

A7. Overview

In this supplementary, we present video demonstrations of the scenes relighted under rich lighting conditions after reconstructed using our method (A8).

Theories We provide an background on neural implicit surface volume rendering and physically based rendering (A9). We provide additional discussion for non-convexity of co-located photometric stereo during geometry initialization (A10), which motivates us to use SfM points during initialization. We provide a proof (A11) that the indirect component of radiance $\sum_{k=2}^{\infty} \mathcal{T}_{\phi}^k(E)(\mathbf{x}, \omega_o)$ is continuous, which shapes our dynamic radiance cache design.

Experiments We perform quantitative evaluation on real data following measured albedo in the wild A12.1. **All measurements and annotations will be released.** we show ablation of surface angle weighting loss (A12.2). We qualitatively compare with natural illumination methods on material properties (A12.3), geometry (A12.4), re-rendering (A12.5). We compare with WildLight [7] on wild-light capture setup (A12.6). Finally, we show additional views from our dataset. (A12.7)

Details We provide additional implementations details on our algorithm (A13) and real dataset construction (A14).

A8. Video Demonstrations

We render all our scenes under novel lighting conditions with our recovered geometry and material parameters using path tracing renderer Mitsuba [46], and denoise with its built-in NVIDIA OptiX denoiser [35]. The videos feature ambient lighting with moving point light sources and cameras to demonstrate the practical applications of our method. They are included as separate mp4 files.

A9. Background

Neural Implicit Surface Volume Rendering Our method is based on neural implicit surface volume rendering. These techniques [45, 51] encode geometry with a Signed Distance Field (SDF), which is represented with a neural network called the geometry network $S_{\Theta_S}(\mathbf{x})$. They then assume the world is semi-transparent, and develop opacity functions, which map SDF values produced by the geometry network at point \mathbf{x} to transparency $w(S_{\Theta_S}(\mathbf{x}))$.

Once transparency is defined, the rest of their system is often similar to NeRF [29]. Rather than modeling the physics of light, they assume the world directly emits light. The amount of light, or radiance, can be represented with a radiance field, where each scene point is assigned distinct radiance values at every outgoing direction. The radiance field is typically implemented with a neural network called the color network $C_{\Theta_C}(\mathbf{x}, \mathbf{v}) \rightarrow \mathbf{c}$, where \mathbf{c} is the radiance and \mathbf{v} is the outgoing direction.

With \mathbf{o} as camera position, t as distance along the camera ray, $\mathbf{p}(t)$ as 3D positions along the camera ray, \mathbf{n} the surface normal (gradient of SDF $S_{\Theta_S}(\mathbf{x})$), \mathbf{f} the feature vector from the geometry network, we can render the pixels of a view $P(\mathbf{o}, \mathbf{v})$ with the following equation. More details can be found in NeuS [45].

$$P(\mathbf{o}, \mathbf{v}) = \int_0^{+\infty} w_{\theta_S}(p(t)) C_{\Theta_C}(\mathbf{p}(t), \mathbf{n}, \mathbf{v}, \mathbf{f}) dt \quad (8)$$

Rendering Equation and Inverse Neural Radiosity. InvNeRad [17] is an inverse rendering method that leverages a technique for self-supervised training of a radiance cache. InvNeRad is built on top of the rendering equation [21], which defines the outgoing radiance of the scene recursively:

$$L(\mathbf{x}, \omega_o) = E(\mathbf{x}, \omega_o) + \int_{\mathcal{H}^2} F(\mathbf{x}, \omega_i, \omega_o) L(r(\mathbf{x}, \omega_i), -\omega_i) d\omega_i^\perp \quad (9)$$

where \mathcal{H}^2 denotes the hemisphere in the direction of surface normal, and $d\omega_i^\perp$ is the differential projected solid-angle measure. \mathbf{x} is a surface point and ω_i, ω_o are the incoming and outgoing directions respectively. E is the emitter radiance distribution. $L(\mathbf{x}, \omega_o)$ is the outgoing radiance. $F(\mathbf{x}, \omega_i, \omega_o)$ is a bidirectional reflectance distribution function (BRDF). $r(\mathbf{x}, \omega_i)$ is the ray tracing operator returning the closest surface intersection of ray (\mathbf{x}, ω_i) .

The equation can be written more compactly in the operator form, where \mathcal{T}_{ϕ} is the light transport operator based on scene parameters ϕ , representing the integral on L .

$$L(\mathbf{x}, \omega_o) = E(\mathbf{x}, \omega_o) + \mathcal{T}_{\phi}(L)(\mathbf{x}, \omega_o) \quad (10)$$

InvNeRad introduces radiance cache L_{θ} represented as a neural network with parameter set θ , and after bouncing the ray only once, it queries the cache to collect the contribution of the rest of the path. To train the radiance cache, InvNeRad is built on the self-supervision technique introduced by NeRad [16], which minimizes the radiometric prior loss. The radiometric prior loss encourages the radiance cache L_{θ} to satisfy the rendering equation by substituting actual radiance L with the radiance cache, and minimizing the difference between left hand side and right hand side of Eq. 10.

$$\mathcal{L}_{\text{prior}}(\theta) = \|L_{\theta}(\mathbf{x}, \omega_o) - (E(\mathbf{x}, \omega_o) + \mathcal{T}_{\phi}(L_{\theta})(\mathbf{x}, \omega_o))\|. \quad (11)$$

A10. Non-Convexity of Co-Located Photometric Stereo

We found that co-located photometric stereo suffers from non convex local minimums. Intuitively, when the reconstructed scene gets larger, there is often significant difference in intensity of each scene point at different scene locations, and such difference in intensity tends to dominate

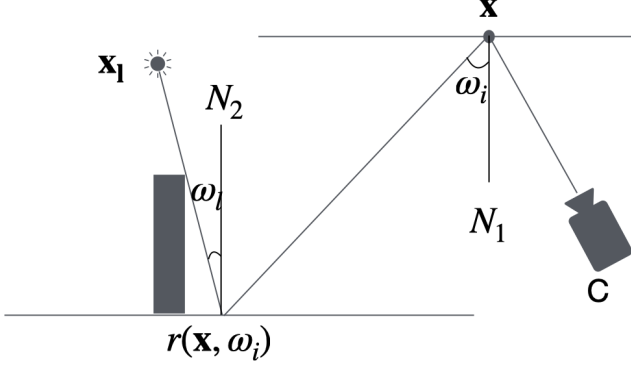


Figure 6. A simple scene for demonstrating continuity and differentiability of the indirect component.

loss, while the error caused by incorrect depth estimates will be relatively small. Therefore, the optimization process will focus on explaining photometric cues, or the effect of shading on surfaces, before moving toward reducing error due to depth estimates. Unlike traditional photometric stereo, photometric cues are fundamentally ambiguous in a co-located light and camera setup, as there is no information about correspondence between pixels in different images. The optimization process will pick a random geometric explanation that is consistent with all photometric cues, and drift far away from correct geometry. By the time error from depth estimates becomes more significant, the optimization process is stuck in a bad local minimum and unable to escape.

A11. Continuity of Indirect Component of Radiance

In this section, we will show the continuity of the indirect component: $\sum_{k=2}^{\infty} \mathcal{T}_{\phi}^k(E)(\mathbf{x}, \omega_o, \mathbf{x}_1)$. If $\sum_{k=2}^{\infty} \mathcal{T}_{\phi}^k(E)(\mathbf{x}, \omega_o, \mathbf{x}_1)$ is continuous, then the term can be approximated with a sufficiently large neural network by the universal approximation theorem.

Our strategy focuses on using mathematical induction to prove every single bounce $\mathcal{T}_{\phi}^k(E)(\mathbf{x}, \omega_o, \mathbf{x}_1)$ is continuous w.r.t. flashlight position for $k \geq 2$.

Before getting started, we make a couple of assumptions. The scene is illuminated by single isotropic point light source. We assume scene geometry \mathbb{S} consists of a finite number of objects, which are 2-manifold smoothly embedded in \mathbb{R}^3 . The physical dimension of the entire scene is finite. We also assume the flashlight energy is finite. We assume flashlight position \mathbf{x}_1 is not in the closure of any object. We differentiate between cast shadows versus attached shadows, which are surfaces oriented away from the light. We model attached shadows as an effect of BRDF $F(\mathbf{x}, \omega_i, \omega_o) = 0$ when $\omega_i \cdot \omega_o < 0$ for non-transparent objects. $\mathbb{V}(\mathbf{x}_1 \leftrightarrow \mathbf{x})$ is the binary visibility function of

light \mathbf{x}_1 from scene point \mathbf{x} , which is 0 if \mathbf{x}_1 is not visible from \mathbf{x} and 1 otherwise. Recall we assume SV-BRDF $F(\mathbf{x}, \omega_i, \omega_o)$ is a continuous function. Therefore, illumination change caused by attached shadows is smooth relative to flashlight positions. When a surface is both occluded and oriented away from light, we model the occluded part as a cast shadow.

The base case is that the second bounce $\mathcal{T}_{\phi}^2(E)(\mathbf{x}, \omega_o, \mathbf{x}_1)$ is continuous and differentiable wrt. flashlight position. For illustrative purposes, we present a simple scene in Fig. 6.

We write down the second bounce only rendering equation at point \mathbf{x} , where L^1 is the radiance of the scene under only direct illumination, \mathbf{x}_1 is the emitter location, and $F(\mathbf{x}, \omega_i, \omega_o)$ is the BRDF function which we assume to be a smooth function. $r(\mathbf{x}, \omega_i)$ is the ray casting operator which finds the first intersection of a ray starting at point \mathbf{x} toward direction ω_i . We denote the direction at $r(\mathbf{x}, \omega_i)$ to light location as ω_l . We denote the surface normal at \mathbf{x} as N_1 , and the surface normal at $r(\mathbf{x}, \omega_i)$ as N_2

$$\mathcal{T}_{\phi}^2(E)(\mathbf{x}, \omega_o, \mathbf{x}_1) = \int_{\mathcal{H}^2} F(\mathbf{x}, \omega_o, \omega_i)(\omega_i \cdot N_1) \frac{\mathbb{V}(r(\mathbf{x}, \omega_i) \leftrightarrow \mathbf{x}_1) F(r(\mathbf{x}, \omega_i), \omega_i, \omega_l)(\omega_l \cdot N_2)}{\|\mathbf{x}_1 - r(\mathbf{x}, \omega_i)\|^2} d\omega_i^{\perp} \quad (12)$$

To make the notation more compact, we define $f(\mathbf{x}, \omega_o, \omega_i, \mathbf{x}_1)$ as the following.

$$f(\mathbf{x}, \omega_o, \omega_i, \mathbf{x}_1) = F(\mathbf{x}, \omega_o, \omega_i)(\omega_i \cdot N_1) \frac{F(r(\mathbf{x}, \omega_i), \omega_i, \omega_l)(\omega_l \cdot N_2)}{\|\mathbf{x}_1 - r(\mathbf{x}, \omega_i)\|^2} \quad (13)$$

$$\mathcal{T}_{\phi}^2(E)(\mathbf{x}, \omega_o, \mathbf{x}_1) = \int_{\mathcal{H}^2} \mathbb{V}(r(\mathbf{x}, \omega_i) \leftrightarrow \mathbf{x}_1) f(\mathbf{x}, \omega_o, \omega_i, \mathbf{x}_1) d\omega_i^{\perp} \quad (14)$$

To show the continuity of $\mathcal{T}_{\phi}^2(E)(\mathbf{x}, \omega_o, \mathbf{x}_1)$ in light position, we want to show that for any sequence $\mathbf{x}_{1n} \rightarrow \mathbf{x}_1$, we have the following.

$$\int_{\mathcal{H}^2} \mathbb{V}(r(\mathbf{x}, \omega_i) \leftrightarrow \mathbf{x}_{1n}) f(\mathbf{x}, \omega_o, \mathbf{x}_{1n}) d\omega_i^{\perp} \rightarrow \int_{\mathcal{H}^2} \mathbb{V}(r(\mathbf{x}, \omega_i) \leftrightarrow \mathbf{x}_1) f(\mathbf{x}, \omega_o, \mathbf{x}_1) d\omega_i^{\perp} \quad (15)$$

However, since $f(\mathbf{x}, \omega_o, \mathbf{x}_1)$ is continuous, we will be focusing on the behavior of $\mathbb{V}(\mathbf{x}_1 \leftrightarrow \mathbf{x}')$ under the sequence. Finally we will examine the behavior of the integral.

Our overall goal is to show the set of points where the visibility function can change is only the cast shadow boundary as $\mathbf{x}_{1n} \rightarrow \mathbf{x}_1$.

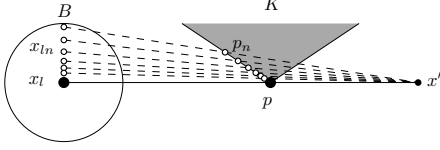


Figure 7. In unoccluded case, the light visibility does not change except the boundary.

Here we introduce a couple of definitions that will be used throughout the proof. Consider a point $\mathbf{x}' = r(\mathbf{x}, \omega_i)$ reachable from secondary bounce location \mathbf{x} in direction ω_i . Consider the points along the ray segment $\mathbf{x}_1 - \mathbf{x}'$:

$$p(\mathbf{x}_1, \mathbf{x}', t) = \mathbf{x}_1 + t(\mathbf{x}' - \mathbf{x}_1) \quad (16)$$

Lemma A11.1. *Assume \mathbf{x}' is not on the shadow boundary of any object in the scene with respect to the light source \mathbf{x}_1 . Additionally suppose that \mathbf{x}' is unoccluded. We wish to prove that there exists a neighborhood around \mathbf{x}_1 such that if \mathbf{x}_1 is perturbed to a position in that neighborhood, \mathbf{x}' is still unoccluded from its perspective.*

Proof. Suppose to the contrary. Then there exists a sequence of positions \mathbf{x}_{1n} which converge to \mathbf{x}_1 such that the ray from each \mathbf{x}_{1n} to \mathbf{x}' always passes through some object in the scene. Since \mathbf{x}_1 is not in the closure of any object, there exists a ball B around \mathbf{x}_1 of positive radius that does not intersect the geometry of the scene. Hence an infinite amount of these \mathbf{x}_{1n} lie in that ball B . Consider only the \mathbf{x}_{1n} in this ball. Since the scene only has a finite amount of objects, an infinite amount of the occluders defined for the \mathbf{x}_{1n} must be the same object. That is to say, there exists an object K such that for an infinite amount of the \mathbf{x}_{1n} the ray from \mathbf{x}_{1n} to \mathbf{x}' always hits this object. Note this object must be strictly between \mathbf{x}_1 and \mathbf{x}' since we are only considering the \mathbf{x}_{1n} in B . Now consider only the \mathbf{x}_{1n} whose ray to \mathbf{x}' hits K . Let the sequence of points they first hit on K be p_n .

Since p_n is on a compact manifold smoothly embedded in R^3 ,

Since the p_n are on a compact space and get infinitely close to the segment from \mathbf{x}_1 to \mathbf{x}' , by the compactness of K there must be some point p on K along this line. Hence p intersects the line segment from \mathbf{x}_1 to \mathbf{x}' , and so \mathbf{x}' must be on the shadow boundary. This contradicts what we assumed about \mathbf{x}' . See Figure 7 for an illustration. \square

Lemma A11.2. *Assume \mathbf{x}' is not on the shadow boundary of any object in the scene with respect to the light source \mathbf{x}_1 . Additionally suppose that \mathbf{x}' is occluded. We wish to prove that there exists a neighborhood around \mathbf{x}_1 such that if \mathbf{x}_1 is perturbed to a position in that neighborhood, \mathbf{x}' is still occluded from its perspective.*

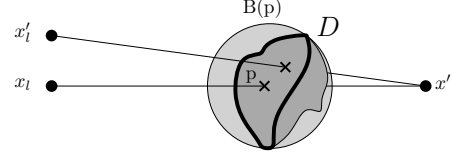


Figure 8. In occluded case, the light visibility does not change except the boundary.

Proof. Since \mathbf{x}' is occluded but not on the shadow boundary of any object in the scene with respect to the light source \mathbf{x}_1 , then the line segment from \mathbf{x}_1 to \mathbf{x}' intersects the relative interior of some surface K at a point, p , strictly between \mathbf{x}' and \mathbf{x}_1 .

Since p is in the relative interior of K then there exists a closed ball around p , $B(p)$, whose radius is smaller than $\frac{1}{2} \min(d(\mathbf{x}_1, p), d(p, \mathbf{x}'))$, such that $D = B(p) \cap K$ is in the relative interior of K . Since the topological circle ∂D (the relative boundary of D) is compact, let θ be the minimum angle $\angle \mathbf{x}_1 \mathbf{x}' y$ makes for any y on ∂D . Note that $\theta > 0$ because p is the unique point for which it is equal zero and p is not on ∂D .

Choose $B(\mathbf{x}_1)$ to be a ball around \mathbf{x}_1 such that it has small enough radius so as not to intersect D and such that for any z in $B(\mathbf{x}_1)$, $\angle \mathbf{x}_1 \mathbf{x}' z$ makes a smaller angle than θ . Since for all $\mathbf{x}'_1 \in B(\mathbf{x}_1)$, $\angle \mathbf{x}_1 \mathbf{x}' \mathbf{x}'_1 < \theta$, then the segment from \mathbf{x}'_1 to \mathbf{x}' passes through the hole of the topological circle ∂D and hence intersects D before it reaches \mathbf{x}' , leaving \mathbf{x}' occluded. And so if \mathbf{x}_1 is perturbed in that ball $B(\mathbf{x}_1)$, \mathbf{x}' is still occluded from it by D . See Figure 8 for an illustration. \square

With Lemma A11.1 and Lemma A11.2, we can show the second bounce radiance is indeed continuous. Denote S as the shadow boundary. We conclude for any sequence $\mathbf{x}_{1n} \rightarrow \mathbf{x}_1$, we have the following.

$$\forall \mathbf{x}' \notin S \quad \mathbb{V}(\mathbf{x}_{1n} \leftrightarrow \mathbf{x}') \rightarrow \mathbb{V}(\mathbf{x}_1 \leftrightarrow \mathbf{x}') \quad (17)$$

First we note $f(\mathbf{x}, \omega_o, \omega_i, \mathbf{x}_{1n})$ is continuous. Moreover, since $\mathbb{V}(r(\mathbf{x}, \omega_i) \leftrightarrow \mathbf{x}_1) f(\mathbf{x}, \omega_o, \omega_i, \mathbf{x}_1)$ is bounded (all its components are bounded) by a constant, say by M , we can define $g(\mathbf{x}, \omega_o, \omega_i) = M$, and $|f(\mathbf{x}, \omega_o, \omega_i, \mathbf{x}_1)| < g(\mathbf{x}, \omega_o, \omega_i)$. By the dominated convergence theorem, as

$\mathbf{x}_{\text{In}} \rightarrow \mathbf{x}_1$, we have the following.

$$\begin{aligned}
& \mathcal{T}_\phi^2(E)(\mathbf{x}, \omega_o, \mathbf{x}_{\text{In}}) \\
&= \int_{\mathcal{H}^2} \mathbb{V}(\mathbf{x}_{\text{In}} \leftrightarrow \mathbf{x}') f(\mathbf{x}, \omega_o, \omega_i, \mathbf{x}_{\text{In}}) d\omega_i^\perp \\
&= \int_{\mathcal{H}^2 \setminus S} \mathbb{V}(\mathbf{x}_{\text{In}} \leftrightarrow \mathbf{x}') f(\mathbf{x}, \omega_o, \omega_i, \mathbf{x}_{\text{In}}) d\omega_i^\perp \\
&\rightarrow \int_{\mathcal{H}^2 \setminus S} \mathbb{V}(\mathbf{x}_1 \leftrightarrow \mathbf{x}') f(\mathbf{x}, \omega_o, \omega_i, \mathbf{x}_1) d\omega_i^\perp \\
&= \int_{\mathcal{H}^2} \mathbb{V}(\mathbf{x}_1 \leftrightarrow \mathbf{x}') f(\mathbf{x}, \omega_o, \omega_i, \mathbf{x}_1) d\omega_i^\perp \\
&= \mathcal{T}_\phi^2(E)(\mathbf{x}, \omega_o, \mathbf{x}_1)
\end{aligned}$$

We conclude $\mathcal{T}_\phi^2(E)(\mathbf{x}, \omega_o, \mathbf{x}_1)$ is continuous with respect to flashlight position \mathbf{x}_1 , which is our base case for our proof by mathematical induction.

For the inductive case on the N -th bounce, we can assume the $(N-1)$ -th bounce is continuous. The radiance is the following.

$$\mathcal{T}_\phi^N(E)(\mathbf{x}, \omega_o, \mathbf{x}_1) = \int_{\mathcal{H}^2} F(\mathbf{x}, \omega_i, \omega_o) \mathcal{T}_\phi^{N-1}(E)(\mathbf{x}, \omega_o, \mathbf{x}_1) d\omega_i^\perp \quad (18)$$

It is easy to notice $\mathcal{T}_\phi^N(E)(\mathbf{x}, \omega_o, \mathbf{x}_1)$ is the integration of a continuous brdf F and continuous function $\mathcal{T}_\phi^{N-1}(E)(\mathbf{x}, \omega_o, \mathbf{x}_1)$. Therefore, we conclude $\mathcal{T}_\phi^N(E)(\mathbf{x}, \omega_o, \mathbf{x}_1)$ is continuous.

A12. Additional Comparisons

In this section, we provide additional comparisons against state of the art methods.

A12.1. Measured Albedo in the Wild 2.0

Method	Albedo Intensity (MSE $\times 10^2$)			Albedo Chromaticity (Delta-E)		
	Coffee table	Shoe rack	Window still	Coffee table	Shoe rack	Window still
IRGS	<u>0.09</u>	<u>0.18</u>	1.56	9.1714	3.2322	9.0345
NeRO	4.86	0.31	1.47	4.7499	7.5724	4.0790
IRON	0.14	3.75	<u>0.54</u>	4.9089	5.1342	4.0915
WildLight	0.85	0.26	1.01	3.7362	<u>1.3104</u>	<u>3.8298</u>
Ours	0.07	0.07	0.10	6.1491	1.0236	2.6620

Table 4. Quantitative evaluation of albedo on real data following measured albedo in the wild [48]. We report albedo intensity, which measures intensity difference with ground-truth albedo in the grayscale component, and albedo chromaticity, which measure chromaticity difference with ground-truth albedo.

To quantitatively compare the quality of albedo on real data, we follow measured albedo in the wild (MAW) [48] to collect ground-truth measured albedo labels on our real datasets *Coffee Table*, *Shoe Rack* and *Window Sill*.

MAW contains few images per scene and is better suited to evaluate single image inverse rendering techniques. On the other hand, our dataset contains densely captured images for each scene and is suitable for evaluating multi-view inverse rendering methods. The measurements and annotations will be released together with our dataset upon acceptance.

Following MAW, we evaluate algorithms with albedo intensity in MSE and albedo chromaticity in Delta-E, and report the performance in Tab. 4. We found that similar to the results on synthetic data, our method significantly outperforms natural illumination baselines in albedo estimation, highlighting the importance of co-located light & camera setup in constraining the ambiguous inverse rendering problem.

A12.2. Surface Angle Weighting loss

In Fig. 9, we compare our surface angle weighting loss to state-of-the-art loss weighting schemes, which includes Ref-NeuS Loss [11] and Adaptive Huber loss from neural-pbir [42].

Ref-NeuS [11] is based on variation between corresponding pixels across images, and is unsuitable to use on co-located light and camera capture setup, as pixel intensity changes due to different light position. Therefore, we compare Ref-NeuS on natural illumination against our geometry initialization stage (stage 1) on co-located light and camera setup. We found Ref-NeuS tend to reconstruct concave regions poorly, potentially due to down-weighting of errors in such regions.

We also compare our full stage 1 with a variant without such surface angle weighting loss, we found the variant without such surface angle weighting loss produce artifacts in regions with specular inter-reflections, while our surface angle weighting loss prevents such artifacts. We also included Adaptive Huber from Neural PBIR [42] for comparison, which we found is unable to prevent these artifacts.

A12.3. Natural Illumination Methods Material Comparisons on Real Data

In Fig. 10, we show comparisons against state of art natural illumination methods NeRO [25] and IRGS [14] on real data. We found that our method significantly outperforms the natural illumination baselines, especially in material properties.

A12.4. Natural Illumination Methods Geometry Comparisons

We compare the geometry of our method with natural illumination methods in Fig. 10. We do not claim we achieve better than state-of-the-art natural illumination geometry, but only comparable.

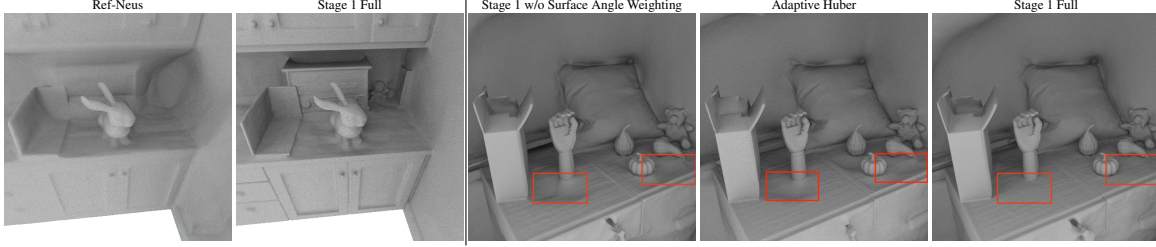


Figure 9. **Ablation of surface angle weighting.** Left: we compare Ref-NeuS with our geometry initialization stage. Ref-NeuS loss is unsuitable for co-located light & camera setup, so Ref-NeuS is trained on natural illumination. Ref-NeuS tends to produce artifacts in concave regions, potentially due to downweighting in these regions. Right: Red box highlights areas where specular inter-reflection causes artifacts in geometry without surface angle weighting. Such errors become significantly less pronounced in our full geometry initialization stage with surface angle weighting loss. Adaptive Huber from neural-pbir [42] does not prevent such artifacts.

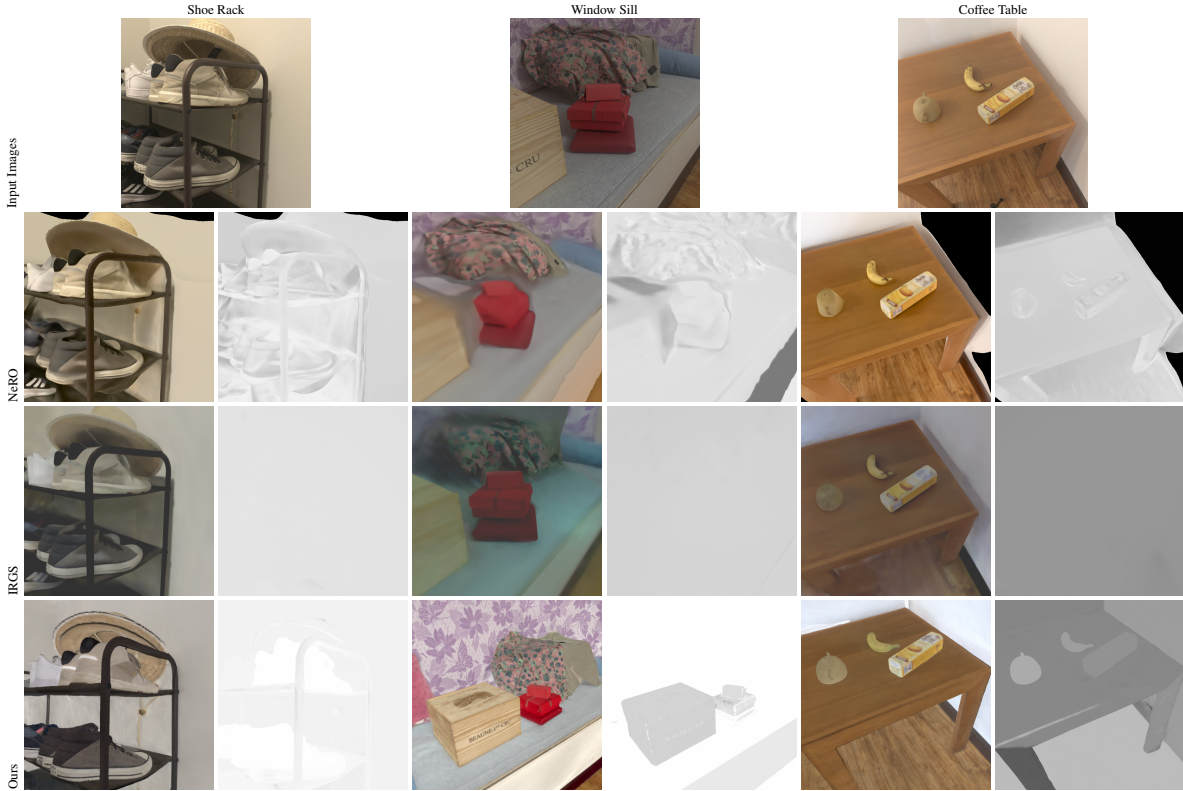


Figure 10. **Qualitative comparison of reflectance estimation on real scenes.** We present estimated albedo, and roughness in validation views for the real scenes. Our method produces significantly better albedo, roughness w.r.t. natural illumination methods as co-located capture setup provides additional constraints. (Zoom in for better visualization.)

A12.5. Qualitative Comparisons on re-rendering

In Fig. 12, we visualize the re-rendering performance of colocated and natural illumination methods. We found all methods perform well in re-rendering as long as geometry is reasonably reconstructed. Nevertheless, natural illumination methods often are unable to recover accurate material properties due to inherent ambiguity in inverse rendering.

A12.6. WildLight Capture Setup

In the main paper, we show WildLight results as a darkroom co-located light and camera method. In Figure 13, we show qualitative results of WildLight on WildLight style capture setup (co-located light and camera under ambient natural illumination). We found WildLight [7] still fail to converge under prominent inter-reflections.

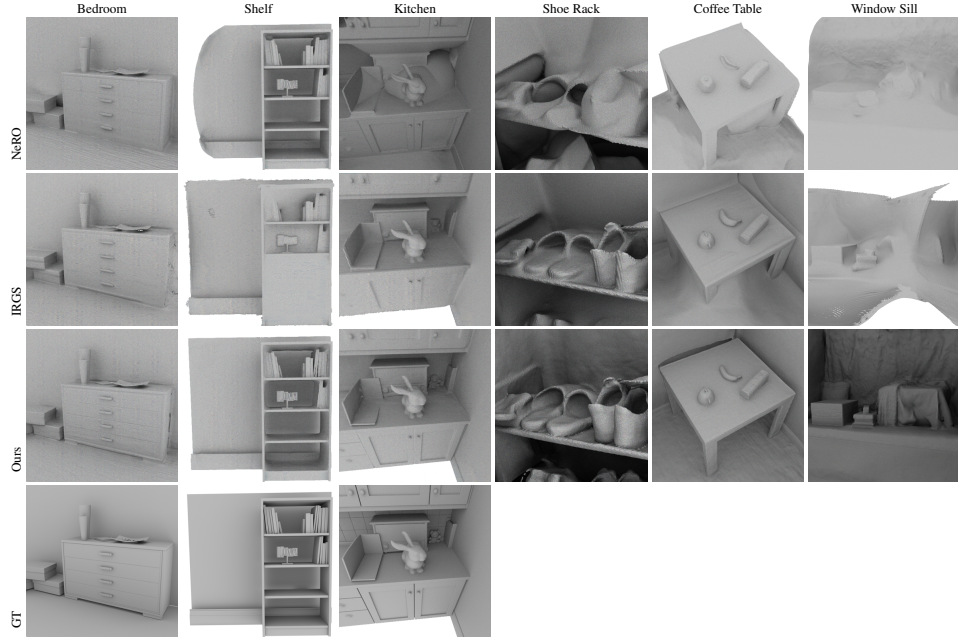


Figure 11. **Qualitative comparison of geometry with co-located light & camera methods.** The geometry of our method is comparable or better than natural illumination methods.

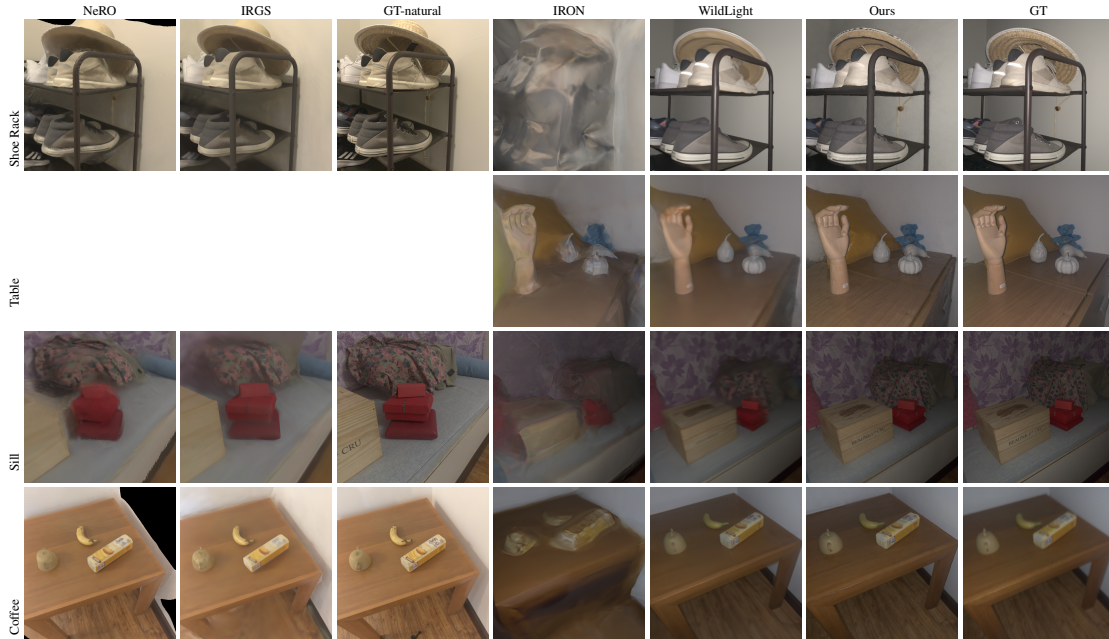


Figure 12. **Qualitative comparison of re-rendering on real scenes.** We present re-rendering in validation views for the real scenes. Our method produces better re-rendering w.r.t. IRON [55] and WildLight [7] as we are able to model global illumination. Our re-rendering is comparable to natural illumination methods, but co-located setup allow us to recover better material properties.

A12.7. Additional Views in Dataset

Fig. 18, Fig. 19, Fig. 20.

We show additional views from our dataset on re-rendering, albedo and roughness in Fig. 14, Fig. 15, Fig. 16, Fig. 17,

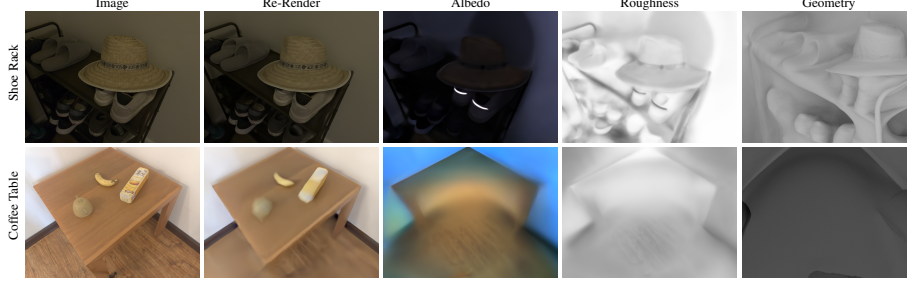


Figure 13. Qualitative results of WildLight on Co-located Light and Camera under ambient natural illumination.

A13. Implementation Details

Here we described additional details of implementation.

A13.1. Stage 1 Geometry Initialization

With our synthetic dataset, the images can contain “background pixels” where the primary ray from the camera does not intersect with any scene geometry during rendering. Since we are using environment map for our scenes, the values of these pixels are undefined, and we use a per image binary mask to ignore these pixels during training. Consequently, we do not put any supervision in the background region. To prevent the network from producing arbitrary values for the background, we adopt mask loss commonly used by prior works [45, 50]. NeuS [45] defines unbiased weights $w_{k,i}$ along the k -th camera ray based on the underlying signed distance field. Denote $\hat{O}_k = \sum_{i=1}^n w_i$ as the sum of weights along the k -th camera ray, $M_k = \{0, 1\}$ as the value of the binary mask on the k -th pixel, and BCE as the binary cross entropy loss, we have the following equation.

$$L_{\text{mask}} = \text{BCE}(M_k, \hat{O}_k) \quad (19)$$

Such mask loss is only used for synthetic data, and not used for real data.

Stage 1 is trained with ADAM optimizer, batch size of 512 rays, learning rate of 5×10^{-4} .

A13.2. Stage 2 Physically Based Rendering

To integrate the outer integral, we need to sample N 3d points $p(t)$ along camera ray, which is importance sampled similar to NeuS[45]. For every sampled points, we evaluate the light transport operator \mathcal{T}_ϕ by sampling M incoming light directions w_i through bsdf sampling [36]. Finally, the radiance cache L_θ is queried at intersection points $L_\theta(r(p(t), \omega_i))$. In our experiments, we set N to be 128 as NeuS, and M to be 1.

As noted by Dejan Azinovic [1], naive automatic differentiation gives incorrect result, and it is important to draw independent samples for Monte Carlo estimator of gradient

of reconstruction loss against output $\frac{\partial \mathcal{L}_{\text{recons}}}{\partial y}$ and gradient of output against parameters $\frac{\partial \hat{y}}{\partial \phi}$, which we follow.

When we use principled BRDF [6]. We set all fixed parameters to zeros, except for “specular” (which sometimes called “specular albedo”), which we set to 0.6 for synthetic scenes and 0.5 for real scenes.

Stage 2 is trained with ADAM optimizer at batch size of 512 rays, learning rate of 5×10^{-5} . We downscale the gradient against SDF network S_{θ_S} by 10^{-1} as we found it improves stability.

Since at the beginning of stage 2, only the geometry is initialized, but not the material properties and radiance cache. To initialize the material properties and radiance cache, we freeze the geometry at the beginning of stage 2 until the material properties and radiance cache become initialized, then we train all parameters jointly.

Sampling Radiometric Prior Loss One important design decision of the radiometric prior loss in the inverse rendering system is how to sample points for evaluating the radiometric prior loss. We evaluate on all points $p(t)$ sampled along the camera ray during volume rendering.

We also include an extra bounce prior similar to InvNeRad. To adapt such loss for volume geometry while keeping computational cost manageable, we only take 1 point out of points sampled along primary camera ray during physically based volume rendering, and sample $N = 128$ points along its income light direction, we we apply extra bounce prior loss on these sampled points.

To ensure thorough coverage of the radiance cache at all location under every light location, when calculating the extra bounce prior radiosity loss, we evaluate both the radiance cache and direct/indirect illumination term under a randomly sampled light location from training data.

A13.3. Stage 3 Material Optimization

Here we provide additional details regarding the final stage of our system. We optimize for BRDF on fixed mesh geometry extracted from geometry network from stage 2. The algorithm is similar to invNeRAD [17], where we intersect rays from camera with mesh geometry of the scene, and render under radiance cache global illumination. However,

we use our dynamic radiance cache instead of naive radiance cache, and the dynamic radiance cache is trained same as stage 2, except instead of evaluating the loss on volume points, we can only evaluate on surface points.

We train our final with ADAM optimizer at learning rate of 5×10^{-4} with batch size of 512.

A14. Real Data Capture Setup

We capture all of our real data using an iPhone XS Max and an iPhone 11 Pro. We capture all the image with ProCamera app on iOS as raw dng file. During capture, we keep manual and fixed white balance, focus and exposure. For co-located capture, we also keep flashlight constantly on through the capture session. We process the raw files with RawPy [38], which is a python interface around libraw [26]. We perform structure-from-motion reconstruction using pixel perfect sfm [24]. We apply camera undistortion parameters estimated by pixel perfect sfm to our captured images. We found that both iPhone XS Max and iPhone 11 Pro experience significant vignetting. To calibrate for vignetting, we use a piece of white paper on a sunny day under direct sunlight as the calibration target. We model the vignetting as 6-th degree even order polynomial [13], and apply vignetting correction accordingly. We store all final processed images as 16-bit unsigned png images with linear response curve without any gamma curve applied, which are used for all following experiments.

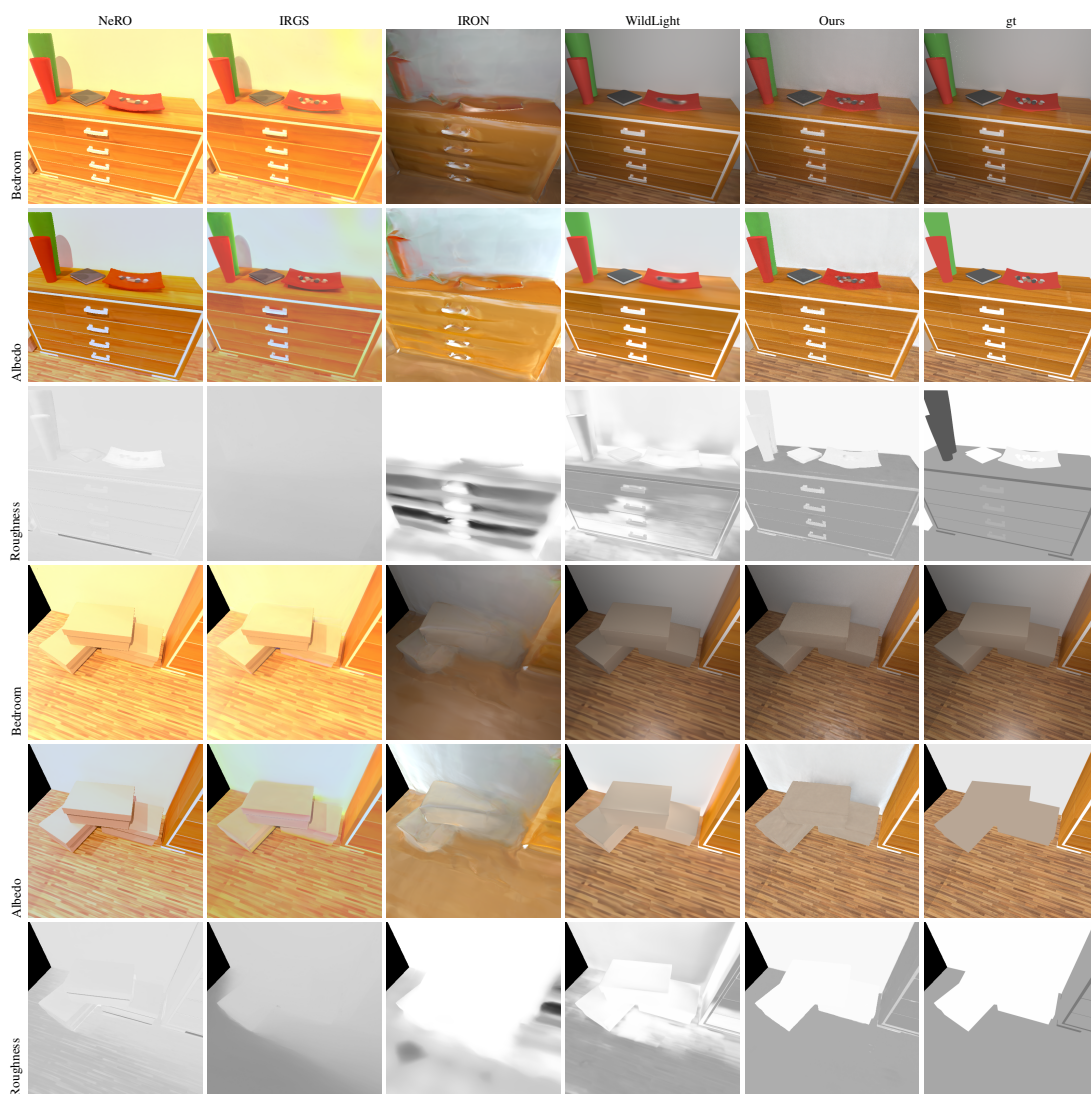


Figure 14. Qualitative comparison on synthetic scene *bedroom*.

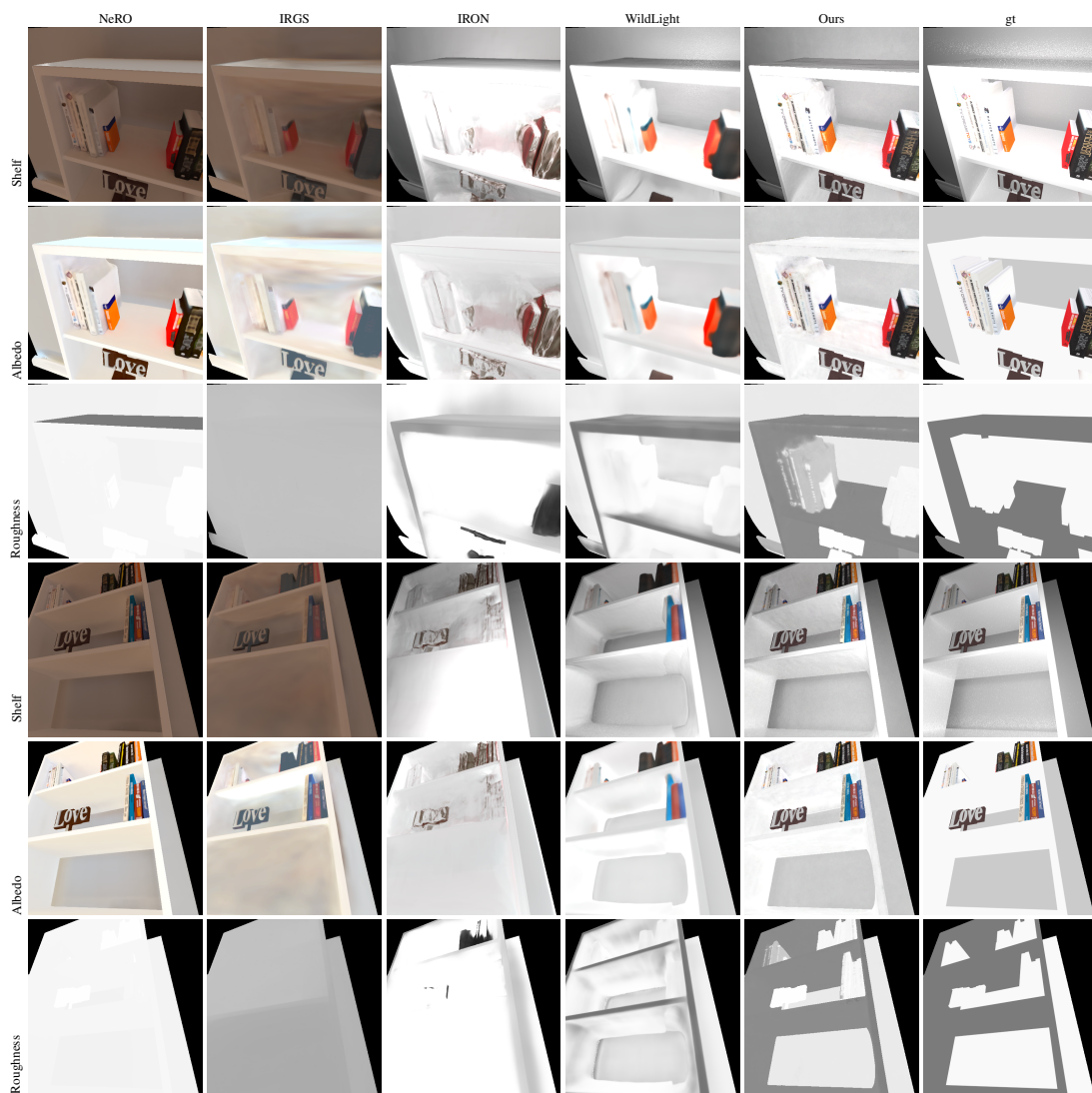


Figure 15. Qualitative comparison on synthetic scene *shelf*.



Figure 16. Qualitative comparison on synthetic scene *kitchen counter*.

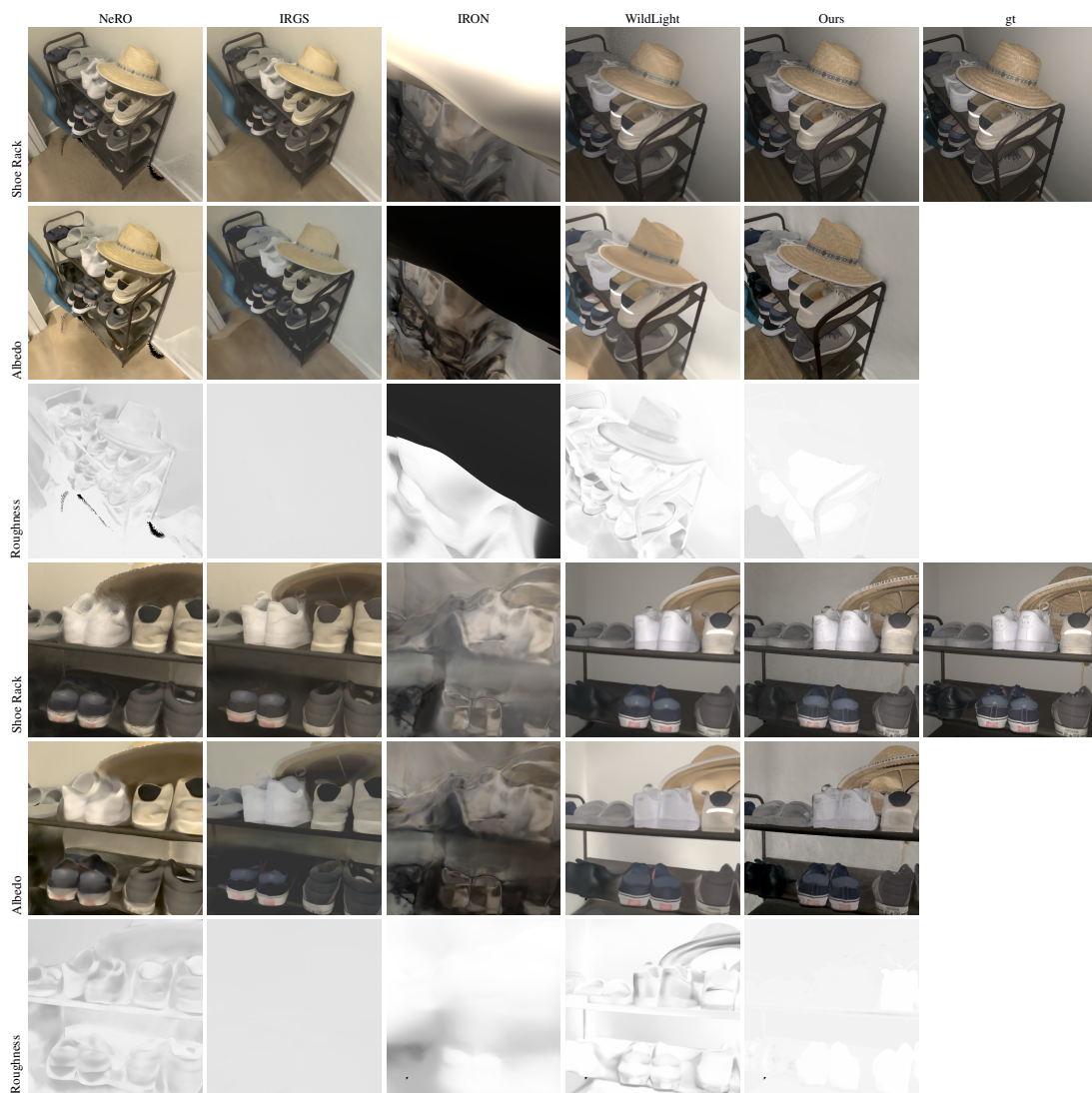


Figure 17. Qualitative comparison on real scene *shoe rack*.

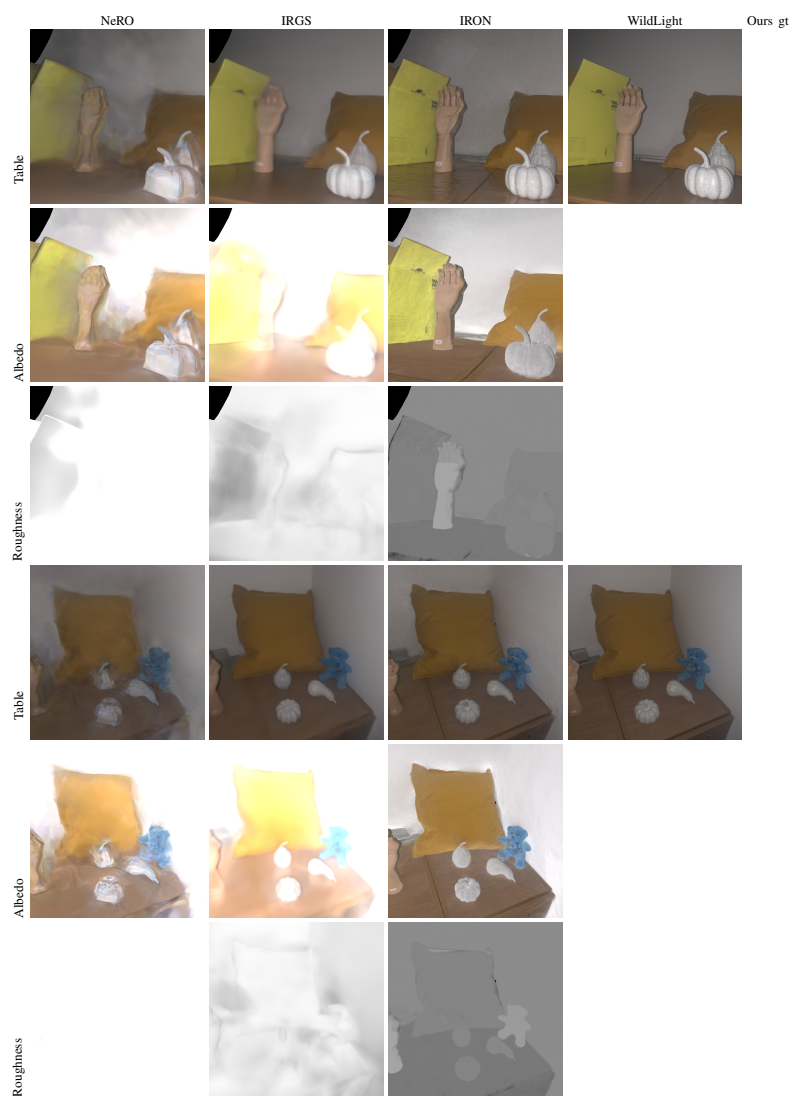


Figure 18. Qualitative comparison on real scene *table*.



Figure 19. Qualitative comparison on real scene *window sill*.

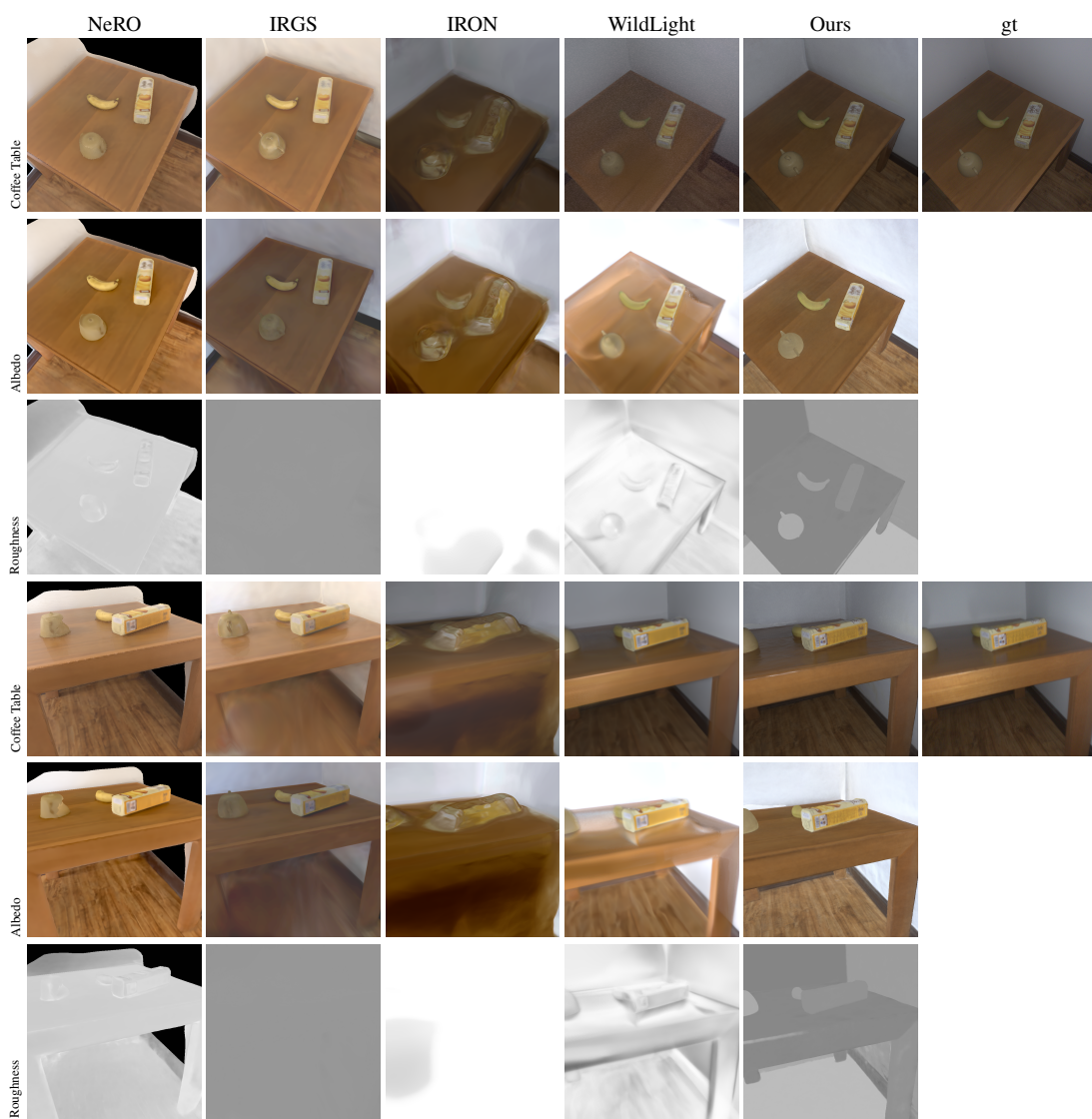


Figure 20. Qualitative comparison on real scene *coffee table*.