

Supplementary Material for Gaussian Splatting in Mirrors: Reflection-aware Rendering via Virtual Camera Optimization

Zihan Wang¹

zihan.1.wang@aalto.fi

Shuzhe Wang¹

shuzhe.wang@aalto.fi

Matias Turkulainen^{1,3}

matias.turkulainen@aalto.fi

Junyuan Fang¹

junyuan.fang@aalto.fi

Juho Kannala^{1,2}

juho.kannala@aalto.fi

¹ Aalto University

² University of Oulu

³ ETH Zurich

1 Method

1.1 Details of plane estimation process

In Section 2.3, we illustrate how to estimate the mirror plane equation by leveraging the predicted depth map, normal map, and pixel-level mirror region mask. Here is a more detailed explanation of the plane estimation process.

We first randomly select a view containing a large mirror region and obtain the depth and normal values in the mirror region using the provided mask. Next, the world coordinates \hat{C} are calculated using the estimated depth map and camera parameters. Each 3D coordinate is associated with a normal value from the corresponding pixel on the normal map \hat{N}_M . We then compute the mean of the normal values as the normal of the mirror plane, denoted as \mathbf{n} . Finally, we fit the plane equation using \mathbf{n} and the 3D points from \hat{C} in a RANSAC [1] loop to obtain the best plane offset \mathbf{o} . During our implementation, a large mirror region means the amount of mirror region pixels in that view should be greater than 30000 pixels and the maximum RANSAC iteration is set to 1000 with the distance threshold for offset is 0.1. The detailed process is presented in Algorithm 1.

Algorithm 1: Plane estimation

Input: Depth map \hat{D} , Normal map \hat{N} , Mirror region mask M , extrinsic matrix $[\mathbf{R}|\mathbf{t}]$, intrinsic matrix \mathbf{K}

Output: Predicted plane equation $\hat{\pi} = [\mathbf{n}, \mathbf{o}]$

- 1 Extract \hat{D}_M, \hat{N}_M from \hat{D}, \hat{N} using M ; // \hat{D}_M, \hat{N}_M are depth and normal map on mirror region
- 2 $\hat{C} \leftarrow \text{ProjectToWorld}(\hat{D}_M, \mathbf{R}, \mathbf{t}, \mathbf{K})$; // \hat{C} is mirror region points in world coordinates
- 3 $\mathbf{n} \leftarrow \text{Mean}(\hat{N}_M)$; // Plane Normal Prediction
- 4 $\mathbf{o} \leftarrow \text{RANSAC}(\mathbf{n}, \hat{C})$; // Plane Offset prediction
- 5 **return** $\hat{\pi} = [\mathbf{n}, \mathbf{o}]$

2 Experiment

2.1 Additional implementation details

During training, the batch size is set to 1 for all scenes. Each batch contains an RGB image, the corresponding mirror region mask, and the camera pose. The maximum width of the image is set to 1600px following the protocol in 3D-GS [14]. During testing, we resize the rendered images to match the shape used by Mirror-NeRF [14] for its synthetic and real datasets. In our dataset, we resize all rendered images to 400×400 pixels to compute the evaluation metrics.

The official implementation of 3D-GS assumes that the c_x and c_y values are always at the center of the image. However, this assumption can lead to blurry images in real scenes captured by phone cameras. To address this issue, we extend the 3D-GS baseline by incorporating c_x and c_y as inputs when calculating the intrinsic matrix for both Mirror-NeRF’s real dataset and our dataset.

2.2 Additional details of our dataset

We propose a dataset that encompasses scenes of different scales and various mirror shapes. *Corridor* features a large mirror on the wall. This scene is relatively straightforward for reflection-based methods as it provides sufficient planar surfaces from the wall, doors, and ground. *Recovery Room* is captured in a small-scale room. In this scene, a small circular mirror is placed on the sofa. It is more complex than the *Corridor*, containing several pieces of furniture such as a sofa, television, and desk. However, most objects have planar or smooth surfaces, which facilitates depth and normal regularization based on planar surface assumptions. *Work Space* includes various irregular objects like cables, clothes, poster rolls, and small items on the desk. A medium-sized mirror is placed along the door. The lack of planar surfaces can result in incomplete 3D reconstruction and challenges in depth and normal estimation, making it difficult to predict and optimize the virtual camera pose. The diversity in scene complexity and mirror shapes in this dataset provides a robust testbed for evaluating reflection-based methods in different environments.

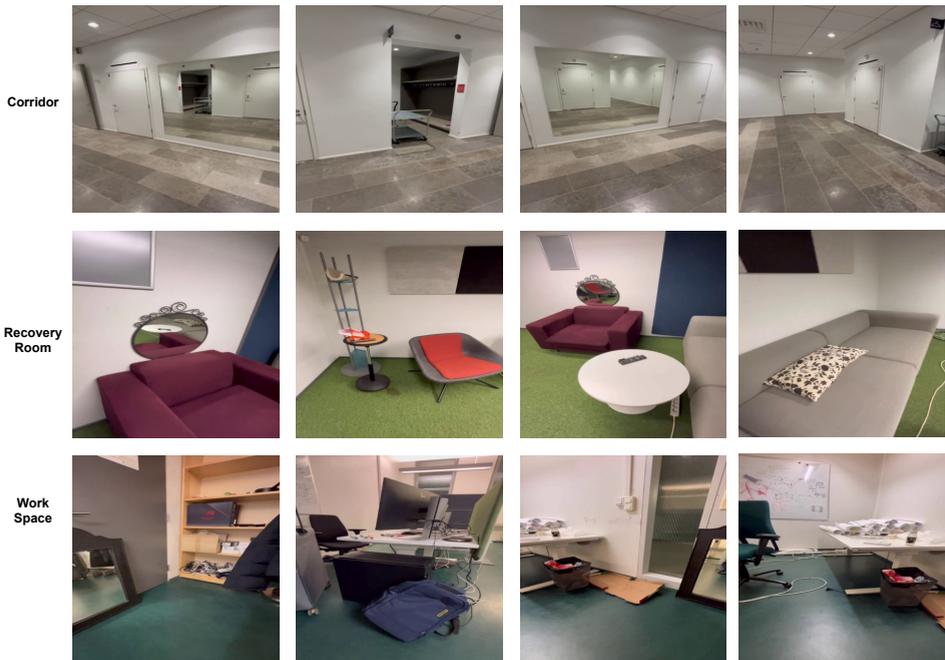


Figure 1: The visualization of our dataset.

Methods	Corridor (easy)			Recovery Room (medium)			Work Space (complex)		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
3D-GS [10]	25.9	0.845	0.332	28.84	0.917	0.166	26.88	0.884	0.219
Mirror-NeRF [9]	24.61	0.821	0.396	25.65	0.878	0.260	17.51	0.674	0.493
Ours	29.14	0.874	0.291	32.21	0.938	0.139	24.89	0.849	0.288

Table 1: Results on the our real dataset, Our method outperforms 3D-GS on two of the scenes and outperforms Mirror-NeRF in all three scenes. The best result is in **bold**.

2.3 Comparison result on our dataset

We have extended the quantitative results for our dataset by including the performance of Mirror-NeRF [9]. Our findings indicate that Mirror-NeRF struggles to accurately predict the normal and depth of mirror regions in our proposed dataset, resulting in blurry renderings of mirror region. As shown in Table 1, our method outperforms Mirror-NeRF across all three scenes.

Additional visualization results on both Mirror-NeRF’s dataset and our dataset are presented in Fig. 2.

2.4 Ablation study for mirror plane estimation

In mirror plane estimation, as described in the main paper, we randomly select a view with a large mirror region and then use this view to estimate the mirror plane. To further investi-

Scene	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Living_room	43.57 ± 0.321	$0.992 \pm 6.9 \times 10^{-7}$	$0.011 \pm 4 \times 6^{-6}$

Table 2: Ablation study result for different view selection in mirror plane estimation.

gate whether this random selection affects the quantitative results, we conduct an additional ablation study. Specifically, we utilize the *Living Room* scene from the Mirror-NeRF dataset and randomly choose 5 different views for plane estimation. Each view is trained using the same settings as outlined in our paper, and the metrics are computed as presented in Table 2. The results suggest that varying the selected views has minimal impact on the final rendering quality, underscoring the robustness of our approach in accommodating random view selection scenarios.

References

- [1] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [2] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):1–14, 2023.
- [3] Junyi Zeng, Chong Bao, Rui Chen, Zilong Dong, Guofeng Zhang, Hujun Bao, and Zhaopeng Cui. Mirror-nerf: Learning neural radiance fields for mirrors with whitted-style ray tracing. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 4606–4615, 2023.

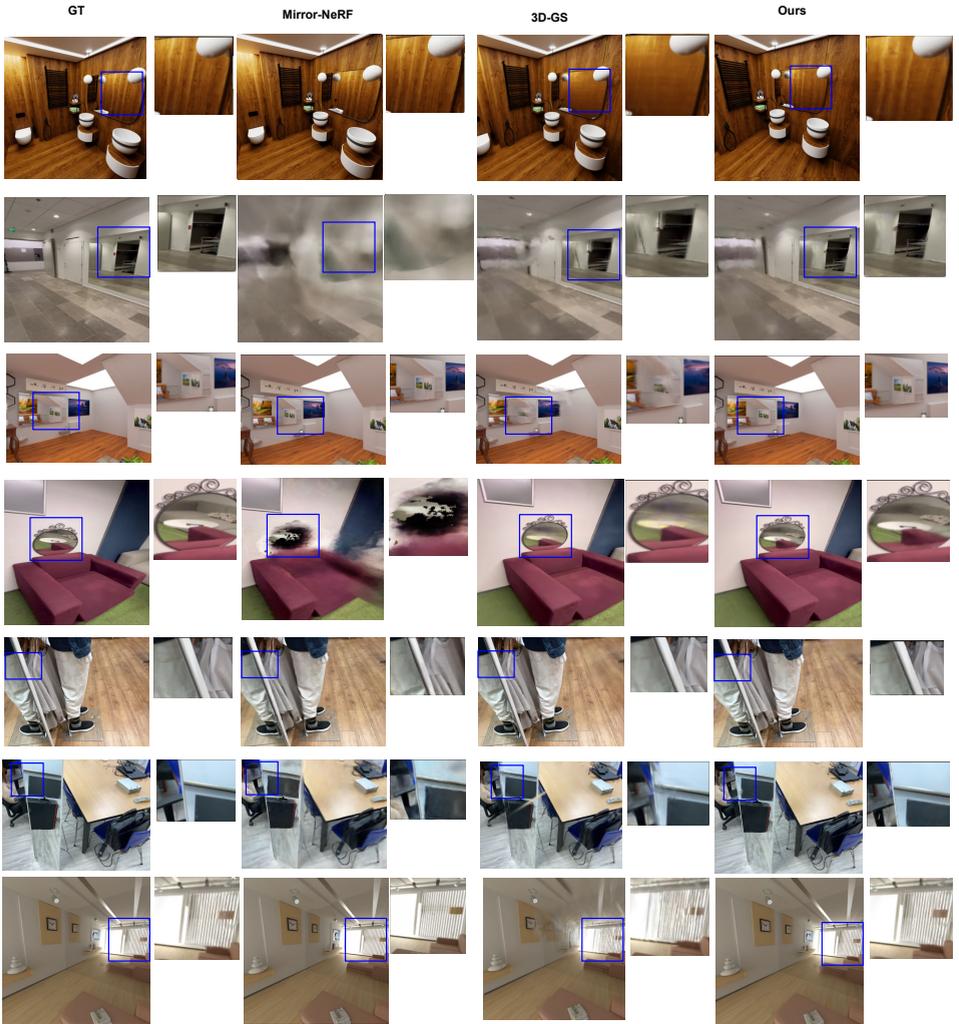


Figure 2: The visual comparison of GT, Mirror-NeRF, 3D-GS, and ours method. The smaller image in the upper right corner of each main images is an enlargement of a mirror region.