ExploreGS: Explorable 3D Scene Reconstruction with Virtual Camera Samplings and Diffusion Priors

Supplementary Material

A. WildExplore Dataset

To evaluate the effectiveness of our method in real-world exploration scenarios, we introduce WildExplore, a new dataset for novel view synthesis task. Fig. 1 visualizes all eight scenes in the dataset, including both train camera trajectories and two exploration (evaluation) camera trajectories.

B. Enhance-Extrapolate Dataset

We create a dataset using 3DGS [1] renderings that simulates artifacts caused by large viewpoint changes. While 3DGS-Enhancer [3] uses 130 scenes from DL3DV-10K [2], we significantly scale up dataset to improve generalization capability. To be specific, we select 1K scenes from DL3DV-10K [2] and sample four subsets per scene using two different strategies:

- **One sequence sampling.** We select a consecutive portion of the scene, covering 10%, 30%, 50%, and 70% for training.
- **Two sequence sampling.** We sample two consecutive sequences per scene, each covering 5%, 15%, 25%, or 35% of the scene, ensuring that the total percentage aligns with the first strategy. Sampled sequences have certain distance to ensure various camera trajectories.

C. Implementation details

Scene initialization. We obtain the mesh using TSDF after training 3DGS on training viewpoints. For the target bounding box, we extract the oriented bounding box tightly surrounding the mesh. We set visibility threshold as 0.5 for occupancy grid estimation, motivated by ExtraNeRF [4].

Virtual view sampling. We discretize $\mathcal{M}$ into elevation and azimuth bins at 30° intervals, totaling in 32 bins. For camera translation, we use a fixed step size equal to the diagonal length of the scene bounding box divided by S. For camera rotation, we apply a fixed angular step of 10°.

Diffusion model. Our video diffusion model is built on DynamiCrafter [6]. We compose the input sequence as 15 test views sampled by methods described at B and one frame from the training set selected as the nearest neighbor to the chosen test frames. For text condition, we obtain a scene description from GPT-4o with a maximum of 70 tokens. We train the diffusion model for 12 days on 4 A100 (40 GB) GPUs.

Curated Nerfbusters Dataset	PSNR↑	SSIM↑	LPIPS↓
Nerfbusters	16.00	**0.506**	0.454
Ours	**16.22**	0.478	**0.433**

Table 1. Quantitative comparision between the original Nerfbusters model and our ExploreGS.

Method	PSNR↑	SSIM↑	LPIPS↓
3DGS	20.16	0.797	0.334
3DGS + Depth	20.19	0.794	0.334
Ours	**21.88**	**0.815**	**0.318**

Table 2. Quantitative comparisons on ScanNet++.

D. Experiments

Visualizations of virtual viewpoints. We provide the examples of sampled virtual viewpoints as shown in Fig. 2. Virtual viewpoints are widely spread in the scene, leading to maximize information gain.

Comparions with Nerfbusters [5]. To further validate the effectiveness of our approach, we provide a comprehensive comparison between Nerfbusters and our method on curated Nerfbusters dataset. As shown in Table 1 and Fig. 5, our method consistently outperforms the original Nerfbusters. Nerfbusters designs diffusion model to take an occupancy grid with artifacts as input and to predict a refined occupancy grid. Since it only removes floating points, it is unable to fill missing information while our diffusion model provides pseudo observations.

Comparison on ScanNet++ [7] For further validations, We evaluate our method and baselines on eight scenes from the ScanNet++ [7] dataset. Evaluation viewpoints of this dataset are placed at arbitrary positions, including large viewpoint changes. As observed in Table 2 and Fig. 3, our method outperform all the baselines.

Ablation study on finetuning methods. We offer additional qualitative results of finetuning methods as illustrated in Fig. 4. Although the quantitative improvements may appear marginal, our approach demonstrates its effectiveness

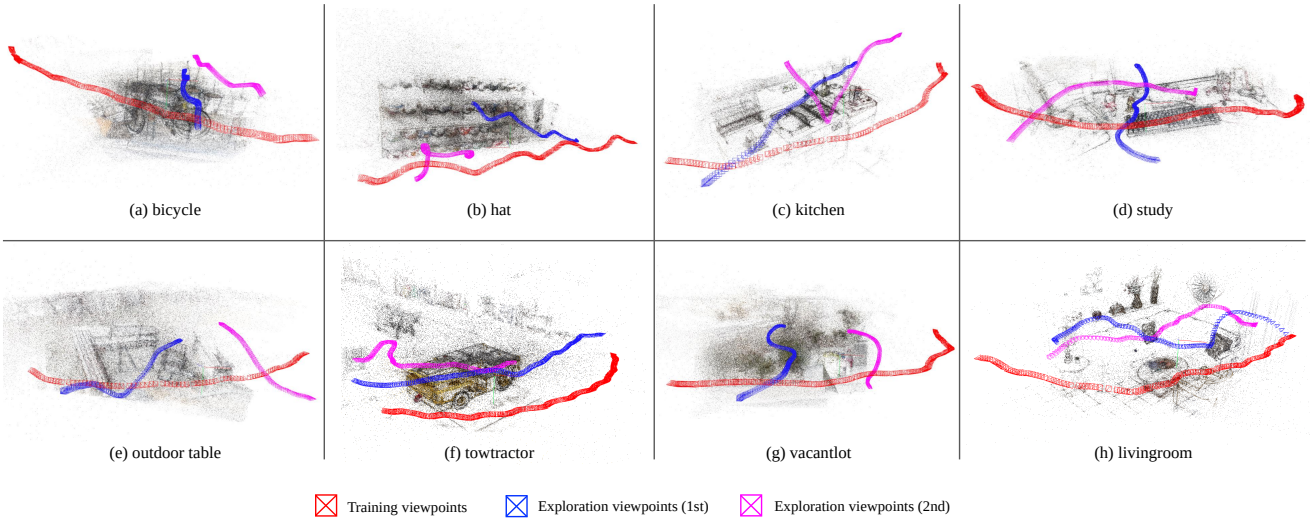

Figure 1. Visualization of our WildExplore dataset for novel view synthesis. We present all eight scenes, showing both train camera trajectories and exploration (evaluation) camera trajectories.

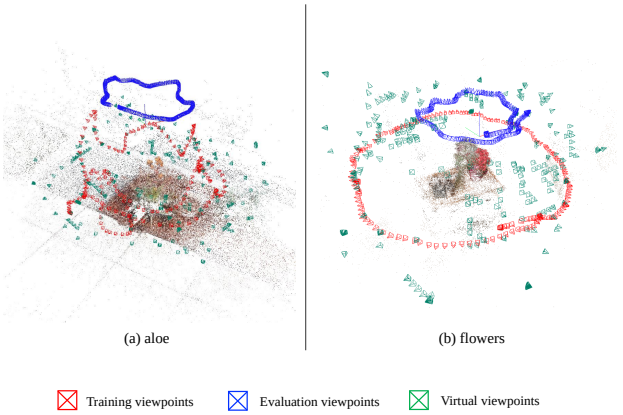

Figure 2. Visualization of virtual camera placements on Curated Nerfbusters scenes.

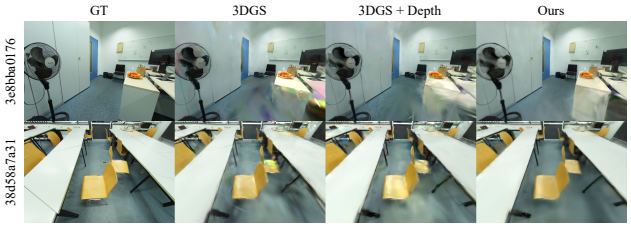

Figure 3. Qualitative results on ScanNet++. Please zoom in for the details.

in visual quality, resulting in sharper results and reduced artifacts.

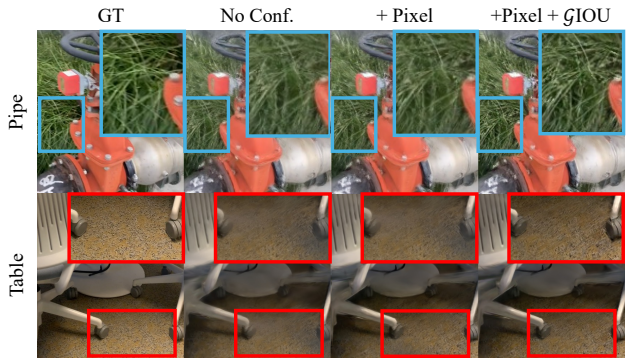

Figure 4. Qualitative ablation results on the finetuning.

E. Future works

Blurry pixels. Our method often produces blurry appearance, mainly due to the limited capacity of the diffusion model. While our diffusion model effectively fills missing regions and removes artifacts, it often struggles to recover high-frequency details (e.g. delicate floor patterns), thereby leading to less sharp pixels in final outputs. A resolution mismatch between the diffusion output and the rendered image may also introduce minor errors. Adopting larger or more improved diffusion backbones can mitigate this issue. Nevertheless, our framework is orthogonal to the choice of diffusion backbone, and can naturally benefit from future advances in diffusion models.

Initial viewpoints selection. For virtual viewpoint sampling, We opt to sample initial viewpoints from training viewpoints at uniform intervals. While it still achieves broad scene coverage, more principled initial view selection can further improve sampling efficiency.

Figure 5. Qualitative comparison between the original Nerfbusters model and our ExploreGS.

References

[1] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 1

[2] Lu Ling, Yichen Sheng, Zhi Tu, Wentian Zhao, Cheng Xin, Kun Wan, Lantao Yu, Qianyu Guo, Zixun Yu, Yawen Lu, Xuanmao Li, Xingpeng Sun, Rohan Ashok, Aniruddha Mukherjee, Hao Kang, Xiangrui Kong, Gang Hua, Tianyi Zhang, Bedrich Benes, and Aniket Bera. Dl3dv-10k: A large-scale scene dataset for deep learning-based 3d vision, 2023. 1

[3] Xi Liu, Chaoyi Zhou, and Siyu Huang. 3dgs-enhancer: Enhancing unbounded 3d gaussian splatting with view-consistent 2d diffusion priors, 2024. 1

[4] Meng-Li Shih, Wei-Chiu Ma, Lorenzo Boyice, Aleksander Holynski, Forrester Cole, Brian Curless, and Janne Kontkanen. Extranerf: Visibility-aware view extrapolation of neural radiance fields with diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20385–20395, 2024. 1

[5] Frederik Warburg, Ethan Weber, Matthew Tancik, Aleksander Holynski, and Angjoo Kanazawa. Nerfbusters: Removing ghostly artifacts from casually captured nerfs, 2023. 1

[6] Jinbo Xing, Menghan Xia, Yong Zhang, Haoxin Chen, Wangbo Yu, Hanyuan Liu, Xintao Wang, Tien-Tsin Wong, and Ying Shan. Dynamicrafter: Animating open-domain images with video diffusion priors, 2023. 1

[7] Chandan Yeshwanth, Yueh-Cheng Liu, Matthias Nießner, and Angela Dai. Scannet++: A high-fidelity dataset of 3d indoor scenes, 2023. 1