

# NEURAL GLOBAL ILLUMINATION FOR INVERSE RENDERING

*Nikolay Patakin   Dmitry Senushkin   Anna Vorontsova   Anton Konushin*

Samsung Research

## ABSTRACT

Rapid progress in scene reconstruction from images should be attributed to the emergence of differentiable renderers. Still, accurate material reconstruction remains a challenge, as it requires modeling indirect light effects. Modern inverse path tracers solve this problem, but are computationally expensive. At the same time, inverse renderers based on real-time graphics ignore indirect light for real-time performance. In this paper, we introduce a novel neural global illumination model, which estimates both direct environment light and indirect light as a surface light field. We build NeGIL, a Monte Carlo differentiable rendering framework based on the proposed model. Our framework effectively handles complex lighting effects (such as inter-reflections) without costly path tracing and facilitates the reconstruction of physically-based spatially-varying materials in an end-to-end manner. Through experiments on the challenging synthetic scenes, we demonstrate that NeGIL significantly outperforms existing light modeling approaches in terms of novel-view synthesis and relighting quality.

*Index Terms*— inverse rendering, neural rendering

## 1. INTRODUCTION

Lately, AR/VR applications have attracted considerable attention to 3D scene reconstruction from multi-view images. This problem is currently addressed with differentiable rendering. Recently emerged, NeRF [1] performs volumetric rendering and represents a scene as a continuous radiance field. While scene geometry can be reconstructed from such a representation, scene appearance is not being decomposed into lighting and materials, making it challenging to render the scene under novel illumination. On the contrary, we aim at modeling light and materials separately, so that our scene representation is compatible with modern real-time rendering engines.

Meanwhile, differentiable versions of computer graphics (CG) rendering approaches have been introduced. In differentiable renderers, a pixel’s color is modeled via the rendering equation [2], which allows disentangling scene appearance into components compatible with graphic engines. Differential rendering is addressed with either path tracing [3] or rasterization [4]. Rasterization approaches leverage real-time approximations inherited from CG: while speeding up

the computation, these approximations poorly handle indirect illumination and specular inter-reflections present in a scene. In path tracing, the rendering integral is estimated via the explicit light transport simulation, which makes it possible to capture global illumination at the expense of high computational cost.

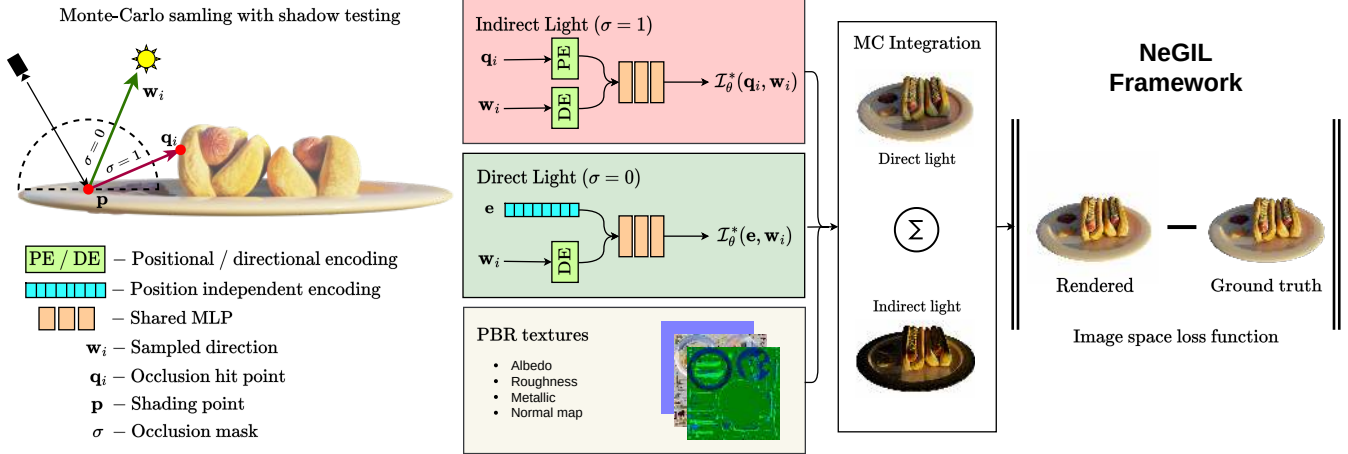
Our goal is to combine the advantages of different rendering approaches in a single method, that would decompose color into lighting and materials and handle indirect light without costly computations. To address this task, we propose a novel neural global illumination model, which captures complex light effects and allows reconstructing physically-based spatially-varying materials. In this work, we model direct light as an environment map and indirect light as a surface light field. The experiments on synthetic scenes with complex illumination and specular inter-reflections demonstrate that our approach recovers materials and lighting more precisely than state-of-the-art approaches.

## 2. RELATED WORK

Nowadays, inverse rendering keeps evolving rapidly. Modern NeRF-based approaches [5, 6] employ inventive strategies to extract materials and lighting from the obtained reconstructions; still, materials contain severe artifacts getting revealed in rendered images. This undesired effect is due to volumetric rendering, which does not restore scene surfaces explicitly. Meanwhile, significant progress has been achieved in differentiable mesh-based rendering (e.g., DIB-R [7], NVDIFFRAST [4], Redner[8], Mitsuba [3]), a competing inverse rendering paradigm capable of decomposing lighting, materials and geometry.

Most real-time inverse rendering approaches ignore global illumination while computing the rendering integral. NVDIFFREC [9] implements a differentiable version of a split-sum image-based lighting approximation [10]. The follow-up work, NVDIFFREC-MC [11], combines image-based lighting with Monte-Carlo integration. These methods do not process indirect light, thereby producing biased renderings.

NeILF [12] is likely the most similar with NeGIL, since both methods perform a neural approximation of incident light. Unlike NeILF, we decompose the incident light into direct and indirect lighting, and parameterize indirect light with an occlusion ray hit position rather than a primary (shading)



**Fig. 1:** Scheme of our rendering framework. Given a point  $p$ , we sample a set of directions  $w_i$  and detect occlusion points  $q_i$  via shadow ray casting. Incident light  $\mathcal{I}_\theta^*(q_i, w_i)$  is calculated as an occlusion-aware combination of direct  $\mathcal{I}_\theta^*(e, w_i)$  and indirect  $\mathcal{I}_\theta^*(q_i, w_i)$  lighting. For direct light, we use a global trainable embedding  $e$  to avoid positional dependency. Both direct and indirect lighting are modeled with a shared MLP that accepts  $q_i$  and  $w_i$  as inputs. A rendered color is a Monte Carlo estimate (Eq. 5). Since the framework is end-to-end differentiable, PBR materials and neural lighting are optimized end-to-end by back-propagating gradients of image-based loss function.

ray hit point considered in NeILF.

### 3. PROPOSED METHOD

Inverse rendering aims at finding the optimal scene parameters via analysis-by-synthesis. Given a set of reference images  $R_{ref}(c)$  with camera poses  $c$ , the desired scene parameters  $\theta$  are derived from the following optimization problem:

$$\arg \min_{\theta} \mathbb{E}_c [\mathcal{L}(R_\theta(c), R_{ref}(c))]. \quad (1)$$

Here,  $R_\theta(c)$  denotes an image rendered from a camera pose  $c$ ;  $\mathcal{L}(\cdot, \cdot)$  is an image space loss function. Following a surface-based rendering paradigm [2], we model a pixel color as a *reflected radiance*: light reflected in a direction  $w_o$  from a shading point  $p$  with a surface normal  $n$ :

$$R_\theta(p, w_o) = \int_{\Omega} \mathcal{I}_\theta(p, w_i) f_\theta(p, w_i, w_o) (w_i \cdot n) dw_i, \quad (2)$$

The above integral is calculated over incident light directions  $w_i$  on a unit hemisphere  $\Omega$ . It includes an explicit incident light model  $\mathcal{I}_\theta(p, w_i)$  and an explicit light-surface interaction model  $f(p, w_i, w_o)$ , usually referred to as a bidirectional reflection distribution function (BRDF).

In our framework, we perform physically-based rendering (PBR), using the Lambertian BRDF to model diffuse reflections and the Cook-Torrance [13] microfacet shading model for a specular lobe. The final BRDF is a parametric function of PBR materials: diffuse albedo  $a$ , metallic  $m$ , and roughness  $\rho$ :

$$f_\theta(p, w_i, w_o) = \frac{k_d(a, m)}{\pi} + \frac{D(\rho)F(a, m)G(\rho)}{4(w_i \cdot n)(w_o \cdot n)} \quad (3)$$

Our core contribution is a neural global illumination model. We explicitly separate direct light  $\mathcal{I}_\theta(e, w_i)$  depending on an incident direction  $w_i$ , and indirect light  $\mathcal{I}_\theta(q_i, w_i)$  depending on an incident direction  $w_i$  and an occlusion point  $q_i$ . Hence, our global illumination model does not depend on a shading point position  $p$ , but is parameterized with an occlusion point  $q_i$ , an incident direction  $w_i$ , and a binary occlusion mask  $\sigma$ :

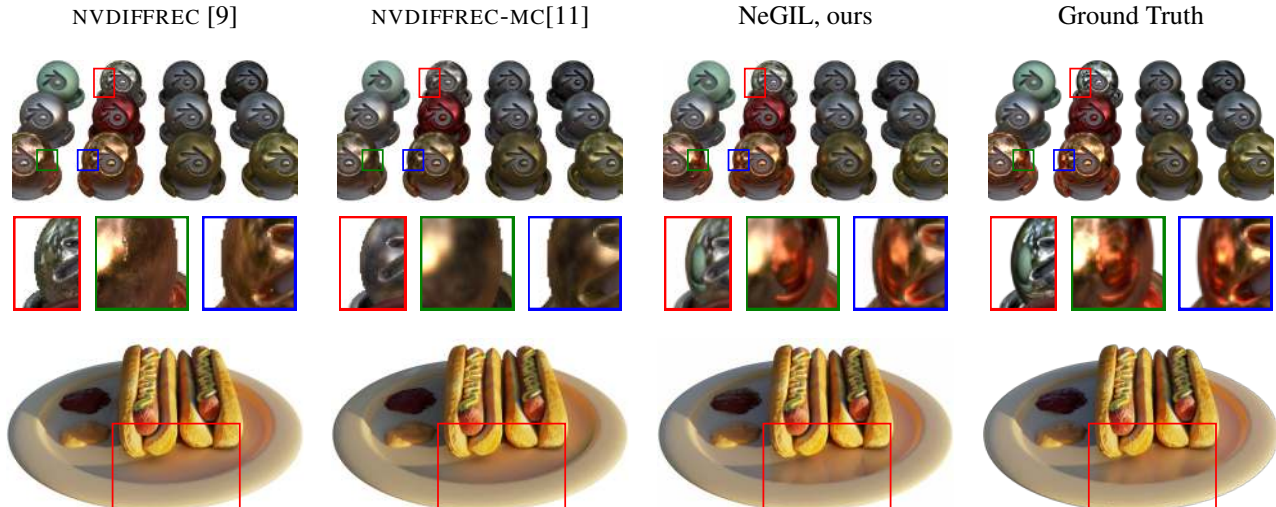
$$\mathcal{I}_\theta(q_i, w_i) = \underbrace{\sigma \cdot \mathcal{I}_\theta(e, w_i)}_{\text{direct}} + \underbrace{(1 - \sigma) \cdot \mathcal{I}_\theta(q_i, w_i)}_{\text{indirect}} \quad (4)$$

Parameterizing incident light with an occlusion point instead of a shading point is based on an assumption of no energy loss on a pass-through the media (absolute permeability); we found this trick to significantly increase the lighting consistency. We also found that both direct and indirect lighting can be modeled with the same shared MLP  $\mathcal{I}_\theta(\cdot, \cdot)$ . This small network expects an encoded position of an occlusion point  $q_i$  and a direction  $w_i$  as inputs. To process position-independent direct light, we leverage an optimizable global encoding  $e$  instead of an actual encoded position.

Eventually, the infinity-dimensional integral (Eq. 2) is approximated via stochastic Monte Carlo integration:

$$R_\theta(p, w_o) = \sum_{w_i \sim p(w)} \frac{\mathcal{I}_\theta(q_i, w_i) f_\theta(p, w_i, w_o) (w_i \cdot n)}{p(w_i)}, \quad (5)$$

Incident directions  $w_i$  are sampled from a unit hemisphere  $\Omega$  according to the multiple importance sampling strategy with cosine and GGX distributions. Following [14], we detach gradients of PDF estimation during optimization.



**Fig. 2:** Novel views of *Materials* and *Hotdog* scenes from the NeRF synthetic benchmark, rendered using NeGIL and baseline approaches. Both NVDIFFREC and NVDIFFREC-MC miss complex light effects, while NeGIL models inter-reflections precisely.

Method	Materials	Hotdog	Mic	Ficus	Chair	Drums	Ship	Lego	Average
NVDIFFREC [9]	28.10	34.04	29.61	30.50	<b>31.05</b>	25.60	28.50	31.17	29.82
NVDIFFREC-MC [11]	25.18	31.40	27.00	27.40	27.60	23.90	24.45	28.00	26.87
NeGIL, ours	<b>31.40</b>	<b>35.56</b>	<b>34.31</b>	<b>31.22</b>	30.19	<b>26.93</b>	<b>30.33</b>	<b>32.87</b>	<b>31.60</b>

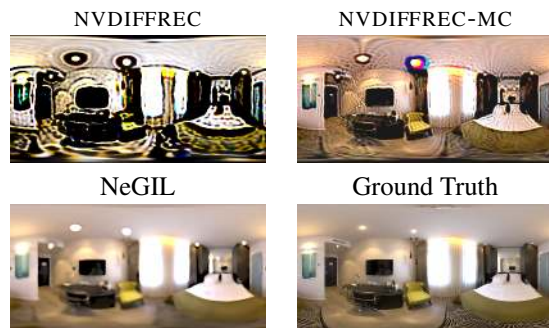
**Table 1:** Results of novel-view synthesis of scenes from the NeRF synthetic benchmark. PSNR, dB $\uparrow$  values are reported. NeGIL ensures higher rendering quality for most scenes, and is superior on average.

## 4. EXPERIMENTS

### 4.1. Experimental Setup

**Datasets.** We evaluate NeGIL against existing baselines in two scenarios. Firstly, we measure the novel-view synthesis quality using the NeRF synthetic benchmark [15]. Secondly, we assess the quality of estimated materials by relighting the obtained assets: for that purpose, we create a synthetic dataset comprising four challenging scenes with inter-reflections. In all experiments, PSNR between rendered and ground truth images is used as a metric.

**Implementation details.** During training, we randomly select a set of  $K = 6$  training images at each iteration and sample  $N = 5000$  rays from each image. Monte Carlo integration is performed based on 128 samples during training and 1024 samples during testing. We minimize a tonemapped log-space  $L_1$  image loss [9], using the Adam [16] optimizer with a learning rate of  $10^{-3}$ . The optimization is performed with PyTorch [17]. Ray tracing and shadow ray casting are accelerated with OptiX hardware, while BRDF operations are implemented as separate CUDA kernels. We employ a mixed-precision MLP [18]. As a result, our method converges in 18 minutes on RTX 3090, making it comparable with the baselines: NVDIFFREC-MC requires about 25 minutes to run, while NVDIFFREC converges in 12 minutes.



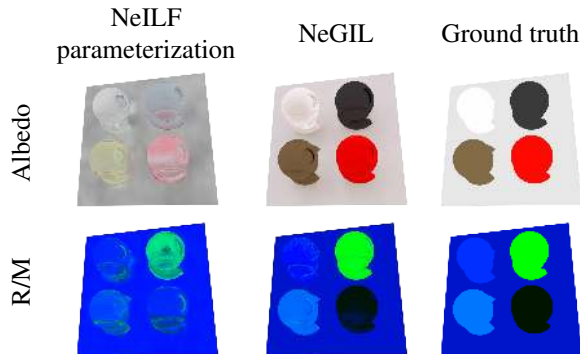
**Fig. 3:** Comparison of estimated direct light on the *Toaster* scene.

### 4.2. Novel-view Synthesis Results

We compare novel-view synthesized images in Fig. 2. We zoom the rendered images to highlight the effects of modeling indirect light. NVDIFFREC-MC is biased: light is not propagated along occluded directions, resulting in dark regions appearing instead of reflections. NVDIFFREC is also incapable to capture inter-reflections, while our approach provides physically correct and visually plausible results. These observations are supported quantitatively (Tab. 1): NeGIL outperforms both NVDIFFREC and NVDIFFREC-MC on all scenes except *Chair*, providing the best overall quality.

Method	Novel-view synthesis, dB $\uparrow$				Relighting, dB $\uparrow$			
	Balls	Toaster	Bells	Materials	Balls	Toaster	Bells	Materials
NVDIFFREC [9]	26.99	23.85	26.25	28.10	25.02	19.20	24.31	21.00
NVDIFFREC-MC [11]	27.25	23.79	25.99	25.70	21.82	19.58	21.61	18.11
NeGIL, ours	<b>35.83</b>	<b>29.20</b>	<b>33.10</b>	<b>31.40</b>	<b>28.46</b>	<b>23.82</b>	<b>28.13</b>	<b>22.69</b>

**Table 2:** Results of novel-view synthesis and relighting. NeGIL outperforms competitors on all four challenging scenes with inter-reflections.



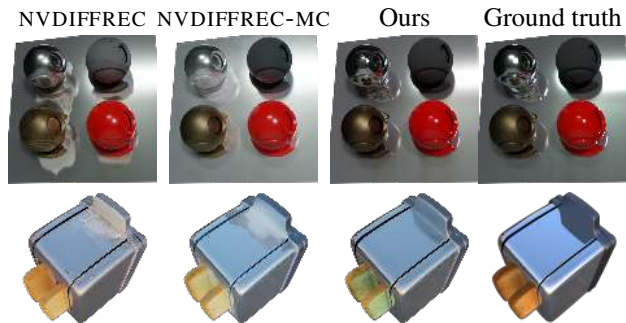
**Fig. 4:** Comparison of materials optimized with different indirect light field parameterizations: as in NeILF [12] and NeGIL, on the *Balls* scene. Our parameterization results in more accurate albedo, roughness/metallic (green/blue) estimates.

### 4.3. Ablation Study

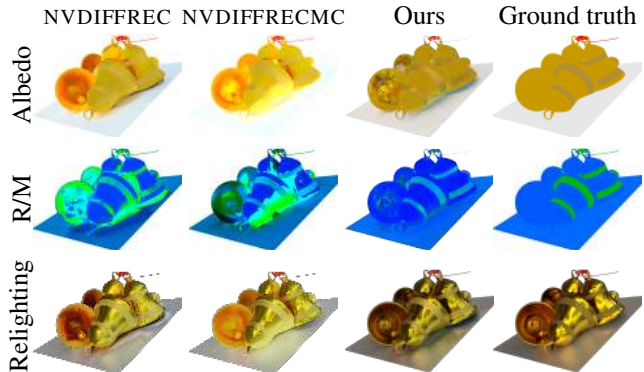
The essential part of our method is the proposed parameterization of an indirect light field. To illustrate that, we compare our method with NeILF [12] parameterization, which does not consider occlusions and models the incident radiance in the shading point  $p$ . On the contrary, we perform an occlusion-aware decomposition of the incident light into direct and indirect components and parameterize the light field in the occlusion point  $q$ . In Fig. 4, we provide a visual comparison of two light field parameterizations. Evidently, the quality of the restored material hugely benefits from the use of our method.

### 4.4. Relighting Results

Eventually, we perform assessment of the reconstructed materials via relighting. We create a challenging synthetic dataset of four scenes (*Balls*, *Bells*, *Toaster*, *Materials*) featuring heavy indirect lighting effects. The training set contains 100 images rendered with the single lighting. The testing set is rendered from 100 camera poses with 8 different environment maps. We render the optimized assets with Blender Cycles. As shown in Tab. 2, NeGIL demonstrates a solid quality gain over baselines in terms of PSNR. By modeling reflections, our approach estimates direct light more precisely (Fig. 3) and produces more plausible materials (Fig. 6), resulting in higher quality of rendered images after relighting (Fig. 5).



**Fig. 5:** Comparison of the relighting quality on *Balls* and *Toaster* scenes. Unlike baselines, NeGIL estimates plausible materials in reflective regions.



**Fig. 6:** Estimated materials: albedo, roughness (green), metallic (blue), and relighted renders of the *Bells* scene. NeGIL estimates roughness and metallic better than baselines.

## 5. CONCLUSION

In this paper, we formulated a novel neural model that represents scene lighting in the form of direct environment light and an indirect surface light field. We used this model to build NeGIL, a differentiable rendering framework that approximates incident radiance and effectively handles recursive light effects without costly computations. We verified NeGIL to reconstruct physically-based materials: albedo, specular roughness, and metallic, in an end-to-end optimization pipeline. Our experiments on challenging synthetic scenes revealed that our approach surpasses the state-of-the-art competitors in a novel-view synthesis and relighting.

## 6. REFERENCES

- [1] 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] James T. Kajiya, “The rendering equation,” in *Proceedings of the 13th annual conference on Computer graphics and interactive techniques*, 1986, pp. 143–150.
- [3] Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, Merlin Nimier-David, Delio Vicini, Tizian Zeltner, Baptiste Nicolet, Miguel Crespo, Vincent Leroy, and Ziyi Zhang, “Mitsuba 3 renderer,” 2022, <https://mitsuba-renderer.org>.
- [4] Samuli Laine, Janne Hellsten, Tero Karras, Yeongho Seol, Jaakko Lehtinen, and Timo Aila, “Modular primitives for high-performance differentiable rendering,” *ACM Transactions on Graphics*, vol. 39, no. 6, 2020.
- [5] 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)*, vol. 40, no. 6, pp. 1–18, 2021.
- [6] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T. Barron, Ce Liu, and Hendrik P.A. Lensch, “Nerd: Neural reflectance decomposition from image collections,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 12684–12694.
- [7] Wenzheng Chen, Huan Ling, Jun Gao, Edward Smith, Jaakko Lehtinen, Alec Jacobson, and Sanja Fidler, “Learning to predict 3d objects with an interpolation-based differentiable renderer,” in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett, Eds. 2019, vol. 32, Curran Associates, Inc.
- [8] Tzu-Mao Li, Miika Aittala, Frédo Durand, and Jaakko Lehtinen, “Differentiable monte carlo ray tracing through edge sampling,” *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, vol. 37, no. 6, pp. 222:1–222:11, 2018.
- [9] Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Müller, and Sanja Fidler, “Extracting triangular 3d models, materials, and lighting from images,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 8280–8290.
- [10] Brian Karis, “Real shading in unreal engine 4,” *SIGGRAPH Course: Physically Based Shading in Theory and Practice*, 2013.
- [11] Jon Hasselgren, Nikolai Hofmann, and Jacob Munkberg, “Shape, light, and material decomposition from images using monte carlo rendering and denoising,” in *Advances in Neural Information Processing Systems*, Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, Eds., 2022.
- [12] Yao Yao, Jingyang Zhang, Jingbo Liu, Yihang Qu, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan, “Neilf: Neural incident light field for physically-based material estimation,” in *European Conference on Computer Vision (ECCV)*, 2022.
- [13] Robert L. Cook and Kenneth E. Torrance, “A reflectance model for computer graphics,” *ACM Transactions on Graphics (ToG)*, vol. 1, no. 1, pp. 7–24, 1982.
- [14] Tizian Zeltner, Sébastien Speierer, Iliyan Georgiev, and Wenzel Jakob, “Monte carlo estimators for differential light transport,” *Transactions on Graphics (Proceedings of SIGGRAPH)*, vol. 40, no. 4, Aug. 2021.
- [15] 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*, vol. 65, no. 1, pp. 99–106, 2021.
- [16] Diederik P. Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” in *Proceedings of the 3rd ICLR*, 2015.
- [17] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala, “Pytorch: An imperative style, high-performance deep learning library,” in *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019.
- [18] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller, “Instant neural graphics primitives with a multiresolution hash encoding,” *ACM Trans. Graph.*, vol. 41, no. 4, pp. 102:1–102:15, July 2022.