Learning Pseudo 3D Guidance for View-consistent Texturing with 2D Diffusion

Appendix

Kehan Li[1,4], Yanbo Fan[2,6⋆], Yang Wu[2], Zhongqian Sun[2], Wei Yang[2]
Xiangyang Ji[5], Li Yuan[1,3,4], and Jie Chen[1,3,4⋆]

[1] School of Electronic and Computer Engineering, Peking University, Shenzhen, China
[2] Tencent AI Lab, Shenzhen, China
[3] Peng Cheng Laboratory, Shenzhen, China
[4] AI for Science (AI4S)-Preferred Program, Peking University, Shenzhen, China
[5] Department of Automation and BNRist, Tsinghua University, Beijing, China
[6] Ant Group, Hangzhou, China

I Limitations

Both the guidance learning and the guided image generation process of the proposed method are based on text-to-image diffusion models for its high visual quality following state-of-the-art methods [1, 3]. However, an inherent flaw of these models is the unawareness about the true 3D structure of the input geometry. Although the depth map is adopted to constrain the image generation with the geometry, we empirically find that the Janus problem appears in a few cases where the geometry is similar from several mutually invisible views. Getting rid of the inherent flaws is still an open question and a possible solution is to incorporate large-scale and high quality 3D data with text annotations. We leave this for future research.

II Speed Comparison

Tab. 1 shows the time spent for generating texture of a 3D shape. Although our method introduces some acceptable additional overhead in the guidance learning stage, it achieves better quality and consistency.

III Additional Ablation Study

Depth-guided SDS. The 3D texturing task requires to generate high-fidelity texture that conforms to the fixed geometry. Since classic text-to-image models are not geometry-aware, using them to perform SDS can lead to results that do not match the geometry, thus interfering with the convergence and quality of multi-view optimization. As shown in Fig. 1, the depth-guided SDS improves the speed and quality due to geometry awareness.

⋆ Corresponding author.

Table 1: Speed comparison.

Method	Time (seconds)		
	Total	**Stage 1**	**Stage 2**
TEXTure [2] SIGGRAPH'23	90	-	-
Text2Tex [1] ICCV'23	1,091	605	486
P3G (*Ours*)	614	466	148

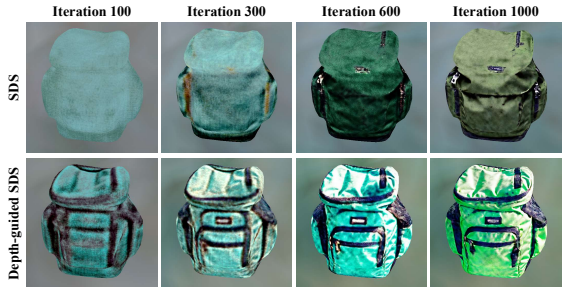

Fig. 1: Ablation on Depth-guided SDS. Due to the introduction of geometry awareness, it improves the speed and quality.

Latent-to-RGB Optimization. Fig. 2 shows the comparison of the latent-to-RGB optimization strategy and the pure RGB optimization strategy. Fig. 3 shows the texture obtained by optimization in latent space and RGB space. Due to the small computational cost, the texture quickly converges in the latent space, but the image resolution limits the sharpness of the texture. Taking it as initialization, the RGB space optimization further improves the texture in detail by high-resolution images, producing precise guidance.

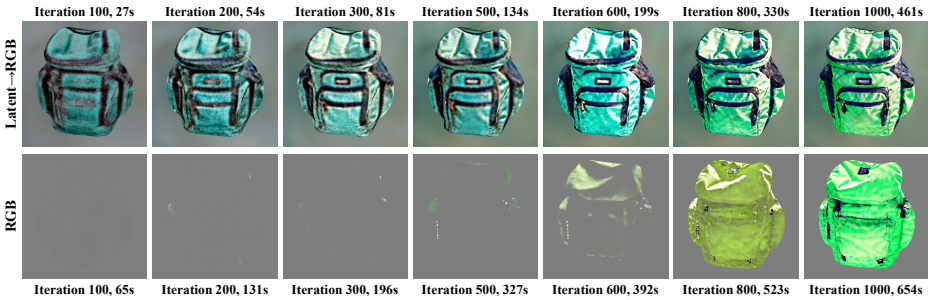

Fig. 2: Comparison on the optimization speed and result quality when different strategies are adopted: Latent-to-RGB *vs.* pure RGB.

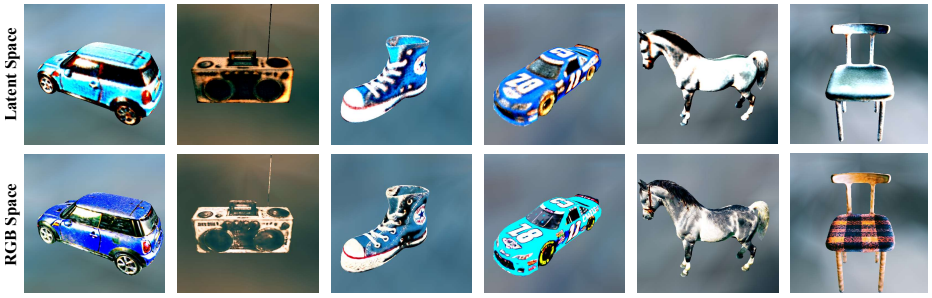

Fig. 3: Visualizations of learned textures *w.r.t.* latent space and RGB space.

IV Additional Qualitative Results

View Selection Strategy. We visualize the multi-view image generation process with the proposed view selection strategy in Fig. 4. With only negligible additional calculations, the view covering as wide an area as possible can be accurately estimated by our strategy, which finally speeds up the generation of the entire texture.

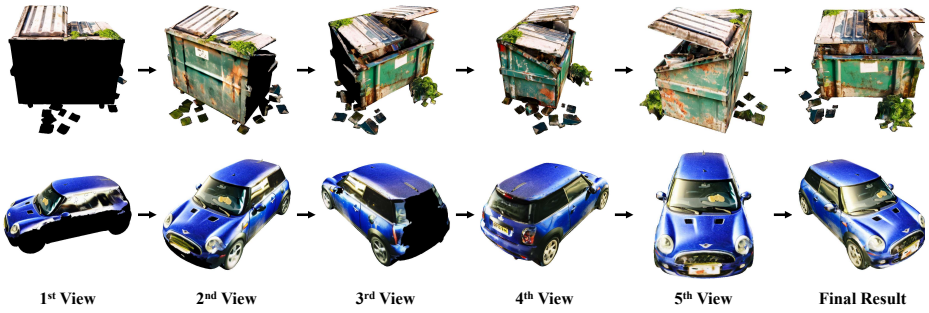

Fig. 4: Visualizations of the view selection strategy.

Consistency Comparison. Some additional comparisons of consistency with previous method are shown in Fig. 5.

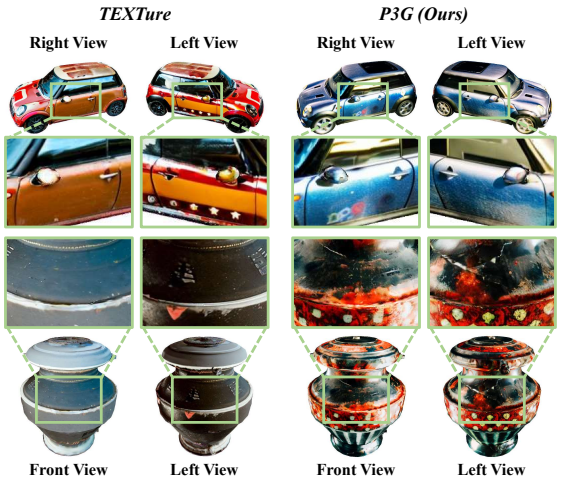

Fig. 5: Consistency comparison. Previous inpainting-based methods fail to ensure long-range consistency. Our P3G resolves it by introducing the guidance.

Quality Comparison. In Figs. 6 to 9 we compare the texture quality of our P3G to TEXTure [2] and Text2Tex [1]. Our approach better responds to the input text, matches geometry, and encourages consistency.

Fig. 6: More qualitative comparisons (group 1).

Fig. 7: More qualitative comparisons (group 2).

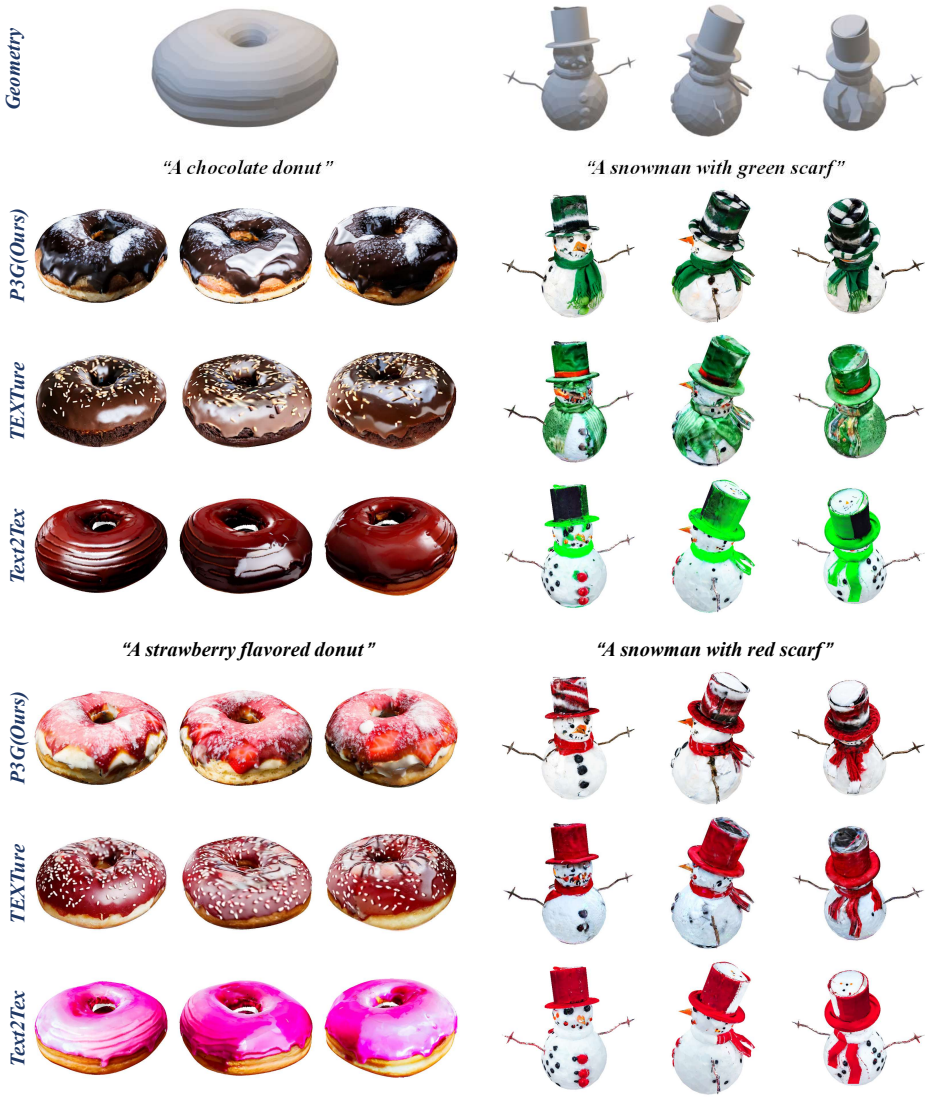

Fig. 8: More qualitative comparisons (group 3).

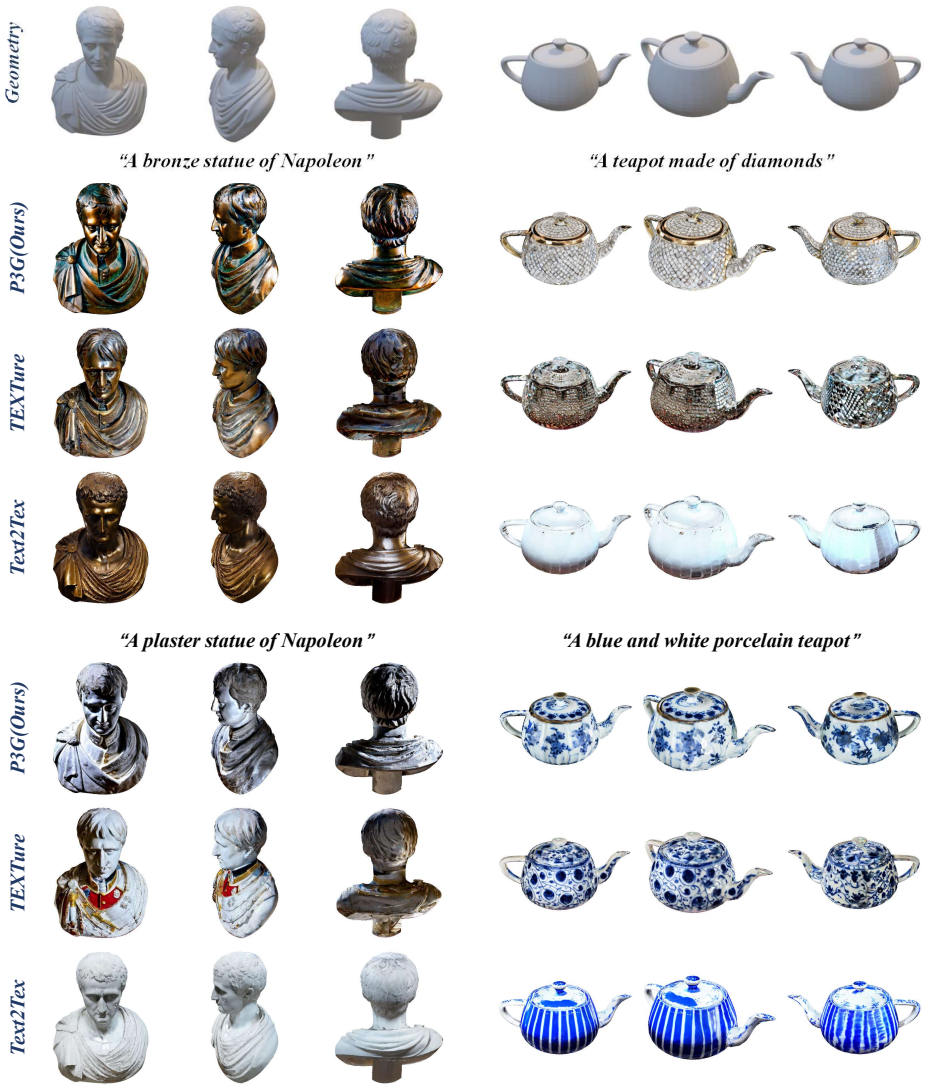

Fig. 9: More qualitative comparisons (group 4).

References

1. Chen, D.Z., Siddiqui, Y., Lee, H.Y., Tulyakov, S., Nießner, M.: Text2tex: Text-driven texture synthesis via diffusion models. arXiv preprint arXiv:2303.11396 (2023)
2. Richardson, E., Metzer, G., Alaluf, Y., Giryes, R., Cohen-Or, D.: Texture: Text-guided texturing of 3d shapes. arXiv preprint arXiv:2302.01721 (2023)
3. Yu, X., Dai, P., Li, W., Ma, L., Liu, Z., Qi, X.: Texture generation on 3d meshes with point-uv diffusion. arXiv preprint arXiv:2308.10490 (2023)