

Appendix: PanoDreamer: Consistent Text to 360-Degree Scene Generation

Supplementary Material

1. Experiment Details

During rendering, as we use pinhole cameras, reducing the camera Field-of-View(FoV) will reduce the distortions. During the projection, we set the camera FOV of the base camera and the supplementary cameras to be 60° . We set the number of base camera to be 80, with each base camera corresponding to 4 supplementary cameras. The resolution of the projected images from base cameras and the supplementary cameras are set to be 512×512 .

During the two-stage 3D gaussian splatting process, for the first 5000 iterations, we use the base set to initialize the 3D Gaussians, and after 5000 iterations, we add the refined supplementary set for the second-stage refinement of the 3D Gaussians. We report the performance of the rendered results on training 10,000 iterations.

2. Further Qualitative Results of Scene Generation

We show more qualitative results of our method on some scenes, shown in Fig. 1. Results show that our method not only generate high-quality rendered images, but also maintains accurate geometry and scene consistency.

3. Further Qualitative Results on Text-to-Panorama Generation

We present more results for text-to-panorama generation in comparison with prior methods [1, 2] in Fig. 2. Compared with previous methods, with LLM guidance, our method shows less duplicated objects and better generation quality.

References

- [1] Omer Bar-Tal, Lior Yariv, Yaron Lipman, and Tali Dekel. Multidiffusion: Fusing diffusion paths for controlled image generation. 2023. 1, 2
- [2] Shitao Tang, Fuyang Zhang, Jiacheng Chen, Peng Wang, and Yasutaka Furukawa. Mvdifffusion: Enabling holistic multi-view image generation with correspondence-aware diffusion. *arXiv*, 2023. 1, 2

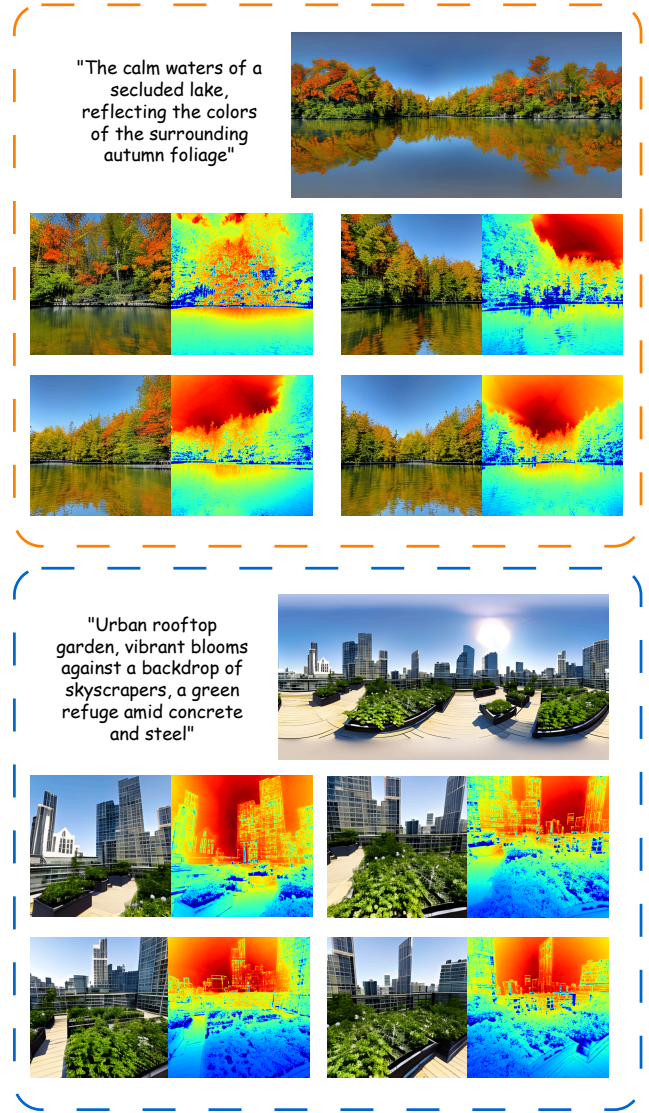


Figure 1. Additional results about scene generation. We show both the rendered images and rendered depth.



Figure 2. Additional comparison of text-to-panorama generation. As panoramas generated by MultiDiffusion [1] and MVDiffusion [2] have both limited vertical FoV, for a fair comparison, we only show our panorama before outpainting.