Reflect3r: Single-View 3D Stereo Reconstruction Aided by Mirror Reflections

Supplementary Material

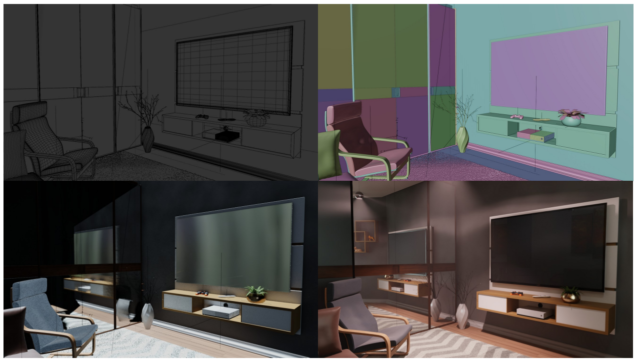

Figure 9. Details of the **synthetic** dataset's Blender modeling. Top left: wireframe view; top right: solid view; bottom left: material view; bottom right: rendered view. The scene is fully customizable, allowing adjustments to the mirror's position and properties, as well as the room setup and lighting, enabling easy extension of the dataset.

7. Implementation Details

We explain more details about the mirror plane recovery in Sec. 5.1. Since DUSt3R lacks direct correspondences in the mirror region, its predicted mirror plane is often inaccurate in position, though the normal direction is generally reliable. Our objective is to recover both the plane's normal and its position. Our goal is to obtain the normal and the position of the plane. The normal is estimated using the method described in Eq. (11). For plane positioning, we first extract edge points in the main view via image-space edge detection and back-project them into 3D. RANSAC is then applied to remove outliers. To improve robustness, we retain the top 10% of edge points ranked by DUSt3R 's confidence scores and randomly select one as a reliable anchor point to finalize the plane position.

8. More Details About the Synthetic Data

We provide the thumbnail of all 16 scenes included in our dataset in Fig. 10.

Figure 9 shows an example Blender scene in our dataset. Each scene in the dataset can be fully customized, including object shapes, room layouts, furniture placement, material properties, lighting conditions, etc. We additionally insert and adjust mirror surfaces with controllable positions and reflectance properties to simulate realistic reflective setups. This design not only ensures diverse and detailed scenes for training and evaluation but also provides a flexible foundation for extending the dataset in future work.

Figure 10. Thumbnails of the 16 fully editable Blender scenes included in our dataset.

9. More Details About the Experiments

9.1. Details of Evaluation Metrics

We cover the math of the evaluation metrics in this section.

Given the ground-truth point cloud $\mathcal{P} = \{\mathbf{p}_i\}_{i=1}^{N_P}$ and the reconstructed point cloud $\mathcal{Q} = \{\mathbf{q}_i\}_{i=1}^{N_Q}$, Chamfer Distance is computed as

$$\frac{1}{2N_P}\sum_{\mathbf{p}\in\mathcal{P}}\min_{\mathbf{q}\in\mathcal{Q}}||\mathbf{p}-\mathbf{q}||_2+\frac{1}{2N_Q}\sum_{\mathbf{q}\in\mathcal{Q}}\min_{\mathbf{p}\in\mathcal{P}}||\mathbf{p}-\mathbf{q}||_2 \tag{17}$$

In our paper, we report the completeness (Comp.) and the accuracy (Accu.) in percentage with a threshold of 1 cm of the point cloud distance. Specifically, let the indicator function be $\mathbf{1}[\cdot]$,

$$\text{Comp.}=\frac{1}{N_P}\sum_{\mathbf{p}\in\mathcal{P}}\mathbf{1}[\min_{\mathbf{q}\in\mathcal{Q}}||\mathbf{p}-\mathbf{q}||_2<1\text{cm}]; \tag{18}$$

$$\text{Accu.}=\frac{1}{N_Q}\sum_{\mathbf{q}\in\mathcal{Q}}\mathbf{1}[\min_{\mathbf{p}\in\mathcal{P}}||\mathbf{p}-\mathbf{q}||_2<1\text{cm}]; \tag{19}$$

$$\text{F1}=\frac{2\cdot\text{Comp.}\cdot\text{Accu.}}{\text{Comp.}+\text{Accu.}}. \tag{20}$$

9.2. More Statistics And Analysis

To further assess robustness, we visualize the score distributions in Fig. 11 using boxplots for completeness, accuracy, and F1 score on the synthetic dataset. Reflect3r achieves the highest median scores across all metrics while also exhibiting the smallest interquartile range and whisker span, indicating low variance and strong stability across scenes. MoGe attains similar median scores to Reflect3r, suggesting that its predicted point clouds are close to the ground-truth, but its results fluctuate significantly, reflecting high variance. Its completeness is notably lower, as MoGe does not leverage the stereo information provided by mirrors and therefore fails to recover occluded regions, leaving parts of the scene uncovered. DUSt3R and VGGT, unable to identify mirrors, generate false geometry in reflective regions and consequently show much higher variability. MASt3R, which cannot handle single-view reconstruction, consistently predicts flat geometry and therefore yields uniformly low scores.

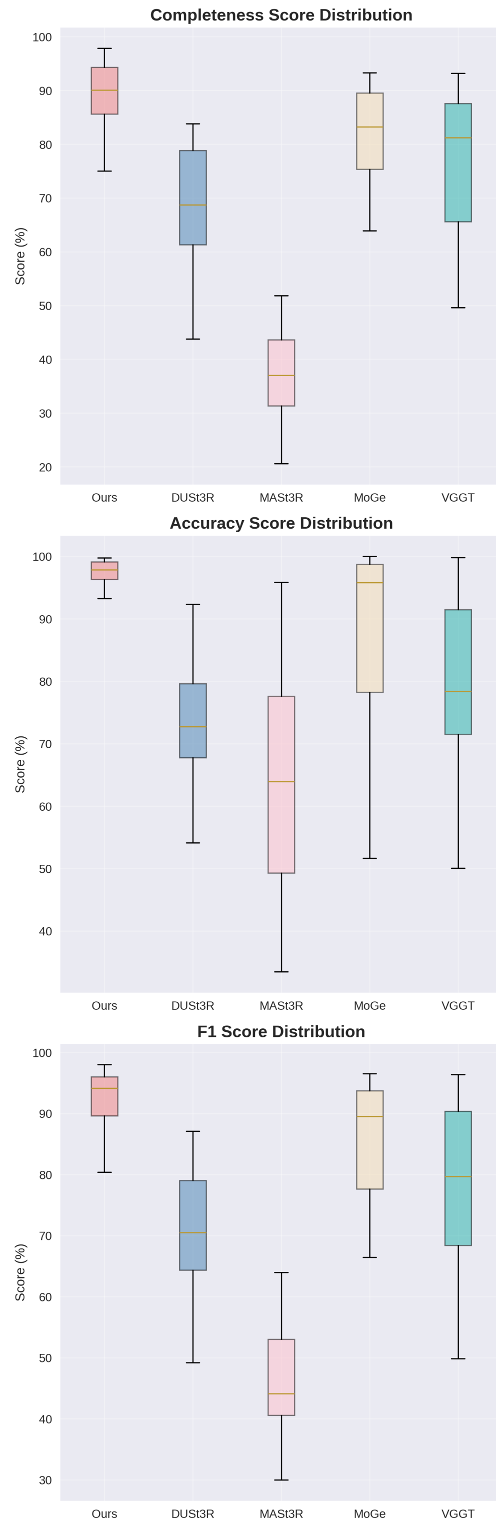

Figure 11. Boxplots of completeness, accuracy, and F1 score on the synthetic dataset.