



Figure 1: generation results of scene-scale models.

1 USER STUDY

For a more comprehensive comparison, we recruited a group of volunteers to compare the visual quality of generated textures in pairs of TEXTure-Ours, Meshy-Ours, Fantasia3D-Ours, and TexFusion-Ours, over the set of 182 examples, resulting in a total of $182 \times 4 = 728$ pairs of comparisons. These 728 pairs were shuffled and divided into 8 sets. To maintain an as-fair-as-possible evaluation process, we recruited 24 volunteers with 17 males and 7 females, aged between 20 and 35, with undergraduate or higher educational degrees. Each volunteer was randomly assigned with one test set, and each test set was tested on three different volunteers. In each pair of comparison within the test set, the volunteer was presented with two textured models captured by a camera rotated around them, with textures generated using the same text prompt but two different methods. Each participant was then asked to choose the one with better quality in terms of four criteria: natural color (NC); less noise and seams (NS); more detailed content (DC); better alignment with text prompt (TP). The results were summarized in Table 1. Our method achieved consistently better results in almost all criteria. This user study reflects the same observations as our other comparisons.

Method	NC	NS	DC	TP
TEXTure	80.31%	87.75%	61.30%	63.51%
Meshy	57.27%	67.43%	43.64%	56.48%
Fantasia3D	68.22%	53.50%	45.93%	57.26%
TexFusion	58.16%	61.89%	57.18%	50.29%

Table 1: The percentage of participants who think our results are better, under four criteria. Denote NC = natural color, NS = less noise and seams, DC = more detailed content, TP = better alignment with text prompt.

2 GENERATION RESULTS

One remarkable feature of 2D LDM is its significant diversity in generation. The generated texture of our method also inherits this diversity from the pre-trained LDM. We show the diverse generation results using the same text prompt in Figure 2.

We also generate textures for scene-scale models. The results are shown in Figure 1. With the default configuration, we successfully generate textures for scene-scale models aligned with text prompts, although the appearance lacks some details. However, by adding more cameras and views with detailed geometry, the quality of generation can be enhanced, albeit at the cost of increased computation.

In Figure 3 and Figure 4, we also provide more results of our method generated in experiments.



New Balance shoe



A shiny fish



An ancient pot, weathered by time



A combat helmet

Figure 2: Diversity generation results with the same prompts.



Figure 3: generation results of scene-scale models.

