GS-2DGS: Geometrically Supervised 2DGS for Reflective Object Reconstruction

Supplementary Material

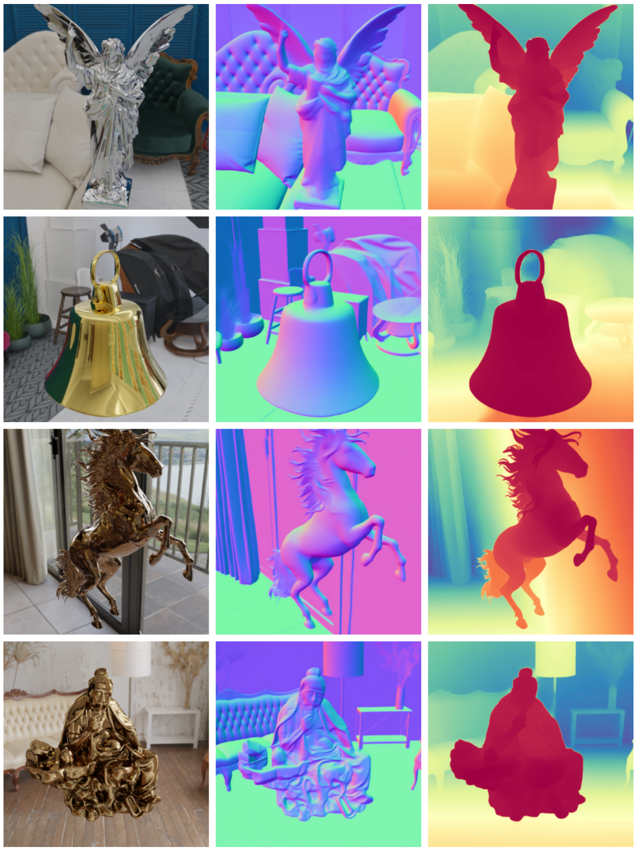

Figure S1. Normal and depth estimation results on the Glossy Blender dataset.

1. Additional Implementation Details

1.1. Representation

We implement GS-2DGS mainly based on the original 2DGS. The 2DGS uses $\{\mathbf{x}, \mathbf{s}, \mathbf{t}\}$ and $\{\alpha, \mathbf{c}\}$ to represent its geometric and volumetric appearance properties respectively, where $\mathbf{x}$ is the position of the Gaussian, $\mathbf{s}$ and $\mathbf{t}$ are the scale and rotation of the axis, α is the opacity and $\mathbf{c}$ is the SH coefficients. Apart from those properties, we add PBR parameters $\{\mathbf{a}, \mathbf{m}, \mathrm{and} \mathbf{r}\}$ to represent the albedo, metalness, and roughness of the Gaussian.

1.2. Usage of Foundation Models

Normal Estimation. We adopt the pre-trained model of Marigold [2] fine-tuned by [5] for normal estimation for its robust performance on the reflective objects. We follow the original code [1] and default pipeline to use the RGB images as input and predict the normal map.

Depth Estimation. For depth estimation, we use the Depth Pro [1] model for its superior performance on the object level prediction. We use the original implementation [2] and follow the original pipeline to predict the depth map. Specially, for the StanfordORB dataset, we use the masked RGB images without the background for better performance.

We show some examples of the normal and depth estimation results in Fig. S1 and Fig. S2.

1.3. StanfordORB dataset

Object	Training Scene	Evaluation Scenes
Baking	scene001	scene002, scene003
Ball	scene003	scene002, scene004
Blocks	scene005	scene002, scene006
Cactus	scene001	scene005, scene007
Car	scene004	scene002, scene006
Chips	scene003	scene002, scene004
Cup	scene006	scene003, scene007
Curry	scene001	scene005, scene007
Gnome	scene003	scene005, scene007
Grogu	scene001	scene002, scene003
Pepsi	scene003	scene002, scene004
Pitcher	scene007	scene001, scene005
Salt	scene007	scene004, scene005
Teapot	scene002	scene001, scene006

Table S1. Training and evaluation scenes for each object in the StanfordORB dataset.

The StanfordORB dataset contains 14 objects each in three scenes with different lighting conditions. Both low dynamic range (LDR) and high dynamic range (HDR) images are offered. For each object, we randomly picked one scene for training and the other two for evaluation as illustrated in Tab. S1. The LDR images are downsampled to 1024×1024 for training.

[1] https://github.com/VisualComputingInstitute/diffusion-e2e-ft

[2] https://github.com/apple/ml-depth-pro

Figure S2. Normal and depth estimation results on the StanfordORB dataset.

2. Additional Experiments

2.1. Results on Glossy Blender dataset

Reconstruction Fig. S3 shows the additional reconstruction results on the Glossy Blender dataset. Our method achieves the best reconstruction quality among all the Gaussian-based methods. Compared to the SDF-based method, we achieve comparable results with NeRO [4] and better than TensoSDF [3]. We also show the normal map results in Fig. S4.

Material Decomposition and Relighting. We show the PBR related material parameters decomposition in Fig. S5 and Fig. S6. We render the corresponding parameter map by alpha blending according to Eq.(3) of the main paper. Compared to other methods, our method can get a more reasonable decomposition of the PBR parameters with less noise. We also show the relighting results of different objects under several environment lights in Fig. S7.

2.2. Results on StanfordORB dataset

Here, we provide more details of the experiment results on the StanfordORB dataset.

Reconstruction In Fig. S8, we show the comparison of reconstructed geometry among different methods. Our method achieves faithful reconstruction results with less noise and smooth surfaces.

Material Decomposition and Relighting. Fig. S9 shows the PBR material decomposition in terms of albedo and roughness. From the results, our method can achieve a more reasonable decomposition of the PBR parameters while GS-IR tends to get high roughness and R3DG's roughness maps are noisy. We also show the relighting results in Tab. S2 and Fig. S10 to demonstrate the relighting quality further.

References

[1] Aleksei Bochkovskii, Amaël Delaunoy, Hugo Germain, Marcel Santos, Yichao Zhou, Stephan R Richter, and Vladlen Koltun. Depth pro: Sharp monocular metric depth in less than a second. *arXiv preprint arXiv:2410.02073*, 2024. 1

[2] Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, and Konrad Schindler. Repurposing diffusion-based image generators for monocular depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9492–9502, 2024. 1

[3] Jia Li, Lu Wang, Lei Zhang, and Beibei Wang. Tensosdf: Roughness-aware tensorial representation for robust geometry and material reconstruction. *ACM Transactions on Graphics (TOG)*, 43(4):1–13, 2024. 2

[4] Yuan Liu, Peng Wang, Cheng Lin, Xiaoxiao Long, Jiepeng Wang, Lingjie Liu, Taku Komura, and Wenping Wang. Nero: Neural geometry and brdf reconstruction of reflective objects from multiview images. *ACM Transactions on Graphics (TOG)*, 42(4):1–22, 2023. 2

[5] Gonzalo Martin Garcia, Karim Abou Zeid, Christian Schmidt, Daan de Geus, Alexander Hermans, and Bastian Leibe. Fine-tuning image-conditional diffusion models is easier than you think. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2025. 1

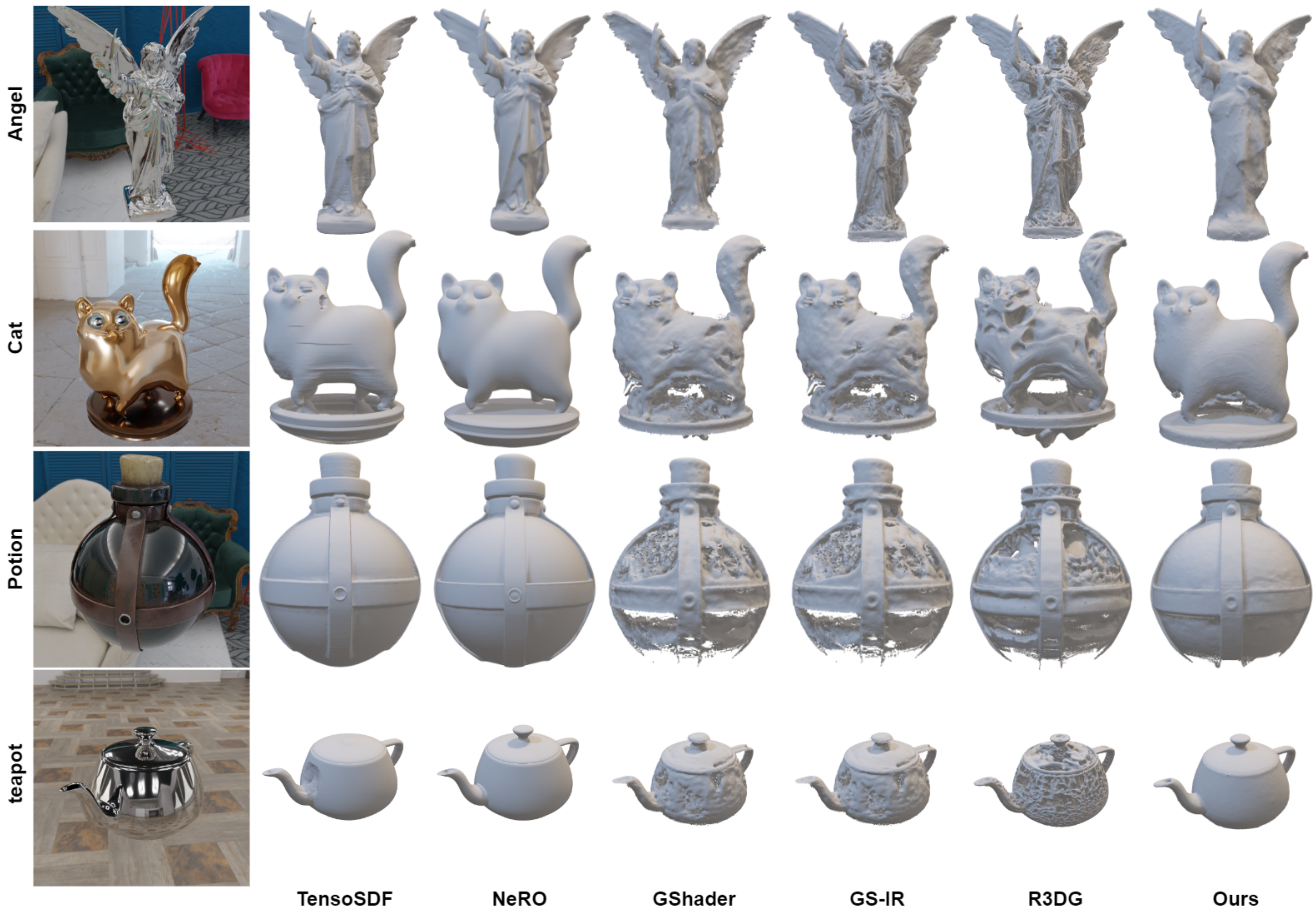

Figure S3. Additional reconstruction results on the Glossy Blender dataset.

Methods	GShader	GS-IR	R3DG	Ours
Baking	23.86 / 0.9586	23.99 / 0.9611	24.71 / 0.9638	24.16 / 0.9593
Ball	22.04 / 0.9219	23.48 / 0.9208	23.19 / 0.9156	25.17 / 0.9357
Blocks	27.09 / 0.9716	28.78 / 0.9707	27.94 / 0.9692	32.16 / 0.9812
Cactus	26.99 / 0.9681	31.94 / 0.9724	30.15 / 0.9650	33.79 / 0.9819
Car	25.13 / 0.9671	26.71 / 0.9670	26.71 / 0.9662	30.57 / 0.9802
Chips	29.94 / 0.9742	28.19 / 0.9673	28.29 / 0.9694	28.36 / 0.9696
Cup	25.90 / 0.9641	26.96 / 0.9592	25.59 / 0.9482	28.93 / 0.9709
Curry	27.37 / 0.9704	30.58 / 0.9666	29.76 / 0.9674	31.77 / 0.9702
Gnome	28.88 / 0.9502	27.92 / 0.9388	26.94 / 0.9254	31.04 / 0.9540
Grogu	25.05 / 0.9709	27.17 / 0.9684	25.17 / 0.9606	25.80 / 0.9737
Pepsi	22.24 / 0.9517	24.46 / 0.9533	22.82 / 0.9490	24.10 / 0.9608
Pitcher	25.55 / 0.9525	27.43 / 0.9545	29.00 / 0.9530	29.53 / 0.9659
Salt	25.64 / 0.9616	23.61 / 0.9289	24.96 / 0.9524	24.74 / 0.9446
Teapot	24.58 / 0.9655	23.86 / 0.9588	23.99 / 0.9575	25.86 / 0.9708
Average	25.73 / 0.9606	26.79 / 0.9563	26.37 / 0.9545	28.28 / 0.9656

Table S2. The relighting quality in terms of PSNR↑ and SSIM↓ in the StanfordORB dataset, we report the average metrics of two evaluation scenes. The comparison shows our method achieves the highest rendering quality.

Figure S4. Rendered normal maps results on the Glossy Blender dataset.

Figure S5. PBR material parameters decomposition on the Glossy Blender dataset.

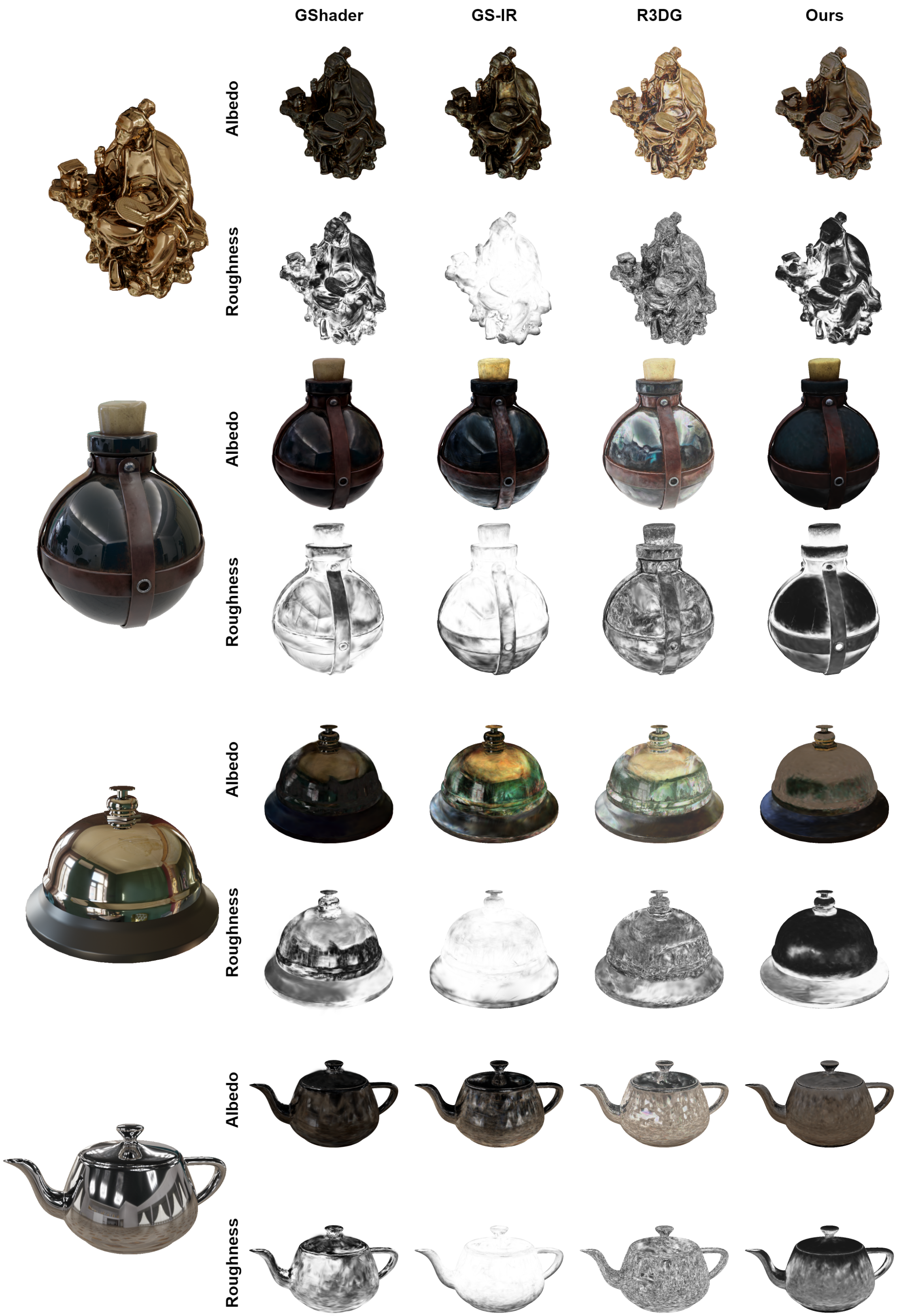

Figure S6. PBR material parameters decomposition on the Glossy Blender dataset.

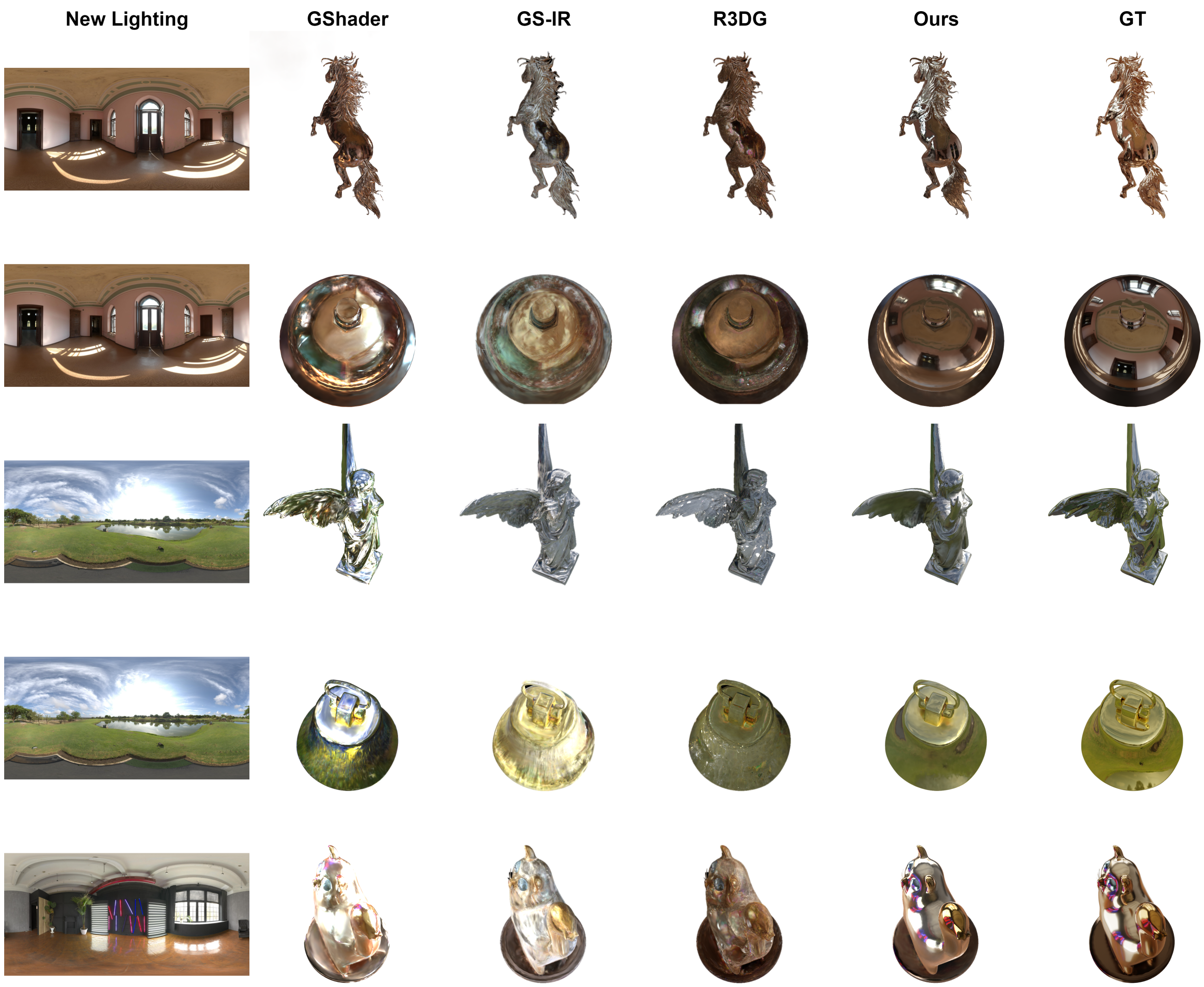

Figure S7. Relighting results on the Glossy Blender dataset.

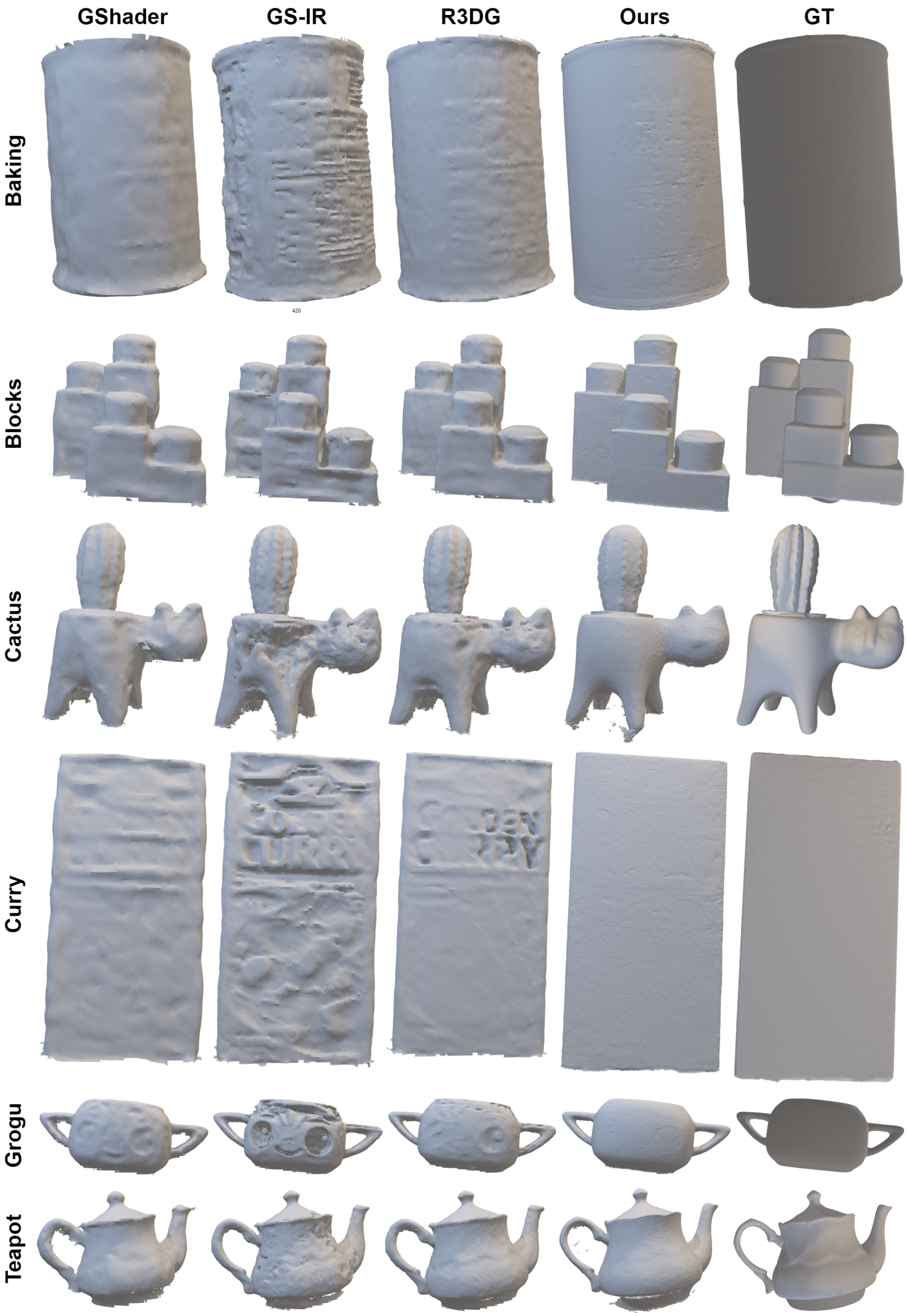

Figure S8. Additional reconstruction results on the StanfordORB dataset.

Figure S9. PBR material decomposition on the StanfordORB dataset.

Figure S10. Relighting results on the StanfordORB dataset.