Appendix

Anonymous ICCV submission

Paper ID 5625

1. Details of Model Structure

In the main part of the paper, we have provided a wraparound description of DG3D. We will provide the details of the improved Generator and the multi-modal discriminators respectively in Sec. 1.1 and Sec. 1.2.

1.1. Improved Generator

For the intuition that the geometry should only be controlled by the geometry latent code and the texture should be able to not only adapt to the changes on the texture latent code, but also to changes in geometry, we utilizes the improved G which is depicted in Fig. 1. To obtain the feature $f \in \mathbb{R}^{32}$ of a surface point $p \in \mathbb{R}$, p is first projected onto each plane. Bilinear interpolation is further performed on the features and finally we sum the interpolated features:

$$f = \sum_i b(\pi_i(p)), \tag{1}$$

where i can be x, y, z planes, $\pi_i(p)$ is the projection of point p on plane i and b denotes bilinear interpolation of the features. The features are further decoded into the color of the point p, signed distance value and vertex offset using three conditional FC layers. The signed distance value and vertex offset are further converted to surfaces by the differentiable Marching Tetrahedron layer.

1.2. Multi-Modal Discriminators

We use three discriminators together to output the probability that a sample is treated as real. Keeping the RGB images discriminator unchanged, the mask images discriminator is extended as the alpha transparency images discriminator by the diffusion-augmented module. We feed the mask image into the diffusion-augmented module and adaptively inject noise into the binary mask image. The output alpha transparency image has pixel values which distribute between 0 and 1, preventing from discontinuous optimization. Additional, we construct a patch-wise discriminator which is used to supervise path-wise detailed textures. The input of the patch-wise discriminator is four stacked patches of the RGB images which is selected by the rule in the main paper. The weight coefficients of the three discriminator are 1, 1, and 0.1 respectively. The combination strategy significantly boosts the texture quality of our generated shapes.

2. More Experimental Results

We provide more visualization results from multi views in Fig. 2 and more interpolation results in Fig. 3 due to the lack of space in the main paper. We also provide a demo video with 2K resolution named as $demo5625.mp4$ to visualize the details of generated shapes.

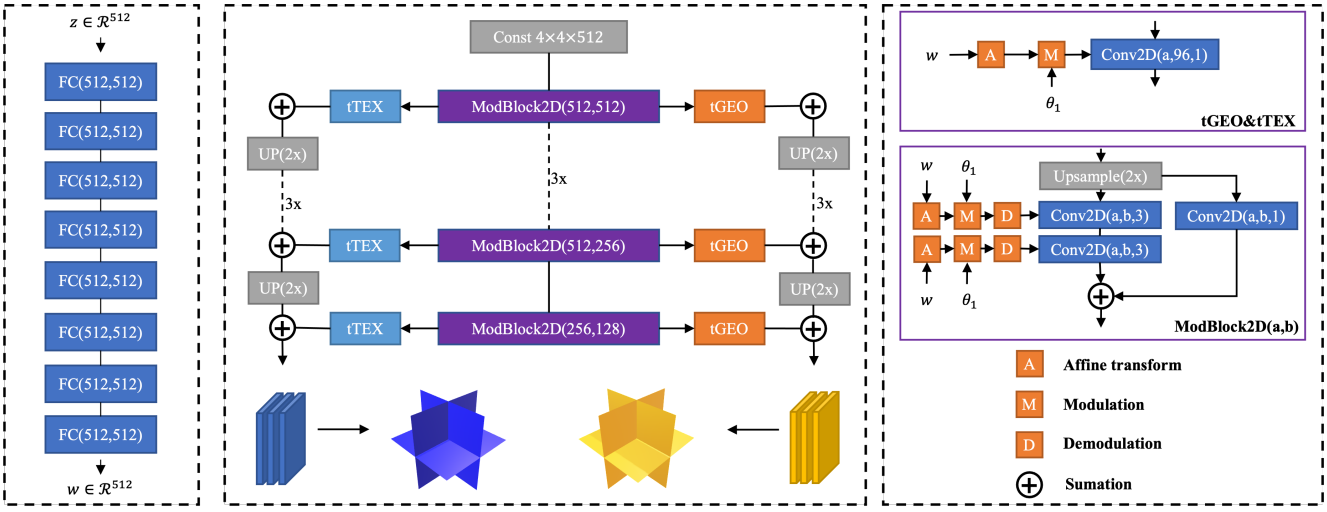

Figure 1: The projection module consists of eight fully connected layers. The main generation module consists of five **ModBlock2D**s and each **ModBlock2D** has a **tTEX** branch and a **tGEO** branch. The output of **tTEX** and **tGEO** will be further upsampled and added to the next layer. The output of the last layer is further reformulate as a texture tri-plane and a geometry tri-plane.

Figure 2: Multi-view images of generated shapes from DG3D.

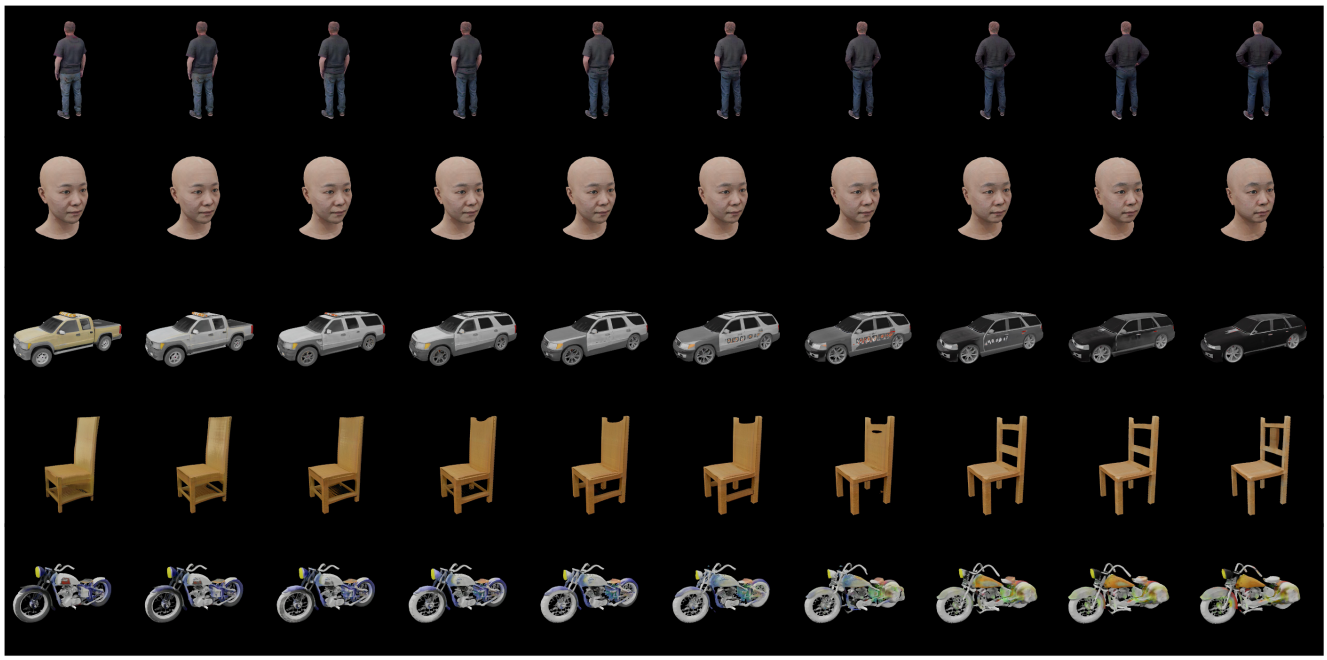

Figure 3: Interpolation results of all datasets in the main paper.