotation tool are more accurate. We also attempt to use CLIP similarity [Radford et al. 2021], but the similarity scores between annotation results from different methods and material render images are very close. We believe this is due to the significant discrepancy between the natural images in CLIP's dataset [Radford et al. 2021] and those in the material domain.

Figure 9 compares our annotation method with He et al.'s method[He et al. 2023] using identical input material. In the rightmost column, we present material rendering images along with their corresponding BRDF maps. The small images on the right are arranged from left to right, top to bottom, representing the roughness map, normal map, base color Map, and metallic Map respectively. The results show our tool offers diverse and rich content from various perspectives, free from a fixed sentence structure.

5.2 Evaluation for DTDMat

To validate the effectiveness of DTDMat, we train the Text2Mat framework proposed by He et al. [He et al. 2023] on DTDMat, which is a text-based material generation network grounded on Stable Diffusion[Rombach et al. 2022]. We compare the material generation capabilities of the Text2Mat framework based on both the original dataset and DTDMat. Additionally, we include the latest and popular material-to-text website named Polycam[pol 2024] in our comparisons.

Figure 10 displays Text2Mat trained DTDMat and make comparisons against original Text2Mat and an online text-to-generation website[pol 2024]. The input text encompasses distinct colors, intrinsic attributes and textures. We put the generated SVBRDF maps in appendix. Here we displays the rendered images using the generated SVBRDF maps in Base Color/Metallic workflow. These results indicate that our dataset covers a broader and more diverse range of information, particularly in terms of intrinsic attributes, texture and color. Our results for the first input accurately capture the diamond appearance, particularly with respect to the leather quality. In contrast, He et al. fail to generate any texture for the leather. Moreover, Polycam is inadequate in representing both texture and intrinsic attributes. The second text input demonstrates our dataset's capability to effectively handle complex texture structures, notably evident in its ability to process 'golden' cases, a capability exclusive to our dataset. Results obtained from Polycam tend to skew towards yellow hues, while those from He et al. exhibit direct recognition failures. The last text input describes a metal material. Our outcomes produce metallic maps, in contrast to others' maps that appear black. Moreover, our results provide a more accurate depiction of the grid structure and other information.

The experimental results demonstrate that DTDMat encompasses a broad spectrum of colors, intrinsic attributes, and texture patterns. This diversity aids GANs in acquiring a deeper understanding of materials, enabling them to generate realistic textures while handling various input categories.

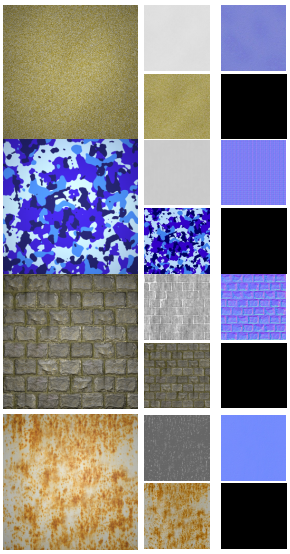
He et al.[He et al. 2023]	Ours	Render and BRDF
light yellow carpet	The rough, tan carpet was flat and bright, adding warmth to the room.	
blue camouflage	This material is a smooth, dark blue fabric with a dim texture, designed in a camouflage pattern.	
gray stone bricks with I-shaped pattern	The material is bumpy and dim in texture, resembling bricks in shape and color. It has a dimgray hue and a unique I-shaped pattern.	
rust old weathered painted rusty metal with pitted pattern	This material is a rough, bright sienna-colored metal with a rusty, mottled appearance.	

Figure 9: Comparision to Text2Mat[He et al. 2023] using identical input material.

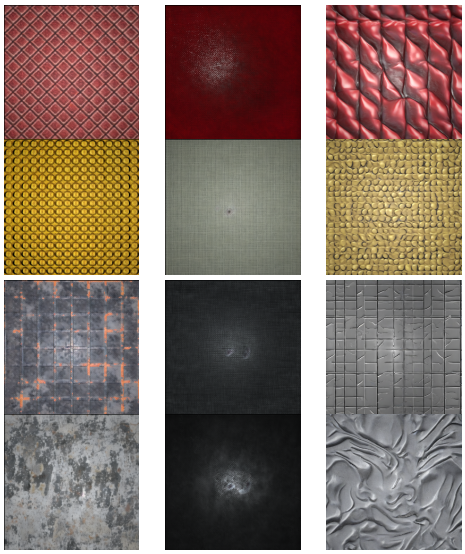
Input Text	Ours	He et al.	Polycam
The material is bumpy, bright, and resembles red with a leather texture. It also has diamond-like appearance.			
The plastic material is bumpy and has a grid pattern. It forms a slight bulge in the shape of a circle. The surface is bright and has a goldenrod color.			
The metal material is darkslategray in color and has a bumpy texture, with a grid pattern etched into its surface.			
The plastic material is bumpy and has a grid pattern. It forms a slight bulge in the shape of a circle. The surface is bright and has a goldenrod color.			

Figure 10: Comparisons against Txt2Mat[He et al. 2023] and an online text-to-generation website[pol 2024].

6 Conclusion and Future Work

We present a comprehensive material dataset with diverse materials and detailed automated textual descriptions. Our tool generates accurate descriptions from material BRDF maps across six aspects. Future work will enhance the tool's versatility and robustness, and explore suitable methods to measure text accuracy in materials. To refine roughness and brightness thresholds, we plan to incorporate a questionnaire to determine appropriate values. We also try to address the issue of uneven distribution across various categories in DTDMat through the fine-tuned diffusion model. Figure 11 shows the rendering results of four materials generated from the input text. The fine-tuned model [He et al. 2023] enhances the material's texture content by adding diamond and circle textures, upon which we can expand our dataset. There are still many challenges and possibilities in the research of the material domain. We will continue to explore and experiment based on DTDMat.

References

2024. ntlk. https://www.nltk.org/.
2024. polycam. https://poly.cam/tools/ai-texture-generator.
Adobe. 2024. Substance 3D Assets. https://substance3d.adobe.com/assets/.
Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. 2014. Describing textures in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3606–3613.
Lennart Demes. 2024. ambientCG. https://ambientcg.com.
Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.
Valentin Deschaintre, Miika Aittala, Fredo Durand, George Drettakis, and Adrien Bousseau. 2018. Single-image svbrdf capture with a rendering-aware deep network. *ACM Transactions on Graphics (ToG)* 37, 4 (2018), 1–15.
Valentin Deschaintre, Diego Gutierrez, Tamy Boubekeur, Julia Guerrero-Viu, and Belen Masia. 2023. The visual language of fabrics. (2023).
Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. *Commun. ACM* 63, 11 (2020), 139–144.
Paul Guerrero, Miloš Hašan, Kalyan Sunkavalli, Radomír Měch, Tamy Boubekeur, and Niloy J Mitra. 2022. Matformer: A generative model for procedural materials. *arXiv preprint arXiv:2207.01044* (2022).
Jie Guo, Shuichang Lai, Chengzhi Tao, Yuelong Cai, Lei Wang, Yanwen Guo, and Ling-Qi Yan. 2021. Highlight-aware two-stream network for single-image SVBRDF acquisition. (2021).
Yu Guo, Cameron Smith, Miloš Hašan, Kalyan Sunkavalli, and Shuang Zhao. 2020. MaterialGAN: Reflectance capture using a generative SVBRDF model. *arXiv preprint arXiv:2010.00114* (2020).
Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
Zhen He, Jie Guo, Yan Zhang, Qinghao Tu, Mufan Chen, Yanwen Guo, Pengyu Wang, and Wei Dai. 2023. Text2Mat: Generating Materials from Text. (2023).
Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems* 33 (2020), 6840–6851.
Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685* (2021).
Yiwei Hu, Paul Guerrero, Milos Hasan, Holly Rushmeier, and Valentin Deschaintre. 2023. Generating Procedural Materials from Text or Image Prompts. In *ACM SIGGRAPH 2023 Conference Proceedings*. 1–11.
Kaizhang Kang, Cihui Xie, Chengan He, Mingqi Yi, Minyi Gu, Zimin Chen, Kun Zhou, and Hongzhi Wu. 2019. Learning efficient illumination multiplexing for joint capture of reflectance and shape. *ACM Trans. Graph.* 38, 6 (2019), 165–1.
Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2017. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196* (2017).
Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2021. Alias-free generative adversarial networks. *Advances in neural information processing systems* 34 (2021), 852–863.
Xiao Li, Yue Dong, Pieter Peers, and Xin Tong. 2019. Synthesizing 3D shapes from silhouette image collections using multi-projection generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5535–5544.
Luping Liu, Yi Ren, Zhijie Lin, and Zhou Zhao. 2022. Pseudo numerical methods for diffusion models on manifolds. *arXiv preprint arXiv:2202.09778* (2022).
Xiaohe Ma, Kaizhang Kang, Ruisheng Zhu, Hongzhi Wu, and Kun Zhou. 2021. Free-form scanning of non-planar appearance with neural trace photography. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–13.
Xiaohe Ma, Xianmin Xu, Leyao Zhang, Kun Zhou, and Hongzhi Wu. 2023. OpenSVBRDF: A Database of Measured Spatially-Varying Reflectance. *ACM Transactions on Graphics (TOG)* 42, 6 (2023), 1–14.
Wes McDermott. 2018. The PBR Guide.
Sean Memery, Osmar Cedron, and Kartic Subr. 2023. Generating Parametric BRDFs from Natural Language Descriptions. 42, 7 (2023), e14980.
Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. 2019. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2337–2346.
Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. arXiv:2103.00020 [cs.CV] https://arxiv.org/abs/2103.00020
Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125* 1, 2 (2022), 3.
Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695.
Sam Sartor and Pieter Peers. 2023. Matfusion: A Generative Diffusion Model for SVBRDF Capture. *SIGGRAPH Asia 2023 Conference Proceedings* (2023).
ShareTextures. 2024. ShareTextures. https://www.sharetextures.com.
Jiaming Song, Chenlin Meng, and Stefano Ermon. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502* (2020).
3D Textures. 2024. 3D Textures. https://3dtextures.me.
Giuseppe Vecchio and Valentin Deschaintre. 2024. MatSynth: A Modern PBR Materials Dataset. *arXiv preprint arXiv:2401.06056* (2024).
Giuseppe Vecchio, Rosalie Martin, Arthur Roullier, Adrien Kaiser, Romain Rouffet, Valentin Deschaintre, and Tamy Boubekeur. 2023. ControlMat: A Controlled Generative Approach to Material Capture. *arXiv preprint arXiv:2309.01700* (2023).
Linxuan Xin, Zheng Zhang, Jinfu Wei, Ge Li, and Duan Gao. 2024. DreamPBR: Text-driven Generation of High-resolution SVBRDF with Multi-modal Guidance. *arXiv preprint arXiv:2404.14676* (2024).
Hu Ye, Jun Zhang, Sibo Liu, Xiao Han, and Wei Yang. 2023. Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arXiv:2308.06721* (2023).
Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3836–3847.
Xilong Zhou, Milos Hasan, Valentin Deschaintre, Paul Guerrero, Kalyan Sunkavalli, and Nima Khademi Kalantari. 2022. Tilegen: Tileable, controllable material generation and capture. In *SIGGRAPH Asia 2022 conference papers*. 1–9.
Ke Zhu and Jianxin Wu. 2021. Residual attention: A simple but effective method for multi-label recognition. In *Proceedings of the IEEE/CVF international conference on computer vision*. 184–193.
Minfeng Zhu, Pingbo Pan, Wei Chen, and Yi Yang. 2019. Dm-gan: Dynamic memory generative adversarial networks for text-to-image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5802–5810.

A Limitations

Our automatic annotation tool has limitations, particularly in handling material images with multiple colors. We shows failure examples in figure 12. The rendered image on the left with color labels generated by our tool and manual labeled annotations from He et al.'s[He et al. 2023]. Our annotation tool misses pattern colors such as red and blue.

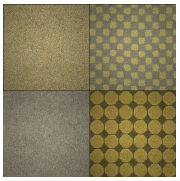
Input Text	Generate Results
The rough, tan carpet was flat and bright, adding warmth to the room.	

Figure 11: The rendering results of four generated materials from the same input text.

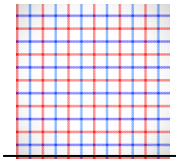
Render	Color Label
	whitesmoke white, blue and red

Figure 12: An example of annotation failure in the color tags of a material.

B DTDMat Source

DTDMat not only includes the open-source datasets MatSynth [Vecchio and Deschaintre 2024], OpenSVBRDF [Ma et al. 2023], and He et al. [He et al. 2023], but also comprises a substantial collection of materials from various open-source websites [Textures 2024], [Demes 2024], [pol 2024], [ShareTextures 2024]. Each material contains seven BRDF maps: diffuse, normal, specular, metallic, roughness, base color, and glossiness. It can be rendered using either the Diffuse/Specular and Base Color/Metallic workflows.

C DTDMat Examples

In this section, we show examples in DTDMat according to texture and intrinsic type labels. For texture, figure 13 displays the distribution of materials across these texture labels. Figure 14 displays an example for each texture label. Figure 15,16,17 show materials for 20 different intrinsic types with text descriptions generated by our automatic annotation tool. The rendering results use the Base Color/Metallic workflow and display four SVBRDF maps: base color, normal, roughness, and metallic, used in the rendering workflow.

D Material Generative Results

Based on the same input text, Figure18 displays generated BRDF maps from the finetuned stable diffusion model, which utilizes our DTDMat, Txt2Mat[He et al. 2023], and Polycam[pol 2024]. The rendered images are using the Base Color/Metallic workflow.

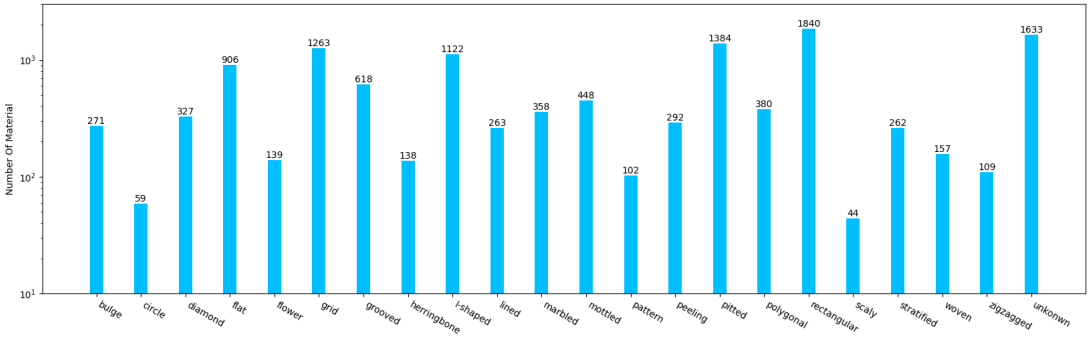


Figure 13: Distribution of materials across textures types. We designed a series of 22 texture labels based on material texture structure for classification purposes.

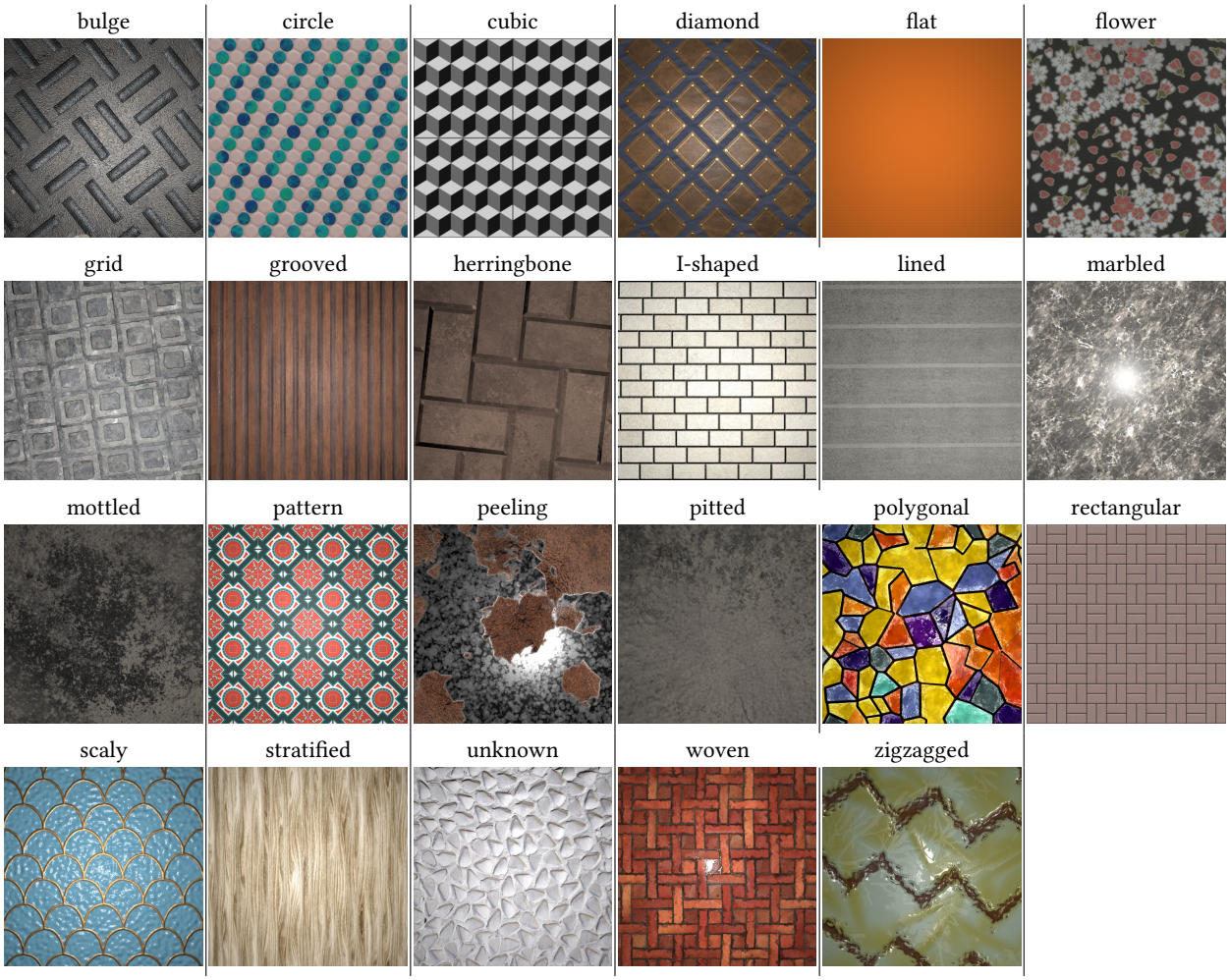


Figure 14: Render examples for texture labels

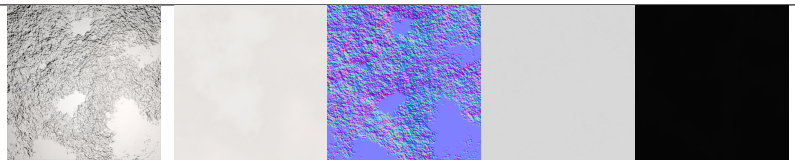
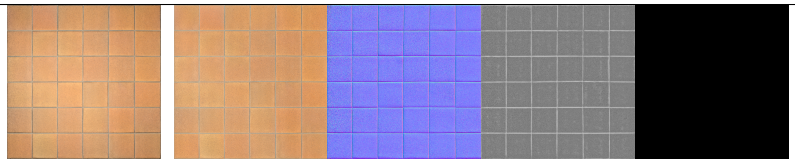
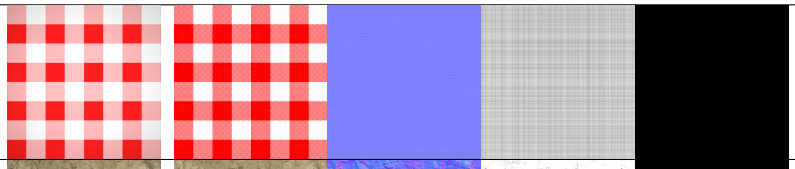
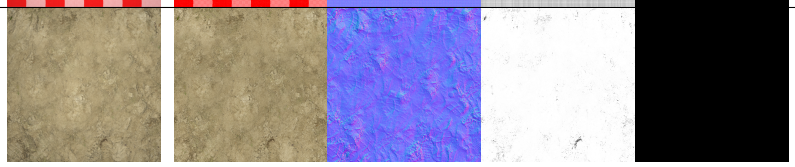
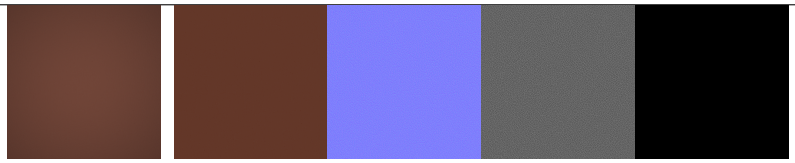
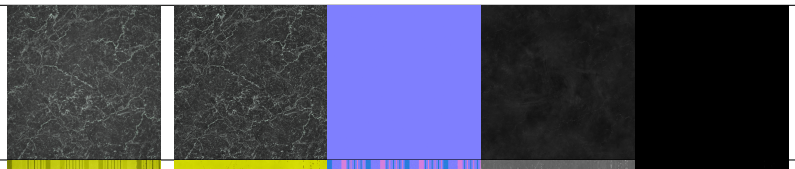
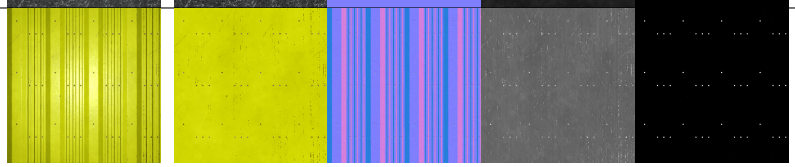
Text	Render	SVBRDF
The material can be described as having a textured surface with a light, gainsboro hue, complemented by subtle lime undertones.		
The material is bumpy to the touch, with a dim appearance in a dim gray color. It is made up of bricks that are I-shaped and blended together seamlessly.		
The rough ceramic tiles in a dark salmon color are surprisingly light, making them perfect for easy installation in a grid pattern.		
The rough red fabric is adorned with a bright grid pattern.		
The material has a rough texture with a bright appearance, featuring a dark khaki color reminiscent of sand. It has a pitted surface, giving it an earthy, grounded feel.		
The leather material is saddlebrown in color with a dim appearance and a bumpy texture. Despite the bumps, there is no visible texture on the surface, giving it a smooth and luxurious feel.		
The marble material had a rough and dim appearance, with a darkslategray coloring that added depth to its marbled texture.		
The marble material had a rough and dim appearance, with a darkslategray coloring that added depth to its marbled texture.		

Figure 15: Examples in DTDMat

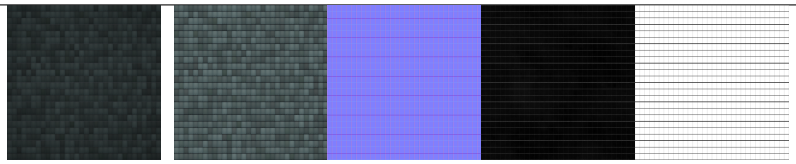
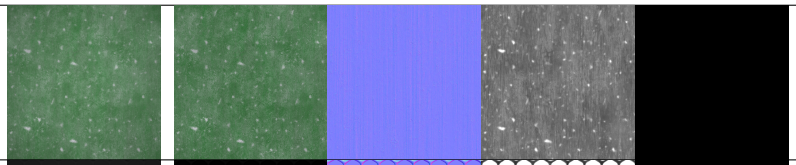
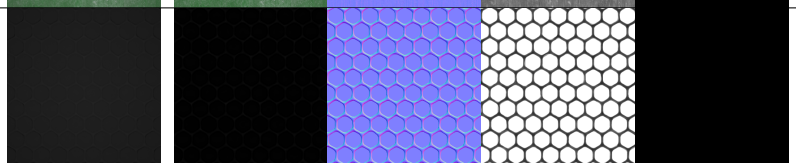
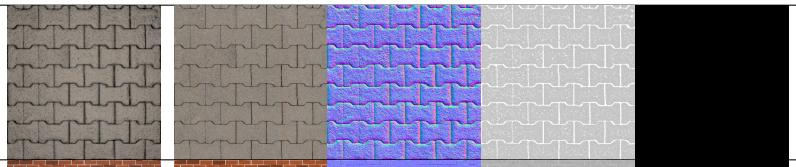
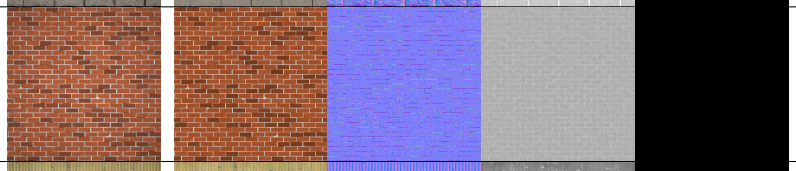
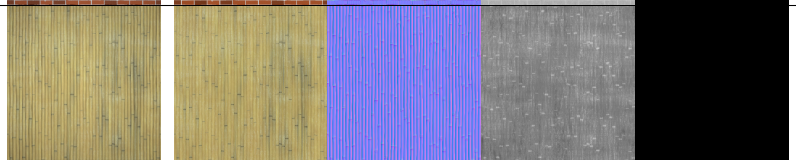
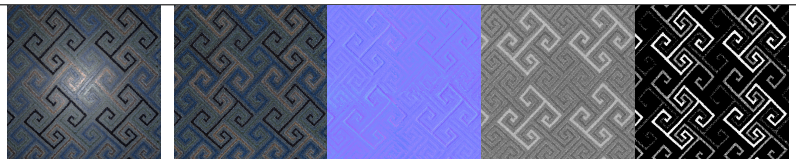
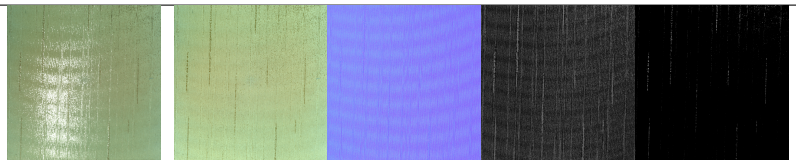
Text	Render	SVBRDF
The material was rough to the touch, with a dim appearance in a dim gray color. It had a miscellaneous texture, resembling a grid pattern.		
The rough dim gray plaster material had a dull and matte appearance, giving it a muted and subdued aesthetic.		
The material is a black plastic with a polygonal shape. It has a bumpy texture and a dim appearance. It is associated with JS technology and has a grid pattern.		
The material is bumpy and gray, with a bright and glossy finish. It is made of stone, in I-shaped and rectangular pieces.		
The material is bumpy and gray, with a bright and glossy finish. It is made of stone, in I-shaped and rectangular pieces.		
The material is rough to the touch, with a bright sheen that catches the light. Its color is a dark khaki, reminiscent of bamboo or wood.		
The material is rough to the touch, with a bright sheen that catches the light. Its color is a dark khaki, reminiscent of bamboo or wood.		
The material is rough to the touch, with a bright sheen that catches the light. Its color is a dark khaki, reminiscent of bamboo or wood.		
The material is smooth to the touch, with a dim sheen that resembles the color dark-slategray. It has a crystal-like appearance, reflecting light in a captivating manner. The overall feel is light, both in weight and in the way it interacts with light.		

Figure 16: Examples in DTDMat

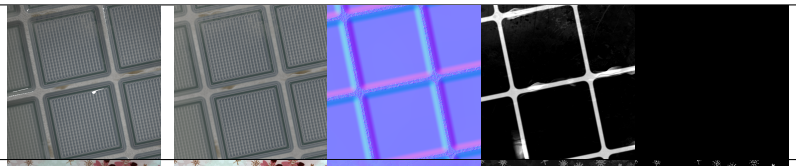
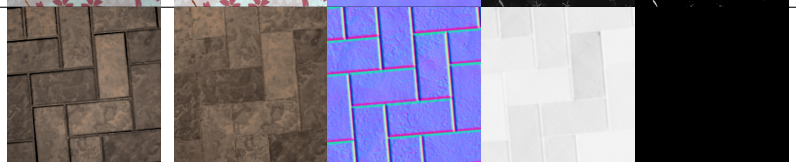
Text	Render	SVBRDF
The material is a light gray glass with a bumpy texture, arranged in a grid pattern resembling blocks.		
The material is bumpy and has a bright Indian red color. It also features a flower pattern.		
The material can be described as having a rough surface with a herringbone texture, featuring a dark slate gray color. It embodies a Japanese style, often used for walls to create a sophisticated and textured aesthetic.		

Figure 17: Examples in DTDMat

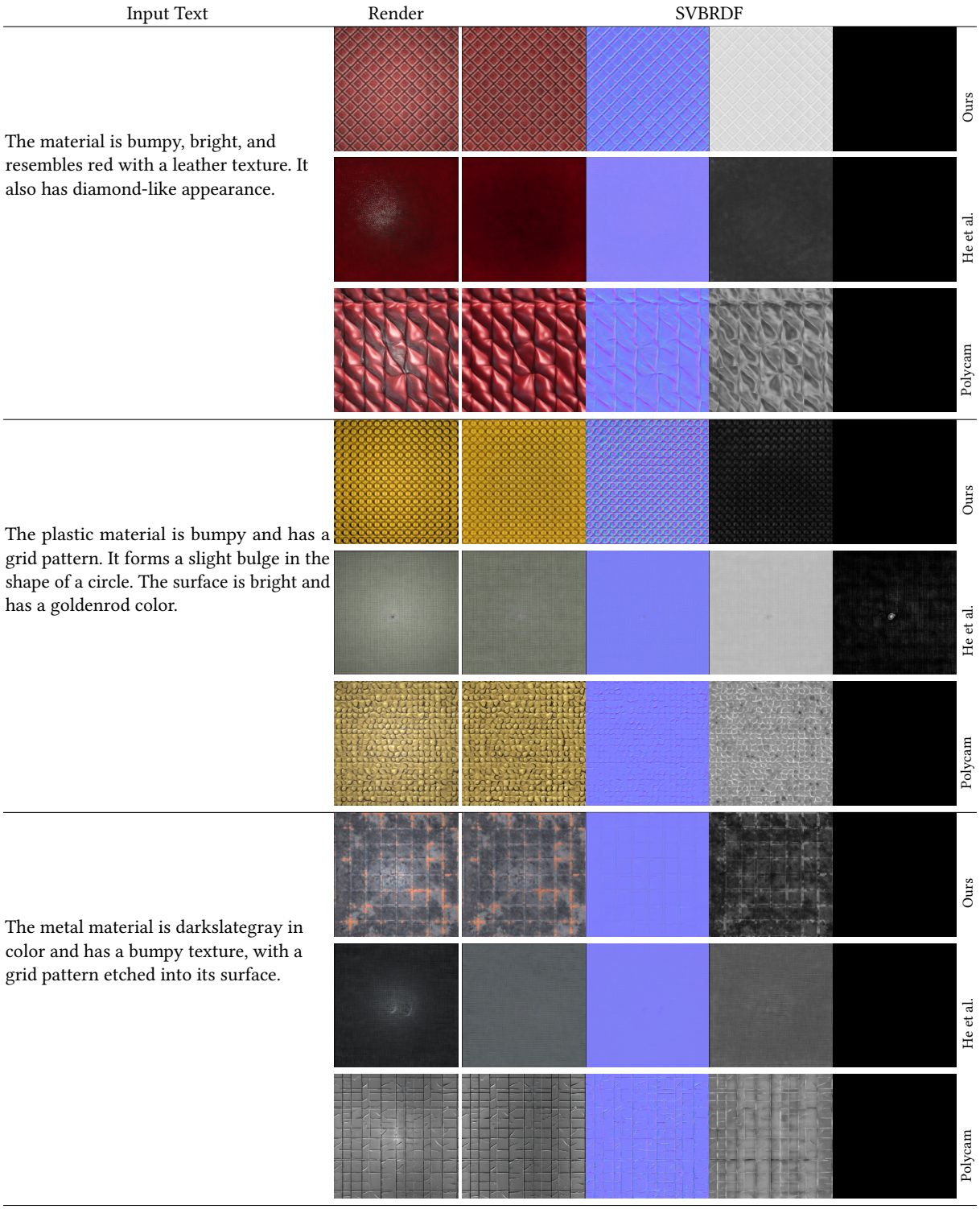


Figure 18: Comparisons against Txt2Mat[He et al. 2023] and an online text-to-generation website[pol 2024].