

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

Georgios Kouros<sup>1</sup> Minye Wu<sup>1</sup> Sushruth Nagesh<sup>2</sup> Xianling Zhang<sup>2</sup> Tinne Tuytelaars<sup>1</sup>

<sup>1</sup> KU Leuven <sup>2</sup> Ford Motor Company

<sup>1</sup>{georgios.kouros, minye.wu, tinne.tuytelaars}@esat.kuleuven.be

<sup>2</sup>{snagesh1, xzhan258}@ford.com

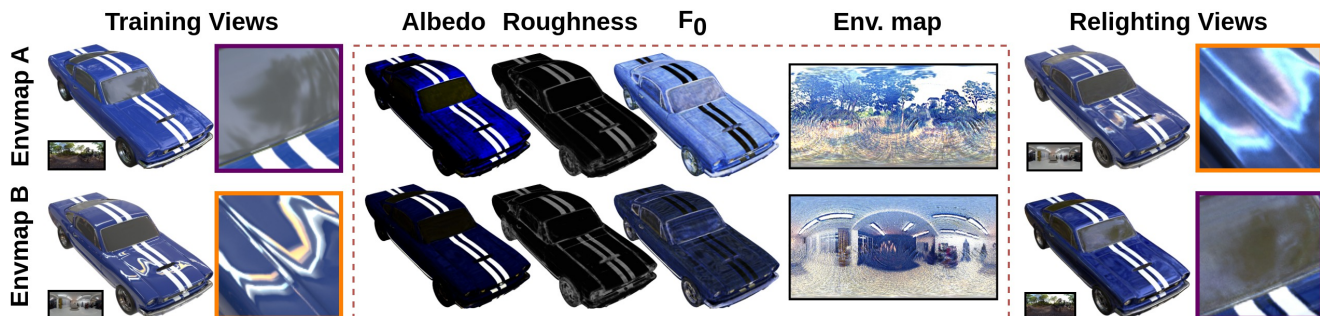


Figure 1. Demonstration of the inherent ambiguity in the task of neural inverse rendering. Training an inverse rendering algorithm like NMF [18] on the same object under different illuminations yields varying material properties that subsequently impact relighting tasks.

## Abstract

*Inverse rendering aims to reconstruct the scene properties of objects solely from multiview images. However, it is an ill-posed problem prone to producing ambiguous estimations deviating from physically accurate representations. In this paper, we utilize Neural Microfacet Fields (NMF), a state-of-the-art neural inverse rendering method to illustrate the inherent ambiguity. We propose an evaluation framework to assess the degree of compensation or interaction between the estimated scene properties, aiming to explore the mechanisms behind this ill-posed problem and potential mitigation strategies. Specifically, we introduce artificial perturbations to one scene property and examine how adjusting another property can compensate for these perturbations. To facilitate such experiments, we introduce a disentangled NMF where material properties are independent. The experimental findings underscore the intrinsic ambiguity present in neural inverse rendering and highlight the importance of providing additional guidance through geometry, material, and illumination priors.*

## 1. Introduction

Inverse rendering goes beyond simple reflection modeling and involves deducing scene properties from given images,

including geometry, lighting, and materials. This long-standing challenge finds applications in various fields, such as scene understanding, image manipulation, AR/VR, autonomous driving, etc. Unfortunately, inferring scene properties solely from multi-view images is an inherently ill-posed inverse problem. This is due to the intricate integral relationships defined by the rendering equation, which governs the behavior of light transport and connects geometry, materials, and illumination. This inherent complexity gives rise to inherent ambiguities in the process.

To investigate the inherent ambiguity in the task of neural inverse rendering, we employ Neural Microfacet Fields (NMF) [18], a state-of-the-art inverse rendering method for efficiently recovering high-fidelity scene geometry, materials, and lighting. However, similar to most inverse rendering methods the estimated material properties are not always correct when using NMF despite producing plausible reconstructions. NMF, for example, may estimate different material properties under different illumination settings, as shown in Fig. 1

To understand the correlation between scene properties and improve their estimation, we perform controlled experiments to discover underlying interactions and compensation mechanisms between them. For our experiments, we employ complex scenes of glossy objects from the shinyblender dataset [28] and demonstrate the inherent ambiguity in inverse rendering through inconsistencies among scene

properties for different illuminations. To understand the reason behind these inconsistencies, we design an experimental methodology based on which, we perturb one scene property and then fine-tune another to observe how the latter adapts and how well it can compensate to return the model closer to baseline reconstruction performance. We also adapt the architecture to disentangle the material properties and their input features to enable independent fine-tuning. This allows us to analyze the strengths and weaknesses of the complementary relationships between parameters, which in turn elucidates the extent of ambiguity during optimization. Our contributions can be summarized as follows:

- We demonstrate the inherent ambiguity of inverse rendering using NMF, by presenting the inconsistencies in estimated scene properties under different illuminations.
- Our key contribution is a framework designed to assess inverse rendering ambiguity, which systematically facilitates the analysis of compensatory dynamics among scene properties.
- Leveraging the established framework, our research conducts controlled experiments to identify interactions and compensation mechanisms between scene properties.

## 2. Related Work

### Neural Radiance Fields

Neural Radiance Fields (NeRF) [20] have emerged as a groundbreaking method in the field of computer graphics and vision, providing a novel approach to synthesizing photorealistic images from sets of posed 2D images using implicit neural representations. Although, highly photorealistic and compact, NeRF is inefficient for training and rendering. Subsequent methods adopted hybrid implicit-explicit representations such as voxel grids [15, 24, 26, 27] or even more compact tensor decompositions [4] or hash-encodings [21] that significantly speed up training and rendering without sacrificing reconstruction quality. Nevertheless, these methods struggle with glossy objects due to the ambiguity in the geometry caused by multi-view inconsistencies. Ref-NeRF[28] addressed this limitation by reparameterizing NeRF’s view-dependent appearance to better handle reflections, without, however, any relighting capabilities. The lighting-aware NeRF object model trained in LANE [13] can be composed into scenes with different lighting, but it does not handle glossy material well.

### Inverse Rendering

Inverse rendering is a complex task that aims to reconstruct the properties of a scene, such as geometry, materials, and lighting from images. Recent inverse rendering methods have incorporated differentiable volume rendering with implicit neural representations, inspired by NeRF

[20], to learn representations of the scene properties including geometry and materials. Early such works have made strong assumptions or require additional priors, such as a priori known illumination [1, 25] or geometry [33] or [1, 7] varying illuminations. Such limitations are addressed in [2, 3, 6, 12, 14, 31] by jointly estimating geometry, materials (e.g. BRDF), and lighting. One common issue with inverse rendering methods involves their limited or inefficient ability to handle interreflections, usually through visibility fields [25, 33] or radiance transfer fields [10, 19]. NMF [18] handles interreflections effectively by casting additional rays from micro-surfaces through the scene using efficient Monte-Carlo sampling with multi-bounce ray tracing. However, similar to most inverse rendering methods, it is susceptible to ambiguity and estimates plausible yet not consistent material properties under different lighting conditions.

Using images with varying illuminations has been shown to reduce ambiguity [2, 3, 5, 30]. In this work, however, we focus on the constant-illumination setting and what priors can be applied to alleviate the ambiguity of the problem. Priors have been previously used to reduce the ambiguity in inverse rendering and thus produce more consistent and plausible geometry, material properties, and illumination. The different kinds of priors can be categorized depending on which scene property they target, into geometry priors, material priors, and illumination priors. Geometry priors can be as simple as smoothing priors [11], object-category priors [23], or a priori known geometry [33]. Material priors usually involve pretraining of implicit neural BRDFs [3, 14] on BRDFs of known materials to constrain the search space to plausible materials. Illumination priors, on the other hand, can similarly involve smoothing priors to eliminate high-frequency noise [16, 33], or pretraining to lead the optimization into plausible illuminations [8, 17].

## 3. Methodology

### Neural Microfacet Fields

NMF [18] is a neural inverse rendering method that is trained on a collection of posed RGB images and estimates a scene representation composed of volume density  $\sigma$ , surface normals  $\hat{n}$ , material properties (BRDFs) including albedo  $\rho$ , roughness  $\alpha$ , and reflectance at the normal incidence angle  $F_0$ . The method also estimates the scene illumination, under a far-field assumption, represented as an environment map. NMF combines volume and surface rendering by treating each 3D point in space as a local micro-facet. This is modeled as a microfacet field represented by a density and a local micro-surface per 3D point. Light is accumulated along rays according to conventional volume rendering, but the outgoing light of every point along a ray is sampled based on its estimated micro-surface.

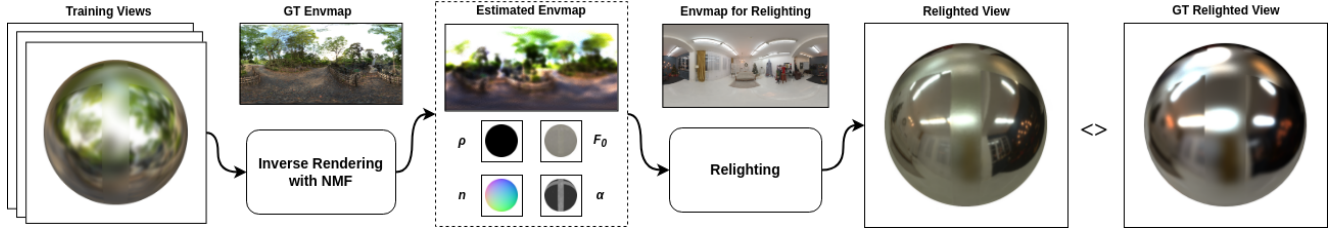


Figure 2. An example of suboptimal estimation of material properties and illumination. The model underestimates the roughness of the ball and overestimates the smoothness of the environment map, which results in sharper reflections when relighting the scene.

**Overall structure.** The geometry and appearance information of the scene is parameterized using TensorRF [4], a low-rank vector-matrix (VM) decomposition of a 4D tensor into multiple compact low-rank tensor components, that ensures storage and training/rendering time efficiency. Specifically, NMF employs one TensorRF model for the density field, noted as  $\mathcal{G}_\sigma$ , and one to represent the spatially varying appearance features of the scene, noted as  $\mathcal{G}_c$ . Therefore, the Neural Microfacet Model is designed to decode these features into RGB values and use volume rendering to accumulate pixel colors.

**Neural Microfacet Model.** The Neural Microfacet Model represents materials through a spatially varying BRDF model corresponding to the sum of the diffuse and specular components of the scene, expressed as

$$f(\hat{\omega}_o, \hat{\omega}_i) = \frac{\rho}{\pi}(1 - F_r(\hat{\mathbf{h}})) + F_r(\hat{\mathbf{h}})f_s(\hat{\omega}_o, \hat{\omega}_i), \quad (1)$$

where  $\rho$  is the RGB albedo,  $F_r$  is the fresnel term,  $\hat{\mathbf{h}}$  is the half vector,  $f_s$  is the specular component, and  $\hat{\omega}_o, \hat{\omega}_i$  are the outgoing and incoming light directions. NMF uses the Schnell approximation of the Fresnel term, calculated as

$$F_r(\hat{\mathbf{h}}) = F_0(\mathbf{p}) + (1 - F_0(\mathbf{p}))(1 - \hat{\mathbf{h}} \cdot \hat{\omega}_o)^5, \quad (2)$$

where  $F_0(\mathbf{p})$  is the specular reflection at the normal incidence for a 3D point  $\mathbf{p}$ .

The specular BRDF, based on the Cook-Torrance model, is expressed as

$$f_s(\hat{\omega}_o, \hat{\omega}_i) = \frac{D(\hat{\mathbf{h}}; \alpha, \hat{\mathbf{n}}, \hat{\omega}_o)G_1(\hat{\omega}_o, \hat{\mathbf{h}})g(\hat{\omega}_o, \hat{\omega}_i)}{4(\hat{\mathbf{n}} \cdot \hat{\omega}_o)(\hat{\mathbf{n}} \cdot \hat{\omega}_i)}, \quad (3)$$

where  $D$  is the Trowbridge-Reitz distribution, an approximation of the material’s roughness,  $G_1$  is the Smith shadow masking function, and  $g$  represents material properties that are not modeled explicitly, but rather implicitly via an MLP.

Albedo  $\rho$ , roughness  $\alpha$ , and normal incidence specular reflectance  $F_0$  are modeled using MLPs. These MLPs are conditioned on the positional encoding of 3D positions and

the appearance features at those positions, and their outputs are passed through sigmoid nonlinearities. The surface normals, on the other hand, are estimated as the negative normalized gradient of the density and are regularized, similar to [28] to penalize back-facing normals and prevent semi-transparent surfaces. For more details about the architecture and rendering pipeline of NMF, we refer the reader to the original paper [18].

### The Ambiguity of Scene Properties

As already discussed in several existing works [2, 3, 17, 18, 33] inverse rendering is a highly ill-posed problem entailing a lot of ambiguity in the estimation of a scene’s properties. Consequently, most works [22, 34] focus on estimating scene properties that produce a plausible reconstruction of the scene, even with suboptimal and physically unrealistic material properties. Although the actual reconstruction may be of high quality, if the geometry and material properties of the scene are not estimated correctly, any downstream relighting task will diverge from the expected behaviour.

Ambiguity intensifies when dealing with a scene under constant lighting conditions rather than varying. Varying lighting conditions allow a model [13] to better disambiguate the geometry, material properties, and illumination of the scene. In the absence of varying lighting conditions, however, the estimated material properties can vary depending on how the scene is illuminated. This raises the question of how various scene properties interact or compensate for each other and the extent of their significance. For example, as shown in Figure 2, the model may underestimate the roughness of the object and produce sharper reflections during relighting tasks. As a result, first, we aim to demonstrate the ambiguity of inverse rendering using NMF as our testbench. Second, we aim to identify potential interactions between the scene properties and how the suboptimal estimation of one can be compensated by another.

To demonstrate the ambiguity and ill-posedness of inverse rendering we investigate how (in)consistent NMF is by generating the same objects under the same poses but different lighting conditions. Then training NMF models on each of those lighting conditions allows us to observe the variance in the estimated material properties.

Given the established ambiguity in inverse rendering, we aim to investigate how the ambiguity of one scene property affects the other properties and how one compensates for the other so that the model achieves good reconstruction even under suboptimal geometry, material, or illumination estimates. To this end, we attempt to isolate the interaction between two scene properties at a time, by first manipulating one scene property, freezing all other properties but one, and then fine-tuning the remaining unfrozen property. For this experiment, we select the view-consistent material properties of the model, namely albedo  $\rho$ , roughness  $\alpha$ , and normal incidence reflectance  $F_0$ , as well as the density  $\sigma$  and the estimated environment map.

**Disentanglement.** As shown in the upper part of Fig. 3, albedo, roughness and  $F_0$  MLPs are conditioned on the same appearance features extracted from a TensorRF representation  $\mathcal{G}_c$ . To facilitate independent fine-tuning of a single material property at a time, we first need to disentangle their inputs, considering that any gradients passing through an MLP would backpropagate to the shared appearance features and hence elicit change to the other properties as well. Simply splitting the feature vector into parts would still be insufficient, since the elements of the feature vector are not completely disentangled. Therefore, we replace the appearance TensorRF with three distinct TensorRFs but with proportionally smaller dimensions (from 24 to 8). The input feature vector of each MLP is also concatenated with the positional encoding (PE) of the 3D position  $x$ . The adapted architecture is shown at the bottom of Fig. 3.

**Properties Tweaking.** To investigate the interaction between scene properties and discover any inherent compensating mechanisms, we first need to manipulate them by tweaking one property at a time before fine-tuning another. We employ three types of applied noise depending on the type of each property split into material, geometry, and illumination noise. Inverse rendering models tend to underestimate or overestimate scene properties like albedo, roughness, and  $F_0$  in NMF. To simulate such effects we apply a simple multiplier  $m$  on the outputs of the MLPs, after the sigmoid, while also clipping the adjusted property so that it falls in the expected range. This entails the limitation that the model cannot fine-tune a property to reach a higher value than the multiplier and also if a property is closed to zero then a multiplier would have little effect. Multiplier values lower or greater than one simulate underestimation or overestimation of the property, respectively. An additional common issue observed in inverse rendering, especially when dealing with shiny objects, is that an overestimation of roughness might result in a rougher more noisy geometry. To simulate this effect, we perturb the geom-

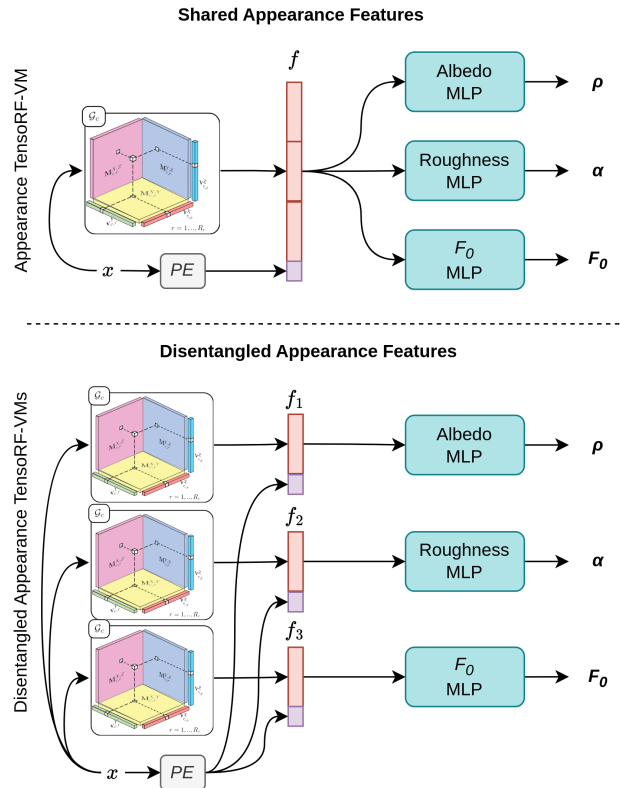


Figure 3. Disentanglement of appearance features to enable independent fine-tuning of albedo  $\rho$ , roughness  $a$ , and  $F_0$ . The single appearance TensorRF is split into three distinct TensorRFs with one third of the original feature dimensionality. This ensures that fine-tuning the albedo, roughness, or  $F_0$  will not impact the input feature space of the other two. The TensorRF visualization was adapted from the original paper [4].

etry by directly applying Gaussian noise on  $\mathcal{G}_\sigma$ , noted as  $N(0, \sigma_d^2)$ , where  $\sigma_d$  stands for the variance of the distribution. On the other hand, underestimating the roughness may cause an estimation of a blurry environment map, which we can easily simulate by applying a Gaussian smoothing filter. We denote the filter as  $G(s_I, \sigma_I)$ , where  $s_I$  and  $\sigma_I$  denote the size and the standard deviation of the Gaussian kernel, respectively.

## 4. Results

### Evaluation Protocol

We evaluate our experiments both qualitatively and quantitatively using conventional novel view synthesis metrics, namely PSNR [9], SSIM [29], and LPIPS [32] for the visual aspect of the reconstruction. For the estimated geometry, we provide the mean angular error (MAE) in degrees between the estimated and ground truth surface normals. Last but not least, we also utilize PSNR to evaluate the quality of the estimated environment map compared to the ground



truth as implemented in the released codebase of NMF[18]. To differentiate between the PSNRs we term the PSNR of the environment map as EPSNR.

### Consistency Across Illuminations

In our first experiment, we evaluate the consistency of inverse rendering in estimating scene properties under different illuminations. To this end, we use two scenes from the shiny-blender dataset, namely *car* and *helmet*, and render each of them under three different illuminations using the released blender models [28]. Then, we fully train NMF models for each object-illumination combination separately. In Fig. 4, we present the estimated scene properties of each object under the different illumination settings as seen in the first column. The second column shows the estimated environment map, followed by the color renderings, albedo, roughness,  $F_0$ , and surface normals. As evident, NMF fails to estimate consistent scene properties across illuminations, with the largest variations occurring in albedo and  $F_0$ . The roughness and the geometry estimates (as viewed through surface normals) appear more consistent, although, there seem to be more errors in cases of high-frequency reflections as with *Illumination #1*. Furthermore, there seems to be bleeding of the albedo of the object into the estimate of the environment map, observed as the blue and yellow tints for the car and helmet, respectively. Based on the results of this experiment, we can safely conclude that inverse rendering under a single illumination, as in the case of NMF, can be an ill-posed problem with a large degree of inherent ambiguity in the estimation of the scene properties.

### Scene Property Interactions and Compensations

To evaluate the interaction between scene properties and their corresponding abilities to compensate for perturbations in other properties, we begin by fully training an NMF model for 30K iterations. Then, we manipulate one of the scene properties, freeze it, and successively unfreeze one other property at a time for fine-tuning. This way we can observe how effectively the latter can compensate for the perturbations introduced in the former.

To avoid the fine-tuning of one property affecting another, we completely disentangle albedo, roughness and  $F_0$  as described in Section 3. In Tab. 1, we present the performance of NMF with shared versus disentangled appearance features across three scenes of the shiny blender dataset [28], namely ball, car, and helmet. Based on these results, we observe a minimal change in performance.

Tab. 2 shows the amount of manipulation applied on each scene property and the resulting performance drop before fine-tuning. The introduced noise to the model was tuned to reach a 15-25% drop in PSNR. The results of the experiments are presented quantitatively for all scenes in Fig. 5

and qualitatively for just the car scene in Fig. 6. The quantitative results per scene and the qualitative results for the ball and helmet scenes are included in the supplementary material.

From the averaged quantitative results for all scenes in Fig. 5, we observe that the model can easily compensate for changes in albedo and  $F_0$ , but struggles to improve performance when there are errors in roughness, density, or the environment map. Errors in density and the environment map specifically appear to be the hardest to recover from by compensating with another property. Moreover, it is evident across all scenes, that the model can better compensate for errors in albedo, roughness, and  $F_0$  by fine-tuning one of them or the environment map but fine-tuning density, on the other hand, provides less improvement and results in worse geometry. Additional results per scene are included in the supplementary material.

Also in Fig. 6, we can observe specific interactions and compensations between some scene properties. For example, underestimating albedo and fine-tuning  $F_0$ , seems to cause a drastic change in the latter, compared to the original unperturbed property. In general, fine-tuning  $F_0$  appears to result in significantly more variation across the experiments compared to the other properties. Albedo and roughness mostly vary in intensity without any observable structural changes when they are fine-tuned. Fine-tuning the geometry, on the other hand, appears to vary most drastically when the roughness is manipulated or the environment map is smoothed. More specifically, a large degree of underestimation in roughness causes the geometry to become rougher itself, observed as the extremely noisy normals with correspondingly low metrics. Overestimation of the roughness, on the other hand, has less effect, although still noticeable, on the geometry. but an even larger effect on the environment map, causing it to become blurrier. At the same time, it seems that overestimating the smoothness of the environment map causes a rougher more noisy geometry as well. Lastly, we also note that the largest PSNR improvement is mostly observed when fine-tuning the environment map, which indicates its ability to compensate for perturbations on other properties.

### Future Research

Alleviating the ill-posedness of inverse rendering is certainly a difficult problem, especially for constant illumination scenes because of the complexity of disambiguating between geometry, materials, and illumination. However, priors can be beneficial to that end and therefore we highlight potential research directions focusing on guiding or regularizing the geometry, materials, or illumination properties of the scene.

As observed in our experiments, suboptimal geometry leads to significant drops in performance that cannot be ad-

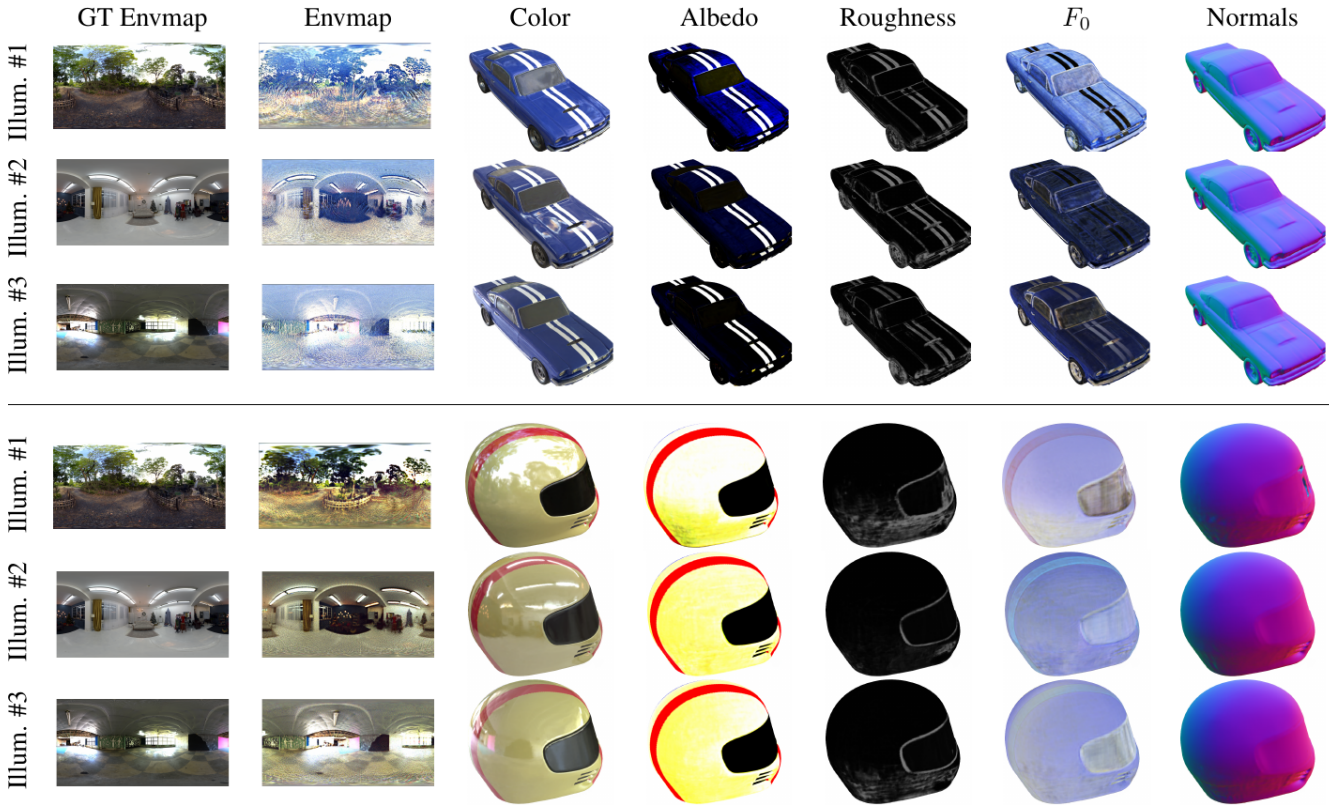


Figure 4. Estimated scene properties of the car and helmet scenes from the shiny-blender dataset under different illuminations. Due to the underconstrained nature of the problem and NMF optimizing solely for reconstruction, it fails to learn consistent scene properties.

equately compensated for by other scene properties. Therefore, accurate geometry estimation is vital to the inverse rendering process, and geometry priors can thus be beneficial in alleviating some ambiguity. Simple smoothing priors can easily be applied to the depth or normals, especially for shiny objects that can be assumed by definition to be locally smooth. However, this requires careful tuning to avoid oversmoothing and losing high-frequency geometry details. More sophisticated geometry priors would be either instance-specific (e.g. pretraining to extract a rough or even fine geometry as an initialization to speed up inverse rendering) or category-specific which would require training on a large amount of objects to generalize across categories and at the same time restrict the geometry search space.

Regarding material priors and based on our experiments, we conclude that the suboptimal material properties can lead to large errors in geometry or illumination (especially for roughness) and even though the model may be able to compensate for such errors towards a better reconstruction by adapting albedo or  $F_0$ , subsequent relighting efforts would be heavily impacted. We also observed inconsistencies across surface points that should have the same material properties. As a result, we advocate for moving forward by utilizing pre-trained BRDFs [14] to restrict the search space

of inverse rendering to real physically plausible materials either explicitly or as a regularization step that penalizes unrealistic materials. At the same time, another promising direction involves semantic-driven priors to restrict the search space but also enforce material consistency among points belonging to the same semantic class or instance. For example, the points on the surface of a metal sphere should be restricted to metal BRDFs, and should also be consistent across the surface.

Finally, based on the outcomes of our experiments, the model struggles to estimate an accurate illumination due to the ambiguity with other scene properties which may cause high-frequency noise, oversmoothing, or bleeding of albedo into the environment map. This indicates a need for illumination priors to improve this ambiguity. Specifically, we propose that such priors should follow the paradigm of [17] and focus on generative approaches to guide the optimization by sampling more natural illuminations or more plausible illumination-material combinations.

## Conclusion

In this work, we investigate the ambiguity and ill-posedness of neural inverse rendering while trying to uncover interactions between the scene properties and potential compen-

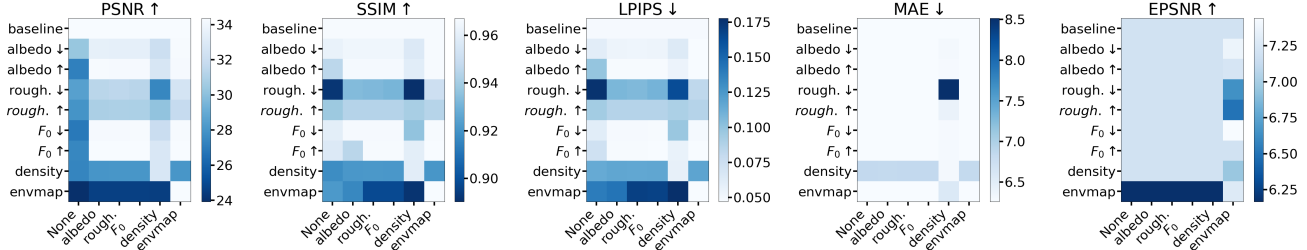


Figure 5. Graphs demonstrating to what extent one scene property can compensate for suboptimal estimation of other properties. The vertical axis corresponds to the manipulated scene property while the horizontal axis corresponds to the fine-tuned property. The arrows  $\uparrow, \downarrow$  on the manipulated property labels refer to overestimating or underestimating the property, respectively. A weighted average is used across the examined scenes (ball, car, helmet) of the shiny-blender dataset.

Table 1. Performance of NMF with shared versus disentangled appearance features on three scenes of the shiny-blender dataset (ball, car, and helmet) averaged over five runs. The table includes the average metric (black) and the standard deviation (blue) per scene and across the scenes. The small performance drop observed with the disentangled representation of the appearance features is deemed acceptable for the purposes of our experiments.

Parameter	PSNR	SSIM	LPIPS	MAE	
shared	ball	$38.73 \pm 0.042$	$0.982 \pm 0.000$	$0.051 \pm 0.000$	$2.492 \pm 0.004$
	car	$30.04 \pm 0.070$	$0.949 \pm 0.001$	$0.033 \pm 0.000$	$7.709 \pm 0.085$
	helmet	$34.18 \pm 0.100$	$0.970 \pm 0.000$	$0.052 \pm 0.000$	$2.415 \pm 0.028$
	Mean	$34.32 \pm 0.050$	$0.967 \pm 0.001$	$0.045 \pm 0.000$	$4.205 \pm 0.041$
disent.	ball	$38.46 \pm 0.032$	$0.976 \pm 0.013$	$0.051 \pm 0.001$	$2.482 \pm 0.004$
	car	$30.20 \pm 0.024$	$0.951 \pm 0.000$	$0.03 \pm 0.000$	$8.552 \pm 0.025$
	helmet	$33.56 \pm 0.166$	$0.964 \pm 0.004$	$0.056 \pm 0.002$	$2.626 \pm 0.061$
	Mean	$34.07 \pm 0.070$	$0.964 \pm 0.006$	$0.046 \pm 0.001$	$4.553 \pm 0.030$

sating mechanisms that allow a model to produce plausible reconstructions with suboptimal geometry, material properties, and illumination. Using Neural Microfacet Fields (NMF), we showcase this ambiguity by demonstrating how the scene properties can vary under different illuminations. Given our proposed evaluation framework, we perform controlled experiments for manipulating one scene property and independently fine-tuning another to discover interactions between them. Our experiments show that the model can easily compensate for suboptimal albedo or  $F_0$  estimates, but not so easily for roughness, geometry, or illumination which can significantly impact results. Finally, we propose future research towards geometry, material, and illumination priors to reduce the ambiguity of inverse rendering tasks and produce more consistent results across different illuminations, and thus more accurate, plausible, and consistent relighting.

## Acknowledgements

We gratefully acknowledge funding support from the Sim2Real2 project, in the context of the Ford-KU Leuven alliance program.

Table 2. Effect of manipulating a single parameter at a time without finetuning. The arrows  $\downarrow, \uparrow$  next to albedo, roughness, and  $F_0$  denote underestimation and overestimation, respectively. Albedo, roughness, and  $F_0$  are perturbed using a multiplier  $m$  at the output of the MLP after the sigmoid nonlinearity, while the geometry and illumination are perturbed by added Gaussian noise  $N(0, \sigma_d^2)$  and Gaussian blurring  $G(s_I, \sigma_I)$ , respectively.

Parameter	Noise	PSNR	SSIM	LPIPS	MAE	EPSNR	
-	-	38.44	0.982	0.051	2.484	5.265	
ball	Albedo $\downarrow$ *	$m = 0$	38.49	0.982	0.051	2.489	5.265
	Albedo $\uparrow$	$m = 1000$	30.53	0.959	0.127	2.484	5.265
	Rough. $\downarrow$	$m = 0.1$	29.40	0.892	0.236	2.484	5.265
	Rough. $\uparrow$	$m = 2.0$	31.87	0.975	0.061	2.484	5.625
	$F_0 \downarrow$	$m = 0.8$	28.69	0.979	0.055	2.484	5.265
	$F_0 \uparrow$	$m = 1.2$	30.79	0.980	0.054	2.484	5.265
	Density	$N(0, 1.05)$	29.05	0.936	0.128	3.142	5.265
	Envmap	$G(301, 300)$	19.88	0.904	0.214	2.484	5.099
car	-	-	30.23	0.951	0.032	8.566	6.771
	Albedo $\downarrow$	$m = 0.4$	26.08	0.943	0.043	8.566	6.771
	Albedo $\uparrow$	$m = 4.0$	24.19	0.933	0.046	7.566	6.771
	Rough. $\downarrow$	$m = 0.0$	27.10	0.914	0.065	8.566	6.771
	Rough. $\uparrow$	$m = 10$	25.93	0.914	0.063	8.566	6.771
	$F_0 \downarrow$	$m = 0.6$	25.78	0.939	0.048	8.566	6.771
	$F_0 \uparrow$	$m = 1.5$	25.56	0.939	0.055	8.566	6.771
	Density	$N(0, 1.5)$	26.64	0.911	0.060	8.970	6.771
Envmap	$G(151, 100)$	26.46	0.922	0.060	8.566	6.189	
helmet	-	-	33.63	0.967	0.054	2.568	8.719
	Albedo $\downarrow$	$m = 0.6$	24.95	0.96	0.077	2.568	8.719
	Albedo $\uparrow$	$m = 100$	26.12	0.945	0.099	2.568	8.719
	Rough. $\downarrow$ *	$m = 0$	28.86	0.871	0.174	2.568	8.719
	Rough. $\uparrow$	$m = 15$	25.34	0.925	0.138	2.568	8.719
	$F_0 \downarrow$	$m = 0.6$	25.82	0.960	0.073	2.568	8.719
	$F_0 \uparrow$	$m = 1.5$	25.25	0.951	0.099	2.568	8.719
	Density	$N(0, 3)$	26.18	0.906	0.128	3.649	8.719
Envmap	$G(151, 100)$	26.22	0.939	0.110	2.568	6.809	

\* Properties are too low for an underestimation experiment to have the desired drop in performance before finetuning, but we still report them for completeness.



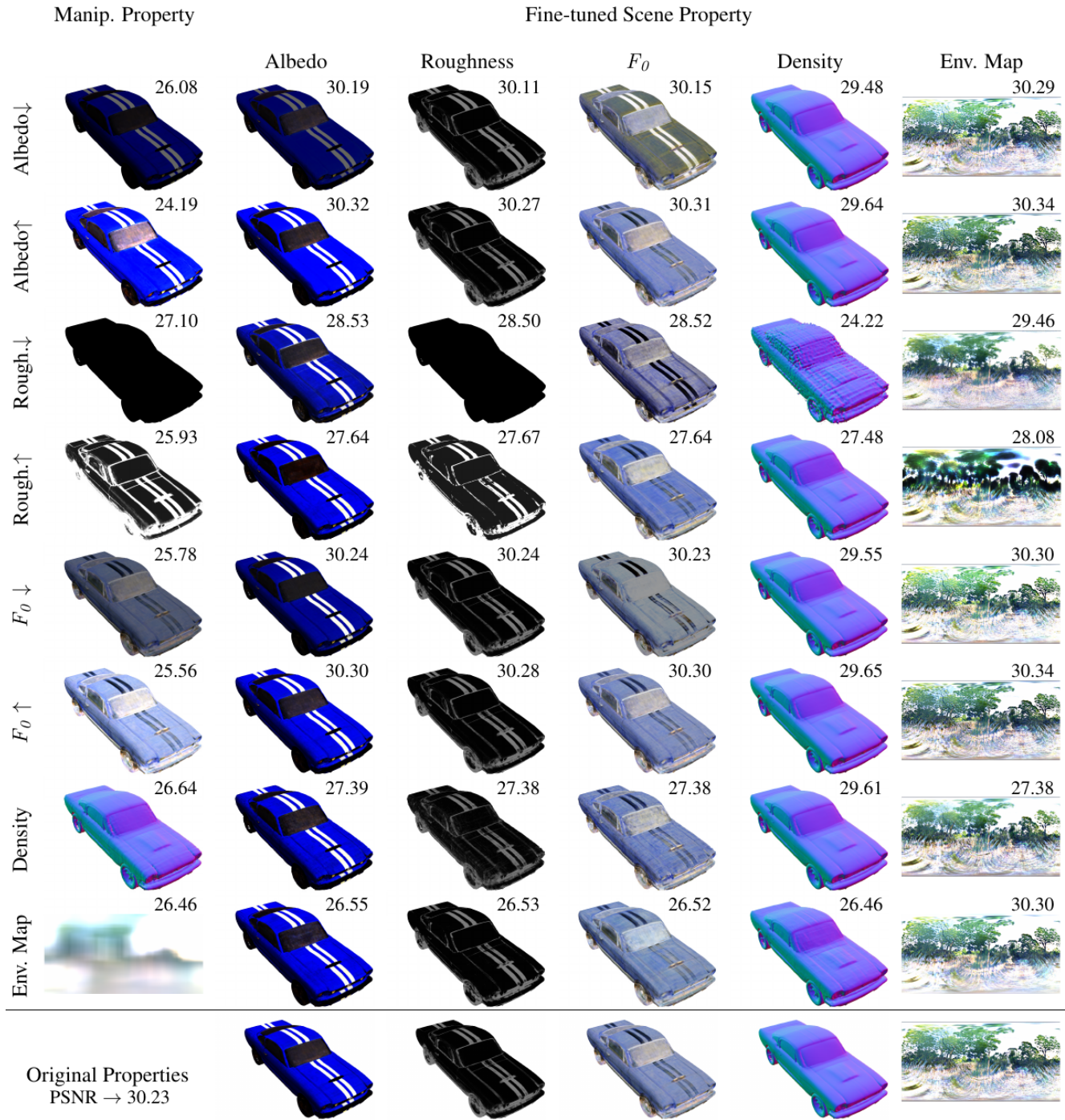


Figure 6. Effect of manipulating one scene property before fine-tuning another for the car scene of the shiny-blender dataset [28]. The arrows ↓, ↑ next to albedo, roughness, and  $F_0$  denote underestimation or overestimation respectively. The bottom row contains the original properties before adding noise and fine-tuning. For every experiment, we also include the PSNR at the top right corner of the corresponding image.



## References

- [1] Sai Bi, Zexiang Xu, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. Deep reflectance volumes: Relightable reconstructions from multi-view photometric images. In *European Conference on Computer Vision (ECCV)*, page 294–311, Berlin, Heidelberg, 2020. Springer-Verlag. 2
- [2] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T. Barron, Ce Liu, and Hendrik P.A. Lensch. Nerd: Neural reflectance decomposition from image collections. In *IEEE International Conference on Computer Vision (ICCV)*, 2021. 2, 3
- [3] Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan T. Barron, and Hendrik P.A. Lensch. Neural-pil: Neural pre-integrated lighting for reflectance decomposition. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 2, 3
- [4] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *European Conference on Computer Vision (ECCV)*, 2022. 2, 3, 4
- [5] Andreas Engelhardt, Amit Raj, Mark Boss, Yunzhi Zhang, Abhishek Kar, Yuanzhen Li, Deqing Sun, Jonathan T. Barron, Hendrik P.A. Lensch, and Varun Jampani. SHINOBI: Shape and Illumination using Neural Object decomposition via Brdf optimization In-the-wild, 2023. 2
- [6] Yue Fan, Ivan Skorokhodov, Oleg Voynov, Savva Ignatyev, Evgeny Burnaev, Peter Wonka, and Yiqun Wang. Factored-neus: Reconstructing surfaces, illumination, and materials of possibly glossy objects, 2023. 2
- [7] Duan Gao, Guojun Chen, Yue Dong, Pieter Peers, Kun Xu, and Xin Tong. Deferred neural lighting: free-viewpoint relighting from unstructured photographs. *ACM Transactions on Graphics (TOG)*, 39(6):258, 2020. 2
- [8] James A D Gardner, Bernhard Egger, and William A P Smith. Rotation-equivariant conditional spherical neural fields for learning a natural illumination prior. In *Advances in Neural Information Processing Systems*, 2022. 2
- [9] Jerry D. Gibson and Al Bovik. *Handbook of Image and Video Processing*. Academic Press, Inc., USA, 1st edition, 2000. 4
- [10] Michelle Guo, Alireza Fathi, Jiajun Wu, and Thomas Funkhouser. Object-centric neural scene rendering, 2020. 2
- [11] Yuan-Chen Guo, Di Kang, Linchao Bao, Yu He, and Song-Hai Zhang. Nerfren: Neural radiance fields with reflections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18409–18418, 2022. 2
- [12] Haiyan Jin, Isabella Liu, Peijia Xu, Xiaoshuai Zhang, Songfang Han, Sai Bi, Xiaowei Zhou, Zexiang Xu, and Hao Su. Tensor: Tensorial inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 2
- [13] Akshay Krishnan, Amit Raj, Xianling Zhang, Alexandra Carlson, Nathan Tseng, Sandhya Sridhar, Nikita Jaipuria, and James Hays. Lane: Lighting-aware neural fields for compositional scene synthesis, 2023. 2, 3
- [14] Ruofan Liang, Huiting Chen, Chunlin Li, Fan Chen, Selvakumar Panneer, and Nandita Vijaykumar. Envidr: Implicit differentiable renderer with neural environment lighting. *arXiv preprint arXiv:2303.13022*, 2023. 2, 6
- [15] Lingjie Liu, Jiatao Gu, Kyaw Zaw Lin, Tat-Seng Chua, and Christian Theobalt. Neural sparse voxel fields. *NeurIPS*, 2020. 2
- [16] Linjie Lyu, Ayush Tewari, Thomas Leimkühler, Marc Habermann, and Christian Theobalt. Neural radiance transfer fields for relightable novel-view synthesis with global illumination. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVII*, page 153–169, Berlin, Heidelberg, 2022. Springer-Verlag. 2
- [17] Linjie Lyu, Ayush Tewari, Marc Habermann, Shunsuke Saito, Michael Zollhöfer, Thomas Leimkühler, and Christian Theobalt. Diffusion posterior illumination for ambiguity-aware inverse rendering. *ACM Transactions on Graphics*, 42(6), 2023. 2, 3, 6
- [18] Alexander Mai, Dor Verbin, Falko Kuester, and Sara Fridovich-Keil. Neural microfacet fields for inverse rendering, 2023. 1, 2, 3, 5
- [19] N. Max. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1(2): 99–108, 1995. 2
- [20] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 2
- [21] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4):102:1–102:15, 2022. 2
- [22] Ava Pun, Gary Sun, Jingkang Wang, Yun Chen, Ze Yang, Sivabalan Manivasagam, Wei-Chiu Ma, and Raquel Urtasun. Neural lighting simulation for urban scenes. In *Advances in Neural Information Processing Systems*, pages 19291–19326. Curran Associates, Inc., 2023. 3
- [23] Konstantinos Rematas, Ricardo Martin-Brualla, and Vittorio Ferrari. Sharf: Shape-conditioned radiance fields from a single view. In *ICML*, 2021. 2
- [24] Sara Fridovich-Keil and Alex Yu, Matthew Tancik, Qinlong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *CVPR*, 2022. 2
- [25] Pratul P. Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T. Barron. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In *CVPR*, 2021. 2
- [26] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In *CVPR*, 2022. 2
- [27] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Improved direct voxel grid optimization for radiance fields reconstruction, 2022. 2
- [28] 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, 2, 3, 5, 8

- [29] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 4
- [30] Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K. Wong. Ps-nerf: Neural inverse rendering for multi-view photometric stereo. In *European Conference on Computer Vision (ECCV)*, 2022. 2
- [31] Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. PhysSG: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 2
- [32] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric, 2018. 4
- [33] Xiuming Zhang, Pratul P. Srinivasan, Boyang Deng, Paul Debevec, William T. Freeman, and Jonathan T. Barron. Nerfactor: neural factorization of shape and reflectance under an unknown illumination. *ACM Trans. Graph.*, 40(6), 2021. 2, 3
- [34] Xianling Zhang, Nathan Tseng, Ameerah Syed, Rohan Bhasin, and Nikita Jaipuria. Simbar: Single image-based scene relighting for effective data augmentation for automated driving vision tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3718–3728, 2022. 3