

Text2Tex: Text-driven Texture Synthesis via Diffusion Models

Supplementary Material

Dave Zhenyu Chen¹ Yawar Siddiqui¹ Hsin-Ying Lee² Sergey Tulyakov² Matthias Nießner¹

¹Technical University of Munich

²Snap Research

<https://daveredrum.github.io/Text2Tex/>

In this supplementary material, we provide the categories of the Objaverse subset in Sec. 1 and the details of user study in Sec. 2. To showcase the effectiveness of the proposed texture synthesis method, we provide additional results and analysis in Sec. 3.

1. Objaverse Subset

Our method is evaluated on a subset of the Objaverse [1] dataset. To construct this subset, we first sample 3 random meshes from each category, where thin or unrecognizable meshes are filtered out. We also manually remove meshes with too simplistic textures and wrong categories. To reduce the processing time, we remove over-triangulated and scanned objects. After this curation, we obtain in total 410 high-quality textured objects across 225 categories. The chosen categories are as follows:

```
# Objaverse subset object categories
["Bible", "CD player", "Lego", "Tabasco sauce",
"aerosol can", "airplane", "alarm clock",
"ambulance", "apple", "apricot", "armchair",
"army tank", "baby buggy", "backpack", "bagel",
"ball", "banana", "banjo", "barrel", "baseball",
"baseball bat", "baseball glove", "basket",
"basketball", "bathtub", "beanbag", "bed",
"bedpan", "bench", "bicycle", "binoculars",
"birdhouse", "birthday cake", "blimp", "boat",
"bookcase", "bottle", "bowl", "bread",
"briefcase", "broccoli", "broom", "bucket",
"bulldog", "bulldozer", "burrito", "butterfly",
"cabinet", "calculator", "camera", "can",
"candle", "canoe", "cappuccino", "carrot",
"chair", "chaise longue", "chalice",
"chocolate bar", "cigarette", "clipboard",
"clock", "clutch bag", "coconut", "coffee maker",
"coffee table", "coffeepot", "comic book",
"compass", "computer keyboard", "cowbell",
"cowboy hat", "crate", "crown", "crucifix",
"cucumber", "cup", "cupcake", "cushion",
"dagger", "deck chair", "deer", "desk", "dog",
"doll", "dolphin", "doughnut", "duffel bag",
"dumbbell", "dumpster", "elephant", "fan",
"faucet", "fighter jet", "file cabinet",
"fire extinguisher", "first-aid kit",
"fish", "flashlight", "forklift", "frog",
"frying pan", "gargoyle", "giant panda",
```

```
"globe", "glove", "goldfish", "goose",
"guitar", "gun", "hair dryer", "hairbrush",
"hamburger", "hammer", "hardback book",
"heart", "helicopter", "helmet", "highchair",
"hippopotamus", "hockey stick", "hourglass",
"hummingbird", "iPod", "jar", "jeep",
"jet plane", "keg", "kettle", "key", "knife",
"ladybug", "lantern", "laptop computer",
"lemon", "lightbulb", "lizard", "machine gun",
"martini", "matchbox", "microscope",
"microwave oven", "milk", "milk can", "minivan",
"money", "motorcycle", "muffin", "mug",
"mushroom", "onion", "ottoman", "pancake",
"pelican", "pen", "pencil", "pencil box",
"piano", "pickup truck", "pie", "pigeon",
"piggy bank", "pillow", "pistol",
"pliers", "polar bear", "police cruiser",
"pool table", "pot", "pretzel",
"pudding", "pumpkin", "race car", "radish",
"rat", "refrigerator", "remote control", "rifle",
"rocking chair", "saltshaker", "saucepan",
"sausage", "school bus", "scissors",
"screwdriver", "sewing machine", "shaker",
"shark", "sheep", "shield", "shoe",
"shopping bag", "shovel", "skateboard",
"snowman", "soccer ball", "sofa",
"sofa bed", "space shuttle", "spider", "stool",
"suitcase", "sunflower", "sunglasses", "sunhat",
"sushi", "sword", "syringe", "table",
"table lamp", "teacup", "teakettle", "teapot",
"teddy bear", "telephone", "television set",
"thermos bottle", "toilet", "toothbrush",
"trailer truck", "trash can",
"tricycle", "truck", "typewriter", "umbrella",
"urn", "vending machine", "videotape", "violin",
"watch", "water cooler", "water faucet",
"watermelon", "wheel", "windmill",
"wrench", "zucchini"]
```

2. User Study Details

We develop a Django-based web application for the user study. In Fig. 1, we show the interface for the questionnaire. We randomly select 5 pairs of textured objects from each baseline and our method. To better visualize the samples, we render multi-view images for those objects from 8 preset viewpoints. After the samples are prepared, we ask

Text2Tex User Study

Each question shows objects textured by two different approaches. Each option image shows the textured object rendered from 8 different viewpoints. Select the one that looks more realistic.

1. Enter your name: *

2. Which of the textured objects best represents "bookcase"? *

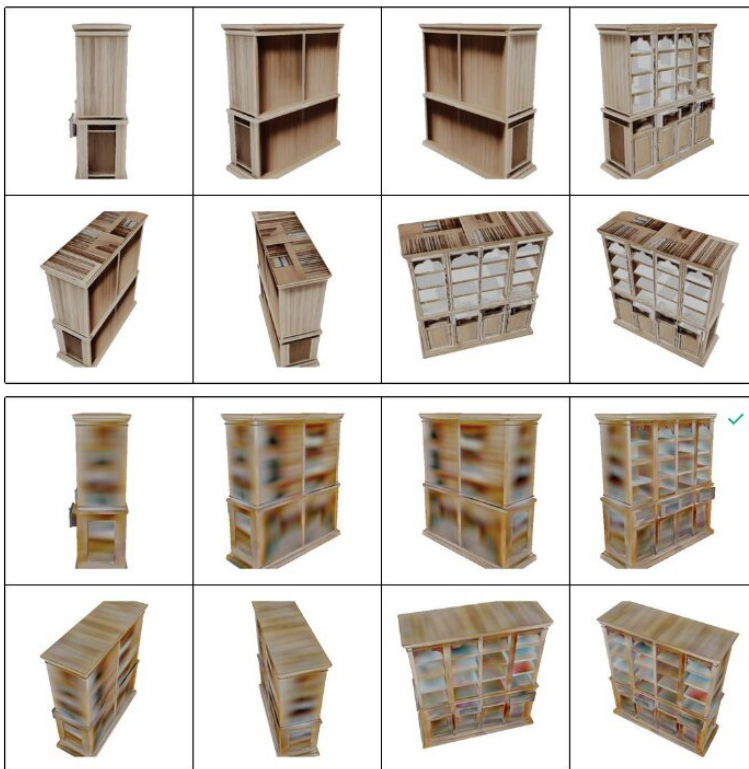


Figure 1: Screenshot of the user study interface.

the users to pick the sample from those pairs that best represents the text prompts. To avoid biases and cheating in this user study, we shuffle the pairs so that there is no positional hint of our method. In the end, we gather 604 responses from 41 participants to calculate the preferences.

3. Additional Qualitative Results

To further showcase the effectiveness of our method, we present additional qualitative results on objects from the Objaverse [1] dataset.

Comparison with the baselines. We show additional comparisons against previous text-driven methods in Fig. 2. In comparison with CLIP-based baselines (CLIPMesh [4]

and Text2Mesh [3]), our results are shown to be more detailed and realistic. Despite the blurry appearance, the results of Latent-Paint [2] still show competitive textures against ours. However, those results fail to capture the structural details of the input geometries. For instance, the scattered marshmallows in the case “a cappuccino” are incorrectly blended with the plate. In contrast, our method presents high-quality textures for the objects, while maintaining the correct structural details for the geometries.

Style the same objects. To show that our method can generate various textures for the same objects, we show different texturing results on the same objects. In Fig. 3, we show textures for the backpack with different colors in the

prompts. All our textures are highly detailed and loyal to the colors, demonstrating the diversity and variety of the potential 3D contents. Moreover, we show that our method is not constrained by simple attributes. As shown in Fig. 4, our method is fully capable of reflecting complicated styles in the texture space, such as “baroque” and “cyberpunk”. This indicates a great potential to stylize more high-quality 3D textures.

Creative texture synthesis. One interesting trait of our method is that it can move beyond certain categories. To show this, we showcase some creative textures on a Porsche in Fig. 5. Given some unrealistic prompts as input, our method is able to properly wrap the appearance on the geometry. For instance, in the case “hippo”, our method aligns the eyes of a hippopotamus to the lamps of the Porsche as they are semantically similar to each other. It is worth mentioning that all textured objects clearly represents the original properties of the geometry (see the wheels of the case “airplane”), while reflecting iconic characteristics of the input prompts (see the crocodile skin of the case “crocodile”).

References

- [1] Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects. *arXiv preprint arXiv:2212.08051*, 2022. 1, 2, 4
- [2] Gal Metzer, Elad Richardson, Or Patashnik, Raja Giryes, and Daniel Cohen-Or. Latent-nerf for shape-guided generation of 3d shapes and textures. In *CVPR*, 2023. 2
- [3] Oscar Michel, Roi Bar-On, Richard Liu, Sagie Benaim, and Rana Hanocka. Text2mesh: Text-driven neural stylization for meshes. In *CVPR*, 2022. 2
- [4] Nasir Mohammad Khalid, Tianhao Xie, Eugene Belilovsky, and Tiberiu Popa. Clip-mesh: Generating textured meshes from text using pretrained image-text models. In *SIGGRAPH Asia*, 2022. 2



Figure 2: Additional qualitative comparison on objects from Objaverse [1] dataset.



Figure 3: Different colors for the backpack. Our method loyally reflects the prompt colors in the textures.



Figure 4: Different styles for the Porsche. Our method is capable of handling complicated styles such as “baroque” and “cyberpunk” without distorting the original properties of the input geometry.



Figure 5: Creative textures for the Porsche with unrealistic prompts. Our method clearly represents the original properties of the geometry, while reflecting iconic characteristics of the input prompts.