

Learning Relighting and Intrinsic Decomposition in Neural Radiance Fields

Yixiong Yang¹, Shilin Hu², Haoyu Wu², Ramon Baldrich¹, Dimitris Samaras², Maria Vanrell¹

¹ Universitat Autònoma de Barcelona, ² Stony Brook University

{yixiong, ramon, maria}@cvc.uab.cat, {shilhu, haoyuwu, samaras}@cs.stonybrook.edu

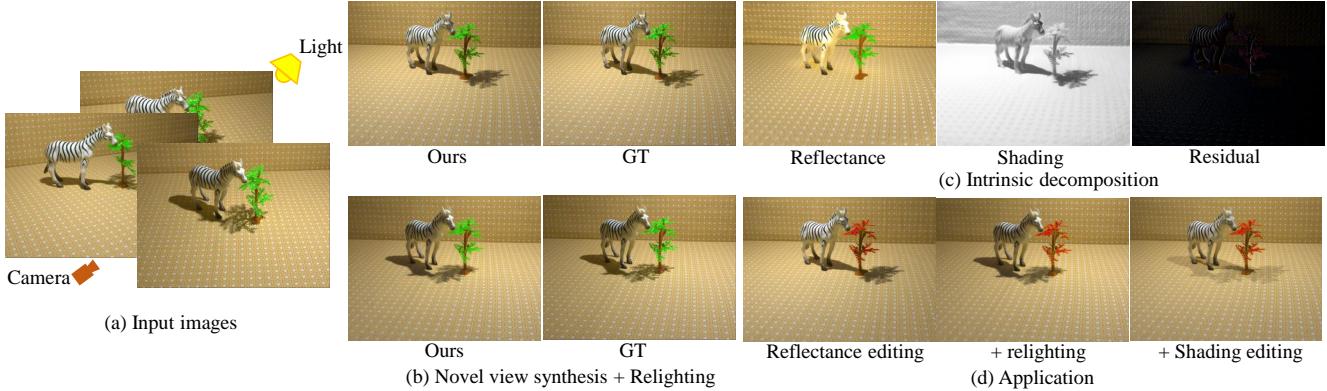


Figure 1. Given images under different camera pose and light positions (a), our method learns the neural radiance fields that enable novel view synthesis and relighting (b), and intrinsic decomposition (c) simultaneously. Other image editing applications (d), such as reflectance editing, reflectance editing + relighting, and shading editing (simulating two lights), can also be employed.

Abstract

The task of extracting intrinsic components, such as reflectance and shading, from neural radiance fields is of growing interest. However, current methods largely focus on synthetic scenes and isolated objects, overlooking the complexities of real scenes with backgrounds. To address this gap, our research introduces a method that combines relighting with intrinsic decomposition. By leveraging light variations in scenes to generate pseudo labels, our method provides guidance for intrinsic decomposition without requiring ground truth data. Our method, grounded in physical constraints, ensures robustness across diverse scene types and reduces the reliance on pre-trained models or hand-crafted priors. We validate our method on both synthetic and real-world datasets, achieving convincing results. Furthermore, the applicability of our method to image editing tasks demonstrates promising outcomes.

1. Introduction

Recent advances in neural rendering have made significant strides in novel view synthesis [20, 24], ranging from small objects to large-scale scenes. Concurrently, there has been an exploration towards scene editing [36], such as recoloring [39] and relighting [21, 40]. To facilitate editing, it often becomes necessary to decompose scenes into editable

sub-attributes. Within the task of scene decomposition into geometry, reflectance, and illumination using neural rendering, two lines of work are particularly noteworthy: inverse rendering and intrinsic decomposition.

The first approach [15, 38, 42, 43] integrates inverse rendering with neural rendering methods for scene decomposition. They often employ the BRDF model [3], to model material properties and jointly optimize geometry, BRDF, and environmental lighting. However, inverse rendering presents a highly ill-posed challenge: separating material properties and illumination in images often yields ambiguous results, and tracing light within scenes is computationally intensive. These factors limit inverse rendering to object-specific scenarios.

The second approach [39], based on intrinsic decomposition [2], aims to provide an interpretable representation of a scene (in terms of reflectance and shading) suitable for image editing. It can be considered a simplified variant of inverse rendering, making it more applicable to a broader range of scenarios, including individual objects and more complex scenes with backgrounds. However, despite simplifications over inverse rendering, previous attempts at applying intrinsic decomposition to neural rendering have shown limited success. This motivates our work in this paper.

Our inspiration is drawn from the idea of using neural rendering to combine relighting and intrinsic decomposi-

tion, aiming not only to enhance the quality of intrinsic decomposition but also to expand editing capabilities. Just as experts in mineral identification illuminate specimens from various angles to reveal their features, varying light source positions are essential for uncovering a scene’s intrinsic details. In fact, the connection between relighting and intrinsic decomposition has been discussed in previous works on 2D images [17, 18], but it has yet to be explored in neural rendering. Additionally, the field of neural rendering has significantly explored relighting [31, 40]. While IntrinsicNeRF [39] has pioneered the integration of intrinsic decomposition within NeRF, they have not utilized relighting or fully leveraged the 3D information available through neural rendering. Instead, we focus on physics-based constraints to enhance the intrinsic decomposition performance.

In this paper, we propose a two-stage method. In the first stage, we train a neural implicit radiance representation to enable novel view synthesis and relighting. Based on the results of this stage, we calculate normals and light visibility for each training image, which allows us to develop a method for generating pseudo labels for reflectance and shading. In the second stage, we treat reflectance and shading as continuous functions parameterized by Multi-Layer Perceptrons (MLPs). During training, we apply constraints based on physical principles and our pseudo labels. Notably, our approach does not depend on any pre-trained models or ground truth data for intrinsic decomposition, yet achieves convincing results, as shown in Fig. 1. Our contributions are summarized as follows:

- We propose a method that integrates relighting with intrinsic decomposition, allowing for novel view synthesis, lighting condition altering, and reflectance editing.
- We propose a method to generate pseudo labels for reflectance and shading through neural fields that integrate multiple lighting conditions.
- Our method, applied to NeRF scenes, operates free from data-driven priors. It factorizes the scene into reflectance, shading, and a residual component, proving effective even in the presence of strong shadows.

2. Related Work

Intrinsic decomposition. Intrinsic decomposition is a classical challenge in computer vision [2], with much of the previous research focused on the 2D image[1, 4, 6, 19]. A key difficulty in this area is the scarcity of real datasets, which need complicated and extensive annotation. This limitation has spurred interest in semi-supervised and unsupervised techniques[17, 18, 22]. IntrinsicNeRF [39] has been a pioneer in applying intrinsic decomposition to neural rendering. Similar to previous unsupervised methods in 2D, it utilizes hand-crafted constraints, including chromaticity and semantic constraints, for guidance. However, these constraints do not accurately reflect physical principles and of-

ten fall short in complex scenarios. Our approach leans on 3D information and physical constraints (e.g., variations in illumination) to achieve superior results.

Relighting. Relighting has recently garnered attention from various perspectives within the field [7]. Data-driven approaches have been explored, with research focusing on portrait scenes [14, 27, 28, 33, 44] and extending to more complex scenarios [8, 13, 26, 30, 35]. Kocsis et al. [16] have also investigated lighting control within diffusion models, enabling the generation of scenes under varying lighting conditions. Meanwhile, relighting has also received widespread attention within the field of neural rendering [10, 32, 34, 40], achieving impressive relighting outcomes within individual scenes.

3. Method

Under the Lambertian assumption, images can be decomposed into reflectance and shading components [2, 4, 9]. However, real-world scenes often require a residual term to account for discrepancies [11, 39]. Thus, we model intrinsic decomposition as follows:

$$I(i, j) = R(i, j) \odot S(i, j) + Re(i, j) \quad (1)$$

where R , S and Re denote Reflectance, Shading and Residual, respectively.

Our method extends implicit neural representation for relighting and intrinsic decomposition. We propose a two-stage approach, illustrated in Fig. 2. In the first stage, we train our model to represent scenes under varying camera positions and lighting conditions, enabling novel view synthesis and relighting. We then apply three steps to generate pseudo labels for reflectance and shading. In the second stage, we expand the model to decompose intrinsics using these pseudo labels as constraints. Our proposed model achieves novel view synthesis, relighting, and intrinsic decomposition simultaneously.

3.1. Stage 1: Learning to Relight

We use 3D hash grids [20, 25] to represent geometry and a small MLP to model color, which accepts the light position as an input. We illustrate Stage 1 in Fig. 2 (top-left), with formulas as follows:

$$sdf = f(\mathbf{x}), \quad \mathbf{c} = \text{MLP}_{color}(\mathbf{x}, \mathbf{d}, \mathbf{l}, \mathbf{feat}) \quad (2)$$

where $f(\cdot)$ is the geometry network that predicts Signed Distance Function and $\text{MLP}_{color}(\cdot)$ is the color network. \mathbf{x} is the spatial position, \mathbf{d} is the view direction, \mathbf{l} is the light position, and \mathbf{feat} is the feature from SDF network. Following [20], the loss for Stage 1 is:

$$\mathcal{L}_{S1} = \mathcal{L}_{RGB} + w_{eik}\mathcal{L}_{eik} + w_{curv}\mathcal{L}_{curv} \quad (3)$$

where \mathcal{L}_{RGB} is the loss of the rendered image, \mathcal{L}_{eik} represents the Eikonal loss [12], and \mathcal{L}_{curv} is the curvature loss. The terms w_{eik} and w_{curv} are the corresponding weights.

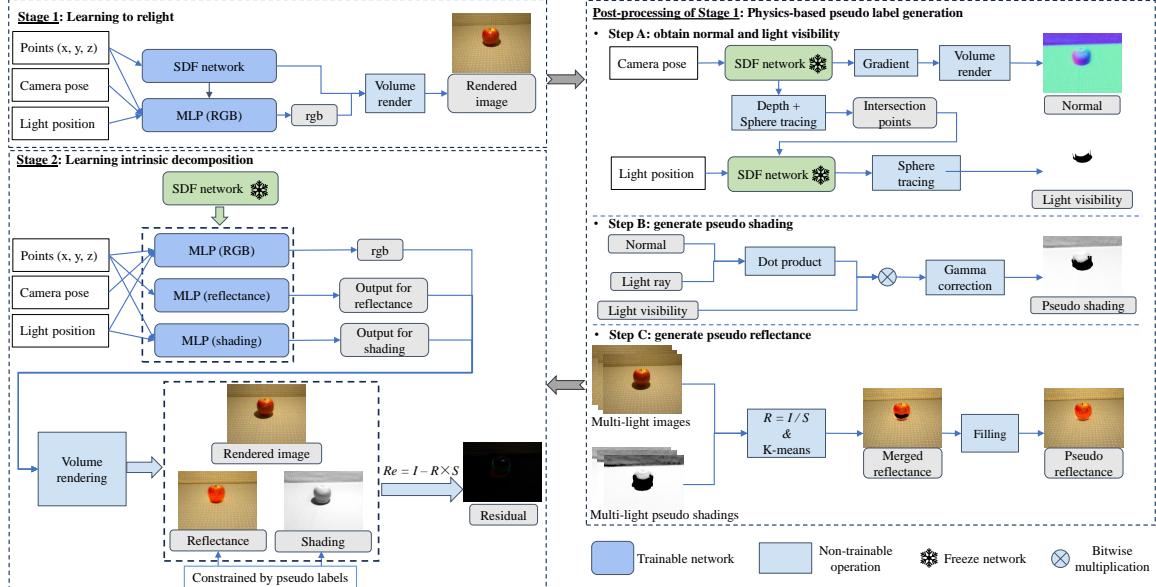


Figure 2. Method Framework: Stage 1 involves learning the neural field with relighting (top left). Post-processing and generating pseudo labels (right). In Stage 2, the learning process continues to learn intrinsic decomposition based on the model trained in Stage 1 and the pseudo labels (bottom left).

3.2. Physics-based Pseudo Label Generation

Our proposed post-processing aims to generate pseudo labels for reflectance and shading in three steps, as illustrated in Fig. 2 (right). Based on the physics modeling of image formation, we start with generating pseudo shading by the normal and light visibility. We then generate pseudo reflectance using multiple images and shadings under different illumination. Details can be found in the supplementary.

Step A. The normals are derived from the SDF network. The geometry network also provides depth information which is used to estimate the intersection points in conjunction with sphere tracing[5]. Light visibility, which indicates whether a point is directly illuminated, is obtained by sphere tracing based on the light position and intersection points.

Step B. The generation of pseudo-shading follows the formula, $S^* = ((\vec{N} \cdot \vec{L}) \cdot V)^\gamma$, where the optimal shading S^* is the multiplication of the light visibility V and the dot product of the normal \vec{N} and the light ray \vec{L} . $(\cdot)^\gamma$ represents for gamma correction. This correction is essential because the human eye's perception of brightness is not linear. Thus, we apply it to accommodate the perceptual effect, yielding to our pseudo shading.

Step C. Our method infers pseudo reflectance from pseudo shading using $R = I/S$. For pseudo labels, this calculation is only applicable in the case of direct illumination. We utilize various lighting conditions to obtain different reflectance values; and by employing K-means [29] along with the confidence related to pseudo shading, we merge to form the most probable reflectance map. Areas lacking direct illumination are filled using a strategy considering pixel distance, normals, and RGB colors, resulting in the final

pseudo reflectance.

3.3. Stage 2: Learning Intrinsic Decomposition

As illustrated in Fig. 2 (bottom-left), we jointly learn the relighting and intrinsic decomposition. Expanding the model from Stage 1, we add two extra MLPs dedicated to generating reflectance and shading outputs, while the geometry network is frozen. Note that, while all MLPs receive SDF feature inputs, the RGB color MLP accepts spatial points, camera pose, and light positions as input, the reflectance MLP only receives spatial points, and the shading MLP takes spatial points and light positions.

After volume rendering, we obtain RGB images, along with reflectance and shading. Subsequently, the residual is derived from Eq. (1). During training, the pseudo labels are used to impose constraints on reflectance and shading.

$$L_{intrinsic} = W_R \cdot \|\hat{R} - R^*\|_1 + W_S \cdot \|\hat{S} - S^*\|_1 \quad (4)$$

where \hat{R} and \hat{S} represent the predicted reflectance and shading, respectively, and R^* and S^* are their corresponding pseudo labels. W_R and W_S represent weight maps for reflectance and shading, derived during pseudo label generation. As demonstrated in [39], the diffuse components dominate the scene, so it is crucial to prevent the training from converging to undesirable local minima ($R = 0, S = 0, Re = I$). Therefore, we introduce a regularization term, $L_{reg} = \|\hat{R}\|_1$, to ensure that the image is primarily recovered through R and S . Finally, the Stage 2 loss is:

$$\mathcal{L}_{S2} = \mathcal{L}_{RGB} + w_{intrinsic} L_{intrinsic} + w_{reg} L_{reg} \quad (5)$$

	Hotdog (Reflectance)			Hotdog (Shading)			Lego (Reflectance)			Lego (Shading)		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
PIE-Net[6]	22.21	0.9208	0.0843	22.92	0.9291	0.1091	20.70	0.8902	0.1127	21.36	0.9016	0.1180
Careaga et al. [4]	19.67	0.9177	0.0894	18.94	0.9346	0.0946	17.78	0.8801	0.1123	19.63	0.9032	0.1068
IntrinsicNeRF*[39]	25.62	0.9620	0.0967	-	-	-	19.00	0.9046	0.1288	-	-	-
Ours	28.87	0.9642	0.0381	27.68	0.9666	0.0330	26.88	0.9472	0.0432	24.45	0.9417	0.0520

Table 1. Quantitative results on the NeRF [24] dataset.

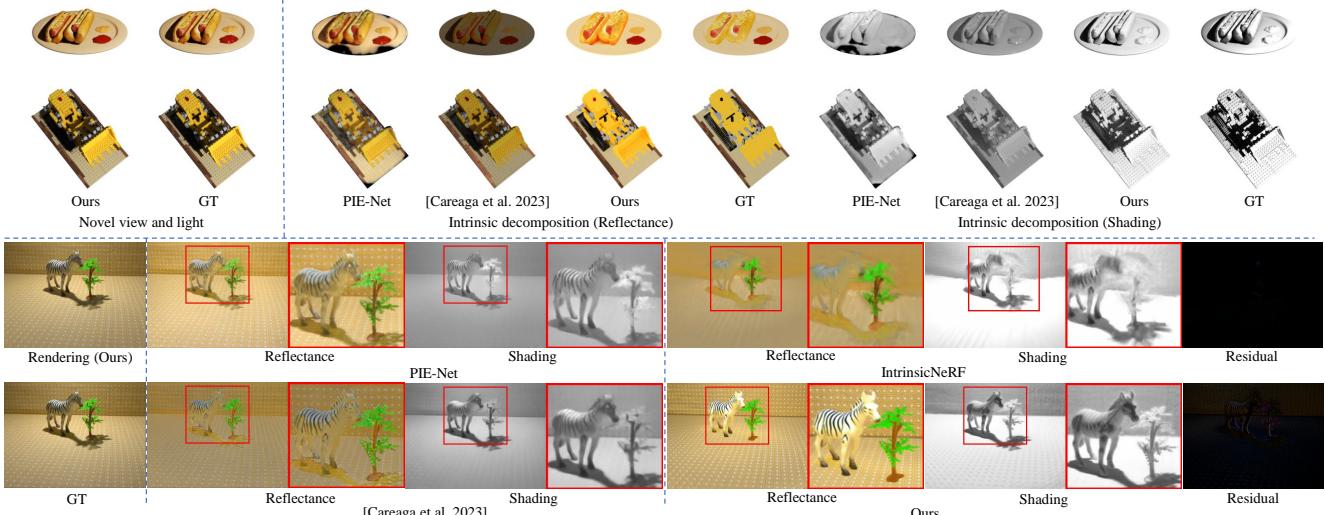


Figure 3. Qualitative comparisons on the NeRF [24] (the first two rows) and ReNe [34] datasets (the last two rows).

4. Experiments

We conduct experiments on both the NeRF [24] (synthetic) and the ReNe [34] (real) datasets. Detailed setup can be found in the supplementary. We compare our method with traditional learning-based methods (PIE-Net [6] and Careaga et al. [4]) and the state-of-the-art neural rendering approach (IntrinsicNeRF [39]).

Tab. 1 displays our method’s quantitative results compared with other methods on the NeRF dataset. Since IntrinsicNeRF struggles with datasets that have lighting variations, we use the numbers from the original publication for comparison. For both the Hotdog and Lego scenes, our approach surpasses others in terms of both reflectance and shading across all metrics.

Fig. 3 presents the qualitative comparison of our method against others on both the synthetic NeRF dataset and the real-world ReNe dataset. On the NeRF dataset, we first showcase the outcomes of our synthesized novel views and lighting conditions on the left, demonstrating results closely aligned with the GT. Then, we display the results of intrinsic decomposition compared to other approaches. It is evident that our results are quite convincing and outperform those of others, with almost no lingering cast shadows in the reflectance. The latter part shows results from the challenging ReNe dataset, characterized by real scenes with backgrounds. Our rendering effects, displayed on the left, closely approximate the GT. Moreover, our method is the only one that achieves credible results in intrinsic de-

composition. In terms of reflectance, the object’s texture edges are sharp, the colors are vibrant, and shadows are accurately eliminated. In contrast, the results from PIE-Net[6] and Careaga et al. [4] are blurry and fail to remove shadows correctly. The other neural rendering method, IntrinsicNeRF [39], also fails to achieve correct decomposition, primarily attributed to the failure in distinguishing intrinsic components and also the difficulty in scene reconstruction.

5. Conclusion

We introduce a neural rendering method that learns relighting and intrinsic decomposition from multi-view images with varying lighting without the intrinsic GT. This approach supports the creation of new views, relighting, and decomposition simultaneously, serving as a versatile tool for editing tasks like reflectance and shading adjustments. Our tests on both synthetic and real-world datasets validate our method’s effectiveness. This method, grounded in basic physical concepts rather than predefined priors, shows promise for more complex scene analyses. In the future, we aim to extend our experiments to explore a more comprehensive set of scenes.

Acknowledgement: Thanks to Hassan Ahmed Sial for his assistance in generating the synthetic scenes. YY, MV and RB were supported by Grant PID2021-128178OB-I00 funded by MCIN/AEI/10.13039/501100011033, and Generalitat de Catalunya 2021SGR01499. YY is supported by China Scholarship Council.

References

- [1] Jonathan T Barron and Jitendra Malik. Shape, illumination, and reflectance from shading. *IEEE transactions on pattern analysis and machine intelligence*, 37(8):1670–1687, 2014. 2
- [2] Harry Barrow, J Tenenbaum, A Hanson, and E Riseman. Recovering intrinsic scene characteristics. *Comput. Vis. Syst.*, 2 (3-26):2, 1978. 1, 2
- [3] Brent Burley and Walt Disney Animation Studios. Physically-based shading at disney. In *Acm Siggraph*, pages 1–7. vol. 2012, 2012. 1
- [4] Chris Careaga and Yağız Aksoy. Intrinsic image decomposition via ordinal shading. *ACM Trans. Graph.*, 2023. 2, 4
- [5] Ziyu Chen, Chenjing Ding, Jianfei Guo, Dongliang Wang, Yikang Li, Xuan Xiao, Wei Wu, and Li Song. L-tracing: Fast light visibility estimation on neural surfaces by sphere tracing. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022. 3, 1
- [6] Partha Das, Sezer Karaoglu, and Theo Gevers. Pie-net: Photometric invariant edge guided network for intrinsic image decomposition. In *IEEE Conference on Computer Vision and Pattern Recognition, (CVPR)*, 2022. 2, 4
- [7] Farshad Einabadi, Jean-Yves Guillemaut, and Adrian Hilton. Deep neural models for illumination estimation and relighting: A survey. In *Computer Graphics Forum*, pages 315–331. Wiley Online Library, 2021. 2
- [8] Majed El Helou, Ruofan Zhou, Sabine Susstrunk, and Radu Timofte. Ntire 2021 depth guided image relighting challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 566–577, 2021. 2
- [9] Qingnan Fan, Jiaolong Yang, Gang Hua, Baoquan Chen, and David Wipf. Revisiting deep intrinsic image decompositions. 2018. 2
- [10] 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
- [11] Elena Garces, Carlos Rodriguez-Pardo, Dan Casas, and Jorge Lopez-Moreno. A survey on intrinsic images: Delving deep into lambert and beyond. *International Journal of Computer Vision*, 130(3):836–868, 2022. 2
- [12] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. In *Proceedings of Machine Learning and Systems 2020*, pages 3569–3579. 2020. 2
- [13] Majed El Helou, Ruofan Zhou, Sabine Süsstrunk, Radu Timofte, Mahmoud Afifi, Michael S Brown, Kele Xu, Hengxing Cai, Yuzhong Liu, Li-Wen Wang, et al. Aim 2020: Scene relighting and illumination estimation challenge. *arXiv preprint arXiv:2009.12798*, 2020. 2
- [14] Andrew Hou, Michel Sarkis, Ning Bi, Yiying Tong, and Xiaoming Liu. Face relighting with geometrically consistent shadows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4217–4226, 2022. 2
- [15] Haian Jin, Isabella Liu, Peijia Xu, Xiaoshuai Zhang, Songfang Han, Sai Bi, Xiaowei Zhou, Zexiang Xu, and Hao Su. Tensoir: Tensorial inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1
- [16] Peter Kocsis, Julien Philip, Kalyan Sunkavalli, Matthias Nießner, and Yannick Hold-Geoffroy. Lightit: Illumination modeling and control for diffusion models. In *CVPR*, 2024. 2
- [17] Louis Lettry, Kenneth Vanhoey, and Luc Van Gool. Unsupervised deep single-image intrinsic decomposition using illumination-varying image sequences. In *Computer Graphics Forum*, pages 409–419. Wiley Online Library, 2018. 2
- [18] Zhengqi Li and Noah Snavely. Learning intrinsic image decomposition from watching the world. In *Computer Vision and Pattern Recognition (CVPR)*, 2018. 2
- [19] Zhengqi Li and Noah Snavely. Cgintrinsics: Better intrinsic image decomposition through physically-based rendering. In *European Conference on Computer Vision (ECCV)*, 2018. 2
- [20] Zhaoshuo Li, Thomas Müller, Alex Evans, Russell H Taylor, Mathias Unberath, Ming-Yu Liu, and Chen-Hsuan Lin. Neuralangelo: High-fidelity neural surface reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1, 2
- [21] Jingwang Ling, Zhibo Wang, and Feng Xu. Shadowneus: Neural sdf reconstruction by shadow ray supervision, 2022. 1
- [22] Yunfei Liu, Yu Li, Shaodi You, and Feng Lu. Unsupervised learning for intrinsic image decomposition from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3248–3257, 2020. 2
- [23] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2018. 2
- [24] 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. *Communications of the ACM*, 65(1):99–106, 2021. 1, 4, 2
- [25] 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
- [26] Lukas Murmann, Michael Gharbi, Miika Aittala, and Fredo Durand. A dataset of multi-illumination images in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4080–4089, 2019. 2
- [27] Thomas Nestmeyer, Jean-François Lalonde, Iain Matthews, and Andreas Lehrmann. Learning physics-guided face relighting under directional light. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5124–5133, 2020. 2
- [28] Rohit Kumar Pandey, Sergio Orts Escalano, Chloe LeGenstre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. Total relighting: Learning to relight portraits for background replacement. 2021. 2

- [29] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. 3
- [30] Densen Puthusseray, Melvin Kuriakose, Jiji C V, et al. Wdrn: A wavelet decomposed relightnet for image relighting. *arXiv preprint arXiv:2009.06678*, 2020. 2
- [31] Viktor Rudnev, Mohamed Elgharib, William Smith, Lingjie Liu, Vladislav Golyanik, and Christian Theobalt. Nerf for outdoor scene relighting. In *European Conference on Computer Vision (ECCV)*, 2022. 2
- [32] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7495–7504, 2021. 2
- [33] Tiancheng Sun, Jonathan T Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul E Debevec, and Ravi Ramamoorthi. Single image portrait relighting. *ACM Trans. Graph.*, 38(4):79–1, 2019. 2
- [34] Marco Toschi, Riccardo De Matteo, Riccardo Spezialetti, Daniele De Gregorio, Luigi Di Stefano, and Samuele Salti. Relight my nerf: A dataset for novel view synthesis and relighting of real world objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20762–20772, 2023. 2, 4
- [35] Li-Wen Wang, Wan-Chi Siu, Zhi-Song Liu, Chu-Tak Li, and Daniel PK Lun. Deep relighting networks for image light source manipulation. *arXiv preprint arXiv:2008.08298*, 2020. 2
- [36] Yuxin Wang, Wayne Wu, and Dan Xu. Learning unified decomposition and compositional nerf for editable novel view synthesis. In *ICCV*, 2023. 1
- [37] 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. 2
- [38] Ziyi Yang, Yanzhen Chen, Xinyu Gao, Yazhen Yuan, Yu Wu, Xiaowei Zhou, and Xiaogang Jin. Sire-ir: Inverse rendering for brdf reconstruction with shadow and illumination removal in high-illuminance scenes. *arXiv preprint arXiv:2310.13030*, 2023. 1
- [39] Weicai Ye, Shuo Chen, Chong Bao, Hujun Bao, Marc Pollefeys, Zhaopeng Cui, and Guofeng Zhang. IntrinsicNeRF: Learning Intrinsic Neural Radiance Fields for Editable Novel View Synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023. 1, 2, 3, 4
- [40] Chong Zeng, Guojun Chen, Yue Dong, Pieter Peers, Hongzhi Wu, and Xin Tong. Relighting neural radiance fields with shadow and highlight hints. In *ACM SIGGRAPH 2023 Conference Proceedings*, 2023. 1, 2
- [41] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 2
- [42] 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 Transactions on Graphics (ToG)*, 40(6):1–18, 2021. 1
- [43] Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. Modeling indirect illumination for inverse rendering. In *CVPR*, 2022. 1
- [44] Hao Zhou, Sunil Hadap, Kalyan Sunkavalli, and David W Jacobs. Deep single-image portrait relighting. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7194–7202, 2019. 2

Learning Relighting and Intrinsic Decomposition in Neural Radiance Fields

Supplementary Material

In this supplementary material, we present the following:

1. More detailed procedures for generating pseudo labels.
2. Specifications of experimental settings.
3. Additional qualitative results.

6. Pseudo Label Generation

Here, we elaborate on the post-processing steps (Fig. 4) in the main paper Sec. 3.2. It starts with generating pseudo shading based on Lambertian reflection principles. Under the assumption that the light intensity and color remain constant, shading can be approximated by the dot product between the normal ray and the light ray. The light ray encompasses both direct/indirect illumination and necessitates light visibility to account for occlusion effects.

6.1. Step A: obtain normal and light visibility

The normals are derived from the SDF network. The geometry network also provides depth information which is used to estimate the intersection points in conjunction with sphere tracing[5]. Light visibility, which indicates whether a point is directly illuminated, is obtained by sphere tracing based on the light position and intersection points.

6.2. Step B: generate pseudo shading

The generation of pseudo-shading follows the formula,

$$S' = ((\vec{N} \cdot \vec{L}) \cdot V)^\gamma \quad (6)$$

where the optimal shading S' is the multiplication of the light visibility V and the dot product of the normal \vec{N} and the light ray \vec{L} . $(\cdot)^\gamma$ represents for gamma correction. This correction is crucial because the human eye's perception of brightness is not linear. Most images we see have undergone gamma correction to accommodate this perceptual effect. Therefore, calculating shading also necessitates gamma correction, yielding to our defined pseudo shading.

6.3. Step C: generate pseudo reflectance

This step entails inferring the most probable pseudo reflectance from the pseudo shading, principally based on the equation $R = I/S$. The approach has two main points that should be noted here.

First, the current pseudo shading only considers direct light. As seen in previous papers [15, 38, 43], solving for indirect light is a complex and computationally expensive process. Our novel approach leverages the trained model to generate multiple versions of images under different lighting conditions, each accompanied by respective

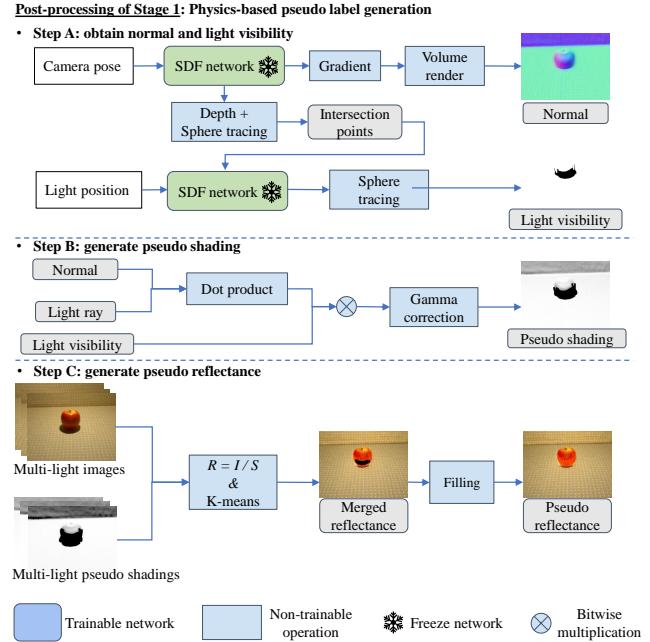


Figure 4. Post-processing and generating pseudo labels.

pseudo shadings. As direct light strengthens on a pixel, the influence of indirect light diminishes, making the reflectance derived from higher pseudo shading values more reliable. We compare the outcomes under multiple lighting conditions and synthesize the most credible reflectance for each pixel based on the intensity of pseudo shading.

Second, the residual term includes specularity and other effects that are not considered in $R = I/S$. Specular highlights, which have high pseudo shading values, do not reflect the object color but rather the light source color (e.g., white reflections). By analyzing different lighting conditions, where highlights typically vanish except under specific angles, we can deduce the object color by selecting the most common reflectance outcomes.

Our implementation employs the K-means algorithm, incorporating the weights of pseudo shading. This approach allows us to achieve a merged reflectance under varied lighting conditions, as shown in the intermediate result at the bottom in Fig. 4. However, some regions within the merged reflectance may appear vacant due to the absence of direct illumination in all lighting conditions. So, we address these areas with a filling strategy. This strategy specifically considers the distance between void and non-void pixels, their normals, and their colors in the RGB image, thereby achieving the final pseudo reflectance.

Additionally, we compute weight maps W_R and W_S for

both pseudo reflectance and pseudo shading based on the edges of pseudo shading and visibility. Areas with higher pseudo shading values, or those further from visibility edges (where visibility calculations may be prone to errors), exhibit greater credibility in their pseudo labels; conversely, areas closer to visibility edges or with lower pseudo shading values are deemed less reliable.

7. Experimental Settings

Datasets. To validate our approach, we conduct experiments on both synthetic and real-world datasets.

For the synthetic dataset, models are obtained from NeRF [24], with lighting configurations borrowed from Zeng et al. [40]. To facilitate quantitative analysis, GT for reflectance, shading, and residuals are rendered in Blender. Each scene comprises 500 images for training, 100 for validation, and 100 for testing, including intrinsic components for each image. Importantly, adhering to the configurations in [40], the settings for lighting and camera poses are managed independently.

The real dataset we use is the ReNe dataset [34], where lighting and camera poses are grid-sampled. This dataset features 2000 images across scenes, captured from 50 different viewpoints under 40 lighting conditions. Following their dataset split, we use 1628 images (44 camera poses \times 37 light positions) for training.

Additionally, given that the settings of lights and cameras are dependent on the former one and grid-sampled in the latter, our proposed method is designed to accommodate both configurations.

Metrics. To evaluate the comparison between predicted images and ground truth (GT), we employ the following metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) [37], and Learned Perceptual Image Patch Similarity (LPIPS) [41].

Implementation details. Our model’s hyperparameters include a batch size of 2048 and each stage was trained for 500k iterations. We implemented the model in PyTorch and used the AdamW [23] optimizer with a learning rate of $1e^{-3}$ for optimization. The experiments can be conducted on a single Nvidia RTX 3090 or A40 GPU. The weights of losses, w_{eik} , w_{curv} , $w_{\text{intrinsic}}$, w_{reg} are set to 0.1, $5e^{-4}$, 1.0, and 1.0, respectively.

8. Additional qualitative results

We present additional results in this section. Fig. 5 displays additional examples comparing our method with others. Fig. 6 - Fig. 11 shows more qualitative results of our method on the ReNe dataset. Furthermore, Fig. 12 - Fig. 14 demonstrate additional qualitative results of our method on real scenes from [40].

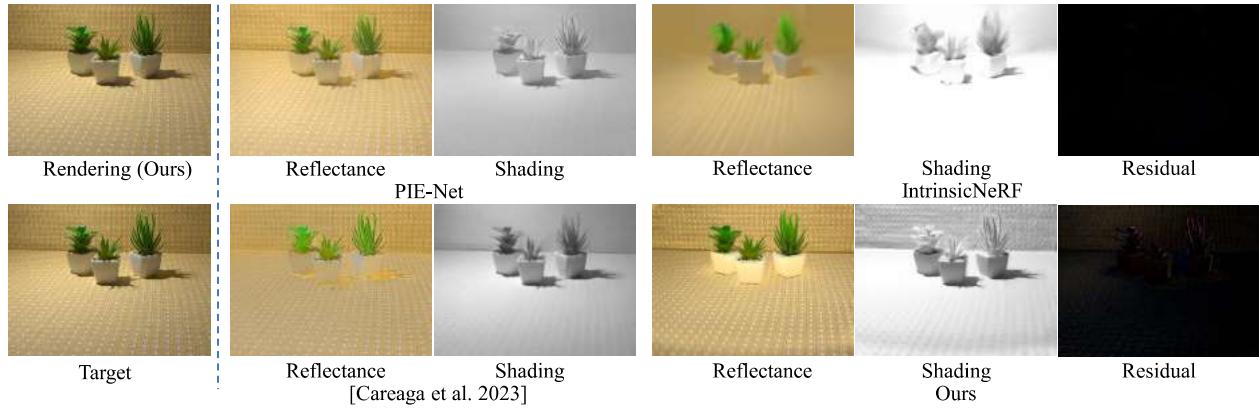


Figure 5. Compare with other methods on the ReNe dataset.

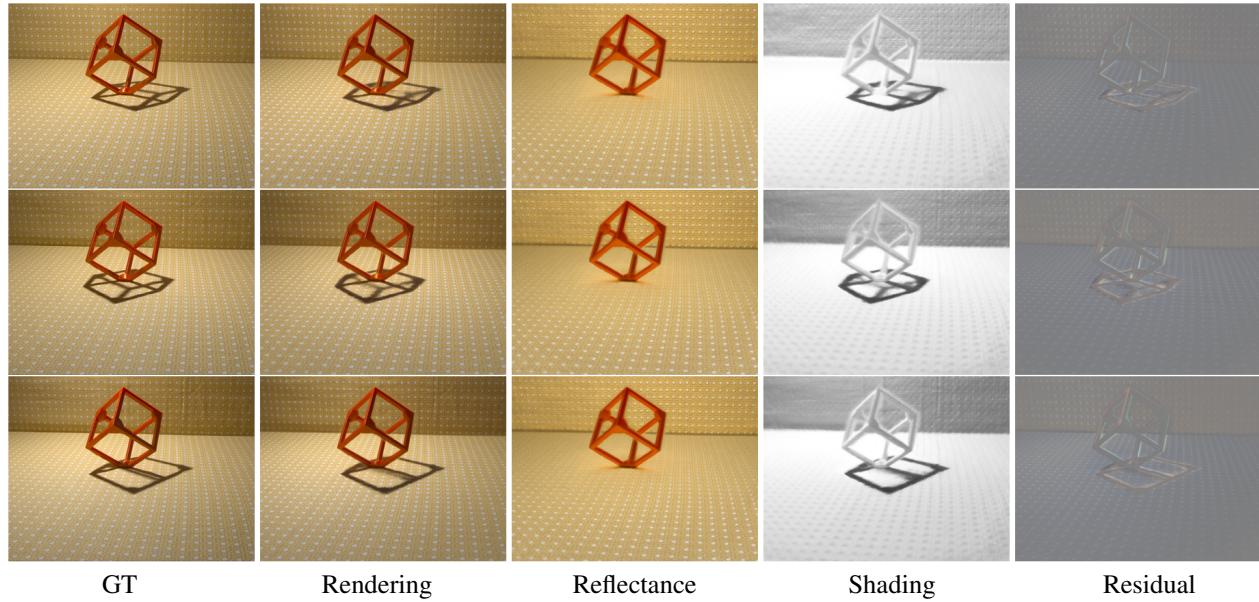


Figure 6. Qualitative results on the ReNe dataset (Cube).

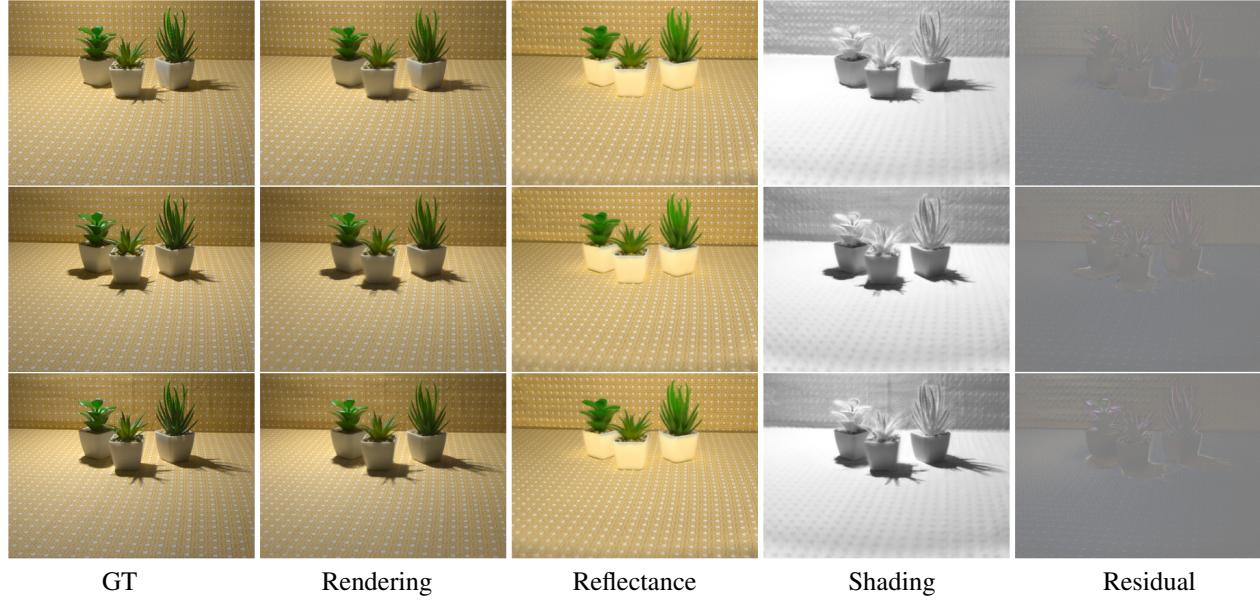


Figure 7. Qualitative results on the ReNe dataset (Garden).

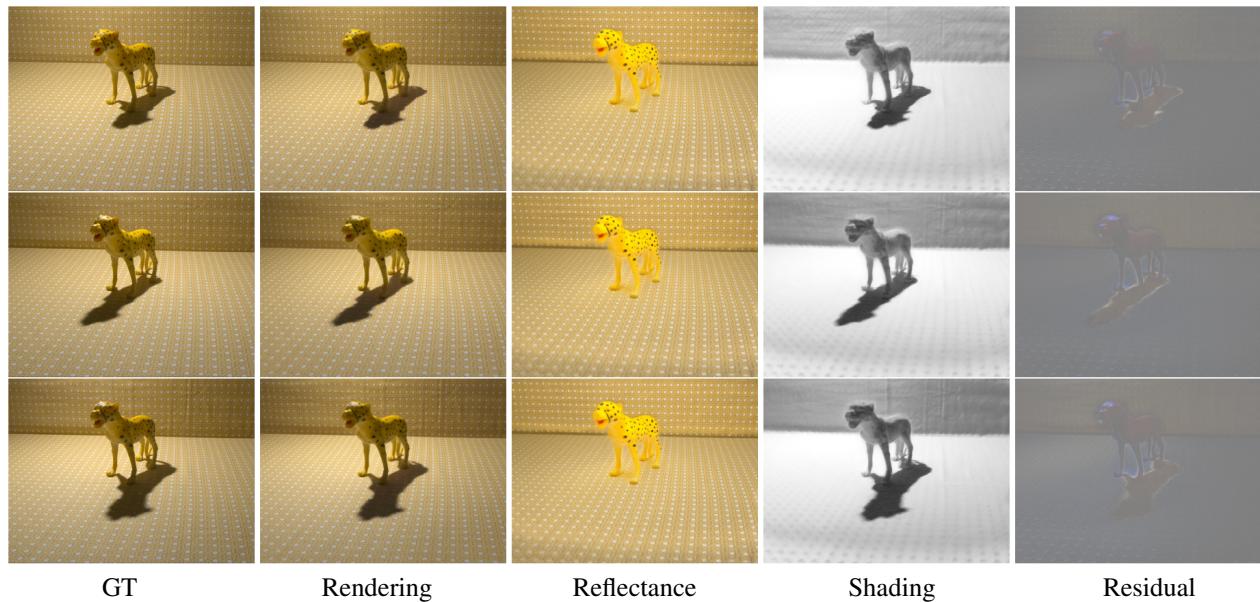


Figure 8. Qualitative results on the ReNe dataset (Cheetah).

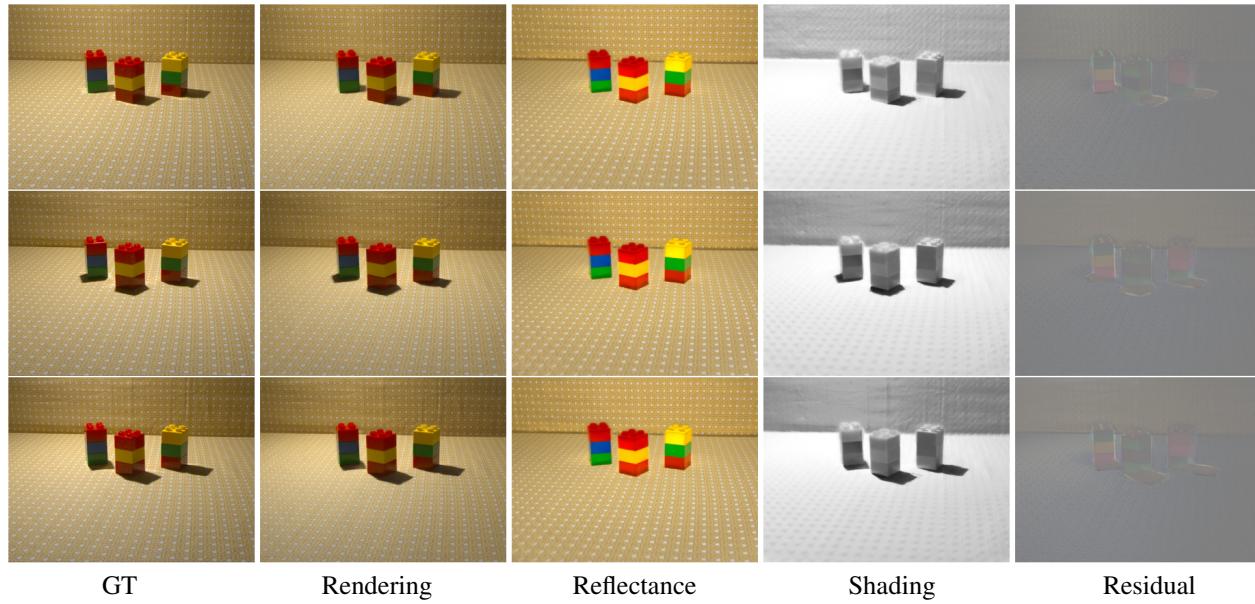


Figure 9. Qualitative results on the ReNe dataset (Lego).

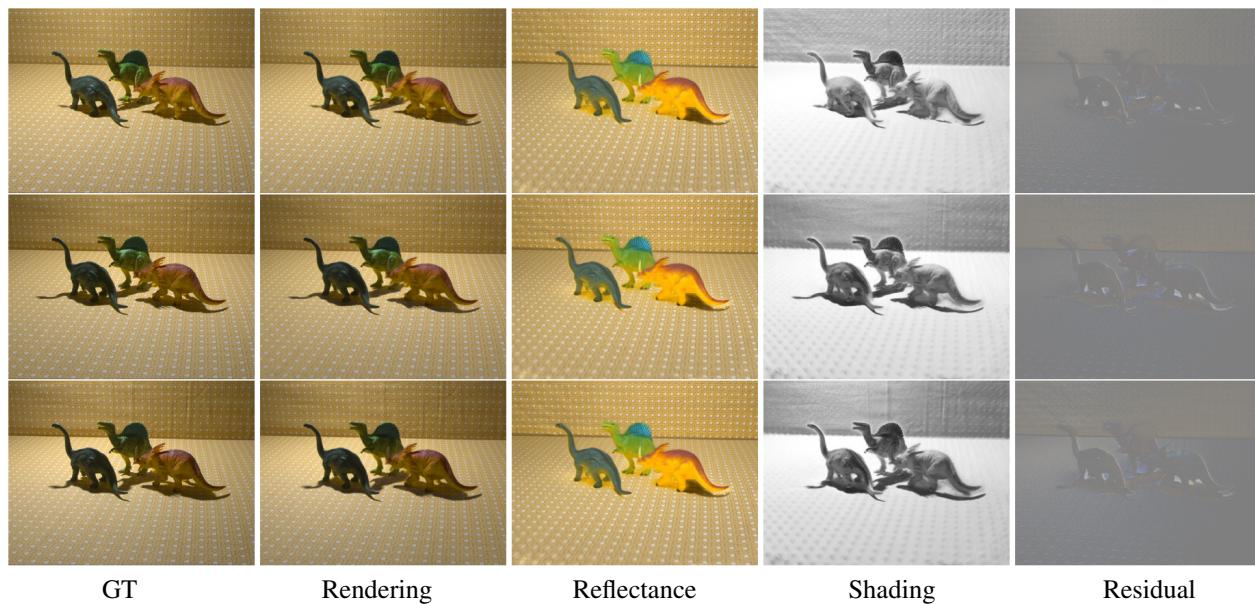


Figure 10. Qualitative results on the ReNe dataset (Dinosaurs).

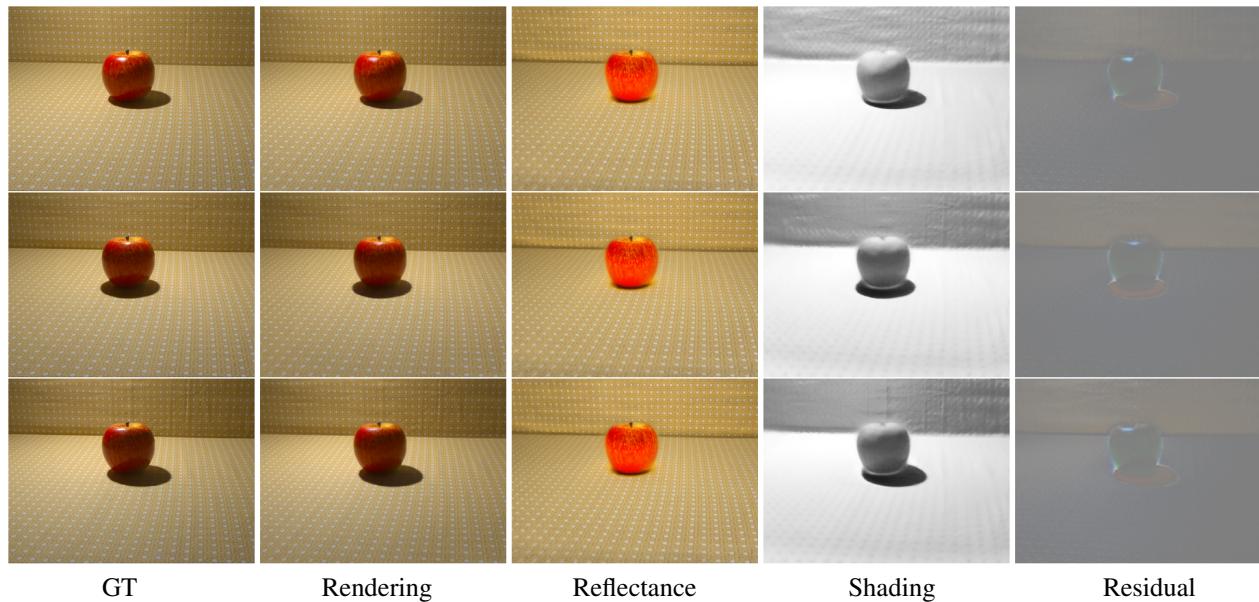


Figure 11. Qualitative results on the ReNe dataset (Apple).

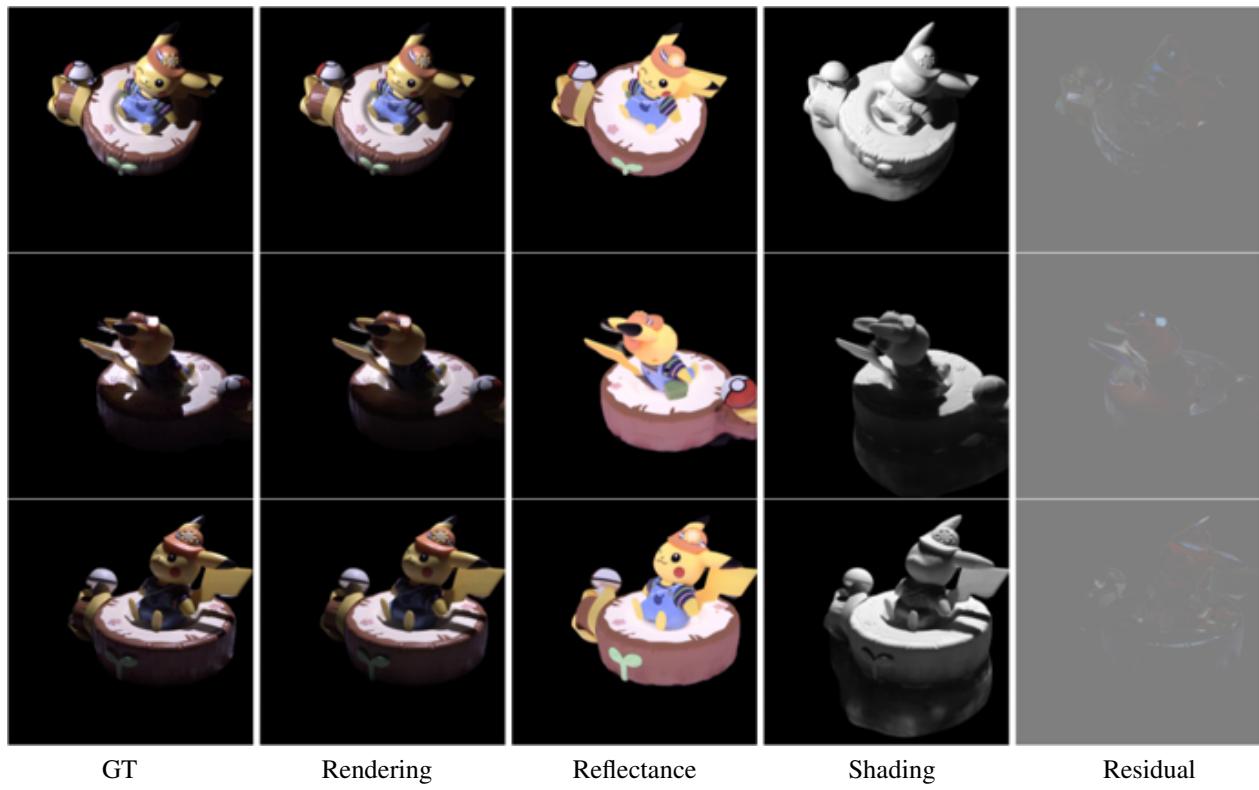


Figure 12. Qualitative results on the real scene (Pikachu).

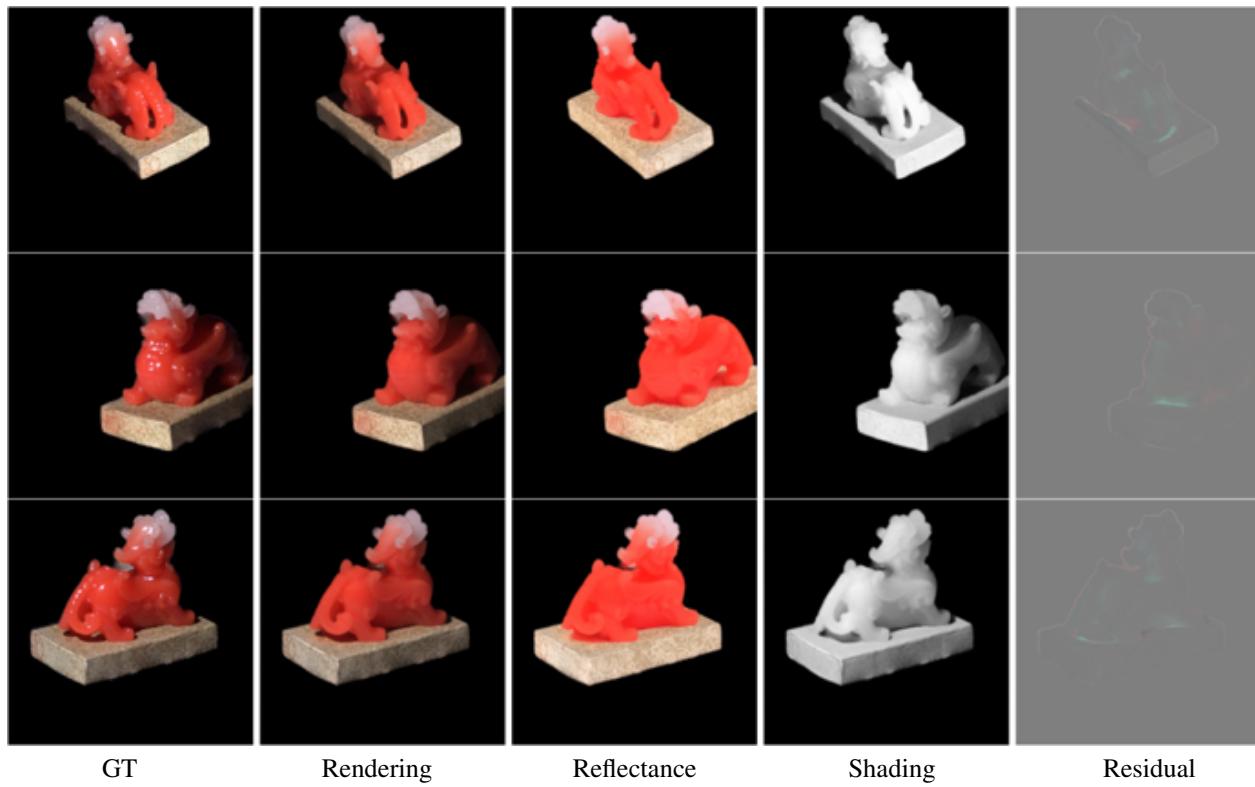


Figure 13. Qualitative results on the real scene (Pixiu).

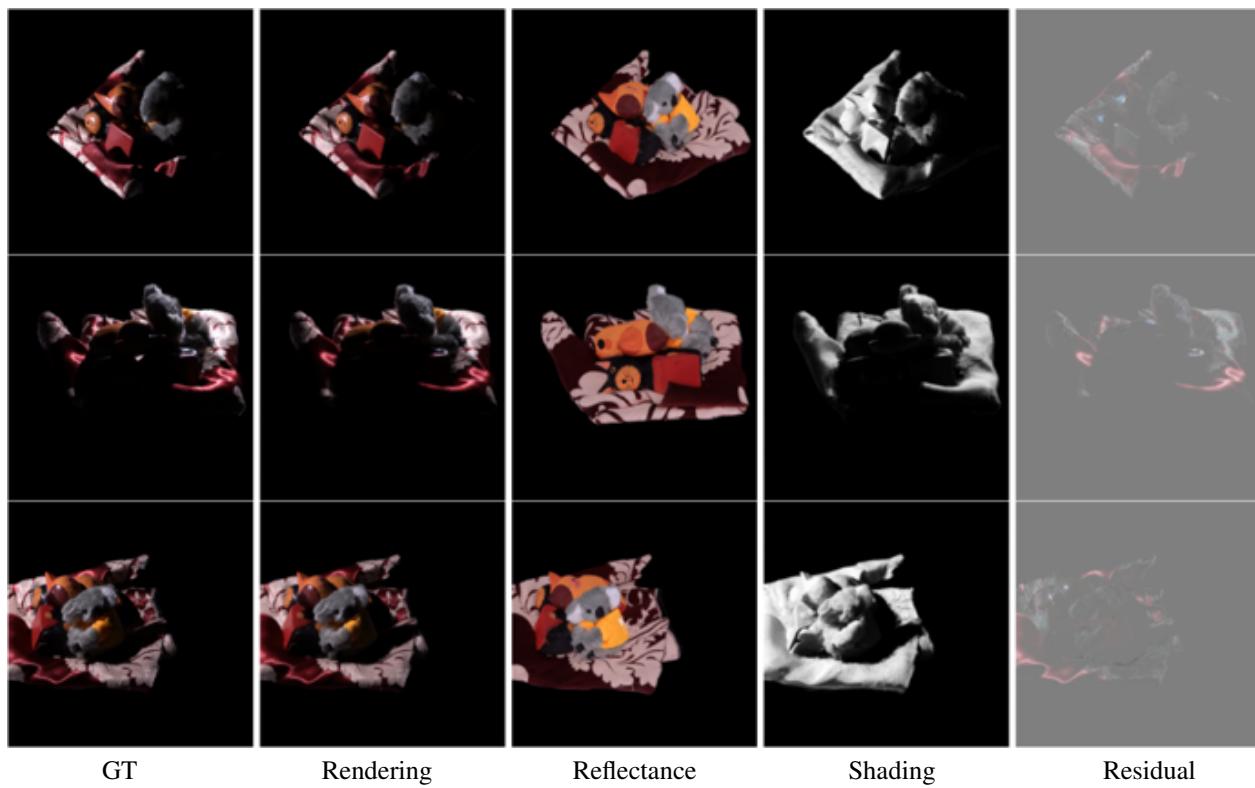


Figure 14. Qualitative results on the real scene (FurScene).