# RI3D: Few-Shot Gaussian Splatting With Repair and Inpainting Diffusion Priors

## Supplementary Material

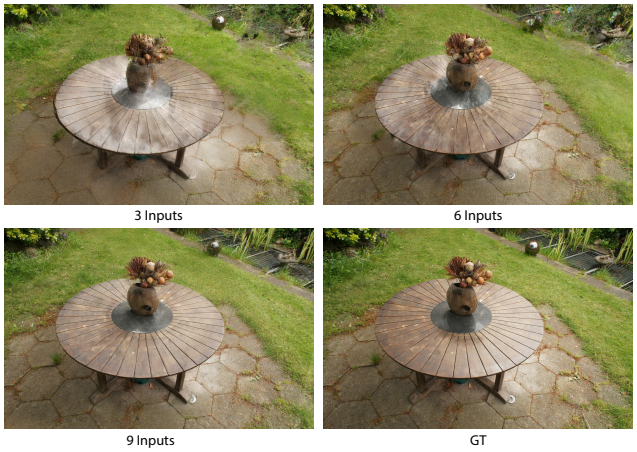


Figure 2. Number of views ablation.

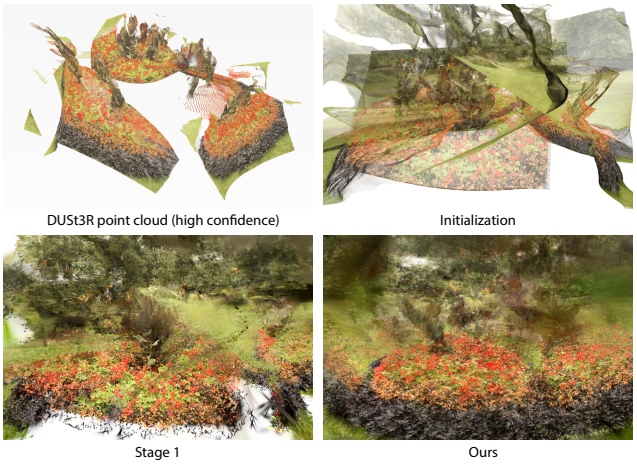


Figure 3. DUSt3R sometimes fails to generate a reasonable point-cloud on ambiguous scenes. This in turn affects our optimization quality.

## 1. Implementation Details

We provide additional implementation details for some of the models and systems described in the main paper. All optimization and fine-tuning are performed on a single Nvidia RTX A5000 GPU, except for Stage 2, which is carried out using two A5000 GPUs.

### 1.1. Repair Model

Similar to GaussianObject [12], the leave-one-out strategy is applied by introducing the left-out view after 6000 iterations of optimization. The optimization continues until iteration 10000, thereby obtaining training pairs for the repair model. We then fine-tune the repair model using these training pairs for 1800 iterations.

### 1.2. Inpainting Model

We adapt the RealFill [9] methodology to fine-tune the Stable Diffusion inpainting model on the sparse input views. Due to unavailability of code by the original authors, we utilize a third-party implementation by Nguyen [5]. Consistent with the referenced GitHub repository, we fine-tune the model for 2000 iterations.

### 1.3. Optimization

We run the **Stage 1** optimization for 4000 iterations, utilizing 8 evenly distributed novel repaired views in addition to the input training images. The repaired views are refreshed every 400 iterations.

Similarly, **Stage 2** optimization runs for 4000 iterations. During this stage, we sample 10 evenly distributed views every 200 iterations. For each sampling cycle, we sequentially inpaint and project every other view (5 views) before rendering and repairing all 10 views. Inpainting is performed up to iteration 2800, after which only the repair process is carried out to address minor artifacts.

## 2. Additional Results

We provide additional results on the **Mip-NeRF 360** [1] and **CO3D** [6] dataset for the 3-, 6- and 9-input setting. For both datasets, we utilize the training cameras provided by ReconFusion [10] in our evaluation. In addition to the numerical results presented by ReconFusion and CAT3D [3], we compare against recent state-of-the-art 3D Gaussian based sparse novel view synthesis methods including FSGS [15], CoR-GS [14] and DNGaussian [4]. In contrast to MipNeRF 360 dataset [1], which contains long range general scenes with large missing regions, most CO3D scenes are close up photos of objects in front of a simple background, e.g., a wooden table or floor. This limits the need for and the performance improvement gained from high-quality inpainting. As shown in Table 1, we outperform previous approaches including ReconFusion in terms of perceptual quality (LPIPS) while being competitive with CAT3D. We also provide visual comparisons in Fig. 5. Our approach generates complete and consistent results compared to other 3DGS-based approaches. We provide video comparisons for both MipNeRF 360 and CO3D in the supplementary video.

We provide an additional ablation result (Fig. 2) illustrating how the number of training views affects the out-

| Method | 3-view PSNR ↑ | 3-view SSIM ↑ | 3-view LPIPS ↓ | 6-view PSNR ↑ | 6-view SSIM ↑ | 6-view LPIPS ↓ | 9-view PSNR ↑ | 9-view SSIM ↑ | 9-view LPIPS ↓ |
|---|---|---|---|---|---|---|---|---|---|
| DiffusioNeRF* [11] | 15.65 | 0.575 | 0.597 | 18.05 | 0.603 | 0.544 | 19.69 | 0.631 | 0.500 |
| FreeNeRF* [13] | 13.28 | 0.461 | 0.634 | 15.20 | 0.523 | 0.596 | 17.35 | 0.575 | 0.561 |
| SimpleNeRF* [8] | 15.40 | 0.553 | 0.612 | 18.12 | 0.622 | 0.541 | 20.52 | 0.672 | 0.493 |
| Zip-NeRF* [2] | 14.34 | 0.496 | 0.652 | 14.48 | 0.497 | 0.617 | 14.97 | 0.514 | 0.590 |
| ZeroNVS* [7] | 17.13 | 0.581 | 0.566 | 19.72 | 0.627 | 0.515 | 20.50 | 0.640 | 0.500 |
| DNGaussian† [4] | 16.95 | 0.497 | 0.463 | 19.59 | 0.601 | 0.425 | 20.68 | 0.647 | 0.408 |
| CoR-GS† [14] | 14.09 | 0.487 | 0.501 | 16.93 | 0.545 | 0.463 | 17.86 | 0.544 | 0.458 |
| FSGS† [15] | 15.92 | 0.544 | 0.443 | 19.75 | 0.641 | 0.351 | 21.25 | 0.677 | 0.316 |
| ReconFusion [10] | 19.59 | 0.662 | 0.398 | 21.84 | 0.714 | 0.342 | 22.95 | 0.736 | 0.318 |
| CAT3D | 20.57 | 0.666 | 0.351 | 22.79 | 0.726 | 0.292 | 23.58 | 0.752 | 0.273 |
| **RI3D** (Ours) | 18.72 | 0.628 | 0.385 | 20.53 | 0.673 | 0.311 | 21.32 | 0.704 | 0.277 |

Table 1. We quantitatively compare our approach against other sparse view synthesis methods on CO3D dataset. PSNR and SSIM measure pixel-wise error and as such ReconFusion and CAT3D score higher in most cases, since these metrics favor blurrier results. We outperform the previous approaches in terms of perceptual quality (LPIPS) while being competitive with CAT3D, a concurrent approach. The numbers for approaches marked by * are directly grabbed from ReconFusion. CAT3D results are also obtained from the original paper. The methods marked by † are initialized using DUSt3R to improve their results.

put quality. As expected, increasing the number of training views leads to better reconstruction quality

## 3. Limitations

As discussed in the main paper, the quality of our dense initialization and consequently the final reconstruction is influenced by the accuracy of the depth maps produced by DUSt3R. In cases where DUSt3R generates highly inaccurate depth estimates, our optimization procedure cannot fully compensate for the resulting misalignments, leading to noticeable ghosting artifacts, as shown in Fig. 3. However, due to the modular design of our method, DUSt3R can be readily replaced with more accurate depth estimation models as they become available which mitigates this limitation.

## References


[1] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. *CVPR*, 2022. 1

[2] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Zip-nerf: Anti-aliased grid-based neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19697–19705, 2023. 2

[3] Ruiqi Gao, Aleksander Holynski, Philipp Henzler, Arthur Brussee, Ricardo Martin-Brualla, Pratul Srinivasan, Jonathan T Barron, and Ben Poole. Cat3d: Create anything in 3d with multi-view diffusion models. *arXiv preprint arXiv:2405.10314*, 2024. 1

[4] Jiahe Li, Jiawei Zhang, Xiao Bai, Jin Zheng, Xin Ning, Jun Zhou, and Lin Gu. Dngaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20775–20785, 2024. 1, 2

[5] Thuan H. Nguyen. Unofficial implementation of realfill. https://github.com/thuanz123/realfill, 2023. 1

[6] Jeremy Reizenstein, Roman Shapovalov, Philipp Henzler, Luca Sbordone, Patrick Labatut, and David Novotny. Common objects in 3d: Large-scale learning and evaluation of real-life 3d category reconstruction. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10901–10911, 2021. 1

[7] Kyle Sargent, Zizhang Li, Tanmay Shah, Charles Herrmann, Hong-Xing Yu, Yunzhi Zhang, Eric Ryan Chan, Dmitry Lagun, Li Fei-Fei, Deqing Sun, et al. Zeronvs: Zero-shot 360-degree view synthesis from a single real image. *arXiv preprint arXiv:2310.17994*, 2023. 2

[8] Nagabhushan Somraj, Adithyan Karanayil, and Rajiv Soundararajan. Simplenerf: Regularizing sparse input neural radiance fields with simpler solutions. In *SIGGRAPH Asia 2023 Conference Papers*, pages 1–11, 2023. 2

[9] Luming Tang, Nataniel Ruiz, Qinghao Chu, Yuanzhen Li, Aleksander Holynski, David E. Jacobs, Bharath Hariharan, Yael Pritch, Neal Wadhwa, Kfir Aberman, and Michael Rubinstein. Realfill: Reference-driven generation for authentic image completion. *ACM Trans. Graph.*, 43(4), 2024. 1

[10] Rundi Wu, Ben Mildenhall, Philipp Henzler, Keunhong Park, Ruiqi Gao, Daniel Watson, Pratul P Srinivasan, Dor Verbin, Jonathan T Barron, Ben Poole, et al. Reconfusion: 3d reconstruction with diffusion priors. *arXiv preprint arXiv:2312.02981*, 2023. 1, 2

[11] Jamie Wynn and Daniyar Turmukhambetov. Diffusionerf: Regularizing neural radiance fields with denoising diffusion models. In *Proceedings of the IEEE/CVF Conference on*

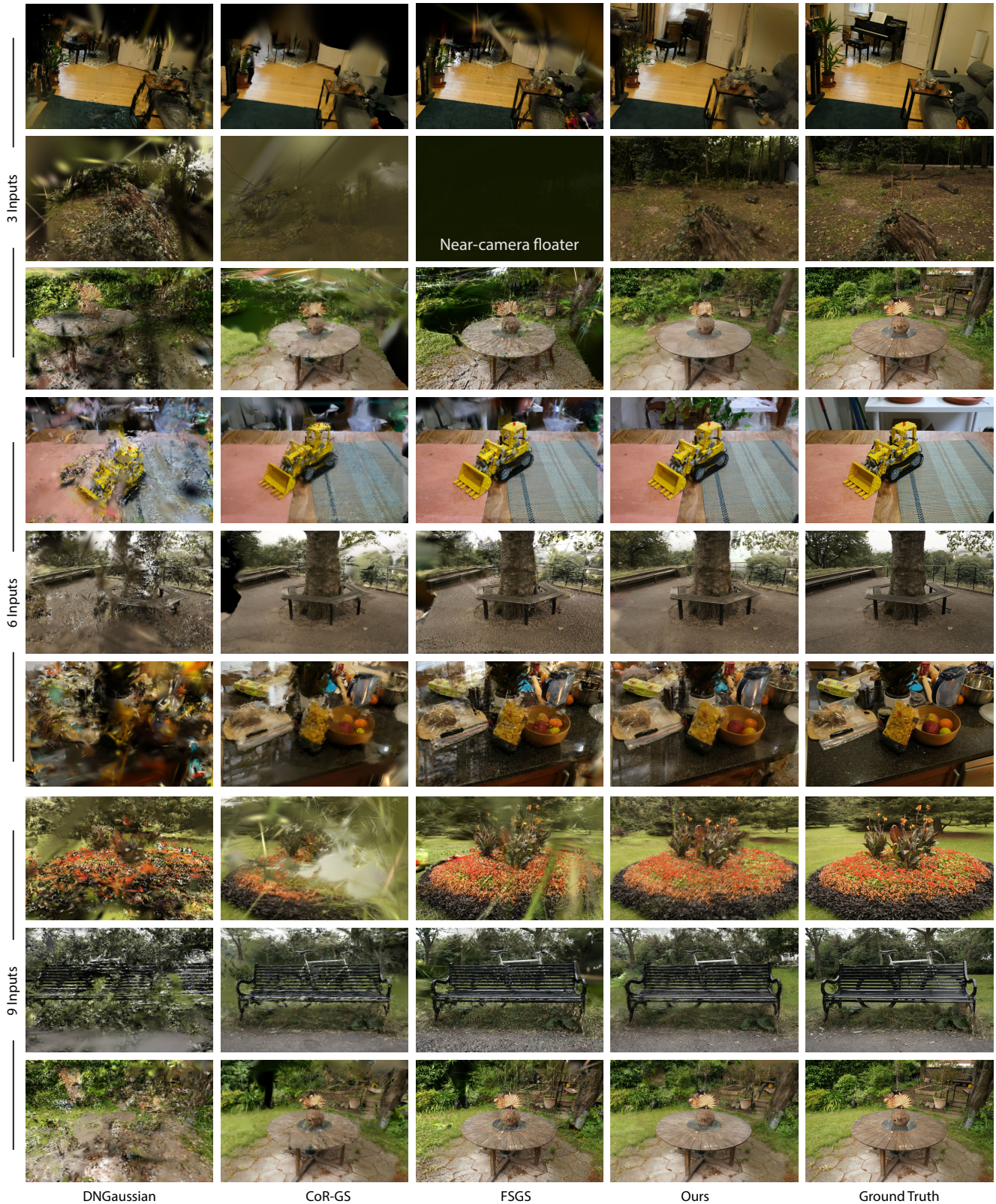


Figure 4. We compare our approach against the other state-of-the-art sparse view synthesis methods on a few scenes from the Mip-NeRF 360 dataset.

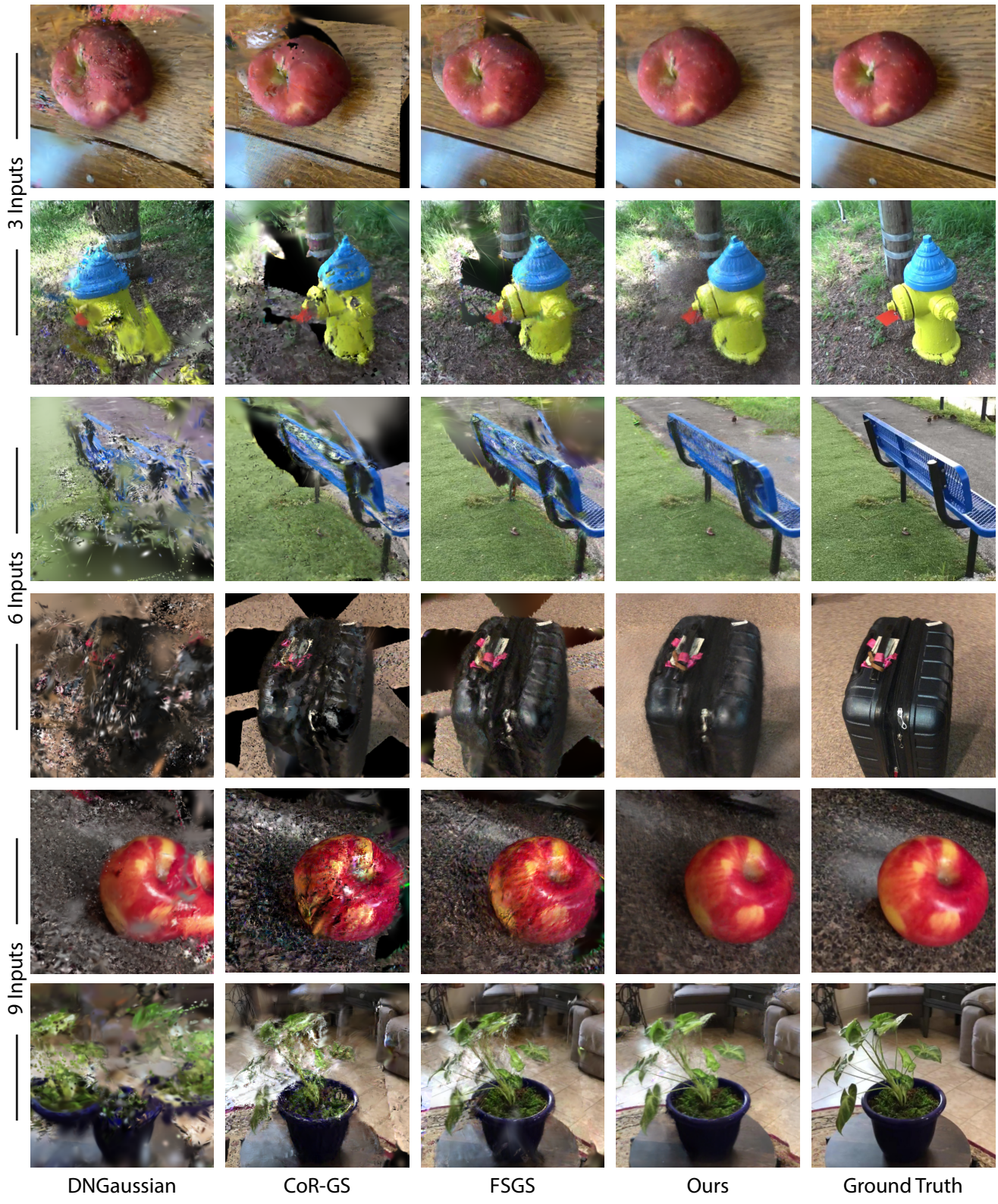


Figure 5. We compare our approach against the other state-of-the-art sparse view synthesis methods on a few scenes from the CO3D dataset.

*Computer Vision and Pattern Recognition (CVPR)*, pages 4180–4189, 2023. 2

[12] Chen Yang, Sikuang Li, Jiemin Fang, Ruofan Liang, Lingxi Xie, Xiaopeng Zhang, Wei Shen, and Qi Tian. Gaussianobject: Just taking four images to get a high-quality 3d object with gaussian splatting. *arXiv preprint arXiv:2402.10259*, 2024. 1

[13] Jiawei Yang, Marco Pavone, and Yue Wang. Freenerf: Improving few-shot neural rendering with free frequency regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8254–8263, 2023. 2

[14] Jiawei Zhang, Jiahe Li, Xiaohan Yu, Lei Huang, Lin Gu, Jin Zheng, and Xiao Bai. Cor-gs: sparse-view 3d gaussian splatting via co-regularization. In *European Conference on Computer Vision*, pages 335–352. Springer, 2025. 1, 2

[15] Zehao Zhu, Zhiwen Fan, Yifan Jiang, and Zhangyang Wang. Fsgs: Real-time few-shot view synthesis using gaussian splatting. *arXiv preprint arXiv:2312.00451*, 2023. 1, 2