SuperMat: Physically Consistent PBR Material Estimation at Interactive Rates

Supplementary Material

1. Physically Based Rendering (PBR) Model

We use the Cook-Torrance Bidirectional Reflectance Distribution Function (BRDF) [3] based on microfacet theory to define the materials and establish the rendering model. Following [1], for a point with coordinates $p \in \mathbb{R}^3$, albedo $a \in \mathbb{R}^3$, metallic $m \in \mathbb{R}$, roughness $r \in \mathbb{R}$, and surface orientation $n \in \mathbb{R}^3$, the PBR result L observed from viewpoint $c \in \mathbb{R}^3$ is given by:

$$L(p,\omega) = a(1-m)\int_\Omega L_i(p,\omega_i)(\omega_i \cdot n)\mathrm{d}\omega_i + \int_\Omega \frac{DFG}{4(\omega \cdot n)(\omega_i, n)} L_i(p,\omega_i)(\omega_i \cdot n)\mathrm{d}\omega_i, \tag{1}$$

where ω represents the direction of the outgoing light from point p to c, *i.e.*, the viewing direction. L_i denotes the incident light from the direction ω_i, and $\Omega = \{\omega_i : \omega_i \cdot n \geq 0\}$ represents the hemisphere of normals. D, F, and G are the distribution, Fresnel, and geometry functions, respectively. For the integral part, the time complexity of Monte Carlo methods is unacceptable. Given that we use ambient lighting as the light source, we compute it efficiently using the split-sum method [9].

2. Additional Details

Implementation Details. We report some implementation details as follows: 1) In the UV refinement one-step model, the input channels are expanded to 8, with the weights of the additional 4 channels initialized to 0. 2) In the ablation experiment, the structure of SuperMat under the "w/o e2e" setting is slightly different. Without single-step inference, we can only train the diffusion model in a denoising task manner, meaning that during the single-step denoising process, the latents that the UNet receives are not encoded from a clean rendered image, but rather a noisy albedo and noisy RM. Therefore, on top of SuperMat, we additionally replicate the *conv_in* and the first *DownBlock* as independent parts of structural expert branches to map the inputs from two different domains to similar distributions. These are then fused in an averaged manner before being fed into the shared modules. 3) In the re-render loss implementation, each time we perform relighting, we randomly select a lighting condition from a set of 50 environment maps, covering nearly all possible lighting scenarios.

Training Details. Both the SuperMat and the UV refinement model are fine-tuned from Stable Diffusion 2.1, while SuperMatMV is built upon SuperMat. We train SuperMat using the AdamW optimizer with a learning rate of $2e-5$ on 8 NVIDIA A800 (80GB) GPUs, with a batch size of 32, for a total of 30 epochs. The UV refinement one-step model is trained with the AdamW optimizer at a learning rate of $2e-5$, also on 8 NVIDIA A800 (80GB) GPUs, with a batch size of 16, for 40 epochs. The images are resized to resolutions of 512×512 and 1024×1024, used for SuperMat and the UV refinement one-step model, respectively. The training setup for SuperMatMV mirrors that of SuperMat, except that SuperMatMV is trained for only 3 epochs using a batch size of 8 for 6-view images.

Image Space Decomposition Baselines. We compare SuperMat and SuperMatMV with 7 other image space decomposition networks. Inverse Indoor Rendering (IIR) [13], Intrinsic Image Diffusion (IID) [6], and RGB→X [12] are scene-level material estimation methods. Derender3D [11] and IntrinsicAnything [2] are adaptable to diverse data but do not generate all material types. StableMaterial [8] and its multi-view version, StableMaterialMV, provide image space material denoising diffusion priors to MaterialFusion and, like SuperMat, focus on decomposing object materials. It is worth noting that, for scene-targeted methods, to ensure a fairer comparison, we re-train a deterministic method, IIR, and a diffusion-based method, RGB→ X, on our dataset.

3. Additional Visualizations

In Figures 2, 3, 4, 5, 6, 7, we present additional results of SuperMat on the Objaverse [4], BlenderVault [8] and DTC [5] test datasets, covering both artist-designed and real-world scanned objects. Fig. 8 illustrates more decomposition results on 3D objects.

4. Additional Experiments

Novel View Synthesis Relighting. We validate our material decomposition pipeline for 3D objects by relighting the decomposed objects under novel lighting and viewpoints and comparing the results with ground-truth relit object images. We select all 14 objects from the StanfordORB dataset [7] and 14 randomly chosen unseen objects from the BlendVault dataset [8], applying appropriate lighting conditions before performing material decomposition. We adapt baseline methods to incorporate the same UV backprojection and blending framework as our approach, allowing them to handle 3D object decomposition. We then render the decomposed models from 60 novel viewpoints under three different environment maps. The quantitative comparison results are presented in Tab. 4, while qualitative results are shown in Fig. 1. Our method significantly outperforms others in both texture completeness and the physical con-

	Albedo			Metallic			Roughness			Relighting		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
L1	26.8102	0.9164	0.0956	22.4801	0.8889	0.1803	23.5855	0.9109	0.1194	26.8610	0.9374	0.0672
SSI	24.2359	0.9063	0.1033	22.2521	0.8701	0.1814	23.7918	0.9139	0.1165	24.7511	0.9266	0.0732
Perceptual	27.0082	0.9151	0.0949	22.8702	0.8669	0.1760	24.1452	0.9145	0.1156	27.2484	0.9374	0.0650

Table 1. Quantitative comparison across different loss functions. We highlight the best results for each metric.

Training Data Type	Albedo				Metallic				Roughness			
	MSE↓	PSNR↑	SSIM↑	LPIPS↓	MSE↓	PSNR↑	SSIM↑	LPIPS↓	MSE↓	PSNR↑	SSIM↑	LPIPS↓
Synthetic	0.0027	27.41	0.9195	0.0885	0.0120	23.80	0.8767	0.1692	0.0084	24.32	0.9152	0.1118
Real-captured	0.0061	24.50	0.8982	0.1049	0.0350	19.40	0.8464	0.2043	0.0113	22.32	0.9075	0.1165
Synthetic+Real-captured	0.0025	27.58	0.9205	0.0867	0.0111	23.77	0.8787	0.1696	0.0074	24.63	0.9154	0.1115

Table 2. Ablation results on training data types in the image space decomposition task. We highlight the best results for each metric.

	Albedo			RM		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Input	13.56	0.5499	0.4227	12.52	0.5484	0.5245
Refined	23.92	0.7537	0.2792	27.31	0.8460	0.1178

Table 3. Quantitative evaluation on UV maps before and after refinement.

	BlenderVault			StanfordORB		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Derender3D	22.52	0.9151	0.1023	17.03	0.8241	0.2012
IIR	21.10	0.9005	0.1046	15.83	0.7944	0.2005
IID	22.10	0.9095	0.0989	16.21	0.8000	0.2059
RGB→X	22.23	0.9115	0.0987	17.23	0.8272	0.1849
IA	23.33	0.9115	0.1047	17.98	0.8286	0.1875
SM	23.36	0.9176	0.0909	16.94	0.8174	0.1801
SMMV	23.57	0.9175	0.0887	17.10	0.8189	0.1785
Ours	27.00	0.9463	0.0629	24.99	0.9281	0.0962

Table 4. Quantitative comparison of novel view synthesis relighting. We highlight the best, second-best, and third-best results for each metric. Here "IA", "SM", "SMMV" represents "IntrinsicAnything", "StableMaterial", "StableMaterialMV" respectively.

sistency of materials, demonstrating its effectiveness in decomposing high-quality materials for 3D objects.

Comparison between Different Loss Functions. Without re-render loss, we experiment with three types of loss functions in the end-to-end framework: L1 loss, shift and scale invariant (SSI) loss [10], and perceptual loss. The performance of models trained with these loss designs is evaluated on the same test dataset in Objaverse [4], and the results are shown in Tab. 1. Among these losses, the perceptual loss, which is ultimately adopted by SuperMat, demonstrates the best performance.

Quantitative Validation of UV Refinement. Tab. 3 presents the quantitative evaluation results of the UV maps generated and blended by SuperMatMV before and after refinement, validating the effectiveness of the UV refinement one-step model.

Ablation on the Training Dataset. For the image space decomposition task, we conduct an ablation study on the composition of the training dataset, testing the SuperMat's performance when trained solely on synthetic data or only on real-captured data. We randomly select 64 objects containing both data types from the training set, totaling 6144 inputs, which are unseen during training but used to evaluate the model's estimations of 3 material types with MSE, PSNR, SSIM, and LPIPS metrics. The results of the experiment are shown in Tab.5, indicating that data diversity helps improve the model's generalization ability.

Detailed Quantitative Comparison. We show the detailed quantitative comparison of the image space decomposition task in Tables 5, 6, 7 and 8.

5. Limitations

Although our method enables fast and physically consistent material decomposition for both images and 3D objects, it still has several limitations. First, converting the diffusion model into a deterministic model significantly improves performance but also sacrifices the benefits of the diffusion process. For instance, in cases with high-frequency details, SuperMat's results may lack sharpness. Additionally, as a deterministic model, SuperMat has reduced flexibility—once an estimation error occurs, rerunning the process will always yield the same incorrect result. Furthermore, due to the constraints of the chosen BRDF model, SuperMat faces challenges in handling transparent materials, multi-layered surfaces, and highly reflective objects.

References

[1] Rui Chen, Yongwei Chen, Ningxin Jiao, and Kui Jia. Fantasia3d: Disentangling geometry and appearance for high-quality text-to-3d content creation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. 1

[2] Xi Chen, Sida Peng, Dongchen Yang, Yuan Liu, Bowen Pan, Chengfei Lv, and Xiaowei Zhou. Intrinsicanything: Learning diffusion priors for inverse rendering under unknown illumination. *arXiv preprint arXiv:2404.11593*, 2024. 1

[3] Robert L Cook and Kenneth E. Torrance. A reflectance model for computer graphics. *ACM Transactions on Graphics (ToG)*, 1(1):7–24, 1982. 1

[4] Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13142–13153, 2023. 1, 2

[5] James Fort. Introducing the digital twin catalog from reality labs research, 2024. https://ai.meta.com/blog/digital-twin-catalog-3d-reconstruction-shopify-reality-labs-research/, Last accessed on 2025-03-07. 1

[6] Peter Kocsis, Vincent Sitzmann, and Matthias Nießner. Intrinsic image diffusion for indoor single-view material estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5198–5208, 2024. 1

[7] Zhengfei Kuang, Yunzhi Zhang, Hong-Xing Yu, Samir Agarwala, Elliott Wu, Jiajun Wu, et al. Stanford-orb: a real-world 3d object inverse rendering benchmark. 2023. 1

[8] Yehonathan Litman, Or Patashnik, Kangle Deng, Aviral Agrawal, Rushikesh Zawar, Fernando De la Torre, and Shubham Tulsiani. Materialfusion: Enhancing inverse rendering with material diffusion priors. *arXiv preprint arXiv:2409.15273*, 2024. 1

[9] Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Müller, and Sanja Fidler. Extracting triangular 3d models, materials, and lighting from images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8280–8290, 2022. 1

[10] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE transactions on pattern analysis and machine intelligence*, 44(3):1623–1637, 2020. 2

[11] Felix Wimbauer, Shangzhe Wu, and Christian Rupprecht. De-rendering 3d objects in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18490–18499, 2022. 1

[12] Zheng Zeng, Valentin Deschaintre, Iliyan Georgiev, Yannick Hold-Geoffroy, Yiwei Hu, Fujun Luan, Ling-Qi Yan, and Miloš Hašan. Rgb $\leftrightarrow$ x: Image decomposition and synthesis using material-and lighting-aware diffusion models. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–11, 2024. 1

[13] Jingsen Zhu, Fujun Luan, Yuchi Huo, Zihao Lin, Zhihua Zhong, Dianbing Xi, Rui Wang, Hujun Bao, Jiaxiang Zheng, and Rui Tang. Learning-based inverse rendering of complex indoor scenes with differentiable monte carlo raytracing. In *SIGGRAPH Asia 2022 Conference Papers*, pages 1–8, 2022. 1

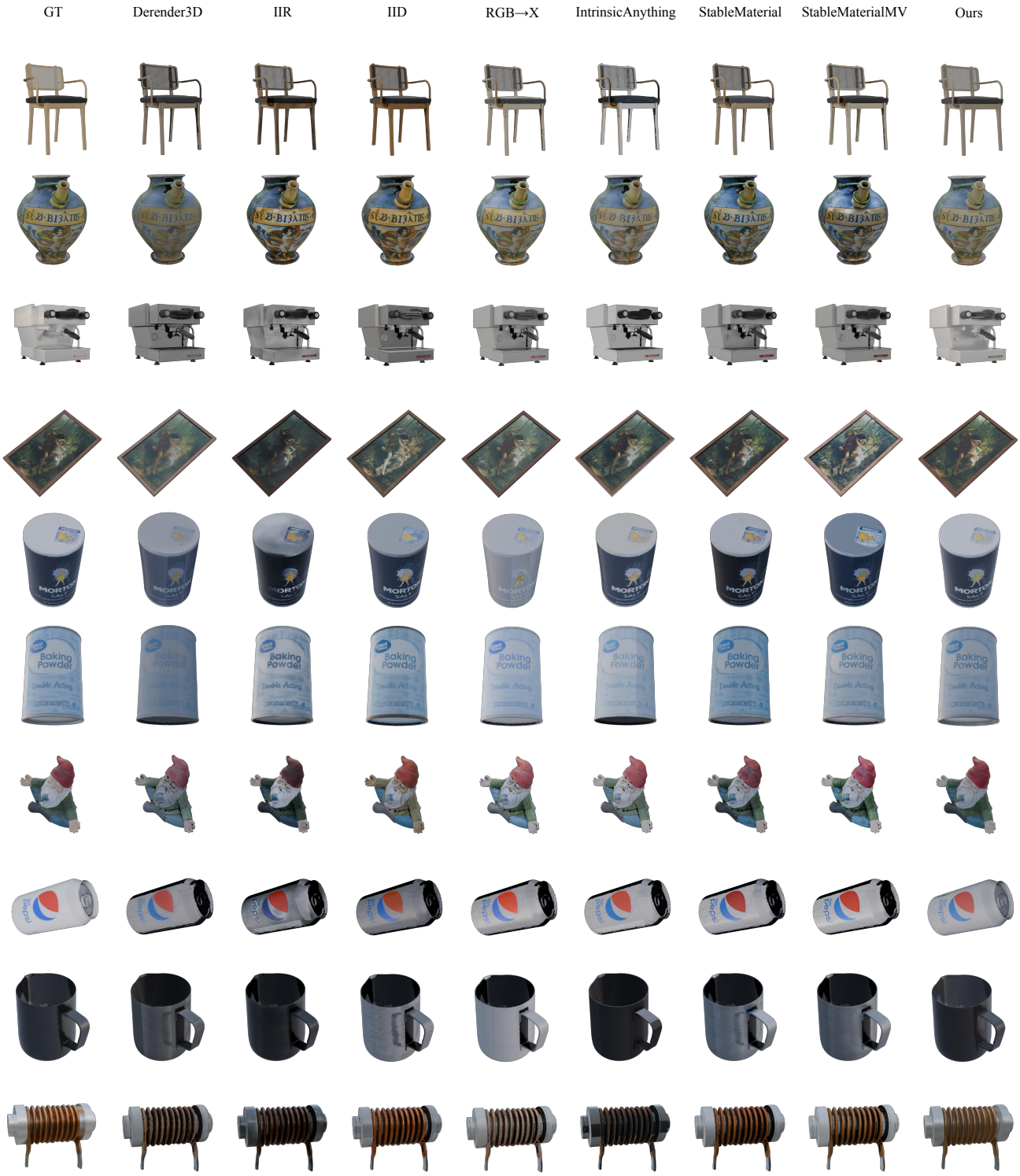

Figure 1. Qualitative comparison of novel view synthesis relighting.

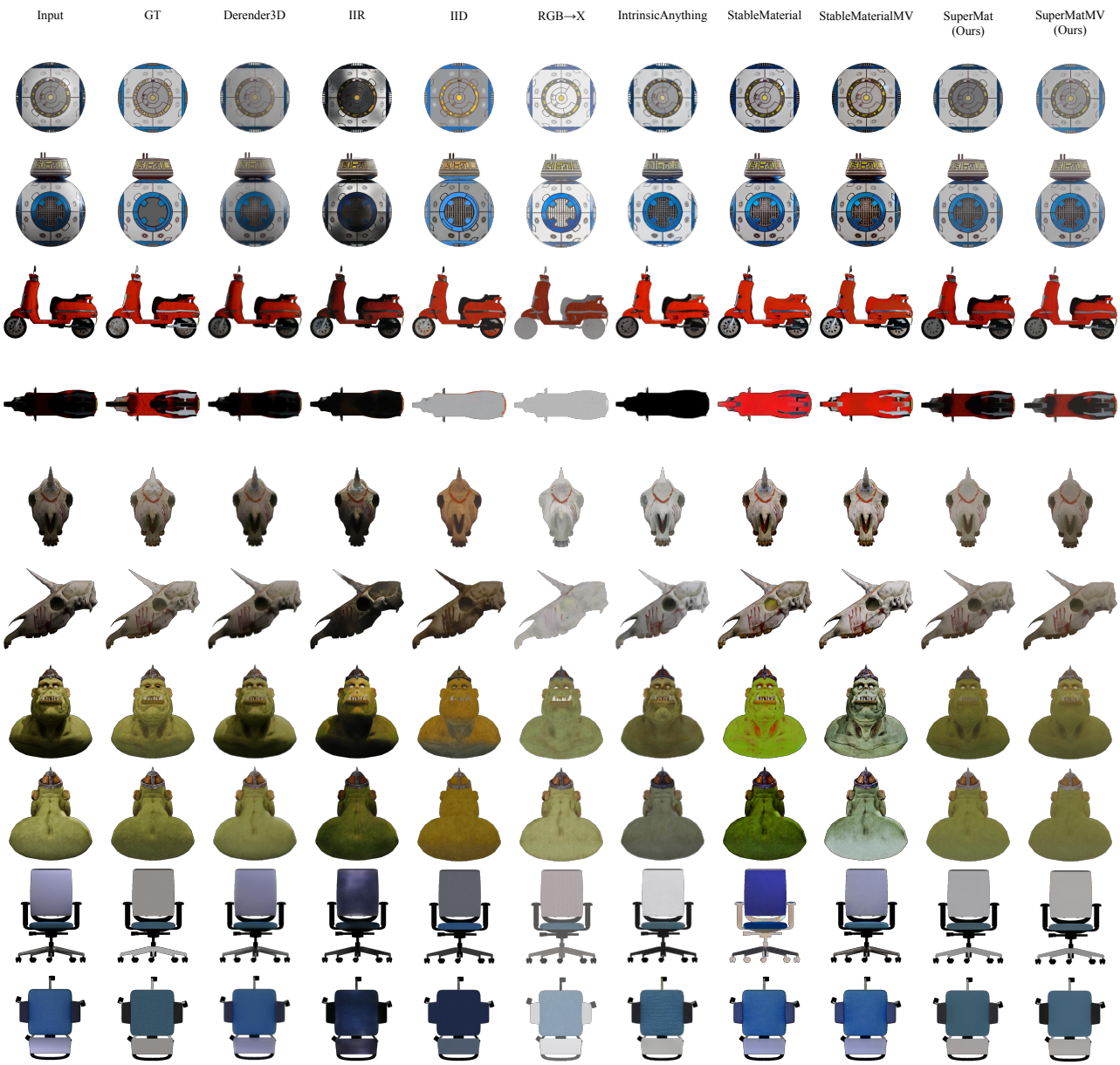

Figure 2. Additional albedo comparison from the Objaverse test dataset.

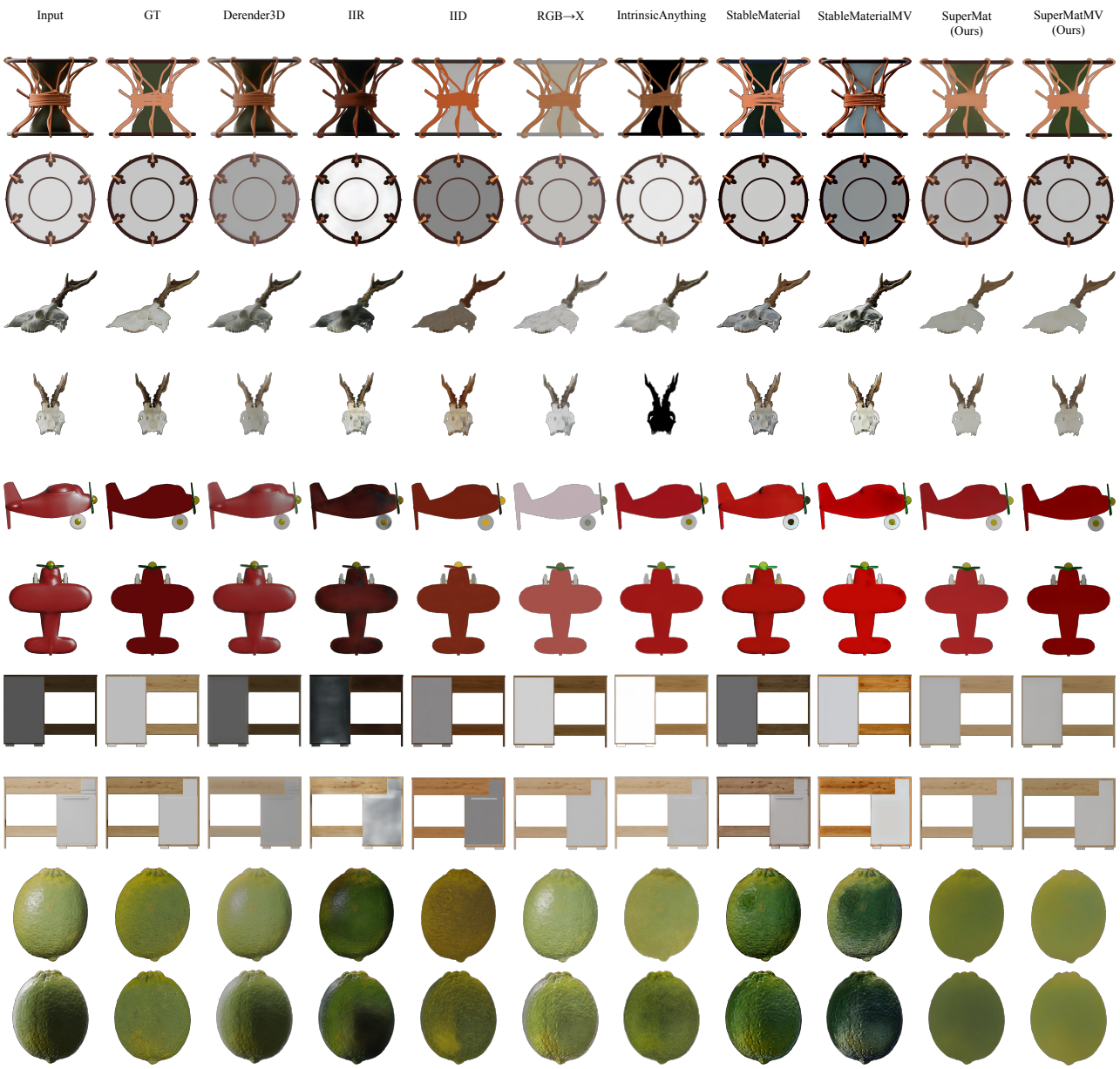

Figure 3. Additional albedo comparison from the BlenderVault test dataset.

Figure 4. Additional albedo comparison from the DTC test dataset.

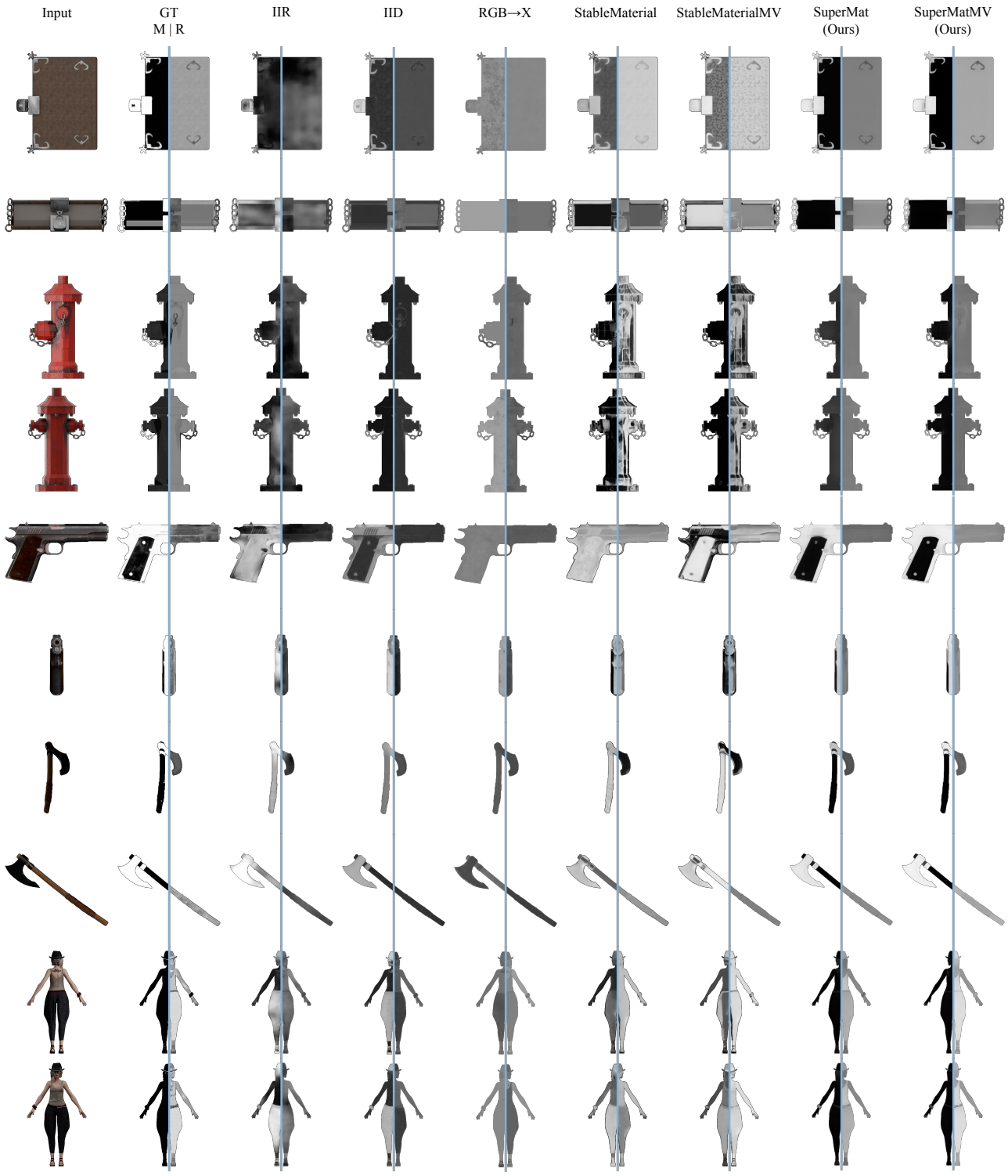

Figure 5. Additional metallic and roughness comparison from the Objaverse test dataset. The metallic maps are shown on the left side (M), while the roughness maps are shown on the right side (R).

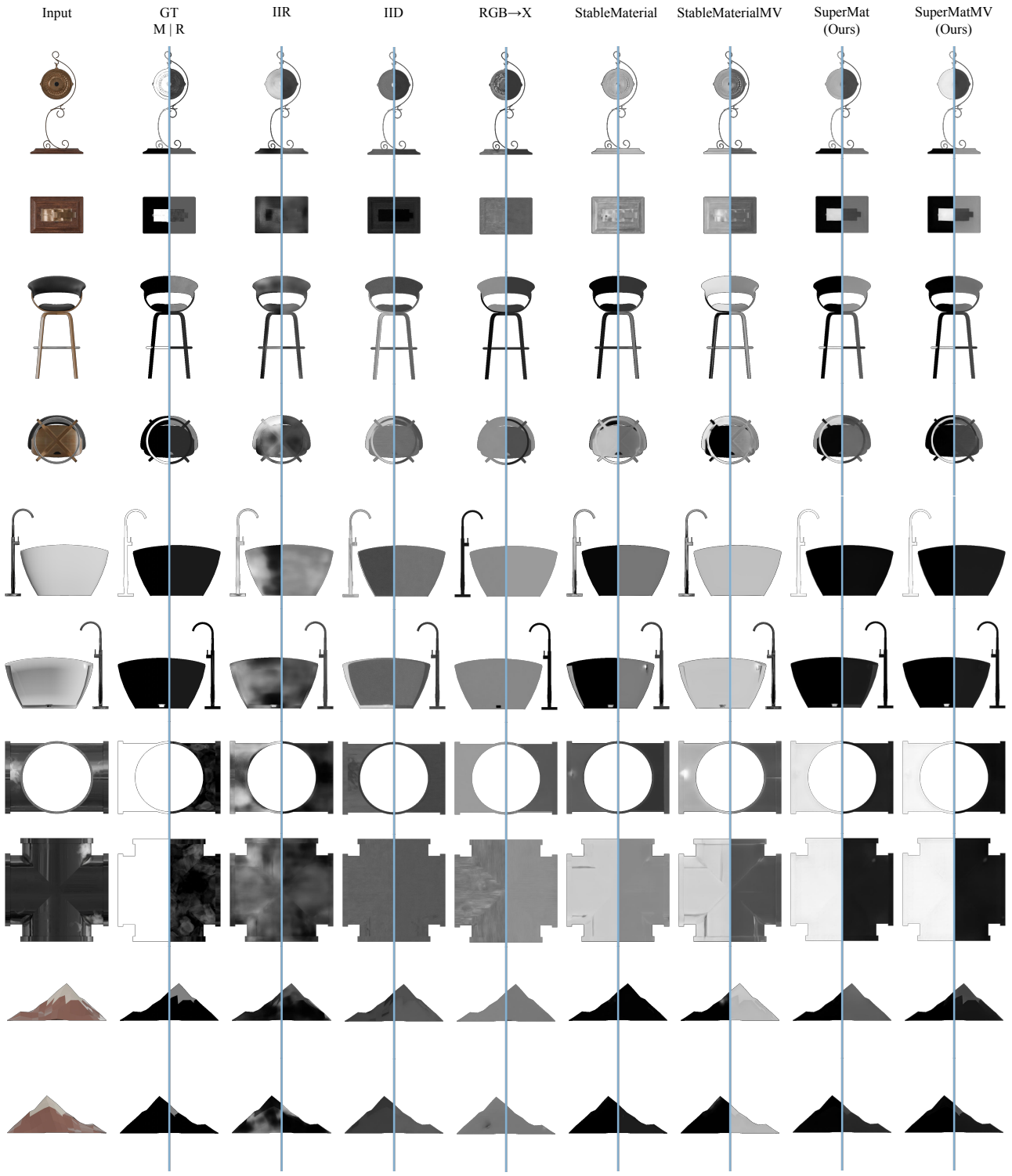

Figure 6. Additional metallic and roughness comparison from the BlenderVault test dataset. The metallic maps are shown on the left side (M), while the roughness maps are shown on the right side (R).

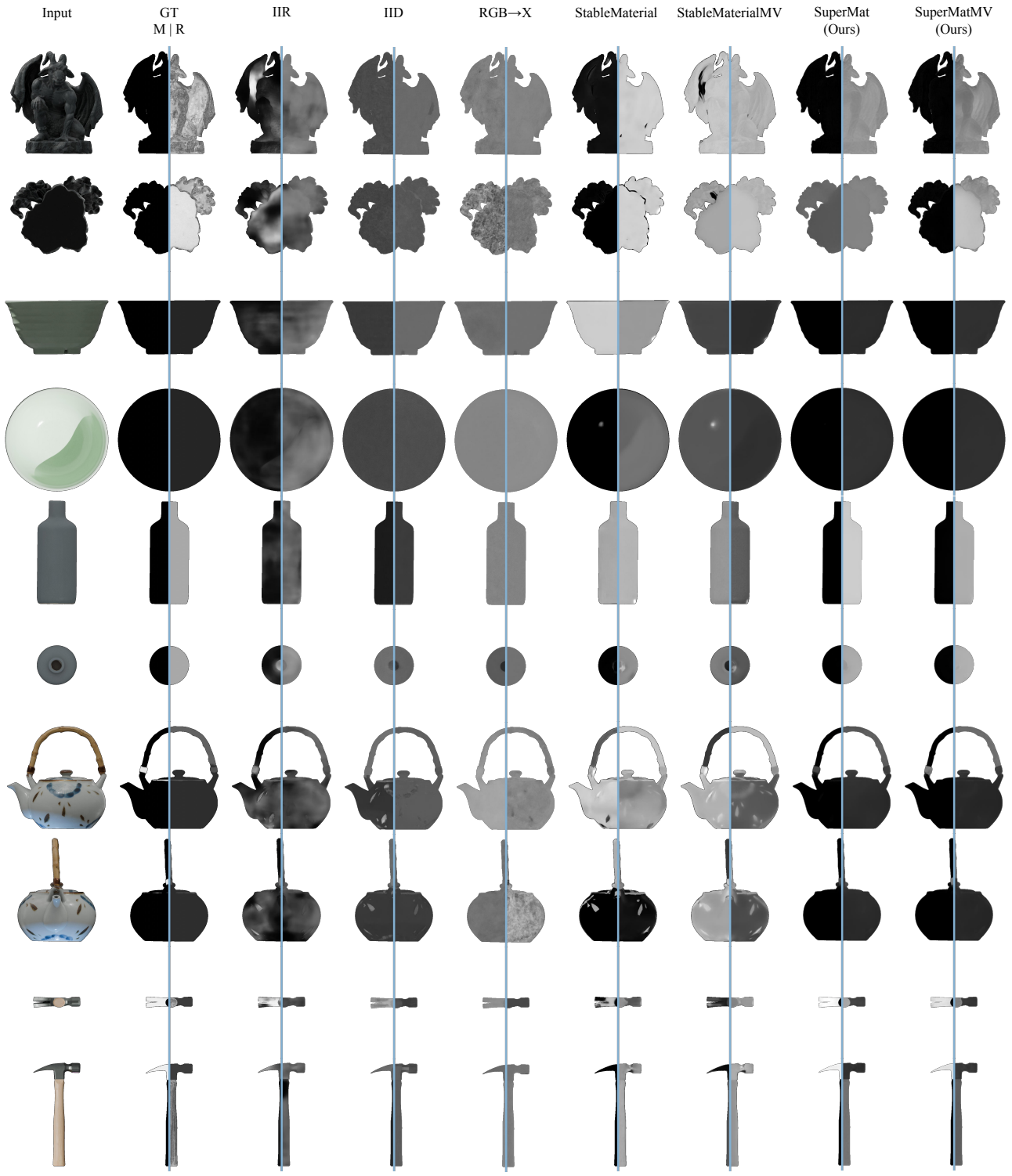

Figure 7. Additional metallic and roughness comparison from the DTC test dataset. The metallic maps are shown on the left side (M), while the roughness maps are shown on the right side (R).

Figure 8. Additional results of decomposition for 3D objects.

	Objaverse				BlenderVault				DTC			
	MSE↓	PSNR↑	SSIM↑	LPIPS↓	MSE↓	PSNR↑	SSIM↑	LPIPS↓	MSE↓	PSNR↑	SSIM↑	LPIPS↓
Derender3D	0.0087	22.16	0.8532	0.1816	0.0104	22.45	0.8919	0.1489	0.0246	18.30	0.8366	0.2157
IIR	0.0100	22.99	0.8741	0.1396	0.0150	22.32	0.8709	0.1616	0.0165	20.50	0.8677	0.1644
IID	0.0088	23.15	0.8847	0.1276	0.0122	22.39	0.8845	0.1395	0.0101	22.15	0.9031	0.1175
RGB→X	0.0073	23.56	0.8905	0.1017	0.0141	21.73	0.8757	0.1332	0.0129	21.60	0.8917	0.1114
IntrinsicAnything	0.0125	21.36	0.8701	0.1550	0.0219	20.24	0.8700	0.1682	0.0122	21.23	0.8963	0.1328
StableMaterial	0.0110	23.78	0.8989	0.1064	0.0143	23.90	0.9061	0.1079	0.0129	22.64	0.9008	0.1086
StableMaterialMV	0.0074	24.41	0.8974	0.0970	0.0089	25.13	0.9071	0.1014	0.0067	24.15	0.9229	0.0855
SuperMat w/o e2e	0.0057	25.11	0.8973	0.1027	0.0102	24.30	0.8935	0.1219	0.0107	23.36	0.8985	0.1038
SuperMat w/o re-render	0.0029	27.01	0.9151	0.0949	0.0047	26.12	0.9152	0.1045	0.0031	26.96	0.9405	0.0725
SuperMat	0.0024	27.66	0.9209	0.0865	0.0044	26.63	0.9171	0.0997	0.0018	28.74	0.9490	0.0597
SuperMatMV	0.0022	28.04	0.9241	0.0832	0.0039	26.75	0.9183	0.1026	0.0023	27.89	0.9443	0.0658

Table 5. Quantitative comparison on albedo. We highlight the best, second-best, and third-best results for each metric.

	Objaverse				BlenderVault				DTC			
	MSE↓	PSNR↑	SSIM↑	LPIPS↓	MSE↓	PSNR↑	SSIM↑	LPIPS↓	MSE↓	PSNR↑	SSIM↑	LPIPS↓
IIR	0.0462	16.87	0.8399	0.2230	0.0472	17.11	0.7688	0.2900	0.0324	19.86	0.6771	0.3604
IID	0.0354	17.31	0.8307	0.2108	0.0431	17.43	0.7742	0.2866	0.0347	17.41	0.6797	0.3895
RGB→X	0.0350	16.81	0.8296	0.2090	0.0549	15.36	0.7634	0.3374	0.0682	13.90	0.6738	0.4617
StableMaterial	0.0493	16.83	0.8398	0.2164	0.0515	20.79	0.8156	0.2379	0.0567	23.26	0.7938	0.2833
StableMaterialMV	0.0452	16.96	0.8411	0.2114	0.0511	18.74	0.8150	0.2590	0.0777	16.15	0.7096	0.3783
SuperMat w/o e2e	0.0373	19.40	0.8672	0.2021	0.0470	20.91	0.8455	0.2764	0.0496	22.05	0.8204	0.3694
SuperMat w/o re-render	0.0135	22.87	0.8669	0.1760	0.0311	22.52	0.8040	0.2343	0.0087	28.23	0.8067	0.2687
SuperMat	0.0109	23.78	0.8785	0.1695	0.0344	23.02	0.8335	0.2328	0.0068	29.65	0.8936	0.2641
SuperMatMV	0.0069	25.38	0.8919	0.1619	0.0324	23.17	0.8437	0.2311	0.0059	29.78	0.8892	0.2656

Table 6. Quantitative comparison on metallic. We highlight the best, second-best, and third-best results for each metric.

	Objaverse				BlenderVault				DTC			
	MSE↓	PSNR↑	SSIM↑	LPIPS↓	MSE↓	PSNR↑	SSIM↑	LPIPS↓	MSE↓	PSNR↑	SSIM↑	LPIPS↓
IIR	0.0257	20.67	0.8902	0.1242	0.0438	18.58	0.8503	0.1754	0.0320	19.93	0.8512	0.1793
IID	0.0204	21.32	0.8863	0.1266	0.0295	19.97	0.8718	0.1586	0.0168	21.51	0.8695	0.1576
RGB→X	0.0159	20.93	0.8705	0.1177	0.0257	19.78	0.8486	0.1780	0.0179	20.48	0.8229	0.1817
StableMaterial	0.0131	21.97	0.9065	0.1064	0.0251	20.94	0.8962	0.1312	0.0256	20.13	0.8612	0.1604
StableMaterialMV	0.0142	21.26	0.8949	0.1081	0.0253	20.49	0.8915	0.1284	0.0233	20.83	0.8651	0.1493
SuperMat w/o e2e	0.0170	21.36	0.8832	0.1221	0.0278	20.48	0.8679	0.1633	0.0246	20.60	0.8352	0.1907
SuperMat w/o re-render	0.0081	24.15	0.9145	0.1156	0.0211	21.73	0.8942	0.1429	0.0086	24.69	0.8961	0.1502
SuperMat	0.0074	24.59	0.9154	0.1114	0.0201	22.39	0.8972	0.1386	0.0053	25.78	0.9046	0.1342
SuperMatMV	0.0074	24.96	0.9165	0.1070	0.0145	23.22	0.8998	0.1297	0.0049	26.34	0.9064	0.1257

Table 7. Quantitative comparison on roughness. We highlight the best, second-best, and third-best results for each metric.

	Objaverse				BlenderVault				DTC			
	MSE↓	PSNR↑	SSIM↑	LPIPS↓	MSE↓	PSNR↑	SSIM↑	LPIPS↓	MSE↓	PSNR↑	SSIM↑	LPIPS↓
IIR	0.0132	23.19	0.8991	0.1060	0.0290	20.96	0.8765	0.1441	0.0245	18.80	0.8897	0.0890
IID	0.0105	23.83	0.9057	0.0971	0.0160	22.31	0.9009	0.1142	0.0149	21.21	0.9066	0.1137
RGB→X	0.0107	22.86	0.8974	0.0864	0.0177	21.43	0.8828	0.1205	0.0207	20.23	0.8780	0.1327
StableMaterial	0.0107	24.08	0.9108	0.0851	0.0203	22.22	0.9022	0.1013	0.0183	21.38	0.8999	0.1074
StableMaterialMV	0.0091	24.02	0.9114	0.0816	0.0154	22.69	0.9019	0.1018	0.0134	22.70	0.9137	0.0966
SuperMat w/o e2e	0.0079	25.81	0.9192	0.0776	0.0171	23.15	0.9031	0.1059	0.0158	22.75	0.9032	0.1057
SuperMat w/o re-render	0.0042	27.25	0.9374	0.0650	0.0109	24.94	0.9270	0.0855	0.0050	27.05	0.9490	0.0600
SuperMat	0.0041	28.01	0.9406	0.0566	0.0101	25.50	0.9300	0.0811	0.0031	29.47	0.9590	0.0460
SuperMatMV	0.0032	28.51	0.9437	0.0566	0.0118	25.41	0.9289	0.0815	0.0041	29.01	0.9567	0.0502

Table 8. Quantitative comparison on relighting. We highlight the best, second-best, and third-best results for each metric.