Supplementary Material
Unveiling the Ambiguity in Neural Inverse Rendering: A Parameter Compensation Analysis

Georgios Kouros[1] Minye Wu[1] Sushruth Nagesh[2] Xianling Zhang[2] Tinne Tuytelaars[1]
[1] KU Leuven [2] Ford Motor Company
[1]{georgios.kouros, minye.wu, tinne.tuytelaars}@esat.kuleuven.be
[2]{snagesh1, xzhan258}@ford.com

In this supplementary material, we present additional results from our experiments regarding the interactions and compensations between scene properties.

1. Quantitative Results Across Illuminations

In Tab. 1, we present the qualitative metrics of the three scenes, ball, car, and helmet, from the shiny-blender dataset [2], trained under different lighting conditions. NMF[1] achieves similar performance across different conditions. Although, not directly comparable, we can deduce from the mean agnular error (MAE) that the model learns consistent geometries across illuminations, which cannot be said for the material properties.

Table 1. Evaluation metrics of the car and helmet scenes from the shiny-blender dataset under three distinct illuminations.

Scene	Illum.	PSNR	SSIM	LPIPS	MAE	EPSNR
car	1	29.78	0.948	0.035	8.489	6.785
	2	29.26	0.945	0.045	8.471	9.218
	3	30.37	0.954	0.033	8.413	7.914
helmet	1	32.85	0.959	0.06	3.059	7.602
	2	34.14	0.97	0.051	2.258	9.718
	3	33.87	0.969	0.052	2.398	8.914

2. Quantitative Results Per Scene

Fig. 1 presents the examined metrics PSNR, SSIM, LPIPS, MAE, and EPSNR for the ball, car, and helmet scenes of the shiny-blender dataset, as well as a weighted average of the metrics across scenes. We utilize a weighted average to avoid scenes with higher metrics e.g. ball, to exert a larger influence on the graphs. To calculate the weighted average, we multiply each value with the corresponding scene average, sum the values across all scenes, and then divide by the sum of the scene averages. The results seem to be consistent across the scenes with some exceptions in the cases where a material property is too low to underestimate and thus has a minimal change in performance e.g. underestimating albedo for the ball scene.

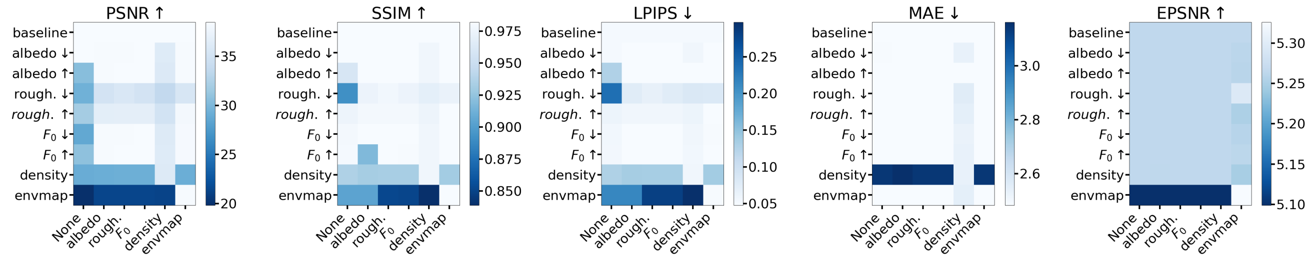
3. Additional Qualitative Results

We also provide the qualitative visualizations of the interaction and compensation experiments for the ball (Fig. 2) and helmet scenes (Fig. 3), where we observe similar outcomes as with the car scene. Regarding the ball scene, the model can recover more easily from errors in albedo or F_0 and slightly less from errors in roughness. Density and illumination perturbations, on the other hand, are the most difficult to recover from unless those scene properties are fine-tuned on their own. Furthermore, it's interesting to note how blurring the environment map affects the roughness, F_0, and geometry (surface normals) when individually fine-tuned. The effect is an increase in roughness, the bleeding of high-frequency environment details in F_0, or a rougher geometry, respectively. One issue with this scene is that the albedo and roughness properties are extremely low, which hinders meaningful underestimation or overestimation experiments while not providing insightful visualizations (note the almost entirely black albedo and roughness images in Fig. 2). Similarly, the helmet scene can better recover from the perturbations in albedo, and F_0 and less so in roughness and environment map.

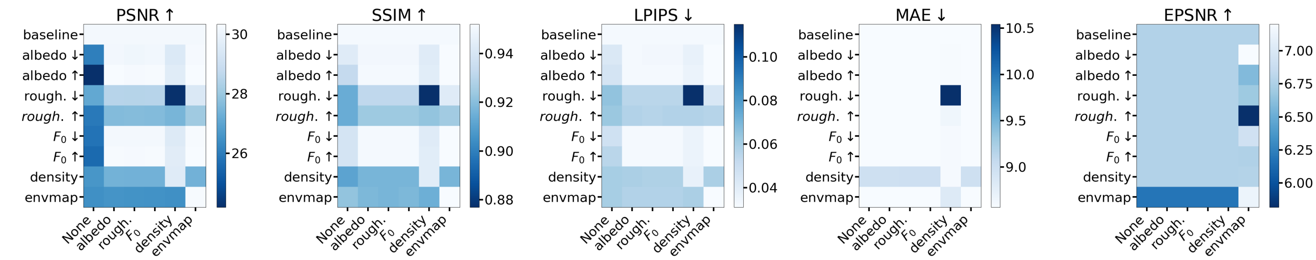
References

[1] Alexander Mai, Dor Verbin, Falko Kuester, and Sara Fridovich-Keil. Neural microfacet fields for inverse rendering, 2023. 1

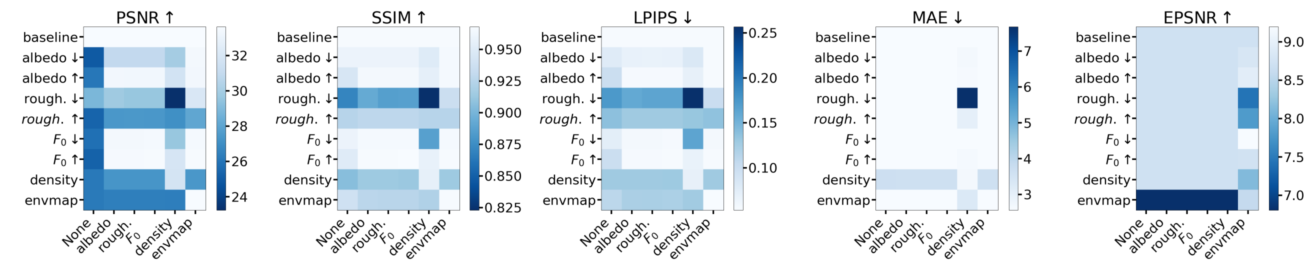
[2] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T. Barron, and Pratul P. Srinivasan. Ref-NeRF: Structured view-dependent appearance for neural radiance fields. *CVPR*, 2022. 1, 3, 4



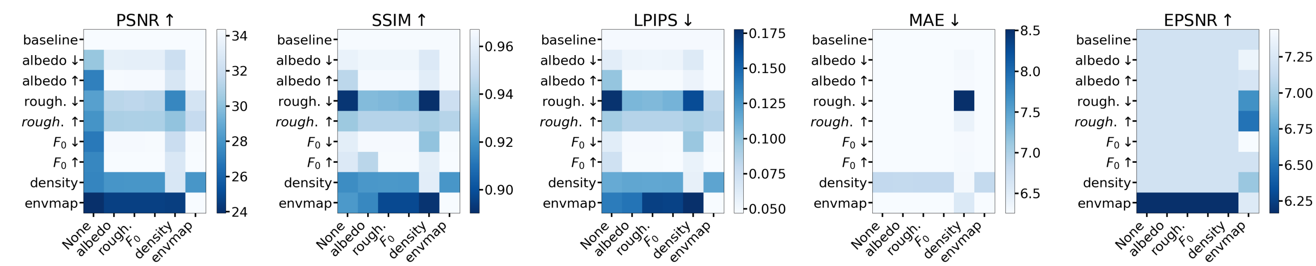
(a) ball



(b) car



(c) helmet



(d) weighted average

Figure 1. Graphs demonstrating how one scene property can compensate for suboptimal estimation of other properties for the ball, car, and helmet scenes of the shiny-blender dataset. We also show the weighted average (d) of the metrics across the scenes. The vertical axis corresponds to the manipulated scene property while the horizontal axis corresponds to the fine-tuned property. The arrows ↑, ↓ on the manipulated property labels refer to overestimating or underestimating the property, respectively. The arrows ↑, ↓ next to the metrics on the title of each plot, correspond with higher or lower is better, respectively.

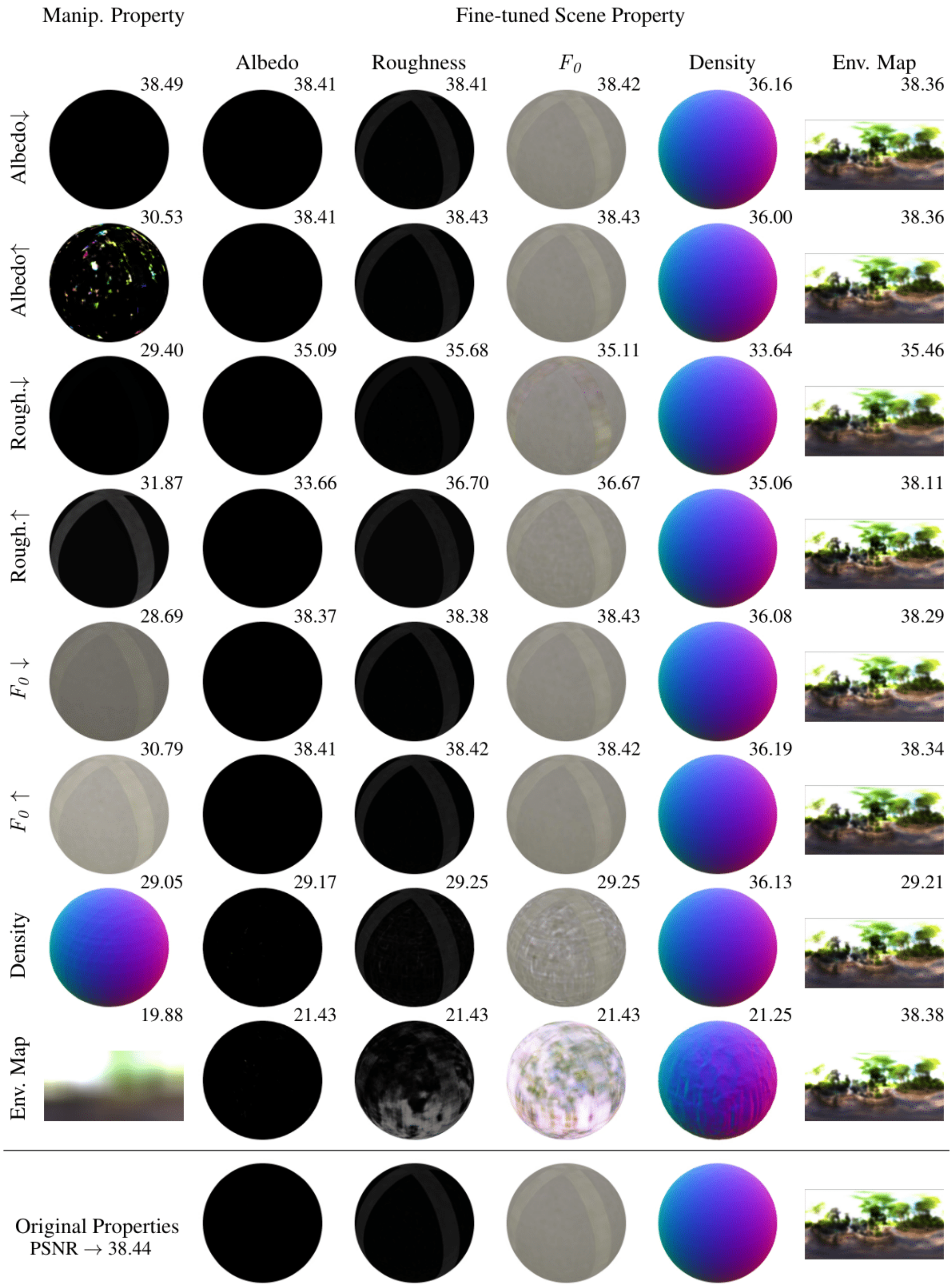


Figure 2. Effect of manipulating one scene property before fine-tuning another for the ball scene of the shiny-blender dataset [2]. The arrows ↓, ↑ next to albedo, roughness, and F_0 denote underestimation or overestimation respectively. The bottom row contains the original properties before adding noise and fine-tuning. For every experiment, we also include the PSNR at the top right corner of the corresponding

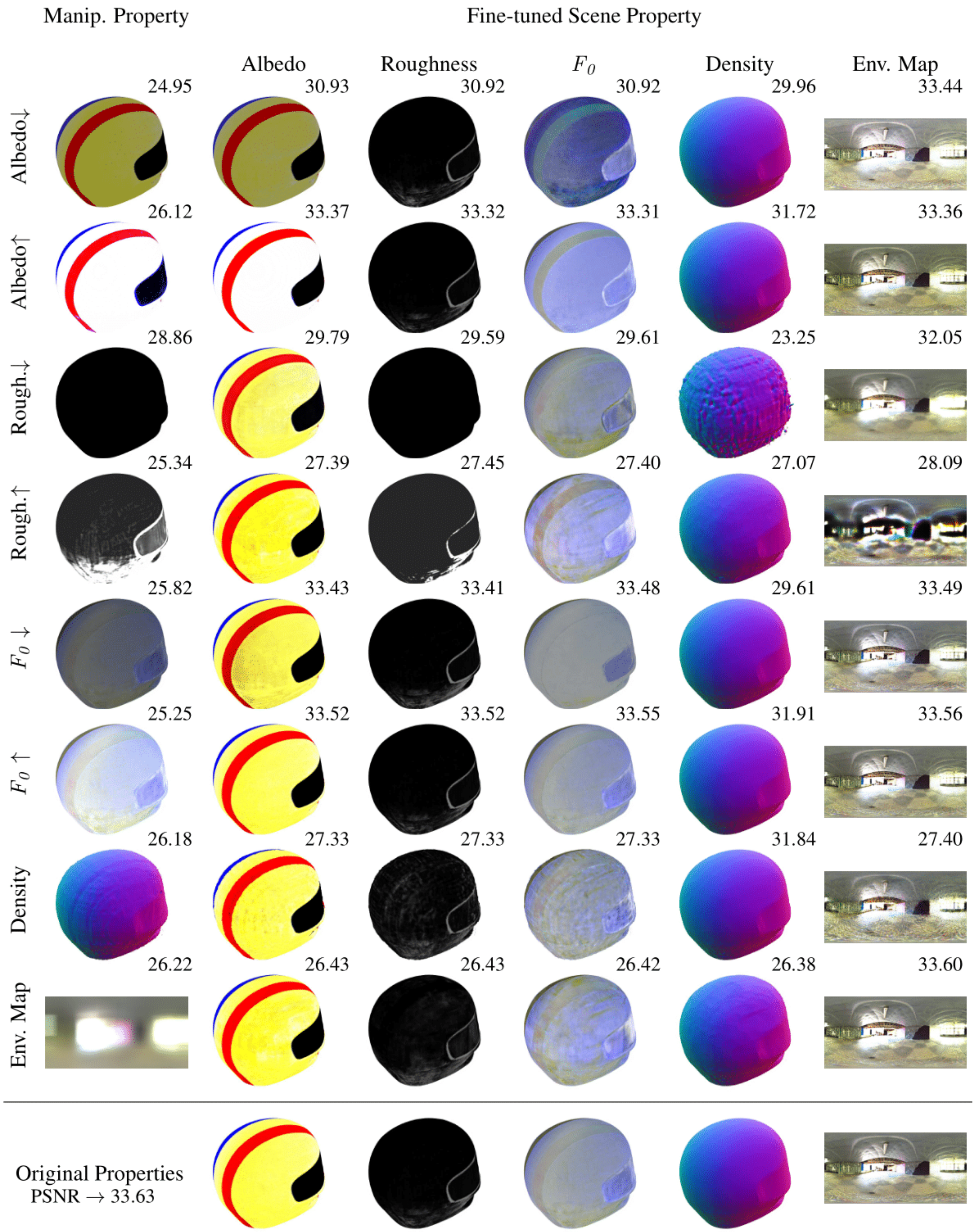


Figure 3. Effect of manipulating one scene property before fine-tuning another for the helmet scene of the shiny-blender dataset [2]. The arrows ↓, ↑ next to albedo, roughness, and F_0 denote underestimation or overestimation respectively. The bottom row contains the original properties before adding noise and fine-tuning. For every experiment, we also include the PSNR at the top right corner of the corresponding image.