Importance Sampling Guided Neural Radiosity

Huangsheng Du, Youcheng Cai*, Yutian Zhu, Peifeng Li, and Ligang Liu

University of Science and Technology of China

Abstract. The rendering equation, as a high-dimensional recursive integral equation, requires recursive expansion and Monte Carlo sampling for its solution, which results in significant computational complexity. Recently, the Neural Radiosity method has been proposed to solve the rendering equation using neural networks, effectively circumventing the need for explicit recursion. However, Neural Radiosity relies on Monte Carlo integration for training, which necessitates a large number of samples for optimization, leading to unstable optimization and a relatively low convergence rate. In this paper, we propose an Importance Sampling Guided Neural Radiosity framework, which systematically integrates Neural Radiosity with importance sampling to improve optimization performance. Firstly, we propose a joint optimization strategy that simultaneously trains the importance sampling module and the Neural Radiosity module. Specifically, our importance sampling module predicts the distribution to effectively enhance the optimization of Neural Radiosity, while the importance sampling module can also achieve rapid convergence through the radiance estimated by Neural Radiosity. Subsequently, we propose an Improved Kullback-Leibler divergence to mitigate the gradient conflict problem in standard KL divergence, thereby further improving convergence performance. Extensive experiments demonstrate that our framework achieves rapid convergence and stable optimization while maintaining high-quality rendering performance. The project code will be released upon acceptance.

Keywords: Rendering · Ray-tracing · Neural Rendering,Importance Sampling, Neural Networks.

1 Introduction

Solving the rendering equation [18] is a fundamental problem in computer graphics, involving the minimization of the discrepancy between the representation of the radiance function and the rendering equation, which is commonly addressed using radiosity approaches. Traditional radiosity approaches typically adapt the radiance function representation dynamically during the optimization process, including progressive meshing [1,21], wavelet radiosity [7], and hierarchical meshless basis functions [20], which often suffer from high storage demands and intensive computation. A noteworthy follow-up is Neural Radiosity [8], which employs neural networks as functional representations to solve the

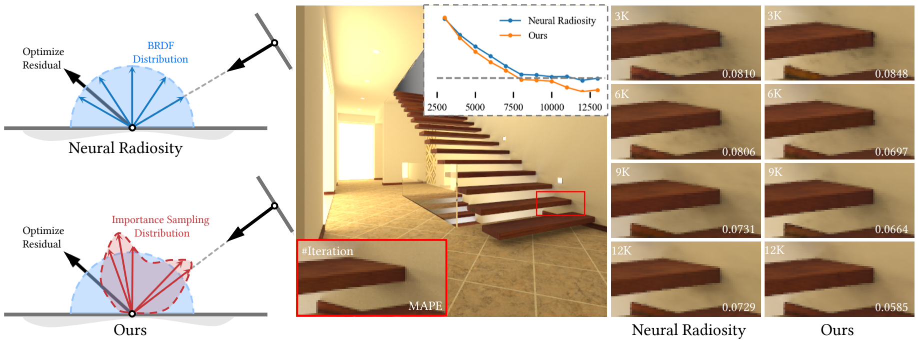

Fig. 1. We propose a novel Importance Sampling Guided Neural Radiosity framework that ensures fast convergence and stable optimization of Neural Radiosity. Our method achieves faster and more stable convergence while maintaining high-quality rendering performance. More results can be seen in the accompanying video. All reference results are rendered with path tracing, using 32,768 samples per pixel for images and 1,024 samples per pixel for videos.

rendering equation, thereby enabling expressive modeling of high-dimensional functions and efficient gradient-based optimization.

Neural Radiosity aims to solve the rendering equation by directly minimizing the norm of its residual $\mathcal{L}(\alpha)$:

$$\mathcal{L}(\alpha) = \left\| L_\alpha(x,\omega_o) - \left(L_e(x,\omega_o) + \int f(x,\omega_o,\omega_i) L_\alpha(x', -\omega_i)\, d\omega_i^\perp\right) \right\|_2^2, \tag{1}$$

where $L_\alpha(x,\omega_o)$ is a neural network that represents the radiance function and is parameterized by α, the remaining term corresponds to the rendering equation, which is typically approximated using Monte Carlo (MC) integration. Consequently, Neural Radiosity inevitably inherits the limitations of Monte Carlo methods, including unstable optimization and relative low convergence rate.

While Monte Carlo (MC) integration provides an unbiased estimate for complex integrals, it suffers from high variance, often necessitating a large number of samples for accurate results. In practice, various techniques have been developed to reduce variance and improve efficiency, such as antithetic sampling [36], quasi-random sampling, and importance sampling [17]. In this work, we focus on the concept of importance sampling and propose employing it to guide the optimization of Neural Radiosity. Importance sampling is commonly used in path guiding [33] to improve the efficiency and robustness of path tracing through the acquisition of improved sampling distributions that guide path construction. Similarly, we employ importance sampling to model an approximation of spatio-directional sampling distribution, which is naturally compatible with the continuous representation in Neural Radiosity. The challenges are twofold. The first is to jointly optimize the importance sampling block and the Neural Radiosity block. Naively training the importance sampling and Neural Radiosity blocks iteratively only leads to reduced training efficiency. The second challenge is the efficient optimization of the importance sampling block. Only rapid and stable optimization of the importance sampling block can ensure fast convergence and stable optimization of Neural Radiosity.

To address the aforementioned challenges, we propose an Importance Sampling Guided Neural Radiosity framework, which combines the strengths of Neural Radiosity and importance sampling to achieve fast convergence and stable optimization, while ensuring high rendering quality. Firstly, we propose a joint optimization strategy that simultaneously trains the importance sampling module (I-Module) and the Neural Radiosity module (N-Module) using stochastic gradient descent. Specifically, the I-Module is trained based on the radiance function approximated by the N-Module using a single-bounce estimation. The N-Module is optimized guided by the spatio-directional sampling distribution approximated by the I-Module. Here, we use a stop-gradient operator between the N-Module and I-Module to ensure computational efficiency in automatic differentiation. Secondly, to further improve the convergence rate, we propose an Improved Kullback-Leibler (IKL) divergence loss to mitigate the gradient conflict problem during I-Module optimization. Specifically, we observe that the negative component of the KL divergence between the target and optimized distributions is detrimental to optimizing their divergence. Thus, we propose an IKL divergence, whose gradient is defined as the sum of the absolute values of the positive and negative components of the KL divergence. This formulation mitigates the gradient conflict problem while retaining the influence of gradient magnitude. To the best of our knowledge, this is the first attempt to integrate a learnable importance distribution with Neural Radiosity. Extensive experiments demonstrate that our Importance Sampling Guided Neural Radiosity achieves state-of-the-art performance even with fewer training samples and iterations, achieving fast convergence and stable optimization.

- We present an Importance Sampling Guided Neural Radiosity framework that jointly optimizes the importance sampling module and the Neural Radiosity module, achieving both efficiency and effectiveness.
- We propose an Improved Kullback-Leibler (IKL) divergence loss to mitigate the gradient conflict problem and further improve the convergence rate.

2 Related Work

2.1 Radiosity Techniques

The early radiosity method [6] was based on the finite element method, by dividing the scene into discrete surface patches and simulating light transport by computing energy exchanges between these patches under the assumption of Lambertian reflection. Subsequently, Immel et al. [13] and Sillion et al. [29] extend radiosity to accommodate surfaces with general Bidirectional Scattering Distribution Functions (BSDFs). Hanrahan et al. [9] propose a rapid hierarchical radiosity algorithm to improve computational efficiency. Zatz [35] apply the

Galerkin method from finite element analysis to radiosity, using high-order polynomial basis functions to represent radiosity over surface regions, improving both accuracy and compactness. To accommodate complex scenes, various methods [1,21,7,20] propose a variety of basis functions for representing the solution to the rendering equation. However, these methods are associated with high storage requirements and significant computational cost.

Recently, Dahm et al. [3] observe structural similarities between reinforcement learning and light transport simulation, and thus use reinforcement learning to progressively learn the origins of light. Later, Hadadan et al. [8] present neural radiosity, which employs neural networks to approximate the solution to the rendering equation. To accommodate dynamic scenes, Coomans et al. [2] propose a real-time capable spatio-temporal cache to approximate a dynamic light field. Su et al. [30] propose high-dimensional grid-based features to encode dynamic scenes, which are able to handle dynamic scenes while maintaining only a moderate increase in computational and storage overhead. Our work aims to further improve the efficiency and effectiveness of neural radiosity for faster convergence and more stable optimization, guided by importance sampling.

2.2 Importance Sampling

Importance sampling is a popular sampling technique used to improve the efficiency of Monte Carlo (MC) integration [32], which enhances the convergence of MC integration by reducing variance [11]. Importance sampling has been commonly used in path guiding. This is because relying solely on BSDF-based sampling is insufficient for effectively handling complex scenes, which leads to a relatively high variance in the Monte Carlo (MC) estimator. Path guiding [16,19,34,28] estimates the distribution of incident radiance using previous samples and facilitates subsequent sampling. Recently, Müller et al. [25] and Dong et al. [5] propose the use of neural networks to approximate the importance sampling distribution, achieving both compactness and efficiency in representation. Our approach is similar to path guiding in that we both use neural networks to approximate the importance sampling distribution; however, we aim to approximate the distribution to stabilize the neural radiosity and train the neural network using only a single-bounce estimation.

2.3 Implicit Neural Representation

Implicit neural representation has achieved remarkable success in NeRF [22] for novel view synthesis, which uses multilayer perceptrons (MLPs) to encode radiance fields and render images via differentiable volume rendering. Subsequently, neural representations have been widely adopted and applied to a variety of tasks. For example, Diolatzis et al. [4] propose a neural renderer that enables direct control over parameters for global illumination and supports interactive inference. Müller et al. [26] propose a neural radiance caching method that employs neural networks to cache radiance for real-time global illumination rendering. In addition, a key factor in their success is the use of positional encoding, which enhances the neural network's ability to represent high-frequency details [31], which has been widely adopted in various fields, including neural radiosity [8].

3 Preliminary

3.1 Monte Carlo Integration

The formalization of light transport is based on the rendering equation [18]:

$$L_o(x,\omega_o) = L_e(x,\omega_o) + \int_{\mathcal{H}} f_s(x,\omega_o,\omega_i) L_i(x,\omega_i) d\omega_i^{\perp}, \tag{2}$$

where L_o denotes the outgoing radiance, L_e denotes the emitted radiance, L_i denotes the incident radiance and f_s denotes the bidirectional reflectance distribution function (BSDF), which are conditioned on position x, the direction of incidence ω_i and the outgoing direction ω_o. In addition, the integral is evaluated over the hemisphere $\mathcal{H}$, and $d\omega_i^{\perp} = \cos\theta d\omega_i$, where θ denotes the angle between ω_i and the surface normal.

Monte Carlo integration is used to obtain an estimate of the scattered integral L_s by averaging multiple samples:

$$\langle L_r(x,\omega_o)\rangle = \mathrm{E}_{\omega_i \sim p(\omega_i|x,\omega_o)} \left[\frac{f_s(x,\omega_o,\omega_i) L_i(x,\omega_i)\cos\theta}{p(\omega_i \mid x,\omega_o)} \right], \tag{3}$$

where $\langle L_r(x,\omega_o)\rangle$ is an unbiased estimate of the scattered radiance $L_s(x,\omega_o)$, and ω_i denotes the incident direction sampled from the directional probability distribution $p(\omega_i|x,\omega_o)$. The variance of the Monte Carlo estimator $\mathrm{V}\left[\langle L_r(x,\omega_o)\rangle\right]$ can be reduced if the sampling distribution $p(\omega_i|x,\omega_o)$ closely matches the shape of the integrand; it can even reach zero if $p \propto f_s L_i \cos\theta$. However, achieving this with predefined distributions such as the BSDF is challenging, as the incident radiance is unknown, which leads to a relatively high variance of the Monte Carlo estimator.

3.2 Neural Radiosity

The goal of Neural Radiosity [8] is to compute solutions to the rendering equation. Specifically, Neural Radiosity learns a radiance function $L_\alpha(x,\omega)$, which is represented by a neural networks with parameters α. Here, the residual of the rendering equation is minimized to optimize the network parameters α,

$$\begin{aligned} r_\alpha(x,\omega_o) = & L_\alpha(x,\omega_o) - L_e(x,\omega_o) \\ & - \int_{\mathcal{H}} f_s(x,\omega_o,\omega_i) L_\alpha(x'(x,\omega_i),-\omega_i) d\omega_i^{\perp}, \end{aligned} \tag{4}$$

where $L_\alpha(x'(x,\omega_i),-\omega_i)$ represents the incident radiance at position x from the incident direction ω_i, it is equivalent to the outgoing radiance in the direction

$-\omega_i$ at position x', which is the intersection of a ray originating from x along ω_i. The loss function of Neural Radiosity is defined as the norm of the residual:

$$\mathcal{L}_{NR}(\alpha) = \|r_\alpha(x, \omega_o)\|_2^2 = \int_{\mathcal{M}} \int_{\mathcal{H}} r_\alpha(x, \omega_o)^2 d\omega_o dx, \tag{5}$$

where the result is performed over all scene surfaces $\mathcal{M}$ and the hemisphere $\mathcal{H}$.

3.3 Importance Sampling

Importance sampling aims to develop a more effective sampling strategy for MC integration to improve efficiency. Radiance-based path guiding methods [5,12] typically use previous radiance estimates to approximate the incident radiance distribution for subsequent sampling. The objective is to learn a distribution $q(\omega_i|x, \omega_o)$ that approximates the target distribution $p(\omega_i|x, \omega_o)$. In the most general setting, $q(\omega_i|x, \omega_o)$ is considered to be proportional to the product of all terms in the integrand of the rendering equation,

$$q_\gamma(\omega_i|x, \omega_o) = f_s(x, \omega_o, \omega_i) L_i(x, \omega_i) \cos\theta F^{-1}, \tag{6}$$

where $F^{-1} = \int_{\mathcal{H}} L_i(x, \omega_i) f_s(x, \omega_o, \omega_i) cos\theta d\omega_i$ is the normalization factor, and q_γ is the distribution function defined by the parameters γ. During optimization, a common approach is to minimize the Kullback–Leibler (KL) divergence between the target distribution $p(\omega_i)$ and the learned distribution $q_\gamma(\omega_i)$,

$$\mathcal{L}_{KL}(p||q; \gamma) = \int_{\mathcal{H}} p(\omega_i) \log \frac{p(\omega_i)}{q_\gamma(\omega_i)} d\omega_i. \tag{7}$$

4 Method

We develop an Importance Sampling Guided Neural Radiosity framework that takes advantage of importance sampling to achieve fast convergence and stable optimization for Neural Radiosity. An overview of our framework is shown in Fig. ?????????2. First, we propose a joint optimization strategy that simultaneously trains the Importance Sampling module (I-Module) and the Neural Radiosity module (N-Module) (Sec. 4.1). The I-Module learns the incident radiance distribution through the outputs of the N-Module, while the N-Module is optimized using the sampling distribution approximated by the I-Module. Next, we propose an Improved Kullback-Leibler (IKL) divergence to further enhance the convergence rate, which mitigates the gradient conflict problem during I-Module optimization (Sec. 4.2).

4.1 Joint Optimization of I-Module and N-Module

Our goal is to solve the rendering equation efficiently and effectively by achieving fast convergence and stable optimization of Neural Radiosity. The rendering

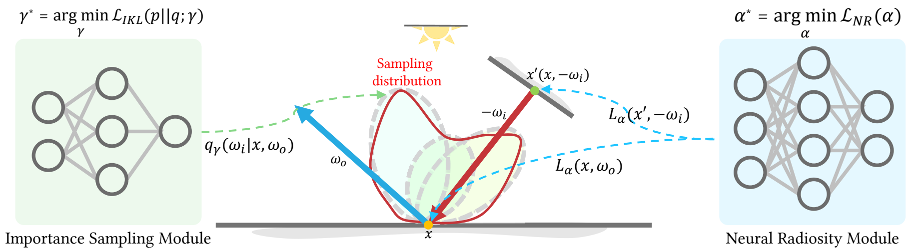

Fig. 2. Overview of the proposed Importance Sampling Guided Neural Radiosity framework. We introduce a joint optimization strategy that simultaneously trains the importance sampling module (I-Module) and the Neural Radiosity module (N-Module) using stochastic gradient descent. Here, the I-Module is trained based on the radiance function approximated by the N-Module, and the N-Module is optimized under the guidance of the importance sampling distribution approximated by the I-Module. In addition, an Improved Kullback-Leibler (IKL) divergence loss $\mathcal{L}_{IKL}$ is proposed to enhance the convergence rate.

equation is a high-dimensional, recursive integral equation that presents two main computational challenges: the recursive expansion of light transport and the numerical evaluation of integrals. Neural radiosity effectively addresses the recursive expansion by self-training; however, the sampling strategy employed for MC integration remains inefficient. Importance sampling develops an approximate sampling distribution for effective MC integration but requires a large number of computationally expensive recursive path-traced samples. Therefore, we propose the joint optimization of the Importance Sampling module (I-Module) and the Neural Radiosity module (N-Module), thereby combining the advantages of both modules. In our optimization stage, we also introduce the stop-gradient operator between the N-Module and I-Module to ensure computational efficiency in automatic differentiation. In addition, the stop-gradient prevents abnormal gradients in one module from affecting the optimization of the other. Since each new iteration incorporates the updated outputs from the other module, the negative impact of the stop-gradient on optimization is effectively mitigated.

Specifically, the N-Module predicts the scattered radiance $L_\alpha(x,\omega_o)$ with learned parameters α, and its MC estimation is guided by the sampling distribution learned by the I-Module. The I-Module learns the sampling distribution $q_\gamma(\omega_i|x,\omega_o)$ with learned parameters γ for the MC estimation, which is informed by the radiance estimates provided by the N-Module.

N-Module Optimization The objective of the N-Module is to predict the scattered radiance $L_\alpha(x,\omega_o)$. We propose to minimize the residual norm $\mathcal{L}_{NR}(\alpha)$ as described in Eq. 5, and the desired solution is,

$$\alpha^* = \arg\min_\alpha \mathcal{L}_{NR}(\alpha). \tag{8}$$

Here, the residual norm is estimated using MC integration, and gradients of the radiance field are computed using automatic differentiation. Due to the self-learning approach used to optimize the network, the neural network appears twice in the gradient computation. Following Coomans et al. [2], we stop the gradient flow of the incident radiance estimates to achieve better performance,

$$\begin{aligned} r_\alpha(x, \omega_o) =& L_\alpha(x, \omega_o) - L_e(x, \omega_o) \\ & - \int_{\mathcal{H}} f_s(x, \omega_o, \omega_i) sg(L_\alpha(x'(x, \omega_i), -\omega_i)) d\omega_i^{\perp}, \end{aligned} \tag{9}$$

where $sg(\cdot)$ denotes the stop-gradient operator, which treats $L_\alpha(x', -\omega_i)$ as a constant with respect to backpropagation.

MC Estimation. The MC estimate of the residual norm $\mathcal{L}_{NR}(\alpha)$ is given by

$$\mathcal{L}_{NR}(\alpha) \approx \frac{1}{M} \sum_{m=1}^{M} \frac{r_\alpha\left(x^m, \omega_o^m\right)^2}{p_{\mathcal{M}}(x^m, \omega_o^m)}, \tag{10}$$

where M is the number of samples, and the surface positions x^m and outgoing directions ω_o^m are sampled according to the probability density $p_{\mathcal{M}}(x^m, \omega_o^m)$. Here, we employ uniform sampling in our experiments. And the integrator term of $r_\alpha\left(x^m, \omega_o^m\right)$ is computed as,

$$\begin{aligned} \text{Int}\{r_\alpha(x, \omega_o)\} =& E_{\omega_i \sim q}\left[\frac{f_s(x, \omega_o, \omega_i) L_i(x, \omega_i) cos\theta}{q(\omega_i \mid \gamma)}\right] \\ \approx & \frac{1}{N} \sum_{n=1}^{N} \frac{f_s(x, \omega_o, \omega_i) L_\alpha(x'(x, \omega_i), -\omega_i) \cos\theta}{q_\gamma(\omega_i | x, \omega_o)}. \end{aligned} \tag{11}$$

Here, the incident direction ω_i is sampled according to the distribution $q_\gamma(\omega_i|x, \omega_o)$.

Gradients. The gradient of the residual $r_\alpha\left(x^m, \omega_o^m\right)$ is estimated by evaluating the hemispherical integral using MC integration,

$$\begin{aligned} & \nabla_\alpha r_\alpha(x^m, \omega_o^m) = L_\alpha(x^m, \omega_o^m) \\ & - \frac{1}{N} \sum_{n=1}^{N} \frac{f_s(x^m, \omega_o^m, \omega_i^{m,n}) sg(L_\alpha(x'(x^m, \omega_i^{m,n}), -\omega_i^{m,n})) \cos\theta}{sg(q_\gamma(\omega_i | x, \omega_o))}. \end{aligned} \tag{12}$$

Here, $\omega_i^{m,n}$, $n \in 1, \cdots, N$, represents a set of N incident direction samples (x^m, ω_o^m). The operator $sg(q_\gamma(\omega_i|x, \omega_o))$ is employed to eliminate the influence of gradients on the I-Module.

I-Module Optimization In Neural Radiosity, the integrator term of the residual is estimated by sampling from the local BSDF distribution, which is inefficient and struggles to sample directions that contribute significantly to the radiance estimate. The objective of the I-Module is to learn a distribution $q(\omega_i|x, \omega_o)$ that approximates the target distribution $p(\omega_i|x, \omega_o)$ for importance sampling of the scattering direction. The Kullback–Leibler (KL) divergence between $p(\omega_i)$ and $q_\gamma(\omega_i)$ is minimized to obtain the optimal parameters γ,

$$\gamma^* = \arg\min_{\gamma} \mathcal{L}_{KL}(p|q;\gamma). \tag{13}$$

Von Mises-Fisher Mixtures. We use the von Mises-Fisher (vMF) distribution [14] as the basis function of $q_\gamma(\omega_i)$, which is defined as,

$$\upsilon(\omega|\mu,\kappa) = \frac{\kappa}{2\pi(1-exp(-2\kappa))}\exp(\kappa(\mu^T\omega-1)), \tag{14}$$

where $\upsilon \in S^2$ and $\kappa \in [0,+\infty)$ define the mean direction and sharpness of the vMF distribution. The vMF distribution possesses strong expressive power with fewer parameters, and supports efficient sampling and computation. The vMF mixture model (VMM) with K vMF components is defined to represent $q_\gamma(\omega_i)$.

$$q(\omega|\gamma) = \sum_{s=1}^{K} \lambda_s \upsilon(\omega \mid \mu_s, \kappa_s), \tag{15}$$

where $(\lambda_s, \mu_s, \kappa_s) \in \gamma$ are the parameters of each vMF component. The I-Module learns the parameters γ, which define the sampling distribution.

Kullback–Leibler Divergence. As recalled in Eq. 6, the objective is to ensure that $q(\omega_i|x,\omega_o)$ is proportional to the full product of all terms in the integrand of the rendering equation, i.e., $q_\gamma(\omega_i|x,\omega_o) \propto f_s(x,\omega_o,\omega_i)L_i(x,\omega_i)|\cos\theta|$. Given the considerable challenge, most research has focused on the simpler setting where $q_\gamma(\omega_i|x,\omega_o) \propto L_i(x,\omega_i)$ [3,24,5], and only a few attempts [10,25] have been made to consider the full product. However, all of these approaches ignore the normalization factor F, since F is unknown. In contrast, within our framework, the normalization factor F can be readily computed from the N-Module. Accordingly, we define the ideal target distribution as $p(\omega_i|x,\omega_o) = f_s(x,\omega_o,\omega_i)L_i(x,\omega_i)|\cos\theta|F^{-1}$, i.e., the normalized form of the integrand, which minimizes the variance of Monte Carlo sampling. The Kullback–Leibler (KL) divergence, as defined in Eq. 7, can then be estimated using a MC estimator with N samples.

$$\mathcal{L}_{KL}(p||q;\gamma) \approx \frac{1}{N}\sum_{n=1}^{N} \frac{p(\omega_i^n)}{\tilde{p}_\gamma(\omega_i^n)} \log\frac{p(\omega_i^n)}{q_\gamma(\omega_i^n)}, \tag{16}$$

where $\tilde{p}$ is a combination of the BSDF importance sampling and the learned sampling distribution.

Gradients. By taking the derivative with respect to γ, we obtain the MC estimate of the gradient,

$$\begin{aligned}\nabla_\gamma \mathcal{L}_{KL}(p(x^m,\omega_i)||q_\gamma(x^m,\omega_i) = \\ -\sum_{n=1}^{N} \frac{sg(L_\alpha(x'(x^m,\omega_i^{m,n})), -\omega_i^{m,n})\nabla_\gamma q_\gamma(\omega_i^n)F^{-1}}{\hat{q}_\gamma(\omega_i^n)q_\gamma(\omega_i^n)}.\end{aligned} \tag{17}$$

The operator $sg(L_\alpha(\cdot))$ is employed to eliminate the influence of gradients on the N-Module.

Discussion Our approach to the joint optimization of the I-Module and N-Module has three distinct advantages. First, this allows us to obtain the normalization factor F, which is computed from the scattered radiance predicted by the N-Module. Taking the normalization factor F into account results in more accurate optimization. The second advantage is that our I-Module training only requires one ray-tracing step per sampled point to gather the predicted scattered emittance at each intersected point, which is different from path-guided methods that require a large number of recursive path-traced samples. Lastly, our approach achieves fast convergence and stable optimization of the N-Module. By being guided by the importance sampling distribution learned by the I-Module, the N-Module is able to efficiently generate samples.

4.2 Improved KL-divergence

Observation. Our approach relies on gradient descent to optimize the model parameters, where the magnitude of the gradient reflects the convergence rate. However, the negative contributions of the integrand in the KL divergence weaken the effectiveness of gradient descent, a phenomenon we refer to as gradient conflict. Specifically, as stated in Eq. 7, the integrand of the KL divergence, $I_{KL} = p(\omega_i) \log \frac{p(\omega_i)}{q_\gamma(\omega_i)} d\omega_i$, is not guaranteed to be non-negative everywhere. During Monte Carlo (MC) estimation, it is possible to sample points where the integrand is negative, i.e., $I_{KL}(\omega|\gamma) < 0$, and after the parameters are optimized via gradient backpropagation ($\gamma \to \gamma'$), the value of the integrand at these points may decrease further, i.e., $I_{KL}(\omega|\gamma') < I_{KL}(\omega|\gamma) < 0$. This indicates that the optimization step moves in a direction that increases the local KL divergence, which contradicts the overall objective of minimizing the global KL divergence. Therefore, we propose an Improved Kullback-Leibler (IKL) by taking the absolute value into the integrand component:

$$\mathcal{L}_{IKL}(p||q;\gamma) = \int_{\mathcal{H}} p(\omega_i) \mathrm{abs}(\log \frac{\mathrm{p}(\omega_\mathrm{i})}{\mathrm{q}_\gamma(\omega_\mathrm{i})}) \mathrm{d}\omega_\mathrm{i}. \tag{18}$$

Interpretation. The KL divergence is guaranteed to be non-negative due to the convexity of the logarithm function, as established by Jensen's inequality. The KL divergence can be decomposed as $D_{KL} = D_{p>q} + D_{p\leq q}$, where $D_{p>q}$ and $D_{p\leq q}$ denote the positive and negative contributions of the integrand,

$$\begin{aligned} D_{p>q} &= \int_{\Omega} p(\omega) \log \frac{p(\omega)}{q(\omega)} \mathbf{1}_{p>q} d\omega, \\ D_{p\leq q} &= \int_{\Omega} p(\omega) \log \frac{p(\omega)}{q(\omega)} \mathbf{1}_{p\leq q} d\omega, \end{aligned} \tag{19}$$

where $\mathbf{1}_{(\cdot)}$ denotes the indicator function, which evaluates to 1 if the condition inside the parentheses is true, and 0 otherwise. The KL divergence equals zero if and only if $p = q$. In this case, both $D_{p>q}$ and $D_{p\leq q}$ are zero. Conversely, if $D_{p>q} = D_{p\leq q} = 0$, it follows that $p = q$, and thus the KL divergence is zero.

Therefore, $D_{KL} = 0$ is equivalent to $D_{p>q} = D_{p\leq q} = 0$, which is also equivalent to $D_{IKL} = 0$.

We use $\mathcal{L}_{IKL}$ as the loss function to optimize the I-Module. Compared to the KL divergence $\mathcal{L}_{KL}$, the integrand of $\mathcal{L}_{IKL}$ is strictly non-negative, which ensures that $\mathcal{L}_{IKL} \geq \mathcal{L}_{IKL} \geq 0$. This property demonstrates better convergence compared to the KL divergence. It should be noted that in our approach, we can obtain an accurate estimation of $p(\omega_i)$ guided by the N-Module, which allows us to compute $\mathbf{1}_{p>q}$ precisely.

The gradient of $\mathcal{L}_{IKL}$ is computed using the MC estimator as,

$$\begin{aligned} &\nabla_\gamma \mathcal{L}_{IKL}(p(x^m, \omega_i)||q_\gamma(x^m, \omega_i) = \\ &\quad -\sum_{n=1}^{N} \frac{sg(L_\alpha(x'(x^m, \omega_i^{m,n})), -\omega_i^{m,n})\nabla_\gamma q_\gamma(\omega_i^n)F^{-1}}{\hat{q}_\gamma(\omega_i^n)q_\gamma(\omega_i^n)}(2 \cdot \mathbf{1}_{p>q} - 1), \end{aligned} \tag{20}$$

where $p(\omega_i|x, \omega_o) = f_s(x, \omega_o, \omega_i)L_\alpha(x, \omega_i)|\cos\theta|F^{-1}$, and $q_\gamma(\omega_i)$ is the predicted by the I-Module.

4.3 Implementation

Our framework is implemented using the Mitsuba 3 renderer [15]. The neural network is implemented using PyTorch [27]. We employ a learning rate of 0.005 using the Adam optimizer for training. All experiments are conducted on an Intel Core i7-14700 CPU and a single NVIDIA RTX 4090 GPU.

N-Module. In our experiments, the N-Module consists of four linear layers, each with a width of 512 and a ReLU activation function. The N-Module takes as input the surface position x, the outgoing direction ω_o, and auxiliary information including the surface normal and albedo, and outputs the radiance as three non-negative RGB values by applying the absolute value. In addition, we employ trainable spatial encoding using a multi-resolution hash grid [23] to model spatial features.

I-Module. The I-Module is composed of three linear layers, each with a width of 64 and a ReLU activation function. The I-Module takes as input the surface position x, the outgoing direction ω_o, the surface normal $\mathbf{n}$, and roughness ρ, and outputs a set of parameters representing a vMF mixture distribution. Here, the spatial position is also encoded by the trainable multi-resolution spatial embedding to enable efficient training. The output of the I-Module is a 32-dimensional vector that parameterizes a mixture of 8 vMF components to model the spatial directional distribution for use as sampling distributions. The normalization factor F is computed based on the luminance of radiance predicted by the N-Module.

5 Experiments

5.1 Comparisons

We compare our method to Neural Radiosity [8]. For all experiments, images are rendered at a resolution of 512 × 512, and image quality is evaluated using

$N = 4$	Neural Radiosity					Ours				
Scene	Step	Time	Memory	MAPE ↓	MSE ↓	Step	Time	Memory	MAPE ↓	MSE ↓
LivingRoom	27k	43m	1.74G	0.024	6.42e-4	27k	51m	1.76G	**0.022**	**5.51e-4**
Bedroom	21k	68m	2.03G	0.051	1.19e-2	21k	74m	2.07G	**0.049**	**1.17e-2**
ModernHall	25k	49m	2.00G	0.033	9.68e-4	25k	60m	2.07G	**0.024**	**9.04e-4**
WoodenStaircase	25k	157m	2.03G	0.040	1.34e-4	25k	162m	2.07G	**0.038**	**1.33e-4**
Bathroom	25k	131m	2.00G	0.181	4.68e-1	25k	154m	2.03G	**0.153**	**4.53e-1**
GrayandWhite	24k	75m	2.00G	0.053	1.76e-2	24k	79m	2.07G	**0.045**	**1.68e-2**
VeachDoor	30k	32m	1.74G	0.081	7.01e-2	30k	44m	1.76G	**0.080**	**6.64e-2**
VeachEgg	29k	48m	2.00G	0.031	2.602	29k	58m	2.03G	**0.027**	**2.600**

Table 1. Quantitative comparisons are conducted with the number of samples set to $N = 4$. Our method achieves the best scores in both MAPE and MSE.

$N = 32$	NeuralRadiosity					Ours				
Scene	Step	Time	Memory	MAPE ↓	MSE ↓	Step	Time	Memory	MAPE ↓	MSE ↓
LivingRoom	30k	84m	10.2G	0.018	3.75e-4	23k	73m	10.5G	0.019	4.68e-4
Bedroom	26k	106m	12.0G	0.046	1.13e-2	16k	72m	12.4G	0.049	1.16e-2
ModernHall	25k	83m	12.0G	0.023	9.17e-4	13k	46m	12.4G	0.028	**8.94e-4**
WoodenStaircase	24k	198m	12.0G	0.041	1.34e-3	15k	130m	12.4G	**0.041**	1.38e-3
Bathroom	20k	143m	12.0G	0.213	5.01e-1	11k	83m	12.3G	**0.207**	5.08e-1
GrayandWhite	24k	96m	12.0G	0.045	2.15e-2	15k	65m	12.4G	0.053	**1.70e-2**
VeachDoor	22k	52m	10.3G	0.063	1.48e-2	12k	33m	10.5G	0.067	**1.39e-2**
VeachEgg	30k	91m	12.0G	0.051	2.81e-2	14k	47m	12.3G	0.051	2.86e-2

Table 2. Quantitative comparisons on the number of samples $N = 32$. Our method achieves comparable scores in terms of MAPE and MSE even with fewer iterations.

MAPE and MRSE. We evaluate the rendering results across 8 scenes for comparison. For a fair comparison, we keep the same training settings, including batch size and grid resolution. To demonstrate the efficiency of our method, we conducted experiments under two sampling settings: $N = 4$ and $N = 32$, where N represents the number of samples of the incident direction.

The quantitative results are shown in Table 1 and Table 2. Our method consistently achieves better efficiency and performance across all scenes. For fewer training samples, our method shows a significant improvement in rendering quality in terms of MAPE and MRSE, which demonstrates the effectiveness of importance sampling. As shown in Table 2, our method requires only half the number of iterations to achieve rendering quality comparable to Neural Radiosity, demonstrating the faster and more stable convergence performance of our method. The quantitative results are illustrated in a series of figures (see Figs. 4 and following), where we provide rendered images of our method and Neural Radiosity at different iteration step, demonstrating that our method can ensure rendering quality while achieving faster convergence performance.

We further compared the performance of our method and Neural Radiosity under equal-time training conditions. We compare our method with Neural Radiosity under the same number of samples and training duration, based on the training step setting of our method listed in Table 2. We compute the residual of both methods. As defined in Eq. 5 a lower residual indicates a solution closer to the true solution of the rendering equation. To reduce the impact of randomness, we sampled 2^{20} space-direction points per scene, and applied 2^{10} Monte Carlo samples per point. The final results are presented in Table 3. Experimental results show that under equal-time conditions, all results of our method, except for the Living Room scene, consistently yield lower residuals. The anomaly observed in the living room scene is mainly due to its relatively simple lighting configuration, where the benefit from importance sampling is offset by the additional training cost. Nevertheless, since our current implementation is built upon Python and the PyTorch framework, the performance of the I-Module can be further improved in future work.

Scene	Living Room	Bed Room	Modern Hall	Wooden Staircase	Bath Room	Gray and White	Veach Door	Veach Egg
Neural Radiosity	9.521e−1	8.473e−3	2.010e−2	4.399e−3	4.669e−2	5.644e−2	8.932e−2	1.220e0
Ours(Equal Time)	9.597e−1	8.198e−3	2.007e−2	4.051e−3	3.968e−2	5.572e−2	8.717e−2	1.078e0
Ours(Equal Samples)	9.491e−1	7.998e−3	1.993e−2	3.941e−3	3.941e−2	5.321e−2	7.564e−2	1.024e0

Table 3. Equal-time and equal-sample comparisons. Although our method takes longer per iteration than Neural Radiosity, it achieves better performance under the same training time. Furthermore, when trained with the same number of samples, our method shows a clear improvement over Neural Radiosity.

5.2 Performance

The performance of the I-Module is crucial for the training efficiency of the N-Module. Our I-Module is extremely lightweight, incurring significantly lower training and inference overhead compared to the N-Module. Here, we analyze the training efficiency in comparison to Neural Radiosity. Due to the additional computational load introduced by the I-Module, our method is approximately 10% slower per iteration than Neural Radiosity. However, because the sampling distribution predicted by the I-Module enables the N-Module to learn the radiance function more effectively, our method can achieve the desired rendering quality using fewer training samples and iterations. This advantage becomes more pronounced in scenes with complex illumination or multiple reflections. As shown in our accompanying video, in challenging cases such as the Veacg Egg scene, our method can more clearly reproduce the caustic patterns on the walls caused by refraction and reflection through the glass egg.

5.3 Ablation Study

In this section, we study the effectiveness of each component in our method. Specifically, we remove the following components and compare the resulting ablated models to the full version: the importance sampling module (I-Module) and

Ablation	MAPE			MSE		
Iteration	w/o I-Module	w/o IKL	Full	w/o I-Module	w/o IKL	Full
4k	0.1133	0.2746	**0.0609**	0.0500	0.2887	**0.0296**
10K	0.0855	0.1077	**0.0554**	**0.0309**	0.0684	0.0322
20K	0.0579	0.0552	**0.0524**	0.0286	**0.0281**	0.0286
30K	0.0511	0.0511	**0.0501**	0.0281	0.0281	**0.0280**

Table 4. Ablation studies on the proposed I-Module and IKL divergence. The best results are highlighted.

Fig. 3. We present the MAPE results during training. The curves demonstrate that incorporating the I-Module leads to faster convergence, while our proposed IKL significantly improves training stability. The convergence plots correspond to the ablation studies: w/o the I-Module, w/o the IKL, and the full model. Left: the convergence plots. Right: error map at step 5000.

the Improved KL-divergence (IKL). We perform ablation studies on the Veach Egg scene and evaluate the rendering results at various training iterations.

Effects of the I-Module. In fact, vanilla Neural Radiosity corresponds to our model without the I-Module. As shown in Table 4 and Fig. 3, the absence of the I-Module leads to a decrease in performance and results in a significant reduction in quality at the same number of training iterations. By incorporating the I-Module, our model achieves faster and more stable convergence performance.

Effects of the IKL. We demonstrate the effectiveness of the IKL divergence by replacing it with the standard KL divergence. As demonstrated in Table 4 and Fig. 3, our IKL exhibits faster convergence compared to the standard KL divergence. Due to the instability during the early stages of training, the standard KL divergence fails to effectively guide the distribution. In contrast, our proposed IKL yields more stable convergence of the guiding distribution, thereby enabling more effective importance sampling.

6 Conclusion

We propose the Importance Sampling Guided Neural Radiosity framework, which efficiently combines Neural Radiosity with importance sampling to achieve rapid convergence and stable optimization. Firstly, we propose a joint optimization strategy that simultaneously trains the importance sampling module and the Neural Radiosity module, effectively leveraging the strengths of both modules while compensating for their respective weaknesses. Subsequently, we propose an Improved Kullback-Leibler (IKL) divergence loss to further enhance the convergence rate. Since our framework can provide an accurate distribution $p(\omega_i)$, we are able to effectively eliminate the negative contributions in the KL divergence. Experimental results demonstrate that our method surpasses Neural Radiosity in both efficiency and effectiveness.

Limitation. Our work has several limitations. First, our approach relies on Neural Radiosity, and therefore inherently inherits its limitation of being unsuitable for dynamic lighting. Future work could explore importance sampling strategies in higher-dimensional spaces to potentially overcome this limitation. Second, although Neural Radiosity has demonstrated promising results in modeling diffuse, specular, and certain glossy effects, it still struggles to accurately capture high-frequency glossy reflections. Identifying a more suitable function space that better approximates the solution to the rendering equation remains an important direction for future research. We aim to address this challenge in future work. Finally, our method facilitates the learning of the radiance function by providing higher-quality samples to the neural network. However, from an optimization perspective, the training process still relies on first-order optimization methods. Future work could explore more efficient optimization frameworks from a mathematical standpoint.

References

1. Cohen, M.F., Chen, S.E., Wallace, J.R., Greenberg, D.P.: A progressive refinement approach to fast radiosity image generation. In: Proceedings of Conference on Computer Graphics and Interactive Techniques. pp. 75–84 (1988)
2. Coomans, A., Dominci, E.A., Döring, C., Mueller, J.H., Hladky, J., Steinberger, M.: Real-time neural rendering of dynamic light fields. In: Computer Graphics Forum. vol. 43, p. e15014 (2024)
3. Dahm, K., Keller, A.: Learning light transport the reinforced way. In: ACM SIGGRAPH Talks, pp. 1–2 (2017)
4. Diolatzis, S., Philip, J., Drettakis, G.: Active exploration for neural global illumination of variable scenes. ACM Transactions on Graphics (TOG) **41**(5), 1–18 (2022)
5. Dong, H., Wang, G., Li, S.: Neural parametric mixtures for path guiding. In: ACM SIGGRAPH Papers. pp. 1–10 (2023)
6. Goral, C.M., Torrance, K.E., Greenberg, D.P., Battaile, B.: Modeling the interaction of light between diffuse surfaces. ACM SIGGRAPH Computer Graphics **18**(3), 213–222 (1984)
7. Gortler, S.J., Schröder, P., Cohen, M.F., Hanrahan, P.: Wavelet radiosity. In: Proceedings of Conference on Computer Graphics and Interactive Techniques. pp. 221–230 (1993)
8. Hadadan, S., Chen, S., Zwicker, M.: Neural radiosity. ACM Transactions on Graphics (TOG) **40**(6), 1–11 (2021)
9. Hanrahan, P., Salzman, D., Aupperle, L.: A rapid hierarchical radiosity algorithm. In: Proceedings of the 18th Annual conference on Computer graphics and interactive techniques. pp. 197–206 (1991)

10. Herholz, S., Elek, O., Schindel, J., Křivánek, J., Lensch, H.P.: A unified manifold framework for efficient brdf sampling based on parametric mixture models. In: Proceedings of the Eurographics Symposium on Rendering: Experimental Ideas & Implementations. pp. 41–52 (2018)
11. Hesterberg, T.: Weighted average importance sampling and defensive mixture distributions. Technometrics **37**(2), 185–194 (1995)
12. Huang, J., Iizuka, A., Tanaka, H., Komura, T., Kitamura, Y.: Online neural path guiding with normalized anisotropic spherical gaussians. ACM Transactions on Graphics (TOG) **43**(3), 1–18 (2024)
13. Immel, D.S., Cohen, M.F., Greenberg, D.P.: A radiosity method for non-diffuse environments. Acm Siggraph Computer Graphics **20**(4), 133–142 (1986)
14. Jakob, W.: Numerically stable sampling of the von mises-fisher distribution on s^2 (and other tricks). Interactive Geometry Lab, ETH Zürich, Tech. Rep **6** (2012)
15. Jakob, W., Speierer, S., Roussel, N., Nimier-David, M., Vicini, D., Zeltner, T., Nicolet, B., Crespo, M., Leroy, V., Zhang, Z.: Mitsuba 3 renderer (2022), `https://mitsuba-renderer.org`
16. Jensen, H.W.: Importance driven path tracing using the photon map. In: Proceedings of the Eurographics Workshop. pp. 326–335 (1995)
17. Jensen, H.W., Arvo, J., Dutre, P., Keller, A., Owen, A., Pharr, M., Shirley, P.: Monte carlo ray tracing. In: ACM SIGGRAPH Papers. vol. 5, p. 340769537 (2003)
18. Kajiya, J.T.: The rendering equation. In: Proceedings of Conference on Computer Graphics and Interactive Techniques. pp. 143–150 (1986)
19. Lafortune, E.P., Willems, Y.D.: A 5d tree to reduce the variance of monte carlo ray tracing. In: Proceedings of the Eurographics Workshop. pp. 11–20 (1995)
20. Lehtinen, J., Zwicker, M., Turquin, E., Kontkanen, J., Durand, F., Sillion, F.X., Aila, T.: A meshless hierarchical representation for light transport. In: ACM SIGGRAPH Papers, pp. 1–9 (2008)
21. Lischinski, D., Tampieri, F., Greenberg, D.P.: Discontinuity meshing for accurate radiosity. IEEE Computer Graphics and Applications **12**(06), 25–39 (1992)
22. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
23. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM transactions on graphics (TOG) **41**(4), 1–15 (2022)
24. Müller, T., Gross, M., Novák, J.: Practical path guiding for efficient light-transport simulation. In: Computer Graphics Forum. vol. 36, pp. 91–100 (2017)
25. Müller, T., McWilliams, B., Rousselle, F., Gross, M., Novák, J.: Neural importance sampling. ACM Transactions on Graphics (ToG) **38**(5), 1–19 (2019)
26. Müller, T., Rousselle, F., Novák, J., Keller, A.: Real-time neural radiance caching for path tracing. ACM Transactions on Graphics (TOG) **40**(4), 1–16 (2021)
27. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch. In: NIPS-W (2017)
28. Ruppert, L., Herholz, S., Lensch, H.P.: Robust fitting of parallax-aware mixtures for path guiding. ACM Transactions on Graphics (TOG) **39**(4), 147–1 (2020)
29. Sillion, F.X., Arvo, J.R., Westin, S.H., Greenberg, D.P.: A global illumination solution for general reflectance distributions. In: Proceedings of Conference on Computer Graphics and Interactive Techniques. pp. 187–196 (1991)

30. Su, R., Dong, H., Ren, J., Jin, H., Chen, Y., Wang, G., Li, S.: Dynamic neural radiosity with multi-grid decomposition. In: SIGGRAPH Asia Papers. pp. 1–12 (2024)
31. Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J., Ng, R.: Fourier features let networks learn high frequency functions in low dimensional domains. Advances in Neural Information Processing Systems **33**, 7537–7547 (2020)
32. Tokdar, S.T., Kass, R.E.: Importance sampling: a review. Wiley Interdisciplinary Reviews: Computational Statistics **2**(1), 54–60 (2010)
33. Vorba, J., Hanika, J., Herholz, S., Müller, T., Křivánek, J., Keller, A.: Path guiding in production. In: ACM SIGGRAPH Courses, pp. 1–77 (2019)
34. Vorba, J., Karlík, O., Šik, M., Ritschel, T., Křivánek, J.: On-line learning of parametric mixture models for light transport simulation. ACM Transactions on Graphics (TOG) **33**(4), 1–11 (2014)
35. Zatz, H.R.: Galerkin radiosity: A higher order solution method for global illumination. In: Proceedings of Conference on Computer Graphics and Interactive Techniques. pp. 213–220 (1993)
36. Zhang, C., Dong, Z., Doggett, M., Zhao, S.: Antithetic sampling for monte carlo differentiable rendering. ACM Transactions on Graphics (TOG) **40**(4), 1–12 (2021)

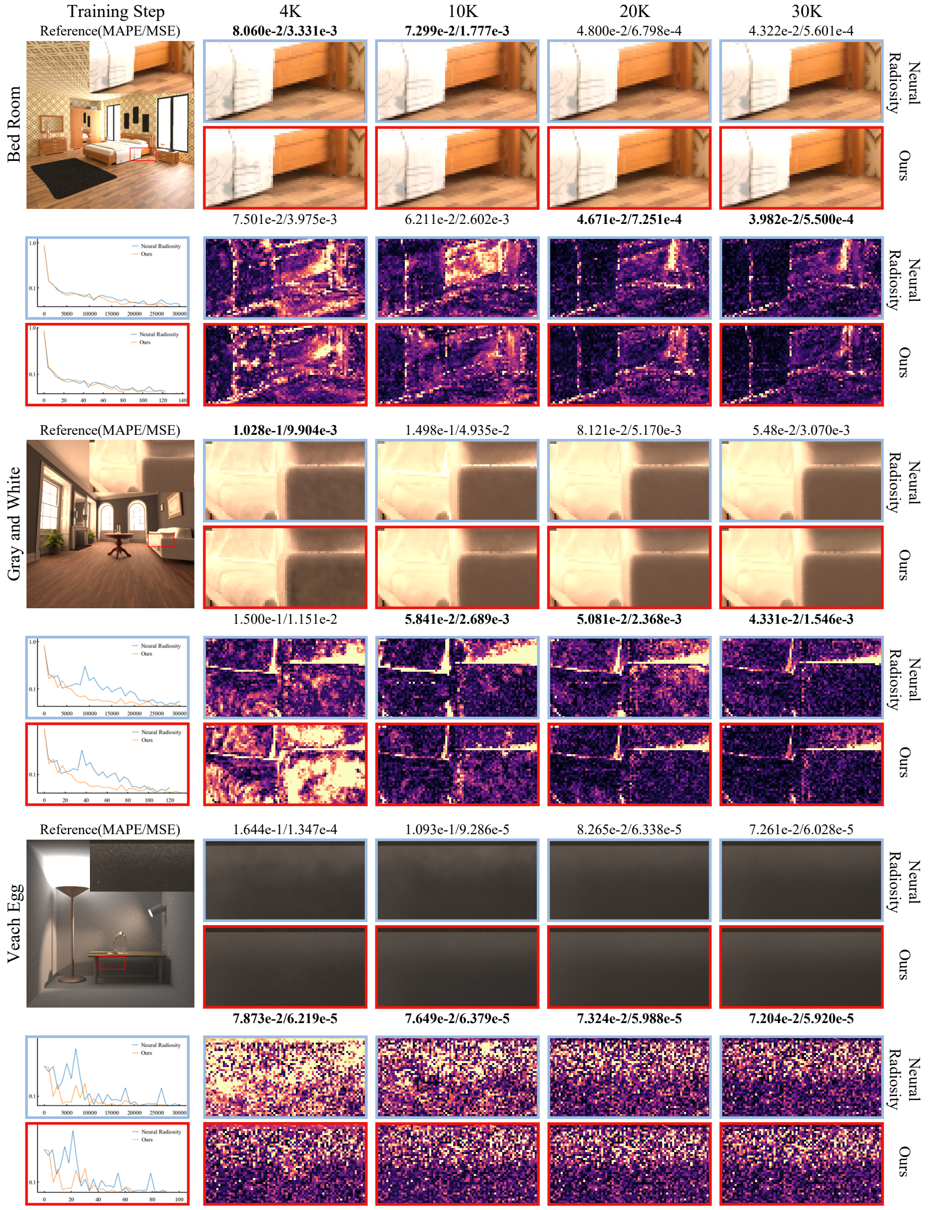

Fig. 4. Qualitative comparisons of our approach($N = 32$) with Neural Radiosity on Bed Room, Gray and White and Veach Egg scenes. Our method produces high-quality rendering results. For the diagram, the blue boxes indicate comparisons under the same number of iterations (measured in steps), while the red boxes correspond to comparisons under equal-time conditions (measured minutes). Zooming in is recommended for better visualization. An accompanying video is provided for dynamic qualitative comparisons.

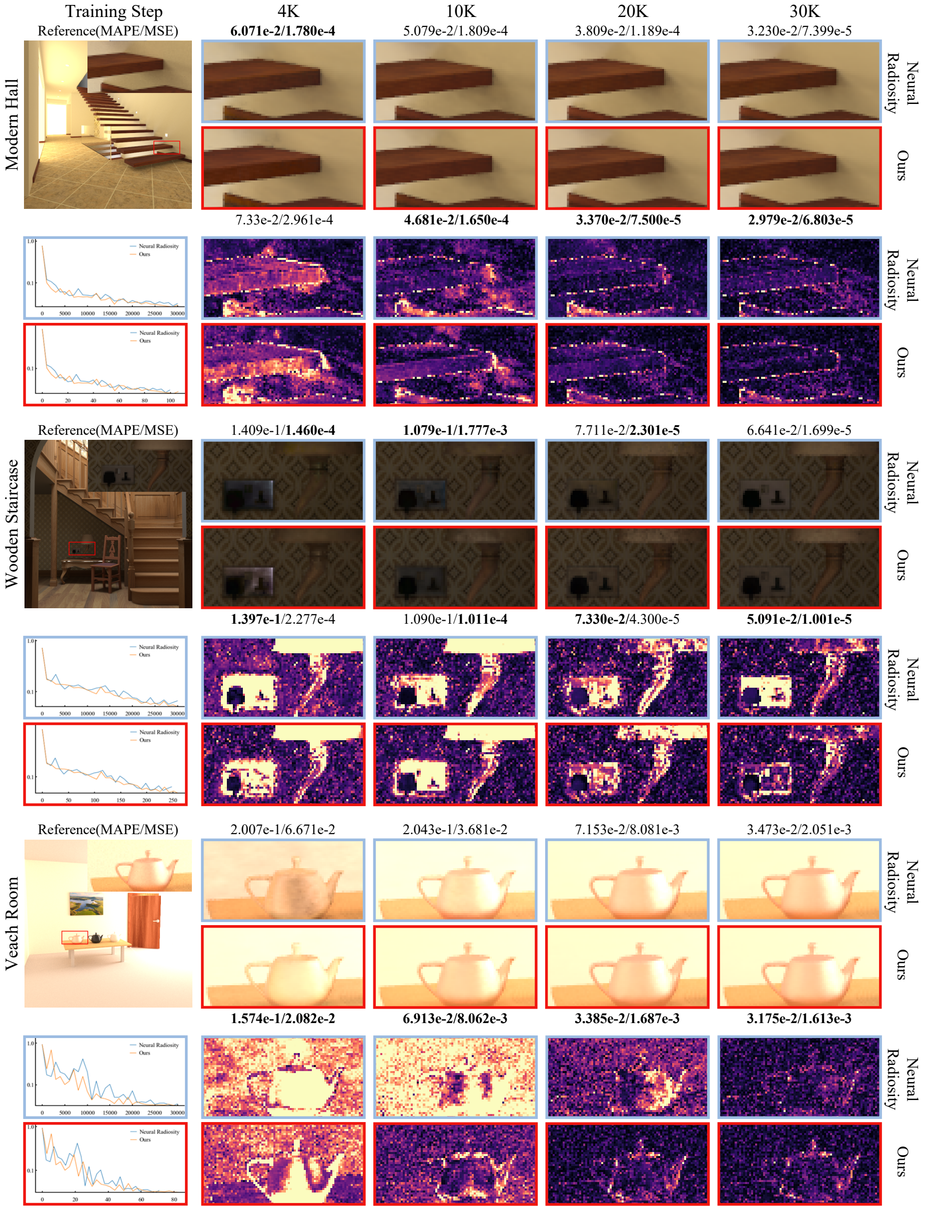

Fig. 5. Qualitative comparisons of our approach($N = 32$) with Neural Radiosity on Modern Hall, Wooden Staircase and Veach Room scenes. Our method produces high-quality rendering results. For the diagram, the blue boxes indicate comparisons under the same number of iterations (measured in steps), while the red boxes correspond to comparisons under equal-time conditions (measured minutes). Zooming in is recommended for better visualization. An accompanying video is provided for dynamic qualitative comparisons.

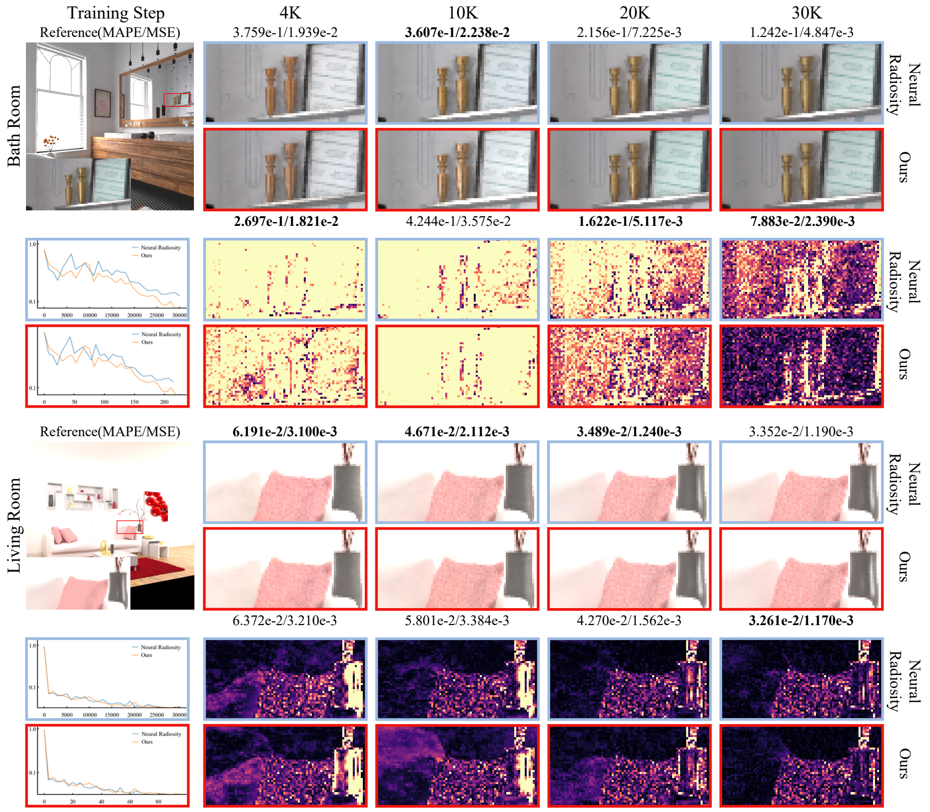

Fig. 6. Qualitative comparisons of our approach($N = 32$) with Neural Radiosity on Bath Room and Living Room scenes. Our method produces high-quality rendering results. For the diagram, the blue boxes indicate comparisons under the same number of iterations (measured in steps), while the red boxes correspond to comparisons under equal-time conditions (measured minutes). Zooming in is recommended for better visualization. An accompanying video is provided for dynamic qualitative comparisons.

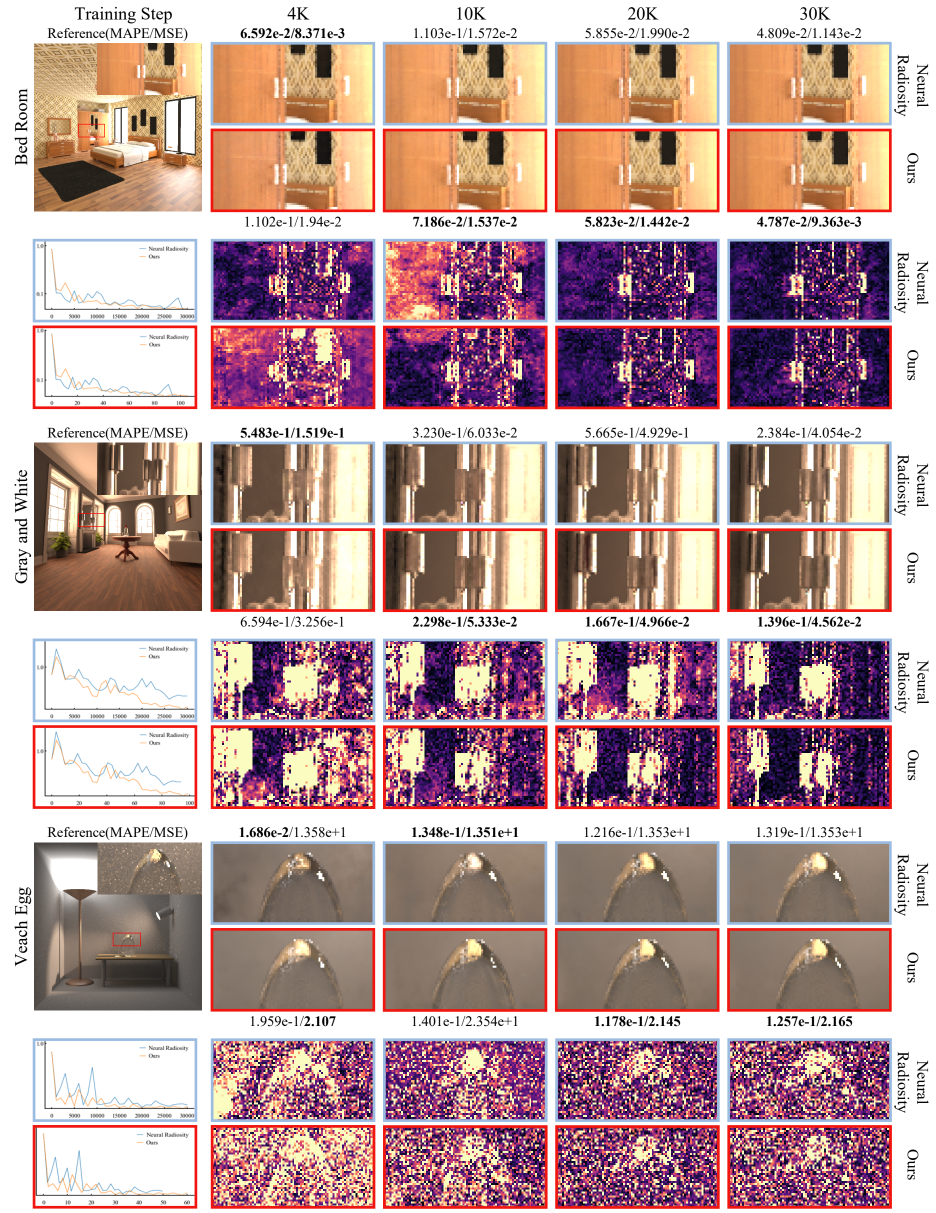

Fig. 7. Qualitative comparisons of our approach($N = 4$) with Neural Radiosity on Bed Room, Gray and White and Veach Egg scenes. Our method produces high-quality rendering results. For the diagram, the blue boxes indicate comparisons under the same number of iterations (measured in steps), while the red boxes correspond to comparisons under equal-time conditions (measured minutes). Zooming in is recommended for better visualization. An accompanying video is provided for dynamic qualitative comparisons.

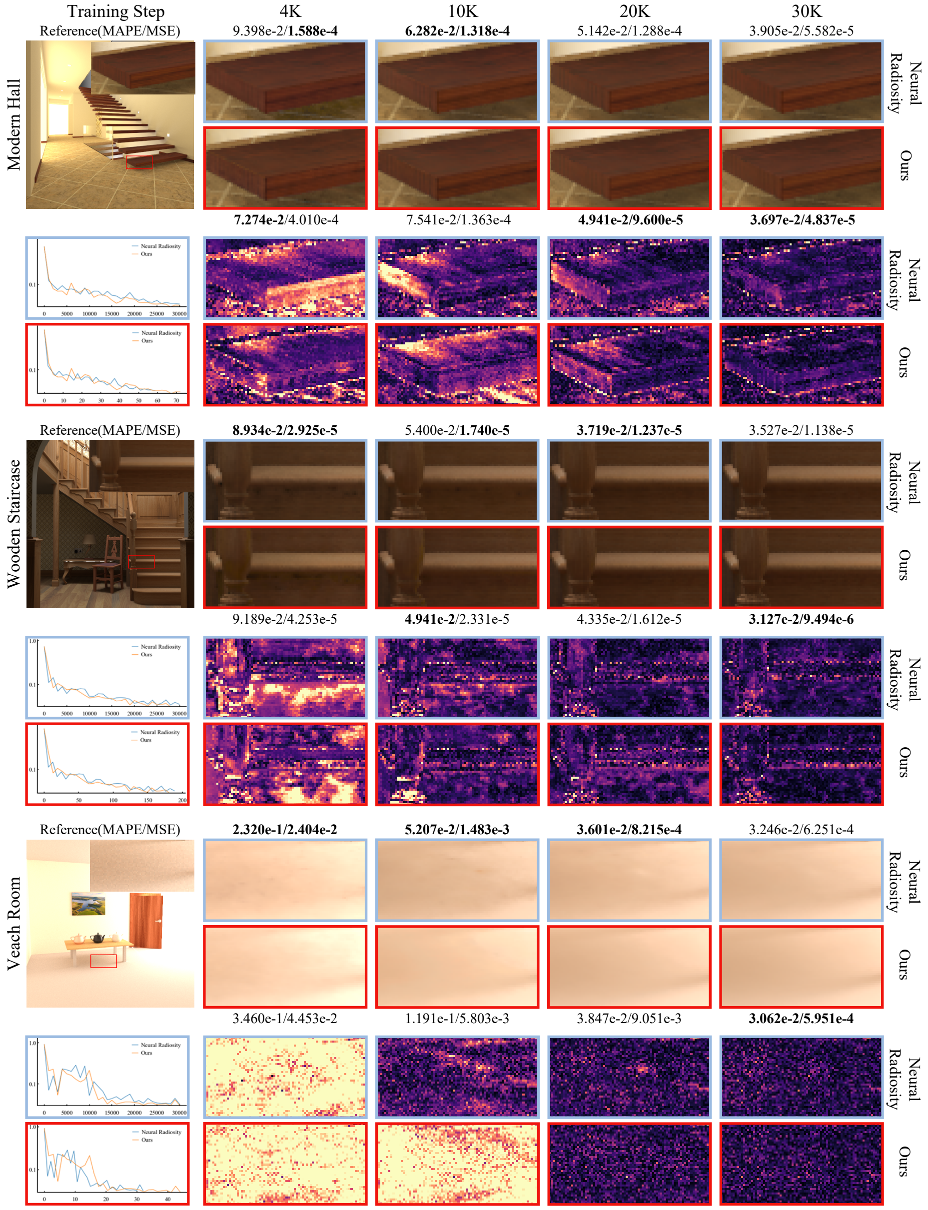

Fig. 8. Qualitative comparisons of our approach($N = 4$) with Neural Radiosity on Modern Hall, Wooden Staircase and Veach Room scenes. Our method produces high-quality rendering results. For the diagram, the blue boxes indicate comparisons under the same number of iterations (measured in steps), while the red boxes correspond to comparisons under equal-time conditions (measured minutes). Zooming in is recommended for better visualization. An accompanying video is provided for dynamic qualitative comparisons.

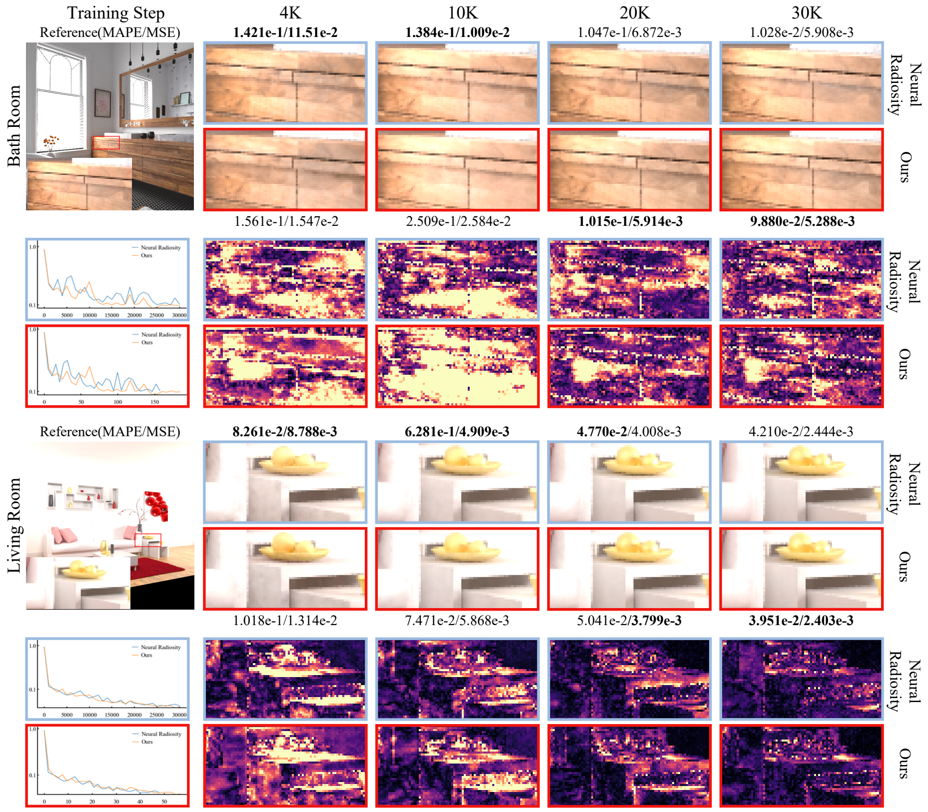

Fig. 9. Qualitative comparisons of our approach($N = 4$) with Neural Radiosity on Bath Room and Living Room scenes. Our method produces high-quality rendering results. For the diagram, the blue boxes indicate comparisons under the same number of iterations (measured in steps), while the red boxes correspond to comparisons under equal-time conditions (measured minutes). Zooming in is recommended for better visualization. An accompanying video is provided for dynamic qualitative comparisons.