

MatSpray: Fusing 2D Material World Knowledge on 3D Geometry

Supplementary Material

Overview

This supplementary material provides extended results, analyses, and implementation details that complement the findings in the main paper. For ease of navigation, the main components are summarized here and referenced through the corresponding section labels.

- **Additional Videos and Real-World Objects (A):** A detailed collection of video comparisons and reconstructions of real objects that highlight the performance and stability of our method relative to earlier approaches.
- **Neural Merger Ablation (B):** An extended analysis of the importance of the final `Softmax` layer in the Neural Merger, supported by qualitative and quantitative evidence.
- **Tone Mapping Analysis (C):** A discussion of the tone mapping behaviour of DiffusionRenderer, how this affects predicted base color, roughness and metallic maps, and why this creates a mismatch when compared to linear ground truth.
- **Implementation Details (D):** A description of our training setup, super sampling strategy, Neural Merger inputs and other practical considerations that are important for stable optimization.

A. Additional Videos and Real-World Objects

Figure 1 shows the thumbnail that links to all additional videos included with this supplementary material. These videos provide an extensive visual comparison of our method with Extended R3DGS [?], IRGS [?], and the forward renderer of DiffusionRenderer [?]. While the main paper presents representative examples, the extended videos

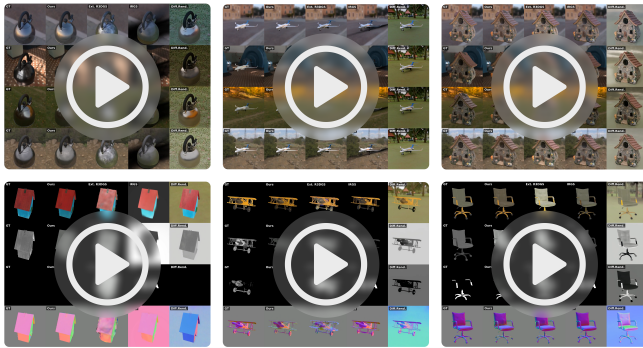


Figure 1. Thumbnail showing six videos that can be viewed [here](#). Three videos show relighting comparison and three show material prediction comparisons.

give a more complete picture of the consistency and stability of our approach, especially compared to the produced material maps of DiffusionRenderer.

Across the set of videos, our method consistently produces reconstructions that remain stable across all viewpoints, without the flickering or structural collapse that can be observed in the other methods. This is particularly visible in objects with complex geometry or pronounced specular highlights such as the Kettle. Extended R3DGS often fails to maintain surface smoothness and yields unstable representations. On the other hand IRGS tends to over-smooth surfaces and tends to bake in specular reflections of metallic objects into its base color. In contrast, our approach maintains coherent structure even under strong lighting variations.

To illustrate this, we provide three relighting videos: [White Golden Airplane](#), [Stone Birdhouse](#), and [Kettle](#). Additionally, three videos visualize predicted material properties: [Yellow Airplane](#), [Birdhouse with Yellow Flower](#), and [Chair](#). These examples show that DiffusionRenderer, despite being trained on its own dataset, still produces inconsistent material maps that vary strongly with camera angle and lighting. Our method mitigates these issues and aligns predictions across views more reliably.

Real-World Objects Figure 2 shows additional real objects reconstructed by our method and by Extended R3DGS and IRGS. Here, each method is evaluated under two relighting settings and compared in base color and normals. The differences are most obvious in the base color: our base color is locally sharp and coherent across the surface, while both baselines exhibit noise, distortions, or view-dependent artifacts. The relighting results further demonstrate that our predicted materials generalize well across lighting conditions, while the other methods still have lighting effects baked into their materials (R3DGS) or tend to be washed out (IRGS).

B. Neural Merger Ablation

The Neural Merger plays a key role in ensuring that the material parameters assigned to each Gaussian remain stable and consistent across all viewpoints. One central element of the Neural Merger is the final `Softmax` layer, which normalizes its output into weights acting as a weighted average of the inputs. Although this layer may ap-

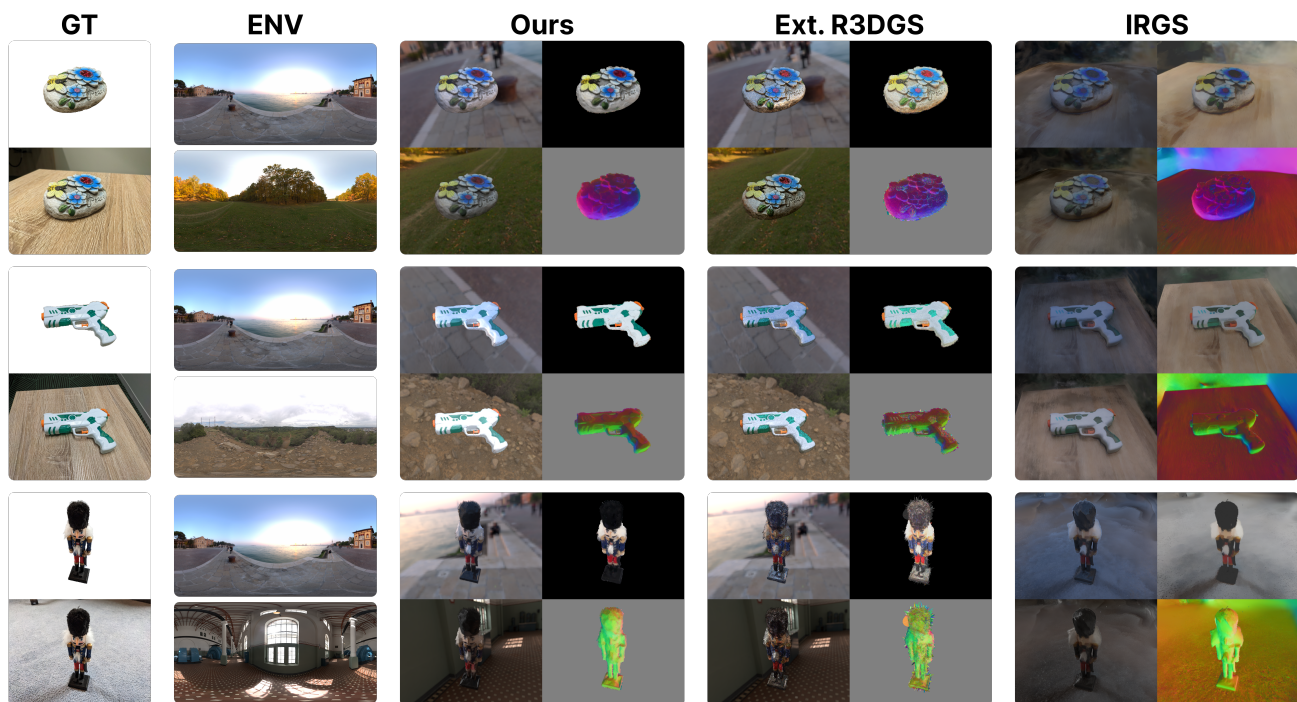


Figure 2. Additional real objects reconstructed with our method, Extended R3DGS [?] and IRGS [?]. The figure includes relighting under two environments, base color and normal maps.

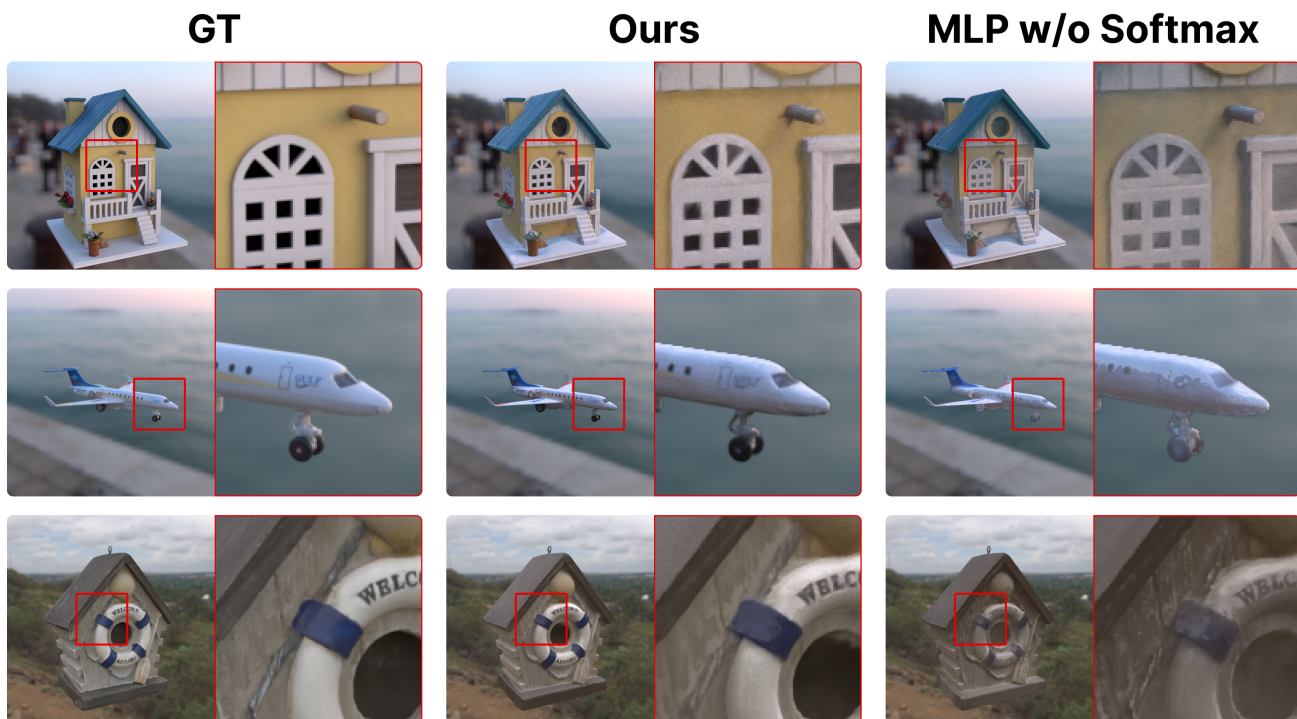


Figure 3. The impact of the `Softmax` layer in the Neural Merger. Without it, lighting and shadow patterns leak into the material maps, leading to inconsistent relighting.

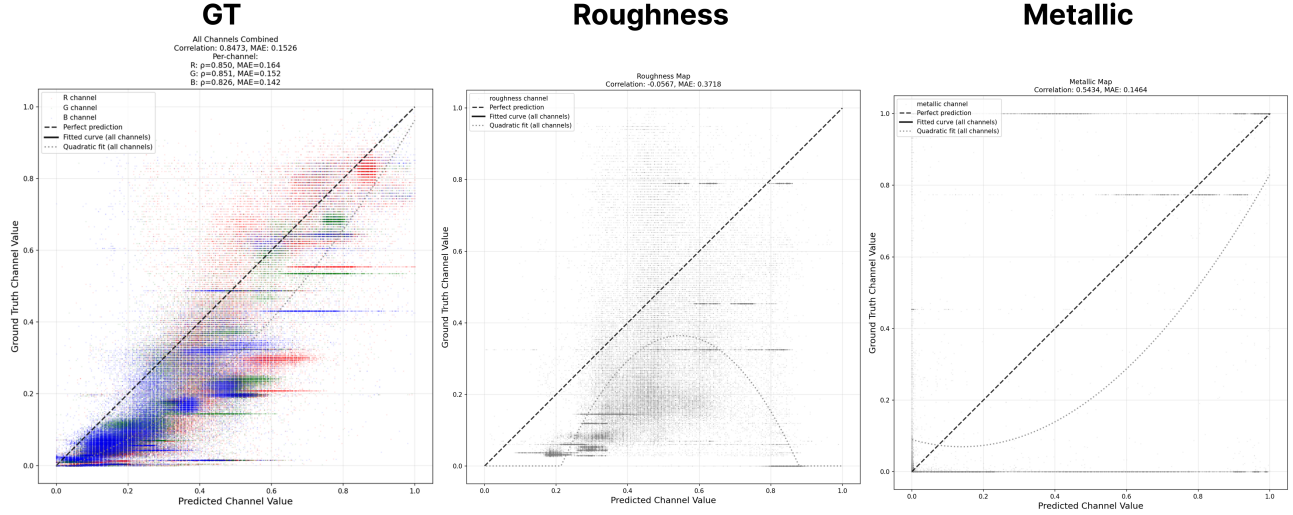


Figure 4. Tone mapping applied by DiffusionRenderer significantly alters the appearance of material maps. The alpha mask removes background content and focuses on the region of interest.

pear to be a small architectural detail, it has a sizable impact on the quality of the final reconstruction.

Without the `Softmax` normalization, the Neural Merger becomes unconstrained and starts to absorb illumination cues directly from the training images. In other words, instead of learning clean, view-independent materials, the MLP blends in signals that correspond to lighting variations and shadows. Because these patterns differ between view-points, the network produces material values that vary from view to view, which leads to inconsistency during rendering. Although this might also be additionally influenced by slight variations in the 2D Diffusion predictions geometry. This behaviour becomes especially problematic under re-lighting, because the embedded shadows and highlights interfere with the simulated lighting and produce unrealistic results.

Figure 3 shows a comparison between the full method, the version without the `Softmax`, and the linear ground truth. The differences become clear when observing fine geometric structures and shadow placement. Without `Softmax`, shadows from the input images appear in the base color maps and the renderings become blurry in high detail areas. These issues are especially visible in the lower birdhouse example, where the version without `Softmax` fails to maintain consistent materials on the swim ring and the surrounding areas.

We further quantify these findings in Table 1, which reports results across all scenes in the dataset. The full model outperforms the version without the `Softmax` across all

metrics, with especially large gains in perceptual similarity (LPIPS). This confirms that the `Softmax`-based normalization is not merely a numerical improvement but a key component that ensures robustness and prevents the network from encoding view-dependent appearance into the materials.

C. DiffusionRenderer Tone Mapping Analysis

One recurring observation in our experiments was that the base color predicted by our method tended to appear darker than the linear ground truth material map. This appeared to be a miss-prediction of the 2D material maps by DiffusionRenderer for a few objects. However, this discoloration appeared in almost all objects that we tested, hinting towards a systemic problem in DiffusionRenderer. Figure 4 illustrates this systemic discoloration of the predicted base color. This indicates that during training DiffusionRenderer was supervised using tone-mapped ground truth images.

Our analysis suggests that DiffusionRenderer employs a filmic or AgX tone mapping curve. These tone-mapping algorithms compress high dynamic range values into the range expected by standard displays, which improves visual

Table 1. **Ablation Study** on all objects. Removing the `Softmax` layer causes the network to encode lighting, which degrades all metrics.

Method	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Full	27.282	0.897	0.080
Without Softmax	24.600	0.874	0.114

quality but complicates the recovery of physically meaningful material parameters. In particular, these tone mappings are not analytically invertible, and even approximate inverse curves introduce errors, especially near shadows or high-lights.

Base color is affected in a predictable way, because tone mapping acts like a softened gamma curve. Applying an inverse gamma of roughly one point eight partially recovers the linear values but cannot undo the full nonlinearity. Roughness is affected more severely, because its values occupy a small part of the zero to one interval, which collapses under tone mapping. Metallic values, on the other hand, remain closer to either zero or one and thus suffer less from compression. These effects explain why our predicted material maps sometimes differ from the linear ground truth as they closely match DiffusionRenderer’s tone-mapped output.

D. Implementation Details

Our experiments were performed on an NVIDIA RTX 4090 GPU with PyTorch, C++ and Optix. To keep the input consistent with the internal resolution of DiffusionRenderer, we render all training views at a resolution of 512×512 pixel. This choice ensures that the reconstruction quality aligns with the scale at which DiffusionRenderer was originally trained. In scenes with strong specular highlights, we disable geometry learning entirely and keep the Gaussian positions fixed, because additional geometric optimization tends to destabilize the representation under these conditions.

The Neural Merger is optimized using a learning rate of zero point zero zero one. Material supervision uses an L1-loss with a weight of 1.0, as we found that this balance

prevents the model from overfitting shadows while still enforcing high fidelity in the material maps. During training, we also apply random view sampling to avoid biasing the model toward any particular viewpoint.

Super Sampling A key technical detail is the use of super sampling during the projection of material values into the Gaussian representation. We employ a 16×16 grid of rays per pixel to ensure that even small or distant Gaussians receive material parameters. With fewer samples, Gaussians are occasionally missed leading to a patchy geometry and a low resolution material parameter transferal. Figure 5 shows an example where a lower sampling rate produces obvious reconstruction defects.

Merger Inputs Finally, Figure 6 illustrates the input to the Neural Merger. The features consist of a NeRF-style positional encoding of the Gaussian location along with the projected base color, roughness and metallic values. The combination of positional encoding and projected materials allows the network to balance local detail with global consistency, which is essential for producing clean results under relighting.

Computation Time Table 2 shows runtime breakdown on the Navi dataset [?]. DiffusionRenderer requires 112 seconds (~ 6 seconds per image) to process the full image set. Gaussian Splatting takes 131 seconds on average (ranging from 64 to 274 seconds depending on object complexity). Normal generation using R3DGS takes 270 seconds on average (247–347 seconds). Material optimization requires 975 seconds on average, extending up to 3,631 seconds (approximately one hour) for complex objects.

In total, our method takes 1,488 seconds (~ 25 minutes) on average, approximately $3.5 \times$ faster than IRGS (5,347 seconds, ~ 89 minutes).

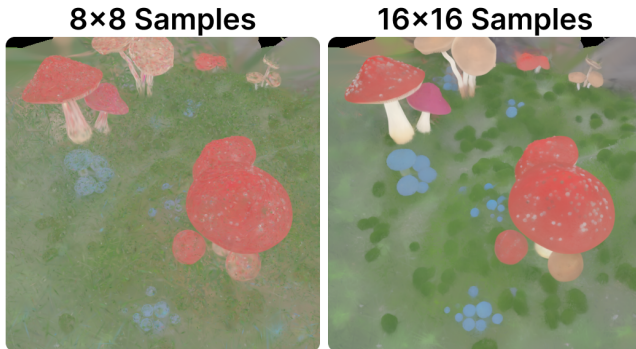


Figure 5. Super sampling avoids missed Gaussians and ensures reliable projection of material supervision. Lower sampling rates cause holes and unstable geometry.

Table 2. **Runtime Comparison** between our method and IRGS on the Navi dataset. All timings are reported in seconds and measured on an NVIDIA RTX 4090 GPU. Our method is approximately $3.5 \times$ faster than IRGS.

Stage	Ours (s)	IRGS (s)
Diffusion Predictions	112	-
Gaussian Splatting	131	2490
Normal Generation (R3DGS)	270	-
Material Optimization	975	2857
Total	1488	5347

Table 3. Quantitative ablation on a shared subset. The table compares four fusion strategies: a scale-invariant supervised baseline, a handcrafted weighting variant, a joint-channel MLP variant, and the final method. The final method achieves the best overall trade-off, with highest PSNR/SSIM and lowest LPIPS.

Method	Scale-Inv.	Suggested Weighting	Combi.	Channels MLP	Ours (Subset)
PSNR	26.1	27.3		26.7	29.1
SSIM	.91	.89		.89	.91
LPIPS	.06	.10		.10	.06

E. Additional Experiments

This section provides additional qualitative and quantitative analyses of the design choices in MatSpray. We compare alternative fusion strategies against the final model and visualize their effect on recovered base color and relighting quality.

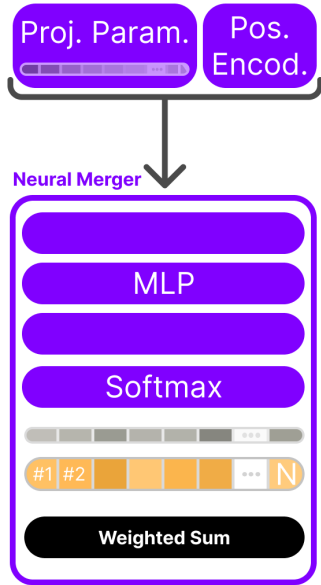


Figure 6. The input to the Neural Merger includes positional codes and projected material parameters.

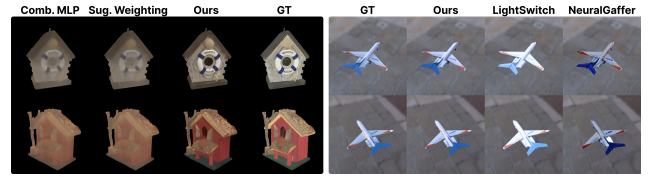


Figure 7. Qualitative comparison of ablated variants and the final model. The left side shows differences in recovered base color for different fusion strategies, highlighting artifacts such as residual baked-in lighting and reduced spatial consistency in weaker variants. The right side shows relighting behavior under novel illumination, where the final model yields cleaner materials and more stable appearance changes.