2DGS-Room: Seed-Guided 2D Gaussian Splatting with Geometric Constraints for High-Fidelity Indoor Scene Reconstruction

Supplementary Material

In this supplementary material, we provide the following components:

- Definitions of the 3D geometry metrics used to evaluate reconstruction quality in Sec. A.
- Additional details of the datasets, training configuration, and the iteration schedule for key modules in Sec. B.
- Additional qualitative results, including mesh comparison, ablation results, and rendering comparison in Sec. C.

A. Definitions of Eevaluation Metrics

We evaluate our method using five widely-used 3D geometry metrics: Accuracy, Completion, Precision, Recall, and F-score, defined in Table 1. These metrics collectively assess the geometric fidelity of the reconstructed point clouds by measuring the alignment between the predicted and ground truth point clouds.

Accuracy measures the average distance between reconstructed points and the ground truth, with smaller values indicating better alignment. Completion assesses how well the reconstruction covers the ground truth, where lower values are better. Precision and Recall evaluate the proportion of points within a set threshold, with higher values indicating better performance. F-score, the harmonic mean of Precision and Recall, provides a balanced measure of reconstruction quality, where higher values reflect superior results.

Metric	Definition
Acc.	$\text{mean}_{c \in C}(\min_{c^* \in C^*} \|\|c - c^*\|\|)$
Comp.	$\text{mean}_{c^* \in C^*}(\min_{c \in C} \|\|c - c^*\|\|)$
Prec.	$\text{mean}_{c \in C}(\min_{c^* \in C^*} \|\|c - c^*\|\| < .05)$
Recall	$\text{mean}_{c^* \in C^*}(\min_{c \in C} \|\|c - c^*\|\| < .05)$
F-score	$\frac{2 \times \text{Prec} \times \text{Recall}}{\text{Prec} + \text{Recall}}$

Table 1. **Definitions of 3D metrics.** c and c^* are the predicted and ground truth point clouds.

B. Additional Implementation Details

Datasets. As described in the main paper, the quantitative evaluation metrics are derived from results tested two datasets. Specifically, we select 8 scenes from the ScanNet dataset: scene0050_00, scene0085_00, scene0114_02, scene0580_00, scene0603_00, scene0616_00, scene0617_00, scene0721_00, and 4 scenes from the ScanNet++ dataset: 8b5caf3398, 8d563fc2cc, 41b00feddb, b20a261fdf.

Training details. For all scenes, our seed-guided optimization is performed between 1,500 and 15,000 iterations. We set $N_g = 100$ for the gradient-guided growth and $N_\alpha = 100$ for the pruning strategy. Depth supervision and normal supervision are applied consistently from the first iteration through to the end of training, providing continuous geometric constraints. The multi-view consistency constraint is introduced after 7,000 iterations, once the foundational structure has been established, to further improve view alignment.

C. Additional Qualitative Results

C.1. Additional Ablation Results

To complement the local detail comparisons in the main paper, we provide additional ablation results focusing on the overall scene structure in Figure 2. These visualizations highlight the contributions of key components, including the seed points guidance, monocular depth supervision, and monocular normal supervision. The multi-view consistency constraints are primarily designed to further mitigate floating artifacts in certain scenarios, which have a limited impact on the overall structure. Therefore, they are not included in these structural comparisons. Their effectiveness is instead reflected in the qualitative results shown in Figure **??** and the quantitative metrics presented in Table **??** of the main paper.

When the seed points guidance strategy is removed, the reconstructed objects appear fused together, with unclear boundaries, compromising the scene's structural clarity.

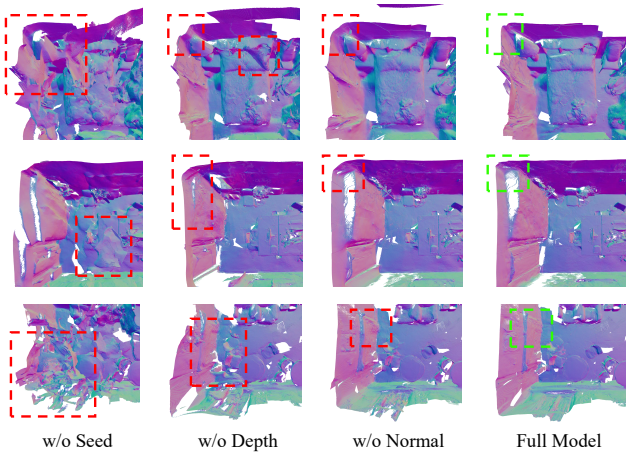

Figure 2. **Additional qualitative results of ablation study.**

Without depth supervision, objects exhibit depth misalignments, leading to unrealistic spatial arrangements. Similarly, excluding normal supervision results in uneven surfaces, especially on planar regions like walls, where visible curvature or misalignment artifacts occur.

C.2. Additional Qualitative Comparison

In addition to the four indoor scenes shown in the main paper, we further include qualitative reconstruction comparison results of the different methods on additional scenes from ScanNet and ScanNet++. As demonstrated in Figure 3, our method significantly outperforms other approaches in capturing global structures, preserving fine-grained details as well as reducing artifacts in textureless regions.

C.3. Rendering Comparison

We also provide extensive rendering results comparing our 2DGS-Room with 2DGS across various scenes and viewpoints from the ScanNet and ScanNet++ datasets in Figures 4, 5, and 6. Rendered RGB, depth, and normal maps are shown for visual comparison. Our method achieves significant improvements in the rendering quality of depth and normal maps, showcasing smoother transitions and more accurate surface details. Furthermore, the quality of the RGB images rendered by our method remains robust and shows clear advantages over 2DGS in challenging scenarios, such as handling fine details and varying lighting conditions. This demonstrates the effectiveness of our method in achieving superior geometric reconstructions while maintaining photometric accuracy.

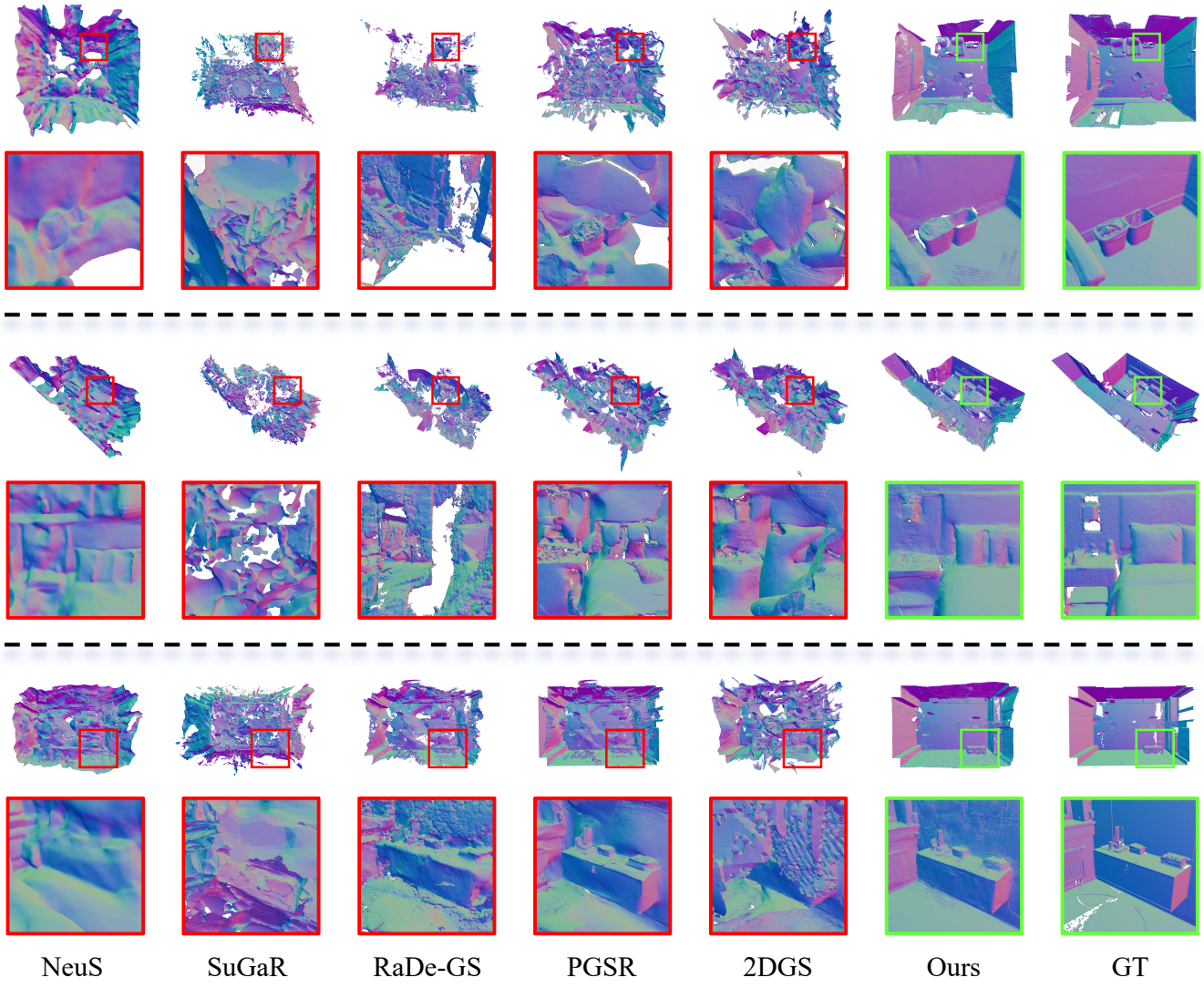

Figure 3. **Additional qualitative reconstruction comparison.** For each indoor scene, the first row is the top view of the whole room and the second row is the details of the masked region.

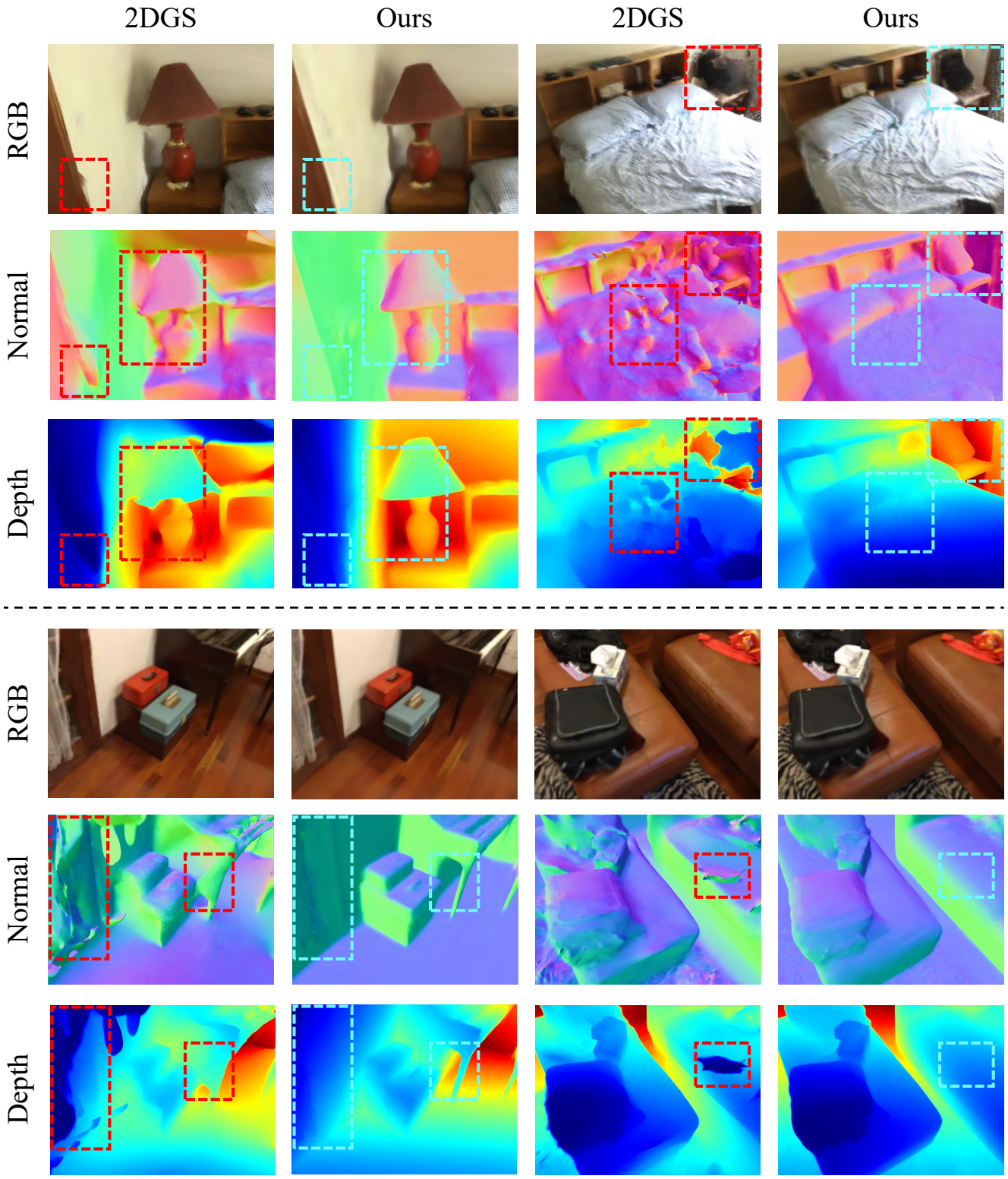

Figure 4. **Rendering comparison on the ScanNet dataset (scene0580 and scene0050).**

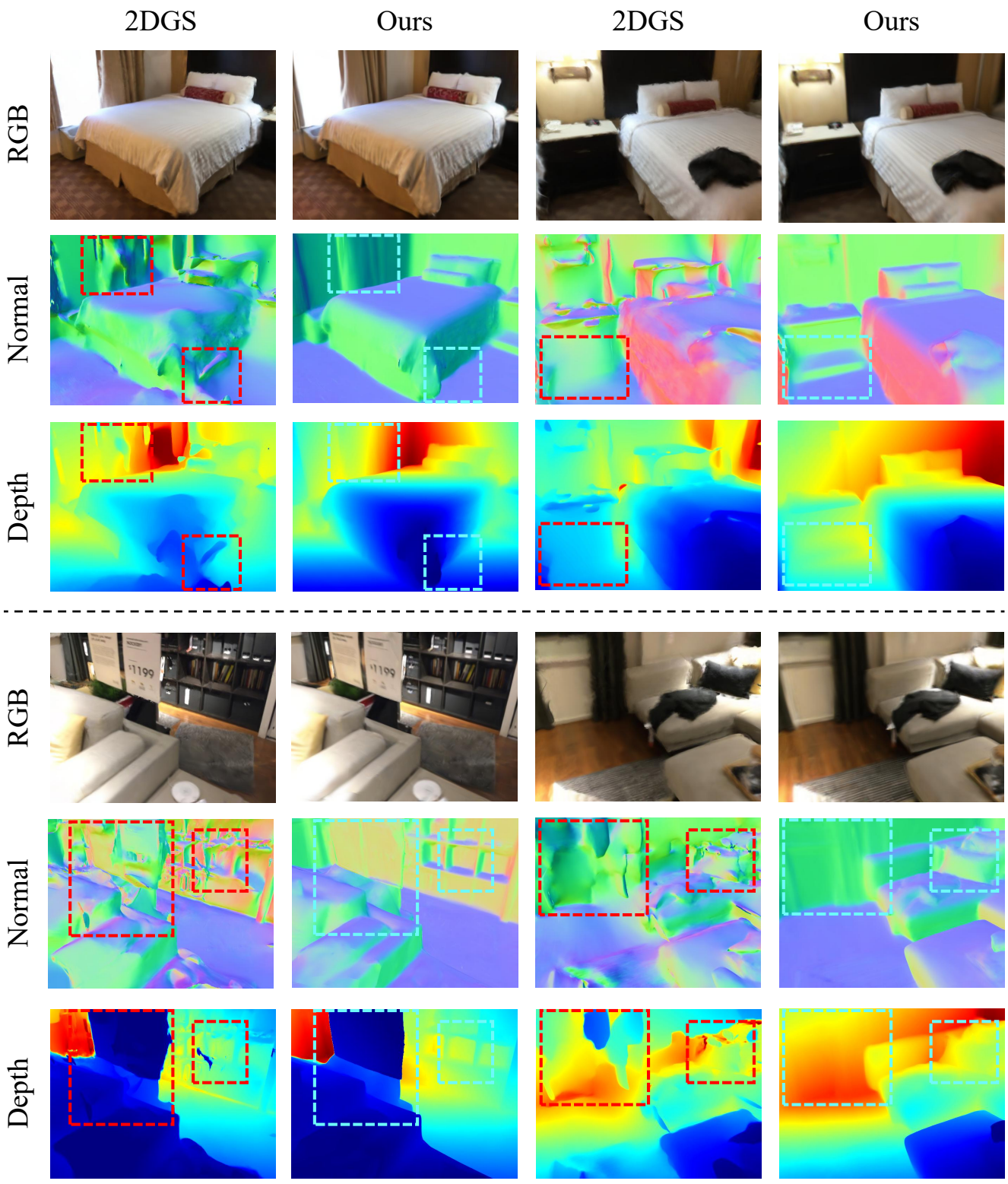

Figure 5. **Rendering comparison on the ScanNet dataset (scene0085 and scene0617).**

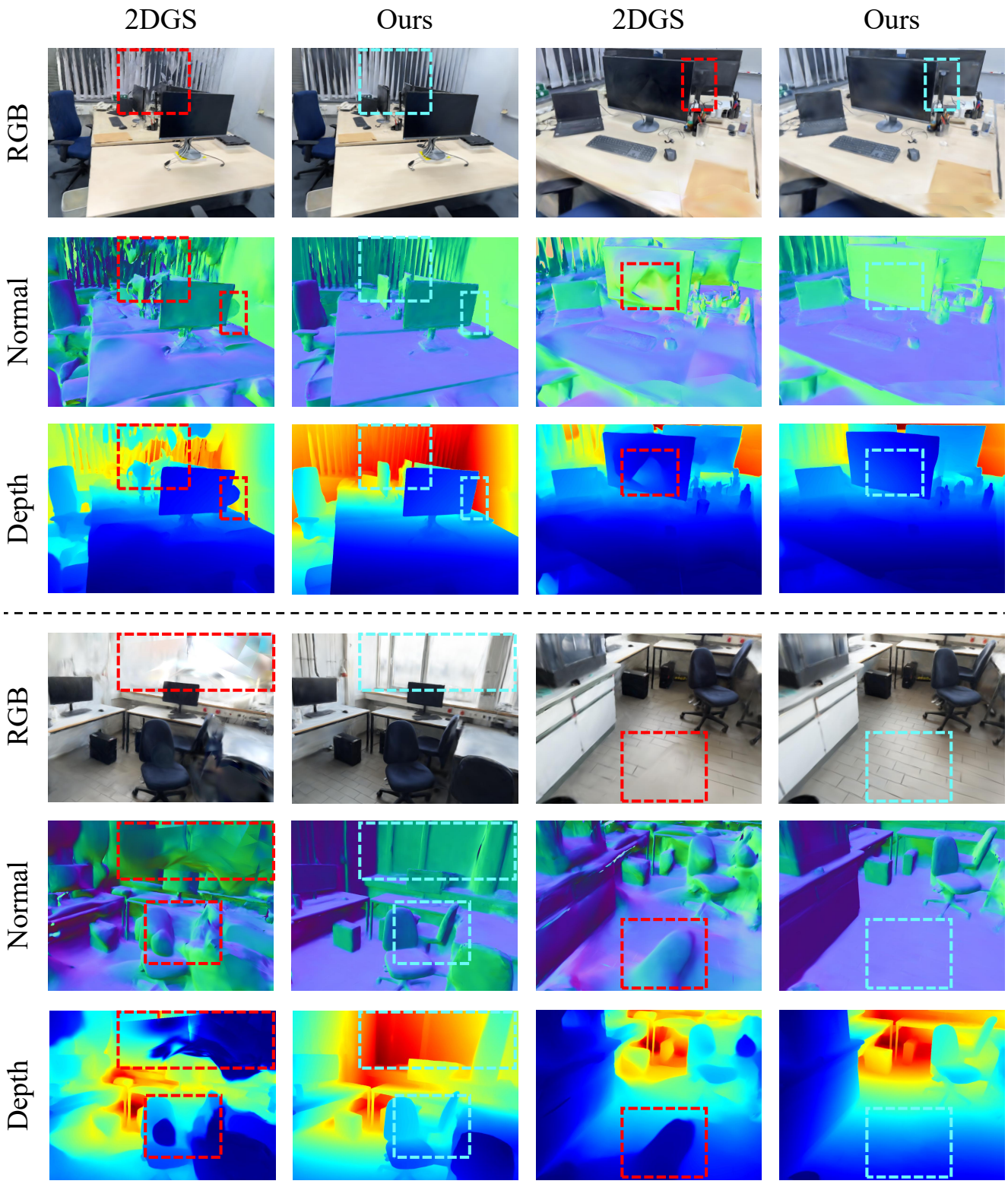

Figure 6. **Rendering comparison on the ScanNet++ dataset (8d563fc2cc and 41b00feddb).**