Supplementary Materials for TEXTRIX: Latent Attribute Grid for Native Texture Generation and Beyond

1. Data Source Clarification

In Section 4 (Implementation Details), we describe the use of the Objaverse [1] and ObjaverseXL [2] datasets. For completeness, we additionally note that the final training corpus also included a curated collection of internal high-quality 3D assets. These internal assets were incorporated to broaden the diversity of texture patterns beyond what is available in current public repositories.

The inclusion of internal data is limited strictly to model training. All quantitative evaluations, those reported in Tables 1, 2, and 3, were conducted entirely on public benchmarks and publicly accessible test sets, ensuring fair, transparent, and reproducible comparison with prior work.

2. Quantitative Comparison and Resolution Analysis for VAE

We isolate the performance of our Attribute VAE to evaluate its effectiveness in compressing and reconstructing 3D surface attributes. We compare our texture reconstruction fidelity against the VAE utilized in TRELLIS [3], a state-of-the-art baseline.

Scalability and Resolution Analysis. A key advantage of our native attribute grid representation is its scalability. To demonstrate this, we conducted comparative experiments with our VAE trained at two distinct grid resolutions: 512^3 and 1024^3. As visualized in Figure 1, we present a side-by-side qualitative comparison.

- **Ours** (512^3)**:** Even at a lower resolution, our model produces results with sharper edges and better color fidelity compared to the baseline.
- **Ours** (1024^3)**:** The higher-resolution model further unlocks the potential of our framework, capturing microscopic high-frequency texture details that are smoothed out at lower resolutions.
- **TRELLIS:** In contrast, the baseline exhibits noticeable blurring and artifacts in complex regions.

3. Ablation Study for Sparse Latent Condition

In the main paper (Table 4), we provided quantitative results validating the effectiveness of our sparse latent condition. To complement that numerical analysis, we present corresponding qualitative comparisons in Figure 2.

Figure 1. **Qualitative comparison of VAE reconstruction fidelity.** We compare our Attribute VAE against the VAE used in TRELLIS [3]. Our model (left) demonstrates a superior ability to reconstruct high-frequency texture details from the latent representation compared to TRELLIS (right).

The visualization clearly illustrates the impact of our conditioning strategy. Without the sparse latent condition ('w.o. Latent Condition'), the model relies solely on global semantic features, which often leads to misalignment with the input view. By projecting the input image into the 3D voxel grid, our proposed method achieves significantly bet-

ter alignment and consistency. The sparse latent condition acts as a robust 3D anchor, ensuring that the generated texture not only respects the semantic content but also maintains precise spatial correspondence with the input view, yielding results that are geometrically consistent across all viewing angles.

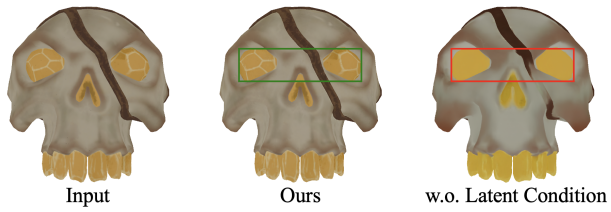

Figure 2. Ablation on Sparse Latent Condition. Our condition method yields more consistent results with the input image.

4. Ablation Study for Rendering Loss

We further investigate the impact of our training objective by comparing our rendering-based reconstruction loss against a more direct MSE (Mean Squared Error) loss computed directly on the sparse color grid (i.e., the "cube" loss).

Although the MSE loss is computationally simpler, it treats all voxels equally, regardless of their visibility or contribution to the final rendered image. This often leads to optimization efforts being wasted on occluded or internal structures that do not affect the visual output. In contrast, our rendering-based loss operates in the image space. This approach is more aligned with human perception, forcing the model to optimize for the final visual projection. As shown in Figure 3 and Table 1, the rendering-based loss yields results that are visually superior, with sharper details and fewer artifacts, and quantitatively better in terms of PSNR and SSIM metrics.

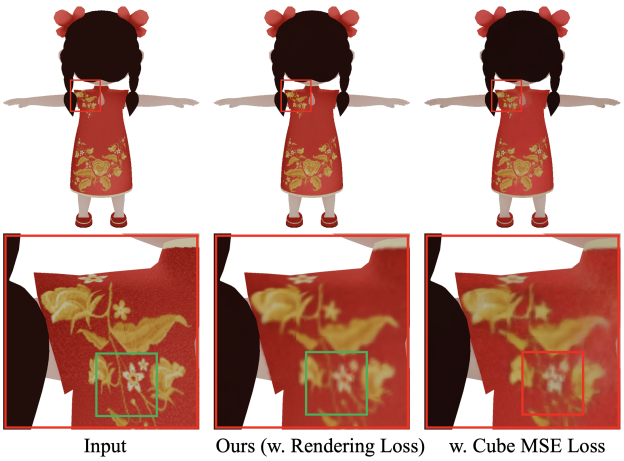

Figure 3. Ablation Study on Rendering Loss. Our VAE produces results with more details compared to cube MSE loss.

Table 1. Ablation study on the **VAE reconstruction loss**. The rendering-based loss achieves superior reconstruction fidelity.

Model Variant (VAE)	SSIM ↑	PSNR ↑	LPIPS ↓
Ours (w. Rendering Loss)	**0.9841**	**33.21**	**0.0246**
w/ Cube MSE Loss	0.9788	31.59	0.0421

5. Limitation

Despite the state-of-the-art results achieved by our framework in both 3D generation and perception tasks, certain limitations remain. Our framework's reliance on a latent attribute grid makes its performance sensitive to the quality and topology of the input mesh[cite: 318]. Specifically, when dealing with non-manifold geometry, severe self-intersections, or meshes with extremely irregular triangulation, the grid-based attribute querying (via trilinear interpolation) may introduce minor noise or discontinuities. While our method is robust to general mesh variations, extremely low-quality geometry can hinder the precision of the generated texture.

6. Demostration for PBR Generation

As discussed in Section 3.4 of the main paper, our Native 3D Attribute Grid is a unified representation capable of storing diverse surface properties. To validate this extensibility, we applied our framework to PBR (Physically Based Rendering) material generation.

Implementation. The training pipeline remains identical to the one used for 3D part segmentation. We simply substitute the training targets: instead of semantic labels or RGB colors, the model is trained to predict PBR attribute maps (specifically metallic, roughness, and normal maps) within the sparse voxel grid.

Results. As illustrated in Figure 4, our model successfully synthesizes coherent PBR materials. We observe that the framework performs particularly well on meshes characterized by distinct material transitions and clear geometric boundaries. In these cases, the sparse attribute grid accurately captures the sharp variations in material properties (e.g., the transition from metal to plastic), demonstrating the potential of TEXTRIX as a unified solution for both texture and material generation.

Figure 4. Visualization for Results of PBR Generation

7. More Results for Single View Generation

Figure 5. Visualization for More Results of Single View Generation.

8. More Results for Multi-View Based Generation

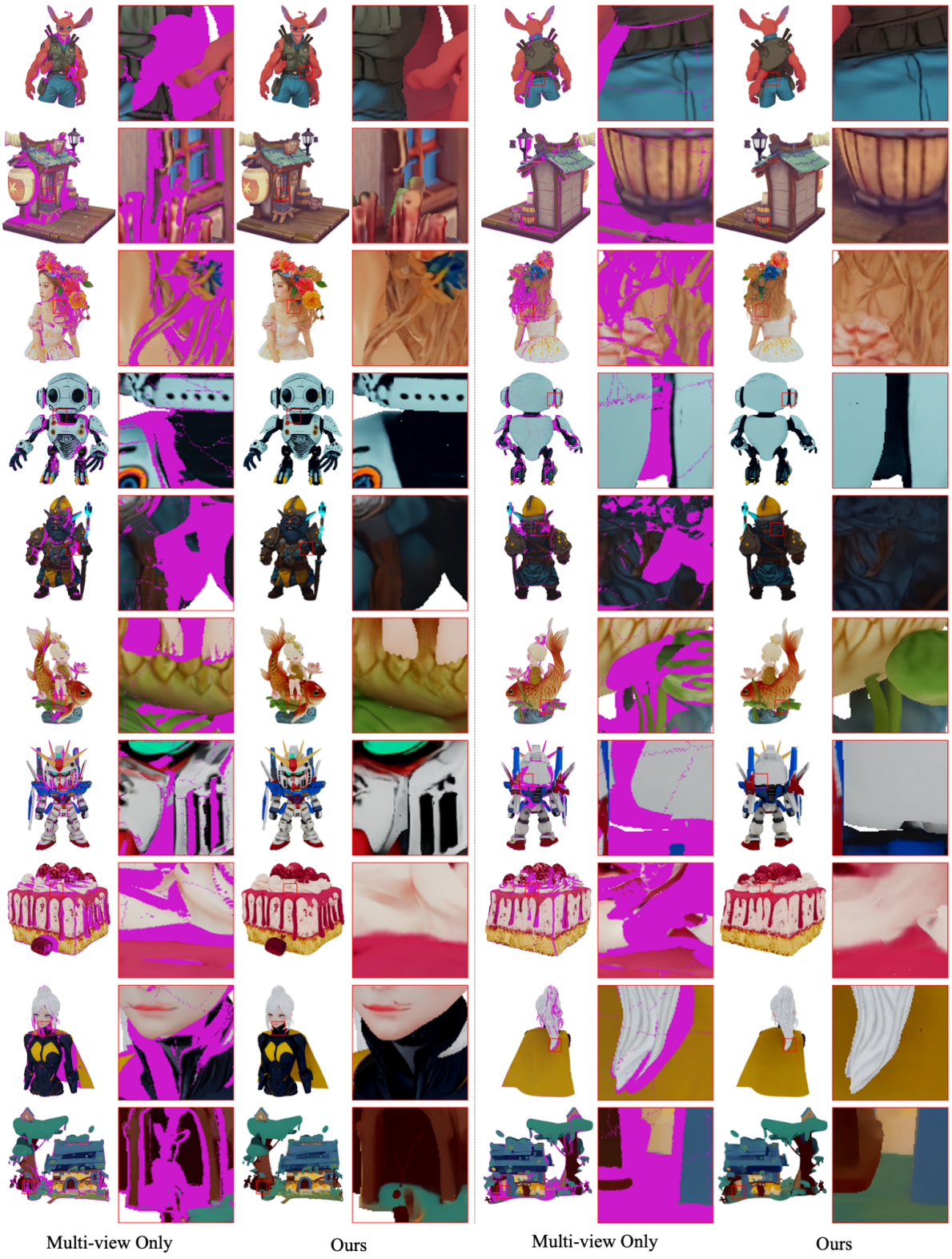

Figure 6. Visualization for More Results of Multi-view Based Generation.

9. More Results for 3D Segmentation

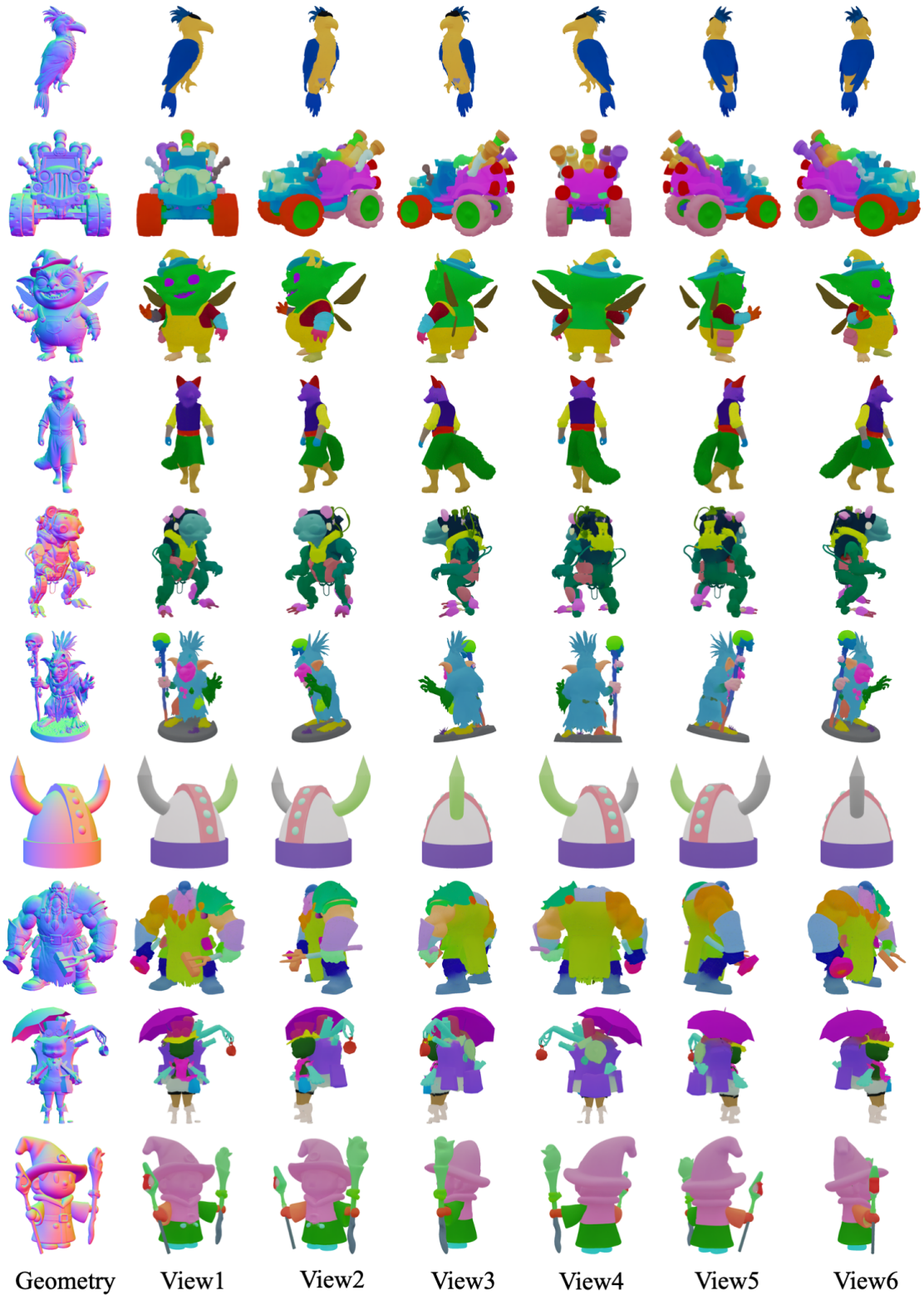

Figure 7. Visualization for More Results of 3D Segmentation.

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] Matt Deitke, Ruoshi Liu, Matthew Wallingford, Huong Ngo, Oscar Michel, Aditya Kusupati, Alan Fan, Christian Laforte, Vikram Voleti, Samir Yitzhak Gadre, Eli VanderBilt, Aniruddha Kembhavi, Carl Vondrick, Georgia Gkioxari, Kiana Ehsani, Ludwig Schmidt, and Ali Farhadi. Objaverse-xl: A universe of 10m+ 3d objects. *arXiv preprint arXiv:2307.05663*, 2023. 1

[3] Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, and Jiaolong Yang. Structured 3d latents for scalable and versatile 3d generation. *arXiv preprint arXiv:2412.01506*, 2024. 1